8513162

THE UNITED STATES OF AMERICA

TO ALL TO WHOM THESE PRESENTS SHALL COME:

UNITED STATES DEPARTMENT OF COMMERCE

United States Patent and Trademark Office

August 05, 2024

THIS IS TO CERTIFY THAT ANNEXED IS A TRUE COPY FROM THE RECORDS OF THIS OFFICE OF THE FILE WRAPPER AND CONTENTS OF:

APPLICATION NUMBER: ***11/483,825***
FILING DATE: ***July 10, 2006***
PATENT NUMBER: ***7883848***
ISSUE DATE: ***February 08, 2011***

Certified by

Katherine Kelly Vidal

Performing the Functions and Duties of the
Under Secretary of Commerce
for Intellectual Property
and Director of the United States
Patent and Trademark Office

8513162

THE UNITED STATES OF AMERICA

TO ALL TO WHOM THESE PRESENTS SHALL COME:

UNITED STATES DEPARTMENT OF COMMERCE

United States Patent and Trademark Office

August 05, 2024

THIS IS TO CERTIFY THAT ANNEXED IS A TRUE COPY FROM THE RECORDS OF THIS OFFICE OF THE FILE WRAPPER AND CONTENTS OF:

APPLICATION NUMBER: ***11/483,825***
FILING DATE: ***July 10, 2006***
PATENT NUMBER: ***7883848***
ISSUE DATE: ***February 08, 2011***

Certified by

Katherine Kelly Vidal

Performing the Functions and Duties of the
Under Secretary of Commerce
for Intellectual Property
and Director of the United States
Patent and Trademark Office

17589 U.S. PTO
071006
112960 U.S. PTO
117483825
071006

Please type a plus sign (+) inside this box [+]

UTILITY PATENT APPLICATION TRANSMITTAL (Only for new nonprovisional applications under 37 CFR 1.53(b))	Attorney Docket No: 22639-3	Total Pages: 89
	First Named Inventor: Olof Ericsson	
	Title: Regulation Analysis by Cis Reactivity, RACR	
	Express Mail Label No.	

APPLICATION ELEMENTS

See MPEP chapter 600 concerning utility patent application contents

ADDRESSED TO: Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

1. ☒ Utility Patent Application Transmittal [Total Pages: 2]
2. ☐ Applicant(s) claim(s) small entity status, 37 CFR 1.27
3. ☒ Specification, Claims and Abstract [Total Pages: 71
4. ☐ Drawing(s) (35 USC 113) [Total Sheets: 9
5. Oath or Declaration [Total Pages: 1]
   a. ☒ Newly executed (original or copy)
   b. ☐ Copy from a prior application (37 CFR 1.63(d)) (for continuation/divisional with Box 21 completed) **[Note Box 6 below]**
      i. ☐ DELETION OF INVENTORS Signed statement attached deleting inventor(s) named in the prior application, see 37 CFR 1.63(d)(2) and 1.33(b).
6. ☐ Incorporation By Reference (useable if Box 5b is checked) The entire disclosure of the prior application, from which a copy of the oath or declaration is supplied under Box 5b, is considered as being part of the disclosure of the accompanying application and is hereby incorporated by reference therein.
7. ☐ Microfiche Computer Program (Appendix)
8. ☒ Nucleotide and/or Amino Acid Sequence Submission (if applicable, all necessary)
   a. ☒ Computer readable copy
   b. ☒ Paper copy (identical to computer copy)
   c. ☒ Statement verifying identity of above copies

Accompanying Application Parts

9. ☒ Assignment Papers (cover sheet & document(s))
10. ☐ 37 CFR 3.73(b) Statement ☒ Power of attorney (when there is an assignee)
11. ☐ English Translation Document (if applicable)
12. ☐ Information Disclosure Statement (IDS/PTO-1449)
    ☐ Copies of Citations
13. ☐ Preliminary Amendment
14. ☒ Return Receipt Postcard (MPEP 503)
15. ☐ Certified Copy of Priority Document(s)
16. ☐ Request and Certification Under 35 U.S.C. §122(b)(2)(B)(i)
17. ☐ Other: _____

17. FEE CALCULATION

CLAIMS	(1) FOR	(2) NUMBER FILED	(3) NUMBER EXTRA	(4) RATE	(5) CALCULATIONS
	TOTAL CLAIMS (37 CFR 1.16(c)	43-20 =	23	x $50 =	$1150.00
	INDEPENDENT CLAIMS (37 CFR 1.16(b)	3- 3 =	0	x $200 =	$00.00
	MULTIPLE DEPENDENT CLAIMS (if applicable) (37 CFR 1.16(d))			+ $_____ =	$-0-
				BASIC FILING FEE (37 CFR 1.16(a))	$300.00
				Size Fee - Each 50 Pages ≥ 100 (37 CFR 1.116(s))	$
				Search Fee (37 CFR 1.116(k))	$500.00
				Examination Fee (37 CFR 1.116(o))	$200.00
				Total of Above Calculations =	$
				Reduction by 50% for filing by small entity (37 CFR 1.27) =	$
				TOTAL =	$2150.00

Alamar Exhibit 1002
Alamar v. Olink

18. ☐ Please charge Deposit Account No. 04-1133 in the amount of $_____.
19. a. ☐ A check in the amount of $_____ is enclosed.
b. ☒ Please charge the amount of $2150.00 to our Visa credit card account. Form PTO-2038 is attached.
20. ☒ The Commissioner is hereby authorized to credit overpayments or charge the following fees to Deposit Account No. 04-1133:
a. ☒ Fees required under 37 CFR 1.16
b. ☒ Fees required under 37 CFR 1.17

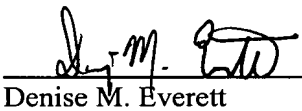
21. **If a CONTINUING APPLICATION**, check appropriate box and supply the requisite information:
☐ Continuation ☐ Divisional ☐ Continuation-in-part (CIP) of prior application No. _____/_____, filed _____.

22. CORRESPONDENCE ADDRESS

☒ Customer Number or Bar Code Label 24256 or ☐ Correspondence Address Below

Name	Denise M. Everett				
	Dinsmore & Shohl LLP				
Address	1900 Chemed Center				
	255 East Fifth Street				
City	Cincinnati	State	OH	Zip Code	45202
Country	USA	Telephone	513-977-8787	Fax	513-977-8141

Respectfully submitted,



Denise M. Everett
Registration No. 47,552
Attorney for Applicant(s)
Dinsmore & Shohl LLP
1900 Chemed Center
255 E. Fifth Street
Cincinnati, OH 45202
(513) 977-8787

CERTIFICATE OF EXPRESS MAILING
"Express Mail" mailing label #: EV 832369757 US
Date of Deposit: July 10, 2006
I hereby certify that this paper or fee is being deposited with the United States Postal Service "Express Mail Post Office to Addressee" service under 37 CFR 1.10 on the date indicated above and is addressed to: Commissioner for Patents; P.O. Box 1450; Alexandria, VA 22313-1450.

Marlene Troxle

Alamar Exhibit 1002
Alamar v. Olink

Regulation Analysis by Cis Reactivity, RACR

Olof Ericsson

RELATED APPLICATION

This application claims priority to U.S. Provisional application Serial No. 60/697,415, filed on July 8, 2005.

FIELD OF THE INVENTION

The invention relates to methods for detecting affinity-based interactions and functional-based interactions, and the relative strengths of those interactions, between molecules of interest. Libraries of interactors are constructed wherein the interactors comprise at least one molecule of interest coupled to at least one unique nucleic acid moiety. The inventive methods allow detection, interactor identification, and determination of relative strengths of all interactions possible within a library of interactors. Kits enabling rapid and efficient application of the inventive methods are also provided.

BACKGROUND OF THE INVENTION

To identify molecular networks, information concerning at least two molecular species must be gathered since there are at least two molecules involved in any interaction. Current high throughput techniques generally resolve one of the interaction partners in a high throughput fashion while the other interaction partner is limited to much fewer species. There is a need for methods which permit high throughput screening and identification of all partners of an interaction.

Alamar Exhibit 1002
Alamar v. Olink

The human genome contains approximately 30,000 genes, disregarding splice variants and post translational modifications this corresponds approximately to 30,000 proteins and $4.5*10^8$ possible protein pairs. In general, if there are n molecules of interest in the library, then there exists $(n^2+n)/2$ possible interaction pairs in the network. The capability of investigating vast libraries and networks like these would be greatly enhanced by the development of molecular interaction-based methods wherein high throughput information may be generated for both partners of the interaction.

To analyze complex libraries a readout platform which enables data acquisition in high throughput fashion is required. Protein analysis via micro-array techniques can be achieved by converting information about protein interactions into nucleic acid-based information, which is highly amenable to analysis with microarrays. Microarray analysis is a conventional platform which enables cost effective data generation in a high throughput manner. A method of protein interaction analysis which is successfully adapted for microarray analysis is highly desirable. The present inventors have developed such an adaptation wherein molecular interactions are detected by detecting a combination of at least two nucleic acid tags, one from each molecule, wherein at least one such combination is required for signal output in one microarray feature.

Several other approaches exist in the art which can be utilized for interaction screening, but which differ from the present inventive methods in significant ways.

Protein microarrays use micro-scale, spatial localization for identification of different molecules. The format allows detection of molecules in high throughput if target molecules in solution are investigated by probes organized on a solid phase. Targets can be labelled with a detectable function and the presence of the detectable function in one array feature reflects the presence of the target in the sample, assuming that no cross reactivity between probes are

Alamar Exhibit 1002
Alamar v. Olink

present. Examples of this approach are expression arrays and antibody arrays. The assumption when using these arrays is that each target in solution only displays affinity to one defined probe on the microarray. When investigating molecular interactions the affinities are not known per definition. To identify molecular interactions with regular microarray analysis one detectable function is required for each target in solution.

The number of detectable functions which can be resolved then becomes limiting. Generally only one or two detectable functions are used although there are approaches which potentially could resolve more they are far away from the number of molecules which can be identified by position in the array. This limits the capacity of information extraction per experiment to the number of detectable functions multiplied by the library size, e.g. if a library has 500 members and the approach can detect two detectable functions 250 arrays need to be performed to extract all possible interaction information. Further on, the microarray platform suffers from the disadvantage that one partner of the protein pair needs to be immobilized on a solid phase which potentially can disturb the interaction and/or the protein conformation. While protein microarrays permit analysis of one member of an interaction in high throughput and can provide information about interactions, the data extraction is cost and labour effective. The analysis is performed on a solid phase and interaction studies are limited by the number of resolvable detectable functions which also limits the possibility of inter- or intra- library interaction screens.

Another known method is the yeast two hybrid system which utilizes transcription factors to investigate protein interactions. Many transcription factors contain two distinct functional regions, one which binds a specific DNA sequence and one which recruits the transcription

Alamar Exhibit 1002
Alamar v. Olink

machinery to activate a proximal gene. Yeast two hybrid systems utilize these features to investigate protein affinity interactions. One protein or protein library is fused with the DNA binding domain forming the "bait" and a second protein or protein library is joined to the activating domain forming the "prey". The fusions are constructed on the DNA level and the two constructs are then co-expressed in yeast. If two proteins which have affinity for each other are expressed in the same cell they will activate transcription. Affinity interaction can then be detected by reporter gene activation due to reconstitution of the transcription factor. The two major approaches for yeast two hybrid interaction analysis is the "array approach" and the "exhaustive screening" approach. In the array approach yeast clones expressing different baits are ordered in an array so that the identity of the protein in each clone can be deducted from position. The array is then mated with yeast expressing one prey protein of interest. The mated clones are then cultured on selective media and protein interactions can be identified without cloning and sequencing from the position of positive clones on the array. The exhaustive screening approach mate one bait clone with a library of prey proteins, the mated library is then cultured on selective media and positive clones are sequenced to identify which prey proteins have interacted with the bait. The major drawback of the exhaustive screening approach is the time and cost of sequencing all the positive clones, which combined with the high intrinsic level of false positives increase the time and cost dramatically, compared to the array approach. Hence, library-library screens are generally not considered.

While the yeast two hybrid systems are advantageous in that they are genetic systems and thus do not require protein synthesis, and they can perform high throughput of one of the interaction partners, sequencing of positive clones is necessary for high throughput library

Alamar Exhibit 1002
Alamar v. Olink

screening approaches, which, combined with an intrinsic high frequency of false positives becomes problematic.

Display techniques such as phage display, ribosome display, RNA display, SELEX, cis display and covalent display are used to isolate affinity binders from libraries. These techniques are based on a large library of potential interaction partners (binders) which are allowed to interact with one partner (target). The target is typically immobilized on a solid phase and then allowed to interact with the library of binders. Members of the library which do not bind to the target are discarded and bound members are regenerated. Then the procedure is iterated until only one or a few library members remain. These are then analysed for affinity towards the target. The different display techniques primarily differ in what molecule type is used to create the library and how the binders are regenerated. Display techniques can be used to identify affinity interactions but these are generally limited to a low number of target molecules. Some display techniques like phage display have the potential to identify interactions between libraries but the analysis of the interactions are done by cloning and sequencing which is very time consuming and labour intensive. A majority of the display techniques like e.g phage display are genetic systems and do not easily adapt to other molecules than proteins.

To enable selection of binders based on libraries other than peptide, protein or nucleic acid, libraries, DNA templated synthesis (DTS) approaches has been developed. In this technique, nucleic acids can be attached to low molecular compounds in order to establish identification of the compound after selection of a library for a desired property. If two compounds are attached to nucleic acids which are complementary, hybridization of these nucleic acids will force the two compounds in proximity and elevate the relative local concentration of the compounds. Thus, the two compounds will be more prone to react than if

Alamar Exhibit 1002
Alamar v. Olink

they were not associated. This approach has also been utilized for multi step reactions, where libraries of compounds joined to nucleic acid tags are directed to serially react with a compound attached to a template nucleic acid. The synthesized library can then be screened with respect to a desired property and selected compounds can then be identified. Further on, the identity of compounds synthesized by DTS can be identified by microarray hybridization of the nucleic acid template which directed the synthesis and to which the product is also attached.

In DTS, two molecules are forced together by using nucleic acids to allow the molecules to interact to a higher degree than if they were not brought in proximity. For example, in the DTS-based method taught by Kanan et. al (Kanan et. al Nature 2004), one of the molecules forced together is attached to a "template" nucleic acid which specifically directs the second compound by direct hybridization of the nucleic acid attached to the second molecule. Thus the "template" nucleic acid contains both tags used to identify the compounds which are forced together. To create interactions between all members of a library with *n* members with this approach, the number of "template" oligonucleotides which have to be synthesized are in the range of $n^2/2$, since one specific oligonucleotide has to be synthesized for each individual interaction on the library.

However, the present invention provides a method whereby the nucleic acid encoding the tags to identify both members of the interaction are formed upon joining of the two molecules. Hence, to investigate all interactions in a library of *n* members the number of oligonucleotides required is in the order of *n*. The combination of molecules in DTS is not combinatorial in the sense that all library members potentially may interact. It becomes combinatorial due to the presence of all specific pairs DNA templates which in turn direct the

Alamar Exhibit 1002
Alamar v. Olink

interactions. If the specificity of these interactions fails, the link between the compound phenotype and the nucleic acid genotype also fails. Thus, in the DTS approach, the link between nucleic acid identification tag and the molecules in the interaction depend on correct hybridization of one tag molecule to the template nucleic acid. If any cross hybridization occurs then the link between the tag and the identity of the interaction is unreliable.

In the current invention the nucleic acid formed by the NAM association encodes the combination of identification sequences. Therefore the link between the identification sequences in the associated nucleic acid molecule and the phenotype of the molecule is more effectively maintained.

The utilization of combinatorial association of different nucleic acid tags instead of pre-synthesis of all different combinations also provides further advantages. The amplification of the nucleic acid can be performed so that only nucleic acids which have been joined to create a pair of molecules are amplified by, e.g., introducing one primer site on each arm. Nucleic acids which remain unassociated will thus not be amplified. In the DTS approach all template nucleic acids can serve as PCR templates even though they have no interaction partner.

A method referred to as "proximity ligation" has been recently described (International Patent Application Serial No. WO0161037, U.S. Patent Application Serial No. US2002051986). According to this method, targets are detected by utilizing two or more binders, e.g. antibodies, with known affinity to the target. The method is based upon the co-localization of the binder pair in the presence of a target. This target brings the binders into proximity, enabling ligation of nucleic acids located on the binder pair. Thus the joining of the nucleic acid becomes elevated in

Alamar Exhibit 1002
Alamar v. Olink

the presence of the target molecule. The nucleic acid can subsequently be quantified and the amount of nucleic acid corresponds to the amount of target.

The primary embodiment of WO0161037 and US2002051986 aims at detection of a defined target. Thus all the affinities in the system are known and there is always at least one molecule, the target, which does not have a nucleic acid attached to it. Moreover, the invention utilizes two or more binders which bind their unlabelled analyte pair-wise in a predefined manner. Several pairs might be used in the same reaction but they are always analyzed in a target specific pair wise fashion, not in a combinatorial fashion.

The current inventive method differs from proximity ligation in that it does not detect or quantify a predefined target in a sample. The present inventive methods do not utilize molecules with predefined affinities but rather the reverse; the inter-molecular affinities are retrieved as a result of the inventive method.

In the current invention the association of the nucleic acid is combinatorial and the novel nucleic acid produced by this association is then identified to yield information concerning the molecular interactions in the library. This enables interrogation of all intra- or inter-library member interactions. The proximity ligation methods always include at least one molecule which is not attached to a nucleic acid label and this "target" is the molecule which is detected. Thus the co-localization/proximity of the binders and thereby the nucleic acid in the assay arises from the presence of a target molecule.

Finally the information gained from the inventions differs significantly. In the present inventive methods, combinations of interacting molecules may be detected and quantified. The proximity ligation art, on the other hand, teaches high throughput screening of inhibitors for a defined binding event. For example, a protein affinity interaction is identified first by some

Alamar Exhibit 1002
Alamar v. Olink

other method and proximity ligation is employed to find an inhibitor for the known affinity interaction. The proximity ligation screening approach involves attaching nucleic acids to the two proteins which participate in the predefined interaction. Then different potential inhibitors are added to the reaction. These potential inhibitors are not labelled with nucleic acids and their action is monitored by observation of a reduced signal from the two labelled proteins. The pre-defined affinity reagents are labelled with nucleic acids and a pre-defined pair of nucleic acids is analyzed.

There is a need in the art for methods capable of investigating affinity and/or functional interactions within libraries of molecules wherein such interactions are not previously established to exist, and methods for quantification of any detected interactions.

SUMMARY OF THE INVENTION

Accordingly, the present invention is directed to methods for *discovering,* as well as detecting molecular interaction networks such as affinity interactions or functional interactions, e.g., enzymes-substrate pairs. Significantly, the invention allows investigation of all possible interactions within or between libraries of molecules, such as proteins, and also permits relative quantification of the detected interactions. Information about the interactions is derived through the use of nucleic acid tags which facilitates high throughput analysis, preferably by microarray analysis, however other analysis platforms may also be used. Typically, each feature on the microarray corresponds to a molecular interaction between two specific molecules in the library.

Each molecule of interest (MOI) in the library is tagged with a nucleic acid tag, referred to herein as a nucleic acid moiety (NAM). If two molecules interact the combination of their

Alamar Exhibit 1002
Alamar v. Olink

nucleic acid tags is then selectively enhanced. The oligonucleotide formed from the combination of two NAMs is referred to herein as an associated oligonucleotide.

One embodiment is directed to a method of detecting an affinity interaction between at least two molecules of interest. The method comprises: (a) forming a plurality of interactors by coupling each molecule of interest with at least one nucleic acid moiety comprising an identification sequence element and at an association element; (b) promoting an association between at least two nucleic acid moieties to form a plurality of unique associated oligonucleotides, wherein each nucleic acid moiety may form more than one unique associated oligonucleotide, and wherein each unique associated oligonucleotide comprises at least two identification sequence elements derived from the at least two nucleic acid moieties; (c) selecting the plurality of associated oligonucleotides; and (d) subjecting the selected associated oligonucleotides to an analysis that permits detection of the at least two identification sequence elements.

According to this embodiment, an "interactor" library is developed. Each interactor consists of the molecule of interest (MOI) tagged with at least one unique nucleic acid moiety (NAM). The NAM comprises, *inter alia*, an identification sequence making downstream detection possible. It is the associated oligonucleotides which are subjected to qualitative and quantitative analysis, as a single associated oligonucleotide corresponds to an interaction between the two MOIs tagged with the NAMs that comprise that associated oligonucleotide. The library of interactors (NAM-tagged MOIs) is first allowed to interact and equilibrate. As the MOIs associate based on any affinity(s) between them, the corresponding NAMs associate in a proximity dependent manner incidental thereto.

Alamar Exhibit 1002
Alamar v. Olink

Another embodiment of the present invention is directed to a method of detecting functional interactions between at least two molecules of interest. The method comprises: (a) forming a plurality of interactors by coupling each molecule of interest with a nucleic acid moiety, the nucleic acid moiety comprising an identification sequence element and an association element, wherein an affinity exists between the nucleic acid moieties; (b) forming a plurality of cis-reactive cells wherein a cis-reactive cell comprises at least two interactors bound in proximity to one another by an associated oligonucleotide formed from the affinity between at least two nucleic acid moieties, wherein the associated oligonucleotide comprises at least two identification elements derived from the at least two nucleic acid moieties; (c) subjecting the plurality of cis-reactive cells to conditions which stimulate a desired functional interaction having a detectable trace; (d) selecting all cis-reactive cells exhibiting the detectable trace; and (e) subjecting the associated oligonucleotides from the cis-reactive cells selected in (d) to an analysis that permits detection of the at least two identification sequence elements.

As in the affinity embodiment, each MOI is tagged with a unique NAM to form an interactor. However, in some functionality-based embodiments, the NAMs are designed to so that a pre-existing affinity exists between them. MOIs are independently tagged with the NAMs. Again, the interactor library is permitted to interact and equilibrate, creating a secondary library of "cis reactive cells" (CRCs) wherein each CRC consists of MOI molecules brought into proximity by the affinity between their respective NAMs, and co-localized by the association of their respective NAMs. Subsequently the function to be investigated is activated. When activated, MOIs will primarily act upon the other members in the CRC due to the proximity created by the physical linkage. In some embodiments it may be desirable to dilute the

Alamar Exhibit 1002
Alamar v. Olink

concentration of CRCs in order to lessen the probability of trans activation of the function. That is, activation between, rather than within, CRCs.

Following functional activation the secondary library is filtered for CRCs which contain MOIs that have been altered by the activation. Selection of positive (that is, altered) CRCs can be done by, e.g., affinity purification of the library of CRCs with respect to a modification introduced by an enzymatic reaction. The CRCs are then identified by the sequence motifs in the NAMs, thus both members of a functional interaction can be identified, for example, pairs of novel enzymes and novel substrates.

BRIEF DESCRIPTION OF THE FIGURES

Figure 1

Schematic description of one embodiment for affinity interaction detection. A) Library morphology; a library with 1, 2, ..., *n* MOIs, each associated with a defined NAM with nucleic acid tags 1, 2, ..., *n*. B). The library is allowed to interact by affinity to induce co-localization of the MOIs and the respective NAMs. C) The association function of the NAMs is activated and NAMs brought into proximity by affinity between their respective MOIs are associated to for a novel associated oligonucleotide comprising nucleic acid from both of the originating NAMs. D) Selective amplification of the different NAM combinations. NAMs which are not associated will not yield any amplification product for subsequent detection.

Figure 2

Schematic description of one embodiment for functional interaction detection. A) Library morphology; a library with 1, 2, ..., *n* species of MOI, each MOI species is associated with a defined NAM with nucleic acid tags 1, 2, ..., *n*. B) Construction of CRCs. The library

Alamar Exhibit 1002
Alamar v. Olink

association function is designed to co-localize the library members in a combinatorial fashion to create all possible CRC variants between the n library members ($(n\text{-}n^2)/2$ unique combinations). C) The functional interaction to be detected is activated and reaction conditions are controlled to restrict the functional interactions to substantially within the CRCs. D) Selection. Selection is carried out with respect to at least one detectable alteration introduced by the functional interaction. E) Amplification of the selected CRCs and subsequent detection and analysis of the nucleic acids.

Figure 3

Schematic description of the library design in the exemplary approach described for affinity interaction. A) The NAMs are produced from a seed PCR product and unique NAMs are synthesized with tailed PCR primers comprising one of a plurality of nucleic acid tags. An association function is located in one primer and a conjugation function is located in the other. B) library morphology after conjugation to MOI. The association element in this case comprises a protruding end which allows selective blunt end ligation.

Figure 4

Schematic description of the library design in the exemplary approach described for detecting a functional interaction. A) NAMs before conjugation to the MOI are depicted, including two separate sub-populations of NAMs comprising potential kinases and substrates respectively, one with the forward primer and one with the reverse primer. Each NAM comprises a PCR primer, identification element and an association element. B) Library morphology after association of the interactors to construct CRCs is illustrated. Each CRC is joined by addition of a polymerase that stabilizes the hybridization of the respective 3′ ends and incorporates the nucleic acid tags and PCR primers into one associated oligonucleotide by polymerization.

Alamar Exhibit 1002
Alamar v. Olink

Figure 5

Non-limiting examples of association of NAMs. Arrows indicate nucleic acid 3′ ends. Figures 5(a - d) illustrate ligation approaches. The structure in 5(e) can be associated with either ligation or polymerization with a polymerase comprising 5′ exonuclease activity. 5(f) can be associated with polymerization. 5(g) illustrates association by introduction of a functional group which reacts to create a native nucleic acid. In 5(h) reactive functions are used which associate the nucleic acids by a link separate from nucleic acid.

Figure 6

Examples of two tag microarray detection. Black sequences represent a general sequence common to all targets and red sequences represent identification tag sequences. 6(a) illustrates hybridization of a target where the two tag elements are separated by a general sequence that is not complementary to the microarray probe. 6(b) illustrates hybridization of a target comprising two identification elements not separated by any spacing or general sequence. In 6(c) the identification tags are located in the 3′and 5′ ends with the hybridized target forming a circular structure that could be ligation-templated by the microarray probe. In 6(d) the target is hybridized to form a circular structure similar to 6(c) but with a general spacing sequence at the ends of the target and in the microarray probe, an detection can be mediated by sandwich hybridization of a third nucleic acid or by direct labelling of the target. Figure 6(e) represents detection of a target with two joined tags by ligation of a third nucleic acid comprising a detectable function. In 6(f) the detection is mediated by polymerization, for example, by a exonuclease deficient polymerase, rather than ligation. 6(g) represents detection of the structure formed in 6(c) by ligation of the circular structure, polymerization and subsequent detection by

Alamar Exhibit 1002
Alamar v. Olink

addition of a third labelled nucleic acid. The polymerization event could also be detected by labelled nucleotides.

Figure 7

This represents results from the example for reaction discovery. Three tag microarrays from three experiments with different reaction conditions are displayed. A higher intensity represents detection of higher concentrations of the tag combination. Each hetero-interaction is represented by two separate microarray features.

Figure 8

Illustration of a single stranded associated oligonucleotide or single stranded amplification product thereof. Two separate single stranded oligonucleotides comprising a type IIS restriction enzyme recognition sites are hybridized to the general sequence flanking the identification elements. The restriction enzyme cleavage occur between the identification element and the general sequence. Figure 8(b) Illustrates the single stranded DNA produced by the restriction enzyme cleavage with identification elements in the 3′ and 5′ end.

Figure 9

Diagrammatic illustration of microarray features representing 3' arm and 5' arm tag combinations resulting from an interaction context. Data is plotted on a grid including the different 3′ and 5′ NAMs with the signal to noise (feature signal/ sum of all feature signals) plotted on the z-axis. The arms with thiol groups are 3′A and 3′B and 5′C and 5′D.

DETAILED DESCRIPTION OF THE INVENTION

Terms & definitions

Alamar Exhibit 1002
Alamar v. Olink

As used herein, the term "combinatorial association" refers to association of a plurality of molecules to create a plurality of combinations of those molecules.

As used herein, the term nucleic acid moiety (NAM) refers to a nucleic acid molecule which is coupled to at least one molecule of interest (MOI). A NAM comprises at least one unique identification nucleic acid sequence so that detection of that sequence infers involvement of the coupled MOI. A NAM also comprises at least one element which enables association with other NAMs. A NAM may be single stranded DNA, double stranded DNA, single stranded RNA, double stranded RNA or combinations thereof. The NAM may contain modified nucleotides or nucleotide analogues. The NAM may be derived from synthetic, biochemical, biological synthesis or combinations thereof.

As used herein, the term "molecule of interest" (MOI) refers to a molecule about which interactive information is desired. An MOI is coupled with at least one NAM. In one embodiment each MOI is coupled with a single unique NAM. Examples of suitable MOIs include, but are not limited to, whole proteins, protein domains, peptides, peptoids, functional enzymes, low molecular weight compounds, small molecules, polymers, non-proteinaceous cellular products, virus particles, cells metabolites, lipids, carbohydrates, nucleic acids, inorganic compounds or combinations thereof. In one specific embodiment the MOI comprises proteins. In any library of MOIs, not all the MOIs may solely comprise nucleic acids. In a more specific embodiment, the MOIs may comprise factors regulating transcription, non-limiting examples include enhancers, repressors, and isolators) and DNA sequences. A library of such MOIs may be investigated for affinity interactions. In very specific embodiments the DNA sequences are derived from genomic sequences or comprise random elements. The DNA sequence MOIs may comprise genes involved in transcriptional regulation and the present inventive methods may be

Alamar Exhibit 1002
Alamar v. Olink

employed to reveal interactions between DNA and proteins to enable, for example, deduction of how the transcriptional units regulate transcription by binding to DNA.

As used herein, the term "small molecule" refers to a non-peptidic, non-oligomeric organic compound either synthesized in the laboratory or found in nature. Small molecules, as used herein, can refer to compounds that are similar to complex natural products which nature has selected through evolution, however, the term "small molecule" is not limited to these compounds. Rather, a small molecule is typically characterized in that it contains several carbon-carbon bonds, and has a molecular weight of less than 1500, although this characterization is not intended to be limiting for the purposes of the present invention. Examples of "small molecules" that occur in nature include, but are not limited to, taxol, dynemicin, and rapamycin. Examples of "small molecules" that are synthesized in the laboratory include, but are not limited to, compounds described in Tan et al., ("Stereoselective Synthesis of over Two Million Compounds Having Structural Features Both Reminiscent of Natural Products and Compatible with Miniaturized Cell-Based Assays" J. Am. Chem. Soc. 1998, 120, 8565) and United States Patent Application 20030082830 "Synthesis of Combinatorial Libraries of Compounds Reminiscent of Natural Products", the entire contents of which are incorporated herein by reference.

As used herein, the term "interactor" refers to at least one MOI coupled with at least one NAM.

As used herein, the term "Cis Reactive Cell" (CRC) refers to an association of at least two interactor moieties, joined by an associated oligonucleotide. In a specific embodiment a CRC comprises two interactors.

As used herein, the term "interactor library" or "library of interactors" refers to a plurality of interactors.

Alamar Exhibit 1002
Alamar v. Olink

As used herein, the term "Cis reactive cell library" or "library of cis reactive cells," or "secondary library" refers to a plurality of CRCs.

As used herein, the term "association of NAMs" refers to an association between the NAM components of interactors which results in a molecule containing nucleic acid sequences originating from the associated NAM molecules. This combination of NAMs is referred to herein as an "associated oligonucleotide" and, in specific embodiments, comprises the identification sequences for two component NAMs. The association may be covalent or non-covalent and occurs between at least two NAMs. In specific embodiments, the association occurs pairwise between two NAMs.

The types of interactions which may be detected with the current invention include, but are not limited to, interactions based on affinity, enzymatic activity including kinase, glycosylation, phosphatase and ubiqutinylation activity, fusion, covalent bond formation, structural alteration, degradation, and/or subunit addition. A person of ordinary skill in the art will recognize that an "affinity-based" interaction may include any interaction which results in non-random assortment of molecules due to any basis other than repulsion.

The current invention aims to retrieve information about molecular interactions. Interactions can be of any type between molecules such as e.g. affinity, enzymatic, structural, degrading, fusion, addition of one (or more) groups. The foundation of the interaction analysis is the increased probability of a reaction between molecules if they are brought in close proximity by molecular co-localization. The co-localized molecules will be prone to interact in *cis* rather then interact with separate molecules in *trans* with the presumption that the relative concentration of the co-localized molecules is higher then the concentration of other molecules in the solution. In the case of detection of affinity interactions the co-localization is introduced

Alamar Exhibit 1002
Alamar v. Olink

by the affinity interaction between the MOIs which then enhances the probability of the respective co-localized nucleic acids to associate. Functional interactions are detected by bringing the MOIs into proximity with one another through an independent assorting means. For example, MOIs may be brought into proximity with one another incidental to affinities between the NAMs, and then co-localized by association of the NAMs to form interactors. This increases the probability that all possible MOI groupings occur and are exposed to the functional trigger. A detectable function is then triggered.

The feature of intra-molecular enhancement of reaction kinetics frequently appears naturally in biological systems like, for example, transcription activation where affinities co-localize the transcription machinery to a gene initialized by binding of a transcription activator. The present inventive methods adapt such an approach in vitro as a molecular technique for analysis of molecular interactions.

According to the present inventive methods, interaction information is essentially encoded into nucleic acids which then are detected by nucleic acid analysis techniques. This is achieved by tagging each MOI with a NAM comprising a unique nucleic acid identification sequence. (In some embodiments more than one MOI may be tagged with the same NAM species). The downstream association of NAMs form a novel oligonucleotide molecule encoding the interaction. This approach allows combinatorial analysis of all potential interactions in a library if all different tag combinations can occur.

The following detailed description of the Interactor Library is divided into the following sections for clarity: components; construction; interaction analysis; filtering; and readout.

Interactor Library

Library components

Alamar Exhibit 1002
Alamar v. Olink

The current invention is based upon the construction of a library of molecules referred to herein as interactors. These library members comprise at least two distinct parts which are coupled, one part is the molecule of interest (MOI) and one part is the nucleic acid moiety (NAM).

Molecule of interest (MOI)

The MOI component comprises molecules including, but not limited to, whole proteins, protein domains, functional enzymes, low molecular weight compounds, polymers, non proteinacious cellular products, virus particles, cells metabolites, lipids, carbohydrates, nucleic acids or combinations thereof. The library may consist of a heterogeneous population of MOI types. Minimally, the MOI component must meet two requirements, it must be coupled with at least one uniquely identifiable NAM, and not all MOIs in the library can be nucleic acids.

Nucleic acid moiety (NAM)

The NAM may comprise single stranded DNA (ssDNA), double stranded DNA (dsDNA) single stranded RNA, double stranded RNA or combinations thereof. It may include any type of modified nucleotides or nucleotide analogues or other types of molecules which can be incorporated into nucleic acids. The coupling between a NAM and an MOI to form one interactor may be non covalent but covalent association is desirable. Each NAM is coupled to at least one MOI and it may associate with other NAMs in the library.

Interactors

One interactor in the library comprises at least one MOI coupled to at least one identifiable NAM. The inventive methods enable retrieval of information concerning interactions between these interactors in vitro. The retrieved data can then be translated into information about the interactions of the MOIs.

Alamar Exhibit 1002
Alamar v. Olink

Library construction / coupling of MOI and NAM

The inventive methods prepare and utilize libraries of interactors as described above. These libraries can be synthesized in several different ways. The construction of the interactor library includes coupling of at least one MOI to at least one NAM. Generally the coupling means may be specific or unspecific with respect to coupling position and couplings may be performed in *trans* and *cis*. In other embodiments an interactor comprises more than one MOI and at least one NAM.

Coupling of MOIs and NAMs in trans

It is possible to construct an interactor library merely by mixing a plurality of MOIs with a plurality of NAMs and permitting them to incubate and couple under conditions that favor coupling. Possible mechanisms of coupling include, but are not limited to, chemical, enzymatic or affinity based mechanisms. Preferably the coupling between the MOI and the NAM is covalent and stable or at least of high affinity. The coupling of the MOI to the NAM may also include addition of one or several types of bi-functional linkers. The coupling may be specific or nonspecific.

Nonspecific coupling includes means wherein the exact position of the NAM might not be controlled or known and wherein several NAMs might couple to one MOI. However, unspecific coupling of the NAM to the MOI might interfere with or alter the interaction phenotype of the MOI. Therefore specific coupling may be desirable when possible. Specific coupling includes methods where the position of the NAM on the MOI is controlled and/or known. To control localization of the NAM a reactive functionality may be introduced on the MOI during or after synthesis of the molecule, whether synthesis is chemical, enzymatical, *in vitro* or *in vivo*. This will secure NAM coupling with the MOI at defined position(s) on the MOI.

Alamar Exhibit 1002
Alamar v. Olink

Depending on the nature of the MOI, specific coupling of the NAM may be performed downstream of MOI synthesis without modification of the MOI if there is a novel site(s) on the MOI to which the coupling can be directed.

Examples of specific coupling means

Generally, specific coupling may be accomplished by introduction or utilization of a site(s) on the MOI to which a specific coupling can be directed. The site can be used for direct coupling of the NAM or coupling to a third molecule which then can be coupled with the NAM. The nature of the association can be of covalent or non covalent as long as the link between the MOI and the NAM is maintained. Different interactors in the library may be produced with different approached for association of the respective MOI and NAM molecules depending on the nature of the interactor.

One example of a specific coupling means is the generation of a specific site by intein splicing. Several examples of this which may be utilized in the present inventive methods are disclosed in Takeda et al. *Bioorg Med Chem Lett* 14, 2407-2410 (2004), Southworth, et al. Biotechniques 27, 110-114, 116, 118-120 (1999), and Burbulis et al. "*Using Protein-DNA Chimeras to Detect and count small numbers of molecules*" 2, 31-37 (2005), all of which are incorporated fully herein by this reference. Generally, this approach creates a site for specific chemical conjugation in either the N-terminal or the C-terminal of a protein. An intein approach which could be used for both MOI purification and conjugation has been commercialized by New England Biolabs under the trademark IMPACT-TWIN. Specific coupling of the NAM to the MOI may also be performed by introduction of a fusion partner to the MOI which has affinity or enzymatic activity towards the NAM or vice versa. A specific example of this is the introduction of the hAGT protein as a fusion partner to the MOI and addition of the hAGT

Alamar Exhibit 1002
Alamar v. Olink

substrate to the NAM. hAGT then covalently binds the substrate on the NAM. This has been used to conjugate different molecules to proteins fused to hAGT as disclosed in Keppler et al. *Methods*, 32, 437-444 (2004), also incorporated fully herein by this reference.

A specific site may be introduced on the MOI to which a function for coupling then can be introduced downstream of synthesis. An example of this approach is the utilization of biotin ligase as taught in Duffy et al. Analytical Biochemistry 262, 122-128 (1998), the disclosure of which is fully incorporated herein by this reference. Briefly, a signal peptide is incorporated into the MOI during synthesis, which then is recognized by the biotin ligase. The biotin ligase recognizes the peptide and adds a biotin to the MOI. This biotin can then be used for coupling of the NAM by introducing an affinity molecule such as streptavidine or avidine on the NAM. The introduction of the biotin may be performed *in vitro* or *in vivo*.

Examples of unspecific coupling means

Unspecific coupling includes the use of chemical coupling to sites on the MOI which yield a coupling between the NAM and the MOI where the exact position of each NAM is uncertain and where the coupling may occur to several different sites on the MOI. These approaches would typically include one reagent (A) which can couple to the NAM and one reagent (B) which can couple to the MOI and a subsequent step where A and B are associated. There is also the possibility to associate A and B first or utilize a molecule which binds both the NAM and the MOI. A person of ordinary skill in the art will appreciate that the diversity of these molecules is quite high and the choice will depend upon the nature of the MOI. Specific nonlimiting examples of functional groups which can be used for this type of coupling include NHS esther, malemide, Azide, Alkyne, amine, or thiol. For more examples see; Double-Agents

Alamar Exhibit 1002
Alamar v. Olink

Cross-Linking Guide(#1600918) from Pierce. Coupling can also be mediated by a reactive group on the NAM which reacts directly with the MOI or vice versa.

Coupling between MOIs and NAMs in cis

Instead of individual construction of the library interactors, the coupling between the NAM and the MOI can be performed during the synthesis of the MOI. The synthesis is controlled so that it will produce coupling with at least one defined NAM. The interactor synthesis can be performed in multiplex, i.e. several interactors can be synthesized in the same reaction. This could potentially speed up the library construction significantly.

Nonlimiting examples of this include the approach used in CIS display or Covalent display as disclosed by Reiersen et al. Nucleic Acids Res 33, e10 (2005), or Odegrip, R. et al *PNAS* 101, 2806-2810 (2004), the disclosures of which are fully incorporated herein by this reference. A library of proteins is produced and during production the protein product is specifically associated to its own DNA by a protein fusion partner. Variants of this approach have been published whereby the coupling is covalent or non covalent. In a modified approach this could be used to create libraries where the protein serves as the MOI and the thereto coupled DNA molecule serves as a NAM.

Another approach is the use of the mRNA display technique disclosed in Wilson et al. PNAS 98, 3750-3755 (2001), fully incorporated herein by this reference. According to this approach a pool of proteins is produced from a pool of mRNA templates and the protein products are coupled with the mRNA they were produced from by a pyromycin modification at the end of the mRNA. The protein then serves as the MOI and the coupled mRNA as the NAM. A modified variant of this exists where the protein product is linked to a cDNA instead of the mRNA.

Alamar Exhibit 1002
Alamar v. Olink

Another approach which creates a physical link between the RNA and the protein is ribosome display, taught by Hanes & Pluckthun in PNAS 94, 4937-4942 (1997), the disclosure of which is fully incorporated herein by this reference. However in this particular coupling technique, coupling is mediated by the ribosome which is a large complex with the potential to interfere with the assay. The design of the coupled nucleic acid could be adapted to the current invention in these systems and the nucleic acid component may be modified post synthesis by, e.g., restriction endonucleases. In the genetic protein based systems the nucleic acid sequence encoding the proteins could be utilized as the nucleic acid identifying sequence for downstream detection. Constant sequences present on all nucleic acids in the library could be utilized for amplification.

A person of ordinary skill in the art will appreciate that any method which utilizes a function which couples the DNA or RNA with the protein product during synthesis could fulfil the purpose of creating libraries of interactors in multiplex employable herein. This has the additional advantage of ensuring integrity between the genotype and the resultant phenotype in display techniques. The important feature in these approaches is that the coupling between the nucleic acid and the MOI occurs in cis. Thus, the binding of the nucleic acid primarily occurs to the protein product synthesized from that nucleic acid. For coupling between DNA and protein this is mediated by the fact that the DNA is attached to the protein product by the transcription and translation machinery.

The desired MOI-NAM coupling may be achieved by a variety of means including fusion of MOIs to nucleic acid binding proteins which bind back to the nucleic acid upon synthesis and addition of molecules to the RNA or DNA which bind the synthesized protein or an attached fusion partner. The binding may be directly or indirectly mediated by a bifunctional linker, that

Alamar Exhibit 1002
Alamar v. Olink

is, an intermediate molecule holding the DNA and protein together. This intermediate molecule may range in size from a small molecule as defined herein, to a very large ribosome. The cis coupling of a nucleic acid to a desired protein could also be mediated by *in vivo* expression of, for example, plasmids, phages or viruses. If the nucleic acid encoding the protein product is utilized as the NAM, constant motifs can be used for amplification purposes and MOI encoding motifs for identification. One specific embodiment of the present invention is directed to methods of detecting affinity interactions between at least two molecules of interest, wherein at least one molecule of interest comprise a protein that is an expression product of a nucleic acid molecule, and wherein at least some of the plurality of interactors is formed by coupling the protein with a nucleic acid moiety comprising an identification sequence element and an association element derived from the nucleic acid molecule. This provides a library of interactors wherein each interactor comprises a molecule of interest comprising a protein that is an expression product of a nucleic acid molecule, coupled with a nucleic acid moiety tag, the tag comprising an association element and an identification sequence element, wherein the identification sequence element and the association element derived from the nucleic acid molecule.

DTS, described in Kanan et.al *Nature* 431, 545-9, herein incorporated fully by this reference, permits multiplex synthesis of MOIs other than proteins. DTS may be utilized to synthesize libraries encoded by nucleic acids and these library members can subsequently be utilized as interactors in the current invention.

Interaction analysis

The present invention provides methods which permit analysis of the interactor library with respect to interactions between the interactor members, and, therefore, by inference,

Alamar Exhibit 1002
Alamar v. Olink

between the MOIs. One embodiment provides analysis based primarily on affinity interactions between the MOIs. Another embodiment provides analysis based primarily on functionally-based interaction between MOIs. Affinity interactions are interactions where two or more members of the library display a binding affinity towards each other. Functional interactions are interactions where two or more members affect each other in some detectable way and may include affinity interactions, though a person of ordinary skill in the art will appreciate that there are functional interactions which display very low or transient affinity.

Affinity interactions

One embodiment of the present invention is directed to a method of detecting an affinity interaction between at least two molecules of interest. The method comprises: (a) forming a plurality of interactors by coupling each molecule of interest with at least one nucleic acid moiety comprising an identification sequence element and at an association element; (b) promoting an association between at least two nucleic acid moieties from different interactors to form a plurality of unique associated oligonucleotides, wherein each nucleic acid moiety may form more than one unique associated oligonucleotide, and wherein each unique associated oligonucleotide comprises at least two identification sequence elements derived from the at least two nucleic acid moieties; (c) selecting the plurality of associated oligonucleotides; and (d) subjecting the selected associated oligonucleotides to an analysis that permits detection of the at least two identification sequence elements.

Broadly, affinity interactions are defined as interactions between two or more molecules that induce a non-random distribution of these molecules in a solution comprising these molecules. Thus, the molecules which display the affinity will more frequently be close to each other then if their distribution were random. It is noted in particular that a non-random

Alamar Exhibit 1002
Alamar v. Olink

distribution per se may be due to a repulsion between the MOIs such that the MOIs are more frequently separated than in a random distribution.

Affinity interactions within the interactor library are analyzed by allowing the library members to mix and equilibrate under conditions which permit affinity interactions. Following this step association of NAMs is actively promoted. The association of NAMs is performed so that association between two NAMs depends on proximity of the NAMs, with the proximity being due to an affinity interaction between their respective MOIs. Thus MOI affinity interaction induced proximity or co-localization of NAMs induces a higher association frequency of those NAMs compared to a control where the MOI affinity is absent. Downstream of the association, the associated NAMs (the associated oligonucleotide) can optionally be amplified and subsequently detected, by, for example, microarray analysis.

A person of ordinary skill in the art will recognize that a stronger affinity between two MOIs generally induces more frequent association of the NAMs. This enables a relative quantification of the interaction affinity strength between interactors and, by inference, the MOIs.

Functional interactions

Another embodiment of the present invention is directed to a method of detecting functional interactions between at least two molecules of interest. The method comprises: (a) forming a plurality of interactors by coupling each molecule of interest with a nucleic acid moiety, the nucleic acid moiety comprising an identification sequence element and an association element, wherein an affinity exists between the nucleic acid moieties; (b) forming a plurality of cis-reactive cells wherein a cis-reactive cell comprises at least two interactors bound in proximity to one another by an associated oligonucleotide formed from the affinity between at

Alamar Exhibit 1002
Alamar v. Olink

least two nucleic acid moieties, wherein the associated oligonucleotide comprises at least two identification elements derived from the at least two nucleic acid moieties; (c) subjecting the plurality of cis-reactive cells to conditions which stimulate a desired functional interaction having a detectable trace; (d) selecting all cis-reactive cells exhibiting the detectable trace; and (e) subjecting the associated oligonucleotides from the cis-reactive cells selected in (d) to an analysis that permits detection of the at least two identification sequence elements.

Functional interactions are defined as interactions between molecules where the nature of at least one of the partners which participate in the interaction deviates from the nature of that partner when the other participant(s) of the functional interaction is/are present in conditions permitting the functional interaction. Functional interactions may include an affinity interaction.

Functional interactions include interactions with functions such as, e.g., kinase, phosphatase, glycosylation, deglycosylation, ubiqutinylation, deubiqutinylation or other paired activities where one member adds or removes a molecule(s) from the other member. Additional nonlimiting examples include interactions with altering activities such as cleavage, fusion, and inducement of a structural alteration in one or more members of the interaction. Significantly and advantageously, the present invention permits identification of both partners of a functional interaction. The criterion for detection of functional interactions by this approach is that the functional interaction is capable of being detected or selectively enhanced to permit detection.

The approach to identify functional interactions is similar to that for identification of affinity interactions but differs in some key features. When employing the previously discussed embodiment directed to affinity interactions, the association function of the NAMs is activated after interactors are permitted to display any affinity interactions which may detectably exist. In the case of detecting functional interactions the MOI association activity is controlled and the

Alamar Exhibit 1002
Alamar v. Olink

interactors in the library are forced together by association of their respective NAMs, NOT primarily between any affinities between the MOIs. One way to achieve this is by introducing an affinity between the different NAMs in the library by hybridization. This will create co-localizations of two or more interactors through the association of the NAMs. The result of the forced NAM association is a secondary library of cis reactive cells (CRCs). A CRC is a co-localization of two or more interactors which are associated by their NAMs (associated oligonucleotide). Thus the MOI components of the same CRC are brought in proximity although the MOIs might not have any intrinsic affinity. However, it is contemplated that in some embodiments the CRCs may be created by relying on intrinsic MOI affinities or a combination of both MOI affinity and NAM affinity.

If two individual MOIs are members of a functional interaction pair and the conditions favour this function, the MOIs in one CRC, being in closer proximity to one another, will interact more frequently with one another than with MOIs in other CRCs. It is important that the interactor concentration in the library be maintained lower than the relative concentration between the MOIs in the CRC.

Once the CRC library is constructed the functional interaction being investigated is stimulated. This can be accomplished by, for example, addition of a substrate or a shift in reaction conditions. The MOIs in the CRC will primarily act on each other in cis since they are physically linked, but they could potentially act in trans with MOIs in other CRCs. Hence, measures are taken to ensure that any functional interactions remain substantially restricted within the CRCs. Such measures may include, but are not limited to: temporal limitation of the functional activity; competition with the reverse functionality; and dilution of the associated library.

Alamar Exhibit 1002
Alamar v. Olink

In one specific embodiment, temporal limitation of the functional interaction is achieved by adding a component essential for the activation in a milieu where this compound is rapidly degraded. The degradation of the component may be chemical or enzymatic. In a further specific embodiment, competitive limitation is achieved by activating the functional interaction in a milieu where it is reversed. The action which reverses the functional interaction could be of chemical or enzymatic nature. If the functional interaction includes addition of a molecular group or alteration of a structure the reaction conditions can be maintained so that the added group hydrolyses or the structure reverts to the initial state quickly. Employing enzymes which remove the group added by the functional interaction may also be an option. If a functional interaction is present within a CRC it might be counteracted by the reaction conditions or by the addition of compounds but the intrinsic functional activity of the CRC will immediately counteract this. On the other hand, if a member of a CRC acts in *trans* on a separate CRC this functional interaction will be reversed by the reaction conditions and the probability of a repeated functional interaction in *trans* is low. Thus, an equilibrium that favour functional interactions within CRCs can be achieved.

In another specific embodiment, the CRC library is diluted subsequent to association of the interactors but prior to activation of the function being investigated. Interactor association will be more efficient at higher concentrations, but this also elevates the risk of *trans* interactions between the CRCs once the function is triggered. Diluting the concentration after formation of the CRCs will lower the concentrations of CRCs in the library while the proximity between MOIs within the CRCs remains.

Alamar Exhibit 1002
Alamar v. Olink

The library is then analyzed with respect to which CRCs display the desired functional interactions. The functional interaction being investigated is limited to those functions that are detectable.

Association of NAM molecules

Association of interactors is a key feature of the current invention in both the embodiment for detection of affinity interactions and the embodiment for detection of functional interactions. The binding of the interactors by their respective NAM association elements forms novel molecules, herein referred to as associated oligonucleotides, comprised of the nucleic acid tags comprising the identification sequences derived from the component interactors. Analysis of the associated oligonucleotide molecule formed from the combination of these tags permits identification of the interactor MOI components.

It is desirable that the NAM association results in a molecule comprising nucleic acid derived from each of the associated interactors. There are several approaches to achieve this and the examples described below and in figure 5 serve to illustrate some approaches, though this should not be construed as excluding other approaches. Association by enzymatic ligation may be achieved as illustrated in figures 5(a - e). Ligation substrates may include single stranded nucleic acid templated by a third nucleic acid, double stranded nucleic acid with protruding ends templated by a third nucleic acid, ligation of single stranded nucleic acids, templated ligation where a separate nucleic acid is introduced into the ligation product or ligation of double stranded nucleic acid with complementary protruding ends. Exemplary association approaches for polymerization include substrates comprising double stranded nucleic acid with complementary protruding 3′ ends, or single stranded nucleic acid with complementary 3′ ends, as illustrated in figures 5(d) and 5(f).

Alamar Exhibit 1002
Alamar v. Olink

Association may also be accomplished via chemical rather than enzymatic ligation. Chemical ligation may be performed by providing nucleic acids with reactive groups on the ends which allow association by chemical reaction. The chemical reaction may result in a nucleic acid of native composition which can serve as, for example, a polymerization substrate, or a nucleic acid derivative which does not serve as polymerization substrate. Two illustrative approaches for chemical association are illustrated in figures 5(g) and (h). Other non-limiting examples of NAM association approaches which could be used are gap fill polymerization and subsequent ligation and gap ligation as described in Lizardi et al. *Nat Genet* 19, 225, 232 (1998), or invader cleavage and subsequent ligation as described in Lyamichev et al, *Science* 260, 778-783 (1993), both of which are fully incorporated herein by reference. In a specific embodiment the nucleic acid is DNA but other nucleic acids may also be used.

The association approach utilized depends on the qualitative assay design. An illustrative example is the introduction of PCR amplification. If PCR amplification of associated NAMs is desired the association may preferentially combine NAMs comprising the forward primer with NAMs comprising the reverse primer to form amplification substrates. Selective association may be achieved by introduction of the forward primer to a portion of the interactor library with the association element in the 3′ end and the reverse primer in another portion of the interactor library with the association element in the 5′ end. This would be desirable since both primer motifs are required for amplification of associated NAMs by PCR.

In one embodiment, selective association approaches which exclude homo-association permit construction of interactor libraries with interactors comprised of more then one NAM per MOI. If homo-association is not excluded, then the homo-associations of NAMs associated to the same MOI be very efficient and potentially render the library useless.

Alamar Exhibit 1002
Alamar v. Olink

In one embodiment, two sub type libraries are constructed wherein each subtype comprises a different NAM and the NAMs of one subtype can associate only with NAMs from the other subtype. If the associated library is enhanced with PCR subsequent to association each NAM sub type comprises one of the PCR primer motifs. In the case of affinity interactions this design also allows association of several NAMs to each MOI, since unique NAMs are unable to associate. For functional interactions, the design allow presentation of all possible interactor combination in one mixing step, such as when interactor libraries comprising different NAM subtypes are pooled. If the NAM association is efficient and/or of high affinity this has the potential to override affinities present between MOIs since each interactor will associate with any interactor comprising a compatible NAM and this may occur before equilibrium is achieved with respect to the individual MOI affinities. Subsequent to construction of CRCs, the interactor library can be diluted to reduce the effects of MOI affinities between CRCs.

Association of nucleic acids in for affinity interaction analysis

For investigation of affinity interactions it is desirable to use NAMs comprising ssDNA or dsDNA. For investigation of affinity interactions it is important that the different NAMs do not display too high of an intrinsic affinity toward one another other. The affinity between the different association elements of the examples illustrated in figures 5(a) to (f) depend on the length of the complementary sequences of the association elements, and the addition of a third nucleic acid where applicable, except in 5(c) where no complementary sequence is present. Low affinity may by achieved by restricting each complementary sequence to 1-10 nucleotides. Association of NAMs by addition of a third templating nucleic acid can advantageously be performed by template addition in high concentration (e.g. >100 fold) compared to the interactor concentration. This will provide template mediated association of NAMs proximal at the

Alamar Exhibit 1002
Alamar v. Olink

moment of addition, while individual NAMs will hybridize to templates individually. Once all NAMs have hybridized to a template nucleic acid, this hybridization can block further template mediated dimerization of NAMs by intrinsic stability. This approach may allow utilization of template nucleic acids with higher stability.

Association of nucleic acids for functional interaction analysis

CRCs are constructed by the association of NAMs. The association may advantageously force different MOIs together and therefore affinity between the different NAMs is desired. In the exemplary illustrations 5(a, b, d, and f) this is achieved by introducing stable complementary sequences in the association elements.

When a third templating nucleic acid is employed for construction of CRCs, this may advantageously be added in equimolar or slightly higher (e.g. <5 fold excess) concentrations compared to the total interactor concentration. This may introduce an equilibrium where a majority of the NAMs are associated.

According to another embodiment, association between NAMs with low intrinsic NAM affinity is employed to construct a CRC library wherein the frequency of each CRC depends on the relative affinities of the interactors. This permits investigation of both affinity interactions and functional interactions with the same library. The CRC library may also be enriched for CRCs subsequent to association to avoid *trans*-interactions from individual interactors.

The association reaction can also be regulated by reaction conditions. For example, in an exemplary library comprised of proteins the protein-protein affinity interactions may be destabilized by introduction of high salt concentrations, which also stabilizes the nucleic acid interactions. Thus the inter-interactor affinities are shifted to depend on the nucleic acids.

Filtering and Selection

Alamar Exhibit 1002
Alamar v. Olink

Affinity interactions

The selection in the affinity embodiment of the inventive methods primarily concerns elimination of interactors which are not associated with any other interactors and maintaining the associated interactors. This can be achieved in several ways. One way is to perform selective amplification of associated NAMs, for example, with PCR, wherein there is one primer on each NAM so that only joined NAMs will produce amplification products. This approach might necessitate two types of NAMs per MOI (one with each primer motif) to enable identification of all interactions in the library including homo interactions.

A second amplification-based selection method is to cut the respective arms with a restriction enzyme and circularize the nucleic acid with a general template. Upon a second ligation, ligated arms will form a circular structure while individual arms will form a linear structure. The circular nucleic acid can then selectively be amplified by rolling circle amplification, a well-known technique in the art. Another way is to immobilize one part of the library to a solid phase and then perform the association of the NAMs. After the association, washing may remove all interactors which have not associated with any immobilized interactors. All molecules may or may not be present in solution or immobilised on the solid phase. The interactors may be immobilized in a spatially organized way or in a random fashion. Following washing of the solid phase the combinations of NAMs may be analyzed with or without amplification. One example of a suitable procedure would be elution of the interactors and amplification of the associated NAMs.

According to another embodiment, associated NAMs are selected by degrading any NAMs which are not associated. An example of this approach would be addition of an exonuclease which degrades any free nucleic acid ends. If the NAMs have been associated with a

Alamar Exhibit 1002
Alamar v. Olink

ligase the associated NAMs will not have any free 5′ or 3′ ends and not be amenable to degradation.

Functional interactions

To detect functional interactions in a pool of CRCs the CRCs which display functional interactions must be selected for detection and/or analysis. It will be appreciated by one of ordinary skill in the art that this step will be modified according to the nature of the functional interaction that is being investigated. As used herein, a "filter" or a "filtering process" is the means employed to selectively separate the desired CRCs. Nonlimiting examples of filters for functional interactions include: affinity purification of the form which has undergone alteration in the functional interaction; detection of the CRCs by labelling with affinity reagents which detect interactors which are included in the functional interaction; and introduction of a modified molecule which becomes incorporated by the functional interaction where the modification permits selective detection. Downstream of the filtering process of CRCs the compositions of associated NAMs are detected and/or quantified.

One example of a suitable filtering process is affinity purification. Filtering of positive CRCs in the case of an interaction which includes addition of a group by one member of the interaction to the other member such as a kinase or a glycosylation interaction between enzyme and substrate may be accomplished by affinity purification of CRCs with respect to the added group. In the specific example of a kinase interaction the library of CRCs can be affinity purified with a phosphate specific antibody. The specificity can be controlled to a phosphorylated amino acid, peptide motif or a whole protein depending on the assay. CRCs which do not contain the modification will be discarded and the CRCs with a modification will be retained and detected. Each CRC with a modification can be assumed to contain two members with the activity

Alamar Exhibit 1002
Alamar v. Olink

screened for, unless a homo interaction occurs. So in the case of kinase activity each selected CRC contains one kinase and one substrate thereof. The affinity filtering could also serve as a negative selector where the undesired molecules remain on the solid phase and the desired ones are collected. As non-limiting examples, this type of selection would be suitable in the case of phosphatase activity, glycosylation / deglycosylation, ubiquitinylation / deubiquitinylation and for any other interaction which serves to add or remove molecules or parts from one of the interaction partners as long as there are a specific affinity reagent for the molecule or part which is added or removed.

If the functional interaction results in a structural change in one of the partners to the interaction, affinity purification can be performed which only recognizes the altered portion. If no affinity reagents are available for the molecule which is added or removed an affinity function could be attached to the group which is added. For example, a protein like ubiquitin may be fused with one interaction partner and the added ubiquitin may be labelled with biotin which may then be affinity purified with streptavidine. However the "filtering" is not necessarily through affinity purification and can also be, for example, a conditional detection of the NAMs. If the detection output is coupled to the result of the interaction, for example, with an affinity reagent for the modification, only the CRCs with positive interaction pairs will be detected.

Another example of a suitable filtering process is conditional detection. The CRCs can be hybridized to a DNA oligonucleotide tag array where all possible combinations of identification sequences in the library are present. The different combinations of associated NAMs are spatially separated by the hybridization. To identify which NAM combinations represent a CRC with a positive interaction pair the array with the CRCs can be stained with a labelled affinity reagent specific for the interaction. However this approach requires that the

Alamar Exhibit 1002
Alamar v. Olink

MOIs are present during detection and thus no amplification of the associated NAMs is possible. A detectable function may also be associated with a modification, more specifically, for example, a fluorescent or radioactive label can be introduced by the modification, which more specifically comprises, in kinase interactions for example, a radioactive phosphate group.

Detection

In one embodiment, detection or readout of the library interactions is achieved by analysis of the NAM combinations, both qualitatively and quantitatively, by nucleic acid analysis. In a more specific embodiment, the nucleic acid composition is analyzed using microarrays. Other embodiments contemplate other known and unknown methods of nucleic acid analysis or include detection of molecules attached to the NAMs or the respective MOIs. The important feature is that at least two bits of information are retrieved, each representing one partner of the interaction.

Nucleic acid based readout

In one specific example, the analysis of interactions present in the library is accomplished by analysis of the combinations of the NAM elements by microarray analysis. The analysis of the associated oligonucleotides is similar both for the affinity interaction assay and functional interaction assay, once the CRCs of the functional interaction assay are filtered/selected for those CRCs which contain functionally positive interactors as described above. In some embodiments, the detection of the associated oligonucleotides is preceded by an optional amplification step.

Amplification methods

To increase the sensitivity and speed up the analysis, the associated oligonucleotides can be amplified by conventional nucleic acid amplification techniques such as, for example, polymerase chain reaction (PCR), rolling circle amplification (RCA), strand displacement

Alamar Exhibit 1002
Alamar v. Olink

amplification (SDA), nucleic acid sequence based amplification (NASBA), RNA transcription or invader assay. If the NAMs comprise RNA, reverse transcription may be used to translate these sequences into DNA. Other embodiments contemplate additional methods, known and unknown, of nucleic acid amplification. In the case of affinity interactions the amplification can also be used as a step to select for associated NAMs.

When PCR is utilized for amplification of associated NAMs, two primer motifs are required. In one specific example, PCR is used for amplification of associated NAMs wherein one primer motif is located on each NAM so that associated NAMs are selectively amplified. However PCR amplification requires two types of NAMs per MOI, one for each respective primer, to enable amplification of all intra-library MOI combinations. If inter-library interactions are detected, the libraries can be designed to comprise separate primer motifs.

RCA amplification of associated NAMs requires synthesis of a circular polymerization substrate. Conversion of the associated oligonucleotide constructed upon NAM association to a circular substrate can be achieved by removing the MOIs by, for example, restriction enzyme digestion and allowing circularization of the associated oligonucleotide by templated intramolecular ligation.

Sequence information retrieval approaches

The pertinent information which must be retrieved from the associated oligonucleotides is the identity of the two identity sequence motifs which provide information about the identity of the MOIs corresponding to that associated oligonucleotide. Thus at least two pieces of information need to be decoded. There are several ways to do this and most of them are either sequencing based or hybridization based approaches.

Hybridization based approaches:

Alamar Exhibit 1002
Alamar v. Olink

The use of microarrays to identify nucleic acid sequences in complex pools is well-known and has been adapted to high throughput platforms, for example, in expression profiling. The utilization of oligonucleotide arrays has increased the specificity and the introduction of *in situ* oligonucleotide synthesis has enabled very high throughput analysis by cheap probe synthesis of very complex microarrays. Currently the major platforms for microarray hybridization to oligonucleotides are fabricated by either robotic deposition of chemically synthesized oligogonucleotides or *in situ* synthesis of the oligonucleotides on the array. Separate platforms for microarray readout also exist, such as the Illumina platform, wherein oligonucleotides are attached to beads which subsequently organize on a solid support by self assembly.

The present inventive methods detect different tag combinations. If a library consists of *n* different interactors, each with a different identification sequence in the NAM, there would be $(n+n^2)/2$ potential combinations including homo-interactions, and $(n^2-n)/2$ potential combinations without homo-interactions. If two subtypes of NAMs are utilized for one MOI library, each subtype with a unique set of identification sequences, the number of tag combinations will be n^2, where each interaction is represented by two separate unique tag combinations except the homo-interactions.

Microarrays with probes for all these nucleic acid tag combinations could be manufactured for readout. The number of probes needed increase exponentially with the number of library members according to the formula above. However *in situ* synthesis of microarrays can produce very complex arrays which could be used for any interaction study since the nucleic acid tag set generally is independent of the MOIs.

Alamar Exhibit 1002
Alamar v. Olink

The general approach to detecting the NAM combinations by microarray based readout involves construction of a microarray wherein all possible tag combinations in the library are present. The associated oligonucleotides, or amplification products thereof, are allowed to hybridize to the array. Preferably the melting temperature of the identity sequence elements should be designed so that the difference between hybridization of one element and two elements is maximized. Other criteria that increase hybridization stringency could also be introduced to maximize specificity, such as, for example, enzymatic discrimination steps. Non-limiting examples of these include ligation and/or polymerization.

The design of the microarray-based dual tag detection includes hybridization of a nucleic acid comprising at least two separate tag or identification sequence motifs to array probes with their respective complements. Illustrative examples of microarray hybridization designs are provided in figure 6. Hybridization of the two identification sequence elements t_x and t_y (red in figure 6) may be accomplished by hybridization with or without a general sequence (black in figure 6) present on the targets. The elements may be hybridized as a continuous sequence (6b)or with a spacing sequence between the elements (Fig 6a). The elements may also be designed to form a circular structure upon hybridization (Figures 6c and d). The specificity of the detection can be enhanced by introduction of enzymatic steps. 6e represents ligation of a general detection oligonucleotide. Incorporation of nucleotides comprising a detectable function is illustrated in 6f and solid phase polymerization and subsequent detection by hybridization of a general detection oligo is illustrated in Figure 6(g).

Specificity

The specificity of the detection of combinations of identifying elements might have to be enhanced if the stringency from the hybridization is insufficient. The hybridization of

Alamar Exhibit 1002
Alamar v. Olink

identification element combinations differs from regular hybridization of single elements. If there are n*2 NAMs (two NAMs per protein) and n^2 array identification element oligonucleotides on the array the different array elements can be referred to as $t_{x,,y}$ where x is the identity of one identification element position and y the identity of the other identification element. Thus x and y can be any element between 1 and *n*. The combination of the arbitrary identifying elements *a* and *b* will thus complementary to the array element $t_{a,b}$ but they will also be partially complementary to the array elements $t_{x,b}$, and $t_{a,y}$.

The specificity of the hybridization should substantially ensure that the combination *a,b* primarily hybridizes to the array element $t_{a,b}$. Cross hybridization might occur to some or all of the array elements $t_{x,b}$, and $t_{a,y}$ to some extent, however. In a complex pool of tag combinations which may have a high dynamic range of the different combinations the discrimination between two species may become difficult. If, for example, the combination *a,b* is present in 10,000 fold excess over the combination *a,c* the nucleic acid from the *a,b* combination might cross hybridize to the array element $t_{a,c}$ which is dedicated for the *a,c* combination. This specificity problem may be solved by, for example, the introduction of enzymatic discrimination steps.

Enzymatic steps have facilitated specific detection of individual nucleotide sequences in the human genome by, for example, utilization of a ligase (see Landegreen et al. *Science* 241, 1077-1080 (1988), Hardenbol et al. *Nat Biotechnol* (2003), Yeakley et al. *Nat Biotechnol* 20, 353-358 (2002), fully incorporated herein by reference), or a polymerase (see Fan et al. *Genome Research* 14, 878-885 (2004), fully incorporated herein by reference). This specificity should be sufficient for a synthetic system where the sequences of the nucleic acid sequences can be designed *in silico*.

Polymerase based discrimination;

Alamar Exhibit 1002
Alamar v. Olink

To increase the selectivity with a polymerase the hybridized nucleic acids can be allowed to serve as primer or template for a polymerase extension reaction including labelled nucleotides. If this is performed with the proper polymerase which lack, for example, 3′exonuclease activity, and the detection of the hybridized molecules depend on this step, the selectivity of detection is enhanced. Thus only nucleic acids which have hybridized to the array and can form a structure which can serve as a substrate for polymerization will be detected (Figure 6f). This approach has been utilized for microarray based discrimination of single nucleotide variations in human genomic DNA (see Gunderson, KL et.al Nat Genet. 2005 May;37(5):549-54, fully incorporated herein by reference).

Ligation based discrimination:

To increase the selectivity of detection by ligation the hybridized nucleic acid can be allowed to act as a template or substrate for ligation. The detectable function can be added on a third oligonucleotide which can be joined to the surface complex by ligation. Only if the hybridization event yields a structure which together with the third oligonucleotide, forms a substrate for ligation, will detection be mediated. The ligation may be performed at the 5′end, the 3′ end, or both ends of the hybridizing nucleic acid (Fig 6e).

Ligation and polymerization based specificity enhancement.

Ligation can also be used in conjunction with polymerization to permit specific detection. If the identifying elements are hybridized so that a circular structure is formed where the 5′ and 3′ end are in juxtaposition to each other, these two ends can be joined by a ligase to form a circular nucleic acid template in the array. This circular nucleic acid can then serve as a template for polymerization from the array nucleic acid 3′ end. If a processive nucleic acid polymerase is used, for example, phi 29 polymerase, a concatemeric polymerization product of several copies

Alamar Exhibit 1002
Alamar v. Olink

of the nucleic acid circle can be produced (see Hatch, A et al Genet Anal. 1999 Apr;15(2):35-40, fully incorporated herein by reference). These copies can then be detected by hybridization of an oligonucleotide attached to a detectable function which hybridizes to the sequence between the identification sequences. This approach only yields signal output if a ligation substrate is formed (Fig 6g). Other non-limiting examples of ligation approaches which could be used in combination with polymerization are gap fill and subsequent ligation as described in Lizardi et al. *Nat Genet* 19, 225, 232 (1998), or invader cleavage and subsequent ligation as described in Lyamichev et al, *Science* 260, 778-783 (1993).

Sequencing based approaches

Sequencing based approaches retrieve the nucleotide sequence of the two identifying elements. These approaches yield a very high complexity of information and thus relatively few nucleotides need to be sequenced to enable identification of a unique element. For example the complexity of five nucleotides is 4^5=1024 different unique sequences. Thus if sequencing approaches are used very short identifying elements can be utilized. The length of the element used for identification can be adjusted to suit the complexity of the library being investigated. Identification of two different tags in a library of 1000 members and 10^6 combinations would require sequencing of at least 10 nucleotides per interaction, and five nucleotides per MOI.

However, when sequencing complex pools of nucleic acids, measures have to be taken to isolate the individual molecules or cloned products thereof. This can be done with conventional techniques like molecular cloning into bacteria and subsequent Sanger sequencing. The sequencing of molecular identification sequences of approximately 10 nucleotides can be speeded up significantly by using approaches similar to serial analysis of gene expression (SAGE) where short sequences are joined together before sequencing and several identifying

Alamar Exhibit 1002
Alamar v. Olink

elements can be sequenced in one sequence read. Sequencing of complex pools can also be performed by high throughput approaches like the 454 platform (www.454.com). This approach includes emulsion based clonal PCR amplification and subsequent sequencing on a solid phase. Approaches like this permit very high throughput analysis of the readout.

Depending on the possible readout length these approaches may be combined with SAGE like polymerization of paired identification elements. Solid phase sequencing of the joined tag motifs or possibly concatemeric products thereof would permit information retrieval of the interaction frequency in digital form. However, the sequences combined by the interaction have to be significantly more salient than the background frequency, otherwise the majority if the sequences retrieved will only be random combinations. Sequence based detection of an interaction which is 1000 fold less frequent than another will require sequencing of 1000 identical sequences of the frequent one per copy of the rare one.

Interactor based detection

In some situations nucleic acid based amplification of the NAM might not be desired. The different combinations of interactors may then be analyzed directly. One advantage of this is that other identification moieties than the actual nucleic acid could be used or the MOIs may be identified directly. Nonetheless, the readout approach requires identification of at least two separate pieces of information from the associated interactors.

One way of retrieving information from two different species is utilization of a nucleic acid tag array for hybridization of the combinations. In this approach the array would only contain individual tags on the array. Thus if there are *n* identifiers in the library the array would consist of n spatially separated oligonucleotides. Subsequent to the NAM association and filtering, which could include selection for combined NAMs, the complex of associated

Alamar Exhibit 1002
Alamar v. Olink

interactors is hybridized to the array. The localization of the interactor complexes will reveal the identity of one of the members and the second member is identified by a separate tag moiety, which may be of another nature than nucleic acid. For example, if fluorescent readout is chosen the second tag could be decoded with sequential combinatorial hybridization as described in Gunderson et al., *Genome Research* 14, 870-877 (2004) disclosed fully herein by this reference.

Another approach could be to identify the second tag by matrix assisted laser desorption / ionisation time of flight mass spectrometry (MALDI TOF MS). According to this method, the identification elements in the NAMs would include molecular "tags" which could be identified with this approach. These tags would preferentially be non nucleic acid molecules associated with the NAMs. Depending on the nature of the MOIs these molecules could be identified by MALDI TOF MS directly without the utilization of a separate molecule for the detection.

EXAMPLES

The following examples are intended to illustrate specific embodiments of the present invention and should not be construed as limiting the invention as defined by the claims. All detection schemes encompassed within the scope of the present invention include combinatorial association of NAMs.

Example 1

Affinity interactions

This example illustrates embodiments directed to detecting affinity interactions.

The individual proteins to be investigated are produced by recombinant expression and purified. These proteins serve as MOIs in the library and are coupled with NAMs to form interactors.

Alamar Exhibit 1002
Alamar v. Olink

The NAMs are double stranded DNA molecules and comprise a primer motif, a nucleic acid identification sequence and an association sequence. To enable amplification of all intra-library combinations, there are two sub populations of proteins created, one with the forward PCR primer and one with the reverse primer in their respective NAMs (Figure 3). The nucleic acid identification elements in the two sub-populations comprised by the same protein are not identical, thus each protein is "encoded" by two nucleic acid identification elements in the library.

The individual NAMs are synthesized from two generic PCR products which comprise the general primer motif and spacer sequence. The nucleic acid identification sequence is introduced by performing a second PCR on the generic PCR product with where one primer contains a conjugation function and the other a 5′ tail encoding the nucleic acid identification element and the association element.

The coupling function is performed by a sticky end which is created by restriction enzyme digest. The protruding end is created by utilizing a restriction endonuclease which produces a non-palindromic protruding end. The NAMs encoding the forward primer are designed to have one polarity of the protruding end and the NAMs encoding the reverse primer are designed to have the complementary protruding end. The design permits association of several identical NAMs to one MOI without homo-ligation between the NAMs. The individual NAM subtypes are gel purified and associated with the MOI by the conjugation element in one of the primers. The proteins are recombinantly expressed with the IMPACT system from New England Biolabs and the conjugation function of the NAM PCR product is a Cysteine residue. C-terminal fusion of the protein in the IMPACT system generates a SO_3^- group at the N-terminal of the protein which then can be specifically conjugated to the Cysteine group on the NAM.

Alamar Exhibit 1002
Alamar v. Olink

However IMPACT may perform poorly on certain members and additional approaches for conjugation may also be utilized to include all desired proteins.

Subsequent to interactor synthesis by individual coupling of the PCR product derived NAMs to the proteins, the individual interactors are pooled in 1X PBS with inclusion of bovine serum albumin and polyadenine. The interactors are allowed to affinity interact and are subsequently ligated by addition of a ligase mix containing ATP, T4 DNA ligase TRIS HCl pH 7,5 and $MgCl_2$. Following ligation the library is heated to deactivate the ligase and an aliquot of the library is transferred to a PCR reaction. PCR is performed with one biotinylated primer and one Cy3 labelled primer. Subsequent to PCR amplification the fluorescent strand is purified by immobilization of the PCR product on dynabeads and elution of the fluorescent strand. The elution is mixed with hybridization buffer and co-hybridized to a micoarray with a control sample. The control sample is treated with protease degradation before ligation and the fluorescent primer is Cy5 labelled. After hybridization the microarray is scanned and analysed with respect to the fluorescent ratios.

The microarray comprises features with all combinations between tags on reverse primer and forward primer NAMs. Thus if 100 proteins are analysed the number of features is 10,000, representing every library interaction in duplicate.

Example 2

Functional interactions

This example illustrates construction of a library of interactors and CRCs and detection of a specific enzyme-substrate functional interaction where one MOI acts as an enzyme and a second MOI is the substrate.

Alamar Exhibit 1002
Alamar v. Olink

Enzyme substrate interactions can be difficult to distinguish with conventional approaches since the enzyme might have low or transient affinity to the substrate. A specific choice of enzyme-substrate pair is kinases and their respective substrates. For screening proteins for kinase and kinase substrate properties a library of interactors is constructed where the MOIs of the library are potential kinases and substrates respectively.

The interactor library comprises two sub-populations, the first comprising potential kinase MOIs and the second comprising potential substrate MOIs. Each sub-population of a potential kinase is conjugated to the 5′ end of a oligonucleotide subset comprising the forward primer motif for downstream amplification, the association element in the 3′ end of the oligonucleotide and one of a plurality of nucleic acid identification elements. Each sub-population of a potential substrate is conjugated to the 5′ end of an oligonucleotide comprising the downstream reverse primer motif, an association element in the 3′ end of the oligonucleotide and one of a plurality of nucleic acid identification sequences (Figure 4).

The NAM association function is a motif for polymerase assisted association at the ends of the oligonucleotides which are not conjugated to the protein. The 3′sequences are designed to be complementary and brings the oligonucleotides comprising the forward primer and the reverse primer together. Upon polymerisation association is stabilised and the respective tags and PCR primers are encoded into the same nucleic acid molecule. The kinases and substrates are proteins and the conjugation between these proteins and their respective oligonucleotides are mediated by expression of the proteins by the IMPACT system and conjugation of oligonucleotides with a Cysteine residue. However proper interactor synthesis may require different expression and conjugation approaches depending on the nature of the protein.

Alamar Exhibit 1002
Alamar v. Olink

Following interactor synthesis CRCs are stabilized by the addition of a polymerase, nucleotides and the appropriate reaction conditions. 20nt of the oligonucleotide 3′ ends of the two interactor sub populations are complementary and subsequent to hybridization the oligonucleotides serve as both polymerisation templates and substrates respectively. The polymerisation encodes the two identification sequences joined by the hybridisation into one nucleic acid molecule and also stabilise the hybridization of the oligonucleotides further. Polymerisation is achieved by addition of klenow fragment exonuclease deficient polymerase, nucleotides and the appropriate reaction buffer.

The interactors are associated through polymerization to form CRCs and the library is diluted so that the concentration of CRCs becomes significantly lower than the relative concentrations between the interactors in the CRC. To regulate the kinase activity apyrase is added to the library before activation of kinase activity by ATP addition. The appropriate relative concentrations of apyrase and ATP might have to be titrated depending on the kinase activity screened for. The apyrase will degrade the ATP added to the CRC library and thus provide a temporal limitation of kinase activity. Subsequent to activation, the library of CRCs is affinity purified with an antibody specific for a phosphorylated amino acid or amino acid sequence. The library can be split into subsets which are investigated with separate affinity reagents. The selected library is then eluted and an aliquot is amplified by PCR comprising one Cy3 labelled primer and one biotinylated primer. The PCR product is immobilized on streptavidine beads and the fluorescent strand is eluted and co-hybridized to a microarray with a control PCR Cy5 labelled product derived from a library which has not been affinity purified. The microarray contains features which are complementary to all kinase and substrate tag combinations in the library, representing all kinase and substrate interactions between the two libraries.

Alamar Exhibit 1002
Alamar v. Olink

Example 3

Reaction discovery

In one specific embodiment screening for reactivity between molecules is provided. A library of molecules is subsequently screened for formation of chemical bonds between the library molecules. The procedure is similar to discovery of affinity interactions, however, in this embodiment, the interaction results in a covalent bond between the molecules. The embodiment enables reactivity screening in different reaction conditions and/or in the presence of catalysts.

The molecules of interest are each tagged with a single stranded DNA NAM to form the interactor library. Two sub types of interactors are constructed, comprising the same molecules of interest, one with free NAM oligonucleotide 3′ ends and one with free NAM oligonucleotide 5′ends. These are constructed so that any free 3′ end can be joined with any free 5′ end by the addition of a DNA oligonucleotide template, complementary to each of the oligonucleotide sub types, by the addition of a ligase and under the appropriate reaction conditions. Each subtype also comprises one PCR primer motif. Oligonucleotides with free 5′ ends comprise the forward primer and oligonucleotides with free 3′ ends comprise the reverse primer. All DNA NAM oligonucleotides also comprise one unique sequence element for identification.

Ten different molecules of interest are investigated for reactivity in three different reaction conditions, thus there are 10*10=100 different possible reactions, each interrogated in three different reaction conditions.

The 10 interactors comprising NAMs with identical association elements are pooled to form two populations, one comprising 10 interactors with free 3′ ends one and comprising 10 interactors with free 5′ ends. These pools are subsequently mixed. The interactors are incubated in concentrations of 10-100nM to allow the chemical reaction to occur. Subsequently the reacted

Alamar Exhibit 1002
Alamar v. Olink

library is diluted down and the template oligonucleotide is added in high concentration together with a ligase mixture. Since the library is screened for formation of covalent bonds, the interaction will not be effected by dilution subsequent to the reaction, however the dilution will reduce the random association. An aliquot of the ligation reaction is transferred to a PCR reaction where the reverse primer contains a 5′ biotin modification and the forward primer a 5′ Cy5 label. The PCR products are then hybridized to a tag microarray containing all different tag combinations. The hybridization to the microarray is designed according to figure 6(a).

Ten different molecules of interest are investigated, M_1, M_2, ..., M_{10}. These molecules are used to create 20 different interactors divided into two pools; one pool with free 3′ ends and one with free 5′ ends. Each molecule of interest is incorporated in two separate interactors and the identifying sequence elements in these interactors are not identical. The reaction between two molecules of interest will promote intra molecular joining of the respective NAMs during the ligation step. These associated oligonucleotides are then detected by micorarray analysis. The microarray comprise 100 different features, each feature representing a combinations between two different tags. All interactions except the homo interactions are represented by two tag combinations on the tag microarray. Subsequent to microarray detection the signal intensity of the microarray features is compared within the microarray and between experiments with different reaction conditions. Three different reaction conditions are investigated, one comprised only of reaction buffer, one comprised of reaction buffer and inclusion of a chemical catalyst, and the last one comprising reaction buffer and a enzyme catalyst, which is presumed to enable reactions between the current substrates.

Results:

Alamar Exhibit 1002
Alamar v. Olink

Results are illustrated in figure 7. The figure schematically illustrates a microarray comprising tags complementary to combinations of identification sequence elements. The tags are designated 1 to 10 where tag 1 corresponds to the interactor comprising molecule of interest M_1 and so forth. Thus the interactions between two molecules can be deducted from the signal intensity of the respective tag combination. In reaction buffer M_6 reacts weakly with molecules of interest M_2 through to M_6, the last interaction representing an homo interaction. When a chemical catalyst is added reactions occur of moderate intensity between M_2-M_8, M_6-M_5, and M_3-M_3, the last one also representing a homo interaction. When the chemical catalyst is exchanged to an enzymatic one reaction of high intensity occur, M_9-M_6. The relative reaction efficiencies of the reactions can be interpreted from the array intensities, where a high intensity corresponds to an efficient reaction.

Example 4

Enhancement for microarray analysis

To facilitate detection of associated oligonucleotides the nucleic acid to be analyzed may be modified to suit the readout platform. This modification may include, e.g., modifying enzymes or biochemical modification. In the specific case of tag microarray mediated detection it is desirable to hybridize single stranded DNA. Non limiting examples of generation of single stranded DNA include: amplifying the associated oligonucleotides with RCA which produces a single stranded amplification product comprised of concatemeric amplicon repeats; and single stranded DNA generation subsequent to PCR amplification by, e.g., utilization of one biotinylated primer and streptavidine affinity purification, PCR with one phosphorylated primer and subsequent lambda exonuclease treatment or performing PCR with one primer in excess, also known as asymmetric PCR or Linear After The exponential (LATE) PCR.

Alamar Exhibit 1002
Alamar v. Olink

Depending on the qualitative design of the microarray detection, the single stranded DNA may be further modified by, e.g, biochemical or enzymatic modification. One specific example of quantitative design is the introduction of a ligation step involving two identification elements on the tag array to form a circular nucleic acid. To enable ligation of the two identification elements to the microarray tag oligonucleotide, a nucleic acid has to be generated that comprises one identification element in the 3′ end and one in the 5′ end. Downstream of single stranded DNA generation the single stranded DNA may be cleaved with a restriction enzyme next to the identification elements. Restriction enzyme digestion of single stranded nucleic acid may be achieved by hybridization of oligonucleotides comprising the restriction enzyme recognition sequence complementary to the amplification products. In one specific embodiment the enzyme comprises a type IIS restriction endonuclease and in a further specific embodiment the Type IIS restriction enzyme comprises MlyI.

One general design embodiment of the NAMs provides an associated oligonucleotide, downstream of association, comprising the association sequence in the middle flanked by two identification elements which, in turn, are flanked by general sequence in which specific examples can be utilized for amplification by, e.g., PCR or RCA. The associated oligonucleotide may be enhanced to comprise a single stranded amplification product by, for example, the non limiting methods described above.

This single stranded amplification product may then comprise the association sequence in the middle flanked by two identification elements which in turn are flanked by a general sequence (Fig 8). The general sequence may be designed to comprise the recognition sequence for a restriction enzyme, more specifically a type IIS restriction endonuclease, and in a more specific example comprises the type IIS restriction enzyme, MlyI. In one specific

Alamar Exhibit 1002
Alamar v. Olink

embodiment of generating a single stranded nucleic acid comprising the identification elements in the 3′ and 5′ ends, the nucleic acid may be designed so the restriction enzyme cleavage occur right between the general sequence contacted by the hybridizing oligonucleotide and the single stranded identification element. In a very specific embodiment, the restriction enzyme comprises MlyI. This results in single stranded nucleic acid with a general sequence in the middle and the identification elements in the 3′ and 5′ ends.

This nucleic acid may be hybridized and/or ligated to a tag microarray complementary to the identification elements, to form a circular nucleic acid as illustrated in figure 5(c). If the nucleic acid is ligated to the tag microarray to form a circular nucleic acid the structure may serve as substrate for RCA amplification. In a specific embodiment, RCA is primed by the microarray tag oligonucleotide and in a more specific embodiment the polymerase is the phi29 polymerase.

EXAMPLE 5

This example demonstrates that encoding interactors with unique tags and detecting the combinations formed therefrom on a combinatorial array may resolve all interactions within a library. A library of interactors is provided and association of a defined library subset is detected by readout of the respective tag combinations. MOIs are thiol groups. Thiol groups can be induced to dimerize pair-wise and form covalent bonds. This may be compared to a reaction screen where reactive compounds are identified or it can be viewed as formation of an affinity interaction with infinite affinity.

Experimental Design

The application of resolving pair wise reacted groups within an interactor library is exemplified. Formation of a covalent bond is interrogated within a library. Dimerization of thiol

Alamar Exhibit 1002
Alamar v. Olink

groups serves as an example. Oligonucleotides with thiol groups will dimerize pairwise if mixed and subsequently allowed to react. Five oligonucleotides with association elements in the 3′ end and 5 oligonucleotides with association elements in the 5′ end are employed. Two of the 3′arms and 2 of the 5′arms are equipped with thiol groups. Subsequent to dimerization interacting MOI are identified by association of the respective NAMs by ligation (Figure 5a). The ligation joins the two tag motifs encoding the thiol groups. The ligated product can also serve as amplification template for PCR, enabling selective amplification of the ligated NAMs. Subsequent to the PCR single stranded DNA is generated by one biotinylated primer and streptavidine support. The buffer for the single stranded DNA is changed by a G50 column. The single stranded DNA is then modified to have the two tag motifs in the ends. This is accomplished by cleaving off the PCR primer sites next to the identification tags. (Figure 8) Cleavage is mediated by annealing two oligonucleotides directing the restriction enzyme cleavage and then adding the restriction enzyme MlyI. The cleaved product will have the identifying elements in the 3′ and 5′ end respectively. These are then added to a micorarray comprising oligonucleotides complementary to the tag combinations. The reporter molecules created by the cleavage hybridize to the array oligonucleotide with the correct tags similar to Figure 6(c). The hybridized DNA is then ligated and a circular substrate forms which can serve as a template for rolling circle replication and detected by general detection reagents (Figure 6(g)).

Experimental Procedure

a. Library Construction via dimerization of thiol modified oligonucleotides: all oligonucleotides are obtained from commercial vendors, including oligonucleotides with thiol groups for reaction screening or amine groups for array manufacturing. Each oligonucleotide comprises one unique tag element and one of two primer sites. A selected subset of the oligonucleotides are also

Alamar Exhibit 1002
Alamar v. Olink

equipped with thiol groups enabling selective dimerization. The mixture comprising thiol modified oligonucleotides is dimerized by addition of 1M DTT, facilitating reduction of the homodimeric oligonucleotides. Subsequent exclusion of the DTT is achieved by a buffer exchange column.

b. Association and selection by ligation and PCR

A total volume of 5µl comprising 100pM of the thiolated (NAM-Thiol-D5, NAM-Thiol-E5, , NAM-Thiol-A3, , NAM-Thiol-B3 and non thiolated oligonucleotides (NAM B5, NAM, C5, NAM F5, NAM C3, Nam D3, NAM E3) is mixed to constitute a final volume of 50µl comprising 1X PCR buffer (available from Invitrogen, Norway), 5pmol of each PCR primer (FWDprime, REWprime), 1.5U Taq platinum (available from Invitrogen, Norway), 3mM $MgCl_2$, 80mM ATP, 100nM of the oligonucleotide LigA, 200nM dNTP and 1U T4 DNA ligase. The mixture is incubated at 95C for 2min and then cycled 30 times between 60C and 95C.

c. Analysis of the two identification sequences in the associated oligonucleotides

Generation of single stranded DNA

Five µl of Streptavidine coated Dynabeads M-280 (Invitrogen, Norway) is washed twice in B&W buffer (1M NaCl, 0.5mM EDTA, 5mM Tris-HCl pH 7.5) and subsequently the reaction is added to the beads and allowed to bind for 10minutes. The beads are then washed twice in 100µl of B&W buffer. Subsequent to washing the beads are incubated with 5µl 0,1M NaOH for 5 minutes. The NaOH supernatant is subsequently removed and mixed with 5µl 0,1M HCl and 5µl 0,1M TRIS-HCl pH8.

Buffer exchange

The eluted and neutralized DNA is added to a G50 column (available from General Electric) and purified according to the manufacturers instructions.

Alamar Exhibit 1002
Alamar v. Olink

Restriction enzyme cleavage

10µl of the column eluate is mixed with 2µl of S8 buffer (20mM Tris Ac pH 8, 50mM KAc, 10mM $(NH_4)_2SO_4$, 10mM $MgAc_2$, 1mM DTT) 5U MlyI (available from New England Biolabs), 10pmol of the oligonucleotides REVprime and MlyIRE, 2µg of BSA and 5.5µl of ddH_2O. The reaction is first incubated for 1h at 37°C and then at 75°C for 15min.

Microarray manufacturing

Microarrays are manufactured by a GMS 417 microarray printer. The oligonucleotides are printed on Codelink™ slides (available from General Electric) in a solution comprising 10µM oligonucleotide, 150mM $NaCO_3$ pH9, and 0.05% SDS. After printing, the slides are incubated for at least 1h in a humid chamber and then blocked with 30ml 70% EtOH, 70ml 1X PBS and 0,235g $NaBH_4$. Slides are washed and dried by centrifugation.

Ligation to microarrays

Following incubation the reaction is mixed with 3µl 10X Ampligase Buffer (Epicentre), 25µg BSA, 5U Ampligase (Epicentre) and 23,5µl ddH_2O. The mixture is added to a microarray slide comprising the oligonucleotides Tag3A5B; Tag3A5C, Tag3A5D, Tag 3A5E, Tag3A5F, Tag3B5C, Tag3B5D, Tag3B5E, Tag3B5F, Tag3C5D Tag3C5E, Tag3C5F, Tag3D5E, Tag3D5F, Tag3E5F. The reaction is incubated on the slide in a silicon rubber chamber (Genome Res. 2000 Jul;10(7):1031-42.) based on a 96 well format at 50°C over night. The slide is then washed in 0.75X TNT buffer (75m M Tris-HCl, pH 7.5, 0.1125 M NaCl, and 0.0375% Tween 20) rinsed in 0.1X SSC and dried by centrifugation.

Rolling circle amplification and detection on microarrays

Alamar Exhibit 1002
Alamar v. Olink

A reaction mixture comprising 1X Phi29 buffer (available from Fermentas), 25μg BSA, 0.2mM dNTP, 10U Phi29 in the final volume of 50μl was added to the slide and incubated for 1h at 37°C. The slide is then washed in 0.75X TNT buffer, rinsed in 0.1X SSC and dried by centrifugation.

The slide is incubated with 50μl comprising 2XSSC, 0.1X SDS and 10nM of the oligonucleotide DET at room temperature. The slide is then washed in 0.75X TNT buffer, rinsed in 0.1X SSC and dried by centrifugation.

Data acquisition and analysis

The slide is scanned in a Genepix™ scanner and the data acquired by Genepix pro 5.0 exported and analysed in Microsoft Excel® (Microsoft Office 2003). Signal-to-Noise was calculated by dividing the Feature signal with the average signal of all features.

d. Results

The microarray features are represented in Figure 9. The grid represents the combination of the different oligonucleotide tag combinations. The four microarray features representing combinations of the interactors equipped with thiol groups show increased signal to noise, due to their interaction. The other tag pairs represented in the library and on the array display low signal to noise, however significant background is visible in one feature due to cross hybridization of the detection oligonucleotide.

Alamar Exhibit 1002
Alamar v. Olink

Claims:

1. A method of detecting affinity interactions between at least two molecules of interest, the method comprising:

   a. forming a plurality of interactors by coupling each molecule of interest with at least one nucleic acid moiety comprising an identification sequence element and at an association element;

   b. promoting an association between at least two nucleic acid moieties from different interactors to form a plurality of unique associated oligonucleotides, wherein each nucleic acid moiety may form more than one unique associated oligonucleotide, and wherein each unique associated oligonucleotide comprises at least two identification sequence elements derived from the at least two nucleic acid moieties;

   c. selecting the plurality of unique associated oligonucleotides; and

   d. subjecting the selected associated oligonucleotides to an analysis that permits detection of the at least two identification sequence elements.

2. The method of detecting affinity interactions between at least two molecules of interest as recited in claim 1, wherein the plurality of interactors comprises more than two interactors.

3. The method of detecting affinity interactions between at least two molecules of interest as recited in claim 1, wherein selecting the plurality of associated oligonucleotides is accomplished via enhancement.

Alamar Exhibit 1002
Alamar v. Olink

4. The method of detecting affinity interactions between at least two molecules of interest as recited in claim 2, wherein enhancement is by PCR amplification.

5. The method of detecting affinity interactions between at least two molecules of interest as recited in claim 1, wherein coupling is covalent or noncovalent.

6. The method of detecting affinity interactions between at least two molecules of interest as recited in claim 5, wherein coupling is covalent.

7. The method of detecting affinity interactions between at least two molecules of interest as recited in claim 1, wherein association is via ligation.

8. The method of detecting affinity interactions between at least two molecules of interest as recited in claim 1, wherein association is via polymerization.

9. The method of detecting affinity interactions between at least two molecules of interest as recited in claim 1, wherein at least one molecule of interest comprises a protein, a protein domain, a peptide, a peptoid, a functional enzyme, a small molecule, a polymer, a non-proteinaceous cellular product, a virus particle, a cell metabolite, a lipid, a carbohydrate, a nucleic acid, an inorganic compound, a reactive compound, or any combination thereof, however, wherein all molecules of interest are not nucleic acids.

Alamar Exhibit 1002
Alamar v. Olink

10. The method of detecting affinity interactions between at least two molecules of interest as recited in claim 9, wherein at least one molecule of interest comprises at least one protein.

11. The method of detecting affinity interactions between at least two molecules of interest as recited in claim 9, wherein the protein is an expression product of a nucleic acid molecule, and further wherein at least some of the plurality of interactors is formed by coupling the protein with a nucleic acid moiety comprising an identification sequence element derived from the nucleic acid molecule.

12. The method of detecting affinity interactions between at least two molecules of interest as recited in claim 11, wherein the nucleic acid molecule comprises ribonucleic acid or deoxyribonucleic acid.

13. The method of detecting affinity interactions between at least two molecules of interest as recited in claim 9, wherein the at least two molecules of interest comprise at least one protein and at least one nucleic acid.

14. The method of detecting affinity interactions between at least two molecules of interest as recited in claim 1, wherein detecting further comprises quantifying and/or determining strength of the affinity interactions.

Alamar Exhibit 1002
Alamar v. Olink

15. The method of detecting affinity interactions between at least two molecules of interest as recited in claim 1, wherein the analysis comprises high throughput screening.

16. The method of detecting affinity interactions between at least two molecules of interest as recited in claim 13, wherein the high throughput screening comprises a micro array analysis.

17. The method of detecting affinity interactions between at least two molecules of interest as recited in claim 1, wherein the analysis comprises nucleic acid sequencing.

18. The method of detecting affinity interactions between at least two molecules of interest as recited in claim 17, wherein the analysis comprises sequencing by synthesis.

19. The method of detecting affinity interactions between at least two molecules of interest as recited in claim 1, wherein promoting an association between at least two nucleic acid moieties to form a plurality of associated oligonucleotides comprises manipulating conditions to maximize cis associations and to minimize trans associations between the at least two nucleic acid moieties.

Alamar Exhibit 1002
Alamar v. Olink

20. A method of detecting functional interactions between at least two molecules of interest, the method comprising:

a. forming a plurality of interactors by coupling each molecule of interest with at least one nucleic acid moiety, the nucleic acid moiety comprising an identification sequence element and an association element;

b. forming a plurality of cis-reactive cells wherein a cis-reactive cell comprises at least two interactors bound in proximity to one another by an associated oligonucleotide formed from the association between at least two nucleic acid moieties, wherein the associated oligonucleotide comprises at least two identification elements derived from the at least two nucleic acid moieties;

c. subjecting the plurality of cis-reactive cells to conditions which stimulate a desired functional interaction having a detectable trace;

d. selecting all cis-reactive cells exhibiting the detectable trace;

e. subjecting the associated oligonucleotides from the cis-reactive cells selected in (d) to an analysis that permits detection of the at least two identification sequence elements.

21. The method of detecting functional interactions between at least two molecules of interest as recited in claim 20, wherein selecting all cis-reactive cells exhibiting the detectable trace is accomplished via enhancement.

Alamar Exhibit 1002
Alamar v. Olink

22. The method of detecting functional interactions between at least two molecules of interest as recited in claim 21, wherein selecting all cis-reactive cells exhibiting the detectable trace is accomplished affinity purification.

23. A method of detecting functional interactions between at least two molecules of interest as recited in claim 20, further comprising enhancing the selected associated oligonucleotides by PCR amplification.

24. A method of detecting functional interactions between at least two molecules of interest as recited in claim 20, wherein coupling is covalent or noncovalent.

25. A method of detecting functional interactions between at least two molecules of interest as recited in claim 24, wherein coupling is covalent.

26. A method of detecting functional interactions between at least two molecules of interest, as recited in claim 20, wherein at least one molecule of interest comprises a protein, a protein domain, a peptide, a peptoid, a functional enzyme, a small molecule, a polymer, a non-proteinaceous cellular product, a virus particle, a cell metabolite, a lipid, a carbohydrate, a nucleic acid, an inorganic compound, a reactive compound, or any combination thereof, however, wherein not all molecules of interest are nucleic acids.

Alamar Exhibit 1002
Alamar v. Olink

27. A method of detecting functional interactions between at least two molecules of interest, as recited in claim 25, wherein at least one molecule of interest comprises at least one protein.

28. The method of detecting functional interactions between at least two molecules of interest as recited in claim 27, wherein the protein is an expression product of a nucleic acid molecule, and further wherein at least some of the plurality of interactors is formed by coupling the protein with a nucleic acid moiety comprising an identification sequence element derived from the nucleic acid molecule.

29. A method of detecting functional interactions between at least two molecules of interest, as recited in claim 26, wherein the at least two molecules of interest comprises at least one protein and at least one nucleic acid.

30. A method of detecting functional interactions between at least two molecules of interest, as recited in claim 20, wherein detecting further comprises quantifying the functional interactions.

31. A method of detecting functional interactions between at least two molecules of interest, as recited in claim 20, wherein the analysis comprises high throughput screening.

Alamar Exhibit 1002
Alamar v. Olink

32. The method of detecting functional interactions between at least two molecules of interest as recited in claim 31, wherein the high throughput screening comprises a micro array analysis.

33. The method of detecting functional interactions between at least two molecules of interest as recited in claim 20, wherein the analysis comprises nucleic acid sequencing.

34. The method of detecting functional interactions between at least two molecules of interest as recited in claim 33, wherein the analysis comprises sequencing by synthesis.

35. The method of detecting functional interactions between at least two molecules of interest as recited in claim 20, wherein association is via ligation.

36. The method of detecting functional interactions between at least two molecules of interest as recited in claim 20, wherein association is via polymerization.

37. The method of detecting functional interactions between the at least two molecules of interest as recited in claim 20, wherein the conditions for stimulating the functional interaction between the at least two molecules of interest comprises manipulating conditions to maximize cis reactivity and to minimize trans reactivity between the at least two molecules of interest.

Alamar Exhibit 1002
Alamar v. Olink

38. A library of interactors wherein each interactor comprises at least one molecule of interest coupled with at least one nucleic acid moiety comprising: a unique identification sequence element; and an association element, wherein each association element may associate with association elements from other interactors in the library to form a plurality of unique associated oligonucleotides, each oligonucleotide comprising at least two unique identification sequence elements derived from at least two different interactors, however wherein all molecules of interest are not nucleic acid.

39. The library of interactors according to claim 38 wherein the at least one molecule of interest comprises protein and the at least one nucleic acid moiety comprises DNA, wherein the DNA may be double or single stranded.

40. The library of interactors according to claim 38 wherein the at least one molecule of interest comprises a protein that is an expression product of a nucleic acid molecule, and further wherein the identification sequence element derives from the nucleic acid molecule.

41. A kit for assaying interactions between at least two molecules of interest, the kit comprising:

a. a library of interactors as recited in claim 38;

Alamar Exhibit 1002
Alamar v. Olink

b. a plurality of oligonucleotides comprising sequence elements which enable specific detection of the unique associated oligonucleotides.

42. The kit according to claim 41, wherein at least some of the plurality of oligonucleotides is immobilized on a solid phase.

43. The kit according to claim 42, wherein the solid phase comprises a microarray.

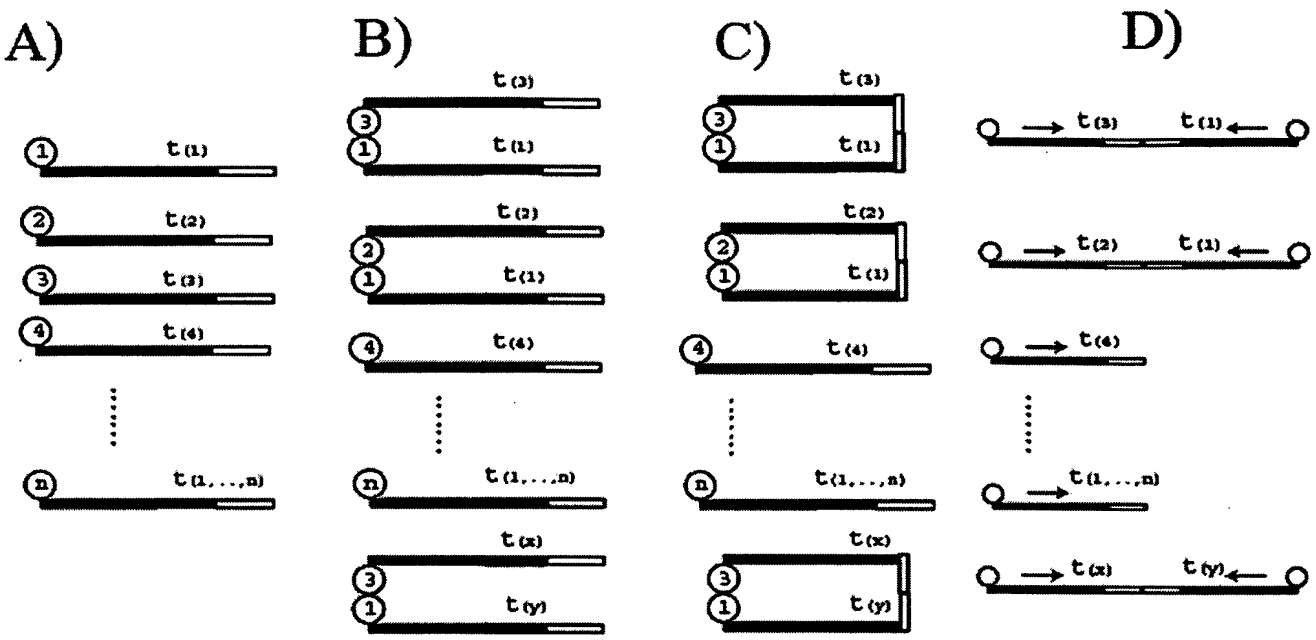
Alamar Exhibit 1002
Alamar v. Olink

Abstract

Methods of detecting affinity interactions between at least two molecules of interest are provided. The method comprises: a. forming a plurality of interactors by coupling each molecule of interest with at least one nucleic acid moiety comprising an identification sequence element and at an association element; b. promoting an association between at least two nucleic acid moieties from different interactors to form a plurality of unique associated oligonucleotides, wherein each nucleic acid moiety may form more than one unique associated oligonucleotide, and wherein each unique associated oligonucleotide comprises at least two identification sequence elements derived from the at least two nucleic acid moieties; c. selecting the plurality of unique associated oligonucleotides; and d. subjecting the selected associated oligonucleotides to an analysis that permits detection of the at least two identification sequence elements. Similar methods directed to detecting functional interactions, libraries of interactors employable in the present methods, and kits comprising those libraries are also provided.

Alamar Exhibit 1002
Alamar v. Olink

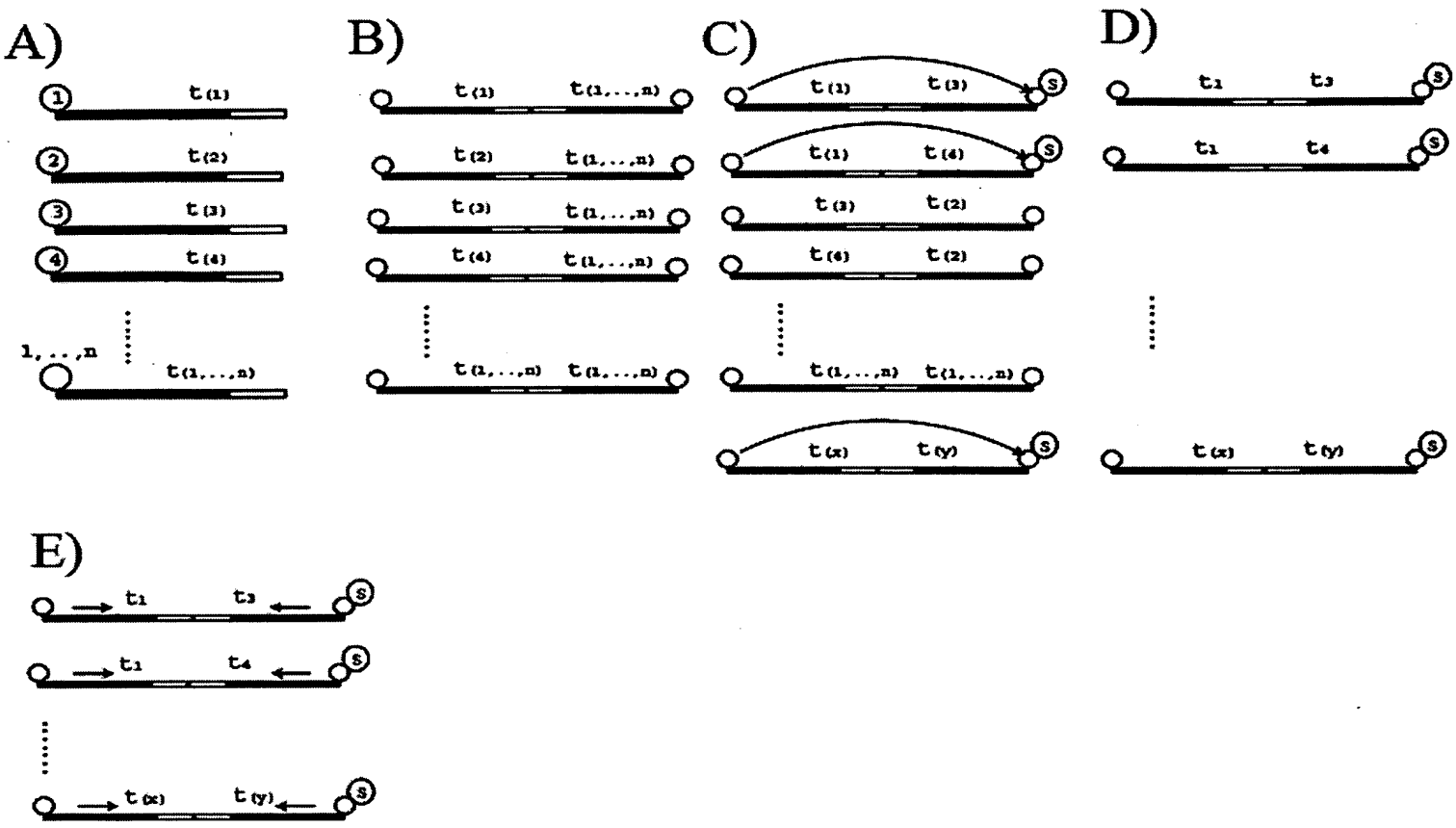
Figure 1



BEST AVAILABLE COPY

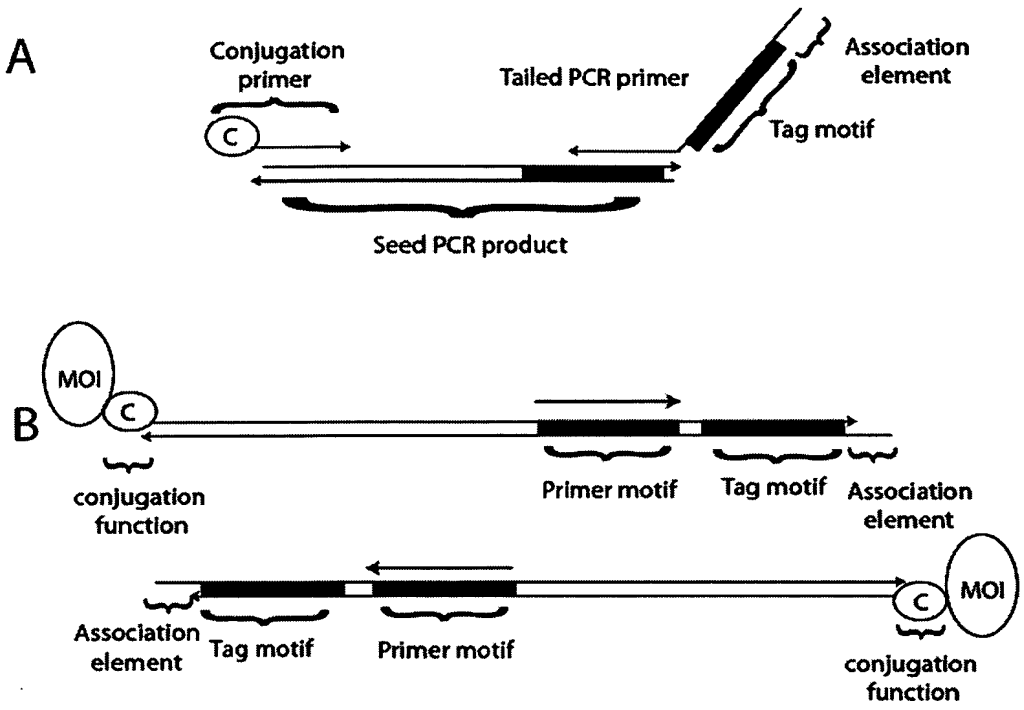
Alamar Exhibit 1002
Alamar v. Olink

Figure 2



Alamar Exhibit 1002
Alamar v. Olink

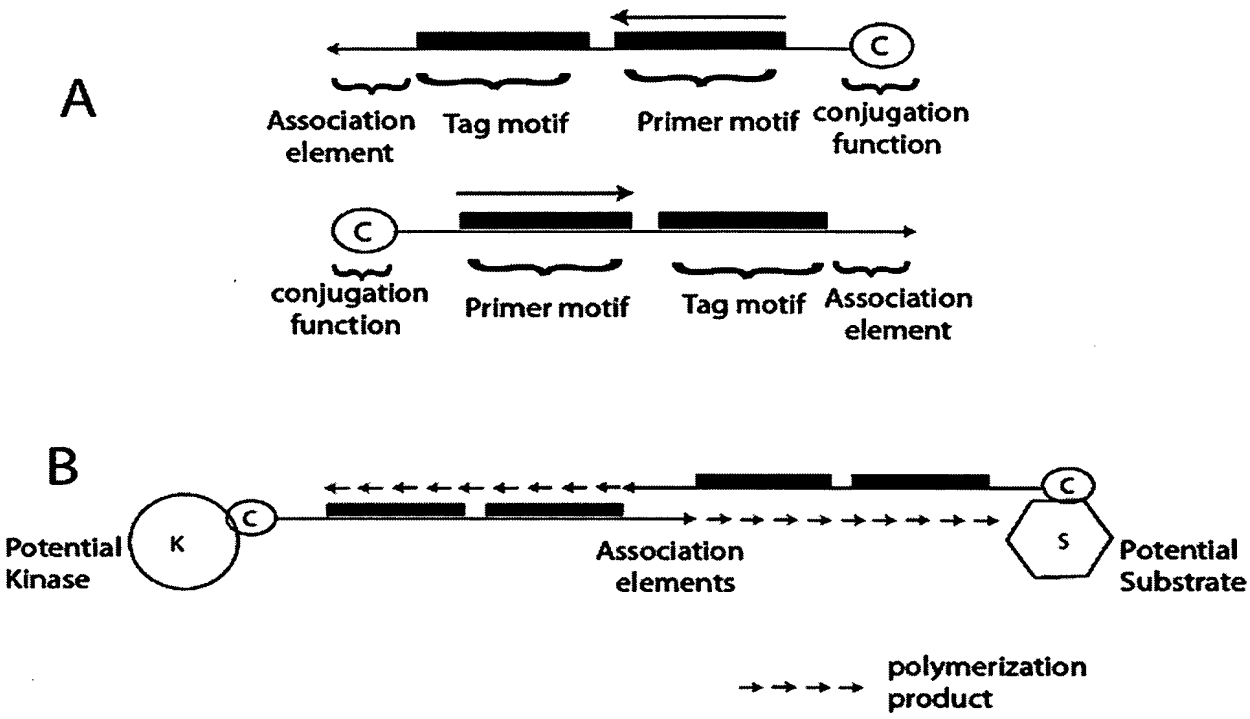
Figure 3



BEST AVAILABLE COPY

Alamar Exhibit 1002
Alamar v. Olink

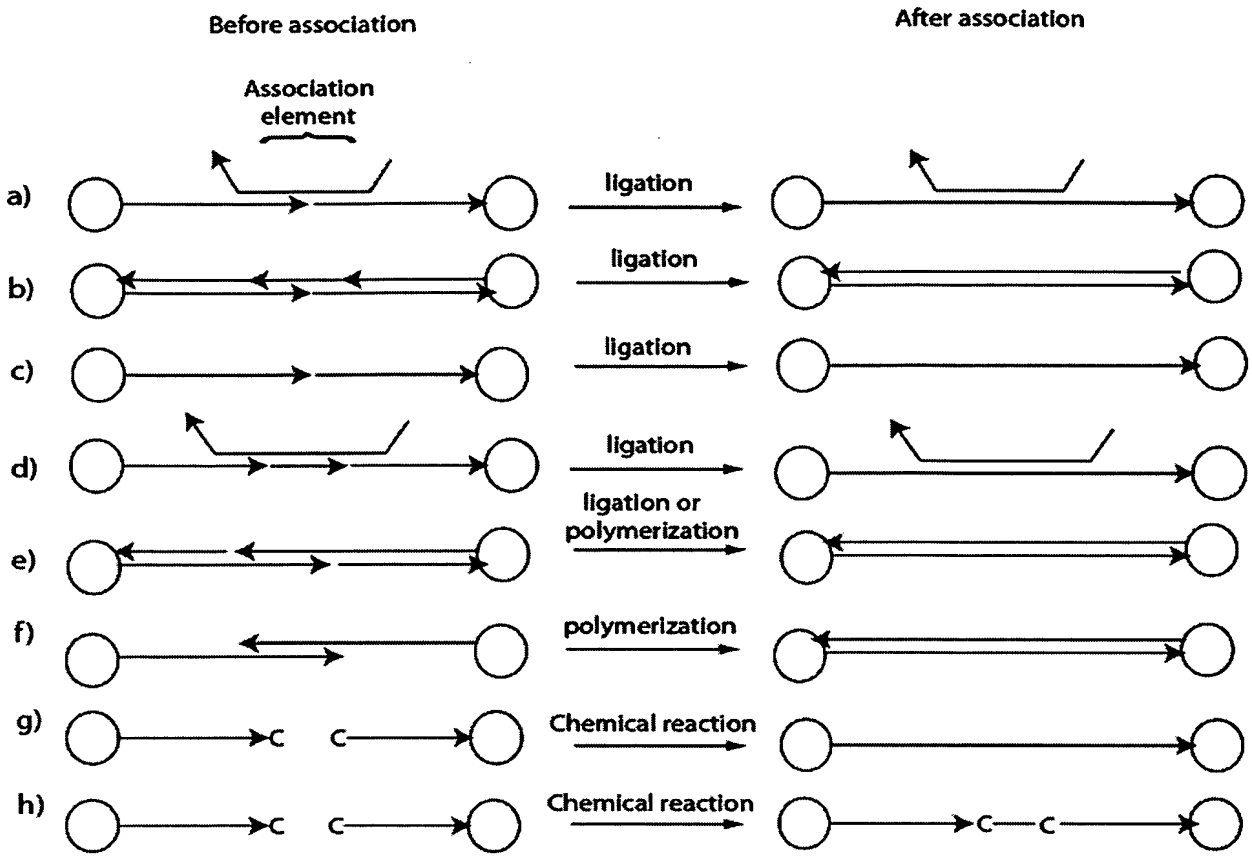
Figure 4



BEST AVAILABLE COPY

Alamar Exhibit 1002
Alamar v. Olink

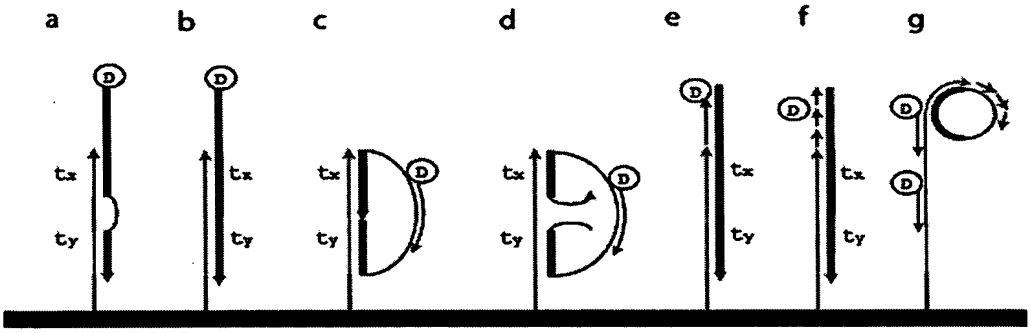
Figure 5



BEST AVAILABLE COPY

Alamar Exhibit 1002
Alamar v. Olink

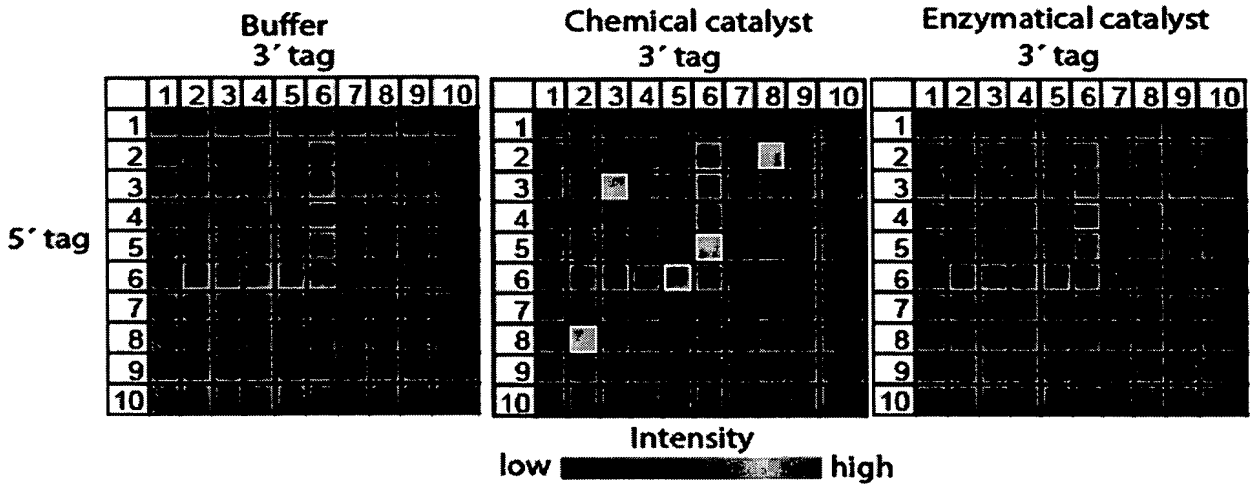
Figure 6



BEST AVAILABLE COPY

Alamar Exhibit 1002
Alamar v. Olink

Figure 7

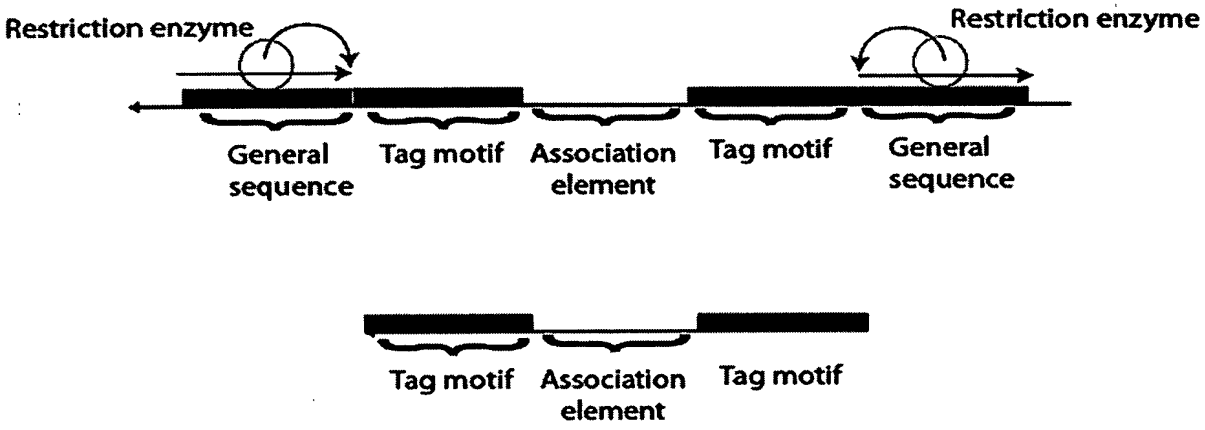


7

BEST AVAILABLE COPY

Alamar Exhibit 1002
Alamar v. Olink

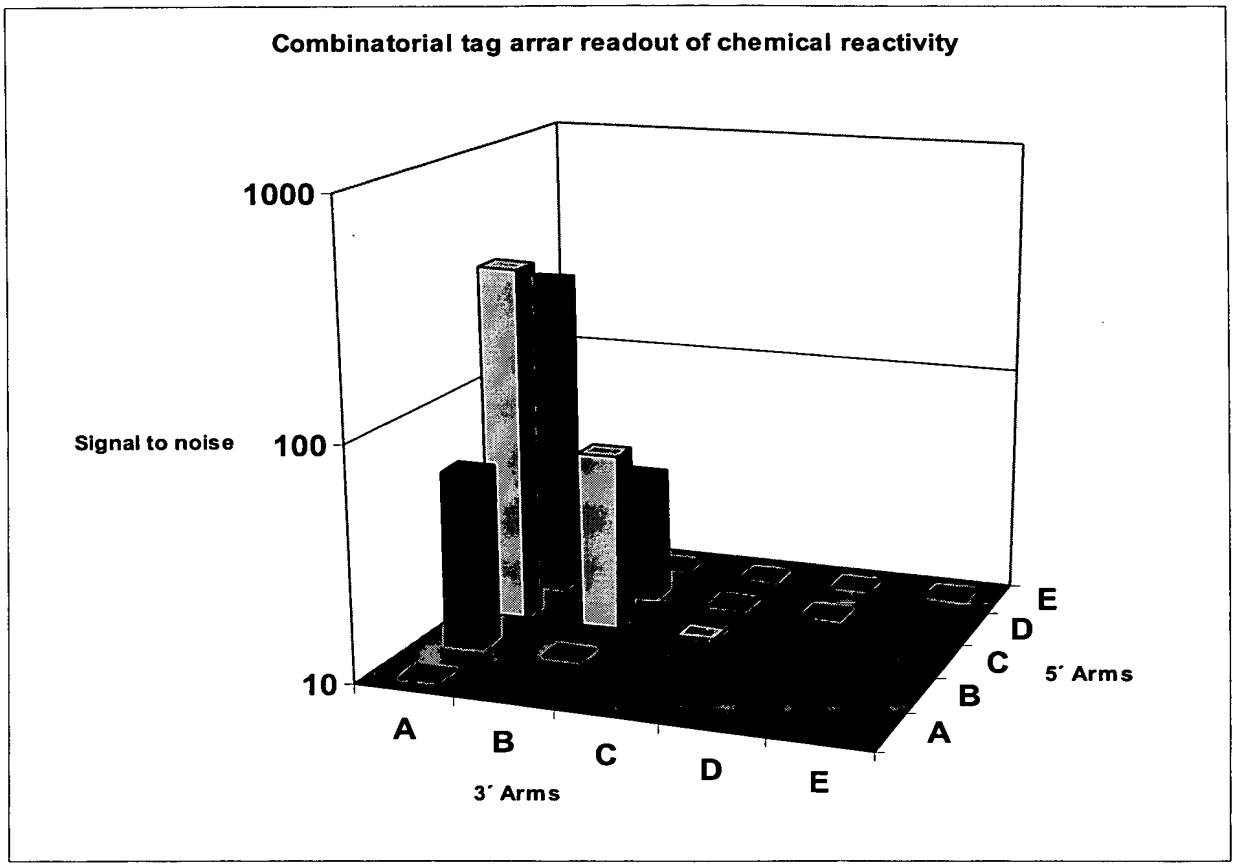
Figure 8



BEST AVAILABLE COPY

Alamar Exhibit 1002
Alamar v. Olink

Figure 9



BEST AVAILABLE COPY

DECLARATION
and
POWER OF ATTORNEY

As a below named inventor, I hereby declare that:

My residence, post office address and citizenship are as stated below next to my name.

I believe I am the original, first and sole inventor of the subject matter which is claimed and for which a patent is sought on the invention entitled "Regulation Analysis by Cis Reactivity, RACR,", the specification of which

[x] is attached hereto.
[ ] was filed on as Application Serial No. and was amended on

I hereby state that I have reviewed and understand the contents of the above identified specification, including the claims.

I acknowledge the duty to disclose information which is material to the examination of this application in accordance with Title 37, Code of Federal Regulations, §1.56(a).

I hereby claim U.S. provisional application or foreign priority benefits under Title 35, United States Code, §119 of any U.S. provisional applications or any foreign application(s) for patent or inventor's certificate listed below and have also identified below any foreign application for patent or inventor's certificate having a filing date before that of the application on which priority is claimed:

Prior U.S. Provisional or Foreign Application(s)				
			Priority Claimed	
Number	Country	Day/Month/Year Filed	Yes	No
60/697,415	U.S.	08/07/2005	x	

I hereby appoint the Dinsmore & Shohl attorneys associated with customer number **24256**, my attorneys, c/o Dinsmore & Shohl LLP, 1900 Chemed Center, 255 East Fifth Street, Cincinnati, Ohio 45202 USA (513) 977-8200, with full power in each of them, of substitution and revocation, to prosecute this application and to transact all business in the Patent and Trademark Office connected therewith.

The undersigned hereby authorizes the above-named U.S. attorneys to accept and follow instructions from Frank B. Dehn & Co. as to any action to be taken in the Patent and Trademark Office regarding this application without direct communication between the undersigned and the aforenamed U.S. attorneys. In the event of a change in the firm or persons from whom instructions may be taken, the aforenamed U.S. attorneys will be so notified in writing by the undersigned.

I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful false statements may jeopardize the validity of the application or any patent issued thereon.

****

Full name of sole or first inventor: Olof Ericsson

Inventor's signature: Olof Ericsson 5 July 2006
Date

Residence: Uppsala
Citizenship: Sweden
Post Office Address: Johannes gatan 30 C, 75233 Uppsala, Sweden

BEST AVAILABLE COPY

Alamar Exhibit 1002
Alamar v. Olink

Docket No: 22639-3

PATENT

CERTIFICATE OF EXPRESS MAILING
"Express Mail" MAILING LABEL #: EV 832369757 US
Date of Deposit: July 10, 2006
I hereby certify that this paper or fee is being deposited with the United States Postal Service addressed to Commissioner for Patents; P.O. Box 1450; Alexandria, VA 22313-1450.

Marlene Troxle

IN THE UNITED STATES PATENT & TRADEMARK OFFICE

Applicant:	Olof Ericsson :	Paper No.:	
Serial No.:	TBD :	Group Art Unit:	TBD
Filing Date:	July 10, 2006 :	Examiner:	TBD

For: **Regulation Analysis by Cis Reactivity, RACR**

STATEMENT REGARDING SEQUENCE LISTING IDENTITY

Mail Stop Sequence
Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

Dear Sir:

Submitted herewith is a substitute computer readable form (CRF) copy of the "Sequence Listing" and a substitute paper copy as required by 37 C.F.R. 1.821(e). Applicant states that the content of the substitute computer readable form copy is identical to the substitute paper copy of the Sequence Listing submitted herewith, in accordance with 37 C.F.R. 1.821(e), (f) or (g), or 1.825(b) or (d).

Respectfully submitted,

Denise M. Everett, Reg. No. 47,552
Attorney for Applicants
Dinsmore & Shohl LLP
1900 Chemed Center
255 East Fifth Street
Cincinnati, Ohio 45202

Alamar Exhibit 1002
Alamar v. Olink

SEQUENCE LISTING

<110> Ericsson, Olof

<120> Regulatioin Analysis by Cis Reactivity, RACR

<130> 26397-24

<150> US 60/697,415
<151> 2005-07-08

<160> 30

<170> PatentIn version 3.3

<210> 1
<211> 50
<212> DNA
<213> artificial

<220>
<223> Tag3A5B

<400> 1
tctctctctc tctctctctc gatgaatgtc gtcgcaacac cgtccatacg 50

<210> 2
<211> 50
<212> DNA
<213> artificial

<220>
<223> Tag3A5C

<400> 2
tctctctctc tctctctctc gatgaatgtc gtcgcgtatc acgttcgtcg 50

<210> 3
<211> 50
<212> DNA
<213> artificial

<220>
<223> Tag3A5D

<400> 3
tctctctctc tctctctctc gatgaatgtc gtcgcctccg taatcgtggt 50

<210> 4
<211> 50
<212> DNA
<213> artificial

<220>
<223> Tag3A5E

<400> 4
tctctctctc tctctctctc gatgaatgtc gtcgcaaccg tatcctcacg 50

Alamar Exhibit 1002
Alamar v. Olink

<210> 5
<211> 50
<212> DNA
<213> artificial

<220>
<223> Tag3A5F

<400> 5
tctctctctc tctctctctc gatgaatgtc gtcgcttacc ctctttcggc 50

<210> 6
<211> 50
<212> DNA
<213> artificial

<220>
<223> Tag3B5C

<400> 6
tctctctctc tctctctctc ggcgcacatg gtatagtatc acgttcgtcg 50

<210> 7
<211> 50
<212> DNA
<213> artificial

<220>
<223> Tag3B5D

<400> 7
tctctctctc tctctctctc ggcgcacatg gtatactccg taatcgtggt 50

<210> 8
<211> 50
<212> DNA
<213> artificial

<220>
<223> Tag3B5E

<400> 8
tctctctctc tctctctctc ggcgcacatg gtataaaccg tatcctcacg 50

<210> 9
<211> 50
<212> DNA
<213> artificial

<220>
<223> Tag3B5F

<400> 9
tctctctctc tctctctctc ggcgcacatg gtatattacc ctctttcggc 50

<210> 10
<211> 50
<212> DNA

Alamar Exhibit 1002
Alamar v. Olink

<213> artificial

<220>
<223> Tag3C5D

<400> 10
tctctctctc tctctctctc tgtgcgactt tctcactccg taatcgtggt 50

<210> 11
<211> 50
<212> DNA
<213> artificial

<220>
<223> Tag3C5E

<400> 11
tctctctctc tctctctctc tgtgcgactt tctcaaaccg tatcctcacg 50

<210> 12
<211> 50
<212> DNA
<213> artificial

<220>
<223> Tag3C5F

<400> 12
tctctctctc tctctctctc tgtgcgactt tctcattacc ctctttcggc 50

<210> 13
<211> 50
<212> DNA
<213> artificial

<220>
<223> Tag3D5E

<400> 13
tctctctctc tctctctctc cgacctaatg ctacgaaccg tatcctcacg 50

<210> 14
<211> 50
<212> DNA
<213> artificial

<220>
<223> Tag3D5F

<400> 14
tctctctctc tctctctctc cgacctaatg ctacgttacc ctctttcggc 50

<210> 15
<211> 50
<212> DNA
<213> artificial

<220>

Alamar Exhibit 1002
Alamar v. Olink

<223> Tag3E5F

<400> 15
tctctctctc tctctctctc gcgcctactg tgtatttacc ctctttcggc 50

<210> 16
<211> 63
<212> DNA
<213> artificial

<220>
<223> NAM-thiol-D5

<400> 16
aagatagcgt acagtcggca aaagagtcga gagctccgta atcgtggtaa aaagtatctg 60

ctt 63

<210> 17
<211> 63
<212> DNA
<213> artificial

<220>
<223> NAM-thiol-E5

<400> 17
aagatagcgt acagtcggca aaagagtcga gagaaccgta tcctcacgaa aaagtatctg 60

ctt 63

<210> 18
<211> 63
<212> DNA
<213> artificial

<220>
<223> NAM-thiol-A3

<400> 18
atgtcgcccg aaaaagatga atgtcgtcgc gagaggactc aaaacaacta ggtacaggta 60

tcg 63

<210> 19
<211> 63
<212> DNA
<213> artificial

<220>
<223> NAM-thiol-B3

<400> 19
atgtcgcccg aaaaaggcgc acatggtata gagaggactc aaaacaacta ggtacaggta 60

tcg 63

<210> 20

Alamar Exhibit 1002
Alamar v. Olink

<211> 83
<212> DNA
<213> artificial

<220>
<223> NAM B5

<400> 20
gtactggctg ataatcttcc aagatagcgt acagtcggca aaagagtcga gagaacaccg 60

tccatacgaa aaagtatctg ctt 83

<210> 21
<211> 83
<212> DNA
<213> artificial

<220>
<223> NAM C5

<400> 21
cccgagcagc cggagcagag aagatagcgt acagtcggca aaagagtcga gaggtatcac 60

gttcgtcgaa aaagtatctg ctt 83

<210> 22
<211> 83
<212> DNA
<213> artificial

<220>
<223> NAM F5

<400> 22
gtcctgtgtc tgaacctgag aagatagcgt acagtcggca aaagagtcga gagttcccct 60

ctttcggcaa aaagtatctg ctt 83

<210> 23
<211> 83
<212> DNA
<213> artificial

<220>
<223> NAM C3

<400> 23
atgtcgcccg aaaaatgtgc gactttctca gagaggactc aaaacaacta ggtacaggta 60

tcgtgtcgac gaggatgaag agg 83

<210> 24
<211> 83
<212> DNA
<213> artificial

<220>
<223> NAM D3

Alamar Exhibit 1002
Alamar v. Olink

<400> 24
atgtcgcccg aaaaacgacc taatgctacg gagaggactc aaaacaacta ggtacaggta 60

tcgctgtggc atgtttgtct aca 83

<210> 25
<211> 83
<212> DNA
<213> artificial

<220>
<223> NAM E3

<400> 25
atgtcgcccg aaaaagcgcc tactgtgtat gagaggactc aaaacaacta ggtacaggta 60

tcggtattgt gatttttaa aag 83

<210> 26
<211> 20
<212> DNA
<213> artificial

<220>
<223> FWD prime

<400> 26
tagttgtttt gagtcctctc 20

<210> 27
<211> 19
<212> DNA
<213> artificial

<220>
<223> REV prime

<400> 27
tcggcaaaag agtcgagag 19

<210> 28
<211> 19
<212> DNA
<213> artificial

<220>
<223> MlyIRE

<400> 28
gagaggactc aaacaacta 19

<210> 29
<211> 29
<212> DNA
<213> artificial

<220>
<223> LigA

<400> 29
aaaaacgggc gacataagca gatacaaaa 29

<210> 30
<211> 20
<212> DNA
<213> artificial

<220>
<223> DET 5' Cy5 modification

<400> 30
cgggcgacat aagcagatac 20

Alamar Exhibit 1002
Alamar v. Olink

PATENT APPLICATION SERIAL NO. _____

U.S. DEPARTMENT OF COMMERCE
PATENT AND TRADEMARK OFFICE
FEE RECORD SHEET

07/12/2006 HLE333 00000082 11483825

01 FC:1011	300.00	OP
02 FC:1111	500.00	OP
03 FC:1311	200.00	OP
04 FC:1202	1150.00	OP

PTO-1556
(5/87)

*U.S. Government Printing Office: 2002 — 489-267/69033

Alamar Exhibit 1002
Alamar v. Olink

PTO/SB/06 (12-04)
Approved for use through 7/31/2006. OMB 0651-0032
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

PATENT APPLICATION FEE DETERMINATION RECORD

Substitute for Form PTO-875 Effective December 8, 2004

Application or Docket Number: 11/083,825

APPLICATION AS FILED – PART I

FOR	(Column 1) NUMBER FILED	(Column 2) NUMBER EXTRA	SMALL ENTITY RATE ($)	SMALL ENTITY FEE ($)	OR	OTHER THAN SMALL ENTITY RATE ($)	OTHER THAN SMALL ENTITY FEE ($)
BASIC FEE (37 CFR 1.16(a), (b), or (c))	N/A	N/A	N/A	150.00		N/A	300.00
SEARCH FEE (37 CFR 1.16(k), (l), or (m))	N/A	N/A	N/A	$250		N/A	$500
EXAMINATION FEE (37 CFR 1.16(o), (p), or (q))	N/A	N/A	N/A	$100		N/A	$200
TOTAL CLAIMS (37 CFR 1.16(i))	43 minus 20 =	* 23	X$ 25 =		OR	X$50 =	1150
INDEPENDENT CLAIMS (37 CFR 1.16(h))	3 minus 3 =	* 0	X100 =			X200 =	
APPLICATION SIZE FEE (37 CFR 1.16(s))	If the specification and drawings exceed 100 sheets of paper, the application size fee due is $250 ($125 for small entity) for each additional 50 sheets or fraction thereof. See 35 U.S.C. 41(a)(1)(G) and 37 CFR 1.16(s).						
MULTIPLE DEPENDENT CLAIM PRESENT (37 CFR 1.16(j))			+180=			+360=	
			TOTAL			TOTAL	1150

* If the difference in column 1 is less than zero, enter "0" in column 2.

APPLICATION AS AMENDED – PART II

AMENDMENT A

	(Column 1) CLAIMS REMAINING AFTER AMENDMENT		(Column 2) HIGHEST NUMBER PREVIOUSLY PAID FOR	(Column 3) PRESENT EXTRA	SMALL ENTITY RATE ($)	SMALL ENTITY ADDITIONAL FEE ($)	OR	OTHER THAN SMALL ENTITY RATE ($)	OTHER THAN SMALL ENTITY ADDITIONAL FEE ($)
Total (37 CFR 1.16(i))	*	Minus	**	=	X$ 25 =		OR	X$50 =	
Independent (37 CFR 1.16(h))	*	Minus	***	=	X100 =		OR	X200 =	
Application Size Fee (37 CFR 1.16(s))									
FIRST PRESENTATION OF MULTIPLE DEPENDENT CLAIM (37 CFR 1.16(j))					+180=		OR	+360=	
					TOTAL ADD'L FEE		OR	TOTAL ADD'L FEE	

AMENDMENT B

	(Column 1) CLAIMS REMAINING AFTER AMENDMENT		(Column 2) HIGHEST NUMBER PREVIOUSLY PAID FOR	(Column 3) PRESENT EXTRA	RATE ($)	ADDITIONAL FEE ($)	OR	RATE ($)	ADDITIONAL FEE ($)
Total (37 CFR 1.16(i))	*	Minus	**	=	X$ 25 =		OR	X$50 =	
Independent (37 CFR 1.16(h))	*	Minus	***	=	X100 =		OR	X200 =	
Application Size Fee (37 CFR 1.16(s))									
FIRST PRESENTATION OF MULTIPLE DEPENDENT CLAIM (37 CFR 1.16(j))					+180=		OR	+360=	
					TOTAL ADD'L FEE		OR	TOTAL ADD'L FEE	

* If the entry in column 1 is less than the entry in column 2, write "0" in column 3.
** If the "Highest Number Previously Paid For" IN THIS SPACE is less than 20, enter "20".
*** If the "Highest Number Previously Paid For" IN THIS SPACE is less than 3, enter "3".
The "Highest Number Previously Paid For" (Total or Independent) is the highest number found in the appropriate box in column 1.

This collection of information is required by 37 CFR 1.16. The information is required to obtain or retain a benefit by the public which is to file (and by the USPTO to process) an application. Confidentiality is governed by 35 U.S.C. 122 and 37 CFR 1.14. This collection is estimated to take 12 minutes to complete, including gathering, preparing, and submitting the completed application form to the USPTO. Time will vary depending upon the individual case. Any comments on the amount of time you require to complete this form and/or suggestions for reducing this burden, should be sent to the Chief Information Officer, U.S. Patent and Trademark Office, U.S. Department of Commerce, P.O. Box 1450, Alexandria, VA 22313-1450. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. **SEND TO: Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450.**

If you need assistance in completing the form, call 1-800-PTO-9199 and select option 2

Alamar Exhibit 1002
Alamar v. Olink

STIC Biotechnology Systems Branch

RAW SEQUENCE LISTING ERROR REPORT

The Biotechnology Systems Branch of the Scientific and Technical Information Center (STIC) detected errors when processing the following computer readable form:

Application Serial Number: 11/483,825
Source: IFW0
Date Processed by STIC: 7/17/06

THE ATTACHED PRINTOUT EXPLAINS DETECTED ERRORS.
PLEASE FORWARD THIS INFORMATION TO THE APPLICANT BY EITHER:

1) **INCLUDING A COPY OF THIS PRINTOUT IN YOUR NEXT COMMUNICATION TO THE APPLICANT, WITH A NOTICE TO COMPLY or,**
2) **TELEPHONING APPLICANT AND FAXING A COPY OF THIS PRINTOUT, WITH A NOTICE TO COMPLY**

FOR CRF SUBMISSION AND PATENTIN SOFTWARE QUESTIONS, PLEASE CONTACT MARK SPENCER, TELEPHONE: 571-272-2510; FAX: 571-273-0221

TO **REDUCE** ERRORED SEQUENCE LISTINGS, **PLEASE** USE THE **CHECKER VERSION 4.4.0 PROGRAM**, ACCESSIBLE THROUGH THE U.S. PATENT AND TRADEMARK OFFICE WEBSITE. SEE BELOW FOR ADDRESS:

http://www.uspto.gov/web/offices/pac/checker/chkrnote.htm

Applicants submitting genetic sequence information electronically on diskette or CD-Rom should be aware that there is a possibility that the disk/CD-Rom may have been affected by treatment given to all incoming mail.
Please consider using alternate methods of submission for the disk/CD-Rom or replacement disk/CD-Rom.
Any reply including a sequence listing in electronic form should NOT be sent to the 20231 zip code address for the United States Patent and Trademark Office, and instead should be sent via the following to the indicated addresses:

1. EFS-Bio (<**http://www.uspto.gov/ebc/efs/downloads/documents.htm**> , **EFS Submission User Manual** - ePAVE)
2. **U.S. Postal Service:** Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450
3. **Hand Carry, Federal Express, United Parcel Service, or other delivery service (EFFECTIVE 01/14/05):** U.S. Patent and Trademark Office, Mail Stop Sequence, Customer Window, Randolph Building, 401 Dulany Street, Alexandria, VA 22314

Revised 01/10/06

Alamar Exhibit 1002
Alamar v. Olink

Raw Sequence Listing Error Summary

ERROR DETECTED	SUGGESTED CORRECTION	SERIAL NUMBER: 11/483,825

ATTN: NEW RULES CASES: PLEASE DISREGARD ENGLISH "ALPHA" HEADERS, WHICH WERE INSERTED BY PTO SOFTWARE

1	Wrapped Nucleics Wrapped Aminos	The number/text at the end of each line "wrapped" down to the next line. This may occur if your file was retrieved in a word processor **after** creating it. Please adjust your right margin to .3; this will prevent "wrapping."
2	Invalid Line Length	The rules require that a line **not exceed** 72 characters in length. This includes white spaces.
3	Misaligned Amino Numbering	The numbering under each 5th amino acid is misaligned. Do **not** use tab codes between numbers; use **space characters**, instead.
4	Non-ASCII	The submitted file was **not** saved in ASCII(DOS) text, as **required** by the Sequence Rules. **Please ensure your subsequent submission is saved in ASCII text.**
5	Variable Length	Sequence(s)____contain n's or Xaa's representing more than one residue. **Per Sequence Rules, each n or Xaa can only represent a single residue.** Please present the **maximum** number of each residue having variable length and indicate in the <220>-<223> section that some may be missing.
6	PatentIn 2.0 "bug"	A "bug" in PatentIn version 2.0 has caused the <220>-<223> section to be missing from amino acid sequences(s) _____. Normally, PatentIn would automatically generate this section from the previously coded nucleic acid sequence. Please manually copy the relevant <220>-<223> section to the subsequent amino acid sequence. **This applies to the mandatory <220>-<223> sections for Artificial or Unknown sequences.**
7	Skipped Sequences (OLD RULES)	Sequence(s) _____ missing. If intentional, please insert the following lines for **each** skipped sequence: (2) INFORMATION FOR SEQ ID NO:X: (insert SEQ ID NO where "X" is shown) (i) SEQUENCE CHARACTERISTICS: (Do not insert any subheadings under this heading) (xi) SEQUENCE DESCRIPTION:SEQ ID NO:X: (insert SEQ ID NO where "X" is shown) This sequence is intentionally skipped Please also adjust the "(ii) NUMBER OF SEQUENCES:" response to **include** the skipped sequences.
8	Skipped Sequences (NEW RULES)	Sequence(s) _____ missing. If **intentional**, please insert the following lines for **each** skipped sequence. <210> sequence id number <400> sequence id number 000
9	Use of n's or Xaa's (NEW RULES)	Use of n's and/or Xaa's have been detected in the Sequence Listing. Per 1.823 of Sequence Rules, use of <220>-<223> is MANDATORY if n's or Xaa's are present. In <220> to <223> section, please explain location of **n** or **Xaa**, and which residue **n** or **Xaa** represents.
10	Invalid <213> Response	Per 1.823 of Sequence Rules, the only **valid** <213> responses are: Unknown, Artificial Sequence, or scientific name (Genus/species). <220>-<223> section is **required** when <213> response is Unknown or is Artificial Sequence. (see item 11 below)
11	Use of <220>	Sequence(s) ___ missing the <220> "Feature" and associated numeric identifiers and responses. Use of <220> to <223> is MANDATORY if <213> "Organism" response is "Artificial Sequence" or "Unknown." Please explain source of genetic material in <220> to <223> section or use "chemically synthesized" as explanation. (See "Federal Register," 06/01/1998, Vol. 63, No. 104, pp. 29631-32), also Sec. 1.823 of Sequence Rules
12	PatentIn 2.0 "bug"	Please do not use "Copy to Disk" function of PatentIn version 2.0. This causes a corrupted file, resulting in missing mandatory numeric identifiers and responses (as indicated on raw sequence listing). Instead, please use "File Manager" or any other manual means to copy file to floppy disk.
13	Misuse of n/Xaa	"n" can **only** represent a single nucleotide; "Xaa" can **only** represent a single amino acid

AMC – STIC Systems Branch – 03/02/06

Alamar Exhibit 1002
Alamar v. Olink

IFWO

RAW SEQUENCE LISTING DATE: 07/17/2006
PATENT APPLICATION: **US/11/483,825** TIME: 08:44:14

Input Set : **A:\1284340_1.txt**
Output Set: **N:\CRF4\07172006\K483825.raw**

```
     3 <110> APPLICANT: Ericsson, Olof
     5 <120> TITLE OF INVENTION: Regulatioin Analysis by Cis Reactivity, RACR
     7 <130> FILE REFERENCE: 26397-24
C--> 9 <140> CURRENT APPLICATION NUMBER: US/11/483,825
C--> 9 <141> CURRENT FILING DATE: 2006-07-10
     9 <150> PRIOR APPLICATION NUMBER: US 60/697,415
    10 <151> PRIOR FILING DATE: 2005-07-08
    12 <160> NUMBER OF SEQ ID NOS: 30
    14 <170> SOFTWARE: PatentIn version 3.3
    16 <210> SEQ ID NO: 1
    17 <211> LENGTH: 50
    18 <212> TYPE: DNA
    19 <213> ORGANISM: artificial
    21 <220> FEATURE:
    22 <223> OTHER INFORMATION: Tag3A5B
    24 <400> SEQUENCE: 1
    25 tctctctctc tctctctctc gatgaatgtc gtcgcaacac cgtccatacg          50
    28 <210> SEQ ID NO: 2
    29 <211> LENGTH: 50
    30 <212> TYPE: DNA
    31 <213> ORGANISM: artificial
    33 <220> FEATURE:
    34 <223> OTHER INFORMATION: Tag3A5C
    36 <400> SEQUENCE: 2
    37 tctctctctc tctctctctc gatgaatgtc gtcgcgtatc acgttcgtcg          50
    40 <210> SEQ ID NO: 3
    41 <211> LENGTH: 50
    42 <212> TYPE: DNA
    43 <213> ORGANISM: artificial
    45 <220> FEATURE:
    46 <223> OTHER INFORMATION: Tag3A5D
    48 <400> SEQUENCE: 3
    49 tctctctctc tctctctctc gatgaatgtc gtcgcctccg taatcgtggt          50
    52 <210> SEQ ID NO: 4
    53 <211> LENGTH: 50
    54 <212> TYPE: DNA
    55 <213> ORGANISM: artificial
    57 <220> FEATURE:
    58 <223> OTHER INFORMATION: Tag3A5E
    60 <400> SEQUENCE: 4
    61 tctctctctc tctctctctc gatgaatgtc gtcgcaaccg tatcctcacg          50
    64 <210> SEQ ID NO: 5
    65 <211> LENGTH: 50
```

pp 1-5

Does Not Comply
Corrected Diskette Needed

what is (are) source(s) of genetic material? (see item 11 on Error Summary Sheet)

Alamar Exhibit 1002
Alamar v. Olink

RAW SEQUENCE LISTING DATE: 07/17/2006
PATENT APPLICATION: US/11/483,825 TIME: 08:44:14

Input Set : A:\1284340_1.txt
Output Set: N:\CRF4\07172006\K483825.raw

<212> TYPE: DNA
<213> ORGANISM: artificial
<220> FEATURE:
<223> OTHER INFORMATION: Tag3A5F
<400> SEQUENCE: 5
tctctctctc tctctctctc gatgaatgtc gtcgcttacc ctctttcggc 50
<210> SEQ ID NO: 6
<211> LENGTH: 50
<212> TYPE: DNA
<213> ORGANISM: artificial
<220> FEATURE:
<223> OTHER INFORMATION: Tag3B5C
<400> SEQUENCE: 6
tctctctctc tctctctctc ggcgcacatg gtatagtatc acgttcgtcg 50
<210> SEQ ID NO: 7
<211> LENGTH: 50
<212> TYPE: DNA
<213> ORGANISM: artificial
<220> FEATURE:
<223> OTHER INFORMATION: Tag3B5D
<400> SEQUENCE: 7
tctctctctc tctctctctc ggcgcacatg gtatactccg taatcgtggt 50
<210> SEQ ID NO: 8
<211> LENGTH: 50
<212> TYPE: DNA
<213> ORGANISM: artificial
<220> FEATURE:
<223> OTHER INFORMATION: Tag3B5E
<400> SEQUENCE: 8
tctctctctc tctctctctc ggcgcacatg gtataaaccg tatcctcacg 50
<210> SEQ ID NO: 9
<211> LENGTH: 50
<212> TYPE: DNA
<213> ORGANISM: artificial
<220> FEATURE:
<223> OTHER INFORMATION: Tag3B5F
<400> SEQUENCE: 9
tctctctctc tctctctctc ggcgcacatg gtatattacc ctctttcggc 50
<210> SEQ ID NO: 10
<211> LENGTH: 50
<212> TYPE: DNA
<213> ORGANISM: artificial
<220> FEATURE:
<223> OTHER INFORMATION: Tag3C5D
<400> SEQUENCE: 10
tctctctctc tctctctctc tgtgcgactt tctcactccg taatcgtggt 50
<210> SEQ ID NO: 11
<211> LENGTH: 50
<212> TYPE: DNA

Alamar Exhibit 1002
Alamar v. Olink

RAW SEQUENCE LISTING DATE: 07/17/2006
PATENT APPLICATION: US/11/483,825 TIME: 08:44:14

Input Set : A:\1284340_1.txt
Output Set: N:\CRF4\07172006\K483825.raw

```
139 <213> ORGANISM: artificial
141 <220> FEATURE:
142 <223> OTHER INFORMATION: Tag3C5E
144 <400> SEQUENCE: 11
145 tctctctctc tctctctctc tgtgcgactt tctcaaaccg tatcctcacg          50
148 <210> SEQ ID NO: 12
149 <211> LENGTH: 50
150 <212> TYPE: DNA
151 <213> ORGANISM: artificial
153 <220> FEATURE:
154 <223> OTHER INFORMATION: Tag3C5F
156 <400> SEQUENCE: 12
157 tctctctctc tctctctctc tgtgcgactt tctcattacc ctctttcggc          50
160 <210> SEQ ID NO: 13
161 <211> LENGTH: 50
162 <212> TYPE: DNA
163 <213> ORGANISM: artificial
165 <220> FEATURE:
166 <223> OTHER INFORMATION: Tag3D5E
168 <400> SEQUENCE: 13
169 tctctctctc tctctctctc cgacctaatg ctacgaaccg tatcctcacg          50
172 <210> SEQ ID NO: 14
173 <211> LENGTH: 50
174 <212> TYPE: DNA
175 <213> ORGANISM: artificial
177 <220> FEATURE:
178 <223> OTHER INFORMATION: Tag3D5F
180 <400> SEQUENCE: 14
181 tctctctctc tctctctctc cgacctaatg ctacgttacc ctctttcggc          50
184 <210> SEQ ID NO: 15
185 <211> LENGTH: 50
186 <212> TYPE: DNA
187 <213> ORGANISM: artificial
189 <220> FEATURE:
190 <223> OTHER INFORMATION: Tag3E5F
192 <400> SEQUENCE: 15
193 tctctctctc tctctctctc gcgcctactg tgtatttacc ctctttcggc          50
196 <210> SEQ ID NO: 16
197 <211> LENGTH: 63
198 <212> TYPE: DNA
199 <213> ORGANISM: artificial
201 <220> FEATURE:
202 <223> OTHER INFORMATION: NAM-thiol-D5
204 <400> SEQUENCE: 16
205 aagatagcgt acagtcggca aaagagtcga gagctccgta atcgtggtaa aaagtatctg  60
207 ctt                                                              63
210 <210> SEQ ID NO: 17
211 <211> LENGTH: 63
212 <212> TYPE: DNA
```


Alamar Exhibit 1002
Alamar v. Olink

RAW SEQUENCE LISTING DATE: 07/17/2006
PATENT APPLICATION: US/11/483,825 TIME: 08:44:14

Input Set : A:\1284340_1.txt
Output Set: N:\CRF4\07172006\K483825.raw

213 <213> ORGANISM: artificial
215 <220> FEATURE:
216 <223> OTHER INFORMATION: NAM-thiol-E5
218 <400> SEQUENCE: 17
219 aagatagcgt acagtcggca aaagagtcga gagaaccgta tcctcacgaa aaagtatctg 60
221 ctt 63
224 <210> SEQ ID NO: 18
225 <211> LENGTH: 63
226 <212> TYPE: DNA
227 <213> ORGANISM: artificial
229 <220> FEATURE:
230 <223> OTHER INFORMATION: NAM-thiol-A3
232 <400> SEQUENCE: 18
233 atgtcgcccg aaaaagatga atgtcgtcgc gagaggactc aaaacaacta ggtacaggta 60
235 tcg 63
238 <210> SEQ ID NO: 19
239 <211> LENGTH: 63
240 <212> TYPE: DNA
241 <213> ORGANISM: artificial
243 <220> FEATURE:
244 <223> OTHER INFORMATION: NAM-thiol-B3
246 <400> SEQUENCE: 19
247 atgtcgcccg aaaaaggcgc acatggtata gagaggactc aaaacaacta ggtacaggta 60
249 tcg 63
252 <210> SEQ ID NO: 20
253 <211> LENGTH: 83
254 <212> TYPE: DNA
255 <213> ORGANISM: artificial
257 <220> FEATURE:
258 <223> OTHER INFORMATION: NAM B5
260 <400> SEQUENCE: 20
261 gtactggctg ataatcttcc aagatagcgt acagtcggca aaagagtcga gagaacaccg 60
263 tccatacgaa aaagtatctg ctt 83
266 <210> SEQ ID NO: 21
267 <211> LENGTH: 83
268 <212> TYPE: DNA
269 <213> ORGANISM: artificial
271 <220> FEATURE:
272 <223> OTHER INFORMATION: NAM C5
274 <400> SEQUENCE: 21
275 cccgagcagc cggagcagag aagatagcgt acagtcggca aaagagtcga gaggtatcac 60
277 gttcgtcgaa aaagtatctg ctt 83
280 <210> SEQ ID NO: 22
281 <211> LENGTH: 83
282 <212> TYPE: DNA
283 <213> ORGANISM: artificial
285 <220> FEATURE:
286 <223> OTHER INFORMATION: NAM F5
288 <400> SEQUENCE: 22

Alamar Exhibit 1002
Alamar v. Olink

RAW SEQUENCE LISTING DATE: 07/17/2006
PATENT APPLICATION: US/11/483,825 TIME: 08:44:14

Input Set : A:\1284340_1.txt
Output Set: N:\CRF4\07172006\K483825.raw

289 gtcctgtgtc tgaacctgag aagatagcgt acagtcggca aaagagtcga gagttcccct 60
291 ctttcggcaa aaagtatctg ctt 83
294 <210> SEQ ID NO: 23
295 <211> LENGTH: 83
296 <212> TYPE: DNA
297 <213> ORGANISM: artificial
299 <220> FEATURE:
300 <223> OTHER INFORMATION: NAM C3
302 <400> SEQUENCE: 23
303 atgtcgcccg aaaaatgtgc gactttctca gagaggactc aaaacaacta ggtacaggta 60
305 tcgtgtcgac gaggatgaag agg 83
308 <210> SEQ ID NO: 24
309 <211> LENGTH: 83
310 <212> TYPE: DNA
311 <213> ORGANISM: artificial
313 <220> FEATURE:
314 <223> OTHER INFORMATION: NAM D3
316 <400> SEQUENCE: 24
317 atgtcgcccg aaaaacgacc taatgctacg gagaggactc aaaacaacta ggtacaggta 60
319 tcgctgtggc atgtttgtct aca 83
322 <210> SEQ ID NO: 25
323 <211> LENGTH: 83
324 <212> TYPE: DNA
325 <213> ORGANISM: artificial
327 <220> FEATURE:
328 <223> OTHER INFORMATION: NAM E3
330 <400> SEQUENCE: 25
331 atgtcgcccg aaaaagcgcc tactgtgtat gagaggactc aaaacaacta ggtacaggta 60
333 tcggtattgt gatttttaa aag 83
336 <210> SEQ ID NO: 26
337 <211> LENGTH: 20
338 <212> TYPE: DNA
339 <213> ORGANISM: artificial
341 <220> FEATURE:
342 <223> OTHER INFORMATION: FWD prime
344 <400> SEQUENCE: 26
345 tagttgtttt gagtcctctc 20
348 <210> SEQ ID NO: 27
349 <211> LENGTH: 19
350 <212> TYPE: DNA
351 <213> ORGANISM: artificial
353 <220> FEATURE:
354 <223> OTHER INFORMATION: REV prime
356 <400> SEQUENCE: 27
357 tcggcaaaag agtcgagag 19
360 <210> SEQ ID NO: 28
361 <211> LENGTH: 19
362 <212> TYPE: DNA
363 <213> ORGANISM: artificial

Please correct these types of error in subsequent sequences

Alamar Exhibit 1002
Alamar v. Olink

VERIFICATION SUMMARY DATE: 07/17/2006
PATENT APPLICATION: **US/11/483,825** TIME: 08:44:15

Input Set : **A:\1284340_1.txt**
Output Set: **N:\CRF4\07172006\K483825.raw**

L:9 M:270 C: Current Application Number differs, Replaced Current Application No
L:9 M:271 C: Current Filing Date differs, Replaced Current Filing Date

Alamar Exhibit 1002
Alamar v. Olink

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NUMBER	FILING OR 371 (c) DATE	FIRST NAMED APPLICANT	ATTORNEY DOCKET NUMBER
11/483,825	07/10/2006	Olof Ericsson	22639-3

CONFIRMATION NO. 8659

FORMALITIES LETTER

24256
DINSMORE & SHOHL, LLP
1900 CHEMED CENTER
255 EAST FIFTH STREET
CINCINNATI, OH 45202

Date Mailed: 07/28/2006

NOTICE TO COMPLY WITH REQUIREMENTS FOR PATENT APPLICATIONS CONTAINING NUCLEOTIDE SEQUENCE AND/OR AMINO ACID SEQUENCE DISCLOSURES

Filing Date Granted

Applicant is given **TWO MONTHS FROM THE DATE OF THIS NOTICE** within which to file the items indicated below to avoid abandonment. Extensions of time may be obtained under the provisions of 37 CFR 1.136(a).

- A copy of the "Sequence Listing" in computer readable form has been submitted. However, the content of the computer readable form does not comply with the requirements of 37 CFR 1.822 and/or 1.823, as indicated on the attached copy of the marked -up "Raw Sequence Listing." Applicant must provide a substitute computer readable form (CRF) copy of the "Sequence Listing" and a statement that the content of the sequence listing information recorded in computer readable form is identical to the written (on paper or compact disc) sequence listing and, where applicable, includes no new matter, as required by 37 CFR 1.821(e), 1.821(f), 1.821(g), 1.825(b), or 1.825(d).

  Applicant is cautioned that correction of the above items may cause the specification and drawings page count to exceed 100 pages. If the specification and drawings exceed 100 pages, applicant will need to submit the required application size fee.

To Download Patentin Software, visit http://www.uspto.gov/web/patents/software.htm
For questions regarding compliance to these requirements, please contact:

- **For Rules Interpretation, call (571) 272-0951**
- **For Patentin Software Program Help, call Patent EBC at 1-866-217-9197 or directly at 703-305-3028 / 703-308-6845 between the hours of 6 a.m. and 12 midnight, Monday through Friday, EST.**
- **Send e-mail correspondence for Patentin Software Program Help @ ebc@uspto.gov**

Replies should be mailed to: Mail Stop Missing Parts
Commissioner for Patents
P.O. Box 1450
Alexandria VA 22313-1450

Alamar Exhibit 1002
Alamar v. Olink

A copy of this notice MUST be returned with the reply.

Office of Initial Patent Examination (571) 272-4000, or 1-800-PTO-9199, or 1-800-972-6382
PART 3 - OFFICE COPY

Alamar Exhibit 1002
Alamar v. Olink

OIPE
SEP 25 2006
PATENT & TRADEMARK OFFICE

Docket No: 22639-3

PATENT

CERTIFICATE OF MAILING

I hereby certify that this paper is being deposited with the United States Postal Service with sufficient postage as first class mail in an envelope addressed to: Mail Stop Amendment; Commissioner for Patents, P.O. Box 1450; Alexandria, VA 22313-1450 on September 21, 2006.

Marlene Troxle

IN THE UNITED STATES PATENT & TRADEMARK OFFICE

Applicant:	Olof Ericsson		
Serial No.:	11/483,825	Group Art Unit:	TBD
Filed:	July 10, 2006	Examiner:	TBD
For:	Regulation Analysis by Cis Reactivity, RACR		

INFORMATION DISCLOSURE STATEMENT

Mail Stop Amendment
Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

Dear Sir:

In accordance with 37 C.F.R. §§ 1.97 and 1.98 Applicant herewith submits certain patent and non-patent references which the Patent & Trademark Office may wish to consider in examining the above-identified application. The references are listed on the attached Form PTO-1449.

Copies of the listed references are also provided. No representation is made or intended that a prior art search has been made or that no better art than that listed is available.

As this Statement is being submitted prior to issuance of the first office action on the merits, it is believed that no fee is required. Nonetheless, in the event this is mistaken or an office action has been mailed in crossing, please charge any fees required, or credit any overpayment, to Deposit Account No. 04-1133.

Alamar Exhibit 1002
Alamar v. Olink

Respectfully submitted,

By: _____

Denise M. Everett
Registration No. 47,552
Attorney for Applicant(s)
DINSMORE & SHOHL LLP
1900 Chemed Center
255 East Fifth Street
Cincinnati, Ohio 45202
(513) 977-8787

Alamar Exhibit 1002
Alamar v. Olink

SEP 25 2006
PATENT & TRADEMARK OFFICE

PTO FORM 4/92												Page 1 of 1
FORM PTO - 1449 LIST OF PATENTS AND PUBLICATIONS FOR APPLICANT'S INFORMATION DISCLOSURE STATEMENT									ATTY DOCKET: 22639-3 APPLICANT: Olof Ericsson FILING DATE: July 10, 2006 FOR:		SERIAL NO.: 11/483,825 GROUP: TBD	
UNITED STATES PUBLISHED PATENT DOCUMENTS												
Exr. Init.		APPL. DOCUMENT NUMBER							PUBL. DATE	NAME	CLASS	SUB CLASS
		02	0	5	1	9	8	6	May 2, 2002	Baez et a.		
		03	0	8	2	8	3	0	May 1, 2003	Schrieber et al.		
FOREIGN PATENT DOCUMENTS												
		DOCUMENT NUMBER							DATE	COUNTRY	CLASS	SUB CLASS

OTHER ART (INCLUDING AUTHOR, TITLE DATE, PERTINENT PAGES, ETC.)	
	Burbulis, I, et al. "Using protein-DNA Chimeras to detect and count small numbers of molecules. Nat Methods. 2005 Jan; 2(1):31-37 e-pub: 2004 Dec. 21
	Duffy, S. et al. "site-specific enzymatic biotinylation of recombinant proteins in Spodoptera frugiperda cells using biotin acceptor peptides. Anal Biochem. 1998 Sep. 10:262(2):122-28
	Fan JB, et al. "A versatile assay for high-throughput gene expression profiling on universal array matrices" Genome Res. 2004 May;14(5):878-85.
	Fredriksson S, Gullberg M, Jarvius J, Olsson C, Pietras K, Gustafsdottir SM, Ostman A, Landegren U. Protein detection using proximity-dependent DNA ligation assays. Nat Biotechnol. 2002 May;20(5):473-7.
	Gullberg M, Gustafsdottir SM, Schallmeiner E, Jarvius J, Bjarnegard M, Betsholtz C, Landegren U, Fredriksson S. Cytokine detection by antibody-based proximity ligation. Proc Natl Acad Sci U S A. 2004 Jun 1;101(22):8420-4. Epub 2004 May 21.
	Gunderson KL, Steemers FJ, Lee G, Mendoza LG, Chee MS. A genome-wide scalable SNP genotyping assay using microarray technology. Nat Genet. 2005 May;37(5):549-54. Epub 2005 Apr 17.
	Gunderson KL, Kruglyak S, Graige MS, Garcia F, Kermani BG, Zhao C, Che D, Dickinson T, Wickham E, Bierle J, Doucet D, Milewski M, Yang R, Siegmund C, Haas J, Zhou L, Oliphant A, Fan JB, Barnard S, Chee MS. Decoding randomly ordered DNA arrays. Genome Res. 2004 May;14(5):870-7. Epub 2004 Apr 12.
	Hanes J, Pluckthun A. In vitro selection and evolution of functional proteins by using ribosome display. Proc Natl Acad Sci U S A. 1997 May 13;94(10):4937-42.
	Hardenbol P, Baner J, Jain M, Nilsson M, Namsaraev EA, Karlin-Neumann GA, Fakhrai-Rad H, Ronaghi M, Willis TD, Landegren U, Davis RW. Multiplexed genotyping with sequence-tagged molecular inversion probes. Nat Biotechnol. 2003 Jun;21(6):673-8. Epub 2003 May 5.
	Hatch A, Sano T, Misasi J, Smith CL. Rolling circle amplification of DNA immobilized on solid surfaces and its application to multiplex mutation detection. Genet Anal. 1999 Apr;15(2):35-40.
	Kanan MW, Rozenman MM, Sakurai K, Snyder TM, Liu DR. Reaction discovery enabled by DNA-templated synthesis and in vitro selection. Nature. 2004 Sep 30;431(7008):545-9.
	Keppler A, Kindermann M, Gendreizig S, Pick H, Vogel H, Johnsson K. Labeling of fusion proteins of O6-alkylguanine-DNA alkyltransferase with small molecules in vivo and in vitro. Methods. 2004 Apr;32(4):437-44.

		Landegren U, Kaiser R, Sanders J, Hood L. A ligase-mediated gene detection technique. Science. 1988 Aug 26;241(4869):1077-80.
		Lizardi PM, Huang X, Zhu Z, Bray-Ward P, Thomas DC, Ward DC. Mutation detection and single-molecule counting using isothermal rolling-circle amplification. Nat Genet. 1998 Jul;19(3):225-32.
		Lyamichev V, Brow MA, Dahlberg JE. Structure-specific endonucleolytic cleavage of nucleic acids by eubacterial DNA polymerases. Science. 1993 May 7;260(5109):778-83.
		Odegrip R, Coomber D, Eldridge B, Hederer R, Kuhlman PA, Ullman C, FitzGerald K, McGregor D. CIS display: In vitro selection of peptides from libraries of protein-DNA complexes. Proc Natl Acad Sci U S A. 2004 Mar 2;101(9):2806-10. Epub 2004 Feb 23.
		Reiersen H, Lobersli I, Loset GA, Hvattum E, Simonsen B, Stacy JE, McGregor D, Fitzgerald K, Welschof M, Brekke OH, Marvik OJ. Covalent antibody display--an in vitro antibody-DNA library selection system. Nucleic Acids Res. 2005 Jan 14;33(1):e10.
		Southworth MW, Amaya K, Evans TC, Xu MQ, Perler FB. Purification of proteins fused to either the amino or carboxy terminus of the Mycobacterium xenopi gyrase A intein. Biotechniques. 1999 Jul;27(1):110-4, 116, 118-20.
		Takeda S, Tsukiji S, Nagamune T. Site-specific conjugation of oligonucleotides to the C-terminus of recombinant protein by expressed protein ligation. Bioorg Med Chem Lett. 2004 May 17;14(10):2407-10.
		Tan DS, Foley MA, Shair MD, Schreiber SL. Stereoselective synthesis of over two million compounds having structural features both reminiscent of natural products and compatible with miniaturized cell-based assays.
		Wilson DS, Keefe AD, Szostak JW. The use of mRNA display to select high-affinity protein-binding peptides. Proc Natl Acad Sci U S A. 2001 Mar 27;98(7):3750-5. Epub 2001 Mar 13.
		Yeakley JM, Fan JB, Doucet D, Luo L, Wickham E, Ye Z, Chee MS, Fu XD. Profiling alternative splicing on fiber-optic arrays. Nat Biotechnol. 2002 Apr;20(4):353-8.

EXAMINER DATE CONSIDERED

EXAMINER: Initial if reference considered, whether or not citation is in conformance with MPEP 609; Draw line through citation if not in conformance and not considered. Include copy of this form with next communication to applicant.

Profiling alternative splicing on fiber-optic arrays

Joanne M. Yeakley[1,2], Jian-Bing Fan[3]*, Dennis Doucet[3], Lin Luo[3,4], Eliza Wickham[3], Zhen Ye[1], Mark S. Chee[3], and Xiang-Dong Fu[1]*

© 2002 Nature Publishing Group http://biotech.nature.com

The human transcriptome is marked by extensive alternative mRNA splicing and the expression of many closely related genes, which may be difficult to distinguish using standard microarray techniques. Here we describe a sensitive and specific assay for parallel analysis of mRNA isoforms on a fiber-optic microarray platform. The method permits analysis of mRNA transcripts without prior RNA purification or cDNA synthesis. Using an endogenously expressed viral transcript as a model, we demonstrated that the assay readily detects mRNA isoforms from as little as 10–100 pg of total cellular RNA or directly from a few cells. Multiplexed analysis of human cancer cell lines revealed differences in mRNA splicing and suggested a potential autocrine mechanism in the development of choriocarcinomas. Our approach may be useful in the large-scale analysis of the role of alternative splicing in development and disease.

Analysis of the human genome has revealed that up to 60% of genes express multiple mRNA isoforms via alternative splicing[1,2]. In several cases, human genes can produce hundreds or thousands of mRNA isoforms[3]. The fly *Dscam* gene is capable of generating ~38,000 mRNA isoforms, twice the predicted number of genes in the entire *Drosophila* genome[4]. Changes in alternative splicing have been linked to cell physiology, developmental regulation, and cancers[5,6]. However, the functional significance of most individual mRNA isoforms remains undefined, and little is known about the establishment of cell type–specific splice choices or their regulation by signaling or developmental cues. Deciphering the biological meaning of alternative splicing thus poses a challenge for genomic research and demands the development of novel approaches[7].

Current microarray technologies are limited in their ability to distinguish and analyze mRNA isoforms. The long probes used in cDNA arrays are not designed to discriminate the subtle sequence differences that can occur in splice variants as a result of the insertion of a few nucleotides or the substitution of highly related exons. Oligonucleotide-based arrays (including exon arrays, tiling arrays, and exon/exon junction arrays) provide partial solutions to the detection of mRNA isoforms (refs 8,9; M. Ares, pers. commun.). However, these approaches may be compromised by cross-hybridization due to closely related sequences. Additionally, conventional microarray analyses tend to rely for increased sensitivity on target amplification methods that are biased toward 3′ UTR sequences, and may therefore require larger amounts of RNA to preserve mRNA isoform representation.

Here we describe a general methodology for profiling alternative splicing. The method permits direct analysis of splice variants without prior RNA purification or cDNA synthesis, allows discrimination of mRNA isoforms and closely related gene transcripts, and requires only sub-nanogram amounts of total RNA for analysis. We have combined this methodology with a high-throughput microarray platform and applied this technology to the characterization of alternative splicing in a panel of human cancer cell lines. Our results revealed cell-specific alternative splicing indicative of distinct mechanisms for growth control.

Results

A general method for profiling alternative splicing. We designed a strategy for large-scale analysis of mRNAs based on a fiber-optic microarray platform[10–12]. In this platform, a self-assembled bead array is formed when a mixture of microspheres (beads) is loaded onto the tip of an etched fiber-optic bundle (Fig. 1A). Each bead carries a specific oligonucleotide sequence (address), and the positions of individual beads on the array are identified through iterative hybridization or "decoding" (M.S.C. and colleagues, unpublished results). Hybridization signals of ~100-fold linear range are recorded by a charge-coupled device (CCD) camera at the opposite end of the fiber bundle, and multiple bundles are arranged in a block to form a matrix of arrays, permitting large-scale parallel analysis.

Alternatively spliced mRNAs have unique sequences at splice junctions, which can be individually analyzed using an addressable approach[13,14]. In our strategy, DNA oligos complementary to alternative exons are each linked to a unique sequence (address) that can hybridize to its complementary strand on the bead array (Fig. 1B), or on any other array containing oligos complementary to selected address sequences. To avoid potential cross-hybridization from single oligos hybridizing across splice junctions, we incorporated a previously described strategy for improving targeting specificity by ligation[15,16]. In this approach, we target any given splicing event using two oligos complementary to exonic sequences flanking the splice junction (Fig. 1B). Aligned oligo pairs can be ligated when bridged by expressed mRNA isoforms. To achieve maximal sensitivity, we designed a signal-amplification step after oligo ligation by linking addressed and common oligos with universal priming sequences for PCR. The use of universal primers and the uniform length of amplicons minimize potential bias during amplification (see below).

Figure 1C summarizes the method, which comprises RNA-mediated annealing, selection, and ligation (RASL). Total cellular RNAs are annealed to a mixture of addressed oligos, common oligos, and

[1]*Department of Cellular and Molecular Medicine, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093.* [2]*Current address: Illumina, Inc., 9885 Towne Centre Drive, San Diego, CA 92121.* [3]*Illumina, Inc., 9885 Towne Centre Drive, San Diego, CA 92121.* [4]*Current address: Johnson & Johnson Pharmaceutical Research & Development, L.L.C., 3210 Merryfield Row, San Diego, CA 92121.* *Corresponding authors (xdfu@ucsd.edu and jfan@illumina.com).*

Alamar Exhibit 1002
Alamar v. Olink

Best Available Copy

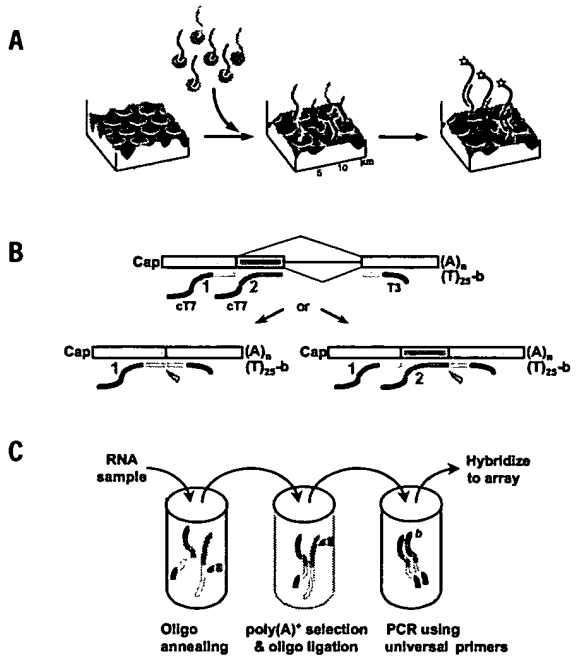


Figure 1. The technology platform and experimental strategy. (A) Assembly of an array of beads in wells. A bundle of optical fibers is chemically etched to create an array of shallow pits. Microspheres, each covalently linked to multiple copies of an oligonucleotide (address), are embedded on the etched surface. The address sequences (colored) are selected on the basis of minimal homology to sequences in DNA databases and for their good behavior in hybridization. These address sequences are used to determine the identity of individual beads in randomly assembled bead arrays, and serve as probes for experimental samples. Hybridization of biotinylated oligos containing complementary sequences to the addresses can be measured by labeled streptavidin (star). (B) Detection of alternative splicing using the addressable approach. Oligos targeted to alternative splice sites are tagged with a specific address (1, 2). Both addressed oligos and common oligos contain universal primer sites (T3 or T7). RNAs are annealed to oligo mixtures and biotinylated oligo-d$(T)_{25}$. Only annealed oligos that are hybridized to the splice junction should be ligated (lightning symbol) and serve as templates for amplification. (C) The RASL procedure. RNA samples are denatured and annealed to oligo mixtures and biotinylated oligo-d$(T)_{25}$ (biotin indicated by orange ball). Alternatively, cell lysis and oligo annealing can be carried out in a single step. After the mixture is transferred to streptavidin-coated tubes (orange), excess oligos are washed off, and annealed oligos are ligated and then eluted to serve as template for PCR. Biotin is incorporated during amplification using biotinylated T7 and unmodified T3 primers. Labeled RASL-PCR products are then hybridized to the bead array and visualized by a CCD camera. Signals from multiple beads carrying identical addresses on the array are combined to derive the final signal output.

biotinylated oligo-dT. Samples are then transferred to streptavidin-coated PCR tubes for poly(A)$^+$ selection. Thus, the assay is transferred to a solid phase, allowing the removal of free oligos that otherwise would be nonspecifically ligated. Juxtaposed oligos are joined by T4 DNA ligase[17]. This step is designed to improve the specificity of the assay because, in principle, only juxtaposed oligos should be ligated, and only ligated products should then be amplified by PCR. One of the universal PCR primers is labeled with dye or biotin for quantitation of the RASL-PCR products on the array. Although the RASL-PCR method is designed to measure mRNA isoforms, the approach should also be adaptable for profiling gene expression by selecting unique targets in individual mRNAs.

Splicing shift detected by RASL. To validate the RASL strategy for detecting splicing changes, we chose the well-characterized alternative splicing of E1A transcripts as a model system. E1A pre-mRNA is spliced using three alternative donors to a common splice acceptor, resulting in 9S, 12S, and 13S mRNAs (Fig. 2A). RT-PCR analysis of RNAs from transfected COS cells showed that all three splice donors were used and that co-transfection with the splicing factor ASF/SF2 shifted splice-site usage toward the proximal 13S splice donor (Fig. 2B), consistent with previous reports[18,19]. The same RNAs were subjected to RASL-PCR, and a specific band was detected only in cells expressing E1A (Fig. 2B). When RASL-PCR products were hybridized to microarrays assembled for this experiment, the fluorescence intensity measured for each splice isoform corresponded to the levels of mRNA isoforms predicted by RT-PCR (Fig. 2C). Quantitation of intensity levels showed that RASL-PCR detected ASF/SF2-dependent decreases in 9S and 12S splicing and an increase in 13S splicing (Fig. 2D). Note that because of kinetic differences in oligo targeting to mRNAs and in hybridization of RASL-PCR products to the array, one can only compare relative changes in the abundance of a given target between different samples, rather than derive the abundance of different isoforms within a single sample. We therefore used RNA from cells transfected with the E1A reporter alone as the baseline (1×) for comparison. Our data show that, although the absolute magnitude of changes induced by ASF/SF2 differed between the RT-PCR and RASL-PCR assays, overall changes in splicing were concordant.

Specificity and sensitivity of the assay. To eliminate the variability and overexpression associated with transient transfection, we chose HEK 293T cells carrying an integrated adenovirus 5 (Ad5) fragment as an endogenous E1A expression system. RT-PCR

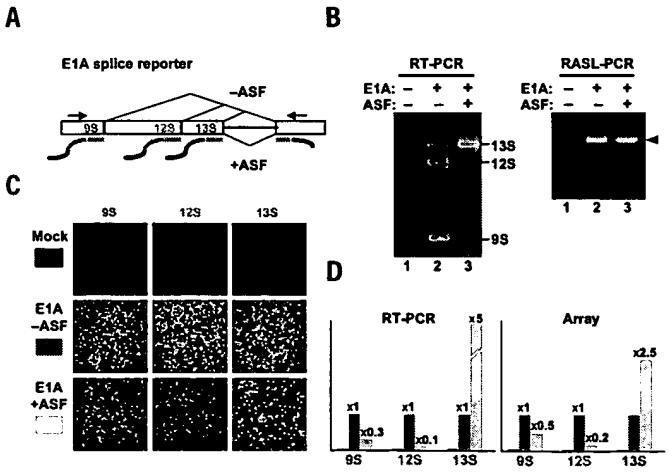


Figure 2. Detection of splicing changes in transfected cells. (A) The diagram illustrates alternative splice donor usage (9S, 12S, 13S) and the effect of ASF/SF2 overexpression on splicing of the E1A reporter. Arrows indicate primers for RT-PCR. (B) Total RNAs from COS cells mock-transfected (1) or transfected with the E1A splicing reporter alone (2) or with ASF/SF2 (3) were analyzed by RT-PCR (left) or RASL-PCR (right, products indicated by the arrowhead). Note that the RASL-PCR products are all of similar size, even though they consist of a population of sequences representing E1A mRNA isoforms. (C) Hybridization images. Multiple bundles, each carrying a single address, were constructed for this experiment. Changes in fluorescence intensity induced by ASF/SF2 are visually detectable. (D) Quantitative changes in alternative splicing measured by RT-PCR are compared to those derived from RASL-PCR. Data are expressed as fold differences relative to E1A isoform expression in the absence of overexpressed ASF/SF2. Black bars, mock transfection; dark-gray bars, minus ASF/SF2; light-gray bars, plus ASF/SF2.

© 2002 Nature Publishing Group http://biotech.nature.com

Alamar Exhibit 1002
Alamar v. Olink

Best Available Copy

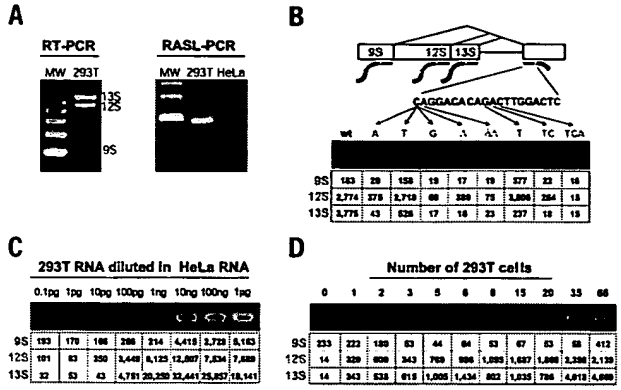


Figure 3. Specificity and sensitivity of the RASL-PCR assay. (A) Expression of endogenous E1A transcripts in 293T cells. These cells predominantly express the 12S and 13S isoforms as detected by RT-PCR (left). RASL-PCR reported E1A expression in 293T, but not in HeLa cells (right). (B) Specificity of RASL-PCR. E1A transcripts in 293T total RNA were assayed using mutant common oligos tested individually in combination with wild-type addressed oligos. Mutations were incorporated either at the 3′ end or 10 nucleotides from the 3′ end of the common oligo. RASL-PCR products as analyzed by agarose gel are shown above the corresponding array data. (C, D) Sensitivity of RASL-PCR. 293T total RNA was progressively diluted in 1 µg of HeLa total RNA to reduce E1A transcript concentration without decreasing overall mRNA complexity. To directly detect E1A signals in individual cells, 293T cells were diluted into individual wells in 96-well plates, counted microscopically, then lysed directly in lysis/binding buffer and processed by RASL-PCR. The numbers in B, C, and D represent the fluorescence (arbitrary units) captured by the CCD camera.

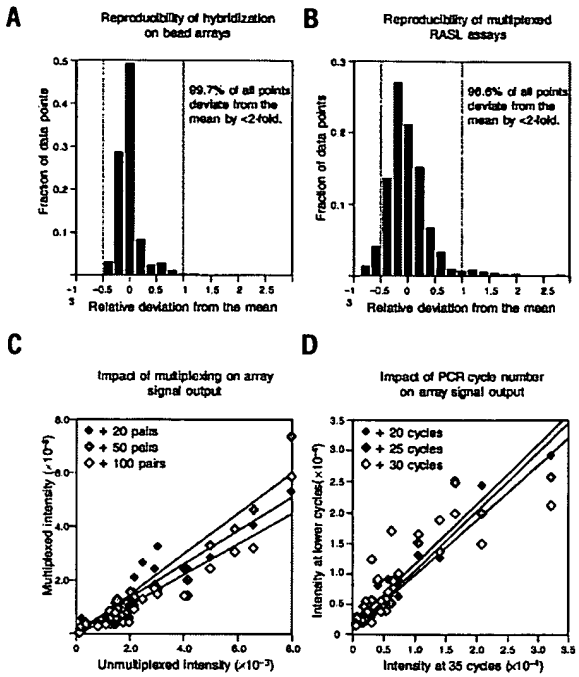


Figure 4. Characterization of multiplexed RASL-PCR. (A) Reproducibility of hybridization from bundle to bundle. The expression of 97 isoforms in 293T cells was analyzed by RASL-PCR under multiplexed conditions and identical aliquots of the RASL-PCR products were applied to 16 individual arrays, yielding 16 signals for each target. For each address, the mean and the difference from the mean for each bundle were calculated. Bars represent the fraction of data points within a given range of signal deviation from the mean. (B) Reproducibility of RASL-PCR. Sixteen parallel multiplexed analyses of 293T total RNA were used to calculate overall reproducibility of the entire experimental procedure. (C) The impact of multiplexing on signal output. The analysis was conducted by monitoring 10 reference mRNAs in the presence of increasing numbers of oligos targeting other mRNAs. (D) The impact of PCR cycle number on isoform representation. RASL products for 97 isoforms in 293T cells were amplified for 20, 25, 30, or 35 cycles. PCR products were normalized for mass before array hybridization by diluting the higher cycle number products. Signal intensities for the lower cycle numbers are plotted against those for 35 cycles.

revealed that these cells expressed roughly equal amounts of 12S and 13S mRNA, with little detectable 9S mRNA (Fig. 3A). RASL-PCR was able to detect E1A transcripts only in 293T RNA, not in HeLa RNA, demonstrating the specificity of the assay (Fig. 3A).

Assay specificity was further examined using the scheme diagrammed in Figure 3B. In each sample, the common oligo for E1A was replaced by a mismatched variant. Mismatches and deletions at the 3′ terminus dramatically reduced the amount of RASL-PCR product by affecting ligation, as previously demonstrated[16], with the exception of the C to T transition, which was likely due to wobble base-pairing in the oligo–mRNA duplex. We note that the 13S isoform was preferentially affected by mutations in the common oligo, possibly because of context-dependent differences in annealing and/or ligation efficiencies. At an annealing temperature of 45°C, the assay appeared relatively insensitive to the presence of a single-base mismatch positioned 10 nucleotides from the splice junction, but two or three mismatches at that position were sufficient for discrimination. Mismatch specificity could be further increased by raising the annealing temperature, but with reduced sensitivity (data not shown). Through titration of the annealing temperature, experimental conditions were determined that permit efficient hybridization of most oligos to their mRNA targets, yet preserve sufficient specificity to distinguish among closely related transcripts.

The sensitivity of the assay was tested by making serial dilutions of 293T total RNA in HeLa total RNA. In this way, the overall sample input was maintained, but the amount of E1A-specific transcript was progressively decreased. As little as 10–100 pg of 293T total RNA in 1 µg HeLa total RNA gave a detectable signal on the microarray (Fig. 3C). A total RNA content of 10 pg roughly corresponds to the RNA content of a single cell. To test whether single cells could be used as source material, we carried out serial dilutions of 293T cells in 96-well plates and counted the number of cells per well by visual inspection. Cells were subjected to direct lysis followed by RASL-PCR. Microarray analysis showed that even <10 cells gave readily detectable E1A signals (Fig. 3D). To estimate E1A expression in 293T cells, we carried out real-time quantitative PCR and found that a single 293T cell expresses about 100 E1A transcripts, consistent with an estimate by northern blotting (data not shown). Together, these results illustrate the sensitivity of the RASL assay and its ability to detect specific signals for well-expressed transcripts in subnanogram amounts of total RNA. Clearly, detection of rarer messages will require larger amounts of starting material.

The 293T RNA titration experiment in Figure 3C showed that the signal intensity started to fall off at higher concentrations (see 12S and 13S signals with 100 ng and 1 µg of 293T total RNA). This observation indicates that the array output is saturated under such conditions. We are continuing to investigate ways to increase the sensitivity and dynamic range of the assay.

Multiplex analysis. Having established experimental conditions using the E1A model, we examined the capacity of RASL to analyze alternative splicing in a multiplexed fashion. For this purpose, we prepared oligo pairs targeting 97 mRNA isoforms from 23 genes (*BAX, BCL2, BCL2L1, BRCA1, CASP2, CDKN2A, CDKN2B, ESR2, FAS, FGFR2, FN1, FOLH1, FYN, ITGB1, KLK3, MUC1, PTPRC, RUNX1, TERT, TP63, TP73, VEGFA*, and *WT1*).

© 2002 Nature Publishing Group http://biotech.nature.com

Alamar Exhibit 1002
Alamar v. Olink

Best Available Copy

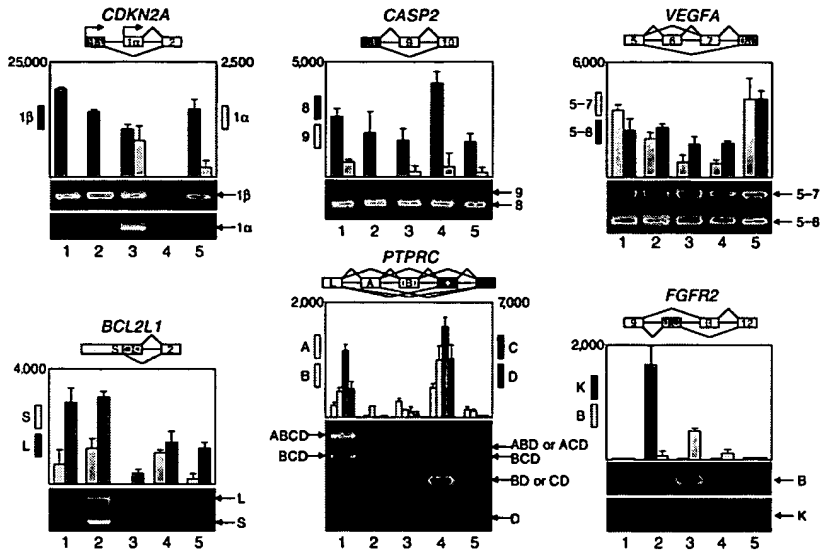


Figure 5. Cell type–specific gene expression and alternative splicing analyzed by RASL-PCR and RT-PCR. Transcripts from six genes analyzed by multiplexed RASL-PCR (bar graph) were chosen for validation by RT-PCR (gel bands). In each case, the structure of known splice variants is diagrammed above. For *PTPRC*, the assay contains address-linked oligos targeting the 5′ ends of exons A–D as well as unaddressed oligos targeting the 3′ ends of exons L to C. This strategy was designed to measure the frequency of alternative exon usage. Expression was examined in triplicate in 5 cell lines (1 = U-937, 2 = HT-29, 3 = 293T, 4 = Jurkat, 5 = HeLa). Error bars indicate standard deviation. Because different addresses cannot be directly compared, some data are scaled for comparison between splice isoforms for *CDKN2A* (1β on the left, 1α on the right) and *PTPRC* (A and B on the left, C and D on the right).

© 2002 Nature Publishing Group http://biotech.nature.com

We first examined the reproducibility and signal output of the assay under multiplexed conditions. When identical RASL-PCR samples were applied to each of 16 arrays in an array matrix, we found that signal variation was minimal from array to array. Statistical analysis indicated that 99.7% of data points fell within 2-fold of their mean (Fig. 4A). When identical RNA samples were processed through RASL-PCR and arrayed, 96.6% fell within 2-fold of their mean (Fig. 4B). Further analysis indicated that the majority of variation in independent RASL-PCR assays is derived from stochastic PCR amplification of rare isoforms (data not shown). Overall, these data demonstrate the reproducibility of the array format.

To determine the impact of multiplexing on signal output, we measured the expression of 10 mRNA isoforms from 5 genes in the presence of increasing numbers of other targeting oligo pairs (Fig. 4C). The signals from the 10 reference targets, assayed in the presence of as many as 100 other targets, remained largely unaltered, indicating that the impact of multiplexing at the current complexity is minimal. In additional experiments, we noted a limitation of our assay when multiplexed RASL-PCR reactions were carried out in the presence of some high-abundance mRNAs, which caused early saturation in PCR amplification. As a result, messages of low abundance were not amplified to a detectable level. Thus, it may be preferable to assay mRNAs in groups according to their relative abundance in cells to detect changes in a quantitative range.

Finally, because RASL products are of similar length and contain universal primer-binding sites, biases during PCR amplification should be minimized. To test this, we varied the number of PCR cycles from 20 to 35 and applied a normalized amount of RASL-PCR products to the array. We found that relative mRNA isoform abundances were only modestly affected by PCR cycle number (Fig. 4D). We therefore routinely use 35 cycles of PCR amplification to maximize sensitivity.

Profiling and validating alternative splicing in cell lines. Using RASL in triplicate under multiplexed conditions, we examined the expression of the 97 isoforms in five human cancer cell lines (Fig. 5). Six expressed genes that were predicted by array analysis to show either cell-specific or uniform splicing patterns were chosen for validation by RT-PCR. A comparison of the results obtained by microarray and RT-PCR showed that the two methods generally agreed, although the magnitude of the signals varied, presumably because both analyses are semiquantitative.

Among the genes we selected for comparison, the p16/p14^{ARF} gene (*CDKN2A*) uses alternative promoters coupled with differential splicing to express two unrelated proteins that control the cell cycle through separate pathways[20]. This locus is a frequent target for mutation or deletion in tumors[21]. Both RASL and RT-PCR revealed consistent and distinct expression profiles of the two isoforms in the tumor cell lines examined (Fig. 5). By contrast, array analysis of these five cell lines did not predict differences in alternative splicing of internal exons in *CASP2* and *VEGFA*. The RT-PCR data were in general agreement. Neither the array nor RT-PCR detected *VEGFA* splicing from exon 5 to 6 in these cells. As for *BCL2L1* (*Bcl-X*), both methods showed some variations in isoform expression among the cell lines. Because *CASP2* and *BCL2L1* each express two isoforms with opposite functions in the regulation of apoptosis[22], it would be interesting to examine whether their expression can be regulated at the splicing level by apoptotic stimuli.

The *PTPRC* (*CD45*) gene encodes a receptor tyrosine phosphatase critical for T-cell signaling and is known to be regulated by alternative splicing in T-cell development[23]. This gene is highly expressed in hematopoietic cells, consistent with its expression in U-937 and Jurkat cells as detected by both RASL and RT-PCR (Fig. 5). However, the expression of distinct isoforms in these two cell types was not observed by RASL, reflecting a flaw in our oligo

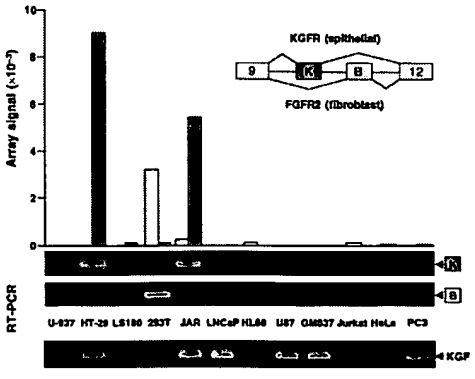


Figure 6. Potential autocrine signaling in human cancer cell lines. The insert shows the cell type–specific splicing pattern of the *FGFR2* gene. Twelve cell lines were analyzed by multiplexed RASL-PCR, and the data for K and B isoforms are presented. The expression of K and B exon-containing transcripts was analyzed by RT-PCR using two separate primer pairs, and these results are presented below the bar graph. RT-PCR was also carried out on the same 12 RNAs to determine which cell types also express KGF (bottom).

Alamar Exhibit 1002
Alamar v. Olink

© 2002 Nature Publishing Group http://biotech.nature.com

design aimed at detection of all possible combinations of alternative exons (see Fig. 5 legend). This problem may be addressed by only using oligo pairs for L-A, L-B, L-C, and L-D isoforms, which would clearly distinguish between U-937 and Jurkat cells. But even this design would not reveal how individual exons are uniquely combined, illustrating a limitation of the current experimental strategy.

Finally, transcripts from the *FGFR2* (fibroblast growth factor receptor 2) gene that give rise to cell-surface receptors with different ligand specificity showed well-defined cell-specific splicing by microarray analysis and by RT-PCR (Fig. 5). Confirmation by RT-PCR of the RASL results for this gene and the other genes described above supports the utility of the RASL method for parallel analysis of alternative splicing under multiplexed conditions.

A potential autocrine mechanism in choriocarcinomas. Our observation of clear cell-specific expression of FGFR2 isoforms allowed us to apply the RASL-PCR method to a biological problem in cancer. The *FGFR2* gene gives rise to two isoforms: the B form encodes FGFR2 in fibroblasts whereas the K form encodes KGFR (keratinocyte growth factor receptor) in epithelial cells[24]. Because fibroblasts secrete KGF and epithelial cells secrete FGF, a paracrine loop is established for growth control[25–27]. Forced expression of KGFR in fibroblasts leads to cellular transformation by the creation of an autocrine loop[24,28]. Such a switch in *FGFR2* splicing is thought to accompany cancer progression in prostate tumors[29,30]. We therefore extended RASL analysis to additional cell lines, including several derived from human prostate cancers (Fig. 6). Although high levels of *FGFR2* expression were detected in three cell lines (HT-29 and JAR expressing *KGFR*, and 293T expressing *FGFR2*), expression of FGFR2 mRNA isoforms was not significant in the two prostate cancer cell lines LNCaP and PC3 under these conditions. Quantitative RT-PCR revealed that all prostate cell lines we examined expressed *FGFR2* transcripts at the level of <0.01 copies per cell (data not shown). Additional investigation is necessary to determine whether such low levels of expression contribute to tumor growth.

The high level of expression of the KGFR mRNA isoform in HT-29 and JAR cells prompted us to examine whether its ligand was also expressed in these cells. RT-PCR indicated that both HT-29 and JAR cells expressed substantial amounts of KGF mRNA, suggesting the potential for autocrine growth stimulation in these cells (Fig. 6). HT-29 cells are derived from a colorectal adenocarcinoma. Human colon cancers were recently shown to overexpress both *KGF* and *KGFR* (ref. 31). JAR cells are derived from a choriocarcinoma. Because normal placental cells exhibit paracrine KGF/KGFR signaling by distinct cell types[32], our results and a similar finding with another choriocarcinoma cell line BeWo (ref. 33) raise the possibility that the establishment of a KGFR/KGF autocrine loop may play a role in human placental tumors. The RASL technology could be applied to a variety of clinical tumors to test the hypothesis that autocrine loops contribute to oncogenic processes in multiple tissues.

Discussion

We have described an approach for parallel analysis of gene expression and alternative splicing using a bead-based fiber-optic microarray platform. The addressable approach allows a wide range of applications using a universal microarray format. Because of its specificity, the assay can be used to discriminate mRNA isoforms from a single gene. In principle, the method is also applicable to monitoring gene expression, especially in discriminating transcripts from closely related genes, or from distinct alleles of the same gene. The sensitivity achieved by the amplification step is especially suitable for monitoring gene expression in microdissected tissues, and the array matrix format permits parallel analysis of many samples.

The current method has some limitations. First, because the splicing assay relies on splice junction sequences, we have little flexibility in selecting the most suitable sequences for hybridization. Second, the RASL-PCR approach targets known sequences, so the method is not a screen for novel transcript expression. Currently, the RASL technology may best be used for large-scale analysis of candidates from genome-wide surveys in various databases. When combined with microdissection, isoform expression may be further assigned to specific cell types in tissue samples. This strategy may be particularly suitable for molecular analysis of solid tumors. Finally, as mentioned earlier, it may not be possible to use RASL to definitively identify mRNA isoforms when multiple regions in a given transcript are subject to alternative splicing. In such cases, RASL may be used to measure the frequency of splice-site utilization, but deducing full-length isoforms will still require cDNA cloning and sequencing.

We expect that the RASL-PCR method will be helpful in elucidating the role and regulation of alternative splicing in development and disease. Our studies revealed cell type–specific gene expression (*PTPRC*), alternative promoter usage (*CDKN2A*), and differential exon selection (*FGFR2*) in a panel of human tumor cell lines. Thus, our method may permit the molecular classification of cancer by profiling both transcription and alternative splicing characteristic of particular tumor types or subtypes. The sensitivity of this microarray technology may be particularly advantageous in addressing questions that require analysis of small numbers of cells, such as the question of molecular diversity in the immune and neuronal systems.

Experimental protocol

Target selection and oligo design. Alternative exons were identified by searching scientific literature and public gene databases. Genes were chosen on the basis of their roles in growth control and relatively simple splicing patterns. Oligos were designed to hybridize to the 20 nucleotides flanking both sides of a splice junction such that upstream and downstream oligos were bridged by mRNA exactly at the splice junction. Exon-specific sequences (20-mers) were fused to address sequences (20- to 24-mers) using an algorithm to minimize intramolecular base-pairing. T7 and T3 priming sequences (23-mers) were also fused to the oligo sequences for signal amplification. Biotinylated T7 primer was used in the amplification to label the strand that hybridized to the array.

RASL-PCR. Oligos complementary to upstream exon sequences were first treated with kinase to allow ligation. Total RNAs (usually 1 μg) or cell lysates were annealed to 100 fmol of each oligo and 2 pmol of biotin-$(dT)_{25}$ in 40 μl in lysis/binding buffer (20 mM Tris-Cl, pH 7.6, 500 mM NaCl, 1 mM EDTA, 0.1% SDS). For experiments with less RNA, the total RNA mass was adjusted to 1 μg by addition of yeast tRNA. Samples were heated for 2 min at 94°C and cooled to 45°C for 5 min before transferring to streptavidin-coated PCR tubes (Roche, Indianapolis, IN). Annealing and selection continued for 1 h at 45°C; then the captured material was washed three times with wash buffer (20 mM Tris-Cl, pH 7.6, 100 mM NaCl, 1 mM EDTA) containing 0.1% Triton X-100, preheated to 45°C. After one wash with wash buffer without Triton or ligation buffer (Fermentas, Hanover, MD), bridged oligos were ligated for 1 h at 37°C using 10 U T4 DNA ligase (Fermentas) in 40 μl ligation buffer. After two more washes with wash buffer, once with and once without Triton, ligated oligos were eluted in 40 μl water at 65°C for 5 min. Between 10% and 100% of the eluant was used as a template for PCR with T3 and biotinylated T7 primers in a 50 μl reaction (15 s denaturing at 94°C, 30 s annealing at 54°C, and 30 s extension at 72°C for 35 cycles). Gels shown represent 10–20% of the PCR reaction.

Microarray assembly, hybridization, and imaging. Microarrays were assembled by loading pools of glass beads (3 or 5 μm in diameter) derivatized with oligonucleotides onto the etched ends of fiber-optic bundles. In the current format, 50,000 optical fibers are hexagonally packed to form an ~1.4 mm diameter bundle. The decoding procedure is based on hybridization of the array to pools of complementary oligos that are fluorescently labeled. By sequential hybridization, an identifying color code is generated for each bead corresponding to a sequence address. The decoding procedure is ~99.9% accurate. For any given bead type within an array, bead-to-bead differences in

Alamar Exhibit 1002
Alamar v. Olink

hybridization fluorescence intensity have a coefficient of variation of 20–30%. On the basis of statistical considerations and computer simulation, 5 beads per address are required for a reliable quantitative estimate of expression levels. In the arrays used in this report, there were ~100 beads for each address. For example, in Figures 4A and 4B there were an average of 116 and 104 beads per address, respectively, with a standard deviation of 41 and 36.

All aqueous steps were done by submerging the bead end of the arrays into wells containing hybridization or wash buffers. Beads were conditioned in 0.1 M NaOH for 1 min followed by neutralization in hybridization buffer (0.1 M potassium phosphate, pH 7.6, 1 M NaCl, 0.1% Tween-20, 5% ethanol). Microarrays were prehybridized with 100 μg/ml herring sperm DNA in hybridization buffer for 30 min at 48°C. After washing in hybridization buffer, 1 to 8 μl of RASL-PCR products were denatured in a final volume of 50 μl hybridization buffer containing 100 μg/ml herring sperm DNA and 0.3 μM each of unlabeled T7 and T3 primers, and hybridized to the arrays for 3 h at 48°C. After hybridization, arrays were washed for 3 min at 48°C with 5 mM sodium phosphate, pH 7.4, 74.5 mM NaCl, 0.5 mM EDTA, 0.01% Triton X-100, 3 min at 48°C with 10 mM sodium phosphate, pH 7.4, 149 mM NaCl, 1 mM EDTA, 0.01% Triton X-100, and 7 min in hybridization buffer.

To quantitate hybridized biotinylated oligos, arrays were blocked in hybridization buffer containing 2 mg/ml BSA for 3 min, and then stained with streptavidin–phycoerythrin conjugate (Molecular Probes, Eugene, OR) at 2 μg/ml in the same buffer for 15 min. After washing three to five times in hybridization buffer, the arrays were imaged at a resolution of 1 μm/pixel using a microscope with 540 nm excitation and 580 nm emission filters and a CCD camera. Data were integrated over 2–5 s, and processed using Ultralite software (Illumina).

Acknowledgments

The authors thank Y.-S. Kwon and other members of the Fu laboratory for their assistance, E. Chudin from Illumina for statistical analyses, and S. Goodison, C. Nelson, H. Li, C. Buckmaster, A. Boyer, M. Ricote, J. Brown, C. Snyder, and E. Brinkman-van der Linden for their gifts of cell lines. D. Che, F. Garcia, and J. Haas provided expert assistance with imaging systems and image analysis. X-D. F. is a Scholar of the Leukemia and Lymphoma Society. This work was supported by grants from the American Cancer Society (J.M.Y) and National Institutes of Health (X-D. F. and M.C.).

Received 25 September 2001; accepted 6 February 2002

1. Lander, E.S. *et al.* Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921 (2001).
2. Venter, J.C. *et al.* The sequence of the human genome. *Science* **291**, 1304–1351 (2001).
3. Graveley, B.R. Alternative splicing: increasing diversity in the proteomic world. *Trends Genet.* **17**, 100–107 (2001).
4. Schmucker, D. *et al. Drosophila* Dscam is an axon guidance receptor exhibiting extraordinary molecular diversity. *Cell* **101**, 671–684 (2000).
5. Lopez, A.J. Alternative splicing of pre-mRNA: developmental consequences and mechanisms of regulation. *Annu. Rev. Genet.* **32**, 279–305 (1998).
6. Smith, C.W. & Valcárcel, J. Alternative pre-mRNA splicing: the logic of combinatorial control. *Trends Biochem. Sci.* **25**, 381–388 (2000).
7. Black, D.L. Protein diversity from alternative splicing: a challenge for bioinformatics and post-genome biology. *Cell* **103**, 367–370 (2000).
8. Shoemaker, D.D. *et al.* Experimental annotation of the human genome using microarray technology. *Nature* **409**, 922–927 (2001).
9. Hu, G.K. *et al.* Predicting splice variant from DNA chip expression data. *Genome Res.* **11**, 1237–1245 (2001).
10. Michael, K.L., Taylor, L.C., Schultz, S.L. & Walt, D.R. Randomly ordered addressable high-density optical sensor arrays. *Anal. Chem.* **70**, 1242–1248 (1998).
11. Ferguson, J.A., Steemers, F.J. & Walt, D.R. High-density fiber-optic DNA random microsphere array. *Anal. Chem.* **72**, 5618–5624 (2000).
12. Walt, D.R. Bead-based fiber-optic arrays. *Science* **287**, 451–452 (2000).
13. Gerry, N.P. *et al.* Universal DNA microarray method for multiplex detection of low abundance point mutations. *J. Mol. Biol.* **292**, 251–262 (1999).
14. Fan, J.B. *et al.* Parallel genotyping of human SNPs using generic high-density oligonucleotide tag arrays. *Genome Res.* **10**, 853–860 (2000).
15. Landegren, U., Kaiser, R., Sanders, J. & Hood, L. A ligase-mediated gene detection technique. *Science* **241**, 1077–1080 (1988).
16. Nilsson, M., Antson, D.O., Barbany, G. & Landegren, U. RNA-templated DNA ligation for transcript analysis. *Nucleic Acids Res.* **29**, 578–581 (2001).
17. Kleppe, K., Van de Sande, J.H. & Khorana, H.G. Polynucleotide ligase-catalyzed joining of deoxyribo-oligonucleotides on ribopolynucleotide templates and of ribo-oligonucleotides on deoxyribopolynucleotide templates. *Proc. Natl. Acad. Sci. USA* **67**, 68–73 (1970).
18. Harper, J.E. & Manley, J.L. Multiple activities of the human splicing factor ASF. *Gene Expression* **2**, 19–29 (1992).
19. Cáceres, J.F., Stamm, S., Helfman, D.M. & Krainer, A.R. Regulation of alternative splicing *in vivo* by overexpression of antagonistic splicing factors. *Science* **265**, 1706–1709 (1994).
20. James, M.C. & Peters, G. Alternative product of the p16/CKDN2A locus connects the Rb and p53 tumor suppressors. *Prog. Cell Cycle Res.* **4**, 71–81 (2000).
21. Liggett, W.H., Jr. & Sidransky, D. Role of the p16 tumor suppressor gene in cancer. *J. Clin. Oncol.* **16**, 1197–1206 (1998).
22. Jiang, Z.H. & Wu, J.Y. Alternative splicing and programmed cell death. *Proc. Soc. Exp. Biol. Med.* **220**, 64–72 (1999).
23. Trowbridge, I.S. & Thomas, M.L. CD45: an emerging role as a protein tyrosine phosphatase required for lymphocyte activation and development. *Annu. Rev. Immunol.* **12**, 85–116 (1994).
24. Miki, T. *et al.* Expression cDNA cloning of the KGF receptor by creation of a transforming autocrine loop. *Science* **251**, 72–75 (1991).
25. Xu, X., Weinstein, M., Li, C. & Deng, C. Fibroblast growth factor receptors (FGFRs) and their roles in limb development. *Cell Tiss. Res.* **296**, 33–43 (1999).
26. De Moerlooze, L. *et al.* An important role for the IIIb isoform of fibroblast growth factor receptor 2 (FGFR2) in mesenchymal–epithelial signaling during mouse organogenesis. *Development* **127**, 483–492 (2000).
27. Hajihosseini, M.K., Wilson, S., De Moerlooze, L. & Dickson, C. A splicing switch and gain-of-function mutation in FgfR2-IIIc hemizygotes causes Apert/Pfeiffer-syndrome-like phenotypes. *Proc. Natl. Acad. Sci. USA* **98**, 3855–3860 (2001).
28. Miki, T. *et al.* Determination of ligand-binding specificity by alternative splicing: two distinct growth factor receptors encoded by a single gene. *Proc. Natl. Acad. Sci. USA* **89**, 246–250 (1992).
29. Yan, G., Fukabori, Y., McBride, G., Nikolaropolous, S. & McKeehan, W.L. Exon switching and activation of stromal and embryonic fibroblast growth factor (FGF)–FGF receptor genes in prostate epithelial cells accompany stromal independence and malignancy. *Mol. Cell. Biol.* **13**, 4513–4522 (1993).
30. Carstens, R.P., Eaton, J.V., Krigman, H.R., Walther, P.J. & Garcia-Blanco, M.A. Alternative splicing of fibroblast growth factor receptor 2 (FGF-R2) in human prostate cancer. *Oncogene* **15**, 3059–3065 (1997).
31. Watanabe, M., Ishiwata, T., Nishigai, K., Moriyama, Y. & Asano, G. Overexpression of keratinocyte growth factor in cancer cells and enterochromaffin cells in human colorectal cancer. *Pathol. Intl.* **50**, 363–372 (2000).
32. Taniguchi, F. *et al.* Paracrine effects of bFGF and KGF on the process of mouse blastocyst implantation. *Mol. Reprod. Dev.* **50**, 54–62 (1998).
33. Taniguchi, F. *et al.* Keratinocyte growth factor in the promotion of human chorionic gonadotropin production in human choriocarcinoma cells. *Am. J. Obstet. Gynecol.* **182**, 692–698 (2000).

© 2002 Nature Publishing Group http://biotech.nature.com

Alamar Exhibit 1002
Alamar v. Olink

The use of mRNA display to select high-affinity protein-binding peptides

David S. Wilson*, Anthony D. Keefe, and Jack W. Szostak†

Howard Hughes Medical Institute and Department of Molecular Biology, Massachusetts General Hospital, Boston, MA 02114

Contributed by Jack W. Szostak, January 16, 2001

We report the use of "mRNA display," an *in vitro* selection technique, to identify peptide aptamers to a protein target. mRNA display allows for the preparation of polypeptide libraries with far greater complexity than is possible with phage display. Starting with a library of $\approx 10^{13}$ random peptides, 20 different aptamers to streptavidin were obtained, with dissociation constants as low as 5 nM. These aptamers function without the aid of disulfide bridges or engineered scaffolds, yet possess affinities comparable to those for monoclonal antibody–antigen complexes. The aptamers bind streptavidin with three to four orders of magnitude higher affinity than those isolated previously by phage display from lower complexity libraries of shorter random peptides. Like previously isolated peptides, they contain an HPQ consensus motif. This study shows that, given sufficient length and diversity, high-affinity aptamers can be obtained even from random nonconstrained peptide libraries. By engineering structural constraints into these ultrahigh complexity peptide libraries, it may be possible to produce binding agents with subnanomolar binding constants.

affinity tag | peptide library | peptide aptamer | Strep-tag | streptavidin

Antibodies and other specific protein-binding reagents are indispensable tools for monitoring protein abundance and activity. With the number of known proteins increasing dramatically as a result of whole-genome-sequencing projects, it has become crucial to find alternatives to traditional time-consuming monoclonal antibody production for generating high-affinity binding reagents. Numerous techniques have been developed to select polypeptides that bind with high affinity and specificity to target proteins (reviewed in ref. 1). The most commonly used such technique is phage display, in which each bacteriophage "displays" a unique peptide or protein on its surface, allowing for the *in vitro* selection of ligands with high affinities (2). As would be expected, the affinity of ligands derived from this process increases with increasing complexity (i.e., diversity) of the starting library (3, 4). The starting complexity of phage display libraries is generally limited to $<10^9$ because of the bacterial transformation requirement. To overcome this limitation, two completely *in vitro* procedures for linking polypeptides to their encoding mRNAs have been developed. Ribosome display achieves this link by stalling the translating ribosome in an *in vitro* translation reaction, thus maintaining the mRNA-ribosome-nascent peptide complexes (5, 6). mRNA display takes advantage of the translation-terminating antibiotic puromycin, which functions by entering the A site of ribosomes and forming a covalent bond with the nascent peptide. By covalently attaching puromycin to the 3′ end of an mRNA, a covalent link between a polypeptide and its encoding message can be achieved *in situ* during *in vitro* translation (7–9). These mRNA-peptide fusions can then by purified and subjected to *in vitro* selection. Here we report the use of this procedure to isolate peptide ligands.

Several different types of library have been used for *in vitro* selections, including random linear or disulfide-constrained peptide libraries and scaffold-based libraries. Nonconstrained linear peptide libraries generally do not yield high-affinity ligands to proteins, except in cases where the protein normally functions in peptide recognition (10, 11). By constraining the peptides, either through disulfide formation or by presenting them on the surface of a known protein scaffold (such as that of the antibody), higher affinities have been achieved (12–21). We show here, however, that even random nonconstrained peptides with high affinity (5–10 nM) to protein targets can be identified if the starting library is of sufficiently high complexity and length, in this case almost 10^{13} peptides, each 88 amino acids long.

Materials and Methods

For a detailed experimental protocol, please refer to the supplemental data (which is published on the PNAS web site, www.pnas.org).

Library Synthesis. The library was constructed, transcribed, ligated to the puromycin-containing linker, translated, fused to the mRNA, purified, and reverse transcribed as described (22). By comparing the ^{35}S counts of the purified, reverse transcribed mRNA-peptide fusions to the [^{35}S]methionine stock, and taking into consideration the total methionine concentration in the translation reaction (10 μM), we estimated the number of displayed peptides in the sample to be 6.7×10^{12}. This number also represents the complexity of the library, because it contains virtually no redundancy (22).

Selection for Streptavidin (SA)-Binding Peptides. The displayed peptides were incubated with immobilized SA (Ultralink Immobilized Streptavidin Plus, about 4 mg/ml; Pierce) in SA-binding buffer [40 mM Tris(hydroxymethyl)aminomethane/300 mM KCl/2 mM EDTA/0.1% Triton X-100/5 mM 2-mercaptoethanol/100 μg/ml BSA/1 μg/ml tRNA, pH 7.4]. The amount of gel used was 0.5 ml in a total volume of 5.5 ml. After incubating for 20 min at room temperature, the contents were loaded onto a chromatography column, drained, washed with 14 column volumes (CV) of SA-binding buffer, and then eluted with 5 successive aliquots (at 10-min intervals) of SA-binding buffer plus 2 mM D-biotin (Sigma). Elution fractions were combined and then PCR-amplified as described (22). This concluded the first round of selection, and the remaining rounds followed the same protocol except that the translation was scaled down 10-fold, and the washing of the SA column was increased (32 CV for round 2; 40 for rounds 3, 4, and 6; 25 for rounds 5 and 7). The SA-binding selection for rounds 5 and 7 were performed directly on the SA-column eluate from the preceding selection rounds, without intervening amplification (the biotin was removed by gel filtration). PCR products amplified after the seventh selection round were cloned and sequenced.

Analysis of Individual Peptide Sequences. To rapidly characterize selected peptides, a method for generating mRNA-peptide

Abbreviations: DTP, DNA-tagged peptide; EMSA, electrophoretic mobility-shift assay; SA, streptavidin.

*Present address: Zyomyx, 3911 Trust Way, Hayward, CA 94545.

†To whom reprint requests should be addressed. E-mail: szostak@molbio.mgh.harvard.edu.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

www.pnas.org/cgi/doi/10.1073/pnas.061028198

Alamar Exhibit 1002
Alamar v. Olink

Best Available Copy

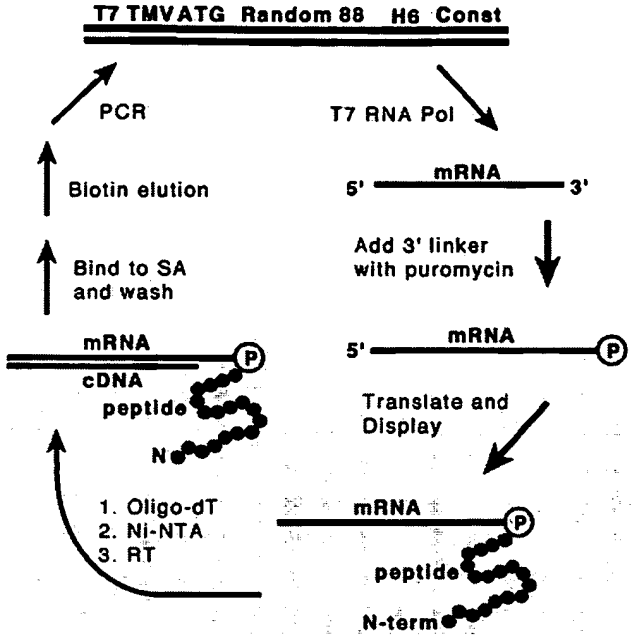


Fig. 1. *In vitro* selection process. Schematic showing the structure of the library and the selection scheme. The DNA library has, from 5′ to 3′, a T7 RNA polymerase promoter (T7), a tobacco mosaic virus translation enhancer (TMV; ref. 36), a start codon (ATG), 88 random amino acids, a hexahistidine tag (H6), and a 3′ constant region (Const). This library is transcribed by using T7 RNA polymerase, after which the puromycin-containing linker is ligated onto the 3′ end of the mRNA. When this template is translated *in vitro*, the nascent peptide forms a covalent bond to the puromycin moiety. The resulting mRNA-peptide covalent fusion molecules are then purified on oligo-dT-cellulose (which anneals to the oligo-dA sequence in the puromycin-containing linker) and Ni-NTA agarose. The mRNA portion of this display construct is then reverse transcribed. The double-stranded DNA/RNA-peptide species are then incubated with the immobilized target protein (SA) and unbound library members are washed off. SA-bound peptides are then displaced with biotin. The eluted molecules are then amplified by PCR, thus completing the first round of selection and amplification.

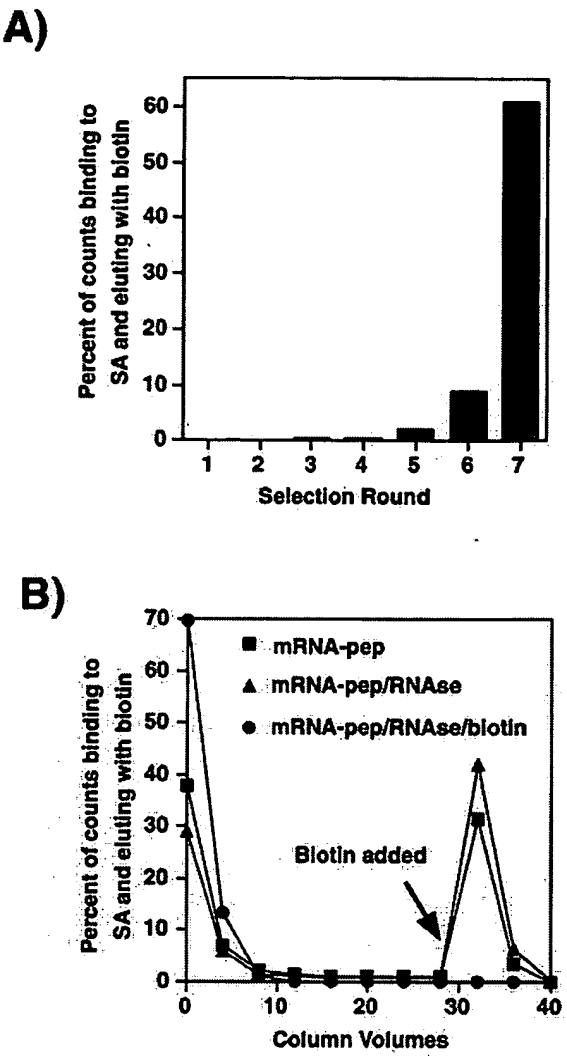


Fig. 2. Progress and result of the selection. (*A*) Fraction of ^{35}S counts from the displayed peptides that bound to SA and eluted with biotin, at each round of selection. (*B*) Elution profile for the peptide library generated from the output of the seventh round of selection. The first fraction represents the flow-through. Biotin was added at the point indicated. The plot compares the binding of the intact reverse-transcribed displayed peptides (mRNA-pep), the same sample treated with RNase A, and the RNase-treated sample applied to an SA column presaturated with biotin (excess biotin was washed away before exposing the library to the matrix).

covalent fusions was used (23). Plasmids containing single inserts were PCR-amplified with the same 5′ PCR primer as described for the library construction (22), and a new 3′ primer that altered the 3′ RNA sequence to ACUGGUCUGGCGGCUG-CAAGCUUGGCACCGGCUAU. This sequence was designed to anneal to the photo-crosslinking linker, which has the sequence 5′-psoralen-TAGCCGGTG-A17-CC-puromycin-3′, in which the underlined positions are 3′-methoxy nucleotides and the remaining ones are deoxy. This new primer changed the encoded constant C-terminal peptide sequence from WSG-GCHHHHHHSSA to WSGGCKLGTGY, of which the last three amino acids may not be translated because the RNA is annealed to the linker. Each DNA template was transcribed and gel purified as described (22), incubated with the psoralen linker, and irradiated at 366 nm. The translation/display reactions and oligo-dT-purification were carried out as above. Finally, RNase A was added to degrade the mRNA (complete degradation was confirmed by SDS/PAGE analysis).

The resulting purified DNA-tagged proteins (DTPs) were analyzed in the SA column-binding assay, in which ≈500 pM ^{35}S-labeled DTPs were mixed with 50 μl of the SA matrix in SA-binding buffer, in a total volume of 300 μl, for 10 min at room temperature (with agitation), after which the contents were loaded onto a chromatography column. The column was drained and then washed with 80 column volumes (CV) of SA-binding buffer, and then eluted with 3 consecutive aliquots (3 CV each) of SA-binding buffer plus 2 mM biotin (over a 15-min period). All fractions were analyzed by scintillation counting to determine the fraction of DTPs that bound SA and eluted with biotin. The *Strep*-tag II (24) template encoded the peptide sequence MSNWSHPQFEKNWSGGCGTGY. The nonselected clone in which two HPQ motifs (separated by 19 amino acids) were introduced encoded the sequence MDEAHPQAGPVDQA-DARLVQQGALQHHPQGDRMMSGGCKLGTGY (the underlined portions are identical to the HPQ regions of clone SB2).

Quantitative analysis of binding was performed by an electrophoretic mobility shift assay (EMSA). DTPs were incubated with varying amounts of SA (Pierce Immunopure Streptavidin) in SA-binding buffer plus 5% glycerol (to increase the density of the solution). After incubating at room temperature for 20 min, the reactions were incubated at 4°C before being loaded onto a 10% polyacrylamide gel containing 2× 90 mM Tris/64.6 mM boric

BIOCHEMISTRY

Alamar Exhibit 1002
Alamar v. Olink

Table 1. Sequences of the 20 clones from the seventh round of selection

Name	No.	Sequence
SB1	3	MDEKTHCTISMNGAVPLVPHH**HPQ**GDPLRLLHRPQPALLVR**HPQ**GDLVALVEHHEGVDRGLVALPELHAEELGEPVGDLVQGPVEQVQGVVDALVWRLPPS
SB2	2	MDEKTHCFHPGDHLVRLVEELQALAEGLQRQGGRQPHRLPRRRPHHLQLLLDEA**HPQ**AGPLRERAHQVDGRLLLQH**HPQ**GDRLLQQPQDHPLELVWRLPPS
SB3	4	MTRRPTASSSSCVRHLLLRQGEHGHQALEDRDKARHVRLVEGDVEVLGGLDRLARARHEAL**HPQ**AGLVHLPLHGGDLGGHLRLVLEA**HPQ**GDRLGLAVHHH
SB4	1	MDEKTHWGISTWRGEPLLH**HPQ**AGRLPLDRRRARHRRILGAEPGGVDHGLRLELLDDHRPLVPDH**HPQ**RGPLQRGDLPQVVPLVRLRHAHVLGLGLAAATIT
SB5	3	MDEKTHWVNVYH**PQ**GDLLVRGHGHDVEALHDQGLHQLDLLVGPPPEVVRALRGEVLGGLRRLVPLDH**PQ**GEALDQARQRPQHLLELHHRALPPALVWRLPPS
SB6	1	MDEKTHWLNNFEELLARLDGLREGEDHPLVLRH**HPQ**GDGLLDQPLGRHRALDGEVREGDRPLDQGGEEDLGALVDDDGEVLDGLVHVGVHVHDPLVCGCHHH
SB7	1	MDEKTHWFGTLNSFPTHWMSAVGNGKIDCSFNMNLSLNHWLSSGHPDGALDDQL**HPQ**GDALVGRDDGVVQALRLEGQHQHRRLAQRRADRHRQLVWRLPPS
SB8	1	MDEKTHCTIELNFSFTHWKLHH**HPQ**GDALLDDGVRPHHPLADEGGGLDQGLGHRRGVVAERLARRDPEVLEGLVERHRGLVPRLRHGGERHAEPLVWRLPPS
SB9	1	MDEKTHCNTGLYDGAADCFNELNKDVAPLVEGRHDLVEGLLLER**HPQ**GDPLVAHRQLVHHPLLGRGERHRRALVPQQEHQPHRLQPVVDLGRRRLVWRLPPS
SB10	1	MDEKTHWHERAQELVGGLLLHD**HPQ**RLLLEPRGPRPLRGLVHERGHQPQPLAGRVEEADGGLLRDGGGELEPLVREGEDHLEPLDDELDAGPRGLVWRLPHHH
SB11	1	MDEKTHWHERVHHLADGLEQ**HPQ**GQRRPLVERHRQVPRGLVRELQHEGLPLEHPAGVHVIRLHQGDDRDVDGLVDGHGRDVRGLEREVGDGPHRLVWRLPPS
SB12	4	MDKDPLLEELEELRERLVH**PQ**GGLLPLRGQVGHDAERLGAEVDDLRGGLLDEPQRAVAGLHHVPHRVGQRLVHEVRELDEGLLDQRDDLRQRLVWRLPPS
SB13	2	MEREDPLDEQLRELREALVD**HPQ**GGAQALHRHDGGEHVPLRRVQHRLQPGLQHHLEPQPLGLLGELQARLQPLAGEHEGDGAGLQRVPGHQGRRLVWRLPPS
SB14	1	MDEKTHRTLSVSLSFNDWLGQTKACWRLVEGLHG**HPQ**GLVREHEVDVLPLAEEVQQVVGGLADGVEQPGGGLLHRAQRVDHPLPDHAGQVLGRLVWRLPPS
SB15	1	MDEKTHWLEDLKGVLKDCLKDLMDFTKDCRSPRVQPQPLLHHDRGEPVPLLREAGRDLGGLGPRAPRQARPLHHGRHDLHEPLVLQD**HPQ**GGPLVCGCHHH
SB16	1	MDEKTHWVLQL**HPQ**GDRLGPRHGGDDVRLVGQGEGVLEGLDGRPRRRRHRLPREDEHRVRALVDQVRDLAERLVEEVDGGVEALRHLGLPQDEPRSGGCHHH
SB17	2	MDEKTHWVGDLQEPLGPLHGGVGEVPGGLVLRH**HPQ**RDRLVDGVGPHGRALARRPHRVVEGLHHLLQRGGERLPPDGPRQLGLLGGELDRADPALVWRLPPS
SB18	1	MDEKTHCAVNVNVGLTHWCHRVAHLQPLDP**HPQ**GDHLRLEPLGHALVDPLVQGVEEVVRPLQLDVGVQRVALVEQVAEVGEGLDHEAGQAHGALVWRLPPS
SB19	1	MDEKTTGWRGGHVVEGLAGELEQLRARLEH**HPQ**GQREPLVQEVEDVDEGLVQDLHGVVAGLLDPVEKLLTDWFKKFKNVSKDCKMTFYLEMYDWSGGCHHH
SB20	1	MNEKTHCKLNFKVNIADWLAEFHGGGQGLLGRRDGVVQRLVDGVQERVERLDRDPGLGDLRLELHHRDHRLRLGGEHLLRDHPLEPDDHLVVGGLVWRLPPS

The "No." indicates the number of times each sequence was observed. The HPQ sequence is in bold type. Defined sequences at the termini are underlined. The six C-terminal residues are not shown.

acid/2.5 mM EDTA (TBE), 0.1% Triton X-100, and 5% glycerol. The gel (prerun for 30 min) and running buffer were precooled to 4°C and run in the cold room (the temperature of the gel was about 20°C) for 45–120 min, depending on the mobility of the particular DTP. The gel was then fixed in 10% acetic acid, 10% methanol for 15 min, transferred to electrophoresis paper (Ahlstrom, Mt. Holly Springs, PA), dried, and analyzed with a PhosphorImager (Molecular Dynamics). The concentration of DTPs was <1 nM in each titration, and thus the K_D can be approximated by the concentration of SA that results in half of the DTPs being mobility-shifted. To determine the K_D, several different measurements were taken in the range of 25–75% of DTPs bound (values outside of this range were unreliable because of background and close proximity of the bound and unbound bands in the gel). The K_D was determined by using the equation $K_D = [SA]^*R$, where R is the ratio of unbound to bound DTPs (ratio of unshifted to shifted band). Independent measurements on gels prepared at different times were used for each clone (the number of different measurements, n, is shown in Table 2). SA concentrations were measured by UV_{282}, by using the molar extinction coefficient of 57,000 per monomer (25). Deletion constructs of clone SB19 were generated by amplification with nested PCR primers.

For the SA-coated plate-based binding assay, the following [^{35}S]methionine labeled protein was generated by *in vitro* translation: MDYKDDDDKMDEKTTGWRGGHVVEGLAG-ELEQLRARLEHHPQGQREPGSHHHHHHMGM. After translation, it was purified under native conditions on FLAG M2 agarose and Ni-NTA agarose as described (22), and then dialyzed into SA-binding buffer. Radiolabeled peptide (100 pM) was then incubated with a range of SA concentrations spanning 10 pM to 10 μM, or with no SA. After incubation at 4°C for 1 hour, these samples were transferred to a Reacti-Bind Streptavidin High Binding Capacity Coated 96-well Plate (Pierce) and incubated at 4°C for 5 min. The supernatant was then counted in a scintillation counter. During this brief incubation, 22–24% of the counts bound to the plate (in the absence of competing SA). This experiment was repeated in duplicate and gave a K_D of 2.4 nM each time. All binding was inhibited by either 10 μM SA or 1 mM biotin.

Results and Discussion

Generation of the Peptide Library. A DNA library encoding polypeptides of length 108 was synthesized as described (22). The library consists of short cassettes (each encoding 11 amino acids) concatamerized together. Forty-four percent of the incorporated cassettes encode a random peptide with a pattern of polar versus nonpolar amino acid side chains compatible with forming an amphipathic α-helix; 45% were similarly patterned to form β-strands; 11% were unpatterned random sequence (22). After concatamerization, the random region is 88 amino acids long, followed by a C-terminal invariant region containing a hexahistidine tag (Fig. 1).

The library was transcribed by T7 RNA polymerase (Fig. 1), after which a "linker" oligonucleotide was added to the 3′ end by using T4 DNA ligase (9, 22). The linker consists of an oligo-dA stretch, followed by 3 triethylene-glycol-phosphate units, followed by the sequence dA-dC-dC-puromycin (9). This puromycin-terminated mRNA was translated *in vitro*, yielding 1.2×10^{14} polypeptides linked via the puromycin moiety to their encoding mRNAs. These mRNA-displayed peptides were then purified on oligo-dT cellulose and Ni-NTA agarose, and the mRNA portion was then reverse transcribed, resulting in a library consisting of 6.7×10^{12} different displayed polypeptides.

Selection of SA-Binding Peptides. We chose SA as a protein target for selection because it does not normally bind to peptides, and because it has been the target of several previous phage display selections (14, 24, 26–31). The mRNA display library was incubated with immobilized SA, washed, and then eluted with biotin (Fig. 1). These molecules were then amplified by PCR to complete the first round of selection. Six subsequent rounds were carried out, and the fraction of the library that bound and eluted from the SA column increased in each round, reaching 61% at round seven (Fig. 2*A*).

The eluate from the seventh round of selection was amplified by PCR. The resulting PCR DNA was used to synthesize a library of displayed peptides. Treatment of the library with RNase A did not reduce the extent of binding/elution from the matrix (Fig. 2*B*), confirming that the peptides, rather than the RNA portion of the library constructs, were responsible for the interaction

Alamar Exhibit 1002
Alamar v. Olink

Table 2. Binding characteristics of selected peptides

Peptide	Structure	% binding and eluting	K_d, nM	Standard deviation (*n*)
Strep-tag II	SNWSHPQFEK	0.04	13,000–72,000	
Nonselected	HPQ 19 HPQ	0.16		
Two HPQ motifs				
SB1	HPQ 19 HPQ	86	50	5.7 (4)
SB2	HPQ 19 HPQ	48	4.8	0.91 (8)
SB3	HPQ 23 HPQ	20		
SB4	HPQ 43 HPQ	49		
SB5	HPQ 52 HPQ	72	110	22 (6)
One HPQ and one similar tripeptide motif				
SB6	HPL 4 HPQ	49		
SB7	HPD 7 HPQ	28		
SB8	HPQ 12 HPL	27		
SB9	HPQ 12 HPL	64		
SB10	HPQ 21 QPQ	15		
SB11	HPQ 28 HPA	68		
SB12	HPQ 30 EPQ	73		
SB13	HPQ 32 EPQ	64		
SB14	HPQ 43 HPL	11		
SB15	QPQ 50 HPQ	44	92	16 (4)
SB16	HPQ 74 LPQ	50		
One HPQ motif				
SB17		8.3		
SB18		58		
SB19		85	10	1.8 (10)
SB19—C4		88	4.9	0.88 (10)
SB19—C4-FLAG			2.4	0.1 (2)
No HPQ motif				
SB20	HPL	34		

First column: clone name (SB1–SB20). For comparison, the *Strep*-tag II peptide (24) and a nonselected sequence with 2 HPQ motifs spaced by 19 residues are also shown. SB-19–C4 is a deletion mutant of peptide SB19 (see text). The clones are grouped according to the number of HPQ and similar tripeptide motifs they possess. Second column: the tripeptide motifs in each clone and the number of amino acid residues separating them. Third column: percentage of peptide binding and specifically eluting from the SA column. Fourth column: K_D, when known, for the interaction between SA and the peptides, as measured in the EMSA (except for the *Strep*-tag II sequence; values from refs. 24 and 33); the K_D for SB10–C4-FLAG was measured by the competition assay; Fig. 4). The standard deviation in the K_D is shown in the fifth column, based on the number of independent measurements (*n*, shown in parentheses).

with SA. Biotin-saturated SA showed no binding to the peptide library (Fig. 2*B*), demonstrating that the interaction of the selected peptides with the SA matrix is specific for the unliganded protein, rather than for any other component of the matrix.

Sequence Analysis of Selected Peptides. Thirty-three randomly chosen clones from the PCR DNA from round seven were chosen for sequencing. Twenty different sequences were observed (Table 1). Surprisingly, all 20 sequences were frame-shifted from the intended frame (frame 1) to frame 3 (achieved by deleting two or adding one nucleotide). The designed pattern of polar and nonpolar residues was therefore discarded, leaving an essentially random sequence. Before the selection, about half of the library members were in frame 1 throughout their entire open reading frames (22). Frame 3 appears to have been enriched over frame 1 because of the increased frequency of the sequence HPQ, which is known to bind SA with micromolar affinity (24, 32). Frame 1 has a low incidence (1:45,000 library members) of the sequence HPQ because of the designed polar/nonpolar pattern. By contrast, frame 3 has a much higher expected incidence of the HPQ sequence (1:64), similar in frequency to that of a library of the same length and with equal mixtures of all four nucleotides at each position (1:193). Frame 2 has a high incidence of stop codons.

Nineteen of the 20 clones have at least one HPQ motif, and five clones contain two such motifs (Table 1). The number of amino acids between the two motifs, when present, ranged from 4 to 74 (Table 2). In addition to the HPQ sequence, some of the flanking positions show a nonrandom composition, but this may simply reflect the patterned structure of the library, which limits the subset of available residues at each position relative to the HPQ core.

Binding Affinities of Peptides. The highest reported affinity for a peptide with an internal HPQ sequence under reducing conditions is in the micromolar range (27, 31–33). We compared our selected clones to that of the decapeptide SNWSHPQFEK ("*Strep*-tag II"), which binds SA with a K_D of 13–72 μM (24, 33). To rapidly screen each of the 20 clones, a method for preparing, tagging, and purifying the peptides was used. First, the mRNA for individual clones was photocrosslinked to a puromycin-containing DNA oligonucleotide (23); these templates were translated *in vitro* with added [^{35}S]methionine and then purified on oligo-dT resin, as above. RNase A was then added to degrade the mRNA portion, leaving peptides fused to a short DNA oligonucleotide. These DTPs were then assayed for binding to immobilized SA and compared with the *Strep*-tag II sequence prepared in the same way.

The results of this analysis are shown in Table 2. All 20 of the selected peptides bound to a much higher degree (200- to 2,200-fold) than the *Strep*-tag II peptide. For another comparison, we introduced 2 HPQ motifs, separated by 19 residues, into an unselected member of the library. It bound only 4-fold better than the single *Strep*-tag II sequence, showing that the presence of 2 HPQ motifs is not sufficient for high-affinity binding.

We measured the dissociation constants of the selected peptides for SA by using an EMSA. The short DNA oligonucleotide tag on the DTPs allows them to migrate in a native gel, and the addition

BIOCHEMISTRY

Alamar Exhibit 1002
Alamar v. Olink

Best Available Copy

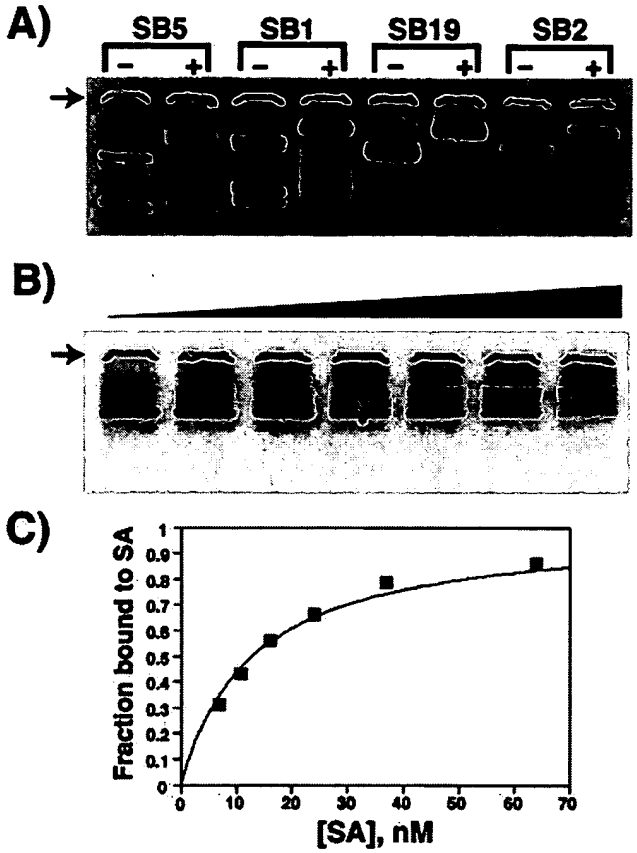


Fig. 3. EMSA analysis of peptide-SA interactions. (*A*) Qualitative demonstration of the binding of four different DNA-tagged peptides to SA. The migration of each clone is shown in the absence (−) and presence (+) of 1 μM SA. Some of the clones show multiple bands, presumably representing different conformations. The arrows show the position of the gel well, which often contains a fraction of the counts. (*B*) Titration of clone SB19 (full-length) with SA. The SA concentration in each lane, from left to right, is: 3.8, 6.6, 10, 15, 23, 35, and 61 nM. (*C*) Curve fitted to the data shown in *B* (the fraction of peptide bound could not be accurately determined for the point with the lowest concentration of SA). Assuming that the peptide is homogeneous and 100% active, the data from this experiment give a K_D of 11 nM (K_D = 10 +/− 1.8 nM from multiple measurements; see Table 2).

of unlabeled SA causes a mobility shift for several of the clones. Examples of these shifts are shown in Fig. 3*A*. Other clones showed either no shift or poorly defined bands, suggesting that the lifetime of these complexes in the gel was too short for detection by this method. We then chose five clones and quantitatively examined their mobility shifts in response to a range of SA concentrations. An example of an SA titration for clone SB19 is shown in Fig. 3 *B* and *C*. The dissociation constants for the clones range from over 100 nM to as low as 5 nM (Table 2). These affinities are comparable to those for monoclonal antibody–antigen interactions.

Dissection of Clone SB19. We further characterized clone SB19, which has a single HPQ motif. A series of deletion constructs were assayed in the SA column-binding assay (Table 3). Construct C4, which retains only the first 38 residues from the selected construct (plus the C-terminal invariant sequence SGGCKLG), retains full binding activity in this assay. Mutating the HPQ motif to HGA reduced the extent of binding by three orders of magnitude (compare construct C4 to M1). Comparison of the N-terminal deletion constructs (N1-N3) suggests that binding determinants are spread throughout the N-terminal 38 residues of the peptide. Construct C4 is therefore the minimal peptide retaining full activity in this assay. EMSA of construct C4 (not shown) confirmed high-affinity SA binding. A fraction (13%) of the peptide was inactive even at SA concentrations >1 μM, but the majority (87%) has an apparent K_D of 4.9 nM.

To demonstrate that this high-affinity binding did not require the DNA tag, we generated the free C4 peptide with a FLAG tag and purified it on a FLAG-affinity matrix. This peptide bound to SA immobilized on a microtiter plate. We measured the binding affinity by competition with varying concentrations of free SA (Fig. 4; K_D = 2.4 +/− 0.1 nM).

Conclusions

SA has been used as a model target protein for numerous phage display selections from libraries of linear peptides ranging in length from 5 to 38 amino acids (24, 26–29, 31). The primary isolates from such selections have dissociation constants of >100 μM, and the HPQ consensus peptide that emerged from these selections was subsequently optimized to the 10-aa-long "*Strep*-tag II" sequence (24). The K_D of this peptide for SA has been measured at 72 μM (24) or 13 μM (33), depending on the method. By contrast, using the ultrahigh complexity libraries accessible by mRNA display, we obtained primary isolates with binding affinities as low as 5 nM, demonstrating that even random nonconstrained peptide libraries can be a source of avid ligands to proteins that do not normally function in peptide binding. The library used in this study differed from previous phage display libraries in its sequence complexity (>10^4-fold more sequences sampled), the length of the random displayed peptides (88 vs. 5–38 amino acids), and an increased incidence of HPQ sequences (about 3-fold per unit length) because of amino acid composition bias. Library diversity is the most dramatic of these differences and is probably the dominant factor

Table 3. Deletion analysis of clone SB19

Type	Sequence	% binding
FL	MDEKTTGWRGGHVVEGLAGELEQLRARLEH**HPQ**GQREPLVQEVEDVDEGLVQDLHGVVAGLLDPVEKLLTDWFKKFKNVSKDCKMTFYLEMYDWSGGCKLG	85
C1	MDEKTTGWRGGHVVEGLAGELEQLRARLEH**HPQ**GQREPLVQEVEDVDEGLVQDLHGVVAGLLDPVEKLLTDWFKKFKNVS MMSGGCKLG	87
C2	MDEKTTGWRGGHVVEGLAGELEQLRARLEH**HPQ**GQREPLVQEVEDVDEGLVQDLHGVVAGLLDPVE MMSGGCKLG	88
C3	MDEKTTGWRGGHVVEGLAGELEQLRARLEH**HPQ**GQREPLVQEVEDVDEGLVQ MMSGGCKLG	89
C4	MDEKTTGWRGGHVVEGLAGELEQLRARLEH**HPQ**GQREP MMSGGCKLG	88
M1	MDEKTTGWRGGHVVEGLAGELEQLRARLEH**HGA**GQREP MMSGGCKLG	0.065
N1	MD GHVVEGLAGELEQLRARLEH**HPQ**GQREP MMSGGCKLG	69
N2	MD EGLAGELEQLRARLEH**HPQ**GQREPLVQEVEDVDEGLVQDLHGVVAGLLDPVEKLLTDWFKKFKNVSKDCKMTFYLEMYDWSGGCKLG	30
N3	M ELEQLRARLEH**HPQ**GQREP MMSGGCKLG	0.058

The full-length (FL), C-terminal deleted (C1–C4), N-terminal deleted (N1–N3), and point mutated (M1) peptide sequences are shown. The "% binding" refers to the performance of these peptides in the SA column-binding assay.

Alamar Exhibit 1002
Alamar v. Olink

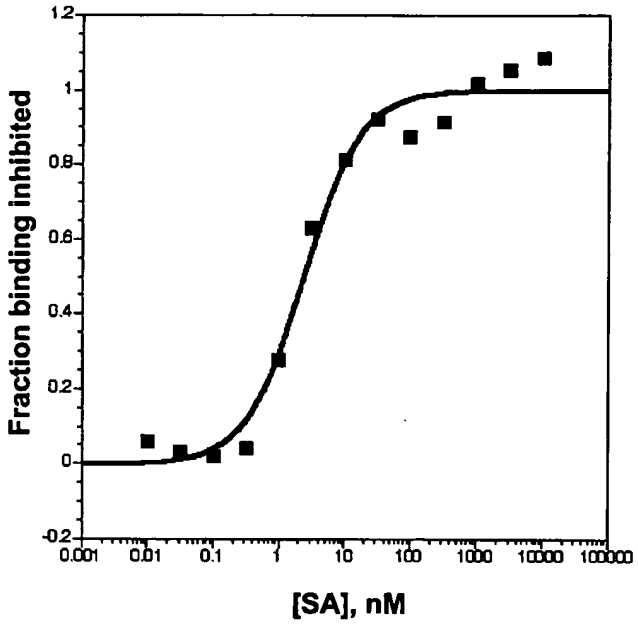


Fig. 4. Binding of peptide SB19-C4-FLAG to immobilized streptavidin. [^{35}S]methionine labeled SB-19 was incubated with a range of concentrations of SA for 1 hour before being transferred to an SA-coated plate and incubated for 5 min. The *y* axis shows the fraction of surface-bound peptide that is competed by the free SA. This experiment was repeated in duplicate and gave a K_D of 2.4 nM each time.

contributing to the isolation of higher-affinity peptides. However, the use of longer peptides may also have been critical, because our 38-aa peptide could only have been isolated from the 38-aa phage display library and not from the other shorter libraries. Our use of longer peptides may also have effectively increased library diversity by another 50-fold, because a given 38-aa peptide could occur in any of 50 registers in the 88-aa random region.

The affinities of the SA-binding peptides could presumably be further improved by evolutionary optimization. The affinity of simple peptide ligands can also be dramatically increased by introducing disulfide constraints (14, 30) or by constraining them as loops within known structural scaffolds (12). By combining these library design features with the high-complexity libraries that can be generated by mRNA display, it may be possible to select subnanomolar binding peptides to proteins.

Structures of HPQ-containing peptides bound to SA have been determined by x-ray crystallography (24, 31, 32, 34). The HPQ motif inserts itself into the biotin-binding cleft and forms hydrogen bonds and hydrophobic interactions with SA (24, 34). The high affinity of the HPQ-containing clones isolated in the present study could result from either of two different factors. First, the flanking amino acids could stabilize the active conformation of the HPQ tripeptide motif; this type of mechanism appears to be responsible for the high-affinity binding of a disulfide-constrained HPQ-containing peptide to SA (K_D = 230 nM; refs. 30, 34). Second, several distinct peptide elements could be interacting with different surfaces on SA, providing additional binding energy from either specific or nonspecific interactions.

These SA-binding peptides may provide useful affinity tags for protein purification. The *Strep*-tag II sequence (SNWSHPQFEK), which binds to SA with a K_D of $\approx$13 μM, has been used for this purpose. Effective column retention requires the use of an SA mutant with a 13-fold higher affinity for the *Strep*-tag II sequence (33, 35). The 38-aa SB19-C4 peptide identified in the present study allows for a one-step purification of proteins from *Escherichia coli* extract by using normal SA and results in a high yield of very pure fusion protein (data not shown).

Complete expression and activity analysis of the proteome will require the use of hundreds of thousands of protein-binding and inhibiting reagents. Traditional methods for generating monoclonal antibodies are expensive and time-consuming; *in vitro* selection of binding peptides from polypeptide libraries may help to meet this need. This study shows that, given sufficient library complexity, even random peptide libraries can be used to generate ligands with strong binding constants (K_D = 5 nM). It should be possible to increase the affinities of peptides for their targets by constraining them as loops in known protein scaffolds (12). Such constrained polypeptides may also have superior properties with respect to solubility and stability against proteolysis.

We thank members of the Szostak lab, especially R. Liu, G. Cho, G. Short, J. Urbach, and J. Pollard for advice, and P. Svec for sequencing. Funding was provided by the Cancer Research Fund of the Damon Runyon–Walter Winchell Foundation, Hoechst AG, the National Aeronautics and Space Administration Astrobiology Institute, and the National Institutes of Health.

1. Colas, P. (2000) *Curr. Opin. Chem. Biol.* **4**, 54–59.
2. Smith, G. P. & Petrenko, V. A. (1997) *Chem. Rev.* **97**, 391–410.
3. Perelson, A. S. & Oster, G. F. (1979) *J. Theor. Biol.* **81**, 645–670.
4. Vaughan, T. J., Williams, A. J., Pritchard, K., Osbourn, J. K., Pope, A. R., Earnshaw, J. C., McCafferty, J., Hodits, R. A., Wilton, J. & Johnson, K. S. (1996) *Nat. Biotechnol.* **14**, 309–314.
5. Hanes, J. & Pluckthun, A. (1997) *Proc. Natl. Acad. Sci. USA* **94**, 4937–4942.
6. Hanes, J., Jermutus, L., Weber-Bornhauser, S., Bosshard, H. R. & Pluckthun, A. (1998) *Proc. Natl. Acad. Sci. USA* **95**, 14130–14135.
7. Roberts, R. W. & Szostak, J. W. (1997) *Proc. Natl. Acad. Sci. USA* **94**, 12297–12302.
8. Roberts, R. W. & Ja, W. W. (1999) *Curr. Opin. Struct. Biol.* **9**, 521–529.
9. Liu, R., Barrick, J. E., Szostak, J. W. & Roberts, R. W. (2000) *Methods Enzymol.* **318**, 268–293.
10. Clackton, T. & Wells, J. A. (1994) *Trends Biotechnol.* **12**, 173–184.
11. Katz, B. A. (1997) *Annu. Rev. Biophys. Biomol. Struct.* **26**, 27–45.
12. Nygren, P. A. & Uhlen, M. (1997) *Curr. Opin. Struct. Biol.* **7**, 463–469.
13. O'Neil, K. T., Hoess R.H., Jackson, S. A., Ramachandran, N. S., Mousa, S. A. & DeGrado, W. F. (1992) *Proteins* **14**, 509–515.
14. McLafferty, M. A., Kent, R. B., Ladner, R. C. & Markland, W. (1993) *Gene* **128**, 29–36.
15. DeLano, W. L., Ultsch, M. H., de Vos, A. M. & Wells, J. A. (2000) *Science* **287**, 1279–1283.
16. Clackson, T., Hoogenboom, H. R., Griffiths, A. D. & Winter, G. (1991) *Nature (London)* **352**, 624–628.
17. Persson, M. A., Caothien, R. H. & Burton, D. R. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 2432–2436.
18. Marks, J. D., Hoogenboom, H. R., Griffiths, A. D. & Winter, G. (1992) *J. Biol. Chem.* **257**, 16007–16010.
19. Orum, H., Andersen, P. S., Oster, A., Johansen, L. K., Riise, E., Bjornvad, M., Svendsen, I. & Engberg J. (1993) *Nucleic Acids Res.* **21**, 4491–4498.
20. Sheets, M. D., Amersdorfer, P., Finnern, R., Sargent, P., Lindqvist, E., Schier, R., Hemingsen, G., Wong, C., Gerhart, J. C. & Marks, J. D. (1998) *Proc. Natl. Acad. Sci. USA* **95**, 6157–6162.
21. de Haard, H. J., van Neer, N., Reurs, A., Hufton, S. E., Roovers, R. C., Henderikx, P., de Bruine, A. P., Arends, J. W. & Hoogenboom, H. R. (1999) *J. Biol. Chem.* **274**, 18218–18230.
22. Cho, G., Keefe, A. D., Wilson, D. S., Liu, R. & Szostak, J. W. (2000) *J. Mol. Biol.* **297**, 309–319.
23. Kurz, M., Gu, K. & Lohse, P. A. (2000) *Nucleic Acids Res.* **28**, E83.
24. Schmidt, T. G. M., Koepke, J., Frank, R. & Skerra, A. (1996) *J. Mol. Biol.* **255**, 753–766.
25. Green, N. M. (1970) *Methods Enzymol.* **18**, 418–424.
26. Devlin, J. J., Panganiban, L. C. & Devlin, P. E. (1990) *Science* **249**, 404–406.
27. Lam, K. S., Salmon, S. E., Hersch, E. M., Hruby, V. J., Kazmierski, W. M. & Knapp, R. J. (1991) *Nature (London)* **354**, 82–84.
28. Kay, B. K., Adey, N. B., He, Y.-S., Manfredi, J. P., Mataragnon, A. H. & Fowlkes, D. M. (1993) *Gene* **128**, 59–65.
29. Schmidt, T. G. M. & Skerra, A. (1993) *Protein Eng.* **6**, 109–122.
30. Giebel, L. B., Cass, R. T., Milligan, D. L., Young, D. C., Arze, R. & Johnson, C. R. (1995) *Biochem. J.* **34**, 15430–15435.
31. Katz, B. A. (1999) *Biomol. Eng.* **16**, 57–65.
32. Weber, P. C., Pantoliano, M. W. & Thompson, L. D. (1992) *Biochem. J.* **31**, 9350–9354.
33. Voss, S. & Skerra, A. (1997) *Protein Eng.* **10**, 975–982.
34. Katz, B. A. (1995) *Biochem. J.* **34**, 15421–15429.
35. Maier, T., Drapal, N., Thanbichler, M. & Bock, A. (1998) *Anal. Biochem.* **259**, 68–73.
36. Sleat, D. E., Gallie, D. R., Jefferson, R. A., Bevan, M. W., Turner, P. C. & Wilson, T. M. (1987) *Gene* **60**, 217–225.

BIOCHEMISTRY

Alamar Exhibit 1002
Alamar v. Olink

Stereoselective Synthesis of over Two Million Compounds Having Structural Features Both Reminiscent of Natural Products and Compatible with Miniaturized Cell-Based Assays

Derek S. Tan, Michael A. Foley, Matthew D. Shair, and Stuart L. Schreiber*

Department of Chemistry & Chemical Biology
Harvard Institute of Chemistry & Cell Biology
Howard Hughes Medical Institute
Harvard University, 12 Oxford St.
Cambridge, Massachusetts 02138

Received May 19, 1998

Understanding the cellular function of a protein usually requires a means to alter the function. This is commonly done by mutating the gene encoding the protein (genetic approach). It can also be done by binding the protein directly with a small molecule ligand (chemical genetic approach). Such ligands, primarily natural products or synthetic variants of them, can either inactivate[1] or activate[2] protein function. For the chemical genetic approach to have its maximal impact, efficient methods of ligand discovery will be required to provide, in the limit, a small molecule partner for every gene product.

Our laboratory has recently described miniaturized cell culture assays for screening large numbers of small molecule ligands, potentially on a genome-wide basis.[3] However, despite its success in natural products synthesis, synthetic chemistry has not yet been applied to the synthesis of vast numbers of compounds with structures both reminiscent of natural products and compatible with miniaturized assays. These features of a synthesis will likely be required in order to discover nonnatural compounds having the binding affinities and specificities characteristic of natural products.

The synthetic strategy we have undertaken is to develop highly efficient multistep syntheses of natural product-like compounds that include several coupling steps and to use split-pool techniques[4] at these steps in order to generate diverse outcomes. This technique requires that one of the coupling substrates be immobilized on a solid support. Our syntheses are further constrained by the water-compatible photolabile supports required when carrying out large numbers ($>10^5$) of miniaturized "nanodroplet" cell culture assays aimed at identifying cell permeable ligands.[3] This presents a considerable synthetic challenge since the structural elements that provide these properties (e.g., poly(ethylene glycol), nitrobenzyl moieties) are incompatible with numerous reagents used in conventional synthetic chemistry.

With these considerations in mind, we first converted shikimic acid, **1**, into both enantiomers of epoxycyclohexenol carboxylic acid **2**,[5] which were then coupled to a photocleavable linker on solid support by standard methods (Figure 1).[6] Treatment of the resin-bound epoxycyclohexenol, **3**, with various nitrone carboxylic acids,[7] **7a−f** and **9b−d**, under esterification conditions yielded tetracyclic compounds **4a−f** and **5b−d** with complete regio- and

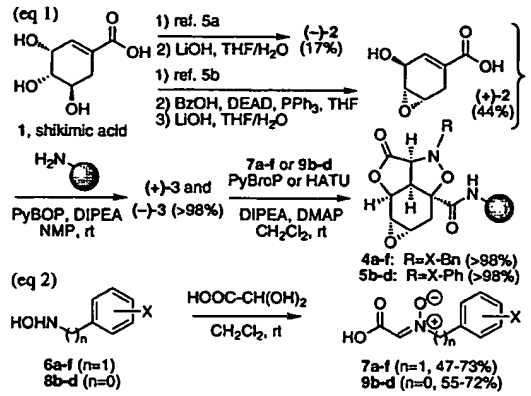


Figure 1. Synthesis of tetracycles and nitrones. a: X = H, b: X = 2-I, c: X = 3-I, d: X = 4-I, e: X = 4-CF_3, f: X = 3,4-$(OMe)_2$.

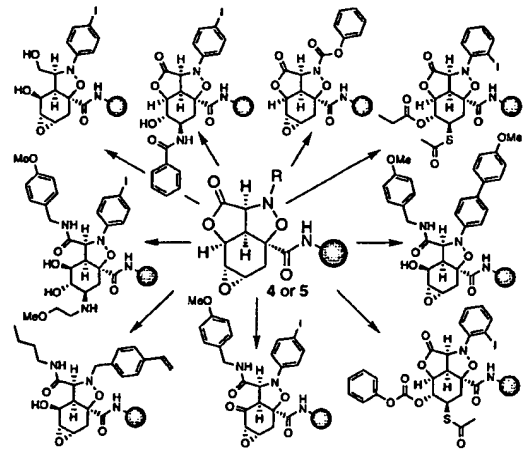


Figure 2. Compounds derived from templates **4** and **5**.

stereoselectivity, presumably via tandem acylation/1,3-dipolar cycloaddition.[8] No intermediate nitrone ester or carboxy isoxazolidine structures were observed. The iodoaryl tetracycles **4b−d** and **5b−d** were selected for further study since the support-bound aryl iodides could be modified with commercially available reagents, avoiding the need for solution-phase syntheses of numerous nitrone acids.

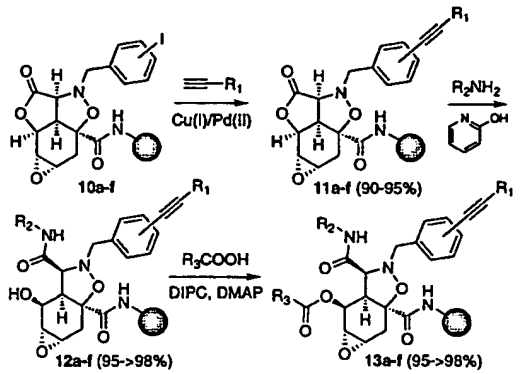
Tetracycles **4b−d** and **5b−d** are rigid, densely functionalized compounds that can undergo further reactions to introduce a variety of functional groups around the central octahydrobenzisoxazole structure, notably, without the use of protecting groups. The iodoaryl groups can serve as substrates for palladium cross-coupling reactions. The electrophilic lactone and epoxide can react with nucleophiles while simultaneously unmasking alcohols for subsequent reactions. Furthermore, reductive N−O bond cleavage would provide two additional handles for functionalization. We developed several such reactions (Figure 2) with product purity ranging from ~50 to ≥98% following photocleavage.[9] Since purification is not practical in a large split-pool synthesis, only reactions generating products in ≥90% purity were acceptable.

The most promising reactions were studied in detail to define the scope, limitations, and optimal conditions for each (Figure 3). Iodobenzyl tetracycles **4b−d** were selected as scaffolds since

(1) See http://www-schreiber.chem.harvard.edu on the Web.
(2) Crabtree, G. R.; Schreiber, S. L. *Trends in Biochem. Sci.* **1996**, *21*, 418−422.
(3) (a) Borchardt, A.; Liberles, S. D.; Biggar, S. R.; Crabtree, G. R.; Schreiber, S. L. *Chem. Biol.* **1997**, *4*, 961−968. (b) You, A. J.; Jackman, R. J.; Whitesides, G. M.; Schreiber, S. L. *Chem. Biol.* **1997**, *4*, 969−975.
(4) (a) Furka, A.; Sebestyén, F.; Asgedom, M.; Dibó, G. *Int. J. Pept. Protein Res.* **1991**, *37*, 487−493. (b) Lam, K. S.; Salmon, S. E.; Hersh, E. M.; Hruby, V. J.; Kazmierski, W. M.; Knapp, R. J. *Nature* **1991**, *354*, 82−84.
(5) (a) Wood, H. B.; Ganem, B. *J. Am. Chem. Soc.* **1990**, *112*, 8907−8909. (b) McGowan, D. A.; Berchtold, G. A. *J. Org. Chem.* **1981**, *46*, 2381−2383. (c) Mitsunobu, O. *Synthesis* **1981**, 1−28.
(6) The resin structure shown represents Tentagel S NH_2, a poly(ethylene glycol)-polystyrene copolymer, loaded with a photocleavable 3-amino-3-*o*-nitrophenylpropionic acid linker. (Brown, B. B.; Wagner, D. S.; Geysen, H. M. *Mol. Div.* **1995**, *1*, 4−12.).
(7) Keirs, D.; Overton, K. *Heterocycles* **1989**, *28*, 841−848.
(8) Tamura, O.; Okabe, T.; Yamaguchi, T.; Gotanda, K.; Noe, K.; Sakamoto, M. *Tetrahedron* **1995**, *51*, 107−118.
(9) Photocleavage products were analyzed by ^1H NMR, FAB-MS, HPLC, and TLC.

S0002-7863(98)01746-6 CCC: $15.00 © 1998 American Chemical Society
Published on Web 08/08/1998

Alamar Exhibit 1002
Alamar v. Olink



with ε-aminocaproic acid spacer
a: R_1=(*ortho*)-*t*-butyl, R_2=4-methoxybenzyl, R_3=isobutyl
b: R_1=(*meta*)-diethoxymethyl, R_2=2,2-dimethoxyethyl, R_3=1-propenyl
c: R_1=(*para*)-*n*-propyl, R_2=4-methoxyphenethyl, R_3=ethyl

without spacer
d: R_1=(*ortho*)-benzyl, R_2=cyclobutyl, R_3=4-methoxybenzyl
e: R_1=(*meta*)-2-propenyl, R_2=allyl, R_3=phenyl
f: R_1=(*para*)-4-chlorophenyl, R_2=2-methoxybenzyl, R_3=isopropyl

Figure 3. Synthesis of 24 demonstration compounds and a 2.18 million compound library.

their palladium-mediated cross-coupling reactions were more efficient than those involving iodophenyl tetracycles **5b–d**. A carefully controlled alkyne coupling reaction was used to convert the aryl iodides (**10**) into aryl alkynes (**11**).[10] Cycloreversion was a persistent problem in uncatalyzed lactone aminolysis; however, 2-hydroxypyridine mediated the efficient reaction of primary amines with the γ-butyrolactone to afford γ-hydroxyamides (**12**).[11] After testing numerous acylating conditions, we found that the unmasked secondary alcohol could be capped reliably with DIPC/DMAP-activated carboxylic acids to give γ-acyloxy amides (**13**).[12]

Split-pool synthesis provides the theoretical means to synthesize the large numbers of molecules likely required for chemical genetic screens, where molecules replace mutations. However, such syntheses present enormous analytical challenges. We have developed a four-step protocol in order to provide maximum confidence that a complex split-pool synthesis of encoded molecules yields the anticipated products in high purity and efficiency. First, to demonstrate the suitability of the entire reaction sequence for library synthesis, we synthesized and fully characterized 24 compounds,[13] **10a–f** through **13a–f**, with the latter providing photocleavage products in acceptable 80–90% purity. Second, potential building blocks at each step were tested by reaction with a single selected substrate. Thus, 50 alkynes were tested by reaction with **10a**, 87 amines with **11a**, and 98 acids with **12a**. Of these, 23 alkynes, 54 amines, and 44 acids reacted with ≥90% conversion and purity (LC-MS). These building blocks, along with a limited number of less optimal candidates (generally ≥70% conversion), were selected for inclusion in library synthesis. Third, to verify effective synthesis in a split-pool format, we generated a small test library with building blocks carefully selected such that the products within each final pool would have unique masses, allowing analysis by LC–MS. Indeed, all of the expected 456 masses were detected.

(10) (a) Sonogashira, K.; Tohda, Y.; Hagihara, N. *Tetrahedron Lett.* **1975**, 4467–4470. (b) Young, J. K.; Nelson, J. C.; More, J. S. *J. Am. Chem. Soc.* **1994**, *116*, 10841–10842. (c) Collini, M. D.; Ellingboe, J. W. *Tetrahedron Lett.* **1997**, *38*, 7963–7966. (d) Odingo, J.; Sharpe, B. A.; Oare, D., Presented at the 213th National Meeting of the American Chemical Society, San Francisco, CA, April, 1997; Paper ORGN 574.

(11) (a) Rony, P. R. *J. Am. Chem. Soc.* **1969**, *91*, 6090–6096. (b) Openshaw, H. T.; Whittaker, N. *J. Chem. Soc. C* **1969**, 89–91.

(12) (a) Atuegbu, A.; Maclean, D.; Nguyen, C.; Gordon, E. M.; Jacobs, J. W. *Bioorg. Med. Chem.* **1996**, *4*, 1097–1106. (b) Hydroxyl groups introduced during the alkyne coupling step were also acylated efficiently under these conditions.

(13) ^{1}H NMR with complete peak assignments by extensive homonuclear decoupling and/or DQF-COSY, TOCSY, and NOESY experiments, FAB-MS, HRMS, HPLC, TLC.

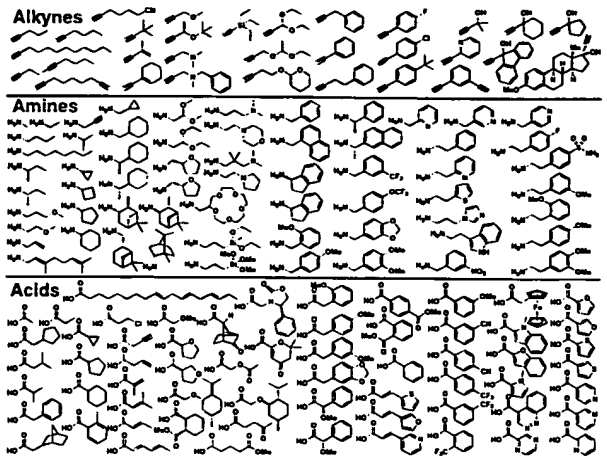


Figure 4. Building blocks used in library synthesis.

Fourth, during the encoded library synthesis below,[14] the effectiveness of carbene tagging was verified by analyzing beads from every pool at each step of the synthesis.

A large encoded library was constructed by split-pool synthesis beginning with attachment of two spacers, ε-aminocaproic acid and glycine, to the photolinker on resin. A third portion of the resin was left without a spacer. The resin was pooled, divided into two portions, and one enantiomer of epoxycyclohexenol carboxylic acid **2** was coupled to each pool. After the resin was pooled and divided into three portions, iodobenzyl nitrone acids **7b–d** were coupled, resulting in a total of 18 tetracyclic scaffolds. The synthesis was completed by reaction with 30 terminal alkynes, then with 62 primary amines, and finally with 62 carboxylic acids (Figures 3 and 4) with an additional skip codon at each step.[15]

This six-step reaction sequence resulted in a collection of compounds calculated to contain 2.18 million distinct, spatially separated, and encoded chemical entities. These synthetic compounds are rigid, stereochemically defined, and structurally diverse, characteristics common to many natural products. Furthermore, the synthesis allows the controlled release of compounds from the individual 90 μ supports into nanodroplets containing engineered cells, features critical to the miniaturized cell-based assays now being used to screen this library for cell permeable, protein-binding ligands.[3] Encouragingly, we have already found that several members of this library activate a reporter gene in mink lung cells.[16] We note, however, that this synthesis relies in part on simple acylation chemistry. Syntheses benefiting from a greater range of mild and selective synthetic methods may facilitate the routine discovery of compounds with protein-binding properties rivaling those of natural products.

Acknowledgment. We thank Dr. A. Tyler, J. Athanasopoulos, and J. Lynch for expert mass spectral support.

Supporting Information Available: Acknowledgments, additional discussion, complete experimental procedures, and analytical data for **2**, **3**, **6b–d**, **7b–d**, **10–13a–f**, building block testing, test library synthesis, large library synthesis, binary encoding (77 pages, print/PDF). See any current masthead page for ordering information and Web access instructions.

JA981746O

(14) Binary encoding was performed by a modified version of the literature procedure (see Supporting Information): Nestler, H. P.; Bartlett, P. A.; Still, W. C. *J. Org. Chem.* **1994**, *59*, 4723–4724.

(15) Combs, A. P.; Kapoor, T. M.; Feng, S.; Chen, J. K.; Daudé-Snow, L. F.; Schreiber, S. L. *J. Am. Chem. Soc.* **1996**, *118*, 287–288.

(16) Compounds with the general structure **13** in Figure 3, with no spacer, R_1 = carbon-based substituents attached to an alkyne substituted at the meta position (see S46 in Supporting Information), R_2 = 3,4-dimethoxybenzyl, and R_3 = methoxymethyl, were found to activate the reporter plasmid p3TPLux in a stably transfected mink lung cell line, both independently and synergistically with TGF-β1 (B. R. Stockwell and S. L. S., unpublished results and *Curr. Biol.* **1998**, *8*, 761–770.).

Alamar Exhibit 1002
Alamar v. Olink

Available online at www.sciencedirect.com
SCIENCE DIRECT®
Bioorganic & Medicinal Chemistry Letters
ELSEVIER
Bioorganic & Medicinal Chemistry Letters 14 (2004) 2407–2410

Site-specific conjugation of oligonucleotides to the C-terminus of recombinant protein by expressed protein ligation

Shuji Takeda, Shinya Tsukiji and Teruyuki Nagamune*

Department of Chemistry and Biotechnology, School of Engineering, The University of Tokyo, Tokyo 113-8656, Japan

Received 24 February 2004; revised 8 March 2004; accepted 9 March 2004

Abstract—We developed a convenient method for synthesizing homogeneous DNA–protein conjugates. The method is based on expressed protein ligation of intein-fusion proteins and oligonucleotides derivatized with a cysteine. A range of cysteinyl oligonucleotides were synthesized by using a new reagent **1** and were successfully applied to expressed protein ligation to attach the oligonucleotides specifically at the C-terminus of a recombinant protein.
© 2004 Elsevier Ltd. All rights reserved.

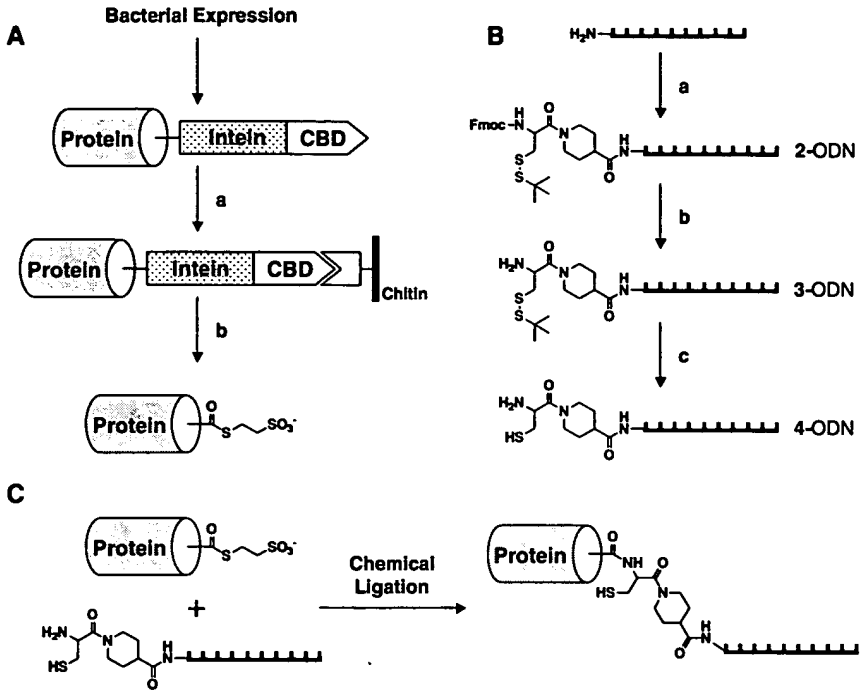
Semisynthetic DNA-protein/enzyme conjugates hold great promise as versatile molecular tools for biotechnology and materials science.[1] By coupling the unique target-specific hybridization properties of DNA molecules with the almost unlimited functionality of proteins, a number of hybrid molecules for a variety of applications have been generated so far. In a pioneering work by Corey and Schultz, an oligonucleotide–*Staphylococcus* nuclease chimera was prepared as a synthetic restriction nuclease.[2] The specific triple-helix formation of the single stranded DNA moiety of the chimera enabled the site-selective cleavage of the plasmid DNA. Recently, Bayley and co-workers constructed a DNA-α-hemolysin nanopore biosensor capable of identifying individual DNA strands with single-base resolution.[3] These and other works[4] clearly demonstrate that the rational introduction of an oligonucleotide binding domain into a protein represents a powerful strategy for the construction of artificial proteins/enzymes, which selectively target (cleave or detect) specific nucleic acids. Such molecules are extremely difficult to design using protein components alone. On the other hand, the accurate and robust hybridization ability of DNA molecules has proven to be useful for the DNA-programmed assembly of proteins into a spatially designed arrangement. This approach has recently attracted considerable attention for the fabrication of protein microarrays,[5] nanometer-sized supramolecular devices,[6] and immunological bioassays.[7]

However, compared to recent significant advances in the total chemical synthesis of oligonucleotide–peptide conjugates by solution-phase[8] or solid-phase synthesis,[9] conjugation strategies for incorporating an oligonucleotide into a larger recombinant protein are still poorly developed. DNA–protein conjugates studied so far have been prepared by three conventional strategies. The first is the direct attachment of a nucleotide oligomer containing a thiopyridyl disulfide to the reactive cysteine residue of a protein via a disulfide exchange reaction.[2,3] Although this strategy can produce a homogeneous protein carrying an oligonucleotide at a specific position, the inherent chemical instability of the disulfide bond against reducing reagents may restrict its broad use. The second strategy uses bifunctional crosslinkers, such as *N*-[γ-maleimidobutyryloxy]succinimide ester (GMBS)[7a] and sulfosuccinimidyl 4-[*p*-maleimidophenyl] butyrate (sulfo-BMPB),[5] to covalently bridge an oligonucleotide and a protein. This approach is applicable to any proteins with no reactive cysteine residue. However, chemical modification of protein surfaces in a nonspecific manner generates conjugates with high heterogeneity. In the third strategy, biotinylated proteins are prepared to produce noncovalent assemblies with biotinylated DNA through the strong interaction with streptavidin.[5,7b] The biotinylation of proteins again relies on random chemical modification. Therefore, development of a novel methodology for efficient and site-selective conjugation of oligonucleotides to proteins

Keywords: Expressed protein ligation; Protein–DNA conjugate; Cysteine modification.
* Corresponding author. Tel.: +81-3-5841-7328; fax: +81-3-5841-8657; e-mail: nagamune@bio.t.u-tokyo.ac.jp

0960-894X/$ - see front matter © 2004 Elsevier Ltd. All rights reserved.
doi:10.1016/j.bmcl.2004.03.023

Alamar Exhibit 1002
Alamar v. Olink



Scheme 1. A. Expression and purification of recombinant proteins containing an α-thioester. (a) Chitin beads equilibrated in buffer A (50 mM Tris–HCl, 500 mM NaCl, 1 mM EDTA, pH 8.5); (b) Thiolysis in buffer B (50 mM Tris–HCl, 500 mM NaCl, 1 mM EDTA, 50 mM MESNA, pH 8.5), 4 °C, 24 h; B. Synthesis of cysteinyl-oligonucleotides. (a) **1** (70–80 equiv) in a borate buffer (10 mM, pH 8.5) containing 50% DMF; (b) aqueous ammonia, 37 °C, 12 h; (c) 2 M DTT aq, rt, 1 h; C. Conjugation of 4-ODNs to proteins at their C-terminus. Buffer B, 4 °C, 48 h.

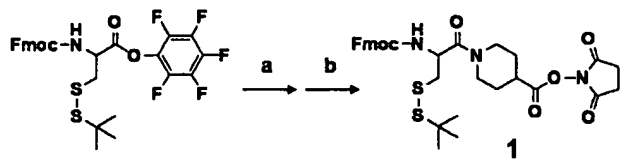
is a challenge that would allow further investigation into the scope of DNA–protein hybrids.

Here we report a facile and efficient method for coupling synthetic oligonucleotides with recombinant proteins specifically at their C-terminus. Our strategy is based on expressed protein ligation (EPL), in which a recombinantly expressed protein containing a C-terminal α-thioester reacts with a synthetic peptide containing an N-terminal cysteine to form a native peptide bond (Scheme 1).[10] This technique has previously been used for the chemical semisynthesis of a variety of proteins.[11]

For chemoselective ligation with C-terminal α-thioesters, oligonucleotides require a cysteine moiety to be incorporated. However, no convenient route for introduction of a cysteine or its analogue at the end of oligonucleotides has been reported until recently.[12] For this purpose, we designed a new reagent **1**, a Fmoc-protected dipeptide carrying active *N*-hydroxysuccinimide (NHS) ester, for solution-phase coupling with amine-functionalized oligonucleotides. Isonipecotic acid was used as a connecter to prevent facile epimerization of the cysteine moiety and rapid hydrolysis of the NHS ester. One advantage of this approach is the simplicity of preparation of the cysteine-appended oligonucleotides, as described below. A second advantage is that, since synthetic 5′- and 3′-amino-oligonucleotides are now available at a low cost, **1** allows recombinant proteins to be conjugated at either end of DNA strands.

Compound **1** was synthesized as shown in Scheme 2 in a straightforward manner from commercially available materials. Coupling of the pentafluorophenyl ester of *N*-α-Fmoc-*S*-*tert*-butylsulfenylcysteine with 1.0 equiv of isonipecotic acid was accomplished under standard conditions. The intermediate was then reacted with NHS using HBTU to yield a white solid **1** in 38% yield for the two steps.[13]

Synthesis of cysteine-modified oligodeoxynucleotides, **4**-ODN, is outlined in Scheme 1B. The reaction mixture of 5′-amino-ODNa (5′-NH_2-$(CH_2)_3$-O-$(CH_2)_2$-pODN, Table 1) and excess (ca. 80 equiv) of **1** was incubated in a borate buffer (10 mM, pH 8.5) containing 50% DMF at room temperature for 10 h, and purified by reversed-phase HPLC (RP-HPLC).[14] RP-HPLC indicates that the coupling reaction takes place cleanly with no significant side reactions except moderate hydrolysis of the NHS ester group of compound **1**. Fmoc protection of the resulting conjugate **2**-ODNa was removed in aqueous ammonia at 37 °C for 12 h, followed by lyophilization, to obtain the conjugate **3**-ODN. Subsequently, **3**-ODNa was treated with 2 M DTT solution at room temperature for 1 h to deprotect the *S*-*tert*-butylsulfenyl



Scheme 2. (a) Isonipecotic acid (1 equiv), DIEA (2 equiv), DMF, rt, 6 h; (b) NHS (1 equiv), HBTU (1.2 equiv), DIEA (1 equiv), DMF, rt, 5 h.

Alamar Exhibit 1002
Alamar v. Olink

Best Available Copy

Table 1. Sequences of ODNs used in this study

ODN	Sequence	Modification
a	AAAAAAAAAAAA	5′
b	AAAAAAAAAAAA	3′
c	TTTTTTTTTTTT	5′
d	TTTTTTTTTTTT	3′
e	CCCCCCCCCCCC	5′
f	CCCCCCCCCCCC	3′
g	TTTCCCAGTCACGACGTTGTA	5′
h	GTGGATAACCGTATTACCGCC	5′
i	TTTTTGATGAATTCGGA	5′
j	TTTTTTCCGAATTCATC	5′
k	TCCTGTGTGAATTTGTTATCCGCT	5′
l	TTTTTTTTTTTTTTCACAGCTGAG-GATAGGACAT	5′

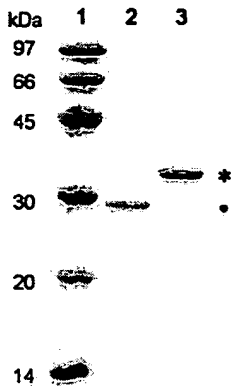


Figure 1. SDS-PAGE analysis of ECFP-ODNc conjugate. Lane 1, molecular weight marker; 2, ECFP-thioester alone; 3, Reaction mixture. The ECFP-thioester and the ligation product are denoted by ● and *, respectively.

group, yielding the desired conjugate **4**-ODNa. These two deprotection steps also produce no by-products as evaluated by RP-HPLC. All products were satisfactorily characterized by MALDI-TOF mass spectrometry.[15] It should be noted that, owing to the clean reaction, highly pure **2**-ODNs and subsequent **3**-ODNs can be easily prepared by purification with size-exclusion chromatography using a NAP-5 column (Amersham Bioscience), in total yields over 60%. With this procedure, several **3**-ODNs with various lengths and sequences were efficiently synthesized (Table 1), indicating the general applicability of this conjugation chemistry.

Recombinant protein α-thioesters can be generated by thiolysis of the corresponding protein–intein fusion where the intein has a mutation to block the final step of protein splicing (Scheme 1A).[16] As a demonstration of our semi-synthetic strategy, we attempted to ligate the **4**-ODNs to the C-terminus of a green fluorescent protein variant, ECFP. The full-length ECFP sequence was cloned into a commercially available intein expression system, pTWIN1 vector (New England Biolabs). Subsequent overexpression in *Escherichia coli* resulted in the generation of a protein chimera in which the desired ECFP fragment was fused at its C-terminus to a *Mycobacterium xenopi* GyrA intein linked to a chitin binding domain (CBD). The recombinant fusion protein was purified by affinity chromatography on chitin beads and was then cleaved using 2-mercaptoethanesulfonic acid (MESNA) to afford the ECFP-thioester with >95% homogeneity as indicated by SDS-PAGE analysis.

Next we investigated EPL between the ECFP-thioester and the **4**-ODNc to give the covalent DNA–protein conjugate (Scheme 1C).[17] Because *S-tert*-butylsulfenyl groups are smoothly released in situ in the ligation buffer containing MESNA, **3**-ODNc was directly used in the following EPL experiments. The ligation reaction was initiated by addition of **3**-ODNc to the ECFP-thioester solution obtained above and was monitored by SDS-PAGE.[18] It was clearly demonstrated that the ligation proceeds to approximately 80% completion after incubation at 4 °C for two days (Fig. 1). MALDI-TOF-MS analysis of the ligation product after agarose gel purification indicated a mass consistent with the expected product.[19] Furthermore, we have successfully applied this strategy to couple ECFP and 3′-cysteinyl-ODN (**b**,**d** and **f**), and the C1 domain of Protein G and 5′-cysteinyl-ODN (**a**, **b** and **l**) (data not shown).

In summary, we demonstrated for the first time the site-specific conjugation of oligonucleotides to recombinant proteins at their C-terminus by expressed protein ligation. We are now applying this method to the synthesis of other DNA–protein hybrids for future applications.

Acknowledgements

We would like to thank Profs. Kazuhiko Saigo and Takeshi Wada for generously allowing us to use their facilities and fruitful discussion in the initial stage of this study. We are grateful to all members of the Saigo Laboratory, particularly Dr. Yasuhiro Ishida for helpful assistance. We thank Tsutomu Tanaka for helpful discussions.

References and notes

1. (a) Niemeyer, C. M. *Chem. Eur. J.* **2001**, *7*, 3189; (b) Niemeyer, C. M. *Trends Biotechnol.* **2002**, *20*, 395.
2. Corey, D. R.; Pei, D.; Schultz, P. G. *J. Am. Chem. Soc.* **1989**, *111*, 8523.
3. Howorka, S.; Cheley, S.; Bayley, H. *Nat. Biotechol.* **2001**, *19*, 636.
4. Saghatelian, A.; Guckian, K. M.; Thayaer, D. A.; Ghadiri, M. R. *J. Am. Chem. Soc.* **2003**, *125*, 344.
5. (a) Niemeyer, C. M.; Sano, T.; Smith, C. L.; Cantor, C. R. *Nucleic Acids Res.* **1994**, *22*, 5530; (b) Niemeyer, C. M.; Boldt, L.; Ceyhan, B.; Blohm, D. *Anal. Biochem.* **1999**, *268*, 54.
6. (a) Keren, K.; Krueger, M.; Gilad, R.; Ben-Yoseph, G.; Sivan, U.; Braun, E. *Science* **2002**, *297*, 72; (b) Yan, H.; Park, S. H.; Finkelstein, H.; Reif, J. H.; LaBean, T. H. *Science* **2003**, *301*, 1882; (c) Keren, K.; Berman, R. S.;

Alamar Exhibit 1002
Alamar v. Olink

Buchstab, E.; Sivan, U.; Braun, E. *Science* **2003**, *301*, 1380; (d) Li, H.; Park, S. H.; Reif, J. H.; LaBean, T. H.; Yan, H. *J. Am. Chem. Soc.* **2004**, *126*, 418.

7. (a) Schweizer, B.; Roberts, S.; Grimwande, B.; Shao, W.; Wang, M.; Fu, Q.; Shu, Q.; Laroche, I.; Zhou, Z.; Tchernev, V. T.; Christiansen, J.; Velleca, M.; Kingsmore, S. F. *Nat. Biotechnol.* **2002**, *20*, 359; (b) Niemeyer, C. M.; Wacker, R.; Adler, M. *Nucleic Acids Res.* **2003**, *31*, e90.
8. (a) Zanta, M. A.; Belguise-Valladier, P.; Behr, J. P. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 91; (b) Harrison, J. P.; Balasubramanian, S. *Nucleic Acids Res.* **1998**, *26*, 3136; (c) Bongartz, J.-P.; Aubertin, A.-M.; Milhaud, P. G.; Lebleu, B. *Nucleic Acids Res.* **1994**, *22*, 4681.
9. (a) Kubo, T.; Morikawa, M.; Ohba, H.; Fujii, M. *Org. Lett.* **2003**, *5*, 2623; (b) Stetsenko, D. A.; Gait, M. J. *Bioconjugate Chem.* **2001**, *12*, 576; (c) Stetsenko, D. A.; Malakhov, A. D.; Gait, M. J. *Org. Lett.* **2002**, *4*, 3259; (d) Peyrottes, S.; Mestre, B.; Burlina, F.; Gait, M. J. *Tetrahedron* **1998**, *54*, 12513; (e) Napoli, L. D.; Messere, A.; Montesarchio, D.; Piccialli, G.; Benedetti, E.; Bucci, E.; Rossi, F. *Bioorg. Med. Chem.* **1999**, *7*, 395; (f) Schwope, I.; Bleczinski, C. F.; Richert, C. *J. Org. Chem.* **1999**, *64*, 4749.
10. (a) Muir, T. W. *Annu. Rev. Biochem.* **2003**, *72*, 249; (b) Hofmann, R. M.; Muir, T. W. *Curr Opin. Biotech.* **2002**, *13*, 297; (c) Cotton, G. J.; Muir, T. W. *Chem. Biol.* **1999**, *6*, R247.
11. Very recently, Niemeyer and co-workers reported the synthesis of PNA–protein conjugates using EPL. However, the synthesis of DNA–protein conjugates by EPL has not yet been reported. Lovrinovic, M.; Seidel, R.; Wacker, R.; Schroeder, H.; Seitz, O.; Engelhard, M.; Goody, R. S.; Niemeyer, C. M. *Chem. Commun.* **2003**, 822.
12. Stetsenko, D. A.; Gait, M. J. *J. Org. Chem.* **2000**, *65*, 4900.
13. The compound **1** reported herein was satisfactorily characterized by ^{1}H NMR and MALDI-TOF-MS spectrometry. ^{1}H NMR ($CDCl_3$): δ 1.34 (s, 9H), 1.94–2.10 (m, 4H), 2.82 (s, 4H), 2.88–2.97 (m, 4H), 3.20–3.42 (m, 2H), 3.99 (m, 1H), 4.22 (d, 2H), 4.37 (t, 1H), 5.03 (q, 1H), 5.68 (t, 1H), 7.26 (t, 2H), 7.34 (t, 2H), 7.59 (d, 2H), 7.76 (d, 2H). MALDI-TOF-MS (dithranol): calcd for $[M+Na]^+$ = 662.20, obsd 662.00.
14. HPLC was carried out on a μBondasphere C18 column (5 μ, 150×3.8 mm, Waters) eluted with 0.1 M ammonium acetate buffer (pH 7.0) containing 0–90% acetonitrile in a linear gradient over 60 min at a flow rate of 1 mL/min, detected at 260 nm.
15. Each product of the modification steps of ODN **a** and ODN**c** was identified by MALDI-TOF-MS (Matrix: 3-hydroxypicolinic acid). ODN **a**: **2**-ODN**a**, calcd for $[M+H]^+$ = 4399.0, obsd 4401.0; **3**-ODN**a**, calcd for $[M+H]^+$ = 4179.0, obsd 4179.8; **4**-ODN**a**, calcd for $[M+H]^+$ = 4190.9, obsd 4090.2. ODN**c**: **2**-ODN**c**, calcd for $[M+H]^+$ = 4292.4, obsd 4289.9; **3**-ODN**c**, calcd for $[M+H]^+$ = 4070.8, obsd 4070.0; **4**-ODN**c**, calcd for $[M+H]^+$ = 3982.7, obsd. 3982.8.
16. Southworth, M. W.; Amaya, K.; Evans, T. C.; Xu, M. Q.; Perler, F. B. *BioTechniques* **1999**, *27*, 110.
17. All reactions were carried out in a total volume of 10 μL of buffer B (50 mM Tris–HCl, 500 mM NaCl, 1 mM EDTA, 50 mM MESNA, pH 8.5). Final concentrations of ECFP and **3**-ODNs were 13 and 100 μM, respectively.
18. All samples were boiled in a loading buffer containing 2-mercaptoethanol to completely cleave disulfide bonds prior to SDS-PAGE analysis.
19. The reaction solution was subjected to 3% agarose gel electrophoresis. The gel was run in TEA buffer for 1.5 h. The product, ECFP-ODN**c**, was extracted from the gel using Freezen'N Squeeze (Bio-Rad). MALDI-TOF-MS (Matrix: sinapic acid): ECFP-thioester: calcd for $[M+Na]^+$ = 27014.3, obsd 27016.0; ECFP-ODN**c**: calcd for [M+2-mercaptoethanesulfonic acid]$^+$ = 30975.04, obsd 30955.0. DTT can be used after the ligation step to reduce any unwanted disulfide bonds. Nyanguile, O.; Dancik, C.; Blackmore, J.; Mulgrew, K.; Kaleko, M.; Stevenson, S. C. *Gene Therapy* **2003**, *10*, 1362.

Alamar Exhibit 1002
Alamar v. Olink

has been described that the AF2 domain of the estrogen receptor, present in GalER, can acquire a repressive activity through interaction with C/EBP (7). This finding is of particular interest considering that the Gal4-responsive promoter used in this system does contain a typical C/EBP target CCAAT box. It has also recently been shown that tamoxifen can promote the binding of co-repressors to the ER (6). Elucidating the molecular mechanism underlying the 4OH-tamoxifen-induced repression would be of fundamental interest. In the context of this work, this feature makes it possible to shift from a tight repression to a tight induction of the expression by using a single transcription factor (see Figure 3 for a synthetical scheme of the system). While only one out of the two activities of induction and repression is exploited at the same time in the other conditional expression systems, they are reunited in the present one, as a result of the dual activity of 4OH-tamoxifen. Taken together, the absence of Gal4-RE-binding factors and Gal4 consensus recognition sites in higher eukaryotes, the lack of cellular toxicity of 4OH-tamoxifen and of estradiol and the strong repressive activity of 4OH-tamoxifen-bound GalER make this system particularly useful for finely regulating transgene expression. As an ultimate test, we plan to use this system for controlling the expression of toxic transgenes. In conclusion, the improved system presented here allows the reconciliation of most requirements for an ideal inducible system; specifically, a strong activity upon induction and, more importantly, a low activity before induction.

REFERENCES

1. **Bonnerot, C., D. Rocancourt, P. Briand, G. Grimber and J.F. Nicolas.** 1987. A beta-galactosidase hybrid protein targeted to nuclei as a marker for developmental studies. Proc. Natl. Acad. Sci. USA *84*:6795-6799.
2. **Braselmann, S., P. Graninger and M. Busslinger.** 1993. A selective transcriptional induction system for mammalian cells based on Gal4-estrogen receptor fusion proteins. Proc. Natl. Acad. Sci. USA *90*:1657-1661.
3. **Chen, C. and H. Okayama.** 1987. High-efficiency transformation of mammalian cells by plasmid DNA. Mol. Cell Biol. *7*:2745-2752.
4. **Gossen, M. and H. Bujard.** 1992. Tight control of gene expression in mammalian cells by tetracycline-responsive promoters. Proc. Natl. Acad. Sci. USA *89*:5547-5551.
5. **Lascombe, I., P. Mougin, C. Vuillermoz, G.L. Adessi and M. Jouvenot.** 1995. Gene transfer into subcultured endometrial cells using lipofection. BioTechniques *20*:88-91.
6. **Lavinsky, R.M., K. Jepsen, T. Heinzel, J. Torchia, T.M. Mullen, R. Schiff, A. Del-Rio, M. Ricote et al.** 1998. Diverse signaling pathways modulate nuclear receptor recruitment of N-CoR and SMRT complexes. Proc. Natl. Acad. Sci. USA *95*:2920-2925.
7. **Stein, B. and M.X. Xang.** 1995. Repression of the interleukin-6 promoter by estrogen receptor is mediated by NF-κB and C/EBPβ. Mol. Cell Biol. *15*:4971-4979.
8. **Sutherland, R.L., L.C. Murphy, M. San Foo, M.D. Green, A.M. Whybourne and Z.S. Krozowski.** 1980. High-affinity anti-estrogen binding site distinct from the estrogen receptor. Nature *288*:273-275.
9. **Webster, N.J., S. Green, J.R. Jin and P. Chambon.** 1988. The hormone-binding domains of the estrogen and glucocorticoid receptors contain an inducible transcription activation function. Cell *54*:199-207.
10. **Wu, J.D., H.C. Hsueh, W.T. Huang, H.S. Liu, H.W. Leung, Y.R. Ho, M.T. Lin and M.D. Lai.** 1997. The inducible lactose operator-repressor system is functional in the whole animal. DNA Cell Biol. *16*:17-22.
11. **Zhang, J., I. Zamir and M.A. Lazar.** 1997. Differential recognition of liganded and unliganded thyroid hormone receptor by retinoid X receptor regulates transcriptional repression. Mol. Cell Biol. *17*:6887-6897.

This work was funded by Education Nationale (DRED), Centre National de la Recherche Scientifique (CNRS) and supported by the Association pour la Recherche contre le Cancer (ARC). *Address correspondence to Prof. Denis Michel, UPRESA 6026 CNRS-Université de Rennes 1, Campus de Beaulieu, 35042 Rennes Cedex, France. Internet: denis.michel @univ-rennesl.fr*

Received 8 September 1998; accepted 5 April 1999.

T. Sourisseau, Y. Le Dréan, G. Salbert, F. Flamant[1] and D. Michel
CNRS
Université de Rennes 1
Rennes
[1]CNRS-ENS de Lyon
Lyon, France

Purification of Proteins Fused to Either the Amino or Carboxy Terminus of the *Mycobacterium xenopi* Gyrase A Intein

BioTechniques 27:110-120 (July 1999)

ABSTRACT

The Mycobacterium xenopi *gyrase A mini-intein has been engineered to yield a controllable N-terminal or C-terminal, single-splice-junction autocleavage element. When combined with an affinity tag, these modified mini-inteins can be used to purify target proteins after a single combined chromatography/cleavage step. Cleavage at the intein N terminus was induced with thiol reagents, while cleavage at the intein C terminus was induced by a temperature shift to 16°–25°C. Different preferences for the residue immediately preceding the intein were observed during thiol-induced, N-terminal splice-junction cleavage of the* M. xenopi *gyrase A mini-intein vs. the* Saccharomyces cerevisiae *vacuolar ATPase, subunit A (VMA) intein present in the IMPACT™ purification system. Furthermore, the* M. xenopi *gyrase A mini-intein C-terminal autocleavage vector allows isolation of polypeptides with N-terminal cysteine residues that are active in the Intein Mediated Protein Ligation method of protein semisynthesis.*

INTRODUCTION

Many single-step systems are available to over-express and purify proteins fused to affinity tags, such as the *Escherichia coli* maltose binding protein (MBP), the *Schistosoma japonicum* glutathione S-transferase (GST), His-tag, chitin binding domain (CBD), antibody epitope tags etc. (7). The subsequent proteolytic cleavage of the fusion protein to separate the target protein from the affinity tag can be problematic in these systems (7). Proteases are often inefficient, especially when the cleavage site is buried within the fusion

Alamar Exhibit 1002
Alamar v. Olink

protein. Frequently, one has to optimize the protease reaction to balance separation from the fusion partner with cleavage at internal sites in the target protein. Many vectors also leave additional amino acids (aa) linked to the target protein after protease digestion. After successful cleavage of the target from the affinity tag, a second purification step is required to separate the protease and the fusion protein affinity tag away from the target protein.

Several of these problems have been solved in the IMPACT™ modified *Saccharomyces cerevisiae* (Sce) vacuolar ATPase, subunit A (VMA) intein-based purification systems (1,2). Inteins are protein-splicing elements that mediate cleavage of the peptide bonds linking them to N-terminal and C-terminal flanking regions (termed exteins). Inteins excise themselves from precursor proteins and ligate the flanking exteins. The mechanism of protein splicing has been extensively reviewed (6,10). Most inteins begin with Ser or Cys and end with Asn or Gln (9). Splicing is initiated by an N-S or N-O acyl shift of Cys1 or Ser1 to form a thioester or ester linkage between the N-extein and the intein. If subsequent steps in the splicing pathway are blocked, the precursor is stalled as a (thio)ester intermediate. The (thio)ester bond at the N-terminal splice junction is more reactive than a normal peptide bond and can be easily cleaved with sulfur or nitrogen nucleophiles, such as dithiothreitol (DTT) (1) or hydroxylamine (14). The C-terminal splice junction is cleaved when the intein C-terminal Asn cyclizes to form a succinimide (13). If Cys1 or Ser1 are mutated to a residue other than Cys, Ser or Thr, then splicing and N-terminal splice-junction cleavage are blocked, while C-terminal splice-junction cleavage can still occur (2,14). The challenge in using inteins for protein engineering is in learning how to control these reactions so that a precursor can be purified and autocleavage initiated at will. Previously, the Sce VMA intein was modified for use as a cleavage vector for both N-terminal (1,4) and C-terminal (2) splice-junction cleavage. In both instances, the modified Sce VMA intein was fused to the *Bacillus circulans* CBD, and the autocleavage reaction was initiated by treatment with a reducing agent, such as DTT. Some of the advantages of intein-based vectors are: *(i)* no exogenous proteases are required, *(ii)* autocleavage is specific to the intein-target protein junction, *(iii)* cloning sites are available so that the target protein can be purified with no additional vector-derived aa and *(iv)* the fusion partner remains bound to the resin after autocleavage, so no further steps are necessary to purify the target protein. However, a small peptide is released along with the target protein when

Alamar Exhibit 1002
Alamar v. Olink

using the pIMC vectors (with the target protein fused to the Sce VMA intein C terminus), and this peptide can be removed by dialysis along with any remaining thiol reagent (2). Intein-based purification systems are limited by the fact that each intein has preferred proximal sequences that promote or inhibit splicing and single-splice junction cleavage (4,12). It would therefore be useful to have several intein-based purification vectors, since each intein has different preferred flanking residues. It might also be advantageous to have an intein C-terminal splice-junction cleavage vector that does not require thiol treatment and that does not generate a small fusion partner peptide.

In this study, we describe N-terminal and C-terminal splice-junction cleavage vectors based on the *Mycobacterium xenopi* gyrase A (Mxe GyrA) intein. The Mxe GyrA intein is a natural mini-intein (198 aa), which lacks a central intein endonuclease domain. As a bacterial protein, it can be very highly expressed in *E. coli* (12). Previous studies have also demonstrated that splicing and single-splice-junction cleavage of the Mxe GyrA intein can be controlled by temperature; in the absence of native GyrA extein sequences, the Mxe GyrA intein remains as a precursor at 37°C and can then be induced to splice at 16°–23°C (12).

MATERIALS AND METHODS

Construction of Mxe GyrA Intein Clones

The Mxe GyrA intein (I or X), with Ala preceding the intein (at the -1 position), was previously cloned between the *E. coli* MBP (M) and a fragment of *Dirofilaria immitis* paramyosin (P) to form pMXP(Ala), which was previously called pMIP(Ala) (12). Ala(-1) in pMXP(Ala) was mutated to 19 other amino acids by cassette replacement. Each cassette also contained an Ile2Val mutation, which did not affect splicing of Mxe GyrA MXP(Ala) when compared with the native Mxe GyrA intein sequence. pMBXP was constructed by cloning the *B. circulans* CBD (B) (1) between MBP and the intein in pMXP(Ser), with Ser at the -1 position. pMBXt was generated from pMBXP by replacing paramyosin with *E. coli* thioredoxin (3). A C-terminal splice-junction cleavage vector, pMBX1 (Figure 1), including a multiple cloning site (MCS) polylinker at the C terminus of the Mxe GyrA mini-intein and Ser preceding the N terminus of the mini-intein, was constructed by removing a *Sap*I site from pMBXP and replacing the paramyosin fragment with a MCS cassette containing unique *Sap*I GCTCTTC(N1/4), *Nco*I (CCATGG), *Not*I (GCGGCCGC) and *Age*I (ACCGGT) cloning sites. All cloning enzymes and oligonucleotides were from New England Biolabs (Beverly, MA, USA) and used according to the manufacturer's instructions. All polymerase chain reaction (PCR) fragments were sequenced in both directions by the New England Biolabs DNA Sequencing

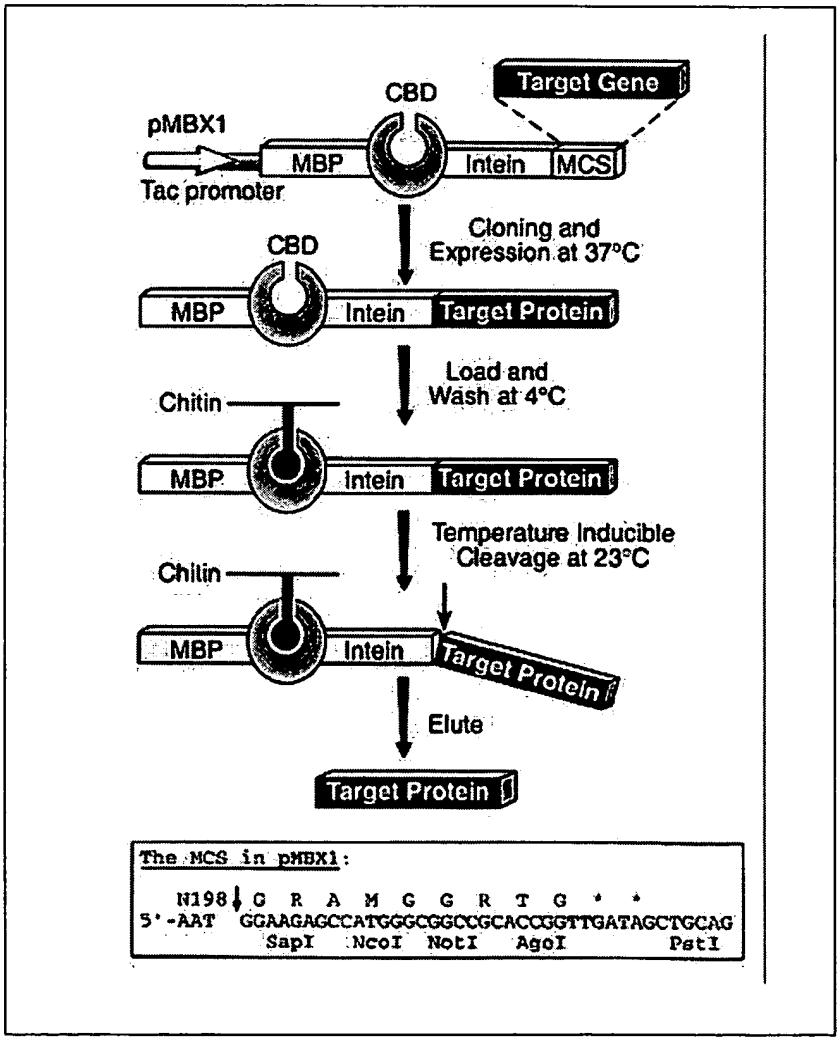


Figure 1. Purification scheme for the pMBX1 C-terminal splice-junction cleavage vector. A target protein is cloned into the pMBX1 MCS such that its N terminus is in-frame with the Mxe GyrA mini-intein C terminus. After expression at 37°C, the cleared cell lysate is loaded onto a chitin resin at 4°C, and unbound proteins are washed away. Autocleavage at the intein C-terminal splice junction is induced by shifting the temperature of the entire column to room temperature overnight. The target protein is then eluted in a suitable buffer for the target protein (pH 6.0–8.0). The pMBX1 MCS is shown with the single letter aa code above the DNA sequence and restriction enzyme sites below. N198 is the C terminus of the intein, and cleavage occurs after this aa. When a target protein is cloned into the *Sap*I site, the native sequence is obtained after autocleavage. Intein, Mxe GyrA intein with Ser at the -1 position; *, stop codon; vertical arrow, site of autocleavage.

Alamar Exhibit 1002
Alamar v. Olink

Core Facility. For details of vector construction, contact the authors.

Expression and Purification

All plasmids were transformed into *E. coli* strain ER2267 and incubated at 37°C in LB medium supplemented with 100 µg/mL ampicillin as previously described (1). The cultures were grown to mid log (optical density $[OD]_{600}$ of 0.6) and induced with 0.4 mM isopropy-B-D-thiogalactoside (IPTG), after which they were further incubated at 37°C for 2 h or 19°C overnight. All samples were electrophoresed in 10%–20% sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) gels (Novex, San Diego, CA, USA). All gels were stained with Coomassie® Blue, digitized with a Microtek Scanmaker III (Microtek International, Taiwan, R.O.C.) and analyzed with NIH Image 1.51 software (**http://rsb.info.nih.gov/nih-image**) for quantitation as described (11). Protein concentrations were determined by the Bradford assay (Bio-Rad, Hercules, CA, USA).

In vitro DTT cleavage studies were performed with precursor protein purified over amylose resin (New England Biolabs) or with lysates from cultures induced at 37°C. Samples were incubated in vitro at room temperature (23°C) in 0.5 M NaCl, 20 mM phosphate buffer, pH 7.2, in the presence or absence of 50 mM DTT. Cleavage reactions were sampled at 0, 1, 2, 5, 10, 30, 60, 120 and 240 min and overnight. The rate of DTT cleavage was determined from the slope of the line obtained by plotting the percent N-terminal splice-junction cleavage products generated vs. time while in the linear range of the curve. The percentage of precursor (MXP), spliced product (MP), N-terminal splice-junction cleavage products (M+XP) and C-terminal splice-junction cleavage products (MX+P) was determined from at least two separate experiments.

For purification of target proteins fused to the intein C terminus (Figure 1), the cell pellet from 500 mL of culture was resuspended in 25 mL Column Buffer (20 mM Tris-HCl, pH 7.4, 0.5 M NaCl, 0.1% Triton® X-100, 1 mM EDTA), sonicated and centrifuged to yield a cleared lysate. Five milliliters of soluble extract were diluted to a final volume of 25 mL with Column Buffer (30 mg total protein), loaded onto 10 mL chitin resin (New England Biolabs) at 4°C and washed with 200 mL of Column Buffer. To induce C-terminal splice-junction cleavage, the entire column containing the bound precursor (MBXP or MBXt) was removed from the refrigerator, incubated for 18 h at room temperature (23°C) to induce the cleavage reaction and then transferred back to 4°C for 2 h. Free paramyosin or thioredoxin was eluted with Column Buffer.

RESULTS AND DISCUSSION

The pTXB1 and pTXB2 Mxe GyrA intein N-terminal splice-junction cleavage vectors for purification of proteins whose C terminus is fused to the intein N terminus have been previously described (5). In this system, mutation of the Mxe GyrA intein C-terminal Asn to Ala blocked C-terminal splice-junction cleavage, while allowing thiol-induced cleavage at the Mxe GyrA intein N terminus at 4°C overnight. The pTXB1 and pTXB2 vectors have a T7 promoter (T), followed by different MCSs, the modified Mxe GyrA intein (X), a CBD affinity tag (B) and a stop codon (5). These vectors were successfully used to generate proteins with C-terminal thioesters for use in Intein Mediated Protein Ligation (IPL), a protein semisynthesis technique (5). In the present study, 20 naturally occurring aa were cloned immediately preceding a splicing-competent Mxe GyrA intein (at the -1 position) in the MXP context (with the MBP preceding the intein and paramyosin following the intein) (12) to determine preferred residues for thiol-induced cleavage. In all cases, except with MXP(Asp), expression at 37°C resulted in predominantly precursor protein, while expression at 19°C yielded differing amounts of spliced product (Table 1). MXP precursors synthesized at 37°C could be cleaved by DTT at their N-terminal splice junction with differing rates, although most -1 aa resulted in >70% N-terminal splice-junction cleavage product after overnight incubation with 50 mM DTT (Table 1).

Alamar Exhibit 1002
Alamar v. Olink

Table 1. The Effect of the -1 Residue on Splicing and Cleavage of the Mxe GyrA Intein

-1 Amino Acid	% MXP Precursor and Products Produced In Vivo				DTT Cleavage In Vitro	
	Spliced Products 19°C	N-Terminal Cleavage 37°C	C-Terminal Cleavage 37°C	MXP Precursor 37°C	Rate Product Formation 23°C	% Cleaved Overnight 23°C
Ser	0	0	37	63	0.04	24
Gly	5	0	23	76	0.48	99
His	7	5	12	81	1.91	86
Thr	0	6	17	77	0.24	91
Asp	0	100	12	0	ND[a]	ND[a]
Ile	3	2	0	98	0.12	64
Trp	54	0	0	98	2.48	83
Leu	16	24	1	74	2.73	95
Lys	41	4	5	77	2.33	86
Ala	10	0	11	88	0.43	93
Arg	36	11	2	84	3.51	77
Met	30	15	3	74	2.88	81
Tyr[b]	83	0	0	99	2.17	73
Phe	86	0	0	98	2.91	72
Pro	1	0	4	96	0.08	39
Val	0	5	1	94	0.18	81
Asn	0	0	2	98	0.40	96
Gln	15	0	0	100	0.63	97
Cys	13	21	3	72	0.48	97
Glu	0	1	0	98	0.10	45

The other 19 naturally occurring aa were substituted for the wild type Tyr preceding the Mxe GyrA intein at the -1 position.
The molar percent of precursor and products present after in vivo induction is listed for proteins expressed at the temperatures indicated.
Samples induced at 37°C were treated with 50 mM DTT at 23°C, and cleavage at the intein N terminus was quantitated during overnight time courses.
The rate of product formation (M + XP) per minute and the percent substrate cleaved after the overnight incubation in DTT are listed in the last two columns.
[a]ND, not determined.
[b]wild type -1 aa.

The exceptions were those precursors whose rate of product formation was <0.13 per min. Some percentage of each precursor might have failed to cleave because of misfolding or aggregation. Note that DTT cleavage can also be affected by intein mutations or extein residues beyond the -1 position (1,4).

In studying the effect of the -1 aa on N-terminal splice-junction cleavage, we observed that both Ser and Gly precursors yielded mostly C-terminal splice-junction cleavage products (>80%) in vivo when induced at 19°C, but not at 37°C (data not shown and Table 1). Moreover, these Ser and Gly precursors synthesized at 37°C could be induced to efficiently cleave the C-terminal splice junction when incubated at 19°C. These properties suggested that constructs with Ser and Gly at the -1 position would be good candidates for temperature-controllable, C-terminal splice-junction cleavage vectors. All other precursors yielded much less C-terminal splice-junction cleavage product after the temperature shift (data not shown). The MXP(Ser) precursor was modified to generate an Mxe GyrA intein-based, C-terminal splice-junction cleavage vector for protein purification when the N terminus of the target protein is fused to the C terminus of the intein (Figure 1). A CBD (B) was inserted between MBP (M) and the Mxe GyrA intein (X), with Ser at the -1 position, to generate

Alamar Exhibit 1002
Alamar v. Olink

pMBXP. Paramyosin was later replaced with a MCS to generate pMBX1. If target genes are cloned into the *Sap*I site in pMBX1, the purified target protein will have no additional vector-derived aa at its N terminus (Figure 1). To purify the target protein, soluble lysates from cultures induced at 37°C for 2 h are loaded onto either amylose or chitin resin at 4°C. The columns are then washed to remove unbound protein. Autocleavage at the intein C terminus is induced by overnight incubation at room temperature, followed by elution of the target protein. Chitin columns are preferred because of the high affinity of the CBD. However, amylose can be used to easily purify the precursor after eluting with maltose. Proteins bound to chitin can be released by treatment with 1% SDS or other denaturants (1).

Two target proteins (paramyosin and thioredoxin) were purified using the pMBX system. When induced at 37°C for 2 h, pMBXP yielded 37 ± 5 mg precursor (99 kDa) per L culture upon elution from amylose resin (data not shown) and 10 ± 0.5 mg of purified paramyosin (29 kDa) after elution from chitin resin (Figure 2). Thioredoxin (t) was cloned in place of paramyosin to yield MBXt. Purification on chitin resin yielded 2 mg of thioredoxin (11 kDa) per L culture (Figure 2). Although some in vivo C-terminal splice-junction cleavage was observed, levels of expression were so high that 2–10 mg of purified target proteins could still be obtained. Varying induction times and minimizing preparation and storage times all reduced the amount of cleavage observed before loading on the column (data not shown). It might be possible in the future to reduce in vivo C-terminal splice-junction cleavage without compromising in vitro C-terminal splice-junction cleavage by mutating residues within the intein.

The chitin resin (1) and the pMBX1 autocleavage system are amenable to many buffer systems. Similar results were obtained with buffers at pH 6.0–8.0, while autocleavage of MBXt was reduced by 60% at pH 9.0–9.5 and by 35% at pH 5.0. The optimal temperature for in vitro autocleavage at the

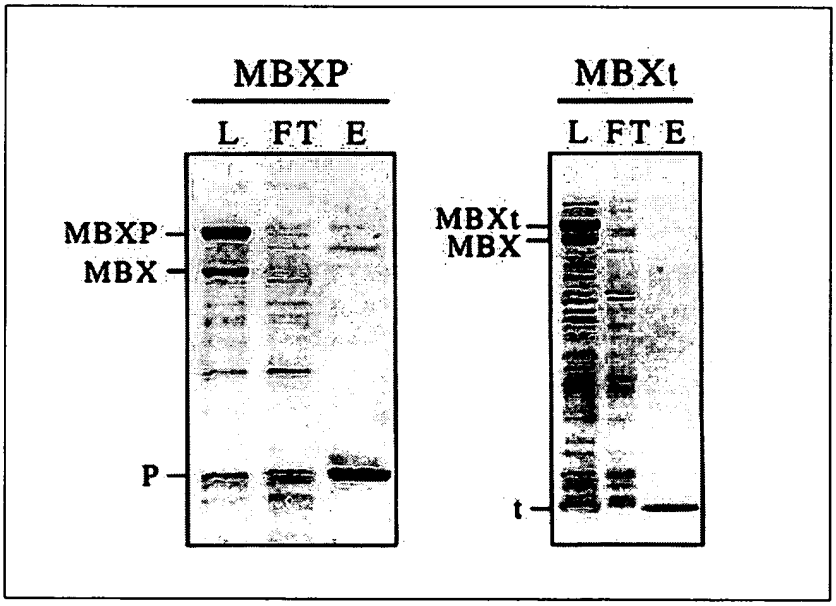


Figure 2. Purification of paramyosin and thioredoxin using the Mxe GyrA intein C-terminal fusion system. Paramyosin (P) and thioredoxin (t) were cloned as fusions in the pMBX system (see Figure 1). Soluble crude MBXP and MBXt extracts from cells induced at 37°C for 2 h (lane L) were loaded onto a chitin column at 4°C. After collecting the flow through (lane FT), the column was washed to remove unbound protein. The column was then incubated overnight at 23°C to activate Asn cyclization, which results in C-terminal splice-junction cleavage and release of free paramyosin or thioredoxin. Free paramyosin or thioredoxin were then eluted from the column after autocleavage (lane E). Samples were visualized on Coomassie® blue-stained SDS-PAGE. MBXP, paramyosin-containing precursor; MBXt, thioredoxin-containing precursor; MBX, the fusion partner after C-terminal splice-junction cleavage; P, paramyosin; t, thioredoxin.

Alamar Exhibit 1002
Alamar v. Olink

Mxe GyrA intein C terminus was 23°–30°C. Autocleavage was 20% less at 16°C, compared to 23°C, and was reduced by >75% at 37°–42°C. Autocleavage at 4°C was very slow; however, if the column was incubated at 4°C for 3 days before elution, similar amounts of purified target protein were obtained with pMBXt.

The Mxe GyrA intein C-terminal purification method is simple and straightforward (Figure 1). The pMBX1 Mxe GyrA intein C-terminal vector yields a purified target protein that is eluted into the desired buffer free of any other enzyme, peptide or chemical. pMBX1 combines the advantages of previous intein purification systems with the high expression levels of soluble protein observed with MBP (7,11). The newly developed IPL method provides a means of joining polypeptides that has great potential for protein engineering and selective protein modification (5,6,8). Unlike the Sce VMA intein-based, C-terminal splice-junction cleavage system (2), the pMBX1 vector works efficiently when Cys is at the N terminus of the target protein (data not shown) and can thus be used to generate the ligation partner with an N-terminal Cys required for IPL.

In summary, the Mxe GyrA intein has been modified to yield vectors for purification of proteins fused to the intein N or C terminus. pMBX1 provides a second method for inducing C-terminal splice-junction cleavage to complement the existing method (2), especially when thiol treatment is incompatible with the target protein. The pTXB1 and pTXB2 Mxe GyrA intein N-terminal splice-junction cleavage system provides an alternative to the Sce VMA intein IMPACT system for purifying proteins fused to the intein N terminus. The system has different preferences for acceptable sequences at the intein-target protein junction, expanding the repertoire of allowable target protein C-terminal residues compatible with intein N-terminal autocleavage vectors. For example, precursors with Asn, Cys and Pro at the -1 position were not cleaved by DTT in the Sce VMA intein system (1,4), but were in the Mxe GyrA intein system. Little in vivo N-terminal splice-junction cleavage was observed in the Mxe GyrA intein system when His, Thr, Arg or Glu were present at the -1 position, although significant amounts of in vivo cleavage were observed with similar Sce VMA intein precursors (1,4). In both cases, Asp preceding the intein resulted in nearly complete in vivo cleavage at the intein N terminus.

REFERENCES

1. **Chong, S., F.B. Mersha, D.G. Comb, M.E. Scott, D. Landry, L.M. Vence, F.B. Perler, J. Benner, R.B. Kucera, C.A. Hirvonen, J.J. Pelletier, H. Paulus and M.Q. Xu.** 1997. Single-column purification of free recombinant proteins using a self-cleavable affinity tag derived from a protein splicing element. Gene *192*:271-281.
2. **Chong, S., G.E. Montello, A. Zhang, E.J. Cantor, W. Liao, M.Q. Xu and J. Benner.** 1998. Utilizing the C-terminal cleavage activity of a protein splicing element to purify recombinant proteins in a single chromatographic step. Nucleic Acids Res. *26*:5109-5115.
3. **Chong, S., Y. Shao, H. Paulus, J. Benner, F.B. Perler and M.Q. Xu.** 1996. Protein splicing involving the *Saccharomyces cerevisiae* VMA intein. The steps in the splicing pathway, side reactions leading to protein cleavage, and establishment of an in vitro splicing system. J. Biol. Chem. *271*:22159-22168.
4. **Chong, S., K.S. Williams, C. Wotkowicz and M.Q. Xu.** 1998. Modulation of protein splicing of the *Saccharomyces cerevisiae* vacuolar membrane ATPase intein. J. Biol. Chem. *273*:10567-10577.
5. **Evans, T.C., J. Benner and M.-Q. Xu.** 1998. Semisynthesis of cytotoxic proteins using a modified protein splicing element. Protein Sci. *7*:2256-2264.
6. **Gimble, F.S.** 1998. Putting protein splicing to work. Chem. Biol. *5*:R251-R256.
7. **LaVallie, E.R. and J.M. McCoy.** 1995. Gene fusion expression systems in *Escherichia coli*. Curr. Opin. Biotechnol. *6*:501-506.
8. **Muir, T.W., D. Sondhi and P.A. Cole.** 1998. Expressed protein ligation: a general method for protein engineering. Proc. Natl. Acad. Sci. USA *95*:6705-6710.
9. **Perler, F.B.** 1999. InBase: the New England Biolabs intein database. Nucleic Acids Res. *27*:346-347.
10. **Perler, F.B., M.-Q. Xu and H. Paulus.** 1997. Protein splicing and autoproteolysis mechanisms. Curr. Opin. Chem. Biol. *1*:292-299.
11. **Southworth, M.W., E. Adam, D. Panne, R. Byer, R. Kautz and F.B. Perler.** 1998. Control of protein splicing by intein fragment reassembly. EMBO J. *17*:918-926.
12. **Telenti, A., M. Southworth, F. Alcaide, S. Daugelat, W.R. Jacobs, Jr. and F.B. Perler.** 1997. The *Mycobacterium xenopi* GyrA protein splicing element: characterization of a minimal intein. J. Bacteriol. *179*:6378-6382.
13. **Xu, M., D.G. Comb, H. Paulus, C.J. Noren, Y. Shao and F.B. Perler.** 1994. Protein splicing: an analysis of the branched intermediate and its resolution by succinimide formation. EMBO J. *13*:5517-5522.
14. **Xu, M. and F.B. Perler.** 1996. The mechanism of protein splicing and its modulation by mutation. EMBO J. *15*:5146-5153.

We thank Eric Adam and Lixin Chen for helpful discussions and Donald G. Comb for support and encouragement. Address correspondence to Dr. Francine B. Perler, New England Biolabs, Inc., 32 Tozer Road, Beverly, MA 01915, USA. Internet: perler@neb.com

Received 17 December 1998; accepted 22 March 1999.

Maurice W. Southworth, Kensey Amaya, Thomas C. Evans, Ming-Qun Xu and Francine B. Perler
New England Biolabs
Beverly, MA, USA

Alamar Exhibit 1002
Alamar v. Olink

Published online January 14, 2005

Nucleic Acids Research, 2005, Vol. 33, No. 1 e10
doi:10.1093/nar/gni010

Covalent antibody display—an *in vitro* antibody-DNA library selection system

Herald Reiersen*, Inger Løbersli, Geir Å. Løset, Else Hvattum, Bjørg Simonsen, John E. Stacy, Duncan McGregor[1], Kevin FitzGerald[1], Martin Welschof, Ole H. Brekke and Ole J. Marvik

Affitech AS, Oslo Research Park, Gaustadalleen 21, N-0349 OSLO, Norway and [1]Isogenica Ltd, Barbraham CB2 4AT, UK

Received September 13, 2004; Revised and Accepted December 15, 2004

Best Available Copy

ABSTRACT

The endonuclease P2A initiates the DNA replication of the bacteriophage P2 by making a covalent bond with its own phosphate backbone. This enzyme has now been exploited as a new *in vitro* display tool for antibody fragments. We have constructed genetic fusions of *P2A* with single-chain antibodies (*scFvs*). Linear DNA of these fusion proteins were processed in an *in vitro* coupled transcription–translation mixture of *Escherichia coli* S30 lysate. Complexes of scFv–P2A fusion proteins covalently bound to their own DNA were isolated after panning on immobilized antigen, and the enriched DNAs were recovered by PCR and prepared for the subsequent cycles of panning. We have demonstrated the enrichment of scFvs from spiked libraries and the specific selection of different anti-tetanus toxoid scFvs from a V-gene library with 50 million different members prepared from human lymphocytes. This covalent antibody display technology offers a complete *in vitro* selection system based exclusively on DNA–protein complexes.

INTRODUCTION

Antibodies are protein ligands with a wide range of biomedical applications. They have been displayed and evolved successfully by different *in vivo*- or *in vitro*-based selection methods using, for example, phage particles, yeast or bacteria cells (1–3). One limitation of the *in vivo* selection systems is the library size that could be generated. A large library is considered to be important to obtain high-affinity ligands. However, the efficiency of transfer of DNA into cells often limits the library size to 10^9–10^{10} members (3–6). *In vitro*-based antibody selection methods such as ribosome display or mRNA display have proven to be successful in the construction and selection of libraries with a high diversity and complexities (potentially up to 10^{14} different members) (5–9). Both technologies are based on an initial selection of stabilized mRNA–protein complexes (4–10).

All the present antibody selection systems only indirectly link phenotype and genotype via cell membranes, phage–protein coat, mRNA–ribosomes, protein–puromycin–mRNA complexes or water/oil emulsions (1–3,7,8,11,12). In nature, however, there is a protein from the bacteriophage P2 that offers a direct DNA–antibody selection system. This phage replicates by attaching its early gene product P2A to its own DNA. The first genetic evidence of the *cis*-activity *in vivo* was obtained when the wild-type P2 phage did not complement mutations in *P2A* (13). P2A initiates the rolling circle replication of the P2 phage *in vivo*. Its catalytic tyrosine residue (Y454) makes a single-stranded-specific nick at the viral origin of replication, located at base pair 1860 inside the 2.3 kb *P2A* gene, and forms a covalent bond with the 5′-phosphate group of the coding strand (14–16) (Figure 1a).

A single chain antibody (scFv) can be genetically fused to the P2A protein creating the smallest imaginable antibody selection particle: a protein and its gene (Figure 1b). Covalent antibody display (CAD) exploits the demonstrated *in vivo cis*-activity of P2A in an *in vitro* selection system: a fusion protein of P2A and an scFv antibody binds to the same molecule of DNA from which it has been expressed. Following *in vitro* coupled transcription and translation, the P2A protein makes a covalent link between scFv genotype and scFv phenotype, by producing a stable protein–DNA complex (14–17). P2A may thus be exploited to select scFvs from a library by using only *in vitro* methods. These antibody–DNA complexes can be isolated with standard affinity selection strategies. Specific

*To whom correspondence should be addressed. Tel: +47 22958877; Fax: +47 22958358; Email: herald@affitech.com
Correspondence may also be addressed to Martin Welschof. Email: m.welschof@affitech.com

The online version of this article has been published under an open access model. Users are entitled to use, reproduce, disseminate, or display the open access version of this article for non-commercial purposes provided that: the original authorship is properly and fully attributed; the Journal and Oxford University Press are attributed as the original place of publication with the correct citation details given; if an article is subsequently reproduced or disseminated not in its entirety but only in part or as a derivative work this must be clearly indicated. For commercial re-use permissions, please contact journals.permissions@oupjournals.org.

© 2005, the authors

Alamar Exhibit 1002
Alamar v. Olink

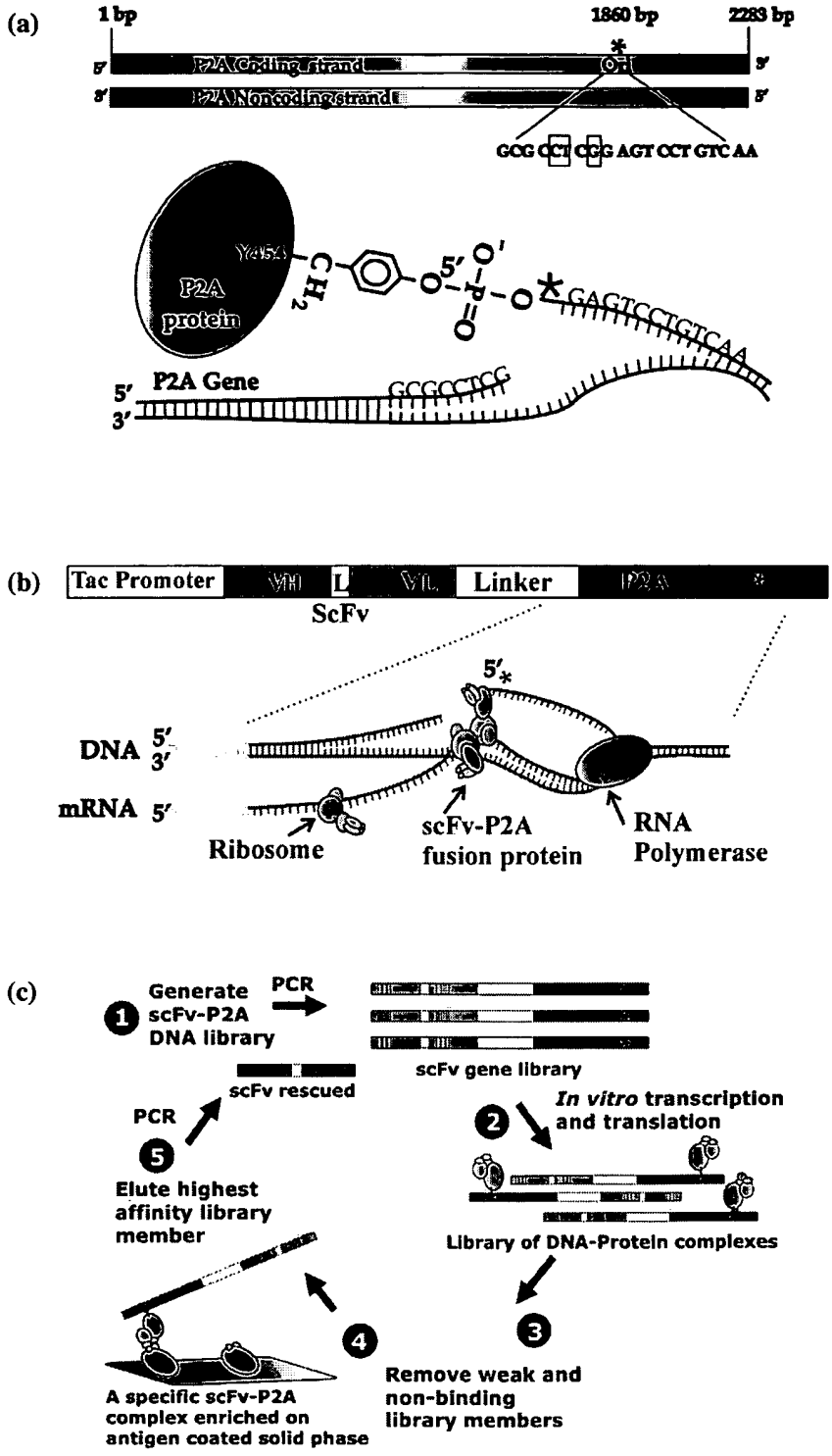


Figure 1. (a) The endonuclease P2A (761 residues, 86.3 kDa) makes a single-stranded site-specific nick at Ori of replication (. . . CCT CGG, *), located inside its own gene at position 1860, and becomes covalently attached (via Y454) to the 5′ phosphate of its own DNA (14–16). (b) CAD is the exploitation of P2A to select antibodies or antibody fragments genetically fused to it. In prokaryotes, the transcription and translation are coupled, and the start of translation normally takes place before the transcription is finished (17). The figure is a model of the complex formation among DNA, RNA polymerase, ribosomes, mRNA and nascent P2A–scFv fusion proteins (colours match the gene products). The P2A protein part of the P2A–scFv fusion protein indirectly attaches the genotype of the scFv covalently to its phenotype. (L = linker). (c) Selection cycle for CAD. An scFv repertoire is assembled with the *P2A* gene using PCR methods and fresh *P2A* and tac-promoter (1). Protein–DNA complexes are being produced in an *in vitro E.coli* S30 coupled transcription–translation mixture (2) and selected on the immobilized target (3 and 4). Retained members are eluted and DNA for the next cycle is prepared using PCR (5). Figures are not drawn to scale.

Alamar Exhibit 1002
Alamar v. Olink

complexes are enriched, eluted and rescued by PCR amplification (Figure 1c).

In the present study, we have demonstrated the suitability of P2A for specific selection of scFvs. Fusion proteins of scFv–P2A were expressed and DNA–antibody complexes were specifically recovered on antigen-coated solid phase. In addition, we have applied this technology to select antibodies from spiked and medium complex libraries. We propose that CAD can be exploited as a complete and independent *in vitro* antibody display tool for affinity selections.

MATERIALS AND METHODS

PCR cloning and assembly

The scFvs anti-phOx [phOx, 4-ethoxymethylene-2-phenyl-2-oxazoline-5-one (18); anti-phOx (19)] and anti-DOB [DOB, 2,5-dimethoxy-4-bromo-amphetamine (20); in house made anti-DOB (unpublished data), specific against DOB] were fused to either the N-terminal or C-terminal position of P2A with standard PCR cloning techniques, attaching a GSGSGS linker containing suitable flanking restriction sites (EcoRI, NotI, XhoI or NcoI) and two stop codons at the 3′ end (Figure 2). A vector tacP2aHa (5926 bp) containing the *P2A–HA* gene under the control of a tac promoter was supplied by Isogenica Ltd. *Pfu* Turbo DNA polymerase (Stratagene) was applied for generating GS-linker-scFv products for cloning. The PCR mixture was composed of 200 µM dNTP mixture, 30 ng vector DNA-template, 0.6 µM primers GsDOB3F (aaattaaaactcgagggttctggctccggttccatggcggaagtgcagctggtgc; XhoI and NcoI sites are underlined and GSGSGS sequence is in italics)/EcoRIPhog (ttttttgaattctatca*gtgatggtgatggtgatg*tgagtttagg; EcoRI site is underlined and HHHHHH sequence is in italics) or GsPhOxF (aaattaaaactcgagggttctggctccggttccatggcccaggtgcagctggtgc; XhoI and NcoI sites are underlined and GSGSGS sequence is in italics)/EcoRIPhog in the Pfu reaction buffer. Initial denaturation was at 94°C for 5 min followed by 30 cycles at 94°C (1 min), 63°C (1 min) and 72°C (2 min 30 s) in an Eppendorf Mastercycler Gradient PCR machine. The final elongation step was at 72°C for 5 min. The GS-linker-P2A for cloning scFvs at the 5′ of P2A was produced by *Pwo* Polymerase (Roche, Norway) at an annealing temperature of 65°C [primers: GsP2af (aaattaaatgcggccgcgggttctggctccggttccatggccgttaaagcctccggg; NotI and NcoI sites are underlined and GSGSGS sequence is in italics)/Lanrb (gtggataaccgtattaccgcc)]. The Pwo PCR programme was as follows: 94°C for 2 min, with initial 10 cycles at 94°C (30 s), 65°C (1 min) and 72°C (2 min), then 20 cycles at 94°C (30 s), 65°C (1 min) and 72°C (2 min) increasing the extension time with 5 s per cycle, and with a final elongation step at 72°C for 7 min. To clone small scFv libraries (scFv)*n* into the *scFv–GSGSGS–P2A* vector via NcoI/NotI sites, it was necessary to delete the second NcoI site at the beginning of P2A (Figure 2). This site was removed with the primers GSP2Afnew (aaattaaatgcggccgcgggttctggctccggttctatggccgttaaagcctccggg; NotI site is underlined, GSGSGS sequence is in italics) and Lanrb and *Pwo* polymerase as described previously. Standard methods were applied for heat transformation of

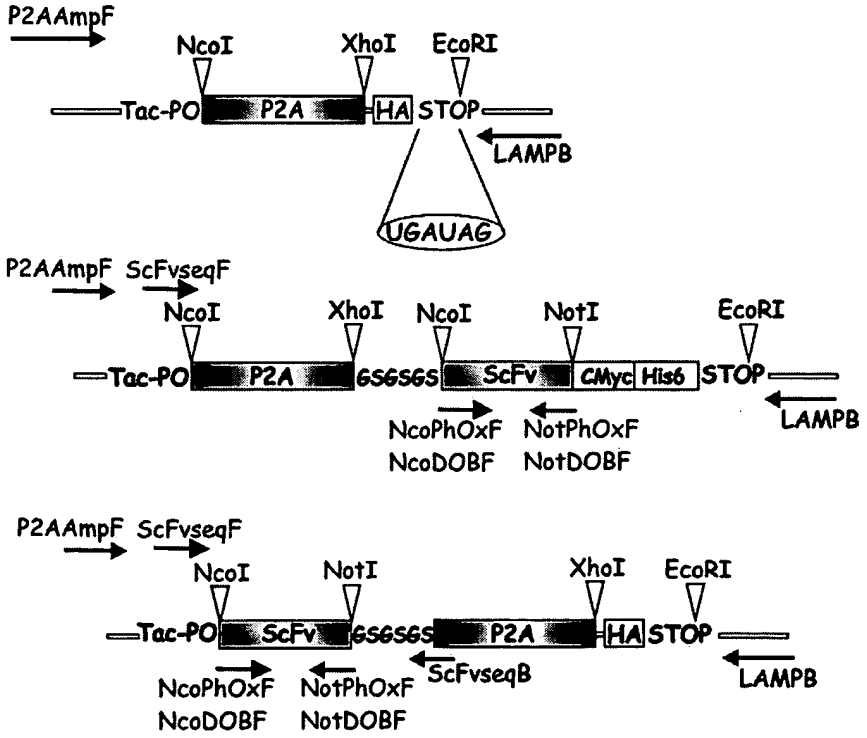


Figure 2. Primer and restriction sites for rescue PCR and assembly of DNA for next rounds of CAD. Top: *P2A–HA* construct; middle: *P2A–scFv*; and bottom: *scFv–P2A*. HA, haemagglutinin peptide epitope of human influenza virus (YPYDVPDYA); *c*-Myc, myc oncogene epitope (EQKLISEEDL); His_6, hexahistidine sequence (HHHHHH).

Alamar Exhibit 1002
Alamar v. Olink

plasmid DNA into chemical competent *Escherichia coli* DH5-α or Origami cells (Novagen). The transformed cells were grown in SOC medium at 37°C for 1 h shaking at 260 r.p.m. and plated on SOC/ampicillin (100 μg/ml) agar dishes for 15–20 h at 37°C.

Input linear DNA of *P2A–scFv* and *scFv–P2A* clones for coupled transcription and translation were made by either *Pwo* or *Tgo* DNA polymerase (Roche, Norway) with a hot-start technique. Preparations of plasmids (ligation mixtures or minipreps; 0.5 μg) were used as template for PCR, and the primers P2AampF (gcttcagtaagccagatgctac, 30 pmol) and LAMPB (tacaccgaactgagata cctac, 30 pmol) were chosen for producing the ~4 kb DNA fragment comprising tac-promoter, *scFv* and *P2A* (Figure 2). The PCR programme for *Tgo* DNA polymerase was as follows: 94°C (2 min), cycling 30 times at 94°C (30 s), 65°C (1 min) and 72°C (3 min) followed by a 7 min incubation at 72°C. The linear DNA fragments were separated on agarose gel and purified with Qiaquick gel extraction kit (Qiagen). Extra washing steps were necessary in order to remove all the salt components inhibiting transcription and translation. The products were adjusted to a high-concentration sample (1–3 μg/μl) by alcohol precipitation and applied in coupled transcription and translation.

In vitro transcription-translation

Reaction buffer (2.5×) and *E.coli* [strain SL-119, (21)] S30 extract (22) with 1 mM DTT (Sigma) were applied. For a 50 μl reaction in an Eppendorf tube, the following components were assembled on ice: S30 2.5× reaction buffer (20 μl), *E.coli* S30 extract (15 μl), ~1 mM L-methionine (Sigma), 1 μM protein disulfide isomerase (PDI; Sigma), 40 U RNasin (Promega), nuclease-free water, 0.1% BSA and DNA (1–2.5 μg). Generally, 5–10 μg of DNA was added for each round of selection. Transcription and translation reaction was started by adding DNA (1–5 μg per 50 μl reaction) and processed for 60 min at 30°C. After 40 min of incubation, 2 mM of oxidized and reduced glutathione were added. The reaction was stopped by diluting the mixture in ice-cold panning buffer and incubated with an immobilized target. The transcription and translation extract from Promega (*E.coli* S30 extract for linear templates) was also applied, but it contains an excess of DTT, which inhibits the folding of antibodies.

Affinity selections

Dynabeads M-280 Tosylactivated (Dynal Biotech, Oslo, Norway) were coated with anti-haemagglutinin (anti-HA), or with the antigens phOx–BSA (18) or DOB–BSA (20) (6 μg antibody or 20 μg hapten–BSA conjugate per mg beads; 20 mg beads/ml) in 0.1 M sodium carbonate, pH 9.6 at 37°C for 20 h on an Eppendorf thermomixer at 1200 r.p.m. Then, they were washed three times in phosphate-buffered saline (PBS), pH 7.4 with 0.1% BSA at 20°C and blocked for another 20 h at 20°C (1200 r.p.m.) in 0.3 M Tris–HCl, pH 8.8 with 0.1% BSA. Finally, they were washed in PBS, pH 7.4 and stored at a concentration of 20 mg beads per ml. For the E7 scFv, an scFv specific for an outer membrane epitope of *Neisseria meningitidis* (E7), the outer membrane vesicles (OMV) made from the bacteria, were applied as antigen (23). Immuno tubes (Maxisorp, Nunc, Denmark) were coated with antigens in PBS (pH 7.4) overnight at 20°C on a blood mixer. The affinity matrices were pre-incubated with panning buffer SuperBlock (Pierce) added ~0.2 mg/ml herring sperm DNA (Sigma), 0.25% (w/v) heparin and 1% Tween-20 (6,24,25) for 1–2 h at 20°C on a blood mixer, and then rinsed three times in PBS before panning mixture was added.

DNA–protein complexes were captured by first diluting the transcription and translation mixture into panning buffer, then incubating on an affinity coated matrix for 2–4 h rotation at 4°C or at 20°C on a blood mixer. The enriched complexes were thoroughly washed in PBS with 0.1% Tween-20, panning buffer (including an incubation step at 20°C for 30 min) and in PBS.

The DNA was eluted using proteinase K (50 μg/ml; Sigma) in TE buffer, pH 8.0 for 45 min at 50°C (26) and ethanol precipitated in seeDNA (Amersham Pharmacia Biotech).

PCR-rescue, detection of P2A–scFv fusions and processing scFv clones

If the P2A–scFv fusions have been expressed and purified, the enriched DNA of the *P2A–scFv* construct can be identified and amplified by PCR. The seeDNA precipitated DNA was dissolved in 10 μl of nuclease-free sterile MQ-water and 1–5 μl was used for each PCR-screen. If the input DNA in the transcription and translation mixture was made by PCR, it was not possible to use the same primers P2AampF and LAMPB in the PCR mixture as was done for producing it (Figure 2). This was probably due to the partial degradation of DNA after incubation. However, by using primer combinations NcoDOBF-(atggcggaagtgcagctggtg)/NotDOBB(gccgcgaagacagatggtgc), NcoPhOxF(atggcccaggtgcagctggtgc)/NotPhOxB (gccgcacctaggacggtgacc) or scFvseqF (cctgttgacaattaatcatggctcg)/scFvseqB (cctgcggcaaatgctgacgg) (Figure 2), they generated products comprising just the *scFv* gene and short flanking regions (~0.8–1 kb). For primer combinations NcoDOBF/NotDOBB and NcoPhOxF/NotPhOxB, an annealing temperature at 68°C was chosen, and 55°C for scFvseqF/scFvseqB. For detection purposes, DynaZyme II DNA polymerase (Finnzymes, Finland) was used, and for amplification of new products for next round of selection, *Pwo* polymerase (Roche) was chosen.

The recovered and amplified *scFv* DNA from each round of CAD was either cloned into an assay expression vector pHOG21 (having a lac promoter and *c*-myc/His_6-tags) (27) or assembled with tac promoter and *P2A* gene using restriction/ligation methods into vector tacP2aHa (NcoI/NotI sites; Figure 2). DNA for the next round of transcription and translation was generated by PCR using ligation mixture as mentioned above. For the expression of soluble scFvs in 96 deep-well plates, the pHOG21 expression vector constructs were transformed into XL-1 Blue cells and induced in Luria–Bertani (LB) medium containing 100 μg/ml of ampicillin (A) with 100 μM isopropyl-β-D-thiogalactopyranoside (IPTG) at 30°C and shaken for 16 h. The expressed scFvs were detected using enzyme-linked immunosorbent assay (ELISA), or were assayed as described below.

SDS–PAGE, western blotting and enhanced chemiluminescence (ECL) detection

The samples, added SDS-sample buffer with 2-mercaptoethanol, were separated on a 12.5% SDS–PAGE criterion

Alamar Exhibit 1002
Alamar v. Olink

precasted gel (Bio-Rad), electro blotted onto nitrocellulose membranes, washed (PBS), blocked with Blotto (1% dry milk in PBS), and incubated in anti-*c*-Myc (diluted 1:5000 in Blotto; Invitrogen) or anti-HA (1:2000 in Blotto; Roche) solutions. The membrane was washed (PBS with 0.1% Tween-20) and incubated with detection antibody [rabbit anti-mouse-horseradish peroxidase (HRP); DAKO, Denmark]. The HRP activity on membrane was visualized in ECL-western blotting reagents RPN 2106, Hyperfilm ECL (both Amersham Pharmacia Biotech) and a film developer.

ELISA and filter screening

Maxisorp 96-well plates (Nunc, Denmark) were coated with antigen, and expressed scFvs were detected via their *c*-Myc tags using antibodies/enzymes as described above. 3,3′,5,5′-Tetramethylbenzidene (TMB) was used as a substrate. Filter screening of 3040 bacterial colonies was performed as described previously (28). To prepare target filters (0.45 μm pore size cellulose nitrate; Schleicher & Schuell Protan BA 85), these were cut to 22 × 7 cm^2 and coated with 10 ml of phOx BSA or DOB BSA at 0.1 mg/ml. Generally, clones were high-density gridded with a Qpix robot (Genetix, New Milton, UK) on a BSA-coated filter surfacing a 22 × 22 cm^2 LB agar bioassay plate containing 30 μg/ml of tetracycline (T), 100 μg/ml of ampicillin (A) and 0.1 M glucose (G), and grown for 16 h at 37°C. Then, the BSA filter with clones was lifted on top of a sandwich with an antigen-coated filter in the middle and fresh LB agar plate with 100 μM IPTG and A at the bottom, and induced for 16 h at 30°C. The active scFvs will leak from the periplasm of the colonies to the antigen-coated filter underneath. This filter was developed like a western membrane as described above.

RESULTS AND DISCUSSION

Constructing a basic scFv–P2A model system

The A protein of the P2 phage can be used as a potential antibody display tool (Figure 1b and c). To test its suitability, we chose two well-characterized scFvs against the haptens: phOx (18) and DOB (20). The scFvs were genetically fused to P2A at two different positions connected by a $(GS)_3$-linker and expressed from a tac-promoter. The scFv N-terminal to P2A with an HA-tag: tacPO-scFv-$(GS)_3$-P2A-HA and scFv C-terminal to P2A with *c*-Myc tag: tacPO-P2A-$(GS)_3$-scFv-cmyc-His_6 were constructed (Figure 2). The fusion proteins and P2A–HA alone were translated in a standard *E.coli* S30 extract for linear templates (Promega) and sharp bands corresponding to the ~90 kDa P2A–HA and the ~110 kDa P2A–scFv proteins were clearly visible, showing efficient *in vitro* expression (Figure 3a). Estimating the yield of protein, the sensitivity of ECL (10^{-11} M) indicated a protein band (~1 ng) where the input of DNA was 1 μg (Figure 3a). This suggests that ~3% of input DNA produces proteins, which would require around 30–40 copies of each scFv to be present in the library to ensure isolation of any unique scFv members.

Antibodies are normally functionally folded in oxidizing or slightly reducing environments (6,24). To allow RNA polymerase activity, which requires a reducing environment, a specially made transcription and translation mixture was prepared (see Materials and Methods). Immediately after transcription and translation, the protein–DNA complexes were enriched onto target antigen-coated beads, eluted and detected by PCR.

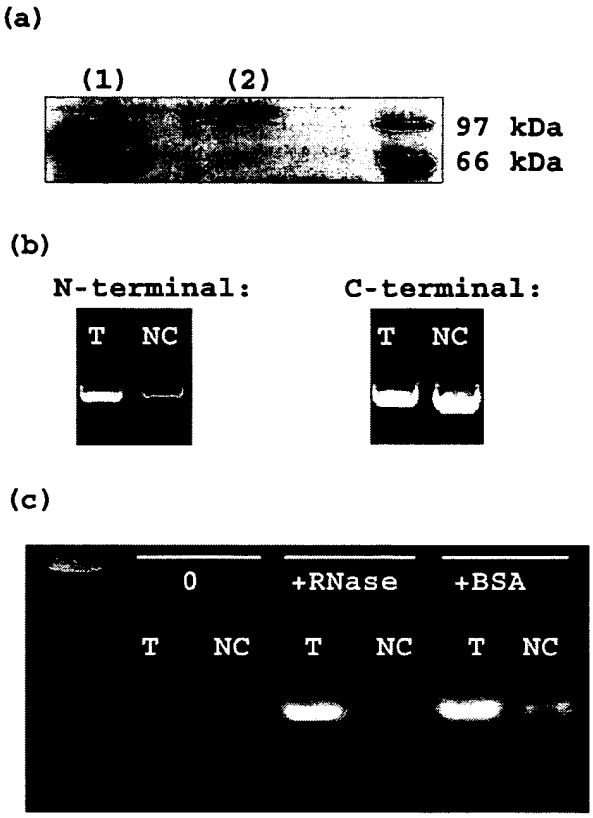


Figure 3. (a) Western / ECL identification of P2A-HA (lane 1) and P2A-anti-DOB3-c-Myc (lane 2) proteins expressed in a coupled transcription-translation mixture (Promega *E.coli* S30 extract for linear templates). The proteins were separated on 12.5% criterion polyacrylamide gel (Bio-Rad), blotted onto nitrocellulose membrane and detected by anti-c-Myc/anti-HA and rabbit anti-mouse IgG-HRP/ECL. (b) Agarose gel of rescue PCR of DNA from *anti-phOx-P2A* (left panel) or *P2A-anti-phOx* (right panel) protein–DNA complexes selected onto target T (phOx–BSA-coated Dynabeads) or negative control NC (BSA-coated Dynabeads). The protein–DNA complexes were generated in *E.coli* S30 lysate added 1 μM PDI and 1 mM DTT. They were eluted with proteinase K digestion, alcohol precipitated and detected by PCR. (c) Agarose gel of rescue PCR of DNA from scFv(E7)–P2A protein–DNA complexes selected on *N.meningitidis* outer membrane vesicles (T) or superblock (Pierce, NC). The effect of RNase (Promega) treating samples immediately after transcription–translation (at 30°C, 10 U) or adding 0.1% BSA during transcription and translation. The DNA bands from no additions/post-treatments are also shown (0).

The display of functional scFv was evaluated at both the N- and C-terminus of P2A. Figure 3b shows the recovery of rescued DNA for N- and C-terminal position of *anti-phOx scFv* relative to *P2A*. For the N-terminal position (*scFv–P2A*), there was more DNA associated with phOx–BSA than BSA beads (target PCR signal > background PCR signal). This illustrates that the DNA enrichment was well correlated with antibody activity. Similar results were also obtained for the anti-DOB fusion proteins (data not shown). In contrast, the C-terminal fusions (*P2A–scFv*) showed no specific enrichment. It is possible that the linker length was too short to accommodate proper folding and functionality of these fusions, or they were structurally destabilized. Premature

Alamar Exhibit 1002
Alamar v. Olink

termination of the translation may also occur if P2A nicks DNA before scFv is synthesized. This may occur more frequently if the scFv is at the C-terminal end of P2A rather than at the N-terminal end.

Different additives improve the functionality of covalent antibody display

Adjusting the redox potential of the transcription and translation mixture generated active fusion proteins associated with DNA, and adding BSA increased the yield of DNA recovered by PCR and improved the signal-to-noise ratio relative to no addition (Figure 3c). BSA may stabilize enzymes and components in the transcription and translation mixture during *in vitro* synthesis, even though BSA also contributes to a stronger unspecific adsorption of DNA (Figure 3c). Furthermore, the samples were treated with RNase after the protein–DNA complexes were formed (Figure 3c). A similar pre-treatment has been applied for ribosome display, using DNase to remove DNA (29). RNase increased the signal-to-noise ratio confirming that mRNA is not involved during selection (as for ribosome display). The enriched active species therefore consist of pure DNA–protein complexes. In addition, the coupled transcription and translation mixture of an *anti-phOx-P2A-HA* DNA construct (Figure 2, bottom) was incubated in 6.3 M urea before it was diluted in panning buffer and enriched on anti-HA coated tubes. Still, this harsh treatment made it possible to specifically enrich its DNA on anti-HA-coated tubes (data not shown), and thus did not release the anti-phOx-P2A-HA protein from its mother DNA. This again confirms that the anti-phOx-P2A-HA protein is covalently bound to DNA as shown previously for P2A–HA proteins (13–16).

Enrichment experiments

To demonstrate the recovery of specific antibody DNA–protein complexes from a complex mixture, specific enrichment experiments were performed. In a population where there is only one copy of a specific *P2A-scFv* gene among, e.g. 10^7 other different non-active *P2A-scFv* genes, the *cis*-activity is related to the probability of recovering the specific *P2A* gene product binding to its own mother DNA. If the P2A protein only finds it own mother DNA molecule, a direct enrichment of P2A protein–DNA complexes is productive. However, all selection systems contain a certain amount of noise (non-relevant gene products) raising the threshold required to detect rare binders from a large diverse library early in the selection process (round 1–2). If there were no *cis*-activity, the amount of randomness involved would possibly block the enrichment of the specific gene and gene product, even after 10–20 cycles. On the other hand, an early enrichment of active library members from a very diverse mixture should demonstrate that there is a *cis*-directed selection.

An scFv antibody library from human lymphocytes (PBL) was constructed by RT–PCR using *IGHV-D-J*, *IGKV-J* and *IGLV-J* genes. The library having a diversity of $n = 5.0 \times 10^7$ different clones was cloned as an N-terminal fusion to P2A, *(scFv)n–P2A*. DNA of *anti-phOx–P2A* or *anti-DOB–P2A* fusions were spiked into the library at a ratio of 1:25 000 (anti-phOx) and 1:1000 and 1:100 000 (anti-DOB) and subjected to a complete covalent display selection cycle (Figure 1c). Two spiking strategies were employed. For the selection of anti-DOB scFvs spiked into the *(scFv)n–P2A* library at a ratio of $1{:}10^3$, 4.8 ng of *anti-DOB–P2A DNA* (1.1×10^9 molecules) and 5 µg (1.1×10^{12} molecules) of library DNA were used. Lowering the same spiking ratio to $1{:}10^5$, 48 pg of *anti-DOB–P2A* (1.1×10^7 DNA molecules) was added into 5 µg (1.1×10^{12} DNA molecules) of library DNA. For the selection of anti-phOx, 400 pg of *anti-phOx–P2A* DNA was spiked into 10 µg of *(scFv)n–P2A* library (1:25 000). The recovered DNA from each round was analysed using filter screen as shown for anti-phOx in Figure 4a. As expected, there were no positive candidates before the selection of 3040 clones in the unselected library. However, after round 1–3 (R1–R3) of CAD, the number of enriched clones increased from 6 (R1), 49 (R2) to 82 clones (R3). The enrichment after each round is also illustrated in Figure 4b. The clones were MvaI digested (30) and their restriction fragment length polymorphism (RFLP) patterns confirmed their *anti-phOx–P2A* identity. Enrichment was also seen after analysing 1:1000 and 1:100 000 spiking of *anti-DOB–P2A* into the same library, respectively. After R1 of CAD 28 positive *anti-DOB–P2A* clones ($1{:}10^3$ spiking)

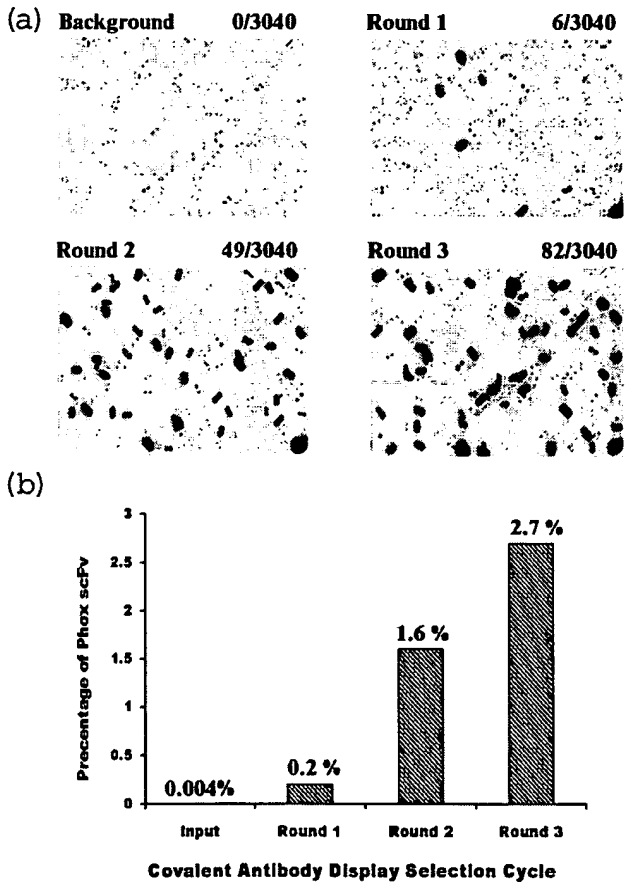


Figure 4. (a) Filter screen of expressed scFvs derived from rescued DNA cloned from each round of CAD selection. Spiking 1:25 000 of *anti-phOx–P2A* DNA into *(scFv)n–P2A* library with diversity of $n = 50$ million unique scFvs, the fraction of positives from a random selection of 3040 clones is shown. (b) Enrichment obtained from each round of selection (total of 800-fold). R0–R1: 50-fold; R1–R2: 8-fold; and R2–R3: 2-fold.

Alamar Exhibit 1002
Alamar v. Olink

were identified from a random sample of 3040 clones (14-fold enrichment), and similarly 8 clones (~300-fold enrichment) for the 1:10^5 spiking (data not shown). On studying the results above, where the ratio of scFv–PLA spiked with other irrelevant scFv–PLA DNA molecules is 1:25 000, the likelihood that the correct scFv is randomly chosen, is very low. The probability of selecting 82 positive clones from a random sample of 3040 candidates with a population of 5×10^7 and having 1 positive among 25 000 clone is as low as 10^{-13} [P, $N(0,1)$]. This argues strongly that the one given molecule of P2A should bind to its own molecule of DNA that directed its synthesis.

The translation complex formed among DNA, mRNA, ribosomes and proteins is common in prokaryotes. This implies the production of extra mRNAs and an overexpression of the P2A protein. Liu and Haggård-Ljungquist (15) describe the inability to complement an amber A mutant by the overexpression of normal A protein on a second plasmid. This indicates high-fidelity *cis*-activity. The authors suggest either compartmentalization of 'A' protein or rapid inactivation of A as the mechanism for controlling *cis*-activity. It may be that the strong *cis*-effect is due to the coupling of P2A with single-stranded DNA that is available only during the melting of double-stranded DNA with RNA polymerase, and that other free P2A proteins may be inactive for nicking. P2A cannot bind to double-stranded DNA (15), and thus, free scFv–P2A molecules should not be able to bind to untranscribed double-stranded DNA. It is also possible to purify P2A without covalently bound DNA, suggesting that P2A can be released from the ribosome without binding to DNA. The experiments, however, demonstrate a specific recovery of protein–DNA complexes. The downward trend during selections (Figure 4) may be explained wherein, for each round of selection, more active scFv–P2A fusion proteins are formed. It is most likely that the majority of these protein complexes were not bound to DNA. However, this does not influence the *cis*-activity since only single-stranded DNA is nicked by P2A. On the other hand, it affected the panning. The free, uncomplexed scFv–P2A fusion proteins compete with the corresponding DNA–protein complexes. Thus, the more the concentration of scFv–P2A protein fusions is increased, the less the DNA–protein complexes are isolated. Panning a more complex library (e.g. $n = 10^{14}$) should not however involve this type of problem since the initial concentration of expressed active proteins is very low and should not compete with DNA–protein complexes. This was also seen from the spiking experiments. Starting with a more diverse library (1:10^5), the initial enrichment from R0 to R1 was larger (300-fold) compared to less diverse spikings 1:2.5 $\times$ 10^3 and 1:10^3 that only had 50- and 14-fold enrichments, respectively. The CAD method has to be developed further to deal with these technical problems, which probably only arise at the final enrichment cycles. It may be possible to first purify all DNA before selection on protein complexes. The initial isolation of DNA could be performed with biotinylated DNA and SA-beads before subjecting it to an ordinary panning against active protein.

Panning tetanus toxoid scFvs from a 5×10^7 library

The human antibody library, (scFv)*n*–P2A, was also selected against tetanus toxoid (TTx). The serum antibody titre of the donor against TTx was found to be about 1:25 000, implying that it should be possible to identify anti-TTx scFv antibodies derived from donor lymphocytes. For each round of selection against TTx, 10 μg of DNA of the *(scFv)n–P2A* library was applied as an input for coupled transcription and translation. Indeed, as shown in Figure 5, after one round (R1) of CAD several anti-TTx scFvs were identified by ELISA. The total yield of available *anti-TTx–P2A* DNA after each round of selection was estimated from *anti-phOx–P2A* DNA standards at known concentrations (pg range). The input DNA was correlated with the corresponding PCR yields of *anti-TTx scFv* DNA versus *anti-phOx DNA* standards on an agarose gel. After the first round of selection (R1), the amount of input DNA for PCR (recovered DNA) was ~0.2 pg, which equals a DNA recovered/input ratio of 1/50 000 000. After two and three rounds of selection (R2 and R3), the estimated ratio increased to 1/2 500 000 and 1/420 000, respectively. The CAD-enriched library was also analysed by filter screen with no scFvs isolated before selection (background, 3040 colonies screened). However, after R1 of CAD, the number of positives increased to 26 (26/3040). Most of these clones bound BSA (TTx-5, Figure 5). Nevertheless, some anti-TTx scFvs not cross-reacting with BSA were confirmed by ELISA. After R2 and R3, other scFvs were rescued and TTx-12 isolated after two rounds of CAD was specific for TTx. Clones were also analysed by RFLP. The R1 clones were unique. However, TTx-8 and TTx-12 (both R2 clones) had identical RLFP patterns and DNA sequence (Table 1). The same HCDR3 sequence also appeared for TTx-17 (R3 clone), although there was a difference in their LCDR3 region (Table 1), suggesting a sequence enrichment during R1–R3 of CAD. Furthermore, the HCDR3 sequences in this study were aligned with known human anti-TTx antibody sequences [Table 1 and (31,32)]. The residues at the C-terminal region of HCDR3 of TTx-12/TTx-17 (R2/R3 clones, respectively) were very similar to 6-ATT (33), suggesting a selection on the same epitope of TTx. Similarly, most of the HCDR3 sequences compared in Table 1 have a positively charged N-terminal (at position 105 or 106) and an Asp residue at position 116. For the

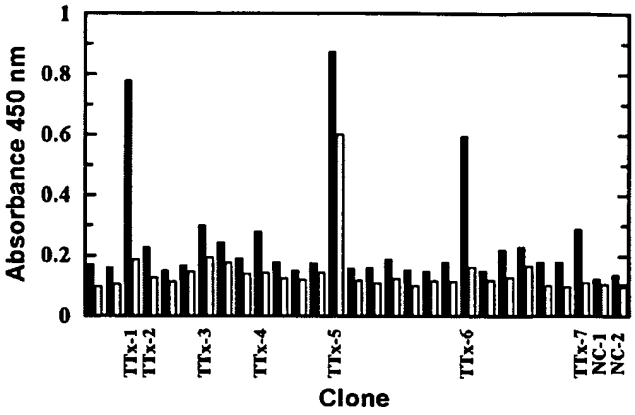


Figure 5. Panning an scFv library. (scFv)*n*–P2A ($n = 50$ million unique clones) against tetanus toxoid. ELISA of R1 of CAD. A total of 92 clones were randomly picked and assayed (ELISA). Only clones with $A_{450} > 0.15$ are shown (anti-TTx scFv, closed bar; BSA, open bar).

Alamar Exhibit 1002
Alamar v. Olink

Table 1. Comparison of human anti-TTx CDR3 regions

Clone	R	Name (IMGT-score)	HCDR3 Sequence	Name (IMGT-score)	KCDR3/LCDR3 Sequence	Reference
TTx-1	R1	IGHV3-74*03 (1227)	ARYCSGGSCLYAFDI	IGLV1-44*01 (1153)	AVWDDS - - LNGPV	–
TTx-6	R1	IGHV1-2*02 (1236)	ARGHGSGSQPRRGLPKFDL	IGKV4-1*01 (1269)	QQYYN - - - - APLT	–
TTx-22	R1	IGHV3-74*01 (1263)	ARVLGV - - EHPLG	IGLV1-44*01 (1237)	AAWDDS - - LSNPV	–
TTx-38	R1	IGHV2-5*10 (1278)	VHRRYFDRSPHFAS	IGLV1-51*01 (1252)	GTWDSS - - LSARV	–
TTx-12	R2	IGHV1-18*01 (1290)	VRLVPKRTATLHYYIDV	IGLV1-51*01 (1202)	GAWDGS - - LREAV	–
TTx-17	R3	IGHV1-18*01 (1290)	VRLVPKRTATLHYYIDV	IGLV1-51*01 (1193)	AAWDDS - - LNGPV	–
CH8	n.a.	IGHV	ETSVRRDYRDYPMTDY	IGLV	QQSYS - - -TPPRT	(34)
CH5	n.a.	IGHV	RSVWGE - - HYFDY	IGLV	QQYNS - - - -WPLT	(34)
6-ATT	n.a.	IGHV4-B*01 (1138)	AKVALMKWFGYYFDS	IGKV1-9*01 (1179)	QQYSS - - - SPPVT	(33)
B7-15A2	n.a.	IGHV1	ARVLFQQLVLYAPFDI	IGLV	QSYDS - - - SLSAR	(35)

TTx-sequences (*IGHV-D-J*, *IGKV-J* and *IGLV-J* genes) were classified and aligned using IMGT/V-QUEST (http://imgt.cines.fr), and the positions 105–117 for CDR3 sequences are shown (31,32). The gene names are the human germline equivalent. IMGT-scores > 1179 reflects at least 95% nucleotide sequence identity (33). The 6-ATT clone is from *Macaca fascicularis* Fab (33). R = Round of CAD. na, not available. The basic residues are in blue, the acidic residues are in red, and serines and threonines are in green.

Table 2. Comparison of different *in vitro* display technologies

Property	Phage (3)[a]	Ribosomal (7)[a]	mRNA (7,8)[a]	IVC (11,12)[a]	CAD (this study)	CIS (36)[a]
Covalent link	No	No	Yes (synthetic)	No	Yes (inherent)	No
Nucleic acid selected	DNA	mRNA	mRNA/cDNA	DNA	DNA	DNA
Antibody display	scFv	scFv	scFv[b]	No	scFv	scFv
Typical e.f.	10^2–10^4	10^1–10^3	10^1–10^3	10^1–10^3	10^1–10^2	10^3
Library size	10^9–10^{11}	10^{12}–10^{13}	10^{12}–10^{13}	10^9–10^{10}/ml	10^7	10^{12}
Potential size	10^{11}	$>10^{13}$	$>10^{13}$	$>10^{10}$	$>10^{13}$	$>10^{13}$
Making libraries	Time consuming	Fast	Fast	Fast	Fast	Fast
Stability	Stable	Stable in Mg^{2+}	Covalent	Stable	Covalent	>2 days

Abbreviations: e.f., enrichment factor per round of selection; and IVC, *in vitro* compartmentalization.
[a]Represent the reference numbers.
[b]Via *IGHV-D-J*, *IGKV-J* and *IGL-V-J* genes.

light chains, they all contain a Ser in the middle of the loop (position 109 or 110).

CAD as a new *in vitro* antibody display tool

We have demonstrated the successful selections from two independent spiking experiments (phOx and DOB fusions) and from an immunized scFv library with a complexity of 10^7. The method should thus be well suited for *in vitro* scFv selections from libraries. However, in order to fully demonstrate its potential, a more diverse library should be screened. P2A can be applied as a general panning tool for antibody library selections at the N-terminal end of the endonuclease. The advantages for such a selection system are numerous, compared to the existing phage display and *in vitro* mRNA display methods. CAD is the only antibody selection method providing a direct covalent link between an *scFv* gene and its protein (14–16), increasing the chemical stability of the panning complex. Table 2 compares the more common methods of *in vitro* selection, and all methods have their advantages and disadvantages. The new DNA-based *cis*-display methods of panning are however very promising. Generally, CAD may facilitate a very rapid library generation using PCR methods with increased library complexity size (potentially up to 10^{13} scFv members) and a fast cycle time. In a few hours, unique scFvs can be enriched, isolated and directly amplified for the next rounds of selection (Figure 1c). The system should be very well suited for scaling up and automation, and may be a future standard for *in vitro* antibody selections. However, at present, the CAD methods must be made more robust to overcome the product inhibition of uncomplexed fusion proteins without DNA. CAD offers a covalent antibody selection tool as an alternative to the latest non-covalent *in vitro* display selection method of peptides and proteins with the RepA protein (36).

ACKNOWLEDGEMENTS

We thank Dr David Andrews (McMaster University, Hamilton, Canada) for providing *E.coli* S30 lysates and Dr Terje Michaelsen (The Norwegian Institute of Public Health, Oslo, Norway) for TTx-antigen. We also thank Professor Björn Lindqvist (Department of Biology, University of Oslo, Norway) for inspiration and initial discussions. Funding to pay the Open Access publication charges for this article was provided by Affitech AS.

REFERENCES

1. Li,M. (2000) Applications of display technology in protein analysis. *Nat. Biotechnol.*, **18**, 1251–1256.
2. Lee,S.Y., Choi,J.H. and Xu,Z. (2003) Microbial cell-surface display. *Trends Biotechnol.*, **21**, 45–52.
3. Hoogenboom,H.R., de Bruïne,A.P., Hufton,S.E., Hoet,R.M., Arends,J.-W. and Roovers,R.C. (1998) Antibody phage display technology and its applications. *Immunotechnology*, **4**, 1–20.
4. Keefe,A.D. and Szostak,J.W. (2001) Functional proteins from a random-sequence library. *Nature*, **410**, 715–718.

Alamar Exhibit 1002
Alamar v. Olink

5. Wilson,D.S., Keefe,A.D. and Szostak,J.W. (2001) The use of mRNA display to select high-affinity protein-binding peptides. *Proc. Natl Acad. Sci. USA*, **98**, 3750–3755.
6. Schaffitzel,C., Hanes,J., Jermutus,L. and Plückthun,A. (1999) Ribosome display: an *in vitro* method for selection and evolution of antibodies from libraries. *J. Immunol. Methods*, **231**, 119–135.
7. Lipovsek,D. and Plückthun,A. (2004) *In-vitro* protein evolution by ribosome display and mRNA display. *J. Immunol. Methods*, **290**, 51–67.
8. Takahashi,T.T., Austin,R.J. and Roberts,R.W. (2003) mRNA display: ligand discovery, interaction analysis and beyond. *Trends Biochem. Sci.*, **28**, 159–165.
9. Hanes,J. and Plückthun,A. (1997) *In vitro* selection and evolution of functional proteins by using ribosome display. *Proc. Natl Acad. Sci. USA*, **94**, 4937–4942.
10. Amstutz,P., Forrer,P., Zahnd,C. and Plückthun,A. (2001) *In vitro* display technologies: novel developments and applications. *Curr. Opin. Biotechnol.*, **12**, 400–405.
11. Tawfik,D.S. and Griffiths,A.D. (1998) Man-made cell-like compartments for molecular evolution. *Nat. Biotechnol.*, **16**, 652–656.
12. Yonezawa,M., Doi,N., Kawahashi,Y., Higashinakagawa,T. and Yanagawa,H. (2003) DNA display for *in vitro* selection of diverse peptide libraries. *Nucleic Acids Res.*, **31**, e118.
13. Lindahl,G. (1970) Bacteriophage P2: replication of the chromosome requires a protein which acts only on the genome that coded for it. *Virology*, **42**, 522–533.
14. FitzGerald,K. (2000) *In vitro* display technologies–new tools for drug discovery. *Drug Discov. Today*, **5**, 253–258.
15. Liu,Y. and Haggård-Ljungquist,E. (1994) Studies of bacteriophage P2 DNA replication: localization of the cleavage site of the A protein. *Nucleic Acids Res.*, **22**, 5204–5210.
16. Odegrip,R. and Haggård-Ljungquist,E. (2001) The two active-site tyrosine residues of the A protein play non-equivalent roles during initiation of rolling circle replication of bacteriophage P2. *J. Mol. Biol.*, **308**, 147–163.
17. Voet,D. and Voet,J.G. (1995) Control of transcription in prokaryotes. In Voet,D and Voet,J.G. (eds.), *Biochemistry*, 2nd edn. John Wiley & Sons Inc., NY, Section 29-3, pp. 930–932.
18. Marks,J.D., Griffiths,A.D., Malmqvist,M., Clackson,T.P., Bye,J.M. and Winter,G. (1992) By-passing immunization: building high affinity human antibodies by chain shuffling. *Biotechnology (NY)*, **10**, 779–783.
19. Braunagel,M. and Little,M. (1997) Construction of a semisynthetic antibody library using trinucleotide oligos. *Nucleic Acids Res.*, **25**, 4690–4691.
20. Cody,J.T. (1990) Cross-reactivity of amphetamine analogues with Roche Abuscreen radioimmunoassay reagents. *J. Anal. Toxicol.*, **14**, 50–53.
21. Lesley,S.A., Brow,M.A. and Burgess,R.R. (1991) Use of *in vitro* protein synthesis from polymerase chain reaction-generated templates to study interaction of *Escherichia coli* transcription factors with core RNA polymerase and for epitope mapping of monoclonal antibodies. *J. Biol. Chem.*, **266**, 2632–2638.
22. Lesley,S.A. (1995) Preparation and use of *E.coli* S-30 extracts. *Methods Mol. Biol.*, **37**, 265–278.
23. Fredriksen,J.H., Rosenqvist,E., Wedege,E., Bryn,K., Bjune,G., Frøholm,L.O., Lindbak,A.-K., Møgster,B., Namork,E., Rye,U. *et al.* (1991) Production, characterization and control of MenB-vaccine 'Folkehelsa': an outer membrane vesicle vaccine against group B meningococcal disease. *NIPH Annal.*, **14**, 67–80.
24. Osbourn,J. and Holet,T. (2001) Improvements to ribosome display. Patent WO0175097.
25. Schatz,P.J., Cull,M.G., Miller,J.F., Stemmer,W.P.C. and Gates,C.M. (1998) Peptide library and screening method. US Patent 6,156,511.
26. Sambrook,J. and Russel,D. (eds.), *Molecular Cloning*. 3rd edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, Vol. 3, pp. A1.11 and A4.
27. Kipriyanov,S.M., Moldenhauer,G. and Little,M. (1997) High level production of soluble single chain antibodies in small-scale *Escherichia coli* cultures. *J. Immunol. Methods*, **200**, 69–77.
28. de Wildt,R.M., Mundy,C.R., Gorick,B.D. and Tomlinson,I.M. (2000) Antibody arrays for high-throughput screening of antibody-antigen interactions. *Nat. Biotechnol.*, **18**, 989.
29. Coia,G., Pontes-Braz,L., Nuttall,S.D., Hudson,P.J. and Irving,R.A. (2001) Panning and selection of proteins using ribosome display. *J. Immunol. Methods*, **254**, 191–197.
30. Huang,Q.R., Morris,D. and Manolios,N. (1997) Identification and characterisation of polymorphisms in the promoter region of the human Apo-1/Fas (CD95) gene. *Mol. Immunol.*, **34**, 577–582.
31. Giudicelli,V., Chaume,D. and Lefranc,M.-P. (2004) IMGT/V-QUEST, an integrated software program for immunoglobulin and T cell receptor V-J and V-D-J rearrangement analysis. *Nucleic Acids Res.*, **32**, W435–W440.
32. Lefranc,M.P., Pommie,C., Ruiz,M., Giudicelli,V., Foulquier,E., Truong,L., Thouvenin-Contet,V. and Lefranc,G. (2003) IMGT unique numbering for immunoglobulin and T cell receptor variable domains and Ig superfamily V-like domains. *Dev. Comp. Immunol.*, **27**, 55–77.
33. Chassagne,S., Laffly,E., Drouet,E., Hérodin,F., Lefranc,M.-P and Thullier,P. (2004) A high-affinity macaque antibody Fab with human-like framework regions obtained from a small phage display immune library. *Mol. Immunol.*, **41**, 539–546.
34. Chin,J., Sohn,Y., Lee,S.H., Park,Y.I. and Choi,M.J. (2003) Production of neutralizing human monoclonal antibody directed to tetanus toxin in CHO cell. *Biologicals*, **31**, 45–53.
35. Faber,C., Shan,L., Fan,Z.-C., Guddat,L.W., Furebring,C., Ohlin,M., Borrebaeck,C.A.K. and Edmundson,A.B. (1998) Three-dimensional structure of a human Fab with high affinity for tetanus toxoid. *Immunotechnology*, **3**, 253–270.
36. Odegrip,R., Coomber,D., Eldridge,B., Hederer,R., Kuhlman,P.A., Ullman,C., FitzGerald,K. and McGregor,D. (2004) CIS display: *in vitro* selection of peptides from libraries of protein–DNA complexes. *Proc. Natl Acad. Sci. USA*, **101**, 2806–2810.

Alamar Exhibit 1002
Alamar v. Olink

This Page is Inserted by IFW Indexing and Scanning Operations and is not part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☐ BLACK BORDERS
- ☐ IMAGE CUT OFF AT TOP, BOTTOM OR SIDES
- ☑ FADED TEXT OR DRAWING
- ☐ BLURRED OR ILLEGIBLE TEXT OR DRAWING
- ☐ SKEWED/SLANTED IMAGES
- ☐ COLOR OR BLACK AND WHITE PHOTOGRAPHS
- ☐ GRAY SCALE DOCUMENTS
- ☐ LINES OR MARKS ON ORIGINAL DOCUMENT
- ☐ REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY
- ☐ OTHER: _____

IMAGES ARE BEST AVAILABLE COPY.
As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.

Alamar Exhibit 1002
Alamar v. Olink

Best Available Copy

Cytokine detection by antibody-based proximity ligation

Mats Gullberg*, Sigrún M. Gústafsdóttir*, Edith Schallmeiner*, Jonas Jarvius*, Mattias Bjarnegård†, Christer Betsholtz†, Ulf Landegren*, and Simon Fredriksson*‡

*The Beijer Laboratory, Department of Genetics and Pathology, Rudbeck Laboratory, Uppsala University, Se-75185 Uppsala, Sweden; and †Department of Medical Biochemistry, Göteborg University, Göteborg SE-405 30, Sweden

Edited by Larry Gold, SomaLogic, Inc., Boulder, CO, and approved April 19, 2004 (received for review January 23, 2004)

Efficient and precise detection techniques, along with extensive repertoires of specific binding reagents, will be needed to meet the challenges of proteome analyses. The recently established proximity ligation mechanism enables sensitive high-capacity protein measurements by converting the detection of specific proteins to the analysis of DNA sequences. Proximity probes containing oligonucleotide extensions are designed to bind pairwise to target proteins and to form amplifiable tag sequences by ligation when brought in proximity. In our previous report, both the ligatable arms and the protein binders were DNA molecules. We now generalize the method by providing simple and convenient protocols to convert any polyclonal antibodies or matched pair of monoclonal antibodies to proximity probe sets through the attachment of oligonucleotide sequences. Sufficient reagent for >100,000 proximity ligation assays can be prepared from 1 μg of antibody. The technique is applied to measure cytokines in a homogenous test format with femtomolar detection sensitivities in 1-μl samples, and we exemplify its utility in situations when only minute sample amounts are available.

The availability of total genome sequence information provides an overview of the proteins potentially present in an organism. It must now be a high priority to devise reagent sets and analytic procedures that can provide insights into the role of all these proteins in molecular processes and pathological alterations thereof (1). We have recently established the proximity ligation assay, which is promising as a general mechanism for highly specific and sensitive detection of proteins, singly or in parallel, in solution or localized in tissue and cell samples (2).

In proximity ligation, binding of pairs of specific protein-binding reagents to the same target-protein molecule brings oligonucleotides attached to the binding reagents in proximity. Next, a mixture is added that contains all components required for ligation of the oligonucleotides and for amplification and detection of the ligation products. This mixture includes a connector oligonucleotide added in molar excess, which hybridizes to the ends of nearby DNA strands, guiding their prompt ligation (Fig. 1). Thereby, proximity ligation allows proteins to be represented as amplifiable information carrying DNA strands through a highly specific mechanism that depends on dual recognition of target molecules. Excellent sensitivity is ensured by the great increase in reactivity of ligatable ends on coincident target binding through increased relative concentration in combination with amplified DNA detection by real-time PCR, enabling the measurement of very few ligation products. Proximity ligation can also be performed by using a solid phase format and, due to its proximity-dependent signal, it has displayed higher sensitivity than another DNA-based protein detection assay, immuno-PCR (2, 3).

We have previously shown that proximity ligation by using pairs of protein-binding DNA aptamers provides advantages for protein analyses. These assays have proven of particular value under conditions where exceptional sensitivity, not offered by traditional techniques, is required (4–6). Unfortunately, only a limited number of suitable aptamers are currently available (7), limiting the utility of the procedure. Here we describe simple and efficient protocols to prepare proximity probes using readily available batches of polyclonal antibodies or matched pairs of monoclonal antibodies. We further present means to predict assay performance and report the use of these reagents for sensitive cytokine detection in minute samples derived from serum, conditioned media, or tissue lysate.

Materials and Methods

Antibodies. Affinity-purified polyclonal antibodies and targets for IL-2 and IL-4 assays were from R & D Systems, whereas reagents for the vascular endothelial growth factor (VEGF) and homodimer of platelet-derived growth factor B chain (PDGF-BB) assays were from PeproTech (Rocky Hill, NJ). Monoclonal antiinsulin antibodies M1–1 and M2–2 and human insulin were generous gifts from Mercodia (Uppsala).

Covalent Antibody–Oligonucleotide Conjugation. Monoclonal insulin antibodies were incubated with a 30-fold excess of succinimidyl 4-[*p*-maleimidophenyl]butyrate (SMPB) (Pierce) freshly prepared in DMSO in a total volume of 35 μl of PBS at room temperature (RT) for 30 min. Reactions were separated twice over a G-50 spin column to remove excess SMPB and eluted into 3 μl of 10× PBS plus EDTA. Oligonucleotides (B)5′-SH and (B)3′-SH (see Table 1, which is published as supporting information on the PNAS web site) in a 3:1 ratio to the amount of antibody were pretreated with 10 mM DTT for 30 min, diluted to 30 μl, purified three times on G-50 microspin columns, and mixed with the activated antibody. Coupling proceeded for 2 h at RT or overnight at 8°C. Purification was performed as follows. The antibody and antibody–oligonucleotide conjugates were precipitated with an equal volume of saturated ammonium sulfate at RT for 1 h or +8°C overnight. The reaction mix was centrifuged at 6,500 × *g* for 30 min, and the supernatant was discarded. Free antibody was removed by applying the reconstituted sample to a 1-ml Q-Sepharose (Amersham Pharmacia Biosciences) column and washing with 6 ml of 50 mM Tris·HCl, pH 7.5/300 mM NaCl. The antibody–oligonucleotide conjugate and remaining free oligonucleotides were eluted with 6 ml of 50 mM Tris·HCl, pH 7.5/1.5 M NaCl into a tube precoated with 1% BSA. The sample was concentrated on a YM-100 microcon concentrator (Millipore) with three consecutive additions of 2 ml of PBS with 5 mM EDTA, also removing remaining oligonucleotides. Coupling efficiencies were ≈1–10%, as judged by the amount of oligonucleotide present after purification of the conjugates. Gel analysis of radiolabeled and purified probes showed that >75% of all oligonucleotides were coupled to the

This paper was submitted directly (Track II) to the PNAS office.

Abbreviations: STV, streptavidin; RT, room temperature; PDGF-BB, homodimer of platelet-derived growth factor B chain; VEGF, vascular endothelial growth factor.

‡To whom correspondence should be sent at the present address: Stanford Genome Technology Center, Stanford University, 855 California Avenue, Palo Alto, CA 94304. E-mail: simon.fredriksson@stanford.edu.

© 2004 by The National Academy of Sciences of the USA

www.pnas.org/cgi/doi/10.1073/pnas.0400552101

Alamar Exhibit 1002
Alamar v. Olink

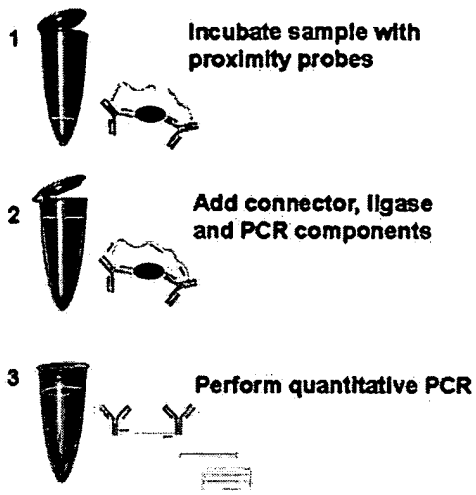


Fig. 1. The principal steps of the proximity ligation assay. Step 1, incubation of sample with proximity probe pair (≈1 h); step 2, addition of all components required for ligation and detection by quantitative PCR (≈5 min ligation time); step 3, quantitative PCR (≈2 h).

antibodies (data not shown). The retentate was collected and diluted in 0.2% BSA in 1× PBS/5 mM EDTA/0.1% sodium azide and stored at +8°C or aliquoted and frozen at −20°C.

Proximity-Probe Construction Through Streptavidin (STV)–Biotin Linkage. Maleimide-derivatized STV (0.5 nmol) (Sigma) was coupled to 2 nmol of DTT-reduced oligonucleotides (A)3′-SH or (A)5′-SH in 50 μl of phosphate-buffered saline with 5 mM EDTA. Reduction was performed in a 50-μl volume of 50 mM freshly prepared DTT with 2% trietylamine for 10 min at RT. Excess DTT was removed by centrifugation through a prespun 3-ml gel filtration column with a 10% (wt/vol) slurry of G-50 (Amersham Pharmacia) at 1,500×*g* for 1 min. The eluate was directly combined with the STV for 2 h at 37°C. The reaction was quenched by the addition of 0.5 μl of 0.5 M 2-mercaptoethanol.

Crude conjugates were purified from free DNA by protein precipitation with 1 volume of saturated ammonium sulfate at RT for 1 h, then centrifuged at +4°C at 13,000×*g* for 30 min. The supernatant was removed, the material redissolved in phosphate-buffered saline with 5nM EDTA, and the protein precipitation process was repeated once more. This was followed by DNA precipitation with 0.1 volume of 3 M NaAc, pH 4.6/10 mM $Mg(Ac)_2$/2 volumes of ice-cold 95% ethanol to remove free protein. After 2 h at RT, the sample were centrifuged at 13,000×*g* for 30 min at +4°C. The final product was dissolved in 50 μl of 10 mM Tris·HCl, pH 7.4, with 0.02% sodium azide and stored at +4°C. The protein content of the purified product was estimated by the NanoOrange kit (N-6666, Molecular Probes) by using the included BSA standard and an Applied Biosystems 7700 fluorometer. Oligonucleotides were quantified by absorbance measurements at 260 nm with a NanoDrop spectrophotometer. Product purity was analyzed by separating the samples on an Agilent (Palo Alto, CA) Bioanalyzer 2100 with the mRNA pico kit. Oligonucleotide sequences are shown in Table 1.

Biotinylated antibodies were combined with the STV–oligonucleotide conjugates in 10-μl volumes with 50 nM of each component for 1 h at RT in phosphate-buffered saline with 5nM EDTA (PBSE) with 0.1% BSA. These were then diluted to 1 nM in PBSE with 1% BSA and 0.02% sodium azide and stored separately at +4°C for the 3′ and 5′ proximity probes. Upon use, these solutions were further diluted to a concentration of 100 pM in PBS with 1% BSA/16 μg/ml sheared polyA bulk nucleic acid (Sigma), and with 1 mM free biotin added to quench any further biotin–STV interaction. 3′ and 5′ proximity probes were then combined and further diluted into 5-μl incubations with a 1-μl sample. Samples were prepared by diluting targets in PBS with 1% BSA. The IL-2 and IL-4 assays were performed by using 25 pM proximity probes and the VEGF assay with 5 pM probes.

Theoretical Calculations. The concentration of free probe [*P*], free target [*T*], and probe–protein complexes [*PT*] at equilibrium was calculated based on the formula for calculating the dissociation constant (K_d), Eq. **1**, and knowing the concentration of probes, target protein, and the K_d for the interaction.

$$K_d = \frac{[P]^*[T]}{[PT]}. \qquad [1]$$

By multiplying the fraction of target-binding site 1 that is bound by probe 1 with the fraction of site 2 bound by probe 2, the fraction of target proteins simultaneously bound by the two proximity probes ($fT_{bound*2}$) is calculated according to Eq. **2**.

$$fT_{bound*2} = \frac{[P_1T]}{[T]}*\frac{[P_2T]}{[T]}. \qquad [2]$$

To correct for impurities, account must be taken of free oligonucleotides that do not contribute to signal but give rise to background and of free antibodies that could reduce signal but not affect the background. The situation can be described by using Eq. **3**, where f_{P1conj} and f_{P2conj} denote the fraction of the two antibodies having oligonucleotides attached. To estimate the signal over background, consideration must also be given to the background caused by free oligonucleotides, which can be experimentally determined.

$$\%T_{bound*2} = f_{P1conj}\frac{[P_1T]}{[T]}*f_{P2conj}\frac{[P_2T]}{[T]}*100. \qquad [3]$$

Because PDGF-BB is a homodimeric protein and both proximity probes are directed against the same epitope, only 50% of target proteins bound by two proximity probes can be expected to involve reagents with both free 5′ and 3′ ends and capable of generating a signal through ligation. For simplicity, ligation efficiency of such trimeric complexes is assumed to be 100%. Signal over background in Fig. 2 was calculated for 20 pM of probes (directed against PDGF and thrombin) or 5 pM or probes (directed against insulin), as used in the corresponding experiments.

Homogenous Proximity Ligation. Except where indicated, assays were performed by incubating samples with proximity probes in 5-μl incubations for 1 h, before addition of a 45-μl mix containing components required for probe ligation and real-time PCR as described (2). The mix contained 50 mM KCl, 20 mM Tris·HCl (pH 8.4), 2.5 mM $MgCl_2$, 0.4 units of T4 DNA ligase (Amersham Pharmacia Biosciences), 400 nM connector oligonucleotide, 80 μM ATP, 0.2 mM dNTPs, 0.5 μM primers, 200 nM probe for the 5′ nuclease assay, and 1.5 units of platinum *Taq*DNA polymerase (Invitrogen). After a 5-min ligation reaction at RT, the reactions were transferred to the real-time PCR instrument for temperature cycling: 95°C for 2 min and then 95°C for 15 sec and 60°C for 60 sec, repeated 45 times (Applied Biosystems PRISM 7700 or 7000). Oligonucleotide sequences are shown in Table 1.

Cell Culture. Human 293T embryonic kidney cells were cultured in DMEM cell culture medium (Invitrogen) with 10% FBS (Invitrogen). For measurement of VEGF-A protein levels, 200,000 cells were seeded per well in six-well plates. After 24 h, cells were transferred to an airtight humidified chamber and

Alamar Exhibit 1002
Alamar v. Olink

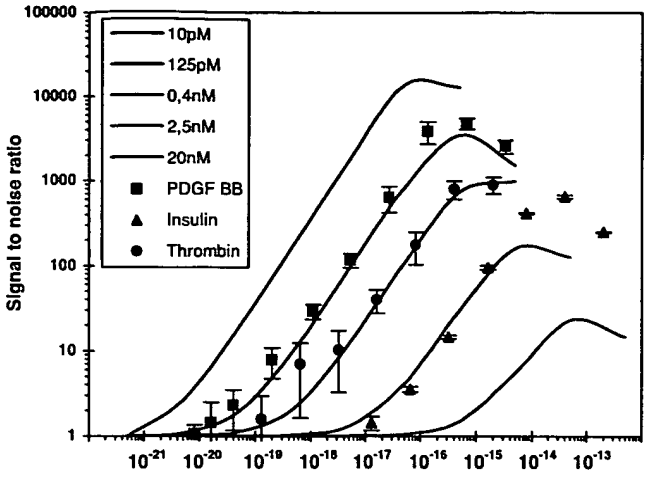


Fig. 2. Triplicate measurements of dilutions of PDGF-BB (squares), thrombin (filled circles), and insulin (triangles), with standard error of the mean, compared to theoretical standard curves for probes with K_d, from left to right, of 10 pM, 125 pM, 0.4 nM, 2.5 nM, and 20 nM, respectively (lines).

flushed with a 95% nitrogen/5% carbon dioxide gas mixture until the O_2 level reached 0.5%, as measured by a PacIII oxygen instrument (Dräger, Partille, Sweden). Cells were then incubated at 37°C for another 24 h in hypoxia. As a control, cells were maintained in normoxia (20.9% O_2). Conditioned media were collected and analyzed for VEGF levels.

VEGF ELISA. The R & D Systems DuoSet ELISA Development Kit was used according to the manufacturer's instructions and within the reported working range of 0.375–95 pM VEGF in 100-μl samples.

Results

Preparation of Proximity Probes. We have established two procedures to convert small amounts of antibodies to proximity probes and used these for highly sensitive homogenous protein detection. Oligonucleotides with free 3′ or 5′ ends were attached directly to antibodies by using the bifunctional crosslinker succinimidyl 4-[*p*-maleimidophenyl]butyrate. Alternatively, oligonucleotides were first coupled to STV for subsequent binding to biotinylated antibodies. Covalent protein–oligonucleotide conjugates were purified from unreacted components before use, because DNA–protein conjugation is typically of low yield. In the case of the STV–oligonucleotide conjugates, purification was performed by using two selective precipitations, first of protein material by using ammonium sulfate and then nucleic acid using NaAc/ethanol. Quantitative analysis of the two purified conjugates revealed that on average around one oligonucleotide was attached per STV molecule [7.7 μM STV with 6.8 μM oligonucleotide-A(3′SH) and 2.2 μM STV and 3.5 μM oligonucleotide-B(5′SH)]. Microfluidic capillary electrophoresis performed on an Agilent BioAnalyzer revealed that 90% of all DNA was attached to STV for the purified 3′-STV and 80% for the 5′-STV (Fig. 5, which is published as supporting information on the PNAS web site). Because no chromatography steps are involved, it is easy to scale the process for batch purification of many different STV–oligonucleotide conjugates. Once prepared, batches of purified STV–oligonucleotide conjugates represent universal reagents that are simply combined with any biotinylated antibodies of the desired specificity to form proximity probes ready for use without further purification. The prepared proximity probes are then stable over several months of storage.

Correlation Between Probe-Affinity and Assay Sensitivity. The proportion of target proteins bound by a pair of proximity probes at equilibrium can be estimated if the concentration of reagents and the K_d for the interactions are known. By taking into account the background signal observed in the absence of target proteins, these calculations provide estimates of signal over background ratios for various target concentrations, representing theoretical standard curves. The background was empirically measured by varying the concentration of two ligatable oligonucleotide [(B)-3′ and (B)-5′) in 5 μl] incubations, ligated and amplified with sequence-system B. As expected, increasing the concentration of one of the probes five times resulted in a 5-fold increase (4.57 ± 0.62) in background, whereas a 5-fold increase of both probes yielded an ≈25-fold higher background (23.4 ± 3.2-fold).

In Fig. 2, we display estimated standard curves, assuming probe–target interactions with the indicated dissociation constants. These estimates are compared with experimental results from detection of PDGF-BB, thrombin, and insulin. The PDGF-BB aptamers have a reported affinity of 129 ± 11 pM (8), whereas the thrombin aptamers are ≈1 nM (9, 10). The PDGF-BB and thrombin data using SELEX aptamers are from Fredriksson *et al.* (2). Proximity ligation signals increase linearly with increasing target up to a point where the probability of each target molecule being bound by two probes decreases. This point depends on the affinity of the particular probes used and their concentration. Also included are data generated by using two anti-insulin monoclonal antibodies that form a proximity probe pair after covalent succinimidyl 4-[*p*-maleimidophenyl]butyrate coupling of oligonucleotides directly to the antibodies (K_d ≈10 nM; P. Lindstedt, personal communication). The proximity ligation assay for insulin has a sensitivity of 30 pM in 1-μl samples, whereas the detection limit using these antibodies in a 25-μl ELISA is 6 pM (standard assay; Mercodia, Uppsala, Sweden) or 0.42 pM (ultrasensitive assay, Mercodia). The PDGF-BB experimental data closely match the 125 pM theoretical standard curve, whereas the thrombin and insulin data fit the expected results for curves calculated for reagents with a K_d of 0.4 and 2.5 nM, respectively. Probe affinities and assay performance are thus strongly correlated, demonstrating that proximity ligation reactions can also be used to estimate affinities of biomolecular interactions. Moreover, the method could be used to study the dynamics of binding by revealing on and off rates (S.M.G., S.F., and U.L., unpublished data), as well as to characterize inhibitors of protein–protein interactions (S.F. and U.L., unpublished data).

Theoretical Setup of Proximity Ligation Assays and Reagent Purity. To determine whether the concentrations of proximity probes applied in the incubation should be adjusted on the basis of their affinities, we calculated the expected signal over background over a range of dissociation constants (see Table 2, which is published as supporting information on the PNAS web site). We found that with probes having K_d values between 0.1 and 10 nM, it is suitable to use a fixed low amount of both probes. However, enough probes should be used in the assay to generate a stable protein-independent background in a range where real-time PCR offers high precision. This is achieved with ≈50–500 amplicons, corresponding to 5–25 pM of the proximity probes in a 5-μl incubation volume, ligated and amplified in 50 μl.

Reagent purity is of importance for assay performance. Impurities derived from proximity probe generation, such as free antibodies and free oligonucleotides, should be removed by purification. High levels of free antibody are expected to reduce the signal by blocking probe binding, but lower levels are not harmful because the assay operates below target saturating conditions. By contrast, free oligonucleotides as well as proximity probes with inactive protein binders reduce assay performance by raising the background (Table 3, which is published as

Alamar Exhibit 1002
Alamar v. Olink

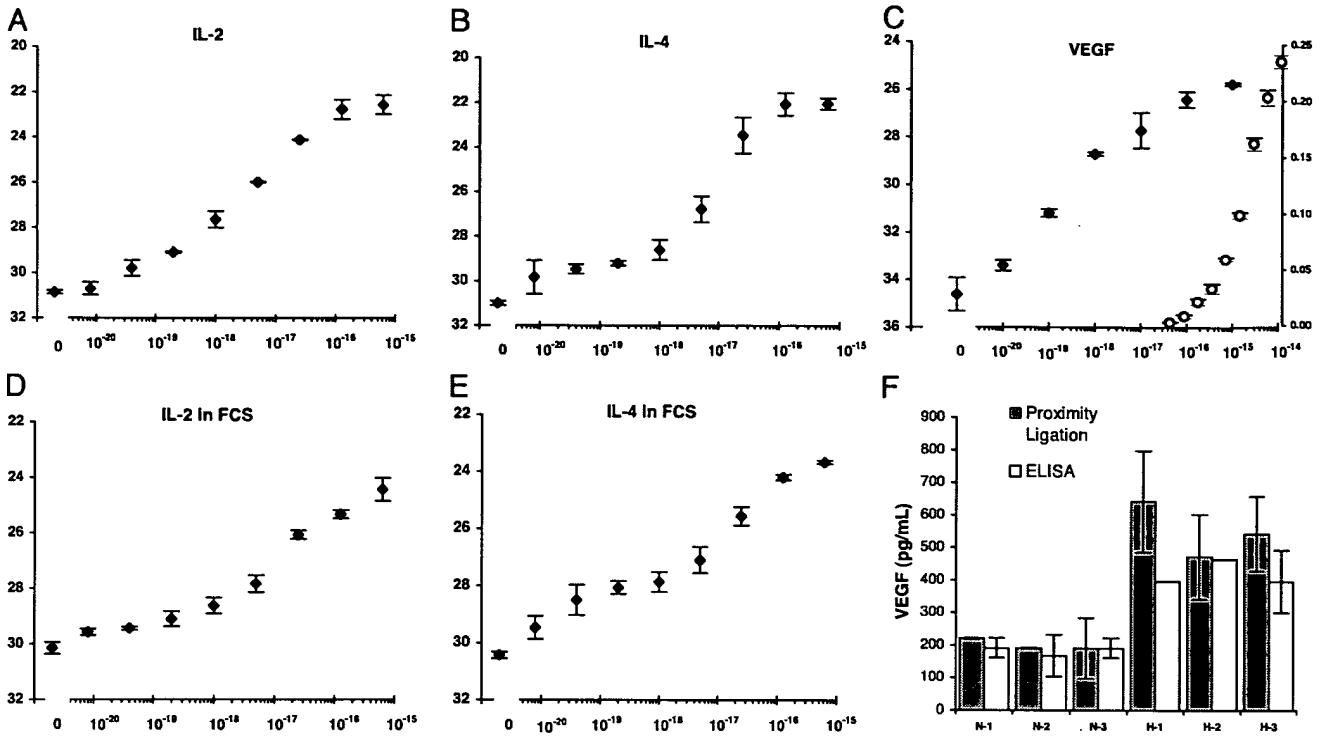


Fig. 3. Cytokine detection by polyclonal antibody-based proximity ligation. (*A* and *B*) Analyses of IL-2 and IL-4 in PBS. (*D* and *E*) Analyses were performed in the presence of 100% FCS. Comparison of proximity ligation- (diamonds) and ELISA-based (circles) detection of VEGF is shown. *x* axes display the molar amount of target protein present in 1-μl samples for proximity ligation and 100-μl samples for ELISA; *y* axes display the cycle threshold values from real-time PCR assays or absorbance at 450 nm for the ELISA. A cycle threshold value (C_T) of 30 corresponds to 250 ligation events. (*F*) Comparison between VEGF measurement by proximity ligation and ELISA. Triplicate samples of conditioned media of cultures grown under either normoxia (N 1–3) or hypoxia (H 1–3) culture conditions. Shown are triplicate measurements with standard deviations for *A*, *B*, *D*, and *E* and duplicates for *C* and *F*.

supporting information on the PNAS web site). It is also important that the oligonucleotides used in proximity probes are full length, and that their sequences have been selected to avoid secondary structures that could prevent hybridization of the connector oligonucleotide and formation of interprobe hybrids.

Affinity-Purified Polyclonal Antibodies. Because polyclonal antisera typically contain antibodies against different epitopes of the same antigen, we assumed proximity probe pairs could be formed by dividing such sera into two portions and coupling oligonucleotides with free 5′ and 3′ ends to the two portions. Accordingly, affinity-purified and biotinylated polyclonal antibodies directed against two monomeric proteins, IL-2 and IL-4, and one homodimeric protein, VEGF, were split in two aliquots each, which were combined with two preparations of STV conjugated to oligonucleotides with ligatable ends. As shown in Fig. 3, all assays displayed low femtomolar sensitivities in 1-μl samples, thus detecting tens of zeptomol of protein. Because the polyclonal antibodies comprise a heterogeneous population of probes of various affinities, the proportionality of increasing signal with increasing target concentration is not as prominent as when monoclonal probes are used. The IL-2 and IL-4 assays were also performed in the presence of 100% FCS without influencing the limit of detection. Compared to a commercial ELISA, the VEGF assay detected 5,000-fold less protein molecules, and it exhibited a wider dynamic range. Cell culture supernatants of embryonic kidney cells, induced to overexpress VEGF by hypoxia and normoxic controls, were analyzed for VEGF content by proximity ligation and ELISA (Fig. 3*f*). The amount of VEGF was estimated, by reference to standard curves, with a precision of 21% coefficient of variance for proximity ligation and 16% for ELISA.

A number of lysed glomeruli tissue samples from conditional PDGF-BB knockout mice and controls were analyzed by proximity ligation by using a polyclonal antibody-based assay and compared to previously obtained results using DNA aptamers (Fig. 4). Samples and aptamer-based data are from Bjarnegård *et al.* (6). Results using antibody- and aptamer-based proximity ligation assays correlate well. By contrast, the sensitivity of conventional ELISAs is insufficient to measure the low amounts of cytokines in some of these minute samples. In the antibody-based assay, the intercalating dye SYBR green I was substituted for the TaqMan probe, illustrating its utility as alternative reporter.

Although it is possible to predict the assay performance for directly conjugated proximity probes of known affinity, this is not true for reagents having oligonucleotides attached via STV conjugates. The latter reagents can form multivalent complexes due to the tetrameric structure of STV and the multiple biotins groups coupled per antibody (11, 12). Accordingly, each proximity probe can involve several antibodies and STV molecules, increasing the apparent affinity by cooperative effects.

Discussion

We show herein that the readily available protein-binding reagents, mono- and polyclonal antibodies, can be conveniently converted to proximity probes. Polyclonal reagents eliminate the need to select matched antibody pairs toward each target protein, while maintaining the element of dual recognition to ensure high specificity. The greater the affinity of the binding reagents, the higher the assay sensitivity, because the assays are

147

Alamar Exhibit 1002
Alamar v. Olink

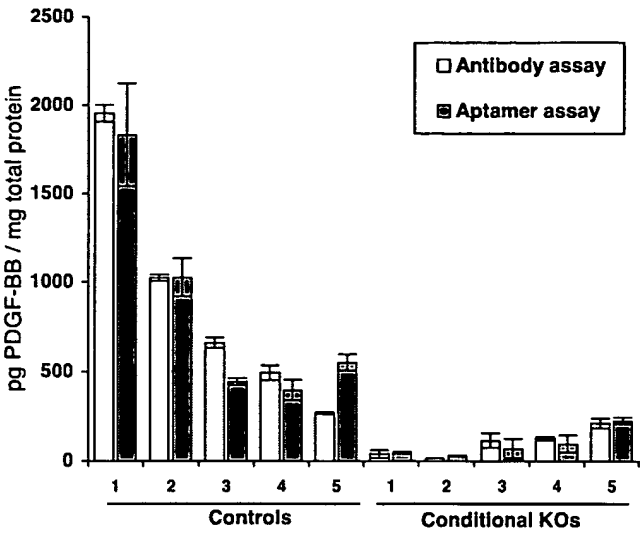


Fig. 4. Proximity ligation analysis of PDGF-BB from lysed mouse glomeruli tissues of controls (1–5) and conditional knock-outs (1–5) using DNA aptamers or polyclonal antibodies as affinity reagents.

performed below target-saturating conditions to minimize background. Antibodies subjected to *in vitro* maturation can reach exceptionally high affinities with K_d of tens of picomolar (13, 14) or even femtomolar (15). Any 10-fold increase in affinity of both members of a proximity probe pair is expected to result in a 100-fold increase in sensitivity (compare Fig. 2).

Homogenous proximity ligation is suitable for automation in high-throughput applications, because no washing steps are required but only two sequential additions to the incubation of first the sample and then a ligation/PCR mixture. The high assay sensitivity allows 1-μl sample aliquots to be monitored by proximity ligation, reducing sample consumption and enabling analysis of samples available only in very small amounts and previously not analyzable by traditional techniques, as illustrated by the mouse glomeruli assays shown herein. Also, 1,000-fold less antibody is used per assay compared to standard ELISAs, and because all assays perform favorably at similar reagent concentrations, new assays do not require extensive optimization.

The precision of proximity ligation is currently at the level of real-time PCR detection, but improved quantitative detection strategies for nucleic acids may offer a further increase in precision (16–18). Multiplexed detection is the goal for many technologies under development, especially for antibody-based microarrays. As more detection reactions are performed in parallel, the issue of antibody crossreactivity becomes an increasing problem limiting scalability (19). Proximity ligation offers a possible solution to this problem if unique ligation junctions are used for each cognate proximity probe pair. The proximity ligation mechanism is also being adapted for *in situ* detection of target proteins and for high-throughput protein–protein interaction studies (O. Söderberg, M.G., M. Emilsson, F. Bahram, L. G. Larsson, and U.L., unpublished work; S.M.G., S.F., and U.L., unpublished work). Finally, by encoding different proteins as amplifiable DNA tag sequences through proximity ligation, parallel analyses may be possible, allowing standard oligonucleotide arrays to be used for absolute or relative measurements of large sets of proteins.

We thank Stefan Wennström and Maria Hägg (Rudbeck Laboratory, Uppsala University) for kindly providing the VEGF ELISA data and the conditioned media samples. This work was supported by grants from the Wallenberg and Beijer Foundations; Polysaccharide Research AB (Uppsala); the Swedish Foundation for Strategic Research; the Graduate Research School in Genomics and Bioinformatics; and the Swedish Research Council for Medicine, Natural Sciences and Technology.

1. Gullberg, M., Fredriksson, S., Taussig, M., Jarvius, J., Gústafsdóttir, S. & Landegren, U. (2003) *Curr. Opin. Biotechnol.* **14**, 82–86.
2. Fredriksson, S., Gullberg, M., Jarvius, J., Olsson, C., Pietras, K., Gústafsdóttir, S. M., Ostman, A. & Landegren, U. (2002) *Nat. Biotechnol.* **20**, 473–477.
3. Sano, T., Smith, C. L. & Cantor, C. R. (1992) *Science* **258**, 120–122.
4. Rollman, O., Jensen, U. B., Ostman, A., Bolund, L., Gústafsdóttir, S. M. & Jensen, T. G. (2003) *J. Invest. Dermatol.* **120**, 742–749.
5. Lindblom, P., Gerhardt, H., Liebner, S., Abramsson, A., Enge, M., Hellstrom, M., Backstrom, G., Fredriksson, S., Landegren, U., Nystrom, H. C., *et al.* (2003) *Genes Dev.* **17**, 1835–1840.
6. Bjarnegård, M., Enge, M., Norlin, J., Gústafsdóttir, S., Fredriksson, S., Abramsson, A., Takemoto, M., Gustafsson, E., Fassler, R. & Betsholtz, C. (2004) *Development (Cambridge, U.K.)* **131**, 1847–1857.
7. Brody, E. N., Willis, M. C., Smith, J. D., Jayasena, S., Zichi, D. & Gold, L. (1999) *Mol. Diagn.* **4**, 381–388.
8. Green, L. S., Jellinek, D., Jenison, R., Ostman, A., Heldin, C. H. & Janjic, N. (1996) *Biochemistry* **35**, 14413–14424.
9. Tasset, D. M., Kubik, M. F. & Steiner, W. (1997) *J. Mol. Biol.* **272**, 688–698.
10. Tsiang, M., Gibbs, C. S., Griffin, L. C., Dunn, K. E. & Leung, L. L. (1995) *J. Biol. Chem.* **270**, 19370–19376.
11. Niemeyer, C. M., Adler, M., Pignataro, B., Lenhert, S., Gao, S., Chi, L., Fuchs, H. & Blohm, D. (1999) *Nucleic Acids Res.* **27**, 4553–4561.
12. Niemeyer, C. M., Adler, M., Gao, S. & Chi, L. (2001) *Bioconjugate Chem.* **12**, 364–371.
13. Chen, Y., Wiesmann, C., Fuh, G., Li, B., Christinger, H. W., McKay, P., de Vos, A. M. & Lowman, H. B. (1999) *J. Mol. Biol.* **293**, 865–881.
14. Hanes, J., Schaffitzel, C., Knappik, A. & Pluckthun, A. (2000) *Nat. Biotechnol.* **18**, 1287–1292.
15. Boder, E. T., Midelfort, K. S. & Wittrup, K. D. (2000) *Proc. Natl. Acad. Sci. USA* **97**, 10701–10705.
16. Ding, C. & Cantor, C. R. (2003) *Proc. Natl. Acad. Sci. USA* **100**, 3059–3064.
17. Nilsson, M., Gullberg, M., Dahl, F., Szuhai, K. & Raap, A. K. (2002) *Nucleic Acids Res.* **30**, e66.
18. Dahl, F., Baner, J., Gullberg, M., Mendel-Hartwig, M., Landegren, U., Nilsson, M. (2004) *Proc. Natl. Acad. Sci. USA* **101**, 4548–4553.
19. Abbott, A. (2002) *Nature* **415**, 112–114.

Alamar Exhibit 1002
Alamar v. Olink

This Page is Inserted by IFW Indexing and Scanning Operations and is not part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☐ BLACK BORDERS
- ☐ IMAGE CUT OFF AT TOP, BOTTOM OR SIDES
- ☑ FADED TEXT OR DRAWING
- ☐ BLURRED OR ILLEGIBLE TEXT OR DRAWING
- ☐ SKEWED/SLANTED IMAGES
- ☐ COLOR OR BLACK AND WHITE PHOTOGRAPHS
- ☐ GRAY SCALE DOCUMENTS
- ☐ LINES OR MARKS ON ORIGINAL DOCUMENT
- ☐ REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY
- ☐ OTHER: _____

IMAGES ARE BEST AVAILABLE COPY.
As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.

Alamar Exhibit 1002
Alamar v. Olink

Downloaded from www.genome.org on September 14, 2006

GENOME RESEARCH

A Versatile Assay for High-Throughput Gene Expression Profiling on Universal Array Matrices

Jian-Bing Fan, Joanne M. Yeakley, Marina Bibikova, Eugene Chudin, Eliza Wickham, Jing Chen, Dennis Doucet, Philippe Rigault, Baohong Zhang, Richard Shen, Celeste McBride, Hai-Ri Li, Xiang-Dong Fu, Arnold Oliphant, David L. Barker and Mark S. Chee

Genome Res. 2004 14: 878-885
Access the most recent version at doi:10.1101/gr.2167504

References This article cites 25 articles, 12 of which can be accessed free at:
http://www.genome.org/cgi/content/full/14/5/878#References

Article cited in:
http://www.genome.org/cgi/content/full/14/5/878#otherarticles

Email alerting service Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or **click here**

Notes

To subscribe to *Genome Research* go to:
http://www.genome.org/subscriptions/

© 2004 Cold Spring Harbor Laboratory Press

Alamar Exhibit 1002
Alamar v. Olink

Downloaded from www.genome.org on September 14, 2006

A Versatile Assay for High-Throughput Gene Expression Profiling on Universal Array Matrices

Jian-Bing Fan,[1,3] Joanne M. Yeakley,[1] Marina Bibikova,[1] Eugene Chudin,[1] Eliza Wickham,[1] Jing Chen,[1] Dennis Doucet,[1] Philippe Rigault,[1] Baohong Zhang,[1] Richard Shen,[1] Celeste McBride,[1] Hai-Ri Li,[2] Xiang-Dong Fu,[2] Arnold Oliphant,[1] David L. Barker,[1] and Mark S. Chee[1]

[1]*Illumina, Inc., San Diego, California 92121, USA;* [2]*Department of Cellular and Molecular Medicine, University of California, San Diego, La Jolla, California 92093, USA*

We report a flexible, sensitive, and quantitative gene-expression profiling system for assaying more than 400 genes, with three probes per gene, for 96 samples in parallel. The cDNA-mediated annealing, selection, extension and ligation (DASL) assay targets specific transcripts, using oligonucleotides containing unique address sequences that can hybridize to universal arrays. Cell-specific gene expression profiles were obtained using this assay for hormone-treated cell lines and laser-capture microdissected cancer tissues. Gene expression profiles derived from this assay were consistent with those determined by qRT-PCR. The DASL assay has been automated for use with a bead-based 96-array matrix system. The combined high-throughput assay and readout system is accurate and efficient, and can cost-effectively profile the expression of hundreds of genes in thousands of samples.

Microarray technology has been widely used in large-scale genomic research, especially in whole-genome gene expression profiling (Lockhart et al. 1996; Lipshutz et al. 1999, Hughes et al. 2001; Holloway et al. 2002; Nuwaysir et al. 2002). Comparison of gene expression patterns in different cell types, developmental stages, and disease states should enable the discovery of characteristic gene-expression patterns that can be associated with functionally important states. For example, microarray-based tumor classification (Golub et al. 1999; Perou et al. 2000; Welsh et al. 2001), as well as treatment response and clinical outcome prediction (Dhanasekaran et al. 2001; West et al. 2001; van 't Veer et al. 2002) have been demonstrated in many cancer types. Now, the challenges are in validating molecular profiles derived from genome-wide microarray experiments in large clinical sample sets, and translating these findings to the clinic (Chuaqui et al. 2002). Such a transition would have profound implications for both basic research and clinical medicine, but is currently hampered by a lack of suitable technologies. Whole-genome arrays are expensive and cumbersome for large numbers of samples. Although quantitative RT–PCR (qRT-PCR) has been in routine use to assay one or a few genes, more flexible and high-throughput validation approaches are required to enable the screening of hundreds of potential biomarkers in large sets of clinical samples.

We previously developed a ligation-based RNA assay for parallel analysis of ~100 mRNA splicing variants (Yeakley et al. 2002). The assay, dubbed RASL (for RNA-mediated annealing, selection, and ligation), permits analysis of mRNA transcripts without prior RNA purification or cDNA synthesis. Here, we describe improvements to the RASL assay for broader RNA-profiling applications. The new DASL assay differs in important respects from the RASL assay. First, the DASL assay converts the RNA target to cDNA, without sacrificing sensitivity. Using a cDNA template offered technical advantages such as enabling analysis of significantly degraded mRNAs, and facilitated the automation of the assay. Second, the DASL assay uses locus-specific oligonucleotide extension and ligation, which resulted in increased specificity compared with the oligonucleotide ligation scheme used in RASL.

The DASL assay is currently multiplexed to detect over 1200 sequence targets simultaneously, and is carried out on 96 samples in parallel, using a matrix of optical fiber bundles (the Sentrix array matrix; Oliphant et al. 2002). Because of its multiplexed assay format, universal-array readout, and high sample throughput, the DASL assay system fills a gap between the existing RNA-profiling technologies of qPCR and gene-specific oligonucleotide or cDNA microarrays, and is complementary to those approaches.

RESULTS

The DASL Assay

A diagram illustrating the biochemistry of the DASL assay is shown in Figure 1. First, total RNA is converted to cDNA using biotinylated primers (both oligo-d$(T)_{18}$ and random hexamers). The biotinylated cDNA is then attached to a streptavidin solid support and assay oligonucleotides are annealed to their target sequences in the cDNA. A pair of oligonucleotides is annealed to each target site. Three to 10 sites are targeted per gene, and up to 1536 oligonucleotide pairs are multiplexed together in a single reaction. High locus specificity is achieved in two ways, first, by the requirement that both members of an oligonucleotide pair must hybridize in close proximity for an assay signal to be generated, and second, by the removal of excess and mishybridized oligonucleotides by washing after the annealing step.

Annealed upstream oligonucleotides are extended and ligated to their corresponding downstream oligonucleotides to create PCR templates that can be amplified with common primers (P1 and P2). In contrast, the RASL assay is based on locus-specific oligonucleotide ligation (Yeakley et al. 2002). We have found that locus-specific oligonucleotide extension and ligation increases signal-to-noise ratio (Abravaya et al. 1995; Fan et al. 2003). Another advantage is that placement of a gap between the

[3]**Corresponding author.**
E-MAIL jfan@illumina.com; FAX (858) 202-4680.
Article and publication are at http://www.genome.org/cgi/doi/10.1101/gr.2167504.

14:878–885 ©2004 by Cold Spring Harbor Laboratory Press ISSN 1088-9051/04; www.genome.org

Alamar Exhibit 1002
Alamar v. Olink

Downloaded from www.genome.org on September 14, 2006

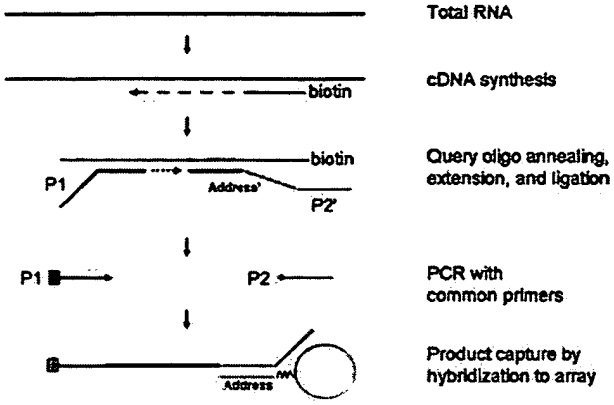


Figure 1 DASL: a cDNA-based assay for RNA profiling. RNA is converted to cDNA, using biotinylated oligo-d(T)$_{18}$ and random hexamers, and immobilized to a streptavidin-coated solid support. Two oligonucleotides are designed to interrogate each target site on the cDNA. The upstream oligonucleotide consists of two parts, the gene-specific sequence and a universal PCR primer sequence (P1) at the 5′-end. The downstream oligonucleotide consists of three parts, the gene-specific sequence, a unique address sequence which is complementary to a capture sequence immobilized on the array, and a universal PCR primer sequence (P2′) at the 3′-end. A single address sequence is uniquely associated with a single target site. The upstream oligonucleotide hybridizes to the targeted cDNA site, extends and ligates to its corresponding downstream oligonucleotide to create a PCR template that can be amplified with universal PCR primers (P1 and P2). The PCR products, which are fluorescently labeled by incorporation of the 5′-labeled primer P1, are hybridized to capture sequences on the beads in the array, and fluorescence intensity is measured for each bead.

upstream and downstream oligonucleotides (typically 1 to 20 bases) provides flexibility in positioning the probes to avoid unfavorable sequences. Both RASL and DASL formats use one PCR primer pair to amplify all of the targets and generate amplicons of ~100 bp. This uniformity results in a relatively unbiased amplification of the PCR template population (Yeakley et al. 2002).

Impact of Multiplexing Level and RNA Input

To assess the impact of different multiplexing levels, we performed experiments with query oligos targeting 281, 665, and 1141 sequence sites. The lower multiplexed oligonucleotide pools represented subsets of the 1141-plex pool. Our results showed that all of the 1141-plex-specific genes were detected only in the 1141-plex reaction, but not in the 281-plex and 665-plex reactions, indicating that the assay is very specific. Consistent expression profiles were obtained with all three multiplexing levels for the shared genes as follows: 281-plex versus 1141-plex ($R^2 = 0.95$) and 665-plex versus 1141-plex ($R^2 = 0.98$). Currently, the number of sequences that can be measured is limited by the number of unique address sequences on each array.

We have systematically measured DASL assay performance as a function of RNA input. Using an oligonucleotide pool targeting 270 mouse genes related to immunological pathways, we assayed total RNAs isolated from two mouse cell lines, EL-4 (T-cell) and A-20 (B-cell). Assay performance was assessed at 1212-plex with RNA input at 100, 75, 50, 25, 10, and 5 ng of total RNA from each cell line. As summarized in Table 1, decreasing the input RNA from 100 to 5 ng resulted in a ~10% reduction (78% vs. 70%) in gene expression detectability and approximately twofold reduction (95% vs. 41%) of genes for which smaller than a twofold difference could be detected between the two cell lines. In general, adequate performance was obtained with 25 ng of total RNA under these conditions.

Because the assay output depends on the relative abundance of the PCR templates (ligated query oligonucleotides), the minimum amount of RNA needed as input to the assay varies according to the expression levels of the targeted transcripts. Therefore, assay sensitivity depends on which transcripts are targeted by the query oligonucleotide pool.

Assay Dynamic Range and Differential Expression Detection

We derived dose-response relationships by spiking different amounts of six synthetic RNAs into a background of mouse total RNA. The six synthetic RNAs were pooled using a permuted matrix experimental design, and each pool was spiked into a background of 100 ng of mouse total RNA. The six synthetic RNAs were transcribed in vitro from plasmids containing prokaryotic gene sequences (*cat*, *cre*, *gus*, *lacZ*, *luc*, and *bla*, respectively) with attached poly(A)$_{30}$ tails. The spiked RNAs were assayed along with the 270 mouse immunology genes (see above). The amount of the synthetic RNAs spiked into the DASL assay ranged from 1×10^4 to 1×10^7 molecules, with a threefold difference for a given RNA in adjacent pools. As a control, the *bla* RNA was kept constant in all pools at 3×10^5 molecules. Each experiment was done in six replicates in order to generate statistics. This experimental design allowed us to measure assay performance with regard to dose-response and fold-difference detection over a range of target concentrations. As shown in Figure 2, the DASL assay has a dynamic range of at least 2.5 orders of magnitude (from 3×10^4 to 1×10^7 molecules or more), in which an average of 1.8-fold difference detection (range 1.3–2.4) was obtained with 95% confidence.

In interpreting these data, it is important to realize that the assay output for one target site is a function of the fraction of the amplifiable pool (total PCR template) that comprises that target. We use a fixed PCR cycle number for all samples, and the PCR reagents are shared among all amplicons in each reaction. As a result, the limit of detection for a given transcript in the DASL assay depends on its abundance relative to other transcripts being targeted in the assay. Given this caveat, data from RNA dilution studies suggest that the assay can quantitatively detect $<1 \times 10^4$ RNA molecules at ~1200-plex levels. If fewer targets are measured, better sensitivity can be achieved (see Discussion).

Table 1. DASL Performance as a Function of RNA Input

RNA Input	100 ng	75 ng	50 ng	25 ng	10 ng	5 ng
% genes "on"	78	76	74	73	70	70
% genes with <2-fold difference detectable	95	80	84	82	59	41
Correlation (R^2) compared to 100 ng	1.000	0.986	0.983	0.970	0.946	0.918

Varying levels of total RNAs from A-20 and EL-4 cells were input to DASL assays. Values shown for detection and reproducibility are the average of those for each cell line. The two cell lines were compared for the proportion of genes differentially expressed as a function of input.

Alamar Exhibit 1002
Alamar v. Olink

Downloaded from www.genome.org on September 14, 2006

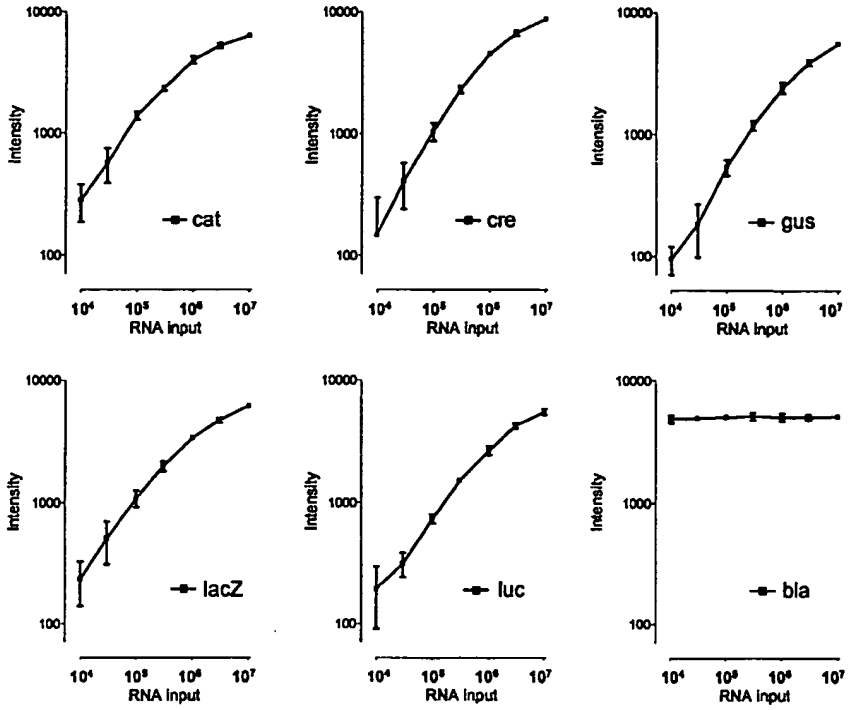


Figure 2 DASL assay performance, dynamic range and differential expression detection. Intensity data from six synthetic RNAs (*cat, cre, gus, lacZ, luc,* and *bla*) are shown as log intensity vs. log RNA input (molecules) in the DASL assay. The RNA input is given as molecules of spiked transcript per DASL reaction. The error bars represent the 95% confidence interval of intensity values for six replicate assays.

DASL Assay Is Consistent With qRT–PCR and Other Array-Based Analysis

Quantitative RT–PCR was used to validate the array experiments. The abundance of 22 housekeeping and cell-specific genes was determined in three total RNA samples (A-20, EL-4, and a 1 : 1 mixture of the two) using both methods. The DASL assay was carried out at 1212-plex and compared with single-plex real time RT–PCR with SYBR Green detection. Figure 3 shows the correlation between the log-fold differences in transcript abundance determined by qRT-PCR and DASL. Although there is some variation, the overall correlation between the two methods shows that the DASL assay can reliably report differences in expression between samples. It is worthwhile to point out that the logarithmic-fold difference in transcript abundance as determined by qRT–PCR was larger than that determined by DASL. This type of underestimating bias has been reported previously for both oligonucleotide arrays and cDNA arrays (Yuen et al. 2002).

In further experiments comparing the DASL assay to other expression techniques, we have compared expression profiles obtained with the DASL assay to those obtained with standard gene-specific probe arrays for the 270 mouse immunology genes, using a common set of RNA samples. Similar results (Speirman rank correlation = 0.83) were obtained with the two assay platforms with regard to the fold differences detected among the samples.

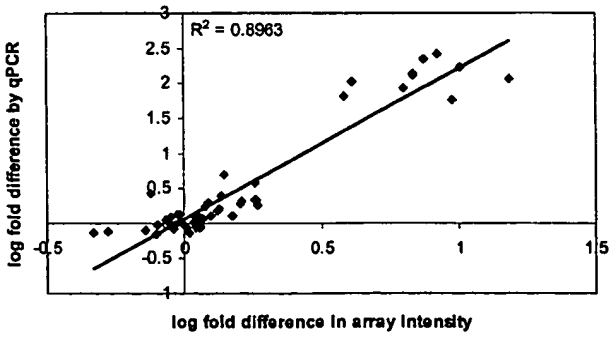


Figure 3 Correlation of DASL with qRT–PCR. The logarithmic fold difference in abundance in pairwise comparisons between A-20, EL-4, and a 1 : 1 mix of the two was estimated for expressed genes in both the DASL assay (fold difference in array intensity, *x*-axis) and qRT–PCR (fold difference in abundance derived from crossover threshold, *y*-axis).

Cell-Specific Response to Hormone Treatment

To determine the sensitivity of the DASL assay to biologically relevant changes in expression, 212 human prostate-related genes (Dhanasekaran et al. 2001; Welsh et al. 2001) were monitored in prostate cell lines treated with androgen or a solvent control. Consistent expression patterns among the cell lines were observed in the technical replicates ($R^2 = 0.99$) and biological replicates ($R^2 = 0.98$). As shown in Figure 4A, there was no detectable impact of androgen on gene expression in cell lines DU-145, RWPE-1, and PC-3. LAPC-4 cells (Fig. 4A, bottom, left) showed a very limited response; four of the 212 genes exceeded a threshold for a significant difference (95% confidence limit). However, in LNCaP cells, 29 genes were down-regulated, and 44 genes were up-regulated as a result of androgen treatment (Fig. 4A, bottom, right). These results are consistent with the known insensitivity of PC-3 cells to androgen and the sensitivity of LNCaP cells to androgen treatment (Mitchell et al. 2000).

The assay intensity data were used to cluster the individual samples on the basis of their expression patterns for the 212 prostate-related genes. As shown in Figure 4B, differences in expression patterns measured in the DASL assay were sufficient to cluster samples from the same cell line together, regardless of androgen response (agglomerative coefficient = 0.98). For samples from the same cell type +/− androgen treatment, two patterns were observed. For cell lines DU-145, RWPE-1, PC-3, and LAPC-4, clustering did not distinguish samples according to treatment as is indicated by the insignificant cluster distance, that is, the Height shown in Figure 4B. In contrast, the androgen-sensitive cell line LNCaP showed a clear segregation. Furthermore, the LNCaP cell line was most similar in expression to the normal prostate tissue samples, consistent with published reports (Mitchell et al. 2000). Therefore, the results of the DASL analysis were concordant with previously observed biological differences, indicating that the assay is sufficiently reproducible and sensitive to sample differences to allow useful comparisons between biological samples.

Alamar Exhibit 1002
Alamar v. Olink

Downloaded from www.genome.org on September 14, 2006

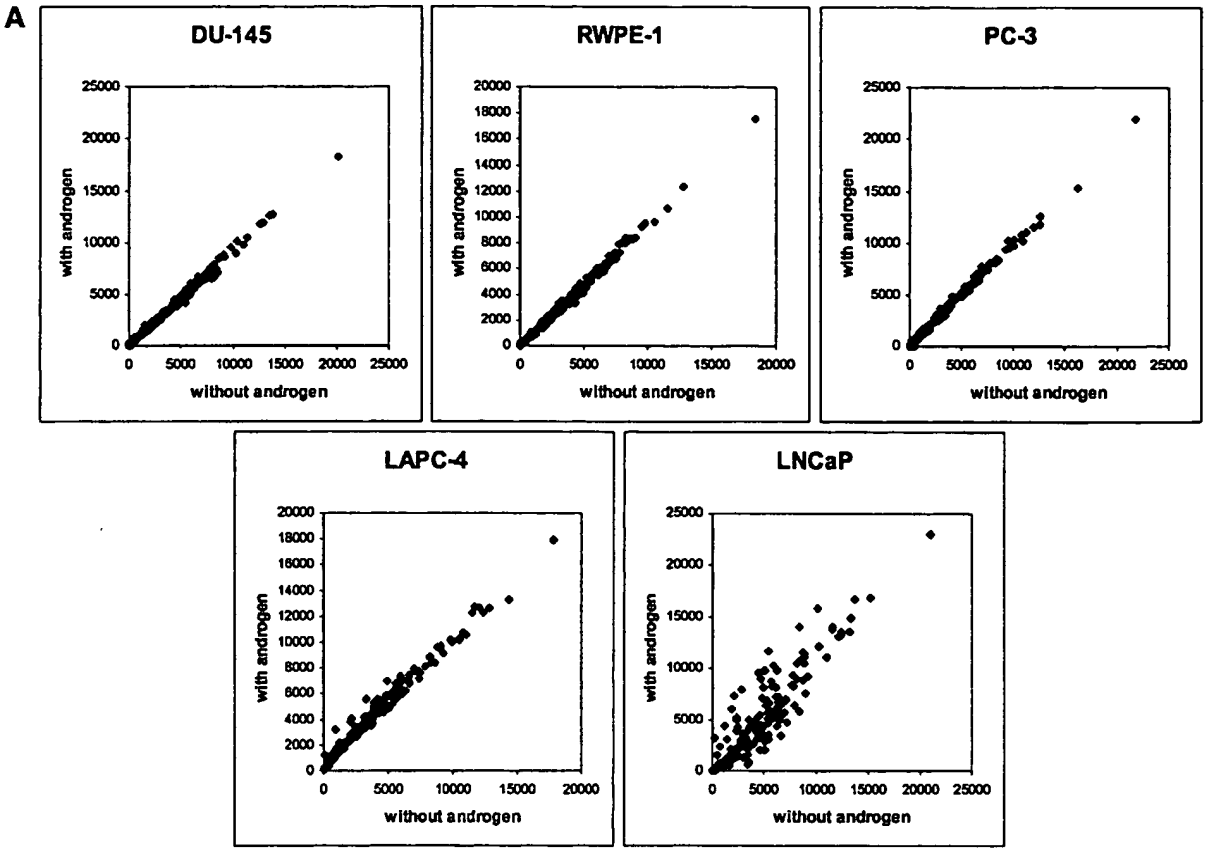


Figure 4 (Continued on next page)

Analysis of Laser-Capture Microdissected Samples

We measured gene expression of the 212 prostate-related genes in laser-capture microdissected (LCM) human prostate samples to determine whether expression data could be reliably derived from such samples (Fig. 5). LCM samples containing different cell numbers were dissected from nearby areas of a sample of benign prostatic hyperplasia. For each LCM sample, three replicate experiments were carried out, each with independent cDNA synthesis, DASL assay processing, and array hybridization. Reproducible results were obtained from as few as 22 microdissected cells' worth of RNA (R^2 = 0.97, 0.96, 0.90, and 0.94 for 176, 88, 44, and 22 cells, respectively). In addition, gene expression profiles obtained with 176, 88, 44, and 22 LCM cells correlate well (R^2 = 0.91, 0.87, and 0.76 for 176 vs. 88, 176 vs. 44, and 176 vs. 22, respectively). Further, 72% of the transcripts detected in the 176-cell samples were also detectable in the 22-cell samples. The two highest expressors in these samples were prostate-specific antigen and prostatic acid phosphatase, consistent with published reports (Mitchell et al. 2000). Varied expression levels were observed for these two genes among the triplicates with lower input (44 cells and 22 cells), consistent with the increased noise observed with low RNA inputs (Fig. 2).

Gene Expression Profiling in Degraded RNA Samples

The DASL assay uses random priming in the cDNA synthesis, and therefore does not depend on an intact poly(A) tail for T7-oligo-d(T) priming (Phillips and Eberwine 1996). In addition, the assay requires a relatively short target sequence of about 50 nucleotides for query oligonucleotide annealing, suggesting that the assay may perform well with significantly degraded RNAs. To test this possibility, RNA aliquots isolated from PC-3 and LNCaP prostate cell lines were heated at 95°C for 5, 10, 20, 40, and 60 min. The RNA was substantially degraded by heating, as shown in the Bioanalyzer traces shown in Figure 6. All RNA samples were converted to cDNA and analyzed by both the DASL assay and qPCR. The fraction of detectable transcripts remaining after degradation was estimated using qRT-PCR with three different primer pairs targeting ~180 bp fragments in the housekeeping genes EF1A1, TAX1BP1, and AGPAT1. As shown in Table 2, incubation at 95°C for 1 h resulted in the loss of ~99.9% of amplifiable species. The same cDNAs were used in a DASL assay measuring the expression of the prostate genes, and highly reproducible results were obtained, with good correlation between technical replicates (Table 2). More than 87% of the genes were detected in the PC-3 sample incubated at 95°C for 1 h compared with the intact RNA sample (Table 2). Correlation between expression profiles in intact RNA and degraded RNA samples was lower, with R^2 ranging from 0.9 to 0.6, probably due to different rates of RNA degradation for different transcripts. Very similar results were obtained for the LNCaP RNA (data not shown).

DISCUSSION

We have developed an automated method for gene expression profiling of >1200 sequence targets (400 genes at 3 probes per gene), for 96 samples in parallel. The assay has a 2.5 log dynamic range, over which an average 1.8-fold difference in RNA abun-

Alamar Exhibit 1002
Alamar v. Olink

Downloaded from www.genome.org on September 14, 2006

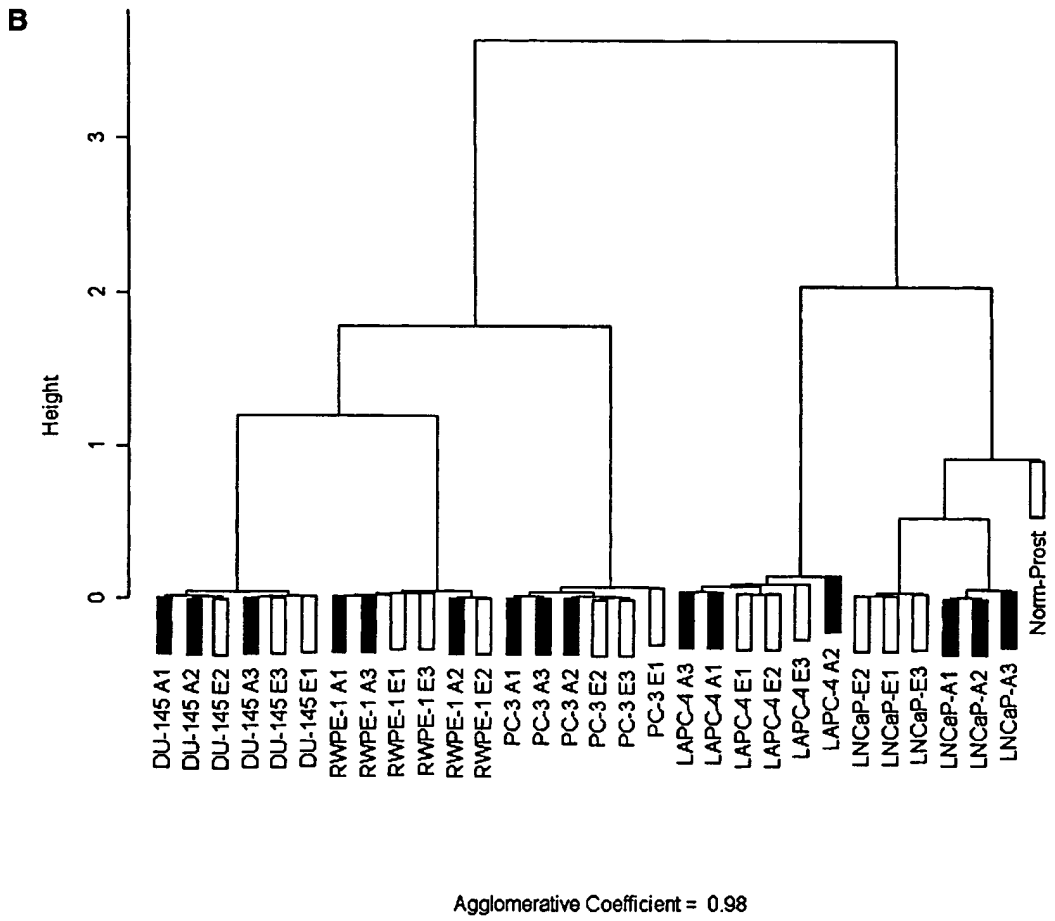


Figure 4 Cell-specific responses to androgen treatment. (*A*) Expression of 212 human prostate-related genes was monitored in prostate cell lines treated with androgen or vehicle. Data for each condition were plotted to illustrate the impact of treatment with androgen (*y*-axis) vs. vehicle (*x*-axis). (*B*) Agglomerative clustering using data from the 212 genes. For each cell line, three randomly selected biological replicates are presented for both androgen-treated (filled box) and untreated (open box) conditions. The distance between subclusters (*y*-axis, the height) measures the divergence of their expression patterns.

dance can be detected with 95% confidence. The assay is highly sensitive, permitting microarray analysis of <100 cells. Furthermore, the DASL assay on the Sentrix array matrix platform allows high-throughput monitoring of hundreds of genes in hundreds to many thousands of samples. We have used this assay for several human, mouse, *Arabidopsis*, and maize (Shou et al. 2004) studies, monitoring as many as 400 genes. In one *Arabidopsis* pathogen response study, gene expression analyses of 388 virus-responsive genes were performed on 540 RNA samples isolated from a variety of wild-type, mutant, and transgenic plants at different developmental stages. This study could have readily been scaled to thousands of samples, had that many been available.

In this study, we systematically designed 3, 4, 5, and up to 10 probes for various sets of genes, and assessed the effect of the probe number on assay quantitation. Our results showed that three optimally designed probes performed comparably to four or more probes with regard to their ability to detect expressed genes as well as differential expression among samples. Expression profiles generated with three probes correlate well with those generated with four probes ($R^2 = 0.99$). Further lowering the probe number negatively impacted assay reproducibility. Three probes per gene also allow monitoring of probe concordance in reporting gene expression changes.

Several expression assays with similar features have been described, and the published results support the capability of the DASL assay to profile gene expression (Hsuih et al. 1996; Nilsson et al. 2001) as well as alternative splicing (Yeakley et al. 2002) and allele-specific expression (Baner et al. 2003; M. Bibikova, unpubl.). Compared with other expression analysis methods, the DASL assay is most like a multiplexed version of RT–PCR, in that both assays utilize cDNA synthesis and PCR amplification. However, DASL offers much higher multiplexing capacity due to the use of common PCR primers, and the array-based readout. A key aspect of the DASL assay design is the incorporation of an address sequence, so that labeled products can be read out on a universal array. This approach offers substantial flexibility, because changing the array content is accomplished simply by reassigning the address sequences. As a result, one can refine gene sets iteratively, if desired, because no custom beads or arrays need to be developed.

Another important aspect of the DASL assay is that its sensitivity depends on the gene set. Thus, the limit of detection is more related to the distribution of targeted transcripts than to

Alamar Exhibit 1002
Alamar v. Olink

Downloaded from www.genome.org on September 14, 2006

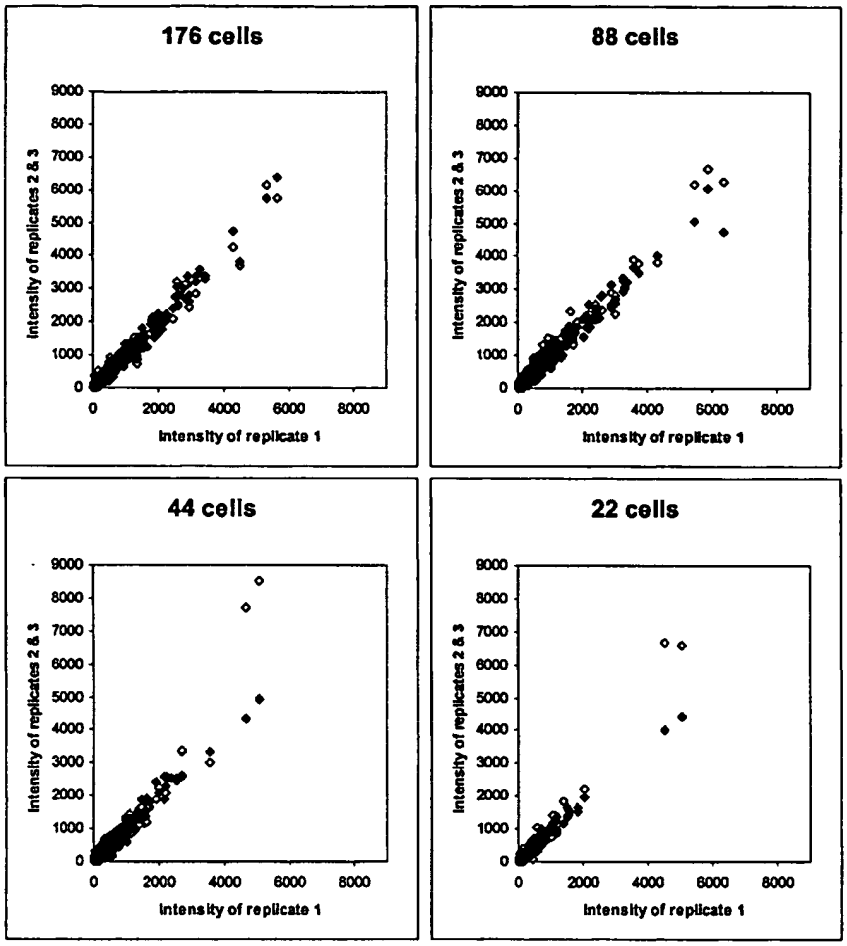


Figure 5 Reproducible gene-expression profiling of laser-captured samples. Total RNAs from laser-captured samples from different numbers of cells were assayed in triplicate. Cell number indicates the number of cells' worth of RNA input to each DASL assay. Correlation between individual replicates is shown by plotting the intensities for two replicates against the intensities of the third.

the mass of input RNA. Nonetheless, as a first approximation for ~400 genes, the assay requires fairly small amounts of starting material; 25 ng of total RNA from complex transcriptomes is sufficient, significantly less than the 5 to 10 µg amounts required by many other microarray-based methods. This level of sensitivity should be sufficient for DASL analyses of RNAs from cell-based screens in 96-well microtiter plates. Multiplexed gene expression analysis has emerged as a viable and efficient approach for high-throughput drug target validation and drug discovery (Johnson et al. 2002). We are developing protocols for expression profiling of cell lysates using the DASL assay, and preliminary data suggest that this approach should be feasible.

Given the competition for PCR reagents in the amplification step, in principle it is possible to construct query oligonucleotide sets that exceed the assay's dynamic range. For example, if a query oligonucleotide set targeted very rare transcripts together with highly abundant species, competition during PCR could squelch the rare transcripts' signal. One possibility for extending the dynamic range could be a strategy in which normalization of differentially expressed genes' representation is accomplished using different PCR primers for abundant transcripts. Our preliminary results using different T_ms for PCR primers targeting abundant transcripts and varying the annealing temperature during PCR showed that this strategy improved assay sensitivity for the less abundant transcripts (data not shown). Of course, using such a strategy requires prior knowledge about expression levels, but if necessary, query oligonucleotide pools could be designed to accommodate vastly different expression levels.

Because the DASL assay targets specific transcripts, it should be insensitive to the complexity of the untargeted RNAs. Consistent with this idea, a cell-dilution experiment showed that cell-specific expression could be detected in RNA from a mixture of one cell in 100 using the DASL assay. Further, selective targeting of transcripts specific to one cell type allowed the detection of one cell against the background of 1000 cells of a different type. These data suggest that DASL can be used to detect and analyze different cell types within a heterogeneous mixture.

A particularly attractive feature of the DASL assay is its tolerance of RNA degradation. The input RNA can be degraded to an average of 100–200 nucleotide fragments and still give robust results (Fig. 6). We have shown that the DASL assay can be used on LCM samples. The assay can also be used on RNA samples isolated from formalin-fixed, paraffin-embedded tissues, with good reproducibility ($R^2 = 0.95$; M. Bibikova, D. Talantov, E. Chudin, J. Yeakley, Y. Wang, and J. Fan, in prep.). Formalin-fixed archival tissues represent an invaluable resource for gene expression analysis, as they are the most widely available materials for studies of human disease. The ability to analyze gene expression in archived tissues will allow not only prospective but also retrospective analyses, because clinical follow up is already available. Therefore, the DASL assay used with the Sentrix array matrix may facilitate research in correlating gene expression profiles with clinical parameters, with the aim of developing biomarkers for use in disease classification and treatment.

METHODS

Assay Probe Design and Test

As shown in Figure 1, for each targeted site, two locus-specific sequences are designed for a T_m of from 57 to 62°C as described previously (Fan et al. 2003). Each target site is designed not to cross exon boundaries. Query oligonucleotides are tested for assay performance using a standard set of samples, including pooled reference RNAs (Stratagene or BD Clontech) for the species of interest. Serial dilutions of RNAs test the dose-response performance of query oligonucleotides.

RNA Samples

RNA samples were isolated from mouse cell lines (EL-4, T-cell and A-20, B-cell) and human prostate cancer cell lines (DU-145,

Downloaded from www.genome.org on September 14, 2006

 Best Available Copy

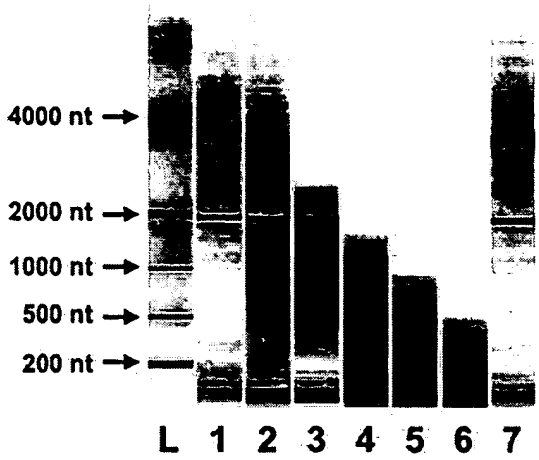


Figure 6 Agilent Bioanalyzer image of PC-3 RNA samples after incubation at 95°C for various periods of time. (L) RNA size ladder; (lanes *1–7*) intact RNA; (lanes *2–6*) 5, 10, 20, 40, and 60 min incubation at 95°C.

RWPE-1, PC-3, LAPC-4, and LNCaP), using a standard TriZol method (Invitrogen). Human prostate samples for LCM were obtained through a collaboration between the UCSD Cancer Center and X.-D. Fu. Synthetic RNAs were transcribed in vitro from plasmids containing prokaryotic gene sequences with attached poly(A) tails.

Real-Time Quantitative RT–PCR

Real-time quantitative RT–PCR analyses were performed on the ABI Prism 7900HT sequence detection system (Applied Biosystems). PCR primers were designed to amplify ~150 bp fragments. Each reaction contained 5 µL of SYBR Green PCR Master Mix (Applied Biosystems), 1 µL of cDNA template, and 250 nM each forward and reverse primer in a total reaction volume of 10 µL. The PCR consisted of an initial enzyme activation step at 95°C for 12 min, followed by 40 cycles of 95°C for 20 sec, 54°C for 20 sec, and 72°C for 30 sec. To assess the final PCR product, a melting curve was generated using a ramp from 60 to 95°C (Applied Biosystems).

cDNA Synthesis and DASL Process

RNA samples were normalized by OD_{260}. Quality testing included analysis by capillary electrophoresis using a Bioanalyzer (Agilent). For most experiments, the following processes were performed on a Tecan Genesis Workstation 150 (Tecan). First, a 20-µL reverse transcription reaction containing a reaction mix (MMC; Illumina) and total RNA (up to 1 µg) was incubated at room temperature for 10 min, and then at 42°C for 1 h. After cDNA synthesis, the remainder of the assay was identical to the GoldenGate assay (Oliphant et al. 2002; Fan et al. 2003), using Illumina-supplied reagents and conditions (BeadLab User Manual, Illumina). Briefly, the cDNA was immobilized on paramagnetic beads and washed. For most experiments, 10% of the immobilized cDNA was processed through the remaining steps. Pooled query oligonucleotides were annealed to the cDNA under a controlled hybridization program, and then washed to remove excess or mishybridized oligonucleotides. Hybridized oligonucleotides were then extended and ligated to generate amplifiable templates. A PCR reaction was performed with fluorescently labeled universal PCR primers. Single-stranded PCR products were prepared by denaturation, then hybridized to a Sentrix array matrix (Yeakley et al. 2002; Fan et al. 2003). The array hybridization was conducted under a temperature gradient program, and arrays were imaged using a BeadArray Reader 1000 scanner (Illumina; Barker et al. 2003). Cy3 labeling was used for all expression analyses.

Array Image Processing and Signal Extraction

Image analysis and data extraction software were as described previously (Oliphant et al. 2002; Galinsky 2003). Briefly, each sequence type is represented by an average of 30 beads on the array. Bead signals are computed with weighted averages of pixel intensities, and local background is subtracted. Sequence-type signal is calculated by averaging corresponding bead signals with outliers removed (using median absolute deviation).

Array Data Normalization

Each oligonucleotide pool contains sequence types designated as negative and positive controls. Negative controls are designed not to have any significant complementarity to naturally occurring sequences. Positive controls are designed to target synthetic RNAs spiked into the sample in fixed quantity. Two major algorithms were used for data normalization. The first, the cubic spline method, was used when the number of up-regulated and down-regulated genes was approximately the same for each point of the signal range (Workman et al. 2002). The normalization uses quantiles of sequence-type signals to fit smoothing B-splines. With the second algorithm, the positive controls method, normalization coefficients (a,b) were computed for each array using iteratively reweighted least-squares fit $y_v = ay_x + b$. Here, y_v, y_x are vectors of intensities of probes corresponding to positive controls on virtual (average of all arrays) and given arrays, respectively.

Expression Analysis

Detection *P*-values were computed using a normal model based on signals of negative controls. For differential expression scores, we dynamically constructed an error model assuming that rank invariant probes (<3% in relative rank change between condition and reference groups) are not differentially expressed. In addition, probes corresponding to negative controls were selected as not differentially expressed. For selected probes, we defined a vector $\vec{S}$ – of separation values for probes identified in the previous step. Here, elements of $\vec{S}$ are defined as follows:

$$S_i = \frac{|\mu_{ref} - \mu_{cond}|}{\sqrt{\frac{\sigma^2_{ref}}{n_{ref}} + \frac{\sigma^2_{cond}}{n_{cond}}}}$$

Table 2. DASL Performance as a Function of RNA Degradation Level

Duration of treatment at 95°C	0 min	5 min	10 min	20 min	40 min	60 min
% RNA detectable by RT–PCR	100	53.02	22.90	5.47	0.75	0.13
Correlation between technical replicates, R^2	1.00	1.00	1.00	0.99	0.98	0.94
Correlation with undegraded RNA, R^2	1.00	0.91	0.83	0.77	0.71	0.61
% of genes detected, compared to undegraded RNA	100	99	98	97	94	88

Alamar Exhibit 1002
Alamar v. Olink

Downloaded from www.genome.org on September 14, 2006

where μ denotes mean probe signal in the reference and condition groups, respectively. The σ denotes standard deviation associated with μ. When either the reference or condition groups contain replicate arrays, σ was computed across replicates. Otherwise, the standard deviation of individual bead intensities carrying a particular probe was used. A score for an individual probe was computed as $\mathrm{sgn}(\mu_{ref} - \mu_{cond}) \times 10 \times \log_{10}(p)$, where p is given by a normal model $N(\mu_S, \sigma_S(I) + r)$ with μ_S being the median of $\overline{S}$, r being a regularization constant, and intensity-dependent variance σ_S being determined through the exponential fit of $|S_i - \mu_S| - A\exp(\lambda I)$. To compute a differential expression score for the gene, we computed scores for all probes corresponding to a given gene, and reported an average score. In addition, we reported a concordance score equal to

$$\frac{|n_+ - n_-|}{|n_+ + n_-|}$$

where $n_{+/-}$ is the number of probes showing up/down-regulation. To determine minimal resolvable fold change, we used piecewise linear approximation of intensity versus concentration. Concentration levels were considered resolvable if corresponding ranges of intensities did not overlap.

Clustering Algorithm

A matrix of correlation coefficients between array signals was computed. Agglomerative nesting was applied using the Agnes function in the R package with Ward's method.

ACKNOWLEDGMENTS

We thank Steven Barnard, Scott Butler, Diping Che, Todd Dickinson, Francisco Garcia, Todd Rubano, Chanfeng Zhao, and Lixin Zhou for their dedicated efforts and contributions to the development of the bead chemistry, imaging systems, and image analysis tools, bioinformatics, and process automation, and Tim McDaniel, Shawn Baker, and Ken Kuhn for useful discussions. This work was supported in part by grants from the National Institutes of Health (R33 CA88351 and R44 CA83398).

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC section 1734 solely to indicate this fact.

REFERENCES

Abravaya, K., Carrino, J.J., Muldoon, S., and Lee, H.H. 1995. Detection of point mutations with a modified ligase chain reaction (Gap-LCR). *Nucleic Acids Res.* **23:** 675–682.

Baner, J., Isaksson, A., Waldenstrom, E., Jarvius, J., Landegren, U., and Nilsson, M. 2003. Parallel gene analysis with allele-specific padlock probes and tag microarrays. *Nucleic Acids. Res.* **31:** e103.

Barker, D.L., Theriault, G., Che, D., Dickinson, T. Shen, R., and Kain, R. 2003. Self-assembled random arrays: High-performance imaging and genomics applications on a high-density microarray platform. *Proc. SPIE* **4966:** 1–11.

Chuaqui, R.F., Bonner, R.F., Best, C.J., Gillespie, J.W., Flaig, M.J., Hewitt, S.M., Phillips, J.L., Krizman, D.B., Tangrea, M.A., Ahram, M., et al. 2002. Post-analysis follow-up and validation of microarray experiments. *Nat. Genet.* **32:** S509–S514.

Dhanasekaran, S.M., Barrette, T.R., Ghosh, D., Shah, R., Varambally, S., Kurachi, K., Pienta, K.J., Rubin, M.A., and Chinnaiyan, A.M. 2001. Delineation of prognostic biomarkers in prostate cancer. *Nature* **412:** 822–826.

Fan, J.B., Oliphant, A., Shen, R., Kermani, B., Garcia, F., Gunderson, K.L., Hansen, M., Steemers, F., Butler, B.L., Deloukas, P., et al. 2003. Highly parallel SNP genotyping. *Cold Spr. Harb. Symp. Biol.* **68:** (in press)

Galinsky, V.L. 2003. Automatic registration of microarray images II: Hexagonal grid. *Bioinformatics* **19:** 1832–1836.

Golub, T.R., Slonim, D.K., Tamayo, P., Huard, C., Gaasenbeek, M., Mesirov, J.P., Coller, H., Loh, M.L., Downing, J.R., Caligiuri, M.A., et al. 1999. Molecular classification of cancer: Class discovery and class prediction by gene expression monitoring. *Science* **286:** 531–537.

Holloway, A.J., van Laar, R.K., Tothill, R.W., and Bowtell, D.D. 2002. Options available—from start to finish—for obtaining data from DNA microarrays II. *Nat. Genet.* **32:** S481–S489.

Hsuih, T.C.H., Park, Y.N., Zaretsky, C., Wu, F., Tyagi, S., Kramer, F.R., Sperling, R., and Zhang, D.Y. 1996. Novel, ligation-dependent PCR assay for detection of hepatitis C virus in serum. *J. Clin. Microbiol.* **34:** 501–507.

Hughes, T.R., Mao, M., Jones, A.R., Burchard, J., Marton, M.J., Shannon, K.W., Lefkowitz, S.M., Ziman, M., Schelter, J.M., Meyer, M.R., et al. 2001. Expression profiling using microarrays fabricated by an ink-jet oligonucleotide synthesizer. *Nat. Biotechnol.* **19:** 342–347.

Johnson, P.H., Walker, R.P., Jones, S.W., Stephens, K., Meurer, J., Zajchowski, D.A., Luke, M.M., Eeckman, F., Tan, Y., Wong, L., et al. 2002. Multiplex gene expression analysis for high-throughput drug discovery: screening and analysis of compounds affecting genes overexpressed in cancer cells. *Mol. Cancer Ther.* **1:** 1293–1304.

Lipshutz, R.J., Fodor, S.P., Gingeras, T.R., and Lockhart, D.J. 1999. High density synthetic oligonucleotide arrays. *Nat. Genet.* **21:** 20–24.

Lockhart, D.J., Dong, H., Byrne, M.C., Follettie, M.T., Gallo, M.V., Chee, M.S., Mittmann, M., Wang, C., Kobayashi, M., Horton, H., et al. 1996. Expression monitoring by hybridization to high-density oligonucleotide arrays. *Nat. Biotechnol.* **14:** 1675–1680.

Mitchell, S., Abel, P., Ware, M., Stamp, G., and Lalani, E. 2000. Phenotypic and genotypic characterization of commonly used human prostatic cell lines. *British Journal of Urology Int.* **85:** 932–944.

Nilsson, M., Antson, D.-O., Barbany, G., and Landegren, U. 2001. RNA-templated DNA ligation for transcript analysis. *Nucleic Acids Res.* **29:** 578–581.

Nuwaysir, E.F., Huang, W., Albert, T.J., Singh, J., Nuwaysir, K., Pitas, A., Richmond, T., Gorski, T., Berg, J.P., Ballin, J., et al. 2002. Gene expression analysis using oligonucleotide arrays produced by maskless photolithography. *Genome Res.* **12:** 1749–1755.

Oliphant, A., Barker, D.L., Stuelpnagel, J.R., and Chee, M.S. 2002. BeadArray technology: Enabling an accurate, cost-effective approach to high-throughput genotyping. *Biotechniques* **32:** S56–S61.

Perou, C.M., Sorlie, T., Eisen, M.B., van de Rijn, M., Jeffrey, S.S., Rees, C.A., Pollack, J.R., Ross, D.T., Johnsen, H., Akslen, L.A., et al. 2000. Molecular portraits of human breast tumours. *Nature* **406:** 747–752.

Phillips, J. and Eberwine, J.H. 1996. Antisense RNA amplification: A linear amplification method for analyzing the mRNA population from single living cells. *Methods* **10:** 283–288.

Shou, H., Bordallo, P., Fan, J.B., Yeakley, J.M., Bibikova, M., Sheen, J., and Wang, K. 2004. Expression of an active tobacco mitrogen activated protein kinase kinase kinase enhances freezing tolerance in transgenic maize. *Proc. Natl. Acad. Sci.* **101:** 3298–3303.

van 't Veer, L.J., Dai, H., van de Vijver, M.J., He, Y.D., Hart, A.A., Mao, M., Peterse, H.L., van der Kooy, K., Marton, M.J., Witteveen, A.T., et al. 2002. Gene expression profiling predicts clinical outcome of breast cancer. *Nature* **415:** 530–536.

Welsh, J.B., Zarrinkar, P.P., Sapinoso, L.M., Kern, S.G., Behling, C.A., Monk, B.J., Lockhart, D.J., Burger, R.A., and Hampton, G.M. 2001. Analysis of gene expression profiles in normal and neoplastic ovarian tissue samples identifies candidate molecular markers of epithelial ovarian cancer. *Proc. Natl. Acad. Sci.* **98:** 1176–1181.

West, M., Blanchette, C., Dressman, H., Huang, E., Ishida, S., Spang, R., Zuzan, H., Olson, J.A., Marks, J.R., and Nevins, J.R. 2001. Predicting the clinical status of human breast cancer by using gene expression profiles. *Proc. Natl. Acad. Sci.* **98:** 11462–11467.

Workman, C., Jensen, L.J., Jarmer, H., Berka, R., Gautier, L., Nielser, H.B., Saxild, H.H., Nielsen, C., Brunak, S., and Knudsen, S. 2002. A new non-linear normalization method for reducing variability in DNA microarray experiments. *Genome Biol.* **3:** RESEARCH0048.

Yeakley, J.M., Fan, J.B., Doucet, D., Luo, L., Wickham, E., Ye, Z., Chee, M.S., and Fu, X.D. 2002. Profiling alternative splicing on fiber-optic arrays. *Nat. Biotechnol.* **20:** 353–358.

Yuen, T., Wurmbach, E., Pfeffer, R.L., Ebersole, B.J., and Sealfon, S.C. 2002. Accuracy and calibration of commercial oligonucleotide and custom cDNA microarrays. *Nucleic Acids Res.* **30:** e48.

Received November 13, 2003; accepted in revised form February 11, 2004.

Alamar Exhibit 1002
Alamar v. Olink

nature genetics

TECHNICAL REPORTS

Best Available Copy

© 2005 Nature Publishing Group http://www.nature.com/naturegenetics

A genome-wide scalable SNP genotyping assay using microarray technology

Kevin L Gunderson[1], Frank J Steemers[1,4], Grace Lee[1,4], Leo G Mendoza[2] & Mark S Chee[3]

Oligonucleotide probe arrays have enabled massively parallel analysis of gene expression levels from a single cDNA sample. Application of microarray technology to analyzing genomic DNA has been stymied by the sequence complexity of the entire human genome. A robust, single base–resolution direct genomic assay would extend the reach of microarray technology. We developed an array-based whole-genome genotyping assay that does not require PCR and enables effectively unlimited multiplexing. The assay achieves a high signal-to-noise ratio by combining specific hybridization of picomolar concentrations of whole genome–amplified DNA to arrayed probes with allele-specific primer extension and signal amplification. As proof of principle, we genotyped several hundred previously characterized SNPs. The conversion rate, call rate and accuracy were comparable to those of high-performance PCR-based genotyping assays.

Array technology has enabled whole-genome gene-expression studies of tens of thousands of genes across a single sample. Bottlenecks in sample preparation have precluded the application of whole-genome arrays to genotype analysis. This is unfortunate because the number of genotyping assays required for comprehensive association studies far outstrips the capacity of current approaches. The human genome contains >10 million common SNPs. A smaller number of well-selected 'tagging' SNPs, which act as markers of more extensive patterns of common variation, or haplotypes, can be used to map most of the genetic variation between individuals[1–4]. Preliminary estimates indicate that ~200,000–300,000 tagging SNPs will be required to map most of the variation in the genome, depending on the population to be studied[4–6]. Some parts of the genome may not be amenable to this approach and may require a much higher density of SNPs. Therefore, even the most conservative estimates of the number of SNP markers required for comprehensive genetic association studies indicate that genotyping of several hundred thousand SNPs per individual across hundreds to many thousands of samples would be required to map a causal variant by linkage disequilibrium. The development of a direct array-based SNP genotyping assay and single-tube sample preparation would make it possible to genotype hundreds of thousands of SNPs efficiently. This would eliminate the need for PCR amplification, which may require many separate reactions for whole-genome studies[7,8].

Although human genomic DNA (gDNA) has been used successfully in cDNA, BAC and oligonucleotide array comparative genomic hybridization experiments[9–11], specific hybridization of the entire human genome to a high-density oligonucleotide probe microarray providing single-base resolution suitable for genotyping has not been reported. Sequence complexity is a barrier, because genomes of less complex organisms have been successfully genotyped on arrays. Winzeler *et al.* fragmented and hybridized the 12-Mb yeast genome to 25-mer oligonucleotide probe arrays designed for gene-expression profiling and discovered and genotyped biallelic markers[12]. A similar array-based oligonucleotide hybridization approach was used by Borevitz *et al.* to characterize nucleotide variation in *Arabidopsis* strains (120-Mb genome)[13]. Recent experiments indicate that genotyping directly from human gDNA is possible. Oligonucleotide probes are capable of sequence-specific hybridization in the context of the entire human genome[14]. Storhoff *et al.* specifically detected a single human locus directly from gDNA (~200 fM concentration) using gold nanosphere detection[15]. The potential of genotyping directly from gDNA without intermediate amplification has been demonstrated by collecting genotype information from a single human locus from gDNA captured on a microsphere support, albeit with low signal-to-noise ratio[16]. More recently, Chen *et al.* similarly demonstrated genotyping of a single SNP locus directly from gDNA[17]. Though encouraging, robust array-based genotyping with these approaches has not yet been accomplished.

Given this genomic complexity challenge, the most common strategy is to reduce complexity by amplifying portions of the genome by PCR. This can be done in a random or semirandom fashion, for example, by using restriction enzyme–based adapter ligation PCR[18]. This representational approach, used to detect changes in genomic copy number, has also been applied to genotyping of 10^4–10^5 or more SNPs from a single sample preparation[19,20]. Highly multiplexed PCR-based approaches have recently been developed to genotype specific loci of interest, enabling targeted genotyping of ~10^3–10^4 or more SNPs from a single sample[21,22]. These targeted genotyping approaches

[1]Illumina, Inc., 9885 Towne Centre Dr., San Diego, California 92121, USA. [2]Ambion, Inc., 2130 Woodward, Austin, Texas 78744, USA. [3]Prognosys Biosciences, Inc., 4215 Sorrento Valley Blvd., Suite 105, San Diego, California 92121, USA. [4]These authors contributed equally to this work. Correspondence should be addressed to K.L.G. (kgunderson@illumina.com).

Published online 17 April 2005; doi:10.1038/ng1547

Alamar Exhibit 1002
Alamar v. Olink

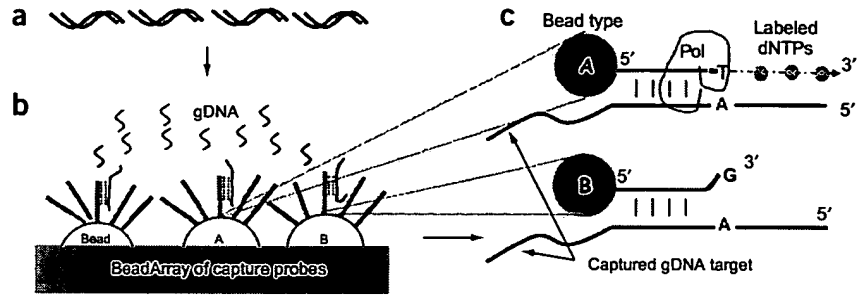


Figure 1 WGG on DNA arrays. (a) WGA to generate large amounts of amplified complex gDNA. (b) Hybridization of the WGA product to a specific and sensitive oligonucleotide probe array (50-mers). (c) An array-based ASPE reaction that scores the captured SNP targets by incorporating multiple biotin-labeled dNTP nucleotides into the appropriate allelic probe, followed by a sensitive detection and signal amplification step to read the incorporated labels. For a given SNP on a given strand, ASPE uses two different allele-specific bead types whose capture sequence is identical except for the 3′ terminal base. The base is chosen such that the probes on one bead type match allele A and those on the other bead type match allele B. Polymerase extension occurs preferentially from matched 3′ termini, enabling appropriate scoring of the SNP.

include a pre-PCR enzymatic SNP scoring step, addition of universal priming sites by ligation and subsequent universal primer PCR. Today, almost all high-throughput genotyping is done using one of these methods. The ability of these methods to scale to 10^5–10^6 genotypes per sample is uncertain.

Here we describe a new array-based whole-genome genotyping (WGG) assay with high signal-to-noise ratio, which allows for accurate and robust genotyping in the context of full genomic complexity. The WGG assay unlocks the potential of microarray technology to read hundreds of thousands to millions of SNPs per genome in a single array experiment. It mirrors an array-based gene-expression experiment in using single-tube sample preparation with virtually unlimited assay multiplex scalability, dependent only on the number of features on the array. Using this approach, it is now feasible to read a vast amount of genetic information directly from the genome using simple, easily automated procedures.

RESULTS

Assay design

The design of the assay is shown in **Figure 1**. It uses a combination of well-proven technologies: (i) whole-genome amplification (WGA) to generate sufficient DNA for hybridization; (ii) hybridization capture of the amplified genomic loci to a specific and sensitive oligonucleotide probe array; (iii) array-based enzymatic SNP scoring assay; and (iv) sensitive signal amplification for readout.

It is crucial to hybridize a sufficiently high concentration (partial concentration of 2–3 pM) of WGA gDNA to the array. We started with an input of ~100–200 ng of gDNA to generate 100–150 μg (1,000–1,500× amplification) of high-complexity amplified product[23].

We assessed the representation bias of the WGA reactions by comparing assay intensities between unamplified and amplified gDNA across 480 nonpolymorphic loci (**Fig. 2**). More than 95% of the loci had signal intensities within threefold of that of the unamplified sample, well within the dynamic range of our assay. Therefore, we estimate that >95% of the unique portion of the genome is accessible with our approach. Moreover, the amplification bias was highly reproducible from sample to sample ($R^2 = 0.98$), ensuring that a set of selected functional assays will result in reproducible amplification from experiment to experiment. Furthermore, reproducible amplification bias is important for future DNA copy and loss-of-heterozygosity applications.

After amplification, we hybridized WGA samples to arrays of 50-base capture probes (actual oligonucleotide probes are 75 bases long, of which 25 bases are used for decoding). We chose this length as a compromise to maximize signal intensity, minimize inter-locus intensity variation and synthesize full-length probes. After hybridization, we scored SNPs using an allele-specific primer extension (ASPE) assay, chosen for its proven high discrimination and the economy of using the capture probes themselves to query the SNP. The ASPE approach also enables us to query all SNP categories with a single color on a single array. We accomplished this by using two bead types (A and B) per assay whose probe sequences differed only at the 3′ terminal base (opposite SNP site), creating allelic discrimination in the polymerase extension step. Probe sequences on a particular bead type will be extended and labeled (multiple incorporation events) only when hybridized to a perfectly matched allelic target. The genotype state of a given SNP locus (AA, AB or BB) is determined by the intensity ratio between the two corresponding bead types, $\theta = (2/\pi)^*\arctan(B/A)$.

Assay optimization

For assay optimization, we designed ASPE assays for a set of 96 nonpolymorphic loci (a subset of the 480 loci) randomly selected from throughout the human genome. We used an artificial mismatch probe and an arbitrarily chosen 3′ mismatch base. Monitoring the assay discrimination (perfect match-to-mismatch ratio) of the

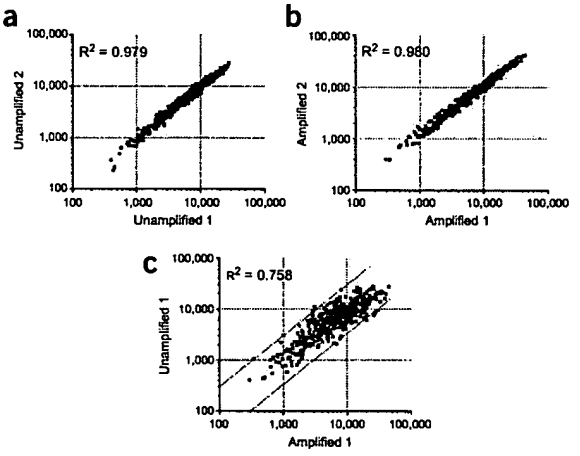


Figure 2 WGA representation. Assay intensities from the 480 nonpolymorphic controls were used to assess representation bias of amplified DNA. Scatter plots of (a) unamplified gDNA versus unamplified gDNA, (b) amplified gDNA versus amplified gDNA and (c) unamplified gDNA versus amplified gDNA. The dashed lines denote threefold changes. The results indicate that >95% of the loci in the amplified product are within threefold of the copy number of the unamplified sample. Furthermore, the scatter plot of amplified gDNA versus amplified gDNA, b, indicates that the amplification bias is consistent between two different amplified samples.

© 2005 Nature Publishing Group http://www.nature.com/naturegenetics

Alamar Exhibit 1002
Alamar v. Olink

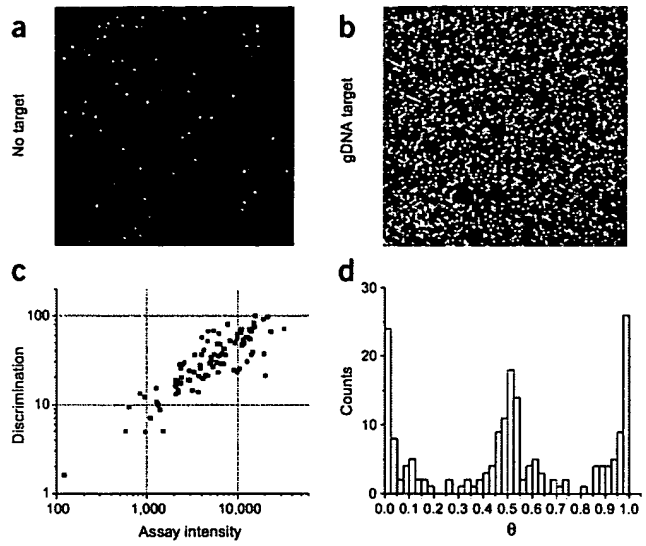


Figure 3 WGG on Sentrix array matrix using the WGG feasibility array. (a) The WGG assay was carried out on an array that went through hybridization without any DNA target. This 'no target control' assesses the self-extension of the probes on the array. (Most of the bright beads are detection controls included on every array to validate the signal amplification.) (b) The WGG assay was carried out on an array hybridized with amplified gDNA. (c) A scatter plot of discrimination ratio (match/mismatch) versus assay intensity across the 96 nonpolymorphic ASPE controls. The discrimination improves with assay intensity and has a median of ~30. (d) Histogram of the θ values for 186 SNP assays from a single sample. The histogram has three peaks corresponding to the three possible genotype categories.

nonpolymorphic controls in a Sentrix array matrix format (96 arrays) enabled us to optimize the WGA, hybridization and primer extension reactions[24]. Additional assays monitored locus representation (480 correctly matched nonpolymorphic sequences) and genotyping quality (186 ASPE linkage panel SNP assays).

Analysis of the nonpolymorphic ASPE controls showed a high discrimination ratio (median >30:1 or θ > 0.97) and minimal self-extension (Fig. 3). This level of discrimination also correlated with good histogram cluster separation in the θ values across the 186 linkage panel SNPs (Fig. 3c). We used this set to genotype 32 DNA samples in triplicate (except two samples genotyped in duplicate). Of the 186 assays, 176 (94.6%) were successful on the basis of cluster separation score (CSS) and visual inspection. The genotyping data on the 176 successful assays had a call rate of 99.7% (16,501 of 16,544), reproducibility of 99.96% (16,362 of 16,368) and concordance of 99.97% (7,918 of 7,920, genotyping concordance data available on only 16 of 32 samples). These results indicate that genotyping was highly accurate and reproducible, with a very low dropout rate, for the complexity (~3 × 10^9 bp) of the human genome.

Validation on HapMap quality control SNPs

We arbitrarily chose the 186 SNP assays from a set of 2,500 linkage panel SNPs empirically selected to perform well in the GoldenGate assay. Therefore, we tested a second set of SNPs that had not undergone empirical filtering. The HapMap project used a set of 1,500 randomly chosen SNPs for quality control purposes (The International HapMap Consortium, unpublished data). We designed assays for 1,485 of the 1,500 SNPs on both strands (excluding 15 insertion-deletion polymorphism SNPs for design reasons). We chose the best strand, as determined by the CSS, for each of the 1,485 assays and used it to genotype 95 samples (30 triplicate samples and 5 duplicate DNAs) from the CEPH plate of the HapMap project (**Fig. 4**). We assessed assay quality by measuring reproducibility in duplicate samples, heritability in parent-child trios and concordance of genotype calls with the GoldenGate assay. As a result of the random selection process, the 1,500-SNP panel contained a large number of ill-defined SNPs (multiple-hit loci, poor sequence, etc.) and nonpolymorphic loci. To avoid systematic errors due to confusion between a nonpolymorphic locus and a multiple-hit locus (resulting from cross-hybridization to multiple sequences in the genome), we defined successful assays as those in which both SNP alleles were detected (*i.e.*, the minor and the major allele). We currently use this approach in product development. Using this metric, we found that 819 of 1,500 SNPs had a minor allele. For comparison, the GoldenGate assay found that 864 of 1500 SNPs had a minor allele. This indicates that the WGG assay successfully develops assays at ~95% the rate of the GoldenGate platform (819 of 864). Of these 819 'successful' loci, genotype calls across the 95 DNA samples (also used in training) generated a call rate of 99.7%, a reproducibility of 99.99%, an inheritance of 99.97% and a concordance of 99.9% (**Table 1**). Notably, the accuracy of the WGG assay is similar to that of the GoldenGate assay, despite the fact that the complexity of a WGG sample is 50,000 times higher when hybridized to the array.

DISCUSSION

Genotyping on probe arrays is limited by sample preparation and assay design constraints and not by array technology. The inherent mismatch between assay scalability (low multiplex) and array scalability (high multiplex) led us to develop a WGG approach.

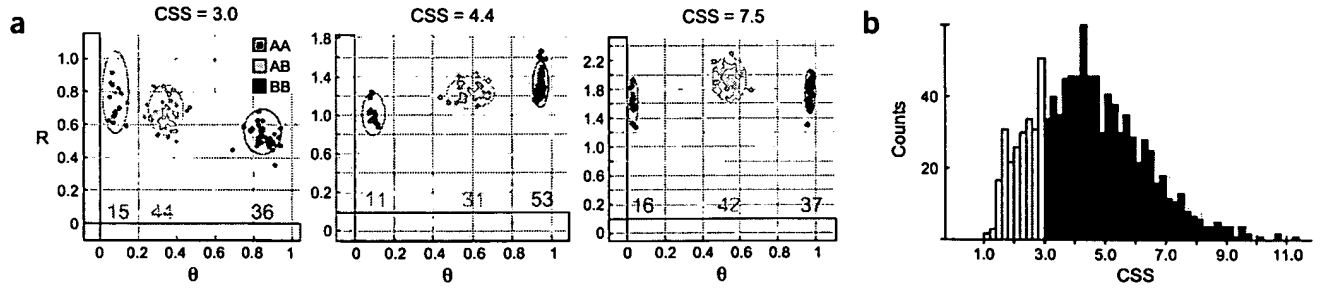


Figure 4 Genotyping of HapMap quality control SNPs. (a) Representative genotype plots illustrating different quality CSSs. This metric measures the statistical separation between genotype clusters and is used as a gauge for a successful assay. Loci with CSSs >3.0 are generally successful and provide reliable genotyping calls. (b) Histogram of CSSs for the 1,485 HapMap quality control loci. The median CSS across 1,485 assays was 4.4, and >76% of the assays had CSSs >3.0.

© 2005 Nature Publishing Group http://www.nature.com/naturegenetics

Alamar Exhibit 1002
Alamar v. Olink

Table 1 Genotyping quality

Parameter	Value	Counts
Assay success*	95%	819 of 864
Call rate	99.5%	68,807 of 68,970
Reproducibility	99.99%	8,189 of 8,190
Heritability	99.97%	83,766 of 83,790
Concordance	99.9%	137,456 of 137,614

*Assay success rate computed relative to the GoldenGate assay.

© 2005 Nature Publishing Group http://www.nature.com/naturegenetics

Inherent in our assay design is the capacity for unlimited multiplexing, as each assay includes the full complexity of the human genome. Therefore, assay quality will be the same whether the assay queries a single SNP or one million SNPs. We carried out a proof-of-principle experiment on several hundred SNPs to demonstrate the concept. As a result of this new assay, the number of bases that can be analyzed in the unique portion of human gDNA is limited only by the number of probes in the array.

The WGG assay is modular in design. Each of the four key steps (WGA, array hybridization, array-based genotyping and signal amplification) can be replaced with alternatives. For instance, our current WGA procedure works well on relatively intact DNA (>1–2 kb) but may not work well with highly degraded DNA (*e.g.*, from formalin-fixed paraffin-embedded samples). An alternative WGA method tolerant of sample degradation could be used[25]. Similarly, the enzymatic SNP scoring step can be replaced by alternatives such as single-base extension, ligation and allelic oligonucleotide hybridization. Finally, the signal amplification step can be replaced with alternative signal amplification schemes with enhanced sensitivity[26,27]. This modular design allows improvements to be quickly and easily incorporated into the assay.

We achieved a high signal-to-noise ratio by hybridizing relatively high concentrations (~2–3 pM) of WGA gDNA and by minimizing 'background' signal generated from interprobe and intraprobe self-extension. Self-extension was greatly suppressed by the addition of single-stranded binding protein to the extension reaction (data not shown). We achieved high specificity, in the presence of the entire genome, by combining stringent hybridization with on-array enzymatic discrimination. Enzymatic-based genotyping assays such as primer extension and ligation have previously been used directly on array surfaces to achieve high levels of discrimination in genotyping captured PCR products[28–32].

The BeadArray technology also contributes to the high quality of the WGG assay. First, each bead type is present in ~30 copies on average, and so each measurement is repeated multiple times, enabling removal of outliers and increased precision[24]. Second, all beads of a particular type are made at the same time, ensuring consistency. This eliminates an important source of variability between the two allele-specific probes, because ASPE-based genotyping uses a ratiometric comparison of the bead type intensities. Finally, we used 5′ immobilization of oligonucleotide probes synthesized in the 3′–5′ direction, yielding intact 3′ termini suitable for primer extension–based assays. Therefore, the exceptional performance of the WGG assay derives from a combination of design elements involving the assay as well as the readout platform.

In comparison to other high-throughput genotyping technologies, WGG combines virtually unlimited multiplexing with freedom of SNP selection. This will be particularly important in assaying custom sets of SNPs, including haplotype-tagging SNPs[33], coding SNPs and other high-value SNPs. Furthermore, the ability to choose SNPs of interest increases the power of association studies relative to random SNPs, particularly with respect to maximizing linkage disequilibrium with markers in feature-rich regions, such as genes and conserved regions[34]. This contrasts with the PCR-based complexity reduction approach, which, though capable of genotyping tens of thousands to hundreds of thousands of SNPs from a single sample or array, suffers from a limited ability to choose SNPs because it requires the SNP to be present in the amplified representation[19,20,35].

Other recently introduced genotyping technologies, such as the GoldenGate assay from Illumina and the Molecular Inversion Probe assay from Parallele, enable researchers to choose almost any SNP of interest and to multiplex to relatively large numbers (1,000–10,000)[21,22]. The ability to multiplex to even higher levels will, at some point, be limited by assay complexity and the amount of material generated for each individual locus in the final PCR reaction. These assays either do not allow genotyping of targeted SNPs or do not scale intrinsically like the WGG assay.

As a next step to realizing the full potential of the WGG assay, we are developing a single array to genotype >100,000 loci from a one-tube reaction. This is made possible by the feature density of our arrays of >288,000 bead types per slide. The BeadChip platform can be scaled to even higher densities. For instance, decreasing bead diameter from the current 3 μm to 1.5 μm (and reducing center-to-center spacing accordingly) enables use of >1 million bead types per BeadChip.

In addition to standard SNP genotyping for association studies, these high-resolution genotyping arrays are applicable to genomic profiling (DNA copy and loss-of-heterozygosity measurements) and allelic expression measurements (cDNA genotyping). Partly as a result of this new assay, we anticipate a future in which massively parallel direct genomic analysis is used routinely in applications ranging from the exploration of genome function to molecular medicine.

METHODS

ASPE Sentrix array design. We obtained all the array data using Illumina's Sentrix BeadArray matrix. Oligonucleotide probes on the beads were 75 bases in length; 25 bases at the 5′ end were used for decoding[24] and the remaining 50 bases were locus-specific. We immobilized the oligonucleotides on activated beads using a 5′ amino group[24]. The WGG feasibility array contained probes for 186 SNP assays (186 probe pairs, allele A and allele B), 384 nonpolymorphic assays (match only), 96 nonpolymorphic discrimination assays (96 probe pairs, match and mismatch), 96 yeast loci (96 probe pairs, match and mismatch) and 7 control bead types (0, 1, 10, 100, 1,000, 10^4 and 10^5 labels per bead). We chose the set of 186 SNP assays randomly from a linkage panel of 2,500 autosomal SNP using an arbitrary strand. The HapMap study used four Sentrix array types (~750 assays per array) with assays designed to both strands. We chose the best strand, as determined by the CSS, for analysis of the HapMap quality control SNPs.

WGA sample preparation. We subjected DNA samples to WGA using commercially available reagents (Illumina MP1 and AMM). We carried out 100-μl WGA reactions yielding ~150 μg of product from 100 ng of input gDNA, indicating that amplification was greater than 1,000-fold. We obtained similar results using Repli-g kits (Qiagen) in accordance with the manufacturer's instructions using 100 ng of input gDNA. We carried out fragmentation by adding 9.5 μl of calcium chloride (5 mM), 30.5 μl of water and 0.025 U DNaseI (Invitrogen) to 50 μl of Repli-g product. We incubated the reaction for 15 min at 37 °C. We inactivated DNaseI by adding 10 μl of 0.1% SDS and heating it to 95 °C for 10 min. The average yield was ~50 μg per reaction, and the average product size after digestion, estimated by gel analysis, was ~100–200 bases (data not shown). We pooled two reactions to generate sufficient material for hybridization. After WGA, we purified the amplified products from nucleotides and primers by ethanol precipitation or by Montage ultra-

Alamar Exhibit 1002
Alamar v. Olink

filtration plates (Millipore) and quantified them by a picoGreen assay (Molecular Probes) or by measuring absorbance at 260 nm. We resuspended the precipitated or concentrated DNA at ~5–6 μg μl^{-1} in 1× hybridization buffer (1 M NaCl, 100 mM potassium-phosphate buffer (pH 7.5) and 0.1% Tween 20) supplemented with 20% formamide. We used sonicated human placental gDNA (Sigma) as unamplified DNA. In these studies, all gDNA samples were from high-quality DNA obtained from the Coriell Institute. The analysis of highly degraded DNA has not yet been undertaken.

Array hybridization. We denatured the resuspended WGA product at 95 °C for 5 min and then exposed it to the Sentrix array matrix, which we mated to a microtiter plate, submerging the fiber bundles in 15 μl of hybridization sample. We incubated the entire assembly for 14–18 h at 48 °C with shaking. After hybridization, we washed arrays in 1× hybridization buffer and 20% formamide at 48 °C for 5 min.

ASPE-based detection. Before carrying out the array-based primer extension reaction, we washed Sentrix array matrices for 1 min with wash buffer (33.3 mM NaCl, 3.3 mM potassium phosphate and 0.1% Tween-20, pH 7.6) and then incubated them for 15 min in 50 μl of ASPE reaction buffer (Illumina EMM, containing polymerase, a mix of biotin-labeled and unlabeled nucleotides, single-stranded binding protein, bovine serum albumin and appropriate buffers and salts) at 37 °C. After the reaction, we immediately stripped the arrays in freshly prepared 0.1 N NaOH for 2 min and then washed and neutralized them twice in 1× hybridization buffer for 30 s. We detected the biotin-labeled nucleotides incorporated during primer extension using a sandwich assay similar to that previously described[36]. We blocked the arrays at room temperature for 10 min in 1 mg ml^{-1} bovine serum albumin in 1× hybridization buffer and then washed them for 1 min in 1× hybridization buffer. We then stained the arrays with streptavidin-phycoerythrin solution (1× hybridization buffer, 3 μg ml^{-1} streptavidin-phycoerythrin (Molecular Probes) and 1 mg ml^{-1} bovine serum albumin) for 10 min at room temperature, washed the arrays with 1× hybridization buffer for 1 min and then counterstained them with an antibody reagent (10 μg ml^{-1} biotinylated antibody to streptavidin (Vector Labs) in 1× PBST (137 mM NaCl, 2.7 mM KCl, 4.3 mM sodium phosphate, 1.4 mM potassium phosphate and 0.1% Tween-20) supplemented with 6 μg ml^{-1} goat normal serum) for 20 min. After counterstaining, we washed the arrays in 1× hybridization buffer and restained them with streptavidin-phycoerythrin solution for 10 min. We washed the arrays one final time in 1× hybridization buffer before imaging them in 1× hybridization buffer on a custom CCD-based BeadArray imaging system[24]. We extracted intensities using custom image analysis software.

Data analysis and genotype calls. We carried out genotype analysis using Illumina's GenCall software (version 1.0.14), which compared intensities between probes for allele A and allele B across a large number of samples to create archetypal clustering patterns. These patterns allowed the genotyping data to be assigned membership to clusters using a probabilistic model and allowed assignment of a corresponding GenCall score. For example, data points falling between two clusters were assigned a low probability score of being a member of either cluster and had a correspondingly low GenCall score.

We initially assessed cluster quality by evaluating the CSS, a measure of statistical separation between clusters. It is defined as

$$\mathrm{CSS} = \min\left(\frac{|\theta_{AB} - \theta_{AA}|}{|\sigma_{AB} + \sigma_{AA}|}, \frac{|\theta_{AB} - \theta_{BB}|}{|\sigma_{AB} + \sigma_{BB}|}\right).$$

Loci with cluster scores around the cutoff of 3.0 were visually evaluated and the training clusters refined by manual intervention. We chose the cutoff value of 3.0 for the CSS on the basis of our experience in minimizing strand concordance errors. We generally observe accurate genotyping when the CSS is >3.0. Loci with questionable clusters were scored as unsuccessful and excluded from further analysis.

For the analysis of concordance of HapMap quality control SNPs with the GoldenGate assay, we used 726 of 819 SNP assays. Of the 93 SNP assays that we excluded, 92 were not called by the GoldenGate assay, and one (rs3778464) had high systematic discordance (170 of 190 calls were discordant) with GoldenGate. We computed the heritability by tabulating inheritance errors across autosomal SNPs (798 of 819) on the set of trios (35) in the CEPH sample plate of the HapMap. Heritability was defined as (total genotypes called – inheritance errors)/total genotypes called.

URL. More information on the HapMap project is available at http://www.hapmap.org/.

ACKNOWLEDGMENTS

We thank our Illumina colleagues in manufacturing for providing oligos and arrays and L. Zhou for generating the nonpolymorphic control sequences used in assay development. This work was supported in part by a US National Institutes of Health National Cancer Institute grant to K.L.G.

AUTHORS' CONTRIBUTIONS

K.L.G. conceived the experimental approach, oversaw the design and analysis of experiments and, together with M.S.C., wrote the paper. F.J.S., G.L. and L.G.M. carried out the experiments, analyzed data and identified and implemented a number of technical improvements.

COMPETING INTERESTS STATEMENT

The authors declare competing financial interests (see the *Nature Genetics* website for details).

Received 18 October 2004; accepted 10 March 2005
Published online at http://www.nature.com/naturegenetics/

1. The International HapMap Consortium. The International HapMap Project. *Nature* **426**, 789–796 (2003).
2. Johnson, G.C. *et al.* Haplotype tagging for the identification of common disease genes. *Nat. Genet.* **29**, 233–237 (2001).
3. Daly, M.J., Rioux, J.D., Schaffner, S.F., Hudson, T.J. & Lander, E.S. High-resolution haplotype structure in the human genome. *Nat. Genet.* **29**, 229–232 (2001).
4. Gabriel, S.B. *et al.* The structure of haplotype blocks in the human genome. *Science* **296**, 2225–2229 (2002).
5. Judson, R., Salisbury, B., Schneider, J., Windemuth, A. & Stephens, J.C. How many SNPs does a genome-wide haplotype map require? *Pharmacogenomics* **3**, 379–391 (2002).
6. Stephens, J.C. *et al.* Haplotype variation and linkage disequilibrium in 313 human genes. *Science* **293**, 489–493 (2001).
7. Kwok, P.Y. & Chen, X. Detection of single nucleotide polymorphisms. *Curr. Issues Mol. Biol.* **5**, 43–60 (2003).
8. Wang, D.G. *et al.* Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. *Science* **280**, 1077–1082 (1998).
9. Pinkel, D. *et al.* High resolution analysis of DNA copy number variation using comparative genomic hybridization to microarrays. *Nat. Genet.* **20**, 207–211 (1998).
10. Pollack, J.R. *et al.* Genome-wide analysis of DNA copy-number changes using cDNA microarrays. *Nat. Genet.* **23**, 41–46 (1999).
11. Carvalho, B., Ouwerkerk, E., Meijer, G.A. & Ylstra, B. High resolution microarray comparative genomic hybridisation analysis using spotted oligonucleotides. *J. Clin. Pathol.* **57**, 644–646 (2004).
12. Winzeler, E.A. *et al.* Direct allelic variation scanning of the yeast genome. *Science* **281**, 1194–1197 (1998).
13. Borevitz, J.O. *et al.* Large-scale identification of single-feature polymorphisms in complex genomes. *Genome Res.* **13**, 513–523 (2003).
14. Wu, D.Y., Nozari, G., Schold, M., Conner, B.J. & Wallace, R.B. Direct analysis of single nucleotide variation in human DNA and RNA using in situ dot hybridization. *DNA* **8**, 135–142 (1989).
15. Storhoff, J.J. *et al.* Diagnostic detection systems based on gold nanoparticle probes. in *Biomedical Applications of Micro- and Nanoengineering* Proc. SPIE vol. 4937 (ed. Nicolau, D.V.) 1–7 (SPIE, Bellingham, Washington, 2002).
16. Rao, K.V. *et al.* Genotyping single nucleotide polymorphisms directly from genomic DNA by invasive cleavage reaction on microspheres. *Nucleic Acids Res.* **31**, e66 (2003).
17. Chen, Y., Shortreed, M.R., Peelen, D., Lu, M. & Smith, L.M. Surface amplification of invasive cleavage products. *J. Am. Chem. Soc.* **126**, 3016–3017 (2004).
18. Lucito, R. *et al.* Genetic analysis using genomic representations. *Proc. Natl. Acad. Sci. USA* **95**, 4487–4492 (1998).
19. Kennedy, G.C. *et al.* Large-scale genotyping of complex DNA. *Nat. Biotechnol.* **21**, 1233–1237 (2003).
20. Matsuzaki, H. *et al.* Parallel genotyping of over 10,000 SNPs using a one-primer assay on a high-density oligonucleotide array. *Genome Res.* **14**, 414–425 (2004).
21. Hardenbol, P. *et al.* Multiplexed genotyping with sequence-tagged molecular inversion probes. *Nat. Biotechnol.* **21**, 673–678 (2003).

© 2005 Nature Publishing Group http://www.nature.com/naturegenetics

163
Alamar Exhibit 1002
Alamar v. Olink

22. Fan, J.B. *et al.* Highly parallel SNP genotyping. *Cold Spring Harbor Symposia on Quantitative Biology LXVIII*, 69–78 (CSHL, Woodbury, New York, 2003).
23. Dean, F.B. *et al.* Comprehensive human genome amplification using multiple displacement amplification. *Proc. Natl. Acad. Sci. USA* **99**, 5261–5266 (2002).
24. Gunderson, K.L. *et al.* Decoding randomly ordered DNA arrays. *Genome Res.* **14**, 870–877 (2004).
25. Wang, G. *et al.* DNA amplification method tolerant to sample degradation. *Genome Res.* **14**, 2357–2366 (2004).
26. Bobrow, M.N., Harris, T.D., Shaughnessy, K.J. & Litt, G.J. Catalyzed reporter deposition, a novel method of signal amplification. Application to immunoassays. *J. Immunol. Methods* **125**, 279–285 (1989).
27. Hacker, G.W. High performance Nanogold-silver in situ hybridisation. *Eur. J. Histochem.* **42**, 111–120 (1998).
28. Shumaker, J.M., Metspalu, A. & Caskey, C.T. Mutation detection by solid phase primer extension. *Hum. Mutat.* **7**, 346–354 (1996).
29. Gunderson, K.L. *et al.* Mutation detection by ligation to complete n-mer DNA arrays. *Genome Res.* **8**, 1142–1153 (1998).
30. Pastinen, T., Kurg, A., Metspalu, A., Peltonen, L. & Syvanen, A.C. Minisequencing: a specific tool for DNA analysis and diagnostics on oligonucleotide arrays. *Genome Res.* **7**, 606–614 (1997).
31. Pastinen, T. *et al.* A system for specific, high-throughput genotyping by allele-specific primer extension on microarrays. *Genome Res.* **10**, 1031–1042 (2000).
32. Erdogan, F., Kirchner, R., Mann, W., Ropers, H.H. & Nuber, U.A. Detection of mitochondrial single nucleotide polymorphisms using a primer elongation reaction on oligonucleotide microarrays. *Nucleic Acids Res.* **29**, E36 (2001).
33. Consortium, T.I.H. The International HapMap Project. *Nature* **426**, 789–796 (2003).
34. Simpson, C.L. *et al.* MaGIC: a program to generate targeted marker sets for genome-wide association studies. *Biotechniques* **37**, 996–999 (2004).
35. Di, X. *et al.* Dynamic model-based algorithms for screening and genotyping over 100K SNPs on oligonucleotide microarrays. *Bioinformatics* advance online publication, 19 January 2005 (doi:10.1093/bioinformatics/bti275).
36. Pinkel, D., Straume, T. & Gray, J.W. Cytogenetic analysis using quantitative, high-sensitivity, fluorescence hybridization. *Proc. Natl. Acad. Sci. USA* **83**, 2934–2938 (1986).

© 2005 Nature Publishing Group http://www.nature.com/naturegenetics

Alamar Exhibit 1002
Alamar v. Olink

This Page is Inserted by IFW Indexing and Scanning Operations and is not part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☐ BLACK BORDERS
- ☐ IMAGE CUT OFF AT TOP, BOTTOM OR SIDES
- ☑ FADED TEXT OR DRAWING
- ☐ BLURRED OR ILLEGIBLE TEXT OR DRAWING
- ☐ SKEWED/SLANTED IMAGES
- ☐ COLOR OR BLACK AND WHITE PHOTOGRAPHS
- ☐ GRAY SCALE DOCUMENTS
- ☐ LINES OR MARKS ON ORIGINAL DOCUMENT
- ☐ REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY
- ☐ OTHER: _____

IMAGES ARE BEST AVAILABLE COPY.
As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.

Alamar Exhibit 1002
Alamar v. Olink

ANALYTICAL BIOCHEMISTRY 262, 122–128 (1998)
ARTICLE NO. AB982770

Site-Specific, Enzymatic Biotinylation of Recombinant Proteins in *Spodoptera frugiperda* Cells Using Biotin Acceptor Peptides[1]

Siobain Duffy,* Kwei-Lan Tsao,† and David S. Waugh*,[2]

*ABL-Basic Research Program, NCI-Frederick Cancer Research and Development Center, P.O. Box B, Frederick, Maryland 21702-1201; and †Roche Research Center, Hoffmann La-Roche, Inc., Kingsland Street, Nutley, New Jersey 07110

Received January 28, 1998

Site-specific, enzymatic biotinylation of recombinant proteins can be exploited to circumvent many problems associated with the use of biotinylating reagents *in vitro* and to overcome some of their inherent limitations. Additionally, biotinyl proteins can be purified to near-homogeneity in a single step under native conditions. Here we report that a biotin acceptor peptide (BAP) substrate for *Escherichia coli* biotin holoenzyme synthetase (BirA) can be used to label recombinant proteins with biotin in *Spodoptera frugiperda* (Sf9) cells, and we describe a collection of baculovirus transfer vectors specifically designed for this purpose. These BioBac vectors will greatly expand the range of proteins to which this technology can be applied. © 1998 Academic Press

Key Words: **enzymatic biotinylation; baculovirus; biotin acceptor peptide; BirA.**

Biotin (vitamin H) is frequently conjugated to proteins so that they can be immobilized on avidin- or streptavidin-coated surfaces for various kinds of ligand-binding experiments. A multitude of chemical reagents have been developed for this purpose (1). However, although each reagent is specific for a certain type of functional group, none is inherently site selective. For example, a reagent that only reacts with primary amino groups can potentially modify any of the lysine side chains in a protein, which are likely to be scattered over its surface. Biotinylation at multiple sites increases the potential for direct and indirect inactivation of the protein by the biotinylating reagent. Direct inactivation can occur when some of the reactive groups (or the neighboring atoms) are crucial for the biological activity of the protein. Although this problem can sometimes be overcome by choosing a different reagent, regardless of which reagent is used, some derivatized proteins are likely to be inactivated indirectly due to immobilization in a sterically unfavorable orientation. Additionally, because biotinylation with chemical reagents typically yields a heterogeneous mixture of reaction products, ensuring batch-to-batch reproducibility can be a formidable problem.

All cells contain a small number of proteins to which biotin is covalently attached as a prosthetic group (2, 3). Biotin carboxyl carrier protein (BCCP),[3] a subunit of acetyl-CoA carboxylase, is the only biotinyl protein in *Escherichia coli*. Biotin is coupled via an amide linkage to the ϵ-amino group of Lys122 in BCCP by the enzyme biotin holoenzyme synthetase (BirA) in an ATP-dependent reaction (4, 5). Recombinant proteins can be labeled with biotin at a single, predetermined site, albeit indirectly, by producing them as BCCP fusion proteins in *E. coli* (6–10). This approach circumvents several problems with chemical reagents. First, enzymatic biotinylation at a specific site ensures that all molecules will be immobilized in the same (bioactive) orientation, thereby resulting in the highest possible specific activity. Additionally, since the chemistry is performed *in vivo*, there are no hazardous reagents to handle and dispose of. Finally, the specificity of the

[1] By acceptance of this article, the publisher or recipient acknowledges the right of the U.S. Government and its agents and contractors to retain a nonexclusive royalty-free license in and to any copyright covering the article.

[2] To whom correspondence should be addressed. Fax: (301) 846-7148. E-mail: waughd@ncifcrf.gov.

[3] Abbreviations used: BCCP, biotin carboxyl carrier protein; BirA, biotin holoenzyme synthetase; BAP, biotin acceptor peptide; TBS, Tris-buffered saline.

0003-2697/98 $25.00
Copyright © 1998 by Academic Press
All rights of reproduction in any form reserved.

Alamar Exhibit 1002
Alamar v. Olink

enzyme allays any concerns about reproducibility or heterogeneity of the reaction products.

A further advantage of biotinylation *in vivo* is that the biotin moiety can be used as an affinity handle to purify a biotinyl protein. Although extremely harsh (i.e., denaturing) conditions are required to disrupt the interaction between biotin and tetrameric avidin or streptavidin, a variety of physically (11), chemically (12), and genetically (13) modified forms of these proteins have been described that still bind biotin specifically, but with weaker affinity. Biotinyl proteins can be released from these weaker-binding forms of avidin and streptavidin either by effecting a modest change in the pH or by including millimolar concentrations of biotin in the elution buffer.

The 87 C-terminal residues of BCCP (amino acids 70–156) comprise its minimal biotin acceptor domain; shorter fragments are inactive, as are some point mutants with amino acid substitutions that lie far from the biotinyl lysine (6, 14). These observations suggest that BirA recognizes a specific tertiary structure adopted by BCCP rather than a simple peptide sequence. Yet, peptidic substrates for BirA have been isolated from random libraries by *in vitro* selection (15). Curiously, these biotin acceptor peptides (BAPs) bear little resemblance to the natural substrates. However, like the biotin acceptor domain of BCCP, BAPs are efficiently biotinylated in *E. coli* when appended to either terminus of a recombinant protein (15–17). Because of their small size (ca. 20 residues), these peptides are far more appealing than BCCP as molecular tags.

Although it has been reported that some eukaryotes are capable of biotinylating BCCP fusion proteins (18–20), thus far the use of BAPs has been restricted to proteins that can be produced in a soluble, active form in *E. coli*. In this report, we demonstrate that a BAP can also be used to label recombinant proteins with biotin in *Spodoptera frugiperda* (Sf9) cells, and we describe a collection of baculovirus transfer vectors, called BioBac vectors, that are specifically designed for this purpose.

MATERIALS AND METHODS

Insect Cell Culture

Sf9 cells were used for the routine propagation of wild-type and recombinant baculoviruses. The cells were cultivated in IPL-41 medium (Life Technologies, Gaithersburg, MD) supplemented with 10% fetal bovine serum (Life Technologies), 0.01% pluonic F-68 solution (JRH Biosciences, Lenexa, KS), 0.4% yeastolate solution (JRH Biosciences), and 50 μM D-biotin.

General Recombinant DNA Methods

Purification of plasmid DNA, restriction endonuclease digestions, fragment isolations, and ligations were performed essentially as described (21). *E. coli* was transformed with plasmid DNA by electroporation (22). The DNA sequences of the relevant regions of all recombinant plasmids were confirmed experimentally.

Construction of Baculovirus Transfer Vectors for the Production of Biotinyl Proteins in Insect Cells

The *Bam*HI/*Hin*dIII fragment of pDW363 (16) that encompasses the *E. coli* BirA coding sequence was ligated with the *Bam*HI/*Hin*dIII vector fragment of p2Bac (Invitrogen, San Diego, CA) to construct pDW440. Two synthetic oligodeoxyribonucleotides (Nos. 436 and 437) were annealed and then ligated directly with the *Bsi*WI/*Sac*I vector backbone of pDW440 to construct pDW445. The sequence of the synthetic DNA cassette can be deduced from Fig. 4B. The same procedure was used to construct pDW446, using instead a different pair of synthetic oligonucleotides (Nos. 438 and 439). The sequence of this DNA cassette can also be deduced from Fig. 4B.

Construction of Donor Vectors for the Production of Recombinant Baculoviruses by Site-Specific Recombination in E. coli.

The 528-bp *Eco*RV fragment of p2Bac (Invitrogen) that includes the p10 promoter was isolated and inserted into the unique *Sna*BI site in pFastBac-1 (Life Technologies). The insert is oriented such that transcription initiating from the p10 and polyhedron promoters is in opposite (divergent) directions. This dual promoter derivative of pFastBac-1 is called pDW462. pDW464 and pDW465 were constructed by ligating the *Hin*dIII/*Sac*I fragments of pDW445 and pDW446, respectively, with the *Hin*dIII/*Sac*I vector backbone of pDW462.

Construction of pDW449 and pDW450

pDW449 was constructed by ligating the *Stu*I/*Bam*HI fragment of pDW342 (16) with the *Stu*I/*Bgl*II vector backbone of pDW445. pDW450 was constructed by cleaving pDW449 with *Bam*HI and *Hin*dIII, blunting the ends with T4 DNA polymerase and deoxynucleotide triphosphates, and then recircularizing the vector fragment to delete the BirA coding sequence.

Expression of E. coli BirA and the BAP–maltose-binding protein (MBP)–Raf Fusion Protein

Recombinant baculoviruses were constructed from transfer vectors pDW449 and pDW450 by recombination with baculovirus DNA *in vivo*, using the Bacu-

Alamar Exhibit 1002
Alamar v. Olink

loGold kit in accordance with the manufacturer's instructions (Pharmingen, San Diego, CA). Putative recombinants were purified twice by plaque assay (23), and high-titer stocks of the recombinant viruses were prepared by standard methods (23). Sf9 cells were infected with recombinant baculoviruses at a multiplicity of infection of approximately 1 in 75-cm^2 tissue culture flasks and incubated at 25°C. After 72 h, the cells were dislodged from the plates, collected by centrifugation, washed once with Tris-buffered saline (TBS), and lysed with a Dounce homogenizer. Samples of the total intracellular protein from cells infected with bDW449 or bDW450 were prepared by combining 50 μl of the crude lysate with 50 μl of 2× SDS sample buffer (24), heating the samples at 90°C for 4 min, and centrifuging them for 5 min at top speed in a microcentrifuge prior to electrophoresis. The remainder of the crude lysates were centrifuged at 15,000*g* for 15 min to remove the insoluble debris, and then samples of the soluble intracellular protein were prepared from the supernatants as described above. Samples of the total intracellular protein from uninfected cells and from cells infected with wild-type baculovirus (Pharmingen) also were prepared as described above.

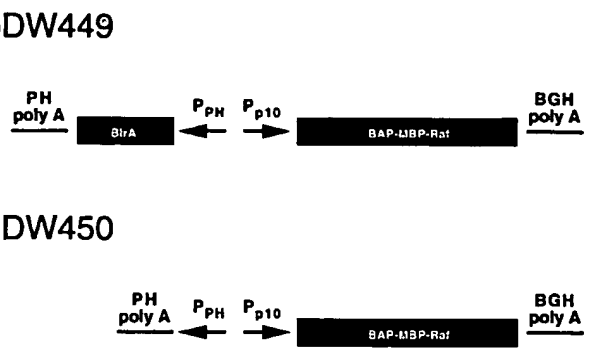


FIG. 1. Recombinant baculovirus vectors bDW449 and bDW450. Only the relevant features of the recombinant baculoviruses are shown. bDW449 produces both *E. coli* BirA, under polyhedron promoter control, and an MBP–Raf_{55-132} fusion protein with a biotin acceptor peptide joined to its N-terminus (BAP–MBP–Raf), under p10 promoter control. The two genes are transcribed in divergent orientation, as depicted. bDW450 produces the BAP–MBP–Raf fusion protein but not BirA. Abbreviations: PH polyA, polyadenosine sequence from the baculovirus polyhedron gene; BGH polyA, polyadenosine sequence from the bovine growth hormone gene; P_{PH}, polyhedron promoter; P_{p10}, p10 promoter.

SDS-PAGE and Western Blot Analysis

The total or soluble intracellular protein from an equivalent number of cells (ca. 2 × 10^4 cells/lane) was resolved by SDS–PAGE (12% Tris–glycine gel; Novex, San Diego, CA) and either stained with Coomassie brilliant blue or transferred to a nitrocellulose membrane by using a Trans-Blot SD apparatus (Bio-Rad, Hercules, CA) at 20 V for 1 h at 4°C. After the transfer was complete, the membrane was blocked in TBS plus 0.2% Tween 20 (TBS plus Tween) with 5% (w/v) dry nonfat milk at room temperature with gentle agitation. After 1 h, the recommended amount of an avidin–horseradish peroxidase conjugate (No. 170-6528, Bio-Rad) was added and the incubation was continued for another hour. Next, the membrane was washed twice for 5 min each with TBS plus Tween and then once with TBS. The Western blot was developed with the ECL chemiluminescence detection kit (Amersham, Arlington Heights, IL) in accordance with the manufacturer's instructions.

Affinity Purification of BAP–MBP–Raf Fusion Proteins

The cells from five 75-cm^2 tissue culture plates infected with either bDW449 or bDW450 were recovered by centrifugation, washed once with TBS, resuspended in 5 ml of 20 mM Tris–HCl (pH 7.6), 100 mM NaCl, 1 mM EDTA with 4-(2-aminoethyl)benzenesulfonyl fluoride (50 μM), leupeptin (1 μg/ml), aprotinin (5 μg/ml), and pepstatin A (1 μg/ml), and frozen at −80°C. The cell suspension was thawed on ice and the cells were disrupted with a Dounce homogenizer. The crude lysate was centrifuged at 15,000*g* for 15 min. Meanwhile, 2 ml of amylose resin (New England Biolabs, Beverly, MA) was equilibrated in 20 mM Tris–HCl, 100 mM NaCl, 1 mM dithiothreitol, and 1 mM EDTA (Buffer A). The soluble extracts were incubated with amylose resin for 30 min at 4°C. The resin was washed 10 times for 5 min each with 30 ml of Buffer A, and the fusion protein was eluted in 2 ml of Buffer A containing 10 mM maltose for 30 min at 4°C. The same procedure was used to purify biotinylated BAP–MBP–Raf fusion protein from cells infected with bDW449 (Fig. 3), except 5 ml of UltraLink immobilized monomeric avidin (Pierce, Rockford, IL) was used instead of amylose resin, and the bound protein was eluted with 5 ml of Buffer A containing 2 mM D-biotin instead of 10 mM maltose.

RESULTS

BAPs Can Be Used to Label Recombiant Proteins with Biotin in S. frugiperda Cells

To evaluate the utility of the BAP as a substrate for enzymatic biotinylation in insect cells, we first constructed recombinant baculoviruses bDW449 and bDW450 (Fig. 1). bDW449 directs the expression of two recombinant proteins. One of these, termed BAP–MBP–Raf, is a tripartite fusion protein consisting of *E. coli* MBP with the Ras-binding domain of human Raf kinase (amino acids 55–132) appended to its C-terminus and a BAP joined to its N-terminus. BAP–

Alamar Exhibit 1002
Alamar v. Olink

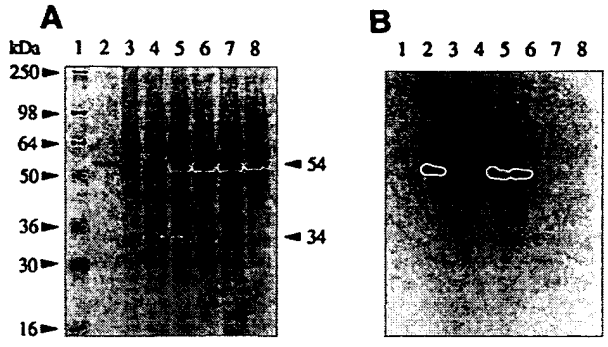


FIG. 2. Expression of recombinant proteins and Western blot analysis. Samples of either the total or soluble intracellular proteins from infected or uninfected cells were prepared and resolved by SDS–PAGE (see Materials and Methods). (A) 12% acrylamide, Tris–glycine gel stained with Coomassie brilliant blue. (B) Western blot of an identical gel, probed with a streptavidin–HRP conjugate. Lanes: (1) See-Blue molecular size standards (Novex); (2) pure biotinylated BAP–MBP–Raf fusion protein; (3) total protein from uninfected Sf9 cells; (4) total protein from Sf9 cells infected with wild-type baculovirus; (5) total protein from Sf9 cells infected with bDW449; (6) soluble protein from Sf9 cells infected with bDW449; (7) total protein from Sf9 cells infected with bDW450; and (8) soluble protein from Sf9 cells infected with bDW450.

MBP–Raf is readily modified by BirA in *E. coli,* but because it accumulates to comprise a large fraction of the intracellular protein, simultaneous overproduction of BirA is required to achieve complete biotinylation *in vivo* (16). We anticipated that a similar situation might arise in insect cells, so we designed bDW449 to overproduce *E. coli* BirA along with the fusion protein. A second recombinant baculovirus, bDW450, produces only the BAP–MBP–Raf fusion protein substrate and not BirA. This construct allowed us to determine whether an endogenous enzyme in *S. frugiperda* is capable of recognizing and modifying the BAP.

Approximately 72 h after infection with bDW449 or bDW450, Sf9 cells were harvested and lysed. Samples of the total and soluble intracellular proteins were prepared and resolved by SDS–PAGE (Fig. 2A). Both recombinant viruses produced a large amount of the 54-kDa BAP–MPB–Raf fusion protein (lanes 5–8). The identity of this protein was corroborated by Western blot analysis with anti-MBP antiserum (data not shown) and by its ability to bind to immobilized amylose (see below). Another very abundant protein with the expected electrophoretic mobility of BirA (ca. 34 kDa) was produced in Sf9 cells infected with bDW449 but not in those cells infected with bDW450. N-terminal sequencing of this band, following its electrophoretic transfer to a polyvinylidene diflouride membrane, yielded the sequence Met-Lys-Asp-Asn, which matches that of BirA. A protein with approximately the same electrophoretic mobility as BirA was produced in cells infected with wild-type baculovirus (lane 4), but not in uninfected cells (lane 3). This protein is probably the product of the polyhedron gene, which is not present in either recombinant baculovirus. BAP–MBP–Raf and BirA were both very abundant and soluble, easily constituting the majority of the intracellular protein in the infected cells.

To determine whether any of the BAP–MBP–Raf fusion protein was biotinylated, we used a streptavidin–HRP conjugate to perform a Western blot of the same experiment (Fig. 2B). A strong chemiluminescent signal was evident in the lanes containing samples prepared from cells that were overproducing *E. coli* BirA (lanes 5 and 6). As expected, pure biotinylated BAP–MBP–Raf protein also reacted with this probe (lane 2). There was no indication, however, of any biotinylated fusion protein in the cells that did not also overproduce BirA (lanes 7 and 8). Although longer exposures revealed the presence of several endogenous biotinyl proteins and a baculovirus-encoded biotinyl protein, there was still no evidence that any biotinylated BAP–MBP–Raf fusion protein was produced under these conditions (data not shown). Additionally, no BAP–MBP–Raf fusion protein could be detected among the products eluted from monomeric avidin resin after incubation with an extract from bDW450-infected cells (data not shown). Hence, unlike the *E. coli* BCCP domain (18, 19), the BAP evidently is not a substrate for enzymatic biotinylation in Sf9 cells.

To determine what fraction of the BAP–MBP–Raf fusion protein was biotinylated in Sf9 cells infected with bDW449, we purified some of the fusion protein from cells infected with either pDW449 or pDW450 by affinity chromatography using amylose resin, which binds to the MBP domain, and then analyzed these samples by electrospray mass spectrometry as previously described (16). The results indicated that all of the fusion protein in cells infected with bDW449 had been modified with a single biotin, whereas none of the fusion protein was biotinylated in cells infected with bDW450 (data not shown). Additionally, we analyzed samples of the soluble extract from cells infected with bDW449 by SDS–PAGE before and after extraction with immobilized monomeric avidin (Fig. 3). Virtually all of the BAP–MBP–Raf, but none of the other proteins, was depleted from the extract after incubation with monomeric avidin resin (lane 3). Moreover, the biotinylated BAP–MBP–Raf fusion protein could be quantitatively recovered from the resin by elution with a buffer containing 2 mM biotin (lane 4). A small amount of BirA copurified with the BAP–MBP–Raf fusion protein under these conditions, probably because of its weak affinity for biotin (25) and overabundance in the crude cell extract.

Alamar Exhibit 1002
Alamar v. Olink

Best Available Copy

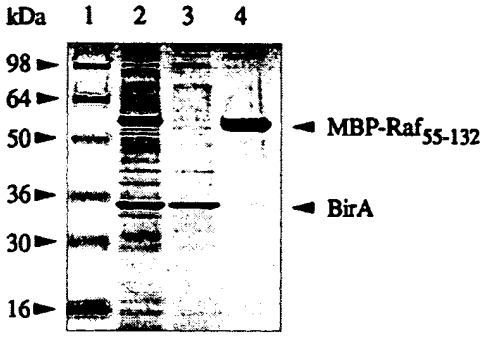


FIG. 3. Affinity purification of biotinylated BAP–MBP–Raf fusion protein using monomeric avidin resin. A soluble extract from Sf9 cells infected with bDW449 was prepared and incubated with monomeric avidin resin in batch mode (see Materials and Methods). Samples of protein before and after incubation with the monomeric avidin were resolved by SDS–PAGE (12% acrylamide, Tris–glycine gel) and stained with Coomassie brilliant blue. Lanes: (1) See-Blue molecular size standards (Novex); (2) soluble extract; (3) soluble extract after incubation with monomeric avidin resin; and (4) protein eluted from monomeric avidin resin.

Baculovirus Transfer Vectors for the Production of Biotinyl Proteins in S. frugiperda Cells

These results encouraged us to design and construct four baculovirus transfer vectors, termed BioBac vectors, that can be used to produce biotinyl proteins in insect cells (Fig. 4). These BioBac vectors are of two general types (Fig. 4A). pDW445 and pDW446 are conventional transfer vectors that work by homologous recombination with baculovirus DNA *in vivo*. Because they recombine at the polyhedron locus, these vectors are compatible with a variety of methods that are commonly used to identify or enrich for recombinant baculoviruses, including those employed by the BaculoGold (Pharmingen) and BlueBac (Invitrogen) kits. pDW464 and pDW465, on the other hand, were designed to use transposon-mediated, site-specific recombination to construct viral genomes in *E. coli*, which are subsequently transfected into insect cells (26). Like bDW449 (see above), all four vectors also produce *E. coli* BirA from the polyhedron promoter.

Each general type of vector is available in two different forms, enabling the BAP to be appended to either terminus of a recombinant protein (Fig. 4B). The BAP encoded by these BioBac vectors has been described previously (15, 16). In pDW445 and pDW464, which are designed to express recombinant proteins with a BAP appended to their N-terminus, translation initiates at the methionine codon within the unique *Nco*I site. This initiator met codon is considered to be in an optimum context for translation initiation in eukaryotes (27). The presence of unique *Stu*I and *Bgl*II sites makes it possible to clone N-terminal BAP fusions that were initially assembled for expression in bacteria (16) directly into the BioBac vectors without using PCR. Thus, recombinant proteins that are not expressed in a soluble, active form in *E. coli* can be rapidly prepared for baculovirus expression. A variety of other unique restriction sites is also available for the *de novo* cloning of inserts.

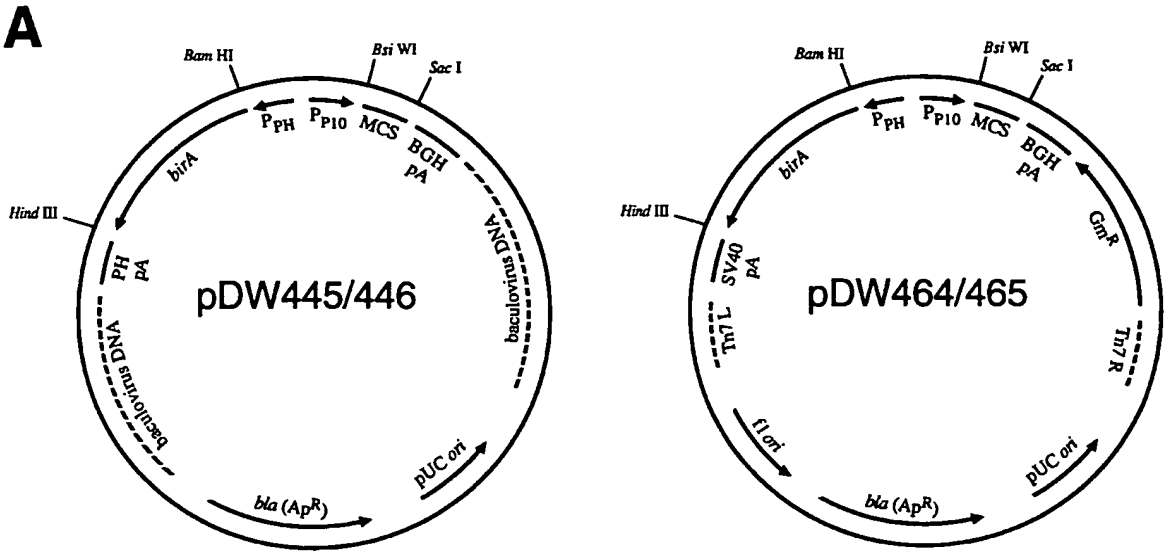
DISCUSSION

Using a model protein with a BAP appended to its N-terminus, we have demonstrated that it is possible to achieve complete biotinylation of a very abundant intracellular protein in Sf9 cells, albeit only in the presence of *E. coli* BirA. While the requirement for coexpression of BirA might be considered a disadvantage by some, we do not view it as such. No difference in the yield of the BAP–MBP–Raf fusion protein was observed in the presence or absence of BirA, which was also very abundant in the infected cells (Fig. 2A). Thus, even with the additional metabolic burden imposed by the expression of BirA, the cells were still capable of producing a very large amount of the fusion protein substrate. Besides, based on our prior experience with the bacterial system (16), we suspect that coproduction of an enzyme catalyst would be required to achieve efficient (i.e., complete) biotinylation of an abundant protein, regardless of whether it is fused to a BAP or a BCCP domain. In the few instances in which BCCP fusion proteins have been produced in *S. frugiperda* and other eukaryotes, the yield was comparatively low and the efficiency of biotinylation was not precisely determined (18–20).

It is curious that although both the BAP and the biotin acceptor domain of BCCP are substrates for *E. coli* BirA, only the latter is biotinylated by the endogenous *S. frugiperda* enzyme. This result suggests that the BAP does not simply mimic the same features of BCCP that are recognized by the *S. frugiperda* homolog of BirA. Rather, the peptide may exploit novel interactions with *E. coli* BirA that are not available in the *S. frugiperda* enzyme. In this regard, it would be of considerable interest to solve the crystal structure of the BAP/BirA complex and compare the conformation of the BAP with that of the BCCP domain.

As discussed above, one of the principal advantages of site-specific biotinylation *in vivo* is that it enables recombinant proteins to be immobilized in a directed orientation on avidin- or streptavidin-coated surfaces. Apart from this, however, enzymatic biotinylation *in vivo* also has some noteworthy advantages as a general method for the affinity purification of proteins. Biotinyl proteins bind tightly and specifically to immobilized monomeric avidin, which is commercially available and easy to prepare. Yet, the interaction is weak enough ($K_d \sim 5 \times 10^{-8}$ M) (28) that the biotinyl proteins can be efficiently displaced from the resin by millimolar concentrations of free biotin in solution. In

Alamar Exhibit 1002
Alamar v. Olink



B

pDW445/pDW464

H2N-M A G G L N D I F E A Q K I E W H E D T G G A P G P S R I L A R S E L

CGTACGCACCATGGCAGGAGGCCTGAATGATATCTTTGAAGCACAAAAAATTGAATGGCATGAAGATACTGGTGGCGCGCCCGGGCCCTCTAGAATTCTGGCCAGATCTGAGCTC

BsiWI NcoI StuI *AscI SmaI ApaI XbaI EcoRI MscI BglII SacI*

pDW446/pDW465

V A R I L E A R P G G L N D I F E A Q K I E W H E D T-COOH

CGTACGTGGCCAGAATTCTAGAGGCGCGCCCGGGAGGCCTGAATGATATCTTTGAAGCACAAAAAATTGAATGGCATGAAGATACGTAAAGATCTGAGCTC

BsiWI MscI EcoRI XbaI AscI SmaI StuI *BglII SacI*

FIG. 4. BioBac baculovirus transfer vectors for the production of recombinant biotinyl proteins in insect cells. (A) Vectors are shown in schematic form (not to scale), with some of the unique restriction sites indicated. In pDW445 and pDW446, regulons for the production of recombinant proteins are flanked by long segments of baculovirus DNA, which promote homologous recombination *in vivo*. In pDW464 and pDW465, these regulons are flanked instead by sequences from the ends of the bacterial transposon Tn7, enabling recombinant baculoviruses to be constructed by site-specific recombination in *E. coli* (26). Abbreviations: PH pA, BGH pA, and SV40 pA, polyadenosine sequences from the baculovirus polyhedron gene, bovine growth hormone gene, and simian virus SV40, respectively; P_{PH}, polyhedron promoter; P_{p10}, p10 promoter; MCS, multiple cloning site; *birA, E. coli* biotin holoenzyme synthetase gene; pUC *ori* and f1 *ori*, origins of replication from the bacterial plasmid pUC19 (30) and the filamentous phage f1, respectively; *bla*, β-lactamase gene (confers resistance to ampicillin); Tn7L and Tn7R, left and right ends of the bacterial transposon Tn7, respectively; Gm^R, gene conferring resistance to the antibiotic gentamycin. (B) BAP coding sequences and cloning sites contained in the BioBac transfer vectors. The nucleotide sequence between the unique *Bsi*WI and *Sac*I sites in pDW445 and pDW464 (top) or in pDW446 and pDW465 (bottom) is presented. All of the indicated restriction sites are unique (i.e., occur only once) in these vectors. The amino acid translation is shown in single-letter code over the nucleotide sequence. The lysine (K) that is biotinylated by BirA is enclosed by a black box. In pDW445 and pDW464, translation initiates at the methionine codon that is coincident with the *Nco*I site, and the BAP precedes the multiple cloning sites. In pDW446 and pDW465, the initiation codon must be provided by the cloned insert, and translation terminates at the end of the BAP coding sequence.

contrast to the similar size polyhistidine tag (29), we have found that it is usually possible to obtain nearly pure preparations of recombinant protein in a single chromatographic step under native (nondenaturing) conditions using the BAP. Hence, a unique attribute of the BAP is that it combines the advantages of small size and tight, specific binding under native conditions. Furthermore, unlike proteinaceous affinity handles, such as MBP, glutathione *S*-transferase, thioredoxin, or BCCP, the BAP seems less likely to interfere with the folding or biological activity of a fusion partner.

The BioBac vectors we have described in this report are the first general vectors for the production of biotinyl proteins in insect cells. Based on the results obtained with bDW449, we expect that these vectors will be capable of directing the high-level expression of

Alamar Exhibit 1002
Alamar v. Olink

recombinant proteins in insect cells, and that even very abundant proteins will be biotinylated with high efficiency. This represents a significant advance in terms of the range of proteins to which this technology can be applied, since many intracellular proteins that cannot be expressed in a soluble, active form in *E. coli* can be produced in Sf9 cells. Additionally, nearly all proteins can be elongated at one end or the other without compromising their biological activity. Thus, it would seem that now almost any intracellular protein can be labeled with biotin at a single, predetermined site with high efficiency *in vivo*.

ACKNOWLEDGMENTS

We thank our colleagues at Hoffmann-La Roche, Inc., for DNA sequencing support (J. Duker and D. Larigan) and LC-ESMS analysis (B. DeBarbieri and H. Michel). This research was sponsored in part by the National Cancer Institute, DHHS, under contract with ABL.

REFERENCES

1. Bayer, E. A., and Wilchek, M. (1990) *Methods Enzymol.* **184,** 138–160.
2. Samols, D., Thornton, C. G., Murtif, V. L., Kumar, G. K., Haase, F. C., and Wood, H. G. (1988) *J. Biol. Chem.* **263,** 6461–6464.
3. Lindqvist, Y., and Schneider, G. (1996) *Curr. Opin. Struct. Biol.* **6,** 798–803.
4. Otsuka, A. J., and Abelson, J. (1978) *Nature* **276,** 689–694.
5. Prakash, O., and Eisenberg, M. A. (1979) *Proc. Natl. Acad. Sci. USA* **76,** 5592–5595.
6. Cronan, J. (1990) *J. Biol. Chem.* **265,** 10327–10333.
7. Reed, K. E., and Cronan, J. E. (1991) *J. Biol. Chem.* **266,** 11425–11428.
8. Weiss, E., Chatellier, J., and Orfanoudakis, G. (1994) *Protein Expr. Purif.* **5,** 509–517.
9. Jander, G., Cronan, J. E., and Beckwith, J. (1996) *J. Bacteriol.* **178,** 3049–3058.
10. Tatsumi, H., Fukuda, S., Kikuchi, M., and Koyama, Y. (1996) *Anal. Biochem.* **243,** 176–180.
11. Kohanski, R. A., and Lane, M. D. (1990) *Methods Enzymol.* **184,** 194–200.
12. Morag, E., Bayer, E. A., and Wilchek, M. (1996) *Anal. Biochem.* **243,** 257–263.
13. Sano, T., and Cantor, C. R. (1995) *Proc. Natl. Acad. Sci. USA* **92,** 3180–3184.
14. Nenortas, E., and Beckett, D. (1996) *J. Biol. Chem.* **271,** 7559–7567.
15. Schatz, P. J. (1993) *Bio/Technology* **11,** 1138–1143.
16. Tsao, K.-L., DeBarbieri, B., Michel, H., and Waugh, D. S. (1996) *Gene* **169,** 59–64.
17. Tucker, J., and Grisshammer, R. (1996) *Biochem. J.* **317,** 891–899.
18. Berliner, E., Mahtani, H. K., Karki, S., Chu, L. G., Cronan, J. E., and Gelles, J. (1994) *J. Biol. Chem.* **269,** 8610–8615.
19. Lerner, C. G., and Chiang Saiki, A. Y. (1996) *Anal. Biochem.* **240,** 185–196.
20. Ackerman, S. H., Martin, J., and Tzagoloff, A. (1992) *J. Biol. Chem.* **267,** 7386–7394.
21. Sambrook, J., Fritsch, E. G., and Maniatis, T. (1989) Molecular Cloning: A Laboratory Manual (Ford, N., Nolan, C., and Ferguson, M., Eds.), 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
22. Dower, W. J., Miller, J. F., and Ragsdale, C. W. (1988) *Nucleic Acids Res.* **16,** 6127–6145.
23. O'Reilly, D. R., Miller, L. K., and Luckow, V. A. (1994) Baculovirus Expression Vectors: A Laboratory Manual, Oxford Univ. Press, New York, NY.
24. Coligan, J. E., Dunn, B. M., Plooegh, H. L., Speicher, D. W., and Wingfield, P. T. (1995) Current Protocols in Protein Science, Wiley, New York, NY.
25. Xu, Y., and Beckett, D. (1996) *Biochemistry* **35,** 1783–1792.
26. Luckow, V. A., Lee, S. C., Barry, G. F., and Olins, P. O. (1993) *J. Virol.* **67,** 4566–4579.
27. Kozak, M. (1994) *Biochimie* **76,** 815–821.
28. Green, N. M., and Toms, E. J. (1973) *Biochem. J.* **133,** 687–698.
29. Hochuli, E., Bannwarth, W., Dobeli, H., Gentz, R., and Stuber, D. (1988) *Bio/Technology* **6,** 1321–1325.
30. Yanisch-Perron, C., Vieira, J., and Messing, J. (1985) *Gene* **33,** 103–119.

Alamar Exhibit 1002
Alamar v. Olink

Protein detection using proximity-dependent DNA ligation assays

Simon Fredriksson[1], Mats Gullberg[1], Jonas Jarvius[1], Charlotta Olsson[1], Kristian Pietras[2], Sigrún Margrét Gústafsdóttir[1], Arne Östman[2], and Ulf Landegren[1]*

The advent of *in vitro* DNA amplification has enabled rapid acquisition of genomic information. We present here an analogous technique for protein detection, in which the coordinated and proximal binding of a target protein by two DNA aptamers promotes ligation of oligonucleotides linked to each aptamer affinity probe . The ligation of two such proximity probes gives rise to an amplifiable DNA sequence that reflects the identity and amount of the target protein. This proximity ligation assay detects zeptomole (40×10^{-21} mol) amounts of the cytokine platelet-derived growth factor (PDGF) without washes or separations, and the mechanism can be generalized to other forms of protein analysis.

© 2002 Nature Publishing Group http://biotech.nature.com

Sensitive and specific methods are required to study concentration, location and colocalization, and secondary modification of the large sets of proteins that are predicted from genome sequence data—stringent demands that are difficult to meet with current techniques. By contrast, methods in use for detection of specific nucleic acid sequences, such as PCR, offer excellent sensitivity and specificity[1]. This is only partially a consequence of the strong and predictable affinity between complementary nucleic acid strands. Of greater importance is the requirement that pairs of probes recognize a given target sequence, and that only this dual interaction can initiate an amplified detection signal. Such a requirement for dual and proximate recognition is also a frequent element of cellular recognition reactions in signal transduction and gene regulation[2].

We have developed an analogous method for *in vitro* analysis of proteins and other macromolecules, termed proximity ligation. The method depends on the simultaneous and proximate recognition of target molecules by pairs of affinity probes, giving rise to an amplifiable detection signal. The molecule we targeted was the homodimer of the platelet-derived growth factor B-chain (PDGF-BB), a cytokine with growth- and differentiation-promoting effects[3]. As affinity probes we used so-called DNA aptamers, a class of oligonucleotide reagents obtained through a process of *in vitro* selection for affinity to a target molecule[4,5], and shown to be suitable for analytical applications[6,7]. A DNA aptamer with affinity for PDGF-BB (ref. 8) was extended by additional sequence elements at either the 5′ or the 3′ end, forming a proximity probe pair. When pairs of probes bind PDGF-BB, the free ends of their sequence extensions are brought sufficiently close to hybridize together to a subsequently added connector oligonucleotide, allowing the ends to be joined by enzymatic DNA ligation. The detected protein molecules thus promote the ligation reaction by bringing together the free ends of the proximity probe pairs. The ligation products can then be replicated by nucleic acid amplification through PCR, while unreacted probes remain silent (Fig. 1).

Results

Detection of PDGF-BB by proximity ligation. On the basis of crude molecular modeling, we selected DNA extensions of the affinity probes of around 40 nucleotides each. However, we found that the length of the extensions could be varied over a considerable range with negligible effects on ligation efficiency and nonspecific signal (data not shown). The method thus also has the potential to allow analysis of large proteins.

The connector oligonucleotide was designed to base-pair to ten nucleotides at the free end of each member of the proximity probe pair. A short hybridizing length and 5′ and 3′ end sequences that do not base-pair to the proximity probes were selected to prevent the connector oligonucleotide from giving rise to ligation-independent amplification products by acting as a primer and/or template for amplification. Longer connector oligonucleotides of 14 + 14 or 16 + 16 hybridizing nucleotides yielded false PCR products (data not shown). The connector oligonucleotide was added in large molar excess over proximity probes to ensure efficient formation of ligation substrates.

We carried out homogeneous assays for PDGF-BB, requiring no washes or separations, by first preincubating proximity probes with the samples at room temperature for 1 h or at 37°C for 15 min. We then added a mixture that included all reagents required for both probe ligation and quantitative PCR with real-time detection[9], thereby also diluting the reaction. Finally, we transferred the reactions to a fluorometric PCR instrument for detection of ligated molecules.

To optimize assay parameters and investigate reaction mechanisms, we preincubated a proximity probe pair with a fixed amount of PDGF-BB or with no PDGF-BB. Then we added variable amounts of connector oligonucleotide and allowed it to act as template for the ligation of proximity probe pairs for 5, 25, or 125 min before PCR with real-time detection (Fig. 2A). With addition of high concentrations of connector oligonucleotide, colocalized ends of proximity probes were joined rapidly and extended incubation resulted in little increase in either signal or background. This lack of further ligation at high connector concentrations renders the assay quite robust and

[1]*The Beijer Laboratory, Department of Genetics and Pathology, Rudbeck Laboratory, Se-75185 Uppsala, Sweden.* [2]*The Ludwig Institute for Cancer Research, Uppsala Biomedical Centre, Box 595, Se-75123 Uppsala, Sweden.* **Corresponding author (Ulf.Landegren@genpat.uu.se).*

Alamar Exhibit 1002
Alamar v. Olink

Best Available Copy

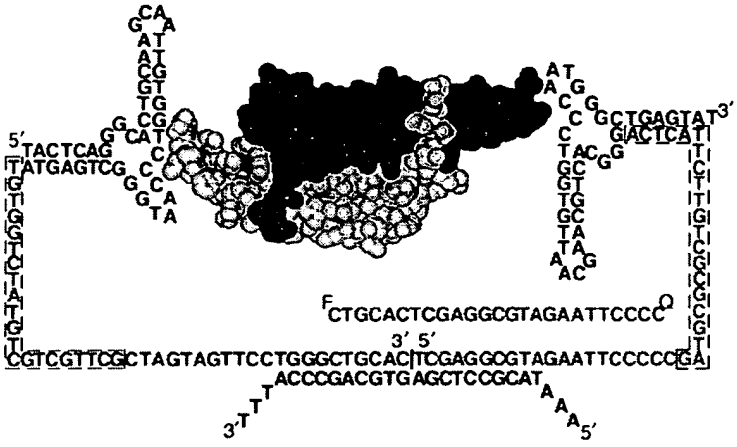


Figure 1. Schematic view of the homodimeric PDGF-BB (ref. 23) bound by two aptamer-based proximity probes, A1 and A2, for detection by proximity ligation. The sequence of the aptamer 41t specific for the PDGF B-chain[8] is shown in black, sequence extensions to be joined by ligation upon hybridization to a common connector oligonucleotide are shown in blue (A1) and red (A2), and primer sites for PCR are boxed. A probe for real-time detection of PCR products via the 5′-nuclease assay is shown in green with a 5′ FAM fluorophore and a 3′ TAMRA quencher.

probably reflects the hybridization of nearly all free proximity probes to one connector oligonucleotide each, inhibiting the formation of new ligation substrates. A connector oligonucleotide concentration of 400 nM was selected for subsequent experiments, and ligation was allowed to proceed for 5 min before amplification.

Next, we varied preincubation volumes and concentrations of proximity probes. All samples were then brought to a final volume of 50 µl by adding the ligation and amplification mix (Fig. 2B). Increased detection sensitivity resulted if PDGF-BB-independent ligation events were limited by using a minimal amount of probes, and by preincubating the proximity probes and sample in the smallest possible volume to promote binding. In further experiments we used 20 pM proximity probes in a preincubation volume of 5 µl. The very low probe concentration during the ligation phase of the assay keeps target-independent ligation events to a minimum, and requires a mere 100 attomoles of proximity probes per assay.

By assuming that the ends of oligonucleotide pairs that have bound a protein molecule are constrained within a sphere of a radius defined by the size of the protein (~60 Å) and taking into account the maximal length of the nucleotide extensions (5 Å per nucleotide), we estimate that under these experimental conditions local concentration of ligatable ends of the proximity probes increases by a factor of at least 10^6 upon target binding.

As few as 24,000 PDGF-BB molecules were reproducibly detected over background plus two standard deviations by proximity ligation, ~1,000-fold fewer than were detected by a sandwich enzyme-linked immunosorbent assay (ELISA) for the same protein (Fig. 3). The linear range for quantitation extended over a >1,000-fold concentration range, with sufficient precision to distinguish 2-fold dilutions of the protein. The error associated with the analysis is of the same order as that of quantitative PCR with real-time detection alone, calculated for the number of amplicons generated: 35% coefficient of variation (CV) for proximity ligation (averaged in Fig. 3) and 32% CV for 88 replicates of the diluted amplicon analyzed by real-time PCR.

Under the reported conditions, approximately 1 PDGF molecule in 25 had bound 2 different proximity probes that were subsequently joined by ligation. The reason that not all protein molecules trigger a ligation reaction is probably a consequence of several factors: most importantly, at 20 pM concentrations of proximity probe, well below the reported K_d for the aptamers (~100 pM), only a minority of proteins are expected to have bound one proximity probe, and even fewer will have bound two. Moreover, only one in two such complexes of target protein with two proximity probes involve probe pairs with one 5′ end and one 3′ end available for ligation. Finally, the ligation efficiency of juxtaposed proximity probe ends could be <100% as a result of imperfections of oligonucleotide synthesis.

The aptamers have a reported 500-fold lower affinity for the PDGF A-chain compared with the B-chain, and accordingly the PDGF-AA homodimer was not detected in the assay and neither were the PDGF-BB homologs PDGF-CC (ref. 10) and PDGF-DD (refs 11,12) (Fig. 3). Furthermore, the PDGF-AB heterodimer was not detected, demonstrating that the assay is truly proximity dependent. The detection reaction could be specifically inhibited through the addition of a fragment of the PDGF β-receptor[13], known to compete with the aptamer for binding to PDGF-BB (ref. 8). PDGF-BB was also analyzed in the presence of fetal calf serum (FCS), cerebrospinal fluid (CSF), and Eagle's minimal essential medium (EMEM) to demonstrate the applicability of the assay in realistic biological samples and cell culture medium (Fig. 4). Closely similar results were obtained in biological samples and in defined media (compare Figs 3 and 4).

The concentration of PDGF-BB in a human serum sample was demonstrated by homogeneous proximity ligation to be 0.2 nM (6 ng/ml), compared with a value of 0.13 nM as determined by ELISA. By contrast, the PDGF-BB content in the corresponding plasma sample was <1 pM as determined by ELISA and <0.2 pM using proximity ligation. The proximity ligation assay further revealed a concentration of 1 pM PDGF-BB in conditioned medium from a 36 h culture of the human anaplastic thyroid carcinoma cell line SW-1736 (ref. 14), shown by northern blot analysis to

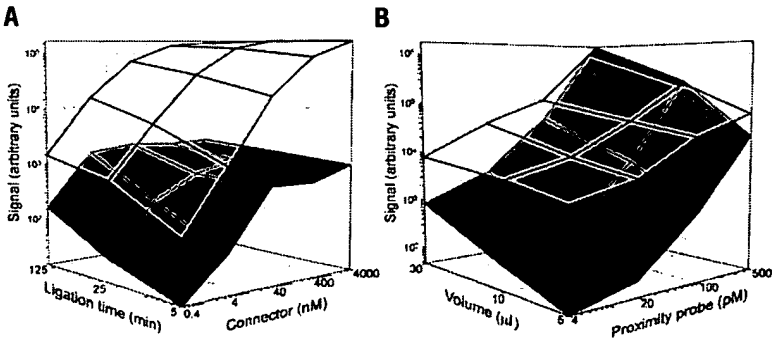


Figure 2. Optimization of reaction conditions for the proximity ligation assay. (A) Effect of connector oligonucleotide concentration and time of ligation on signal and background. (B) Influence of preincubation volume and probe concentration. The upper, yellow layers connect the signals recorded in reactions with 5.6 attomoles of PDGF-BB, while the lower, red layers represent the target-independent background, expressed in arbitrary units. Averages of duplicate measurements are shown.

© 2002 Nature Publishing Group http://biotech.nature.com

Alamar Exhibit 1002
Alamar v. Olink

© 2002 Nature Publishing Group http://biotech.nature.com

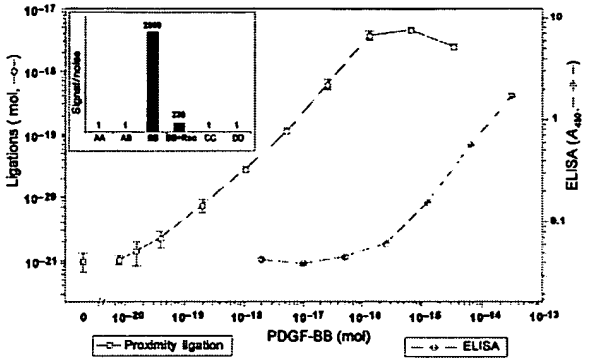


Figure 3. Detection of a dilution series of PDGF-BB by proximity ligation (solid line) and by sandwich ELISA (dashed line). Average values of triplicate and duplicate determinations, respectively, are shown with s.e.m. In the inset panel, results are shown of analyses of 3 fmol PDGF-AA, -AB, -BB, -BB with 1.5 pmol of the PDGF β-receptor, PDGF-CC, and -DD.

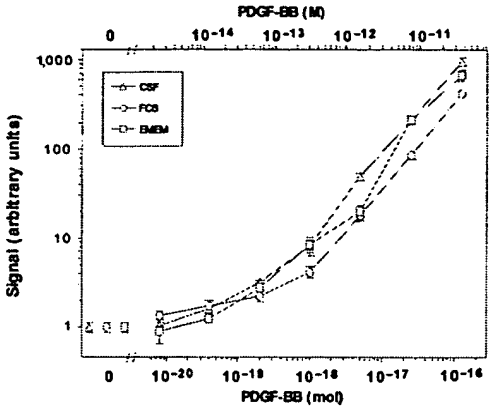


Figure 4. Homogeneous detection of PDGF-BB in the presence of complex biological fluids and cell culture media: FCS, EMEM, and human CSF. The top *x*-axis indicates concentration of PDGF-BB, and the corresponding molar amounts are shown at the bottom *x*-axis. Signals are normalized to a scale of arbitrary units of duplicate measurements with indicated s.e.m.

transcribe the PDGF-B gene (N. Heldin, personal communication). This corresponds to an estimated production of 1,500 cytokine molecules per cell per day.

Proximity ligation detection of human α-thrombin. We also designed proximity probes for detection of human α-thrombin. A pair of DNA aptamers binding two distinct sites[15,16] were extended with sequences for ligation and amplification to form a proximity probe pair (Fig. 5). The sensitivity of detection illustrates that the proximity ligation strategy is also effective when using affinity probes for two distinct determinants on one protein, and that the procedure should also be useful for detecting interactions between different protein molecules.

Solid-phase proximity ligation. The homogeneous proximity ligation assay is simple to standardize and to carry out for large sets of samples. It eliminates the need for extensive, carefully controlled washes as required in solid-phase assays. However, for samples containing potential inhibitors of ligation or amplification, or in which the concentration of analyte is exceedingly low, solid-support immobilization of the protein can serve both to purify and to concentrate the samples. It also permits removal of unbound probes through washes before the ligation step, further reducing target-independent probe ligation. By immobilizing PDGF-BB through binding to antibody-coated reaction wells, we detected low-femtomolar concentrations of the protein using the same pair of proximity probes as was used in the homogeneous assay (Fig. 6). This compares with a low-picomolar detection threshold in the ELISA (Fig. 3).

PCR has previously been used in solid-phase protein assays for sensitive detection of single affinity probes labeled with DNA strands (immuno-PCR; ref. 17). To compare this strategy with proximity ligation, variable amounts of PDGF-BB were bound to antibodies immobilized in microtiter wells. This permitted us to compare detection of binding by a single aptamer having an oligonucleotide extension amplifiable by PCR, with the use of a proximity probe pair that could be amplified only upon ligation (Fig. 6). Across a range of probe concentrations and washing conditions, immuno-PCR resulted in a considerably higher background, and thereby a reduced sensitivity compared with proximity ligation. In the immuno-PCR approach, any nonspecifically bound probes can contribute to background signals, whereas proximity ligation requires proximity of pairs of probes to yield a signal.

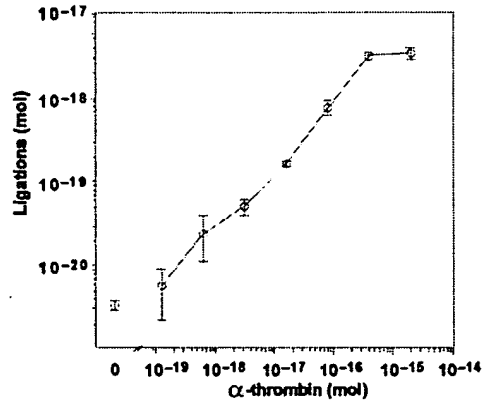


Figure 5. Detection of human α-thrombin by homogeneous proximity ligation. The *x*-axis indicates molar amounts of protein. Averages of duplicate measurements are shown with s.e.m.

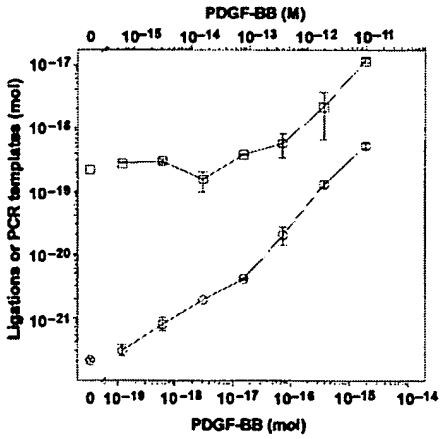


Figure 6. Dilution series of PDGF-BB analyzed in a solid-phase assay, either by proximity ligation (circles) or by aptamer-based immuno-PCR (squares). The top *x*-axis indicates concentration of PDGF-BB analyzed with the corresponding molar amounts at the bottom *x*-axis. Results are represented as averages of duplicate measurements with s.e.m.

Alamar Exhibit 1002
Alamar v. Olink

© 2002 Nature Publishing Group http://biotech.nature.com

Discussion

There are advantages to using aptamers in proximity ligation in that the attachment of DNA extensions is trivial, isolation of aptamers can be automated and scaled up[18], and high-affinity reagents are regularly obtained, as demonstrated in a recent study of 100 protein-specific aptamers[19]. Moreover, this class of affinity probes can be shared in the research community by publishing the nucleotide sequence of appropriate reagents, permitting standardization of protein assays between labs. Nonetheless, most currently available reagents for protein binding are themselves proteins. Preliminary results demonstrate that monoclonal antibodies are suitable as proximity probes after coupling of oligonucleotides (M.G., unpublished results), and that heteroantisera, affinity probes prepared by phage/ribosome display techniques[20], or other types of affinity reagents should also be useful.

The proximity-dependent ligation assay converts the detection of any macromolecule for which suitable pairs of affinity probes exist into a more convenient reaction for detecting nucleic acids. The use of a DNA reporter for the detection reactions permits simultaneous analysis of large sets of proteins by encoding tag sequences in the amplified segment for separate detection, for example by hybridization to a standard oligonucleotide array. Proximity ligation should also enhance the sensitivity and specificity of localized *in situ* detection of proteins, and it could provide a valuable means to investigate secondary modifications of specific proteins, as well as the interaction of pairs of biomolecules in macromolecular complexes.

Experimental protocol

Detection of PDGF-BB by proximity ligation. Unless otherwise indicated, 1 µl aliquots, containing 5.6 attomoles PDGF-BB diluted in 137 mM NaCl, 10.1 mM Na_2HPO_4, 1.8 mM KH_2PO_4, pH 7.4, 2.7 mM KCl, 1 mM $MgCl_2$, and 1% BSA, were added to optical tubes (Applied Biosystems (ABI), Foster City, CA) containing 20 pM of the proximity probes A1 and A2 in a total volume of 5 µl (50 mM KCl, 10 mM Tris-HCl, pH 8.3, 3.3 mM $MgCl_2$, 0.1% BSA). Upon addition of the combined mix for ligation and amplification, the samples contained 50 mM KCl, 10 mM Tris-HCl, pH 8.3, 1.5 mM $MgCl_2$, 0.4 units T4 DNA ligase (Amersham Biosciences, Uppsala, Sweden), 400 nM connector oligonucleotide, 80 µM ATP, carboxy-x-rhodamine (ROX) internal fluorescence standard, 0.2 mM dNTPs, 0.5 µM primers (forward: 5′-ATGTGGTCTATGTCGTCGTTCG-3′; reverse: 5′-TGAGTAAGAACAGCGCGCAT-3′), 50 nM probe for the 5′ nuclease assay, and 1.5 units AmpliTaq Gold polymerase (ABI) in 50 µl. After a 5 min ligation reaction at room temperature, the reactions were transferred to the real-time PCR instrument for temperature cycling: 10 min at 95°C, and then 15 s at 95°C and 60 s at 60°C, repeated 45 times (ABI PRISM 7700). Signal-to-noise values represent the number of detected ligation products divided by the number of ligation products arising in a sample without PDGF-BB. The number of ligation products was calculated from a standard curve of diluted amplicons.

In experiments for which the concentration of connector oligonucleotide was varied, we took precautions to avoid an influence by the complementarity between the 5′ nuclease probe and the connector oligonucleotide. After a preincubation with proximity probes in a 5 µl volume, a 35 µl ligation mix was added and the ligation reaction was terminated by heating at 95°C for 10 min, followed by the addition of reagents for PCR to a final volume of 50 µl. Furthermore, separate standard curves of diluted amplicons were prepared for each connector concentration.

PDGF-BB detection in biological samples. Before use, FCS was heated to 65°C for 20 min to inactivate the α2-macroglobulin that binds PDGF-BB, blocking the binding of proximity probes[21]. Reactions (5 µl, with 3.4 µl sample and 20 pM proximity probes) were incubated at 37°C for 20 min with 16 µg/ml poly(A) oligonucleotide. In the PCR, $MgCl_2$ (3.5 mM) was used.

PDGF-BB quantitation in biological samples. PDGF-BB was quantified in serum and plasma by adding known amounts of PDGF-BB to a diluted sample and calculating the endogenous concentration from the increase in signal. The specificity of the assay was confirmed by blocking probe binding with an excess of soluble PDGF β-receptor, bringing the signal to background levels. The ELISA was done according to the supplier's instructions (R&D Systems, Minneapolis, MN). The incubation mix also contained 16 µg/ml poly(A) oligonucleotide and PCR was done with 3.5 mM $MgCl_2$.

PDGF-BB expression by the SW-1736 cell line, grown in 2 ml EMEM supplemented with 1% FCS, was assayed by homogeneous proximity ligation of 1 µl samples, diluted 5-fold in PBS with 0.1% BSA, and quantified by reference to analyses of known concentrations of the protein diluted in growth medium. A total of 300,000 cells were plated and after 36 h, 514,000 cells were counted. Specificity of protein detection was confirmed by blocking the signal with an excess of PDGF β-receptor.

Detection of thrombin by proximity ligation. Experimental procedures were identical to those for analysis of PDGF-BB except as indicated below. The proximity probes were Thr1 (5′-CAGTCCGTGGTAGGGCAGGTTGGGGTGACTTCGTGGAACTATCTAGCGGTGTACGTGAGTGGGCATGTAGCAAGAGG-3′) and Thr2 (5′-P-GTCATCATTCGAATCGTACTGCAATCGGGTATTAGGCTAGTGACTACTGGTTGGTGAGGTTGGGTAGTCACAAA-3′), aptamer sequences underlined. The connector oligonucleotide was 5′-AAGAATGATGACCCTCTTGCTAAAA-3′, and the 5′ nuclease probe was 5′-TET TGTACGTGAGTGGGCATGTAGCAAGAGG-3′-TAMRA (TET, tetrachlorofluorescein; TAMRA, carboxytetramethylrhodamine). Samples were diluted in 100 mM 2-(*N*-morpholino)-ethanesulfonic acid (MES) buffer, pH 6.5, and 1% BSA, and preincubated with proximity probes Thr1 (15 pM) and Thr2 (20 pM) in 100 mM NaCl, 50 mM Tris-HCl, pH 7.5, 1 mM $MgCl_2$. The PCR buffer contained 1.9 mM $MgCl_2$. Primers for PCR were 5′-GTGACTTCGTGGAACTATCTAGCG-3′ and 5′-AATACCCGATTGCAGTACGATTC-3′.

Detection of PDGF-BB by solid-phase proximity ligation and immuno-PCR. An anti-PDGF antiserum immunoglobulin fraction[22] (500 ng) in a 20 µl volume was added to optical tubes for real-time PCR, followed by blocking with 1% BSA. Up to 200 µl of sample was then incubated together with 5 nM of proximity probes or with 0.1 nM of immuno-PCR reagent. Unbound probes were removed by washes (200 µl) using a multichannel pipette (proximity ligation 8 times with 3 flushes per wash, immuno-PCR 8 times with 4 flushes per wash (washes 1 and 2, PBS with 0.02% SDS; washes 3–7, PBS; wash 8, H_2O)). A ligation mix (50 mM KCl, 10 mM Tris-HCl, pH 8.3, 1.5 mM $MgCl_2$, 0.15 mM ATP, 50 nM connector oligonucleotide, 2 U T4 DNA ligase in 20 µl) was added and the reactions were kept at 30°C for 5 min and then at 80°C for 20 min. Next, a PCR mix was added to a final volume of 50 µl, containing ROX internal fluorescence standard, 0.1 mM dNTPs, 0.2 µM primers, 50 nM probe for the 5′ nuclease assay, and 1 unit AmpliTaq Gold polymerase.

Acknowledgments

Mats Nilsson and Carl-Henrik Heldin offered valuable comments on the manuscript. Olli Leppänen and Nils-Erik Heldin kindly supplied the PDGF β-receptor fragment and the SW-1736 cell line, respectively. Frida Berg contributed to the connector oligonucleotide studies. S.M.G. was funded by a Norfa stipend. The work was supported by the Beijer and Wallenberg Foundations, the Technical and Medical Research Councils of Sweden, the Swedish Cancer Fund, and by Polysaccharide Research AB (Uppsala).

Received 2 July 2001; accepted 22 February 2002

Alamar Exhibit 1002
Alamar v. Olink

1. Saiki, R.K. *et al.* Enzymatic amplification of β-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science* **230**, 1350–1354 (1985).
2. Hartwell, L.H., Hopfield, J.J., Leibler S. & Murray A.W. From molecular to modular cell biology. *Nature* **402**, c47–52 (1999).
3. Heldin, C.H. & Westermark, B. Mechanism of action and *in vivo* role of platelet-derived growth factor. *Physiol. Rev.* **79**,1283–1316 (1999).
4. Tuerk, C. & Gold, L. Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science* **249**, 505–510 (1990).
5. Ellington, A.D. & Szostak, J.W. *In vitro* selection of RNA molecules that bind specific ligands. *Nature* **346**, 818–822 (1990).
6. Jayasena, S.D. Aptamers: an emerging class of molecules that rival antibodies in diagnostics. *Clin. Chem.* **45**, 1628–1650 (1999).
7. Ellington, A.D. & Robertsson, M.P. *In vitro* selection of an allosteric ribozyme that transduces analytes to amplicons. *Nat. Biotechnol.* **17**, 62–66 (1999).
8. Green, L.S. *et al.* Inhibitory DNA ligands to platelet-derived growth factor B-chain. *Biochemistry* **45**, 14413–14424 (1996).
9. Heid, C.A., Stevens, J., Livak, K.J. & Williams, P.M. Real time quantitative PCR. *Genome Res.* **6**, 986–994 (1996).
10. Li, X. *et al.* PDGF-C is a new protease-activated ligand for the PDGF α-receptor. *Nat. Cell Biol.* **2**, 302–309 (2000).
11. Bergsten, E. *et al.* PDGF-D is a specific, protease-activated ligand for the PDGF β-receptor. *Nat. Cell Biol.* **5**, 512–516 (2001).
12. LaRochelle, W.J. *et al.* PDGF-D, a new protease-activated growth factor. *Nat. Cell Biol.* **5**, 517–521 (2001).
13. Leppänen, O. *et al.* Predimerization of recombinant platelet-derived growth factor receptor extracellular domains increases antagonistic potency. *Biochemistry* **39**, 2370–2375 (2000).
14. Dahlman, T. *et al.* Fibrosis in undifferentiated (anaplastic) thyroid carcinomas: evidence for a dual action of tumor cells in collagen type I synthesis. *J. Pathol.* **191**, 376–386 (2000).
15. Macaya, R.F. *et al.* Structural and functional characterization of potent antithrombotic oligonucleotides possessing both quadruplex and duplex motifs. *Biochemistry* **34**, 4478–4492 (1995).
16. Tasset, D.M., Kubik, M.F. & Steiner, W. Oligonucleotide inhibitors of human thrombin that bind distinct epitopes. *J. Mol. Biol.* **272**, 688–698 (1997).
17. Sano, T., Smith, C.L. & Cantor, C.R. Immuno-PCR: very sensitive antigen detection by means of specific antibody-DNA conjugates. *Science* **258**,120–122 (1992).
18. Drolet, D.W., Jenison, R.D., Smith, D.E., Pratt, D. & Hicke, B.J. A high throughput platform for systematic evolution of ligands by exponential enrichment (SELEX). *Comb. Chem. High Throughput Screen.* **2**, 271–278 (1999).
19. Brody, E.N. *et al.* The use of aptamers in large arrays for molecular diagnostics. *Mol. Diagn.* **4**, 381–388 (1999).
20. Li, M. Applications of display technology in protein analysis. *Nat. Biotechnol.* **18**, 1251–1256 (2000).
21. Bonner, J.C. & Osornio-Vargas, A.R. Differential binding and regulation of platelet-derived growth factor A and B chain isoforms by α2-macroglobulin. *J. Biol. Chem.* **27**, 16236–16242 (1995).
22. Thyberg, J., Ostman, A., Backstrom, G., Westermark, B. & Heldin, C.H. Localization of platelet-derived growth factor (PDGF) in CHO cells transfected with PDGF A- or B-chain cDNA: retention of PDGF-BB in the endoplasmic reticulum and Golgi complex. *J. Cell Sci.* **97**, 219–229 (1990).
23. Oefner, C., D'Arcy, A., Winkler, F.K., Eggimann, B. & Hosang, M. Crystal structure of human platelet-derived growth factor BB. *EMBO J.* **11**, 3921–3926 (1992).

© 2002 Nature Publishing Group http://biotech.nature.com

Alamar Exhibit 1002
Alamar v. Olink

Downloaded from www.genome.org on September 14, 2006

Best Available Copy

GENOME RESEARCH

Decoding Randomly Ordered DNA Arrays

Kevin L. Gunderson, Semyon Kruglyak, Michael S. Graige, Francisco Garcia, Bahram G. Kermani, Chanfeng Zhao, Diping Che, Todd Dickinson, Eliza Wickham, Jim Bierle, Dennis Doucet, Monika Milewski, Robert Yang, Chris Siegmund, Juergen Haas, Lixin Zhou, Arnold Oliphant, Jian-Bing Fan, Steven Barnard and Mark S. Chee

Genome Res. 2004 14: 870-877; originally published online Apr 12, 2004;
Access the most recent version at doi:10.1101/gr.2255804

References This article cites 45 articles, 17 of which can be accessed free at:
http://www.genome.org/cgi/content/full/14/5/870#References

Article cited in:
http://www.genome.org/cgi/content/full/14/5/870#otherarticles

Email alerting service Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or **click here**

Notes

To subscribe to *Genome Research* go to:
http://www.genome.org/subscriptions/

© 2004 Cold Spring Harbor Laboratory Press

Alamar Exhibit 1002
Alamar v. Olink

Downloaded from www.genome.org on September 14, 2006

Decoding Randomly Ordered DNA Arrays

Kevin L. Gunderson,[1,3] Semyon Kruglyak,[1,3] Michael S. Graige,[1] Francisco Garcia,[1] Bahram G. Kermani,[1] Chanfeng Zhao,[1] Diping Che,[1] Todd Dickinson,[1] Eliza Wickham,[1] Jim Bierle,[1] Dennis Doucet,[1] Monika Milewski,[2] Robert Yang,[1] Chris Siegmund,[1] Juergen Haas, Lixin Zhou,[1] Arnold Oliphant,[1] Jian-Bing Fan,[1] Steven Barnard,[1] and Mark S. Chee[1,4]

[1]*Illumina, Inc., San Diego, California 92121, USA;* [2]*Genomics Institute of the Novartis Research Foundation, San Diego, California 92121, USA*

We have developed a simple and efficient algorithm to identify each member of a large collection of DNA-linked objects through the use of hybridization, and have applied it to the manufacture of randomly assembled arrays of beads in wells. Once the algorithm has been used to determine the identity of each bead, the microarray can be used in a wide variety of applications, including single nucleotide polymorphism genotyping and gene expression profiling. The algorithm requires only a few labels and several sequential hybridizations to identify thousands of different DNA sequences with great accuracy. We have decoded tens of thousands of arrays, each with 1520 sequences represented at ~30-fold redundancy by up to ~50,000 beads, with a median error rate of $<1 \times 10^{-4}$ per bead. The approach makes use of error checking codes and provides, for the first time, a direct functional quality control of every element of each array that is manufactured. The algorithm can be applied to any spatially fixed collection of objects or molecules that are associated with specific DNA sequences.

Microarray technology, devised for the analysis of complex biological systems, uses the ability of a DNA strand to hybridize specifically to its complement to extract 1000s of measurements at a time from a single sample (Watson and Crick 1953; Southern et al. 1992; Pease et al. 1994; Schena et al. 1995; Chee et al. 1996; Lockhart et al. 1996; Lockhart and Winzeler 2000). Although relatively new, this technology has enabled a variety of important applications, for example, genome-wide quantitative analysis of gene expression and large-scale single nucleotide polymorphism (SNP) discovery and genotyping (Chee et al. 1996; Lockhart et al. 1996; Wang et al. 1998; Fan et al. 2003; Hardenbol et al. 2003; Kennedy et al. 2003; Yvert et al. 2003). Microarrays are also beginning to play a role in the reinvention of cancer classification and drug discovery (Johnson et al. 2002; van't Veer et al. 2002).

Conventional microarrays are manufactured by spotting or synthesizing probes at known locations on a two-dimensional substrate (Fodor et al. 1991; Schena et al. 1995; Holloway et al. 2002). The significance of our novel approach is that it enables the production of randomly assembled arrays in which the location of a probe is initially unknown (Michael et al. 1998). Random bead loading combined with decoding avoids the need for physical addressing of each element and thus achieves unprecedented levels of miniaturization and very high packing densities by using relatively simple bulk processes (Figs. 1, 2). For example, a typical spotted array with 100-μm center-to-center spacing has ~400-fold lower packing density, and a photolithographically synthesized array (Fodor et al. 1991) with 11-μm center-to-center spacing has about fourfold lower density than that of the arrays described here. The random assembly of 300-nm-diameter beads in 500-nm wells has been reported, a density ~40,000 times higher than that of a typical spotted microarray (Michael et al. 1998).

Although randomly assembled arrays were recognized from the outset as a potentially revolutionary approach to microarray technology, the initial attempts to determine the location and identity of beads could only distinguish a few codes, limiting the usefulness of the approach (Michael et al. 1998). These initial attempts at decoding relied on dye impregnation of beads, but this approach suffered from variability in quantitation, lack of stability, and other problems. A number of other schemes that aim to encode particles directly with combinatorial codes generated by mixtures or spatial arrangements of optical signaling molecules have similar issues (Fulton et al. 1997; Han et al. 2001; Lockhart and Trulson 2001; Nicewarner-Pena et al. 2001; Braeckmans et al. 2002; Chan et al. 2002). We sidestepped the need for complex dye chemistries and painstaking labeling processes by devising a novel, highly efficient decoding algorithm that uses the specificity and reversibility of DNA hybridization.

RESULTS

Design of DNA-Based Decoding

Our algorithm uses sequential hybridizations of dye-labeled oligonucleotides, or decoders, complementary to bead sequences to create a combinatorial decoding scheme for arrays. It is distinct from sequencing by hybridization (SBH), which has been used successfully to characterize sequences de novo by hybridization to all *n*-mers or a well-chosen subset, typically in the range of 4- to 10-mers (Drmanac et al. 1996, 1998; Gunderson et al. 1998; Brenner et al. 2000). Our approach uses longer sequences, each designed to hybridize to a defined target with high specificity. It is capable of decoding, with high accuracy, many 1000s of bead types. Each bead type is defined by a unique DNA sequence that is recognized by a complementary decoder.

To illustrate, we show an example of decoding eight different bead types. We use two fluorescent labels, or states (green and red) in combination with three sequential hybridizations, or

[3]**These authors contributed equally to this work.**
[4]**Corresponding author.**
E-MAIL mchee@illumina.com; FAX (858) 202-4680.
Article and publication are at http://www.genome.org/cgi/doi/10.1101/gr.2255804. Article published online before print in April 2004.

14:870–877 ©2004 by Cold Spring Harbor Laboratory Press ISSN 1088-9051/04; www.genome.org

Alamar Exhibit 1002
Alamar v. Olink

Downloaded from www.genome.org on September 14, 2006

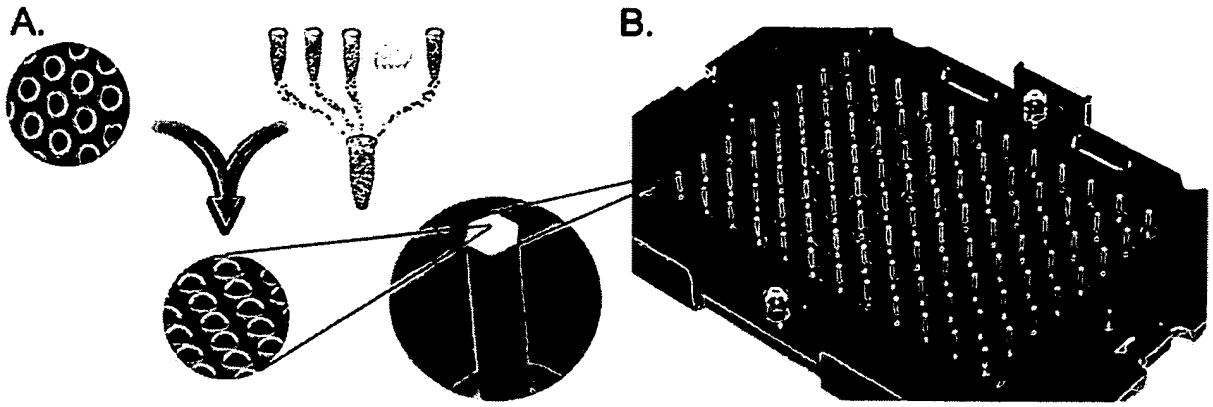


Figure 1 Assembly of a random array. (*A*) Creation of a bead pool and assembly into ~3-μm-diameter wells etched in optical fiber bundles. Once a bead pool is made, it is relatively straightforward to assemble and decode large numbers of arrays. Each array contains ~50,000 beads distributed among 1520 bead types, so that each bead type is represented at ~30-fold redundancy. Scanning electron micrographs are shown of an unassembled and an assembled array containing one bead per well. (*B*) Because individual arrays are only ~1.4 mm in diameter, they can easily be arranged into a 96-array matrix, designed for parallel analysis of samples in standard microtiter plates.

stages. The sequential hybridization process is illustrated for a single bead (bead type 2 of eight) in Figure 2A. In each of the three stages, the bead is "colored" by hybridization to a fluorescently labeled decoder oligonucleotide (Fig. 2A,B). In practice, all beads in the array are labeled simultaneously at each stage, by exposure to a pooled set of decoders, so that the process is intrinsically parallel and efficient.

The combinatorial assignment of green and red within each pool of eight decoders is shown in Figure 2C. There are three decoder pools in total, one for each stage. The stage 1 pool has the first four decoders colored green and the last four colored red. The decoders in subsequent stages are labeled so that after three stages each bead type is assigned a unique three-bit color code. Note that the sequences of the decoders are unchanged from stage to stage; only the fluorescent labels are varied. The bead circled in Figure 2B has the color signature (GRG), or 010 code in binary representation, in which G = 0 and R = 1. Its sequence can be identified as sequence 2 by referring to the color-lookup table in Figure 2C. Although assignment of codes to sequences is unambiguous after three stages, additional stages can be added for error checking purposes (last column of Fig. 2C) to be described below.

In this simple fashion, the eight bead types are decoded with three stages and two color labels. The approach scales exponentially. If there are N bead types and k distinguishable labels, or states, then the number of stages required is $S = \lceil \log_k N \rceil$. Thus, a large number of sequences can be decoded by using only a few labels across a few stages. For instance, four states (colors) combined with eight decoding stages enables up to 65,536 (4^8) different bead types to be decoded.

There are a number of ways of creating different states. Distinct fluorescent labels can be used as described in Figure 2, or the intensity levels of fluorescent labels can be varied to create gray-scale states. We use a process that decodes 1520 different bead types by using three states: two fluorescent "ON" states (FAM and CY3 fluorescent labels) and one nonfluorescent "OFF" state. The logarithmic relationship between the number of bead types and the number of decode stages shows that the 1520 bead types can be decoded in only $\lceil \log_3 1520 \rceil = 7$ stages. In practice, we use an additional stage to enable error checking. Without error checking, a single mistake in resolving a color label in any of the decode stages will lead to misclassification of a bead. This could result from, for example, a weak fluorescent signal or a speck of dust on the array.

Error Checking

The error-checking concepts that we have developed, and that underlie the robustness of our hybridization-based approach, are based on algorithms from digital information theory (Shannon 1948a,b; Hamming 1986). Although they are well known in other fields, they have not, to our knowledge, previously been applied to microarray manufacture or, indeed, more generally to DNA hybridization. With functional testing of every element of every array, we achieve a level of quality control that is unprecedented in microarray manufacture. As a result, we overcome some of the challenges in quality control that can plague the manufacture of ordered microarrays, which can also suffer from random sources of error (Hubbell and Pevzner 1999; Battaglia et al. 2000; Taylor et al. 2001; Sengupta and Tompa 2002; Shearstone et al. 2002; Hessner et al. 2003).

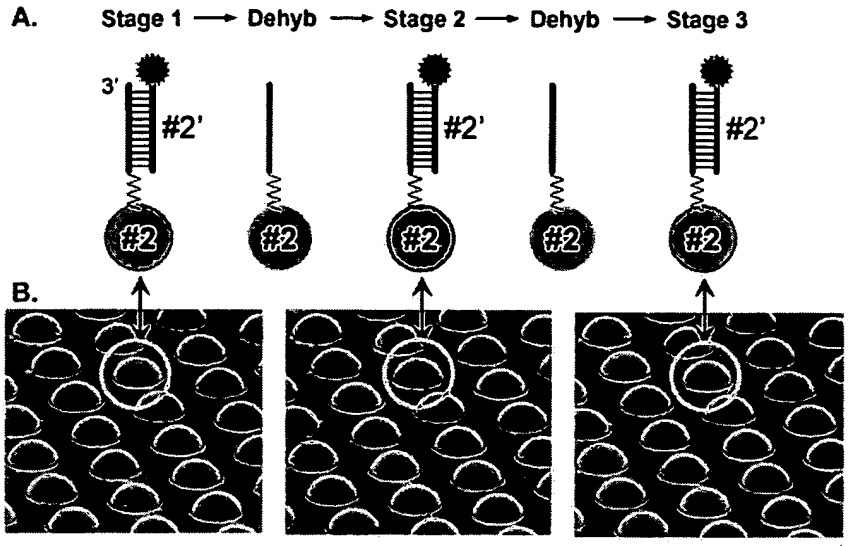
The simplest form of error checking we have used has a parity bit and is illustrated in Figure 2C. The assigned four-bit codes all have an even parity bit sum and are termed valid codes. An error in a single decode stage is a one-bit error that creates an odd parity, or invalid code. Although the parity-based approach is effective, we have optimized error checking by designing a more advanced scheme that assigns codes in a way that takes into account inherent biases in error rates and enables estimation of the misclassification rate.

In our implementation of three-state decoding, the most common errors are transitions from an ON state, one or two, to the OFF state, zero. Less common are transitions from OFF to ON. Transitions from one ON state to the other ON state are extremely rare. This can be explained by the fact that such transitions require two simultaneous classification errors: a mistaken ON in one color channel and a mistaken OFF call in the other color channel. We used these biases in error rates by designing the decode stages so that every valid code has a fixed number of OFF states. For example, if there are 1520 bead types and we use three-state eight-stage decoding, then each valid code would be designed to be OFF in exactly two stages and ON in exactly six stages (the actual scheme used is a slight variant on this design; see Methods). An example of a valid code would be **21110210**.

With this scheme, it is theoretically impossible to misclassify a code through any number of occurrences of the most common error type: a transition from an ON state to an OFF state. The following events can lead to misclassification: a transition from one ON state to the other, or multiple stage errors with at

Alamar Exhibit 1002
Alamar v. Olink

Downloaded from www.genome.org on September 14, 2006



C.

		Stage 1	Stage 2	Stage 3	Signature	Code	Parity Code
Sequences	0				GGG	000	0000
	1				GGR	001	0011
	2				GRG	010	0101
	3				GRR	011	0110
	4				RGG	100	1001
	5				RGR	101	1010
	6				RRG	110	1100
	7				RRR	111	1111

Figure 2 Decoding process. (*A*) The sequential hybridization process is illustrated for a single bead, of bead type 2. In stage 1, a complementary decoder hybridizes to the oligonucleotide probe that is attached to the bead (for details of the procedure, see Methods). The decoder is labeled with a fluorophore (green in stage 1, red in stage 2, and green in stage 3). The fluorescent signal is read by imaging the entire array. The array is then dehybridized, and the process is repeated for two more stages. (*B*) A scanning electron micrograph of an array of beads, artificially colored to represent three sequential hybridization stages. The images, taken collectively, reveal a combinatorial code for each bead. Note that the bead circled in yellow has the color signature GRG or code 010. (*C*) Colors, or states, are assigned to individual decoder sequences in each stage to produce a unique combination across stages. This signature, or code, identifies each bead type. As indicated in the parity code column, an extra decoding stage (data not shown) can be performed to provide an error checking parity bit. After three stages of decoding, all the beads are uniquely identified by their color.

least one ON-to-OFF transition and at least one OFF-to-ON transition. Both events are extremely rare. The space of all possible color codes is divided into three categories: used valid, unused valid, and invalid (Table 1). The unused valid codes represent codes that could be assigned to bead types but are not currently in use. Monitoring the number of beads decoding to this category allows an estimation of the true misclassification rate, as described in the Methods section. With little extra cost to the overall process, this error checking scheme monitors the rate of single state errors, minimizes the number of misclassified beads, and permits the number of misclassified beads, one of the key determinants of array quality, to be estimated.

Decoding of Randomly Assembled Arrays

By using this approach, arrays of 1520 different bead types (~50,000 total beads) were decoded 96 at a time in the Sentrix array matrix format. Representative examples from the decoding are shown in Figure 3, and summary statistics are presented in Table 2. Figure 3A shows a set of hybridization images from all eight stages of the decoding process. High signal to noise allows the code of the circled bead to be read by eye. The code is **11012202**, with FAM = 1, Cy3 = 2, and OFF = 0. Furthermore, we can identify this as a valid used code according to the error checking scheme. Figure 3B shows an example of decoding data for a population of beads from a single array imaged in the FAM channel for one out of the eight decode stages. The clear separation between the two modes in the histogram indicates that most beads can be classified unambiguously as OFF or ON. Information of this type is the input for histogram-based or stage-by-stage decoding, which is used for quality control of individual stages. Actual decoding is carried out by using a core-by-core classification algorithm, which makes use of the fluorescence intensity profile of each bead, measured in FAM and Cy3 channels, across all eight decode stages (Fig. 3C) to classify a bead as ON or OFF. In the example of Figure 3C, the code is **02212110**.

By using the core-by-core algorithm, we have decoded many 10s of 1000s of arrays with a median random error rate of $<1 \times 10^{-4}$ per bead (Table 2). The rate was estimated by using the error checking scheme summarized in Table 1. To get a more direct measure of random error rates, we decoded a matrix of 96 arrays twice and considered beads that were decoded to used codes in

Table 1. Distribution and Use of 6561 Codes in the Three-State, Eight-Stage System

Type	Number	Purpose
Used valid codes	1520	Assigned to actual beads.
Unused valid codes	272	Detect rare two-stage errors. Estimate the number of misclassification events.
Invalid codes	4769	Detect single stage errors. Monitor overall decode process; useful for detecting occasional catastrophic failures.

Alamar Exhibit 1002
Alamar v. Olink

Downloaded from www.genome.org on September 14, 2006

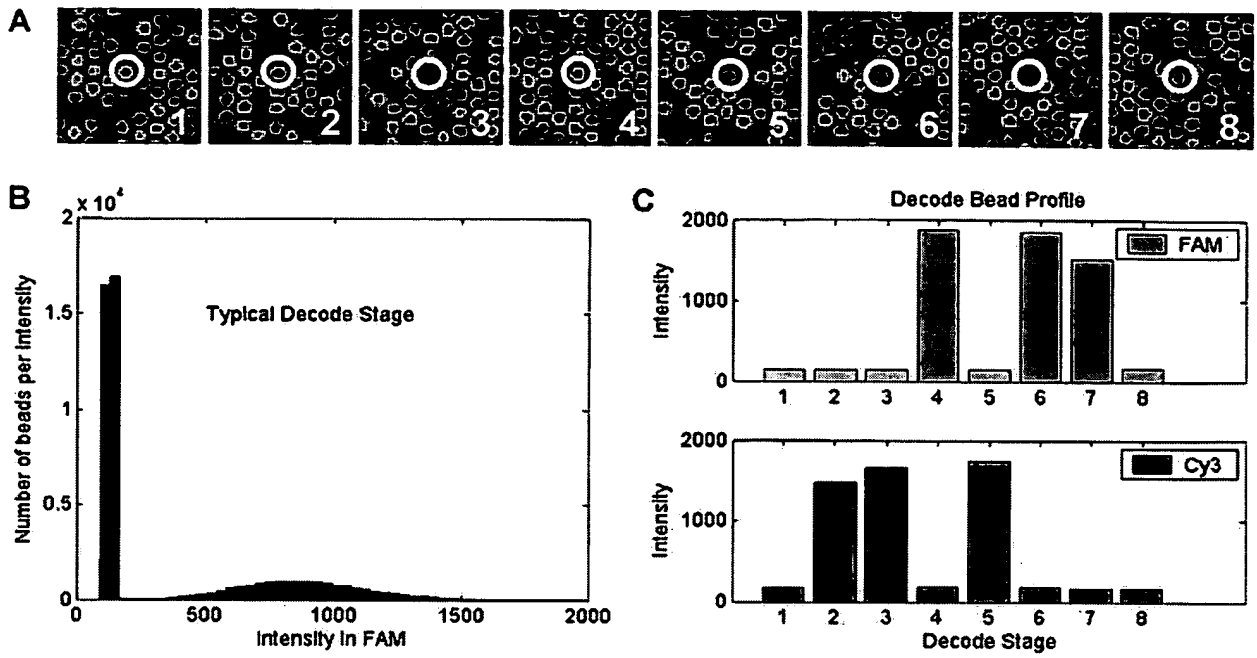


Figure 3 (*A*) Decoding images from eight sequential stages (numbered). Each image is a false-color composite of the FAM and Cy3 grayscale images from each stage. A small region (<0.2%) of a single array is shown. The circled bead is one of up to ~4.8 million in a single 96-array matrix. It has the code **11012202**. A total of 1728 images, each ~5 Mb in size and containing ~1.7 million pixels, are collected in order to decode an array matrix. (*B*) Histogram of bead intensities for a single stage. The low intensity peak includes beads in the OFF state and empty wells. The higher intensity peak includes beads in the ON state. (*C*) The eight-stage intensity profile of a typical individual bead (code = **02212110**) in the FAM and Cy3 channels (~100 counts of intensity are from camera DC offset).

both decoding events (average of 44,912 ± 1098 beads/array). The misclassification rate was then estimated by dividing the number of discrepant calls by the total number of calls. This is an upper bound on the rate, as the errors are distributed among the two decoding events. The mean misclassification rate obtained in this way was 3.8×10^{-5} with a 95% confidence interval of zero to 1.8×10^{-4}. The results are consistent with the estimates obtained in Table 2. This analysis does not account for any systematic misclassification errors, but functional tests (e.g., genotyping comparison studies with other technologies) have not identified any systematic misclassification (data not shown).

Error Rate Impact

The error rate of $<1 \times 10^{-4}$ per bead has a negligible impact on assay accuracy because of the ~30-fold average redundancy and fivefold minimum redundancy of each bead type (Fig. 4). The measured decoding error rate translates to a median of fewer than five misclassified beads per array. Therefore, the chances of more than one random error affecting a given bead type are very low ($\sim 5 \times 10^{-6}$). Given the fact that outlier removal is used in our downstream processes, we estimate that a 100-fold higher average random error rate would have to occur in order to affect our assays (the calculation assumes that the assay is affected if any one bead type on the array has >10% of beads mis-decoded). If we did not use an extra stage for error checking and the number of possible codes was equal to the number of used codes, then any single stage error would lead to misclassification of a bead. Based on the average number of invalid codes generated by our process, we estimate that the average misclassification rate would increase 86-fold. The average array would still perform well, but some fraction of arrays would have unacceptably high misclassification rates.

A fundamental difference between randomly assembled arrays and conventional ordered arrays is that the number of beads (or probes) of each type is intrinsically a random variable with a Poisson sampling distribution for the former (Fig. 4) and is fixed and defined for the latter. Each randomly assembled array is effectively unique, having different numbers and arrangements of beads from array to array yet decoded by a single universal process. This notion is accepted for "liquid arrays" (Fulton et al. 1997), but is fairly radical in the field of spatially fixed microarrays. Nevertheless, the advantage of having multiple beads of each type is that the analytical assay precision is increased by both outlier removal and averaging of replicates (using subsampling of 1 to 20 beads (N) per bead type indicate that the extracted data conform to the theoretical rule of the error of the mean decreasing as $\sqrt{N}$; P. Ng, pers. comm.). At the same time, the random distribution of beads minimizes the chance of any local problem affecting the overall result, increasing robustness of the system. The only added requirement of using "unique" arrays is that the analytical data extraction step must use a uniquely defined template for each particular array. This was easy to implement as part of our data extraction software.

DISCUSSION

Our work provides a high-performance alternative to conventional microarrays. It also expands the reach of microarray assays. For example, we have used a highly miniaturized array format to construct a 96-array matrix for processing many microarray experiments cost-effectively, for ~1500 assays at a time. This provides much needed statistical power that is difficult and prohibitively expensive to obtain by using conventional microarrays, and has the potential to speed the transition of microarray-based assays to large-scale clinical application.

Alamar Exhibit 1002
Alamar v. Olink

Downloaded from www.genome.org on September 14, 2006

Table 2A. Quality Measures for a Random Sampling of 100 Recently Manufactured Array Matrices, Comprising a Total of 9600 Arrays.

	Average beads in used	Average beads in unused	Average beads in invalid
All arrays	46,576	0.14	153
Array matrix 1	48,653	0.21	202
Array matrix 2	46,872	0.29	106
Array matrix 3	48,500	0.04	70
Array matrix 4	45,367	0.09	90
Array matrix 5	44,898	0.85	252
Array matrix 6	46,549	0.77	172
Array matrix 7	46,480	0.20	106
Array matrix 8	46,596	0.17	88
Array matrix 9	46,837	0.06	84
Array matrix 10	48,260	0.04	89

The first row of the table provides averages over all arrays, and other rows provide data on a subset of 10 randomly chosen array matrices. The results shown are based on raw data, prior to the application of any quality control filters.

Table 2B. Summary Statistics for the 9600 Arrays

	Worst percentile	Mean	Best percentile
Bead types decoded	1520	1520	1520
Decode efficiency	97.6%	99.6%	99.9%
Misclassification rate	1.4×10^{-4}	1.2×10^{-5}	0

Decode efficiency is the number of valid signatures assigned, divided by the total number of beads in the array.

We have used decoded arrays to create new assays for large-scale genotyping (Fan et al. 2003) and gene expression (Yeakley et al. 2002; Fan et al. 2004) using a universal array format. We have also developed gene-specific probe sets for gene expression, and demonstrated limits of detection, precision, and dynamic range that are similar or superior to those obtained with conventional microarrays (M. Chee, K. Kuhn, S. Baker, and T. McDaniel, pers. comm.). Therefore, the technology is versatile and capable of conducting all the types of assays that are currently read out on conventional microarrays. For example, a wide variety of alternative genotyping assays could in principle be implemented on our platform (Gerry et al. 1999; Pastinen et al. 2000; Hardenbol et al. 2003). As a result, this work provides the foundation for a host of important existing applications in genomics, as well as new applications that are under development. Our technology is particularly effective at making large numbers of arrays of a given type, and is complementary to microarray technologies that are useful for screening large amounts of sequence with relatively few arrays (Chee et al. 1996; Patil et al. 2001; Nuwaysir et al. 2002).

Importantly, our technology has been proven in highly demanding and competitive large-scale genomics applications, in particular SNP genotyping (Fan et al. 2003). For example, the majority of the genotyping being carried out for the International HapMap Project, which aims to create a detailed map of common genetic variation across the human genome, is being carried out by using randomly assembled arrays manufactured by our decoding process, in conjunction with a new highly multiplexed genotyping assay that we also developed (Fan et al. 2003). The genotypes we have generated for the HapMap project are publicly available (www.hapmap.org). We routinely achieve a combination of call rates of >99.9% together with accuracy of 99.7% to 99.9%, with a capacity to process over a million genotypes per day that is scalable to much higher levels (Fan et al. 2003).

Random arrays have been particularly useful for the accurate, high-throughput, and cost-effective analysis of large numbers of samples for ~1500 assays at a time, a need that was not met by conventional arrays. However, if random arrays are to realize their full potential, the capacity of the decoding scheme must be increased to allow the analysis of 10s or 100s of 1000s of assays per sample. Our approach to decoding is flexible and scalable. We have used a single decode sequence per bead, which works well for decoding 1000s of bead types. The decoding of 100,000s of bead types by the current process would require the synthesis of 100,000s of fluorescently labeled decoder oligos—an expensive and time consuming task. This bottleneck can be overcome by using multiple decoding sequences on each bead type such that if pairs of sequences are used, up to 10^6 (1000×1000) unique combinatoric sequences can be formed from just 2000 primary decoder sequences. By using this combinatoric approach, the decoder pools would only contain a complexity of 1000 to 2000 different sequences rather than 10^6 sequences.

Finally, the decoding algorithm is general and can, in principle, be applied to any spatially fixed collection of objects or molecules that are associated with specific DNA sequences. In genomics, the classification and characterization of large collections of sequences is often a key step in the analysis of complex biological systems. For example, a library of DNA clones is traditionally searched for a single gene of interest by hybridization to a labeled DNA probe (Sambrook 1989). The approach described here would allow a search for 1000s of genes at a time, while distinguishing closely related members of gene families and perhaps alternative splice forms. Similarly, electrophoretic separations of complex mixtures of nucleic acids are often probed to characterize a gene or its mRNA (Southern 1975; Alwine et al. 1979; Liang and Pardee 1992); this can also be parallelized. Fi-

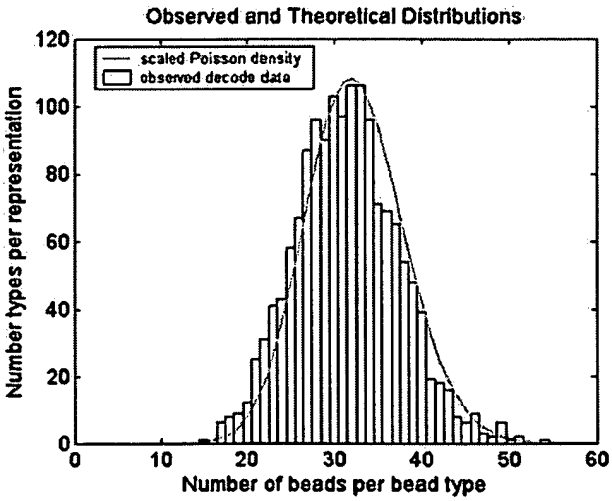


Figure 4 Bead representation histogram from a representative decoded array overlaid with a scaled Poisson density function. The loading of each array is a sampling of beads from a near infinite bead pool. Almost $N \sim 50{,}000$ beads are sampled, and the probability, p, that a sampled bead belongs to a particular type is approximately one divided by the number of bead types. Because N is large, p is small, and the initial bead pool is near infinite, the number of beads from each bead type is well modeled by a Poisson distribution with mean N_p. With 1520 bead types, we average >30 replicate beads per bead type, and the probability that any bead type has fewer than five replicate beads is extremely low ($\sim 5.5 \times 10^{-6}$).

Alamar Exhibit 1002
Alamar v. Olink

Downloaded from www.genome.org on September 14, 2006

nally, fluorescence in situ hybridization (FISH), a powerful class of methods with many applications, could be applied to 1000s of genes simultaneously. To illustrate, FISH with combinatorially labeled oligonucleotide probes has been used to measure transcription from 10 genes in a single cell (Levsky et al. 2002). The decoding strategy we describe could potentially allow transcription to be measured for all genes in a single cell.

In conclusion, we have developed a new and scalable way to make a novel type of microarray. The highly miniaturized arrays have performed well in a variety of applications. The decoding approach used to make them is both accurate and robust and can, in principle, be used to identify not only DNA sequences on beads but also other collections of DNA sequences.

METHODS

Preparation of Oligonucleotide-Linked Beads and Bead Pools

A set of 1536 universal capture oligonucleotide sequences was synthesized. Each sequence was individually immobilized on activated beads, as follows, to create 1536 bead types. Silica beads ~3 μm in diameter (Bangs Laboratories, Inc.) were amino functionalized by incubation in 2.5% 3-amino-propyl-trimethoxysilane (Aldrich) in ethanol for 1 h at RT and then activated by reaction with 2% 2,4,6-trichloro-1,3,5-triazine (Aldrich) in acetonitrile for 2 h at room temperature. Synthetic oligonucleotides labeled at the 5′ terminus with a primary amine were covalently attached to the activated beads by overnight reaction at 50°C in a solution of 3 M NaCl and 100 mM sodium carbonate (pH 11). All reagents were of highest purity grade. Empirical measurements indicated that, on average, each bead carried on the order of 10^6 oligonucleotides. Following quality assessment, 16 of the bead types were discarded due to low signal-to-noise ratios. The remaining bead types were combined to create a pool containing 1520 functional bead types, each representing a unique capture sequence. The sequences were designed to serve as noninteracting decodable address sequences in addition to their function as probes that capture assay products from solution. They were selected to be 22 to 24 bases long with minimal cross-complementarity, similar GC content and Tm, no runs of a single base longer than five, and low similarity to human genomic sequences.

Preparation of Oligonucleotides and Pools Used in the Decode Process

Three sets of 1520 oligonucleotides complementary to the bead sequences were synthesized (Illumina Oligator synthesis). One set was unlabeled (corresponding to the OFF state), another with FAM, and a third with Cy3. The dye labels were incorporated during oligonucleotide synthesis by using 5′ phosphoramidites (Trilink Biotechnologies). After synthesis and deprotection, each oligonucleotide type was individually purified by using 96-well reverse-phase cartridge purification to yield labeled products >75% in purity. These sets of oligonucleotides were used to create eight unique decode pools, according to the strategy described below. Each pool contained 1520 decode sequences, each at a concentration of 10 nM.

Design of Decoder Pools

The strategy for the design of decoder pools to enable error checking, described in the main text, included the condition that all valid codes have exactly two OFF states, which allows for a maximum of 1792 codes. To enable the decoding of more bead types, we pooled the decoders so that valid codes have either exactly two OFF states or exactly five OFF states. The transition events that lead to misclassification are the same as described above, except that errors of the type (ON to OFF) or (OFF to ON) occurring in three different stages could also lead to a misclassification event. Three errors from ON to OFF could occur if a bead falls out of the array in the middle of decoding. To eliminate this source of error, all codes ending in three or more OFF states were removed from the set of valid codes. Finally, we eliminated any codes that do not have at least one ON state in each of the color channels that we use. The fact that each bead must show specific signal in two color channels during the decode process provides an additional quality check. The total number of valid color codes in this scheme is 2012.

Assembly of Array Matrices

Array matrices were manufactured in the following way: 96 optical fiber bundles were set in a rigid frame, in an 8 × 12 matrix that matches the layout of a standard 96-well microtiter plate. Each ~1.4-mm-diameter hexagonally packed bundle contains 49,777 individual glass optical fibers fused together into a monolithic unit. The bundles were polished at both ends to a tolerance of ~1 μm planarity across the entire matrix. One end served as the imaging surface and was used for the collection of fluorescence intensity data. The opposite end was etched in weak acid to create a well ~3 μm in diameter at the end of each optical fiber. The fiber bundles were loaded by pipetting ~0.4 μL (0.12 mg) beads in ethanol onto the end of each bundle, allowing the ethanol to evaporate, and removing excess beads. The assembled arrays were quite stable. Following eight stages of decoding and an overnight analytical hybridization (high salt at 48°C to 55°C), bead retention was typically >97%.

Decoding

Following bead assembly, arrays were hybridized to pools of 1520 decoders, each at 10 nM in decode buffer (600 mM NaCl, 60 mM potassium phosphate, 0.06% Tween-20, and 40% formamide at pH 7.6), for 12 min at room temperature. Following hybridization, three 1-min washes in wash buffer (167 mM NaCl, 16.7 mM potassium phosphate, 0.017% Tween-20 at pH 7.6) were used to remove unbound oligonucleotides. The arrays were then imaged at 1.0-μm resolution by using a 12-bit, 2000 × 3000-pixel CCD camera (Quantix 36E, Roper Scientific) and a standard achromatic 0.3 NA microscope imaging objective (field of view [FOV] = 2 mm) in a custom-engineered high-throughput imaging system (Barker et al. 2003). The imaging system uses an X, Y, Z stage assembly and positioning and autofocus software to collect images from all 96 arrays of a Sentrix array matrix automatically. Fluorescence excitation was performed by using a 300-W xenon arc lamp (ILC technology) and appropriate emission/dichroic/excitation filter sets. The FAM and Cy3 color channels were imaged separately. The light intensity at the sample was ~20 to 35 mW. This imaging design minimizes spectral cross-talk between dyes because excitation and emission are done on peak. After imaging, the arrays were dehybridized by dipping into 0.1 N NaOH for 1 min and were then neutralized in decode buffer. The process of hybridization, washing, image collection, and dehybridization was carried out until eight hybridization stages were completed. The arrays were washed in water and ethanol, dried in nitrogen, and sealed with desiccant in a foil package.

Image Processing and Data Extraction

A core corresponds to an etched well in the optical fiber bundle and has a high probability of containing a bead. We developed algorithms and custom image processing software to discover the locations of cores in decode images and extract intensity data from them (Galinsky 2003a,b). The software identified, aligned, and indexed the cores in a process termed registration. The shift, rotation, and scale of each image were determined relative to a template image as part of this process. After registration, the core indices and locations were matched across the set of decode images. Intensity information for each core was then obtained by averaging a three-by-three-pixel region centered on the core. The diameter of each core is approximately three microns, and each pixel is a one micron square. Many 1000s of decoding images are collected in a day in the production facility. Thus, an automated image analysis pipeline was created to facilitate high-throughput processing.

Alamar Exhibit 1002
Alamar v. Olink

Downloaded from www.genome.org on September 14, 2006

Decoding Algorithms

We describe two algorithms used to assign decode signatures to beads. The first is called stage-by-stage decoding. In this approach, an intensity histogram is computed for each color channel of each decode stage. The histogram is bimodal because each stage and color channel will have beads that are in the ON state and beads that are in the OFF state. An intensity value that separates the ON state from the OFF state is chosen, and all cores are assigned a zero or one, depending on whether their intensity value is less than or greater than the separation threshold. The values assigned for each stage and color channel are combined into the decode signature. This algorithm is highly effective when the OFF state population is well separated from the ON state population. Further, the method is insensitive to shifts caused by stage-to-stage variability in intensity.

Pseudocode for Stage–by–Stage Decoding

For each stage do
  For each color channel do

1. *Input the N core intensity values.*
2. *Compute intensity histogram using 100 bins.*
3. *Locate the bin with the maximum height peak within the first 25 bins, and the bin with the maximum height peak within the last 75 bins.*
4. *Find the bin with the minimum height peak between the two maxima.*
5. *Determine C, the median intensity of the minimum height peak.*
6. *Label all cores with intensity lower than C with zero, and all cores with intensity greater than C as one.*

Tabulate the zero and one stage and color information into the decode signatures.

The second algorithm is called core-by-core decoding. For each bead and color channel, we consider the eight intensity values across decode stages. The values are sorted, and the greatest relative intensity increase is determined. This is the separation between the ON and OFF states for the core. The same procedure is repeated for all beads in both color channels. The results are combined to give the decode signatures.

Pseudocode for Core–by–Core Decoding

For each core do
  For each color channel do

1. *Input the m intensity values for the core and color channel across stages.*
2. *Let $I_1 \ldots I_m$ be the sorted intensity values.*
3. *Let $J_k = (I_{k+1} - I_k)/I_k$, $k = 1 \ldots m - 1$ be the relative intensity jump from stage k to stage k + 1.*
4. *If the greatest relative jump occurs between stage K and K + 1, then assign zero to the core in stages 1 ... K, and assign one to the core in stages K + 1 ... m.*

Tabulate the zero and one core and color information into the decode signatures.

The two methods give virtually identical results. In practice, we use the core-by-core method to decode the arrays, and obtain quality control information from the histograms. As part of the decoding process, quantitative metrics for array quality are output automatically and can be stored in a database. All processing and quality metric generation takes ~20 min on a standard personal computer for a matrix of 96 arrays.

Estimation of Misclassification Rate

There are 1792 eight-stage color codes that have exactly two OFF states. A subset of these codes, (e.g., 1520) can be randomly selected and used in the design of decoder pools. The remaining 272 codes are unused. If a bead is misclassified (identified as the wrong bead type), the error cannot be observed because the number of beads of each type is random. However, because the used and unused color codes have the same form and are partitioned randomly, the number of beads that decode to unused color codes can be used to estimate the random misclassification rate. The estimate is given by $[B_{unused}/(B_{used} + B_{unused})] \times [(C_{used} - 1)/C_{unused}]$, where B_i is the number of beads decoding to category i and C_i is the total number of color codes in category i. In other words, the estimate is the relative number of beads decoding to unused codes scaled by the ratio of the number of codes in the used and unused categories. In the current example, $C_{used} = 1520$ and $C_{unused} = 272$, but the computation is analogous for any other decode space design.

ACKNOWLEDGMENTS

We are grateful to our Illumina colleagues in array manufacturing, process development, and engineering for invaluable technical assistance, and to Bob Kain and David Barker for numerous helpful discussions and insights. This work was supported in part by National Institutes of Health grants R44 HG02003-01, R21 HG01911, and R43 CA81952 to M.S.C.

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC section 1734 solely to indicate this fact.

REFERENCES

Alwine, J.C., Kemp, D.J., Parker, B.A., Reiser, J., Renart, J., Stark, G.R., and Wahl, G.M. 1979. Detection of specific RNAs or specific fragments of DNA by fractionation in gels and transfer to diazobenzyloxymethyl paper. *Methods Enzymol.* **68:** 220–242.

Barker, D.L., Therault, G., Che, D., Dickinson, T., Shen, R., and Kain, R. 2003. Self-assembled random arrays: High-performance imaging and genomics applications on a high-density microarray platform. *Proc. SPIE* **4966:** 1–11.

Battaglia, C., Salani, G., Consolandi, C., Bernardi, L.R., and De Bellis, G. 2000. Analysis of DNA microarrays by non-destructive fluorescent staining using SYBR green II. *Biotechniques* **29:** 78–81.

Braeckmans, K., De Smedt, S.C., Leblans, M., Pauwels, R., and Demeester, J. 2002. Encoding microcarriers: Present and future technologies. *Nat. Rev. Drug Discov.* **1:** 447–456.

Brenner, S., Johnson, M., Bridgham, J., Golda, G., Lloyd, D.H., Johnson, D., Luo, S., McCurdy, S., Foy, M., Ewan, M., et al. 2000. Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays. *Nat. Biotechnol.* **18:** 630–634.

Chan, W.C., Maxwell, D.J., Gao, X., Bailey, R.E., Han, M., and Nie, S. 2002. Luminescent quantum dots for multiplexed biological detection and imaging. *Curr. Opin. Biotechnol.* **13:** 40–46.

Chee, M., Yang, R., Hubbell, E., Berno, A., Huang, X.C., Stern, D., Winkler, J., Lockhart, D.J., Morris, M.S., and Fodor, S.P. 1996. Accessing genetic information with high-density DNA arrays. *Science* **274:** 610–614.

Drmanac, S., Stavropoulos, N.A., Labat, I., Vonau, J., Hauser, B., Soares, M.B., and Drmanac, R. 1996. Gene-representing cDNA clusters defined by hybridization of 57,419 clones from infant brain libraries with short oligonucleotide probes. *Genomics* **37:** 29–40.

Drmanac, S., Kita, D., Labat, I., Hauser, B., Schmidt, C., Burczak, J.D., and Drmanac, R. 1998. Accurate sequencing by hybridization for DNA diagnostics and individual genomics. *Nat. Biotechnol.* **16:** 54–58.

Fan, J.-B., Oliphant, A., Shen, R., Kermani, B.G., Garcia, F., Gunderson, K.L., Hansen, M., Steemers, F., Butler, S.L., Deloukas, P., et al. 2003. Highly parallel SNP genotyping. *Cold Spring Harbor Symp. Biol.* **68:** (in press).

Fan, J-B., Yeakley, J.M., Bibikova, M., Chudin, E., Wickham, E., Chen, J., Doucet, D., Rigault, P., Zhang, B., Shen, R., et al. 2004. A versatile assay for high-throughput gene expression profiling on universal array matrices. *Genome Res.* (this issue).

Fodor, S.P.A., Read, J.L., Pirrung, M.C., Stryer, L., Lu, A.T., and Solas, D. 1991. Light-directed, spatially addressable parallel chemical synthesis. *Science* **251:** 767–773.

Fulton, R.J., McDade, R.L., Smith, P.L., Kienker, L.J., and Kettman Jr., J.R. 1997. Advanced multiplexed analysis with the FlowMetrix system. *Clin. Chem.* **43:** 1749–1756.

Galinsky, V.L. 2003a. Automatic registration of microarray images, I: Rectangular grid. *Bioinformatics* **19:** 1824–1831.

———. 2003b. Automatic registration of microarray images, II: Hexagonal grid. *Bioinformatics* **19:** 1832–1836.

Gerry, N.P., Witowski, N.E., Day, J., Hammer, R.P., Barany, G., and

Alamar Exhibit 1002
Alamar v. Olink

Downloaded from www.genome.org on September 14, 2006

Barany, F. 1999. Universal DNA microarray method for multiplex detection of low abundance point mutations. *J. Mol. Biol.* **292:** 251–262.
Gunderson, K.L., Huang, X.C., Morris, M.S., Lipshutz, R.J., Lockhart, D.J., and Chee, M.S. 1998. Mutation detection by ligation to complete *n*-mer DNA arrays. *Genome Res.* **8:** 1142–1153.
Hamming, R.W. 1986. *Coding and information theory.* Prentice-Hall, Inc., Englewood Cliffs, NJ.
Han, M., Gao, X., Su, J.Z., and Nie, S. 2001. Quantum-dot-tagged microbeads for multiplexed optical coding of biomolecules. *Nat. Biotechnol.* **19:** 631–635.
Hardenbol, P., Baner, J., Jain, M., Nilsson, M., Namsaraev, E.A., Karlin-Neumann, G.A., Fakhrai-Rad, H., Ronaghi, M., Willis, T.D., Landegren, U., et al. 2003. Multiplexed genotyping with sequence-tagged molecular inversion probes. *Nat. Biotechnol.* **21:** 673–678.
Hessner, M.J., Wang, X., Khan, S., Meyer, L., Schlicht, M., Tackes, J., Datta, M.W., Jacob, H.J., and Ghosh, S. 2003. Use of a three-color cDNA microarray platform to measure and control support-bound probe for improved data quality and reproducibility. *Nucleic Acids Res.* **31:** e60.
Holloway, A.J., van Laar, R.K., Tothill, R.W., and Bowtell, D.D. 2002. Options available—from start to finish—for obtaining data from DNA microarrays II. *Nat. Genet.* **32(Suppl):** 481–489.
Hubbell, E. and Pevzner, P.A. 1999. Fidelity probes for DNA arrays. *Proc. Int. Conf. Intell. Syst. Mol. Biol.* 113–117.
Johnson, P.H., Walker, R.P., Jones, S.W., Stephens, K., Meurer, J., Zajchowski, D.A., Luke, M.M., Eeckman, F., Tan, Y., Wong, L., et al. 2002. Multiplex gene expression analysis for high-throughput drug discovery: Screening and analysis of compounds affecting genes overexpressed in cancer cells. *Mol. Cancer Ther.* **1:** 1293–1304.
Kennedy, G.C., Matsuzaki, H., Dong, S., Liu, W.M., Huang, J., Liu, G., Su, X., Cao, M., Chen, W., Zhang, J., et al. 2003. Large-scale genotyping of complex DNA. *Nat. Biotechnol.* **21:** 1233–1237.
Levsky, J.M., Shenoy, S.M., Pezo, R.C., and Singer, R.H. 2002. Single-cell gene expression profiling. *Science* **297:** 836–840.
Liang, P. and Pardee, A.B. 1992. Differential display of eukaryotic messenger RNA by means of the polymerase chain reaction. *Science* **257:** 967–971.
Lockhart, D.J. and Trulson, M.O. 2001. Multiplex metallica. *Nat. Biotechnol.* **19:** 1122–1123.
Lockhart, D.J. and Winzeler, E.A. 2000. Genomics, gene expression and DNA arrays. *Nature* **405:** 827–836.
Lockhart, D.J., Dong, H., Byrne, M.C., Follettie, M.T., Gallo, M.V., Chee, M.S., Mittmann, M., Wang, C., Kobayashi, M., Horton, H., et al. 1996. Expression monitoring by hybridization to high-density oligonucleotide arrays. *Nat. Biotechnol.* **14:** 1675–1680.
Michael, K.L., Taylor, L.C., Schultz, S.L., and Walt, D.R. 1998. Randomly ordered addressable high-density optical sensor arrays. *Anal. Chem.* **70:** 1242–1248.
Nicewarner-Pena, S.R., Freeman, R.G., Reiss, B.D., He, L., Pena, D.J., Walton, I.D., Cromer, R., Keating, C.D., and Natan, M.J. 2001. Submicrometer metallic barcodes. *Science* **294:** 137–141.
Nuwaysir, E.F., Huang, W., Albert, T.J., Singh, J., Nuwaysir, K., Pitas, A., Richmond, T., Gorski, T., Berg, J.P., Ballin, J., et al. 2002. Gene expression analysis using oligonucleotide arrays produced by maskless photolithography. *Genome Res.* **12:** 1749–1755.
Pastinen, T., Raitio, M., Lindroos, K., Tainola, P., Peltonen, L., and Syvanen, A.C. 2000. A system for specific, high-throughput genotyping by allele-specific primer extension on microarrays. *Genome Res.* **10:** 1031–1042.
Patil, N., Berno, A.J., Hinds, D.A., Barrett, W.A., Doshi, J.M., Hacker, C.R., Kautzer, C.R., Lee, D.H., Marjoribanks, C., McDonough, D.P., et al. 2001. Blocks of limited haplotype diversity revealed by high-resolution scanning of human chromosome 21. *Science* **294:** 1719–1723.
Pease, A.C., Solas, D., Sullivan, E.J., Cronin, M.T., Holmes, C.P., and Fodor, S.P. 1994. Light-generated oligonucleotide arrays for rapid DNA sequence analysis. *Proc. Natl. Acad. Sci.* **91:** 5022–5026.
Sambrook, J.E.A. 1989. *Molecular cloning: A laboratory manual.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
Schena, M., Shalon, D., Davis, R.W., and Brown, P.O. 1995. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* **270:** 467–470.
Sengupta, R. and Tompa, M. 2002. Quality control in manufacturing oligo arrays: A combinatorial design approach. *J. Comput. Biol.* **9:** 1–22.
Shannon, C.E. 1948a. A mathematical theory of communication. *Bell System Technical J.* **27:** 379–423.
———. 1948b. A mathematical theory of communication. *Bell System Technical J.* **27:** 623–656.
Shearstone, J.R., Allaire, N.E., Getman, M.E., and Perrin, S. 2002. Nondestructive quality control for microarray production. *Biotechniques* **32:** 1051–1052, 1054, 1056–1057.
Southern, E., Maskos, U., and Elder, R. 1992. Analyzing and comparing nucleic acid sequences by hybridization to arrays of oligonucleotides: Evaluation using experimental models. *Genomics* **13:** 1008–1017.
Southern, E.M. 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J. Mol. Biol.* **98:** 503–517.
Taylor, E., Cogdell, D., Coombes, K., Hu, L., Ramdas, L., Tabor, A., Hamilton, S., and Zhang, W. 2001. Sequence verification as quality-control step for production of cDNA microarrays. *Biotechniques* **31:** 62–65.
Van't Veer, L.J., Dai, H., van de Vijver, M.J., He, Y.D., Hart, A.A., Mao, M., Peterse, H.L., van der Kooy, K., Marton, M.J., Witteveen, A.T., et al. 2002. Gene expression profiling predicts clinical outcome of breast cancer. *Nature* **415:** 530–536.
Wang, D.G., Fan, J.-B., Siao, C.-J., Berno, A., Young, P., Sapolsky, R., Ghandour, G., Perkins, N., Winchester, E., Spencer, J., et al. 1998. Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. *Science* **280:** 1077–1082.
Watson, J.D. and Crick, F.H.C. 1953. Molecular structure of nucleic acid: A structure for deoxyribose nucleic acid. *Nature* **171:** 737–738.
Yeakley, J.M., Fan, J.B., Doucet, D., Luo, L., Wickham, E., Ye, Z., Chee, M.S., and Fu, X.D. 2002. Profiling alternative splicing on fiber-optic arrays. *Nat. Biotechnol.* **20:** 353–358.
Yvert, G., Brem, R.B., Whittle, J., Akey, J.M., Foss, E., Smith, E.N., Mackelprang, R., and Kruglyak, L. 2003. Trans-acting regulatory variation in *Saccharomyces cerevisiae* and the role of transcription factors. *Nat. Genet.* **35:** 57–64.

WEB SITE REFERENCES

www.hapmap.org; International HapMap Project.

Received December 8, 2003; accepted in revised form January 29, 2004.

This Page is Inserted by IFW Indexing and Scanning Operations and is not part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☐ BLACK BORDERS
- ☐ IMAGE CUT OFF AT TOP, BOTTOM OR SIDES
- ☑ FADED TEXT OR DRAWING
- ☐ BLURRED OR ILLEGIBLE TEXT OR DRAWING
- ☐ SKEWED/SLANTED IMAGES
- ☐ COLOR OR BLACK AND WHITE PHOTOGRAPHS
- ☐ GRAY SCALE DOCUMENTS
- ☐ LINES OR MARKS ON ORIGINAL DOCUMENT
- ☐ REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY
- ☐ OTHER: _____

IMAGES ARE BEST AVAILABLE COPY.
As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.

Alamar Exhibit 1002
Alamar v. Olink

Using protein-DNA chimeras to detect and count small numbers of molecules

Ian Burbulis[1], Kumiko Yamaguchi[1], Andrew Gordon[1], Robert Carlson[2] & Roger Brent[1]

© 2005 Nature Publishing Group http://www.nature.com/naturemethods

We describe general methods to detect and quantify small numbers of specific molecules. We redirected self-splicing protein inteins to create 'tadpoles', chimeric molecules comprised of a protein head covalently coupled to an oligonucleotide tail. We made different classes of tadpoles that bind specific targets, including *Bacillus anthracis* protective antigen and the enzyme cofactor biotin. We measured the amount of bound target by quantifying DNA tails by T7 RNA polymerase runoff transcription and real-time polymerase chain reaction (PCR) evaluated by rigorous statistical methods. These assays had a dynamic range of detection of more than 11 orders of magnitude and distinguished numbers of molecules that differed by as little as 10%. At their low limit, these assays were used to detect as few as 6,400 protective antigen molecules, 600 biotin molecules and 150 biotinylated protein molecules. In crudely fractionated human serum, the assays were used to detect as few as 32,000 protective antigen molecules. Tadpoles thus enable sensitive detection and precise quantification of molecules other than DNA and RNA.

Numerous questions in biology depend on detection and quantification of very small numbers of specific molecules mixed among large numbers of other molecules. For example, understanding how genetically identical cells behave differently will require quantification of differences in the amounts and modification states of small amounts of regulatory proteins. Similarly, early detection of many diseases from clinical samples will be facilitated by the ability to detect and quantify small numbers of signature molecules. The mainstay methods for quantification of biological molecules other than nucleic acids remain enzyme-linked immunoabsorption assays[1] (ELISAs) and their predecessors, radioimmune assays[2] (RIAs), developed in 1971 and 1959, respectively. One limitation to the sensitivity and accuracy of these assays has been the means used to quantify the target bound by the antibody.

One approach to target quantification that promises great sensitivity has been to bring specific DNA sequences to the target and then detect and count the bound sequences by nucleic acid hybridization or amplification approaches. The idea of detecting target binding by DNA amplification dates from 1992 (ref. 3), as does the use of DNA tags to detect small numbers of target bound antibodies[4]. These experiments demonstrated the application of the extraordinary sensitivity of PCR detection to non-nucleic acid molecules. Since these initial developments, other investigators have described detection of DNA–linked antibodies bound to targets by PCR[5–8], T7 runoff transcription[9] and hybridization[10]. For a number of reasons, including in many cases the molecular heterogeneity of the antibodies and the lack of precise chemical control of the site(s) of DNA attachment, assays based on DNA–linked antibodies have not gained wide use. We envisioned that the ability to effect well-controlled synthesis of defined protein-DNA conjugates in large quantities might enable wider development of target-binding and amplification measurement methods.

Here, we devised the means to covalently link chemically homogeneous affinity proteins and DNA molecules by controlled site-specific chemistry. The chemistry is catalyzed by inteins, protein domains that excise from polypeptides and join the remaining amino and carboxyl moieties to form a mature protein[11]. After excision, the amino terminal moiety acquires a single thioester residue at its C terminus. This thioester can couple with synthetic peptides in *trans*[12]. It has been used to couple single-chain antibodies to fluorescent labels[13] and has recently been shown to join proteins to cysteine-conjugated polyamide nucleic acids[14].

We used a collection of modified inteins to generate thioester-tagged recombinant antibodies as well as other affinity proteins and coupled them to synthetic oligonucleotides. The resulting chimeric tadpole molecules were composed of a protein head and a DNA tail. We adapted absorption and crosslinking, PCR, T7 amplification and statistical methods to use tadpoles to detect and count small numbers of different molecules, such as biotin and native proteins, including the 'protective antigen' (PA) subunit of the *B. anthracis* toxin. These methods enabled detection and quantification of non-nucleic acid molecules with great precision, great dynamic range and, in some cases, sensitivity near the limits of PCR.

RESULTS

Tadpole synthesis

To explore the diversity of protein structures suitable for the tadpole heads, we directed the synthesis of three classes of affinity protein moieties in *Escherichia coli*: streptavidin, a protein that specifically binds the vitamin biotin[15], single-chain antibodies

[1]The Molecular Sciences Institute, 2168 Shattuck Avenue, Berkeley, California 94704, USA. [2]Present address: University of Washington, Department of Electrical Engineering, Box 352500, Seattle, Washington 98195, USA. Correspondence should be addressed to I.B. (burbulis@molsci.org).

PUBLISHED ONLINE 21 DECEMBER 2004; DOI:10.1038/NMETH729

Alamar Exhibit 1002
Alamar v. Olink

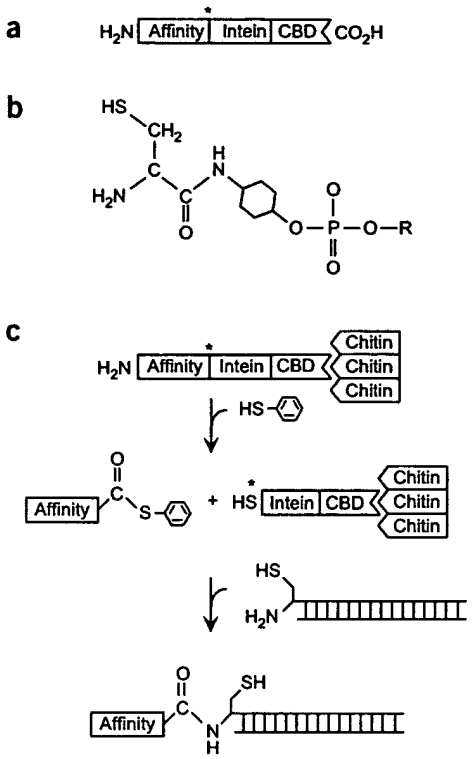


Figure 1 | Design and synthesis of chimeric detector molecules. (**a**) An illustration of the typical three-part fusion proteins containing, from N to C terminus, an affinity domain with 6-His sequence, an intein and the CBD, except for the streptavidin fusion, which lacked the 6-His tag. An asterisk denotes the active-site cysteine. (**b**) Cysteine-like residue incorporated at the 5′ terminus of the forward strand of the DNA tail (represented by 'R') that serves to covalently couple with the reactive thioester on the affinity protein. (**c**) Thioester-dependent ligation of protein to DNA. Chitin-purified recombinant intein fusion proteins are reacted with 0.2% thiophenol, 1 mM TCEP and modified DNA *in situ*. Thiophenol helps promote the hydrolysis of the peptide backbone to yield a reactive thioester at the C terminus of the affinity protein. Next, the thiol moiety on the synthetic DNA tail displaces the thiophenol[12], which then undergoes an N-S acyl shift[11] resulting in an amide linkage between protein and DNA.

© 2005 Nature Publishing Group http://www.nature.com/naturemethods

containing disulfide linkages[16] and thioredoxin (TrxA) peptide aptamers[17] (**Supplementary Table 1** online). We expressed these affinity moieties as fusion proteins that comprised, from N to C terminus, an affinity protein, an intein and the 52-residue chitin binding domain (CBD) from *Bacillus circulans* WL-12 Chitinase A1[18] (**Fig. 1a**). For the DNA components, we synthesized double-stranded (ds) oligonucleotides containing a cysteine-like residue[19] at one 5′ end (**Fig. 1b** and **Supplementary Table 2** online). We purified the three-part fusion proteins on chitin columns[20] and activated intein excision and DNA coupling by filling the columns with buffer containing the reducing agent, thiophenol and cysteine-modified DNA (**Fig. 1c**). In the subsequent reaction, the thiophenol allows the intein to hydrolyze the peptide backbone, excising the affinity protein from the fusion and placing a reactive thioester at its C terminus. The thiol moiety on the cysteine in the synthetic DNA tail displaces the thiophenol[12] and undergoes an N-S acyl shift[11], resulting in an amide bond between the protein and DNA. We allowed the cleavage-coupling reactions to run overnight and purified the protein-DNA chimeras from the column eluate. The resulting tadpole molecules have an affinity protein head covalently linked by an amide to a DNA tail (**Fig. 1c**).

Tadpole characterization

The first tadpoles we created contained the 'core' domain of streptavidin[21] linked to a 149-mer DNA tail. We purified them by anion-exchange and hydrophobic-interaction chromatography to apparent homogeneity (effectively >95% purity) as judged by the presence of a single SYBR gold (Molecular Probes) stained band on native acrylamide gels. A number of observations were consistent with the idea that these molecules comprised a single affinity protein linked to a single DNA moiety (**Fig. 2**). On native gels stained for DNA, digestion with protease K caused a shift in the mobility from that expected for a single protein-DNA chimera to that of the free DNA tail. On size exclusion columns, the tadpoles' apparent molecular weight was as expected for a molecule containing a single protein and a single DNA moiety. As assayed from the change in mobility on native gels in the presence of biotin (and by ELISAs, data not shown), both the anti-biotin tadpole and the free protein head recognized biotin, which indicates the attachment of DNA did not interfere with the binding activity of the protein domain. Parental fusion proteins not treated with thiophenol or Tris(2-carboxyethyl)phosphine hydrochloride (TCEP) did not form detectable amounts of this species, consistent with the idea that linkage occurs at the C-terminal thioester resulting from the intein reaction (**Fig. 2a**).

Biotin quantification

We used these tadpoles to detect and count small numbers of biotin molecules *in vitro*. We determined the biotin content of biotinylated bovine serum albumin (BSA; Pierce) and carbonic anhydrase preparations (data not shown) using 2-(4′-hydroxyazobenzene)-benzoic acid (HABA) in spectrophotometric assays[22] as described in Methods. We diluted biotinylated BSA into mixtures that contained an excess of BSA coupled to two other dyes into wells on a standard ELISA dish. We blocked the wells with casein and calf thymus DNA and added to them anti-biotin tadpoles. We washed the wells and quantified tadpoles that remained bound by loading the wells with T7 polymerase and ribonucleotide triphosphates and allowing run-off transcripts to accumulate[23]. We quantified the amount of RNA by measuring the fluorescence intensity of a stained band on denaturing polyacrylamide gels. The signals for biotinylated BSA were linear over six orders of magnitude, whereas controls containing only NDA- and Texas Red–coupled proteins gave no signal above background (**Fig. 3a**). In parallel experiments, we quantified biotin immobilized on carbonic anhydrase with equivalent sensitivity as that for biotinylated BSA (data not shown), indicating that the measurement of biotin was independent of the protein substrate. These results show that a tadpole can detect and quantify small numbers of an organic molecule against a background of very large numbers of molecules of similar mass.

To determine the smallest number of biotin molecules detectable using these tadpoles, we adapted real-time PCR[24]. We serially diluted biotinylated BSA into BSA, covalently coupled proteins in the dilution mixtures to epoxy-modified magnetic beads[25,26] (Dynal Biotech) and blocked the beads as above. We incubated the beads with tadpoles, washed the beads to remove non-bound

Alamar Exhibit 1002
Alamar v. Olink

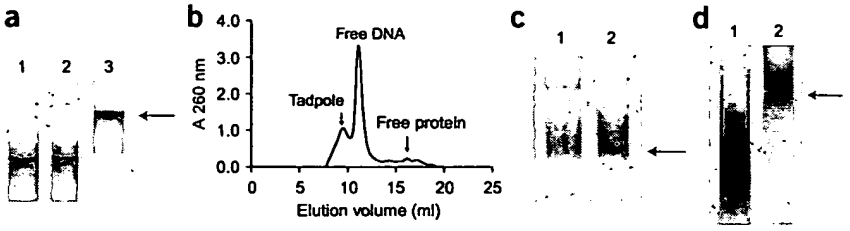


Figure 2 | Characterization of anti-biotin tadpole. (a) To monitor the coupling of DNA to the protein head we resolved sample before and after conjugation on a native 12% Tris-borate-EDTA polyacrylamide gel. Lane 1 is the DNA tail alone, lane 2 is DNA tail added to the streptavidin-intein fusion protein without thiophenol activation and lane 3 is the DNA tail added to the intein fusion protein in the presence of thiophenol and TCEP. (b) Superdex-200 size exclusion chromatography of the streptavidin tadpole with absorbance monitored at 260 nm. The arrows indicate the conjugated tadpole, free DNA and free protein in the elution. The mobility of the streptavidin tadpole is consistent with one DNA tail attached to a single protein. (c) Confirmation that the streptavidin tadpole is composed of protein and DNA. Samples were digested with proteinase K and resolved on a gel as above. Lane 1 is mock-treated tadpole and lane 2 is tadpole digested with proteinase K. The arrow indicates the increase in electrophoretic mobility from a low-mobility tadpole conjugate to a size consistent with a free DNA tail. (d) Confirmation that the streptavidin tadpole bind biotinylated BSA *in vitro*. Lane 1 is the tadpole alone and lane 2 is the tadpole in the presence of biotinylated BSA. The arrow indicates the reduced electrophoretic mobility in the presence of biotinylated BSA as compared to tadpole alone, and this is consistent with the formation of a complex between the tadpole and the biotinylated BSA.

© 2005 Nature Publishing Group http://www.nature.com/naturemethods

species and counted the number of target-tadpole complexes by including the beads in real-time PCRs (**Supplementary Table 3** online). We quantified bound tadpoles by comparing the reaction cycle at which the accumulated fluorescent signal surpassed an arbitrary threshold to a calibration curve generated from known numbers of tadpoles (**Fig. 3b**).

We found that standard linear regression methods typically used to determine numbers of template molecules in real-time PCR experiments were inadequate to determine the actual limit of detection for these tadpole-based assays. The small amount of tadpole binding that occurred on the beads in the absence of target ligand resulted in a non-linear signal output at the transition between background binding to the beads and specific binding to the ligand. At these concentrations, the use of linear regression resulted in less precise measurements and frequently exaggerated the sensitivity of the measurement.

To better quantify the data, we constructed a 'confidence belt' for the measurements according to published methods[27,28]. We defined the limits (area) of our confidence belt as two standard deviations of the PCR cycle measurement (**Fig. 4**). For this work and for other applications of this method, we assumed that when testing an unknown sample, researchers would perform measurements in triplicate. Given the boundary limits of our confidence belt, we evaluated the estimated number of target molecules present in a sample that will result in a mean number of PCR cycles falling within the specified limits in 95 out of 100 measurements. As long as measurement results are distributed normally (**Supplementary Fig. 1** online), this approach is valid across the entire dynamic range of the assay, including the lowest values, as well as the linear range.

Using this technique to evaluate the PCR data, we detected as few as 600 biotin molecules (or 150 biotinylated BSA molecules) with >95% confidence (**Fig. 4**). In our hands, the detection limit for this tadpole-based assay was 10^9-fold lower than for a conventional ELISA (**Supplementary Fig. 2** online); by combining PCR and T7 polymerase data, the dynamic range of these measurements spanned 11 orders of magnitude. In the linear range, the smallest difference between the numbers of molecules these assays could distinguish was consistently below 27%. The resolution of these measurements is relatively high; for example, these methods can distinguish between 50,000 and 64,000 molecules with >97% confidence.

Synthesis of tadpoles to quantify native proteins

We then made tadpoles to quantify native proteins. First, we made tadpoles that contained TrxA peptide aptamers that bound human E2F1[29] and CDK2[17]. In bead-based real-time PCR assays performed as above, these tadpoles allowed quantification of

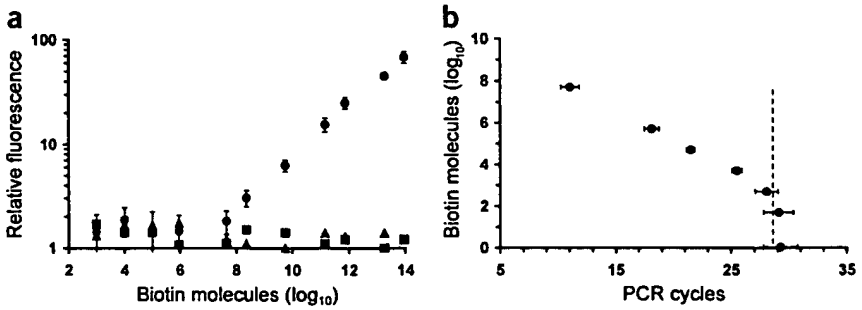


Figure 3 | Quantification of biotinylated BSA diluted in an excess of nonspecific organic molecules. (a) The fluorescence signal of SYBR green-stained RNA, is graphed as a function biotin quantity. The circles, triangles and squares represent biotinylated, NDA-labeled and Texas Red-labeled BSA, respectively. Error bars represent 2 s.d. The measurements were performed in triplicate. (b) Quantification of biotin using real-time PCR. The *x*-axis shows the PCR cycle at which the change in fluorescence surpassed a designated threshold. The *y*-axis shows the number of molecules of target-bound tadpole tails determined by reference to a calibration curve using known numbers of free tadpoles. Circles represent the mean of five replicate measurements and the error bars represent 2 s.d. The dashed line indicates the lower limit of detection for this assay: the smallest number of molecules from a single triplicate measurement that could be distinguished from zero with >97% confidence.

Alamar Exhibit 1002
Alamar v. Olink

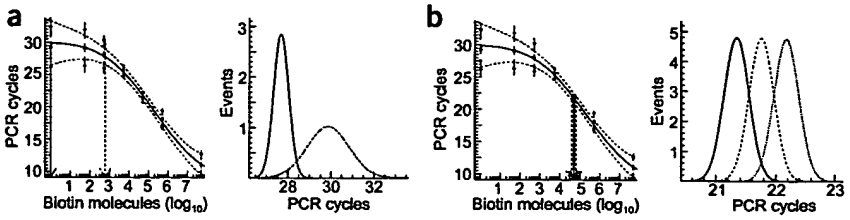


Figure 4 | Statistical analysis of biotin quantification. **(a)** Limit of detection of biotin measurement. Shown in the left graph is the plot of data with the mean PCR cycles/*N* biotin molecules as the solid line and the confidence belt, corresponding to 2 s.d., as dashed lines. These dashed lines define the confidence belt we used to calculate the limit of detection and precision of measurement. The solid arrow extending from the curve to the *x*-axis represents the signal measured in an assay with zero biotin molecules. The dashed arrow extending from the curve to the *x*-axis represents the intercept corresponding to the limit of detection illustrated in the right plot. The right plot illustrates the PCR cycle distribution, corresponding to the arrow in the left graph, representing a mean value (solid line) that is distinguished from the zero distribution (dashed line), given our confidence belt. **(b)** Precision of biotin measurement in the linear range of our assay. The left graph is the same as in **a**, except that the arrow corresponds to the intercept where the distributions in the right graph were evaluated. The right plot illustrates the PCR cycle distributions in the linear range of the assay where the standard deviation is low.

© 2005 Nature Publishing Group http://www.nature.com/naturemethods

recombinant E2F1 and CDK2 over a 10^6-fold range, but presumably due to the relatively weak K_d (~1 μM) of the TrxA aptamers, the limit of detection in these experiments was only marginally lower than in ELISAs using existing antibodies (data not shown). Because of these findings, we chose to explore the use of single-chain variable fragments[16] (scFvs) from mammalian antibodies as affinity reagents. We made tadpoles to quantify PA, a protein subunit of anthrax toxin[30]. The tadpole heads were scFvs that bound PA with subnanomolar affinity[31], and the tails were 82-nucleotide dsDNAs (**Supplementary Table 2** online). The use of scFvs as tadpole heads presented substantial challenges (see Methods) because scFvs require disulfide bonds to form correctly, but the intein-mediated protein-DNA ligation requires reducing chemistry.

We expressed and purified soluble, full-length three-part fusion protein in the cytosol of *E. coli* Origami B (DE3), which carries lesions in *trxB, gor* and an unspecified suppressor mutation[32]. For this protein, we used an intein from *Mycobacterium xenopi gyrA* chosen after pilot experiments revealed that it gave efficient fusion protein hydrolysis even for the large scFv moiety. We treated the purified fusion with thiophenol, TCEP and the 82-mer to catalyze the excision-ligation reaction. We eluted the tadpole from the column and allowed the scFv disulfide bonds to reform by slowly dialyzing away the TCEP. We characterized the anti-PA tadpole by the methods used to evaluate the anti-biotin tadpole and measured its affinity for PA by evanescent wave experiments. The whole tadpole bound PA with a K_d of 4.5 nM, and the free scFv tadpole head bound with a K_d of 0.46 nM, a good match to the published affinity[31]. Reduced binding by the complete tadpole compared to the free scFv was largely due to slower association with the target (**Supplementary Table 4** online), possibly caused by the presence of the DNA tail; consistent with this idea, the off rate for the complete tadpole was also twofold slower.

Quantification of PA protein

We diluted protective antigen and coupled dilution mixtures to carboxylate-modified magnetic beads, as above. We blocked beads, probed with tadpoles for one hour, washed and quantified bound tadpoles by real-time PCR, as above (**Fig. 5a**). These measurements had a 10^9-fold dynamic range and a 10^6-fold linear range. We evaluated the PCR data for these PA assays as we did for the biotin measurements, using a confidence-belt[27,28] consisting of two standard deviations. At the low limit, the smallest number of molecules that could be distinguished with confidence from zero was 6,400 (**Fig. 6a**). In the linear range, the smallest difference in numbers of molecules these assays distinguished was about 10%, enabling for example 5.0×10^8 molecules to be distinguished from 5.5×10^8 molecules with >95% confidence (**Fig. 6b**). In our hands, this tadpole-based assay was thus considerably (10^9-fold) more sensitive than conventional ELISAs using goat anti-PA IgG (List Biologicals) or purified anti-PA scFv (**Supplementary Fig. 3** online).

Quantification of PA from human serum

We quantified PA in human serum. Unfractionated serum coupled to carboxylate beads bound tadpoles nonspecifically, regardless of the specificity of the protein head (data not shown). To circumvent this effect, we diluted PA into human serum and crudely fractionated it by running each dilution mixture on a 2-ml hydrophobic resin column (T-gel; Pierce). We eluted bound protein at low salt, coupled 2 μg of it to beads, blocked the beads, probed with tadpole,

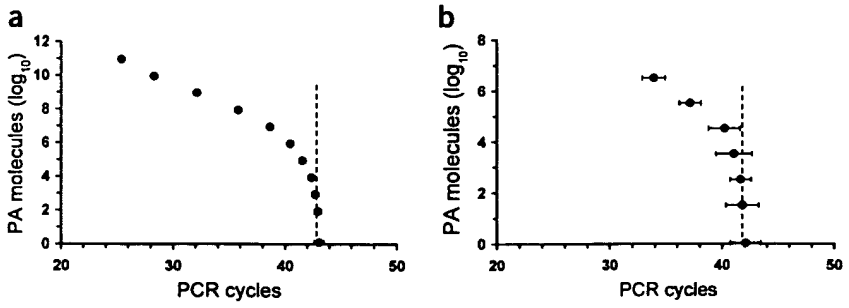


Figure 5 | Quantification of recombinant PA diluted in BSA or human serum. **(a)** Dilution in BSA. Circles represent the mean of ten separate measurements and the error bars represent 2 s.d. The *x*- and *y*-axes indicate the PCR cycles and numbers of PA molecules respectively, and the dashed line indicates the limit of detection, as in **Figure 3**. **(b)** Dilution in human serum. The *x*- and *y*-axes indicate the PCR cycles and numbers of PA molecules respectively, and the dashed line indicates the limit of detection. The circles represent the mean of eight separate measurements and the error bars represent 2 s.d.

Alamar Exhibit 1002
Alamar v. Olink

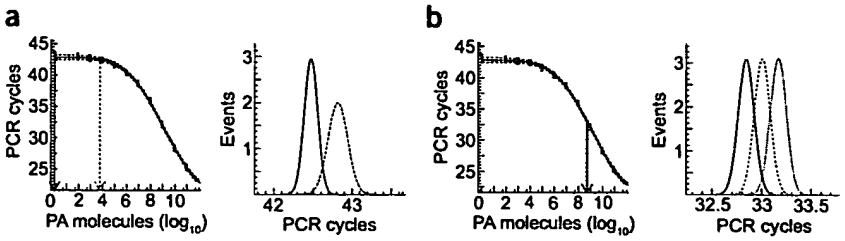


Figure 6 | Statistical analysis of PA quantification. (a) Limit of detection of PA measurement. The left plot shows the measurement data and error boundaries corresponding to 2 s.d. The dashed arrow represents the intercept corresponding to the limit of detection, as depicted in the right graph. The right plot illustrates the PCR cycle distribution at the limit of detection (solid line) compared to the distribution corresponding to zero molecules (dashed line). (b) Precision of PA measurement in the linear range. The left plot is the same as in a except that the arrow represents the intercept where the distributions in the right graph are evaluated. The right plot illustrates calculated PCR cycle distributions in the linear range of the assay where the standard deviation is low.

© 2005 Nature Publishing Group http://www.nature.com/naturemethods

washed and quantified PA-bound tadpole as above. From serum, the precision of the measurement was lower than that in buffered BSA (the smallest difference in the numbers of molecules distinguished ranged from 55% to 98%), and the limit of detection corresponded to about 32,000 molecules (**Figs. 5b**). Tadpole-based detection from serum was thus about 10 times less sensitive than tadpole based detection in buffered BSA but about 200 times more sensitive than a corresponding ELISA of PA in serum using goat anti-PA (data not shown).

DISCUSSION

The intein-mediated chemistry used to assemble tadpoles from molecularly homeogeneous synthetic DNA and recombinant protein subunits is exceedingly efficient, precise and well controlled. In our hands, the chemistry allowed generation of milligram quantities. The resulting molecules were molecularly homogeneous: they had a precise (1:1) protein DNA stoichiometry with and the DNA was attached to the protein at a site chosen to not interfere with protein function. The molecules were made from a recombinant protein and synthetic DNA, allowing, as opposed to antibodies made in animals, new affinity proteins to be selected quickly *in vitro*, fusion proteins containing these affinity moieties to be expressed and purified rapidly and synthesis of the tadpoles to be scaled up rapidly. The resulting tadpoles are chemically stable for long periods at −20 °C.

We presented methods to measure the amounts of non-nucleic acid target molecules present in samples with a sensitivity approaching that for detecting nucleic acid targets by PCR. We adapted a statistical treatment[27,28], normally used to analyze low-incidence events, to evaluate real-time PCR data. This statistical treatment enabled us to distinguish the mean of a measured distribution from the mean of other measured distributions or of distribution from samples containing zero target molecules. This statistical treatment is valid throughout the linear and non-linear ranges of the data curve. Use of calibration standards internal to each PCR may further improve measurement precision, and the use of second–affinity reagent capture methods[33] ('sandwich methods') and methods dependent on binding by two different tadpoles (such as proximity-dependent ligation assays[5]) should further improve sensitivity, perhaps allowing detection and quantification of extremely small amounts of molecules from crude mixtures.

The antibodies now used in ELISAs are typically generated by the immune systems of vertebrate animals. These antibodies recognize limited sets of protein epitopes. But tadpole heads can be chosen from the entire range of possible affinity protein scaffolds, and affinity molecules based on those scaffolds can then be selected from combinatorial libraries *in vitro*. It should therefore be possible to make tadpoles against molecules including proteins, carbohydrates, fatty acids and organic chemicals that the vertebrate immune system does not normally recognize.

The assays are accurate and fast, requiring as little as three hours to measure the number of molecules present. The materials and skills needed to synthesize tadpoles are widely available, and the infrastructure, including real-time PCR machines for quantitative measurements, is widely deployed. Thus, these reagents might be applied to problems in biology, medicine, environmental monitoring and industry.

We also note that a reaction mix consisting of target, tadpole and polymerase can be considered a 'molecular amplifier'; viewed in this manner, tadpole-containing 'devices' may have utility beyond detection. Moreover, hybridized by their DNA tails, or oligomerized by their protein heads, tadpoles may also be useful in assembly and self-assembly of other nanoscale structures of defined geometry and function.

METHODS

Fusion-protein expression plasmids. Plasmids directed the synthesis of three-part fusion proteins that contained, from N to C terminus, an affinity domain with 6-His sequence, an intein and the CBD, except for the streptavidin fusion, which lacked the 6-His-tag. PCR products encoding affinity proteins were cloned into *Eco*RI and *Hin*dIII sites in the parent vectors pTYB1, pTWIN1 or pTWIN2[20]. The plasmids created, inteins used, *E. coli* expression host strains and DNA tail sequences are listed in **Supplementary Table 1** online. In initial attempts to express the scFv fusion proteins under oxidizing conditions, we targeted fusion proteins containing two model scFvs, 1H and 14B7[31], for export to the periplasm. Under these expression conditions, the proteins were hydrolyzed at the affinity protein-intein junction, even after changing residues at the hydrolysis junction[34] and co-expression of the chaperonin Skp1p[35–37] (data not shown).

DNA synthesis. We synthesized oligonucleotides on a controlled-pore glass (CPG) matrix with 2,500 Å pore size (Glen Research) using conventional phosphoramidite-based chemistries[38,39]. We incorporated a cysteine-like residue into the 5′ terminus of the forward strand of each dsDNA tail during synthesis. After synthesis, we treated the oligonucleotide-CPG matrix with 20% piperidine in dimethylformamide (DMF), washed with 10 ml of DMF, washed with 10 ml of acetonitrile and released oligonucleotides

Alamar Exhibit 1002
Alamar v. Olink

from the CPG with 30% ammonium hydroxide. We deprotected samples at 65 °C for 1 h and purified DNA tails by anion-exchange chromatography. The DNA tail for the streptavidin tadpole (**Supplementary Table 2** online) contained a T7 gene 10 promoter that directed transcription away from the affinity protein head; other tails did not contain the T7 promoter. We activated the oligonucleotide for covalent coupling by reduction with 1 mM TCEP for 30 min prior to incubation with the activated thioester on the recombinant proteins.

Intein-mediated protein-DNA ligation. We expressed recombinant intein fusion proteins in strains listed in **Supplementary Table 1** online and purified these fusions from the soluble fraction of *E. coli* lysates using a chitin column. After thorough washing, we passed the cysteine-modified DNA over the column in the presence of 0.2% thiophenol and 1 mM TCEP. We incubated the equilibrated column at 22 °C for 16 to 20 h, collected the column eluate and exchanged the reaction buffer to phosphate-buffered saline (PBS) or Tris-buffered saline (TBS) using NAP-25 columns (Pharmacia). To allow the tadpoles containing scFv domains to refold into an active form, we exchanged into buffer containing 1 mM TCEP and allowed refolding by slowly dialyzing out the TCEP. We purified tadpoles to apparent homogeneity (as judged by observing a single band on SYBR gold–stained polyacrylamide gels) by serial ion-exchange, hydrophobic and affinity chromatography.

Characterization of tadpoles. To monitor the coupling of DNA to the protein head we resolved samples from before and after conjugation in native 12% Tris-borate-EDTA polyacrylamide gels, stained DNA with SYBR gold, imaged with a STORM 860 scanning fluorimeter, and displayed the image using ImageQuant software (Molecular Dynamics). To verify the composition of the tadpoles we resolved conjugates in a Superdex-200 size exclusion column (Pharmacia) and monitored absorbance at 260 nm. To confirm that tadpoles contain a protein head, we digested with 0.1 μg of tadpoles with proteinase K and examined the sample by PAGE as above. To test ligand-binding activities, we incubated 0.1 μg of tadpole and 0.5 μg of specific ligand for 1 h at 22 °C and resolved reaction mixtures on native 12% polyacrylamide gels as above.

Quantification of biotinylated BSA diluted in an excess of nonspecific organic molecules. We verified the biotin content of our biotinylated BSA (Pierce) and carbonic anhydrase preparations using 2-(4′-hydroxyazobenzene)-benzoic acid (HABA) in soluble spectrophotometric assays essentially as described[22]. To quantify the amount of biotin *in vitro*, we diluted 100 μg of biotinylated BSA (containing 4 biotins per BSA) into NDA- and Texas red-coupled BSA and bound to polystyrene microtiter plates. We blocked the plates with 0.2% casein, 0.01% Tween and 10 μg/ml calf thymus DNA in PBS, incubated with 100 nM streptavidin tadpole, washed away unbound species and amplified bound tadpole with 100 U of T7 RNA polymerase and 100 μM of each NTP for 4 h at 37 °C. We resolved synthesized RNA on a denaturing gel, stained with SYBR gold[40] (Molecular Probes) and scanned the gel with 100 micron resolution at a photomultiplier tube voltage of 1,000 mV on a STORM 860. We computed signal values using ImageQuant software and graphed relative signal as a function of biotinylated BSA.

Quantification of biotin using real-time PCR. We covalently attached biotinylated BSA to glycidyl ester[41] (epoxy-modified) M-270 magnetic beads (Dynal Biotech) in PBS at pH 8.5 for 3 h at 30 °C, and unreacted epoxy groups were quenched with 50 mM ethanolamine for 2 h. By measuring free protein concentration[42] before and after coupling and by direct ELISA of the beads using mouse anti-biotin (Pierce) and goat anti-mouse HRP conjugate (Pierce), we determined that each bead immobilized approximately 50 fg of biotinylated BSA (data not shown). We serially diluted these biotinylated BSA-labeled beads into dilutions of unlabeled beads to create samples containing a total of 1.0×10^7 beads. The samples were blocked with 0.2% casein, 0.01% Tween and 10 μg/ml calf thymus DNA in PBS in 100 μl at 22 °C for 1 h and probed with 100 nM anti-biotin tadpole in 100 μl at 22 °C for 20 min. We quickly washed beads 5× with 500 μl blocking buffer for each wash at 22 °C and quantified the amount of bound tadpole by real-time PCR[24] using specific primers (**Supplementary Table 3** online) at a concentration of 0.2 μM with SYBR green[43] in 50 μl reactions.

Quantification of recombinant PA diluted into BSA. We diluted recombinant PA into 0.1 mg/ml BSA and covalently-coupled 10 μg of total protein from the dilutions to 1 μm diameter carboxylate-modified 'MyOne' magnetic beads (Dynal Biotech) using 1-ethyl-3-(3-dimethylaminopropyl) carbodimide[25] (EDC) and N-hydroxy-succinimide[26] (NHS) for 2 h at 22 °C and quenched with 50 mM ethanolamine as above. Under these conditions, ~10 fg of total protein was immobilized to each bead. We confirmed the numbers of PA molecules on the beads by Bradford analysis[42] of the protein solutions before and after coupling and by direct ELISA of the beads using goat anti-PA (List Biologicals) and donkey anti-goat HRP conjugates (Pierce) (data not shown). We blocked beads with 0.2% casein, 0.01% Tween and 10 μg/ml calf thymus DNA in 10 mM Tris pH 8, 0.2 M NaCl, probed samples with 100 nM anti-PA tadpole, washed using blocking buffer and heated to 95 °C for 5 minutes to release bound tadpoles. We quantified tadpole tail by real-time PCR using primers that hybridized to the oligonucleotide tail (**Supplementary Table 3** online) and SYBR green fluorescent dye.

Quantification of recombinant PA diluted into human serum. We diluted PA into 100 μl aliquots of EDTA-treated human serum and fractionated each dilution over 2 ml of T-gel (Pierce). We eluted bound protein with low salt buffer and covalently-coupled 2 μg of protein from the eluate of each column to carboxylate-modified 'MyOne' magnetic beads as above. We assayed these beads for PA using the anti-PA tadpole as above.

Note: Supplementary information is available on the Nature Methods website.

ACKNOWLEDGMENTS

The authors thank R. Fu for technical assistance with early real-time PCR experiments, L. Lok for invaluable advice on statistical analysis, O. Resnekov for suggesting T-gel for fractionating serum, C. Cantor for the gene encoding the streptavidin core subunit, J. Maynard for plasmids encoding anti-PA scFvs, purified scFvs and antibodies, and A. Coleman-Lerner, K. Benjamin, A. Gordon, L. Lok, A. Mendelsohn, O. Resnekov, E. Serra and R. Yu for useful discussions and comments on the manuscript. Some support for this work came from grants from the National Science Foundation and Defense Advanced Research Projects Agency, but main support came from the Center for Genomic Experimentation and Computation, a National Institutes of Health Center of Excellence in Genomic Science (P50 HG02370).

© 2005 Nature Publishing Group http://www.nature.com/naturemethods

Alamar Exhibit 1002
Alamar v. Olink

© 2005 Nature Publishing Group http://www.nature.com/naturemethods

AUTHORS' CONTRIBUTIONS

I.B.B. conceived the intein-mediated protein-DNA ligation synthesis and performed the experiments described here. K.Y. helped with nucleic acid constructions, protein expression, and in troubleshooting real-time PCR. A.G. provided the statistical methods used to quantify real-time PCR data. R.C. articulated the importance of controlled covalent linkage of nucleic acids to affinity molecules for biological measurements and provided input into initial experiments. R.B. input into the experiments and their quantification. R.B. and I.B. wrote the manuscript and guarantee the integrity of its findings.

COMPETING INTERESTS STATEMENT

The authors declare competing financial interests (see the *Nature Methods* website for details).

Received 6 November; accepted 24 November 2004
Published online at http://www.nature.com/naturemethods/

1. Engvall, E. & Perlman, P. Enzyme-linked immunosorbent assay (ELISA). Quantitative assay of immunoglobulin G. *Immunochemistry* **8**, 871–874 (1971).
2. Yalow, R.S. & Berson, S.A. Assay of plasma insulin in human subjects by immunological methods. *Nature* **184** Suppl. 21, 1648–1649 (1959).
3. Brenner, S. & Lerner, R.A. Encoded combinatorial chemistry. *Proc. Natl. Acad. Sci. USA* **89**, 5381–5383 (1992).
4. Sano, T., Smith, C.L. & Cantor, C.R. Immuno-PCR: very sensitive antigen detection by means of specific antibody-DNA conjugates. *Science* **258**, 120–122 (1992).
5. Fredriksson, S. *et al.* Protein detection using proximity-dependent DNA ligation assays. *Nat. Biotechnol.* **20**, 473–477 (2002).
6. Hendrickson, E.R., Truby, T.M., Joerger, R.D., Majarian, W.R. & Ebersole, R.C. High sensitivity multianalyte immunoassay using covalent DNA-labeled antibodies and polymerase chain reaction. *Nucleic Acids Res.* **23**, 522–529 (1995).
7. McKie, A., Samuel, D., Cohen, B. & Saunders, N.A. A quantitative immuno-PCR assay for the detection of mumps-specific IgG. *J. Immunol. Methods* **270**, 135–141 (2002).
8. Niemeyer, C.M. *et al.* Self-assembly of DNA-streptavidin nanostructures and their use as reagents in immuno-PCR. *Nucleic Acids Res.* **27**, 4553–4561 (1999).
9. Zhang, H.T., Kacharmina, J.E., Miyashiro, K., Greene, M.I. & Eberwine, J. Protein quantification from complex protein mixtures using a proteomics methodology with single-cell resolution. *Proc. Natl. Acad. Sci. USA* **98**, 5497–5502 (2001).
10. Nam, J.M., Thaxton, C.S. & Mirkin, C.A. Nanoparticle-based bio-bar codes for the ultrasensitive detection of proteins. *Science* **301**, 1884–1886 (2003).
11. Perler, F.B. *et al.* Intervening sequences in an Archaea DNA polymerase gene. *Proc. Natl. Acad. Sci. USA* **89**, 5577–5581 (1992).
12. Muir, T., Sondhi, D. & Cole, P. Expressed protein ligation: A general method for protein engineering. *Proc. Natl. Acad. Sci. USA* **95**, 6705–6710 (1998).
13. Sydor, J.R., Mariano, M., Sideris, S. & Nock, S. Establishment of intein-mediated protein ligation under denaturing conditions: C-terminal labeling of a single-chain antibody for biochip screening. *Bioconjug. Chem.* **13**, 707–712 (2002).
14. Lovrinovic, M. *et al.* Synthesis of protein-nucleic acid conjugates by expressed protein ligation. *Chem. Commun. (Camb.)* **7**, 822–823 (2003).
15. Green, N.M. & Melamed, M.D. Optical rotatory dispersion, circular dichroism and far-ultraviolet spectra of avidin and streptavidin. *Biochem. J.* **100**, 614–621 (1966).
16. Huston, J.S. *et al.* Protein engineering of antibody binding sites: recovery of specific activity in an anti-digoxin single-chain Fv analogue produced in *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* **85**, 5879–5883 (1988).
17. Colas, P. *et al.* Genetic selection of peptide aptamers that recognize and inhibit cyclin-dependent kinase 2. *Nature* **380**, 548–550 (1996).
18. Watanabe, T. *et al.* The roles of the C-terminal domain and type III domains of chitinase A1 from Bacillus circulans WL-12 in chitin degradation. *J. Bacteriol.* **176**, 4465–4472 (1994).
19. Stetsenko, D.A. & Gait, M.J. New phosphoramidite reagents for the synthesis of oligonucleotides containing a cysteine residue useful in peptide conjugation. *Nucleosides Nucleotides Nucleic Acids* **19**, 1751–1764 (2000).
20. Chong, S. *et al.* Single-column purification of free recombinant proteins using a self-cleavable affinity tag derived from a protein splicing element. *Gene* **192**, 271–281 (1997).
21. Sano, T., Pandori, M.W., Chen, X., Smith, C.L. & Cantor, C.R. Recombinant core streptavidins. A minimum-sized core streptavidin has enhanced structural stability and higher accessibility to biotinylated macromolecules. *J. Biol. Chem.* **270**, 28204–28209 (1995).
22. Green, N.M. A Spectrophotometric Assay for Avidin and Biotin Based on Binding of Dyes by Avidin. *Biochem. J.* **94**, 23C–24C (1965).
23. Van Gelder, R.N. *et al.* Amplified RNA synthesized from limited quantities of heterogeneous cDNA. *Proc. Natl. Acad. Sci. USA* **87**, 1663–1667 (1990).
24. Higuchi, R., Fockler, C., Dollinger, G. & Watson, R. Kinetic PCR analysis: real-time monitoring of DNA amplification reactions. *Biotechnology* **11**, 1026–1030 (1993).
25. Timkovich, R. Detection of the stable addition of carbodiimide to proteins. *Anal. Biochem.* **79**, 135–143 (1977).
26. Staros, J.V., Wright, R.W. & Swingle, D.M. Enhancement by N-hydroxysulfosuccinimide of water-soluble carbodiimide- mediated coupling reactions. *Anal. Biochem.* **156**, 220–222 (1986).
27. Neyman, J. Outline of a theory of statistical estimation based on the classical theory of probability. *Philos. Trans. R. Soc. London* **A236**, 333–380 (1937).
28. Feldman, G.J. & Cousins, R.D. Unified approach to the classical statistical analysis of small signals. *Phys. Rev.* **57**, 3873–3889 (1998).
29. Fabbrizio, E. *et al.* Inhibition of mammalian cell proliferation by genetically selected peptide aptamers that functionally antagonize E2F activity. *Oncogene* **18**, 4357–4363 (1999).
30. Leppla, S.H. Anthrax toxin edema factor: a bacterial adenylate cyclase that increases cyclic AMP concentrations of eukaryotic cells. *Proc. Natl. Acad. Sci. USA* **79**, 3162–3166 (1982).
31. Maynard, J.A. *et al.* Protection against anthrax toxin by recombinant antibody fragments correlates with antigen affinity. *Nat. Biotechnol.* **20**, 597–601 (2002).
32. Prinz, W.A., Aslund, F., Holmgren, A. & Beckwith, J. The role of the thioredoxin and glutaredoxin pathways in reducing protein disulfide bonds in the *Escherichia coli* cytoplasm. *J. Biol. Chem.* **272**, 15661–15667 (1997).
33. Salmon, S.E., Mackey, G. & Fudenberg, H.H. Sandwich" solid phase radioimmunoassay for the quantitative deterimination of human immunoglobulins. *J. Immunol.* **103**, 129–137 (1969).
34. Chen, L., Benner, J. & Perler, F.B. Protein splicing in the absence of an intein penultimate histidine. *J. Biol. Chem.* **275**, 20431–20435 (2000).
35. Schafer, U., Beck, K. & Muller, M. Skp, a molecular chaperone of gram-negative bacteria, is required for the formation of soluble periplasmic intermediates of outer membrane proteins. *J. Biol. Chem.* **274**, 24567–24574 (1999).
36. Hayhurst, A. & Harris, W.J. *Escherichia coli* skp chaperone coexpression improves solubility and phage display of single-chain antibody fragments. *Protein Expr. Purif.* **15**, 336–343 (1999).
37. Bothmann, H. & Pluckthun, A. Selection for a periplasmic factor improving phage display and functional periplasmic expression. *Nat. Biotechnol.* **16**, 376–380 (1998).
38. Caruthers, M.H. *et al.* New chemical methods for synthesizing polynucleotides. *Nucleic Acids Symp. Ser.* **7**, 215–223 (1980).
39. Caruthers, M.H. *et al.* Deoxyoligonucleotide synthesis via the phosphoramidite method. *Gene Amplif. Anal.* **3**, 1–26 (1983).
40. Tuma, R.S. *et al.* Characterization of SYBR Gold nucleic acid gel stain: a dye optimized for use with 300-nm ultraviolet transilluminators. *Anal. Biochem.* **268**, 278–288 (1999).
41. Zacharieva, E.I., Georgieva, M.P., Kabaivanov, V.S. & Popov, D.V. Preparation and investigation of cross-linked copolymers containing antitumour agents. *Biomaterials* **4**, 197–200 (1983).
42. Bradford, M.M. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* **72**, 248–254 (1976).
43. Garner, D.L., Johnson, L.A., Yue, S.T., Roth, B.L. & Haugland, R.P. Dual DNA staining assessment of bovine sperm viability using SYBR-14 and propidium iodide. *J. Androl.* **15**, 620–629 (1994).

Alamar Exhibit 1002
Alamar v. Olink

Proc. Natl. Acad. Sci. USA
Vol. 94, pp. 4937–4942, May 1997
Biochemistry

Best Available Copy

In vitro selection and evolution of functional proteins by using ribosome display

JOZEF HANES AND ANDREAS PLÜCKTHUN*

Biochemisches Institut, Universität Zürich, Winterthurerstrasse 190, CH-8057 Zurich, Switzerland

Communicated by Larry Gold, Nexstar Pharmaceuticals, Inc., Boulder, CO, March 7, 1997 (received for review November 7, 1996)

ABSTRACT **We report here a system with which a correctly folded complete protein and its encoding mRNA both remain attached to the ribosome and can be enriched for the ligand-binding properties of the native protein. We have selected a single-chain fragment (scFv) of an antibody 10^8-fold by five cycles of transcription, translation, antigen-affinity selection, and PCR. The selected scFv fragments all mutated *in vitro* by acquiring up to four unrelated amino acid exchanges over the five generations, but they remained fully compatible with antigen binding. Libraries of native folded proteins can now be screened and made to evolve in a cell-free system without any transformation or constraints imposed by the host cell.**

Evolutionary methods may bring refinement to protein engineering that is beyond the powers and accuracy of rational design today. Evolution can be defined as a succession of "generations," cycles of genetic diversification, followed by Darwinian selection for a desired phenotypic property. In classic experiments, nucleic acids have evolved and been selected for physical properties (1) *in vitro*, and in this case, the substance conferring the phenotype was identical to the genetic material. Oligonucleotide ligands, usually single-stranded RNA, have been identified for many targets by SELEX (2, 3), in which a synthetic DNA library is transcribed, and the RNA is selected for binding, reverse transcribed, and amplified over several rounds. However, all experiments with proteins as the carrier of the phenotype, clearly of much broader applicability, have so far relied on living cells for effecting the coupling between gene and protein, either directly or by the production of phages or viruses (4), since no *in vitro* method to link phenotype and genotype has been available for native proteins.

Since the DNA library as the information carrier for encoded protein diversity has to be transformed or transfected into bacterial or eukaryotic cells, the available diversity is severely limited by the low efficiency of DNA uptake (5). Furthermore, in each generation, the DNA library has to be first ligated into a replicatable genetic package, a process in which diversity is again decreased. Finally, many promising variants may be selected against in the host environment. Only a few studies (6) have carried protein optimization through more than one generation by using methods such as phage display, since this requires repeatedly switching between *in vitro* diversification and *in vivo* screening—a laborious process.

A series of studies have shown that specific mRNAs can be enriched by immunoprecipitation of polysomes (7–9). Recently, Mattheakis and co-workers (10) reported an affinity selection of a short peptide from a library by using polysomes, to connect genotype and phenotype *in vitro*.

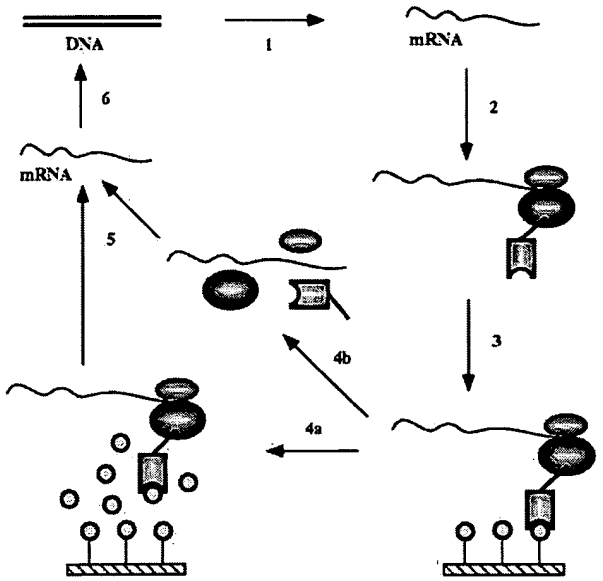


FIG. 1. Principle of *in vitro* ribosome display for screening native protein (scFv) libraries for ligand (antigen) binding. Step 1, a DNA scFv library is first amplified by PCR, whereby a T7 promoter, ribosome-binding site, and stem–loops are introduced, and then transcribed to RNA. Step 2, after purification, mRNA is translated *in vitro* in an *Escherichia coli* S-30 system in the presence of different factors enhancing the stability of ribosomal complexes and improving the folding of the scFv antibodies on the ribosomes. Translation is stopped by cooling on ice, and the ribosome complexes are stabilized by increasing the magnesium concentration. Step 3, the desired ribosome complexes are affinity selected from the translation mixture by binding of the native scFv to the immobilized antigen. Unspecific ribosome complexes are removed by intensive washing. Step 4, the bound ribosome complexes can then be dissociated by EDTA (step 4b), or whole complexes can be specifically eluted with antigen (step 4a). Step 5, RNA is isolated from the complexes. Step 6, isolated mRNA is reverse transcribed to cDNA, and cDNA is then amplified by PCR. This DNA is then used for the next cycle of enrichment, and a portion can be analyzed by cloning and sequencing and/or by ELISA or RIA.

We used Mattheakis' concept to develop a system that is compatible with the folding of whole proteins to their native structure on the ribosome, introducing a series of new features which proved to be essential to make the system efficient enough for whole proteins (Fig. 1). Since the protein can assume its native three-dimensional structure only after it is completely synthesized, the translation has to go to the very

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Copyright © 1997 by THE NATIONAL ACADEMY OF SCIENCES OF THE USA
0027-8424/97/944937-6$2.00/0
PNAS is available online at http://www.pnas.org.

Abbreviations: scFv, single-chain fragment of an antibody; V_L, variable domain of light chain; V_H, variable domain of heavy chain; scFvhag, anti-hemagglutinin scFv 17/9; scFvAL2, anti-β-lactam antibody AL2; T3Te, early terminator of phage T3; lpp term, terminator of *E. coli* lipoprotein; PDI, protein disulfide isomerase; VRC, vanadyl ribonucleoside complexes.

*To whom reprint requests should be addressed. e-mail: plueckthun@biocfebs.unizh.ch.

195

Alamar Exhibit 1002
Alamar v. Olink

end of the mRNA, and only the last few ribosomes on the mRNA can be expected to contain native protein able to bind to the ligand. We show here that an scFv single-chain fragment of an antibody, which requires its correctly assembled three-dimensional structure to bind the antigen (a hydrophilic peptide), can be enriched 10^8-fold by ribosome display, and its sequence "evolves" during the process

METHODS

Construction of scFv Antibody Fragments. The anti-hemagglutinin scFv (scFvhag) construct c7 was amplified by PCR in two steps from an fd phage displaying the scFv 17/9 (11, 12) containing a $(Gly_4Ser)_3$ linker, using in the first step oligonucleotides ON1 (5′-TTTCCCGAATTGTGAGCGGATAACAATAGAAATAATTTTGTTTAACTTTAAGAAGGAGATATATCCATGGACTACAAAGA-3′), which introduces a ribosome-binding site, and ON2 (5′-TTTAAGCAGCTCGATAGCAGCAC-3′). In the second step ON3 (5′-ATATATGTCGACGAAATTAATACGACTCACTATAGGGAGACCACAACGGTTTCCCGAATTGTG-3′), which introduces the T7 promoter and the 5′-loop, and ON4 (5′-AGACCCGTTTAGAGGCCCCAAGGGGTTATGGAATTCACCTTTAAGCAGCTC-3′), which introduces a modified lipoprotein terminator (lpp term) loop with stop codon removed, were used. The spacer, fused C-terminally to the scFv, is derived from amino acids 211–299 of gene *III* of filamentous phage M13 mp19 (12), and the translated lpp term adds another 24 amino acids.

The other constructs were prepared by PCR from c7 using the following primer pairs: c1, ON5 (5′-TCAGTAGCGACAGAATCAAG-3′) and ON6 (5′-GAAATTAATACGACTCACTATAGGGTTAACTTTAGAAGGAGGTATATCCATGGACTACAAAGA-3′), which introduces the T7 promoter and ribosome-binding site without the 5′-stem–loop, the spacer is derived from amino acids 211–294 of gene *III*; c2, ON4 and ON6, the same spacer as c7; c3, ON6 and ON7 (5′-GGCCCACCCGTGAAGGTGAGCCTCAGTAGCGACAG-3′), the spacer is derived from amino acids 211–294 of gene *III* following 7 amino acids of the translated early terminator of phage T3 (T3Te); c4, ON3 and ON5, the same spacer as c1; c5, ON3 and ON7, the same spacer as c3; c6, ON2 and ON3, and a spacer derived from amino acids 211–299 of gene *III* following the first 4 amino acids of lpp term; c8, ON3 and ON8 (5′-TTTAAGCAGCTCATCAAAATCACC-3′), and a spacer derived from amino acids 211–264 of gene *III* following the first 4 amino acids of lpp term; c9, ON3 and ON4, and a spacer derived from amino acids 211–264 of gene *III* following 24 amino acids of lpp term, using construct c8 as a template.

The anti-β-lactam scFvAL2 construct $[V_L\text{-}(Gly_4Ser)_4\text{-}V_H]$ was amplified by PCR in two steps from plasmid pAK202 (13), using in the first step oligonucleotides ON1 and ON7 and in the second step ON3 and ON7. The spacer is derived from amino acids 240–294 of gene *III* of filamentous phage M13 mp19 (13), and the translated T3Te adds another 7 amino acids.

Northern Analysis. RNA electrophoresis and transfer to a Nytran membrane (Schleicher & Schuell) were carried out as described (14) with a Turboblotter (Schleicher & Schuell). Hybridization was performed for 12 hr at 60°C (15). Hybridization was carried out with the oligonucleotide ON9 (5′-ACATGGTAACTTTTTCACCAGCGGTAACGG-3′), which anneals to the V_L region of scFvhag mRNA. This oligonucleotide was labeled by 3′ tailing with digoxigenin-11-dUTP/dATP by using the DIG Oligonucleotide Tailing Kit (Boehringer Mannheim). Washing conditions were as follows: 2× SSC/0.5% SDS for 5 min at room temperature; 2× SSC/0.5% SDS twice for 30 min at 60°C; and 0.1× SSC for 10 min at room temperature. The hybridized oligonucleotide probe was detected by using the DIG DNA Labeling and Detection Kit with the chemiluminescent substrate CSPD (Boehringer Mannheim) and exposure to x-ray film, or with the fluorimetric substrate AttoPhos (Boehringer Mannheim) and analysis using a fluoroimager (Molecular Dynamics).

***In Vitro* Translation of scFv Antibody Fragments.** *In vitro* translation in an *E. coli* S-30 system was performed according to Chen and Zubay (16) with several modifications. Translation was carried out for 10 min at 37°C in a 110-μl reaction mixture that contained the following components: 50 mM Tris·HOAc at pH 7.5, 30 mM NH_4OAc, 12.3 mM $Mg(OAc)_2$, 0.35 mM each amino acid, 2 mM ATP, 0.5 mM GTP, 1 mM cAMP, 0.5 mg/ml *E. coli* tRNA, 20 μg/ml folinic acid, 100 mM KOAc, 30 mM acetylphosphate (17), 1.5% polyethylene glycol 8000, 33 μg/ml rifampicin, 1 mg/ml vanadyl ribonucleoside complexes (VRC), 23 μl of *E. coli* MRE600 extract (16), and 90 μg/ml mRNA.

Affinity Selection of Ribosome Complexes and Isolation of RNA. The translation was stopped by adding $Mg(OAc)_2$ to a final concentration of 50 mM and cooling on ice (18). The samples were diluted 4-fold with ice-cold washing buffer [50 mM Tris·HOAc, pH 7.5/150 mM NaCl/50 mM $Mg(OAc)_2$/0.1% Tween 20] and centrifuged for 5 min at 4°C at 10,000 × *g* to remove insoluble components. Microtiter plates coated with hemagglutinin-transferrin conjugate were prewashed with ice-cold washing buffer, and the supernatant from the centrifuged translation mixture was applied (200 μl per microtiter well) and the plate was gently shaken for 1 hr in the cold room. After five washes with ice-cold washing buffer, the retained ribosome complexes were dissociated with ice-cold elution buffer (100 μl per well; 50 mM Tris·HOAc at pH 7.5/150 mM NaCl/10 mM EDTA/50 μg/ml *E. coli* tRNA) for 10 min in the cold room, and released mRNA was recovered by precipitation with ethanol or by isolation using the Rneasy kit (Qiagen).

Reverse Transcription–PCR and *in Vitro* Transcription. Reverse transcription was performed using Superscript reverse transcriptase (GIBCO/BRL) according to the supplier's recommendation. PCR was performed using *Taq* DNA polymerase (GIBCO/BRL) in the presence of 5% (vol/vol) dimethyl sulfoxide (4 min at 94°C, followed by 3 cycles of 30 sec at 94°C, 30 sec at 37°C, and 2 min at 72°C, followed by 10 similar cycles at 60°C instead of 37°C, 20 similar cycles at 60°C instead of 37°C with elongation at 72°C prolonged by 15 sec per cycle and finished by 10 min at 72°C). PCR products were analyzed by agarose gel electrophoresis and purified from the gel and reamplified, if the amount and quality was not sufficient, or directly used for transcription without additional purification. *In vitro* transcription was performed as described (19).

***In Vitro* Coupled Transcription–Translation and RIA.** After the fifth round of ribosome display, PCR products were cloned into the vector pTFT74 (20). *In vitro* coupled transcription–translation in the S-30 *E. coli* system was performed using 50 μg/ml plasmid DNA and conditions similar to those described above with the following modifications. A coupled transcription–translation was carried out for 30 min at 37°C, and the reaction mixture was supplemented with T7 RNA polymerase at 2000 units/ml and 0.5 mM UTP and TTP. The mixture contained [^{35}S]methionine at 50 μCi/ml (1 μCi = 37 kBq) and each amino acid except methionine at 0.35 mM. After translation, the reaction mixture was diluted 4-fold with PBS and bound to immobilized hemagglutinin peptide in a microtiter well. After 60 min of incubation with gentle shaking, microtiter wells were washed five times with PBS containing 0.1% Tween 20, and bound radioactive protein was eluted with 0.1 M triethylamine. Eluted protein was quantified in a scintillation counter.

RESULTS AND DISCUSSION

Model System and Quantification of Yields of Affinity Selection. As a model system, we used scFv fragments of antibodies (21), in which the variable domain of the light chain

Alamar Exhibit 1002
Alamar v. Olink

Best Available Copy

(V_L) is connected by means of a flexible linker to the variable domain of the heavy chain (V_H). To tether the folded protein to the ribosome and not interfere with folding we fused a spacer to the C terminus of the scFv fragment. Since the antibody domains must form disulfide bonds, and the RNA polymerase requires 2-mercaptoethanol for maximal stability, we performed transcription in a separate reaction and optimized the conditions for oxidative protein folding during translation (22) (see below).

Polypeptide release is an active process, requiring three polypeptide release factors in *E. coli* (23–25) and one ribosome recycling factor which releases the mRNA (26). The consequence of release factor binding is normally the hydrolysis of the peptidyl-tRNA between the ribose and the last amino acid by the peptidyltransferase center of the ribosome (25). Our system is devoid of stop codons, and thus a fraction of the polypeptide may not be hydrolyzed off the tRNA and remain attached to the ribosome and thus be available for affinity selection.

Translation is stopped by cooling on ice, and the ribosome complexes are further stabilized against dissociation by 50 mM magnesium acetate (18). Chloramphenicol at 50 μM, assumed to also induce ribosome stalling (10, 27), did not improve the selection properties (data not shown). The whole complex, consisting of synthesized protein, ribosome, and mRNA was then bound to the affinity matrix, washed, and eluted with competing ligand, or the ribosomes were dissociated with EDTA (Fig. 1). We did not find it necessary to preparatively isolate polysomes at any stage. The efficiency of ribosome display was found to be two orders of magnitude lower when using coupled *in vitro* transcription–translation (data not shown).

We developed the ribosome display system in two stages, first by engineering the gene structure for *in vitro* transcription, translation, and folding, then by optimizing the translation reaction itself. Each test started with 10 μg of input mRNA; this corresponds to $\approx 1.5 \times 10^{13}$ molecules. The input mRNA was subjected to a single round of affinity selection: translation *in vitro*, capture of ribosome complexes on immobilized target ligand, and release of mRNA. The released mRNA was then quantified by Northern analysis.

Effect of mRNA Structure on Yield. In designing the ribosome display system, we first engineered the flanking regions of the scFv gene (Fig. 2*A*). The gene must be transcribed very efficiently from the PCR product, and its mRNA must be stable against nucleases. For the 5′ end, we used the T7 promoter and the natural T7 gene *10* upstream region, which encodes a stem–loop structure directly at the beginning of the mRNA (29). At the 3′ end, we fused a spacer region of 57–116 amino acids to the reading frame of the scFv to tether the emerging folded polypeptide to the putative polypeptide channel of the ribosome and to give it enough distance to not interfere with folding. This spacer encodes at the RNA level a 3′ stem–loop, either the terminator of the *E. coli* lipoprotein (29) (lpp term) or the early terminator of phage T3 (30) (T3Te), to increase the stability of mRNA against exonucleases. The mRNA is therefore protected against exonucleases at both ends, much like a natural mRNA would be, but does not contain a translational stop codon. In a direct comparison of flanking regions (Fig. 2*B*) after one round of translation and selection, we obtained the best result for constructs possessing a 5′-loop (from T7-g10), a 3′-loop (from lpp term), and the longest spacer, 116 amino acids. The yield of mRNA after one round of ribosome display improved from less than 0.001% of input mRNA to 0.015% (15 times).

Effect of Translation Conditions on Yield. We then tested the effect of various compounds present during translation. Nucleases were found to be efficiently inhibited by VRC, which should act as transition state analogs (31). VRC at 1 mg/ml maximized the yield of isolated mRNA from ribosome com-

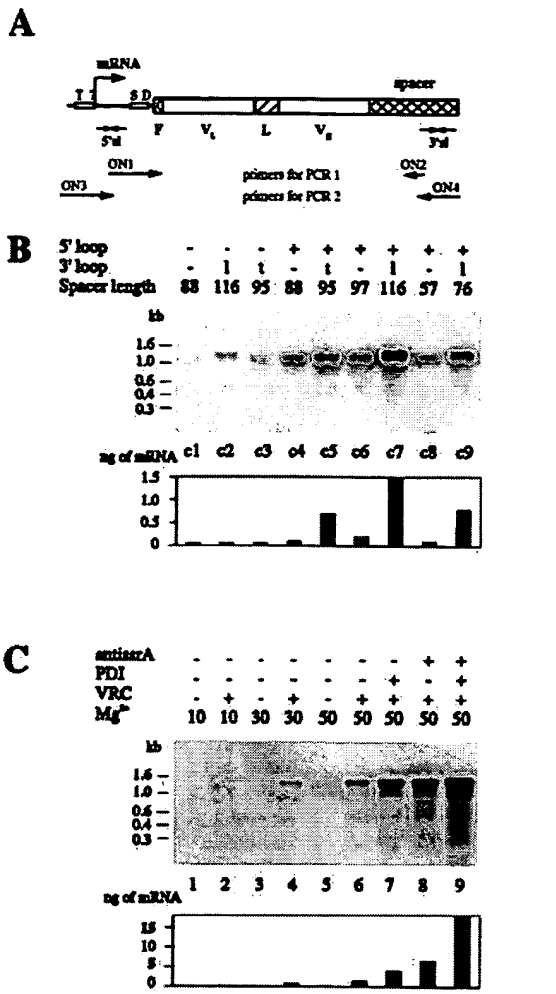


FIG. 2. (*A*) Schematic drawing of the scFv construct used for ribosome display. T7 denotes the T7 promoter, SD the ribosome-binding site, F the 3-amino-acid-long FLAG (28) with N-terminal methionine, V_L and V_H the variable domains of the scFv fragment, L the linker, spacer the part of protein construct connecting the folded scFv to the ribosome, and 5′sl and 3′sl the stem–loops on the 5′ and 3′ ends of the mRNA. The arrow indicates the transcriptional start. The strategy for the PCR amplification is shown. (*B*) The amount of the scFvhag mRNA bound to the polysome complexes is influenced by the secondary structure of its ends and the length of the spacer connecting the folded scFv to the ribosome. Different constructs of scFvhag mRNA were used for one cycle of ribosome display (constructs c1–c9). None of them contained a stop codon. Each species was tested separately. After affinity selection of the ribosome complexes from 100 μl of translation mixtures, mRNAs were isolated and analyzed by Northern hybridization. The presence of the 5′ stem–loop is labeled + or −; that of the 3′ stem–loop is labeled as − when absent, l when derived from lpp term, or t when derived from T3Te. The spacer length indicates the number of amino acids following the scFv and connecting it to the ribosome, including the translated stem–loop. The lengths of RNA in kb are derived from RNA molecular weight marker III (Boehringer Mannheim). The bar graph shows the quantified amounts of mRNA from fluoroimager analysis. (*C*) Effect of additives on the amount of mRNA bound to the ribosome complexes. The mRNA of the scFvhag construct c7 was used for one cycle of ribosome display. Samples in *B* lane c7 and *C* lane 6 are identical. When indicated, 3.5 mM anti-*sstrA* oligonucleotide ON10 (5′-TTAAGCTGCTAAAGCGTAGTTTTCGTCGTTTGCGACTA-3′), 35 mg/ml protein disulfide isomerase (PDI), or 1 mg/ml VRC was included in the translation mixture. Translations were stopped by addition of $Mg(OAc)_2$ to the final concentration indicated (in mM) and by cooling on ice. After affinity selection of the ribosome complexes from 100 μl of translation mixtures, mRNAs were isolated and analyzed by Northern hybridization. The bar graph shows the quantified amounts of mRNA from fluoroimager analysis.

Alamar Exhibit 1002
Alamar v. Olink

Best Available Copy

plexes after one round of affinity selection when they were present during translation (Fig. 2C), even though protein synthesis was partially inhibited (data not shown), and omitting VRC led to a severalfold-decreased efficiency of the ribosome display (Fig. 2C). In contrast, Rnasin (10) had no effect on the efficiency of the system (data not shown). We did not find evidence for significant proteolytic degradation of the scFv synthesized under these conditions, since prolonged incubation of released product (up to 300 min) did not alter the electrophoretic pattern (data not shown).

From a systematic study on the *in vitro* translation of soluble scFv fragments in the presence of molecular chaperones and disulfide-forming catalysts (22), we found that binding activity is obtained if and only if disulfide formation and rearrangement are allowed to take place during translation and folding. The strong beneficial effect of PDI was verified for the ribosome display system, in which the protein is not released. It can be seen that PDI improves the performance of the ribosome display for scFv fragments 3-fold (Fig. 2C) and thus catalyzes the formation and isomerization of disulfide bonds on the ribosome-bound protein.

Recently, a peptide tagging system was discovered in *E. coli*, whereby proteins translated from mRNA, devoid of a stop codon, are modified and released from the ribosomes by the addition of a C-terminal tag, encoded by *ssrA* RNA, and thereby marked for degradation (32). We thus investigated the inhibition of the action of this RNA by an antisense oligonucleotide complementary to the tag-coding sequence of *ssrA* RNA. Indeed, a 4-fold higher efficiency of ribosome display is visible in the presence of anti-*ssrA* oligonucleotide (Fig. 2C), and the molecular weight of the longest protein product is decreased, presumably by preventing the attachment of the degradation tag (Fig. 3). Combining PDI and the anti-*ssrA* oligonucleotide led to a 12-fold-increased efficiency of the ribosome display system (Fig. 2C).

By the combination of proper mRNA secondary structure and various compounds present during translation we could increase the yield of mRNA after one round of affinity selection 200 times—from less than 0.001% to 0.2% of input mRNA. This number expresses the combined efficiency of covalent attachment to the ribosome, protein folding, ligand binding, ribosome capture, and RNA release and amplification.

Specific Enrichment of Target mRNA Through Multiple Rounds of Selection. In a test of the optimized system, we investigated how well a mixture of two proteins can be enriched for function. Two scFv antibody mRNAs, constructed identically according to Fig. 2A, both possessing the 5′ and the 3′ T3Te loop, one encoding the anti-hemagglutinin scFv 17/9 (11) (scFvhag), the other the anti-β-lactam antibody AL2 (13) (scFvAL2), were mixed at a ratio of $1:10^8$. Their PCR products differ slightly in length, because of differences mainly in the spacer length, and can thus easily be distinguished (Fig. 4A). After five cycles according to Fig. 1, undergoing selection on immobilized hemagglutinin-peptide, 90% of the ribosome complexes contained scFvhag. We can thus conclude that the enrichment is about 2 orders of magnitude per cycle under these conditions.

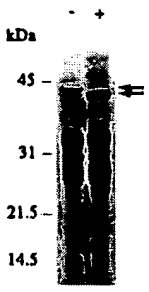


FIG. 3. Anti-*ssrA* antisense oligonucleotide ON10 (Fig. 2C) decreases the molecular mass of the longest protein species (arrows). *In vitro* translation was performed using [^{35}S]methionine and scFvhag c7 mRNA. Reactions were carried out in the absence (−) or presence (+) of 3.5 mM oligonucleotide ON10. An SDS/PAGE of translation products is shown.

To verify that enrichment really occurred through affinity selection, we tested enrichment of a 1:1 mixture of both mRNAs on an irrelevant surface and saw no change from the input ratio (Fig. 4B). Furthermore, from the identical 1:1 mRNA mixture either antibody could be enriched, depending on which antigen was immobilized (Fig. 4B).

Analysis of scFv Antibody Fragments After Affinity Selection. After the fifth round of selection, the PCR products were ligated into a vector, which was used to transform *E. coli*, and single clones were analyzed. Of 20 clones sequenced, 18 had the scFvhag sequence and 2 had the scFvAL2 sequence, demonstrating that the 10^8-fold enrichment was successful. Of the 18 scFvhag clones, 13 gave ELISA signals within 43–102% of wild type and were inhibitable by soluble hemagglutinin peptide, 2 were reduced to 14% and 18%, and 3 were significantly reduced in binding (less than 10%), probably a result of mistakes from the last PCR amplification (Table 1). Thus, the selective pressure to maintain antigen binding, executed by binding and elution from immobilized antigen, is clearly operating, albeit in the context of an ongoing genetic diversification through PCR errors.

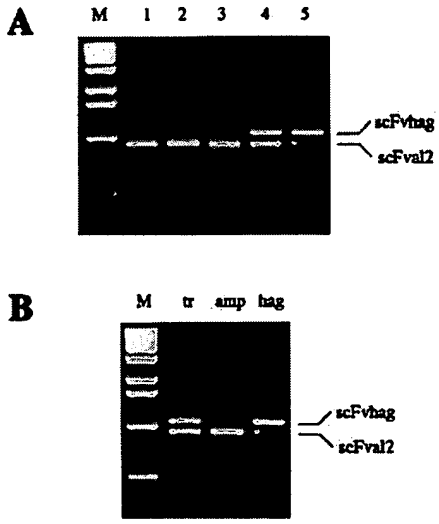


FIG. 4. (A) Enrichment of the scFvhag ribosome complexes from mixtures with scFvAL2 by ribosome display. mRNA of scFvhag c5 was diluted 10^8 times with mRNA of scFvAL2, and the mixture was used for ribosome display. After affinity selection of scFvhag ribosome complexes, mRNA was isolated and reverse transcribed to cDNA using oligonucleotide ON5, and cDNA was amplified by PCR using oligonucleotides ON3 and ON7 and analyzed by agarose electrophoresis. Lanes 1–5 are PCR products amplified after the first to fifth cycles of the ribosome display, respectively. Lane M is the 1-kb DNA ladder (GIBCO/BRL) used as molecular weight markers. PCR products corresponding to scFvhag and scFvAL2 are labeled. (B) Enrichment of either scFvhag c5 or scFvAL2 ribosome complexes by ribosome display as a function of immobilized antigen. mRNAs of scFvhag and scFvAL2 were mixed in a 1:1 ratio and used for one cycle of ribosome display. After affinity selection, mRNA was isolated, reverse transcribed, amplified by PCR, and analyzed by agarose electrophoresis as in A. The same 1:1 mixture was affinity selected on immobilized transferrin (tr) as a control, ampicillin-transferrin conjugate (amp), or hemagglutinin-peptide-transferrin conjugate (hag). PCR products corresponding to scFvhag and scFvAL2 are labeled.

Alamar Exhibit 1002
Alamar v. Olink

Table 1. Mutations in scFvhag fragments selected by ribosome display *in vitro*

Clone no.	Mutation V_L	Mutation V_H	Mutation linker	Relative RIA signal	Relative RIA signal +hag	Inhibition, %
12	K45R	P41R		102	6	94
10		S30P		101	5	95
6*				100	4	96
2*				96	5	95
7		G16D		89	5	95
3	Y49H			86	5	95
9	E55G, E105G	A23V	S5P	86	5	94
1	T20A		G12E	80	3	96
18	N(27d)D	D10G, T50N		76	5	93
4	V13A, K28R	V12A		72	6	91
13	V58I	Y79C		65	8	87
16	L83R	I51V, K75E, A113V		63	5	92
14	K30E	D61G	S5P	43	6	85
17	F71S			18	6	66
8	E17G, K18E, K30E			14	6	56
5	L11P		G13D	9	5	51
11	S(27e)F, N90S	V48I		9	7	22
15	G64D, S77A	G15S		6	6	2

Mutations in the scFv of the antibody 17/9 (11) are numbered according to Kabat *et al.* (33) in the variable domains, and sequentially in the linker. Clones are sorted according to a relative RIA signal. Wild-type clones 2 and 6 are marked by asterisks. RIAs were normalized to the same amount of full-length protein and inhibited by 0.1 mM hemagglutinin peptide (denoted as +hag), to verify binding specificity.

The sequence analysis showed that the clones contained between 3 and 7 base changes, with 90% transitions and 10% transversions, in good agreement with the known error properties of *Taq* DNA polymerase (34). Each clone has gone through a total of 165 cycles of PCR, interrupted by five phenotypic selections for binding, and an error rate of 3.3×10^{-5} can be calculated under these conditions, where only mutations that have survived the selection are counted.

At the protein level, the selected clones carry between 0 and 4 exchanged amino acids, distributed over V_L, V_H, and the linker. All mutations are independent of each other (Table 1) and give, therefore, an indication of a large range of neutral mutations compatible with function, and perhaps even of improvements the system has selected for. While the exact properties of the selected molecules require further analysis, we note the selection of a proline at the beginning of the linker, which may facilitate the required turn formation (35), and the selection of acidic residues, which are known to increase solubility (36, 37).

Conclusions. We have shown in this study that it is possible to carry out phenotypic selection for ligand binding with a complete, native protein molecule *in vitro* by using ribosome display, in this example with a disulfide containing single-chain fragment of an antibody. Functional disulfide-containing protein can be made *in vitro*, when oxidative folding is suitably catalyzed (22). Since binding of the cognate ligand (antigen) is dependent on the correct three-dimensional structure of the protein (antibody) and its sufficient stability, many molecular properties of a protein can now be selected for. The method described here, ribosome display, does not use any cells at any step, and thus, to our knowledge, constitutes the first evolution and selection system for folded proteins that takes place completely *in vitro*. No transformation is necessary, and very large libraries become accessible in a single step. While large repertoires have been created for phage display as well (38), many repeated electroporations are necessary, since transformation cannot be scaled up. Furthermore, complex libraries must often be created by repeating this process for the subsequent ligation of different fragments (38). In ribosome display, all replication and amplification, as well as the introduction of the promoter and the 3′ tether, are carried out by PCR *in vitro*, and thus this method is conveniently interfaced with novel, elegant methods for obtaining even more genetic diversity, such as sexual PCR (39), error-prone PCR (40), or PCR using oligonucleotides made with mixtures of trinucleotide codons (41). Thus, ribosome display is an ideal method for the evolution of proteins through many generations, avoiding the tedious alternation between *in vitro* and *in vivo* steps, which would be required to achieve the same with phage display. Conversely, the error rate can also be decreased in ribosome display by the use of proof-reading polymerases, if ribosome display is used merely as a screening tool. We believe that with this technology described here very fundamental questions of protein structure and evolution become accessible for study, and that it has great utility in lead compound identification and optimization.

We thank L. Jermutus for many helpful discussions and suggestions in the development of the method, L. Ryabova for helpful discussions on *in vitro* translation, D. Desplancq for the scFv 17/9, and A. Krebber for the scFvAL2. This work was supported by Schweizerischer Nationalfonds Grant 3100-037717.93/1.

1. Saffhill, R., Schneider-Bernloehr, H., Orgel, L. E. & Spiegelman, S. (1970) *J. Mol. Biol.* **51**, 531–539.
2. Gold, L., Polisky, B., Uhlenbeck, O. & Yarus, M. (1995) *Annu. Rev. Biochem.* **64**, 763–797.
3. Irvine, D., Tuerk, C. & Gold, L. (1991) *J. Mol. Biol.* **222**, 739–761.
4. Phizicky, E. M. & Fields, S. (1995) *Microbiol. Rev.* **59**, 94–123.
5. Dower, W. J. & Cwirla, S. E. (1992) in *Guide to Electroporation and Electrofusion*, eds. Chang, D. C., Chassy, B. M., Saunders, J. A. & Sowers, A. E. (Academic, San Diego), pp. 291–301.
6. Yang, W. P., Green, K., Pinz-Sweeney, S., Briones, A. T., Burton, D. R. & Barbas, C. F., 3rd (1995) *J. Mol. Biol.* **254**, 392–403.
7. Schechter, I. (1973) *Proc. Natl. Acad. Sci. USA* **70**, 2256–2260.
8. Payvar, F. & Schimke, R. T. (1979) *Eur. J. Biochem.* **101**, 271–282.
9. Kraus, J. P. & Rosenberg, L. E. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 4015–4019.
10. Mattheakis, L. C., Bhatt, R. R. & Dower, W. J. (1994) *Proc. Natl. Acad. Sci. USA* **91**, 9022–9026.
11. Schulze-Gahmen, U., Rini, J. M. & Wilson, I. A. (1993) *J. Mol. Biol.* **234**, 1098–1118.
12. Krebber, C., Spada, S., Desplancq, D. & Plückthun, A. (1995) *FEBS Lett.* **377**, 227–231.
13. Krebber, A., Bornhauser, S., Burmester, J., Honegger, A., Willuda, J., Bosshard, H. R. & Plückthun, A. (1997) *J. Immunol. Methods* **201**, 35–55.

Alamar Exhibit 1002
Alamar v. Olink

14. Goda, S. K. & Minton, N. P. (1995) *Nucleic Acids Res.* **23**, 3357–3358.
15. Church, G. M. & Gilbert, W. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 1991–1995.
16. Chen, H. Z. & Zubay, G. (1983) *Methods Enzymol.* **101**, 674–690.
17. Ryabova, L. A., Vinokurov, L. M., Shekhotsova, E. A., Alakhov, Y. B. & Spirin, A. S. (1995) *Anal. Biochem.* **226**, 184–186.
18. Holschuh, K. & Gassen, H. G. (1982) *J. Biol. Chem.* **257**, 1987–1992.
19. Pokrovskaya, I. D. & Gurevich, V. V. (1994) *Anal. Biochem.* **220**, 420–423.
20. Ge, L., Knappik, A., Pack, P., Freund, C. & Plückthun, A. (1995) in *Antibody Engineering*, ed. Borrebaeck, C. A. K. (Oxford Univ. Press, New York), pp. 229–266.
21. Huston, J. S., Levinson, D., Mudgett-Hunter, M., Tai, M. S., Novotny, J., Margolies, M. N., Ridge, R. J., Bruccoleri, R. E., Haber, E., Crea, R. & Oppermann, H. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 5879–5883.
22. Ryabova, L., Desplancq, D., Spirin, A. & Plückthun, A. (1997) *Nature Biotechnology* **15**, 79–84.
23. Grentzmann, G., Brechemier-Baey, D., Heurgue-Hamard, V. & Buckingham, R. H. (1995) *J. Biol. Chem.* **270**, 10595–10600.
24. Tuite, M. F. & Stansfield, I. (1994) *Mol. Biol. Rep.* **19**, 171–181.
25. Tate, W. P. & Brown, C. M. (1992) *Biochemistry* **31**, 2443–2450.
26. Janosi, L., Shimizu, I. & Kaji, A. (1994) *Proc. Natl. Acad. Sci. USA* **91**, 4249–4253.
27. Moazed, D. & Noller, H. F. (1987) *Nature (London)* **327**, 389–394.
28. Knappik, A. & Plückthun, A. (1994) *BioTechniques* **17**, 754–761.
29. Studier, F. W., Rosenberg, A. H., Dunn, J. J. & Dubendorff, J. W. (1990) *Methods Enzymol.* **185**, 60–89.
30. Reynolds, R., Bermudez-Cruz, R. M. & Chamberlin, M. J. (1992) *J. Mol. Biol.* **224**, 31–51.
31. Berger, S. L. (1987) *Methods Enzymol.* **152**, 227–234.
32. Keiler, K. C., Waller, P. R. & Sauer, R. T. (1996) *Science* **271**, 990–993.
33. Kabat, E. A., Wu, T. T., Perry, H. M., Gottesmann, K. S. & Foeller, C. (1991) in *Sequences of Proteins of Immunological Interest* (U.S. Dept. Health Human Services), 5th Ed., Vol. I, pp. 151 and 464.
34. Keohavong, P. & Thilly, W. G. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 9253–9257.
35. Tang, Y., Jiang, N., Parakh, C. & Hilvert, D. (1996) *J. Biol. Chem.* **271**, 15682–15686.
36. Dale, G. E., Broger, C., Langen, H., D'Arcy, A. & Stüber, D. (1994) *Protein Eng.* **7**, 933–939.
37. Knappik, A. & Plückthun, A. (1995) *Protein Eng.* **8**, 81–89.
38. Vaughan, T. J., Williams, A. J., Pritchard, K., Osbourn, J. K., Pope, A. R., Earnshow, J. C., McCafferty, J., Hodits, R. A., Wilton, J. & Johnson, K. S. (1996) *Nat. Biotechnol.* **14**, 309–314.
39. Stemmer, P. (1994) *Nature (London)* **370**, 389–391.
40. Cadwell, R. C. & Joyce, G. F. (1994) *PCR Methods Appl.* **3**, S136–S140.
41. Virnekäs, B., Ge, L., Plückthun, A., Schneider, K. C., Wellnhofer, G. & Moroney, S. E. (1994) *Nucleic Acids Res.* **22**, 5600–5607.

Alamar Exhibit 1002
Alamar v. Olink

nature biotechnology

© 2003 Nature Publishing Group http://www.nature.com/naturebiotechnology

Multiplexed genotyping with sequence-tagged molecular inversion probes

Paul Hardenbol[1,3], Johan Banér[2], Maneesh Jain[1,3], Mats Nilsson[2], Eugeni A Namsaraev[1,3], George A Karlin-Neumann[1,3], Hossein Fakhrai-Rad[1,3], Mostafa Ronaghi[1], Thomas D Willis[1,3], Ulf Landegren[2] & Ronald W Davis[1]

We report on the development of molecular inversion probe (MIP) genotyping, an efficient technology for large-scale single nucleotide polymorphism (SNP) analysis. This technique uses MIPs to produce inverted sequences, which undergo a unimolecular rearrangement and are then amplified by PCR using common primers and analyzed using universal sequence tag DNA microarrays, resulting in highly specific genotyping. With this technology, multiplex analysis of more than 1,000 probes in a single tube can be done using standard laboratory equipment. Genotypes are generated with a high call rate (95%) and high accuracy (>99%) as determined by independent sequencing.

The availability of large collections of SNPs along with recent large-scale linkage disequilibrium mapping efforts[1] have brought the promise of personalized whole-genome association studies to the field of human genetics. To achieve this goal, methodologies that permit screening of hundreds of thousands of SNPs will be needed to implement such large-scale association studies on a routine basis. These methods not only will have to be inexpensive per SNP screened, but will need to consume very little genomic DNA—that is, no more than is typically obtained from a patient's blood sample. In addition, such technologies should ideally require minimal investment in infrastructure so that the technology can be made broadly available.

The challenge of genotyping the approximately 150 molecules of a given SNP locus present in 1 ng of genomic DNA is commonly met by PCR amplification of the locus before genotyping is done[2]. However, an increase in the number of target sequences for simultaneous amplification by PCR quickly leads to unmanageable levels of cross-reaction among primer pairs[3,4], whereas parallel hybridization on microarrays[5,6] lacks the specificity and sensitivity required to genotype large genomes directly.

There are only a limited number of genotyping technologies with sufficient specificity to identify an SNP from genomic DNA without prior PCR amplification. Flap endonucleases have been used to generate a sequence-specific endonuclease cascade in an isothermal fashion that can be assessed with FRET probes[7,8]. However, this technology is not readily multiplexed for high-throughput applications. Padlock probes are linear oligonucleotides, whose two ends can be joined by ligation when they hybridize to immediately adjacent target sequences[9]. As shown before[10–12], padlock probes provide sufficient specificity analyze SNPs directly, without previous amplification of the target sequences.

Unlike amplification strategies such as PCR and the Invader assay that require two specific primers, cross-reactive padlock probes can easily be distinguished from the desired circular products by methods such as exonucleolysis[9]. This offers the opportunity to add a complex pool of padlock probes to individual DNA samples to investigate large sets of genes in parallel, without a concomitant increase in the risk of cross-reactivity between different probes.

Here we present a strategy that combines DNA detection specificity and sensitivity with the potential to analyze large numbers of target sequences in parallel. Sets of padlock probes with universal tag sequences were reacted with target DNA, molecularly inverted, amplified together and identified in a multiplex analysis yielding more than 1,000 genotypes simultaneously. Using molecular inversion probes, the information content of the SNPs was reformatted into tag sequences that could be detected using a universal oligonucleotide detection array[13]. We report the application of this technique at unprecedented levels of multiplexing, resulting in a lowering of the scale, cost and sample requirements of high-throughput genotyping. The approach retained high accuracy through multiple hybridization and enzymatic processing events, and provided inherent quality control checking.

RESULTS

Selection for circularized probes using exonucleases

Most genotyping methods require PCR amplification of the region spanning the sequence variation. However, when sets of n PCR primer pairs are combined in one reaction to evaluate n target sequences, any of the $2n^2 + n$ possible pairwise primer combinations may give rise to nonspecific amplification products[3]. With padlock probes the corresponding cross-reactive ligation products create linear dimeric molecules, easily distinguished from circularized

[1]Stanford Genome Technology Center, Stanford University, 855 California Avenue, Palo Alto, California 94304, USA. [2]The Beijer Laboratory, Department of Genetics and Pathology, Rudbeck Laboratory, Se-751 85 Uppsala, Sweden. [3]Present address: ParAllele BioScience 384 Oyster Point Blvd Suite 8, S. San Francisco, California 94080, USA. Correspondence should be addressed to M.R. (mostafa@stanford.edu).

Alamar Exhibit 1002
Alamar v. Olink

Best Available Copy

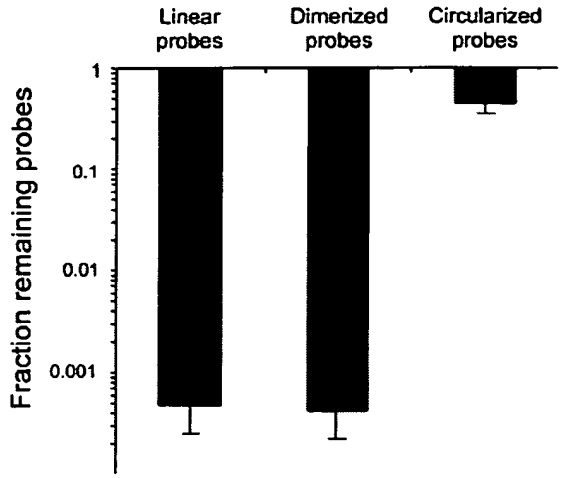


Figure 1 Selection for circularized padlock probes. Effect of exonuclease on linear monomer, dimer or on circularized padlock probes were measured by real-time PCR. Dimerized probes were produced using a ligation template that allowed two different padlock probes to be joined. The results were converted to numbers of molecules by reference to a standard dilution series. The fractions of remaining probe were calculated by dividing each reaction by the respective starting number. Error bars denote s.d. of the ratios from eight reactions.

© 2003 Nature Publishing Group http://www.nature.com/naturebiotechnology

probes by exonucleolytic degradation[9,14]. The exonuclease treatment protocol reduces the number of such linear monomeric and dimeric molecules by almost three orders of magnitude with negligible effects on circularized probes as measured by real-time PCR (Fig. 1). The removal of unreacted probes further reduces ligation-independent amplification events that may otherwise occur through accidental priming or templating of polymerization by the large number of linear probes (data not shown).

Molecular inversion probe (MIP) genotyping

Initially we combined pairs of padlock probes specific for alternate alleles in SNP loci. This permitted parallel genotyping of several loci in a single reaction before amplification and identification of the reaction products on tag arrays (Fig. 2b). Before increasing the multiplexing level, we redesigned the padlock probes to be locus-specific to avoid the need for balancing allele-specific probes at every locus (Fig. 2a). With this strategy only one probe was required per locus. To achieve this, the polymorphic nucleotide at the 3′ end of the probe was left out, creating a gap between the probe ends. This gap was then filled in four separate allele-specific polymerization (A, C, G and T in four different tubes) and ligation reactions[15]. Next, the probes were released from the genomic DNA by removing the uracil residues between primer sequences to avoid topological inhibition of the polymerization reaction[16]. The oligonucleotide probe undergoes a unimolecular rearrangement before amplification (Fig. 2b). Each probe contains a unique 20-base tag sequence that is complementary to a sequence on an Affymetrix GenFlex Tag Array. The tags are selected to be similar in melting temperature (T_m) and base composition, and maximally orthogonal in sequence complementarity. These tags amplify and hybridize under a single set of conditions with minimal cross-hybridization to each other and to other features on the microarray.

After amplification, the products are hybridized on four DNA microarrays and the components are decoded by measuring the fluorescence signals at the corresponding complementary tag site on the DNA array (Fig. 3a). An image of 938 amplified probes hybridized to a DNA array is shown (Fig. 3b). Four intensity values for each probe are generated. The two values for the expected allelic bases are compared to determine whether the sample is homozygous or heterozygous for the given SNP, and the two non-allele bases are compared to the allele bases to determine the signal-to-noise ratio (SNR) for the probe (Table 1). The two non-allele bases serve as internal controls that are used to reduce incorrect genotype calls owing to missing, degraded or noisy probes.

Assay performance

To investigate the performance of the method, probes were generated for 1,121 SNPs from the SNP consortium (TSC) database (http://snp.cshl.org) for a 16-megabase region on chromosome 6 centered on the linkage peak for IgA nephropathy[17] (Table 2). Markers were selected from the database based on map position. Of the 1,121 probes, 183 (16%) were inactive during a single synthesis step, possibly owing to such problems as errors in the database, probe design, or failures of oligonucleotide synthesis, probe synthesis or the assay itself. In a pilot study, 25 different individuals were genotyped with the 938 active probes for a total of 23,450 assays. We successfully called 21,336 full genotypes (two chromosomes) and 1,746 half genotypes (single chromosome) (95%) with a median SNR of 16.7 for allele-specific signal to non-allele signal. Half genotypes are reported when the identity of only one of two chromosomes is certain. A cluster plot of data of four of the probes used to genotype 25 individuals is shown (Fig. 4). No substantial differences were seen in the call rates of probes designed for all allele combinations (Table 3).

Accuracy was determined through independent sequencing. 1,517 loci were genotyped in a 1,517-probe multiplex analysis with ten individuals. Forward and reverse Sanger sequencing was performed on a subset (129) of PCR amplicons of 1,517 loci amplified from the same 10 individuals. Conservative reads were made manually with the identity of the forward and reverse loci blinded at the time of sequence interpretation. Accuracy of Sanger sequencing was measured by comparing reads for which the sequence of both strands existed. 359 of 367 sequence pairs were identical, for an

Table 1 Data generated from the first 10 probes from individual NA17203

Probe ID	Allele	Base call[a]	Signal A[b]	Signal G	Signal C	Signal T	SNR[c]
2,515	A/G	G/G	139	1,472	216	202	6.8
2,516	A/G	A/A	437	21	30	31	14.1
2,517	A/G	A/G	1,538	1,494	95	94	16.2
2,518	A/G	A/G	343	474	39	30	12
2,519	A/G	A/A	3,574	39	51	65	55.2
2,520	A/G	G/G	147	1,702	175	172	9.8
2,521	A/G	G/G	59	1290	45	38	28.5
2,522	A/G	A/G	478	382	110	87	4.4
2,523	A/G	G/G	36	1,234	49	62	19.9
2,524	A/G	G/G	62	1,492	59	115	13

[a]A base-call is made if the SNR is at least 3, and the ratio of the higher allele signal to the lower allele signal is >6:1 for homozygous calls and <2:1 for heterozygous calls. [b]Signals are normalized fluorescence intensity values produced by the microarray scanner. [c]The SNR was calculated for each marker by dividing the maximum allele signal (A and G) by the maximum signal in the two background channels (C and T).

Alamar Exhibit 1002
Alamar v. Olink

Best Available Copy

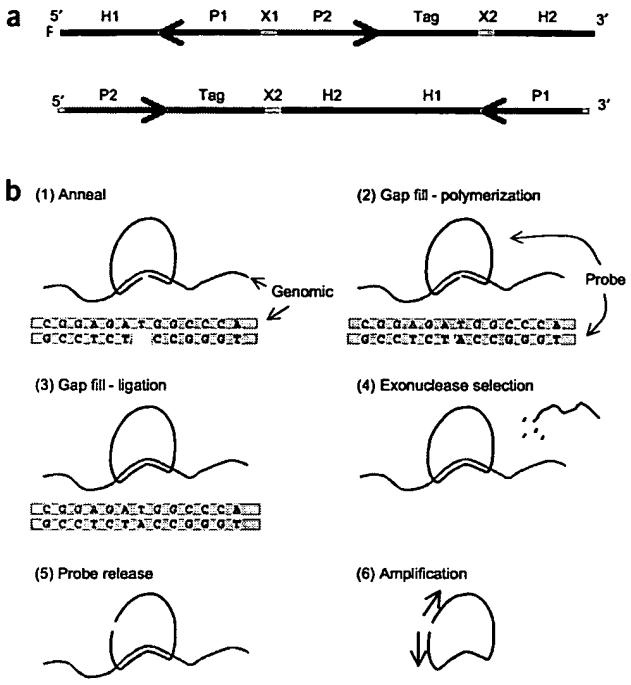


Figure 2 Molecular inversion probes. (a) Unreacted probe (top) and inverted probe (bottom). A single probe is used to detect both alleles of each SNP and consists of seven segments: two regions of homology to target genomic DNA, H1 and H2 (unique to each probe) at the termini of the probe, two PCR primer regions common to all probes, one bar code specific for each locus and two common cleavage sites, X1 and X2. Successfully reacted probes are amplified using primers P1 and P2. A universal detection tag sequence, one of 16,000, is for array detection of amplified probe. Cleavage sites X1 and X2 are used to release the circularized probe from genomic DNA and for post-amplification processing, respectively. (b) Enzymatic probe inversion. (1) A mixture of Genomic DNA, 1,000 or more probes, and thermostable ligase and polymerase is heat-denatured and brought to annealing temperature. Two sequences targeting each terminus of the probe hybridize to complementary sites in the genome, creating a circular conformation with a single-nucleotide gap between the termini of the probe. (2) Unlabeled dATP, dCTP, dGTP or dTTP, respectively, is added to each of the four reactions. In reactions where the added nucleotide is complementary to the single-base gap, DNA polymerase adds the nucleotide and (3) DNA ligase closes the gap to form a covalently closed circular molecule that encircles the genomic strand to which it is hybridized. (4) Exonucleases are added to digest linear probes in reactions where the added nucleotide was not complementary to the gap and excess linear probe in reactions where circular molecules were formed. The reactions are then heated to inactivate the exonucleases. (5) To release probes from genomic DNA, uracil-*N*-glycosylase is added to depurinate the uracil residues in the probes. The mixture is then heated to cleave the molecule at the abasic site and release it from genomic DNA. (6) PCR reagents are added, including a primer pair common to all probes. The reactions are then subjected to thermal cycling, with the result that only probes circularized in the allele-specific gap-fill reaction are amplified.

© 2003 Nature Publishing Group http://www.nature.com/naturebiotechnology

accuracy of 97.8%. The accuracy of the 359 agreeing sequence pairs was therefore 99.95%. In the genotyping data set, 312 full genotypes and 23 half genotypes were identical with the Sanger sequence pairs (643/647 chromosome comparisons), for an accuracy of 99.4%. Similar accuracy was achieved when the data were compared with pyrosequencing[18,19] data on different sets of SNPs. The four discordant genotypes were found in a single locus (probe 2101) in four individuals. Because the sequencing data were nonpolymorphic and subsequent sequencing of these loci with newly generated amplicons was concordant with our genotyping results, it is likely that the original amplification for Sanger sequencing was contaminated with other template DNA or PCR product from another individual.

To measure the reproducibility of MIP genotyping, an individual was assayed three times with 888 markers (5,328 chromosomes). 5,006 of 5,011 chromosome comparisons were concordant (99.9%) (Table 2).

We investigated the effect of increasing the multiplexing level tenfold. The performance of 75 probes either in a 75-probe multiplexed reaction or embedded in a 938-probe multiplex reaction was compared on the same individual's DNA (Table 3). The average call rate in seven repetitions of the same individual for the 75-probe multiplex was 92.6%. Call rate for the same 75 probes in the 938-probe multiplex was 93.4% (average of 25 individuals). The assay conditions were identical in every respect except the number of probes added.

Because DNA array costs represent a substantial fraction of the overall cost of this method, we compared four-chip–one-color detection to two-chip–two-color detection in otherwise identical experiments. The dyes were carboxyfluorescein directly coupled to the labeling oligonucleotide, and phycoerythrin that was coupled to the labeling oligonucleotide via biotin-streptavidin in post–chip hybridization staining[20]. Call rate and SNR in the two-chip–two-color experiment (96.1% and 30, respectively) were very similar to those in the four-chip–one-color experiment (95.8% and 31).

Table 2 Performance metrics of the genotyping assay by molecular inversion probes

Measured parameter	Result
Design rate[a]	84%
Call Rate[b]	95%
Concordance with Sanger sequencing[c]	>99.4%
Repeatability[d]	99.9%
Highest multiplex level	1,517
Average SNR[e]	17
Genomic DNA used / SNP[f]	2 ng

[a]183 of 1,121 probes failed to generate data. [b]An average of 891 of 938 probes called per individual for 25 experiments. [c]Two of 396 chromosomes were discordant with pyrosequencing. [d]5,006 of 5,011 chromosome comparisons were concordant. [e]Average of the ratio of maximum allele signal to maximum non-allele signal of called probes. [f]2 μg genomic DNA used to genotype 1,121 markers per individual.

DISCUSSION

The MIP genotyping method described here has several advantages over alternative techniques. No singleplex PCR amplification is required before mutation detection, thereby reducing labor and expense. PCR is applied only after mutation detection, at which time all molecular inversion probes are converted to standard-length oligonucleotides of similar sequence composition and common primers. This results in a high degree of multiplexing capacity. We have not observed any change in performance in multiplexing from a single probe up to 1,500 probes and speculate that a further increase to 10,000 probes might be possible because sufficient signal is generated in the assay to support that many probes. The data

Alamar Exhibit 1002
Alamar v. Olink

Best Available Copy

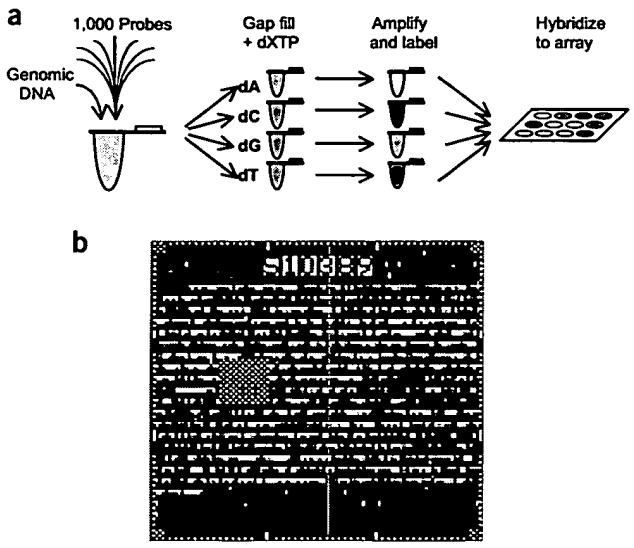


Figure 3 Process flow and array image. (a) Genotyping process flow. 1,000 or more probes are mixed with genomic DNA and gap-fill enzymes (see **Fig. 2**). The reaction is split into four tubes and one of four unmodified nucleotides is added. Reactions are subsequently amplified and a label is added. Reactions are combined and hybridized to the microarray. Relative intensities of two expected allele bases and two background bases indicate genotype and probe performance. (b) Data from 938 amplified probes hybridized to a GenFlex universal DNA array. The relative base incorporation is measured by the fluorescence signals at the corresponding complementary tag site on the DNA array.

© 2003 Nature Publishing Group http://www.nature.com/naturebiotechnology

presented here were generated using four microarrays per sample. Currently we use two microarrays with two-color detection per sample as previously described[20], and we obtain equivalent call rates and SNRs. In theory, genotyping 16,000 markers with this method would require 44 reactions and 2 oligonucleotide arrays (1,500-probe multiplex with 16,000 element Affymetrix Tag 3 array using two-color detection). Thus only a very modest infrastructure is needed to use this approach: a small number of thermocyclers, microarray washing instruments and microarray scanners. This compares very favorably with the robotic infrastructure and detection instrumentation required to set up thousands of PCR reactions and analyze the results.

The intramolecular nature of the MIP genotyping allows higher multiplexing than any other current approach because only the self-self interacting molecules are amplified, while cross-interactions are greatly suppressed. This should allow the current DNA usage of 2 ng per SNP reaction (2 μg/1,000 probes) to be further reduced to 0.2 ng per SNP reaction (2 μg/10,000 probes) as the degree of multiplexing is increased to 10,000 probes.

Several levels of intrinsic specificity are built into this assay. First, the dual recognition sequences at the 3′ and 5′ ends of probes are physically constrained to interact locally. A molecular inversion probe hybridizes much more quickly than two independent probes because the second recognition sequence hybridizes instantaneously after the first. As a result, probe-genomic complexes form at probe concentrations that do not favor nonspecific cross-interactions between probes. Specificity is then increased by the action of the gap-fill enzymes. DNA polymerase selectively extends the correct nucleotide, and DNA ligase ligates only perfectly hybridized DNA. An error requires both misextension and misligation to occur. Probes that have undergone the correct interaction and circle formation are further selected by exonuclease treatment before amplification. Finally, the tag sequences are selected to achieve high hybridization specificity and thereby eliminate cross-talk at the detection step. The synergism of the individually optimized steps comprising the MIP genotyping results in the high degree of multiplexing described here.

An unusual aspect of the approach is the built-in quality control of SNR through monitoring of the background allele channels. Bi-allelic markers such as SNPs have only two possible base alleles. Because this assay monitors all four base possibilities, the SNR is measured with each call and suspect calls can be efficiently discarded.

Molecular inversion requires a single probe per marker, reducing the requirement for probe synthesis. Moreover, any damage or loss of performance of that probe will affect both alleles equally and will therefore not lead to spurious genotypes such as can occur with allele-specific oligonucleotides.

For molecular inversion technology, the rate at which a functional probe is generated from an SNP chosen at random from a database in a single synthesis attempt is 84%. The rate at which all functional probes produce high-quality data over many individuals is 95%. As mapping and cSNP (SNPs that are found in exons) discovery efforts proceed, it will be increasingly important to assay a particular SNP rather than any SNP within a region. This will place increasing emphasis on the ability of a given technology to assay any SNP.

Cost is a fundamental driver for the development of alternative SNP genotyping technologies. There are three main costs associated with SNP genotyping: probe cost, assay cost and detection cost. Although molecular inversion probes are longer than PCR primers, the total number of unique bases that must be synthesized for each probe is comparable to that for a PCR-based approach and much lower than for methods that require allele-specific oligonucleotides, such as the oligonucleotide ligation assay[21]. The locus-specific probes do not require any fluorescent or modified bases and are

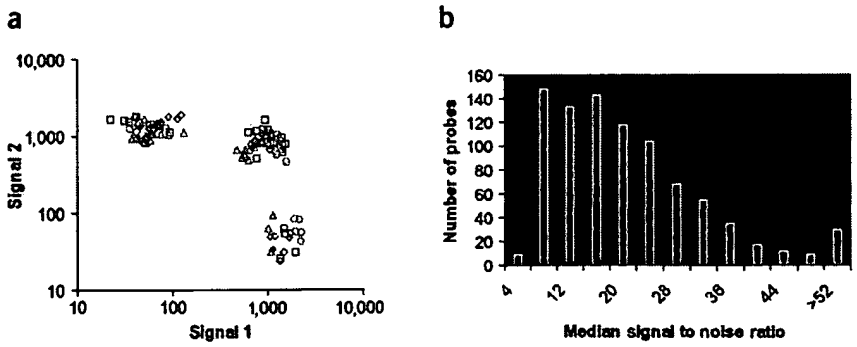


Figure 4 Assay performance. (a) Fluorescence signal from four markers tested on 25 individuals in 32 experiments in a 1,121-probe multiplex assay. Markers 1, 2 and 3 are A/G alleles and marker 4 is a C/G allele. A and C signal is plotted as signal 1 and G signal as signal 2. (b) Median ratio of maximum allele signal to maximum background (non-allele) signal for 938 probes.

Alamar Exhibit 1002
Alamar v. Olink

Table 3 Data on allele type, median call rate and SNR obtained by large-scale genotyping

Allele	No. tested	Average call rate	Median SNR
A/C	77	94%	19
A/G	349	95%	18
A/T	52	94%	22
C/G	90	94%	19
C/T	302	95%	22
G/T	68	94%	21

therefore inexpensive to synthesize. Also, owing to the kinetic advantage of intramolecular interactions, only 12 amol of each probe are used in a single assay. Typical synthesis scales of 1 nmol thus represent millions of assays worth of material. These probes will thus persist as a valuable resource for subsequent genotyping. Assay costs are amortized by the high degree of multiplexing involved, resulting in a very inexpensive assay in the current format (<$0.01 per assay). Detection costs are dominated by the cost of the microarray. To minimize this cost, arrays should be fully occupied and larger arrays used to amortize the cost of the arrays over more genotypes. Such arrays are now available from multiple suppliers either on or off the shelf (GenFlex and Tag3 arrays; Affymetrix) or on a service basis (Agilent, Amersham, NimbleGen), and the cost per array feature ranges from less than $0.01 to $0.10. Taken together, we believe these advantages amount to an enabling advance in genotyping technology that will allow the power of whole-genome analysis to be realized.

METHODS

Oligonucleotide synthesis. Example of a probe constructed for NCBI SNPcluster ID: rs 1389629 (TSC0559094) 5′-TGGATCCCATTATCCTCCATTACGGCTCAACGTTCCTATTCGGTTUUUTTGCAAATGTTATCGAGGTCCGGCACGCACAGGTTATGAATCTCTTTAAACTCCCACAGTGAGGAGC-3′. Molecular inversion probes were ordered from ParAllele BioScience and Eurogentec. Other oligonucleotides were synthesized by Operon, Inc.

Samples. All samples were obtained from the Coriell Institute human variation collection of African American (HD100AA) and Caucasian (HD100CAU) genomic DNA.

Exonucleolysis assay. Circularization of 3 fmol MIP probe A (5′-TGATGGACGTCTGGAAAGCAACCGAAGCTTGTGCGCGTGCAGGGTCACCAGCAGGCATGAGCCCGGTCAACTTCAAGCTCCTAAGCCGGCAGGCAATGCACAGCACCG-3′) was done using as template 6 fmol of the oligonucleotide target 5′-TGCTTTCCAGACGTCCATCACGGTGCTGTGCATTGCCTGC-3′, in 20 mM Tris-HCl pH 8.5, 100 mM KCl, 9 mM $MgCl_2$, 1 mM dithiothreitol, 0.1% Triton X-100 and 4 units Ampligase (Epicentre). Reactions were thermally cycled ten times between 95 °C for 30 s and 58 °C for 1 min. Exonucleolysis was performed by addition of 8 units each of exonuclease I and exonuclease III (New England Biolabs) and incubation at 37 °C for 3 h, followed by denaturation at 95 °C for 10 min. Linear or circularized forms of A were amplified using primers F1 (5′-TGATGGACGTCTGGAAAGCAA-3′) and R1 (5′-CGGTGCTGTGCATTGCCTGCC-3′), or F2 (5′-CACGCGCACAAGCTTCGG-3′) and R2 (5′-CAGGGCACCAGCAGGCA-3′), respectively. Dimerization of the two padlock probes A (sequence above) and B (5′-TGTTCACTGCTGGCCTCCGCAAGCGTGTAGTGTCCGTCGAATAT-TCGTTCTGCAGCATCGCACACAGAAGGTCGATTGCTAGGTGACTGCCACCCAAGGGG-3′) was done using the template oligonucleotide 5′-GCGGAGGCCAGCAGTGAACACGGTGCTGCATTGCCTGC-3′, and the ligation product was amplified with primers F2 and R2. Real-time amplification was done with an ABI Prism 7700 thermocycler (Applied Biosystems) in a total volume of 30 µl containing 2.5 µl buffer A (Applied Biosystems), 2 mM $MgCl_2$, 0.2 mM of each dNTP, 0.3 µM of each primer, 0.8 µM TaqMan probe 5′-FAM-CCCGGTCAACTTCAAGCTCCTAAGCC-TAMRA-3′ and 0.6 units AmpliTaq Gold (Applied Biosystems). The temperature profile was 95 °C for 10 min followed by 50 cycles of denaturation at 95 °C for 30 s and annealing and extension at 58 °C for 45 s.

Molecular inversion probe assay. Denaturing and annealing: four identical reactions containing 400 ng of genomic DNA, 12 amol each of 1,000 probes, 0.0625 units Ampligase (Epicentre) and 0.5 units Stoffel fragment DNA polymerase (Applied Biosystems) in 9 µl of 20 mM Tris-HCl (pH 8.3), 25 mM KCl, 10 mM $MgCl_2$, 0.5 mM NAD and 0.01 % Triton X-100 were incubated for 4 min at 20 °C, 5 min at 95 °C and 15 min at 60 °C. Gap-fill reaction: 1 µl of each of four nucleotides was added to the four reactions and incubated for 10 min at 60 °C and then 1 min at 37 °C. Exonuclease selection: 10 units exonuclease I and 200 units exonuclease III (United States Biochemical) in a 2-µl volume were added and the mixture incubated for 14 min at 37 °C, 2 min at 95 °C and 1 min at 37 °C. Uracil depurination and cleavage: 2 units of uracil-*N*-glycosylase (New England Biolabs) were added in 25 µl of 1.6 mM $MgCl_2$, 10 mM Tris-HCl (pH 8.3), 50 mM KCl and incubated for 9 min at 37 °C and 20 min at 95 °C. Amplification: 2 units of AmpliTaq Gold (Applied Biosystems), 16 pmol primer 1 (5′-CCGAATAGGAACGTTGAGCCGT-3′), and 16 pmol primer 2 (5′-GCAAATGTTATCGAGGTCCGGC-3′) in 25 µl of 1.6 mM $MgCl_2$, 10 mM Tris-HCl (pH 8.3), 50 mM KCl and 112 µm dNTP were preactivated for 10 min at 95 °C and then added to the genotyping reactions. The reactions were amplified in 28 cycles of 95 °C for 20 s, 65 °C for 45 s and 72 °C for 10 s. Sample processing: 20 units of exonuclease I and 10 units *Dra*I (New England Biolabs) were incubated with 60 µl of each amplification product at 37 °C for 1 h and then 80 °C for 30 min.

Microarray hybridization. Approximately 1.25 pmol of each amplified and processed product were hybridized overnight at 39 °C to a GenFlex Tag Array (Affymetrix) DNA array with 55 µl 2× MES, 2.2 µl 50× Denhardt buffer, 110 pmol primer 1 complement (5′-ACGGCTCAACGTTCCTATTCGG-3′), 110 pmol primer 2-FAM (5′-FAM-GCAAATGTTATCGAGGTCCGGC-3′), 1.1 fmol (each) GenFlex control oligonucleotide (Affymetrix). Microarrays were washed and scanned as recommended by the manufacturer (Affymetrix). Data analysis was performed on the raw signal data for each tag feature generated by the Affymetrix image analysis software.

Data analysis. Four signals were generated for each genotype, one for each base reaction. The raw signal was background subtracted and then normalized using the GenFlex hybridization controls that were hybridized in equimolar amounts to each microarray. Base-calls were performed as follows: a small systematic noise was added to all four channels (A,C,G,T). Measured signals were adjusted as follows: $S_{adjusted} = \sqrt{(S_{measured}^2 + (7 \times \sigma)^2)}$, where σ = 3 times the standard deviation of features that were not used in a particular experiment (array noise). The SNR was calculated for each marker by dividing the maximum signal in the two expected allele channels by the maximum signal in the two background channels. The two criteria for calling genotypes were a signal to background ratio of at least 3:1, and the ratio of the higher allele signal to the lower allele signal greater than 6:1 for homozygous calls and less than 2:1 for heterozygous calls.

ACKNOWLEDGEMENTS

We thank K. Tran, T. Brundage, J. Patterson, E. Prakash and C. Bruckner for technical assistance and T. Jones and P. Oefner for helpful discussions. This work was supported by NIH HG00205. The work in Uppsala was supported by the Beijer and Wallenberg Foundations, the Research Councils of Sweden for natural science and for medicine, the Swedish Cancer Fund and Polysaccharide Research AB (Uppsala).

COMPETING INTERESTS STATEMENT

The authors declare competing financial interests (see the *Nature Biotechnology* website for details).

Received 18 October 2002; accepted 21 February 2003
Published online 5 May 2003; doi:10.1038/nbt821

© 2003 Nature Publishing Group http://www.nature.com/naturebiotechnology

Alamar Exhibit 1002
Alamar v. Olink

1. Patil, N. *et al.* Blocks of limited haplotype diversity revealed by high-resolution scanning of human chromosome 21. *Science* **294**, 1719–1723 (2001).
2. Syvanen, A.C. Accessing genetic variation: genotyping single nucleotide polymorphisms. *Nat. Rev. Genet.* **2**, 930–942 (2001).
3. Nilsson, M. *et al.* Making ends meet in genetic analysis using padlock probes. *Hum. Mutat.* **19**, 410–415 (2002).
4. Elnifro, E.M., Ashshi, A.M., Cooper, R.J. & Klapper, P.E. Multiplex PCR: optimization and application in diagnostic virology. *Clin. Microbiol. Rev.* **13**, 559–570 (2000).
5. Lockhart, D.J. *et al.* Expression monitoring by hybridization to high-density oligonucleotide arrays. *Nat. Biotechnol.* **14**, 1675–1680 (1996).
6. Schena, M., Shalon, D., Davis, R.W. & Brown, P.O. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* **270**, 467–470 (1995).
7. Lyamichev, V. *et al.* Polymorphism identification and quantitative detection of genomic DNA by invasive cleavage of oligonucleotide probes. *Nat. Biotechnol.* **17**, 292–296 (1999).
8. Eis, P.S. *et al.* An invasive cleavage assay for direct quantitation of specific RNAs. *Nat. Biotechnol.* **19**, 673–676 (2001).
9. Nilsson, M. *et al.* Padlock probes: circularizing oligonucleotides for localized DNA detection. *Science* **265**, 2085–2088 (1994).
10. Lizardi, P.M. *et al.* Mutation detection and single-molecule counting using isothermal rolling-circle amplification. *Nat. Genet.* **19**, 225–232 (1998).
11. Faruqi, A.F. *et al.* High-throughput genotyping of single nucleotide polymorphisms with rolling circle amplification. *BMC Genomics* **2**, 4 (2001).
12. Antson, D.O., Isaksson, A., Landegren, U. & Nilsson, M. PCR-generated padlock probes detect single nucleotide variation in genomic DNA. *Nucleic Acids Res.* **28**, E58 (2000).
13. Shoemaker, D.D., Lashkari, D.A., Morris, D., Mittmann, M. & Davis, R.W. Quantitative phenotypic analysis of yeast deletion mutants using a highly parallel molecular bar-coding strategy. *Nat. Genet.* **14**, 450–456 (1996).
14. Zhang, D.Y., Brandwein, M., Hsuih, T.C. & Li, H. Amplification of target-specific, ligation-dependent circular probe. *Gene* **211**, 277–285 (1998).
15. Abravaya, K., Carrino, J.J., Muldoon, S. & Lee, H.H. Detection of point mutations with a modified ligase chain reaction (Gap-LCR). *Nucleic Acids Res.* **23**, 675–682 (1995).
16. Banér, J., Nilsson, M., Mendel-Hartvig, M. & Landegren, U. Signal amplification of padlock probes by rolling circle replication. *Nucleic Acids Res.* **26**, 5073–5078 (1998).
17. Gharavi, A.G. *et al.* IgA nephropathy, the most common cause of glomerulonephritis, is linked to 6q22-23. *Nat. Genet.* **26**, 354–357 (2000).
18. Fakhrai-Rad, H., Pourmand, N. & Ronaghi, M. Pyrosequencing: an accurate detection platform for single nucleotide polymorphisms. *Hum. Mutat.* **19**, 479–485 (2002).
19. Ronaghi, M. Pyrosequencing sheds light on DNA sequencing *Genome Res.* **11**, 3–11 (2001).
20. Fan, J.B. *et al.* Parallel genotyping of human SNPs using generic high-density oligonucleotide tag arrays. *Genome Res.* **10**, 853–860 (2000).
21. Landegren, U., Kaiser, R., Sanders, J. & Hood, L. A ligase-mediated gene detection technique. *Science* **241**, 1077–1080 (1988).

© 2003 Nature Publishing Group http://www.nature.com/naturebiotechnology

Alamar Exhibit 1002
Alamar v. Olink

ELSEVIER

Genetic Analysis: Biomolecular Engineering
15 (1999) 35–40

GENETIC ANALYSIS
Biomolecular Engineering

Rolling circle amplification of DNA immobilized on solid surfaces and its application to multiplex mutation detection

Anson Hatch *, Takeshi Sano, John Misasi, Cassandra L. Smith

Center for Advanced Biotechnology and Departments of Biomedical Engineering and Biology, Boston University, 36 Cummington Street, Boston, MA 02215, USA

Accepted 7 August 1998

Abstract

A new method of amplifying short DNA molecules immobilized on a solid support has been developed. This method uses a solid-phase rolling circle replication reaction, termed rolling circle amplification (RCA). The probe consists of a single-stranded DNA primer anchored at the 5′ terminus to a solid support and a single stranded DNA template hybridized to the immobilized primer. Here, DNA ligase was used to circularize the template, and DNA polymerase I was used to extend the immobilized primer in a rolling circle replication reaction. This method was used to identify a known polymorphism in BRCA1 exon 5. These results demonstrate that RCA offers considerable promise to facilitate effective mutation screening of DNA using a solid-phase format.
© 1999 Elsevier Science B.V. All rights reserved.

Keywords: Rolling circle amplification; DNA; Multiplex mutation detection

1. Introduction

The necessity of examining stretches of DNA for polymorphisms has rapidly increased with the rate of discovery of diseased genes. Due to the large number of mutations possible in genomic DNA and their diverse relationship to disease, a rapid, comprehensive, and cost-effective method of scanning these genes is essential for diagnostic and research purposes. A reduction in cost two orders of magnitude below the cost of direct sequencing may be necessary for effective genomic screening [1]. Many methods have been developed to screen DNA for mutations, which can be divided into two general categories; solid-phase methods using probes bound to a solid support, and methods requiring size fractionation by electrophoresis and other separation methods.

Examples of methods using size fractionation include single-strand conformation polymorphisms (SSCP) [2], denaturing gradient gel electrophoresis (DGGE) [3], combined chain reactions (CCR) [4], and enzyme mismatch cleavage [5,6]. Electrophoretic methods are commonly used to screen genes for mutations because of their ease of use and simplicity. An enzyme or chemical reaction is used to specifically identify a mutation followed by gel electrophoresis to analyze the results. The use of electrophoresis requires a large amount of DNA for analysis, usually obtained through PCR, and requires great effort to process large numbers of samples. These methods can be anywhere from 80 to 100% accurate in detecting a mutation, but they are less than an order of magnitude below the cost of direct sequencing. Electrophoretic separation may never reach the accuracy and cost requirements necessary for effective genomic screening.

Solid-phase scanning of genomic DNA, though complex, holds more promise in reducing cost and providing complete accuracy. It is not clear that any

* Corresponding author.

1050-3862/99/$ - see front matter © 1999 Elsevier Science B.V. All rights reserved.
PII S1050-3862(98)00014-X

Alamar Exhibit 1002
Alamar v. Olink

solid-phase method yet provides the cost-savings and accuracy necessary for effective genomic screening [1]. Solid-phase methods include: sequencing by hybridization (SBH) [7–11], trapped oligonucleotide nucleotide incorporation (TONI) [12], multiplex allele-specific diagnostic assay (MASDA) [13], and solid-phase chemical cleavage (SPCC) [14]. Solid-phase approaches allow the screening of large numbers of DNA sequences in parallel, which can significantly reduce the time and materials required for detecting mutations. These methods may also be automated, which could further reduce cost and increase throughput. Current solid-phase methods are limited by specificity. Conventional SBH is not highly accurate (based on physical hybridization alone). Additional refinements such as competition strategies [9,10], excessive quantities of probes and/or target [10,15], and the use of enzymes [7,10,15,16], are necessary to increase the specificity of a solid-phase method.

Another limiting factor in solid-phase screening is signal intensity. Solid-phase methods currently rely on the use of large numbers of probes isolated at different locations on a solid support as well as high concentrations of target material, which is inefficient and costly. PCR amplification can increase a small number of probes or targets by several orders of magnitude in solution, but PCR is difficult to incorporate into a solid-phase approach because the amplified DNA is not immobilized on the surface [17]. A strategy that amplifies the signal from small numbers of initial probes could significantly increase the effectiveness of solid-phase methodology.

Here, a new method of amplifying short DNA molecules immobilized to a solid support is described, termed rolling circle amplification (RCA). RCA targets specific DNA sequences through hybridization and a DNA ligase reaction. The circular product is then used as a template in a rolling circle replication reaction.

2. Materials and methods

2.1. Oligonucleotides

Oligonucleotides (Table 1) were purchased from Operon Technologies (Alameda, CA). DNA was labeled using T4 polynucleotide kinase. A 10-pmol DNA template was labeled with ^{32}P in a 50-μl volume containing 70 mM Tris–HCl (ph 7.6)/10 mM $MgCl_2$/5 mM dithiothreitol/0.2 μM [γ-^{32}P]ATP (0.8 mCi/mmol, 1 Ci = 37 GBq; Amersham). The 5′ terminus of the DNA template was phosporylated by adding 10 U of T4 polynucleotide kinase (New England Biolabs) and incubating at 37°C for 3 h.

2.2. Immobilization of DNA probe complex

A mixture of a 5′ biotinylated priming oligonucleotide (5 pmol) and template oligonucleotide (10 pmol) containing 22 bases complementary to the 3′ terminus of the priming oligonucleotide (Figs. 1 and 2) was heated to 95°C and slowly cooled to room temperature to allow formation of a double stranded complex. The probes were immobilized on streptavidin coated paramagnetic beads (0.25 mg, Dynal, Oslo). The beads were washed three times with 50 μl of TE (10 mM Tris–HCl, pH 8/1 mM EDTA (ethylene diamine tetraacetic acid)) to remove unbound probes and finally suspended in 50 μl of TE. Binding of the probe complex using 5′ labeled template was evaluated by Cerenkov counting. Counts of the initial supernatant containing unbound template were compared to counts of washed beads containing immobilized probes to assess binding efficiency.

2.3. Ligation of immobilized probes

Two target sequences of BRCA1 from exon 5 were used. One target (N) contained the normal sequence and a second target (M) contained a G for T substitution in codon 64 (Table 1). The template sequence contained ten bases at the 5′ terminus and ten bases at the 3′ terminus complementarily juxtaposed to the normal sequence of BRCA1 exon 5 (Fig. 2). The 5′ terminus of the template DNA was complementary to the position of the point mutation in codon 64. Beads with 5 pmol immobilized probes were washed with 50 μl of 1 × T4 DNA ligase buffer containing 50 mM Tris–HCl (pH 7.5)/10 mM $MgCl_2$/10 mM dithiothreitol/1 mM ATP/25 μg/ml bovine serum albumin. Beads and immobilized probes were suspended in 50 μl of 1 × T4

Table 1
Oligonucleotides

Name	Sequence
Primer	5′-b-$(T)_{20}$GCGGCCGC$(T)_{15}$-CGCCAGGGTTTTCCCAGTCACGAC-3′
Template	5′-ATAAAGGACATT-GAATTCGTCGTGACTGGGAAAACC-CTGGGATCCTTTCATTCTTAC-3′
Target N	5′-CACAGTGTCCTTTATGTAAGAAT-GATATAAC-3′
Target M	5′-CACAGTGTCCTTTAGGTAAGAAT-GATATAAC-3′

The primer oligonucleotide has a biotinylated 5′ terminus. The template oligonucleotide contains 22 central bases (underlined) complementarily to the 3′ terminus of the primer sequence. The ten (underlined) 5′ and 3′ terminal bases of the template DNA are complementary juxtaposed to the normal sequence of BRCA1 exon 5. Target N contains the normal sequence of BRCA1 exon 5 and target M contains a G for T substitution indicated in boldface type.

Alamar Exhibit 1002
Alamar v. Olink

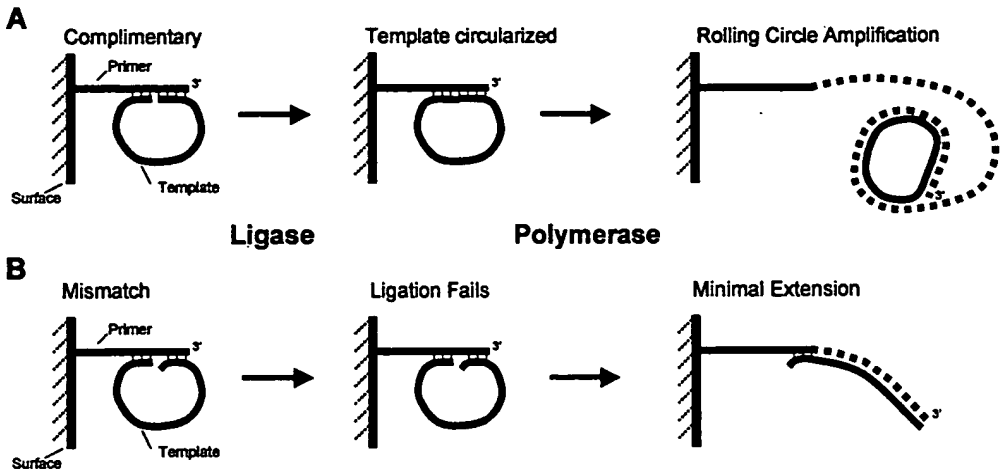


Fig. 1. Solid-phase rolling circle amplification. (A) Template is perfectly complementary at the 3′ and 5′ termini to the primer sequence. Ligase circularizes the template, and polymerase enzyme extends the primer in a rolling circle replication reaction. (B) The template strand is not perfectly complementary, at the 3′ and 5′ termini, to the DNA primer sequence. Ligation fails, and polymerase extends the primer to complement the linear template.

DNA ligase buffer containing 10 pmol target DNA. Mixtures were incubated at room temperature for 15 min with 40 U of T4 DNA ligase (New England Biolabs). Ligation reactions were terminated by adding 1 μl of 0.5 M EDTA. The beads were washed three times with 50 μl of TE and finally suspended in 50 μl of TE. Samples were incubated with 1000-fold excess biotin to binding capacity and heated to 75°C for 30 min to displace the biotinylated product from the beads. The displaced product was fractionated using denaturing polyacrylamide gel electrophoresis and the results quantified using the Storm 860 phosphoimager (Molecular Dynamics).

2.4. Rolling circle amplification of immobilized probes

Beads containing product of the ligation reaction were washed once with 50 μl of 1 × extension buffer (10 mM Tris–HCl (pH 7.5)/5 mM $MgCl_2$/7.5 mM dithiothreitol) and suspended in a final volume of 50 μl

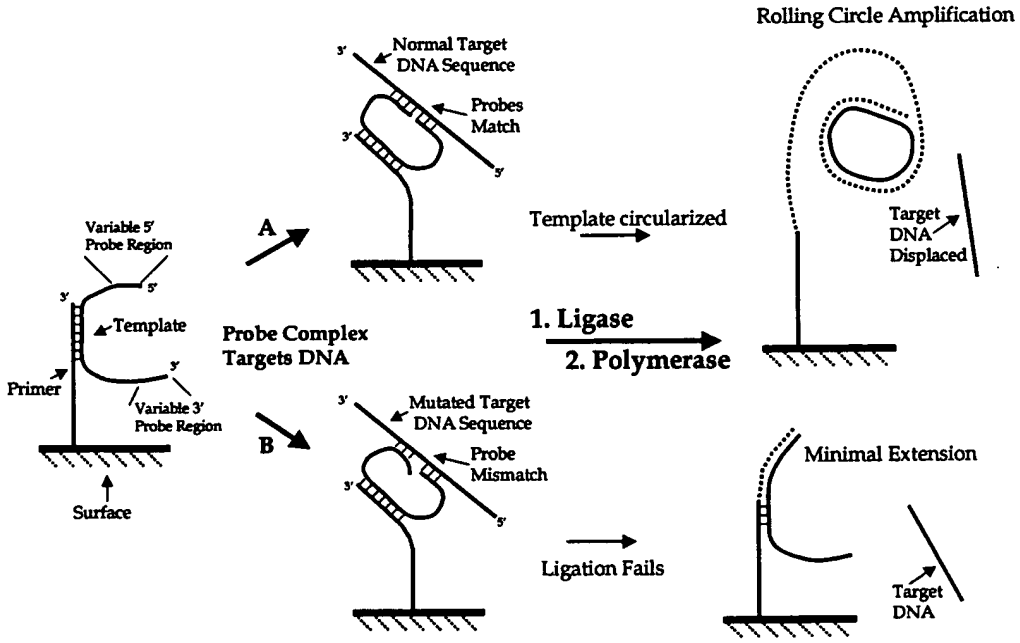
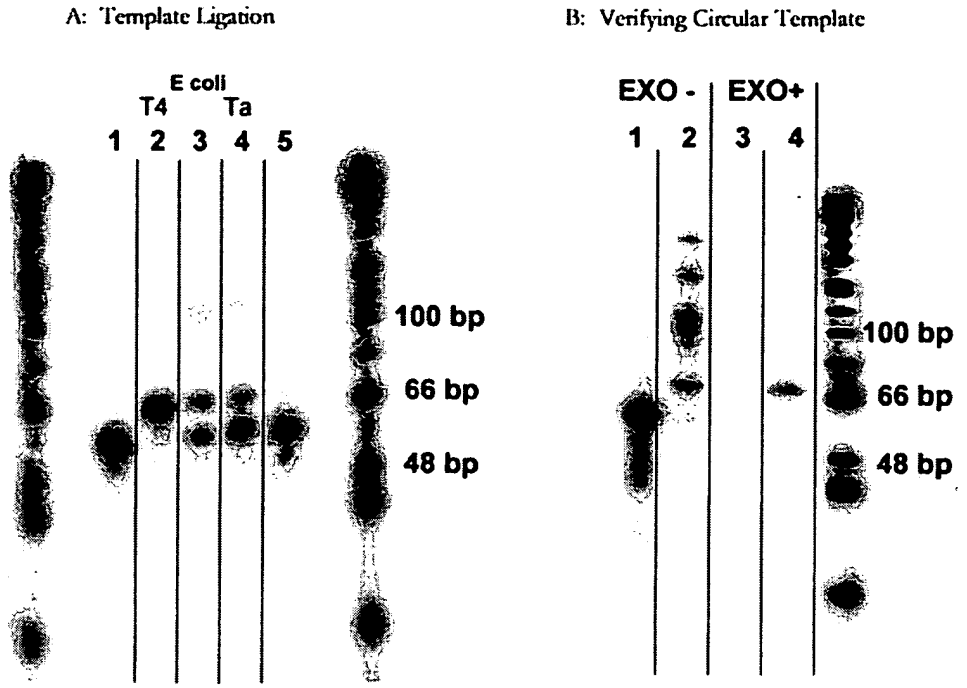


Fig. 2. Applying RCA to probe for polymorphisms. (A) Template is perfectly complementary at the 3′ and 5′ termini to the normal target DNA sequence. Ligase circularizes the template, and polymerase enzyme extends the primer in a rolling circle replication reaction. (B) The template strand is not perfectly complementary at the 3′ and 5′ termini to the mutated target DNA sequence. Ligation fails, and polymerase extends the primer complementary to the linear template.

Alamar Exhibit 1002
Alamar v. Olink



Best Available Copy

Fig. 3. Ligation of template DNA. (A) Linear 57-base DNA template (1,5) is circularized by ligase (2, T4; 3, *E. coli*, 4, Taq). Circular template is observed as a band shifted upward closer to the 66-base size standard. (B) Template circularization is verified by treating linear template (1) and ligated product under conditions generating several bands of product (2) with exonuclease (3, treated template; 4, treated ligation product).

1 × extension buffer containing 50 μM dATP/50 μM dTTP/50 μM dGTP/50 μM dCTP/2 μM [α-^{32}P]dCTP (0.8 mCi/mmol, 1 Ci = 37 GBq; Amersham). The 3′ terminus of the priming strand was extended by adding 2.5 U of DNA polymerase I *Escherichia coli* (Boehringer Mannheim) to the 50-μl reaction and incubating at 37°C for 3 h while mixing. Reactions were terminated by adding 1 μl of 0.5 M EDTA. The beads with bound product were washed once with 50 μl of TE and then transferred to another Eppendorf tube. The beads were then washed four times with 50 μl of TE to remove any unincorporated [α-^{32}P]dCTP. The incorporated radioactivity was measured by Cerenkov counting. Product was displaced from the beads by incubation with 1000-fold excess biotin to binding capacity and heating to 75°C for 30 min. The displaced product was analyzed as described above.

3. Results

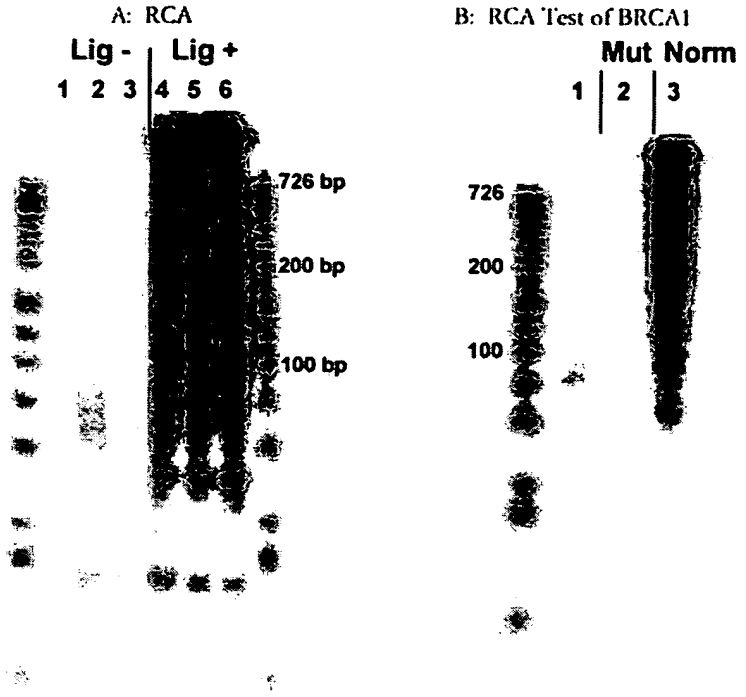
3.1. Principles of RCA

Solid-phase RCA was developed to provide an effective method of detecting constituents of a solution. A simple scheme outlining the use of rolling circle amplification is shown in Fig. 1. A recognition step is used to generate a complex consisting of a DNA primer duplexed with a circular template bound to a surface. A polymerase enzyme is then used to amplify the bound complex. RCA uses small DNA probes that are amplified to provide an intense signal.

These principles were used to design a detection system to recognize polymorphisms in DNA (Fig. 2). The probes consist of a DNA primer anchored at the 5′ terminus to a solid support and a DNA template hybridized to the primer (Fig. 2A). The 3′ and 5′ ends of the template DNA anneals juxtaposed to the target DNA (Fig. 2B). This provides discrimination of target DNA with and without mismatches. DNA ligase is then used to circularize a matched DNA template perfectly duplexed with the target DNA. The ligation provides enhanced discrimination of perfectly matched target DNAs from those containing mismatches, because DNA ligase is sensitive to single-base mismatches near the ligation site [18–20].

DNA polymerase can then extend the anchored primer in a rolling circle replication type reaction (Fig. 2C). Extension of DNA in a rolling circle replication process has previously been achieved in solution using a short DNA primer and a small circular DNA template [21,22]. This type of DNA amplification can significantly extend the primer, generating multiple tandem repeat copies of the circular DNA template. However, if the target DNA contains a mutation in the probed region, ligation fails, and minimal extension of the primer would occur.

Alamar Exhibit 1002
Alamar v. Olink



Best Available Copy

Fig. 4. Rolling circle amplification product. (A) RCA of DNA template directly ligated to a primer immobilized to streptavidin coated beads. Ligated template (4–6) allowed rolling circle replication generating long strands seen by incorporation of ^{32}P-labeled dNTPs. (B) RCA test for a targeted mutation in the sequence of BRCA1 (codon 64/G for T substitution). Conditions containing the normal target sequence (3) were significantly amplified. Conditions containing the mutated target sequence (2) were minimally amplified. Control containing normal target not exposed to ligase enzyme (1) was also minimally amplified.

3.2. Immobilization of RCA probes

Close to 100% of the DNA probes can be captured to the magnetic beads in the form of primer/template probe complex, determined by Cerenkov counts of the immobilized product. The template strand was in a 2:1 ratio to the primer and 49% of the labeled template was bound to the washed magnetic beads. This information is useful for interpreting the amplification system in later steps.

3.3. Ligation of template DNA

A known mutation in the sequence of *BRCA1* gene (exon 5/codon 64/G for T substitution) [23] was targeted as a test of RCA. In these model experiments streptavidin coated magnetic beads were used as the solid support because the beads can be easily manipulated and retain the biotinylated probe complex even under harsh conditions. The template DNA was radiolabeled at the 5′ terminus and hybridized to the target DNA. T4 DNA ligase was found to efficiently (> 70%) circularize a 57-base DNA template bound to a 67-base anchored DNA primer, when a 40-base target N (normal BRCA1 sequence) was introduced. T4 DNA ligase was not efficient (< 2%) in ligating probes which were exposed to target M, containing a point mutation. Ligation of the template was observed as a shift in the electrophoresis pattern of radio-labeled template following ligation (Figs. 3 and 4A). The shifted band was verified to contain circular DNA by exposing the product to exonuclease, which degraded linear template and not the shifted band (Fig. 4B).

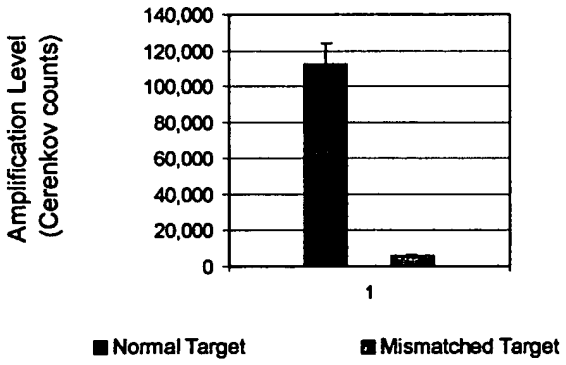


Fig. 5. Test of a BRCA1 exon 5 mutation using RCA. Amplification level of RCA to detect the presence of a mutation in BRCA1. Conditions with the normal target DNA strand were amplified 19-fold greater than conditions with target containing a single base substitution. Measurements were based on Cerenkov counts of [α-^{32}P]dCTP incorporated during the RCA extension reaction.

Alamar Exhibit 1002
Alamar v. Olink

3.4. Extension of immobilized DNA primer

The amount of ^{32}P-labeled nucleotides incorporated into the extension product was evaluated to quantify the level of amplification. DNA polymerase I amplified probe complexes containing ligated and non-ligated template. However, the bound extension products of DNA probes exposed to the target N, the perfectly matched target, were amplified 19 times greater than probes exposed to target M, containing a base mismatch (Fig. 5). Gel analysis of displaced extension product demonstrated that rolling circle replication generates strands of DNA much longer than the original primer anchored to the solid support (Fig. 3).

4. Discussion

RCA amplifies the solid-phase signal using small initial DNA probes immobilized to the solid support. Since the probes are fixed to a solid support, different groups of probes can be fixed at known locations on a surface to generate a testing array capable of screening many unique sequences in parallel. This technology could be highly modular if unique primers are bound to each array location, allowing the addition of different sets of complimentary template probes.

RCA technology is versatile enough to make use of many different methods of DNA recognition to detect the amplification level of the immobilized DNA probes. Although incorporation of radio-labeled nucleotides was used to detect signal amplification in initial testing, other approaches could be used, including direct incorporation of fluorescent-labeled nucleotides in the amplification process, or hybridization of probes to the amplified product.

RCA combines the specificity of hybridization and ligation with the amplification possible through rolling circle replication to achieve an effective solid-phase method of screening target DNA. RCA can offer several advantages for mutation detection. It provides highly specific DNA sequence recognition and intense signal amplification from small starting probes. Using RCA, highly accurate parallel screening of DNA for many mutations may be possible.

Acknowledgements

This work was supported by grant DAMD17-97-I-7310 from the DOA to C.L.S.

References

[1] Cotton RGH. Detection of unknown mutations in DNA: a Catch-22. Am J Hum Genet 1996;59:289–91.
[2] Chen X, Baumstark T, Steger G, Riesner D. High resolution SSCP by optimization of the temperature by transverse TGGE. Nucleic Acids Res 1995;23:4524–5.
[3] Nissen H, et al. Diagnostic strategy, genetic diagnosis and identification of new mutations in intermittent porphyria by denaturing gradient gel electrophoresis. Hum Mutat 1997;9(2):122–30.
[4] Bi W, Stambrook PJ. CCR: a rapid and simple approach for mutation detection. Nucleic Acids Res 1997;25(14):2949–51.
[5] Lu A, Hsu I. Detection of single DNA base mutations with mismatch repair enzymes. Genomics 1992;14:249–55.
[6] Youil R, Kemper BW, Cotton RGH. Screening for mutations by enzyme mismatch cleavage with T4 endonuclease VII. Proc Natl Acad Sci USA 1995;92:87–91.
[7] Broude NE, Sano T, Smith CL, Cantor CR. Enhanced DNA sequencing by hybridization. Proc Natl Acad Sci USA 1994;91: 3072–6.
[8] Cronin MT, et al. Cystic fibrosis mutation detection by hybridization to light-generated DNA probe arrays. Hum Mutat 1996;7: 244–55.
[9] Hacia JG, Brody LC, Chee MS, Fodor SPA, Collins FS. Detection of heterozygous mutations in BRCA1 using high density oligonucleotide arrays and two-colour fluorescence analysis. Nat Genet 1996;14:441–7.
[10] Kozal MJ, et al. Extensive polymorphisms observed in HIV-1 clade B protease gene using high-density oligonucleotide arrays. Nat Med 1996;2(7):753–9.
[11] Lockhart DJ, et al. Expression monitoring by hybridization to high-density oligonucleotide arrays. Nat Biotechnol 1996;14: 1675–80.
[12] Prezant TR, Fischel-Ghodsian N. Trapped-oligonucleotide nucleotide incorporation (TONI) assay, a simple method for screening point mutations. Hum Mutat 1992;1:159–64.
[13] Shuber AP, et al. High throughput parallel analysis of hundreds of patient samples for more than 100 mutations in multiple disease genes. Hum Mol Genet 1997;6(3):337–47.
[14] Hansen LL, Justesen J, Kruse TA. Sensitive and fast mutation detection by solid-phase chemical cleavage. Hum Mutat 1996;7: 256–63.
[15] Syvanen A, Landegren U. Detection of point mutations by solid-phase methods. Hum Mutat 1994;3:172–9.
[16] Shumaker JM, Metspalu S, Caskey CT. Mutation detection by solid-phase primer extension. Hum Mutat 1996;7:346–54.
[17] Khudyakov YE, Gaur L, Singh J, Patel P, Fields HA. Primer specific solid-phase detection of PCR products. Nucleic Acids Res 1994;22(7):1320–1.
[18] Landegren U, Kaiser R, Sanders J, Hood L. A ligase-mediated gene detection technique. Science 1988;241:1077–80.
[19] Barany F. Genetic disease detection and DNA amplification using cloned thermostable ligase. Proc Nat Acad Sci USA 1991;88:189–93.
[20] Nilsson M, Malmgren H, Samiotaki M, Kwiatkowski M, Chowdhary BP, Landegren U. Padlock probes: circularizing oligonucleotides for localized DNA detection. Science 1994;265:2085–8.
[21] Fire A, Xu S. Rolling replication of short DNA circles. Proc Natl Acad Sci USA 1995;92:4641–5.
[22] Liu D, Daubendiek SL, Zillman MA, Ryan K, Kool ET. Rolling circle DNA synthesis: small circular oligonucleotides as efficient templates for DNA polymerases. J Am Chem Soc 1996;118:1587–94.
[23] Castilla LH, et al. Mutations in the *BRCA1* gene in families with early-onset breast and ovarian cancer. Nat Genet 1994;8(4):387–91.

Alamar Exhibit 1002
Alamar v. Olink

Best Available Copy

Reaction discovery enabled by DNA-templated synthesis and *in vitro* selection

Matthew W. Kanan, Mary M. Rozenman, Kaori Sakurai, Thomas M. Snyder & David R. Liu

Department of Chemistry and Chemical Biology, Harvard University, 12 Oxford Street, Cambridge, Massachusetts 02138, USA

Current approaches to reaction discovery focus on one particular transformation. Typically, researchers choose substrates based on their predicted ability to serve as precursors for the target structure, then evaluate reaction conditions[1–6] for their ability to effect product formation. This approach is ideal for addressing specific reactivity problems, but its focused nature might leave many areas of chemical reactivity unexplored. Here we report a reaction discovery approach that uses DNA-templated organic synthesis[7–10] and *in vitro* selection to simultaneously evaluate many combinations of different substrates for bond-forming reactions in a single solution. Watson–Crick base pairing controls the effective molarities of substrates tethered to DNA strands; bond-forming substrate combinations are then revealed using *in vitro* selection for bond formation, PCR amplification and DNA microarray analysis. Using this approach, we discovered an efficient and mild carbon–carbon bond-forming reaction that generates an enone from an alkyne and alkene using an inorganic palladium catalyst. Although this approach is restricted to conditions and catalysts that are at least partially compatible with DNA, we expect that its versatility and efficiency will enable the discovery of additional reactions between a wide range of substrates.

A reaction discovery system capable of simultaneously evaluating many combinations of substrates for bond-forming reactivity in a single solution must meet several requirements. First, the system must organize complex substrate mixtures into discrete pairs that can react (or not react) without affecting the reactivity of the other substrate pairs. Second, the system requires a general method for separating reactive substrate pairs from unreactive pairs. Last, the reactive substrate pairs must be identified efficiently.

Recent developments in DNA-templated organic synthesis[7–10] indicate that DNA annealing can organize many substrates in a single solution into DNA sequence-programmed pairs. To this end, we prepared two pools of DNA-linked substrates, with n substrates in pool A and m substrates in pool B. Each substrate in pool A is covalently linked to the 5′ end of a set of DNA oligonucleotides containing one 'coding region' (uniquely identifying that substrate) and one of m different 'annealing regions' (Fig. 1a). Each of the m substrates in pool B is attached to the 3′ end of an oligonucleotide containing a coding region that uniquely identifies the substrate and complements one of the m annealing regions in pool A.

When pools A and B are combined in a single aqueous solution at nanomolar concentrations, Watson–Crick base pairing organizes the mixture into $n \times m$ discrete pairs of substrates attached to complementary sequences. Only substrates linked to complementary oligonucleotides experience effective molarities in the millimolar range[8]; substrates linked to non-complementary oligonucleotides experience nanomolar solution concentrations and hence do not react with each other at a significant rate. This effective molarity-based design enables reactions that might otherwise be suppressed by the preferential dimerization of one or both substrates. The possibility of interference by the structure of DNA during reactions between substrates is minimized by using long and flexible substrate–DNA linkers[11].

The separation of reactive pairs of substrates is based on selection concepts used in the directed evolution of catalytic RNA and DNA[12,13] (Fig. 1b). We covalently linked each pool B substrate to its corresponding oligonucleotide by a linker containing a biotin group and a disulphide bond (Fig. 1b and Supplementary Information). After incubation under a set of chosen reaction conditions, followed by cleavage of the disulphide bonds, only pool A sequences encoding bond formation between a pool A and pool B substrate remain covalently linked to biotin. Streptavidin affinity selection of the resulting solution separates biotinylated from non-biotinylated sequences. In contrast to many existing reaction discovery screens[1–5] focused on a particular reaction type, our selection-based approach does not depend on any specific substrate or product property but instead can identify all substrate pairs capable of forming a covalent bond under the reaction conditions.

Reactive substrate pairs are identified by amplification of the selected sequences by polymerase chain reaction (PCR) followed by DNA microarray analysis (see below). Because PCR amplification is extremely sensitive, femtomole quantities of substrates are sufficient for the entire reaction discovery process.

We prepared pool A and pool B containing 12 substrates each (Fig. 2a), representing 144 heterocoupling combinations (see Supplementary Methods and Supplementary Fig. S1 for details). We also prepared DNA-linked substrates to enable the detection of homocoupling of any of the 24 different substrates, bringing the total number of unique substrate combinations to 168 (Fig. 2a). The 24 substrates in Fig. 2a were chosen to represent simple functional groups commonly encountered in organic molecules. New reactions forming bonds between simple functionalities are of special interest because they might provide more accessible alternatives to coupling reactions that require more complex substrates. Although our approach requires the preparation of DNA-linked substrates, a single nanomole-scale preparation of each pool provided sufficient material for more than 1,000 reaction discovery experiments that could collectively evaluate more than 168,000 combinations of substrates and reaction conditions.

To test the ability of our system to detect a single reactive combination of substrates out of 168 possibilities, we combined pool A and pool B in the presence of Cu(I), conditions known to promote a cycloaddition between a terminal alkyne (A5 in Fig. 2a) and an azide (B9 in Fig. 2a)[14]. Pool A members (2 pmol total) were combined with 2 pmol total of pool B members (12 fmol of material encoding each possible substrate combination) in the presence of 500 μM Cu(I) in a total volume of 12 μl. After 10 min at 25 °C, the salts were removed and disulphide linkages were cleaved with 0.1 M tris-carboxyethylphosphine hydrochloride. A control experiment lacking Cu(I) but identical in all other respects was also performed.

Sequences encoding bond-forming substrate pairs were captured with streptavidin-linked magnetic particles and amplified by PCR with a DNA primer labelled with the cyanine fluorophore Cy3. For comparison, an aliquot of the pool A sequences before selection was amplified by PCR with a Cy5-labelled primer. The Cy3-labelled and Cy5-labelled PCR products were combined and hybridized to a DNA microarray containing all 168 possible reaction-encoding sequences. The ratios of Cy3 (green) to Cy5 (red) fluorescence for all array locations were calculated and ordered by rank, and spots with green/red fluorescence ratios significantly higher than the majority of spots (in the experiments below, ratios above 1.5) were considered to be positive. A pre-quantified internal standard (bottom right corner of the array images in Fig. 2b) corresponding to a moderate level of reactivity was used as a positive control and as a reference for comparing different arrays.

For the experiment performed in the presence of Cu(I), the array spot corresponding to the combination of the alkyne (A5) and the azide (B9) was the sole spot that had a significant green/red fluorescence ratio (A5 + B9 = 8.5, standard = 2.8; see Fig. 2b and Supplementary Data). In contrast, the experiment performed

©2004 Nature Publishing Group

Alamar Exhibit 1002
Alamar v. Olink

in the absence of any added metal yielded no green array spots other than the standard (Fig. 2b). These results show that the above method can detect a single bond-forming substrate combination from a complex mixture of 168 combinations, and confirm the high chemoselectivity characteristic of the Sharpless-modified Huisgen cycloaddition reaction[15].

Cu(I) damages DNA by promoting radical-mediated processes[16]. Indeed, we observed 48% degradation of a DNA oligonucleotide exposed for 10 min to the Cu(I)-containing conditions (Supplementary Data). The successful detection of the alkyne–azide cycloaddition under conditions that degrade DNA shows that reaction conditions do not need to be fully DNA compatible, because the extreme sensitivity of PCR amplification[17] enables bond formation to be revealed even if a significant fraction of the total DNA in an experiment is destroyed.

To validate this reaction discovery system further, we performed a selection for bond formation after exposure to organic reagents (1-(3-(dimethylamino)propyl)-3-ethylcarbodiimide hydrochloride and *N*-hydroxysuccinimide at pH 6.0) known to promote the coupling of amines with carboxylic acids[18]. Microarray analysis revealed one positive; this spot indeed corresponded to the combination of the carboxylic acid from pool A (A10) and the amine from pool B (B12) (A10 + B12 = 6.9, standard = 15.6; see Fig. 2b and Supplementary Data).

After successfully 'rediscovering' known bond-forming reactions, we examined the reactivity of a simple Pd(II) salt. Inorganic Pd salts and complexes containing Pd are known to mediate diverse coupling reactions[19], indicating that Pd can activate a variety of organic functional groups. Selection for bond formation in the presence of 500 μM Na_2PdCl_4 for 1 h at 37 °C resulted in five strong positives (A7+B3, A5+B3, A4+B8, A5+B5 and A5 homocoupling) with significant green/red fluorescence ratios (3.6 to 2.6; standard = 2.9). In addition, we observed five weaker positives (A9+B3, A8+B3, A8+B8, A5+B8 and A5+B9) with lower, but possibly significant, green/red fluorescence ratios ranging from 1.9 to 1.6 (Fig. 2b and Supplementary Data). One positive representing the combination of aryl iodide (A7) and acrylamide (B3) is consistent with the well-known Heck reaction[20,21] (assuming the formation of some Pd(0) under the reaction conditions). Other positives could be rationalized with mechanistic steps precedented in known Pd-mediated chemistries. For example, the spot indicating bond formation between an olefin (A4) and an aryl boronic acid (B8) is consistent with transmetallation of Pd(II) by the aryl boronic acid[22] followed by the insertion of an olefin into the Pd–aryl bond and subsequent β-hydride elimination[23].

To determine whether these array results reflect genuine bond-forming events, we examined the putative reactions corresponding to the above ten spots in separate DNA-templated reactions. Denaturing polyacrylamide gel electrophoresis (PAGE) analysis indicated that all five strong positives and three of the five weak positives corresponded to authentic DNA-templated reactions, whereas two weak positives (A9+B3 and A5+B9) showed little or no product formation (Fig. 3 and Supplementary Fig. S2). We observed no product formation in control reactions in which Na_2PdCl_4 was omitted, or in which either of the reactive substrates was replaced with an unreactive alkane group. These findings are

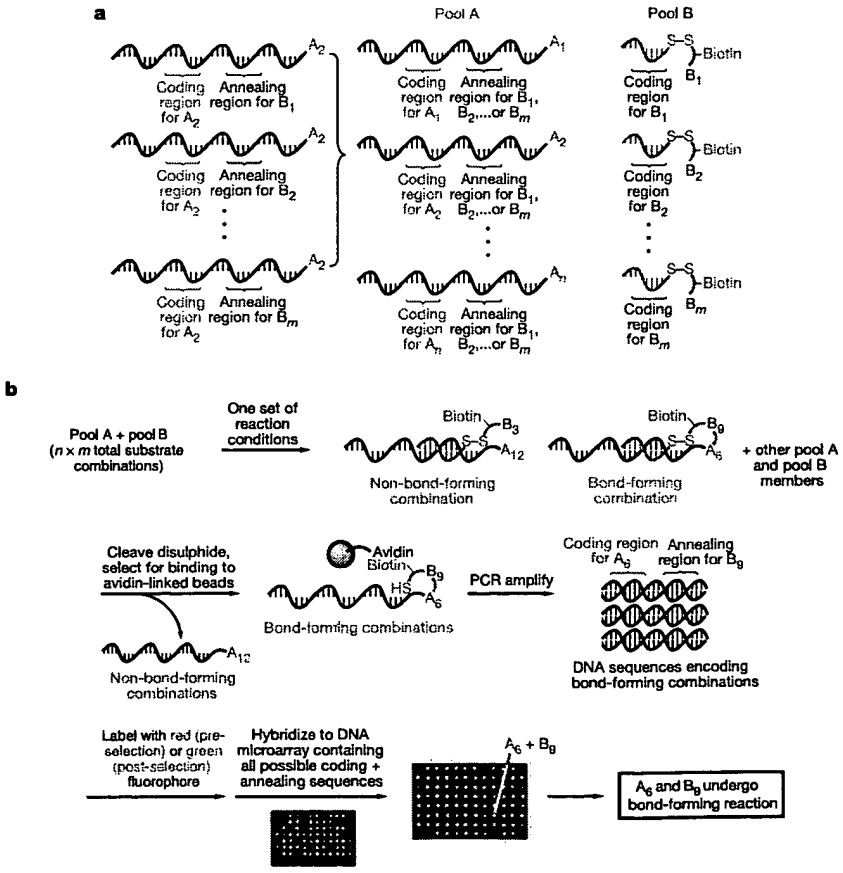


Figure 1 Key elements of a new approach to reaction discovery. **a**, Two pools of DNA-linked organic functional groups that associate each of $n \times m$ substrate combinations with a unique DNA sequence. **b**, A general one-pot selection and analysis method for the detection of bond-forming reactions between DNA-linked substrates.

©2004 Nature Publishing Group

Alamar Exhibit 1002
Alamar v. Olink

consistent with the detection of Pd-dependent reactions that couple substrate groups and do not couple the functionality present in DNA.

We further characterized the PAGE-validated Pd-mediated DNA-templated reactions using matrix assisted laser desorption ionization–time-of-flight (MALDI–TOF) mass spectrometry (Fig. 3). The observed product masses are consistent with covalent bond formation. We suggest possible structures consistent with the observed masses for each of the reaction products but note that other products or mixtures of products are also consistent with the data. Together with the PAGE analysis, these results show that this reaction discovery system reliably and efficiently reveals bond-forming events even under conditions that cause multiple combinations of substrates to react.

Increasing the stringency of the reaction conditions by decreasing the temperature to 25 °C and quenching the reaction after 20 min decreased the number of strong positives to four (A5+B5, A5+B3, A4+B8 and A5 homocoupling; green/red fluorescence ratios = 3.7 to 2.7; standard = 3.5) and the number of weak positives to two (A5+B8 and A5+B9; fluorescence ratios = 1.9 and 1.7; see Fig. 2b and Supplementary Data). Consistent with these results, the five authentic bond-forming reactions among this set (all four strong positives and weak positive A5+B8) showed substantial product formation by PAGE analysis when performed under the 25 °C conditions, whereas the three additional reactions listed in Fig. 3 but not appearing as positives in the 25 °C experiment (A8+B8, A8+B3 and A7+B3) showed significantly weaker product formation (Fig. 3). Varying the stringency of the reaction conditions before selection can therefore efficiently distinguish substrate combinations on the basis of their level of reactivity.

The above results suggested that the Pd(II)-mediated carbon–carbon bond formation between a simple terminal alkyne and

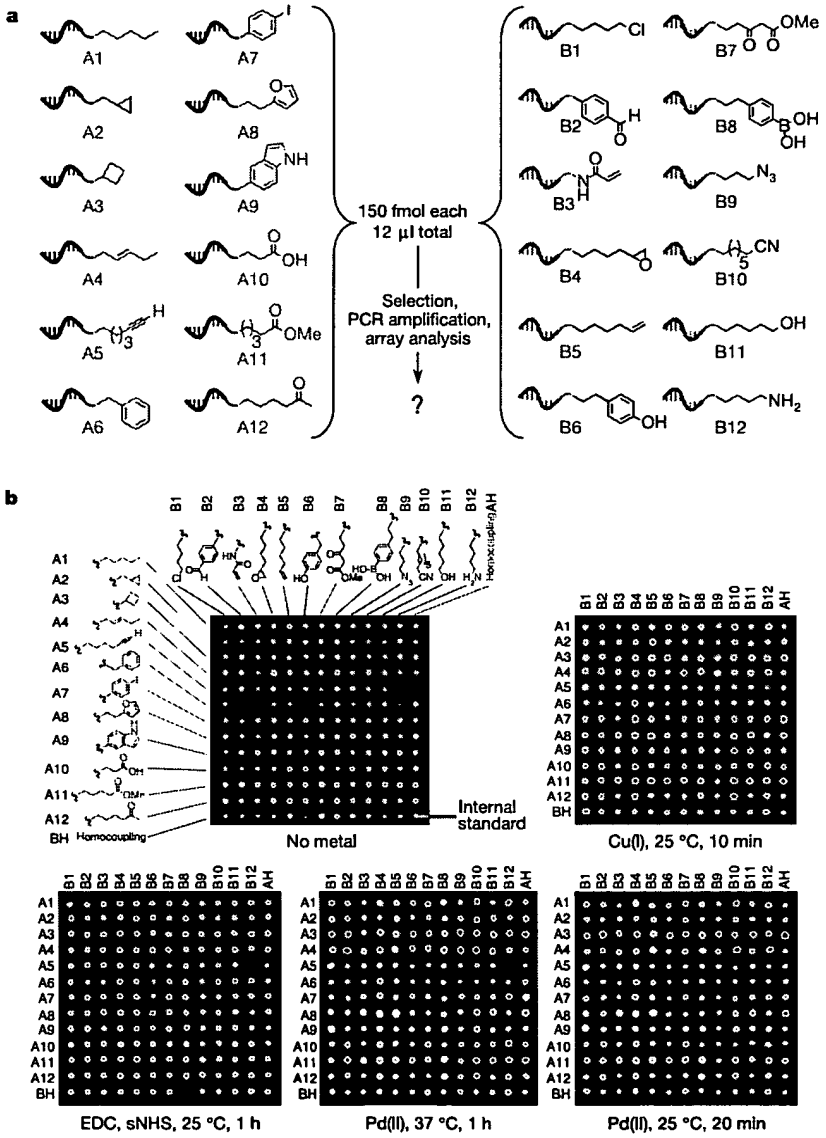


Figure 2 Results from reaction discovery selections and analysis. **a**, Pool A and pool B substrates used in this work. **b**, Qualitative results of reaction discovery selections after exposure to the reaction conditions listed below each array image. Spots that are significantly green suggest bond formation between the corresponding substrates (quantitative fluorescence ratios in Fig. 3 and the Supplementary Data are used for actual interpretations). The 840 reaction possibilities in these five experiments were evaluated by one researcher in two days. See Supplementary Methods for detailed reaction conditions.

©2004 Nature Publishing Group

Alamar Exhibit 1002
Alamar v. Olink

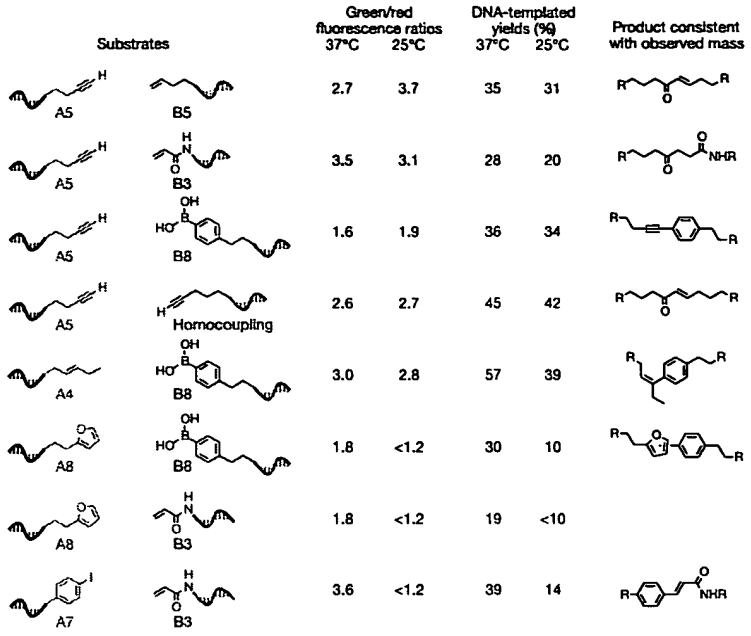
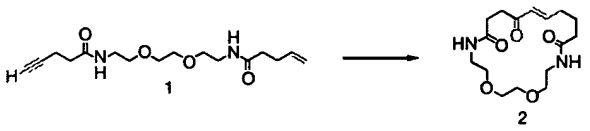


Figure 3 Characterization in a DNA-templated format of array positives resulting from exposure to 500 μM Na_2PdCl_4 at 37 °C for 1 h, or at 25 °C for 20 min. Putative reactions were screened by PAGE and MALDI–TOF mass spectrometric analysis (Supplementary Information). Template architectures and reaction conditions were chosen to match those used in the selection rather than to maximize product yields. Product structures other than those proposed are possible. The green/red fluorescence ratio for the internal standard at 37 °C and 25 °C is 2.9 and 3.5, respectively. For A8 + B3, reactivity was not sufficient to obtain reliable product mass characterization.



Entry	Metal(s)	Solvent	Conditions	Isolated yield
a	1 equiv. Na_2PdCl_4	1 M NaCl in H_2O	25 °C, 15 h	86%
b	5 mol% Na_2PdCl_4 1 equiv. $CuCl_2$	100 mM NaCl in H_2O	25 °C, 2 h	90%
c	5 mol% Na_2PdCl_4 1 equiv. $CuCl_2$	9:1 THF: H_2O	25 °C, 4 h	91%
d	15 mol% Na_2PdCl_4 1 atm O_2	9:1 THF: H_2O	25 °C, 14 h	73%
e	1 equiv. $CuCl_2$	100 mM NaCl in H_2O	25 °C, 4 h	0%
f	1 equiv. CuCl	100 mM NaCl in H_2O	25 °C, 4 h	0%

Figure 4 Characterization of a new alkyne–alkene macrocyclization reaction in a non-DNA-templated format. The macrocyclic enone product (2) was characterized by ^{1}H-NMR, ^{13}C-NMR, COSY, UV–visible spectrometry and high-resolution mass electrospray (Supplementary Information). We speculate that product formation proceeds through the following sequence: soft deprotonation of the alkyne to form a Pd(II)–alkynyl intermediate; insertion of the alkene into the Pd-alkyne bond; β-hydride elimination to form a conjugated enyne; Pd(II)-catalysed hydration of the alkyne to form an enol π-allyl Pd complex; enol tautomerization and π-allyl Pd protonation to generate the *trans*-enone.

terminal alkene (A5+B5) proceeds efficiently to generate a possible enone product (Fig. 2b and Fig. 3), prompting a detailed investigation of this reaction. We synthesized a small-molecule substrate (1) to study a non-DNA-templated, intramolecular version of this alkyne–alkene coupling reaction on a multi-milligram scale (Fig. 4). The addition of 1 to one equivalent of Pd(II) in 1 M aqueous NaCl over 15 h followed by purification by reverse-phase high-performance liquid chromatography provided the 20-membered macrocyclic *trans*-enone 2 as a single olefin stereoisomer in 86% isolated yield (Fig. 4, entry a). The structure of 2 was confirmed by ^{1}H-NMR, ^{13}C-NMR, homonuclear correlated spectroscopy (COSY) and high-resolution mass spectrometry (Supplementary Data). To our knowledge, this is the first example of macrocyclic enone formation from a simple alkyne–alkene precursor. The finding also shows that a reaction discovered by a selection for DNA-templated covalent bond formation can operate in a non-DNA-templated format on a much larger (10^9-fold) scale.

Because the formation of an enone from an alkyne and an alkene represents an oxidative coupling, we proposed that the reaction could be performed with catalytic quantities of Pd by introducing an oxidant such as $CuCl_2$ + air to reoxidize Pd(0) to Pd(II) (ref. 24). Indeed, the addition of 1 to 5 mol% Pd and 1 equivalent of $CuCl_2$ over 2 h in water at 25 °C provided enone 2 as the sole observed product in 90% isolated yield (Fig. 4, entry b). Significantly, this reaction maintains its high efficiency in solvent containing 9:1 tetrahydrofuran:water (Fig. 4, entry c). Product was formed in similar yields, although at a slower rate, using 1 atm O_2 instead of $CuCl_2$ to reoxidize Pd(0) (Fig. 4, entry d). Control reactions with $CuCl_2$ or CuCl alone yielded no observed product formation (Fig. 4, entries e and f).

The discovery of this alkyne–alkene coupling reaction suggests the value of searching a large number of substrate combinations for unexpected reactions. While aqueous Pd(II) has been known for more than 40 years to oxidize alkenes rapidly to ketones[24,25], which are unreactive towards alkynes under these conditions (Fig. 2b and Supplementary Data below Supplementary Fig. S4), our approach to reaction discovery revealed that carbon–carbon bond formation between alkynes and alkenes can outcompete alkene oxidation. Although other enone-forming coupling reactions such as the Horner–Wadsworth–Emmons reaction[26] or aldol condensation are known, the mild reaction conditions, simple hydrocarbon starting materials and high efficiency of the transformation discovered here might render it an attractive alternative for addressing

©2004 Nature Publishing Group

Alamar Exhibit 1002
Alamar v. Olink

some macrocyclization problems in organic synthesis. In addition, the compatibility of this reaction with both organic and aqueous solvents might facilitate the synthesis of highly functionalized macrocyclic enones, of which there are numerous known biologically active examples[27].

Six of the eight observed heterocoupling reactions in Fig. 3 involve a substrate that couples with itself under the reaction conditions (reactions involving either the alkyne or the aryl boronic acid; see Supplementary Data below Supplementary Fig. S4). In a DNA-templated format, heterocoupling without competitive homocoupling is possible because a substrate experiences its sequence-programmed heterocoupling partner at a much higher effective concentration than the concentration of another identical substrate molecule.

Biological macromolecules have previously been used to address specific problems in chemical reactivity[28–30]. In contrast, our approach uses the ability of nucleic acids to direct effective molarities and undergo *in vitro* selection and amplification to reveal bond-forming reactivity in a general manner. Once DNA-linked substrate pools are prepared, a single researcher can evaluate thousands of combinations of substrates and reaction conditions in a two-day experiment. Our approach requires water-soluble catalysts and aqueous reaction conditions that are at least partially compatible with DNA; however, as illustrated by the alkyne–alkene coupling reaction, some of the discovered reactions will proceed in a non-DNA-templated format in organic solvents. We expect that using this approach in a broad examination of reaction conditions including transition-metal complexes, Lewis acids, mild oxidants or reductants, and organic reagents will lead to the discovery of additional bond-forming reactions between simple and relatively unreactive functional groups. □

Received 11 June; accepted 3 August 2004; doi:10.1038/nature02920.

1. Stambuli, J. P. & Hartwig, J. F. Recent advances in the discovery of organometallic catalysts using high-throughput screening assays. *Curr. Opin. Chem. Biol.* **7**, 420–426 (2003).
2. Reetz, M. T. Combinatorial and evolution-based methods in the creation of enantioselective catalysts. *Angew. Chem. Int. Edn Engl.* **40**, 284–310 (2001).
3. Stambuli, J. P., Stauffer, S. R., Shaughnessy, K. H. & Hartwig, J. F. Screening of homogeneous catalysts by fluorescence resonance energy transfer. Identification of catalysts for room-temperature Heck reactions. *J. Am. Chem. Soc.* **123**, 2677–2678 (2001).
4. Taylor, S. J. & Morken, J. P. Thermographic selection of effective catalysts from an encoded polymer-bound library. *Science* **280**, 267–270 (1998).
5. Lober, O., Kawatsura, M. & Hartwig, J. F. Palladium-catalyzed hydroamination of 1,3-dienes: a colorimetric assay and enantioselective additions. *J. Am. Chem. Soc.* **123**, 4366–4367 (2001).
6. Evans, C. A. & Miller, S. J. Proton-activated fluorescence as a tool for simultaneous screening of combinatorial chemical reactions. *Curr. Opin. Chem. Biol.* **6**, 333–338 (2002).
7. Calderone, C. T., Puckett, J. W., Gartner, Z. J. & Liu, D. R. Directing otherwise incompatible reactions in a single solution by using DNA-templated organic synthesis. *Angew. Chem. Int. Edn Engl.* **41**, 4104–4108 (2002).
8. Gartner, Z. J. & Liu, D. R. The generality of DNA-templated synthesis as a basis for evolving non-natural small molecules. *J. Am. Chem. Soc.* **123**, 6961–6963 (2001).
9. Gartner, Z. J., Kanan, M. W. & Liu, D. R. Expanding the reaction scope of DNA-templated synthesis. *Angew. Chem. Int. Edn Engl.* **41**, 1796–1800 (2002).
10. Rosenbaum, D. M. & Liu, D. R. Efficient and sequence-specific DNA-templated polymerization of peptide nucleic acid aldehydes. *J. Am. Chem. Soc.* **125**, 13924–13925 (2003).
11. Li, X. & Liu, D. R. Stereoselectivity in DNA-templated organic synthesis and its origins. *J. Am. Chem. Soc.* **125**, 10188–10189 (2003).
12. Wilson, D. S. & Szostak, J. W. *In vitro* selection of functional nucleic acids. *Annu. Rev. Biochem.* **68**, 611–647 (1999).
13. Joyce, G. F. Directed evolution of nucleic acid enzymes. *Annu. Rev. Biochem.* **73**, 791–836 (2004).
14. Rostovtsev, V. V., Green, L. G., Fokin, V. V. & Sharpless, K. B. A stepwise huisgen cycloaddition process: copper(I)-catalyzed regioselective 'ligation' of azides and terminal alkynes. *Angew. Chem. Int. Edn Engl.* **41**, 2596–2599 (2002).
15. Wang, Q. *et al.* Bioconjugation by copper(I)-catalyzed azide–alkyne [3+2] cycloaddition. *J. Am. Chem. Soc.* **125**, 3192–3193 (2003).
16. Burrows, C. J. & Muller, J. G. Oxidative nucleobase modifications leading to strand scission. *Chem. Rev.* **98**, 1109–1152 (1998).
17. Kramer, M. F. & Coen, D. M. in *Current Protocols in Molecular Biology* Vol. 2 (eds Ausubel, F. M. *et al.*) 15.1.1–15.1.3 (Wiley, New York, 2001).
18. Bailey, P. D., Collier, I. D. & Morgan, K. M. in *Comprehensive Organic Functional Group Transformations* Vol. 5 (eds Katritzky, A. R., Meth-Cohn, O. & Rees, C. W.) 257–307 (Pergamon, New York, 1995).
19. Tsuji, J. *Palladium Reagents and Catalysts* (Wiley, New York, 1995).
20. Heck, R. F. Palladium-catalyzed vinylation of organic halides. *Org. React.* **27**, 345–390 (1982).
21. Li, C.-J. & Chan, T.-H. *Organic Reactions in Aqueous Media* (Wiley, New York, 1997).
22. Miyaura, N. & Suzuki, A. Palladium-catalyzed cross-coupling reactions of organoboron compounds. *Chem. Rev.* **95**, 2457–2483 (1995).
23. Crabtree, R. H. *The Organometallic Chemistry of the Transition Metals* (Wiley, New York, 2001).
24. Smidt, J. *et al.* Olefinoxydation mit palladiumchlorid-katalysatoren. *Angew. Chem.* **74**, 93–102 (1962).
25. Smidt, J. *et al.* Katalytische umsetzungen von olefinen an platinmetall-verbindungen: das consortium-verfahren zur herstellung von acetaldehyd. *Angew. Chem.* **71**, 176–182 (1959).
26. Wadsworth, W. S. J. & Emmons, W. D. Utility of phosphonate carbanions in olefin synthesis. *J. Am. Chem. Soc.* **83**, 1733–1738 (1961).
27. Staunton, J. & Weissman, K. J. Polyketide biosynthesis: a millennium review. *Nat. Prod. Rep.* **18**, 380–416 (2001).
28. Kohli, R. M., Walsh, C. T. & Burkart, M. D. Biomimetic synthesis and optimization of cyclic peptide antibiotics. *Nature* **418**, 658–661 (2002).
29. Breslow, R. Biomimetic chemistry and artificial enzymes: catalysis by design. *Acc. Chem. Res.* **28**, 146–153 (1995).
30. Schultz, P. G. & Lerner, R. A. Completing the circle. *Nature* **418**, 485 (2002).

Supplementary Information accompanies the paper on **www.nature.com/nature.**

Acknowledgements We thank P. Kehayova for the initial analysis of early reaction discovery selections, and C. Bailey (Bauer Center for Genomics Research) for assistance with DNA microarray preparation and analysis. We thank DNA Software for assistance with screening coding and annealing sequences. This work was supported by the National Institutes of Health, the Office of Naval Research, and the Arnold and Mabel Beckman Foundation. M.W.K. and T.M.S. are NSF Graduate Research Fellows. M.M.R. is an NDSEG Graduate Research Fellow. K.S. is a Helen Hay Whitney Postdoctoral Research Fellow.

Competing interests statement The authors declare competing financial interests: details accompany the paper on **www.nature.com/nature.**

Correspondence and requests for materials should be addressed to D.R.L. (liu@chemistry.harvard.edu).

Photosynthetic microbial mats in the 3,416-Myr-old ocean

Michael M. Tice & Donald R. Lowe

Department of Geological and Environmental Sciences, Stanford University, Stanford, California 94305, USA

Recent re-evaluations of the geological record of the earliest life on Earth have led to the suggestion that some of the oldest putative microfossils[1] and carbonaceous matter were formed through abiotic hydrothermal processes[2,3]. Similarly, many early Archaean (more than 3,400-Myr-old) cherts have been reinterpreted as hydrothermal deposits rather than products of normal marine sedimentary processes[2,4,5]. Here we present the results of a field, petrographic and geochemical study testing these hypotheses for the 3,416-Myr-old Buck Reef Chert, South Africa. From sedimentary structures and distributions of sand and mud, we infer that deposition occurred in normal open shallow to deep marine environments. The siderite enrichment that we observe in deep-water sediments is consistent with a stratified early ocean[6,7]. We show that most carbonaceous matter was formed by photosynthetic mats within the euphotic zone and distributed as detrital matter by waves and currents to surrounding environments. We find no evidence that hydrothermal processes had any direct role in the deposition of either the carbonaceous matter or the enclosing sediments. Instead, we conclude that photosynthetic organisms had evolved and were living in a stratified ocean supersaturated in dissolved silica[8,9] 3,416 Myr ago.

The Buck Reef Chert (BRC) is a 250–400-m-thick unit of carbonaceous and ferruginous chert that regionally marks the base of the Kromberg Formation in the Barberton greenstone belt, South Africa[10] (Fig. 1). Along the west limb of the Onverwacht anticline, it includes three main facies reflecting deposition under progressively deeper-water conditions: first, a basal evaporite facies 5–40 m thick that interfingers with the underlying felsic sandstone

©2004 Nature Publishing Group

217

Alamar Exhibit 1002
Alamar v. Olink

This Page is Inserted by IFW Indexing and Scanning Operations and is not part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☐ BLACK BORDERS
- ☐ IMAGE CUT OFF AT TOP, BOTTOM OR SIDES
- ☑ FADED TEXT OR DRAWING
- ☐ BLURRED OR ILLEGIBLE TEXT OR DRAWING
- ☐ SKEWED/SLANTED IMAGES
- ☐ COLOR OR BLACK AND WHITE PHOTOGRAPHS
- ☐ GRAY SCALE DOCUMENTS
- ☐ LINES OR MARKS ON ORIGINAL DOCUMENT
- ☐ REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY
- ☐ OTHER: _____

IMAGES ARE BEST AVAILABLE COPY.
As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.

Alamar Exhibit 1002
Alamar v. Olink

Available online at www.sciencedirect.com
SCIENCE DIRECT®
Methods 32 (2004) 437–444
METHODS
www.elsevier.com/locate/ymeth

Best Available Copy

Labeling of fusion proteins of O^6-alkylguanine-DNA alkyltransferase with small molecules in vivo and in vitro

Antje Keppler,[a,b] Maik Kindermann,[a,b] Susanne Gendreizig,[a,b] Horst Pick,[b] Horst Vogel,[b] and Kai Johnsson[a,b,*]

[a] *Institute of Molecular and Biological Chemistry, Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne CH-1015, Switzerland*
[b] *Institute of Biomolecular Sciences, Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne CH-1015, Switzerland*

Accepted 6 October 2003

Abstract

The in vivo and in vitro labeling of fusion proteins with synthetic molecules capable of probing and controlling protein function has the potential to become an important method in functional genomics and proteomics. We have recently introduced an approach for the specific labeling of fusion proteins, which is based on the generation of fusion proteins with the human DNA repair protein O^6-alkylguanine-DNA alkyltransferase (hAGT) and the irreversible reaction of hAGT with O^6-benzylguanine derivatives. Here, we report optimized protocols for the synthesis of O^6-benzylguanine derivatives and the use of such derivatives for the labeling of different hAGT fusion proteins in vivo and in vitro.
© 2003 Elsevier Inc. All rights reserved.

1. Introduction

Studying protein function in vivo and in vitro often relies on the expression of the protein of interest as a fusion protein with an additional polypeptide fused to its N or C terminus. This so-called tag equips the resulting fusion with a unique property that can be exploited to study certain functions of the protein of interest in vivo and in vitro. In particular, the use of autofluorescent proteins as protein tags has yielded insights into protein functions and biological processes that could not have been achieved with other existing technologies [1].

Other well-established fusion protein-based approaches to study protein function and protein–protein interactions in vivo include the two-hybrid system and the split ubiquitin system [2,3]. However, in these systems, the protein fusion domains are restricted to properties that can be genetically encoded, and each of those protein tags is tailor-made for just one particular application. The latter limitation is particularly problematic if genome-wide studies are to be conducted. One possible way to overcome these restrictions is through the use of protein tags that mediate the labeling of the fusion protein with small synthetic molecules conferring unique properties. Provided that the mechanism of protein labeling is sufficiently promiscuous so that different functionalities can be attached to the tag, but at the same time specific enough so that only the fusion protein is labeled with the synthetic molecule, such an in vivo labeling of fusion proteins should address a number of shortcomings of the more traditional approaches.

The prototype for the in vivo labeling of fusion proteins with small synthetic molecules is based on the formation of stable complexes between biarsenical compounds such as the fluorescein derivative FlAsH and peptides containing a tetracysteine motif [4–7]. In this interaction, each of two As^{III}-atoms binds cooperatively to a pair of vicinal cysteines, and biarsenical fluorophores with different excitation and emission wavelengths have been synthesized and used for in vivo applications. The wide applicability of such an approach is highlighted by the finding that one of the synthesized biarsenical compounds, a resorufin derivative abbreviated as ReAsH, can be used for the localized photoconversion of diaminobenzidine into an electron-dense precipitate [5,7]. This allows performing high resolution

* Corresponding author. Fax: +41-21-693-9365.
E-mail address: kai.johnsson@epfl.ch (K. Johnsson).

1046-2023/$ - see front matter © 2003 Elsevier Inc. All rights reserved.
doi:10.1016/j.ymeth.2003.10.007

Alamar Exhibit 1002
Alamar v. Olink

electron microscopy imaging of cells expressing tetracysteine-tagged proteins labeled with ReAsH, thereby complementing the fluorescence studies.

Our laboratory has recently developed an approach for the specific labeling of fusion proteins in vivo that is based on exploitation of the activity of the DNA repair protein O^6-alkylguanine-DNA alkyltransferase (AGT) [8]. AGT recognizes O^6-alkylated guanine in DNA, and irreversibly transfers the alkyl group to one of its reactive cysteine residues, yielding repaired DNA and alkylated AGT. Human AGT (hAGT) also reacts with the cell-permeable nucleobase O^6-benzylguanine (BG), and we have recently shown that this reaction can be exploited for the specific in vivo labeling of hAGT fusion proteins with BG derivatives substituted at the 4-position of the benzyl ring (Fig. 1A). The high specificity of the labeling reaction and its independence of the nature of the label should allow this approach to complement the in vivo labeling of fusion proteins based on the tetracysteine motif.

To a large extent the efficiency of in vivo labeling of hAGT depends on the cell-permeability of the substrate in question. In addition to in vivo labeling, hAGT fusion proteins can also be efficiently labeled in vitro with a large variety of labels and can be selectively immobilized on surfaces displaying BG [9]. In this article, we present improved protocols for the synthesis of BG derivatives as well as representative examples for the labeling of hAGT fusion proteins in vivo and in vitro.

2. Description of method

The application of the hAGT tag for the labeling of fusion proteins in vivo and in vitro comprises two steps: (i) the synthesis of the BG derivate, which contains the desired molecular probe and (ii) the generation and the labeling of the hAGT fusion protein of interest. A wide variety of spectroscopic and chemical probes commonly used in protein chemistry are commercially available as activated esters and all these molecules can be directly coupled to the free amine of BG derivatives **1** or **2** (Fig. 1B), followed by a purification step. As the rate of the reaction of hAGT is not significantly influenced by substituents attached to the amines of **1** or **2**, there are practically no restrictions with respect to the nature of

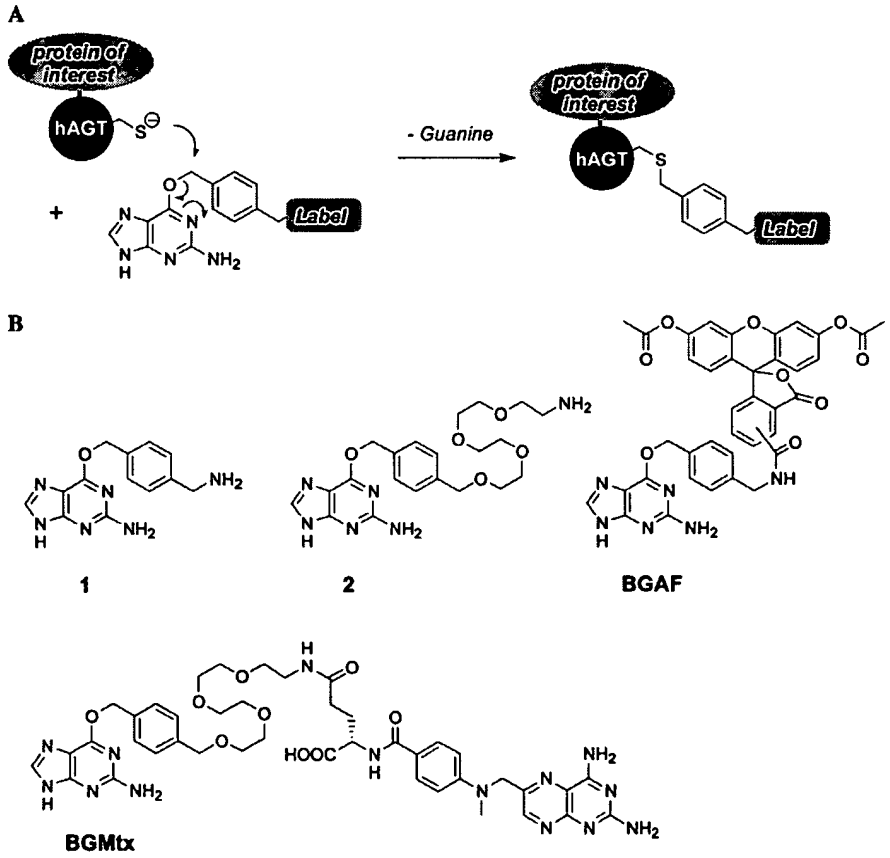


Fig. 1. Labeling of hAGT fusion proteins. (A) General mechanism for the labeling of hAGT fusion proteins with synthetic molecules using BG derivatives. (B) BG derivatives described within this work. Free amines **1** and **2** are the central precursors for the synthesis of BG derivatives, BGAF and BGMtx are derivatives used for the labeling of fusion proteins with either fluorescein (BGAF) or methotrexate (BGMtx).

Alamar Exhibit 1002
Alamar v. Olink

the label for in vitro applications. In vivo applications depend on the cell-permeability of the label and might also be hampered by unspecific interactions of the label within the living cell.

2.1. Synthesis of BG derivatives

For the preparation of BG derivatives, currently two possibilities exist: the label is attached to the free amine of either compound **1** or compound **2**, the latter possessing an additional tetraethylene glycol linker (Fig. 1B). In our hands, derivatives prepared from **1** possess higher cell permeabilities than those prepared from **2**, making these preferable for in vivo studies. In contrast, the solubility and handling of derivatives prepared from **2** are usually better than those prepared from **1** and they are therefore preferred for in vitro applications. Both amines **1** and **2** readily react with activated carboxyl esters, including the commercially available *N*-hydroxysuccinimide (NHS) esters of a large number of spectroscopic and chemical probes. A detailed synthesis of **2** has been recently published [9] and we present here an optimized procedure for the preparation of **1** (Fig. 2). As a representative example for the reaction of **1** or **2** with an activated ester, we report an improved procedure for the derivatization of **1** with membrane permeable fluorescein diacetate, yielding BGAF (Fig. 2).

2.2. 2,2,2-Trifluoro-N-(4-hydroxymethyl-benzyl)-acetamide 5

4.17 g (29.39 mmol) trifluoroacetic acid ethyl ester was added dropwise to a solution of 3.36 g (24.49 mmol) of 4-(aminomethyl)-benzyl alcohol and 3.0 ml triethylamine in 30 ml dry methanol. The reaction mixture was stirred for 1 h and diluted with 50 ml ethyl acetate and 50 ml water. The aqueous layer was extracted with ethyl acetate and the combined organic layers were washed with saturated NaCl and dried over $MgSO_4$. After removal of the solvents in vacuo, the crude product was purified by flash column chromatography using an

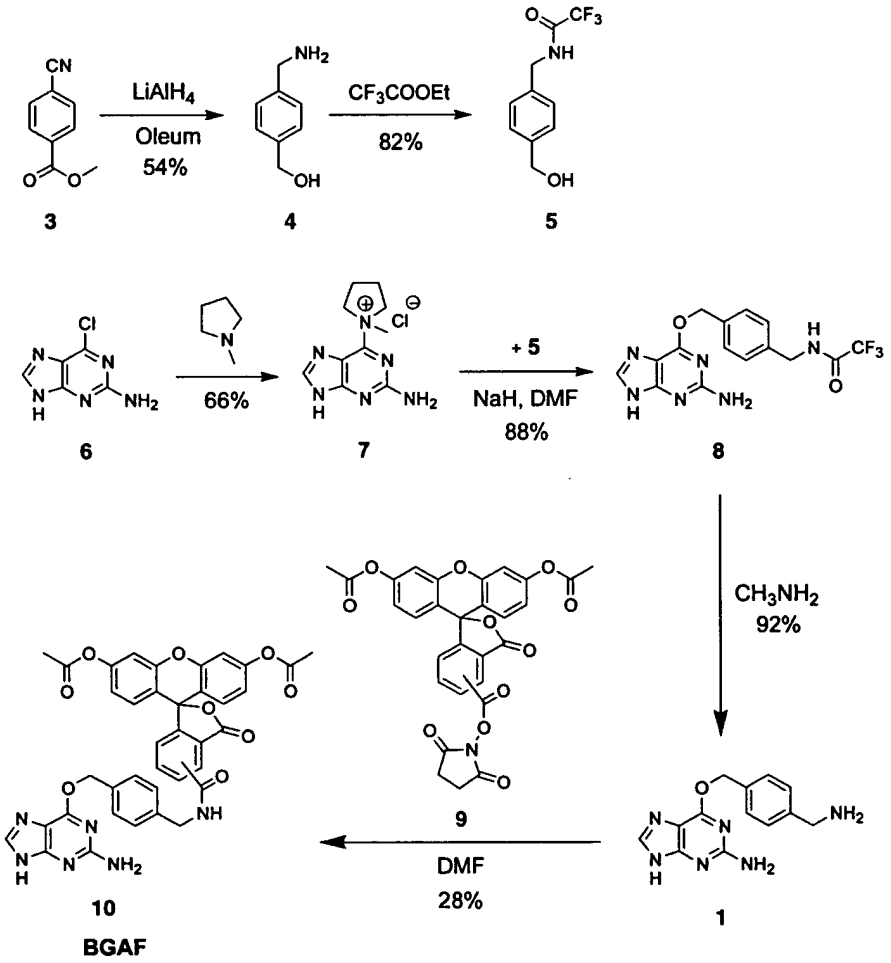


Fig. 2. Synthetic scheme for the preparation of 1 and BGAF 10.

Alamar Exhibit 1002
Alamar v. Olink

eluant mixture of ethyl acetate:petrolether (1:2). Yield: 4.68 g (20.08 mmol, 82%). $R_f = 0.26$ (ethyl acetate:petrolether (1:1)). ^{1}H-NMR ($CDCl_3$): $\delta = 1.74$ (br s, 1H, –OH), 4.53 (d, $J = 5.8$ Hz, 2H, –CH_2–NH–), 4.71 (s, 2H, –CH_2–OH), 6.60 (br s, 1H, –NH–CO–), 7.30 (m, 4H, Ar) ppm. MS (CI) m/z(%) = 234 (2.0 $[M+H]^+$), 233 (8.8), 215 (33.0), 202 (14.2), 91 (18.6).

2.3. N-[4-(2-Amino-9H-purin-6-yloxymethyl)-benzyl]-2,2,2-trifluoro-acetamide 8

1.80 g (7.72 mmol) of 2,2,2-trifluoro-*N*-(4-hydroxymethyl-benzyl)-acetamide was dissolved in 15 ml dry DMF under argon atmosphere and 555 mg (23.15 mmol) NaH was added in three portions over 5 min. As much as 1.97 g (7.72 mmol) of 1-(2-amino-7*H*-purin-6-yl)-1-methyl-pyrrolidinium chloride and 0.1 g (0.82 mmol) 4-(dimethylamino)pyridine (DMAP) were then added and the solution was stirred for 2 h at room temperature. One milliliter of water was added carefully to quench all excess NaH and the solvent was removed in vacuo. The crude product was dissolved in methanol, adsorbed on SiO_2 (5 g), and purified by flash column chromatography with dichloromethane:methanol (95:5). Yield: 2.5 g (6.83 mmol, 88%). $R_f = 0.27$ (dichloromethane:methanol (10:1)). ^{1}H-NMR (DMSO-d_6): $\delta = 4.40$ (d, $J = 4.9$ Hz, 2H, –CH_2–NH–CO–), 5.46 (s, 2H, –CH_2–O–), 6.29 (s, 2H, –NH_2), 7.4 (m, 4H, Ar), 7.81 (s, 1H, H8), 10.01 (s, 1H, –NH–CO–), and 12.42 (1H, H9) ppm. MS (CI) m/z(%) = 367 (100 $[M+H]^+$), 269 (8.21), 152 (12.1).

2.4. O^6-(4-Amino-methyl-benzyl)guanine 1

2.0 g (5.46 mmol) *N*-[4-(2-amino-9*H*-purin-6-yloxymethyl)-benzyl]-2,2,2-trifluoro-acetamide was dissolved under argon atmosphere in 10 ml dry methanol and treated with methylamine (33% in ethanol, 30 ml) at room temperature for 24 h. The solvents and excess methylamine were removed in vacuo. The residue was redissolved in methanol, adsorbed on SiO_2 (3.0 g), and purified by flash column chromatography with dichloromethane:methanol (5:1). After removal of solvents in vacuo, a pale yellow solid was obtained. Yield: 1.35 g (5.02 mmol, 92%). $R_f = 0.02$ (dichloromethane:methanol (5:1)). ^{1}H-NMR (DMSO-d_6): $\delta = 7.82$ (s, 1H, H8 guanine), 7.39 (dd, 4H, $J = 8.19$, 36.5 Hz, C_6H_4), 6.27 (s, 2H, NH_2 guanine), 5.45 (s, 2H, C_6H_4–CH_2–O–), and 3.73 (s, 2H, C_6H_4–CH_2– NH_2). MS (ESI) 271.36 $[M^+ + H^+]$.

2.5. BGAF 10

56.7 mg (0.21 mmol) of O^6-(4-aminomethyl-benzyl)guanine and 117.0 mg (0.21 mmol) of 5(6)-carboxyfluorescein diacetate-*N*-succinimidyl ester (Fluka) were stirred in 1.5 ml dry *N*,*N*-dimethylformamide (DMF) at room temperature. After 10 min, the solution became clear. Stirring was continued for 1 h and the crude product precipitated in 50 ml diethyl ether/*n*-pentane 4:1. The precipitate was redissolved in methanol, adsorbed on SiO_2 (100 mg), and purified by flash column chromatography (gradient chloroform:methanol (40:1 to 20:1)). The product was dissolved in a minimal volume of chloroform and precipitated in diethyl ether to give a colorless solid. Yield: 42 mg (0.058 mmol, 28%). $R_f = 0.38$ (dichloromethane:methanol (10:1)). MS(ESI) 713.35 $[M+H]^+$.

The product can be dissolved in dimethyl sulfoxide (DMSO), stored in aliquots at −20 °C, and should be protected from light. While BG itself is very stable, the diacetate of fluorescein is susceptible to hydrolysis in aqueous solutions. For a typical in vivo labeling experiment (1 ml of 5 μM BGAF), about 4 μg of BGAF is required.

2.6. Labeling of hAGT fusion proteins in vivo and in vitro

The experimental details for the in vivo and in vitro labeling of hAGT fusion proteins depend on the host in which the fusion protein is to be expressed and labeled. In yeast and bacteria, no background labeling is observed since the endogenous AGTs do not show significant reactivity against BG derivatives [8]. In mammalian cells, however, background labeling of the endogenous AGTs has to be taken into account [8]. This problem can be circumvented by using AGT-deficient cell lines such as CHO-9 neo C5 and HeLa MR [8,10]. We have recently described the generation of two hAGT mutant proteins that possess significantly increased activity against BG derivatives and which also lead to a more efficient labeling of hAGT fusion proteins in vivo and in vitro [11]. These proteins possess the mutations Asn157Gly and Ser159Glu (^{GE}hAGT), and Asn157Gly, Ser159Glu, and Trp160Ala (^{GEA}hAGT). It is recommended that these mutant proteins are used for all labeling applications and the genes are available from our laboratory.

Proteins or polypeptides can be fused to either the N or the C terminus of hAGT without significantly affecting the activity of hAGT. Of note, hAGT possesses five cysteine residues of which Cys145 is the site of the alkyl transfer required for labeling, Cys5 and Cys35 participate in the creation of a Zn^{2+}-binding site, and for Cys62 and Cys151 no obvious structural or functional role is known [12]. Care has therefore to be taken when attempting to express and label hAGT fusion proteins in oxidizing environments as oxidation of at least Cys145 will lead to inactivation of the hAGT fusion protein. Despite the presence of functionally or structurally important cysteine residues, we have shown that hAGT fusion proteins can be labeled on the surface of filamentous phage and when displayed on the outer

Alamar Exhibit 1002
Alamar v. Olink

surface of mammalian cells [11]. Nevertheless, the potential problem arising from oxidizing one or several of the cysteines of hAGT should be kept in mind when attempting to label hAGT fusion proteins in an oxidizing environment. Another potential pitfall can occur during labeling of hAGT fusion proteins: the alkylation of Cys145 is known to induce a conformational change (unfolding) that triggers the eventual degradation of alkylated hAGT [12]. In our hands, the fluorescence signal obtained after labeling of a cytosolic hAGT fusion protein (hAGT-βGal) in CHO and HeLa cells decreased by about 10% per hour. This indicates that the degradation of hAGT fusion proteins after labeling should not interfere with most labeling experiments.

While in the following section the fluorescence labeling of hAGT fusion proteins with fluorescein is described, the in vivo labeling with other fluorophores such as tetramethylrhodamine (TMR) is also possible. For these experiments, the corresponding BG derivative is prepared by coupling of the amine **1** to the NHS ester of TMR (available from Molecular Probes). The possibility to use different colors in live cells is a distinct advantage over commonly used autofluorescent proteins.

2.7. Fluorescence labeling of hAGT fusion proteins with the human estrogen receptor α

As a novel and representative example of our technique, we describe here the labeling of a nuclear protein, the human estrogen receptor α (hERα). hERα belongs to the superfamily of nuclear hormone receptors, which comprise a large family of ligand-activated transcription factors. ERs bind to the hormone response elements of target genes like progesterone receptor, c-fos, c-jun or c-myc and regulate their transcription [13,14]. They are involved in the development of normal functions of female reproductive tract, secondary sex characteristics, and in reproductive behavior [15]. The protein structure of the ERs is composed of three distinct domains: an N-terminal domain (A/B), a central DNA-binding domain (C), and a C-terminal ligand-binding domain (E). Upon ligand binding, the ligand-binding domain (E) mediates the dimerization of receptor monomers [16] and binds to transcription coactivators and corepressors [17].

The cellular localization of hERα has been studied mostly by characterizing fusion proteins of hERα with autofluorescent proteins. The vast majority of GFP-hERα molecules are located in the nucleus of the cell but a small amount of receptor molecules undergo dynamic shuttling between the cytoplasm and the nucleus [18,19]. The trafficking is reduced in the presence of 17β-estradiol (E2) or the antagonist tamoxifen. Within the nucleus, GFP-hERα has a diffuse distribution in the nucleus and is excluded from the nucleoli. However, ligand binding significantly changes the distribution of GFP-hERα: 10–20 min after addition of ligand a characteristic hyperspeckled pattern of receptor molecules can be observed [20]. The intranuclear mobility of GFP-hERα was also studied by fluorescence recovery after photobleaching (FRAP), revealing that E2 and tamoxifen treatment increase the immobile fraction of receptor molecules. The subcellular distribution and the intranuclear dynamics of hERα may well influence its transcriptional activity. Here, we show how the fluorescence labeling of hAGT-hERα fusion proteins can be used to study the cellular distribution of the fusion protein in the presence and absence of the antagonist 4-hydroxytamoxifen.

2.8. Construction of fusion proteins and mammalian cell culture

Using standard PCR techniques, ^{GE}hAGT was fused to the N-terminus of the hERα by a linker of 9 amino acids (SRAQASNSA). The construct was cloned into the commercially available mammalian expression vector pEYFP-C1 (Clontech) encoding the enhanced cyano fluorescent protein using its *Nhe*I and *Xho*I sites. For control experiments, the fusion protein EYFP-hERα was expressed using the same vector.

To avoid background staining from endogenous hAGT, a hAGT-deficient HeLa cell line, HeLa MR, was used in all following experiments [10]. HeLa MR cells were grown in DMEM/F12 (Dulbecco's modified Eagle's medium; Gibco-BRL, Rockville, USA). The medium was supplemented with 2.2% (v/v) hormone-free fetal calf serum (Gibco-BRL) and 300 μg/ml G418. Depletion of hormone ligands for steroid receptor activation studies was achieved by treating the serum for 30 min with 10% charcoal. The cultures were kept in an incubator at 37 °C, 5% CO_2, and were split in regular time intervals. Sixteen to 20 h before transfection, exponentially growing cells were seeded (2×10^5 cells/well) on sterile microscope coverslides placed in a 6-well plate. The cells were transfected using calcium phosphate-precipitated DNA. Five hundred microliters of a 250 mM $CaCl_2$ solution containing 25 μg of plasmid DNA was added to an equal volume of a solution containing 140 mM NaCl, 1.4 mM Na_2HPO_4, and 50 mM Hepes, pH 7.05, at 23 °C. One minute after mixing the two solutions at room temperature, this transfection mixture was added to the cell culture medium. For each milliliter of cell medium, 100 μl of calcium phosphate-DNA precipitate (containing 2.5 μg of plasmid DNA) was added. After 4 h at 37 °C in the incubator, the transfection medium was replaced by 10% (v/v) glycerol in DMEM (2 ml/well). Immediately after a "shock" incubation period of 1 min, the glycerol solution was aspirated. Cells were washed twice with 2 ml DMEM, per wash, per well. Finally, 2 ml DMEM

Alamar Exhibit 1002
Alamar v. Olink

medium containing 2.2% (v/v) fetal calf serum was added to each well. Incubation as above was continued for 20 h before the labeling experiments.

2.9. In vivo labeling of hAGT-fusion proteins and fluorescence studies

Adherent HeLa MR cells transiently expressing the relevant hAGT-hERα constructs were transferred to a microscope chamber (400 µl volume, room temperature). To label the hAGT-hERα with fluorescein, the medium was replaced by 400 µl PBS containing 5 µM BGAF and the cells were incubated for 5 min. BGAF, prepared as described above, contains the cell-permeable diacetate of fluorescein, which is readily hydrolyzed to fluorescein by esterases within the cell. The cells were then washed three times for 1 min at room temperature with 400 µl PBS to remove excess dye before covering the cells again with 400 µl hormone-free medium. The careful execution of the washing steps is important for decreasing the background signal from excess dye.

The mean fluorescence intensity of the nucleus and the cytoplasm of 10 different cells was measured 30 min after the last washing step using laser scanning fluorescence microscopy (Zeiss LSM510, Zeiss, Oberkochen, Germany) (Fig. 3). Detection of fluorescence signals of fluorescein (excitation at 488 nm) was achieved by using a 63× water (1,2 numerical aperture) objective and appropriate filters (Zeiss filter set number 15). Scanning speed and laser intensity were adjusted to avoid photobleaching of the fluorescent probe and damage or morphological changes of the cells. Fluorescence signal intensities in optical slice images of cells were evaluated using Zeiss LSM software.

Transfected cells expressing the respective hAGT-hERα fusion protein could be easily distinguished from non-transfected cells by their prominent nuclear fluorescence due to fluorescence-labeled hAGT-hERα (Fig. 3A). The number of transfected HeLa MR in these experiments was about 5% of the total cell number. No significant fluorescence signal above background in the cytoplasm (resulting from fluorescence-labeled hAGT-hERα or free dye) could be detected, the ratio of the relative fluorescence from nucleus compared to cytoplasm being greater than 200. A similar ratio was measured for EYFP-hERα under these conditions, indicating that the nuclear localization of hAGT-hERα is not significantly influenced by the hAGT tag. The fusion protein was distributed evenly throughout the nucleoplasm but was excluded from the nucleoli as has been described before [20].

To study the intranuclear redistribution of the hERα fusion protein dependent on the presence of ligand, 100 nM of 4-hydroxytamoxifen was added to the medium and the cells were imaged 45 min later (Fig. 3C). The addition of 4-hydroxytamoxifen to the medium caused a redistribution of hAGT-hERα from a diffuse pattern to a hyperspeckled distribution [20]. Although the images shown here were taken 45 min after ligand addition, first effects of ligand binding on the nuclear distribution of the receptor fusion protein can already be seen after 20 min.

2.10. Labeling of hAGT fusion proteins in vitro

The specific labeling of hAGT fusion proteins with synthetic molecules in vitro, for example in cell extracts, also offers a variety of interesting applications. Proteins can be labeled with probes such as biotin for affinity purification or with spectroscopic probes for interaction assays. Since in these experiments the cell permeability of the label is no longer a prerequisite for a successful

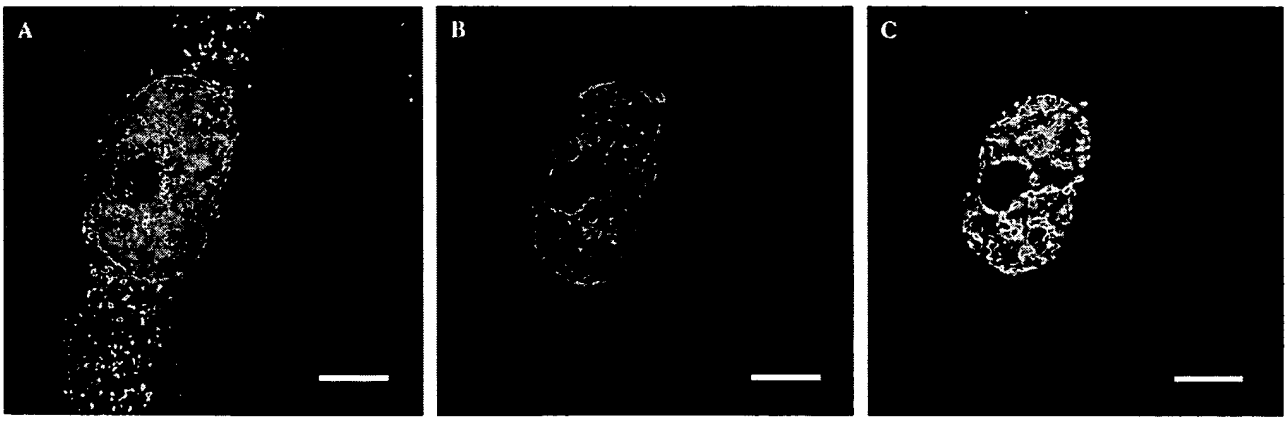


Fig. 3. Intracellular distribution of hAGT-hERα inside HeLa cells visualized by confocal microscopy after labeling with BGAF. (A) Overlay of transmission and fluorescence channel (Ex. 488 nm/Em. 510 nm). (B) In hormone-free growth medium, the receptor is exclusively found in the nucleus. (C) The intranuclear distribution is distinctly altered upon addition of 4-hydroxytamoxifen. The intranuclear disposition of the receptor changes from a more or less uniform to a distinctly punctuate accumulation. The scale bars represent 10 µm.

Alamar Exhibit 1002
Alamar v. Olink

labeling, a much broader spectrum of functionalities can be covalently attached to the protein of interest. As a representative example we report here the selective labeling of two different hAGT fusion proteins in crude cell extracts of *Saccharomyces cerevisiae* with the synthetic affinity label methotrexate using BGMtx (Fig. 1B).

2.11. Labeling of hAGT fusion proteins in crude cell extracts of Saccharomyces cerevisiae

The bacterial DNA-binding protein LexA and the enzyme 3-orotic acid decarboxylase (Ura3) from yeast were expressed as hAGT fusion proteins in the L40 strain of the yeast *S. cerevisiae*. For the construction of the LexA-hAGT fusion, the hAGT gene was subcloned into the vector pHybLex/Zeo (Invitrogen). In the resulting vector, the 49 kDa LexA-hAGT fusion protein is expressed constitutively in yeast L40 strains. In the construct hAGT-Ura3, the stop codon of the hAGT gene is replaced by codons for Arg and Ser and is followed by the HA-tag as an additional linker sequence. The HA-tag is directly followed by the Ura3 gene. The construct was subcloned into a yeast expression vector in which the gene for hAGT-Ura3 is under control of the p_{CUP1} promoter and which codes for Trp1 as an auxotrophic marker. Both plasmids were transformed into yeast L40.

A culture of yeast expressing the corresponding fusion protein in minimal medium lacking tryptophan (initial OD_{600} of about 0.2) was incubated at 30 °C until an OD_{600} of about 0.9–1.0. For the expression of hAGT-Ura3, $CuSO_4$ was added to a concentration of 100 μM. Cells were harvested by centrifugation and a pellet from 1 ml cell culture was resuspended in 10 μl yeast protein extraction reagent (Y-PER; Pierce) containing a protease inhibitor cocktail (AEBSF, 1,10-phenanthroline, Pepstatin A, E-64; Sigma) and BGMtx at a concentration of 100 μM. It should be noted that during cell lysis certain lysis buffers or proteases can affect the activity of hAGT. For example, EDTA can strip hAGT from its structurally important Zn^{2+}-ion, thereby potentially decreasing the efficiency of subsequent labeling steps. After 40 min incubation at room temperature, 15 μl of 2× SDS sample buffer was added to the crude cell lysate and the mixture was heated to 95 °C for 5 min. Fifteen microliters of each sample was applied to SDS–PAGE (12% polyacrylamide) followed by Western blot analysis using an anti-MTX antibody (Assay Designs) and an anti-rabbit antibody peroxidase conjugate (Sigma) as a secondary antibody.

As can be seen from the Western blot, only the hAGT fusion protein is labeled to a significant extent in the cell extract (Fig. 4). As the molecular probe that is attached to the protein of interest can be freely chosen, the in vitro labeling of hAGT fusion proteins should become important for a variety of functional assays.

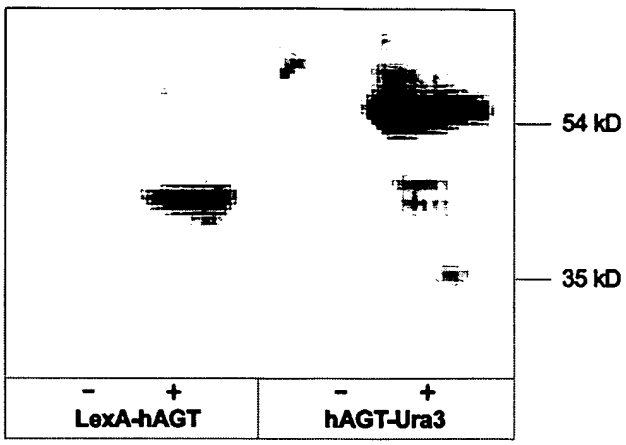


Fig. 4. Labeling of hAGT fusion proteins with BGMtx in total extracts of yeast cells. Shown is a Western blot of labeled protein, which is probed by an anti-MTX antibody. The + sign indicates addition of 100 μM BGMtx to cell extract; the − sign indicates that no substrate was added. The expression level of the two fusion proteins under these conditions was estimated to be below 1% of total protein.

3. Concluding remarks

Equipping fusion proteins with functionalities through chemical labeling that cannot be genetically encoded opens up new ways to study proteins in vivo and in vitro. The usefulness of such an approach critically depends both on the specificity of the labeling step and on the ease with which a wide variety of different functionalities can be attached to the protein of interest. The labeling of fusion proteins of the human O^6-alkylguanine-DNA alkyltransferase (hAGT) fulfills these two requirements as the labeling step possesses a high specificity for hAGT but shows practically no limitations with respect to the nature of the label that can be attached to the protein of interest. This should make the hAGT-based labeling of fusion proteins for a variety of biological problems an attractive alternative to more conventional approaches.

Acknowledgments

We thank Jan Barnikow for critical reading of the manuscript. This work was supported by the Swiss National Science Foundation and the Bundesamt fuer Bildung und Wissenschaft (project *X-TB*).

References

[1] J. Zhang, R.E. Campbell, A.Y. Ting, R.Y. Tsien, Nat. Rev. Mol. Cell Biol. 3 (2002) 906–918.
[2] S. Fields, O. Song, Nature 340 (1989) 245–246.
[3] N. Johnsson, A. Varshavsky, Proc. Natl. Acad. Sci. USA 91 (1994) 10340–10344.

Alamar Exhibit 1002
Alamar v. Olink

[4] B.A. Griffin, S.R. Adams, R.Y. Tsien, Science 281 (1998) 269–272.
[5] G. Gaietta, T.J. Deerinck, S.R. Adams, J. Bouwer, O. Tour, D.W. Laird, G.E. Sosinskyl, R.Y. Tsien, M.H. Ellisman, Science 296 (2002) 503–507.
[6] J. Nakanishi, T. Nakajima, M. Sato, T. Ozawa, K. Tohda, Y. Umezawa, Anal. Chem. 73 (2001) 2920–2928.
[7] S.R. Adams, R.E. Campbell, L.A. Gross, B.R. Martin, G.K. Walkup, Y. Yao, J. Llopis, R.Y. Tsien, J. Am. Chem. Soc. 124 (2002) 6063–6076.
[8] A. Keppler, S. Gendreizig, T. Gronemeyer, H. Pick, H. Vogel, K. Johnsson, Nat. Biotechnol. 21 (2003) 86–89.
[9] M. Kindermann, N. George, N. Johnsson, K. Johnsson, J. Am. Chem. Soc. 125 (2003) 7810–7811.
[10] B. Kaina, G. Fritz, S. Mitra, T. Coquerelle, Carcinogenesis 12 (1991) 1857–1867.
[11] A. Juillerat, T. Gronemeyer, A. Keppler, S. Gendreizig, H. Pick, H. Vogel, K. Johnsson, Chem. Biol. 10 (2003) 313–317.
[12] D.S. Daniels, C.D. Mol, A.S. Arvai, S. Kanugula, A.E. Pegg, J.A. Tainer, EMBO J. 19 (2000) 1719–1730.
[13] P.M. Yen, Physiol. Rev. 81 (2001) 1097–1142.
[14] N.J. McKenna, B.W. O'Malley, Ann. N Y Acad. Sci. 949 (2001) 3–5.
[15] M. Muramatsu, S. Inoue, Biochem. Biophys. Res. Commun. 270 (2000) 1–10.
[16] W.H. Landschulz, P.F. Johnson, S.L. McKnight, Science 240 (1988) 1759–1764.
[17] K.B. Horwitz, T.A. Jackson, D.L. Bain, J.K. Richer, G.S. Takimoto, L. Tung, Mol. Endocrinol. 10 (1996) 1167–1177.
[18] H. Htun, L.T. Holth, D. Walker, J.R. Davie, G.L. Hager, Mol. Biol. Cell. 10 (1999) 471–486.
[19] P. Maruvada, C.T. Baumann, G.L. Hager, P.M. Yen, J. Biol. Chem. 278 (2003) 12425–12432.
[20] D.L. Stenoien, M.G. Mancini, K. Patel, E.A. Allegretto, C.L. Smith, M.A. Mancini, Mol. Endocrinol. 14 (2000) 518–534.

Alamar Exhibit 1002
Alamar v. Olink

This Page is Inserted by IFW Indexing and Scanning Operations and is not part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☐ BLACK BORDERS
- ☑ IMAGE CUT OFF AT TOP, BOTTOM OR SIDES
- ☐ FADED TEXT OR DRAWING
- ☐ BLURRED OR ILLEGIBLE TEXT OR DRAWING
- ☐ SKEWED/SLANTED IMAGES
- ☐ COLOR OR BLACK AND WHITE PHOTOGRAPHS
- ☐ GRAY SCALE DOCUMENTS
- ☐ LINES OR MARKS ON ORIGINAL DOCUMENT
- ☐ REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY
- ☐ OTHER: _____

IMAGES ARE BEST AVAILABLE COPY.
As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.

Alamar Exhibit 1002
Alamar v. Olink

Yeast, A. H. Rose and J. S. Harrison, Eds. (Academic Press, New York, 1969), vol. 1, pp. 385–460.
6. M. S. Esposito, *Proc. Nat. Acad. Sci. U.S.A.* 75, 4436 (1978).
7. H. Roman, *Cold Spring Res. Conserap.* 45, 211 (1980).
8. F. Fabre, *Nature* 272, 795 (1978).
9. R. E. Malone, J. E. Gafin, M. S. Esposito, *Curr. Gen.* 1, 241 (1980).
10. M. Meselson and C. Radding, *Proc. Natl. Acad. Sci. U.S.A.* 72, 358 (1975).
11. C. Bruschi and M. Esposito, *ibid.* 80, 7566 (1983).
12. R. Holliday, *Genet. Res.* 5, 282 (1964).
13. J. W. Szostak, T. L. Orr-Weaver, R. J. Rothstein, *Cell* 33, 25 (1983).
14. M. Esposito and J. Wagstaff, in *The Molecular Biology of the Yeast Saccharomyces*, J. N. Strathern, E. W. Jones, J. R. Broach, Eds. (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1982), vol. 1, pp. 341–370.
15. T. L. Orr-Weaver and J. W. Szostak, *Microbiol. Rev.* 49, 33 (1985).
16. A. Guidry and M. Fitzgerald-Hayes, *Mol. Cell. Biol.* 7, 68 (1987).
17. D. Kotkland, L. Rutledge, M. Fitzgerald-Hayes, L. H. Harrwell, *Cell* 48, 801 (1987).
18. S. Fogel, R. K. Mortimer, K. Lusnak, in *The Molecular Biology of the Yeast Saccharomyces*, J. N. Strathern, E. W. Jones, J. R. Broach, Eds. (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1982), vol. 1, pp. 289–339.
19. J. D. Boeke, F. Lacroute, G. R. Fink, *Mol. Gen. Genet.* 197, 345 (1984).
20. S. Jinks-Robertson and T. D. Petes, *Genetics* 114, 731 (1986).
21. M. Lichten, R. H. Borts, J. E. Haber, *ibid.* 115, 233 (1987).
22. F. Sherman, G. R. Fink, J. B. Hicks, *Methods in Yeast Genetics* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1983).
23. H. Ito, Y. Fukuda, K. Murata, A. Kimura, *J. Bacteriol.* 153, 163 (1983).
24. B. J. Rothstein, *Methods Enzymol.* 101, 202 (1983).
25. L. Clarke and J. Carbon, *Nature* 305, 23 (1983).
26. E. M. Southern *J. Mol. Biol.* 98, 503 (1975).
27. P. A. Brown and J. W. Szostak, *Methods Enzymol.* 101, 278 (1983).
28. We thank M. Fitzgerald-Hayes for providing the GEN1 probes, J. Szostak for providing for the *TCM1* gene on plasmid pSZ414 and S. Jinks-Robertson for providing plasmid pSR7; Nina Brown for technical assistance; S. Piccologlou, and D. Morrison for helpful comments about the manuscript; and W. O. Godfriedsen and Lea Pharmaceutical Products for the gift of nichodomin. Supported by NIH grant GM24110 from the National Institute of General Medical Sciences.

28 March 1988; accepted 13 June 1988

A Ligase-Mediated Gene Detection Technique

ULF LANDEGREN, ROBERT KAISER, JANE SANDERS, LEROY HOOD

An assay for the presence of given DNA sequences has been developed, based on the ability of two oligonucleotides to anneal immediately adjacent to each other on a complementary target DNA molecule. The two oligonucleotides are then joined covalently by the action of a DNA ligase, provided that the nucleotides at the junction are correctly base-paired. Thus single nucleotide substitutions can be distinguished. This strategy permits the rapid and standardized identification of single-copy gene sequences in genomic DNA.

DNA ANALYSIS IS ATTAINING increasing importance for the diagnosis of disease caused by single-gene defects as well as for the detection of infectious organisms (*1*). Moreover, a number of genes, predominantly those encoded in the major histocompatibility complex, have been found to be associated with an increased susceptibility to a variety of disease states (*2*). Of a total of approximately 2000 defined human genetic loci (*3*), approximately 100 have currently been studied at the DNA level for their role in genetic disease (*4*). A number of genetic diseases are caused by alleles present in the population at relatively high frequencies, perhaps because of selective advantages to the heterozygous carriers (*5*). The ongoing characterization of disease-causing or disease-associated gene sequences makes large-scale screening for carrier status and genetic counseling a possibility. It may also sharpen the diagnostic accuracy for diseases such as autoimmune conditions where the susceptibility may be influenced by defined alleles. Such prospects are currently limited by the cumbersome nature of the available DNA detection methods.

A majority of polymorphisms in the human genome are caused by point mutations that involve one or a few nucleotides. Current DNA analysis procedures capable of detecting the substitution of a single nucleotide are based on differential denaturation of mismatched probes as in allele-specific oligonucleotide hybridization (*6*) or denaturing gradient gel electrophoresis (*7*). Alternatively, the sequence of interest can be investigated for polymorphisms that affect the recognition by a restriction enzyme (*8*) or that will allow ribonuclease A (RNase A) to cleave at mismatched nucleotides of an RNA probe hybridized to a target DNA molecule (*9*). Although denaturing gradient gel or RNase A can survey long stretches of DNA for mismatched nucleotides, they are estimated to detect only about half of all mutations that involve single nucleotides (*7*, *9*). Similarly, less than half of all point mutations give rise to gain or loss of a restriction enzyme cleavage site (*10*). The only existing technique capable of identifying any single nucleotide difference, short of DNA sequence analysis, is allele-specific oligonucleotide hybridization. This technique involves immobilizing separated (*6*) or enzymatically amplified (*11*) fragments of target DNA, hybridizing with oligonucleotide probes, and washing under carefully controlled conditions to discriminate single nucleotide mismatches.

We have devised a strategy that permits the facile distinction of known sequence variants differing by as little as a single nucleotide. The approach combines the ability of oligonucleotides to hybridize to the sequence of interest and the potential of a DNA-specific enzyme, T4 DNA ligase, to distinguish mismatched nucleotides in a DNA double helix (Fig. 1). Two oligonucleotide probes are permitted to hybridize to the denatured target DNA such that the 3′ end of one oligonucleotide is immediately adjacent to the 5′ end of the other. The ligase can then join the two juxtaposed oligonucleotides by the formation of a phosphodiester bond, provided that the nucleotides at the junction are correctly base-paired with the target strand. The ligation event thus positively identifies sequences complementary to the two oligonucleotides. A heterozygous sample is therefore scored as positive for both alleles. The joining of the oligonucleotides may be conveniently demonstrated, for instance, by labeling one of the oligonucleotides with biotin and the other one with ^{32}P. After the ligation reaction, the biotinylated oligonucleotides are allowed to bind to streptavidin immobilized on a solid support. Radioactive oligonucleotides that have become ligated to biotinylated oligonucleotides remain on the support after washing and are detected by autoradiography.

The gene encoding human β globin was selected as a model system to test the technique. There are two relatively frequent alleles, β^S and β^C, each differing from the normal allele, β^A, by a single nucleotide substitution in positions 2 and 1, respectively, of codon six (Figs. 2 and 3) (*12*). Subjects homozygous for the β^S allele suffer from sickle cell anemia. Moreover an increased risk of sudden death during exertion has been observed among individuals heterozygous for β^S (*13*).

The ligase-mediated gene detection procedure was used to distinguish β^A and β^S genes in equivalent amounts of DNA present in cells, in cloned DNA, and in genomic DNA (Fig. 2). One of two synthetic oligonucleotides (B131 or B132), specific for each of the alleles, was used in conjunction

Division of Biology, California Institute of Technology, Pasadena, CA 91125.

Best Available Copy

Supplied by The British Library - "The world's knowledge"

Alamar Exhibit 1002
Alamar v. Olink

with another oligonucleotide (P133) hybridizing immediately 3′ to either of the other two oligonucleotides on the target DNA strand. All of the synthetic oligomers used in this study are 20 nucleotides long. The ability of T4 DNA ligase to join the variable, 3′ nucleotide of the allele-specific oligonucleotides to the 5′ terminus of the invariant oligonucleotide was assessed by capturing any ligated product on streptavidin-agarose beads. The beads were filtered and washed to remove unbound oligonucleotides, and then the filter with trapped beads was exposed to x-ray film. The 10^5 nucleated cells used for one assay were obtained from ~0.5 ml of blood. The cells were used in the assay without DNA purification, by first making the DNA accessible for the ligase-mediated analysis by sequential additions of a nonionic detergent (Triton X-100) and a protease (trypsin). The DNA was denatured with alkali and then soybean trypsin inhibitor was added to prevent proteolysis of the added ligase.

The described ligation reactions were performed at 37°C, ~25 K below the melting temperature of the hybridized oligonucleotides, permitting the use of standardized assay conditions independent of the particular sequence investigated. The observed specificity is a consequence of the requirement for the simultaneous hybridization of both oligonucleotides in a precisely juxtaposed position. Although both oligonucleotides are likely to hybridize to numerous sequences in the DNA sample, they are unlikely to do so in the appropriate head-to-tail fashion except where the proper target sequence is present. In addition, we have found that the ligation reaction requires that the two terminal nucleotides on either side of the junction of the two oligonucleotides be engaged in correct base-pairing. This requirement further suppresses incorrect ligation events.

To determine whether any type of single nucleotide mismatch could be distinguished from correct base-pairing with the present method, we used four synthetic target molecules representing a segment of the β-globin gene, each with a different nucleotide in the first position of the sixth codon. Two of the sequences are derived from the β^A and β^C alleles of the β-globin gene. The other two sequences represent the other possible nucleotides occupying the variant position. Four pairs of oligonucleotides were designed to specifically identify one of the target molecules. Four oligonucleotide probes, each with a different nucleotide in the 3′ terminal position and complementary to one of the target molecules, were separately assayed for their ability to be ligated to an invariant oligonucleotide that hybrid-

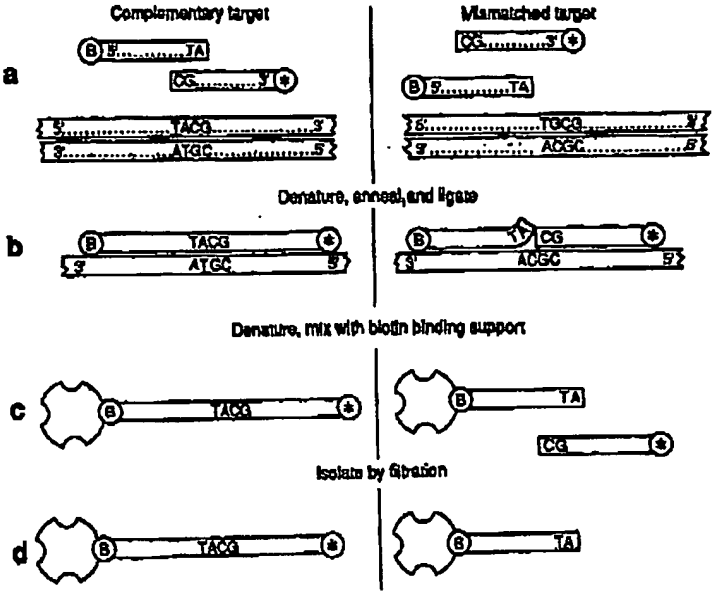


Fig. 1. A diagram depicting gene detection through the ligation of hybridized oligonucleotide probes. Target DNA is denatured and mixed with oligonucleotides and ligase. The ligase joins pairs of oligonucleotides annealed head to tail if they are correctly base-paired at the junction. Radioactively labeled oligonucleotides (*) are immobilized and detected by autoradiography only if ligated to biotinylated oligonucleotides (B) that can be bound to streptavidin on a solid support.

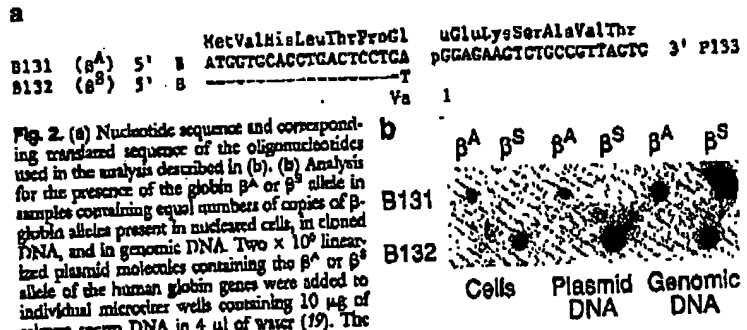


Fig. 2. (a) Nucleotide sequence and corresponding translated sequence of the oligonucleotides used in the analysis described in (b). (b) Analysis for the presence of the globin β^A or β^S allele in samples containing equal numbers of copies of β-globin alleles present in nucleated cells, in cloned DNA, and in genomic DNA. Two × 10^5 linearized plasmid molecules containing the β^A or β^S allele of the human globin genes were added to individual microliter wells containing 10 μg of salmon sperm DNA in 4 μl of water (*19*). The microtiter plates were centrifuged and the supernatants removed. To the resuspended cell pellet was added 1 μl of 10% Triton X-100 and 1 μl of trypsin at 2 μg/μl. The samples were incubated at 37°C for 30 min and were denatured with alkali as above. The pH was neutralized and 1 μl of soybean trypsin inhibitor (Sigma, 10 μg/μl) was added. Each well received 140 fmol of biotinylated oligonucleotides B131 or B132 (*20*), specific for the globin β^A and β^S genes, respectively, and 1.4 fmol of oligonucleotide P133, 5′ end-labeled with [γ-^{32}P]adenosine triphosphate (ATP) and polynucleotide kinase to a specific activity of 5 × 10^8 Cerenkov cpm/μg and purified over a Nensorb column (Du Pont Biotechnology Systems). T4 DNA ligase (0.05 Weiss unit, Collaborative Research) was added in 2 μl of 5× ligase buffer to a final volume of 10 μl containing 50 m*M* tris-HCl (*pH* 7.5), 10 m*M* $MgCl_2$, 150 m*M* NaCl (including 50 m*M* added during denaturation), 1 m*M* spermidine, 1 m*M* ATP, 5 m*M* dithiothreitol, and 100 ng of bovine serum albumin per microliter. The reagents were mixed by briefly centrifuging the microtiter plates before incubating at 37°C and 100% humidity for 5 hours. The ligated oligonucleotides were denatured by the addition of 1 μl of 1.1*M* NaOH and incubated for 10 min at 37°C. After the incubation, 1 μl of 1.1*M* HCl and 2 μl of 10% SDS were added. Three microliters of a 15% (v/v) suspension of streptavidin-coated agarose beads (Bethesda Research Laboratories) was then added, and the plate was incubated on a shaking platform at room temperature for 5 min. The contents of the wells were transferred to a dot blot manifold (Schleicher and Schnell) with a Whatman filter paper no. 4. In order to reduce nonspecific binding of the labeled oligonucleotides, the filter papers had been boiled and the beads diluted in 0.5% (v/v) dry nonfat milk, 1% SDS, and salmon sperm DNA (100 μg/ml). The beads (2*t*) were washed under suction in the manifold with 3 ml of 1% SDS and 1 ml of 0.1*M* NaOH per sample, with a 96-tip dispenser (Vacu-pette/96, Culture Tek). The filters were wrapped in plastic wrap and autoradiographed for 3 days at −70°C with one enhancing screen (Du Pont).

Best Available Copy

Supplied by The British Library - "The world's knowledge"

Alamar Exhibit 1002
Alamar v. Olink

ized immediately 3′ to the first oligonucleotide. These reagents permit studying the effect on ligation by any of the 16 possible base pairs, the 4 correct Watson-Crick pairs and 12 mismatched pairs, in an invariant sequence context. Under the appropriate conditions, only nucleotides engaged in correct base-pairing were efficiently joined by ligation (Fig. 3). Parameters that affected the nucleotide specificity were the salt concentration and the amount of enzyme added relative to the DNA concentration. Higher salt concentration and lesser amounts of enzyme than those found to be optimal for discrimination resulted in loss of signal. The above experiment cannot exclude the possibility that the identification of mismatched nucleotides may be influenced by the surrounding sequence, although we have not yet encountered any evidence for such effects.

Although autoradiographic techniques are relatively simple to implement, a gene detection assay based on the use of fluorescent rather than radioactive probes would have the advantages of safe handling, more stable reagents, and rapid access to the results, and would allow for multicolor analysis by using fluorophores with different emission spectra. In general, conventional organic fluorophores are less sensitive labels than ^{32}P. Thus we increased the amount of target DNA before the detection assay with the polymerase chain reaction (*14*). With this procedure a segment of DNA can be exponentially amplified by repeated cycles of enzymatic synthesis of new strands from two oligonucleotide primers, one with a sequence derived upstream and the other in the opposite orientation downstream of the segment of interest. Genomic DNA was obtained from three human cell lines, MOLT-4, which is homozygous for the β^A-globin allele; SC-1, homozygous for the β^S allele; and GM2064, in which the β-globin locus has been deleted (*15*). The appropriate segment of the β-globin gene was amplified in 25 cycles from 1 μg of genomic DNA from each cell line. We used 3-μl aliquots, equivalent to 24 ng of genomic DNA for the assay. Two oligonucleotides, specific for the β^A and β^S alleles and differentially 5′-labeled with one of two fluorophores, were present at equal concentrations. The amount of each of these oligonucleotides that became ligated to a third oligonucleotide hybridizing downstream of the other two was determined by separating the reaction products on an 8% polyacrylamide gel and analyzing the band migrating as a 40-nucleotide oligomer (the size of two ligated oligonucleotides) for the relative contribution by the two different fluorophores [model 370A DNA sequencer, Applied Biosystems, Foster City, California (*16*)]. No signal was observed when the β-globin gene had been deleted in the cell from which the DNA was obtained, whereas only the correct fluorophore-labeled oligonucleotide was ligated when the cells harbored the β^A or β^S alleles

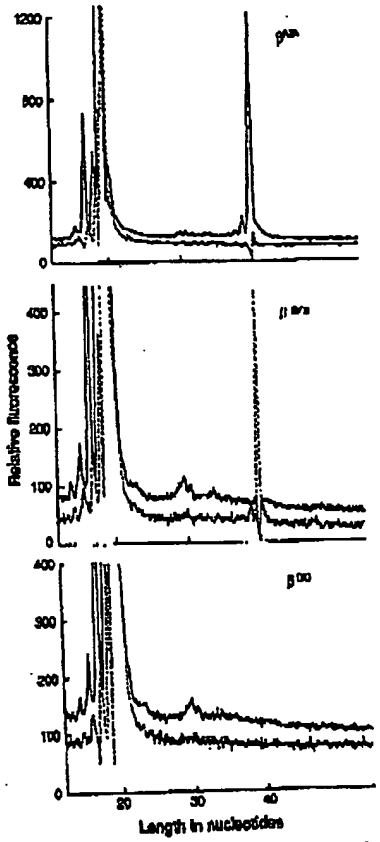


Fig. 4. Demonstration of the presence of the β^A and β^S alleles of the β-globin gene in amplified genomic DNA by probes labeled with fluorescent dyes. A 120-bp segment of the β-globin gene was amplified with the polymerase chain reaction as described (*16*) in 25 cycles starting with 1 μg of genomic DNA from the cell lines MOLT-4, SC-1, and GM2064 ($\beta^{A/A}$, $\beta^{S/S}$, and $\beta^{0/0}$, respectively) in 100 μl. Three microliters of each amplified sample was added to an Eppendorf tube, denatured by alkali, neutralized, and incubated with 14 fmol each of oligonucleotide 131 labeled with carboxy-fluorescein (Molecular Probes) (CF131) (——) and oligonucleotide 132 labeled with carboxy-2′,7′-dimethoxy-4′,6′-dichlorofluorescein (CD132) (– – –), and 14 fmol of nonradioactively 5′ phosphorylated oligonucleotide P133 (for sequences, see Fig. 2). The reaction conditions were essentially as described in Fig. 2, but 0.5 Weiss units of T4 DNA ligase was added to each assay. At the end of the 3-hour incubation, the samples were ethanol precipitated, taken up in 50% formamide, and loaded on a sequencing gel in an ABI 370A automated DNA sequencer. The fluorescence signal was processed to distinguish the partially overlapping emission spectra of the two fluorophores and to determine the relative contribution of each fluorophore to the signal.

a

```
B128  (βA)  5' B  GATGGTCCACCTGACTCCTG pAGGAGAAGTCTGCCGTTACT 3' P129
B134  (βC)  5' B  ------------------A
B136        5' B  ------------------T
B137        5' B  ------------------C

172   (βA)  3'    CTACCACCTGGACTGAGGACTCCTCTTCAGACGGCAATGA   5'
138   (βC)  3'    ------------------T----------------------  5'
139         3'    ------------------A----------------------  5'
140         3'    ------------------G----------------------  5'
```

b

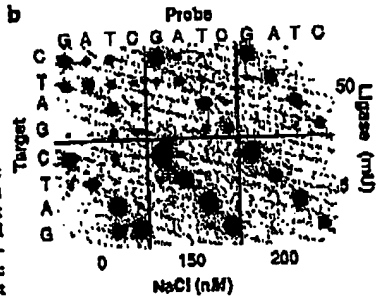


Fig. 3. (a) Nucleotide sequence of the oligonucleotides used in the analysis described in (b). (b) Correct identification of four target molecules, differing by single-nucleotide substitutions in one position. Letters refer to the variable nucleotides in the probe and target sequences. As target molecules, 40-nucleotide oligomers, derived from the β-globin gene sequence, were synthesized. The oligonucleotides 172, 138, 139, and 140 are of identical sequence except in a central position where each target molecule includes a different nucleotide. Four 20-nucleotide biotinylated oligomers, B128, B134, B136, and B137, differing only in their 3′ nucleotide position, were designed to hybridize to the 3′ half of the target molecules such that the variant position of the probe reagents corresponds to that of the target molecules. Each of the biotinylated oligonucleotides was used in conjunction with oligonucleotide P129, 5′ end-labeled with ^{32}P and hybridizing immediately 3′ to the biotinylated probes on the target strands. The assays were performed essentially as described in the legend to Fig. 2, but 2×10^4 copies of one of the target molecules were added to each well with 10 μg of salmon sperm DNA. Each well further received one of the biotinylated oligonucleotides together with oligonucleotide P129. The final NaCl and ligase concentrations were varied as indicated.

Best Available Copy

Supplied by The British Library - "The world's knowledge"

Alamar Exhibit 1002
Alamar v. Olink

Best Available Copy

(Fig. 4). This strategy could be generalized to the simultaneous analysis of several loci. For each set of two labeled, allele-specific oligonucleotides and one unlabeled, the latter is given a nonhybridizing 3′ sequence extension of a unique length. This results in different migration rates for the ligation products, characteristic of each locus.

In contrast to gene detection techniques based on immobilizing the target DNA, such as DNA blots, the hybridization reported here was performed in solution and in a small volume, which reduced the time required for hybridization (*17*). It also obviated the step of immobilizing the target DNA. Both ligation and binding of the biotinylated oligonucleotides are efficient and rapid steps that should permit quantitative detection of target molecules. In general, there are three rate-limiting steps in gene detection techniques. The first is sample preparation, which can be greatly simplified as demonstrated here. The second is the time required for the probes to anneal to the target sequence. This is a function of the concentration of the probe and can be reduced considerably. The third and most time-consuming step in the present technique is signal detection by autoradiography. A sufficiently sensitive fluorescent detection method (*18*) should drastically reduce this time, permitting the development of a rapid, automated gene detection procedure.

REFERENCES AND NOTES

1. C. T. Caskey, *Science* 236, 1223 (1987).
2. A. Svejgaard *et al.*, *Immunol. Rev.* 70, 193 (1983); J. L. Tiwari and P. I. Terasaki, *HLA and Disease Associations* (Springer-Verlag, New York, 1985).
3. V. A. McKusick, *Mendelian Inheritance in Man* (Johns Hopkins Univ. Press, Baltimore, ed. 7, 1986).
4. D. N. Cooper and J. Schmidtke, *Lancet* I, 273 (1987).
5. J. I. Rotter and J. M. Diamond, *Nature* 339, 289 (1987).
6. B. J. Conner *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* 80, 278 (1983); N. M. Whitley, M. W. Hunkapiller, A. N. Glazer, European Patent application 0 186 494 A2 (1986).
7. R. M. Myers *et al.*, *Nature* 313, 495 (1985).
8. R. F. Geever *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* 78, 5081 (1987).
9. R. M. Myers, Z. Lerin, T. Maniatis, *Science* 230, 1242 (1985); B. Wentz, F. Yamamoto, C. Almoguera, M. Perucho, *Proc. Natl. Acad. Sci. U.S.A.* 82, 7578 (1985).
10. T. Hunkapiller, personal communication.
11. R. K. Saiki, T. L. Bugawan, G. T. Horn, K. B. Mullis, H. A. Erlich, *Nature* 324, 163 (1986).
12. R. M. Winslow and W. F. Anderson, in *The Metabolic Basis of Inheritance*, J. B. Stanbury, J. B. Wyngaarden, D. S. Fredrickson, J. L. Goldstein, M. S. Brown, Eds. (McGraw-Hill, New York, 1983), pp. 1666–1710.
13. J. A. Kark, D. M. Posey, H. R. Schumacher, C. J. Ruehle, *N. Engl. J. Med.* 317, 781 (1987).
14. R. K. Saiki *et al.*, *Science* 230, 1350 (1985).
15. R. K. Saiki, N. Arnheim, H. A. Erlich, *Biotechnology* 3, 1008 (1985).
16. L. M. Smith *et al.*, *Nature* 321, 674 (1986); C. Connell *et al.*, *BioTechniques* 5, 342 (1987).
17. J. G. Wetmur and N. Davidson, *Mol. Biol.* 31, 349 (1968).
18. E. Soini and H. Rojola, *Clin. Chem.* 29, 65 (1983).
19. Genomic DNA was purified from guanidinium HCl-solubilized cells as described (D. Boswell, *Anal. Biochem.* 162, 463 (1987)] and resuspended by boiling before adding 7 μg in 4 μl of water per assay well. The plasmid and genomic DNA samples were denatured by adding 1 μl of 0.5*M* NaOH and incubating for 10 min at 37°C before restoring the pH with 1 μl of 0.5*M* HCl. Alternatively, samples of uncleaned blood cells were used directly as a source of DNA for the analysis. Cells (10^6), obtained by Ficoll-Hypaque (Pharmacia) flotation, were added in 50 μl of phosphate-buffered saline to each well.
20. The oligonucleotides were assembled by the phosphoramidite method [S. J. Horvath, J. R. Firca, T. Hunkapiller, M. W. Hunkapiller, L. Hood, *Methods Enzymol.* 154, 314 (1987)] on an Applied Biosystems model 380A DNA synthesizer and purified either by polyacrylamide gel electrophoresis or reversed-phase high-pressure liquid chromatography (HPLC). Biotinylation was performed by reacting a biocin *N*-hydroxysuccinimide ester (Enzotin, Enzo) with a 5′ aminohexylthymidine residue incorporated in the oligonucleotide [L. M. Smith, S. Fung, T. J. Hunkapiller, M. W. Hunkapiller, L. Hood, *Nucleic Acids Res.* 13, 2399 (1985)]. The product was purified by reversed-phase HPLC.
21. The size of the area on which the beads were deposited was reduced by interposing a 3-mm-thick plexiglass disk with conical holes with diameters of 5 mm on the upper surface and 3 mm on the lower.

22. The authors acknowledge a stipend from the Knut and Alice Wallenberg Foundation to U.L. and support from NSF grant BNS 87 14416, Defense Advanced Research Projects Agency grant N00014-86K-0753, Upjohn Company, and Applied Biosystems, Inc. The oligonucleotides were synthesized by S. J. Horvath and the fluorescence data were analyzed by C. Dodd. R. K. Saiki provided plasmids and samples of genomic DNA obtained from cell lines. J. Korenberg and K. Tanaka made available blood samples from sickle cell patients. The *N*-hydroxysuccinimide ester of carboxy-2′,7′-dimethoxy-4′,5′-dichlorofluorescein was provided by M. W. Hunkapiller. We acknowledge discussions with B. Korber, B. Poplin, A. Kamb, N. Lan, L. Smith, R. Barth, V. A. McKusick, J. Richards, and M. Simon.

11 April 1988; accepted 13 June 1988

Amyloid Protein Precursor Messenger RNAs: Differential Expression in Alzheimer's Disease

M. R. PALMERT, T. E. GOLDE, M. L. COHEN, D. M. KOVACS, R. E. TANZI, J. F. GUSELLA, M. F. USIAK, L. H. YOUNKIN, S. G. YOUNKIN*

In situ hybridization was used to assess total amyloid protein precursor (APP) messenger RNA and the subset of APP mRNA containing the Kunitz protease inhibitor (KPI) insert in 11 Alzheimer's disease (AD) and 7 control brains. In AD, a significant twofold increase was observed in total APP mRNA in nucleus basalis and locus ceruleus neurons but not in hippocampal subicular neurons, neurons of the basis pontis, or occipital cortical neurons. The increase in total APP mRNA in locus ceruleus and nucleus basalis neurons was due exclusively to an increase in APP mRNA lacking the KPI domain. These findings suggest that increased production of APP lacking the KPI domain in nucleus basalis and locus ceruleus neurons may play an important role in the deposition of cerebral amyloid that occurs in AD.

ALZHEIMER'S DISEASE (AD) IS characterized pathologically by large numbers of senile plaques and neurofibrillary tangles throughout the cerebral cortex and hippocampus. Senile plaques consist of clusters of degenerating neurites surrounding an amyloid core composed of 5- to 10-nm fibrils that stain metachromatically with Congo red. In many cases of AD, amyloid fibrils are also found in vessel walls (*1*). A 4.2-kD polypeptide, referred to as A4 or the β protein, has been isolated from the amyloid fibrils found in senile plaques (*2*) and vessel walls (*3*) of patients with AD. There is evidence that A4 may also be a component of the paired helical filaments found in neurofibrillary tangles (*4*).

The gene encoding A4, which is located on chromosome 21 (*5*), produces at least three mRNAs (Fig. 1) referred to as APP_{695}, APP_{751}, and APP_{770} (*6–8*). APP_{695}, the mRNA that was initially identified (*5*), encodes an amyloid protein precursor (APP), 695 amino acids in length, that includes A4 at positions 597 to 638. APP_{751} is identical to APP_{695}, except for a 168-nucleotide insert (*6–8*). This insert, previously referred to as HL124i (*7*), would introduce 56 amino acids carboxyl terminal to Arg^{288} and convert Val^{289} into an isoleucine. APP_{770} is identical to APP_{751}, except for a 57-nucleo-

M. R. Palmert, T. E. Golde, M. F. Usiak, L. H. Younkin, S. G. Younkin, Division of Neuropathology, Institute of Pathology and Department of Pharmacology, Case Western Reserve University School of Medicine, Cleveland, OH 44106.
M. L. Cohen and D. M. Kovacs, Division of Neuropathology, Institute of Pathology, Case Western Reserve University School of Medicine, Cleveland, OH 44106.
R. E. Tanzi and J. F. Gusella, Neurogenetics Laboratory, Massachusetts General Hospital and Department of Genetics and the Program in Neuroscience, Harvard Medical School, Boston, MA 02115.

*To whom correspondence should be addressed at the Department of Pharmacology.

Supplied by The British Library - "The world's knowledge"

Alamar Exhibit 1002
Alamar v. Olink

Best Available Copy

Mutation detection and single-molecule counting using isothermal rolling-circle amplification

Paul M. Lizardi[1], Xiaohua Huang[2], Zhengrong Zhu[2], Patricia Bray-Ward[2], David C. Thomas[3] & David C. Ward[2]

Rolling-circle amplification (RCA) driven by DNA polymerase can replicate circularized oligonucleotide probes with either linear or geometric kinetics under isothermal conditions. In the presence of two primers, one hybridizing to the + strand, and the other, to the – strand of DNA, a complex pattern of DNA strand displacement ensues that generates 10^9 or more copies of each circle in 90 minutes, enabling detection of point mutations in human genomic DNA. Using a single primer, RCA generates hundreds of tandemly linked copies of a covalently closed circle in a few minutes. If matrix-associated, the DNA product remains bound at the site of synthesis, where it may be tagged, condensed and imaged as a point light source. Linear oligonucleotide probes bound covalently on a glass surface can generate RCA signals, the colour of which indicates the allele status of the target, depending on the outcome of specific, target-directed ligation events. As RCA permits millions of individual probe molecules to be counted and sorted using colour codes, it is particularly amenable for the analysis of rare somatic mutations. RCA also shows promise for the detection of padlock probes bound to single-copy genes in cytological preparations.

© 1998 Nature America Inc. • http://genetics.nature.com

Nucleic-acid amplification technology has greatly increased our ability to ask detailed questions about genotype or transcriptional phenotype in small biological samples, and has provided the impetus for many significant advances in biology, especially in the field of genetics. Recently, the utility of circularizable oligonucleotides, called 'padlock probes', was demonstrated in the detection of repeated alphoid sequences in metaphase chromosomes[1]. The high sequence specificity of padlock probes, and their potential for the discrimination of point mutations *in situ*[2] prompted us to explore the amplification of circular DNA. The ability of small circular oligonucleotides to serve as templates for DNA polymerases has been documented[3,4]. We have devised alternative methods for ligation-dependent circularization of padlock probes, as well as a method employing preformed circular probes, all of which can be exploited for allele discrimination. These methods have been coupled to isothermal nucleic-acid amplification reactions based on a rolling-circle replication mechanism.

We used a novel geometric hyperbranched RCA (HRCA) reaction to detect point mutations in small amounts of human genomic DNA in solution. Furthermore, linear RCA, using preformed circles as universal amplification templates, was used as a sensitive reporter system to visualize single hybridization/ligation events on surfaces that contain site-addressable probes. The linear RCA reporter system provides a new paradigm for genetic analysis at the molecular level, using a variety of formats ranging from surface-immobilized DNA to cytological specimens.

Linear amplification of allele-specific circularized probes

The circularizable probes described[1] contain two adjacent probe sequences of 20 bases. To increase specificity, we chose to use circularizable probes that hybridize to the target leaving a small gap of 6–10 nt. This gap can be filled by a short, allele-specific oligonucleotide with a 5′ phosphate (Fig. 1*a*) that fits exactly within the gap, completing a stacked duplex structure that is ligated with DNA ligase to form a closed padlock probe (Fig. 1*b*). Alternatively, the gap may be filled by DNA polymerase (see below).

We designed an allele-discriminating probe for a 46-nt target sequence in the *CFTR G542X* gene locus (Fig. 2*a*). The circularizable probe and either of the two alternative allele-specific gap oligonucleotides, each 8 nt, were incubated with an artificial DNA target to investigate sequence discrimination in a ligation reaction. The formation of the circularized oligonucleotide was assessed by its characteristic electrophoretic mobility (Fig. 2*b*). Using wild-type target, the slow-migrating circularized oligonucleotide is observed only when the corresponding wild-type gap oligonucleotide is present. The converse results were obtained using mutant target DNA. The ligation reaction was also strictly dependent on correct sequence complementarity of the gap probe sequence for another three probes of similar design (at two other loci in *CFTR*, and at one locus in *OTC*, data not shown).

An 18-base oligonucleotide complementary to the circularizable probe (Fig. 1*b*) was added after probe ligation to serve as a primer for RCA (Fig. 1*c*). To catalyze this linear RCA reaction, we used the DNA polymerase of phage ø29 (ref. 5), a highly processive enzyme that displays strand-displacing activity in the absence of additional proteins or cofactors[6]. We followed the time course of primer extension by gel electrophoresis in a denaturing alkaline agarose gel. After 10 min, the DNA product was larger than the 23-kb DNA marker (Fig. 3). The observed reaction rate is approximately 53 nt per second, a value consistent with previously published data[6] on the replication of single-stranded M13 DNA by ø29 DNA polymerase. Sequence analysis of cloned RCA products revealed that the amplified DNA contained the sequence expected for each different target-dependent, allele-specific, circle-ligation event (see Methods).

[1]*Department of Pathology, and* [2]*Department of Genetics, Yale University School of Medicine, 333 Cedar St., New Haven, Connecticut 06520, USA.* [3]*Oncor, Inc., 209 Perry Parkway, Gaithersburg, Maryland 20877, USA. Correspondence should be addressed to P.M.L. (e-mail: paul.lizardi@yale.edu) or D.C.W. (e-mail: ward@biomed.med.yale.edu).*

Alamar Exhibit 1002
Alamar v. Olink

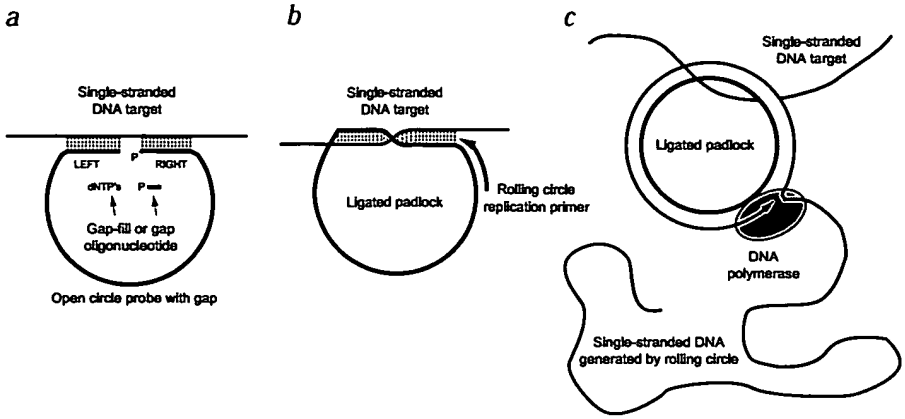


Fig. 1 Probe circularization by ligation, and amplification by a rolling circle-reaction. ***a***, Circularizable probe with a small gap, which is to be filled by binding and ligation of a small phosphorylated oligonucleotide, or by DNA polymerase incorporation of dNTP's proceeding from the 3'-OH end of the probe, and terminating at the junction with the 5' end of the probe, with concomitant ligation by DNA ligase. ***b***, Ligated (padlock) probe, and binding of complementary primer for rolling circle amplification. The primer 3' end is located five or six bases away from the last paired base in the hybridized probe arm. ***c***, Rolling-circle amplification of a padlock probe, catalyzed by a strand displacing DNA polymerase.

Allele discrimination using HRCA and genomic DNA

An alternative probe design for a 46-nt target sequence in the *CFTR G542X* gene locus (Fig. 2*c*) was used in this experiment. It consisted of an 18-base probe sequence at the 3′ end of the circularizable probe, and a 21-base probe sequence at the phosphorylated 5′ end. A gap of seven bases is formed, which will be filled by extension of the 3′-hydroxyl end of the probe using *Thermus flavus* DNA polymerase[7]. The gap-fill reaction, followed by ligation, generates a circular probe harbouring a faithful copy of a small segment of the target sequence. Using artificial DNA targets, the formation of circles after incubation in the presence of DNA polymerase and DNA ligase was found to be quantitative as assessed by gel electrophoresis (Fig. 2*d*). As expected, when polymerase was added in the absence of ligase, the probe was extended, but not circularized.

To facilitate allele discrimination in genomic DNA, we designed a second oligonucleotide primer to bind specifically to the tandem DNA generated by RCA of the gap-fill probe. This primer, which is not complementary to the first primer, will bind to each complementary sequence in the tandem single-stranded DNA, initiating sequential primer extension reactions (Fig. 4). As each extending primer runs into the product of a downstream primer, strand displacement will ensue, generating single-stranded tandem repeats of the sequence of the original circularized probe. This displaced strand will in turn contain multiple binding sites for the first RCA primer. Thus, alternate-strand copying and strand displacement processes generate a continuously expanding pattern of DNA branches connected to the original circle. Strand displacement also generates a discrete set of free DNA fragments comprising double-stranded pieces of the unit length of a circle, and multiples thereof. We call this expanding cascade of strand displacement and fragment-generation events 'DNA hyperbranching', and the special rolling circle amplification driven by two primers 'Hyperbranched-RCA' (HRCA).

HRCA reactions were performed using the exonuclease(–) variant of Vent DNA polymerase, in the presence of phage T4 gene 32 protein[8] at 62–66.5 °C. When seeded with known amounts of preformed circular oligonucleotides, the reactions generate a large DNA output within 60–90 min. The size distribution of the amplified DNA comprises a ladder of bands starting at unit circle length, and extending in discrete increments to several thousand nucleotides (Fig. 5*b*), as predicted by the DNA hyperbranching

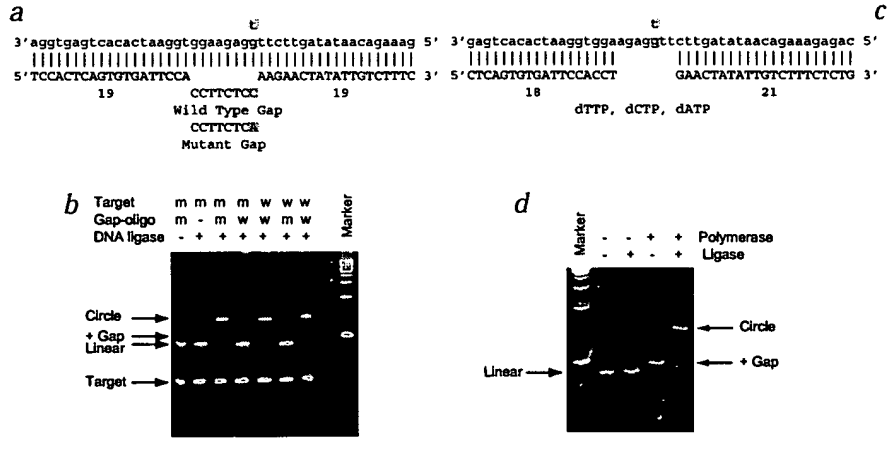


Fig. 2 Ligation of circularizable probes by Ampligase, a thermostable DNA ligase. ***a***, Sequence of the hybridizing arms of a circularizable probe, and two alternative 8-base gap probes designed for the *CFTR G542X* locus. The probe consists of a 19-base probe sequence at the 3′ end, and a 19-base probe sequence at the phosphorylated 5′ end. The probe is 85 nt (see sequence in Methods), and 47 nt of the sequence, constituting the non-hybridized backbone, are arbitrary. An 8-base oligonucleotide that is phosphorylated at the 5′ end is designed to fit in the gap by base pairing with the target. The base at the 3′ end of the 8-nt gap probe is either C or A, corresponding to the complement of the wild type or mutant allele, respectively. The artificial target was 53 nt, but only 46 bases are shown in the illustration. Published studies on *T. thermophilus* DNA ligase[12] have demonstrated that discrimination is highest when the mismatched base is at the 3′ end of the oligonucleotide to be ligated, as is the case for the probes shown here. ***b***, Analysis by gel electrophoresis of circularizable probes ligated with artificial DNA targets in the presence of different gap oligonucleotides. The gels were standard 8% polyacrylamide in Tris-Borate EDTA containing 8M urea, and staining was performed with SYBR-Green II (Molecular Probes). Letters indicate mutant (m) or wild type (w) sequences. The marker lane contains a commercial DNA ladder, where the fastest-migrating band is 100 bases. The ligation efficiency of the non-cognate gap oligonucleotide is low even for the 5′ terminus, which in all cases contains correct sequence complementarity (notice the very low amount of product of size 85+8=93 nt, indicated by the +Gap arrow). This may be explained by the ligation temperature, which is higher than the T_m of the gap oligonucleotide and may hinder the formation of a stable double helix for mismatched DNA. ***c***, Sequence of the hybridizing arms of a circularizable gap-fill probe designed for the *CFTR G542X* mutation locus. The artificial target was 53 nt, but only 46 bases are shown in the illustration. The complete sequence of the 89-nt circularizable probe is in the Methods section. Gap-filling occurs in the presence of dTTP, dCTP and dATP. ***d***, Analysis by gel electrophoresis of an 89-base circularizable probe that was gap-filled and ligated with artificial DNA targets in the presence or absence of DNA polymerase and DNA ligase, as indicated by plus (+) and minus (–) signs. The first lane contains a 100-bp DNA marker. Electrophoresis was performed as in (*b*).

© 1998 Nature America Inc. • http://genetics.nature.com

Alamar Exhibit 1002
Alamar v. Olink

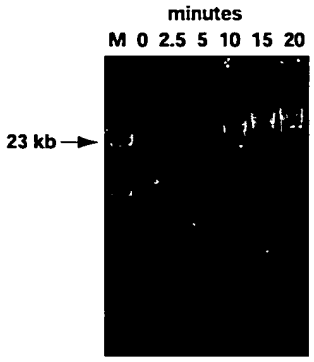


Fig. 3 Analysis of RCA products. Time course of amplification of circularized probes, catalyzed by ø29 DNA polymerase. Circularized padlock probes were prepared as in Fig. 2*b*. Aliquots of the RCA reaction were taken at the times indicated, denatured in alkaline buffer and analysed on a 0.7% alkaline agarose gel. The lane marked as 'M' contained phage lambda DNA digested with restriction endonuclease *Hin*dIII. The largest DNA fragment of 23 kb is indicated by the arrow.

diagram (Fig. 4). The data demonstrates that the HRCA reaction catalyzed by exo(–) Vent DNA polymerase can be initiated with as few as 20 molecules of closed circles, and after 90 min, produces sufficient material for detection. We estimate that each individual circle in the reactions initiated with 20 molecules generated at least 5×10^9 copies of the 96-base repeat during HRCA.

In order to demonstrate the use of HRCA in gene detection, we analysed genomic DNA from tissue-culture cells of defined genotype in an allele-specific amplification assay. Human lymphocytes that were either wild-type, homozygous or heterozygous for the *G542X* locus were used as a DNA source. An assay was performed where *G542X* probes of the gap-fill design (Fig. 2*c*) were extended and ligated using genomic DNA targets, and amplified using HRCA. The amplification reactions employed one common primer, complementary to the backbone sequence of the probe, and either of two allele-specific reverse primers, designed to be selective for each of the two alternative alleles (Fig. 5*a*). The discrimination of single-base changes using allele-specific primers is well documented for PCR (ref. 9). We used primers that contain a deliberately mismatched base at position –3 relative to the 3′-hydroxyl end, in addition to the 3′-terminal discriminating base, in order to improve specificity. After amplification, the DNA was cleaved with a restriction endonuclease whose recognition sequence occurs once in the probe, to generate a single band of repeat-sized fragments. Amplified DNA was generated only in those reactions in which the correct allele-specific primer was used (Fig. 5*c*). In heterozygous DNA samples, the intensity of the band of amplified DNA was somewhat lower, as expected for the halved gene dosage. In a sample containing a mixture of 90% wild-type DNA and 10% mutant DNA, a weaker band corresponding to the mutant allele was observed with good reproducibility in different experiments (lane 11, Fig. 5*c*). We used HRCA for the amplification of another 10 different DNA targets with excellent results, using either Vent DNA polymerase or Bst large fragment DNA polymerase (D.C.T., unpublished data).

Detection of individual ligation events by single-molecule counting

RCA can be used as a reporter system for quantifying hybridization/ligation events on a glass surface by single-molecule analysis. We prepared slides containing an oligonucleotide probe (P1) specific for a 39-base sequence adjacent to the G542X locus of the CFTR gene. The probe contained a free 5′ phosphate, and was bound covalently to the glass surface *via* a reactive 3′-amino group. Two additional probes (P2wt, P2mu) were designed, capable of being ligated to either the wild-type or mutant locus, with precise base stacking continuity with the 5′ end of the P1 probe. An allele-specific base was located at a 3′-hydroxyl terminus in these probes, for optimal discrimination in the ligation step; the opposite end of these probes comprises a coded primer sequence (corresponding to one of two alternative primers) with a free 3′-OH terminus, obtained by reversal of backbone polarity during chemical synthesis. The ligation of P1 with the allele-specific probes P2wt or P2mu (Fig. 6) generates a surface-bound oligonucleotide with a free 3′ terminus, competent for coded priming of an RCA reaction. We used two different pre-formed circular DNA templates for RCA signal coding, one designed to be complementary to the primer sequence of P2wt, the other to the P2mu primer. One of these two circular molecules will enable RCA of its cognate primer, in the event that the complementary primer becomes covalently bound to the surface by a ligation reaction. In the absence of ligation, no priming can occur. The DNA generated by RCA is labelled with

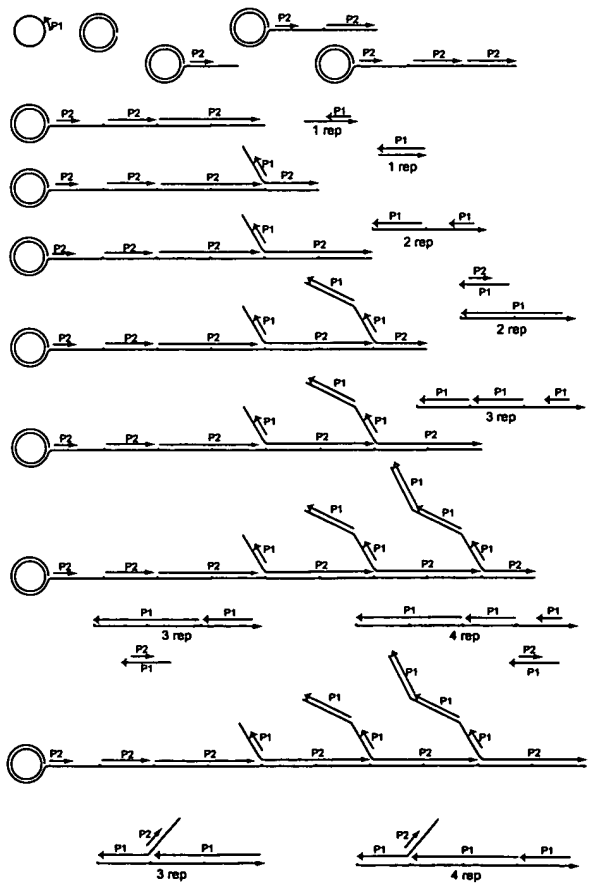


Fig. 4 Rolling-circle amplification of a circularized probe using two primers. The first primer (P1) initiates an RCA reaction, and the reverse primer (P2) binds to each tandem repeat generated by the rolling circle. Multiple priming events are initiated by P2 as the original RCA strand elongates. As these priming events elongate and generate displaced DNA strands, new priming sites for the first primer (P1) are generated. To follow the sequence of strand displacement events, note that as the reverse primer P2 binds to the fifth repeat, the primer at the third repeat begins to displace a branch; subsequently, as P2 binds to the seventh repeat, the elongating strand at the fifth repeat begins to displace a branch, and so forth. By the time a reverse primer binds to the tenth repeat, the DNA product already contains three growing branches. New primer extension events initiated in released DNA molecules also generate branches, as shown at the bottom of the figure. As the displaced DNA becomes completely double-stranded, it accumulates in fragments of unit length containing one, two, three or four repeats (shown as: 1 rep, 2 rep, 3 rep, 4 rep). Thus, in the presence of a circular template, the two primers generate a self-propagating, ever-increasing pattern of alternating strand-displacement, branching and DNA fragment release events, which we call hyperbranching.

© 1998 Nature America Inc. • http://genetics.nature.com

Alamar Exhibit 1002
Alamar v. Olink

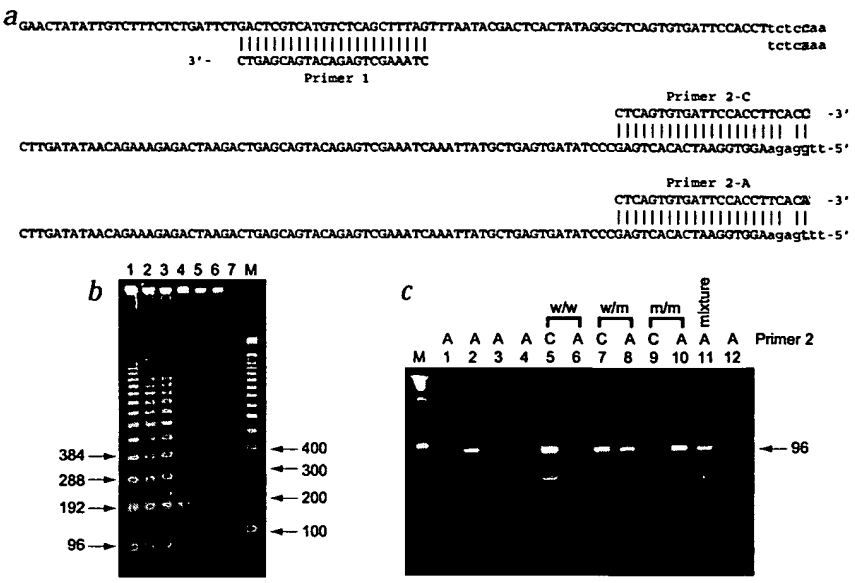


Fig. 5 HRCA of circularized DNA and its use in allele detection. *a*, Design of primers used to amplify an 89-base probe that had been extended by DNA polymerase copying of 7 nt (shown in small type) from the *G542X* target region, and circularized by DNA ligase. Primer 1 is the same for both reactions, and primer 2-C and primer 2-A are allele-specific reverse primers, binding to unique sequences in the complementary strands generated by copying of the circular probe. The calculated T_m of primer 1 is 68 °C, whereas the calculated T_m for the 20 bases that hybridize upstream of the –3 mismatch in primers 2-C and 2-A is 66.8 °C, using 50 mM salt and 0.1 μM oligonucleotide as parameters for the nearest neighbour T_m calculations[19]. *b*, Reaction products generated at 65.5 °C with different inputs of artificially made circular DNA, and analysed in a non-denaturing 2% agarose gel. Lanes 1–7 were seeded, respectively, with 2×10^5, 2×10^4, 2×10^3, 2×10^2, 2×10^1, 2×10^1 (repeat), or zero molecules of *G542X* circularized mutant probe. Lane 8 contains a ladder of markers in multiples of 100 bp. *c*, Detection of the *G542X* point mutation. DNA samples from homozygous wild type, heterozygous or homozygous mutant cultured lymphocytes were denatured by heating for 4 min at 96 °C, and mixed with the gap-fill open circle probe for the *G542X* locus shown in 6A. After filling with *T. flavus* DNA polymerase and ligation with Ampligase, a 4-μl aliquot of the ligated material was incubated in an HRCA reaction containing the appropriate set of two primers. The net DNA input for each amplification reaction consisted of 28 ng of genomic DNA, which corresponds to 17,000 copies of the gene locus. Amplified DNA was cleaved with *AluI* and analysed by electrophoresis in a non-denaturing 8% acrylamide gel. Letters indicate mutant (m) or wild type (w) DNA haploid complements. The lane labelled M contains a 100-bp marker DNA. Other lanes contain HRCA reactions initiated with different inputs. Lanes: 1, zero DNA input; 2, 2×10^4 closed circles; 3, 2×10^3 closed circles; 4, ligation with phage lambda DNA; 5, 6, ligation with GM07828(+/+) DNA; 7, 8, ligation with GM11497B(+/–) DNA; 9, 10, ligation with GM11496(–/–) DNA; lane 11, ligation with an artificial 10:1 mixture of GM07828(+/+) DNA and GM11496(–/–) DNA; lane 12, ligation in the absence of genomic DNA. The faint band below the main 96-base Alu fragment corresponds with the primer-induced deletion fragment repeats, which comprise a full unit repeat minus the 22-base segment that separates the 5′ ends of the first primer and the reverse primer (Fig. 6*a*). Four different batches of T4 Gene 32 protein have been used successfully for HRCA reactions. Although the batch used in (*c*) was less stimulatory than the other three batches, it still allowed detection of the low-abundance mutant allele in lane 11.

© 1998 Nature America Inc. • http://genetics.nature.com

fluorescent DNP-oligonucleotide tags that hybridize at multiple sites in the tandem DNA sequence. The 'decorated' DNA, labelled by specific encoding tags, is then condensed into a small object by cross-linking with a multivalent anti-DNP IgM. The wild-type specific primer generates RCA products which can hybridize to fluorescein-labelled DNP-oligonucleotide tags, whereas the mutant RCA products hybridize to Cy3-labelled DNP-oligonucleotides. The acronym for the process of condensation of amplified circles after hybridization of encoding tags is CACHET.

We performed an assay to measure the ratio of mutant to wild type strands at the G542X locus in genomic DNA samples that had been constructed to simulate the presence of rare somatic mutations. DNA mixed in different ratios was amplified by PCR, and hybridized on slides with immobilized P1 probes, in the presence of an equimolar mixture of P2wt and P2mu probes in solution. After ligation of the P2 probes, signals were generated by RCA-CACHET and imaged as described in Methods. The images show many hundreds of fluorescent dots (Fig. 7*a–d*) with a dia-

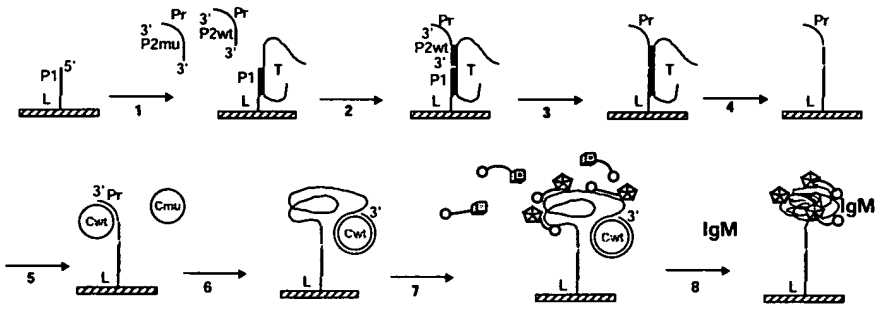


Fig. 6 Design of the RCA-CACHET ligation-dependent assay using immobilized DNA probes. A derivatized glass surface contains an oligonucleotide probe (P1) which is immobilized *via* a spacer (L), bound covalently to the glass. P1 is designed to form 39 bp with the G542X target, and the 5′ terminus of P1 contains a 5′-phosphate to permit ligation. This orientation is preferred because it eliminates the possibility of nonspecific priming by the 3′ end of P1, which could otherwise interact with the circular oligonucleotide templates used for RCA. A set of two allele-specific oligonucleotide probes (P2mu and P2wt) that are linked to different primer sequences (Pr, green or red) is allowed to hybridize with a DNA target (T). These probes, present in solution, are designed to hybridize to a 20-base sequence of the target adjacent to P1, with their 3′ end precisely in stacking contact with the 5′ end of P1, so that P1 and P2 may be ligated. P2-wt contains a 3′-terminal 'G', whereas P2mu contains a 3′-terminal 'T'. Both P2wt-Pr and P2mu-Pr contain at the opposite end a sequence that does not hybridize with the target, so that it may serve as a primer. Therefore, these molecules are synthesized with reversed backbones, and have two 3′ ends. After hybridization of the cognate probe to target, which in the case shown is a wild-type sequence, a thermostable DNA ligase catalyzes the joining of P2wt-Pr to the immobilized P1 probe. Subsequently, the targets, excess probes and any other molecules that are not covalently linked to the solid support are removed by stringent washing. A mixture of two types of circular oligonucleotides, Cwt and Cmu, are hybridized to the primer (Pr, green), which in the case illustrated is complementary to Cwt. The primer is then extended by RCA, using a circular CTwt oligonucleotide as a template. The elongated DNA molecule is then decorated by hybridization of DNP-oligonucleotide tags that harbour either fluorescein or Cy3 fluorescent labels. In the case shown, only the green tags are competent for binding, as the amplified circle only contains sequences complementary to the green tags. The amplified DNA product is finally condensed with anti-DNP IgM, forming a small globular DNA:IgM aggregate that contains green fluorescent tags.

Alamar Exhibit 1002
Alamar v. Olink

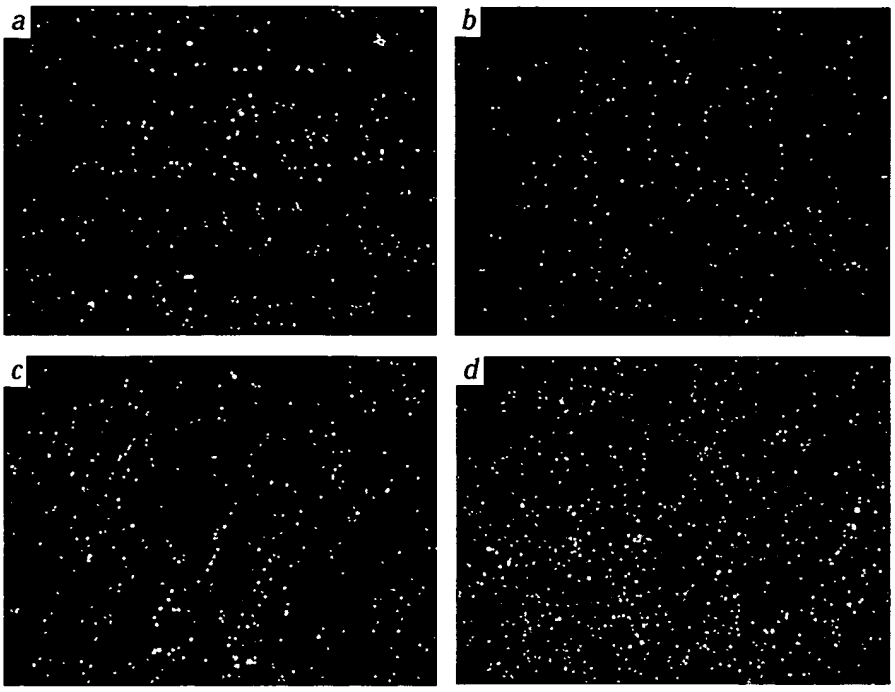


Fig. 7 Detection of individual ligated probe molecules on glass slides by RCA-CACHET. *a–d*, Fluorescent imaging of individual signals generated by surface-immobilized probes. Experiments with different PCR amplicons are shown in each panel, as follows: (*a*) mu:wt = 0:1; (*b*) mu:wt = 1:1; (*c*) mu:wt = 1:25; (*d*) mu:wt = 1:100. Only a small area from each dot on the array is presented. The fluorescent signals from RCA-CACHET products tagged with fluorescein-labelled and Cy3-labelled detector oligos were acquired separately in grey scale using two filter sets (excitation/emission: 455 ± 35 nm/530 ± 15 nm bandpass filters for FITC and 546 ± 5 nm/570 ± 5 nm bandpass interference filters for Cy3). Each image shown is the superimposition of two separate images, with FITC and Cy3 signals pseudocoloured in green and red, respectively. The orange dot in Fig. 6*d* is generated by superimposition of a green signal and a red signal, and probably arises from two RCA extensions occurring in close proximity. *e*, Table of observed counts of individual fluorescent signals pooled from images of several microscope fields.

e

	Ratio mu/wt	Red counts	Green counts	Total counts	Cy3/FITC ratio
a	0:1	9	4613	4622	1:513
b	1:1	2093	2315	4408	1:1.1
c	1:25	107	2758	2865	1:26
d	1:100	46	4799	4845	1:104

© 1998 Nature America Inc. • http://genetics.nature.com

meter of 0.2–0.5 μ, which are generated by single molecules of condensed DNA. The ratio of Cy3-labelled to fluorescein-labelled dots (Fig. 7*e*) corresponds closely with the known ratio of mutant to wild type strands, down to a value of 1/100. A few red signals were generated by pure wild-type DNA (ratio = 1/513), and we interpret these as resulting from mismatch ligation events, which are expected to occur with a frequency of 1/500 to 1/1500 when using wild-type *Thermus Thermophilus* DNA ligase[10].

Detection of padlock probes in cytological samples

As linear RCA was shown to permit analysis of single DNA primers immobilized on glass, we tested this amplification reaction for the detection of padlock probes *in situ*. In spite of the efficiency of surface RCA, padlock probes represent a serious challenge for amplification. Differential chromatin condensation and probe accessibility within a cellular milieu could present obstacles for a rolling-circle reaction. Indeed, attempts to detect the *CFTR* locus in metaphase chromosomes by RCA of ligated padlock probes were unsuccessful. To test the feasibility of padlock probe detection on deproteinized DNA, we ligated the G542X padlock probe (Fig. 2*a*) on a sample of wild-type human genomic DNA which had been immobilized and denatured on a polylysine coated glass slide, at a density of approximately 480 haploid genomes per square mm. An RCA reaction was carried out for 22 min *in situ* to amplify the ligated padlock probes, incorporating BUDR as a hapten into the ssDNA product. The ssDNA was detected by binding of biotinylated anti-BUDR IgG, followed by fluorescent labelling with FITC-avidin. Although most of the ssDNA molecules generated by RCA were condensed into small compact objects, we observed many extended molecules, which were several microns long, lying on the surface of the polylysine-coated glass slide, which contains residual positive charges (Fig. 8*a*). The same padlock probe, with a gap oligonucleotide specific for the wild-type *G542X* locus, was ligated on salt-extracted nuclei prepared by the 'halo' method[11,12]. By carrying out RCA for only 15 min to generate shorter DNA strands, and using the same hapten incorporation and labelling protocol, discernible strands of RCA product originating from padlock probes ligated on the perinuclear DNA halo were observed with very low frequency (Fig. 8*b*). On standard slides, which are not charged, many small, discrete signals were observed inside nuclei or on the perinuclear DNA halo. More than 99.9% of these signals consisted of DNA that was tightly condensed into a small fluorescent dot,

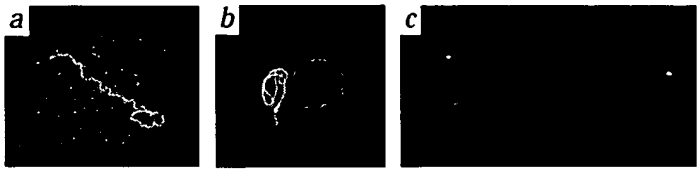


Fig. 8 Detection of padlock probes amplified by RCA on cytological preparations. *a*, Observation of non-condensed rolling circle DNA product from a padlock probe specific for the *CFTR G542X* wt locus, labelled with BUDR as a hapten, lying on the surface of a polylysine-coated slide. *b,c*, Observation of partly condensed (*b*) and fully condensed (*c*) rolling-circle amplification signals generated by *CFTR G542X* wt padlock probes that had been hybridized to nuclear 'halo' cytological preparations (see text for details). Images were photographed with a 63× objective using a Zeiss epifluorescence microscope with a cooled CCD camera. The frequency of nuclei displaying one or two RCA signals in halo preparations of wild-type cells was 197/2182 for the *G542X* locus wt probes, 1936/2573 for *deltaF508* locus wt probes. When we probed the X-linked *OTC* with a wild-type probe, we observed only one or zero signals per male cell nucleus, and no male nuclei with two signals (data not shown).

Alamar Exhibit 1002
Alamar v. Olink

due to the cross-linking effect of avidin:biotin interactions (Fig. 8*c*). Not all nuclei displayed signals, and most nuclei contained a single signal; only a few nuclei had two signals. The frequency of nuclei displaying one or two RCA signals in several halo preparations that contained at least 2000 cells per slide is provided (legend, Fig. 8). Although this initial data is highly encouraging, the variable detection efficiency of probes specific for different loci suggests that our current protocols for *in situ* RCA do not provide optimal reagent access to all regions in chromatin.

Discussion

We have demonstrated three alternative strategies for allele discrimination utilizing ligation of circularizable DNA probes and rolling-circle replication. All three strategies employ RCA or HRCA for signal generation, and the assays on surfaces additionally employ the novel CACHET principle for signal condensation and identification. The polymerase mediated 'gap-fill' reaction is preferred over the 'gap-probe' ligation reaction for solution studies of complex genomes. This is because the gap oligonucleotides, which cannot be washed away in a solution assay, must be used at relatively high concentrations for the ligation step and often interfere with the HRCA reaction, inducing the formation of amplicon artifacts. The gap-ligation reaction is considered preferable for allele analysis of DNA in cytological specimens. Here, the specificity of ligation should be enhanced relative to probes without a gap because three different sequence recognition events and two independent ligation events must occur before 'padlock' closure. The third assay method, the ligase-mediated extension of an oligonucleotide linked to a solid surface, provides a novel approach to quantify hybridization/ligation events for single molecules and to score rare somatic mutations.

There are other potential uses for the gap-fill reaction, in which target-complementary sequences are incorporated into circles by copying and covalent closure. In principle, any DNA sequence thus captured into a circular DNA may be amplified by RCA or HRCA. We envision applications of this reaction in situations where it is desirable to interrogate the sequence incorporated into a padlock probe at some point after RCA. Additional experiments will be required to determine if longer sequences, such as microsatellite repeats, can be copied into circularizable probes for amplification and analysis. As there is little likelihood that rolling-circle amplification will modify the number of repeats incorporated into a circularized probe[3], direct measurement of repeat size would be feasible.

For reasons that we do not yet understand, ø29 DNA polymerase is a poor catalyst for HRCA. HRCA is catalyzed by exo(–) Vent DNA polymerase, by the large fragment of Bst DNA polymerase[13] and by the Sequenase 2.0 variant of T7 DNA polymerase. It had been shown by others[3,4] that Sequenase supports rolling-circle amplification of circular oligonucleotides, albeit relatively slowly. The efficiency of RCA and HRCA reactions is increased by the addition of proteins that bind ssDNA. The addition of *E. coli* single-strand binding protein (SSB) stimulates Sequenase-catalyzed HRCA, whereas phage T4 gene-32 protein stimulates Vent exo(–) catalyzed HRCA. In contrast, Bst large fragment DNA polymerase catalyzes HRCA efficiently in the absence of protein cofactors. Although Sequenase is a suitable catalyst for linear RCA, its rapid polymerization rate in the presence of SSB is not necessarily optimal for HRCA. Many potential priming sites may not be used, as they are rapidly occupied by competing extension reactions. Paradoxically, the theoretical HRCA reaction yield should be somewhat higher at relatively slower polymerization rates (in the range of 16.5 nt per second, as reported for Vent DNA polymerase[14]), because this rate permits a more efficient utilization of priming sites. Our current data demonstrate that the HRCA reaction is most efficient when catalyzed by thermostable strand-displacing enzymes. Additional work will be required to reveal which DNA polymerase is the ideal catalyst for HRCA, although the reaction yield already compares very favourably with PCR.

It is worth noting other unusual features of HRCA. When allele-specific primers are used, there are some mis-priming events, but the strands generated by such events will incorporate the incorrect base only in the primer itself, and not in the several hundred tandem copies that result from each primer extension. Thus, in contrast with PCR, non-specific priming is a non-propagating event in HRCA. For applications that do not require allele-specific primers, two invariant primers may be designed, so that circles with different DNA inserts are replicated with a universal primer set, facilitating reproducible amplification kinetics.

The amplification of small circularized oligonucleotides by linear RCA is rapid, technically simple and the amplified DNA will not diffuse away from the site of synthesis. To date, we have used over 20 different circularized oligonucleotides in the range of 60–110 bases, and have not observed any variability in amplification due to sequence content. RCA products may be detected by incorporating haptens or fluors directly, and it is possible to use circles with biased base compositions to obtain differential labelling. A larger range of labelling combinations is attainable by RCA-CACHET. The two-circle/two-primer signal generating system (Fig. 6) is extensible to any number of probes, and retains the use of the same pair of circular amplicons. Two different allele-discriminating primers are required for each mutational locus being assayed. Although photolithographic DNA microarray technology enables massively parallel assays for mutation detection, altered bases are only detectable if they constitute a sizable fraction of the DNA population. Thus, such arrays are well suited for the detection of germline mutations, but not rare somatic mutations. Ligation-mediated RCA extends the utility of assays based on surface-bound oligonucleotides by permitting the detection of infrequent mutations in the presence of a large excess of wild-type DNA. The detection limit for such mutations will be determined by the stringency of the DNA ligation reaction, which could be improved by using new ligase variants[10]. The single-molecule counting approach promises to be both sensitive and linear in its response to target concentration. Individual immunoglobulins tagged with rolling-circle primers can be detected by RCA-CACHET. This approach is being used by our laboratory to detect single molecules of antibody bound to antigen (Z.Z., unpublished observations). A logical extension of the RCA-CACHET signal detection paradigm is the use of fluorophore combinations designed for multiparametric colour coding[15], thereby markedly increasing the number of targets that could be analysed simultaneously.

RCA shows promise as a reporter system for padlock probes bound to naked DNA fibers on solid surfaces, and should be extensible to *in situ* probing with multiple probes using RCA-CACHET. Detection of padlock probes by RCA works with variable efficiency in partially deproteinized cytological preparations such as salt-extracted halo preps, but not as well in intact nuclei. These observations suggest that the extent of chromatin condensation and the fixation methods used for cytological preparations influence the efficiency of the RCA reaction. Thus, the optimization of the RCA method for allele discrimination *in situ* of single-copy genes will require further work. Nonetheless, the absence of nonspecific signals in the solution HRCA assays, together with the observation that the *in situ* RCA signals are generally limited to one or two fluorescent dots per nucleus, supports the idea that padlock probes amplified by RCA will provide sufficient sensitivity and specificity for the cytological detection of single-copy genes in the context of the total human genome.

© 1998 Nature America Inc. • http://genetics.nature.com

Alamar Exhibit 1002
Alamar v. Olink

© 1998 Nature America Inc. • http://genetics.nature.com

Methods

DNA oligonucleotides. Oligonucleotides containing a phosphate group at the 5′ end were purchased from the Yale University Critical Technologies facility. The sequences are: *CFTR G542X* phosphorylated open circle probe (85 bases) for the gap oligo method, 5′–AAGAACTATATTGTCTTTCATTCTTGCATG-GTCACACGTCGTTCTAGTACGCTTCTAGTACGCTTTTCCACTCAGTG-TGATTCCA–3′; *G542X* wild type phosphorylated gap probe, 5′–CCTTCTCC–3′; *G542X* mutant phosphorylated gap probe, 5′–CCTTCTCA–3′; primer for rolling circle reaction, 542X.P1.18, 5′–ACGACGTGTGACCATGCA–3′. For the polymerase gap-fill method, the phosphorylated open circle probe (89 bases) was 5′–GAACTATATTGTCTT-TCTCTGATTCTGACTCGTCATGTCTCAGCTTTAGTTTAATACGACTCA CTATAGGGCTCAGTGTGATTCCACCT–3′. Primers for the HRCA reactions performed with gap-fill probe were: first primer, 542X.P1-23, 5′–CTAAAGCTGAGACATGACGAGTC–3′; alternative second primers for allele discrimination, 542X.P2-23-C, 5′–CTCAGTGTGATTCCACCTTC–3′ or 542X.P2-23-A, 5′–CTCAGTGTGATTCCACCTTCACA–3′; *G542X* artificial wild type target, 5′–TTGCAGAGAAAGACAATATAGTTCTTGGAGAAG-GTGGAATCACACTGAGTGGA–3′; *G542X* artificial mutant target, 5′–TTGCAGAGAAAGACAATATAGTTCTTTGAGAAGGTGGAATCACACTG-AGTGGA–3′. Phosphorylated open circle probe specific for base number 114 of intron 9 of *OCT*: 5′–GAGGAGAATAAAAGTTTCTCATAAGACTCG-TCATGTCTCAGCAGCTTCTAACGGTCACTAATACGACTCACTATAGG-TTCTGCCTCTGGGAACAC–3′. *OTC* wild-type gap probe was 5′–TAGT-GATC–3′. Primer for RCA, OTC.P.18, 5′–GCTGAGACATGACGAGTC–3′. Oligonucleotides for the RCA-CACHET method, carbon spacers in parenthesis: phosphorylated P1-G542X, 5′–GAGAAGGTGGAATCACACTGAGTG-GAGGTCAACGAGCAATTTTTTTTTTT-(C7-NH2)–3′; P2wt, 3′–GTT-CTTGATATAACAGAAAGTTTT-(C18)-TTTTTATGATCACAGCTGAGGA-TAGGACATGCGA–3′; P2mu, 3′–TTTCTTGATATAACAGAAAGTTTT-(C18)-TTTTTACGTCGTCCGTGCTAGAAGGAAACACGCA–3′; pre-made amplification circles with cognate detector tags, Cwt, 5′–CGCATGTCCTAT-CCTCAGCTGTGATCATCAGAACTCACCTGTTAGACGCCACCAGCTCC-AACTGTGAAGATCGCTTAT–3′; detector tag Fl-det1c-dinitrophenol (DNP), FITC-TCAGAACTCACCTGTTAG-3′-DNP; detector tag Fl-det1d-DNP, FITC-ACTGTGAAGATCGCTTAT–3′-DNP; Cmu, 5′–GCGTGTTTC-CTTCTAGCACGGACGACGTATATGATGGTACCGCAGCCAGCATCACC-AGACTGAGTATCTCCTATCACT–3′; detector tag Cy3-det2b-DNP, Cy3-TATATGATGGTACCGCAG–3′-DNP; detector tag Cy3-det2c-DNP, Cy3-TGAGTATCTCCTATCACT–3′-DNP.

Enzymes, antibodies and cell lines. Phage ø29 DNA polymerase was a gift from M. Salas and B. Moffett. *T. thermophilus* DNA ligase was a gift from F. Barany. Some ligation reactions were performed with Ampligase DNA (Epicentre Technologies). Other proteins and enzymes used were *T. flavus* DNA polymerase (Molecular Biology Resources), Sequenase 2.0 (Amersham), phage T4 gene 32 protein (Amersham), Vent (exo-) DNA polymerase (New England Biolabs), T7 RNA polymerase (Ambion) and biotinylated anti-bromouridine deoxyriboside (BUDR) IgG (Zymed). Anti-DNP IgM was from American Research Products. Cell lines of known genotypes for the *G542X* locus (GM07828[+/+], GM11497B[+/–], GM11496[–/–]) were purchased from Corriell Cell Repositories.

Probe ligations using gap oligonucleotides or gap-filling. Reactions employing gap oligonucleotides for the detection of artificial targets contained 150 nM open circle probe, 1300 nM gap probe, 420 nM artificial target in 20 mM Tris•HCl (pH 8.3), 50 mM KCl, 0.5 mM NAD, 10 mM $MgCl_2$ and 0.01% Triton X-100. Reactions with genomic DNA targets were performed in the same buffer, using 50 nM open circle probe and 140 nM gap oligonucleotide. *T. thermophilus* DNA ligase was used at a concentration of 0.4 Units/µl, in a 40 µl reaction. Ligation was performed for 1 h at 52 °C, which is higher than the calculated T_m for the gap oligonucleotide. The circularizable *G542X* probe and an equimolar mixture of the two allele-specific gap oligonucleotides were ligated in the presence of the mutant sequence artificial DNA target. Circular DNA resulting from the ligation of *G542X* gapped probes were amplified with ø29 DNA polymerase, and the amplified ssDNA was cloned and sequenced. The resulting sequence contained only T at the position of the *G542X* mutation in 20 independent clones, indicating that only the mutant gap oligonucleotide had been ligated and amplified. With the wild-type template, the RCA product contained only G at the *G542X* site in 20 clones (data not shown). Thus, the amplified material contained faithful replicas of a small segment of the target DNA. For combined gap-filling and ligation reactions, the open-circle probe concentration was 135 nM when artificial targets were used, and 70 nM for genomic DNA targets. The reaction was performed in 50 mM (N-[2-Hydroxyethyl]pyperazine)-N′-3-propanesulfonic acid (EPPS)-KOH pH 7.8, 25 mM $MgCl_2$, 0.25 mM NAD, 400 µM dATP, dCTP, dTTP, 0.075 U/µl *T. flavus* DNA polymerase, 0.37 U/µl Ampligase DNA ligase. Artificial targets were at a final concentration of 420 nM. For genomic DNA targets, the DNA concentration was 850 ng in 40 µl. The gap-fill reaction was incubated at 60 °C for 45 min, and 4 µl were used for analysis by HRCA. The circular oligonucleotides used two-colour RCA-CACHET were ligated using an artificial guide oligonucleotide and standard ligation conditions.

Rolling-circle reactions. Reactions contained 10 nM circularized probes, 130 nM rolling-circle primer, in 50 mM Tris•HCl (pH 7.5), 10 mM $MgCl_2$, 200 µg/ml acetylated BSA (Amersham) 400 µM dATP, dGTP, dCTP and dTTP. Phage T4 gene-32 protein was present at a concentration of 40 ng/µl (approximately 1142 nM), and phage ø29 DNA polymerase was added last at a concentration of 6.3 ng/µl (approximately 92 nM). The reactions were incubated at 31 °C for periods ranging from 2–30 min.

RCA-CACHET mutation detection assays on glass slides. The synthetic oligonucleotides used are listed above. The oligonucleotide P1-G542X was immobilized over an area of 1 mm in diameter on the surface of a glass slide activated with reactive groups. The slides were coated with 4-aminobutyldimethylmethoxysilane and derivatized with 1,4-phenylene-diisothiocyanate, using a published protocol[16] with minor modifications. Covalent coupling was obtained by reaction of a primary amine attached to the 3′ end of the P1-G542X oligonucleotide. Genomic DNA preparations obtained from human cell lines that were either wild type or homozygous mutant at the *CFTR-G542X* locus were mixed in different pre-determined ratios (mutant/wild type equal to 1:0, 1:1, 1:25 and 1:100) and then amplified by PCR as described[17]. PCR amplicons were used as targets for allele discrimination assays on glass slides that contained immobilized P1 molecules. The DNA was denatured at 98 °C for 5 min immediately before use. Hybridization/ligation was carried out in 7.5 µl of 20 mM Tris•Cl, 50 mM KCl, 10 mM $MgCl_2$, 10 mM DTT, 1 mM NAD, 200 ug/ml BSA and 0.5 unit/µl of Ampligase (Epicentre). Target concentration was in the 0.1–0.2 nM range. Probe-primers P2-wt and P2-mu were 1 µM each. All enzymatic reactions took place by confining the reaction volume over the DNA dot with a small silicone rubber O-ring sealed with rubber cement. The slide was covered with a second slide, forming a sandwich filled with wet filter paper perforated at the positions of the O-rings. This sandwich was placed on an aluminum block in a pre-heated moisture chamber. Ligation took place at 62 °C for 2 h. Washes were performed at 70 °C, as follows: twice for 3 min in 75% formamide + 2× SSPE (1×SSPE: 150 mM NaCl, 10 mM NaH_2PO_4, 1 mM EDTA, pH 7.4), 0.1% SDS, 1 min in 0.05% Triton X-100, 2 min with 6× SSPE, 0.05% Triton X-100, 1 min in 0.01% Triton X-100 and finally a brief rinse in water. Hybridization of signal-generating circular oligonucleotides (Cwt and Cmu, see oligonucleotide section) was performed using a concentration of 170 nM for each circle in 6 µl of Buffer I (67 mM Tris•Cl, pH 8.0, 16.7 mM $MgCl_2$, 84 mM NaCl, 8 mM DTT, 0.0167% Triton X-100). After 30 min at 45 °C, the slide was placed on ice, 2 µl of Buffer II (containing 20 µM *E. coli* SSB, 25 mM DTT, and 1.5 mM each of dATP, dCTP, dGTP and dTTP) was added, and incubation proceeded for 5 min at 37 °C. Two µl of 5 µM Sequenase Version 2.0 DNA polymerase was added, followed by incubation at 37 °C for 20 min. Stringent washes at 70 °C were as follows: twice for 3 min in 75% formamide, 2×SSC, 0.1% SDS (1×SSC: 0.15 M NaCl, 0.015 M sodium citrate, pH 7.4), 1 min in 0.2×SSC + 0.05% Triton X-100, 2 min in 4×SSC, 0.05% Triton X-100. Each dot was covered with 7.5 µl of detection tag oligonucleotides (0.5 µM each of Fl-det1c-DNP, Fl-det1d-DNP, Cy3-det2b-DNP,Cy3-det2c-DNP in 2×SSC + 0.05% Triton X-100 + 0.5 mg/ml degraded herring sperm DNA) and incubated at 37 °C for 20 min. The slide was washed 4 times for 5 min with 2×SSC + 0.1% Tween 20, and 4 times for 5 min with 4×SSC + 0.05% Triton X-100, rinsed once with 2×SSC and drained. 7.5 µl of condensation solution (33 µM mouse anti-DNP IgM in 2× SSC, 0.1% Tween 20, 0.5% BSA, 1 mg/ml degraded herring sperm DNA) was added, and the slide was incubated at 37 °C for 15 min and washed twice for 5 min in 2×SSC, 0.1% Tween 20 at RT, drained and covered with Pro-long antifade (Molecular Probes Inc.) under a cover slip. Fluorescent imaging was performed using a Zeiss epifluorescence microscope equipped with a Photo-

Alamar Exhibit 1002
Alamar v. Olink

metrics cooled CCD camera. For an oligonucleotide spot of 1 mm in diameter, the maximum number of non-overlapping green or red signals is about 80,000 using our CACHET conditions. Optimization of reaction parameters may permit even higher signal densities without cross-talk.

HRCA reactions. A 4-µl aliquot of a ligation reaction was incubated in a volume of 35 µl containing 20 mM Tris•HCl (pH 8.8), 10 mM KCl, 2.7 mM $MgSO_4$, 5% v/v DMSO, 0.1% Triton X-100, 400 µM dATP, dGTP, dCTP, dTTP and 900 nM each of two specific primers (542X.P1-23, combined with either 542X.P2-23-C, or 542X.P2-23-A). Phage T4 gene-32 protein (Amersham) was present at a concentration of 38 ng/µl, (approximately 1085 nM). After combining all these materials at RT, the reactions were placed in ice, Vent (exo-) DNA polymerase (New England Biolabs) was added at a final concentration of 0.32 units/µl, and the reactions were incubated at 92 °C for 3 min, then at 65.5 °C for 90 min. The conditions described here represent experimentally determined optima for primer concentration (operational range 600–1200 nM), magnesium ion concentration (operational range 2.2–3.0 mM when using 400 µM dNTPs) and temperature (62–66.5 °C for Vent DNA polymerase, 60–65 °C for Bst DNA polymerase). Primers should have a T_m slightly higher than the HRCA reaction temperature. Four different batches of T4 Gene 32 protein have been used successfully; one of the four batches was less stimulatory than the other three, but still satisfactory for exponential amplification (Fig. 5*c*).

In situ detection of padlock probes by RCA. Human DNA was bound on polylysine slides as described[18]. Human peripheral lymphocytes were processed as halo preparations as described[11,12]. Padlock probe ligation *in situ* was performed for 2 h at 52 °C by incubating 18 µl of reaction mixture under a cover slip sealed with rubber cement as follows: 150 nM phosphorylated *G542X* 89-mer probe, 1200 nM wild type *G542X* gap probe, in 20 mM Tris•HCl (pH 8.3), 50 mM KCl, 0.5 mM NAD, 10 mM $MgCl_2$, 150 µg/ml acetylated BSA, 0.01% Triton X-100 and 0.7 U/µl Ampligase DNA ligase. The slide was washed twice for 5 min at 50 °C in 2×SSC, 25% formamide, then washed for 2 min in 50 mM Tris-Cl, pH 7.5, 40 mM KOAC, 10 mM $MgCl_2$, 10 mM DTT, 100 µg/ml BSA to remove formamide. An RCA reaction was carried out in a volume of 12 µl under a coverslip using a reaction buffer consisting of 50 mM Tris•HCl (pH 7.5), 10 mM $MgCl_2$, 200 µg/ml acetylated BSA 400 µM dATP, dGTP, dCTP and dTTP and a rolling-circle primer (600 nM oligonucleotide 542X.P1.18). Phage T4 gene-32 protein was present at a concentration of 40 ng/µl (approximately 1142 nM), and phage ø29 DNA polymerase was added at a concentration of 6.3 ng/µl (approximately 92 nM). The RCA reactions were incubated at 37 °C for either 22 min (for polylysine slides) or 15 min (for halo slides). Each slide was then washed twice for 5 min in 2×SSC with 20% formamide at 25 °C, and twice for 4 min in SSC-BSAT (2×SSC, 2.8% BSA, 0.12% Tween-20) at 37 °C. Slides were incubated for 20 min at 37 °C in SSC-BSAT, containing 5 µg/ml biotinylated Anti-BUDR IgG. The slides were washed three times for 5 min in SSC-BSAT to remove excess antibody, and incubated for 30 min at 37 °C in SSC-BSAT containing 5 µg/ml FITC-Avidin. After three washes in SSC-BSAT, the slides were stained for 2 min in 2×SSC, 0.1 µg/ml 4'6-diamidino-2-phenlindole, washed for 10 min in 1.5×SSC, 0.01% TWEEN-20 at RT and prepared for imaging by covering with 20 µl of antifade under a cover slip.

GenBank accession number. *OCT*, HUMOTC09.

Acknowledgements

We are indebted to M. Salas for a generous gift of ø29 DNA polymerase, and for advice regarding the use of the enzyme. We also thank B. Moffett and Amersham for research samples of this enzyme. F. Barany kindly provided samples of T. thermophilus DNA ligase. C. Radding and his group kindly provided us with E. coli SSB. We thank E. Winn-Deen for critical reading of the manuscript. This work was supported in part by a research grant from ONCOR, Inc., to P.M.L. and a grant from the National Institutes of Health to D.C.W. (HG00272).

1. Nilsson, M. *et al.* Padlock probes: Circularizing oligonucleotides for localized DNA detection. *Science* **265**, 2085–2088 (1994).
2. Nilsson, M. *et al.* Padlock probes reveal single-nucleotide differences, parent of origin and *in situ* distribution of centromeric sequences in human chromosomes 13 and 21. *Nature Genet.* **16**, 252–255 (1997).
3. Fire, A. & Xu, S.Q. Rolling replication of short DNA circles. *Proc. Natl Acad. Sci. USA* **92**, 4641–4645 (1995).
4. Liu, D., Daubendiek, S. L., Zillman, M. A., Ryan, K. & Kool, E. T. Rolling circle DNA synthesis: Small circular oligonucleotides as efficient templates for DNA polymerases. *J. Am. Chem. Soc.* **118**, 1587–1594 (1996).
5. Blanco, L. & Salas, M. Characterization and purification of a phage ø29-encoded DNA polymerase required for the initiation of replication. *Proc. Natl Acad. Sci. USA* **81**, 5325–5329 (1984).
6. Blanco, L. *et al.* Highly efficient DNA synthesis by the phage ø29 DNA polymerase. *J. Biol. Chem.* **264**, 8935–8940 (1989).
7. Abravaya, K., Carrino, J.J., Muldoon, S. & Lee, H. Detection of point mutations with a modified ligase chain reaction. *Nucleic Acids Res.* **23**, 675–682 (1995).
8. Schwarz, K. Improved yields of long PCR products using gene 32 protein. *Nucleic Acids Res.* **18**, 1079 (1990).
9. Newton, C.R. *et al.* Analysis of any point mutation in DNA. The amplification refractory mutation system (ARMS). *Nucleic Acids Res.* **17**, 2503–2516 (1989).
10. Luo, J., Bergstrom, D.E. & Barany, F. Improving the fidelity of *Thermus thermophilus* DNA ligase. *Nucleic Acids Res.* **24**, 3071–3078 (1996).
11. Vogelstein, B., Pardoll, D.M. & Coffey, D.S. Supercoiled loops and eucaryotic DNA replication. *Cell* **22**, 79–85 (1980).
12. Gerdes, M.G, Carter, K.C., Moen, P.T. Jr & Lawrence, J.B. Dynamic changes in the higher-level chromatin organization of specific sequences revealed by *in situ* hybridization in nuclear halos. *J. Cell Biol.* **126**, 289–304 (1994).
13. Aliotta, J.M. *et al.* Thermostable Bst DNA polymerase lacks a 3'-5' proofreading exonuclease activity. *Genet. Anal.* **12**, 185–195 (1996).
14. Kong, H., Kucera, R. & Jack, W. Characterization of a DNA polymerase from the hyperthermophile Archea *Thermococcus litoralis*. *J. Biol. Chem.* **268**, 1965–1975 (1993).
15. Speicher, M.R., Ballard, S.G. & Ward, D.C. Karyotyping human chromosomes by combinatorial multicolor FISH. *Nature Genet.* **12**, 368–375 (1996).
16. Guo, Z., Guilfoyle, R.A., Thiel, A.J., Wang, R. & Smith, L.M. Direct fluorescence analysis of genetic polymorphisms by hybridization with oligonucleotide arrays on glass supports. *Nucleic Acids Res.* **22**, 5456–5465 (1994).
17. Heinoven, P., Itia, A., Torresani, T. & Lovgren, T. Simple triple-label detection of seven cystic fibrosis mutations by time-resolved fluorometry. *Clin. Chem.* **43**, 1142–1150 (1997).
18. Schena, M., Shalon, D., Davis, R.W. & Brown, P. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* **270**, 467–470 (1995).
19. Rychlik, W. & Rhoads, R.E. A computer program for choosing optimal oligonucleotides for filter hybridization, sequencing, and *in vitro* amplification of DNA. *Nucleic Acids Res.* **17**, 8543–8551 (1989).

© 1998 Nature America Inc. • http://genetics.nature.com

Alamar Exhibit 1002
Alamar v. Olink

This Page is Inserted by IFW Indexing and Scanning Operations and is not part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☐ BLACK BORDERS
- ☐ IMAGE CUT OFF AT TOP, BOTTOM OR SIDES
- ☑ FADED TEXT OR DRAWING
- ☐ BLURRED OR ILLEGIBLE TEXT OR DRAWING
- ☐ SKEWED/SLANTED IMAGES
- ☐ COLOR OR BLACK AND WHITE PHOTOGRAPHS
- ☐ GRAY SCALE DOCUMENTS
- ☐ LINES OR MARKS ON ORIGINAL DOCUMENT
- ☐ REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY
- ☐ OTHER: _____

IMAGES ARE BEST AVAILABLE COPY.
As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.

Alamar Exhibit 1002
Alamar v. Olink

Best Available Copy

chap. 6). Thus, deformation mechanisms for the postglacial rebound in the upper mantle near the deglaciated region are similar to those involved in mantle convection. A large difference between the two processes is the strain magnitude. Strain involved in the postglacial rebound is small ($\sim 10^{-6}$), and transient rather than steady-state creep might occur (J. Weertman, *Philos. Trans. R. Astron. Soc. London Ser. A* 288, 9 (1978)). However, recent analysis of transient creep in olivine suggests that the transient period in olivine is much shorter than that of most metals, and so the transient creep proposed by Weertman may not be important [S. Karato, in (3), pp. 176–208].

53. The ice sheet was interpreted to have a parabolic profile. The ice thickness at the load center and the time of deglaciation were adjusted to give the best fit to the RSL data at the center of the load. A thickness of 3500 m was chosen in models A, B, C, D, and E and a thickness of 1600 m was chosen in model F. We chose a deglaciation time of 12,000 years before present except in model E, where 8,000 years before present was chosen. For each model, the first-order approximations to the sea level equations were solved, with the effects of eustatic sea level change, unloading of the ice, and loading of the ocean floor taken into account. As demonstrated by Wu and Peltier in (*51*), such an approximation is valid for RSL in the last 7000 years.
54. This result is consistent with that of W. R. Peltier [*J. Geophys. Res.* 89, 11303 (1984)], who used the RSL data in the Laurentide region, but not with that of M. Nakada and K. Lambeck (*Geophys. J. Int.* 96, 497 (1989)], who used the RSL data in and near Australia where a much thinner lithosphere (50 to 100 km) was inferred. This discrepancy may be due in part to the difference in the geothermal gradient between the two regions, because the recent high-resolution seismic tomography indicates a colder upper mantle in the Laurentide than in Australia [D. L. Anderson, T. Tanimoto, Y.-S. Zhang, *Science* 255, 1645 (1992)].
55. This observation is consistent with some of the findings of Peltier, Wu, and Yuen in (*21*) but not with those of Nakada and Lambeck in (*23*). Part of this discrepancy may result from the regional difference in geothermal regimes (*54*).
56. J. Ravenaugh and T. H. Jordan, *J. Geophys. Res.* 96, 19781 (1991).
57. D. L. Kohlstedt and M. S. Weathers, *ibid.* 85, 6269 (1980).
58. Y.-S. Zhang and T. Tanimoto, *Nature* 355, 45 (1992).
59. T. Kawasaki, *Lithos* 20, 263 (1987).
60. P. van Keken, D. A. Yuen, A. van den Berg, *Earth Planet. Sci. Lett.* 112, 179 (1992).
61. An alternative way to get nearly isoviscous upper mantle with a large activation volume is to invoke a large decrease in activation volume with pressure. Borch and Green in (*39*) proposed that in the relation between strain rate and temperature and pressure

$$\dot{\epsilon} = A\sigma^n \exp[-gT_m(P)/T]$$

in which g is a constant and $T_m(P)$ is the melting temperature, this second variable should be identified as the solidus rather than the melting temperature of the constituting mineral or minerals. The solidus of peridotite changes with pressure with a large curvature [E. Takahashi, *J. Geophys. Res.* 91, 9367 (1986)], and so the presence of nearly isoviscous upper mantle would be interpreted from this assumption. However, this assumption contradicts the experimental observations. If the solidus rather than the melting temperature determines rheology, then creep rate in olivine should increase by a factor of 10^4 to 10^5 when a small amount of pyroxenite is added, because the solidus will be reduced by 500 to 600 K. The observed increase in strain rate by the addition of orthopyroxene is only a factor of about 3 [Q. Bai, S. J. Mackwell, D. L. Kohlstedt, *J. Geophys. Res.* 96, 2441 (1991)]. Therefore, the change in activation volume with pressure under the upper mantle conditions appears to be small, and so an activation volume as high as 27 $cm^3\ mol^{-1}$ leads to unacceptably high viscosities in the deep upper mantle (Fig. 2).
62. Sidereal age is based on the sidereal day—the time required for the Earth to rotate once about its axis; E. Bard, B. Hamelin, R. G. Fairbanks, A. Zindler, *Nature* 345, 405 (1990).
63. We thank D. A. Yuen, W. R. Peltier, and D. L. Kohlstedt for reading the manuscript and for stimulating discussions. Supported by the National Science Foundation (EAR-0017811) (S.K.) and by the Natural Sciences and Engineering Research Council (Canada) (P.W.).

RESEARCH ARTICLE

Structure-Specific Endonucleolytic Cleavage of Nucleic Acids by Eubacterial DNA Polymerases

Victor Lyamichev,* Mary Ann D. Brow, James E. Dahlberg†

Previously known 5′ exonucleases of several eubacterial DNA polymerases have now been shown to be structure-specific endonucleases that cleave single-stranded DNA or RNA at the bifurcated end of a base-paired duplex. Cleavage was not coupled to synthesis, although primers accelerated the rate of cleavage considerably. The enzyme appeared to gain access to the cleavage site by moving from the free end of a 5′ extension to the bifurcation of the duplex, where cleavage took place. Single-stranded 5′ arms up to 200 nucleotides long were cleaved from such a duplex. Essentially any linear single-stranded nucleic acid can be targeted for specific cleavage by the 5′ nuclease of DNA polymerase through hybridization with an oligonucleotide that converts the desired cleavage site into a substrate.

Important functions of several DNA polymerases (DNAPs) include the removal of nucleotides from the 5′ and 3′ ends of DNA chains (*1*). For example, the 5′ exonuclease activities located in the NH_2-terminal domains of DNAP from *Escherichia coli* (DNAP-Ec1) and *Thermus aquaticus* (DNAP-Taq) can participate in (i) removal of the RNA primers of lagging strand synthesis during replication and (ii) the removal of damaged nucleotides during repair (*1*, *2*). Although mononucleotides predominate among the digestion products, short oligonucleotides (≤12 nucleotides) can also be observed, implying that these so-called 5′ exonucleases can function endonucleolytically (*3*, *4*). Thus, we like to call these activities 5′ nucleases.

For removal of primers or damaged nucleotides, a 5′ nuclease must be able to cleave RNA and DNA strands regardless of their sequences. However, indiscriminate cleavage of the nucleic acids would be lethal. We propose that this problem is resolved by having the 5′ nuclease recognize its substrate by structure rather than sequence. An appropriate structure would be the junction where the two strands of a duplex separate into single-stranded arms; the arm with a free 5′ end would be recognized as the displaced strand that is to be cleaved from the remaining duplex (Fig. 1A). To avoid ambiguity, we refer to the top and bottom strands of the complex as the substrate and template strands, respectively. A primer that would be extended during nick translation is paired to the 3′ arm of the template strand. To test our model of recognition and cleavage by the 5′ nuclease, we used the substrate and template strands shown in Fig. 1A, covalently joined at the end of the duplex, for experimental convenience. We also used this model to develop a method for cleaving single-stranded nucleic acids efficiently at desired sites with extreme sequence specificity.

The 5′ nuclease of DNAP-Taq. Incubation of the structure shown in Fig. 1A with DNAP-Taq (*5*) and deoxynucleoside triphosphates (dNTPs) resulted in the re-

The authors are in the Department of Biomolecular Chemistry, University of Wisconsin School of Medicine, 1300 University Avenue, Madison, WI 53706.
*On leave from the Institute of Molecular Genetics, Moscow, Russia.
†To whom correspondence should be addressed.

Supplied by The British Library - "The world's knowledge"

Alamar Exhibit 1002
Alamar v. Olink

lease of a few fragments whose mobilities indicated that they contained most or all of the unpaired 5′ arm (Fig. 1B). Direct sequence analysis confirmed that cleavage occurred near the end of the duplex, releasing the 5′ extension. The released fragments terminated with 3′ hydroxyl groups, as evidenced by their ability to be extended by terminal deoxynucleotidyl transferase (6). The fragments were not released by incubation with the Stoffel fragment (Stf), a derivative of DNAP-Taq lacking the 5′ nuclease domain (7).

Neither primer nor dNTPs were required for cleavage (Fig. 2A, lanes 1 through 6), consistent with the fact that the polymerization and 5′ nuclease functions of DNAP-Taq, like DNAP-Ec1, are in separate domains (2, 3). Nuclease activity required magnesium or manganese ions, and neither zinc nor calcium ions supported the cleavage reaction (lanes 7 to 10). The reaction occurred over a broad temperature range, from 25° to 85°C, with the rate of cleavage increasing at higher temperatures. The pH optimum was about pH 9 (measured at 23°C) although the reaction still occurred at pH 6.5 (6).

Roles of the primer. Although the primer was not required for cleavage, it did influence both the site and the rate of cleavage of the hairpin. The change in the site of cleavage (Fig. 2A) apparently resulted from a change in the substrate structure, and from alteration of the substrate-enzyme interaction. In the absence of primer, sequences such as those indicated in Fig. 1A by underlining could pair, forming extended duplexes (8); cleavage at the ends of these duplexes would release shorter fragments (Fig. 2A, lanes 5 and 6). Addition of excess primer (Fig. 2A, lanes 3 and 4) or incubation at elevated temperatures (Fig. 2B) disrupted these short extensions of the duplex and resulted in a longer 5′ arm and, hence, longer cleavage products.

The 3′ end of the primer also helped to fix the precise site of cleavage. In the absence of primer, cleavage occurred at the ends of the substrate duplexes (either the extended or shortened forms) between the first and second base pairs (23- and 25-nucleotide bands, Fig. 2C, lane 1). Likewise, cleavage also occurred at the end of the duplex when the primer extended up to the base of the duplex (Fig. 2C, lane 2). However, when a gap of four or six nucleotides existed between the 3′ end of the primer and the end of the substrate duplex, the cleavage site was shifted four to six nucleotides in the 5′ direction (Fig. 2C, lanes 3 and 4) (9). The primer-directed shifting of the site of cleavage suggests that precise orientation of the 5′ nuclease on the substrate is dominated by the interaction of the polymerization domain of DNAP-Taq with the primer.

The effect of added primer on the rate of cleavage is evident in the extent of cleavage observed in lanes 1 and 2 of Fig. 2C (10). The influence of primer on cleavage rates

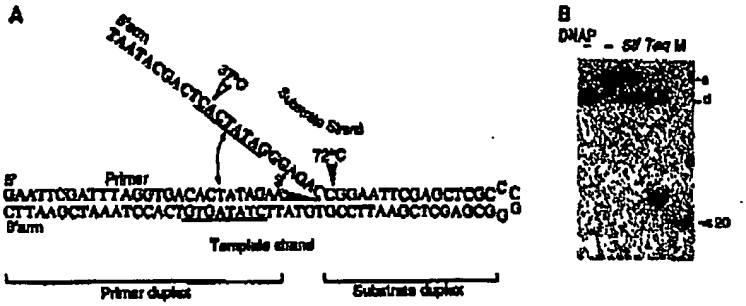


Fig. 1. Cleavage of a DNA hairpin by DNAP-Taq. (A) The proposed hairpin structure of a substrate for the 5′ nuclease of DNAP-Taq. One of three primers used with this substrate is shown, leaving a gap of four nucleotides between the 3′ arm and the bifurcation; the other primers leave no gaps or gaps of six nucleotides. The major cleavage sites at 37° and 72°C without primer are indicated by arrowheads; the filled one is also the cleavage site with a full-length primer. Complementary DNA sequences in 3′ and 5′ arms are underlined. This molecule was generated in a polymerase chain reaction (PCR) catalyzed by DNAP-Stf (22), a variant of DNAP-Taq lacking the 5′ nuclease; the molecule was labeled on the 5′ arm by the use of a ^{32}P-labeled primer. The nomenclature indicated is used throughout the text to describe various regions of the structure. (B) Endonucleolytic cleavage products of the 5′ nuclease of DNAP-Taq. The 5′ end-labeled hairpin was formed by snap-back of complementary sequences after denaturation of the PCR-generated duplex. The hairpin and associated primer shown in (A) were incubated with each dNTP at 60 μM for 5 minutes at 55°C alone or in the presence of either DNAP-Stf or DNAP-Taq. The products were resolved by electrophoresis (5). The mobilities of uncleaved substrates that were in the snap-back form (denoted as s) or that had been converted to duplex (d) by oligonucleotide extension are indicated. A 20-nucleotide size marker is also indicated.

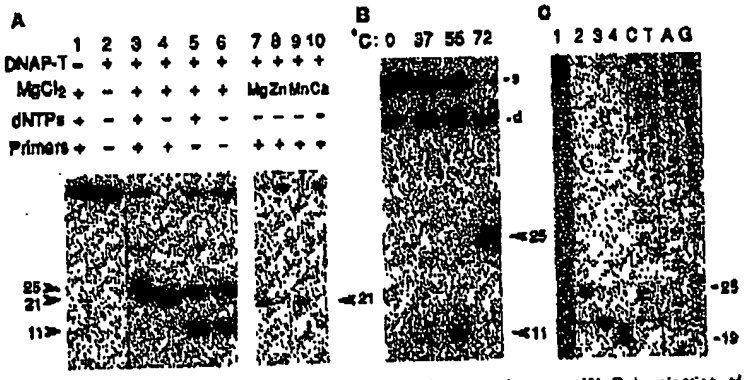


Fig. 2. Influence of reaction components and conditions on cleavage. (A) Determination of components required for cleavage (5). The 5′ end-labeled hairpin DNA shown in Fig. 1A was incubated for 30 minutes at 55°C, in 20 mM KCl, and in the presence of the indicated components. When included, dNTPs were present at 40 μM each. The primer left a gap of four nucleotides between its 3′ end and the first base pair of the hairpin duplex. The experiment in lane 7 is a repeat of lane 4, to serve as a marker; alternative cations were substituted at 1.5 mM. (A to C) Products were resolved by electrophoresis (5), and their lengths, in nucleotides, are indicated. (B) Effect of temperature on the site of cleavage in the absence of added primer. Reactions were incubated in the absence of KCl for 10 minutes at the indicated temperatures. The mobilities of the uncleaved substrate that either had (d) or had not (s) reannealed with its unlabeled complement (5) are indicated. (C) Effect of primer 3′ end locations on the sites of cleavage. The hairpin was incubated at 55°C for 10 minutes, with 50 mM KCl. In the absence (lane 1) or presence of primers that extended to various positions on the 3′ arm. The strong band in lane 2 was produced in the presence of a primer that extended to the end of the substrate duplex. In lanes 3 and 4, shorter primers were used, leaving gaps of four and six nucleotides, respectively, between the 3′ end of the primer and the end of the substrate duplex. The right side shows a dideoxynucleotide sequence ladder of the substrate, as a marker.

Best Available Copy

Supplied by The British Library - "The world's knowledge"

Alamar Exhibit 1002
Alamar v. Olink

Best Available Copy

was accentuated when both reactions were run in 50 mM KCl. In the presence of primer, the rate of cleavage was optimal at about 50 mM (Fig. 3A), but it approached zero at 150 mM KCl (6). In the absence of primer, the maximum rate was at about 20 mM KCl, and the reaction was almost completely inhibited at 50 mM KCl (Fig. 2C, lane 1, and Fig. 3B); also, the reactions (Fig. 3B) were incubated 15 times longer than the reactions shown in Fig. 3A.

The 5′ and 3′ arms. A potential substrate similar to that in Fig. 1A could be formed in a region of DNA containing a single-stranded gap provided that the gap contained self-complementary sequences. Since primer extension across a structured gap could lead to cleavage of the template strand, such a structure must be distinguished from a displaced arm in a bona fide cleavage substrate. A significant difference between in vivo complexes and the substrate that we used is the length of the 5′ and 3′ arms. Nick translation occurs without cleavage on templates with long 3′ arms, and therefore it seemed likely that the distinguishing feature would be the free end of the 5′ arm. As would be expected with this model, substrates that lack a free 5′ end such as circular M13 DNA, but not linear M13 DNA, are resistant to cleavage even in the presence of added primers and pilot oligonucleotides (see below) (6).

The importance of a free 5′ end on cleavage was investigated with the use of DNA hairpins similar to the one in Fig. 1A, but with progressively longer 5′ arms. In the presence of primer and 50 mM KCl, cleavage of a 5′ extension that is 27 nucleotides long was essentially complete within 2 minutes at 55°C (Fig. 4A, lanes 1 to 4). In contrast, cleavage of molecules with 5′ arms of 84 and 188 nucleotides was only about 90 and 40 percent complete after 20 minutes (lanes 5 to 8 and 9 to 12, respectively) in the presence of a molar excess (2.5 times greater) of enzyme.

Three possible explanations to account for this reduction in the efficiency of cleavage were tested. First, if available enzyme could bind nonspecifically to long 5′ arms, it might be tied up in nonproductive complexes. However, this is not the case since exposure of the enzyme to a fourfold molar excess of substrate with long 5′ arms prior to addition of substrate with short arms did not inhibit the cleavage of the shorter arms (Fig. 4B). Second, if the 5′ nuclease were active only when it was bound to a free 5′ end, then the reduction in the local concentration of such ends near the cleavage site, because of longer arms, might reduce the efficiency of the cleavage. This explanation is unlikely, since increasing the concentration of 5′ ends by the addition of a fortyfold molar excess of nonspecific oligonucleotides did not increase the efficiency of cleavage of long 5′ arms (6).

A third model postulates that the appropriate structure for cleavage is achieved only when the 5′ arm is threaded through the enzyme; in contrast to the clamp associated with *E. coli* DNAP-III (*11*), such a mechanism in DNAP-Taq would keep inappropriate DNA strands away from the 5′ nuclease site rather than hold DNA on the enzyme. Longer 5′ arms would be likely to contain structures that could bind the enzyme and impede movement of the strand through the enzyme. Consistent with this model, cleavage was inhibited by an oligonucleotide complementary to the 31 nucleotides at the 5′ end of the 84-nucleotide arm, but not by a nonspecific oligonucleotide (Fig. 4C). In addition, a 268-nucleotide 5′ arm was cleaved efficiently when the temperature was raised to 72°C and the KCl was omitted from the reaction (6), both being measures that would reduce adventitious structures on long arms. Because gapped templates, such as might exist during gap repair, replication, or transcription, do not have free 5′ ends that could be threaded to the cleavage domain of the enzyme, they would not be cleaved.

Cleavage does not appear to be inhibited by long 3′ arms, at least up to 2 kilobases (6). At the other extreme, very short 3′ arms can support cleavage in a primer-independent reaction. A complex containing a template strand with no 3′ arm did not elicit cleavage of the DNA substrate (Fig. 4D, lane 2), but a complex with the same template extended by a single mismatched dideoxynucleotide did support

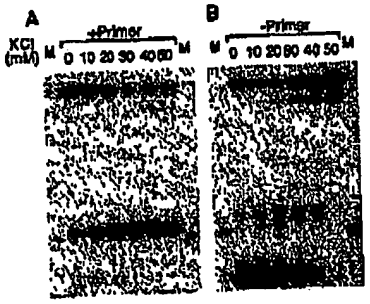


Fig. 3. Effect of KCl on cleavage in the presence and absence of primer. Reactions (5) were incubated at 55°C for 2 minutes or for 30 minutes in the presence (A) or absence (B) of a primer, respectively. The primer left a four-nucleotide gap. Undigested substrate is near the top of the gel. M is a 5′ end-labeled 19-nucleotide oligonucleotide. Substrates and cleavage products were resolved by gel electrophoresis (5).

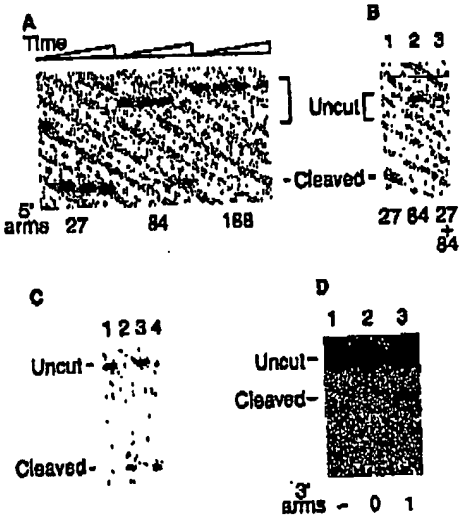


Fig. 4. Effect of arm length on the efficiency of cleavage. (A) Cleavage of substrates with different 5′ arm lengths. A substrate was produced by inserting the sequence shown in Fig. 1A into the Sma I site of plasmid pUC19. After digestion with Bam HI, the DNA was labeled at the 3′ end by filling in the ends with [α-^{32}P]dNTPs, and was recut with Kpn I, Bst NI, or Bgl I to create 5′ arms of 27, 84, or 188 nucleotides, respectively. Cleavage (5) was at 55°C for 0, 2, 6, and 20 minutes (left to right). (A to D) Substrates and cleavage products were resolved by electrophoresis (5). (B) Lack of interference by long 5′ arms. Two substrates, with 5′ extensions of 27 and 84 nucleotides, were incubated separately (lanes 1 and 2), or the shorter substrate was added to a reaction containing the longer substrate just prior to the initiation of cleavage (5) by addition of Mg^{2+} (lane 3). Incubations were at 55°C for 1 minute, and the DNAP-Taq was reduced to 0.05 unit in a 10-μl reaction mixture. (C) Inhibition of cleavage by structure in the 5′ arm. The substrate hairpin with a 5′ arm of 84 nucleotides was incubated for 20 minutes at 55°C with DNAP-Taq without any additional oligonucleotide (lane 2) or with an oligonucleotide that was complementary to 31 nucleotides at the 5′ end of the 84-nucleotide arm (lane 3), or a nonspecific oligonucleotide (lane 4). Material incubated without enzyme or oligonucleotide (lane 1) is shown as a marker. (D) Cleavage directed by a 3′ arm of one nucleotide. A non-hairpin PCR fragment (*22*) of 70 base pairs was heat-denatured and then incubated with DNAP-Taq in 20 mM KCl at 55°C for 30 minutes either alone (lane 1) or in the presence of an oligonucleotide, shown as the DNA pilot in Fig. 7. This latter molecule was either fully paired to the substrate (lane 2) or was extended by a single mismatched dideoxyadenosine (lane 3).

Supplied by The British Library - "The world's knowledge"

Alamar Exhibit 1002
Alamar v. Olink

cleavage of the substrate, albeit inefficiently (lane 3).

The 5′ nuclease activities of other DNAPs. The hairpin substrate shown in Fig. 1A was used to monitor 5′ nucleases of other DNAPs under conditions reported to be optimal for DNA synthesis by each enzyme (*12*). DNAPs from the eubacteria *Thermus thermophilus* (Tth) and *Thermus flavus* (Tfl) cleaved the substrate at the same places as did DNAP-Taq, both in the presence and absence of primer (Fig. 5A). As is the case with DNAP-Taq, the sites of cleavage by these enzymes were influenced by the location of the 3′ end of the primer. No 5′ nuclease activity was detected for the DNAPs from the archaebacteria *Pyrococcus furiosus* (Pfu) and *Thermococcus litoralis* (Tli) (*13*).

DNAP-Ec1 also cleaved the bifurcated duplex endonucleolytically (Fig. 5B), with the precise site of cleavage fixed by the location of the 3′ end of the primer (*14*). At 37°C, in the absence of added primer (lane 3), the major product released by DNAP-Ec1 was the same 11-nucleotide fragment released by DNAP-Taq from the extended duplex (Fig. 2). Addition of a primer that left a gap of four nucleotides (lane 4) greatly increased the extent of cleavage, but the products were primarily di- and trinucleotides; these would be generated if the sequence A-A-T-A-C near the free end of the 5′ arm were paired with the 3′ arm, adjacent to the bifurcation. A primer extending right up to the bifurcation directed cleavage between the first two base pairs in the hairpin duplex (lane 5). The diagrams to the right of the photograph depict ways that the complex could fold to

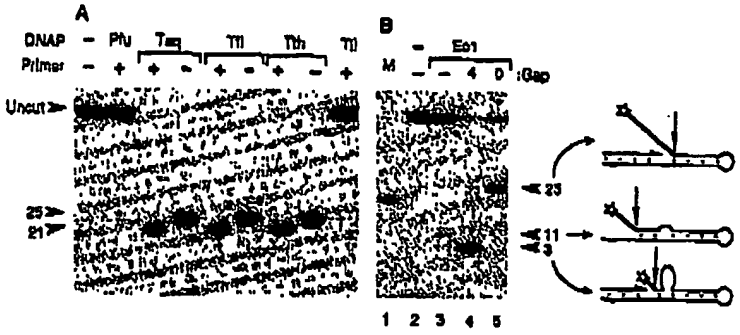


Fig. 5. Survey of several DNAPs for their ability to cleave the hairpin substrate in the presence (+) or absence (−) of added primer. (A) Endonucleases of DNAPs of several thermophilic bacteria (*12*). The 5′ end-labeled substrate and the primer shown in Fig. 1A were incubated at 55°C for 10 minutes with the enzymes indicated in the presence of primer (+) or at 72°C for 30 minutes in the absence of primer (−). The products were resolved by electrophoresis (*5*), and the lengths of the products, in nucleotides, are indicated. (B) Endonucleolytic cleavage by the 5′ nuclease of DNAP-Ec1. DNAP-Ec1 was incubated for 5 minutes at 37°C with the substrate in the absence of primer (lane 3) or with primers that left gaps of four or no nucleotides (lanes 4 and 5). No enzyme was added to the incubation in lane 2. A 20-nucleotide marker oligonucleotide is shown (lane 1); the mobilities of the fragments indicated on the right were determined by coelectrophoresis of known markers (Fig. 2) on the same gel (not shown) and by homochromatography. The diagram on the right shows possible base-paired structures that might form between the 5′ and 3′ arms of the hairpin, and which could give rise to the observed products.

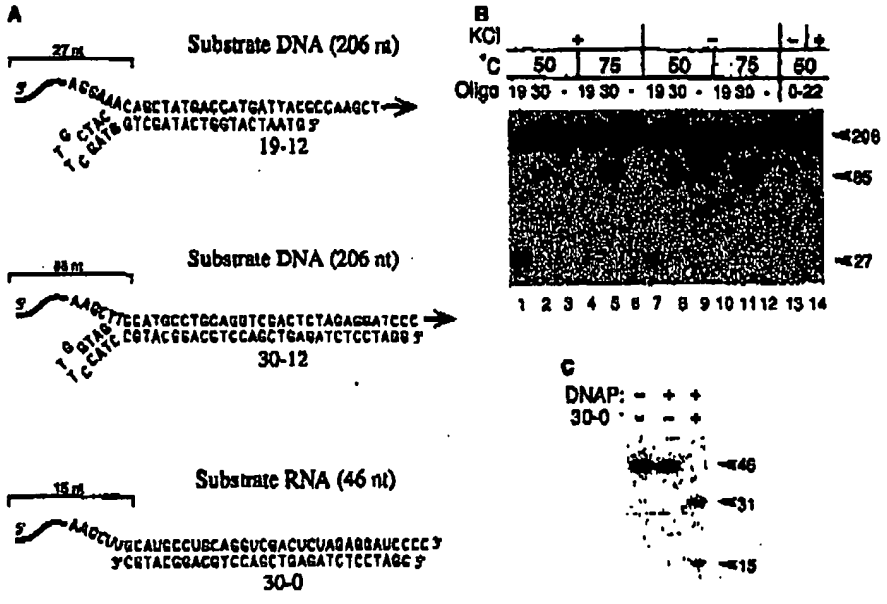


Fig. 6. Specific cleavage of nucleic acids targeted by pilot oligonucleotides. (A) Structures of pilot oligonucleotides base-paired to their complements in the substrate DNA or RNA. Two different sequences were used, which would pair to two different sites in the targeted nucleic acids, making complexes with different stabilities and distinguishable digestion products. The 206-nucleotide DNA substrate was 5′ end-labeled during PCR amplification of the polylinker region of plasmid pGEM3Z (*22*) with the use of the "−47 forward" and "−48 reverse" sequencing primers. (B and C) Products were resolved by electrophoresis (*5*), and their lengths, in nucleotides, are indicated. (B) Cleavage (*5*) of single-stranded DNA directed by pilot oligonucleotides 19–12 or 30–12 or an unrelated oligonucleotide 0–22. Incubations were for 10 minutes at 50° or 75°C, in the presence or absence of 50 mM KCl, as indicated. The 20-μl reaction mixtures contained 10 fmol of single end-labeled substrate DNA, 1 unit of DNAP-Taq (estimated to be 50 fmol) and 5 pmol of pilot oligonucleotide. The DNAs were heated to 95°C for 1 minute to denature the double-stranded substrate DNA, enzyme that had been heated was added, and the reactions were cooled to their final incubation temperatures as indicated. (C) Cleavage of RNA by DNAP-Taq in the presence of 30–0. The RNA substrate, made by T7 RNA polymerase in the presence of [α-^{32}P]UTP, corresponded to a truncated version of the DNA substrate used in (B). Reaction conditions resembled those used in (B), with 50 mM KCl and incubation was for 40 minutes at 55°C.

Best Available Copy

Supplied by The British Library - "The world's knowledge"

Alamar Exhibit 1002
Alamar v. Olink

Best Available Copy

give rise to the observed products. It is unclear whether DNAP-Taq would behave similarly at 37°C.

Since the amino acid sequences of the 5′ nuclease domains of DNAP-Ec1 and DNAP-Taq are about 34 percent similar, it is not surprising that they cleave the substrate DNA by similar mechanisms. It is unknown if other DNAPs and their associated 5′ exonucleases behave like the 5′ nucleases of eubacteria. The 5′ nuclease domain of DNAP-Taq shares about 19 percent homology with the 5′ exonuclease encoded by gene 6 of bacteriophage T7 (*15*). This nuclease, which is not covalently attached to a DNAP polymerization domain, is also able to cleave DNA endonucleolytically both in the presence and absence of primer, at a site similar or identical to the site that is cut by the 5′ nucleases described above (*6*). We have not yet tested eukaryotic DNAPs or associated exonucleases (*16*) for 5′ nuclease activity, but it has been observed that α-type DNAPs are homologous to the enzymes of the archaebacteria (*13*), which showed no 5′ nuclease activity (Fig. 5A).

Transcleavage of DNA and RNA. To learn whether a 5′ nuclease could be directed to cleave efficiently at any specific sequence of our choosing, we hybridized a partially complementary oligonucleotide (called a pilot) to sequences at the desired point of cleavage; the noncomplementary part of the pilot oligonucleotide provided a structure analogous to the 3′ arm of the template, whereas the 5′ region of the targeted nucleic acid became the 5′ arm. A primer was provided by designing the 3′ region of the pilot so that it would fold on itself, creating a short hairpin with a stabilizing tetra-loop (*17*). Two potential pilot oligonucleotides are shown in the structures at the top of Fig. 6A. Oligonucleotides 19–12 and 30–12 have 19 or 30 nucleotides, respectively, that are complementary to different sequences in the target DNA and should melt off their complements at about 50° and 75°C; both pilots have 12 nucleotides at their 3′ ends, which act as 3′ arms with base-paired primers attached.

Cleavage of the substrate DNA occurred in the presence of the pilot oligonucleotide 19–12 at 50° (Fig. 6B, lanes 1 and 7) but not at 75°C (lanes 4 and 10); in the presence of oligonucleotide 30–12, cleavage was observed at both temperatures. Cleavage did not occur at 80°C (*6*) or in the absence of added oligonucleotides (lanes 3, 6, or 12) except at 50°C where adventitious structures in the substrate allowed primer-independent cleavage in the absence of KCl (Fig. 6B, lane 9). A nonspecific oligonucleotide with no complementarity to the DNA target (but with the potential to form a hairpin with a stem of nine base pairs and a tetraloop at its 3′ end), called 0–22, did not direct cleavage at 50°C, either in the absence or presence of 50 mM KCl (lanes 13 and 14). Thus, the specificity of the cleavage reactions can be controlled by the extent of complementarity to the target and by the conditions of incubation.

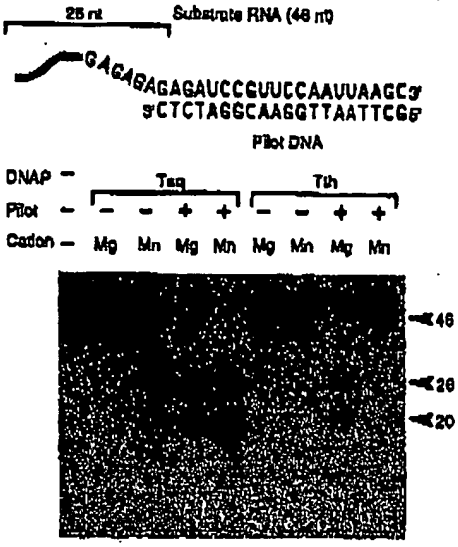


Fig. 7. Differential sensitivity of DNAPs to divalent cations during RNA cleavage. A 48-nucleotide-long RNA substrate, labeled during synthesis with [α-^{32}P]UTP, was incubated with either DNAP-Taq or DNAP-Tth in the presence (+) or absence (–) of the pilot oligonucleotide shown at the top of the figure. The reactions were as in Fig. 6C, with either 1 mM $MgCl_2$ or 1 mM $MnCl_2$, as indicated, and incubation was at 55°C for 20 minutes. The site of cleavage was deduced from the electrophoretic mobilities of the products in denaturing polyacrylamide gel electrophoresis (*5*).

An RNA version of the targeted sequence (Fig. 6A, bottom) was cleaved in a comparable place, in a reaction that was dependent on the presence of a pilot oligonucleotide resembling 30–12. In the case of RNA cleavage, a 3′ arm on the pilot oligonucleotide was not required (Fig. 6C). In contrast to previously described ribonuclease (RNase) H, which would be expected to cut the RNA in several places along the 30-base pair RNA-DNA duplex, the 5′ nuclease of DNAP-Taq is a structure-specific RNase H that cleaves at a single site opposite the 3′ end of the pilot DNA. It is surprising that an oligonucleotide lacking a 3′ arm can act as a pilot in directing cleavage of an RNA target since such oligonucleotides are unable to direct cleavage of DNA targets (Fig. 4C).

Both DNAP-Taq or DNAP-Tth cleaved RNA when incubated with an appropriate pilot oligonucleotide (Fig. 7, top) in the presence of Mg^{2+}. However, DNAP-Tth was unable to cleave the RNA in the presence of Mn^{2+}. The specificity of the 5′ nuclease of DNAP-Tth for Mg^{2+} may contribute to the ability of this enzyme to function as a reverse transcriptase in the presence of Mn^{2+} (*18*).

Some applications. An understanding of the 5′ nuclease activity has several practical consequences. Most exciting is the creation of cleavage reagents that can cut single-stranded polynucleotides in a highly specific manner at chosen sequences (*19*). Because cleavage depends on the ability of the 5′ nuclease to recognize substrates not by their sequence but by their structure, essentially any sequence can be targeted for cleavage by annealing it to an appropriate pilot oligonucleotide. By carrying out cleavage under very stringent hybridization conditions, such as at 75°C, a high degree of specificity of the site of cleavage is achieved (Fig. 6B), and the effects of long 5′ arms can be minimized. Sequences in one or both complementary strands of denatured DNA can be targeted for cleavage by the use of appropriate pilot oligonucleotides. Furthermore, use of conditions that favor cleavage may facilitate optimization of allele-specific PCR (*20*), so that primers with unpaired 3′ ends would act as pilot oligonucleotides to direct selective cleavage of unwanted templates.

If the 3′ extension of a pilot oligonucleotide forms a stable hairpin sequence (*17*), the pilot carries its own primer and can function under primer-dependent conditions (for example, in the presence of 50 mM KCl) (Fig. 6); this feature both greatly increases cleavage efficiency, and reduces the number of unwanted cleavages at regions of secondary structure in the target nucleic acid. Whether single-stranded RNA or DNA binding proteins can facilitate cleavage of long 5′ arms through disruption of secondary structures remains to be determined.

The 5′ nucleases of thermostable DNAPs may also be useful in probing for sequences or sequence variants in nucleic acid preparations. For example, the presence of a particular DNA might be detected

Supplied by The British Library - "The world's knowledge"

Alamar Exhibit 1002
Alamar v. Olink

by using that molecule as the template strand for cleavage of an added substrate probe, analogous to the method of Holland *et al.* (*4*). In this case, however, cleavage of the probe need not be coupled to synthesis, so it can be applied to unamplified DNA fragments or in post-PCR analysis. Additional demands could be made of the system, such as requiring that the template strand also bind a primer adjacent to the probe, thereby adding another level of stringency. Other less direct sequence probing applications of the 5′ nuclease could involve the production of nucleic acid fragments of optimal size for analyses such as single-strand conformation polymorphism (*21*).

REFERENCES AND NOTES

1. A. Kornberg and T. Baker, *DNA Replication* (Freeman, New York, ed. 2, 1992); I. R. Lehman, *Enzymes* 14, 15 (1976); C. M. Joyce and T. A. Steitz, *Trends Biochem. Sci.* 12, 288 (1987).
2. F. C. Lawyer *et al.*, *J. Biol. Chem.* 264, 6427 (1989); M. J. Longley, S. E. Bennett, D. W. Mosbaugh, *Nucleic Acids Res.* 18, 7317 (1990).
3. P. Setlow and A. Kornberg, *J. Biol. Chem.* 247, 232 (1972); R. C. Lundquist and B. M. Olivera, *Cell* 31, 53 (1982).
4. P. M. Holland, R. D. Abramson, R. Watson, D. H. Gelfand, *Proc. Natl. Acad. Sci. U.S.A.* 88, 7276 (1991).
5. Synthetic oligonucleotides were obtained from The Biotechnology Center at the University of Wisconsin–Madison, the DNAP-Taq was from Promega Corporation and Perkin-Elmer–Cetus; DNAP-Stf was from Perkin-Elmer–Cetus. Unless otherwise specified, the cleavage reactions included 10 fmol of heat-denatured, end-labeled substrate DNA (with the unlabeled strand also present) and 0.5 unit of DNAP-Taq (estimated to be about 25 fmol) in a total volume of 10 μl of 10 mM tris-Cl, pH 8.5, 50 mM KCl, 1.5 mM $MgCl_2$, and 0.2 μM primer oligonucleotide (complementary to the 3′ arm). Alternate concentrations of KCl, as well as practice times and temperatures used are indicated in individual figures and legends. Unless otherwise indicated, the primer used was that in Fig. 1A. Reactions were initiated at the final reaction temperature by the addition of either the $MgCl_2$ or enzyme. All reactions were stopped at incubation temperature by the addition of 8 μl of 95 percent formamide with 20 mM EDTA and 0.05 percent marker dyes per 10 μl of reaction volume. To resolve single-stranded from reduplexed uncut material, some samples were not heated prior to gel loading (Figs. 1B, 2B, and 3); all other samples were heated to 80°C for 2 minutes before denaturing polyacrylamide gel electrophoresis. Gels were at 10 or 20 percent acrylamide (19:1 cross-linked), with 7 M urea, in 89 mM tris-borate, pH 8.3, 2.8 mM EDTA. T_m calculations listed were made with the Oligo primer analysis software from National Biosciences, Inc., Plymouth, MN. These were determined with 0.25 μM as the DNA concentration, at either 15 or 65 mM total salt (the 1.5 mM $MgCl_2$ in all reactions was given the value of 15 mM salt for these calculations).
6. V. Lyamichev, M. A. D. Brow, J. E. Dahlberg, data not shown.
7. H. A. Erlich, D. Gelfand, J. J. Sninsky, *Science* 252, 1643 (1991).
8. Presumably the short extension of the duplex was stabilized by being in the same molecule, near to the hairpin.
9. In these experiments, primer-independent cleavages, such as those observed in Fig. 2A, lanes 3 and 4, and Fig. 2B, became relatively insignificant because of the increased efficiency of primer-directed cleavage.
10. Primers worked equally well if they had natural 2′ deoxy- or chain-terminating 2′,3′-dideoxynucleotide ends (*6*).
11. X. P. Kong, R. Onrust, M. O'Donnell, J. Kuriyan, *Cell* 69, 425 (1992).
12. DNAP-Eo1 was obtained from Promega Corporation, the DNAP of *Pyrococcus furiosus* [DNAP-Pfu, M. Bergseid *et al.*, *Strategies* (Stratagene) 4, 34 (1991)] was from Stratagene, the DNAP of *Thermococcus litoralis* [DNAP-Tli, Vent$_R$(exo-), F. B. Perler *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* 89, 5577 (1992)] was from New England Biolabs, the DNAP of *Thermus flavus* [DNAP-Tfl, A. S. Kaledin, A. G. Slyusarenko, S. I. Gorodetskii, *Biokhimiya* 46, 1576 (1981)] was from Epicentre Technologies, and the DNAP of *Thermus thermophilus* [DNAP-Tth, N. Carballeira, M. Nazabal, J. Brito, O. Garcia, *Biotechniques* 9, 276 (1990); (*18*)] was from U.S. Biochemical; 0.5 unit of each DNA polymerase was assayed in a 20-μl reaction, with either the buffers supplied by the manufacturers for the primer-directed reactions, or 10 mM tris-Cl, pH 8.5, 1.5 mM $MgCl_2$, and 20 mM KCl for primer-independent reactions. Reactions with DNAP-Eo1 were in 5 mM tris-Cl, pH 7.5, 5 mM $MgCl_2$, and 0.1 mM dithiothreitol, and reaction mixtures were at 37°C before the addition of enzyme.
13. The DNAPs from *Thermococcus litoralis* and *Pyrococcus furiosus* share little sequence homology with eubacterial enzymes (J. Ito and D. K. Braithwaite, *Nucleic Acids Res.* 19, 4045 (1991); E. J. Mathur, M. W. W. Adams, W. N. Callen, J. M. Cline, *ibid.*, p. 6952; see also Perler *et al.*, in (*12*)).
14. The 3′ exonuclease of DNAP-Eo1 can quickly destroy the single-stranded product. Therefore, 100 pmol of an unrelated single-stranded oligonucleotide was added to the reaction to competitively inhibit the digestion of the cleavage product and the 3′ arm without interfering with the 5′ nuclease release of the 5′ arm.
15. J. J. Dunn and F. W. Studier, *J. Mol. Biol.* 166, 477 (1983).
16. T. Lindahl, J. A. Gally, G. M. Edelman, *Proc. Natl. Acad. Sci. U.S.A.* 62, 597 (1969); A. M. Pedrini and L. Goosenan, *J. Biol. Chem.* 258, 1536 (1983); Y. Ishimi, A. Claude, P. Bullock, J. Hurwitz, *ibid.* 263, 19723 (1988); J. J. Crute and I. R. Lehman, *ibid.* 264, 19266 (1989); M. Goulian, S. H. Richards, C. J. Heard, B. M. Bigsby, *ibid.* 265, 18461 (1990); G. Siegal, J. J. Turchi, T. W. Myers, R. A. Bambara, *Proc. Natl. Acad. Sci. U.S.A.* 89, 9377 (1992).
17. V. P. Antao, S. Y. Lai, I. Tinoco, Jr., *Nucleic Acids Res.* 19, 5901 (1991).
18. T. W. Myers and D. H. Gelfand, *Biochemistry* 30, 7661 (1991).
19. Although the precise site of cleavage in a primer-directed reaction is usually between the first two base pairs of the substrate-template duplex, we have noticed that cleavage in certain complexes may be shifted by one or two nucleotides, even if there is no gap between the primer and the bifurcation (*5*).
20. U. B. Gyllensten and H. A. Erlich, *Proc. Natl. Acad. Sci. U.S.A.* 85, 7652 (1988).
21. M. Orita *et al.*, *Genomics* 5, 874 (1989); K. Hayashi, *PCR Methods Appl.* 1, 34 (1991).
22. R. K. Saiki *et al.*, *Science* 239, 487 (1988); K. B. Mullis and F. A. Faloona, *Methods Enzymol.* 155, 335 (1987). Polymerase chain reactions comprised 1 ng of plasmid target DNA, 5 to 25 pmol of each primer (one of which was 5′-labeled), 40 μM each dNTP, and 2.5 units of DNAP-Taq or DNAP-Stf, in a 50-μl solution of 10 mM tris-Cl (pH 8.3). The DNAP-Taq reactions included 50 mM KCl and 1.5 mM $MgCl_2$, while the DNAP-Stf reactions included 10 mM KCl and 4 mM $MgCl_2$. The temperature profile was 95°C for 60 seconds, 55°C for 1 minute, and 72°C for 1 minute, through 30 cycles. Plasmid pGEM3Z was from Promega Corporation. The sequence shown in Fig. 1A cannot be amplified with DNAP-Taq but can be made with DNAP-Stf.

23. We thank D. Gelfand for discussions relating to the experiments in Fig. 1B, M. Lund for discussions and comments, and M. Wickens for critically reading the manuscript. Supported by NIH grant 30220 (J.E.D.).

9 September 1992; accepted 17 March 1993

Best Available Copy

Supplied by The British Library - "The world's knowledge"

Alamar Exhibit 1002
Alamar v. Olink

Best Available Copy

CIS display: *In vitro* selection of peptides from libraries of protein–DNA complexes

Richard Odegrip, David Coomber, Bill Eldridge, Rosemarie Hederer, Philip A. Kuhlman, Christopher Ullman, Kevin FitzGerald*, and Duncan McGregor

Isogenica Ltd., Babraham Hall, Babraham CB2 4AT, United Kingdom

Communicated by Aaron Klug, Medical Research Council, Cambridge, United Kingdom, January 12, 2004 (received for review December 2, 2003)

We describe the development of an *in vitro* library selection system (CIS display) that exploits the ability of a DNA replication initiator protein (RepA) to bind exclusively to the template DNA from which it has been expressed, a property called cis-activity. A diverse peptide library is created by ligation of DNA fragments of random sequence to a DNA fragment that encodes RepA. After *in vitro* transcription and translation, a pool of protein–DNA complexes is formed where each protein is stably associated with the DNA that encodes it. These complexes are amenable to the affinity selection of ligands to targets of interest. Here we show that RepA is a highly faithful cis-acting DNA-binding protein and demonstrate that libraries encoding $>10^{12}$ random 18-mer peptides can be constructed and used to isolate peptides that bind specifically to disparate targets. The use of DNA to encode the displayed peptides offers advantages over *in vitro* peptide display systems that use mRNA.

The physical linkage of populations of peptides to their encoding nucleic acids to create large, diverse display libraries has provided a rich source of ligands to a wide range of target molecules. Enrichment of ligands from these libraries is achieved by the coselection of target-binding peptides along with their associated encoding nucleic acids, which allows the subsequent identification of the selected peptide sequences. This approach has been most widely exemplified by using the phage display system (1–4). However, phage display library construction requires the insertion of the library DNA into bacterial cells, and the efficiency of bacterial transformation usually restricts library sizes to the 10^9 to 10^{10} range, although *in vivo* constructed libraries $>10^{10}$ have been reported (5, 6). Similarly, the display of peptides on bacterial or yeast surfaces or the display of peptides directly on their encoding plasmids all require a transformation step, which imposes a limit to the library sizes that can be constructed (7–9).

In general, it is accepted that a correlation exists between library size and the affinity of ligands that can be isolated from them (10). This correlation has prompted the development of display technologies in which the size-limiting transformation step is unnecessary, allowing ever-larger display libraries to be constructed. These new technologies enable higher affinity ligands to be obtained through the sampling of an increased structural repertoire, which is made possible through the generation of libraries that are up to four orders of magnitude larger than those that can be constructed for phage display. Examples of *in vitro* display technologies include "*in vitro* virus" (11) or mRNA display (12–14) and ribosome display (15–19). A common limitation with these technologies is that mRNA is used as the library-encoding nucleic acid, which may be prone to rapid degradation. The selected mRNA also requires a reverse-transcription step before amplification.

Existing DNA-based *in vitro* selection systems are based on emulsion encapsulation of DNA and are limited to libraries of 10^9 to 10^{10} per ml (20, 21). Here we describe an alternate DNA-based approach that does not require any compartmentalization of the library-encoding nucleic acid. The system, termed CIS display, exploits the high-fidelity cis-activity that is exhibited by a group of bacterial plasmid DNA-replication initiation proteins typified by RepA of the R1 plasmid (22, 23). In this context, cis-activity refers to the property of the RepA family of proteins to bind exclusively to the template DNA from which they have been expressed. Other cis-acting proteins exist, such as P2A, which exhibits cis-binding properties similar to RepA (24). In contrast to P2A, RepA does not bind covalently to its target DNA (25, 26). Also, RepA and P2A differ considerably in size (RepA is 33 kDa and P2A is 86kDa), making RepA more attractive for the display of polypeptide libraries.

R1 plasmid replication is initiated through the binding of RepA to the plasmid origin of replication (*ori*). *Ori* is separated from the RepA-coding sequence by a DNA element termed *CIS*. This element is thought to be critical in controlling the cis-activity of RepA (Fig. 1; ref. 27). The consensus model for cis-activity is that the *CIS* element, which contains a rho-dependent transcriptional terminator, causes the host RNA polymerase to stall. This delay allows nascent RepA polypeptide emerging from translating ribosomes to bind transiently to *CIS*, which in turn directs the protein to bind to the adjacent *ori* site (28, 29).

By genetically fusing peptide libraries to the N terminus of the RepA protein, we can achieve a direct linkage of peptides to the DNA molecules that encode them; thus, the link between genotype to phenotype that is the common feature of display technologies is established (Fig. 1). In this study we demonstrate that *in vitro* synthesized RepA has the high-fidelity cis-activity required to allow the display and selection of target-specific ligands. We show >1,000-fold enrichment of peptides in each round of affinity selection against a relevant target and then report the selection of peptide ligands to two well characterized antibodies (anti-P53 DO1 and anti-FLAG M2) and to lysozyme from large, diverse libraries.

Materials and Methods

CIS Display Libraries Construction. All enzymes were purchased from New England Biolabs. All PCR reactions contained 12.5 pmol of each of the primers, 2.5 units of TaqDeepVent DNA polymerase mixture (20:1), 250 μM dNTP (Roche Diagnostics) and 1× ThermoPol Buffer per 50-μl PCR reaction. PCR reactions were carried out on a Techne Techgene PCR machine for one cycle of 1 min and 45 s at 94°C, followed by 20–30 cycles at 94°C, 15 s; at 60°C, 30 s; at 72°C, 1–3 min, followed by a final extension of 5 min at 72°C. The *tac*-NNB-*RepA-CIS-ori* PCR construct (sequence is provided in *Data Set 1*, which is published as supporting information on the PNAS web site) was prepared by appending an 18-mer NNB library (where N is any nucleotide, and B is either C, T, or G to minimize the frequency of TAG stop codons) to the *tac* promoter by PCR and then ligating it to the *RepA-CIS-ori* region followed by PCR amplification. The *RepA-CIS-ori* region was amplified by PCR from the R1 plasmid

Abbreviation: ACTH, corticotropin.

*To whom correspondence should be addressed. E-mail: kevin.fitzgerald@isogenica.com.

© 2004 by The National Academy of Sciences of the USA

www.pnas.org/cgi/doi/10.1073/pnas.0400219101

Alamar Exhibit 1002
Alamar v. Olink

Best Available Copy

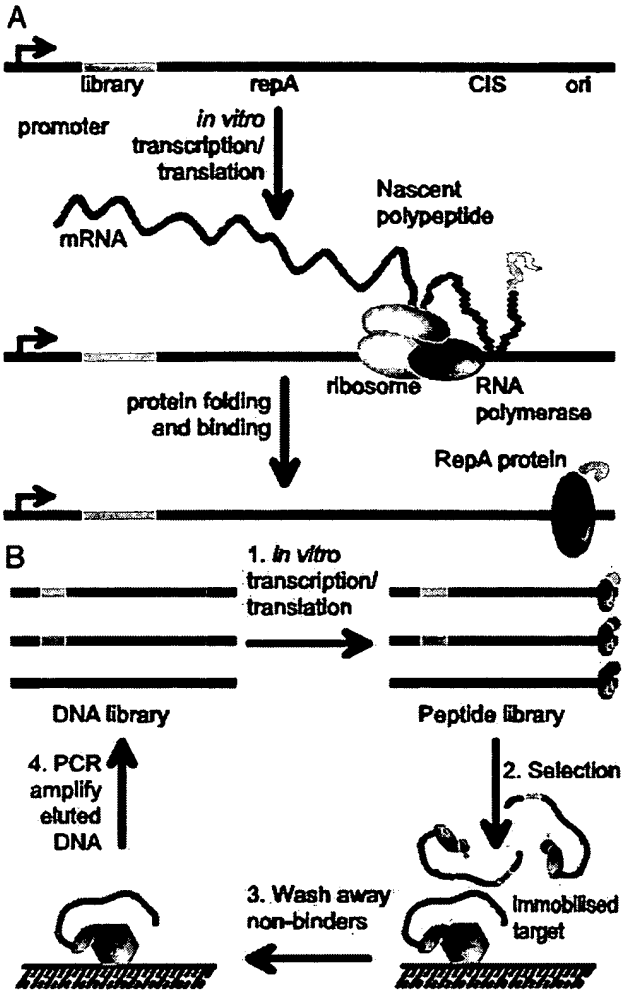


Fig. 1. The principle of the CIS display technology. (*A*) Template DNA encoding an N-terminal library peptide is ligated to the RepA gene. *In vitro* transcription is initiated at the promoter and pauses when the RNA polymerase reaches the CIS element. Concurrent translation produces the RepA protein, which transiently interacts with the CIS element, thereby forcing its subsequent binding to the adjacent *ori* sequence. This process establishes a faithful linkage between a template DNA and the expressed polypeptide that it encodes. (*B*) CIS display selections begin with the construction of a peptide-encoding DNA library followed by *in vitro* transcription/translation to form a pool of protein–DNA complexes (step 1). The library pool is incubated with an immobilized target (step 2), nonbinding peptides are washed away (step 3), and the retained DNA that encodes the target-binding peptides is eluted and amplified by PCR (step 4), to form a DNA library ready for the next round of selection. After three to five rounds of selection, recovered DNA is cloned into an appropriate expression vector for the identification of individual target-binding peptide sequences.

[obtained from Public Health Laboratory Service, London, U.K. (ECO K12 J53 R1), GenBank accession no. V00351, for the R1 plasmid] by using the primers BSPREPAFOR and ORIREV408 (for all oligonucleotide sequences, see Table 1). The library was constructed by PCR with a *tac* promoter as template and the primers TACFARUP and NTERM18-mer. Of each of the two reaction products ≈5 μg were digested and ligated with 50 units each of *Not*I and *Psp*OmI and 2,000 units of T4 DNA ligase in a 300-μl reaction volume, also containing 5 mM ATP (Sigma), 0.1 mg/ml acetylated BSA, and 1× NEB Buffer 4. The amount of full-length ligated product was quantified on an agarose gel, giving a library size of $>10^{12}$ molecules. The resulting template was further amplified by PCR by using the nesting primers TACFAR1 and ORIREV108 to yield enough template for multiple *in vitro* transcription/translation reactions.

V5-RepA and Cκ-RepA Construction. The human Cκ DNA was amplified by PCR from plasmid pDM6 (30) by using primers TACCKFOR and NOTICKREV, followed by assembly to the *tac* promoter and digestion and ligation to the *RepA-CIS-ori* region. The *tac* promoter was amplified by using primers TACFARUP and TACREV. The V5 epitope (GKPIPNPLLGLDST; ref. 31) DNA was appended to the *RepA-CIS-ori* region by PCR with primers V5REPAFOR and ORIREV408 followed by overlapping PCR assembly to the *tac* promoter (see above).

Construction of M13 gpVIII Phagemid Vector. The M13 gene 8 was PCR-amplified with primers ECORIGENE8 and NOTIGENE8 followed by digestion with *Eco*RI and *Not*I. Gel-purified product was ligated into a similarly digested plasmid pDM6 (30) and subsequently electroporated into electrocompetent TG-1 cells.

Affinity Selection. *In vitro* transcription and translation was performed in an *Escherichia coli* (strain SL119; ref. 32) S-30 lysate system (33) for up to 30 min at 30°C and then diluted 10-fold with blocking buffer (2–4% Marvel, 0.1 mg/ml herring sperm DNA, 2.5 mg/ml heparin, in PBS or TBS). Typically, 2–4 μg of linear DNA was added per 50 μl of S-30 lysate reaction. When performing library selections 20 μg of library DNA was added to 250 μl of S-30 lysate reaction in the first round; in subsequent rounds, 5 μg of library DNA was added in 100 μl of S-30 lysate reaction.

Solid-Phase Selection. Targets were immobilized on either 4 ml of NUNC Star Immunotubes (Fisher Scientific), Maxisorp plates (VWR Scientific) (both coated overnight at 4°C), or captured by M-280 streptavidin-coated magnetic beads (Dynal, Oslo). Coating concentrations were 50 μg/ml for lysozyme (Sigma), 10 μg/ml for anti-V5 antibody, anti-human κ-chain antibody (ABCAM, Cambridge, U.K.), DO1 (Oncogene Research, Boston), and anti-corticotropin (anti-ACTH) antibody (Immunologicals Direct, Oxford) in PBS. Immobilized targets were washed twice with PBS and blocked for 1 h at room temperature with Blocking Buffer, and then washed twice with PBS. The diluted transcription/translation reactions were incubated for 1 h at room temperature before washing 6 to 12 times with PBS/0.1%-Tween-20, followed by 6–12 washes with PBS (filled and empty). DNA was eluted with 500 μl of PB solution (Qiagen, Crawley, U.K.) and purified by using QIAquick PCR purification kit (Qiagen). Half of the eluted material was added to a recovery PCR reaction where the N-terminal library region was amplified by using a nested primer for each round (TACFAR2 in round 1, TACFAR3 in round 2, and so forth up to TACFAR5). All recovered PCR product was reattached to the *RepA-CIS-ori* by restriction ligation (as above) and further amplified, as described above, but with the respective nesting primer to produce input DNA for the next round of selection.

In Solution Selection. Two micrograms of biotinylated anti-FLAG M2 antibody (Sigma) or 25 μg of biotinylated lysozyme (the level of biotin incorporation was determined to be five biotin molecules per lysozyme molecule by using the 2-(4′-hydroxyazobenzene)benzoic acid method; EZ-Link biotinylation kit, Pierce) was added to the diluted transcription/translation reactions (see *Solid-Phase Selection*) and allowed to bind for 2 h with occasional shaking. Bound library complexes were captured onto- 50 μl of washed and blocked streptavidin-coated magnetic beads (Dynal, Oslo) for 15 min, then washed four times with 1 ml of PBS/0.1% Tween 20 and two times with 1 ml of PBS, before elution with PB solution and processing as above.

Alamar Exhibit 1002
Alamar v. Olink

Best Available Copy

Table 1. Oligonucleotides (presented 5′–3′)

BSPREPAFOR, CTGGAGATGGCATCAAGGGCCCCAACTGATCTTCACCAAACGTATTACC
ORIREV408, CGTAAGCCGGTACTGATTGA
TACFARUP, CAGTTGATCGGCGCGAGATT
NTERM18mer, ACATACCGTCATGCGGCCGCTGATCCTCCTCCCCC(VNN)$_{18}$GGCCATGGTAGATCCTGTTTC
TACFAR1, CGCCAATCAGCAACGACTGT
ORIREV108, GGGCTTTGTGGTTTCAGTTC
TACCKFOR, CAGGAAACAGGATCTACCATGACTGTGGCTGCACCATCTGTCTTC
NOTICKREV, CGTTTGGTGAAGATCAGTTGCGGCCGCTGATCCTCCTCCCCCTCCCCTGTTGAAGCTCTTTGTG
TACREV, GGTAGATCCTGTTTCCTGTGTG
V5REPAFOR, CACAGGAAACAGGATCTACCATGGCCGGAAAACCTATCCCAAACCCTCTCCTAGGACTGGATT CAACGGGGGGAGGAGGATCAGCGGCCGCAACTGATCTTCACCAAACG
ECORIGENE8, CGCCGGAATTCTTATCAGCTTGCTTTCGAGG
NOTIGENE8, CTGCAGTAATAGGCGGCCGCAGGGGGAGGAGGGTCCGCTGAGGGTGACGATCCCGCA
TACFAR2, AACGTGGCTGGCCTGGTTCA
TACFAR3, GATAAGAGACACCGGCATAC
TACFAR4, GGCGCTATCATGCCATACCG
TACFAR5, ACCATTCGGCTAGCGATGAC
TAC6, CCCCATCCCCCTGTTGACAATTAATC
NOTIRECREV, GGTGAAGATCAGTTGCGGCCGCTGATCCTCCTC

All oligonucleotides were purchased from Sigma Genosys (Pampisford, U.K.). The means of oligonucleotide purification were as follows: <35 bases, desalting; 35–50 bases, reverse-phase cartridge purification; >50 bases, PAGE purification.

ELISA Screening of Selected Peptides. After each round of selection, recovered DNA was PCR-amplified with primers TAC6 and NOTIRECREV and purified and digested with *Not*I and *Nco*I. The DNA was then ligated into a similarly digested M13 gpVIII phagemid vector, transformed into *E. coli* TG-1 cells, and plated on 2% glucose, 2× TY, 100 μg/ml ampicillin plates. Individual colonies were grown for the production of phage particles as described (32). NUNC Maxisorp plates were coated with 100 ng per well of either anti-human Cκ antibody, anti-V5 antibody (ABCAM, Cambridge), anti-FLAG M2 antibody (Sigma), 1 μg/well of DO1 antibody (Oncogene Research), or 5 μg per well of lysozyme (Sigma) in PBS overnight at 4°C. The horseradish peroxidase-conjugated anti-M13 secondary antibody (Amersham Pharmacia Biotech) was diluted 1:5,000 in 2% Marvel in PBS. ELISA assays were performed as described (30). The assay was developed with SureBlue TMB peroxidase substrate (Insight Biotechnology, Middlesex, U.K.) and read at 450 nm.

Results

RepA binds to DNA in cis when expressed *in vitro* from a linear DNA template. To demonstrate cis-activity, two linear *RepA* DNA constructs containing different-sized inserts, the 14-aa V5 peptide tag (31) and the 105-aa human Cκ Ig constant domain, were mixed in a 1:1 ratio and translated *in vitro*; the resulting protein–DNA complexes were subjected to affinity selection against three immobilized antibodies. In this selection binding of complexes to anti-Cκ antibodies resulted in the specific recovery of Cκ-*RepA* DNA but not the V5-*RepA* DNA and *vice versa*, as distinguished by their size in agarose gel electrophoresis (Fig. 2*A*). Binding of complexes to an irrelevant antibody (anti-ACTH) resulted in the recovery of background levels of DNA at the original ratio. These results clearly demonstrate that RepA acts in cis *in vitro* with high fidelity.

High Level of Enrichment During Sequential Rounds of Affinity Selection. To assess the suitability of CIS display under library selection conditions and to estimate the enrichment levels possible with the system, Cκ-*RepA* DNA was mixed with V5-*RepA* DNA at a ratio of either $1:10^8$ or $1:10^{10}$ (Cκ/V5) and then subjected to four rounds of selection either against an anti-Cκ antibody or against a negative control antibody target (anti-FLAG M2). DNA recovery after each round of selection with either target was assessed by PCR amplification (Fig. 2 *B* and *C*). Only two rounds of selection against anti-Cκ were required to enrich Cκ-RepA complexes from the $1:10^8$ starting ratio so that the Cκ-*RepA* DNA was visible in the products from the PCR. Because the recovered DNA corresponding to Cκ-RepA comprised ≈1% (calculated by densitometry by using the gel documentation program GENETOOLS, SynGene, Cambridge, U.K.) of the total DNA amplified at this stage of the selection process (Fig. 2*B*, lane 2), this percentage suggests that an enrichment of ≈1,000-fold per round had been achieved under these conditions. For the $1:10^{10}$ starting ratio, a third round of selection was required before the Cκ-RepA product was visible (Fig. 2*C*, lane 3), again suggesting 1,000-fold enrichment per round. As expected, further rounds of selection continued to increase the relative yield of the Cκ-*RepA* PCR product.

Selection Against Anti-P53, Anti-FLAG Antibody, and Lysozyme. To further validate the technology we selected an 18-mer peptide-RepA library against two antibody targets; the monoclonal anti-P53 antibody DO1 and the monoclonal anti-FLAG antibody M2. DO1 binds to a region within the N terminus of p53 from amino acid residues 11–25 (34). M2 is known to bind with 15 nM affinity to the FLAG epitope sequence (DYKDDDDK) and have a known preference for the sequence DYKXXD, where X is any amino acid (35, 36). A library of 10^{12} DNA molecules was used as the initial input in the first rounds of selection against DO1 and M2. The recovered DNA was amplified by PCR and used to prepare input DNA for the next round of selection. After each round, recovered DNA was cloned into an M13 gpVIII phagemid vector and tested for specific binding to antibody. Despite the well known sequence and insert-size bias of gpVIII, the pVIII fusion partner was chosen to allow antibody capture of all possible binding peptides. We are investigating both monovalent bacterial expression systems (such as maltose-binding protein) and *in vitro* expression systems for the characterization of selected peptides.

Eleven ELISA-reactive peptides from the fifth round of selection against DO1 were sequenced, and two different peptide sequences were identified that had homology to the P53 epitope (Fig. 3*A*). In the selection against M2, 23 specific binding clones were sequenced, and eight different peptide sequences identified. These peptides showed significant homology to the

Alamar Exhibit 1002
Alamar v. Olink

Best Available Copy

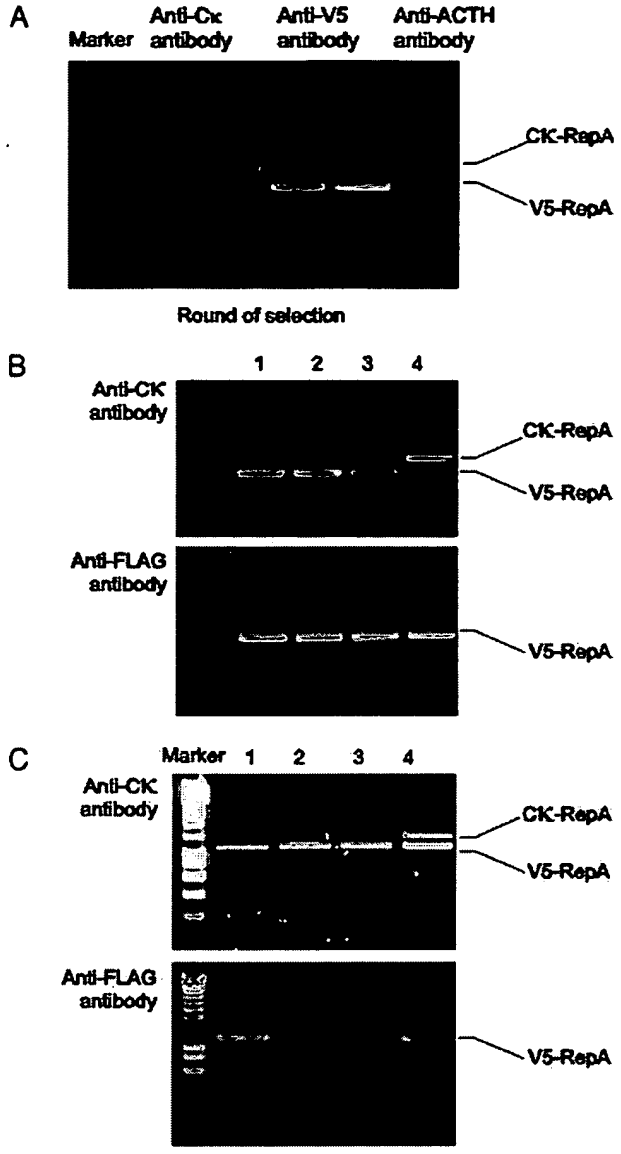


Fig. 2. Demonstration of target-specific ligand selection by using CIS display. (*A*) A 1:1 mixture of Cκ-RepA DNA and V5-repA DNA was prepared, transcribed and translated *in vitro*, and selected against either anti-Ck antibody, anti-V5 antibody, or anti-ACTH antibody. Recovered DNA was amplified with universal primers and separated by agarose gel electrophoresis on the basis of the size difference between the Cκ domain and the V5 peptide tag. The amount of DNA specifically recovered when by using the anti-Cκ and anti-V5 antibodies is shown in duplicate along with a single negative control (anti-ACTH) to reflect background recovery. (*B*) A $1:10^8$ dilution of Cκ-RepA DNA into V5-repA DNA was prepared and subjected to four rounds of selection against either the anti-Cκ antibody or the anti-FLAG antibody as a negative control. The four rounds of selection are indicated along with the positions of the two recovered PCR products. (*C*) A $1:10^{10}$ dilution of Cκ-RepA DNA into V5-repA DNA was prepared and subjected to four rounds of selection against either the anti-Cκ antibody or the anti-FLAG antibody as a negative control. The four rounds of selection are indicated along with the positions of the two recovered PCR products. Marker refers to the DNA Hyperladder (Bioline, London).

FLAG sequence (Fig. 3*B*). All sequences contained the required Y2 and K3 and the preferred D6; only one did not contain D1. Other similarities among the selected peptides were identified. A bias existed for Tyr or Asn at position −1, and Pro and Ser

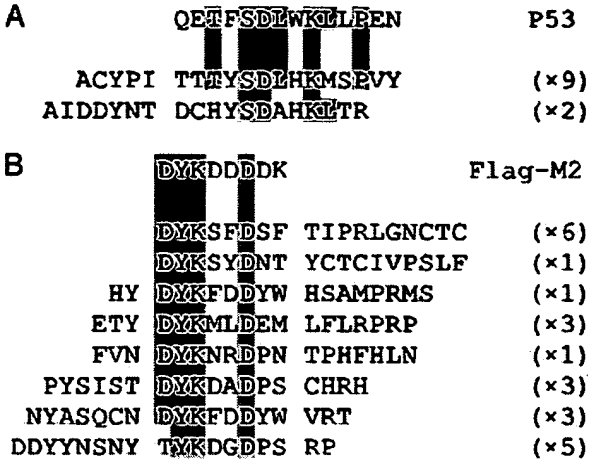


Fig. 3. Homology between selected binding peptides and respective epitope sequences. The sequences of 11 DO1 (*A*) and 23 M2-binding peptides (*B*) are aligned with the P53-DO1 and FLAG-M2 epitope sequences, respectively. Clones from the DO1 selection are taken from round 5, and the M2 peptides originate from rounds 4 and 5. Regions of homology are shaded, and the frequency with which each clone was identified is shown in parentheses.

were prevalent at positions 7 and 8 . In both selections, >90% of clones screened were binding to the target after five rounds, whereas in round 4, the in-solution anti-FLAG selection was still saturating, and the solid-phase anti-p53 selection contained ≈15% of clones binding to the target.

Finally, to demonstrate that the method was capable of deriving peptide ligands to nonantibody targets, a selection was performed against lysozyme. After five rounds of selection, 12% of those clones that were tested were observed to be reactive against lysozyme. The specificities of two representative clones were assessed by ELISA (Fig. 4) by using ovalbumin, BSA, and anti-V5 antibody as negative control targets. The ELISA signals against the lysozyme were between 6- and 12-fold greater than those observed against the control targets.

Discussion

The establishment of a physical linkage between a polypeptide and its encoding nucleic acid is central to the development of

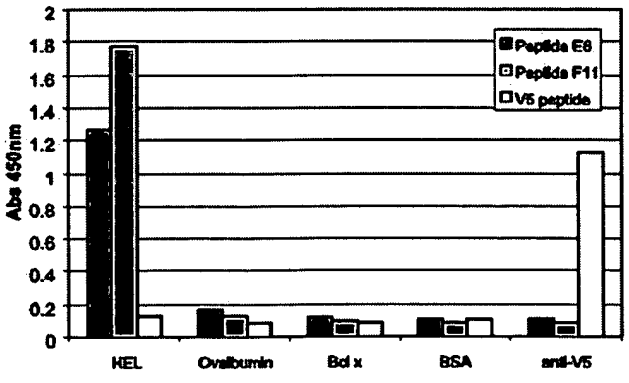


Fig. 4. ELISA of peptides selected against lysozyme both on solid-phase and in-solution selection. After five rounds of selection, peptides expressed on phage as M13 gpVIII fusions were checked for specificity of binding to lysozyme (50 μg/ml), ovalbumin (50 μg/ml), Bcl-X (10 μg/ml), BSA (200 μg/ml), and anti-V5 antibody (10 μg/ml). Two examples of the selected clones are shown; E6 and F11, from solid-phase and in-solution selection, respectively. The V5 peptide fused to gpVIII was used as a both a negative control (against lysozyme) and a positive control (against anti-V5 antibody) for the assay.

Alamar Exhibit 1002
Alamar v. Olink

Best Available Copy

display technologies. Of equal importance is that each nucleic acid within a complex pool of nucleic acids can associate exclusively with its own encoded polypeptide. We have shown that a cis-acting DNA-binding protein, RepA, can facilitate this association through the simple *in vitro* transcription and translation of appropriately constructed DNA templates.

The highly faithful cis-activity of RepA is shown here by the exclusive, target-specific retention of DNA that encoded relevant peptide moieties (Fig. 2*A*) and by the 1,000-fold target-specific enrichment of the appropriate DNA in each round of selection against a control target (Fig. 2 *B* and *C*). A possible explanation for the very low level of trans-activity of RepA is the rapid inactivation of nonbound RepA that has been reported (24).

We have further shown the utility of the CIS display system by achieving successful affinity selection of peptide ligands to the well characterized antibodies DO1 and M2 and to a nonantibody target, lysozyme. The results from these selections compare well with those that have been achieved by using ribosome display, mRNA display, and emulsion-based selections (16, 17, 21, 37). However, CIS display has several advantages over these alternative *in vitro* selection systems. It is likely that the use of DNA to encode the displayed peptides provides advantages over RNA-based *in vitro* selection methods. Very large libraries can be rapidly constructed and screened without separating transcription and translation steps and without purification of the protein–DNA complexes before selection. Also, control of divalent cation concentration is not required, and the complexes do not require incubation under sterile or ribonuclease-free conditions. These simplified selection conditions are in contrast to the tight control of these parameters that is normally required when using RNA-based display technologies. Indeed, a method of protecting the mRNA of ribosome display particles using RNA-binding proteins has recently been published in a patent application (38), and protocols for the conversion of the library-encoding mRNA to cDNA in an attempt to stabilize mRNA display complexes before selection have also been reported (39). The high inherent stability of RepA–DNA complexes has been confirmed by incubating *in vitro* expressed Cκ-RepA with an anti-Cκ antibody for 48 h before elution and recovery of associated DNA. In this experiment no reduction in DNA yield was observed compared with that obtained using standard selection conditions (not shown). Also, preliminary data suggest that we are able to display and demonstrate folding of larger proteins, such as svFv fragments, which further enhances the use of this technology (not shown). Possibly, the most variable factor in the technology is the quality of the library DNA itself. Sequence analysis of cloned unselected library showed that, with different batches of library encoding oligonucleotides, the level of deletions and insertions varied remarkably, e.g., one batch contained 75% deletions and insertions, whereas another contained none. Many further developments of the system can be envisaged. For example, C-terminal display may be possible, but we have avoided testing this at present because of the possibility of disrupting the cis-activity of RepA. In addition, we have found it difficult to fully quantify functional library size, as it seems to depend on the lysate activity, DNA quality, and the method for measuring protein–DNA linkage.

CIS display combines the robustness and ease of use of phage display with the added power and flexibility of *in vitro* operation. Furthermore, the ease with which the protein–DNA complexes are produced and selections performed should make CIS display well suited to automation, allowing the parallel selection of high-affinity, fine-specificity ligands to multiple targets.

We thank Dr. David Andrews (McMaster University, Hamilton, Canada) and his research group for provision of S30 lysates and Dr. Andrew Carter at Institute of Food Research, Norwich, for his assistance.

1. Smith, G. P. (1985) *Science* **228,** 1315–1317.
2. Scott, J. K. & Smith, G. P. (1990) *Science* **249,** 386–390.
3. Cwirla, S. E., Peters, E. A., Barrett, R. W. & Dower, W. J. (1990) *Proc. Natl. Acad. Sci. USA* **87,** 6378–6382.
4. Devlin, J. J., Panganiban, L. C. & Devlin, P. E. (1990) *Science* **249,** 404–406.
5. Griffiths, A. D., Williams, S. C, Hartley, O., Tomlinson, I. M., Waterhouse, P., Crosby, W. L., Kontermann, R. E., Jones, P. T., Low, N. M., Allison, T. J., *et al.* (1994) *EMBO J.* **13,** 3245–3260.
6. Sidhu, S. S., Lowman, H. B., Cunningham, B. C. & Wells, J. A. (2000) *Methods Enzymol.* **328,** 333–363.
7. Fuchs, P., Breitling, F., Dubel, S., Seehaus, T. & Little, M. (1991) *Bio/Technology* **9,** 1369–1372.
8. Boder, E. T. & Wittrup, K. D. (1997) *Nat. Biotechnol.* **15,** 553–557.
9. Cull, M. G., Miller, J. F. & Schatz, P. J. (1992) *Proc. Natl. Acad. Sci. USA* **89,** 1865–1869.
10. Perelson, A. S. & Oster, G. F. (1979) *J. Theor. Biol.* **81,** 645–670.
11. Nemoto, N., Miyamoto-Sato, E., Husimi, Y. & Yanagawa, H. (1997) *FEBS Lett.* **414,** 405–408.
12. Roberts, R. W. & Szostak, J. W. (1997) *Proc. Natl. Acad. Sci. USA* **94,** 12297–12302.
13. Keefe, A. D. & Szostak, J. W. (2001) *Nature* **410,** 715–718.
14. Kreider, B. L. (2000) *Med. Res. Rev.* **20,** 212–215.
15. Palacios, R. & Schimke, R. T. (1973) *J. Biol. Chem.* **248,** 1424–1430.
16. Mattheakis, L. C., Bhatt, R. R. & Dower W. J. (1994) *Proc. Natl. Acad. Sci. USA* **91,** 9022–9026.
17. Gersuk, G. M., Corey, M. J., Corey, E., Stray, J. E., Kawasaki, G. H. & Vessella, R. L. (1997) *Biochem. Biophys. Res. Commun.* **232,** 578–582.
18. Amstutz, P., Forrer, P., Zahnd, C. & Pluckthun, A. (2001) *Curr. Opin. Biotechnol.* **12,** 400–405.
19. Hanes, J. & Pluckthun, A. (1997) *Proc. Natl. Acad. Sci. USA.* **94,** 4937–4942.
20. Sepp, A., Tawfik, D. S. & Griffiths, A. D. (2002) *FEBS Lett.* **532,** 455–458.
21. Yonezawa, M., Doi, N., Kawahashi, Y., Higashinakagawa, T. & Yanagawa, H. (2003) *Nucleic Acids Res.* **31,** e118.
22. Masai, H., Kaziro, Y. & Arai, K. (1983) *Proc. Natl. Acad. Sci. USA* **80,** 6814–6818.
23. Nikoletti, S., Bird, P., Praszkier, J. & Pittard, J. (1988) *J. Bacteriol.* **170,** 1311–1318.
24. FitzGerald, K. (2000) *Drug Discovery Today* **5,** 253–258.
25. Odegrip, R. O. & Haggard-Ljungquist, E. (2001) *J. Mol. Biol.* **308,** 147–163.
26. Giraldo, R. & Diaz, R. (1992) *J. Mol. Biol.* **228,** 787–802.
27. Masai, H. & Arai, K. (1988) *Nucleic Acids Res.* **16,** 6493–6514.
28. Praszkier, J. & Pittard, A. J. (1999) *J. Bacteriol.* **181,** 2765–2772.
29. Praszkier, J., Murthy, S. & Pittard, A. J. (2000) *J. Bacteriol.* **182,** 3972–3980.
30. McGregor, D. P. & Robins, S. P. (2001) *Anal. Biochem.* **294,** 108–117.
31. Southern, J. A., Young, D. F., Heaney, F., Baumgartner, W. K. & Randall, R. E. (1991) *J. Gen. Virol.* **72,** 1551–1557.
32. Lesley, S. A., Brow, M. A. D. & Burgess, R. R. (1991) *J. Biol. Chem.* **266,** 2632–2638.
33. Lesley, S. A. (1995) *Methods Mol. Biol.* **37,** 265–278.
34. Stephen, C. W., Helminen, P. & Lane, D. P. (1995) *J. Mol. Biol.* **248,** 58–78.
35. Slootstra, J. W., Kuperus, D., Plückthun, A. & Meloen, R. H. (1997) *Mol. Divers.* **2,** 156–164.
36. Wegner, G. J., Lee, H. J. & Corn, R. M. (2002) *Anal. Chem.* **74,** 5161–5168.
37. Baggio, R., Burgstaller, P., Hale, S. P., Putney, A. R., Lane, M., Lipovsek, D., Wright, M. C., Roberts, R. W., Liu, R., Szostak, J. W. & Wagner, R. W. (2002) *J. Mol. Recognit.* **15,** 126–134.
38. Holtet, T. & Osbourn, J. (2001) Int. Patent Appl. WO0175097.
39. Kurz, M., Gu, K., Al-Gawari, A. & Lohse, P. A. (2001) *Chembiochem* **2,** 666–672.

Alamar Exhibit 1002
Alamar v. Olink

IFW
Bof-Seq

OIPE IAP55
SEP 25 2006

UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NUMBER	FILING OR 371 (c) DATE	FIRST NAMED APPLICANT	ATTORNEY DOCKET NUMBER
11/483,825	07/10/2006	Olof Ericsson	22639-3

CONFIRMATION NO. 8659

24256
DINSMORE & SHOHL, LLP
1900 CHEMED CENTER
255 EAST FIFTH STREET
CINCINNATI, OH 45202

FORMALITIES LETTER

Date Mailed: 07/28/2006

NOTICE TO COMPLY WITH REQUIREMENTS FOR PATENT APPLICATIONS CONTAINING NUCLEOTIDE SEQUENCE AND/OR AMINO ACID SEQUENCE DISCLOSURES

Filing Date Granted

Applicant is given **TWO MONTHS FROM THE DATE OF THIS NOTICE** within which to file the items indicated below to avoid abandonment. Extensions of time may be obtained under the provisions of 37 CFR 1.136(a).

- A copy of the "Sequence Listing" in computer readable form has been submitted. However, the content of the computer readable form does not comply with the requirements of 37 CFR 1.822 and/or 1.823, as indicated on the attached copy of the marked -up "Raw Sequence Listing." Applicant must provide a substitute computer readable form (CRF) copy of the "Sequence Listing" and a statement that the content of the sequence listing information recorded in computer readable form is identical to the written (on paper or compact disc) sequence listing and, where applicable, includes no new matter, as required by 37 CFR 1.821(e), 1.821(f), 1.821(g), 1.825(b), or 1.825(d).

  Applicant is cautioned that correction of the above items may cause the specification and drawings page count to exceed 100 pages. If the specification and drawings exceed 100 pages, applicant will need to submit the required application size fee.

To Download Patentin Software, visit http://www.uspto.gov/web/patents/software.htm
For questions regarding compliance to these requirements, please contact:

- **For Rules Interpretation, call (571) 272-0951**
- **For Patentin Software Program Help, call Patent EBC at 1-866-217-9197 or directly at 703-305-3028 / 703-308-6845 between the hours of 6 a.m. and 12 midnight, Monday through Friday, EST.**
- **Send e-mail correspondence for Patentin Software Program Help @ ebc@uspto.gov**

Replies should be mailed to: Mail Stop Missing Parts
Commissioner for Patents
P.O. Box 1450
Alexandria VA 22313-1450

Alamar Exhibit 1002
Alamar v. Olink

A copy of this notice MUST be returned with the reply.

Office of Initial Patent Examination (571) 272-4000, or 1-800-PTO-9199, or 1-800-972-6382
PART 2 - COPY TO BE RETURNED WITH RESPONSE

Alamar Exhibit 1002
Alamar v. Olink

Docket No: 22639-3 PATENT

CERTIFICATE OF MAILING

I hereby certify that this paper is being deposited with the United States Postal Service with sufficient postage as first class mail in an envelope addressed to: Mail Stop Missing Parts, Commissioner for Patents; P.O. Box 1450; Alexandria, VA 22313-1450 on September 22, 2006.

Marlene Troxle

SEP 25 2006 PATENT & TRADEMARK OFFICE

IN THE UNITED STATES PATENT & TRADEMARK OFFICE

Applicant:	Olof Ericsson	:	Paper No.:
Serial No.:	11/483,825	:	Group Art Unit: tbd
Filing Date:	July 10, 2006	:	Examiner: tbd

For: **Regulation Analysis by Cis Reactivity, RACR**

STATEMENT REGARDING SEQUENCE LISTING IDENTITY

Mail Stop Missing Parts
Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

Dear Sir:

Submitted herewith is a substitute computer readable form (CRF) copy of the "Sequence Listing" and a substitute paper copy as required by 37 C.F.R. 1.821(e). The Sequence Listing has been revised to correct the <220> entries related to the respective sources of the genetic material in light of the <213> "organism" responses of "Artificial Sequence," throughout the Sequence Listing. Applicants state that the content of the substitute computer readable form copy is identical to the substitute paper copy of the Sequence Listing submitted herewith and includes no new matter, in accordance with 37 C.F.R. 1.821(e), (f) or (g), or 1.825(b) or (d).

Respectfully submitted,

Denise M. Everett, Reg. No. 47,552
Attorney for Applicants
Dinsmore & Shohl LLP
1900 Chemed Center
255 East Fifth Street
Cincinnati, Ohio 45202

Alamar Exhibit 1002
Alamar v. Olink

OIPE IAP55
SEP 25 2006
PATENT & TRADEMARK OFFICE

SEQUENCE LISTING

<110> Ericsson, Olof

<120> Regulation Analysis by Cis Reactivity, RACR

<130> 22639-3

<140> US 11/483,825
<141> 2006-07-10

<150> US 60/697,415
<151> 2005-07-08

<160> 30

<170> PatentIn version 3.3

<210> 1
<211> 50
<212> DNA
<213> artificial

<220>
<223> chemically synthesized

<400> 1
tctctctctc tctctctctc gatgaatgtc gtcgcaacac cgtccatacg 50

<210> 2
<211> 50
<212> DNA
<213> artificial

<220>
<223> chemically synthesized

<400> 2
tctctctctc tctctctctc gatgaatgtc gtcgcgtatc acgttcgtcg 50

<210> 3
<211> 50
<212> DNA
<213> artificial

<220>
<223> chemically synthesized

<400> 3
tctctctctc tctctctctc gatgaatgtc gtcgcctccg taatcgtggt 50

<210> 4
<211> 50
<212> DNA
<213> artificial

<220>
<223> chemically synthesized

<400> 4

Alamar Exhibit 1002
Alamar v. Olink

tctctctctc tctctctctc gatgaatgtc gtcgcaaccg tatcctcacg 50

<210> 5
<211> 50
<212> DNA
<213> artificial

<220>
<223> chemically synthesized

<400> 5
tctctctctc tctctctctc gatgaatgtc gtcgcttacc ctctttcggc 50

<210> 6
<211> 50
<212> DNA
<213> artificial

<220>
<223> chemically synthesized

<400> 6
tctctctctc tctctctctc ggcgcacatg gtatagtatc acgttcgtcg 50

<210> 7
<211> 50
<212> DNA
<213> artificial

<220>
<223> chemically synthesized

<400> 7
tctctctctc tctctctctc ggcgcacatg gtatactccg taatcgtggt 50

<210> 8
<211> 50
<212> DNA
<213> artificial

<220>
<223> chemically synthesized

<400> 8
tctctctctc tctctctctc ggcgcacatg gtataaaccg tatcctcacg 50

<210> 9
<211> 50
<212> DNA
<213> artificial

<220>
<223> chemically synthesized

<400> 9
tctctctctc tctctctctc ggcgcacatg gtatattacc ctctttcggc 50

Alamar Exhibit 1002
Alamar v. Olink

<210> 10
<211> 50
<212> DNA
<213> artificial

<220>
<223> chemically synthesized

<400> 10
tctctctctc tctctctctc tgtgcgactt tctcactccg taatcgtggt 50

<210> 11
<211> 50
<212> DNA
<213> artificial

<220>
<223> chemically synthesized

<400> 11
tctctctctc tctctctctc tgtgcgactt tctcaaaccg tatcctcacg 50

<210> 12
<211> 50
<212> DNA
<213> artificial

<220>
<223> chemically synthesized

<400> 12
tctctctctc tctctctctc tgtgcgactt tctcattacc ctctttcggc 50

<210> 13
<211> 50
<212> DNA
<213> artificial

<220>
<223> chemically synthesized

<400> 13
tctctctctc tctctctctc cgacctaatg ctacgaaccg tatcctcacg 50

<210> 14
<211> 50
<212> DNA
<213> artificial

<220>
<223> chemically synthesized

<400> 14
tctctctctc tctctctctc cgacctaatg ctacgttacc ctctttcggc 50

<210> 15
<211> 50
<212> DNA

Alamar Exhibit 1002
Alamar v. Olink

<213> artificial

<220>
<223> chemically synthesized

<400> 15
tctctctctc tctctctctc gcgcctactg tgtatttacc ctctttcggc 50

<210> 16
<211> 63
<212> DNA
<213> artificial

<220>
<223> chemically synthesized

<400> 16
aagatagcgt acagtcggca aaagagtcga gagctccgta atcgtggtaa aaagtatctg 60

ctt 63

<210> 17
<211> 63
<212> DNA
<213> artificial

<220>
<223> chemically synthesized

<400> 17
aagatagcgt acagtcggca aaagagtcga gagaaccgta tcctcacgaa aaagtatctg 60

ctt 63

<210> 18
<211> 63
<212> DNA
<213> artificial

<220>
<223> chemically synthesized

<400> 18
atgtcgcccg aaaaagatga atgtcgtcgc gagaggactc aaaacaacta ggtacaggta 60

tcg 63

<210> 19
<211> 63
<212> DNA
<213> artificial

<220>
<223> chemically synthesized

<400> 19
atgtcgcccg aaaaaggcgc acatggtata gagaggactc aaaacaacta ggtacaggta 60

tcg 63

<210> 20
<211> 83
<212> DNA
<213> artificial

<220>
<223> chemically synthesized

<400> 20
gtactggctg ataatcttcc aagatagcgt acagtcggca aaagagtcga gagaacaccg 60

tccatacgaa aaagtatctg ctt 83

<210> 21
<211> 83
<212> DNA
<213> artificial

<220>
<223> chemically synthesized

<400> 21
cccgagcagc cggagcagag aagatagcgt acagtcggca aaagagtcga gaggtatcac 60

gttcgtcgaa aaagtatctg ctt 83

<210> 22
<211> 83
<212> DNA
<213> artificial

<220>
<223> chemically synthesized

<400> 22
gtcctgtgtc tgaacctgag aagatagcgt acagtcggca aaagagtcga gagttcccct 60

ctttcggcaa aaagtatctg ctt 83

<210> 23
<211> 83
<212> DNA
<213> artificial

<220>
<223> chemically synthesized

<400> 23
atgtcgcccg aaaaatgtgc gactttctca gagaggactc aaaacaacta ggtacaggta 60

tcgtgtcgac gaggatgaag agg 83

<210> 24
<211> 83
<212> DNA
<213> artificial

Alamar Exhibit 1002
Alamar v. Olink

<220>
<223> chemically synthesized

<400> 24
atgtcgcccg aaaaacgacc taatgctacg gagaggactc aaaacaacta ggtacaggta 60

tcgctgtggc atgtttgtct aca 83

<210> 25
<211> 83
<212> DNA
<213> artificial

<220>
<223> chemically synthesized

<400> 25
atgtcgcccg aaaaagcgcc tactgtgtat gagaggactc aaaacaacta ggtacaggta 60

tcggtattgt gatttttttaa aag 83

<210> 26
<211> 20
<212> DNA
<213> artificial

<220>
<223> chemically synthesized

<400> 26
tagttgtttt gagtcctctc 20

<210> 27
<211> 19
<212> DNA
<213> artificial

<220>
<223> chemically synthesized

<400> 27
tcggcaaaag agtcgagag 19

<210> 28
<211> 19
<212> DNA
<213> artificial

<220>
<223> chemically synthesized

<400> 28
gagaggactc aaacaacta 19

<210> 29
<211> 29
<212> DNA
<213> artificial

<220>
<223> chemically synthesized

<400> 29
aaaaacgggc gacataagca gatacaaaa 29

<210> 30
<211> 20
<212> DNA
<213> artificial

<220>
<223> chemically synthesized

<400> 30
cgggcgacat aagcagatac 20

Alamar Exhibit 1002
Alamar v. Olink