REMARKS

I. Status as Continuation

The present application is being filed as a continuation of U.S. Application No. 16/743,604, which is currently pending. Therefore, the present application is co-pending with a prior-filed nonprovisional patent application in fulfillment of the requirements of 35 U.S.C. §120.

II. Amendments to the Specification

By this preliminary amendment, Applicant has amended the specification to recite the continuing data for the present application and to update the sequence listing name. Applicant submits that these amendments to the specification introduce no new matter.

III. Sequence Listing

Applicant is filing herewith a Sequence Listing in computer readable form in the ASCII text file named: NUGN-001-04US-SeqList-ST25.txt. The Sequence Listing is 3 kilobytes in size, was created on July 21, 2021 in the IBM-PC machine format, and having an operating system compatibility with MS-Windows.

The sequence listing has been updated relative to that filed in U.S. 16/743,604 to include all priority application information for this continuation. Applicant is also filing herewith a Statement of Compliance pursuant to 37 C.F.R. §1.825 indicating that the Sequence Listing contains no new matter. *See, e.g., MPEP (9th ed. Rev. 08.2017), §608.05.*

In accordance with MPEP §2422.03(a), the ASCII text file of the Sequence Listing will serve as both the paper copy and the computer readable form (CRF) of the Sequence Listing. Accordingly, Applicant is not submitting herewith: (1) a second copy of the sequence listing as a PDF file; (2) a statement under 37 C.F.R. §1.821(f) indicating that the paper copy and CRF of the sequence listing are identical; or (3) a request to use a compliant computer readable form of

IDT Exhibit 1020.0001

the sequence listing that is already on file for another application pursuant to 37 C.F.R. §1.821(e).

If there are any questions regarding these remarks, the Examiners are invited and encouraged to contact Applicant's representatives at the telephone number provided.

Dated: July 21, 2021

Respectfully submitted,
BROWN RUDNICK LLP

BROWN RUDNICK LLP
One Financial Center
Boston, MA 02111
Tel: 617-856-8483

/Thomas C. Meyers/
Thomas C. Meyers, Reg. No. 36,989
Attorney for Applicant
Email: tmeyers@brownrudnick.com

IDT Exhibit 1020.0002

AMENDMENTS TO THE SPECIFICATION

Please amend the "CROSS-REFERENCE" section on page 1, paragraph [0001] of the as-filed specification as follows:

CROSS-REFERENCE

This application is a continuation of U.S. Application Serial No. 16/743,604, filed January 15, 2020, which is a continuation of U.S. Application Serial No. 15/406,002, filed January 13, 2017, which is a divisional of U.S. Serial No. 14/540,917, filed November 13, 2014, which claims the benefit of U.S. Provisional Application No. 61/903,826, filed November 13, 2013, which [[is]] are each herein incorporated by reference in [[its]] their entirety.

Please amend the "DESCRIPTION OF THE TEXT FILE SUBMITTED ELECTRONICALLY" section on page 1, paragraph [0002] of the as-filed specification as follows:

DESCRIPTION OF THE TEXT FILE SUBMITTED ELECTRONICALLY

[0002] The contents of the text file submitted electronically herewith are incorporated herein by reference in their entirety: A computer readable format copy of the Sequence Listing (filename: NUGN-001-04US-SeqList-ST25.txt ~~NUGN_001_01US_SeqList_ST25.txt~~, date recorded: July 21, 2021 ~~November 12, 2014~~, file size 3 kilobytes).

IDT Exhibit 1020.0003

PTO/AIA/01 (06-12)
Approved for use through 01/31/2014. 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.

DECLARATION (37 CFR 1.63) FOR UTILITY OR DESIGN APPLICATION USING AN APPLICATION DATA SHEET (37 CFR 1.76)

Title of Invention	COMPOSITIONS AND METHODS FOR IDENTIFICATION OF A DUPLICATE SEQUENCING READ

As the below named inventor, I hereby declare that:

This declaration is directed to:

☐ The attached application, or

☒ United States application or PCT international application number 14/540,917

filed on 11/13/2014.

The above-identified application was made or authorized to be made by me.

I believe that I am the original inventor or an original joint inventor of a claimed invention in the application.

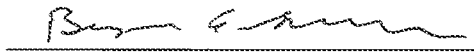
I hereby acknowledge that any willful false statement made in this declaration is punishable under 18 U.S.C. 1001 by fine or imprisonment of not more than five (5) years, or both.

WARNING:

Petitioner/applicant is cautioned to avoid submitting personal information in documents filed in a patent application that may contribute to identity theft. Personal information such as social security numbers, bank account numbers, or credit card numbers (other than a check or credit card authorization form PTO-2038 submitted for payment purposes) is never required by the USPTO to support a petition or an application. If this type of personal information is included in documents submitted to the USPTO, petitioners/applicants should consider redacting such personal information from the documents before submitting them to the USPTO. Petitioner/applicant is advised that the record of a patent application is available to the public after publication of the application (unless a non-publication request in compliance with 37 CFR 1.213(a) is made in the application) or issuance of a patent. Furthermore, the record from an abandoned application may also be available to the public if the application is referenced in a published application or an issued patent (see 37 CFR 1.14). Checks and credit card authorization forms PTO-2038 submitted for payment purposes are not retained in the application file and therefore are not publicly available.

LEGAL NAME OF INVENTOR

Inventor: Ben Schroeder Date (Optional) : _____

Signature: 

Note: An application data sheet (PTO/SB/14 or equivalent), including naming the entire inventive entity, must accompany this form or must have been previously filed. Use an additional PTO/AIA/01 form for each additional inventor.

This collection of information is required by 35 U.S.C. 115 and 37 CFR 1.63. 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 1.11 and 1.14. This collection is estimated to take 1 minute 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.

American LegalNet, Inc.
www.FormsWorkFlow.com

PTO/AIA/01 (06-12)
Approved for use through 01/31/2014. 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.

DECLARATION (37 CFR 1.63) FOR UTILITY OR DESIGN APPLICATION USING AN APPLICATION DATA SHEET (37 CFR 1.76)

Title of Invention	COMPOSITIONS AND METHODS FOR IDENTIFICATION OF A DUPLICATE SEQUENCING READ

As the below named inventor, I hereby declare that:

This declaration is directed to:

☐ The attached application, or

☒ United States application or PCT international application number 14/540,917

filed on 11/13/2014.

The above-identified application was made or authorized to be made by me.

I believe that I am the original inventor or an original joint inventor of a claimed invention in the application.

I hereby acknowledge that any willful false statement made in this declaration is punishable under 18 U.S.C. 1001 by fine or imprisonment of not more than five (5) years, or both.

WARNING:

Petitioner/applicant is cautioned to avoid submitting personal information in documents filed in a patent application that may contribute to identity theft. Personal information such as social security numbers, bank account numbers, or credit card numbers (other than a check or credit card authorization form PTO-2038 submitted for payment purposes) is never required by the USPTO to support a petition or an application. If this type of personal information is included in documents submitted to the USPTO, petitioners/applicants should consider redacting such personal information from the documents before submitting them to the USPTO. Petitioner/applicant is advised that the record of a patent application is available to the public after publication of the application (unless a non-publication request in compliance with 37 CFR 1.213(a) is made in the application) or issuance of a patent. Furthermore, the record from an abandoned application may also be available to the public if the application is referenced in a published application or an issued patent (see 37 CFR 1.14). Checks and credit card authorization forms PTO-2038 submitted for payment purposes are not retained in the application file and therefore are not publicly available.

LEGAL NAME OF INVENTOR

Inventor: Douglas Amorese Date (Optional) :

Signature:

Note: An application data sheet (PTO/SB/14 or equivalent), including naming the entire inventive entity, must accompany this form or must have been previously filed. Use an additional PTO/AIA/01 form for each additional inventor.

This collection of information is required by 35 U.S.C. 115 and 37 CFR 1.63. 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 1.11 and 1.14. This collection is estimated to take 1 minute 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.

American LegalNet, Inc.
www.FormsWorkFlow.com

PTO/AIA/01 (06-12)
Approved for use through 01/31/2014. 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.

DECLARATION (37 CFR 1.63) FOR UTILITY OR DESIGN APPLICATION USING AN APPLICATION DATA SHEET (37 CFR 1.76)

Title of Invention	COMPOSITIONS AND METHODS FOR IDENTIFICATION OF A DUPLICATE SEQUENCING READ

As the below named inventor, I hereby declare that:

This declaration is directed to:

☐ The attached application, or

☒ United States application or PCT international application number 14/540,917

filed on 11/13/2014.

The above-identified application was made or authorized to be made by me.

I believe that I am the original inventor or an original joint inventor of a claimed invention in the application.

I hereby acknowledge that any willful false statement made in this declaration is punishable under 18 U.S.C. 1001 by fine or imprisonment of not more than five (5) years, or both.

WARNING:

Petitioner/applicant is cautioned to avoid submitting personal information in documents filed in a patent application that may contribute to identity theft. Personal information such as social security numbers, bank account numbers, or credit card numbers (other than a check or credit card authorization form PTO-2038 submitted for payment purposes) is never required by the USPTO to support a petition or an application. If this type of personal information is included in documents submitted to the USPTO, petitioners/applicants should consider redacting such personal information from the documents before submitting them to the USPTO. Petitioner/applicant is advised that the record of a patent application is available to the public after publication of the application (unless a non-publication request in compliance with 37 CFR 1.213(a) is made in the application) or issuance of a patent. Furthermore, the record from an abandoned application may also be available to the public if the application is referenced in a published application or an issued patent (see 37 CFR 1.14). Checks and credit card authorization forms PTO-2038 submitted for payment purposes are not retained in the application file and therefore are not publicly available.

LEGAL NAME OF INVENTOR

Inventor: Jonathan Scolnick Date (Optional) : _____

Signature: _____

Note: An application data sheet (PTO/SB/14 or equivalent), including naming the entire inventive entity, must accompany this form or must have been previously filed. Use an additional PTO/AIA/01 form for each additional inventor.

This collection of information is required by 35 U.S.C. 115 and 37 CFR 1.63. 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 1.11 and 1.14. This collection is estimated to take 1 minute 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.

American LegalNet, Inc.
www.FormsWorkFlow.com

Electronic Patent Application Fee Transmittal

Application Number:	
Filing Date:	
Title of Invention:	COMPOSITIONS AND METHODS FOR IDENTIFICATION OF A DUPLICATE SEQUENCING READ
First Named Inventor/Applicant Name:	Douglas A. Amorese
Filer:	Thomas C. Meyers/Sandra Richards
Attorney Docket Number:	NUGN-001/04US 34079/550

Filed as Large Entity

Filing Fees for Utility under 35 USC 111(a)

Description	Fee Code	Quantity	Amount	Sub-Total in USD($)
Basic Filing:				
UTILITY APPLICATION FILING	1011	1	320	320
UTILITY SEARCH FEE	1111	1	700	700
UTILITY EXAMINATION FEE	1311	1	800	800
Pages:				
Claims:				
CLAIMS IN EXCESS OF 20	1202	1	100	100
Miscellaneous-Filing:				
Petition:				

Description	Fee Code	Quantity	Amount	Sub-Total in USD($)
Patent-Appeals-and-Interference:				
Post-Allowance-and-Post-Issuance:				
Extension-of-Time:				
Miscellaneous:				
		Total in USD ($)		**1920**

CLAIMS

1. A method for detecting a duplicate sequencing read from a population of sample sequencing reads comprising:
    a) ligating an adaptor to a 5' end of each nucleic acid fragment of a plurality of nucleic acid fragments from one or more samples, wherein the adaptor comprises:
        (i) an indexing primer binding site;
        (ii) an indexing site;
        (iii) an identifier site; and
        (iv) a target sequence primer binding site;
    b) amplifying the adapter-nucleic acid fragment ligated products;
    c) generating a population of sequencing reads from amplified adapter-nucleic acid fragment ligated products; and
    d) detecting the population of sequencing reads comprising a sequencing read with a duplicate identifier site and target sequence.
2. The method of claim 1, wherein the method further comprises removing from the population of sequence reads the sequencing read with a duplicate identifier site and target sequence.
3. The method of claim 1, wherein the identifier site is sequenced with the indexing site.
4. The method of claim 1, wherein the identifier site is sequenced separately from the indexing site.
5. The method of claim 1, wherein the identifier site is sequenced with the target sequence.
6. The method of claim 1, wherein the identifier site is sequenced separately from the target sequence.
7. The method of claim 1, wherein the adaptor comprises from 5' to 3':

(i) the indexing primer binding site;

(ii) the indexing site;

(iii) the identifier site; and

(iv) the target sequence primer binding site.

8. The method of claim 1, wherein the adaptor comprises from 5' to 3':

(i) the indexing primer binding site;

(ii) the indexing site;

(iii) the target sequence primer binding site; and

(iv) the identifier site.

9. The method of claim 1, wherein the plurality of nucleic acid fragments is generated from more than one sample.

10. The method of claim 9, wherein the nucleic acid fragments from each sample has the same indexing site.

11. The method of claim 10, wherein the sequencing reads are separated based on the indexing site.

12. The method of claim 11, wherein the separation of sequencing reads is performed prior to step d).

13. The method of claim 1, wherein the nucleic acid fragments are DNA fragments, RNA fragments, or DNA/RNA fragments.

14. The method of claim 13, wherein the nucleic acid fragments are genomic DNA fragments or cDNA fragments.

15. The method of claim 1, wherein the indexing site is between 2 and 8 nucleotides in length.

16. The method of claim 1, wherein the indexing site is about 6 nucleotides in length.

17. The method of claim 1, wherein the identifier site is between 1 and 8 nucleotides in length.

18. The method of claim 1, wherein the identifier site is about 8 nucleotides in length.

19. The method of claim 1, wherein the indexing primer binding site is a universal indexing primer binding site.

20. The method of claim 1, wherein the target sequence primer binding site is a universal target sequence primer binding site.

21. A kit comprising a plurality of adaptors, wherein each adaptor comprises:

    (i) an indexing primer binding site;

    (ii) an indexing site;

    (iii) an identifier site; and

    (iv) a target sequencing primer binding site.

Electronic Acknowledgement Receipt

EFS ID:	43309891
Application Number:	17381888
International Application Number:	
Confirmation Number:	6188
Title of Invention:	COMPOSITIONS AND METHODS FOR IDENTIFICATION OF A DUPLICATE SEQUENCING READ
First Named Inventor/Applicant Name:	Douglas A. Amorese
Customer Number:	21710
Filer:	Thomas C. Meyers/Sandra Richards
Filer Authorized By:	Thomas C. Meyers
Attorney Docket Number:	NUGN-001/04US 34079/550
Receipt Date:	21-JUL-2021
Filing Date:	
Time Stamp:	16:13:59
Application Type:	Utility under 35 USC 111(a)

Payment information:

Submitted with Payment	yes
Payment Type	DA
Payment was successfully received in RAM	$1920
RAM confirmation Number	E20217KG14421422
Deposit Account	
Authorized User	
The Director of the USPTO is hereby authorized to charge indicated fees and credit any overpayment as follows:	

File Listing:

Document Number	Document Description	File Name	File Size(Bytes)/ Message Digest	Multi Part /.zip	Pages (if appl.)
1	Sequence Listing (Text File)	NUGN-001-04US-SeqList_ST25.txt	2694	no	-
Warnings:					
Information:					
2	Application Data Sheet	34079_550US_-_Application_Data_Sheet_Fillable_PDF.pdf	1255992 a33f96aa1971c93870efc1ff716f7b413f9459db	no	9
Warnings:					
Information:					
3		NUGN-001-04US-Application.pdf	1878669 a3df885a2d7489a1ca35c04d2a08fe06fafd129a	yes	44

Multipart Description/PDF files in .zip description		
Document Description	**Start**	**End**
Specification	1	40
Claims	41	43
Abstract	44	44

Warnings:					
Information:					
4	Drawings-only black and white line drawings	NUGN-001-04US--Drawings.pdf	4460762 af3766a1bec892c1cef5832db9f5abd0d2b2f43e	no	6
Warnings:					
Information:					
5	Oath or Declaration filed	NUGN-001-01US-Declarations.pdf	339248 f4554724 6f2e9d79527b94c819fc8943da734284	no	3

Warnings:					
Information:					
6		NUGN-001-04US-Prelim-Amnd.pdf	155717 8612268355598 68d8b4b475bf6cb9bf098268833	yes	4

	Multipart Description/PDF files in .zip description		
	Document Description	**Start**	**End**
	Preliminary Amendment	1	1
	Specification	2	2
	Applicant Arguments/Remarks Made in an Amendment	3	4

Warnings:					
Information:					
7	Sequence Listing	NUGN-001-04US-Certification.pdf	113150 79743205b8e2c0d5dbf6f7963a54961613ffe f1b2	no	1
Warnings:					
Information:					
8	Fee Worksheet (SB06)	fee-info.pdf	36783 d7c6a245f2c7f8dc706e21151f2d5364708b9f54	no	2
Warnings:					
Information:					
		Total Files Size (in bytes):	8243015		

This Acknowledgement Receipt evidences receipt on the noted date by the USPTO of the indicated documents, characterized by the applicant, and including page counts, where applicable. It serves as evidence of receipt similar to a Post Card, as described in MPEP 503.

New Applications Under 35 U.S.C. 111

If a new application is being filed and the application includes the necessary components for a filing date (see 37 CFR 1.53(b)-(d) and MPEP 506), a Filing Receipt (37 CFR 1.54) will be issued in due course and the date shown on this Acknowledgement Receipt will establish the filing date of the application.

National Stage of an International Application under 35 U.S.C. 371

If a timely submission to enter the national stage of an international application is compliant with the conditions of 35 U.S.C. 371 and other applicable requirements a Form PCT/DO/EO/903 indicating acceptance of the application as a national stage submission under 35 U.S.C. 371 will be issued in addition to the Filing Receipt, in due course.

New International Application Filed with the USPTO as a Receiving Office

If a new international application is being filed and the international application includes the necessary components for an international filing date (see PCT Article 11 and MPEP 1810), a Notification of the International Application Number and of the International Filing Date (Form PCT/RO/105) will be issued in due course, subject to prescriptions concerning national security, and the date shown on this Acknowledgement Receipt will establish the international filing date of the application.

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

APPLICANT:	Tecan Genomics, Inc.	ART UNIT:	Not Yet Assigned
SERIAL NUMBER:	Continuation of US Serial No. 16/743,604	CONF. NO.:	Not Yet Assigned
FILING DATE:	Herewith	EXAMINER:	Not Yet Assigned
TITLE:	COMPOSITIONS AND METHODS FOR IDENTIFICATION OF A DUPLICATE SEQUENCING READ		

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

FILED ELECTRONICALLY

PRELIMINARY AMENDMENT

Prior to examination of the above U.S. nonprovisional patent application, please amend the claims and the specification and consider the following remarks. Applicants believe that no fees are due with respect to this Preliminary Amendment.

Amendments to the specification begin on page 2.

A **Sequence Listing** is referenced on page 3.

Remarks begin on page 3.

Applicant believes that no fees are due in connection with this Preliminary Amendment. However, Applicant authorizes any required fees to be charged to Deposit Account No. 500369.

PTO/AIA/14 (02-18)
Approved for use through 11/30/2020. 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 contains a valid OMB control number.

Application Data Sheet 37 CFR 1.76	Attorney Docket Number	NUGN-001/04US 34079/550
	Application Number	
Title of Invention	COMPOSITIONS AND METHODS FOR IDENTIFICATION OF A DUPLICATE SEQUENCING READ	

The application data sheet is part of the provisional or nonprovisional application for which it is being submitted. The following form contains the bibliographic data arranged in a format specified by the United States Patent and Trademark Office as outlined in 37 CFR 1.76.
This document may be completed electronically and submitted to the Office in electronic format using the Electronic Filing System (EFS) or the document may be printed and included in a paper filed application.

Secrecy Order 37 CFR 5.2:

☐ Portions or all of the application associated with this Application Data Sheet may fall under a Secrecy Order pursuant to 37 CFR 5.2 (Paper filers only. Applications that fall under Secrecy Order may not be filed electronically.)

Inventor Information:

Inventor 1 Remove

Legal Name

Prefix	Given Name	Middle Name	Family Name	Suffix
	Douglas	A.	Amorese	

Residence Information (Select One) ● US Residency Non US Residency Active US Military Service

City	State/Province	Country of Residence
Los Altos	CA	US

Mailing Address of Inventor:

Address 1	900 Chesapeake Drive
Address 2	
City	Redwood City
State/Province	CA
Postal Code	94063
Country	US

Inventor 2 Remove

Legal Name

Prefix	Given Name	Middle Name	Family Name	Suffix
	Jonathan		Scolnick	

Residence Information (Select One) ◉ US Residency Non US Residency Active US Military Service

City	State/Province	Country of Residence
San Francisco	CA	US

Mailing Address of Inventor:

Address 1	451 Kansas Street #518
Address 2	
City	San Francisco
State/Province	CA
Postal Code	94107
Country	US

Inventor 3 Remove

Legal Name

Prefix	Given Name	Middle Name	Family Name	Suffix
	Ben		Schroeder	

Residence Information (Select One) ◉ US Residency Non US Residency Active US Military Service

IDT Exhibit 1020.0017

PTO/AIA/14 (02-18)
Approved for use through 11/30/2020. 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 contains a valid OMB control number.

Application Data Sheet 37 CFR 1.76	Attorney Docket Number	NUGN-001/04US 34079/550
	Application Number	
Title of Invention	COMPOSITIONS AND METHODS FOR IDENTIFICATION OF A DUPLICATE SEQUENCING READ	

City	San Mateo	State/Province	CA	Country of Residence	US

Mailing Address of Inventor:

Address 1	33 Hayward Ave, Apt 112		
Address 2			
City	San Mateo	**State/Province**	CA
Postal Code	94401	**Country**	US

All Inventors Must Be Listed - Additional Inventor Information blocks may be generated within this form by selecting the **Add** button. Add

Correspondence Information:

Enter either Customer Number or complete the Correspondence Information section below.
For further information see 37 CFR 1.33(a).

☐ **An Address is being provided for the correspondence Information of this application.**

Customer Number	21710		
Email Address	ip@brownrudnick.com	Add Email	Remove Email

Application Information:

Title of the Invention	COMPOSITIONS AND METHODS FOR IDENTIFICATION OF A DUPLICATE SEQUENCING READ		
Attorney Docket Number	NUGN-001/04US 34079/550	**Small Entity Status Claimed**	☐
Application Type	Nonprovisional		
Subject Matter	Utility		
Total Number of Drawing Sheets (if any)	6	**Suggested Figure for Publication (if any)**	

Filing By Reference:

Only complete this section when filing an application by reference under 35 U.S.C. 111(c) and 37 CFR 1.57(a). Do not complete this section if application papers including a specification and any drawings are being filed. Any domestic benefit or foreign priority information must be provided in the appropriate section(s) below (i.e., "Domestic Benefit/National Stage Information" and "Foreign Priority Information").

For the purposes of a filing date under 37 CFR 1.53(b), the description and any drawings of the present application are replaced by this reference to the previously filed application, subject to conditions and requirements of 37 CFR 1.57(a).

Application number of the previously filed application	Filing date (YYYY-MM-DD)	Intellectual Property Authority or Country

Publication Information:

☐ Request Early Publication (Fee required at time of Request 37 CFR 1.219)

☐ **Request Not to Publish.** I hereby request that the attached application not be published under 35 U.S.C. 122(b) and certify that the invention disclosed in the attached application **has not and will not** be the subject of an application filed in another country, or under a multilateral international agreement, that requires publication at eighteen months after filing.

PTO/AIA/14 (02-18)
Approved for use through 11/30/2020. 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 contains a valid OMB control number.

Application Data Sheet 37 CFR 1.76	Attorney Docket Number	NUGN-001/04US 34079/550
	Application Number	
Title of Invention	COMPOSITIONS AND METHODS FOR IDENTIFICATION OF A DUPLICATE SEQUENCING READ	

Representative Information:

Representative information should be provided for all practitioners having a power of attorney in the application. Providing this information in the Application Data Sheet does not constitute a power of attorney in the application (see 37 CFR 1.32).
Either enter Customer Number or complete the Representative Name section below. If both sections are completed the customer Number will be used for the Representative Information during processing.

Please Select One:	● Customer Number	US Patent Practitioner	◯ Limited Recognition (37 CFR 11.9)
Customer Number	21710		

Domestic Benefit/National Stage Information:

This section allows for the applicant to either claim benefit under 35 U.S.C. 119(e), 120, 121, 365(c), or 386(c) or indicate National Stage entry from a PCT application. Providing benefit claim information in the Application Data Sheet constitutes the specific reference required by 35 U.S.C. 119(e) or 120, and 37 CFR 1.78.
When referring to the current application, please leave the "Application Number" field blank.

Prior Application Status	Pending		Remove
Application Number	Continuity Type	Prior Application Number	Filing or 371(c) Date (YYYY-MM-DD)
	Continuation of	16/743604	2020-01-15
Prior Application Status	Pending		Remove
Application Number	Continuity Type	Prior Application Number	Filing or 371(c) Date (YYYY-MM-DD)
16/743604	Continuation of	15/406002	2017-01-13
Prior Application Status	Pending		Remove
Application Number	Continuity Type	Prior Application Number	Filing or 371(c) Date (YYYY-MM-DD)
15/406002	Division of	14/540917	2014-11-13
Prior Application Status	Expired		Remove
Application Number	Continuity Type	Prior Application Number	Filing or 371(c) Date (YYYY-MM-DD)
14/540917	Claims benefit of provisional	61/903826	2013-11-13

Additional Domestic Benefit/National Stage Data may be generated within this form by selecting the **Add** button. Add

Foreign Priority Information:

IDT Exhibit 1020.0019

PTO/AIA/14 (02-18)
Approved for use through 11/30/2020. 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 contains a valid OMB control number.

Application Data Sheet 37 CFR 1.76	Attorney Docket Number	NUGN-001/04US 34079/550
	Application Number	
Title of Invention	COMPOSITIONS AND METHODS FOR IDENTIFICATION OF A DUPLICATE SEQUENCING READ	

This section allows for the applicant to claim priority to a foreign application. Providing this information in the application data sheet constitutes the claim for priority as required by 35 U.S.C. 119(b) and 37 CFR 1.55. When priority is claimed to a foreign application that is eligible for retrieval under the priority document exchange program (PDX)[i] the information will be used by the Office to automatically attempt retrieval pursuant to 37 CFR 1.55(i)(1) and (2). Under the PDX program, applicant bears the ultimate responsibility for ensuring that a copy of the foreign application is received by the Office from the participating foreign intellectual property office, or a certified copy of the foreign priority application is filed, within the time period specified in 37 CFR 1.55(g)(1).

Remove

Application Number	Country[i]	Filing Date (YYYY-MM-DD)	Access Code[i] (if applicable)

Additional Foreign Priority Data may be generated within this form by selecting the **Add** button. Add

Statement under 37 CFR 1.55 or 1.78 for AIA (First Inventor to File) Transition Applications

☐ This application (1) claims priority to or the benefit of an application filed before March 16, 2013 and (2) also contains, or contained at any time, a claim to a claimed invention that has an effective filing date on or after March 16, 2013.
NOTE: By providing this statement under 37 CFR 1.55 or 1.78, this application, with a filing date on or after March 16, 2013, will be examined under the first inventor to file provisions of the AIA.

IDT Exhibit 1020.0020

PTO/AIA/14 (02-18)
Approved for use through 11/30/2020. 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 contains a valid OMB control number.

Application Data Sheet 37 CFR 1.76	Attorney Docket Number	NUGN-001/04US 34079/550
	Application Number	
Title of Invention	COMPOSITIONS AND METHODS FOR IDENTIFICATION OF A DUPLICATE SEQUENCING READ	

Authorization or Opt-Out of Authorization to Permit Access:

When this Application Data Sheet is properly signed and filed with the application, applicant has provided written authority to permit a participating foreign intellectual property (IP) office access to the instant application-as-filed (see paragraph A in subsection 1 below) and the European Patent Office (EPO) access to any search results from the instant application (see paragraph B in subsection 1 below).

Should applicant choose not to provide an authorization identified in subsection 1 below, applicant **must opt-out** of the authorization by checking the corresponding box A or B or both in subsection 2 below.

NOTE: This section of the Application Data Sheet is **ONLY** reviewed and processed with the **INITIAL** filing of an application. After the initial filing of an application, an Application Data Sheet cannot be used to provide or rescind authorization for access by a foreign IP office(s). Instead, Form PTO/SB/39 or PTO/SB/69 must be used as appropriate.

1. Authorization to Permit Access by a Foreign Intellectual Property Office(s)

A. Priority Document Exchange (PDX) - Unless box A in subsection 2 (opt-out of authorization) is checked, the undersigned hereby **grants the USPTO authority** to provide the European Patent Office (EPO), the Japan Patent Office (JPO), the Korean Intellectual Property Office (KIPO), the State Intellectual Property Office of the People's Republic of China (SIPO), the World Intellectual Property Organization (WIPO), and any other foreign intellectual property office participating with the USPTO in a bilateral or multilateral priority document exchange agreement in which a foreign application claiming priority to the instant patent application is filed, access to: (1) the instant patent application-as-filed and its related bibliographic data, (2) any foreign or domestic application to which priority or benefit is claimed by the instant application and its related bibliographic data, and (3) the date of filing of this Authorization. See 37 CFR 1.14(h)(1).

B. Search Results from U.S. Application to EPO - Unless box B in subsection 2 (opt-out of authorization) is checked, the undersigned hereby **grants the USPTO authority** to provide the EPO access to the bibliographic data and search results from the instant patent application when a European patent application claiming priority to the instant patent application is filed. See 37 CFR 1.14(h)(2).

The applicant is reminded that the EPO's Rule 141(1) EPC (European Patent Convention) requires applicants to submit a copy of search results from the instant application without delay in a European patent application that claims priority to the instant application.

2. Opt-Out of Authorizations to Permit Access by a Foreign Intellectual Property Office(s)

☐ A. Applicant **DOES NOT** authorize the USPTO to permit a participating foreign IP office access to the instant application-as-filed. If this box is checked, the USPTO will not be providing a participating foreign IP office with any documents and information identified in subsection 1A above.

☐ B. Applicant **DOES NOT** authorize the USPTO to transmit to the EPO any search results from the instant patent application. If this box is checked, the USPTO will not be providing the EPO with search results from the instant application.

NOTE: Once the application has published or is otherwise publicly available, the USPTO may provide access to the application in accordance with 37 CFR 1.14.

IDT Exhibit 1020.0021

PTO/AIA/14 (02-18)
Approved for use through 11/30/2020. 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 contains a valid OMB control number.

Application Data Sheet 37 CFR 1.76	Attorney Docket Number	NUGN-001/04US 34079/550
	Application Number	
Title of Invention	COMPOSITIONS AND METHODS FOR IDENTIFICATION OF A DUPLICATE SEQUENCING READ	

Applicant Information:

Providing assignment information in this section does not substitute for compliance with any requirement of part 3 of Title 37 of CFR to have an assignment recorded by the Office.

Applicant 1 — Remove

If the applicant is the inventor (or the remaining joint inventor or inventors under 37 CFR 1.45), this section should not be completed. The information to be provided in this section is the name and address of the legal representative who is the applicant under 37 CFR 1.43; or the name and address of the assignee, person to whom the inventor is under an obligation to assign the invention, or person who otherwise shows sufficient proprietary interest in the matter who is the applicant under 37 CFR 1.46. If the applicant is an applicant under 37 CFR 1.46 (assignee, person to whom the inventor is obligated to assign, or person who otherwise shows sufficient proprietary interest) together with one or more joint inventors, then the joint inventor or inventors who are also the applicant should be identified in this section.

Clear

● Assignee	Legal Representative under 35 U.S.C. 117	Joint Inventor
Person to whom the inventor is obligated to assign.	Person who shows sufficient proprietary interest	

If applicant is the legal representative, indicate the authority to file the patent application, the inventor is:

Name of the Deceased or Legally Incapacitated Inventor:

If the Applicant is an Organization check here. ☒

Organization Name	Tecan Genomics, Inc.		
Mailing Address Information For Applicant:			
Address 1	900 Chesapeake Drive		
Address 2			
City	Redwood City	**State/Province**	CA
Country	US	Postal Code	94063
Phone Number		Fax Number	
Email Address			

Additional Applicant Data may be generated within this form by selecting the Add button. Add

Assignee Information including Non-Applicant Assignee Information:

Providing assignment information in this section does not substitute for compliance with any requirement of part 3 of Title 37 of CFR to have an assignment recorded by the Office.

IDT Exhibit 1020.0022

PTO/AIA/14 (02-18)
Approved for use through 11/30/2020. 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 contains a valid OMB control number.

Application Data Sheet 37 CFR 1.76	Attorney Docket Number	NUGN-001/04US 34079/550
	Application Number	
Title of Invention	COMPOSITIONS AND METHODS FOR IDENTIFICATION OF A DUPLICATE SEQUENCING READ	

Assignee 1

Complete this section if assignee information, including non-applicant assignee information, is desired to be included on the patent application publication. An assignee-applicant identified in the "Applicant Information" section will appear on the patent application publication as an applicant. For an assignee-applicant, complete this section only if identification as an assignee is also desired on the patent application publication.

Remove

If the Assignee or Non-Applicant Assignee is an Organization check here. ☐

Prefix	**Given Name**	Middle Name	**Family Name**	Suffix

Mailing Address Information For Assignee including Non-Applicant Assignee:

Address 1			
Address 2			
City		**State/Province**	
Country i		Postal Code	
Phone Number		Fax Number	
Email Address			

Additional Assignee or Non-Applicant Assignee Data may be generated within this form by selecting the Add button. Add

Signature:

Remove

NOTE: This Application Data Sheet must be signed in accordance with 37 CFR 1.33(b). **However, if this Application Data Sheet is submitted with the INITIAL filing of the application and either box A or B is not checked in subsection 2 of the "Authorization or Opt-Out of Authorization to Permit Access" section, then this form must also be signed in accordance with 37 CFR 1.14(c).**

This Application Data Sheet **must** be signed by a patent practitioner if one or more of the applicants is a **juristic entity** (e.g., corporation or association). If the applicant is two or more joint inventors, this form must be signed by a patent practitioner, **all** joint inventors who are the applicant, or one or more joint inventor-applicants who have been given power of attorney (e.g., see USPTO Form PTO/AIA/81) on behalf of **all** joint inventor-applicants.

See 37 CFR 1.4(d) for the manner of making signatures and certifications.

Signature	/Thomas C. Meyers/			Date (YYYY-MM-DD)	2021-07-21
First Name	Thomas	Last Name	Meyers	Registration Number	36,989

Additional Signature may be generated within this form by selecting the Add button. Add

PTO/AIA/14 (02-18)
Approved for use through 11/30/2020. 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 contains a valid OMB control number.

Application Data Sheet 37 CFR 1.76	Attorney Docket Number	NUGN-001/04US 34079/550
	Application Number	
Title of Invention	COMPOSITIONS AND METHODS FOR IDENTIFICATION OF A DUPLICATE SEQUENCING READ	

This collection of information is required by 37 CFR 1.76. 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 23 minutes to complete, including gathering, preparing, and submitting the completed application data sheet 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.**

Privacy Act Statement

The Privacy Act of 1974 (P.L. 93-579) requires that you be given certain information in connection with your submission of the attached form related to a patent application or patent. Accordingly, pursuant to the requirements of the Act, please be advised that: (1) the general authority for the collection of this information is 35 U.S.C. 2(b)(2); (2) furnishing of the information solicited is voluntary; and (3) the principal purpose for which the information is used by the U.S. Patent and Trademark Office is to process and/or examine your submission related to a patent application or patent. If you do not furnish the requested information, the U.S. Patent and Trademark Office may not be able to process and/or examine your submission, which may result in termination of proceedings or abandonment of the application or expiration of the patent.

The information provided by you in this form will be subject to the following routine uses:

1 The information on this form will be treated confidentially to the extent allowed under the Freedom of Information Act (5 U.S.C. 552) and the Privacy Act (5 U.S.C. 552a). Records from this system of records may be disclosed to the Department of Justice to determine whether the Freedom of Information Act requires disclosure of these records.

2. A record from this system of records may be disclosed, as a routine use, in the course of presenting evidence to a court, magistrate, or administrative tribunal, including disclosures to opposing counsel in the course of settlement negotiations.

3 A record in this system of records may be disclosed, as a routine use, to a Member of Congress submitting a request involving an individual, to whom the record pertains, when the individual has requested assistance from the Member with respect to the subject matter of the record.

4. A record in this system of records may be disclosed, as a routine use, to a contractor of the Agency having need for the information in order to perform a contract. Recipients of information shall be required to comply with the requirements of the Privacy Act of 1974, as amended, pursuant to 5 U.S.C. 552a(m).

5. A record related to an International Application filed under the Patent Cooperation Treaty in this system of records may be disclosed, as a routine use, to the International Bureau of the World Intellectual Property Organization, pursuant to the Patent CooperationTreaty.

6. A record in this system of records may be disclosed, as a routine use, to another federal agency for purposes of National Security review (35 U.S.C. 181) and for review pursuant to the Atomic Energy Act (42 U.S.C. 218(c)).

7. A record from this system of records may be disclosed, as a routine use, to the Administrator, General Services, or his/her designee, during an inspection of records conducted by GSA as part of that agency's responsibility to recommend improvements in records management practices and programs, under authority of 44 U.S.C. 2904 and 2906. Such disclosure shall be made in accordance with the GSA regulations governing inspection of records for this purpose, and any other relevant (i.e., GSA or Commerce) directive. Such disclosure shall not be used to make determinations about individuals.

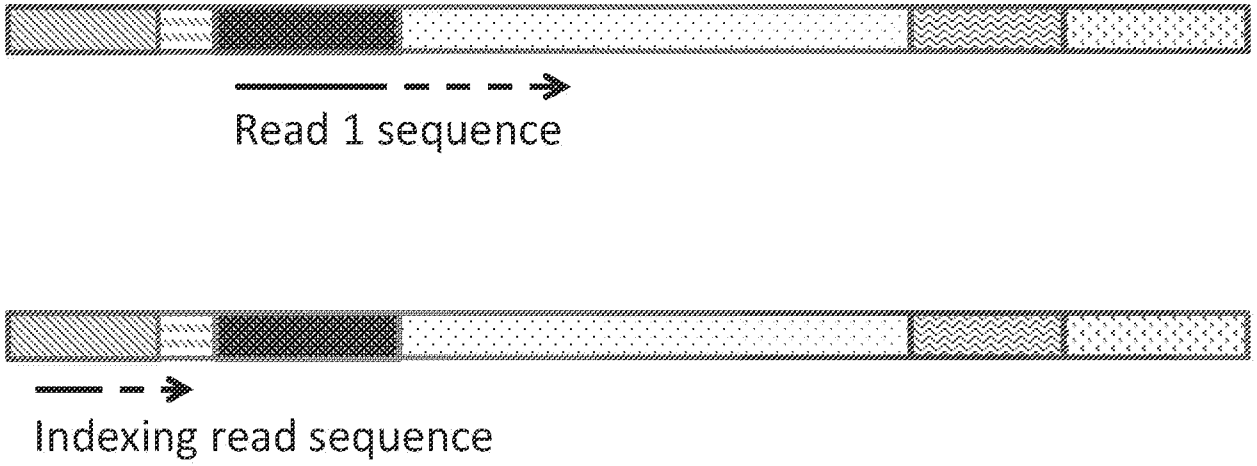
8. A record from this system of records may be disclosed, as a routine use, to the public after either publication of the application pursuant to 35 U.S.C. 122(b) or issuance of a patent pursuant to 35 U.S.C. 151. Further, a record may be disclosed, subject to the limitations of 37 CFR 1.14, as a routine use, to the public if the record was filed in an application which became abandoned or in which the proceedings were terminated and which application is referenced by either a published application, an application open to public inspections or an issued patent.

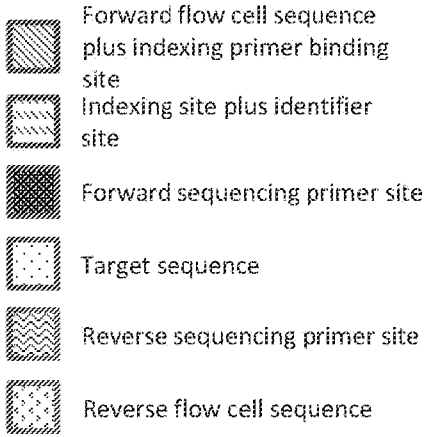
9. A record from this system of records may be disclosed, as a routine use, to a Federal, State, or local law enforcement agency, if the USPTO becomes aware of a violation or potential violation of law or regulation.

IDT Exhibit 1020.0025

Standard DNA Sequencing With Duplication Counter

- Sequence Read 1 (forward read)





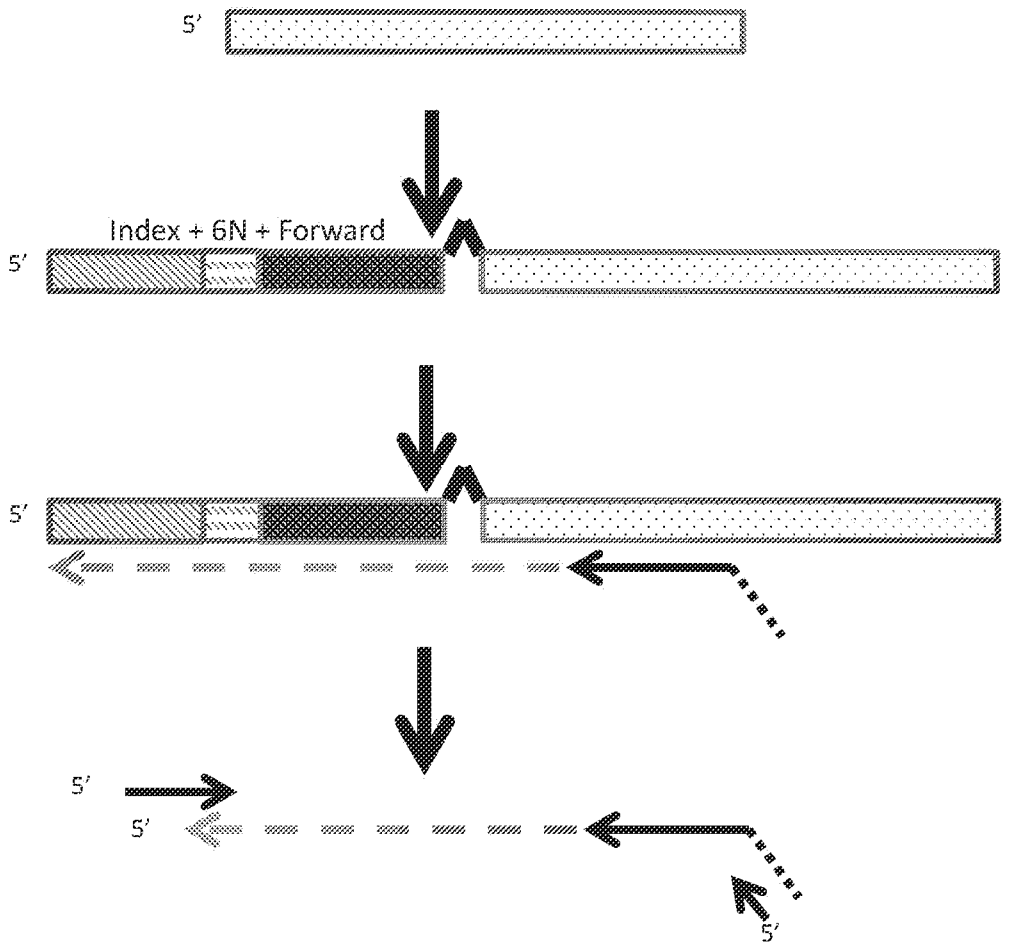
The indexing read results in sequencing of both the index site and the identifier site

- Extension of the indexing read beyond the indexing site and into the nucleotides of the identifier site provides data on which nucleic acid molecule is being sequenced.

Figure 1

IDT Exhibit 1020.0026

Single Primer Extension (SPET) Reaction



gDNA is sheared

Adapters are ligated on to end-repaired DNA. The adapters contain at least a partial indexing sequence priming site, an indexing site, an identifier site, and the forward sequencing priming site. Each gDNA is now labeled with a unique identifier sequence

Targeting oligos are hybridized to gDNA and extended through the adapter containing an identifier site (6N) sequence

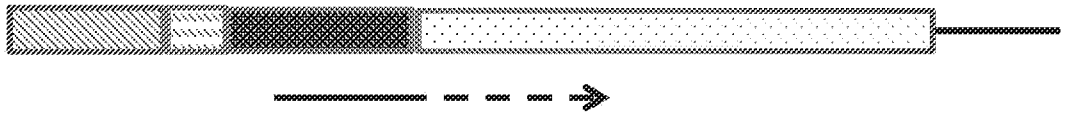
Probes that have extended through an adapter are PCR amplified for sequencing

Figure 2A

IDT Exhibit 1020.0027

SPET Sequencing

- Sequence from random end toward the probe



Read 1 sequence

PCR adds sequence specific primer binding sites and Read 1 of sequencing is performed



Indexing read sequence

Indexing Read performed sequences both the indexing site as well as the identifier site (N6)

- Extending the indexing read past the 6 base indexing site sequence and into the random Ns of the identifier site provides data on which gDNA molecule the extension reaction occurred upon.

Figure 2B

IDT Exhibit 1020.0028

Example Adaptor

An adaptor sequence with necessary priming sites, the index site, and the identifier site

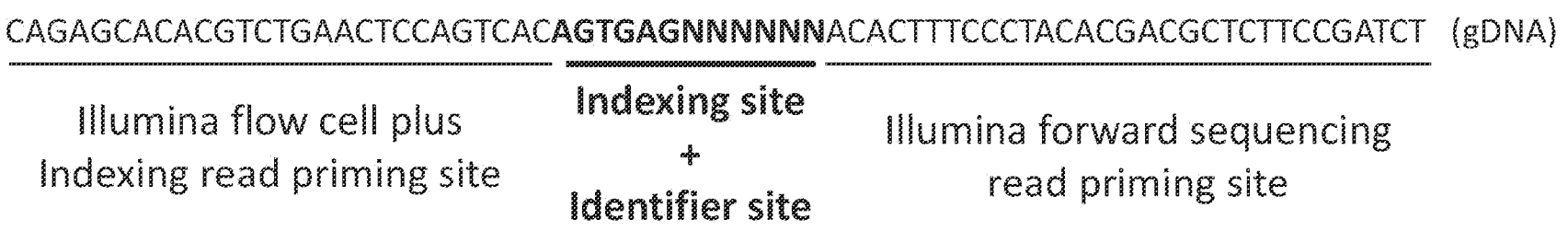
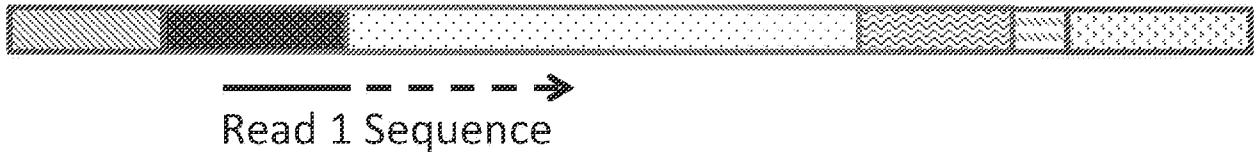


Figure 3

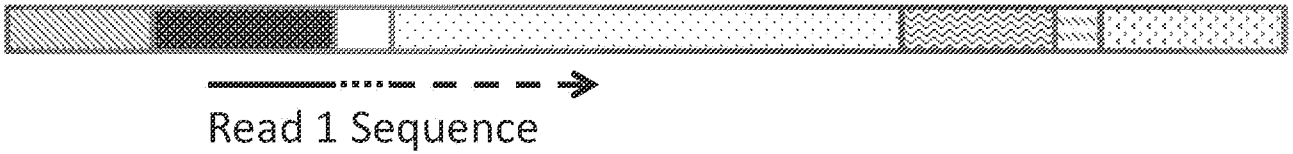
IDT Exhibit 1020.0029

DNA Sequencing With Duplicate Recognition at the insert junction

- Traditional library, sequence read 1 (forward read)



- Duplicate recognition library, sequence read 1 (forward read)



Read 1 initially records the approximately 8 random bases of the identifier site appended to the library sequences, followed by the target sequence

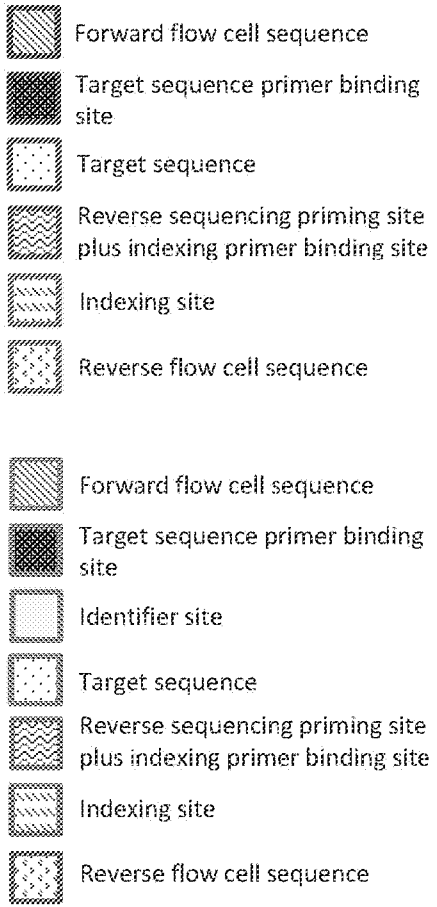


Figure 4

IDT Exhibit 1020.0030

Resolution of apparent duplicate reads by identifier sequence ligation

Libraries were generated from 100ng genomic DNA Covaris fragmented to ~500bp in length. These were enriched with the 344 gene cancer panel, ~12k probes covering 969kb of genomic space and sequenced on the ILMN Miseq. Different numbers of reads were randomly selected for this analysis

Total reads randomly selected	Aligned read	Fold Target Coverage	Total duplicates after identifier analysis (true rate)	Percentage duplicates after identifier analysis (true rate)	Total apparent duplicates with no identifier	Percentage apparent duplicates with no identifier
100K	97,919	2X	1854	1.89	3256	3.33
500K	488,423	7X	39731	8.13	68741	14.07
1M	976,662	13X	145589	14.91	237882	24.36
2M	1,954,673	21X	514848	26.34	764825	39.13
10M	9,771,462	43X	6493577	66.45	7515496	76.91

"Apparent" Duplicates – individual targets within libraries consist of one random end (result of fragmentation) and one common end (initiated with the sequence specific probe). Apparent duplicates are derived from sequencing libraries and using conventional measures of duplicate reads ie, reads were mapped using bowtie and all reads with the same start and end genomic coordinates were counted as duplicates.

"True" duplicates - An identifier sequence is introduced through ligation to differentiate between fragments of DNA that randomly have the same start and end mapping coordinates. The analysis method is described in the specification.

Figure 5

IDT Exhibit 1020.0031

ABSTRACT

The present invention provides methods, compositions and kits for detecting duplicate sequencing reads. In some embodiments, the duplicate sequencing reads are removed.

COMPOSITIONS AND METHODS FOR IDENTIFICATION OF A DUPLICATE SEQUENCING READ

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application claims the benefit of U.S. Provisional Application Serial No. 61/903,826, filed November 13, 2013, which is herein incorporated by reference in its entirety.

DESCRIPTION OF THE TEXT FILE SUBMITTED ELECTRONICALLY

[0002] The contents of the text file submitted electronically herewith are incorporated herein by reference in their entirety: A computer readable format copy of the Sequence Listing (filename: NUGN_001_01US_SeqList_ST25.txt, date recorded: November 12, 2014, file size 3 kilobytes).

FIELD OF THE PRESENT INVENTION

[0003] The present invention relates generally to the field of high throughput sequencing reactions and the ability to discern artifacts arising through sequence duplications from nucleotide molecules that are unique molecules.

BACKGROUND OF THE PRESENT INVENTION

[0004] In RNA sequencing applications, accurate gene expression measurements may be hampered by PCR duplicate artifacts that occur during library amplification. When analyzing RNA sequencing data, when two or more identical sequences are found, it can be difficult to know if these represent unique cDNA molecules derived independently from different RNA molecules, or if they are PCR duplicates derived from a single RNA molecule. In genotyping by

sequencing, duplicate reads can be considered non-informative and may be collapsed down to a single read, thus reducing the number of sequencing reads used in final analysis. Generally, sequencing reads may be determined to be duplicates if both forward and reverse reads have identical starting positions, even though two independently generated molecules can have identical starting positions by random chance. Single primer extension based targeted re-sequencing suffers from an issue in that only one end of a sequencing read is randomly generated, while the other (reverse read) end is generated by a specific probe. This may make it difficult to determine if two reads are duplicates because they have been duplicated by PCR or because by chance they happened to start at the same position.

[0005] In expression analysis studies there may be limited value in doing paired end sequencing since the goal of the experiment is to determine amounts of transcript present as opposed to studying exon usage. In these studies, paired end sequencing adds costs while the only value is in helping distinguish PCR duplicates. The probability of two reads starting in the same position on only one end is higher than the probability of two reads having the same starting position on two ends (forward and reverse read). There is a need for improved methods that allow for low-cost, high throughput sequencing of regions of interest, genotyping or simple detection of RNA transcripts without inherent instrument inefficiencies that drive up sequencing costs due to the generation of unusable or non-desired data reads. The invention described herein fulfills this need. Here, we describe an adaptor approach that allows for the identification of true PCR duplicates and their removal.

[0006] The methods of the present invention provide novel methods for identifying true duplicate reads during sequencing, such as to improve data analysis of sequencing data, and other related advantages.

SUMMARY OF THE PRESENT INVENTION

[0007] The present invention is based, in part, on compositions and methods for discerning duplicate sequencing reads from a population of sequencing reads. The detection and/or removal of duplicate sequencing reads presented herein is a novel approach to increasing the efficacy of evaluating data generated from high throughput sequence reactions, including complex multiplex sequence reactions.

[0008] Accordingly, the present invention provides a method of detecting a duplicate sequencing read from a population of sample sequencing reads, the method comprising ligating an adaptor to a 5' end of each nucleic acid fragment of a plurality of nucleic acid fragments from one or more samples, wherein the adaptor comprises an indexing primer binding site, an indexing site, an identifier site, and a target sequence primer binding site. The ligated adaptor-nucleic acid fragment products can be amplified, thus generating a population of sequencing reads from the amplified adaptor-nucleic acid ligation products. The sequencing reads with a duplicate identifier site and target sequence can then be detected from the population of sequencing reads. The methods can further include the removal of the sequencing reads with the duplicate identifier site and target sequence from the population of sequence reads.

[0009] In some embodiments, the identifier site is sequenced with the indexing site or the target sequence. In further embodiments, the identifier site is sequenced separately from the indexing site or the target sequence.

[0010] In some embodiments, the adaptor comprises from 5' to 3' the indexing primer binding site; the indexing site; the identifier site; and the target sequence primer binding site. In further embodiments, the adaptor comprises from 5' to 3' the indexing primer binding site; the indexing site; the target sequence primer binding site; and the identifier site.

[0011] In some embodiments, the plurality of nucleic acid fragments is generated from more than one sample. In some embodiments, the nucleic acid fragments from each sample have the same indexing site. In some embodiments, the sequencing reads are separated based on the indexing site. In yet other embodiments, the separation of sequencing reads is performed prior to detecting sequence reads with a duplicate identifier site and target sequence.

[0012] In some embodiments, the nucleic acid fragments are DNA fragments, RNA fragments, or DNA/RNA fragments. In further embodiments, the nucleic acid fragments are genomic DNA fragments or cDNA fragments.

[0013] In some embodiments, the indexing site is between 2 and 8 nucleotides in length. In further embodiments, the indexing site is about 6 nucleotides in length. In some embodiments, the identifier site is between 1 and 8 nucleotides in length. In further embodiments, the identifier site is about 8 nucleotides in length.

[0014] In some embodiments, the indexing primer binding site is a universal indexing primer binding site; and in some embodiments, the target sequence primer binding site is a universal target sequence primer binding site.

[0015] The present invention also encompasses embodiments that include a kit comprising a plurality of adaptors, wherein each adaptor comprise an indexing primer binding site; an indexing site, and identifier site, and a target sequencing primer binding site.

BRIEF DESCRIPTION OF THE DRAWINGS

[0016] A better understanding of the novel features of the invention and advantages of the present invention will be obtained by reference to the following description that sets forth

illustrative embodiments, in which the principles of the invention are utilized, and the accompanying drawings of which:

[0017] Figure 1 depicts a schematic of generating sequencing reads of a library including where an indexing primer and a target sequence primer anneal.

[0018] Figure 2A depicts a mechanism of the Single Primer Enrichment Technology (SPET) and how an identifier site is carried over into the final library and how sequencing through the indexing site into the identifier site provides data on which nucleic acid molecule is being identified.

[0019] Figure 2B is a continuation of Figure 2A.

[0020] Figure 3 offers a detailed view of the sequence of an example adaptor, among many envisioned embodiments, and the position of the indexing and identifier sites (SEQ ID NO: 1). "N" refers to any nucleic acid.

[0021] Figure 4 depicts a schematic of two separate sequence libraries, indicating where the indexing primers and the target primers anneal in a traditional library as compared to a library using an identifier site.

[0022] Figure 5 depicts a data table demonstrating the accuracy of resolving true duplicates versus apparent or perceived duplicates using an identifier site in the adaptors.

DETAILED DESCRIPTION OF THE PRESENT INVENTION

[0023] The present invention is based, in part, on compositions and methods for discerning duplicate sequencing reads from a population of sequencing reads. The present invention encompasses methods of detecting sequences duplicated in sequencing applications, and further removal of the duplicated sequence reads. The present invention further encompasses kits

comprising components that would allow for customized applications of the method of detecting and removing duplicated sequence reads in high throughput sequencing reactions. The compositions and methods can be used with various applications for genetic sample analysis, such as RNA sequence analysis, copy number variation analysis, methylation sequencing analysis, genotyping and whole genome amplification.

[0024] Reference will now be made in detail to exemplary embodiments of the invention. While the disclosed methods and compositions will be described in conjunction with the exemplary embodiments, it will be understood that these exemplary embodiments are not intended to limit the present invention. On the contrary, the present disclosure is intended to encompass alternatives, modifications and equivalents, which may be included in the spirit and scope of the present invention.

[0025] Unless otherwise specified, terms and symbols of genetics, molecular biology, biochemistry and nucleic acid used herein follow those of standard treatises and texts in the field, e.g. Kornberg and Baker, *DNA Replication,* Second Edition (W.H. Freeman, New York, 1992); Lehninger, *Biochemistry,* Second Edition (Worth Publishers, New York, 1975); Strachan and Read, Human *Molecular Genetics,* Second Edition (Wiley-Liss, New York, 1999); Eckstein, editor, *Oligonucleotides and Analogs: A Practical Approach* (Oxford University Press, New York, 1991); Gait, editor, *Oligonucleotide Synthesis: A Practical Approach* (IRL Press, Oxford,1984); and the like.

[0026] In some embodiments, the methods disclosed herein is for detecting from a population of sequencing reads a sequencing read, such as a duplicate sequencing read with a duplicate identifier site and target sequence. A duplicate sequencing read can be a sequencing read with

the same identifier site and target sequence as another sequencing read in the population of sequencing reads.

Adaptor

[0027] The present invention provides compositions of adaptors and methods comprising use of an adaptor. An adapter refers to an oligonucleotide sequence, the ligation of which to a target polynucleotide or a target polynucleotide strand of interest enables the generation of amplification-ready products of the target polynucleotide or the target polynucleotide strand of interest. The target polynucleotide molecules may be fragmented or not prior to the addition of adaptors. In some embodiments, a method disclosed herein comprises ligating an adaptor to a 5' end of each nucleic acid fragment of a plurality of nucleic acid fragments from one or more samples.

[0028] Various adaptor designs are envisioned which are suitable for generation of amplification-ready products of target sequence regions/strands of interest. For example, the two strands of the adaptor may be self-complementary, non-complementary or partially complementary. In some embodiments, the adaptor can comprise an indexing primer binding site, an indexing site, an identifier site, and a target sequence primer binding site.

[0029] An indexing primer binding site is a nucleotide sequence for binding a primer for an indexing site. An indexing site is a nucleic acid sequence that acts as an index for multiple polynucleotide samples, thus allowing for the samples to be pooled together into a single sequencing run, which is known as multiplexing. In some embodiments, the indexing site is at least 2, 3, 4, 5, 6, 7, 8, 9, or 10 nucleotides in length. In some embodiments, the indexing site is between 2 and 8 nucleotides in length. In some embodiments, the indexing site is about 6 nucleotides in length.

[0030] An identifier site is a nucleic acid sequence that comprises random bases and is used to identify duplicate sequencing reads. In some embodiments, the identifier site is at least 2, 3, 4, 5, 6, 7, 8, 9, or 10 nucleotides in length. In some embodiments, the identifier site is between 1 and 8 nucleotides in length. In some embodiments, the identifier site is about 8 nucleotides in length. This identifier site can be designed as a set of sequences, or it can be semi-random, or it can be completely random. In addition, this identifier site can be a fixed length, or it can be a variable length. In some embodiments, the identifier sites in a plurality of adaptors are of a fixed length. For example, the identifier sites can all be eight random bases. In another embodiment, the identifier sites in a plurality of adaptors are of a variable length. For example, the identifier sites can range in size from 1 to 8 bases. In yet another embodiment, the identifier sites can be of a defined set of defined sequence. For example, the identifier site of a plurality of adaptors can be one of 96 defined six-base nucleotide sequences.

[0031] A target sequence primer binding site nucleotide sequence for binding a primer for a target sequence. The primer can be used to amplify the target sequence (e.g., a nucleic acid fragment from a sample). Accordingly, in some embodiments, the adaptor comprises an indexing primer binding site and a target sequence primer binding site.

[0032] A primer is a polynucleotide chain, typically less than 200 residues long, most typically between 15 and 100 nucleotides long, but can encompass longer polynucleotide chains. A primer targeting the primer binding sites are typically designed to hybridize to single-stranded nucleic acid strands. In some embodiments, the primers targeting the primer binding sites are designed to hybridize to single-stranded DNA targets. In the case where the sample comprises genomic DNA or other double-stranded DNA, the sample can be first denatured to render the target single stranded and enable hybridization of the primers to the desired sequence regions of

interest. In these embodiments, the methods and compositions described herein can allow for region-specific enrichment and amplification of sequence regions of interest. In some embodiments, the other double-stranded DNA can be double-stranded cDNA generated by first and second strand synthesis of one or more target RNAs.

[0033] In other embodiments, the primers targeting the primer binding sites are designed to hybridize to double-stranded nucleic acid targets, without denaturation of the double stranded nucleic acids. In other embodiments, the primers targeting the primer binding sites are designed to hybridize to a double-stranded DNA target, without denaturation of the dsDNA. In these embodiments, the primers targeting the selected sequence regions of interest are designed to form a triple helix (triplex) at the selected sequence regions of interest. The hybridization of the primers to the double-stranded DNA sequence regions of interest can be carried out without prior denaturation of the double stranded nucleic acid sample. In such embodiments, the methods and compositions described herein can allow for region-specific enrichment as well as strand-specific enrichment and amplification of sequence regions of interest. This method can be useful for generation of copies of strand specific sequence regions of interest from complex nucleic acid without the need to denature the dsDNA input DNA, thus enabling enrichment and analysis of multiplicity of sequence regions of interest in the native complex nucleic acid sample. The method can find use for studies and analyses carried out in situ, enable studies and analysis of complex genomic DNA in single cells or collection of very small well defined cell population, as well as permit the analysis of complex genomic DNA without disruption of chromatin structures.

[0034] The primers of the invention are generally oligonucleotides that are employed in an extension reaction by a polymerase along a polynucleotide template, such as for amplification of a target sequence (e.g., in PCR). The oligonucleotide primer can be a synthetic polynucleotide

that is single stranded, containing a sequence at its 3'-end that is capable of hybridizing with a sequence of the target polynucleotide. In some embodiments, the 3' region of the primer that hybridizes with the target nucleic acid has at least 80%, preferably 90%, more preferably 95%, most preferably 100%, complementarity to a primer binding site.

[0035] In some embodiments, the primer binding site is a binding site for a universal primer. A universal primer is a primer that can be used for amplifying a number of different sequences. In some embodiments, a universal primer is used to amplify different libraries. In some embodiments, an indexing primer binding site is a binding site for a universal indexing primer (i.e., the indexing primer binding site is a universal indexing primer binding site). In some embodiments, the adaptors used for ligating a plurality of nucleic acid fragments have a universal indexing primer binding site. In some embodiments, the universal indexing primer can be used to amplify and/or sequence a number of different indexing sites.

[0036] In some embodiments, a target sequence primer binding site is a binding site for a universal target sequence primer (i.e., the target sequence primer binding site is a universal target sequence primer binding site). In some embodiments, the adaptors used for ligating a plurality of nucleic acid fragments have a universal target sequence primer binding site. In some embodiments, the universal target sequence primer can be used to amplify and/or sequence a number of different target sequences.

[0037] In some embodiments, the adaptor comprises an identifier site 3' to an indexing site. In some embodiments, the adaptor comprises an identifier site 5' to an indexing site. In some embodiments, the adaptor comprises from 5' to 3' an indexing primer binding site, indexing site, identifier site, and target sequence primer binding site. In other embodiments, the adaptor comprises from 5' to 3' an indexing primer binding site, indexing site, identifier site, and target

sequence primer binding site. In yet other embodiments, the adaptor comprises from 5' to 3' an indexing primer binding site, identifier site, indexing site, and target sequence primer binding site.

Samples

[0038] In some embodiments, an adaptor is ligated to a nucleic acid fragment (e.g., the 5' end of the nucleic acid fragment). The nucleic acid fragment can be from a plurality of nucleic acid fragments from one or more samples. The nucleic acid fragment can be RNA, DNA, or complex DNA, for example genomic DNA and PNA, in which case one might use a modified nucleic acid. The nucleic acid fragment may also be cDNA. The cDNA can be generated from RNA, e.g., mRNA.

[0039] The sample can be a biological sample. For example, the sample can be an animal, plant, bacterial, algal, or viral sample. In some embodiments, the sample is a human, rat, or mouse sample. The sample can be from a mixture of genomes of different species such as host-pathogen, bacterial populations and the like. The sample can be cDNA made from a mixture of genomes of different species. In some embodiments, the sample can be from a synthetic source. The sample can be mitochondrial DNA. The sample can be cell-free DNA. The cell-free DNA can be obtained from sources such as a serum or a plasma sample. The sample can comprise one or more chromosomes. For example, if the sample is from a human, the sample can comprise one or more of chromosome 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, X, or Y. In some embodiments, the sample comprises a linear or circular genome. The sample can be plasmid DNA, cosmid DNA, bacterial artificial chromosome (BAC), or yeast artificial chromosome (YAC). The sample can be from more than one individual or organism. The sample

can be double-stranded or single-stranded. The sample can be part of chromatin. The sample can be associated with histones.

[0040] In some embodiments, adaptors are ligated to a plurality of nucleic acid fragments from more than one sample, such as 2, 3, 4, 5, or more samples. In some embodiments, the nucleic acid fragments from each sample have the same indexing site. In some embodiments, a plurality of nucleic acid fragments is generated from a first sample and a second sample and adaptors are ligated to each nucleic acid fragment, in which adaptors ligated to each nucleic acid fragment from the first sample have the same first indexing site and adaptors ligated to nucleic acid fragments from the second sample has the same second indexing site. In some embodiments, the nucleic acid fragments or data associated with the nucleic acid fragments (e.g., sequencing reads) are separated based on the indexing site.

[0041] In some embodiments, a population of nucleic acid fragments generated from a sample is of one or more specific size range(s). In some embodiments, the fragments have an average length from about 10 to about 10,000 nucleotides. In some embodiments, the fragments have an average length from about 50 to about 2,000 nucleotides. In some embodiments, the fragments have an average length from about 100-2,500, 10-1,000, 10-800, 10-500, 50-500, 50-250, or 50-150 nucleotides. In some embodiments, the fragments have an average length less than 10,000 nucleotide, such as less than 5,000 nucleotides, less than 2,500 nucleotides, less than 2,500 nucleotides, less than 1,000 nucleotides, less than 500 nucleotides, such as less than 400 nucleotides, less than 300 nucleotides, less than 200 nucleotides, or less than 150 nucleotides.

[0042] In some embodiments, fragmentation of the nucleic acids can be achieved through methods known in the art. Fragmentation can be achieved through physical fragmentation methods and/or enzymatic fragmentation methods. Physical fragmentation methods can include

nebulization, sonication, and/or hydrodynamic shearing. In some embodiments, the fragmentation can be accomplished mechanically comprising subjecting the nucleic acids in the input sample to acoustic sonication. In some embodiments, the fragmentation comprises treating the nucleic acids in the input sample with one or more enzymes under conditions suitable for the one or more enzymes to generate double-stranded nucleic acid breaks. Examples of enzymes useful in the generation of nucleic acid or polynucleotide fragments include sequence specific and non-sequence specific nucleases. Non-limiting examples of nucleases include DNase I, Fragmentase, restriction endonucleases, variants thereof, and combinations thereof. Reagents for carrying out enzymatic fragmentation reactions are commercially available (e.g., from New England Biolabs). For example, digestion with DNase I can induce random double-stranded breaks in DNA in the absence of Mg^{++} and in the presence of Mn^{++}. In some embodiments, fragmentation comprises treating the nucleic acids in the input sample with one or more restriction endonucleases. Fragmentation can produce fragments having 5′ overhangs, 3′ overhangs, blunt ends, or a combination thereof. In some embodiments, such as when fragmentation comprises the use of one or more restriction endonucleases, cleavage of sample polynucleotides leaves overhangs having a predictable sequence. In some embodiments, the method includes the step of size selecting the fragments via standard methods known in the art such as column purification or isolation from an agarose gel.

[0043] In some embodiments, fragmentation of the nucleic acids is followed by end repair of the nucleic acid fragments. End repair can include the generation of blunt ends, non-blunt ends (i.e sticky or cohesive ends), or single base overhangs such as the addition of a single dA nucleotide to the 3′-end of the nucleic acid fragments, by a polymerase lacking 3′-exonuclease activity. End repair can be performed using any number of enzymes and/or methods known in the art

including, but not limited to, commercially available kits such as the Encore™ Ultra Low Input NGS Library System I. In some embodiments, end repair can be performed on double stranded DNA fragments to produce blunt ends wherein the double stranded DNA fragments contain 5′ phosphates and 3′ hydroxyls. In some embodiments, the double-stranded DNA fragments can be blunt-end polished (or "end-repaired") to produce DNA fragments having blunt ends, prior to being joined to adapters. Generation of the blunt ends on the double stranded fragments can be generated by the use of a single strand specific DNA exonuclease such as for example exonuclease 1, exonuclease 7 or a combination thereof to degrade overhanging single stranded ends of the double stranded products. Alternatively, the double stranded DNA fragments can be blunt ended by the use of a single stranded specific DNA endonuclease, for example, but not limited to, mung bean endonuclease or S1 endonuclease. Alternatively, the double stranded products can be blunt ended by the use of a polymerase that comprises single stranded exonuclease activity such as for example T4 DNA polymerase, or any other polymerase comprising single stranded exonuclease activity or a combination thereof to degrade the overhanging single stranded ends of the double stranded products. In some cases, the polymerase comprising single stranded exonuclease activity can be incubated in a reaction mixture that does or does not comprise one or more dNTPs. In other cases, a combination of single stranded nucleic acid specific exonucleases and one or more polymerases can be used to blunt end the double stranded fragments generated by fragmenting the sample comprising nucleic acids. In still other cases, the nucleic acid fragments can be made blunt ended by filling in the overhanging single stranded ends of the double stranded fragments. For example, the fragments may be incubated with a polymerase such as T4 DNA polymerase or Klenow polymerase or a combination thereof in the presence of one or more dNTPs to fill in the single stranded portions

of the double stranded fragments. Alternatively, the double stranded DNA fragments can be made blunt by a combination of a single stranded overhang degradation reaction using exonucleases and/or polymerases, and a fill-in reaction using one or more polymerases in the presence of one or more dNTPs.

[0044] U.S. Patent Publication Nos. 2013-0231253 A1 and 2014-0274729 A1 further describe methods of generating nucleic acid fragments, methods of modifying the fragments and analysis of the fragments, and are incorporated by reference in their entirety herein.

Ligation of Adaptors

[0045] Ligation of adaptors at the desired end of the sequence regions of interest (e.g., at the 5' or 3' end of a nucleic acid fragment generated form a sample) is suitable for carrying out the methods of the invention. Various ligation modalities are envisioned, dependent on the choice of nucleic acid, nucleic acid modifying enzymes and the resulting ligatable end of the nucleic acid. For example, when a blunt end product comprising the target region/sequence of interest is generated, blunt end ligation can be suitable. Alternatively, where the cleavage is carried out using a restriction enzyme of known sequence specificity, leading to the generation of cleavage sites with known sequence overhangs, suitable ends of the adaptors can be designed to enable hybridization of the adaptor to the cleavage site of the sequence region of interest and subsequent ligation. Ligation also refers to any joining of two nucleic acid molecules that results in a single nucleic acid sequence that can be further modified to obtain the sequence of the nucleic acids in question. Reagents and methods for efficient and rapid ligation of adaptors are commercially available, and are known in the art.

[0046] In some embodiments, the 5′ and/or 3′ end nucleotide sequences of fragmented nucleic acids are not modified or end-repaired prior to ligation with the adapter oligonucleotides of the present invention. For example, fragmentation by a restriction endonuclease can be used to leave

a predictable overhang, followed by ligation with one or more adapter oligonucleotides comprising an overhang complementary to the predictable overhang on a nucleic acid fragment. In another example, cleavage by an enzyme that leaves a predictable blunt end can be followed by ligation of blunt-ended nucleic acid fragments to adapter oligonucleotides comprising a blunt end. In some embodiments, end repair can be followed by an addition of 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 or more nucleotides, such as one or more adenine, one or more thymine, one or more guanine, or one or more cytosine, to produce an overhang. Nucleic acid fragments having an overhang can be joined to one or more adapter oligonucleotides having a complementary overhang, such as in a ligation reaction. For example, a single adenine can be added to the 3′ ends of end-repaired DNA fragments using a template independent polymerase, followed by ligation to one or more adapters each having a thymine at a 3′ end. In some embodiments, adapter oligonucleotides can be joined to blunt end double-stranded nucleic acid fragments which have been modified by extension of the 3′ end with one or more nucleotides followed by 5′ phosphorylation. In some cases, extension of the 3′ end can be performed with a polymerase such as for example Klenow polymerase or any of the suitable polymerases provided herein, or by use of a terminal deoxynucleotide transferase, in the presence of one or more dNTPs in a suitable buffer containing magnesium. In some embodiments, nucleic acid fragments having blunt ends can be joined to one or more adapters comprising a blunt end. Phosphorylation of 5′ ends of nucleic acid fragments can be performed for example with T4 polynucleotide kinase in a suitable buffer containing ATP and magnesium. The fragmented nucleic acid molecules may optionally be treated to dephosphorylate 5′ ends or 3′ ends, for example, by using enzymes known in the art, such as phosphatases.

[0047] In some embodiments, appending the adaptor to the nucleic acid fragments generated by methods described herein can be achieved using a ligation reaction or a priming reaction. In some embodiments, appendage of an adaptor to the nucleic acid fragments comprises ligation. In some embodiments, ligation of the adaptor to the nucleic acid fragments can be following end repair of the nucleic acid fragments. In another embodiment, the ligation of the adaptor to the nucleic acid fragments can be following generation of the nucleic acid fragments without end repair of the nucleic acid fragments. The adaptor can be any type of adaptor known in the art including, but not limited to, conventional duplex or double stranded adaptors in which the adaptor comprises two complementary strands. In some embodiments, the adaptor can be a double stranded DNA adaptor. In some embodiments, the adaptor can be an oligonucleotide of known sequence and, thus, allow generation and/or use of sequence specific primers for amplification and/or sequencing of any polynucleotides to which the adaptor is appended or attached. In some embodiments, the adaptor can be a conventional duplex adaptor, wherein the adaptor comprises sequence well known in the art. In a some embodiments, the adaptor can be appended to the nucleic acid fragments generated by the methods described herein in multiple orientations. In some embodiment, the methods described herein can involve the use of a duplex adaptor comprising double stranded DNA of known sequence that is blunt ended and can bind to the double stranded nucleic acid fragments generated by the methods described herein in one of two orientations. In some embodiments, the adaptor can be ligated to each of the nucleic acid fragments such that each of the nucleic acid fragments comprises the same adaptor. In other words, each of the nucleic acid fragments comprises a common adaptor. In another embodiment, an adaptor can be appended or ligated to a library of nucleic acid fragments generated by the methods described herein such that each nucleic acid fragment in the library of nucleic acid

fragments comprises the adaptor ligated to one or both ends. In another embodiment, more than one adaptor can be appended or ligated to a library of nucleic acid fragments generated by the methods described herein. The multiple adaptors may occur adjacent to one another, spaced intermittently, or at opposite ends of the nucleic acid fragments. In some embodiments, the adaptor can be ligated or appended to the 5′ and/or 3′ ends of the nucleic acid fragments generated by the methods described herein. The adaptor can comprise two strands wherein each strand comprises a free 3′ hydroxyl group but neither strand comprises a free 5′ phosphate. In some embodiments, the free 3′ hydroxyl group on each strand of the adaptor can be ligated to a free 5′ phosphate present on either end of the nucleic acid fragments of the present invention. In this embodiment, the adaptor comprises a ligation strand and a non-ligation strand whereby the ligation strand can be ligated to the 5′ phosphate on either end of the nucleic acid fragment while a nick or gap can be present between the non-ligation strand of the adaptor and the 3′ hydroxyl on either end of the nucleic acid fragment. In some embodiments, the nick or gap can be filled in by performing a gap repair reaction. In some embodiments, the gap repair can be performed with a DNA dependent DNA polymerase with strand displacement activity. In some embodiments, the gap repair can be performed using a DNA-dependent DNA polymerase with weak or no strand displacement activity. In some embodiments, the ligation strand of the adaptor can serve as the template for the gap repair or fill-in reaction. The gap repair or fill-in reaction may comprise an extension reaction wherein the ligation strand of the adaptor serves as a template and leads to the generation of nucleic acid fragments with complementary termini or ends. In some embodiments, the gap repair can be performed using Taq DNA polymerase. In some embodiments, the ligation of the first adaptor to the nucleic acid fragments generated by the

methods described herein may not be followed by gap repair. The nucleic acid fragments may comprise the adaptor sequence ligated only at the 5′ end of each strand.

[0048] Ligation and, optionally gap repair, of the adaptor to the nucleic acid fragments generates an adaptor-nucleic acid fragment complex. In some embodiments, the adaptor-nucleic acid fragment complex can be denatured. Denaturation can be achieved using any of the methods known in the art including, but not limited to, physical, thermal, and/or chemical denaturation. In some embodiments, denaturation can be achieved using thermal or heat denaturation. In some embodiments, denaturation of the adaptor-nucleic acid fragment complex generates single stranded nucleic acid fragments comprising the adaptor sequence at only the 5′end of the nucleic acid fragments. In another embodiment, denaturation of the first adaptor-nucleic acid fragment complex generates single stranded nucleic acid fragments comprising adaptor sequence at both the 5′end and 3′end of the nucleic acid fragments.

Methods of Amplification

[0049] The methods, compositions and kits described herein can be useful to generate amplification-ready products directly from a nucleic acid source for downstream applications such as next generation sequencing, as well as generation of libraries with enriched population of sequence regions of interest. In some embodiments, the adapter-nucleic fragment ligated products, e.g., from the ligation of an adaptor to a 5' end of each nucleic acid fragment of a plurality of nucleic acid fragments from one or more samples, is amplified.

[0050] Methods of amplification are well known in the art. In some embodiments, the amplification is exponential, e.g. in the enzymatic amplification of specific double stranded sequences of DNA by a polymerase chain reaction (PCR). In other embodiments the amplification method is linear. In other embodiments the amplification method is isothermal. In

some embodiments, the amplification is exponential, e.g. in the enzymatic amplification of specific double stranded sequences of DNA by a polymerase chain reaction (PCR).

[0051] Suitable amplification reactions can be exponential or isothermal and can include any DNA amplification reaction, including but not limited to polymerase chain reaction (PCR), strand displacement amplification (SDA), linear amplification, multiple displacement amplification (MDA), rolling circle amplification (RCA), single primer isothermal amplification (SPIA, see e.g. U.S. Pat. No. 6,251,639), Ribo-SPIA, or a combination thereof. In some cases, the amplification methods for providing the template nucleic acid may be performed under limiting conditions such that only a few rounds of amplification (e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30 etc.), such as for example as is commonly done for cDNA generation. The number of rounds of amplification can be about 1-30, 1-20, 1-15, 1-10, 5-30, 10-30, 15-30, 20-30, 10-30, 15-30, 20-30, or 25-30.

[0052] PCR is an in vitro amplification procedure based on repeated cycles of denaturation, oligonucleotide primer annealing, and primer extension by thermophilic template dependent polynucleotide polymerase, resulting in the exponential increase in copies of the desired sequence of the polynucleotide analyte flanked by the primers. The two different PCR primers, which anneal to opposite strands of the DNA, are positioned so that the polymerase catalyzed extension product of one primer can serve as a template strand for the other, leading to the accumulation of a discrete double stranded fragment whose length is defined by the distance between the 5′ ends of the oligonucleotide primers.

[0053] LCR uses a ligase enzyme to join pairs of preformed nucleic acid probes. The probes hybridize with each complementary strand of the nucleic acid analyte, if present, and ligase is

employed to bind each pair of probes together resulting in two templates that can serve in the next cycle to reiterate the particular nucleic acid sequence.

[0054] SDA (Westin et a 2000, Nature Biotechnology, 18, 199-202; Walker et al 1992, Nucleic Acids Research, 20, 7, 1691-1696), is an isothermal amplification technique based upon the ability of a restriction endonuclease such as HincII or BsoBI to nick the unmodified strand of a hemiphosphorothioate form of its recognition site, and the ability of an exonuclease deficient DNA polymerase such as Klenow exo minus polymerase, or Bst polymerase, to extend the 3′-end at the nick and displace the downstream DNA strand. Exponential amplification results from coupling sense and antisense reactions in which strands displaced from a sense reaction serve as targets for an antisense reaction and vice versa.

[0055] Some aspects of the invention utilize linear amplification of nucleic acids or polynucleotides. Linear amplification generally refers to a method that involves the formation of one or more copies of the complement of only one strand of a nucleic acid or polynucleotide molecule, usually a nucleic acid or polynucleotide analyte. Thus, the primary difference between linear amplification and exponential amplification is that in the latter process, the product serves as substrate for the formation of more product, whereas in the former process the starting sequence is the substrate for the formation of product but the product of the reaction, i.e. the replication of the starting template, is not a substrate for generation of products. In linear amplification the amount of product formed increases as a linear function of time as opposed to exponential amplification where the amount of product formed is an exponential function of time.

Downstream Applications

[0056] One aspect of the present invention is that the methods and compositions disclosed herein can be efficiently and cost-effectively utilized for downstream analyses, such as in next generation sequencing or hybridization platforms, with minimal loss of biological material of interest. The methods of the present invention can also be used in the analysis of genetic information of selective genomic regions of interest (e.g., analysis of SNPs or other disease markers) as well as genomic regions which may interact with the selective region of interest. The methods of the present invention may further be used in the analysis of copy number variation as well as differential expression.

Sequencing

[0057] In some embodiments, a population of sequencing reads is generated from the amplified adapter-nucleic fragment ligated products. In some embodiments, a sequencing read comprises an index read, which comprises the sequence of an indexing site. In some embodiments, an index read comprises the sequence of an indexing site and the sequence of an identifier site. For example, the indexing site is sequenced with the identifier site. In some embodiments, the index read does not include the sequence of an identifier site. For example, the indexing site is not sequenced with the identifier site. In some embodiments, a sequencing read comprises the target sequence. In some embodiments, a sequencing read comprises a target sequence and an identifier sequence. For example, the target sequence is sequenced with the identifier site. In some embodiments, the target sequence is not sequenced with the identifier site.

[0058] The methods of the present invention are useful for sequencing by the method commercialized by Illumina, as described U.S. Pat Nos. 5,750,341; 6,306,597; and 5,969,119.

[0059] In general, double stranded fragment polynucleotides can be prepared by the methods of the present invention to produce amplified nucleic acid sequences tagged at one (e.g., (A)/(A′) or both ends (e.g., (A)/(A′) and (C)/(C′)). In some cases, single stranded nucleic acid tagged at one or both ends is amplified by the methods of the present invention (e.g., by SPIA or linear PCR).The resulting nucleic acid is then denatured and the single-stranded amplified polynucleotides are randomly attached to the inside surface of flow-cell channels. Unlabeled nucleotides are added to initiate solid-phase bridge amplification to produce dense clusters of double-stranded DNA. To initiate the first base sequencing cycle, four labeled reversible terminators, primers, and DNA polymerase are added. After laser excitation, fluorescence from each cluster on the flow cell is imaged. The identity of the first base for each cluster is then recorded. Cycles of sequencing are performed to determine the fragment sequence one base at a time.

[0060] In some embodiments, the methods of the invention are useful for preparing target polynucleotides for sequencing by the sequencing by ligation methods commercialized by Applied Biosystems (e.g., SOLiD sequencing). In other embodiments, the methods are useful for preparing target polynucleotides for sequencing by synthesis using the methods commercialized by 454/Roche Life Sciences, including but not limited to the methods and apparatus described in Margulies et al., *Nature (*2005) 437:376-380 (2005); and U.S. Pat. Nos. 7,244,559; 7,335,762; 7,211,390; 7,244,567; 7,264,929; and 7,323,305. In other embodiments, the methods are useful for preparing target polynucleotide(s) for sequencing by the methods commercialized by Helicos BioSciences Corporation (Cambridge, Mass.) as described in U.S. application Ser. No. 11/167,046, and U.S. Pat. Nos. 7,501,245; 7,491,498; 7,276,720; and in U.S. Patent Application Publication Nos. US20090061439; US20080087826; US20060286566; US20060024711;

US20060024678; US20080213770; and US20080103058. In other embodiments, the methods are useful for preparing target polynucleotide(s) for sequencing by the methods commercialized by Pacific Biosciences as described in U.S. Pat. Nos. 7,462,452; 7,476,504; 7,405,281; 7,170,050; 7,462,468; 7,476,503; 7,315,019; 7,302,146; 7,313,308; and US Application Publication Nos. US20090029385; US20090068655; US20090024331; and US20080206764.

[0061] An example of a sequencing technique that can be used in the methods of the provided invention is semiconductor sequencing provided by Ion Torrent (e.g., using the Ion Personal Genome Machine (PGM)). Ion Torrent technology can use a semiconductor chip with multiple layers, e.g., a layer with micro-machined wells, an ion-sensitive layer, and an ion sensor layer. Nucleic acids can be introduced into the wells, e.g., a clonal population of single nucleic can be attached to a single head, and the bead can be introduced into a well. To initiate sequencing of the nucleic acids on the beads, one type of deoxyribonucleotide (e.g., dATP, dCTP, dGTP, or dTTP) can be introduced into the wells. When one or more nucleotides are incorporated by DNA polymerase, protons (hydrogen ions) are released in the well, which can be detected by the ion sensor. The semiconductor chip can then be washed and the process can be repeated with a different deoxyribonucleotide. A plurality of nucleic acids can be sequenced in the wells of a semiconductor chip. The semiconductor chip can comprise chemical-sensitive field effect transistor (chemFET) arrays to sequence DNA (for example, as described in U.S. Patent Application Publication No. 20090026082). Incorporation of one or more triphosphates into a new nucleic acid strand at the 3' end of the sequencing primer can be detected by a change in current by a chemFET. An array can have multiple chemFET sensors.

[0062] Another example of a sequencing technique that can be used in the methods of the provided invention is nanopore sequencing (see e.g., Soni G V and Meller A. (2007) *Clin Chem*

53: 1996-2001). A nanopore can be a small hole of the order of 1 nanometer in diameter. Immersion of a nanopore in a conducting fluid and application of a potential across it can result in a slight electrical current due to conduction of ions through the nanopore. The amount of current that flows is sensitive to the size of the nanopore. As a DNA molecule passes through a nanopore, each nucleotide on the DNA molecule obstructs the nanopore to a different degree. Thus, the change in the current passing through the nanopore as the DNA molecule passes through the nanopore can represent a reading of the DNA sequence.

Data Analysis

[0063] In some embodiments, the sequence reads are used in the analysis of genetic information of selective genomic regions of interest as well as genomic regions which may interact with the selective region of interest. Amplification methods as disclosed herein can be used in the devices, kits, and methods known to the art for genetic analysis, such as, but not limited to those found in U.S. Pat. Nos. 6,449,562, 6,287,766, 7,361,468, 7,414,117, 6,225,109, and 6,110,709.

[0064] In some embodiments, the sequencing reads are used to detect duplicate sequencing reads. In some embodiments, a sequencing read is identified as a duplicate sequencing read when it contains an identifier site and target sequence that is the same as another sequencing read from the same population of sequencing reads.

[0065] In some embodiments, duplicate sequencing reads are differentiated from one another as being true duplicates versus apparent or perceived duplicates. Apparent or perceived duplicates may be identified from sequencing libraries and using conventional measures of duplicate reads (i.e., reads were mapped using bowtie), where all reads with the same start and end nucleic acid coordinates were counted as duplicates. True duplicates may be identified from sequencing

libraries having had an identifier site introduced through ligation to differentiate between fragments of DNA that randomly have the same start and end mapping coordinates.

[0066] In some embodiments, the sequence of the identifier sites from the index read of two nucleic acids generated from any dsDNA may have the same start site, as determined from the sequencing reads of the target sequence of the generated nucleic acid fragments. If the identifier site from the index read of the two nucleic acid fragments are not identical, then the target sequence reads were not generated from the same original dsDNA molecule and, therefore, are no true duplicate reads. The ligation of random sequences onto dsDNA molecules, accompanied by the methods of the invention, allow for the identification of true duplicate reads versus apparent or perceived duplicate reads.

[0067] In some embodiments, the sequence of the identifier sites from the index read of two nucleic acid fragments generated from a genomic DNA (gDNA) molecule that have the same start site, which can be determined from the sequencing reads of the target sequence of the nucleic acid fragments is determined. If the identifier site from the index read of the two nucleic acid fragments are not identical, then the target sequence reads were not generated from the same original gDNA molecule and, therefore, are not true duplicate reads. In another embodiment, an identifier site is inserted at the adapter insert junction. The sequence of the identifier site is carried through the library amplification step. The identifier site is the first sequence read during the forward read. As the identifier sequence is not logically present adjacent to naturally occurring sequence, this uniquely identifies the DNA fragment. Therefore, by ligating random sequences onto the original gDNA, the methods of the invention identify true duplicate reads.

[0068] In some embodiments, a duplicate sequencing read is detected and analyzed. Duplicate reads can be filtered using 'samtools rmdup', wherein reads with identical external coordinates

are removed, only one read with highest mapping quality is retained. After filtering, a filtered set of deduplicated reads can be used in any downstream analysis. Conversely, this filtering step can be skipped, and downstream analysis can be done using the unfiltered reads, including duplicates.

[0069] In some embodiments, sequencing reads are generated from a number of samples. In some embodiments, adaptors are ligated to a plurality of nucleic acid fragments from the samples, in which the nucleic acid fragments from each sample has the same indexing site. In some embodiments, a plurality of nucleic acid fragments is generated from a first sample and a second sample and adaptors are ligated to each nucleic acid fragment, in which adaptors ligated to each nucleic acid fragment from the first sample have the same first indexing site and adaptors ligated to nucleic acid fragments from the second sample has the same second indexing site. In some embodiments, the data associated with the nucleic acid fragments (e.g., sequencing reads) are separated based on the indexing site before sequencing reads of the target sequence and/or identifier site are analyzed. In some embodiments, the nucleic acid fragments or data associated with the nucleic acid fragments (e.g., sequencing reads) are separated based on the indexing site before duplicate sequencing reads are analyzed and/or removed.

[0070] In some embodiments, a method disclosed herein identifies or detects one or more true duplicate(s) with increased accuracy as compared to other methods. For example, in some embodiments, a method disclosed herein identify true duplicates (in contrast to identifying apparent or perceived duplicates) with increased accuracy as compared to other methods. The increased resolution and/or accuracy in identifying one or more true duplicate(s) can provide a considerable contribution to the state of the art in more accurately identifying true duplicates. In some embodiments, a method disclosed herein identifies or detects a true duplicate with

increased efficiency as compared to other methods, such as paired end sequencing. The increase in accuracy, resolution, and/or efficiency in detecting duplicate reads (e.g., true duplicate(s)) can increase confidence in sequencing results, such as for expression and CNV analysis.

Kits

[0071] Any of the compositions described herein may be comprised in a kit. In a non-limiting example, the kit, in a suitable container, comprises: an adaptor or several adaptors, one or more of oligonucleotide primers and reagents for amplification.

[0072] The containers of the kits will generally include at least one vial, test tube, flask, bottle, syringe or other containers, into which a component may be placed, and preferably, suitably aliquotted. Where there is more than one component in the kit, the kit also will generally contain a second, third or other additional container into which the additional components may be separately placed. However, various combinations of components may be comprised in a container.

[0073] When the components of the kit are provided in one or more liquid solutions, the liquid solution can be an aqueous solution. However, the components of the kit may be provided as dried powder(s). When reagents and/or components are provided as a dry powder, the powder can be reconstituted by the addition of a suitable solvent.

[0074] A kit may include instructions for employing the kit components as well the use of any other reagent not included in the kit. Instructions may include variations that can be implemented.

[0075] In some embodiments, the invention provides kits containing any one or more of the elements disclosed in the above methods and compositions. In some embodiments, a kit comprises a composition of the invention, in one or more containers. In some embodiments, the

invention provides kits comprising adapters, primers, and/or other oligonucleotides described herein. In some embodiments, the kit further comprises one or more of: (a) a DNA ligase, (b) a DNA-dependent DNA polymerase, (c) an RNA-dependent DNA polymerase, (d) a forward adapter (e) one or more oligonucleotides comprising reverse adaptor sequence and (f) one or more buffers suitable for one or more of the elements contained in said kit. The adapters, primers, other oligonucleotides, and reagents can be, without limitation, any of those described above. Elements of the kit can further be provided, without limitation, in any of the amounts and/or combinations (such as in the same kit or same container) described above. The kits may further comprise additional agents, such as those described above, for use according to the methods of the invention. For example, the kit can comprise a first forward adaptor that is a partial duplex adaptor as described herein, a second forward adapter, and a nucleic acid modifying enzyme specific for a restriction and/or cleavage site present in the first forward adaptor. The kit elements can be provided in any suitable container, including but not limited to test tubes, vials, flasks, bottles, ampules, syringes, or the like. The agents can be provided in a form that may be directly used in the methods of the invention, or in a form that requires preparation prior to use, such as in the reconstitution of lyophilized agents. Agents may be provided in aliquots for single-use or as stocks from which multiple uses, such as in a number of reaction, may be obtained.

[0076] In some embodiments, the kit comprises a plurality of adaptor oligonucleotides, wherein each of the adaptor oligonucleotides comprises at least one of a plurality of identifier site sequences, wherein each identifier site sequence of the plurality of identifier site sequences differs from every other identifier site sequence in said plurality of identifier site sequences at at least three nucleotide positions, and instructions for using the same. Adapters comprising

different identifier site sequences can be supplied individually or in combination with one or more additional adapters having a different identifier site sequence. In some embodiments, the kit can comprises a plurality of adapter oligonucleotides.

EXAMPLES

Example 1: Identification of Duplicate Sequencing Reads with NuGEN Ovation Target Enrichment Library System

[0077] Sample Description: 100ng of DNA from a human HapMap sample (NA19238) was fragmented to approximately 500 base pair in length by sonication with a Covaris system (Covaris, Inc., Woburn, MA). The resulting DNA was treated with end repair enzyme mix NuGEN R01280 and R01439 (NuGEN Technologies, Inc., San Carlos, CA) according to supplier's recommendation to produce blunt ended DNA fragments.

[0078] Library Generation, Enrichment, and Identifier Site Incorporation: An oligonucleotide with segments, from 5' to 3' of the top strand, 1) an Illumina indexing read priming site such as AGAGCACACGTCTGAACTCCAGTCAC (SEQ ID NO:2), 2) an indexing site, 3) an identifier site having a random 6 base sequence and, 4) a sequence compatible with an Illumina forward sequencing priming site such as TCTTTCCCTACACGACGCTCTTCCGATCT (SEQ ID NO:3) , was annealed to a second oligonucleotide to form a partially double stranded DNA adapter. Five uM of these adapters were ligated onto the end-repaired DNA using Ligase and Ligase reaction buffer from NuGEN's Ovation Ultralow Library System (NuGEN Technologies, Inc., San Carlos, CA) according to supplier's recommendations. Following 30 minutes of incubation at 25°C, the reaction mixture was diluted with water, 0.8X volume of Ampure XP magnetic beads (Agencourt Biosciences Corporation, A Beckman Coulter Company, Beverly, MA) was added and the solution thoroughly mixed. The beads were collected, washed and the ligated

DNA fragments eluted according to manufacturer's recommendations. A pool of targeting probes was annealed to the eluted DNA fragments by initially heating the solution to 95°C then slowly cooling the mixture from 80°C to 60°C by 0.6 degrees/minute. Targeting probes that were specifically annealed were extended with Taq DNA polymerase (New England Biolabs, Inc., Ipswich, MA) according to the manufacturer's protocols. Following extension, the DNA fragments were collected on Agencourt magnetic beads, washed and eluted according to manufacturer's recommendations. These libraries were enriched by 30 cycles of PCR using NuGEN library enrichment primers (Ovation Target Enrichment Library System, NuGEN Technologies, Inc., San Carlos, CA) that also contain the Illumina flow cell sequences (Illumina Inc., San Diego, CA) according to supplier's recommendation.

[0079] The resulting libraries were quantitated by qPCR using a kit provided by KAPA, diluted to 2nM and applied to an Illumina MiSeq DNA Sequencer (Illumina Inc., San Diego, CA). The following series was run: 36 base first read, 14 base second read, and 24 base third read.

[0080] *Data Analysis:* The sequencer output was processed in accordance with manufacturer's recommendation. In order to analyze the data, the indexing read was split into two files. The first file contained the first 8 bases of the indexing reads and is utilized as the library index file for standard library parsing. The other file contains only the random bases and is set aside for further sequence parsing.

[0081] Following our data analysis pipeline of sequence alignment with bowtie aligner (Langmead B. et al., Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* 2009, 10:R2.), duplicate reads were identified by their genomic start positions. At this point, sequencing reads that started at the same genomic position were checked against the random bases file to see if they have the same or different set of random

bases that were ligated to them. Where two sequences with the same starting genomic coordinate had the same set of random bases, they were considered to have come from the same initial DNA ligation event, regardless of which Ovation Target Enrichment targeting probe was used to generate the sequence fragment in question. These sequences, therefore, did not provide unique information about the starting genomic DNA and are considered as one sequencing read for the purpose of variant analysis. Two sequence reads with the same starting genomic coordinate that have different random bases were derived from unique ligation events and were considered to both be valid sequencing reads for the purpose of variant identification. **Figure 5** provides the analysis results demonstrating the identification of the duplicate reads. The use of an identifier site allowed for the determination of the number of true duplications versus the number of apparent or perceived duplicates.

[0082] If the sequences of two libraries are identical, their duplication status is unknown since this could occur by chance in any library. If an identifier sequence is used in combination with the library reads, the status can be determined (duplicates if identical, distinct if different). With the SPET system, one end is common so the probability of two libraries having identical ends increases. These would appear to be duplicate sequences and their true status can be determined by looking at an identifier sequence.

[0083] Over the sampling of all randomly selected reads, the use of identifier sites provided an increased resolution on the presence of true duplicates. When evaluating two million random reads, the apparent duplicates were found to comprise 39% of all reads. However, the true duplicates, identified through the use of an identifier site, were found to comprise only 26% of all reads. The methods employing the use of an identifier site were found to considerably increase the resolution of the true number of duplicates within the pool of reads.

Example 2: Removal of Duplicate Sequencing Reads with 8 Base Identifier Site

[0084] In a standard RNA sequencing library, adaptors are ligated to the ends of double stranded cDNA. These adaptors contain universal sequences that allow for PCR amplification and sequencing on high throughput sequencing machines. The adaptors are synthesized with a large population of additional sequences, in which each additional sequence is an identifier site, at the ligation end. The identifier sites are present at the junction between the adaptor and the cDNA. The sequence read starts with the identifier site and follows with the cDNA sequence.

[0085] This pool of identifier sites is used for the detection of PCR duplicates, as PCR duplicates will contain the identical identifier site, whereas two different cDNA molecules will ligate to two different adaptors containing two different identifier sites. This identifier site is designed as eight random bases introduced onto the end of the adaptor. Sequence reads from libraries made with such adaptors contain the 8 bases of the identifier site followed by the cDNA sequence. Standard PCR duplicate removal software such as PICARD, Markduplicates, and/or SAMtools rmdup is used to identify and remove the PCR duplicates, leaving behind for analysis any instances of multiple cDNA fragments that happen to have the same sequence.

Example 3: Removal of Duplicate Sequencing Reads with Mixture of Random 1-8 Base Identifier Sites

[0086] In a standard RNA sequencing library, adaptors are ligated to the ends of double stranded cDNA. These adaptors contain universal sequences that allow for PCR amplification and sequencing on high throughput sequencing machines. The adaptors are synthesized with a large population of additional sequences, in which each additional sequence is an identifier site, at the

ligation end. The identifier sites are present at the junction between the adaptor and the cDNA. The sequence read starts with the identifier site and follows with the cDNA sequence.

[0087] This pool of identifier sites is used for the detection of PCR duplicates, as PCR duplicates will contain the identical identifier site, whereas two different cDNA molecules will ligate to two different adaptors containing two different identifier sites.

[0088] One to eight random bases are introduced onto the end of the adaptor. Sequence reads from libraries made with such adaptors contain between 1 and 8 bases of the identifier site followed by the cDNA sequence. Standard PCR duplicate removal software such as PICARD, Markduplicates, and/or SAMtools rndup is used to identify and remove the PCR duplicates, leaving behind for analysis any instances of multiple cDNA fragments that happen to have the same sequence.

Example 4: Removal of Duplicate Sequencing Reads with Mixture of 96 Defined 6-Base Identifier Sites

[0089] In a standard RNA sequencing library, adaptors are ligated to the ends of double stranded cDNA. These adaptors contain universal sequences that allow for PCR amplification and sequencing on high throughput sequencing machines. The adaptors are synthesized with a large population of additional sequences, in which each additional sequence is an identifier site, at the ligation end. The identifier sites are present at the junction between the adaptor and the cDNA. The sequence read starts with the identifier site and follows with the cDNA sequence.

[0090] This pool of identifier sites is used for the detection of PCR duplicates, as PCR duplicates will contain the identical identifier site, whereas two different cDNA molecules will ligate to two different adaptors containing two different identifier sites.

[0091] A mixture of 96 defined six-bases sequences are introduced onto the end of the adaptors. Thus, each six-base sequence is an identifier site. Sequence reads from libraries made with such adaptors contain one of the 96 six-base identifier site followed by the cDNA sequence. Standard PCR duplicate removal software such as PICARD, Markduplicates, and/or SAMtools rndup is used to identify and remove the PCR duplicates, leaving behind for analysis any instances of multiple cDNA fragments that happen to have the same sequence.

Example 5: Identification of Duplicate Sequencing Reads in Determining mRNA Expression Levels

[0092] Sample Description: Total RNA is extracted from tumor and normal adjacent tissue for the purpose of finding differences in expression levels of transcripts between the two sample types. 100ng of each sample is converted into cDNA using the USP primers, reaction buffer, and Reverse Transcriptase provided in NuGEN's Encore Complete Library System (NuGEN Technologies, Inc., San Carlos, CA) according to the supplier's recommendations. This is followed by second strand synthesis, again using materials provided in the kit according to recommendations. Double stranded cDNA was prepared using the SuperScript® Double-Stranded cDNA Synthesis Kit from Life Technologies (Carlsbad, CA) according the manufacturer's instructions. DNA was sheared with a Covaris S-series device (Covaris, Inc., Woburn, MA) using the 200 bp sonication protocol provided with the instrument (10% duty cycle, 200 cycles/burst, 5 intensity, 180 seconds). DNA was treated with 1.5 uL 10X Blunting Buffer, 0.5 uL Blunting Enzyme (New England Biolabs, Inc., Ipswich, MA; p/n E1201) and 1.2 uL of 2.5mM of each dNTP mix in a total volume of 15 uL for 30 minutes at 25°C followed by 10 minutes at 70°C.

[0093] *Library Generation*: The DNA fragments were then subjected to end repair using end repair buffers and enzymes provided in NuGEN's Ovation Ultralow Library System (NuGEN Technologies, Inc., San Carlos, CA).

[0094] Forward adaptor 5'

[0095] AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCC GATCTNNNNNNNN (SEQ ID NO:4),

[0096] Reverse adaptor 1) 5'

[0097] CAAGCAGAAGACGGCATACGAGATTCCCTTGTGACTGGAGTTCAGACGTGTG CTCTTCCGATCTNNNNNNNN (SEQ ID NO:5),

[0098] Reverse adaptor 2) 5'

[0099] CAAGCAGAAGACGGCATACGAGATTGAAGGGTGACTGGAGTTCAGACGTGT GCTCTTCCGATCNNNNNNNN (SEQ ID NO:6), and

[00100] Common partner 5' NNNNNNNNAGATCGGAAGAGC (SEQ ID NO:7) were all ordered from IDT (Integrated DNA Technologies, Coralville, IA). The reverse adaptors each contain a unique identifier (underlined) that enables libraries made with these adapters to be distinguished. In this case N represents an equimolar mixture of A, C, G, and T. A mixture of 5uM forward, 5uM reverse, and 10uM common in 10mM MgCl2, 50mM Tris pH 8 was heated to 95C for 5 minutes, then cooled to 20C. Adaptor ligation was performed by addition of 4.5 uL water, 3 uL Adaptor mix (prepared above), 6 uL 5X NEBNext Quick Ligation Reaction Buffer and 1.5 uL Quick T4 DNA Ligase (New England Biolabs, Inc., Ipswich, MA; p/n E6056), followed by incubation for 30 minutes at 25°C followed by 10 minutes at 70°C. Ligation products were purified by adding 70uL water and 80 uL of Ampure XP beads (Agencourt Genomics), washing twice with 70% ethanol and eluting with 20 uL of 10mM Tris pH 8.0.

Library products were amplified in a 50uL PCR containing 0.5uM each of primer (5' AATGATACGGCGACCACCGA (SEQ ID NO:8) , and 5' CAAGCAGAAGACGGCATACGA (SEQ ID NO:9), 10 mM Tris-HCl, pH 8.3, 50 mM KCl, 2 mM MgC12, 0.2mM each dNTP, and 1 unit Taq polymerase. The reaction was cycled 15 times under the conditions 95C for 15 seconds, 60C for 1 minute. PCR products were purified with 1 volume of Ampure XP beads (Agencourt Biosciences Corporation, A Beckman Coulter Company, Beverly, MA) as described above. The library was analyzed by HS DNA Bioanalyzer (Agilent Technologies, Santa Clara, CA) and quantitated with the KAPA Library Quantification Kit (KAPA Biosystems, Wilmington, MA; p/n KK4835) according to the supplied instructions. The resulting libraries are combined and are compatible with standard TruSeq single end or paired Illumina sequencing protocols for GAIIx, MiSeq, or Hi Seq sequencing instruments (Illumina Inc., San Diego, CA). The following series is run; 50 base first read, 6 base second read. A third read is not required for counting purposes or duplication analysis.

[00101] Data Analysis: The sequencer output was processed in accordance with manufacturer's recommendation. The 6 bases of the index read is used for standard library parsing, separating the data files from the two sample types. The first 50 bases of the target sequence read are compared to each other. Any read that has identical sequence to any other read is identified as a duplicate and removed from the population, thus, only a single copy is retained within the file. Once the duplicate reads have been removed, 8 bases are trimmed from each read. The trimmed reads are aligned to a reference genome. Differential expression is then determined by comparing FPKM (fragments per kilobase per million reads) values between libraries utilizing scripts such as cufflinks or cuffdiff (Trapnell et al. 2010, Nature Biotechnology, 28, 511-515; Trapnell et al. 2013, Nature Biotechnology, 31, 46-53).

Example 6: Sequencing of the Identifier Site With the Target Sequence in Paired End Reads

[00102] The Ovation Library System for Low Complexity Samples (NuGEN Technologies, Inc., San Carlos, CA) was used to generate four libraries, each from a single amplicon, following manufacturer's protocol. Purified libraries were mixed and sequenced as a multiplex on the Illumina MiSeq (Illumina Inc., San Diego, CA) to produce 125nt forward, 8nt index 1, 8nt index 2, and 25nt reverse reads. Because all amplicon reads start and end at the same sequence coordinates, the traditional method of marking library PCR duplicates (marking reads that start and end at the same genomic coordinates as duplicates) cannot be used. Instead, the 0-8nt of random sequence contained in the adaptors ligated to the amplicon were treated as an identifier sequence and used to mark duplicates. Any paired end reads that shared the same length and sequence of these random bases with any other paired end read was called a duplicate. The table below shows the results of this duplicate marking.

Table 1.

Library	Reads	Duplicate Reads	Reads from independent molecules
1	454249	376797	77452
2	439367	364001	75366
3	317760	253057	64703
4	476572	398380	78192

[00103] Table 1 demonstrates the accuracy of the method used to differentiate between duplicate reads and reads from truly independent molecules. The population of reads from independent molecules, depicted in the final column of Table 1, represent sequences from independent amplicon molecules used in generating the libraries.

Example 7: Sequencing of the Identifier Site With the Target Sequence in Reduced Representation Bisulfite (RRBS) Libraries

[00104] Reduced representation bisulfite (RRBS) libraries of the human genome are generated through the complete restriction enzyme digestion of 100ng input sample, followed by selection for short fragments. The resulting pool of fragments are ligated to adaptor sequences comprising an indexing site and an identifier site. The identifier sites comprise either 6 or 8 random nucleotides. The sequences are then sequenced to identify the identifier sites, thus revealing the true number of duplicates in the pool. In the absence of identifier sites, the number of apparent or perceived duplicates are greater than the number of true duplicates. The inclusion of an identifier site results in the identification of the number of true duplicates, as compared to the larger number of apparent or perceived duplicates.

[00105] Unless defined otherwise, all technical and scientific terms herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. Although any methods and materials, similar or equivalent to those described herein, can be used in the practice or testing of the present invention, the preferred methods and materials are described herein. All publications, patents, and patent publications cited are incorporated by reference herein in their entirety for all purposes.

[00106] The publications discussed herein are provided solely for their disclosure prior to the filing date of the present application. Nothing herein is to be construed as an admission that the present invention is not entitled to antedate such publication by virtue of prior invention.

[00107] All references, articles, publications, patents, patent publications, and patent applications cited herein are incorporated by reference in their entireties for all purposes. However, mention of any reference, article, publication, patent, patent publication, and

patent application cited herein is not, and should not be taken as an acknowledgment or any form of suggestion that they constitute valid prior art or form part of the common general knowledge in any country in the world.

[00108] While the invention has been described in connection with specific embodiments thereof, it will be understood that it is capable of further modifications and this application is intended to cover any variations, uses, or adaptations of the invention following, in general, the principles of the invention and including such departures from the present disclosure as come within known or customary practice within the art to which the invention pertains and as may be applied to the essential features hereinbefore set forth and as follows in the scope of the appended claims.

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

APPLICANT:	Tecan Genomics, Inc.	ART UNIT:	Not Yet Assigned
SERIAL NUMBER:	Continuation of US Serial No. 16/743,604	CONF. NO.:	Not Yet Assigned
FILING DATE:	Herewith	EXAMINER:	Not Yet Assigned
TITLE:	COMPOSITIONS AND METHODS FOR IDENTIFICATION OF A DUPLICATE SEQUENCING READ		

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

FILED ELECTRONICALLY

STATEMENT OF COMPLIANCE UNDER 37 C.F.R. §1.825

In accordance with 37 C.F.R. §1.825, I hereby certify that the content of the Sequence Listing submitted herewith as an ASCII text file named NUGN-001-04US-SeqList-ST25.txt does not introduce any new matter into the subject application.

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 any such willful false statement may jeopardize the validity of the application or any patent issued therefrom.

Dated: July 21, 2021

Respectfully submitted,
BROWN RUDNICK LLP

BROWN RUDNICK LLP
One Financial Center
Boston, MA 02111
Tel: 212-209-4988

/Robert B. Ruh III/
Robert B. Ruh III, Reg. No. 73,996

DocCode – SEQ.TXT

SCORE Placeholder Sheet for IFW Content

Application Number: 17381888 Document Date: 07/21/2021

The presence of this form in the IFW record indicates that the following document type was received in electronic format on the date identified above. This content is stored in the SCORE database.

Since this was an electronic submission, there is no physical artifact folder, no artifact folder is recorded in PALM, and no paper documents or physical media exist. The TIFF images in the IFW record were created from the original documents that are stored in SCORE.

- Sequence Listing

At the time of document entry (noted above):

- USPTO employees may access SCORE content via DAV or via the SCORE web page.
- External customers may access SCORE content via PAIR using the Supplemental Content tab.

Form Revision Date: March 1, 2019

```
============================================================
Sequence Listing was accepted.
If you need help call the Patent Electronic Business Center at (866)
217-9197 (toll free).
Reviewer: Anjum, Durreshwar
Timestamp: [year=2021; month=7; day=22; hr=8; min=35; sec=35; ms=227; ]
============================================================
```

Validated By CRFValidator v 1.0.5

Application No: 17381888 Version No: 1.0

Input Set:

Output Set:

Started: 2021-07-21 16:14:54.817
Finished: 2021-07-21 16:14:55.585
Elapsed: 0 hr(s) 0 min(s) 0 sec(s) 768 ms
Total Warnings: 47
Total Errors: 0
No. of SeqIDs Defined: 9
Actual SeqID Count: 9

Error code		Error Description
W	213	Artificial or Unknown found in <213> in SEQ ID (1)
W	447	n or Xaa used, for: SEQID(1) on line number 41
W	447	n or Xaa used, for: SEQID(1) on line number 41
W	447	n or Xaa used, for: SEQID(1) on line number 41
W	447	n or Xaa used, for: SEQID(1) on line number 41
W	447	n or Xaa used, for: SEQID(1) on line number 41
W	447	n or Xaa used, for: SEQID(1) on line number 41
W	213	Artificial or Unknown found in <213> in SEQ ID (2)
W	213	Artificial or Unknown found in <213> in SEQ ID (3)
W	213	Artificial or Unknown found in <213> in SEQ ID (4)
W	447	n or Xaa used, for: SEQID(4) on line number 85
W	447	n or Xaa used, for: SEQID(4) on line number 85
W	447	n or Xaa used, for: SEQID(4) on line number 87
W	447	n or Xaa used, for: SEQID(4) on line number 87
W	447	n or Xaa used, for: SEQID(4) on line number 87
W	447	n or Xaa used, for: SEQID(4) on line number 87
W	447	n or Xaa used, for: SEQID(4) on line number 87
W	447	n or Xaa used, for: SEQID(4) on line number 87
W	213	Artificial or Unknown found in <213> in SEQ ID (5)
W	447	n or Xaa used, for: SEQID(5) on line number 107

Input Set:

Output Set:

Started: 2021-07-21 16:14:54.817
Finished: 2021-07-21 16:14:55.585
Elapsed: 0 hr(s) 0 min(s) 0 sec(s) 768 ms
Total Warnings: 47
Total Errors: 0
No. of SeqIDs Defined: 9
Actual SeqID Count: 9

Error code	Error Description
W 447	n or Xaa used, for: SEQID(5) on line number 107
W 447	n or Xaa used, for: SEQID(5) on line number 107
W 447	n or Xaa used, for: SEQID(5) on line number 107
W 447	n or Xaa used, for: SEQID(5) on line number 107
W 447	n or Xaa used, for: SEQID(5) on line number 107 This error has occured more than 20 times, will not be displayed
W 213	Artificial or Unknown found in <213> in SEQ ID (6)
W 213	Artificial or Unknown found in <213> in SEQ ID (7)
W 213	Artificial or Unknown found in <213> in SEQ ID (8)
W 213	Artificial or Unknown found in <213> in SEQ ID (9)

SEQUENCE LISTING

<110> NuGEN Technologies, Inc.

<120> COMPOSITIONS AND METHODS FOR IDENTIFICATION OF A DUPLICATE SEQUENCE READ

<130> NUGN-001/04US 34079/550

<140> US 17/381,888
<141> 2021-07-21

<150> 16/743,604
<151> 2017-01-13

<150> 15/406,002
<151> 2017-01-13

<150> 14/540,917
<151> 2014-11-13

<150> 61/903,826
<151> 2013-11-13

<160> 9

<170> PatentIn version 3.5

<210> 1
<211> 70
<212> DNA
<213> artificial

<220>
<223> Adaptor

<220>
<221> misc_feature
<222> (34)..(39)
<223> n is a, c, g, or t

<400> 1
cagagcacac gtctgaactc cagtcacagt gagnnnnnna cactttccct acacgacgct 60

cttccgatct 70

<210> 2
<211> 26
<212> DNA
<213> artificial

<220>
<223> Priming site

<400> 2

agagcacacg tctgaactcc agtcac 26

<210> 3
<211> 29
<212> DNA
<213> artificial

<220>
<223> forward priming site compliment

<400> 3
tctttcccta cacgacgctc ttccgatct 29

<210> 4
<211> 66
<212> DNA
<213> artificial

<220>
<223> Forward adaptor

<220>
<221> misc_feature
<222> (59)..(66)
<223> n is a, c, g, or t

<400> 4
aatgatacgg cgaccaccga gatctacact ctttccctac acgacgctct tccgatctnn 60

nnnnnn 66

<210> 5
<211> 72
<212> DNA
<213> artificial

<220>
<223> Reverse adaptor

<220>
<221> misc_feature
<222> (65)..(72)
<223> n is a, c, g, or t

<400> 5
caagcagaag acggcatacg agattccctt gtgactggag ttcagacgtg tgctcttccg 60

atctnnnnnn nn 72

<210> 6
<211> 71

<212> DNA
<213> artificial

<220>
<223> Reverse Adaptor

<220>
<221> misc_feature
<222> (64)..(71)
<223> n is a, c, g, or t

<400> 6
caagcagaag acggcatacg agattgaagg gtgactggag ttcagacgtg tgctcttccg 60

atcnnnnnnn n 71

<210> 7
<211> 21
<212> DNA
<213> artificial

<220>
<223> Common partner

<220>
<221> misc_feature
<222> (1)..(8)
<223> n is a, c, g, or t

<400> 7
nnnnnnnnag atcggaagag c 21

<210> 8
<211> 20
<212> DNA
<213> artificial

<220>
<223> Primer

<400> 8
aatgatacgg cgaccaccga 20

<210> 9
<211> 21
<212> DNA
<213> artificial

<220>
<223> Primer

<400> 9

caagcagaag acggcatacg a 21

Electronic Acknowledgement Receipt

EFS ID:	43365410
Application Number:	17381888
International Application Number:	
Confirmation Number:	6188
Title of Invention:	COMPOSITIONS AND METHODS FOR IDENTIFICATION OF A DUPLICATE SEQUENCING READ
First Named Inventor/Applicant Name:	Douglas A. Amorese
Customer Number:	21710
Filer:	Thomas C. Meyers/Sandra Richards
Filer Authorized By:	Thomas C. Meyers
Attorney Docket Number:	NUGN-001/04US 34079/550
Receipt Date:	28-JUL-2021
Filing Date:	
Time Stamp:	15:27:17
Application Type:	Utility under 35 USC 111(a)

Payment information:

Submitted with Payment	no

File Listing:

Document Number	Document Description	File Name	File Size(Bytes)/ Message Digest	Multi Part /.zip	Pages (if appl.)
1		NUGN-001-04US-Amnd-72821.pdf	125699 8faac05c05be990c5d5a2e980da8f2b6716f5d8b	yes	5

	Multipart Description/PDF files in .zip description		
	Document Description	**Start**	**End**
	Preliminary Amendment	1	1
	Claims	2	4
	Applicant Arguments/Remarks Made in an Amendment	5	5
Warnings:			
Information:			
	Total Files Size (in bytes):	125699	

This Acknowledgement Receipt evidences receipt on the noted date by the USPTO of the indicated documents, characterized by the applicant, and including page counts, where applicable. It serves as evidence of receipt similar to a Post Card, as described in MPEP 503.

New Applications Under 35 U.S.C. 111

If a new application is being filed and the application includes the necessary components for a filing date (see 37 CFR 1.53(b)-(d) and MPEP 506), a Filing Receipt (37 CFR 1.54) will be issued in due course and the date shown on this Acknowledgement Receipt will establish the filing date of the application.

National Stage of an International Application under 35 U.S.C. 371

If a timely submission to enter the national stage of an international application is compliant with the conditions of 35 U.S.C. 371 and other applicable requirements a Form PCT/DO/EO/903 indicating acceptance of the application as a national stage submission under 35 U.S.C. 371 will be issued in addition to the Filing Receipt, in due course.

New International Application Filed with the USPTO as a Receiving Office

If a new international application is being filed and the international application includes the necessary components for an international filing date (see PCT Article 11 and MPEP 1810), a Notification of the International Application Number and of the International Filing Date (Form PCT/RO/105) will be issued in due course, subject to prescriptions concerning national security, and the date shown on this Acknowledgement Receipt will establish the international filing date of the application.

REMARKS

By this Preliminary Amendment, Applicant has herein cancelled claims 1-21 and added new claims 22-37. Accordingly, upon entry of these claim amendments, claims 22-37 will be pending, of which claim 22 is the sole independent claim.

Support for new claims 22-37 can be found throughout the originally filed application, including at paragraph [0047] on page 17; paragraph [0064] on page 25; paragraph [0065] on pages 25-26; paragraph [0030] on page 8; paragraph [0029] on page 7; and paragraph [0069] on page 27.

If there are any questions regarding this application, the Office is invited and encouraged to contact Applicant's representatives at the telephone number provided.

Dated: July 28, 2021

Respectfully submitted,
BROWN RUDNICK LLP

BROWN RUDNICK LLP
One Financial Center
Boston, MA 02111
Tel: 617-856-8483

/Thomas C. Meyers/
Thomas C. Meyers, Reg. No. 36,989
Attorney for Applicant
Email: tmeyers@brownrudnick.com

IDT Exhibit 1020.0084

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

APPLICANT:	Tecan Genomics, Inc.	ART UNIT:	Not Yet Assigned
SERIAL NUMBER:	17/381,888	CONF. NO.:	6188
FILING DATE:	July 21, 2021	EXAMINER:	Not Yet Assigned
TITLE:	COMPOSITIONS AND METHODS FOR IDENTIFICATION OF A DUPLICATE SEQUENCING READ		

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

FILED ELECTRONICALLY

PRELIMINARY AMENDMENT

Prior to examination of the above U.S. nonprovisional patent application, please amend the claims and consider the following remarks. Applicant believes that no fees are due in connection with this Preliminary Amendment. However, Applicant authorizes any required fees to be charged to Deposit Account No. 500369.

Amendments to the claims begin on page **2**.

Remarks begin on page **5**.

AMENDMENTS TO THE CLAIMS

Please amend the claims as follows:

1 – 21 (Cancelled).

22. (New) A method for detecting duplicate sequencing reads, the method comprising:

obtaining amplicons each comprising an amplified fragment of a nucleic acid with an appended adaptor, wherein each adaptor comprises an identifier site comprising a plurality of nucleotides unique to the amplified fragment;

sequencing the amplicons to generate sequence reads that include identifier and target sequences; and

identifying sequence reads with identical identifier and target sequences as duplicates.

23. (New) The method of claim 22, further comprising removing a duplicate read from the sequence reads.

24. (New) The method of claim 22, wherein the identifier site is sequenced with the target sequence.

25. (New) The method of claim 22, wherein the identifier site is sequenced separately from the target sequence.

26. (New) The method of claim 22, wherein the amplicons include amplified fragments of cDNA.

IDT Exhibit 1020.0086

27. (New) The method of claim 22, wherein the amplicons include amplified fragments of genomic DNA.

28. (New) The method of claim 22, wherein each adaptor includes a primer binding site.

29. (New) The method of claim 28, wherein the primer binding site is a universal target sequencing primer binding site.

30. (New) The method of claim 22, wherein each unique identifier is at least 6 nucleotides in length.

31. (New) The method of claim 22, wherein the adaptors have been appended to the fragments using a ligation reaction or a priming reaction.

32. (New) The method of claim 22, wherein each unique identifier is at a junction between the adapter and the nucleic fragment.

33. (New) The method of claim 22, further comprising determining the genomic start position for each sequencing read.

34. (New) The method of claim 22, wherein each adaptor further comprises an indexing site.

35. (New) The method of claim 22, wherein each identifier site is at least 2, 3, 4, 5, 6, 7, 8, 9, or 10 nucleotides in length.

IDT Exhibit 1020.0087

36. (New) The method of claim 22, further comprising obtaining a sample comprising the nucleic acid, fragmenting the nucleic acid to yield the fragments of nucleic acid, appending the adaptors to the fragments of nucleic acid, and performing an amplification reaction to obtain the amplicons.

37. (New) The method of claim 22, wherein the amplicons include amplified fragment of a nucleic acid from a first sample with an appended adaptor that includes a first index and amplified fragment of a nucleic acid from a second sample with an appended adaptor that includes a second index, and the method includes separating sequence reads by sample based on the first or second index.

IDT Exhibit 1020.0088

United States Patent and Trademark Office

Office of the Chief Financial Officer

Document Code:WFEE

User :C43106

Sale Adjustment Accounting Date:08/10/2021

Effective Date	Sale Accounting Date	Sale Item Reference Number
07/21/2021	08/10/2021	17381888

Document Number	Fee Code	Fee Code Description	Amount Paid	Payment Method
I202180D59200876	1202	CLAIMS IN EXCESS OF 20	$100.00	DA

United States Patent and Trademark Office

Office of the Chief Financial Officer

Document Code:WFEE

User :C43106

Refund Accounting Date:08/10/2021

Effective Date	Sale Item Reference Number	Refund Total
07/21/2021	17381888	$100.00

Document Number	Fee Code	Fee Code Description	Amount Paid	Payment Method	Account Number
I202180D59200876	1202	CLAIMS IN EXCESS OF 20	$100.00	DA	500369

PATENT APPLICATION FEE DETERMINATION RECORD

Substitute for Form PTO-875

Application or Docket Number: 17/381,888

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			N/A	320
SEARCH FEE (37 CFR 1.16(k), (i), or (m))	N/A	N/A	N/A			N/A	700
EXAMINATION FEE (37 CFR 1.16(o), (p), or (q))	N/A	N/A	N/A			N/A	800
TOTAL CLAIMS (37 CFR 1.16(i))	16 minus 20 =	*			OR	x 100 =	0.00
INDEPENDENT CLAIMS (37 CFR 1.16(h))	1 minus 3 =	*				x 480 =	0.00
APPLICATION SIZE FEE (37 CFR 1.16(s))	If the specification and drawings exceed 100 sheets of paper, the application size fee due is $310 ($155 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).						0.00
MULTIPLE DEPENDENT CLAIM PRESENT (37 CFR 1.16(j))							0.00
* If the difference in column 1 is less than zero, enter "0" in column 2.			TOTAL			TOTAL	1820

APPLICATION AS AMENDED - PART II

		(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($)	OTHER THAN SMALL ENTITY ADDITIONAL FEE($)
AMENDMENT A	Total (37 CFR 1.16(i))	*	Minus	**	=	x =		OR	x =	
	Independent (37 CFR 1.16(h))	*	Minus	***	=	x =		OR	x =	
	Application Size Fee (37 CFR 1.16(s))									
	FIRST PRESENTATION OF MULTIPLE DEPENDENT CLAIM (37 CFR 1.16(j))							OR		
						TOTAL ADD'L FEE		OR	TOTAL ADD'L FEE	

		(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($)	OTHER THAN SMALL ENTITY ADDITIONAL FEE($)
AMENDMENT B	Total (37 CFR 1.16(i))	*	Minus	**	=	x =		OR	x =	
	Independent (37 CFR 1.16(h))	*	Minus	***	=	x =		OR	x =	
	Application Size Fee (37 CFR 1.16(s))									
	FIRST PRESENTATION OF MULTIPLE DEPENDENT CLAIM (37 CFR 1.16(j))							OR		
						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 found in the appropriate box in column 1.

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	GRP ART UNIT	FIL FEE REC'D	ATTY.DOCKET.NO	TOT CLAIMS	IND CLAIMS
17/381,888	07/21/2021		1820	NUGN-001/04US 34079/550	16	1

CONFIRMATION NO. 6188

21710
BROWN RUDNICK LLP
ONE FINANCIAL CENTER
BOSTON, MA 02111

FILING RECEIPT



Date Mailed: 08/11/2021

Receipt is acknowledged of this non-provisional utility patent application. The application will be taken up for examination in due course. Applicant will be notified as to the results of the examination. Any correspondence concerning the application must include the following identification information: the U.S. APPLICATION NUMBER, FILING DATE, NAME OF FIRST INVENTOR, and TITLE OF INVENTION. Fees transmitted by check or draft are subject to collection.

Please verify the accuracy of the data presented on this receipt. If an error is noted on this Filing Receipt, please submit a written request for a corrected Filing Receipt, including a properly marked-up ADS showing the changes with strike-through for deletions and underlining for additions. If you received a "Notice to File Missing Parts" or other Notice requiring a response for this application, please submit any request for correction to this Filing Receipt with your reply to the Notice. When the USPTO processes the reply to the Notice, the USPTO will generate another Filing Receipt incorporating the requested corrections provided that the request is grantable.

Inventor(s)

Douglas A. Amorese, Los Altos, CA;
Jonathan Scolnick, San Francisco, CA;
Ben Schroeder, San Mateo, CA;

Applicant(s)

Tecan Genomics, Inc., Redwood City, CA;

Power of Attorney: None

Domestic Priority data as claimed by applicant

This application is a CON of 16/743,604 01/15/2020 PAT 11098357
which is a CON of 15/406,002 01/13/2017 PAT 10570448
which is a DIV of 14/540,917 11/13/2014 PAT 9546399
which claims benefit of 61/903,826 11/13/2013

Foreign Applications for which priority is claimed (You may be eligible to benefit from the **Patent Prosecution Highway** program at the USPTO. Please see http://www.uspto.gov for more information.) - None.
Foreign application information must be provided in an Application Data Sheet in order to constitute a claim to foreign priority. See 37 CFR 1.55 and 1.76.

Permission to Access Application via Priority Document Exchange: Yes

Permission to Access Search Results: Yes

IDT Exhibit 1020.0092

Applicant may provide or rescind an authorization for access using Form PTO/SB/39 or Form PTO/SB/69 as appropriate.

If Required, Foreign Filing License Granted: 08/10/2021
The country code and number of your priority application, to be used for filing abroad under the Paris Convention, is **US 17/381,888**
Projected Publication Date: 11/18/2021
Non-Publication Request: No
Early Publication Request: No
Title

COMPOSITIONS AND METHODS FOR IDENTIFICATION OF A DUPLICATE SEQUENCING READ

Preliminary Class

Statement under 37 CFR 1.55 or 1.78 for AIA (First Inventor to File) Transition Applications: No

PROTECTING YOUR INVENTION OUTSIDE THE UNITED STATES

Since the rights granted by a U.S. patent extend only throughout the territory of the United States and have no effect in a foreign country, an inventor who wishes patent protection in another country must apply for a patent in a specific country or in regional patent offices. Applicants may wish to consider the filing of an international application under the Patent Cooperation Treaty (PCT). An international (PCT) application generally has the same effect as a regular national patent application in each PCT-member country. The PCT process **simplifies** the filing of patent applications on the same invention in member countries, but **does not result** in a grant of "an international patent" and does not eliminate the need of applicants to file additional documents and fees in countries where patent protection is desired.

Almost every country has its own patent law, and a person desiring a patent in a particular country must make an application for patent in that country in accordance with its particular laws. Since the laws of many countries differ in various respects from the patent law of the United States, applicants are advised to seek guidance from specific foreign countries to ensure that patent rights are not lost prematurely.

Applicants also are advised that in the case of inventions made in the United States, the Director of the USPTO must issue a license before applicants can apply for a patent in a foreign country. The filing of a U.S. patent application serves as a request for a foreign filing license. The application's filing receipt contains further information and guidance as to the status of applicant's license for foreign filing.

Applicants may wish to consult the USPTO booklet, "General Information Concerning Patents" (specifically, the section entitled "Treaties and Foreign Patents") for more information on timeframes and deadlines for filing foreign patent applications. The guide is available either by contacting the USPTO Contact Center at 800-786-9199, or it can be viewed on the USPTO website at http://www.uspto.gov/web/offices/pac/doc/general/index.html.

For information on preventing theft of your intellectual property (patents, trademarks and copyrights), you may wish to consult the U.S. Government website, http://www.stopfakes.gov. Part of a Department of Commerce initiative, this website includes self-help "toolkits" giving innovators guidance on how to protect intellectual property in specific countries such as China, Korea and Mexico. For questions regarding patent enforcement issues, applicants may call the U.S. Government hotline at 1-866-999-HALT (1-866-999-4258).

IDT Exhibit 1020.0093

LICENSE FOR FOREIGN FILING UNDER

Title 35, United States Code, Section 184

Title 37, Code of Federal Regulations, 5.11 & 5.15

GRANTED

The applicant has been granted a license under 35 U.S.C. 184, if the phrase "IF REQUIRED, FOREIGN FILING LICENSE GRANTED" followed by a date appears on this form. Such licenses are issued in all applications where the conditions for issuance of a license have been met, regardless of whether or not a license may be required as set forth in 37 CFR 5.15. The scope and limitations of this license are set forth in 37 CFR 5.15(a) unless an earlier license has been issued under 37 CFR 5.15(b). The license is subject to revocation upon written notification. The date indicated is the effective date of the license, unless an earlier license of similar scope has been granted under 37 CFR 5.13 or 5.14.

This license is to be retained by the licensee and may be used at any time on or after the effective date thereof unless it is revoked. This license is automatically transferred to any related applications(s) filed under 37 CFR 1.53(d). This license is not retroactive.

The grant of a license does not in any way lessen the responsibility of a licensee for the security of the subject matter as imposed by any Government contract or the provisions of existing laws relating to espionage and the national security or the export of technical data. Licensees should apprise themselves of current regulations especially with respect to certain countries, of other agencies, particularly the Office of Defense Trade Controls, Department of State (with respect to Arms, Munitions and Implements of War (22 CFR 121-128)); the Bureau of Industry and Security, Department of Commerce (15 CFR parts 730-774); the Office of Foreign AssetsControl, Department of Treasury (31 CFR Parts 500+) and the Department of Energy.

NOT GRANTED

No license under 35 U.S.C. 184 has been granted at this time, if the phrase "IF REQUIRED, FOREIGN FILING LICENSE GRANTED" DOES NOT appear on this form. Applicant may still petition for a license under 37 CFR 5.12, if a license is desired before the expiration of 6 months from the filing date of the application. If 6 months has lapsed from the filing date of this application and the licensee has not received any indication of a secrecy order under 35 U.S.C. 181, the licensee may foreign file the application pursuant to 37 CFR 5.15(b).

SelectUSA

The United States represents the largest, most dynamic marketplace in the world and is an unparalleled location for business investment, innovation, and commercialization of new technologies. The U.S. offers tremendous resources and advantages for those who invest and manufacture goods here. Through SelectUSA, our nation works to promote and facilitate business investment. SelectUSA provides information assistance to the international investor community; serves as an ombudsman for existing and potential investors; advocates on behalf of U.S. cities, states, and regions competing for global investment; and counsels U.S. economic development organizations on investment attraction best practices. To learn more about why the United States is the best country in the world to develop technology, manufacture products, deliver services, and grow your business, visit http://www.SelectUSA.gov or call +1-202-482-6800.

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	ATTY. DOCKET NO./TITLE
17/381,888	07/21/2021	Douglas A. Amorese	NUGN-001/04US 34079/550

CONFIRMATION NO. 6188

21710
BROWN RUDNICK LLP
ONE FINANCIAL CENTER
BOSTON, MA 02111

PUBLICATION NOTICE

OC000000129893290

Title:COMPOSITIONS AND METHODS FOR IDENTIFICATION OF A DUPLICATE SEQUENCING READ

Publication No.US-2021-0355537-A1
Publication Date:11/18/2021

NOTICE OF PUBLICATION OF APPLICATION

The above-identified application will be electronically published as a patent application publication pursuant to 37 CFR 1.211, et seq. The patent application publication number and publication date are set forth above.

The publication may be accessed through the USPTO's publically available Searchable Databases via the Internet at www.uspto.gov. The direct link to access the publication is currently http://www.uspto.gov/patft/.

The publication process established by the Office does not provide for mailing a copy of the publication to applicant. A copy of the publication may be obtained from the Office upon payment of the appropriate fee set forth in 37 CFR 1.19(a)(1). Orders for copies of patent application publications are handled by the USPTO's Public Records Division. The Public Records Division can be reached by telephone at (571) 272-3150 or (800) 972-6382, by facsimile at (571) 273-3250, by mail addressed to the United States Patent and Trademark Office, Public Records Division, Alexandria, VA 22313-1450 or via the Internet.

In addition, information on the status of the application, including the mailing date of Office actions and the dates of receipt of correspondence filed in the Office, may also be accessed via the Internet through the Patent Electronic Business Center at www.uspto.gov using the public side of the Patent Application Information and Retrieval (PAIR) system. The direct link to access this status information is currently https://portal.uspto.gov/pair/PublicPair. Prior to publication, such status information is confidential and may only be obtained by applicant using the private side of PAIR.

Further assistance in electronically accessing the publication, or about PAIR, is available by calling the Patent Electronic Business Center at 1-866-217-9197.

Office of Data Managment, Application Assistance Unit (571) 272-4000, or (571) 272-4200, or 1-888-786-0101

IDT Exhibit 1020.0095

PTO/AIA/80 (07-17)
Approved for use through 03/31/2021. OMB 0651-0035
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no person is required to respond to a collection of information unless it displays a valid OMB control number

POWER OF ATTORNEY TO PROSECUTE APPLICATIONS BEFORE THE USPTO

I hereby revoke all previous powers of attorney given in the application identified in the attached statement under 37 CFR 3.73(c).

I hereby appoint:

[x] Practitioners associated with Customer Number: 21710

OR

[ ] Practitioner(s) named below (if more than ten patent practitioners are to be named, then a customer number must be used):

Name	Registration Number	Name	Registration Number

As attorney(s) or agent(s) to represent the undersigned before the United States Patent and Trademark Office (USPTO) in connection with any and all patent applications assigned only to the undersigned according to the USPTO assignment records or assignments documents attached to this form in accordance with 37 CFR 3.73(c).

Please change the correspondence address for the application identified in the attached statement under 37 CFR 3.73(c) to:

[ ] The address associated with Customer Number:

OR

[ ]

Firm or individual name		
Address		
City	State	Zip
Country		
Telephone	Email	

Assignee name and address:

Tecan Genomics, Inc.
900 Chesapeake Drive
Redwood City, California 94063

A copy of this form, together with a statement under 37 CFR 3.73(c) (Form PTO/AIA/96 or equivalent) is required to be filed in each application in which this form is used. The statement under 37 CFR 3.73(c) may be completed by one of the practitioners appointed in this form, and must identify the application in which this Power of Attorney is to be filed.

SIGNATURE of Assignee of Record

The individual whose signature and title is supplied below is authorized to act on behalf of the assignee.

Signature	Leonard J. Gorén	Date	6 December 2021
Name	Leonard J. Gorén	Telephone	650-590-3660
Title	General Manager		

Signature		Date	6 December 2021
Name	Rasmus Lind	Telephone	
Title	Vice President Global Sales		

64303617

Electronic Acknowledgement Receipt

EFS ID:	44546595
Application Number:	17381888
International Application Number:	
Confirmation Number:	6188
Title of Invention:	COMPOSITIONS AND METHODS FOR IDENTIFICATION OF A DUPLICATE SEQUENCING READ
First Named Inventor/Applicant Name:	Douglas A. Amorese
Customer Number:	21710
Filer:	Thomas C. Meyers/Sandra Richards
Filer Authorized By:	Thomas C. Meyers
Attorney Docket Number:	NUGN-001/04US 34079/550
Receipt Date:	17-DEC-2021
Filing Date:	21-JUL-2021
Time Stamp:	12:50:45
Application Type:	Utility under 35 USC 111(a)

Payment information:

Submitted with Payment	no

File Listing:

Document Number	Document Description	File Name	File Size(Bytes)/ Message Digest	Multi Part /.zip	Pages (if appl.)
1	Assignee showing of ownership per 37 CFR 3.73	NUGN-001-01US-373-statement.pdf	25972 bd3d403fb79184b5fa9b781f92bd09c459b54347	no	2

Warnings:

Information:					
2	Power of Attorney	Tecan-POA-120621.pdf	83539 d84541658fd7a3f8276760e65ab3d5eee2f29a76	no	1
Warnings:					
Information:					
			Total Files Size (in bytes):	109511	

This Acknowledgement Receipt evidences receipt on the noted date by the USPTO of the indicated documents, characterized by the applicant, and including page counts, where applicable. It serves as evidence of receipt similar to a Post Card, as described in MPEP 503.

New Applications Under 35 U.S.C. 111

If a new application is being filed and the application includes the necessary components for a filing date (see 37 CFR 1.53(b)-(d) and MPEP 506), a Filing Receipt (37 CFR 1.54) will be issued in due course and the date shown on this Acknowledgement Receipt will establish the filing date of the application.

National Stage of an International Application under 35 U.S.C. 371

If a timely submission to enter the national stage of an international application is compliant with the conditions of 35 U.S.C. 371 and other applicable requirements a Form PCT/DO/EO/903 indicating acceptance of the application as a national stage submission under 35 U.S.C. 371 will be issued in addition to the Filing Receipt, in due course.

New International Application Filed with the USPTO as a Receiving Office

If a new international application is being filed and the international application includes the necessary components for an international filing date (see PCT Article 11 and MPEP 1810), a Notification of the International Application Number and of the International Filing Date (Form PCT/RO/105) will be issued in due course, subject to prescriptions concerning national security, and the date shown on this Acknowledgement Receipt will establish the international filing date of the application.

PTO/AIA/96 (08-12)
Approved for use through 11/30/2020. OMB 0651-0031
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.

STATEMENT UNDER 37 CFR 3.73(c)

Applicant/Patent Owner: Douglas A. Amorese et al.

Application No./Patent No.: 17/381,888 Filed/Issue Date: July 21, 2021

Titled: COMPOSITIONS AND METHODS FOR IDENTIFICATION OF A DUPLICATE SEQUENCING READ

Tecan Genomics, Inc. (Name of Assignee) , a Corporation (Type of Assignee, e.g., corporation, partnership, university, government agency, etc.)

states that, for the patent application/patent identified above, it is (choose **one** of options 1, 2, 3 or 4 below):

1. [X] The assignee of the entire right, title, and interest.

2. [ ] An assignee of less than the entire right, title, and interest (check applicable box):

   [ ] The extent (by percentage) of its ownership interest is _____ %. Additional Statement(s) by the owners holding the balance of the interest must be submitted to account for 100% of the ownership interest.

   [ ] There are unspecified percentages of ownership. The other parties, including inventors, who together own the entire right, title and interest are:

   Additional Statement(s) by the owner(s) holding the balance of the interest must be submitted to account for the entire right, title, and interest.

3. [ ] The assignee of an undivided interest in the entirety (a complete assignment from one of the joint inventors was made). The other parties, including inventors, who together own the entire right, title, and interest are:

   Additional Statement(s) by the owner(s) holding the balance of the interest must be submitted to account for the entire right, title, and interest.

4. [ ] The recipient, via a court proceeding or the like (*e.g.*, bankruptcy, probate), of an undivided interest in the entirety (a complete transfer of ownership interest was made). The certified document(s) showing the transfer is attached.

The interest identified in option 1, 2 or 3 above (not option 4) is evidenced by either (choose **one** of options A or B below):

A. [ ] An assignment from the inventor(s) of the patent application/patent identified above. The assignment was recorded in the United States Patent and Trademark Office at Reel _____, Frame _____, or for which a copy thereof is attached.

B. [X] A chain of title from the inventor(s), of the patent application/patent identified above, to the current assignee as follows:

1. From: Douglas A. Amorese, et al. To: Nugen Technologies, Inc.
   The document was recorded in the United States Patent and Trademark Office at
   Reel 053605, Frame 0123, or for which a copy thereof is attached.

2. From: Nugen Technologies, Inc. To: Tecan Genomics, Inc.
   The document was recorded in the United States Patent and Trademark Office at
   Reel 056918, Frame 0037, or for which a copy thereof is attached.

64323374

IDT Exhibit 1020.0099

PTO/AIA/96 (08-12)
Approved for use through 11/30/2020. OMB 0651-0031
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.

STATEMENT UNDER 37 CFR 3.73(c)

3. From: _____ To: _____
The document was recorded in the United States Patent and Trademark Office at
Reel _____ , Frame _____ , or for which a copy thereof is attached.

4. From: _____ To: _____
The document was recorded in the United States Patent and Trademark Office at
Reel _____ , Frame _____ , or for which a copy thereof is attached.

5. From: _____ To: _____
The document was recorded in the United States Patent and Trademark Office at
Reel _____ , Frame _____ , or for which a copy thereof is attached.

6. From: _____ To: _____
The document was recorded in the United States Patent and Trademark Office at
Reel _____ , Frame _____ , or for which a copy thereof is attached.

☐ Additional documents in the chain of title are listed on a supplemental sheet(s).

☒ As required by 37 CFR 3.73(c)(1)(i), the documentary evidence of the chain of title from the original owner to the assignee was, or concurrently is being, submitted for recordation pursuant to 37 CFR 3.11.

[NOTE: A separate copy (i.e., a true copy of the original assignment document(s)) must be submitted to Assignment Division in accordance with 37 CFR Part 3, to record the assignment in the records of the USPTO. See MPEP 302.08]

The undersigned (whose title is supplied below) is authorized to act on behalf of the assignee.

/Thomas C. Meyers/	December 17, 2021
Signature	Date
Thomas C. Meyers	36,989
Printed or Typed Name	Title or Registration Number

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	ATTY. DOCKET NO./TITLE
17/381,888	07/21/2021	Douglas A. Amorese	NUGN-001/04US 34079/550

CONFIRMATION NO. 6188

21710
BROWN RUDNICK LLP
ONE FINANCIAL CENTER
BOSTON, MA 02111

POA ACCEPTANCE LETTER

OC00000130600718

Date Mailed: 12/22/2021

NOTICE OF ACCEPTANCE OF POWER OF ATTORNEY

This is in response to the Power of Attorney filed 12/17/2021.

The Power of Attorney in this application is accepted. Correspondence in this application will be mailed to the above address as provided by 37 CFR 1.33.

Questions about the contents of this notice and the requirements it sets forth should be directed to the Office of Data Management, Application Assistance Unit, at **(571) 272-4000** or **(571) 272-4200** or **1-888-786-0101**.

/agizaw/

IDT Exhibit 1020.0101

PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : C12Q 1/68

A1

(11) International Publication Number: **WO 98/44151**

(43) International Publication Date: 8 October 1998 (08.10.98)

(21) International Application Number: PCT/GB98/00961

(22) International Filing Date: 1 April 1998 (01.04.98)

(30) Priority Data:

9706528.8	1 April 1997 (01.04.97)	GB
9706529.6	1 April 1997 (01.04.97)	GB
9713236.9	23 June 1997 (23.06.97)	GB
9713238.5	23 June 1997 (23.06.97)	GB

(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, GW, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).

(71) Applicant *(for all designated States except US)*: GLAXO GROUP LIMITED [GB/GB]; Glaxo Wellcome House, Berkeley Avenue, Greenford, Middlesex UB6 0NN (GB).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: KAWASHIMA, Eric [US/CH]; Serono Pharmaceutical Research Institute S.A., 14, chemin de Aulx, CH–1228 Plan–les–Ouates (CH). FARINELLI, Laurent [CH/CH]; 15, chemin de Aulx, CH–1256 Troinex (CH). MAYER, Pascal [FR/CH]; Serono Pharmaceutical Research Institute S.A., 14, chemin de Aulx, CH–1228 Plan–les–Ouates (CH).

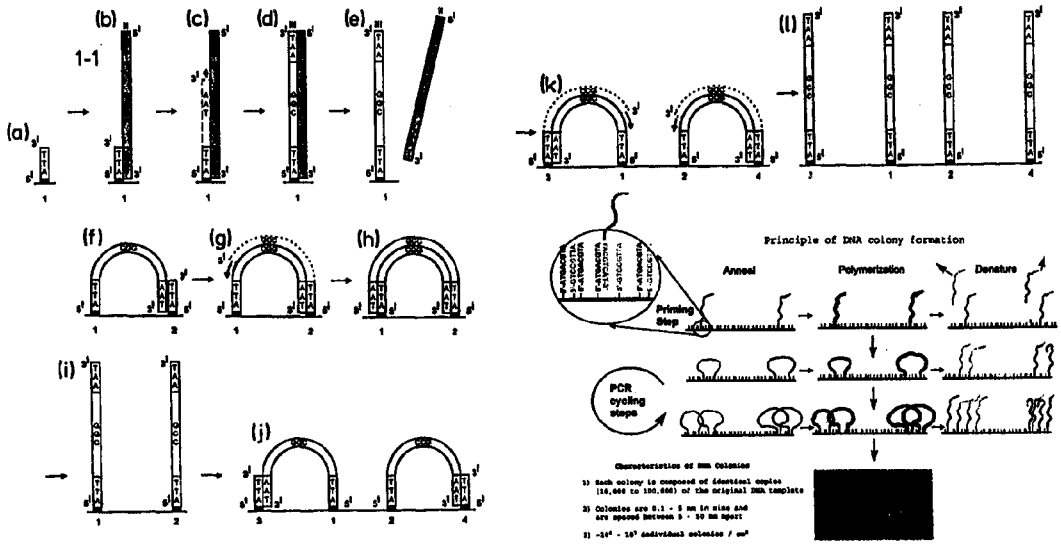
(74) Agent: GLAXO WELLCOME PLC; Glaxo Wellcome House, Berkeley Avenue, Greenford, Middlesex UB6 0NN (GB).

Published

With international search report.

Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: METHOD OF NUCLEIC ACID AMPLIFICATION



(57) Abstract

A nucleic acid molecule can be annealed to an appropriate immobilised primer. The primer can then be extended and the molecule and the primer can be separated from one another. The extended primer can then be annealed to another immobilised primer and the other primer can be extended. Both extended primers can then be separated from one another and can be used to provide further extended primers. The process can be repeated to provide amplified, immobilised nucleic acid molecules. These can be used for many different purposes, including sequencing, screening, diagnosis, *in situ* nucleic acid synthesis, monitoring gene expression, nucleic acid fingerprinting, etc.

IDT Exhibit 1020.0102

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	**ES**	Spain	**LS**	Lesotho	**SI**	Slovenia
AM	Armenia	**FI**	Finland	**LT**	Lithuania	**SK**	Slovakia
AT	Austria	**FR**	France	**LU**	Luxembourg	**SN**	Senegal
AU	Australia	**GA**	Gabon	**LV**	Latvia	**SZ**	Swaziland
AZ	Azerbaijan	**GB**	United Kingdom	**MC**	Monaco	**TD**	Chad
BA	Bosnia and Herzegovina	**GE**	Georgia	**MD**	Republic of Moldova	**TG**	Togo
BB	Barbados	**GH**	Ghana	**MG**	Madagascar	**TJ**	Tajikistan
BE	Belgium	**GN**	Guinea	**MK**	The former Yugoslav Republic of Macedonia	**TM**	Turkmenistan
BF	Burkina Faso	**GR**	Greece			**TR**	Turkey
BG	Bulgaria	**HU**	Hungary	**ML**	Mali	**TT**	Trinidad and Tobago
BJ	Benin	**IE**	Ireland	**MN**	Mongolia	**UA**	Ukraine
BR	Brazil	**IL**	Israel	**MR**	Mauritania	**UG**	Uganda
BY	Belarus	**IS**	Iceland	**MW**	Malawi	**US**	United States of America
CA	Canada	**IT**	Italy	**MX**	Mexico	**UZ**	Uzbekistan
CF	Central African Republic	**JP**	Japan	**NE**	Niger	**VN**	Viet Nam
CG	Congo	**KE**	Kenya	**NL**	Netherlands	**YU**	Yugoslavia
CH	Switzerland	**KG**	Kyrgyzstan	**NO**	Norway	**ZW**	Zimbabwe
CI	Côte d'Ivoire	**KP**	Democratic People's Republic of Korea	**NZ**	New Zealand		
CM	Cameroon			**PL**	Poland		
CN	China	**KR**	Republic of Korea	**PT**	Portugal		
CU	Cuba	**KZ**	Kazakstan	**RO**	Romania		
CZ	Czech Republic	**LC**	Saint Lucia	**RU**	Russian Federation		
DE	Germany	**LI**	Liechtenstein	**SD**	Sudan		
DK	Denmark	**LK**	Sri Lanka	**SE**	Sweden		
EE	Estonia	**LR**	Liberia	**SG**	Singapore		

IDT Exhibit 1020.0103

Method of Nucleic Acid Amplification

This invention relates, *inter alia*, to the amplification of nucleic acids.

Molecular biology and pharmaceutical drug development now make intensive use of nucleic acid analysis (Friedrich, G.A. Moving beyond the genome projects, *Nature Biotechnology* 14, 1234 (1996)). The most challenging areas are whole genome sequencing, single nucleotide polymorphism detection, screening and gene expression monitoring. Currently, up to hundreds of thousands of samples are handled in single DNA sequencing projects (Venter, J.C., H.O Smith, L. Hood, A new strategy for genome sequencing, *Nature* 381, 364 (1996)). This capacity is limited by the available technology. Projects like the "human genome project" (gene mapping and DNA sequencing) and identifying all polymorphisms in expressed genes involved in common diseases imply the sequencing of millions of DNA samples.

With most of the current DNA sequencing technologies, it is simply not possible to decrease indefinitely the time required to process a single sample. A way of increasing throughput is to perform many processes in parallel. The introduction of robotic sample preparation and delivery, 96 and 384 well plates, high density gridding machines (Maier, E., S. Meierewer, A.R. Ahmadi, J. Curtis, H. Lehrach, Application of robotic technology to automated sequence fingerprint analysis by oligonucleotide hybridization, *Journal Of Biotechnology* 35,191 (1994)) and recently the development of high density oligonucleotide arrays (Chee, M., R.Yang, E. Hubbell, A. Berno,X.C. Huang,D. Stern,J. Winkler, D.J. Lockhart,M.S. Morris, and S.P.A. Fodor, Accessing genetic information with high-density DNA arrays, *Science* 274(5287):610-614, (1996)) are starting to bring answers to the demand in ever higher throughput. Such technologies allow up to 50,000~100,000 samples at a time to be processed within days and even hours (Maier, E., Robotic technology in library screening, *Laboratory Robotics and Automation* 7, 123 (1995)).

SUBSTITUTE SHEET (RULE 26)

IDT Exhibit 1020.0104

In most known methods for performing nucleic acid analysis, it is necessary to first extract the nucleic acids of interest (e.g., genomic or mitochondrial DNA or messenger RNA (mRNA)) from an organism. Then it is necessary to isolate the nucleic acids of interest from the mixture of all nucleic acids and usually, to amplify these nucleic acids to obtain quantities suitable for their characterisation and/or detection. Isolating the nucleic fragments has been considered necessary even when one is interested in a representative but random set of all of the different nucleic acids, for instance, a representative set of all the mRNAs present in a cell or of all the fragments obtained after genomic DNA has been cut randomly into small pieces.

Several methods can be used to amplify DNA with biological means and are well known by those skilled in the art. Generally, the fragments of DNA are first inserted into vectors with the use of restriction enzymes and DNA ligases. A vector containing a fragment of interest can then be introduced into a biological host and amplified by means of well established protocols. Usually hosts are randomly spread over a growth medium (e.g. agar plates). They can then replicate to provide colonies that originated from individual host cells.

Up to millions of simultaneous amplification of cloned DNA fragments can be carried out simultaneously in such hosts. The density of colonies is of the order of 1 colony/mm^2. In order to obtain DNA from such colonies one option is to transfer the colonies to a membrane, and then to immobilise the DNA from within the biological hosts directly to the membrane (Grunstein, M. and D.S. Hogness, Colony Hybridization: A method for the isolation of cloned DNAs that contain a specific gene, *Proceedings of the National Academy of Science, USA,* 72:3961 (1975)). With these options however, the amount of transferred DNA is limited and often insufficient for non-radioactive detection.

Another option is to transfer by sterile technique individually each colony into a container (e.g., 96 well plates) where further host cell replication can occur so that more DNA can be obtained from the colonies. Amplified nucleic acids can be recovered from the host

SUBSTITUTE SHEET (RULE 26)

IDT Exhibit 1020.0105

cells with an appropriate purification process. However such a procedure is generally time and labour consuming, and difficult to automate.

The revolutionary technique of DNA amplification using the polymerase chain reaction (PCR) was proposed in 1985 by Mullis et al. (Saiki, R., S. Scharf, F. Faloona, K. Mullis, G. Horn, H. Erlich and N. Arnheim, *Science* 230, 1350-1354 (1985) and is now well known by those skilled in the art. In this amplification process, a DNA fragment of interest can be amplified using two short (typically about 20 base long) oligonucleotides that flank a region to be amplified, and that are usually referred to as "primers". Amplification occurs during the PCR cycling, which includes a step during which double stranded DNA molecules are denatured (typically a reaction mix is heated, e.g. to 95°C in order to separate double stranded DNA molecules into two single stranded fragments), an annealing step (where the reaction mix is brought to e.g., 45°C in order to allow the primers to anneal to the single stranded templates) and an elongation step (DNA complementary to the single stranded fragment is synthesised via sequential nucleotide incorporation at the ends of the primers with the DNA polymerase enzyme).

The above procedure is usually performed in solution, whereby neither the primers nor a template are linked to any solid matrix.

More recently, however, it has been proposed to use one primer grafted to a surface in conjunction with free primers in solution in order to simultaneously amplify and graft a PCR product onto the surface (Oroskar, A.A., S.E. Rasmussen,H.N. Rasmussen,S.R. Rasmussen,B.M. Sullivan, and A. Johansson, Detection of immobilised amplicons by ELISA-like techniques, *Clinical Chemistry* 42:1547 (1996)). (The term "graft" is used herein to indicate that a moiety becomes attached to a surface and remains there unless and until it is desired to remove it.)The amplification is generally performed in containers (e.g., in 96 well format plates) in such a way that each container contains the PCR product(s) of one reaction. With such

IDT Exhibit 1020.0106

methods, some of the PCR product become grafted to a surface of the container having primers therein which has been in contact with the reactant during the PCR cycling. The grafting to the surface simplifies subsequent assays and allows efficient automation.

Arraying of DNA samples is more classically performed on membranes (e.g., nylon or nitro-cellulose membranes). With the use of suitable robotics (e.g., Q-bot™ , Genetix ltd, Dorset BH23 3TG UK) it is possible to reach a density of up to 10 samples/mm^2. Here, the DNA is covalently linked to a membrane by physicochemical means (e.g., UV irradiation). These technologies allow the arraying of large DNA molecules (e.g. molecules over 100 nucleotides long) as well as smaller DNA molecules. Thus both templates and probes can be arrayed.

New approaches based on pre-arrayed glass slides (arrays of reactive areas obtained by ink-jet technology (Blanchard, A.P.and L. Hood, Oligonucleotide array synthesis using ink jets, *Microbial and Comparative Genomics,* 1:225 (1996)) or arrays of reactive polyacrylamide gels (Yershov, G. et al., DNA analysis and diagnostics on oligonucleotide microchips, *Proceedings of the National Academy of Science, USA*, 93:4913-4918 (1996)) allow the arraying of up to 100 samples/mm^2. With these technologies, only probe (oligonucleotide) grafting has been reported. Reported number of samples/mm^2 are still fairly low (25 to 64).

Higher sample densities are achievable by the use of DNA chips, which can be arrays of oligonucleotides covalently bound to a surface and can be obtained with the use of micro-lithographic techniques (Fodor, S.P.A. et al., Light directed, spatially addressable parallel chemical synthesis, *Science* 251:767(1991)). Currently, chips with 625 probes/mm^2 are used in applications for molecular biology (Lockhart, D.J. et al., Expression monitoring by hybridisation to high-density oligonucleotide arrays, *Nature Biotechnology* 14:1675 (1996)). Probe densities of up to 250 000 samples/cm^2 are claimed to be achievable (Chee, M. et al., Accessing genetic information with

IDT Exhibit 1020.0107

high-density DNA arrays, *Science* 274:610 (1996)). Currently, up to 132000 different oligonucleotides can be arrayed on a single chips of approximately 2.5 cm^2. Presently, these chips are manufactured by direct solid phase oligonucleotide synthesis with the 3'OH end of the oligo attached to the surface. Thus these chips have been used to provide oligonucleotide probes which cannot act as primers in a DNA polymerase-mediated elongation step.

When PCR products are linked to the vessel in which PCR amplification takes place, this can be considered as a direct arraying process. The density of the resultant array of PCR products is then limited by the available vessel. Currently available vessels are only in 96 well microtiter plate format. These allow only around ~0.02 samples of PCR products/mm^2 of surface to be obtained.

Using the commercially available Nucleolink™ system obtainable from Nunc A/S (Roskilde, Denmark) it is possible to achieve simultaneous amplification and arraying of samples in containers on the surface of which oligonucleotide primers have been grafted. However, in this case the density of the array of samples is fixed by the size of the vessel. Presently a density of 0.02 samples/mm^2 is achievable for the 96 well plate format. Increasing this density is difficult. This is apparent since, for instance, the availability of 384 well plates (0.08 samples/mm^2) suitable for PCR has been delayed due to technical problems (e.g. heat transfer and capillary effects during filling). It is thus unlikely that orders of magnitude improvements in the density of samples arrayed with this approach can be achieved in the foreseeable future.

The present invention aims to overcome or at least alleviate some of the disadvantages of prior art methods of nucleic acid amplification.

According to the present invention there is provided a method of nucleic acid amplification, comprising the steps of:

IDT Exhibit 1020.0108

A. providing a plurality of primers that are immobilised but that have one end exposed to allow primer extension;
B. allowing a single stranded target nucleic acid molecule to anneal to one of said plurality of primers over part of the length of said single stranded nucleic acid molecule and then extending that primer using the annealed single stranded nucleic acid molecule as a template, so as to provide an extended immobilised nucleic acid strand;
C. separating the target nucleic molecule from the extended immobilised nucleic acid strand;
D. allowing the extended immobilised nucleic acid strand to anneal to one of said plurality of primers referred to in step A) and then extending that primer using the extended immobilised nucleic acid strand as a template, so as to provide another extended immobilised nucleic acid strand; and optionally,
E. separating the annealed extended immobilised nucleic acid strands from one another.

Preferably the method also comprises the step of:
F. using at least one extended immobilised nucleic acid strand to repeat steps D) and E), so as to provide additional extended immobilised nucleic acid strands and, optionally,
G. repeating step F) one or more times.

Desirably the single-stranded target nucleic acid sequence is provided by a method in which said single-stranded target nucleic acid is produced by providing a given nucleic acid sequence to be amplified (which sequence may be known or unknown) and adding thereto a first nucleic acid sequence and a second nucleic acid sequence; wherein said first nucleic acid sequence hybridises to one of said plurality of primers and said second nucleic acid sequence is complementary to a sequence which hybridises to one of said plurality of primers.

The second nucleic acid sequence may be a sequence that is the same as the sequence of one of the plurality of primers. Thus the single-stranded target nucleic acid squence may be provided by a method in which said single-stranded target nucleic acid is produced by providing a given nucleic acid sequence to be amplified (which

IDT Exhibit 1020.0109

sequence may be known or unknown) and adding thereto a first nucleic acid sequence and a second nucleic acid sequence; wherein said first nucleic acid sequence hybridises to one of said plurality of primers and said second nucleic acid sequence is the same as the sequence of one of said plurality of primers.

The first and second nucleic acid sequences may be provided at first and second ends of said single-stranded target nucleic acid, although this is not essential.

If desired a tag may be provided to enable amplification products of a given nucleic acid sequence to be identified.

Colonies

The method of the present invention allows one or more distinct areas to be provided, each distinct area comprising a plurality of immobilised nucleic acid strands (hereafter called "colonies"). These areas can contain large numbers of amplified nucleic acid molecules. These molecules may be DNA and/or RNA molecules and may be provided in single or double stranded form. Both a given strand and its complementary strand can be provided in amplified form in a single colony.

Colonies of any particular size can be provided.

However, preferred colonies measure from 10nm to 100μm across their longest dimension, more preferably from 100nm to 10μm across their longest dimension. Desirably a majority of the colonies present on a surface (i.e. at least 50% thereof) have sizes within the ranges given above.

Colonies can be arranged in a predetermined manner or can be randomly arranged. Two or three dimensional colony configurations are possible. The configurations may be regular (e.g. having a polygonal outline or having a generally circular outline) or they may be irregular.

IDT Exhibit 1020.0110

Colonies can be provided at high densities. Densities of over one colony/mm^2 of surface can be achieved. Indeed densities of over 10^2, over 10^3 or even over 10^4 colonies/mm^2 are achievable using the present invention. In preferred embodiments, the present invention provides colony densities of 10^{4-5} colonies/mm^2, more preferably densities of 10^{6-7} colonies/mm^2, thus offering an improvement of 3 to 4 orders of magnitude relative to densities achievable using many of the prior art methods. It is this property of the invention that allows a great advantage over prior art, since the high density of DNA colonies allows a large number of diverse DNA templates (up to 10^{6-7} colonies/mm^2, to be randomly arrayed and amplified.

Primers

The immobilised primers for use in the present invention can be provided by any suitable means, as long as a free 3'-OH end is available for primer extension. Where many different nucleic acid molecules are to be amplified, many different primers may be provided. Alternatively "universal" primers may be used, whereby only one or two different types of primer (depending upon the embodiment of the invention) can be used to amplify the different nucleic acid molecules. Universal primers can be used where the molecules to be amplified comprise first and second sequences, as described previously. The provision of universal primers is advantageous over methods such as those disclosed in WO96/04404 (Mosaic Technologies, Inc.) where specific primers must be prepared for each particular sequence to be amplified.

Synthetic oligodeoxynucleotide primers are available commercially from many suppliers (e.g. Microsynth, Switzerland, Eurogentech, Belgium).

Grafting of primers onto silanized glass or quartz and grafting of primers onto silicon wafers or gold surface has been described (Maskos, U. and E.M Southern, Oligonucleotide hybridizations on glass supports: a novel linker for oligonucleotide synthesis and hybridization properties of oligonucleotides synthesised in situ, *Nucleic Acids Research* 20(7):1679-84, 1992; Lamture, J.B., et.al. Direct-detection of nucleic-acid hybridization on the surface of a

IDT Exhibit 1020.0111

charge-coupled-device, *Nucleic Acids Research* 22(11):2121-2125, 1994; Chrisey, L.A., G.U. Lee, and C.E. Oferrall, Covalent attachment of synthetic DNA to self-assembled monolayer films, *Nucleic Acids Research* 24(15):3031-3039, 1996).

Grafting biotinylated primers to supports covered with streptavidin is another alternative. This grafting method is commonly used for bio-macromolecules in general.

Non covalent grafting of primers at the interface between an aqueous phase and a hydrophobic phase through an hydrophobic anchor is also possible for the present invention. Such anchoring is commonly used for bio-macromolecules in general (S. Terrettaz et al. : Protein binding to supported lipid membranes, *Langmuir* 9,1361 (1993)). Preferred forms of such interfaces would be liposomes, lipidic vesicles, emulsions, patterned bilayers, Langmuir or Langmuir-Blodgett films. The patterns may be obtained by directed pattering on templates, e.g., silicon chips patterned through micro-lithographic methods (Goves, J.T. et al., Micropatterning Fluid Bilayers on Solid Supports, in *Science* 275,651 (1997)). The patterns may also be obtained by due to the self-assembly properties of "colloids", e.g., emulsions or latex particles (Larsen, A.E. and D.G. Grier, Like charge attractions in metastable colloidal crystallites, *Nature* 385,230 (1997)).

In the above methods, one, two or more different primers can be grafted onto a surface. The primers can be grafted homogeneously and simultaneously over the surface.

Using microlithographic methods it is possible to provide immobilised primers in a controlled manner. If direct synthesis of oligonucleotides onto a solid support with a free 3'-OH end is desired, then micro-lithographic methods can be used to simultaneously synthesise many different oligonucleotide primers (Pirrung, M.C. and Bradley, J.C. Comparison of methods for photochemical phosphoramidite-based DNA- synthesis. *Journal Of Organic Chemistry* 60(20):6270-6276, 1995). These may be provided in distinct areas that may correspond in configuration to colonies to be

IDT Exhibit 1020.0112

formed, (e.g. they may be several nanometers or micrometers across). Within each area, only a single type of primer oligonucleotide need be provided. Alternatively a mixture comprising a plurality of different primers may be provided. In either case, primers can be homogeneously distributed within each area. They may be provided in the form of a regular array.

Where areas initially comprise only one type of immobilised primer they may be modified, if desired, to carry two or more different types of primer. One way to achieve this is to use molecules as templates for primer extension that have 3′ ends that hybridise with a single type of primer initially present and that have 5′ ends extending beyond the 3′ ends of said primers. By providing a mixture of templates with different sequences from one another, primer extension of one type of primer using the mixture of such templates followed by strand separation will result in different modified primers. (The modified primers are referred to herein as "extended" primers in order to distinguish from the "primary" primers initially present on a surface).

One, two or more different types of extended primer can be provided in this manner at any area where primary primers are initially located. Substantially equal portions of different templates can be used, if desired, in order to provide substantially equal proportions of different types of immobilised extended primer over a given area. If different proportions of different immobilised extended primers are desired, then this can be achieved by adjusting the proportions of different template molecules initially used accordingly.

A restriction endonuclease cleavage site may be located within the primer. A primer may also be provided with a restriction endonuclease recognition site which directs DNA cleavage several bases distant (Type II restriction endonucleases). (For the avoidance of doubt, such sites are deemed to be present even if the primer and its complement are required to be present in a double stranded molecule for recognition and/or cleavage to occur.) Alternatively a cleavage site and/or a recognition site may be produced when a primer is extended. In any event, restriction endonucleases can be useful in allowing an immobilised nucleic acid molecule within a colony to

IDT Exhibit 1020.0113

be cleaved so as to release at least a part thereof. As an alternative to using other restriction endonucleases, ribozymes can be used to release at least parts of nucleic acid molecules from a surface (when such molecules are RNA molecules). Other methods are possible. For example if a covalent bond is used to link a primer to a surface this bond may be broken (e.g. by chemical, physical or enzymatic means).

Primers for use in the present invention are preferably at least five bases long. Normally they will be less than 100 or less than 50 bases long. However this is not essential. Naturally occurring and/or non-naturally occurring bases may be present in the primers.

Target Nucleic Acid Molecules

Turning now to target nucleic acid molecules (also referred to herein as "templates") for use in the method of the present invention, these can be provided by any appropriate means. A target molecule (when in single-stranded form) comprises a first part having a sequence that can anneal with a first primer and a second part having a sequence complementary to a sequence that can anneal with a second primer. In a preferred embodiment the second part has the same sequence as the second primer.

The second primer may have a sequence that is the same as, or different from, the sequence of the first primer.

The first and second parts of the target nucleic acid molecules are preferably located at the 3' and at the 5' ends respectively thereof. However this is not essential. The target molecule will usually also comprise a third part located between the first and second parts. This part of the molecule comprises a particular sequence to be replicated. It can be from any desired source and may have a known or unknown (sometimes referred to as "anonymous") sequence. It may be derived from random fractionation by mechanical means or by limited restriction enzyme digestion of a nucleic acid sample, for example.

IDT Exhibit 1020.0114

Further parts of the target molecules may be provided if desired. For example parts designed to act as tags may be provided. A "tag" is defined by its function of enabling a particular nucleic acid molecule (or its complement) to be identified.

Whatever parts are present, target nucleic acid molecules can be provided by techniques known to those skilled in the art of nucleic acid manipulation. For example, two or more parts can be joined together by ligation. If necessary, prior to ligation appropriate modifications can be made to provide molecules in a form ready for ligation. For example if blunt end ligation is desired then a single-strand specific exonuclease such as S1 nuclease could be used to remove single stranded portions of molecules prior to ligation. Linkers and/or adapters may also be used in nucleic acid manipulation. (Techniques useful for nucleic acid manipulation are disclosed in Sambrook *et al*, *Molecular Cloning*, 2nd Edition, Cold Spring Harbor Laboratory Press (1989), for example.)

Once a template molecule has been synthesised it can be cloned into a vector and can be amplified in a suitable host before being used in the present invention. It may alternatively be amplified by PCR. As a further alternative, batches of template molecules can be synthesised chemically using automated DNA synthesisers (e.g. from Perkin-Elmer/Applied Biosystems, Foster City, CA).

It is however important to note that the present invention allows large numbers of nucleic acid molecules identical in sequence to be provided in a colony arising from a single molecule of template. Furthermore, the template can be re-used to generate further colonies. Thus it is not essential to provide large numbers of template molecules to be used in colony formation.

The template can be of any desired length provided that it can participate in the method of the present invention. Preferably it is at least 10, more preferably at least 20 bases long. More preferably it is at least 100 or at least 1000 bases long. As is the case for primers for use in the present invention, templates may comprise naturally occurring and/or non-naturally occurring bases.

SUBSTITUTE SHEET (RULE 26)

IDT Exhibit 1020.0115

Reaction Conditions

Turning now to reaction conditions suitable for the method of the present invention, it will be appreciated that the present invention uses repeated steps of annealing of primers to templates, primer extension and separation of extended primers from templates. These steps can generally be performed using reagents and conditions known to those skilled in PCR (or reverse transcriptase plus PCR) techniques. PCR techniques are disclosed, for example, in "PCR: Clinical Diagnostics and Research", published in 1992 by Springer-Verlag.

Thus a nucleic acid polymerase can be used together with a supply of nucleoside triphosphate molecules (or other molecules that function as precursors of nucleotides present in DNA/RNA, such as modified nucleoside triphosphates) to extend primers in the presence of a suitable template.

Excess deoxyribonucleoside triphosphates are desirably provided. Preferred deoxyribonucleoside triphosphates are abbreviated; dTTP (deoxythymidine nucleoside triphosphate), dATP (deoxyadenosine nucleoside triphosphate), dCTP (deoxycytosine nucleoside triphosphate) and dGTP (deoxyguanosine nucleoside triphosphate). Preferred ribonucleoside triphosphates are UTP, ATP, CTP and GTP. However alternatives are possible. These may be naturally or non-naturally occurring. A buffer of the type generally used in PCR reactions may also be provided.

A nucleic acid polymerase used to incorporate nucleotides during primer extension is preferably stable under the pertaining reaction conditions in order that it can be used several times. (This is particularly useful in automated amplification procedures.) Thus, where heating is used to separate a newly synthesised nucleic acid strand from its template, the nucleic acid polymerase is preferably heat stable at the temperature used. Such heat stable polymerases are known to those skilled in the art. They are obtainable from thermophilic micro-organisms. They include the DNA dependent DNA polymerase known as *Taq* polymerase and also thermostable derivatives thereof. (The nucleic acid polymerase need not however be DNA

SUBSTITUTE SHEET (RULE 26)

IDT Exhibit 1020.0116

dependent. It may be RNA dependent. Thus it may be a reverse transcriptase - i.e. an RNA dependent DNA polymerase.)

Typically, annealing of a primer to its template takes place at a temperature of 25 to 90°C. Such a temperature range will normally be maintained during primer extension. Once sufficient time has elapsed to allow annealing and also to allow a desired degree of primer extension to occur, the temperature can be increased, if desired, to allow strand separation. At this stage the temperature will typically be increased to a temperature of 60 to 100°C. [High temperatures can also be used to reduce non-specific priming problems prior to annealing. They can be used to control the timing of colony initiation, e.g. in order to synchronise colony initiation for a number of samples.] Alternatively, the strands maybe separated by treatment with a solution of low salt and high pH (> 12) or by using a chaotropic salt (e.g. guanidinium hydrochloride) or by an organic solvent (e.g. formamide).

Following strand separation (e.g. by heating), preferably a washing step will be performed. The washing step can be omitted between initial rounds of annealing, primer extension and strand separation, if it is desired to maintain the same templates in the vicinity of immobilised primers. This allows templates to be used several times to initiate colony formation. (It is preferable to provide a high concentration of template molecules initially so that many colonies are initiated at one stage.)

The size of colonies can be controlled, e.g. by controlling the number of cycles of annealing, primer extension and strand separation that occur. Other factors which affect the size of colonies can also be controlled. These include the number and arrangement on a surface of immobilised primers, the conformation of a support onto which the primers are immobilised, the length and stiffness of template and/or primer molecules, temperature and the ionic strength and viscosity of a fluid in which the above-mentioned cycles can be performed.

SUBSTITUTE SHEET (RULE 26)

IDT Exhibit 1020.0117

Uses of Colonies

Once colonies have been formed they can be used for any desired purpose.

For example, they may be used in nucleic acid sequencing (whether partial or full), in diagnosis, in screening, as supports for other components and/or for research purposes (preferred uses will be described in greater detail later on). If desired colonies may be modified to provide different colonies (referred to herein as "secondary colonies" in order to distinguish from the "primary colonies" initially formed).

Surfaces Comprising Immobilised Nucleic Acid Strands

A surface comprising immobilised nucleic acid strands in the form of colonies of single stranded nucleic acid molecules is also within the scope of the present invention.

Normally each immobilised nucleic acid strand within a colony will be located on the surface so that an immobilised and complementary nucleic acid strand thereto is located on the surface within a distance of the length of said immobilised nucleic acid strand (i.e. within the length of one molecule). This allows very high densities of nucleic acid strands and their complements to be provided in immobilised form. Preferably there will be substantially equal proportions of a given nucleic acid strand and its complement within a colony. A nucleic acid strand and its complement will preferably be substantially homogeneously distributed within the colony.

It is also possible to provide a surface comprising single stranded nucleic acid strands in the form of colonies, where in each colony, the sense and anti-sense single strands are provided in a form such that the two strands are no longer at all complementary, or simply partially complementary. Such surfaces are also within the scope of the present invention. Normally, such surfaces are obtained after treating primary colonies, e.g., by partial digestion by restriction enzymes or by partial digestion after strand separation (e.g., after

IDT Exhibit 1020.0118

heating) by an enzyme which digests single stranded DNA), or by chemical or physical means, (e.g., by irradiating with light colonies which have been stained by an intercalating dye e.g., ethidium bromide).

Once single stranded colonies have been provided they can be used to provide double stranded molecules. This can be done, for example, by providing a suitable primer (preferably in solution) that hybridises to the 3′ ends of single stranded immobilised molecules and then extending that primer using a nucleic acid polymerase and a supply of nucleoside triphosphates (or other nucleotide precursors).

Thus surfaces comprising colonies of non-bridged double stranded nucleic acid molecules are also within the scope of the present invention. (The term "non-bridged" is used here to indicate that the molecules are not in the form of the bridge-like structures shown in e.g. figure 1h.)

Using the present invention, small colonies can be provided that contain large numbers of nucleic acid molecules (whether single or double stranded). Many colonies can therefore be located on a surface having a small area. Colony densities that can be obtained may therefore be very high, as discussed *supra*.

Different colonies will generally be comprised of different amplified nucleic acid strands and amplified complementary strands thereto. Thus the present invention allows many different populations of amplified nucleic acid molecules and their complements to be located on a single surface having a relatively small surface area. The surface will usually be planar, although this is not essential.

Apparatuses

The present invention also provides an apparatus for providing a surface comprising colonies of the immobilised nucleic acid molecules discussed *supra*.

Such an apparatus can include one or more of the following:

IDT Exhibit 1020.0119

a) means for immobilising primers on a surface (although this is not needed if immobilised primers are already provided);

b) a supply of a nucleic acid polymerase;

c) a supply of precursors of the nucleotides to be incorporated into a nucleic acid (e.g. a supply of nucleoside triphosphates);

d) means for separating annealed nucleic acids (e.g. heating means);

and

e) control means for co-ordinating the different steps required for the method of the present invention.

Other apparatuses are within the scope of the present invention. These allow immobilised nucleic acids produced via the method of the present invention to be analysed. They can include a source of reactants and detecting means for detecting a signal that may be generated once one or more reactants have been applied to the immobilised nucleic acid molecules. They may also be provided with a surface comprising immobilised nucleic acid molecules in the form of colonies, as described *supra*.

Desirably the means for detecting a signal has sufficient resolution to enable it to distinguish between signals generated from different colonies.

Apparatuses of the present invention (of whatever nature) are preferably provided in automated form so that once they are activated, individual process steps can be repeated automatically.

The present invention will now be described without limitation thereof in sections A to I below with reference to the accompanying drawings.

SUBSTITUTE SHEET (RULE 26)

IDT Exhibit 1020.0120

It should be appreciated that procedures using DNA molecules referred to in these sections are applicable *mutatis mutandis* to RNA molecules, unless the context indicates otherwise.

It should also be appreciated that where sequences are provided in the following description, these are written from 5′ to 3′ (going from left to right), unless the context indicates otherwise.

The figures provided are summarised below:

FIGURE 1 illustrates a method for the simultaneous amplification and immobilisation of nucleic acid molecules using a single type of primer.

FIGURE 2 illustrates how colony growth using a method of the present invention can occur.

FIGURE 3 illustrates the principle of the method used to produce DNA colonies using the present invention. It also illustrates the annealing, elongation and denaturing steps that are used to provide such colonies.

FIGURE 4 is an example of DNA colonies formed by amplification of a specific template with single primers grafted onto a surface.

FIGURE 5 illustrates a method for the simultaneous amplification and immobilisation of nucleic acid molecules using two types of primer.

FIGURE 6 shows actual DNA colonies produced via the present invention.

FIGURE 7 illustrates a method of the simultaneous amplification and immobilisation of nucleic acid molecules when a target molecule is used as a template having internal sequences that anneal with primers.

SUBSTITUTE SHEET (RULE 26)

IDT Exhibit 1020.0121

Figure 8 illustrates a method to synthesise additional copies of the original nucleic acid strands using nucleic acid strands present in colonies. The newly synthesised strands are shown in solution but can be provided in immobilised form if desired.

Figure 9 shows the PCR amplification of DNA from DNA found in the pre-formed DNA colonies.

FIGURE 10 illustrates how secondary primers can be generated from DNA colonies.

FIGURE 11 illustrates how secondary DNA colonies can be generated from secondary primers.

FIGURE 12 illustrates how primers with different sequences can be generated from a surface functionalised with existing primers.

FIGURE 13 depicts methods of preparing DNA fragments suitable for generating DNA colonies.

FIGURE 14 illustrates a method for synthesising cRNA using the DNA colony as a substrate for RNA polymerase.

FIGURE 15 illustrates a preferable method to determine the DNA sequence of DNA present in individual colonies.

FIGURE 16 illustrates a method of determining the sequence of a DNA colony, *de novo*.

FIGURE 17 illustrates the utility of secondary DNA colonies in the assay of mRNA expression levels.

FIGURES 18 and 19 illustrates the use of the secondary DNA colonies in the isolation and identification of novel and rare expressed genes.

IDT Exhibit 1020.0122

A) Scheme Showing The Simultaneous Amplification And Immobilisation Of Nucleic Acid Molecules Using A Single Type Of Primer

Referring now to Figure 1a), a surface is provided having attached thereto a plurality of primers (only one primer is shown for simplicity). Each primer (1) is attached to the surface by a linkage indicated by a dark block. This may be a covalent or a non-covalent linkage but should be sufficiently strong to keep a primer in place on the surface. The primers are shown having a short nucleotide sequence (5'-ATT). In practice however longer sequences would generally be provided.

Figure 1b) shows a target molecule (II) that has annealed to a primer. The target molecule comprises at its 3' end a sequence (5'-ATT) that is complementary to the primer sequence (5'-ATT). At its 5' end the target molecule comprises a sequence (5'-ATT) that is the same as the primer sequence (although exact identity is not required).

Between the two ends any sequence to be amplified (or the complement of any sequence to be amplified) can be provided. By way of example, part of the sequence to be amplified has been shown as 5'-CCG.

In Figure 1c) primer extension is shown. Here a DNA polymerase is used together with dATP, dTTP, dGTP and dCTP to extend the primer (5'-ATT) from its 3' end, using the target molecule as a template.

When primer extension is complete, as shown in Figure 1d), it can be seen that an extended immobilised strand (III) is provided that is complementary to the target molecule. The target molecule can then be separated from the extended immobilised strand (e.g. by heating, as shown in Figure 1e)). This separation step frees the extended, immobilised strand so that it can then be used to initiate a subsequent round of primer extension, as shown in Figures 1f) and 1g)

IDT Exhibit 1020.0123

Here the extended, immobilised strand bends over so that one end of that strand (having the terminal sequence 5'-AAT) anneals with another primer (2, 5'-ATT), as shown in Figure 1f). That primer provides a 3' end from which primer extension can occur, this time using the extended, immobilised strand as a template. Primer extension is shown occurring in Figure 1g) and is shown completed in Figure 1h).

Figure 1i) shows the two extended immobilised strands that were shown in Figure 1h) after separation from one another (e.g. by heating). Each of these strands can then themselves be used as templates in further rounds of primer extension initiated from new primers (3 and 4), as shown in Figures 1j) and 1k). Four single stranded, immobilised strands can be provided after two rounds of amplification followed by a strand separation step (e.g. by heating), as shown in Figure 1l). Two of these have sequences corresponding to the sequence of the target molecule originally used as a template. The other two have sequences complementary to the sequence of the target molecule originally used as a template. (In practice a given immobilised strand and its immobilised complement may anneal once.)

It will therefore be appreciated that a given sequence and its complement can be provided in equal numbers in immobilised form and can be substantially homogeneously distributed within a colony.

Further rounds of amplification beyond those shown in Figure 1 can of course be performed so that colonies comprising large numbers of a given single stranded nucleic acid molecule and a complementary strand thereto can be provided. Only a single template need be used to initiate each colony, although, if desired, a template can be re-used to initiate several colonies.

It will be appreciated that the present invention allows very high densities of immobilised extended nucleic acid molecules to be provided. Within a colony each extended immobilised molecule will be located at a surface within one molecule length of another extended immobilised molecule. Thus position 3 shown in Figure 1l) is within one molecule length of position 1; position 1 is within one molecule

SUBSTITUTE SHEET (RULE 26)

IDT Exhibit 1020.0124

length of position 2; and position 2 is within one molecule length of position 4.

Figure 2 is provided to illustrate how colony growth can occur (using the method described with reference to Figure 1 and to Figure 6 or any other method of the present invention for providing immobilised nucleic acid molecules).

A flat plate is shown schematically in plan view having primers immobilised thereon in a square grid pattern (the primers are indicated by small dots). A regular grid is used solely for simplicity: in many real cases, the positions of the primers might indeed be less ordered or random.

At the position indicated by arrow X a template molecule has annealed to a primer and an initial bout of primer extension has occurred to provide an immobilised, extended nucleic acid strand. Following strand separation, an end of that strand becomes free to anneal to further primers so that additional immobilised, extended nucleic acid strands can be produced. This is shown having occurred sequentially at positions indicated by the letter Y. For simplicity, the primer chosen for annealing is positioned next to the primer carrying the nucleic acid strand: in real cases, the nucleic acid strand could anneal with a primer which is not its next nearest neighbour. However, this primer will obviously be within a distance equal to the length of the nucleic acid strand.

It will be appreciated that annealing at only one (rather than at all) of these positions is required for colony cell growth to occur.

After immobilised, extended, single-stranded nucleic acid molecules have been provided at the positions indicated by letter Y, the resultant molecules can themselves anneal to other primers and the process can be continued to provide a colony comprising a large number of immobilised nucleic acid molecules in a relatively small area.

SUBSTITUTE SHEET (RULE 26)

IDT Exhibit 1020.0125

Figure 3 shows a simplified version of the annealing, elongation and denaturing cycle. It also depicts the typical observations that can be made, as can be seen on the examples shown in figures 4 and 6. The simultaneous amplification and immobilisation of nucleic acids using solid phase primers has been successfully achieved using the procedure described in Examples 1, 2 and 3 below:

EXAMPLE 1

Oligonucleotides, phosphorylated at their 5'-termini (Microsynth GmbH, Switzerland), were grafted onto Nucleolink plastic microtitre wells (Nunc, Roskilde, Denmark). The sequence of the oligonucleotide p57 corresponds to the sequence 5'-TTTTTTCACCAACCCAAACCAACCCAAACC and p58 corresponds to the sequence 5'-TTTTTTAGAAGGAGAAGGAAAGGGAAAGGG. Microtitre wells with p57 or p58 were prepared as follows. In each Nucleolink well, 30 µl of a 160nM solution of the oligonucleotide in 10 mM 1-methyl-imidazole (pH 7.0) (Sigma Chemicals, St.Louis, MO) was added. To each well, 10 µl of 40 mM 1-ethyl-3-(3-dimethylaminopropyl)-carbodiimide (pH 7.0) (Sigma Chemicals) in 10 mM 1-methyl-imidazole, was added to the solution of oligonucleotides. The wells were then sealed and incubated at 50 °C overnight. After the incubation, wells were rinsed twice with 200 µl of RS (0.4N NaOH, 0.25% Tween 20 (Fluka Chemicals, Switzerland)), incubated 15 minutes with 200 µl RS, washed twice with 200 µl RS and twice with 200 µl TNT (100mM TrisHCl pH7.5, 150mM NaCl, 0.1% Tween 20). Tubes were dried at 50°C and were stored in a sealed plastic bag at 4°C.

Colony generation was initiated in each well with 15 µl of priming mix; 1 nanogram template DNA (where the template DNA began with the sequence 5'-AGAAGGAGAAGGAAAGGGAAAGGG and terminated at the 3'-end with the sequence CCCTTTCCCTTTCCTTCTCCTTCT-3'), the four dNTPs (0.2 mM), 0.1% BSA (bovine serum albumin, Boehringer-Mannheim, Germany), 0.1% Tween 20, 8% DMSO (dimethylsulfoxide, Fluka Chemicals, Switzerland), 1X Amplitaq PCR buffer and 0.025 units/µl of AmpliTaq DNA polymerase (Perkin Elmer, Foster City, CA). The priming reaction was a single round of PCR under the following conditions; 94°C for 4 minutes, 60°C for 30 seconds and 72°C for 45 seconds in a

IDT Exhibit 1020.0126

thermocycler (PTC 200, MJ Research, Watertown, MA). Then 100 µl TE buffer (10mM trisHCl, pH 7.5, 1mM EDTA) was used in three successive one minute long washes at 94°C. The DNA colonies were then formed by adding to each well, 20 µl of polymerisation mix, which was identical to the priming mix but lacking the template DNA. The wells were then placed in the PTC 200 thermocycler and colony growing was performed by incubating the sealed wells 4 minutes at 94°C and cycling for 50 repetitions the following conditions: 94°C for 45 seconds, 65°C for 2 minutes, 72°C for 45 seconds. After completion of this program, the wells were kept at 8°C until further use.

A 640 base pair fragment corresponding to the central sequence of the template (but not including the 5'-AGAAGGAGAAGGAAAGGGAAAGGG sequence) was amplified by PCR. The isolated fragment was labelled with biotin-N^4-dCTP (NEN Life Sciences, Boston, MA) and a trace of [α-^{32}P]dCTP (Amersham, UK) using the Prime-it II labelling kit (Stratagene, San Diego, CA) to generate a biotinylated probe.

The biotinylated probe was diluted in to a concentration of 2.5 nM in EasyHyb (Boehringer-Mannheim, Germany) and 15 µl was hybridized to each sample with the following temperature scheme (PTC 200 thermocycler): 94°C for 5 minutes, followed by 500 steps of 0.1°C decrease in temperature every 12 seconds (in other words, the temperature is decreased down to 45°C in 100 minutes). The samples are then washed as follows; 1 time with 2X SSC/0.1% SDS (2X SSC; 0.3M NaCl/0.03M sodium citrate pH7.0/0.001 mg/ml sodium dodecyl sulfate) at room temperature, once with 2X SSC/0.1% SDS at 37°C and once with 0.2X SSC/0.1% SDS at 50°C. The wells are then incubated for 30 minutes with 50 µl of red fluorescent, Neutravidin-coated, 40 nm FluoSpheres(R) (580 nm excitation and 605 nm emission, Molecular Probes Inc., Eugene, OR) in TNT/0.1% BSA. (The solution of microspheres is made from a dilution of 2 µl of the stock solution of microspheres into 1ml of TNT/0.1% BSA, which is then sonicated for 5 minutes in a 50 W ultra-sound water-bath (Elgasonic, Switzerland), followed by filtration through a 0.22 µm filter (Millex GV4). The wells are then counted (Cherenkov) on a Microbeta plate scintillation counter (WALLAC, Turku, Finland).

IDT Exhibit 1020.0127

Excess FluoSpheres(R) are removed by washing for 30 min in TNT/0.1%BSA at room temperature. Images of the stained samples are observed using a 20X objective on an inverted microscope (Axiovert S100TV, Carl Zeiss AG, Oberkochen, Germany) equipped with a Micromax 512x768 CCD camera (Princeton instruments, Trenton, NJ) through a XF43 filter set (PB546/FT580/LP590, Omega Optical, Brattleboro, VT) with a 5 second exposure.

Figure 4 shows the hybridisation results for colony generation on tubes functionalised with either; (a) oligonucleotide p57 or (b) oligonucleotide p58. The control reaction shows very few fluorescent spots, since the sequence of the flanking regions on the template do not correspond to the primer sequences grafted onto the well. In contrast, figure 4b shows the number of fluorescent spots detected when the primers grafted to the wells match the flanking sequences on the initiating DNA template. Calculating the number of fluorescent spots detected and taking into consideration the magnification, we can estimate that there are between 3 and 5 x 10^7 colonies/cm^2. The photos are generated by the program, Winview 1.6.2 (Princeton Instruments, Trenton, NJ) with backgrounds and intensities normalised to the same values.

IDT Exhibit 1020.0128

B) Scheme Showing The Simultaneous Amplification And Immobilisation Of Nucleic Acid Molecules Using Two Different Types of Primer

Referring now to Figure 5, another embodiment of the present invention is illustrated. Here two different immobilised primers are used to provide primer extension.

In this embodiment the target molecule shown is provided with a nucleotide sequence at its 3′ end (AAT-3′) that is complementary to the sequence of a first primer, (5′-ATT, I), which is grafted on the surface, so that annealing with that primer can occur. The sequence (5′-GGT) at the 5′ end of the target molecule, III, corresponds to the sequence (5′-GGT) of a second primer, II, which is also grafted to the surface, so that the sequence which is complementary to the sequence at the 5′ end can anneal with that said second primer. Generally said complementary sequence (5′-ACC) is chosen so that it will not anneal with the first primer (5′-ATT). Unlike the situation described in section A, once the 3′ end of a newly synthesised strand anneals to a primer on the surface, it will have to find a primer whose sequence is different from the sequence it carries at its 5′ end (see the difference between Figures 1f and 5f).

The embodiment shown in Figure 5 has an advantage over the embodiment illustrated in Figure 1 since the possibility of one end of a single stranded target molecule annealing with another end of the same molecule in solution can be avoided and therefore amplification can proceed further. The possibility of annealing occurring between both ends of an immobilised complement to a target molecule can also be avoided.

SUBSTITUTE SHEET (RULE 26)

IDT Exhibit 1020.0129

EXAMPLE 2

A mix of two oligonucleotides which are phosphorylated at the 5'-end (Microsynth GmbH, Balgach, Switzerland) have been grafted on 96 well Nucleolink plates (Nunc, Denmark) as recommended by the manufacturer. The resulting plates has been stored dry at 4°C. The sequence of the primer, P1, was 5'-GCGCGTAATACGACTCACTA, the sequence of the other primer, P2, was 5'-CGCAATTAACCCTCACTAAA. These plates are specially formulated by Nunc, allowing the covalent grafting of 5'-phosphorylated DNA fragments through a standard procedure.

A template has been cloned in a vector (pBlueScript Skminus, Stratagene Inc, San Diego, CA) with the appropriate DNA sequence at the cloning site (i.e., corresponding to P1 and P2 at position 621 and 794 respectively), and 174 bp long linear double stranded DNA template has been obtained by PCR amplification, using P1 and P2. The template PCR product has been purified on Qiagen Qia-quick columns (Qiagen GmbH, Hilden, Germany) in order to remove the nucleotides and the primers used during the PCR amplification.

The purified template (in 50 µl solution containing 1X PCR buffer (Perkin Elmer, Foster City, CA) with the four deoxyribonucleoside triphosphates (dNTPs) at 0.2 mM, (Pharmacia, Uppsala, Sweden) and 2.5 units of AmpliTaq Gold DNA polymerase (Perkin Elmer, Foster City, CA)) has been spread on the support, i.e. on the Nucleolink plates grafted with P1 and P2 (the plates have been rinsed with a solution containing 100 mM TRIS-HCl (pH 7.5), 150 mM NaCl and 0.1% Tween 20 (Fluka, Switzerland) at room temperature for 15 min). This solution has been incubated at 93°C for 9 minutes to activate the DNA polymerase and then 60 cycles (94°C/30 sec.,48°C/30 sec.,72°C/30 sec.) have been performed on a PTC 200 thermocycler. Several different concentrations of PCR template have been tested (approximately 1, 0.5, 0.25, 0.125, 0.0625 ng/µl) and for each sample a control reaction carried out without Taq polymerase has been performed (same conditions as above but without DNA polymerase).

Each sample has been stained with YO-PRO (Molecular Probes, Portland OR), a highly sensitive stain for double stranded DNA. The resulting products have been observed on a confocal microscope using a 40X objective (LSM 410, Carl Zeiss AG, Oberkochen, Germany) with

IDT Exhibit 1020.0130

appropriate excitation (an 488 argon laser) and detection filters (510 low pass filter) (note: the bottom of each well is flat and allows observation with an inverted fluorescence microscope).

In figure 6A, the control well (without added DNA template, panel a) shows only rare objects which can be observed on a blank surface (these objects were useful at this stage for reporting that the focus was correct). These objects have an irregular shape, are 20 to 100 micro-meters in size and have a thickness much larger than the field depth of the observation. In a well where DNA polymerase was present (figure 6A, panel ii), in addition to the objects of irregular shape observed in the control well, a great number of fluorescent spots can be observed. They present a circular shape, they are 1 to 5 micro meters in size and do not span the field of view. The number of spots depends on the concentration of the template used for initiating colony formation. From the observed size of the colonies, one can estimate that more than 10,000 distinct colonies can be arrayed within 1 mm^2 of support.

EXAMPLE 3

Oligonucleotides (Microsynth GmbH, Switzerland) were grafted onto Nucleolink wells (Nunc, Denmark). Oligonucleotide P1 corresponds to the sequence 5'-TTTTTTCTCACTATAGGGCGAATTGG and oligonucleotide P2 corresponds to 5'-TTTTTTCTCACTAAAGGGAACAAAAGCTGG. In each Nucleolink well, a 45 µl of 10 mM 1-methyl-imidazole (pH 7.0) (Sigma Chemicals, St.Louis, MO) solution containing 360 fmol of P1 and 360 fmol of P2 was added. To each well, 15 µl of 40 mM 1-ethyl-3-(3-dimethylaminopropyl)-carbodiimide (pH 7.0) (Sigma Chemicals) in 10 mM 1-methyl-imidazole, was added to the solution of oligonucleotides. The wells were then sealed and incubated at 50°C for 16 hours. After the incubation, wells have been rinsed twice with 200 µl of RS (0.4N NaOH, 0.25% Tween 20), incubated 15 minutes with 200 µl RS, washed twice with 200 µl RS, and twice with 200µl TNT (100mM Tris/HCl pH7.5, 150mM NaCl, 0.1% Tween 20), before they are put to dry at 50°C in an oven. The dried tubes were stored in a sealed plastic bag at 4°C.

IDT Exhibit 1020.0131

Colony growing was initiated in each well with 15 µl of initiation mix (1X PCR buffer, 0.2 mM dNTPs and 0.75 units of AmpliTaq Gold DNA polymerase, 20 nanograms of template DNA, where the template DNA was either S1 DNA or S2 DNA or a mixture of different ratios of S1 DNA and S2 DNA, as indicated in discussion to figure 6B. S1 and S2 are 704 base pair and 658bp fragments, respectively, which have been cloned into pBlueScript Skminus plasmids and subsequently amplified through a PCR using P1 and P2 as primers. The fragments were purified on Qiagen Qia-quick columns (QIAGEN GmbH, Germany) in order to remove the nucleotides and the primers.

Each well was sealed with Cycleseal™ (Robbins Scientific Corp., Sunnyvale, CA), and incubated at 93 °C for 9 minutes, 65°C for 5 minutes and 72°C for 2 minutes and back to 93°C. Then 200 µl TNT solution was used in three successive one minute long washes at 93°C. The initiation mix was then replaced by 15 µl growing mix (same as initiation mix, but without template DNA), and growing was performed by incubating the sealed wells 9 minutes at 93 °C and repeating 40 times the following conditions: 93°C for 45 seconds, 65°C for 3 minutes, 72 °C for 2 minutes. After completion of this program, the wells were kept at 6°C until further use. The temperature control was performed in a PTC 200 thermo-cycler, using the silicon pad provided in the Nucleolink kit and the heated (104°C) lid of the PTC 200.

A 640base pair fragment corresponding to the central sequence of the S1 fragment, but not including the P1 or P2 sequence was amplified by PCR as previously described. The probe was labelled with biotin-16-dUTP (Boehringer-Mannheim, Germany) using the Prime-it II random primer labelling kit (Stratagene, San Diego, CA) according to the manufacturers instructions.

The biotinylated probes were hybridized to the samples in EasyHyb buffer (Boehringer-Mannheim, Germany), using the following temperature scheme (in the PTC 200 thermocycler): 94°C for 5 minutes, followed by 68 steps of 0.5°C decrease in temperature every 30 seconds (in other words, the temperature is decreased down to 60 °C in 34 minutes), using sealed wells. The samples are then washed 3 times with 200µl of TNT at room temperature. The wells are then

IDT Exhibit 1020.0132

incubated for 30 minutes with 50 µl TNT containing 0.1 mg/ml BSA. Then the wells are incubated 5 minutes with 15µl of solution of red fluorescent, Neutravidin-coated, 40 nm FluoSpheres(R) (580 nm excitation and 605 nm emission, Molecular Probes, Portland, OR). The solution of microspheres is made of 2 µl of the stock solution of microspheres, which have been sonicated for 5 minutes in a 50 W ultra-sound water-bath (Elgasonic, Bienne, Switzerland), diluted in 1 ml of TNT solution containing 0.1 mg/ml BSA and filtered with Millex GV4 0.22 µm pore size filter (Millipore, Bedford, MA).

The stained samples are observed using an inverted Axiovert 10 microscope using a 20X objective (Carl Zeiss AG, Oberkochen, Germany) equipped with a Micromax 512x768 CCD camera (Princeton Instruments, Trenton, NJ), using a XF43 filter set (PB546/FT580/LP590, Omega Optical, Brattleboro, VT), and 10 seconds of light collection. The files are converted to TIFF format and processed in the suitable software (PhotoPaint, Corel Corp., Ottawa, Canada). The processing consisted in inversion and linear contrast enhancement, in order to provide a picture suitable for black and white print-out on a laser printer.

The figure 6B shows the results for 3 different ratios of the S1/S2 templates used in the initiating reaction: i) the S1/S2 is 1/0, many spots can be observed, ii) the S1/S2 is 1/10, and the number of spots is approximately 1/10 of the number of spots which can be observed in the i) image, as expected, and iii) the S1/S2 is 0/1, and only a few rare spots can be seen.

IDT Exhibit 1020.0133

C Scheme Showing Simultaneous Amplification And Immobilisation Of Nucleic Acid Molecules When The Target Molecule Contains Internal Sequences Complementary To The Immobilised Primers

Figure 7 is provided to show that the sequences shown at the 5′ and 3′ ends of the target molecule illustrated in Figures 1 and 5 need not be located at the ends of a target molecule.

A target nucleic acid molecule (II) may have a sequence at each (or either) end that is neither involved in annealing with a primer nor in acting as a template to provide a complementary sequence that anneals with a primer (sequence 5′-AAA and sequence 5′-CCC). One of the internal sequences (5′-AAT) is used as a template to synthesise a complementary sequence, III, thereto (5′-TTT), as is clear from figures 7(a) to 7(e).

The sequence 5′-TTT is not however itself used to provide a sequence complementary thereto, as is clear from figures 7(f) to 7(k). It can be seen from Figure 7(l) that only one of the four immobilised strands shown after two rounds of primer extension and a strand separation step comprises the additional sequence 5′-TTT and that no strand comprising a complementary sequence (5′-AAA) to this sequence is present (i.e. only one strand significantly larger than the others is present). After several rounds of amplification the strand comprising the sequence 5′-TTT will represent an insignificant proportion of the total number of extended, immobilised nucleic acid molecules present.

SUBSTITUTE SHEET (RULE 26)

IDT Exhibit 1020.0134

D) Using Nucleic Acid Strands Present in Colonies to Synthesise Additional Copies of Nucleic Acid Strands

Amplified, single stranded nucleic acid molecules present in colonies provided by the present invention can themselves be used as templates to synthesise additional nucleic acid strands.

Figure 8 illustrates one method of synthesising additional nucleic acids using immobilised nucleic acids as a starting point.

Colonies will usually comprise both a given nucleic acid strand and its complement in immobilised form (figure 8a). Thus they can be used to provide additional copies not only of a given nucleic acid strand but also of its complement.

One way of doing this is to provide one or more primers (primers TTA and TGG) in solution that anneal to amplified, immobilised nucleic acid strands present in colonies (figure 8c) provided by the present invention. (These primers may be the same as primers initially used to provide the immobilised colonies, apart from being provided in free rather than immobilised form.) The original DNA colony is denatured by heat to it single-stranded form (figure 8b), allowing primers TTA and TGG to anneal to the available 3' end of each DNA strand. Primer extension, using AmpliTaq DNA polymerase and the four deoxyribonucleoside triphosphates (labelled or unlabeled) can then be used to synthesise complementary strands to immobilised nucleic acid strands or at least to parts thereof (step (iii)).

Once newly formed strands (figure 8d) have been synthesised by the process described above, they can be separated from the immobilised strands to which they are hybridised (e.g. by heating). The process can then be repeated if desired using the PCR reaction, to provide large number of such strands in solution (figure 8e).

SUBSTITUTE SHEET (RULE 26)

IDT Exhibit 1020.0135

Strands synthesised in this manner, after separation from the immobilised strands, can, if desired, be annealed to one another (i.e. a given strand and its complement can anneal) to provide double-stranded nucleic acid molecules in solution. Alternatively they can be separated from one another to provide homogenous populations of single-stranded nucleic acid molecules in solution.

It should also be noted that once single-stranded molecules are provided in solution they can be used as templates for PCR (or reverse PCR). Therefore it is not essential to continue to use the immobilised nucleic acid strands to obtain further amplification of given strands or complementary strands thereto.

It should be noted that where a plurality of colonies are provided and nucleic acid strands in different colonies have different sequences, it is possible to select only certain colonies for use as templates in the synthesis of additional nucleic acid molecules. This can be done by using primers for primer extension that are specific for molecules present in selected colonies.

Alternatively primers can be provided to allow several or all of the colonies to be used as templates. Such primers may be a mixture of many different primers (e.g. a mixture of all of the primers originally used to provide all of the colonies, but with the primers being provided in solution rather than in immobilised form).

EXAMPLE 4

Oligonucleotides (Microsynth GmbH Balgach, Switzerland) were grafted onto Nucleolink wells (Nunc, Denmark). Oligonucleotide P1 corresponds to the sequence 5'-TTTTTTTTTTCACCAACCCAAACCAACCCAAACC and oligonucleotide P2 corresponds to 5'-TTTTTTTTTTAGAAGGAGAAGGAAAGGGAAAGGG. In each Nucleolink well, a 45 µl of 10 mM 1-methyl-imidazole (pH 7.0) (Sigma Chemicals) solution containing 360 fmol of P1 and 360 fmol of P2 was added. To each

IDT Exhibit 1020.0136

well, 15 µl of 40 mM 1-ethyl-3-(3-dimethylaminopropyl)-carbodiimide (pH 7.0) (Sigma Chemicals) in 10 mM 1-methyl-imidazole, was added to the solution of oligonucleotides. The wells were then sealed and incubated at 50 °C for 16 hours. After the incubation, wells have been rinsed twice with 200 µl of RS (0.4N NaOH, 0.25% Tween 20), incubated 15 minutes with 200 µl RS, washed twice with 200 µl RS, and twice with 200µl TNT (100mM Tris/HCl pH7.5, 150mM NaCl, 0.1% Tween 20), before they are put to dry at 50°C in an oven. The dried tubes were stored in a sealed plastic bag at 4°C.

Colony growing was initiated in each well with 15 µl of initiation mix (1X PCR buffer, 0.2 mM dNTPs and 0.75 units of AmpliTaq DNA polymerase, 20 nanograms of template DNA, where the template DNA was either S1 DNA or S2 DNA or a 1/1 mixture of S1 DNA and S2 DNA, as indicated in discussion to Example 3. S1 and S2 are 658 base pair and 704 b.p. fragments, respectively, which have been prepared as described in EXAMPLE 3.

Each well was sealed with Cycleseal™ (Robbins Scientific Corp., Sunnyvale, CA) and incubated at 93°C for 9 minutes, 65°C for 5 minutes and 72°C for 2 minutes and back to 93°C. Then 200 µl TNT solution was used in three successive one minute long washes at 93°C. The initiation mix was then replaced by 15 µl growing mix (same as initiation mix, but without template DNA) and growing was performed by incubating the sealed wells 9 minutes at 93°C and repeating 40 times the following conditions: 93°C for 45 seconds, 65°C for 3 minutes, 72°C for 2 minutes. After completion of this program, the wells were kept at 6°C until further use. The temperature control was performed in a PTC 200 thermo-cycler.

Different treatments where applied to 6 sets (A,B,C,D,E and F)of 3 wells (1,2,3), one prepared with template S1, one with template S1 and template S2 and one prepared with template S2 alone (yielding A1,A2,A3,…,F1,F2,F3). The set A was left untreated, set B has been incubated for 10 minutes with BAL-31 exonuclease (New England Biolabs, Beverly, MA) at 37°C in BAL-31 buffer (BAL-31 essentially digests double stranded DNA which has both ends free), set C has been incubated for 10 minutes with S1 nuclease (Pharmacia, Uppsala, Sweden) at 37°C in S1 buffer (S1 nuclease essentially digests single

IDT Exhibit 1020.0137

stranded DNA), set D, E and F have been incubated with both BAL-31 and S1 nucleases. Reactions were stopped by rinsing the wells with TNT buffer.

PCR (25 cycles, 30 sec. at 94°C , 45 sec. at 60°C , 45 sec. at 72°C) has been performed in the Nucleolink wells with 0.25 µM primers P70 (5'-CACCAACCCAAACCAACCCAAACCACGACTCACTATAGGGCGAA) and P71 (5'-AGAAGGAGAAGGAAAGGGAAAGGGTAAAGGGAACAAAAGCTGGA) in solution in sets A, B, C and D. P70 and P71 are suited for the amplification of both S1 and S2, since primer P70 contains the sequence of primer P1 and p71 contains P2. In the set E wells, PCR has been performed with a set of forward (P150, 5'-GGTGCTGGTCCTCAGTCTGT) and reverse (P151, 5'-CCCGCTTACCAGTTTCCATT) primers which are within S1 and not within S2 so as to produce a 321 bp PCR product, and in the set F wells, PCR has been performed with a set of forward (P152, 5'-CTGGCCTTATCCCTAACAGC) and reverse (P153, 5'-CGATCTTGGCTCATCACAAT) primers which are within S2 and not within S1 so as to produce a 390 bp PCR product. For each of the 18 PCR reactions, 3 µl of solution have been used for gel electrophoresis on 1% agarose in presence of 0.1 µg/ml ethidium-bromide. The pictures of the gels are presented in Figure 9. These pictures show that DNA in the colonies is protected from exonuclease digestion (sets B, C and D as compared to set A), and that both S1 and S2 can be recovered either simultaneously using P1 and P2 (sets A, B, C and D) or specifically (set E and F). In set E and F, where the shorter PCR products are more efficiently amplified than the longer PCR products in sets A, B, C, D, a cross-contamination between the S1 and S2 templates is detectable (see lane E2 and F1).

SUBSTITUTE SHEET (RULE 26)

IDT Exhibit 1020.0138

E) Provision of Secondary Colonies

It is also possible to modify initially formed colonies to provide different colonies (i.e. to provide colonies comprising immobilised nucleic acid molecules with different sequences from those molecules present in the initially formed colonies). Here, the initially formed colonies are referred to as "primary colonies" and the later formed colonies as "secondary colonies". A preliminary procedure is necessary to turn the primary colonies into "secondary primers" which will be suitable for secondary colony generation.

Figure 10 shows how 'secondary primers' are generated using existing primary colonies. As a starting point, the primary colony (figure 10a) is left in the fully hybridised, double-stranded form. A single-strand specific DNA exonuclease might be used to remove all primers which have not been elongated. One could also choose to cap all free 3'-OH ends of primers with dideoxyribonucleotide triphosphates using a DNA terminal transferase (step (i), figure 10b).

Secondly and independently, the DNA molecules forming the colonies can be cleaved by using endonucleases. For example, a restriction enzyme that recognises a specific site within the colony (depicted by the 'RE' arrow in figure 10c) and cleaves the DNA colony (step (ii), figure 10). If desired, the enzymatically cleaved colony (figure 10d) can then be partially digested with a 3' to 5' double-strand specific exonuclease (e.g. *E.coli* exonuclease III, depicted by 'N', step (iii), figure 10d). In any case, the secondary primers are available after denaturation (e.g., by heat) and washing (figure 10e).

Alternatively, the double stranded DNA forming the colonies (figure 10f) can be digested with the double-strand specific 3'-5' exonuclease, which digest only one strand of double stranded DNA. An important case is when the exonuclease digests only a few bases of

IDT Exhibit 1020.0139

the DNA molecule before being released in solution, and when digestion can proceed when another enzyme binds to the DNA molecule (figure 10g). In this case the exonuclease digestion will proceed until there remain only single stranded molecules which, on average, are half the length of the starting material, and are without any complementary parts (which could form partial duplexes) remaining in the single stranded molecules in a colony (figure 10h).

In all cases, these treatments result in single-stranded fragments grafted onto a support which correspond to the sequence of the original template and that can be used for new DNA colony growing if an appropriate new template is provided for colony initiation (figures 10e and 10h).

The result of such a treatment, thus a support holding secondary primers, will be referred to as a "support for secondary colony growing". Templates useful for secondary colony growing may include molecules having known sequences (or complements of such sequences). Alternatively templates may be derived from unsequenced molecules (e.g. random fragments). In either event the templates should be provided with one or more regions for annealing with nucleic acid strands present in the primary colonies.

Figure 11 a-e show how a secondary colony can be generated when an appropriate template (**TP**, figure 11a)is provided for a second round of DNA colony generation on a support for secondary colony growing, holding secondary primers. In this example, treatment of the primary colony as described above has generated the secondary primers, **SP1** and **SP2** (figure 11a). The template **TP**, will hybridise to its complementary secondary primer, **SP1**, and following an extension reaction using a DNA polymerase as described, will be extended as depicted (figure 12b). Following a denaturing (step ii), reannealing (step iii) and DNA polymerase (step iv) cycle, a replica of the original primary colony will be formed (figure 11e).

IDT Exhibit 1020.0140

The maximum size of a secondary colony provided by this embodiment of the present invention is restricted by the size of the primary colony onto which it grows. Several secondary growing processes can be used sequentially to create colonies for specific applications (i.e. a first colony can be replaced with a second colony, the second colony can be replaced with a third colony, etc.)

SUBSTITUTE SHEET (RULE 26)

IDT Exhibit 1020.0141

F) Provision of Extended Primers

Figure 12 shows how extended primers can be generated on an array of oligonucleotides. The same procedure could be applied to a support covered with colonies or secondary primers as described in section E.

In Figure 12a) a support is provided having a plurality of immobilised primers shown thereon. Different immobilised primers are shown present in different regions of the support (represented by squares). Primers having the sequence 5'-AAA are present in one square and primers having the sequence 5'-GGG are present in another square.

Figures 12b) to 12d) show how the initial primers present (initial primers) are modified to give different primers (extended primers). In this example, those initial primers having the sequence 5'-AAA are modified to produce two different types of extended primers, having the sequences 5'-AAAGCC and 5'-AAATAC respectively. This is achieved by the hybridisation of oligonucleotide templates, 5'-GTATTT and 5'-GGCTTT to the primary primers immobilised on the surface (figure 12b), followed by DNA polymerase reaction. Those initial primers having the sequence 5'-GGG are modified to produce two different types of extended primers, having the sequences 5'-GGGTAT and 5'-GGGTAA (figure 12d) in a similar manner.

The technique of producing extended primers is useful for transforming immobilised oligonucleotides provided on a DNA chip or other surface into immobilised primers useful in amplifying a particular target nucleic acid sequence and/or in amplifying a complementary strand thereto.

IDT Exhibit 1020.0142

G) Preparation of nucleic acid fragments

Apparatuses of the present invention can be used for various procedures some of which will be described later on. Nucleic acid fragments for use in colony generation may be prepared differently for the different procedures (referred to herein as "prepared nucleic acids"). Various preparation procedures are described below:

(i) Preparation of random DNA fragments

Here is described a method to prepare DNA originating from one biological sample (or from a plurality of samples) for amplification in the case where it is not necessary to keep track of the origin of the DNA when it is incorporated within a colony.

The DNA of interest is first extracted from the biological sample and cut randomly into "small" pieces (e.g., 50 to 10,000 bases long, but preferentially 500 to 1000 base pairs in length, represented by bar '**I**', figure 13a). (This can be done e.g., by a phenol-chloroform extraction followed by ultrasound treatment, mechanical shearing, by partial digestion with frequent cutter restriction endonucleases or other methods known by those skilled in the art). In order to standardise experimental conditions, the extracted and cut DNA fragments can be size-fractionated, e.g., by agarose gel electrophoresis, sucrose gradient centrifugation or gel chromatography. Fragments obtained within a single fraction can be used in providing templates in order to reduce the variability in size of the templates.

Secondly, the extracted, cut and (optionally) sorted template DNA fragments can be ligated with oligonucleotide linkers (**IIa** and **IIb**, figure 13a) containing the sequence of the primer(s) which have previously been grafted onto a support. This can be achieved, for instance, using "blunt-end" ligation. Alternatively, the template

SUBSTITUTE SHEET (RULE 26)

IDT Exhibit 1020.0143

DNA fragments can be inserted into a biological vector at a site that is flanked by the sequence of the primers that are grafted on the support. This cloned DNA can be amplified within a biological host and extracted. Obviously, if one is working with a single primer grafted to the solid support for DNA colony formation, purifying fragments containing both P1 and P2 primers does not pose a problem.

Hereafter, the DNA fragments obtained after such a suitable process are designated by the expression : "prepared genomic DNA" (**III**, figure 13a).

(ii) Preparation of random DNA fragments originating from a plurality of samples

Here it is described how to prepare DNA originating from a plurality of biological samples in the case where it is necessary to keep track of the origin of the DNA when it is incorporated within a colony.

The procedure is the same as that described in the previous section except that in this case, the oligonucleotide linkers used to tail the randomly cut genomic DNA fragments are now made of two parts; the sequence of the primers grafted onto the surface (P1 and P2, figure 13b) and a "tag" sequence which is different for each sample and which will be used for identifying the origin of the DNA colony. Note that for each sample, the tag may not be unique, but a plurality of tags could be used. Hereafter, we will designate the DNA fragments obtained after such a suitable process by the expression : "tagged genomic DNA" (**III**, Figure 13b).

This tagging procedure can be used for providing colonies carrying a means of identification which is independent from the sequence carried by the template itself. This could also be useful when some colonies are to be recovered specifically (using the procedure given in section D). This could also be useful in the case the recovered colonies are further processed, e.g., by creating new primary

SUBSTITUTE SHEET (RULE 26)

IDT Exhibit 1020.0144

colonies and a cross reference between the original colonies and the new colonies is desired.

(iii) Preparation of DNA fragments corresponding to a plurality of DNA sequences originating from one sample

The DNA of interest can first be extracted from a biological sample by any means known by those skilled in the art (as mentioned supra). Then the specific sequences of interest can be amplified with PCR (step (i), figure 13c) using PCR primers (**IIa** and **IIb**) made of two parts; 1) at the 5'-end, the sequences corresponding to the sequences of primer oligonucleotide(s) that have been grafted onto a surface (P1 and P2)and 2) at the 3'-end, primer sequences specific to the sequence of interest (S1 and S2). Hereafter, we will designate the DNA fragments obtained after such a suitable process by the expression: "prepared DNA" (**III**, figure 13c).

(iv) Preparation of a plurality of DNA fragments originating from a plurality of samples

The procedure is the same as in the previous section except that in this case the DNA primers (**IIa** and **IIb**) used to perform the PCR amplification (step (i), figure 13d) are now made of three parts; 1) the sequence of the primers grafted onto the surface (P1 and P2), 2) a "tag" sequence which is different for each sample and which will be used for the identifying the origin of the DNA colony and 3) primer sequences surrounding the specific sequence of interest (S1 and S2). Note that for each sample, a plurality of tags might be used, as in (ii) *supra*.

Hereafter, we will designate the DNA fragments obtained after such a suitable process by the expression : "tagged DNA" (**III**, figure 13d). Potential uses of tags are the same as in (ii), *supra*.

SUBSTITUTE SHEET (RULE 26)

IDT Exhibit 1020.0145

(v) Preparation of mRNA

The procedure is similar to the procedures described for preparing DNA fragments in the previous sections except that the starting point is to extract mRNA by any means known to those skilled in the art (e.g., by use of commercially available mRNA preparation kits). The mRNA can be copied into double-stranded cDNA by any means known to those skilled in the art (e.g. by using a reverse transcriptase and a DNA polymerase). Certainly, the tags and primers described *supra* can be used in conjunction with the process of double-stranded cDNA synthesis to allow their incorporation into the templates.
Hereafter, we will designate the mRNA fragments obtained after such suitable processes by the expressions: "prepared total mRNA" (cf. "prepared genomic DNA", as described in section (I) *supra*), "tagged total mRNA", (cf. "tagged genomic DNA", as described in section (ii) *supra*), "prepared mRNA" (cf. "prepared DNA", as described in section (iii) *supra*) and "tagged mRNA" (cf. "tagged DNA", as described in section (iv) *supra*).

SUBSTITUTE SHEET (RULE 26)

IDT Exhibit 1020.0146

H) Preferred Detection Assays

In assay procedures of the present invention labels may be used to provide detectable signals. Examples include:

a) a fluorescent group or a energy-transfer based fluorescence system.

b) a biotin based system. In this case colonies can be incubated with streptavidin labelled with a fluorescent group or an enzyme (e.g. fluorescent latex beads coated with streptavidin; streptavidin labelled with fluorescent groups; enzymes for use with the corresponding fluorescence assay).

c) a system based on detecting an antigen or a fragment thereof - e.g. a hapten (including biotin and fluorescent groups). In this case colonies can be incubated with antibodies (e.g. specific for a hapten). The antibodies can be labelled with a fluorescent group or with an enzyme (e.g. fluorescent latex beads coated with the antibody; antibodies labelled with fluorescent groups; antibodies linked to an enzyme for use with a corresponding fluorescence or luminescence assay, etc.).

d) a radio-label (e.g. incorporated by using a 5'-polynucleotide kinase and [γ-^{32}P]adenosine triphosphate or a DNA polymerase and [α-^{32}P or α-^{33}P] deoxyribonucleoside triphosphates to add a radioactive phosphate group(s) to a nucleic acid). Here colonies can be incubated with a scintillation liquid.

e) a dye or other staining agent.

Labels for use in the present invention are preferably attached

a) to nucleic acids

b) to proteins which bind specifically to double stranded DNA (e.g., histones, repressors, enhancers)

and/or

c) to proteins which bind specifically to single stranded DNA (e.g. single-stranded nucleic acid binding protein).

SUBSTITUTE SHEET (RULE 26)

IDT Exhibit 1020.0147

Labelled colonies are preferably detected by:

a) measuring fluorescence.

b) measuring luminescence.

c) measuring radioactivity

d) measuring flow or electric field induced fluorescence anisotropy.

and/or

e) measuring the polymer layer thickness.

Staining agents can be used in the present invention. Thus DNA colonies can be incubated with a suitable DNA-specific staining agent, such as the intercalating dyes, ethidium bromide, YO-YO, YO-PRO (Molecular Probes, Eugene, OR).

With certain staining agents the result can be observed with a suitable fluorescence imaging apparatus.

Examples of particular assays/procedures will now be described in greater detail:

IDT Exhibit 1020.0148

I) Preferred Embodiments of Assays of the Present Invention

(i) Nucleic acid probe hybridisation assay

DNA colonies are first prepared for hybridisation. Then they are hybridised with a probe (labelled or unlabelled). If required, the hybridised probed is assayed, and the result is observed. This can be done with an apparatus of the present invention (e.g. as described *supra*).

Preparation for hybridisation

In a preferred embodiment of the present invention colonies are treated with a DNA restriction endonuclease which is specific either for a sequence provided by a double stranded form of one of the primers originally grafted onto the surface where colonies are formed or for another sequence present in a template DNA molecule (see e.g. figure 16c).

After restriction enzyme digestion, the colonies can be heated to a temperature high enough for double stranded DNA molecules to be separated. After this thermal denaturing step, the colonies can be washed to remove the non-hybridised, detached single-stranded DNA strands, leaving a remaining attached single-strand DNA.

In another embodiment the colonies can be partially digested with a double-strandspecific 3′ to 5′ DNA exonuclease (see section E, figure 10f) which removes one strand of DNA duplexes starting from the 3′ end, thus leaving a part of a DNA molecule in single stranded form.

Alternatively, DNA in colonies can first be heat denatured and then partially digested with an single-strand specific 3′ to 5′ DNA exonuclease which digests single stranded DNA starting from the 3′ end.

SUBSTITUTE SHEET (RULE 26)

IDT Exhibit 1020.0149

A further alternative is simply to heat denature DNA in the colonies.

Hybridisation of the probe

Single-stranded nucleic acid probes (labelled or unlabelled) can be hybridised to single-stranded DNA in colonies at the appropriate temperature and buffer conditions (which depends on the sequence of each probe, and can be determined using protocols known to those skilled in the art).

Assaying of unlabelled hybridised probes

A hybridised probe provided initially in unlabelled form can be used as a primer for the incorporation of the different (or a subset of the different) labelled (or a mix of labelled and unlabelled) deoxyribonucleoside triphosphates with a DNA polymerase. The incorporated labelled nucleotides can then be detected as described *supra*.

Cyclic assaying of labelled or unlabelled probes

Firstly, the DNA colonies can be prepared for hybridisation by the methods described *supra*. Then they can be hybridised with a probe (labelled or initially unlabelled). If required, hybridised labelled probes are assayed and the result is observed with an apparatus as described previously. The probe may then be removed by heat denaturing and a probe specific for a second DNA sequence may be hybridised and detected. These steps maybe repeated with new probes as many times as desired.

Secondly, the probes can be assayed as described *supra* for unlabelled probes, except that only a subset (preferably 1 only) of the different (labelled or unlabelled) nucleotides are used at each cycle. The colonies can then be assayed for monitoring the incorporation of the nucleotides. This second process can be repeated until a sequence of a desired length has been determined.

SUBSTITUTE SHEET (RULE 26)

IDT Exhibit 1020.0150

(ii) *in situ* RNA Synthesis assay

In this embodiment, DNA colonies can be used as templates for *in situ* RNA synthesis as depicted in figure 14. DNA colonies can be generated from templates and primers, such that a RNA polymerase promoter sequence is positioned at one end of the double-stranded DNA in the colony. DNA colonies can then be incubated with RNA polymerase and the newly synthesised RNA (cRNA) can be assayed as desired. The detection can be done non-specifically (e.g., staining) or in a sequence dependent way (e.g., hybridisation).

The DNA template (**I**, figure 14a) to be amplified into a colony is generated by PCR reaction using primers (**IIa** and **IIb**) which have the following four parts; 1) sequence identical to the sequences of the primers grafted onto the surface ('P1' and 'P2'), 2) a "tag" sequence which is different for each sample , a sequence corresponding to a RNA polymerase promoter, ie. the T3, T7 and SP6 RNA promoters, ('RPP', figure 14a) and 4) primer sequences surrounding the specific sequence of interest ('S1' and 'S2'). Hereafter, we will designate the DNA fragments obtained after such a suitable process by the expression: "tagged RNA synthesis DNA" (**III**, figure 14b).

After amplification of the DNA template from the original DNA sample, these templates are used to generate DNA colonies. The DNA colonies (IV, figure 14c) are then incubated with the RNA polymerase specific for the RNA polymerase promoter ('RPP', figure 14c). This will generate a copy of RNA specific for the DNA colony template (Template-cRNA, **V**, figure 14d).

cRNA thus synthesised can be isolated and used as hybridisation probes, as messenger RNA (mRNA) templates for *in vitro* protein synthesis or as templates for *in situ* RNA sequence analysis.

SUBSTITUTE SHEET (RULE 26)

IDT Exhibit 1020.0151

(iii) Methods for sequencing

In another embodiment of the present invention, colonies can be analysed in order to determine sequences of nucleic acid molecules which form the colonies. Since very large numbers of the same nucleic acid molecules can be provided within each colony the reliability of the sequencing data obtained is likely to be very high.

The sequences determined may be full or partial. Sequences can be determined for nucleic acids present in one or more colonies. A plurality of sequences may be determined at the same time.

In some embodiments the sequence of a complementary strand to a nucleic acid strand to be sequenced (or of a part thereof) may be obtained initially. However this sequence can be converted using base-pairing rules to provide the desired sequence (or a part thereof). This conversion can be done via a computer or via a person. It can be done after each step of primer extension or can be done at a later stage.

Sequencing can be done by various methods. For example methods relying on sequential restriction endonuclease digestion and linker ligation can be used. One such method is disclosed in WO95/27080 for example. This method comprises the steps of: ligating a probe to an end of a polynucleotide, the probe having a nuclease recognition site; identifying one or more nucleotides at the end of the polynucleotide; and cleaving the polynucleotide with a nuclease recognising the nuclease recognition site of the probe such that the polynucleotide is shortened by one or more nucleotides.

However in a preferred method of the present invention, amplified nucleic acid molecules (preferably in the form of colonies, as disclosed herein) are sequenced by allowing primers to hybridise with the nucleic acid molecules, extending the primers and detecting the nucleotides used in primer extension. Preferably, after extending a primer by a single nucleotide, the nucleotide is detected before a

IDT Exhibit 1020.0152

further nucleotide is used in primer extension (step-by-step sequencing).

One or more of the nucleotides used in primer extension may be labelled. The use of labelled nucleotides during primer extension facilitates detection. (The term "label" is used in its broad sense to indicate any moiety that can be identified using an appropriate detection system. Preferably the label is not present in naturally-occurring nucleotides.) Ideally, labels are non-radioactive, such as fluorophores. However radioactive labels can be used.

Where nucleotides are provided in labelled form the labels may be the same for different nucleotides. If the same label is used each nucleotide incorporation can be used to provide a cumulative increase of the same signal (e.g. of a signal detected at a particular wavelength). Alternatively different labels may be used for each type of nucleotide (which may be detected at different wavelengths).

Thus four different labels may be provided for dATP, dTTP, dCTP and dGTP, or the same label may be provided for them all. Similarly, four different labels may be provided for ATP, UTP,CTP and GTP, or the same label may be provided for them all). In some embodiments of the present invention a mixture of labelled and unlabelled nucleotides may be provided, as will be described in greater detail later on.

In a preferred embodiment of the present invention the sequencing of nucleic acid molecules present in at least 2 different colonies is performed simultaneously. More preferably, sequencing of nucleic acid molecules present in over 10, over 100, over 1000 or even over 1,000,000 different colonies is performed simultaneously. Thus if colonies having different nucleic acids molecules are provided, many different sequences (full or partial) can be determined simultaneously - i.e. over 10, over 100, over 1000 or even over 1,000,000 different sequences may be determined simultaneously.

If desired, controls may be provided, whereby a plurality of colonies comprising the same nucleic acid molecules are provided. By

IDT Exhibit 1020.0153

determining whether or not the same sequences are obtained for nucleic acid molecules in these colonies it can be ascertained whether or not the sequencing procedure is reliable.

One sequencing method of the present invention is illustrated in figure 17, which is entitled "*in situ* sequencing". On prepared DNA colonies hybridised with an appropriate sequencing primer, cyclic addition of the individual deoxyribonucleoside triphosphates and DNA polymerase will allow the determination of the DNA sequence immediately 3′ to the sequencing primer. In the example outlined in figure 17, the addition of dGTP allows the determination of colony 1 to contain a 'G'. In the second cycle addition of dATP is detected in both colonies, determining that both colonies have an 'A' in the next position. After several repetitions of the addition of single deoxyribonucleoside triphosphates, it will be possible to determine any sequence. For example sequences of at least 10, at least 20, at least 50 or at least 100 bases may be determined.

If colonies are provided initially in a form comprising double-stranded molecules the colonies can be processed to provide single-stranded molecules for use in sequencing as described above. (It should however be noted that double stranded molecules can be used for sequencing without such processing. For example a double stranded DNA molecule can be provided with a promoter sequence and step-by-step sequencing can then be performed using an RNA polymerase and labelled ribonucleotides (cf Figure 16 d)). Another alternative is for a nick to be introduced in a double stranded DNA molecule so that nick translation can be performed using labelled deoxyribonucleotides. and a DNA polymerase with 5′ to 3′ exonuclease activity.)

One way of processing double-stranded molecules present in colonies to provide single-stranded colonies as described later with reference to Figure 19. Here double-stranded immobilised molecules present in a colony (which may be in the form of bridge-like structures) are

IDT Exhibit 1020.0154

cleaved and this is followed by a denaturing step. (Alternatively a denaturing step could be used initially and could be followed by a cleavage step). Preferably cleavage is carried out enzymatically. However other means of cleavage are possible, such as chemical cleavage. (An appropriate cleavage site can be provided in said molecule). Denaturing can be performed by any suitable means. For example it may be performed by heating and / or by changing the ionic strength of a medium in the vicinity of the nucleic acid molecules.

Once single-stranded molecules to be sequenced are provided, suitable primers for primer extension can be hybridised thereto. Oligonucleotides are preferred as primers. These are nucleic acid molecules that are typically 6 to 60, e.g. 15 to 25 nucleotides long. They may comprise naturally and / or non-naturally occurring nucleotides. (However other molecules, e.g. longer nucleic acid strands may alternatively be used as primers, if desired.) The primers for use in sequencing preferably hybridise to the same sequences present in amplified nucleic acid molecules as do primers that were used to provide said amplified nucleic acids. (Primers having the same / similar sequences can be used for both amplification and sequencing purposes).

When primers are provided in solution and are annealed (hybridised) to nucleic acid molecules present in colonies to be sequenced, those primers which remain in solution or which do not anneal specifically can be removed after annealing. Preferred annealing conditions (temperature and buffer composition) prevent non-specific hybridisation. These may be stringent conditions. Such conditions would typically be annealing temperatures close to a primer's Tm (melting temperature) at a given salt concentration (e.g. 50 nM primer in 200 mM NaCl buffer at 55°C for a 20-mer oligonucleotide with 50% GC content). (Stringent conditions for a given system can be determined by a skilled person. They will depend on the base composition, GC content, the length of the primer used and the salt concentration. For a 20 base oligonucleotide of 50% GC, calculated average annealing temperature is 55-60°C, but in practice may vary between 35 to 70°C).

SUBSTITUTE SHEET (RULE 26)

IDT Exhibit 1020.0155

Primers used for primer extension need not be provided in solution, since they can be provided in immobilised form. In this embodiment the primers should be provided in the vicinity of the immobilised molecules to which they are to be annealed. (Such primers may indeed already be present as excess immobilised primers that were not used in amplifying nucleic acid molecules during the formation of colonies.)

The nucleic acid molecules present in colonies to be sequenced will include a sequence that hybridises to the primers to be used in sequencing (preferably under "stringent" conditions). This portion can be added to a given molecule prior to amplification (which molecule may have a totally / partially unknown sequence) using techniques known to those skilled in the art. For example it can be synthesised artificially and can be added to a given molecule using a ligase.

Once a nucleic acid molecule annealed to a primer is provided, primer extension can be performed. RNA or DNA polymerases can be used. DNA polymerases are however the enzymes of choice for preferred embodiments. Several of these are commercially available. Polymerases which lack 3' to 5' exonuclease activity can be used, such as T7 DNA polymerase or the small (Klenow) fragment of DNA polymerase I may be used [*e.g.* the modified T7 DNA polymerase Sequenase™ 2.0 (Amersham) or Klenow fragment (3' to 5' exo-, New England Biolabs)]. However it is not essential to use such polymerases. Indeed, where it is desired that the polymerases have proof-reading activity polymerases lacking 3' to 5' exonuclease activity would not be used.Certain applications may require the use of thermostable polymerases such as ThermoSequenase™ (Amersham) or Taquenase™ (ScienTech, St Louis, MO). Any nucleotides may be used for primer extension reactions (whether naturally occurring or non-naturally occurring). Preferred nucleotides are deoxyribonucleotides; dATP, dTTP, dGTP and dCTP (although for some applications the dTTP analogue dUTP is preferred) or ribonucleotides ATP, UTP, GTP and CTP; at least some of which are provided in labelled form.

IDT Exhibit 1020.0156

A washing step is preferably incorporated after each primer extension step in order to remove unincorporated nucleotides that may interfere with subsequent steps. The preferred washing solution should be compatible with polymerase activity and have a salt concentration that does not interfere with the annealing of primer molecules to the nucleic acid molecules to be sequenced. (In less preferred embodiments, the washing solution may interfere with polymerase activity. Here the washing solution would need to be removed before further primer extension.)

Considering that many copies of molecules to be sequenced can be provided in a given colony, a combination of labelled and non-labelled nucleotides can be used. In this case, even if a small proportion of the nucleotides are labelled (e.g. fluorescence-labelled), the number of labels incorporated in each colony during primer extension can be sufficient to be detected by a detection device. For example the ratio of labelled to non-labelled nucleotides may be chosen so that, on average, labelled nucleotides are used in primer extension less than 50%, less than 20%, less than 10% or even less than 1% of the time (i.e. on average in a given primer extension step a nucleotide is incorporated in labelled form in less than 50%, less than 20%, less than 10 %, or less than 1% of the extended primers.)

Thus in a further embodiment of the present invention there is provided a method for sequencing nucleic acid molecules present in a colony of the present invention, the method comprising the steps of:

a) providing at least one colony comprising a plurality of single stranded nucleic acid molecules that have the same sequences as one another and that are hybridised to primers in a manner to allow primer extension in the presence of nucleotides and a nucleic acid polymerase;

b) providing said at least one colony with a nucleic acid polymerase and a given nucleotide in labelled and unlabelled form under conditions that allow extension of the primers if a complementary base or if a plurality of such bases is present

SUBSTITUTE SHEET (RULE 26)

IDT Exhibit 1020.0157

at the appropriate position in the single stranded nucleic acid molecules present in said at least one colony;

c) detecting whether or not said labelled nucleotide has been used for primer extension by determining whether or not the label present on said nucleotide has been incorporated into extended primers;

Steps b) and c) may be repeated one or more times. Preferably a plurality of different colonies are provided and several different sequences are determined simultaneously.

This further embodiment of the present invention can be used to reduce costs, since relatively few labelled nucleotides are needed. It can also be used to reduce quenching effects.

It is however also possible to use only labelled nucleotides for primer extension or to use a major portion thereof (eg. over 50 %, over 70% or over 90% of the nucleotides used may be labelled). This can be done for example if labels are selected so as to prevent or reduce quenching effects. Alternatively labels may be removed or neutralised at various stages should quenching effects become problematic (e.g. laser bleaching of fluorophores may be performed). However this can increase the number of steps required and it is therefore preferred that labels are not removed (or at least that they are not removed after each nucleotide has been incorporated but are only removed periodically). In other less preferred embodiments, the primer itself and its extension product may be removed and replaced with another primer. If required, several steps of sequential label-free nucleotide additions may be performed before actual sequencing in the presence of labelled nucleotides is resumed. A further alternative is to use a different type of label from that used initially (*e.g.* by switching from fluorescein to rhodamine) should quenching effects become problematic.

In preferred embodiments of the present invention a plurality of labelled bases are incorporated into an extended primer during sequencing. This is advantageous in that it can speed up the sequencing procedure relative to methods in which, once a labelled

IDT Exhibit 1020.0158

base has been incorporated into an extended primer, the label must be removed before a further labelled base can be incorporated. (The plurality of labelled bases may be in the form of one or more contiguous stretches, although this is not essential.)

The present invention therefore also includes within its scope a method for sequencing nucleic acid molecules, comprising the steps of:

a) using a first colony to provide a plurality of single stranded nucleic acid molecules that have the same sequences as one another and that are hybridised to primers in a manner to allow primer extension in the presence of nucleotides and a nucleic acid polymerase;

b) using a second colony to provide a plurality of single stranded nucleic acid molecules that have the same sequences as one another, and that are also hybridised to primers in a manner to allow primer extension in the presence of nucleotides and a nucleic acid polymerase;

c) providing each colony with a nucleic acid polymerase and a given labelled nucleotide under conditions that allow extension of the primers if a complementary base or if a plurality of such bases is present at the appropriate position in the single stranded nucleic acid molecules;

d) detecting whether or not said labelled nucleotide has been used for primer extension at each colony by determining whether or not the label present on said nucleotide has been incorporated into extended primers;

e) repeating steps c) and d) one or more times so that extended primers comprising a plurality of labels are provided.

Preferably the sequences of the nucleic acid molecules present at said first and said locations are different from one another - i.e. a plurality of colonies comprising different nucleic acid molecules are sequenced

IDT Exhibit 1020.0159

In view of the foregoing description it will be appreciated that a large number of different sequencing methods using colonies of the present invention can be used. Various detection systems can be used to detect labels used in sequencing in these methods (although in certain embodiments detection may be possible simply by eye, so that no detection system is needed). A preferred detection system for fluorescent labels is a Charge-Coupled-Device (CCD) camera, which can optionally be coupled to a magnifying device. Any other device allowing detection and, preferably, also quantification of fluorescence on a surface may be used. Devices such as fluorescent imagers or confocal microscopes may be chosen.

In less preferred embodiments, the labels may be radioactive and a radioactivity detection device would then be required. Ideally such devices would be real-time radioactivity imaging systems. Also less preferred are other devices relying on phosphor screens (Molecular Dynamics) or autoradiography films for detection.

Depending on the number of colonies to be monitored, a scanning system may be preferred for data collection. (Although an alternative is to provide a plurality of detectors to enable all colonies to be covered.) Such a system allows a detector to move relative to a plurality of colonies to be analysed. This is useful when all the colonies providing signals are not within the field of view of a detector. The detector may be maintained in a fixed position and colonies to be analysed may be moved into the field of view of the detector (e.g. by means of a movable platform). Alternatively the colonies may be maintained in fixed position and the detection device may be moved to bring them into its field of view.

The detection system is preferably used in combination with an analysis system in order to determine the number (and preferably also the nature) of bases incorporated by primer extension at each colony after each step. This analysis may be performed immediately after each step or later on, using recorded data. The sequence of nucleic

IDT Exhibit 1020.0160

acid molecules present within a given colony can then be deduced from the number and type of nucleotides added after each step.

Preferably the detection system is part of an apparatus comprising other components. The present invention includes an apparatus comprising a plurality of labelled nucleotides, a nucleic acid polymerase and detection means for detecting labelled nucleotides when incorporated into a nucleic acid molecule by primer extension, the detection means being adapted to distinguish between signals provided by labelled nucleotides incorporated at different colonies.

The apparatus may also include temperature control, solvent delivery and washing means. It may be automated.

Methods and apparatuses within the scope of the present invention can be used in the sequencing of:

- unidentified nucleic acid molecules (i.e. *de novo* sequencing);
- and nucleic acid molecules which are to be sequenced to check if one or more differences relative to a known sequence are present (e.g. identification of polymorphisms). This is sometimes referred to as "re-sequencing".

Both *de novo* sequencing and re-sequencing are discussed in greater detail later on (see the following sections (v) and (vi)).

For *de novo* sequencing applications, the order of nucleotides applied to a given location can be chosen as desired. For example one may choose the sequential addition of nucleotides dATP, dTTP, dGTP, dCTP; dATP, dTTP, dGTP, dCTP; and so on. (Generally a single order of four nucleotides would be repeated, although this is not essential.) For re-sequencing applications, the order of nucleotides to be added at each step is preferably chosen according to a known sequence.

Re-sequencing may be of particular interest for the analysis of a large number of similar template molecules in order to detect and identify sequence differences (*e.g.* for the analysis of recombinant plasmids in candidate clones after site directed mutagenesis or more importantly, for polymorphism screening in a population). Differences

SUBSTITUTE SHEET (RULE 26)

IDT Exhibit 1020.0161

from a given sequence can be detected by the lack of incorporation of one or more nucleotides present in the given sequence at particular stages of primer extension. In contrast to most commonly used techniques, the present method allows for detection of any type of mutation such as point mutations, insertions or deletions. Furthermore, not only known existing mutations, but also previously unidentified mutations can be characterised by the provision of sequence information.

In some embodiments of the present invention long nucleic acid molecules may have to be sequenced by several sequencing reactions, each one allowing for determination of part of the complete sequence. These reactions may be carried out at different colonies (where the different colonies are each provided with the same nucleic acid molecules to be sequenced but different primers), or in successive cycles applied at the same colony (where between each cycles the primers and extension products are washed off and replaced by different primers).

(iv) DNA Fingerprinting

This embodiment of the present invention aims to solve the problem of screening a large population for the identification of given features of given genes, such as the detection of single nucleotide polymorphisms.

In one preferred embodiment, it consists in generating tagged genomic DNA (see section G(ii) *supra*). (Thus each sample originating from a given individual sample has been labelled with a unique tag). This tagged DNA can be used for generating primary colonies on an appropriate surface comprising immobilised primers. Several successive probe hybridisation assays to the colonies can then be performed. Between each assay the preceding probe can be removed, e.g. by thermal denaturation and washing.

IDT Exhibit 1020.0162

Advantages of this embodiment of the present invention over other approaches for solving this problem are illustrated in the following example of a potential practical application.

It is intended to detect which part of a gene (of, e.g., 2000 bases in size), if any, is related to a disease phenotype in a population of typically 1,000 to 10,000 individuals. For each individual, a PCR amplification can be performed to specifically amplify the gene of interest and to link a tag and a colony generating primer (refer to section G(iv), preparation of "tagged DNA").

In order to obtain a representative array of sample, one might want to array randomly 500,000 colonies (i.e. 10 times redundancy, so to have only a small probability of missing the detection of a sample). With a colony density of 10,000 colonies per mm^2, a surface of ~7 mm x 7 mm can be used. This is a much smaller surface than any other technology available at present time (e.g. The HySeq approach uses 220 mm x 220 mm for the same number of samples (50,000) without redundancy). The amount of reactants (a great part of the cost) will be proportional to the surface occupied by of the array of samples. Thus the present invention can provide an 800 fold improvement over the presently available technology.

Using an apparatus to monitor the result of the *'in situ'* sequencing or probe hybridisation assays, it should take on the order of 1 to 10 seconds to image a fluorescent signal from colonies assayed using fluorescence present on a surface of ~1 mm^2. Thus, assuming that the bottleneck of the method is the time required to image the result of the assay, it takes of the order of 10 minutes to image the result of an assay on 50,000 samples (500,000 colonies). To provide 200 assays including imaging (on one or several 7 mm x 7 mm surfaces), using the present invention can take less than 36 hours. This represents a 20 times improvement compared to the best method known at present time (HySeq claims 30 days to achieve a comparable task).

SUBSTITUTE SHEET (RULE 26)

IDT Exhibit 1020.0163

Improvements (colony densities 10 times higher and imaging time of 1 second) could allow for much higher throughput and finally the ultimately expected throughput could be about 2000 times faster than the best, not yet fully demonstrated, technology available at present time.

Another advantage of using the present invention lies in the fact that it overcomes the problem arising with individuals who have heterozygous mutations for a given gene. While this problem may be addressed by existing sequencing methods to determine allelic polymorphisms, current high throughput mutation detection methods based on oligonucleotide probe hybridisation may lead to difficulties in the interpretation of results due to an unequal hybridisation of probes in cases of allelic polymorphisms and therefore errors can occur. In this embodiment of the present invention, each colony arises from a single copy of an amplified gene of interest. If an average of 10 colonies are generated for each individual locus, there will be an average of 5 colonies corresponding to one version of a gene and 5 colonies corresponding to the other version of the gene. Thus heterozygotic mutations can be scored by the number of times a single allele is detected per individual genome sample.

(v) DNA resequencing

This embodiment of the present invention provides a solution to the problem of identifying and characterising novel allelic polymorphisms within known genes in a large population of biological samples.

In its preferred embodiment it consists in obtaining tagged DNA (each sample originating from a given individual has been tagged with a unique tag - see section G(iv)). This encoded DNA can then be used for generating primary colonies on an appropriate surface comprising immobilised primers. Several successive assays of probe hybridisation to the colonies can then be performed wherein between each cyclic assay the preceding probe can be removed by thermal denaturation and washing. Preferably, the DNA sequence 3' to a

SUBSTITUTE SHEET (RULE 26)

IDT Exhibit 1020.0164

specific probe may be determined directly by '*in situ* sequencing' (section I(iii), Methods of sequencing).

The advantages of the present invention over other approaches for solving this problem are illustrated in the following example of potential practical application:

It is desired to identify the variability of the sequence of a gene (of, e.g., 2 000 bases in size), if any, in a population of typically 4 000 individuals. It is assumed that a reference sequence of the gene is known. For each individual, a PCR amplification can be performed to specifically amplify the gene of interest and link a tag and a colony generating primer. In order to obtain a representative array of sample, one might want to array randomly 40 000 colonies (i.e. 10 times redundancy, so to have a small probability of missing the detection of a sample). With a colony density of 10 000 colonies per mm^2, a surface of ~2 mm x 2 mm can be used.

Using an apparatus with a CCD camera (having a 2000 x 2000 pixel chip) to monitor the result of the assay, it should take of the order of 10 seconds to image a fluorescent signal from colonies on a surface of 4 mm^2. If it is assumed possible to read at least 20 bases during one round of the assay, this requires 61 imaging steps (3n+1 imaging steps are necessary for reading n numbers of bases). If it is assumed that the bottleneck of the method is the time to image the result of the assay, it takes of the order of 15 minutes to image the result of an assay on 4 000 samples (40 000 colonies). To realise 100 assays (on one or several $2x2mm^2$ surfaces) in order to cover the entire gene of interest, the present invention can allow the whole screening experiment to be performed in approximately one day, with one apparatus. This can be compared to the most powerful systems operational at the present time.

In this embodiment of the present invention with conservative assumptions (colony density, imaging time, size of the CCD chip), a throughput of $3.2x\ 10^6$ bases per hour could be reached, i.e. a 400

SUBSTITUTE SHEET (RULE 26)

IDT Exhibit 1020.0165

fold improvement when compared to the most commonly used system at present time (current DNA sequencers have a typical throughput of the order of 8,000 bases read/hour).

(vi) *de novo* DNA sequencing

This embodiment of the present invention aims to solve the problem of sequencing novel genomes (or parts thereof) with low cost and in short time, where the sequence of the DNA is not known. Genomic DNA can be prepared, either directly from the total DNA of an organism of interest or from a vector into which DNA has been inserted. The prepared genomic DNA (from whatever source) can be used to generate DNA colonies. The DNA colonies can then be digested with a rare-cutting restriction enzyme, whose site is included in the linker, denatured and sequenced.

Figure 16 depicts an example of *de novo* DNA sequencing. In this example, genomic DNA is fragmented into pieces of 100 to 2000 base pairs (see preparation of random DNA fragments, section G(i)). These fragments will be ligated to oligonucleotide linkers (**IIa** and **IIb**, figure 16a) which include sequences specific for the grafted primers on the surface ('P1' and 'P2'), a sequence which is recognised by a rare-cutting restriction nuclease ('RE') and a sequence corresponding to a sequencing primer ('SP'), resulting in templates (**III**, figure 16b). Using this prepared DNA as template for DNA colony formation, one obtains primary colonies (**IV**, figure 16c). These colonies are then digested with the corresponding restriction endonuclease and denatured to remove the non-attached DNA strand (**V**, figure 16d). The sequencing primer (**SP**) is then annealed to the attached single-stranded template (figure 16e). Incorporation and detection of labelled nucleotides can then be carried out as previously described (see section I(iii), Methods of Sequencing).

In this embodiment, the throughput obtainable can be at least 400 times higher than presently available methods.

SUBSTITUTE SHEET (RULE 26)

IDT Exhibit 1020.0166

(vii) mRNA gene expression monitoring

This embodiment of our invention means to solve the problem of monitoring the expression of a large number of genes simultaneously.

Its preferred embodiment is depicted in figure 17.

Firstly, primary colonies are prepared, as depicted in figure 3. In its preferred form, the DNA used for this preparation is 'prepared genomic DNA' or 'tagged genomic DNA', as described in section G(i) and G(iii), respectively, and where the DNA is either from the whole genome of one (or several) organism(s) or from a subset thereof (e.g., from a library of previously isolated genes). In figure 17, the uppercase letters, "A", "B" and "D" represent colonies which have arisen from genes which exhibit high, medium and low expression levels, respectively, and "E" represents colonies arising from non-expressed genes (in real cases, all these situations may not necessarily be present simultaneously).

Secondly, the colonies are treated to turn then into supports (i.e. secondary primers) for secondary colony growing (step i in figure 17a), as described in section E. At this stage (figure 17b), the treated colonies are represented by underlined characters (A, B, D, or E).

Thirdly, (step ii in figure 17b) this support for secondary colony growing is used to regenerate colonies from mRNA (or cDNA) templates extracted from a biological sample, as described in section C. If the template is mRNA, the priming step of colony regeneration will be performed with a reverse transcriptase. After a given number of colony amplification cycles, preferably 1 to 50, the situation will be as depicted in (figure 17c): the colonies corresponding to highly expressed genes (represented by the letter "A") are totally

SUBSTITUTE SHEET (RULE 26)

IDT Exhibit 1020.0167

regenerated, as their regeneration has been initiated by many copies of the mRNA; the colonies corresponding to genes of medium expression levels (represented by the letters "b" and "B"), have been only partially regenerated; only a few of the colonies corresponding to rare genes (represented by the letter "d"), have been partially regenerated; the colonies corresponding to non-expressed sequences (represented by the letter "E"), have not been regenerated at all.

Lastly, (step iii in figure 17c), additional cycles of colony growing are performed (preferably 2 to 50), and the colonies which have not been totally regenerated during the previous steps finally become totally regenerated, "b" becomes "**B**", "d" becomes "**D**" (figure 17d) : the colonies corresponding to genes with high and medium expression levels are all regenerated "**A**" and "B" or "**B**"; the colonies corresponding to genes with low levels of expression are not all regenerated "D" and "D"; the colonies corresponding to non-expressed sequences are not regenerated at all "E".

The relative levels of expression of the genes can be obtained by the following preferred methods:

- Firstly levels of expression can be monitored by following the rate of regeneration of the colonies (i.e., by measuring the amount of DNA inside a colony after different number of colony growing cycles during step (iii)) as the rate at which a colony is regenerated will be linked the number of mRNA (or cDNA) molecules which initiated the regeneration of that colony (at first approximation, the number of DNA molecules after n cycles, noted M(n), in a colony undergoing regeneration should be given by $M(n) = M_0 r^{(n-1)}$, where M_0 is the number of molecules which initiated the regeneration of the colony, r is the growing rate and n is the number of cycles);

- Secondly, levels of expression can be monitored by counting, for each gene, the number of colonies which have been regenerated and comparing this number to the total number of colonies corresponding to that gene. These measurements will generally give access to the

IDT Exhibit 1020.0168

relative expression levels of the genes represented by the colonies. The identification of the colonies is preferably performed by fingerprinting, in a manner essentially similar to embodiment, section I(iv). Note that encoding the DNA samples is not required, but can be considered as an alternative to the direct identification of the DNA in the colonies. This can be of practical interest because with coding, the same codes (thus the same oligonucleotides involved in assaying the code) can be used for any set of genes, whereas without code, a different set of specific oligonucleotides has to be used for each set of genes.

This embodiment of our invention has many advantages if compared to current state of the art including: a very high throughput; no requirement for prior amplification of the mRNA (even though prior amplification is compatible with or invention); small amounts of samples and reactants are required due to the high density of samples with our invention; the presence of highly expressed genes has no incidence on the ability to monitor genes with low levels of expression; the ability to simultaneously monitor low and high levels of expression within the set of genes of interest.

When the initial DNA in the generation of the primary DNA colony is made from the DNA of a whole genome, this embodiment also provides the following features: there is no interference between genes expressed at high level and at low level even though one has not performed specific amplification of the genes of interest. This is a unique feature of the use of this invention: specific amplification is not possible because the initial assumption of this embodiment is to monitor the expression genes which may have not yet been isolated, thus which are unknown, and thus for which no specific (unique) sequences are known and which specific sequences would have been necessary for specific gene amplification. The ability of our invention to perform this type of mRNA expression monitoring is due to the fact that when the primary colonies are prepared, statistically, each piece of the initial genome will be represented by the same number of colonies. Thus, frequent and rare DNA will initiate the same number of colonies (e.g., one colony per added

IDT Exhibit 1020.0169

genome molecule). Quantitative information might be obtained both from frequent and rare mRNAs by monitoring the growing rate of the colonies.

(viii) Isolation and characterisation of novel expressed genes

This embodiment of our invention means to solve the problem of isolating the genes which are specifically induced under given conditions, e.g., in specific tissues, different strains of a given species or under specific activation. A practical example is the identification of genes which are up or down regulated after drug administration.

The preferred embodiment for isolating genes from a specific or activated biological sample (hereafter called target sample) which are up-regulated compared with a reference biological sample (hereafter called reference sample) is depicted in figure 18.

Firstly, primary colonies are prepared (figure 18a). In its preferred form, the DNA used for this preparation is prepared genomic DNA or tagged genomic DNA, as described in sections G(i) and G(ii), respectively, where the DNA is either from the whole genome of one (or several) organism(s) or from a subset thereof (e.g., from a library of previously isolated genes), and where both the primers used for colony generation (hereafter called P1 and P2) contain a endonuclease restriction site. In figure 18a, "A" represents colonies which have arisen from genes expressed in both the reference sample and the target sample, "B" represents colonies which have arisen from genes expressed only in the reference sample, "C" represents colonies which have arisen from genes expressed only in the target sample, and "D" represents colonies arising from non-expressed genes (in real cases, all these situations may not necessarily be present simultaneously).

IDT Exhibit 1020.0170

Secondly, primary colonies are then treated to generate secondary primers as the support for secondary colony growing (step i in figure 18a). At this stage (b), the colonies are represented as underlined characters (A,B,C,D).

Thirdly, (step ii in figure 21b) the secondary primers are used to regenerate colonies using mRNA or cDNA (represented by "mA + mB") extracted from the biological reference sample as a template, as described in G(v). If the template is mRNA, the first elongation step of colony regeneration will be performed with a reverse transcriptase. After enough colony growing cycles, preferably 5 to 100, only the colonies corresponding to genes expressed in the reference sample ("A" and "B") will be regenerated, as depicted in (figure 18c).

In step (iii), the colonies are digested with a restriction enzyme (represented by RE) which recognises a site in the flanking primer sequences, P1 and P2, which are grafted on the support and which were the basis of primary colony generation. Importantly, only the colonies which have been regenerated during step (ii) will be digested. This is because the support for secondary colony growth is made of single stranded DNA molecules, which can not be digested by the restriction enzyme. Only the regenerated colonies are present in a double stranded form, and are digested. After digestion, the situation is the one depicted in figure 18d. The colonies corresponding to the genes expressed in the reference sample have totally disappeared, i.e., they are not even present as a support for secondary colony growth, and the colonies corresponding to genes expressed only in the target sample "C" and the colonies corresponding to non-expressed genes "D" are still present as a support for secondary colony generation.

In step (iv), mRNA (or cDNA) (represented by "mA + mC") extracted from the target sample is used to generate secondary colonies. Because colonies corresponding to mA and mB no longer exist, only the colonies corresponding to mC can be regenerated (ie. Only the mRNA

SUBSTITUTE SHEET (RULE 26)

IDT Exhibit 1020.0171

specifically expressed in the target sample). After sufficient number of colony growing cycles (preferably 5 to 100), the situation is such that only the colonies corresponding to genes expressed specifically in the target sample are regenerated ("C", figure 18e).

In step (v), the regenerated colonies "C" are used to generate copies of the DNA that they contain by performing several (preferably 1 to 20) colony growing cycles in the presence of the primers P1 and P2, as described in section D of the present invention. A PCR amplification is then performed using P1 and P2 in solution (described in section D) and the amplified DNA characterised by classical methods.

The preferred embodiment for isolating genes from a specific or activated biological sample which are less expressed than in a reference biological sample is depicted in figure 19. The different steps involved in this procedure are very similar to those involved in the isolation of gene which are more regulated than in the reference sample, and the notation are the same as in figure 18. The only difference is to inverse the order used to regenerate the colonies: in step (ii), the mRNA used is the one extracted from the target biological sample ("mA + mC") instead of the mRNA extracted form the reference biological sample ("mA + mB"), and in step (iv), the mRNA used is the one extracted from the reference biological sample ("mA + mB") instead of the one extracted from the target sample ("mA + mC"). As a result, only the DNA from colonies corresponding to genes which are expressed in the reference sample but not in the target sample is recovered and amplified ("B", figure 19f).

SUBSTITUTE SHEET (RULE 26)

IDT Exhibit 1020.0172

Claims

1. A method of nucleic acid amplification, comprising the steps of:

A. providing a plurality of primers that are immobilised but that have one end exposed to allow primer extension;
B. allowing a single stranded target nucleic acid molecule to anneal to one of said plurality of primers over part of the length of said single stranded nucleic acid molecule and then extending that primer using the annealed single stranded nucleic acid molecule as a template, so as to provide an extended immobilised nucleic acid strand;
C. separating the target nucleic molecule from the extended immobilised nucleic acid strand;
D. allowing the extended immobilised nucleic acid strand to anneal to one of said plurality of primers referred to in step A) and then extending that primer using the extended immobilised nucleic acid strand as a template, so as to provide another extended immobilised nucleic acid strand; and optionally,
E. separating the annealed extended immobilised nucleic acid strands from one another.

2. A method according to claim 1, further comprising:

F. using at least one extended immobilised nucleic acid strand to repeat steps D) and E), so as to provide additional extended immobilised nucleic acid strands and, optionally,
G. repeating step F) one or more times.

3. A method according to claim 1 or claim 2, wherein said single-stranded target nucleic acid is produced by providing a given nucleic acid sequence to be amplified (which sequence may be known or unknown) and adding thereto a first nucleic acid sequence and a second nucleic acid sequence; wherein said first nucleic acid sequence hybridises to one of said plurality of primers and said second nucleic acid sequence is complementary to a sequence which hybridises to one of said plurality of primers.

SUBSTITUTE SHEET (RULE 26)

IDT Exhibit 1020.0173

4. A method according to any preceding claim; wherein said single-stranded target nucleic acid is produced by providing a given nucleic acid sequence to be amplified (which sequence may be known or unknown) and adding thereto a first nucleic acid sequence and a second nucleic acid sequence; wherein said first nucleic acid sequence hybridises to one of said plurality of primers and said second nucleic acid sequence is the same as the sequence of one of said plurality of primers.

5. A method according to claim 3 or claim 4 wherein said first and second nucleic acid sequences are provided at first and second ends of said single-stranded target nucleic acid.

6. A method according to any of claims 3 to 5, wherein a tag is also added to the given nucleic acid sequence, said tag enabling amplification products of the given nucleic acid sequence to be identified.

7. A method according to any preceding claim wherein the plurality of primers is a plurality of primers that have the same sequence.

8. A method according to any of claims 1 to 6, wherein the plurality of primers comprises at least two different types of primer, one type having a different sequence from another type.

9. A method according to claim 8, wherein the plurality of primers consists of 2^n different types of primer; wherein n is an integer.

10. A method according to claim 8 or claim 9, wherein the different types of primer are present in substantially the same concentrations as one another

11. A method according to any preceding claim, wherein the primers are substantially homogeneously dispersed over a given area

12. A method according to any preceding claim, wherein the primers are located in a predetermined arrangement (e.g. in a grid pattern)

IDT Exhibit 1020.0174

13. A method according to any preceding claim, wherein a supply of nucleotides and a nucleic acid polymerase are used to extend primers

14. A method according to any preceding claim, wherein heating is used to separate annealed nucleic acid strands

15. A method according to claim 14 when dependent upon claim 13, wherein the nucleic acid polymerase is not rendered inactive by the heating conditions used to separate annealed nucleic acid strands.

16. A method according to claim 15, wherein said nucleic acid polymerase is *taq* polymerase, is another polymerase that is derivable from a thermophilic organism; or is a thermostable derivative thereof.

17. A method according to any preceding claim, wherein said primer extension results in the incorporation of one or more detectable labels (e.g. fluorescent labels or radiolabels) into extended immobilised nucleic acid strands.

18. A method according to any preceding claim, further including the step of treating one or more extended immobilised nucleic acid strands so as to release a nucleic acid molecule or a part thereof.

19. A method according to claim 18, wherein said treating consists of cleavage with a restriction endonuclease or with a ribozyme.

20. A method according to any preceding claim, wherein one or more of said primers has a restriction endonuclease recognition site or a ribozyme recognition site or has part of such a site, which part becomes complete when primer extension occurs.

21. A method according to any preceding claim that is automated to allow repeated cycles of nucleic acid amplification.

22. A method according to any preceding claim, when used to amplify a plurality of different nucleic acid sequences

IDT Exhibit 1020.0175

23. A method according to claim 22, when used to amplify a plurality of different nucleic acid sequences simultaneously.

24. A method according to claim 22 or 23, wherein said different nucleic acid sequences are each provided with a first and second nucleic acid sequence as described in any of claims 3 to 5, said first and second nucleic acid sequences being the same for the each of the different nucleic acid sequences.

25. A method according to any of claims 22 to 24, wherein said different nucleic acid sequences are each provided with a different tag so that the different sequences can be distinguished from one another.

26. A plurality of immobilised nucleic acids producable by a method according to any preceding claim.

27. A plurality of immobilised nucleic acids in the form of one or more distinct areas on a surface, each area comprising a plurality of identical nucleic acid strands and a plurality of identical complementary strands thereto; wherein each nucleic acid strand within such an area is located so that another nucleic acid strand is located on the surface within a distance of the length of that strand.

28. A plurality of immobilised nucleic acids according to claim 26 or claim 27, wherein there is at least one distinct area present per mm^2 of surface on which the nucleic acids are immobilised.

29. A plurality of immobilised nucleic acids according to claim 27, wherein the number of distinct areas/mm^2 of surface on which the nucleic acids are immobilised is greater than 1, greater than 10^2, greater than 10^3 or greater than 10^4.

30. The use of a method according to any of claims 1 to 25 or of a plurality of immobilised nucleic acid molecules according to any of claims 26 to 29 in providing nucleic acid molecules for sequencing.

IDT Exhibit 1020.0176

31. The use according to claim 30, wherein sequencing is performed by extending primers hybridised to single stranded nucleic acid molecules and detecting the nucleotides used in primer extension.

32. The use according to claim 31, wherein said primers hybridise to the same sequences as the primers used in the method of claim 1 to provide the immobilised nucleic acid molecules for sequencing.

33. The use according to claim 31, wherein said primers have the same sequences as the primers used in the method of claim 1 to provide the immobilised nucleic acid molecules for sequencing.

34. The use according to claim 30, wherein nicks are provided in double-stranded nucleic acid molecules, nick translation is performed and the nucleotides used in nick translation are detected.

35. The use according to claim 30, wherein an RNA polymerase is used to synthesise an RNA strand from a DNA molecule and the nucleotides used to produce said RNA strand are detected.

36. The use according to any of claims 31 to 35, wherein all or at least some of said nucleotides are labelled nucleotides.

37. The use according to claim 36, wherein said labelled nucleotides all have the same label.

38. The use according to claim 36 or claim 37, wherein said labelled nucleotides are fluorescence labelled

39. The use according to any of claims 36 to 38, wherein a mixture of labelled and non-labelled nucleotides are used.

40. The use according to claim 39, wherein the labelled nucleotides comprise less than 50 % of the nucleotides used.

41. The use according to claim 40, wherein the labelled nucleotides comprise less than 10 % of the nucleotides used.

IDT Exhibit 1020.0177

42. The use according to any of claims 30 to 41, wherein the sequencing is parallel sequencing of nucleic acid molecules present in at least 2 different distinct areas.

43. The use according to any of claims 30 to 41, wherein the sequencing is parallel sequencing of nucleic acid molecules present in over 10 different distinct areas.

44. The use according to any of claims 30 to 41, wherein the sequencing is parallel sequencing of nucleic acid molecules present in over 100 different distinct areas.

45. The use according to any of claims 30 to 41, wherein the sequencing is parallel sequencing of nucleic acid molecules present in over 1000 different distinct areas.

46. The use according to any of claims 30 to 41, wherein the sequencing is parallel sequencing of nucleic acid molecules present in over 1000000 different distinct areas.

47. The use according to any of claims 30 to 46, wherein a plurality of different sequences are determined in parallel.

48. The use of a method according to any of claims 1 to 25 or of a plurality of immobilised nucleic acid molecules according to any of claims 26 to 29 in providing amplified nucleic acid molecules for diagnosis.

49. The use of a method according to any of claims 1 to 25 or of a plurality of immobilised nucleic acid molecules according to any of claims 26 to 29 in providing amplified nucleic acid molecules for screening.

50. The use of a method according to any of claims 1 to 25 or of a plurality of immobilised nucleic acid molecules according to any of

IDT Exhibit 1020.0178

claims 26 to 29 in providing amplified nucleic acid molecules to be used as a support for other components.

51. The use of a method according to any of claims 1 to 25 or of a plurality of immobilised nucleic acid molecules according to any of claims 26 to 29 in generating additional nucleic acid molecules in free (rather than immobilised) form.

52. The use of a method according to claim 51 in *in situ* RNA synthesis

53. The use of a method according to any of claims 1 to 25 or of a plurality of immobilised nucleic acid molecules according to any of claims 26 to 29 in monitoring gene expression

54. The use of a method according to any of claims 1 to 25 or of a plurality of immobilised nucleic acid molecules according to any of claims 26 to 29 in identifying nucleic acid molecules with gene products that are rarely expressed.

55. The use of a method according to any of claims 1 to 25 or of a plurality of immobilised nucleic acid molecules according to any of claims 26 to 29 in identifying heterozygous individuals

56. The use of a method according to any of claims 1 to 25 or of a plurality of immobilised nucleic acid molecules according to any of claims 26 to 29 in nucleic acid fingerprinting

57. An apparatus for performing a method as described in any of claims 1 to 25; comprising a plurality of immobilised primers, a nucleic acid polymerase, a plurality of nucleotides and means for separating annealed nucleic acid strands.

58. An apparatus according to claim 57, wherein the means for separating annealed nucleic acid strands comprises a controlled heating means.

59. An apparatus for analysing a plurality of nucleic acid molecules according to any of claims 26 to 29, wherein said apparatus

IDT Exhibit 1020.0179

comprises a source of reactants and detector means for detecting one or more signals produced after said reactants have been applied to said nucleic acid molecules.

60. An apparatus according to claim 59 wherein said detector means has sufficient resolution to distinguish between the distinct areas referred to in claim 26.

61. An apparatus according to claim 59 or claim 60 comprising a charge coupled device (CCD).

62. An apparatus according claim 61 wherein said charge coupled device (CCD) is operatively connected with a magnifying device (e.g. a microscope).

63. A kit for use in screening, diagnosis or in nucleic acid sequencing; comprising a plurality of immobilised nucleic acid according to any of claims 26 to 29.

64. The invention substantially as hereinbefore described.

IDT Exhibit 1020.0180

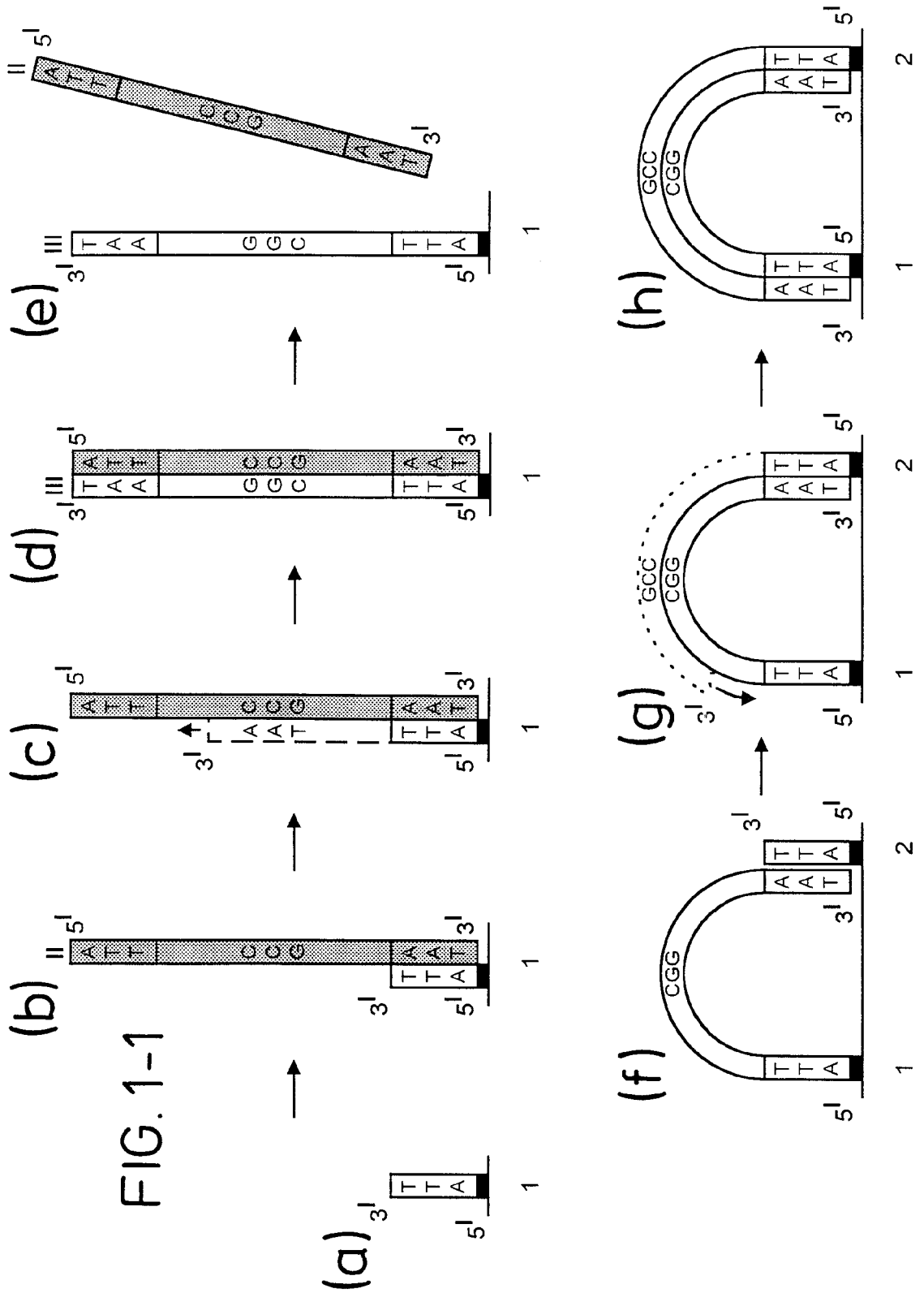
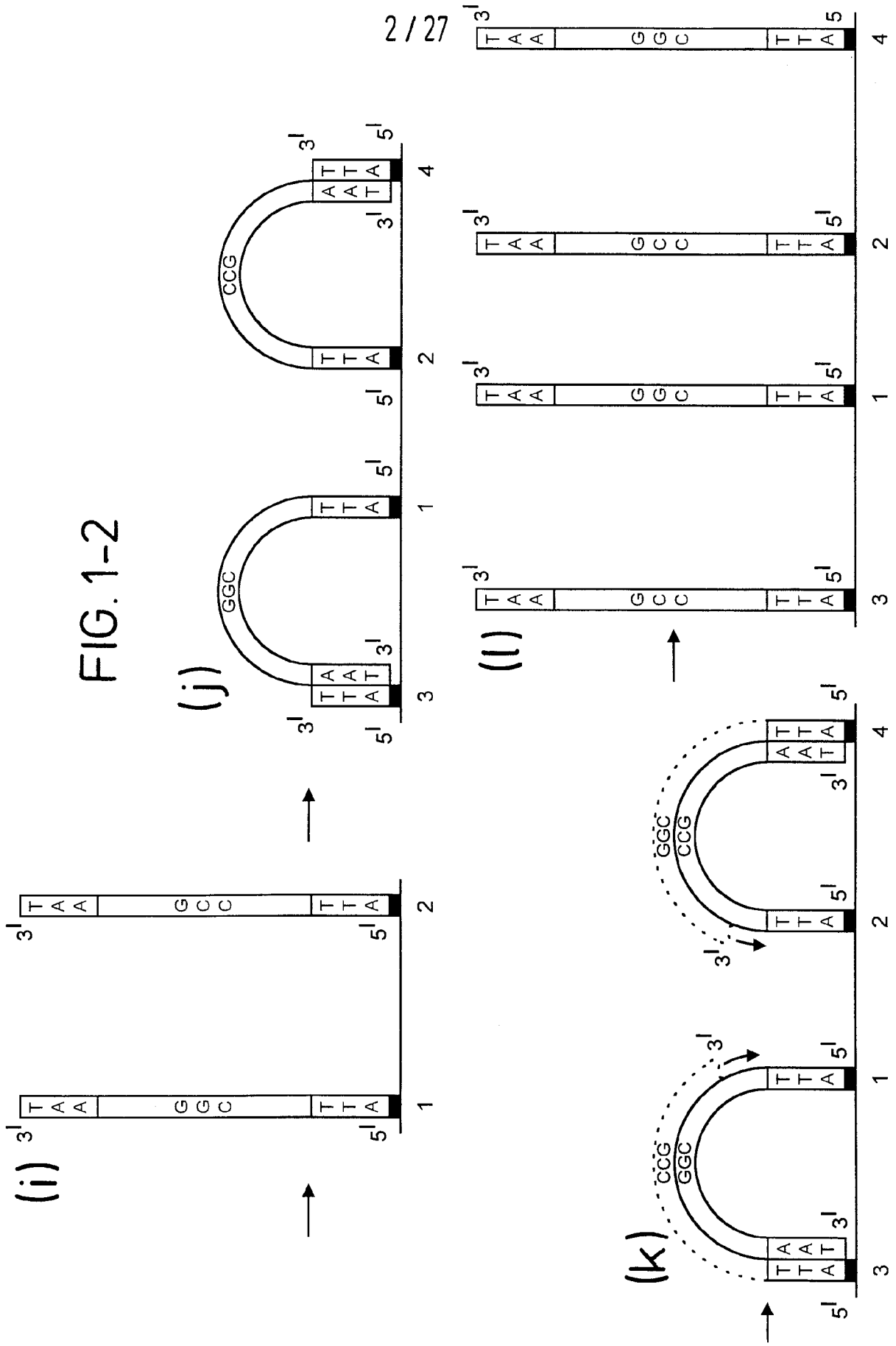


FIG. 1-1

IDT Exhibit 1020.0181

FIG. 1-2



IDT Exhibit 1020.0182

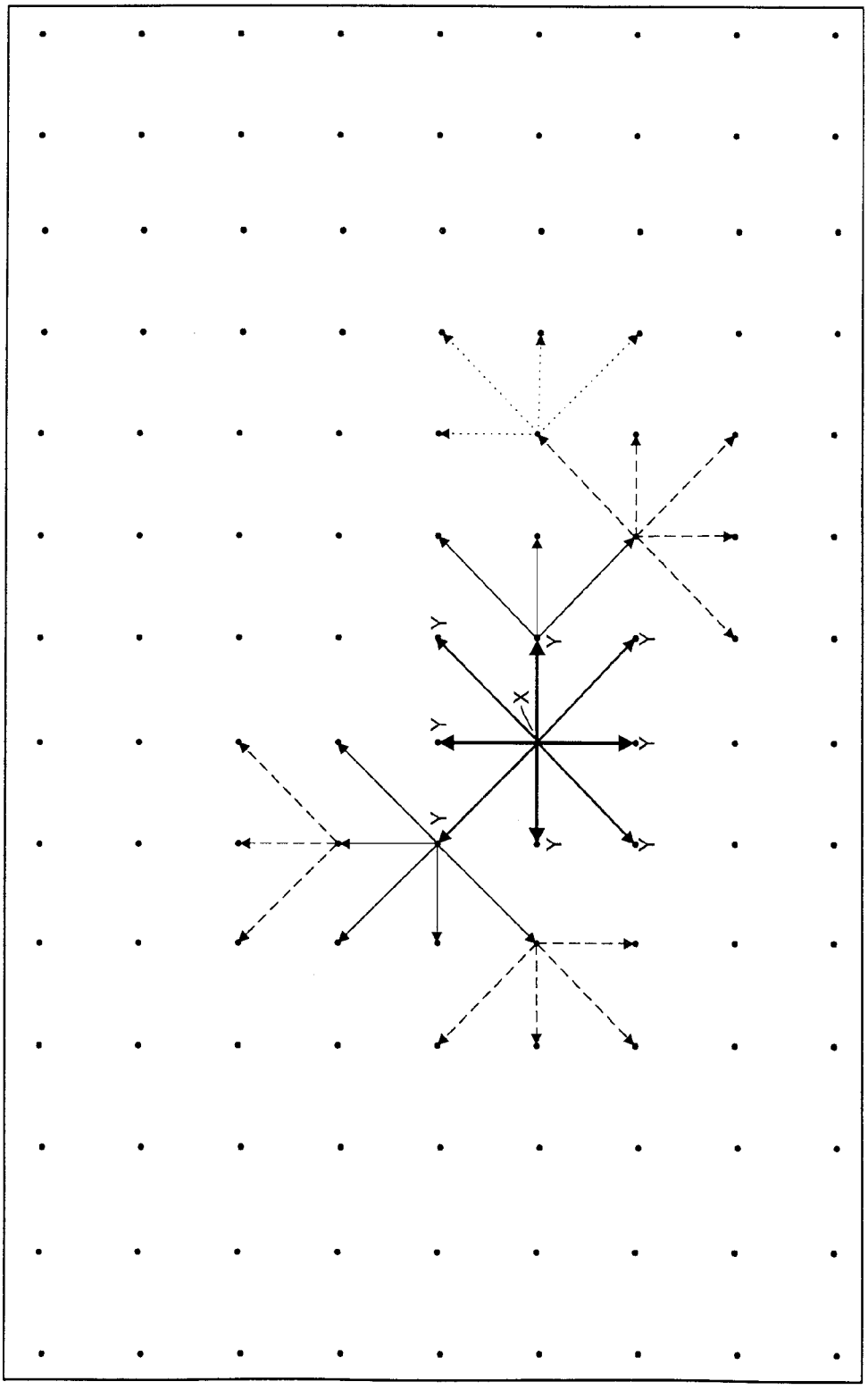
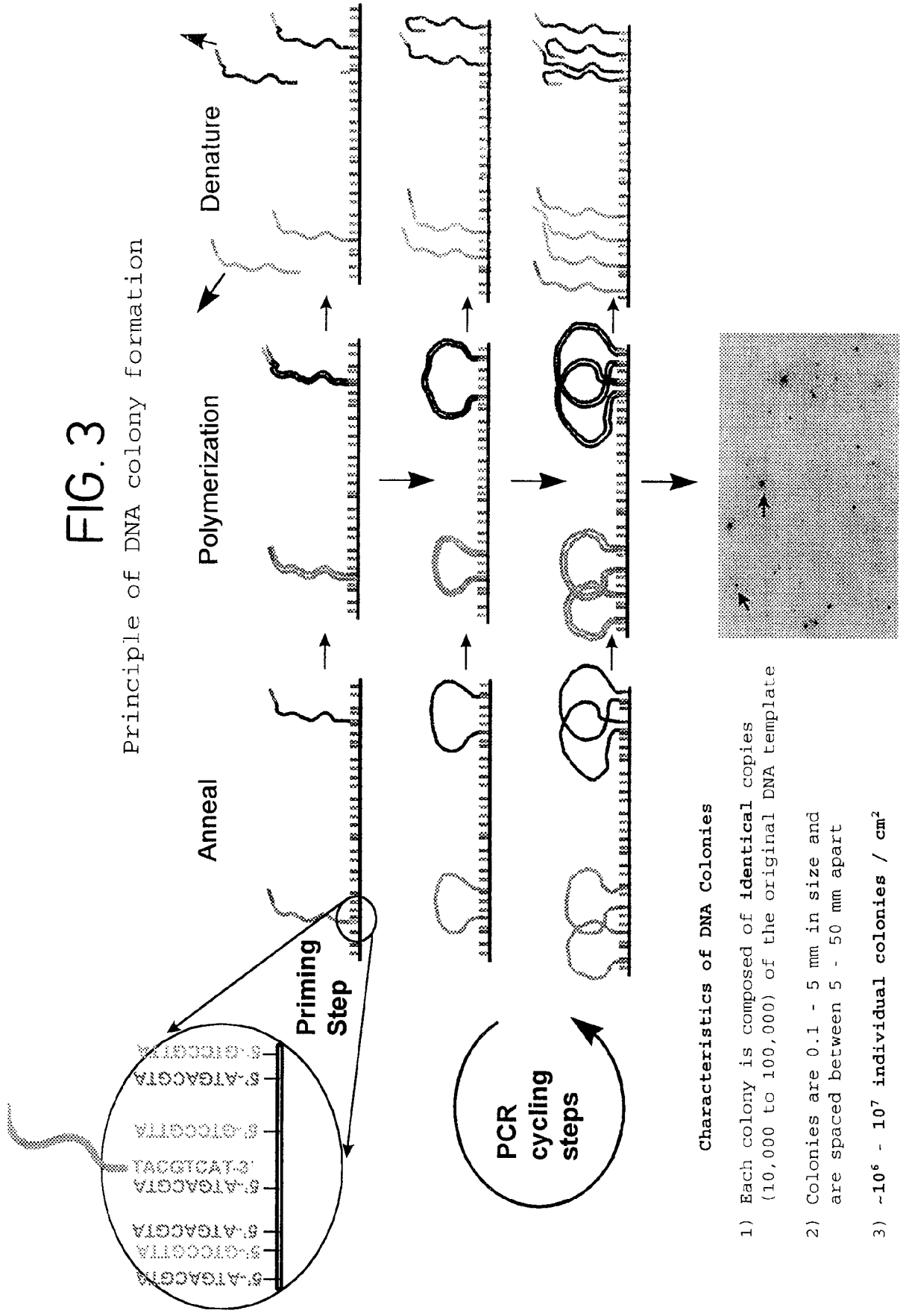


FIG. 2 SHOWS SCHEMATIC VIEW OF COLONY GROWTH

IDT Exhibit 1020.0183

FIG. 3

Principle of DNA colony formation



IDT Exhibit 1020.0184

FIG. 4a

Colony generation on tubes functionalised with oligonucleotide p57 (control). Results of visualisation of fluorescent microspheres. Hybridisation with probe detects 0.10 fmol/tube.

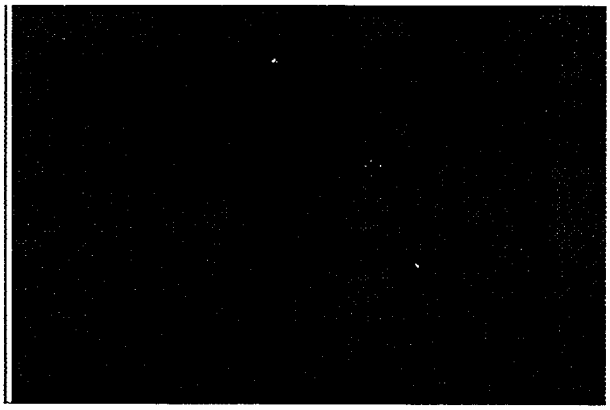
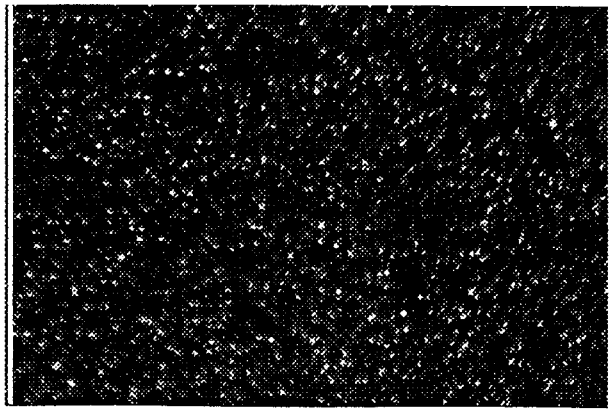


FIG. 4b

Colony generation on tubes functionalised with oligonucleotide p58. Results of visualisation of fluorescent microspheres. Hybridisation with probe detects 6.2 fmol/tube



IDT Exhibit 1020.0185

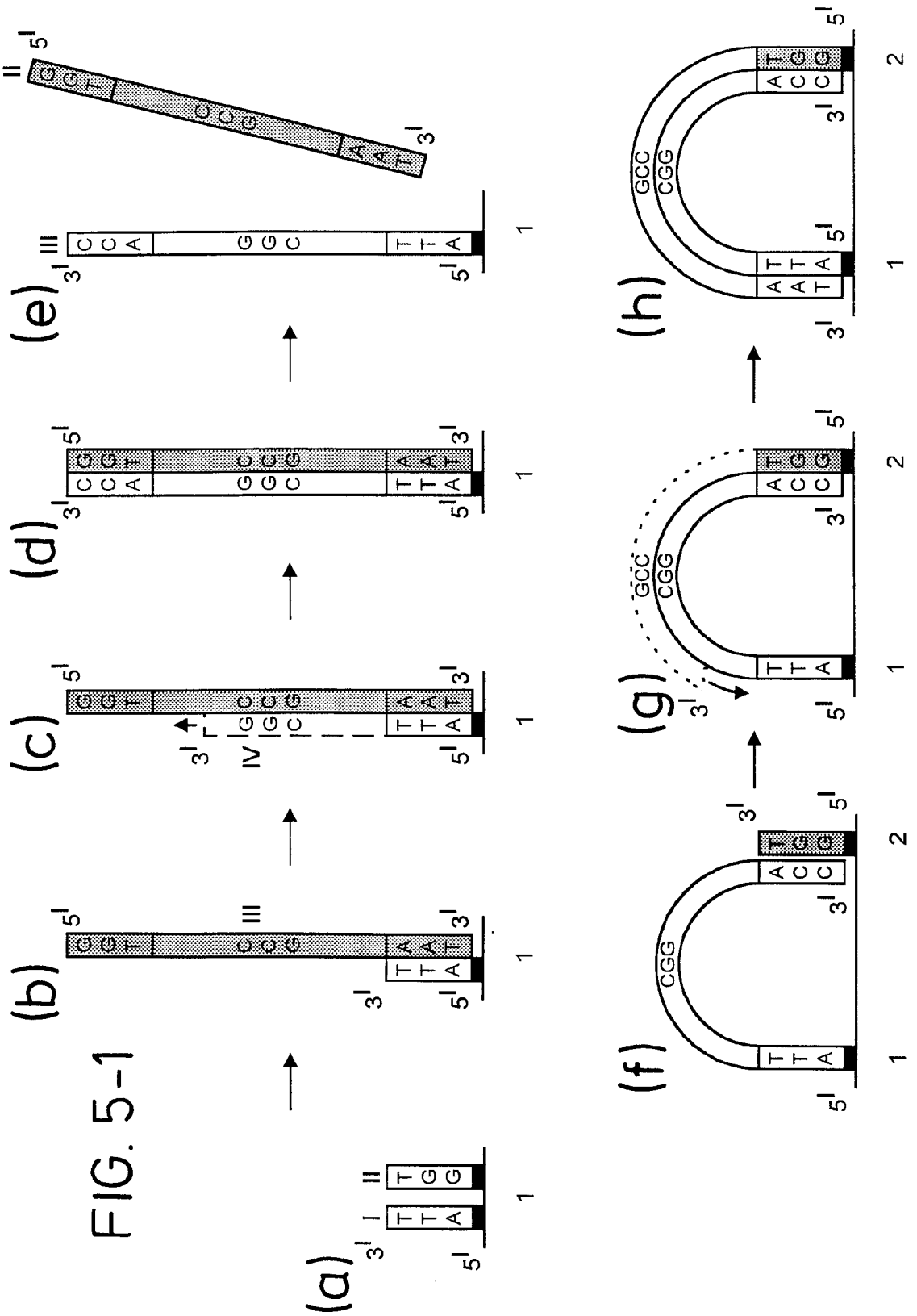
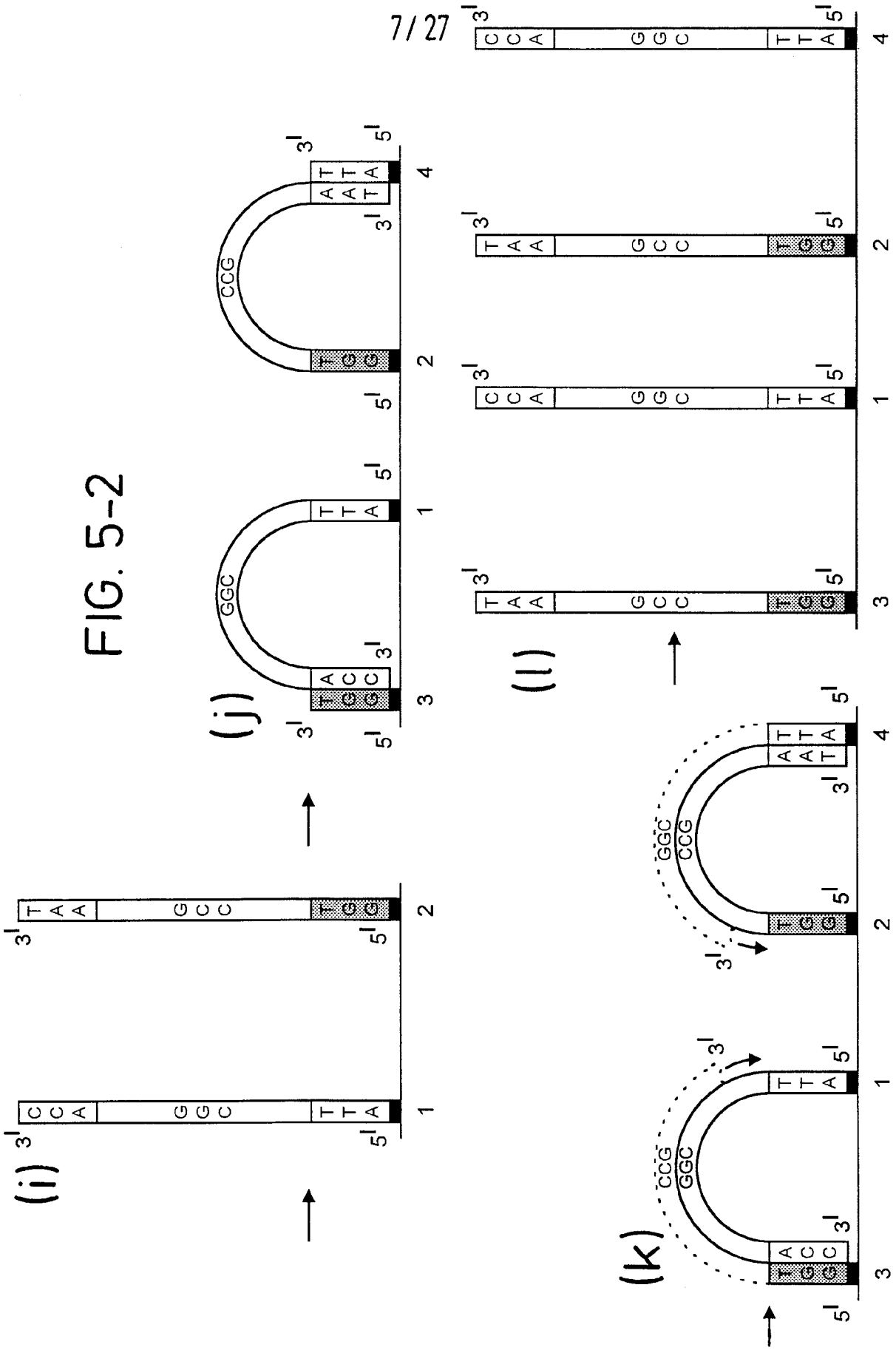


FIG. 5-1

IDT Exhibit 1020.0186

FIG. 5-2



IDT Exhibit 1020.0187

FIG. 6(a)

DNA amplification - detection of colonies with DNA intercalating dyes

i) control



ii) with template

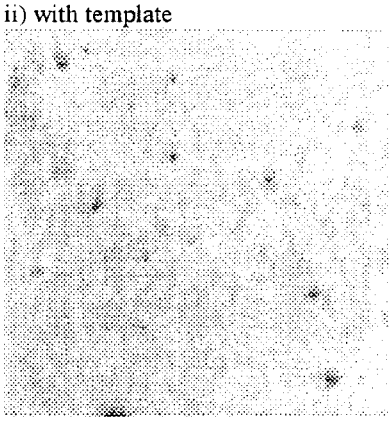
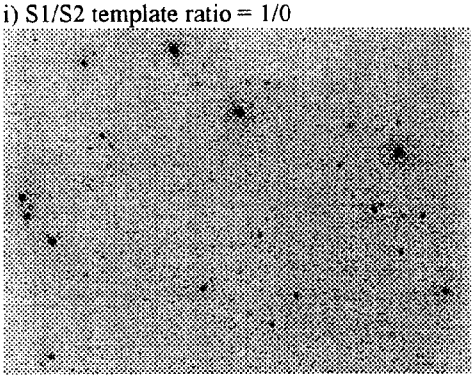


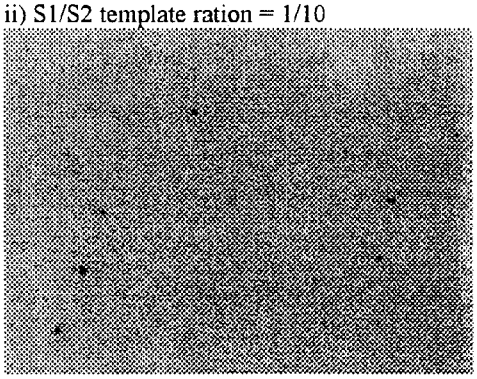
FIG. 6(b)

DNA amplification; detection of colonies by specific probe hybridisation

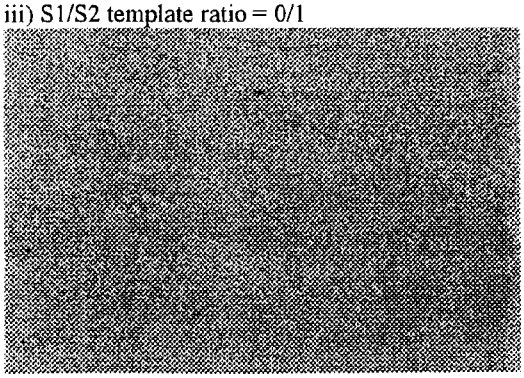
i) S1/S2 template ratio = 1/0



ii) S1/S2 template ration = 1/10



iii) S1/S2 template ratio = 0/1



IDT Exhibit 1020.0188

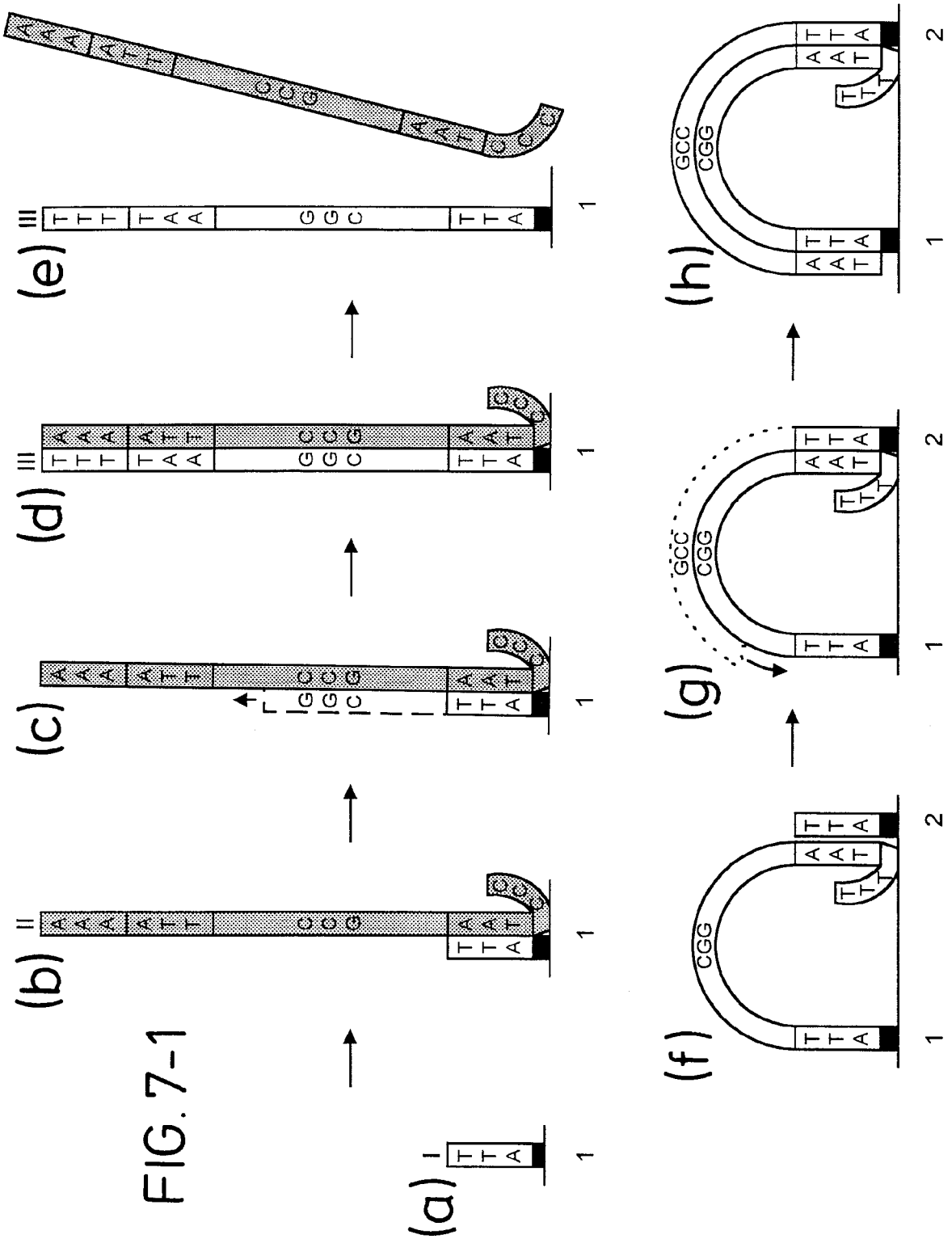
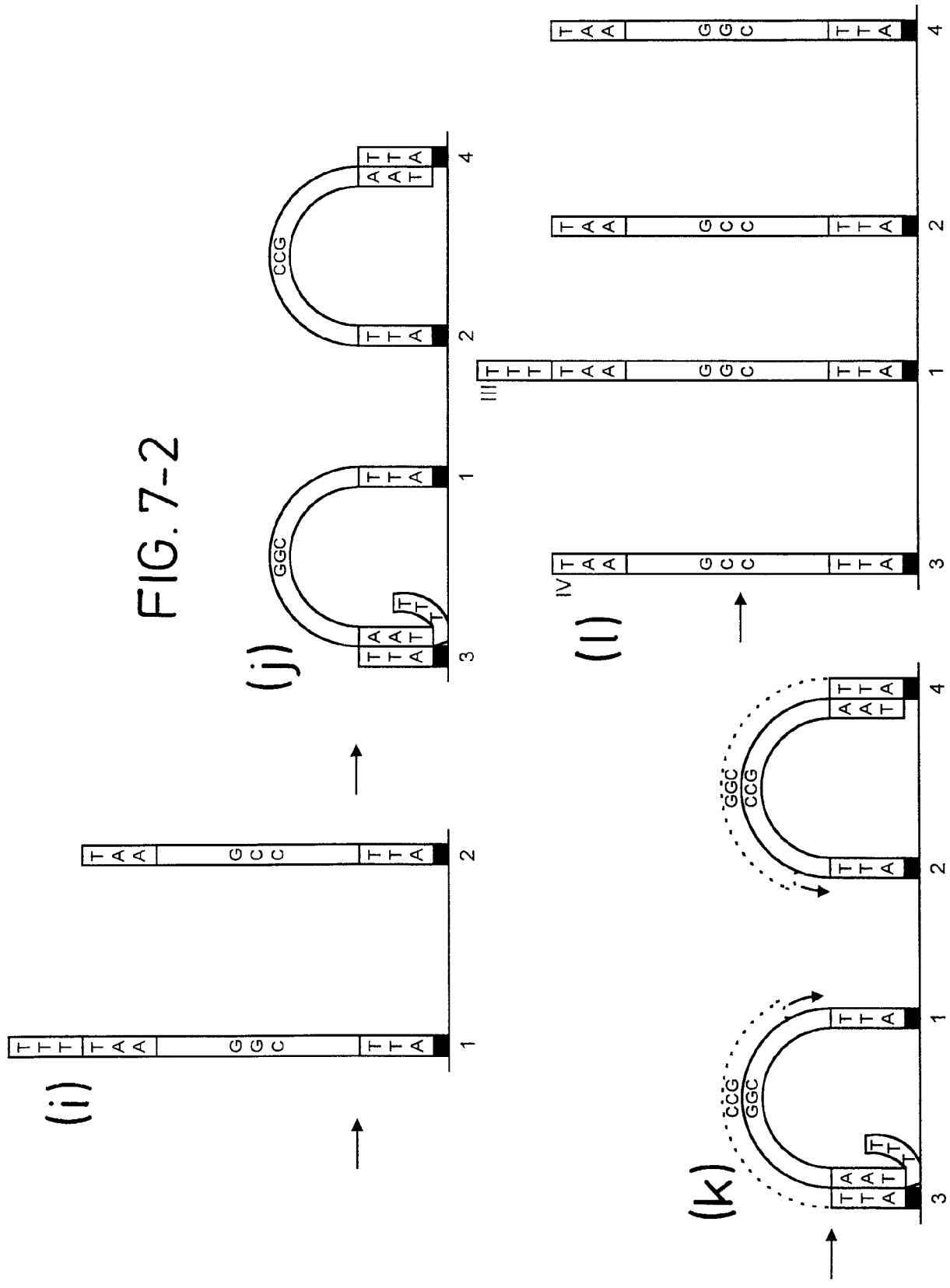


FIG. 7-1

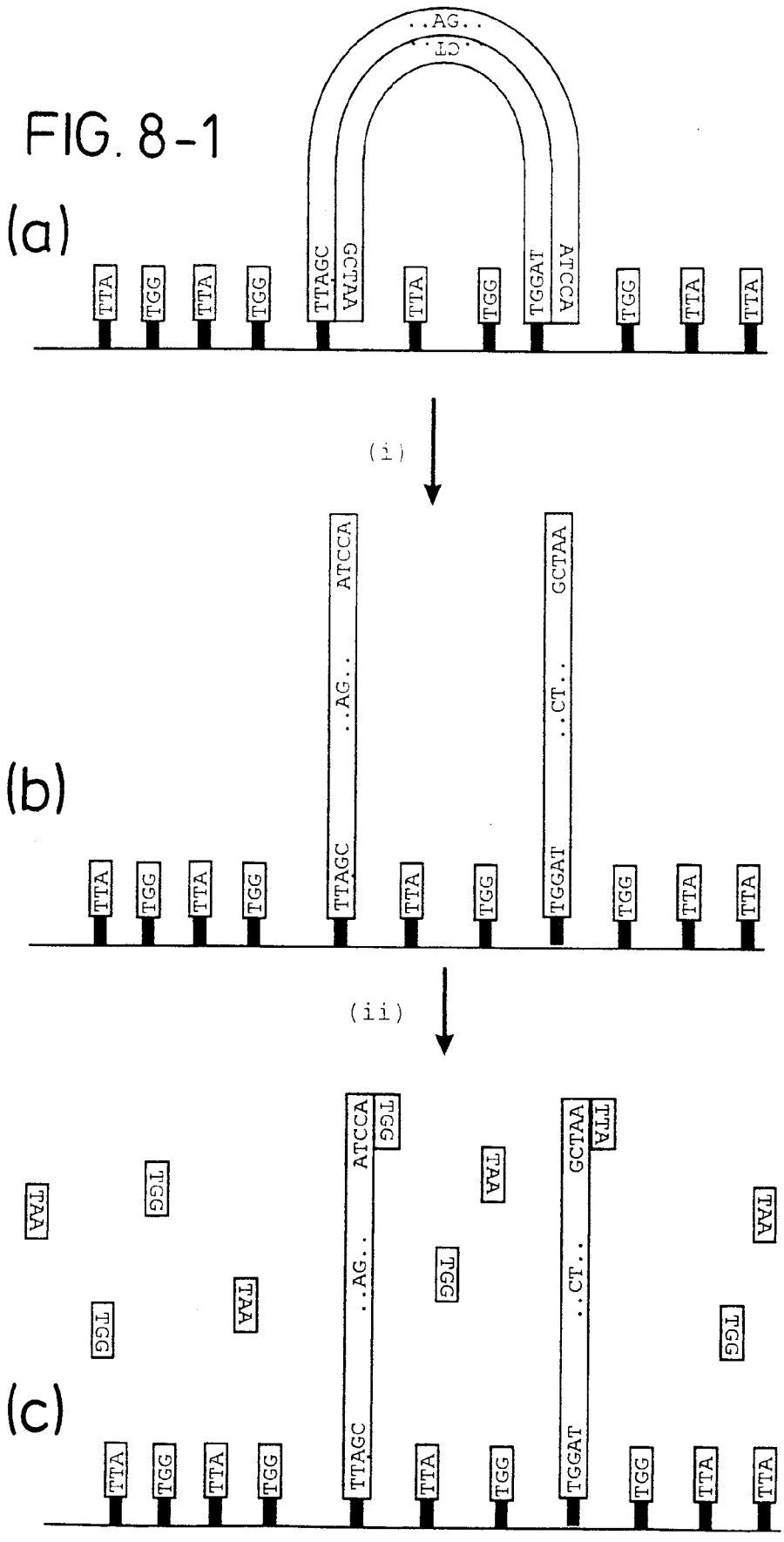
IDT Exhibit 1020.0189

FIG. 7-2



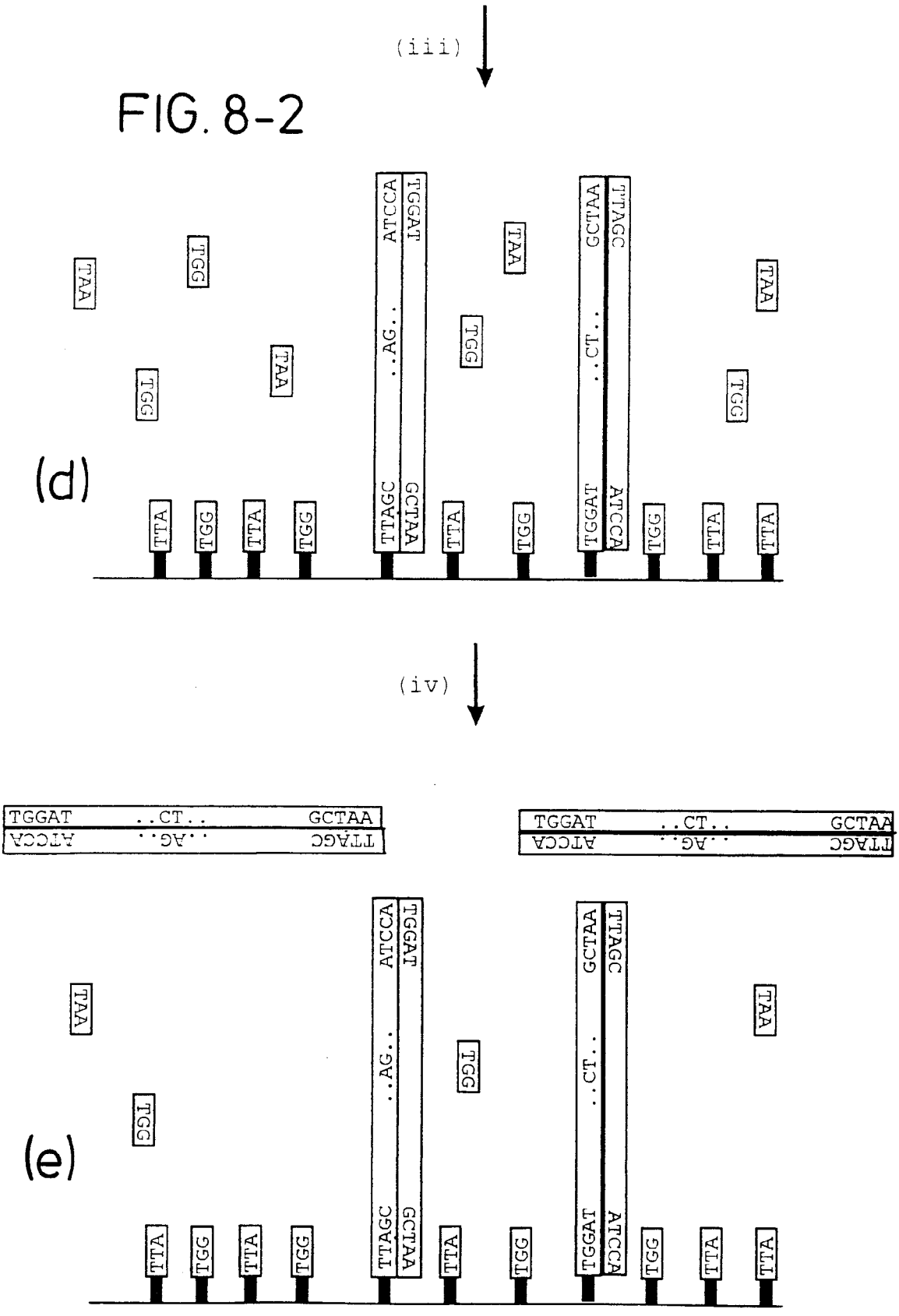
IDT Exhibit 1020.0190

FIG. 8-1



IDT Exhibit 1020.0191

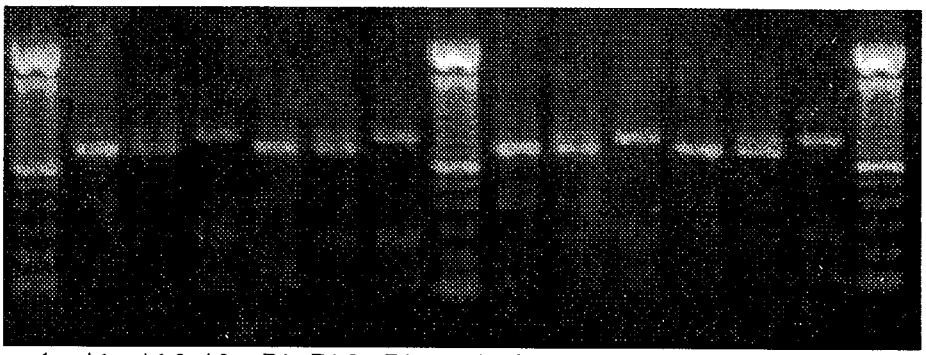
FIG. 8-2



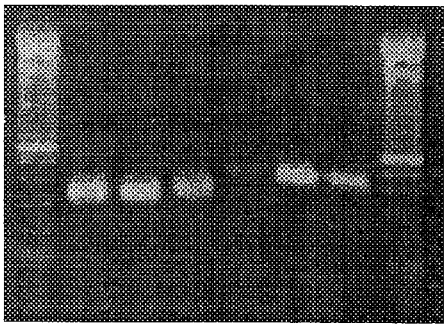
IDT Exhibit 1020.0192

FIG. 9

A)

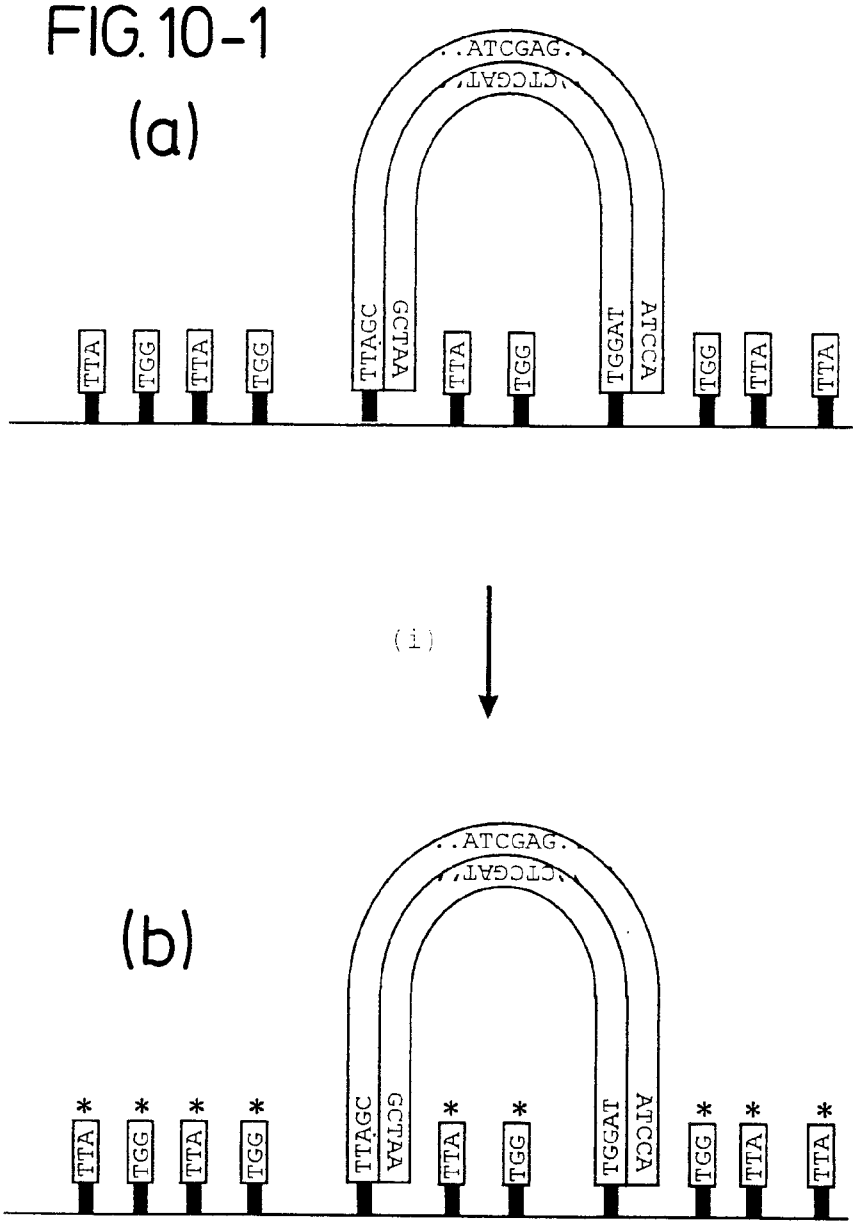


B)



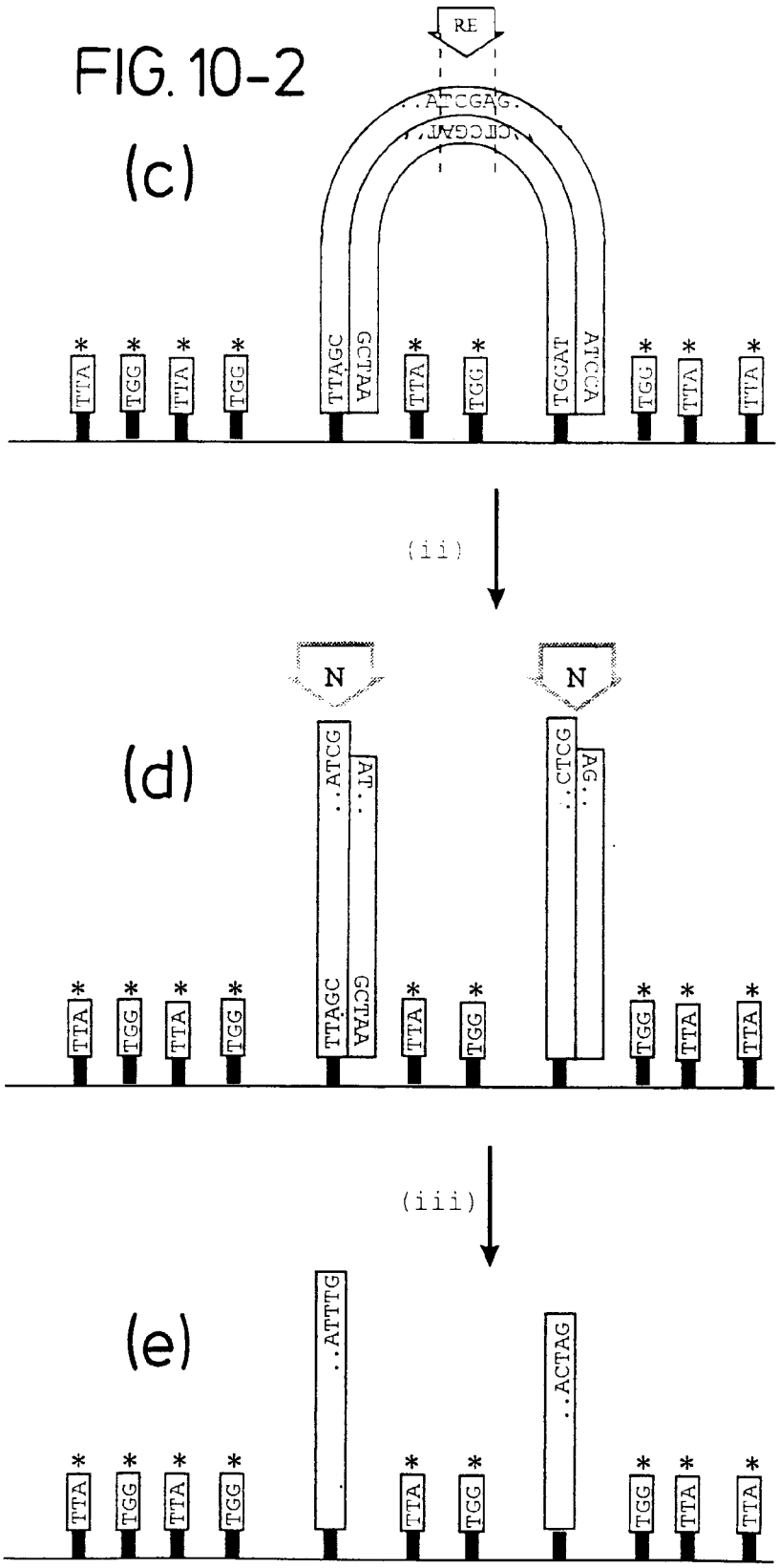
IDT Exhibit 1020.0193

FIG. 10-1



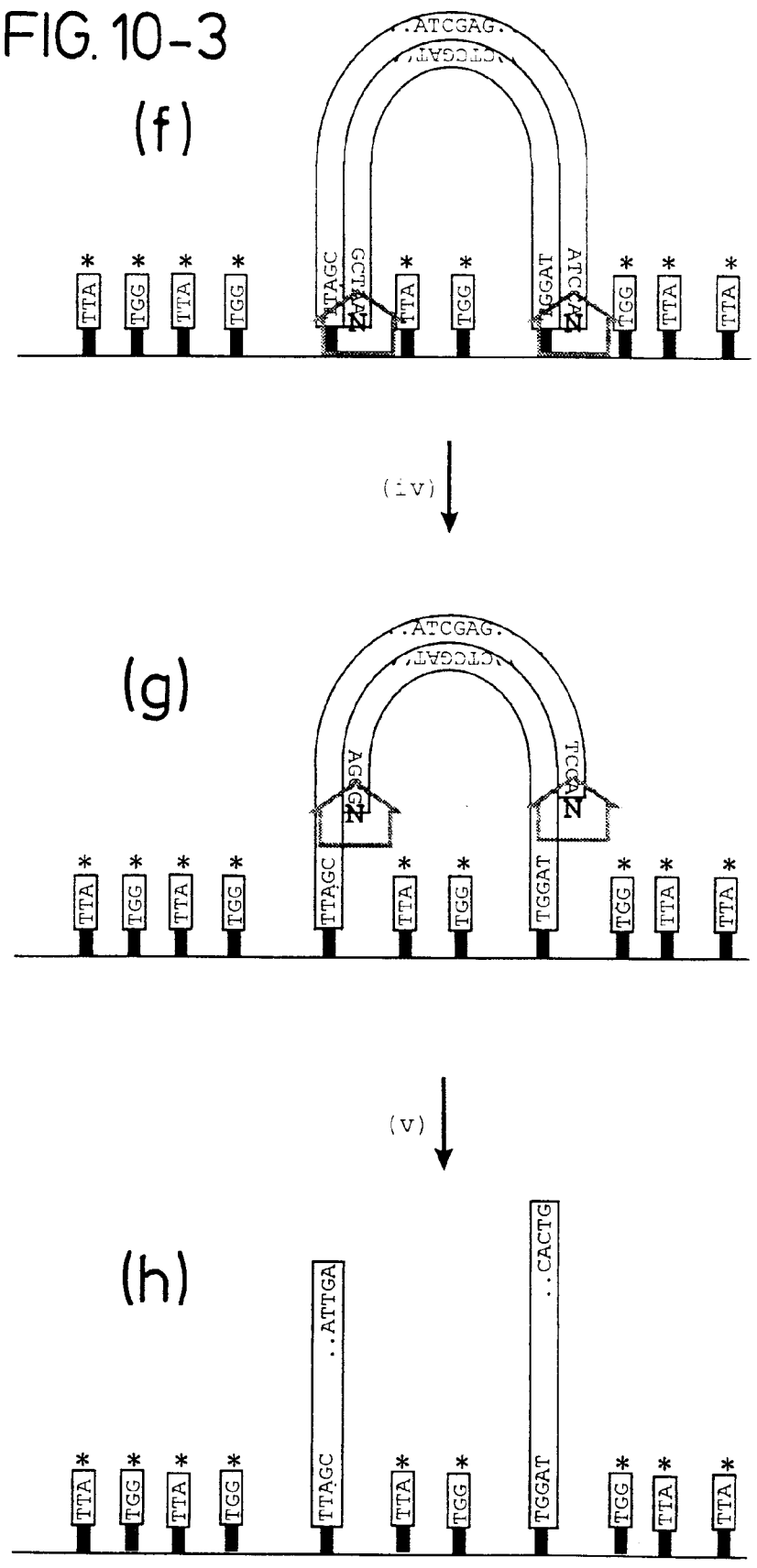
IDT Exhibit 1020.0194

FIG. 10-2



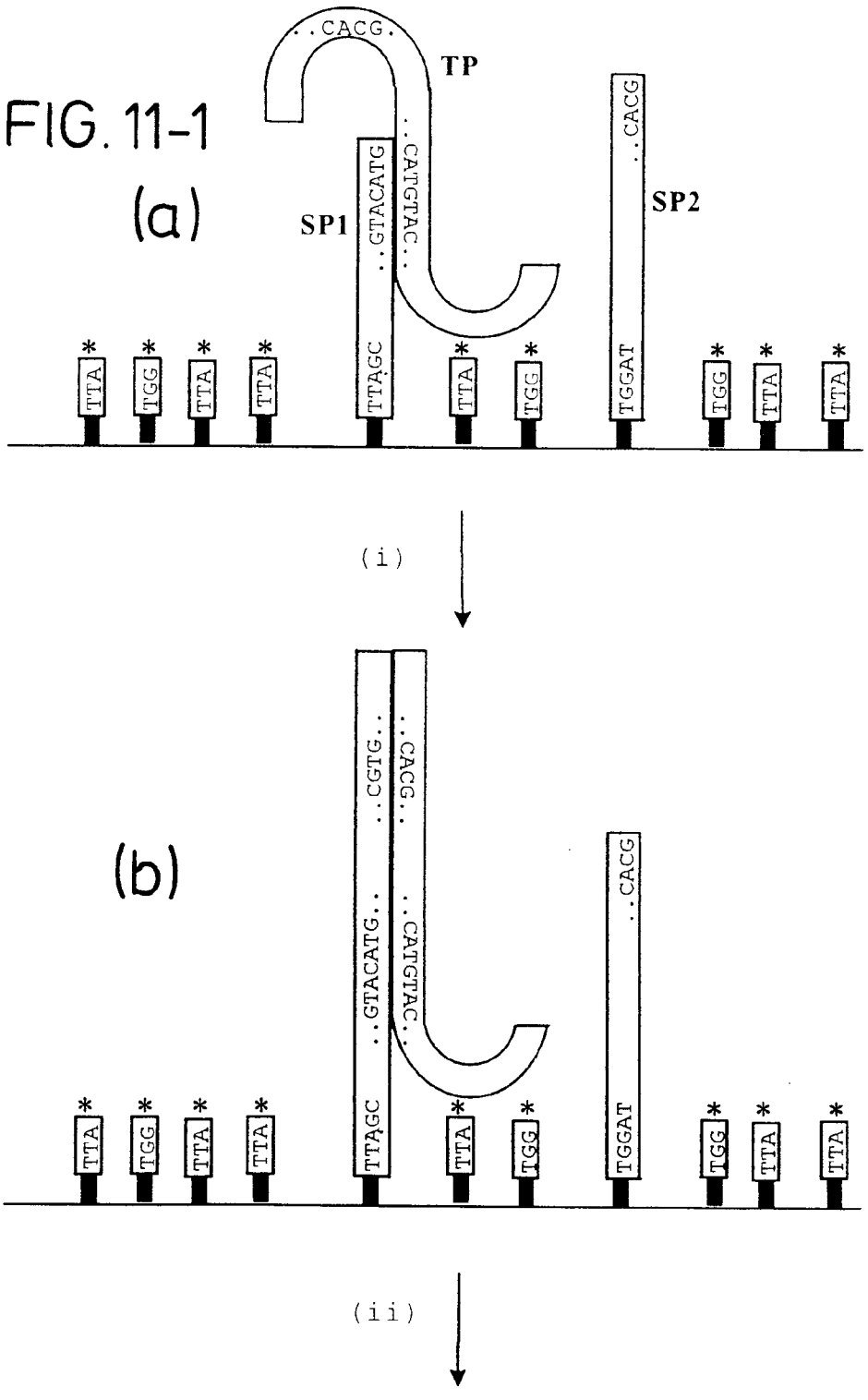
IDT Exhibit 1020.0195

FIG. 10-3



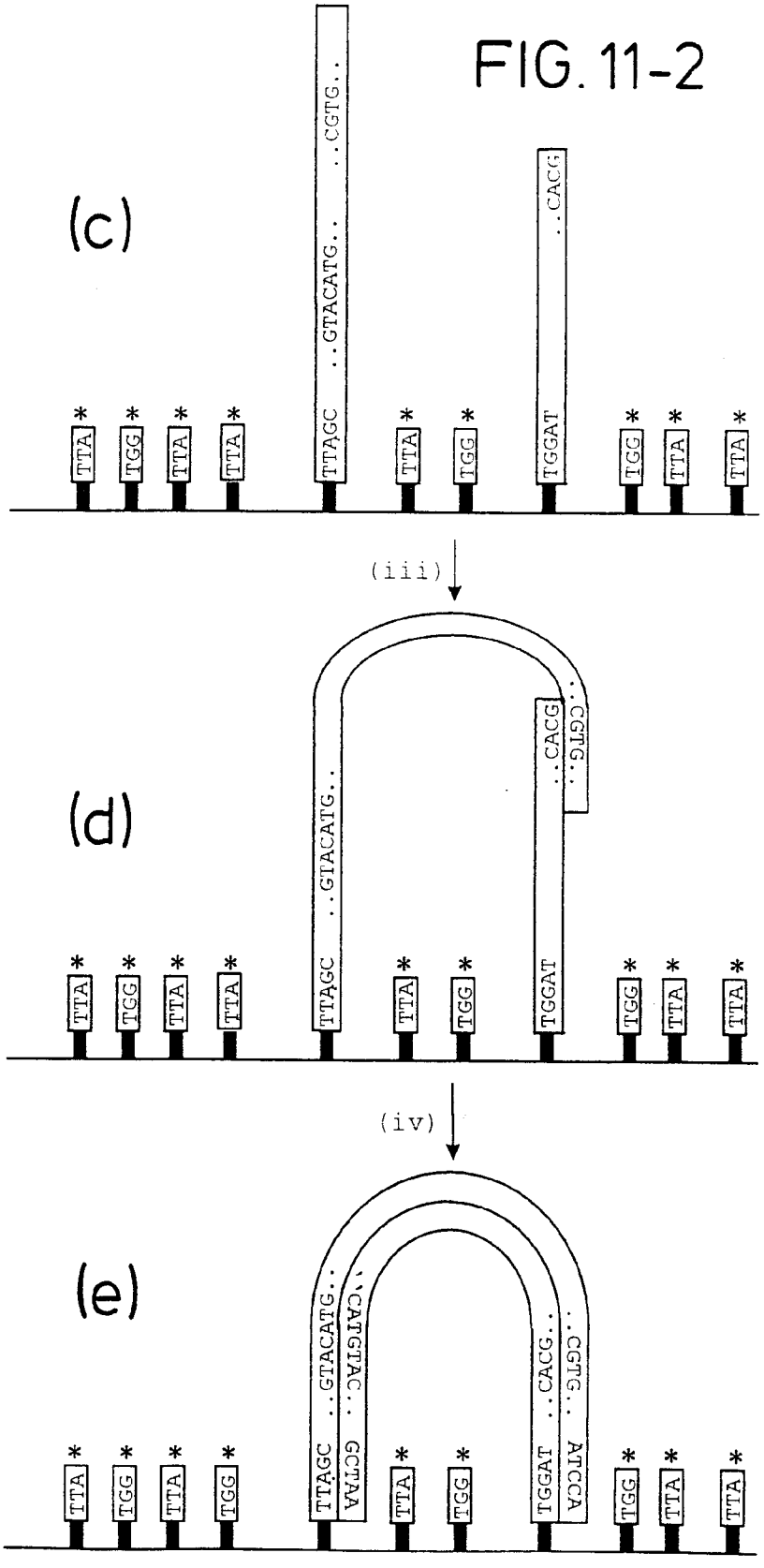
IDT Exhibit 1020.0196

FIG. 11-1



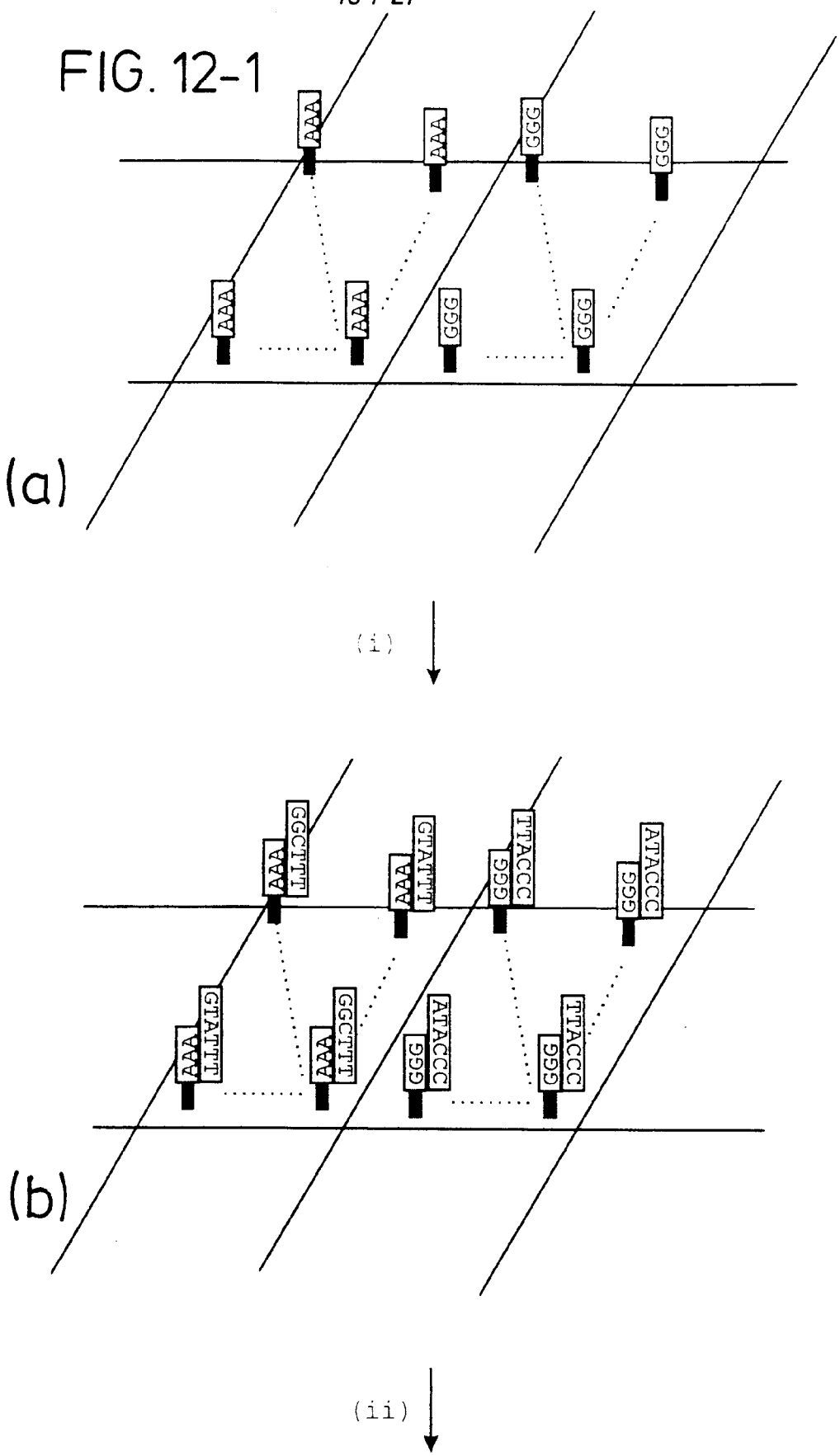
IDT Exhibit 1020.0197

FIG. 11-2



IDT Exhibit 1020.0198

FIG. 12-1



IDT Exhibit 1020.0199

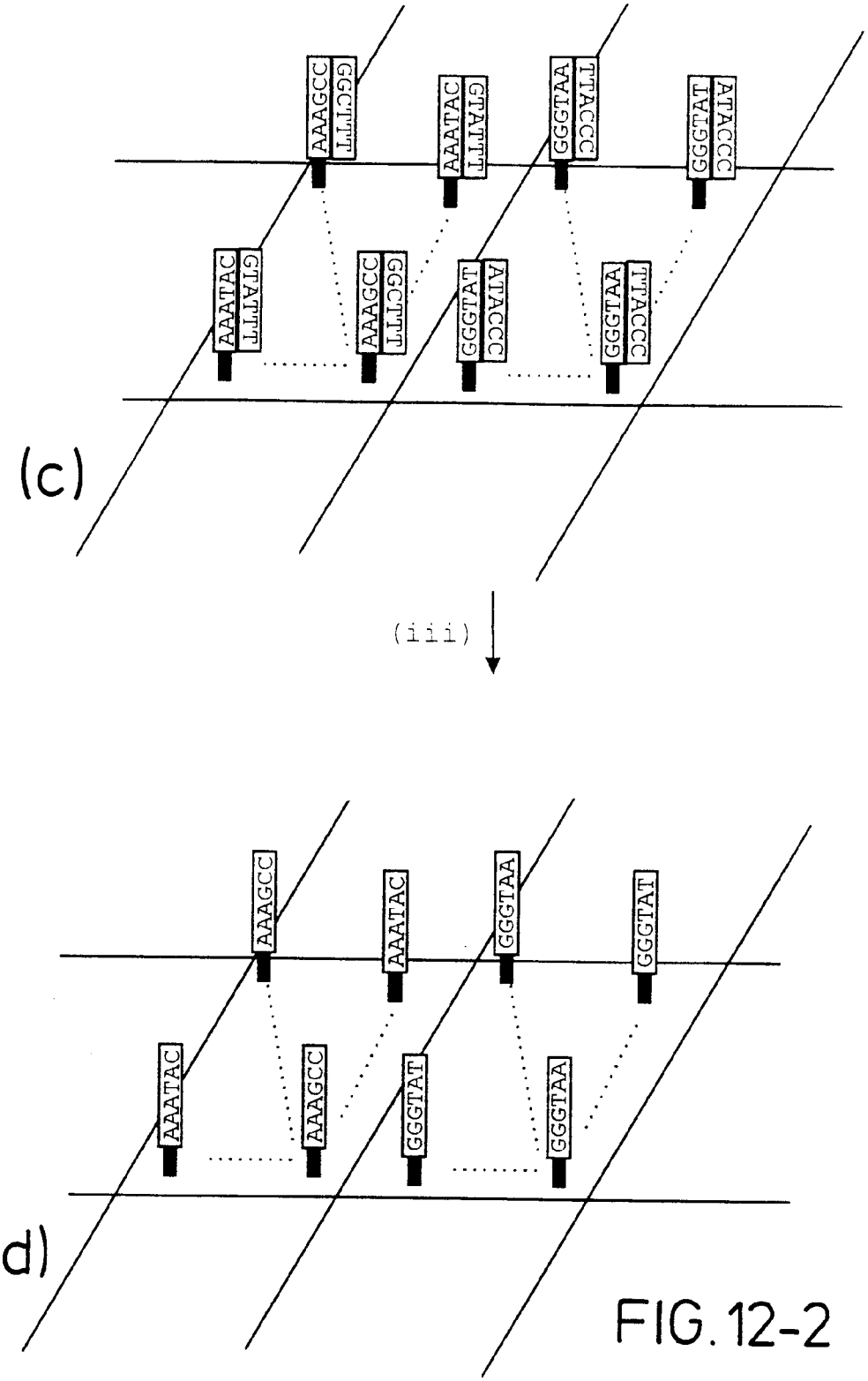


FIG. 12-2

IDT Exhibit 1020.0200

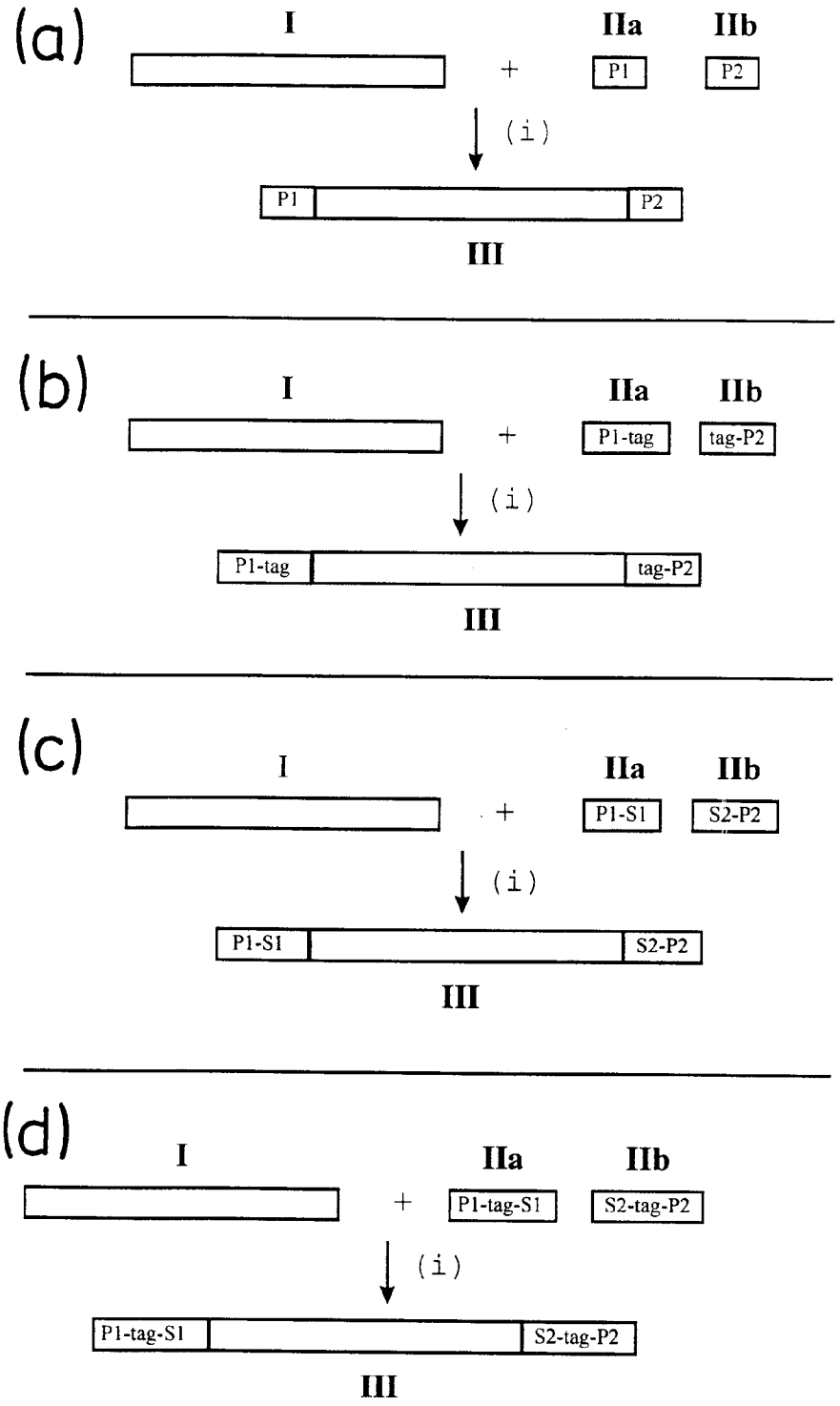


FIG. 13

IDT Exhibit 1020.0201

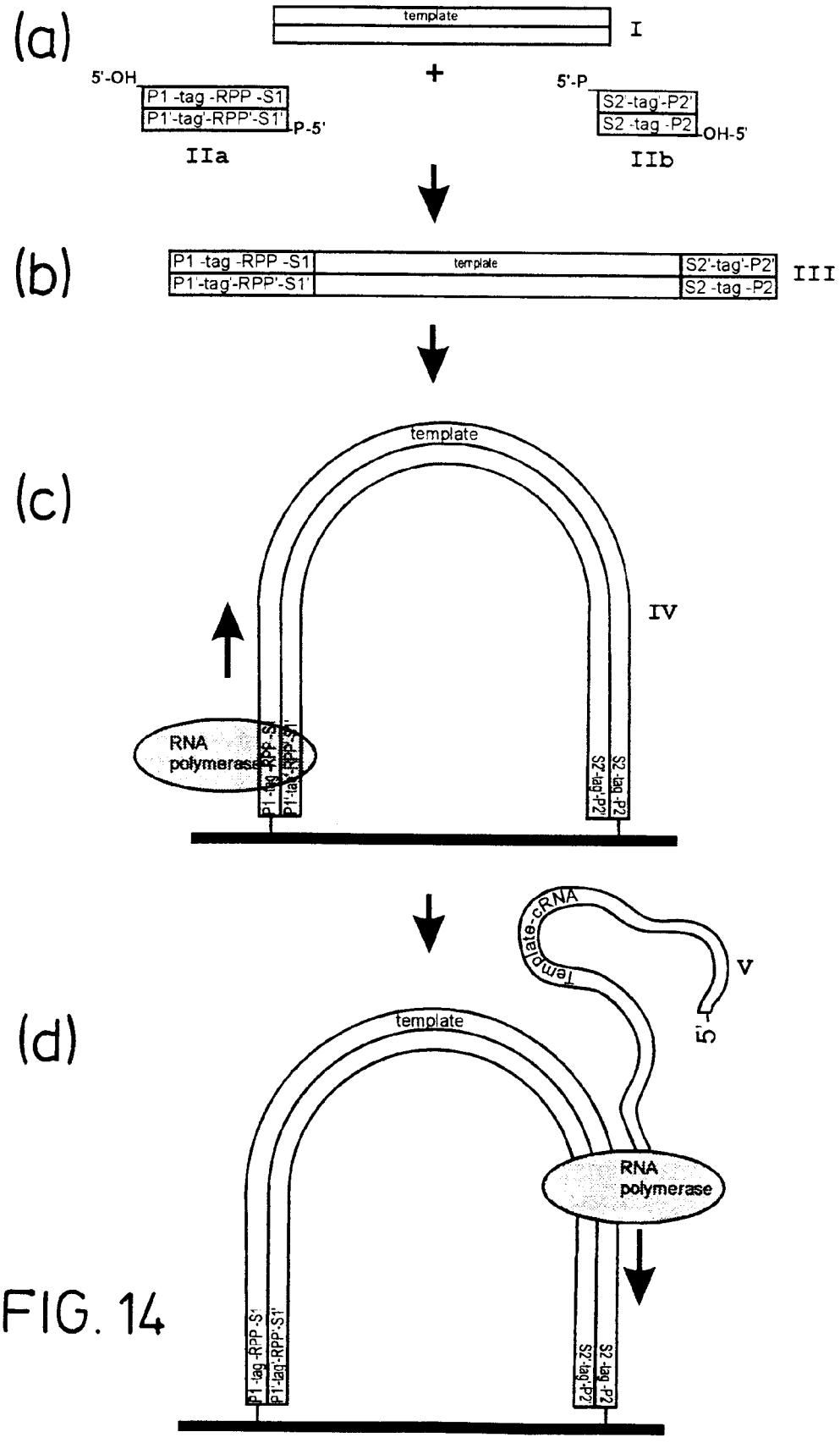
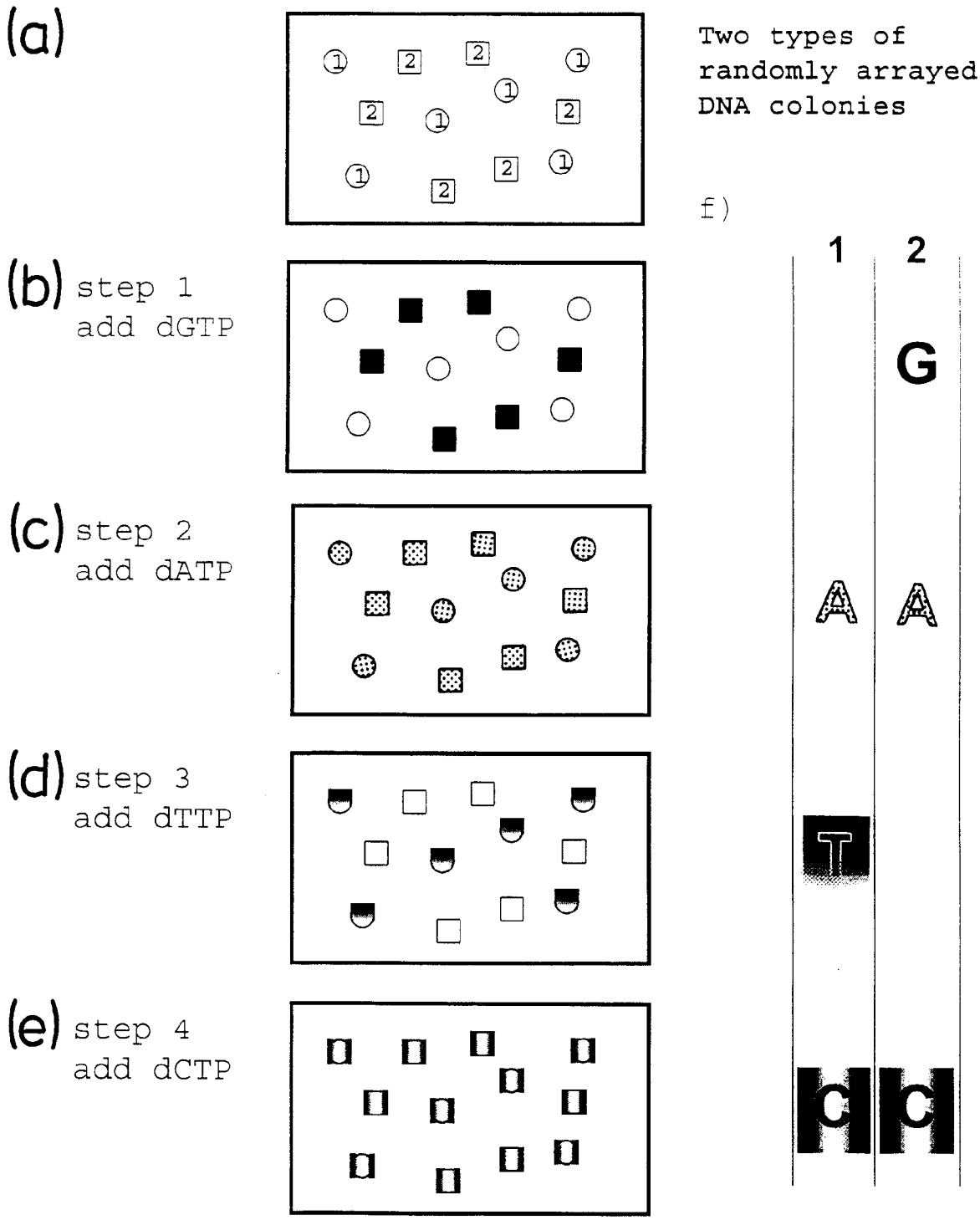


FIG. 14

IDT Exhibit 1020.0202

FIG. 15

in situ Sequencing



IDT Exhibit 1020.0203

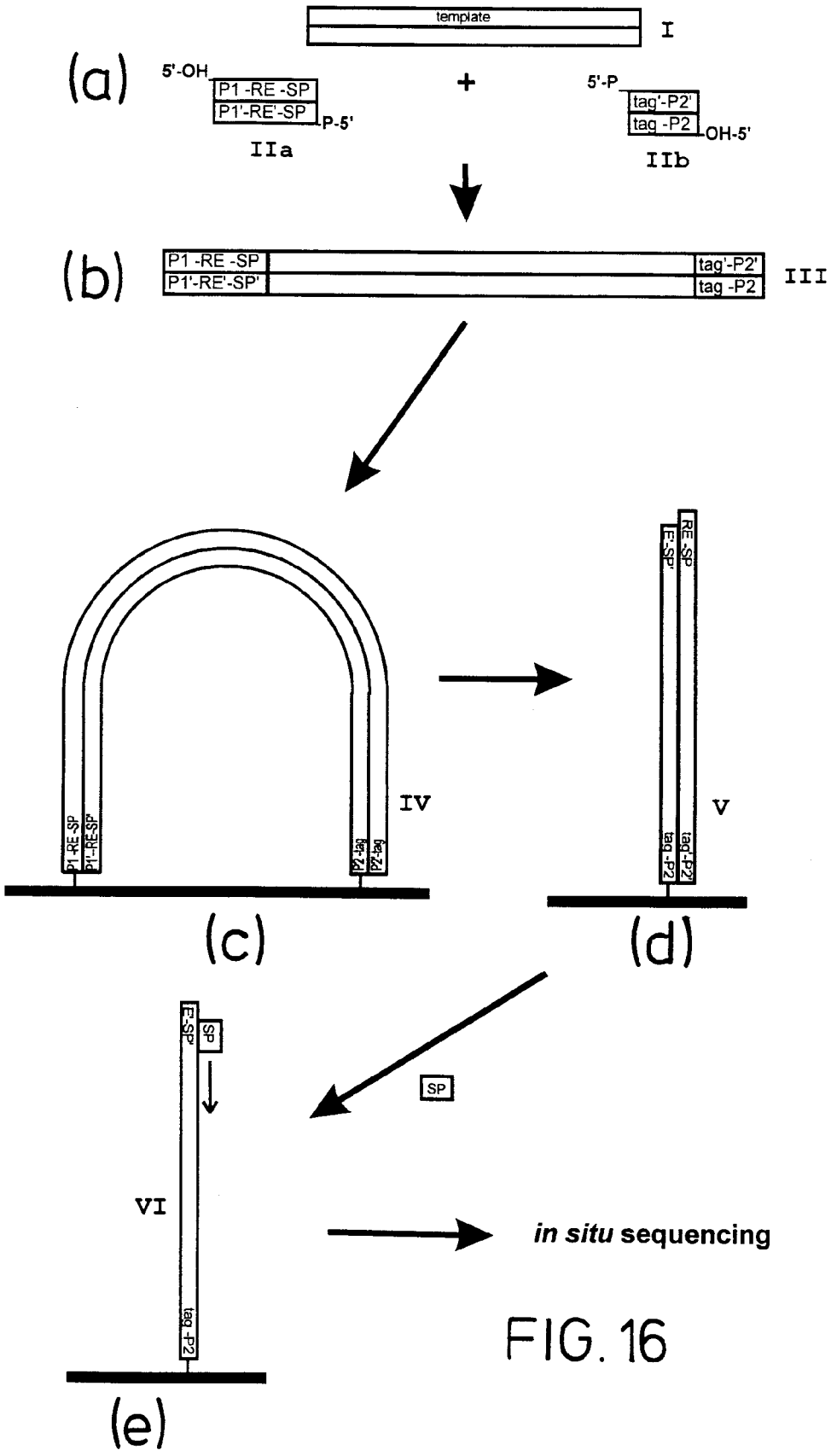


FIG. 16

IDT Exhibit 1020.0204

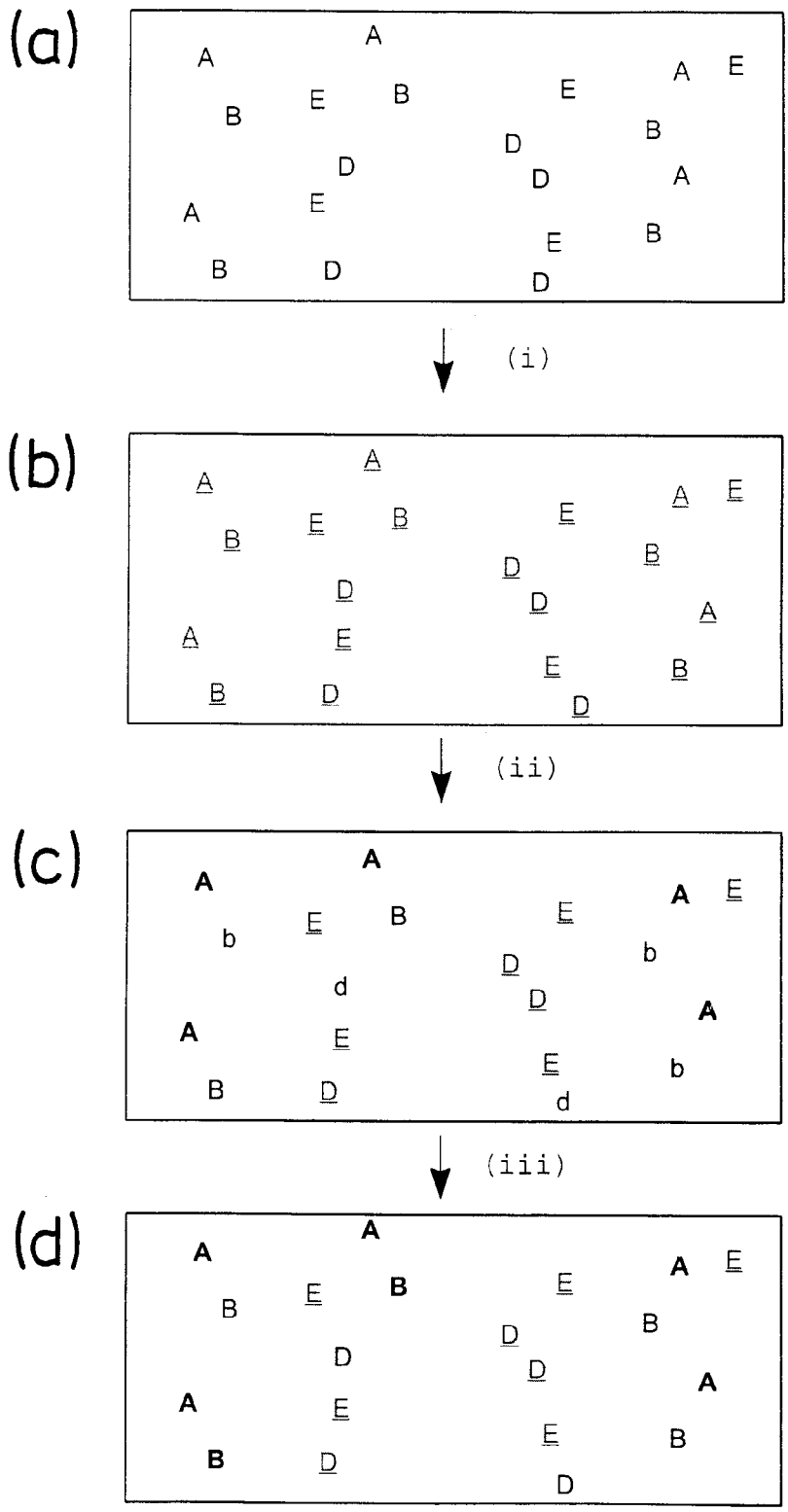


FIG. 17

IDT Exhibit 1020.0205

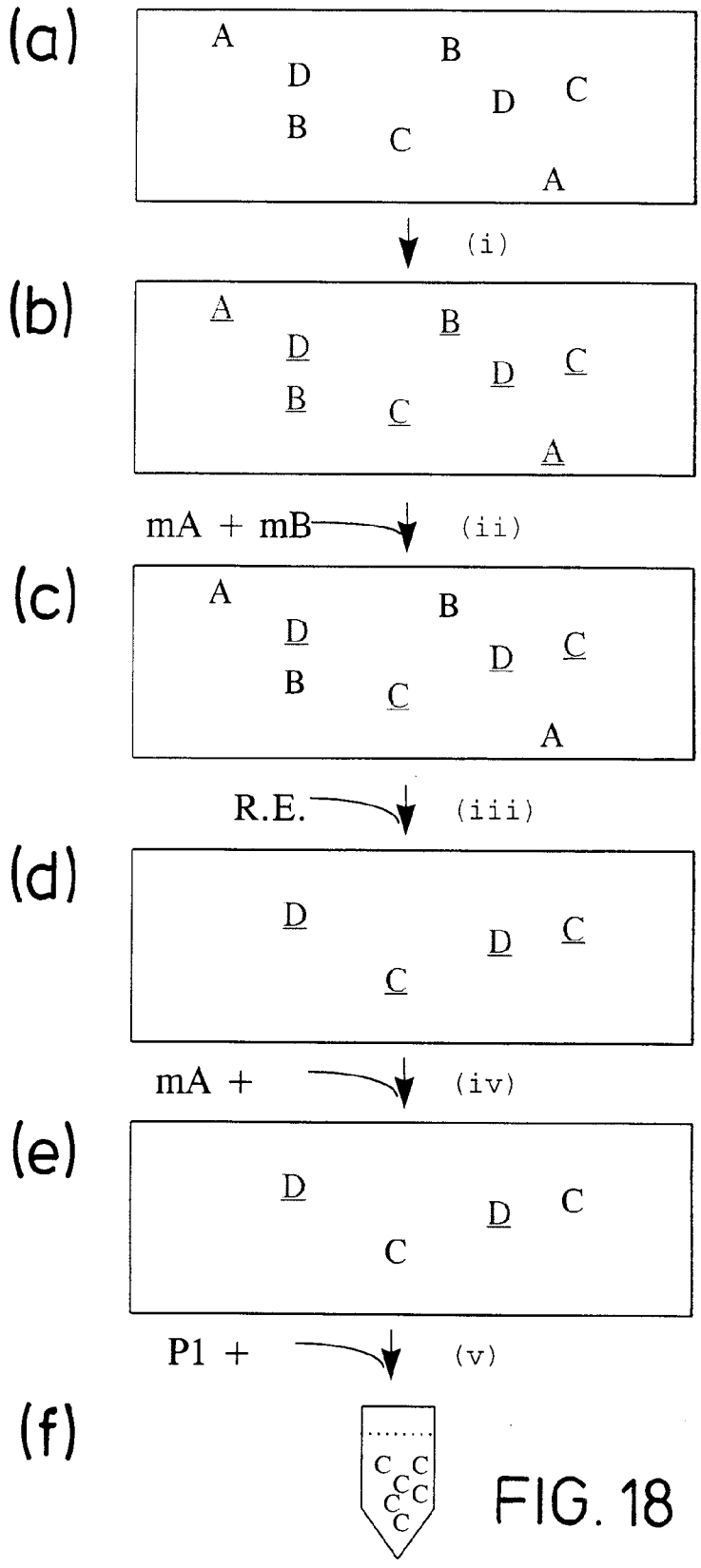


FIG. 18

IDT Exhibit 1020.0206

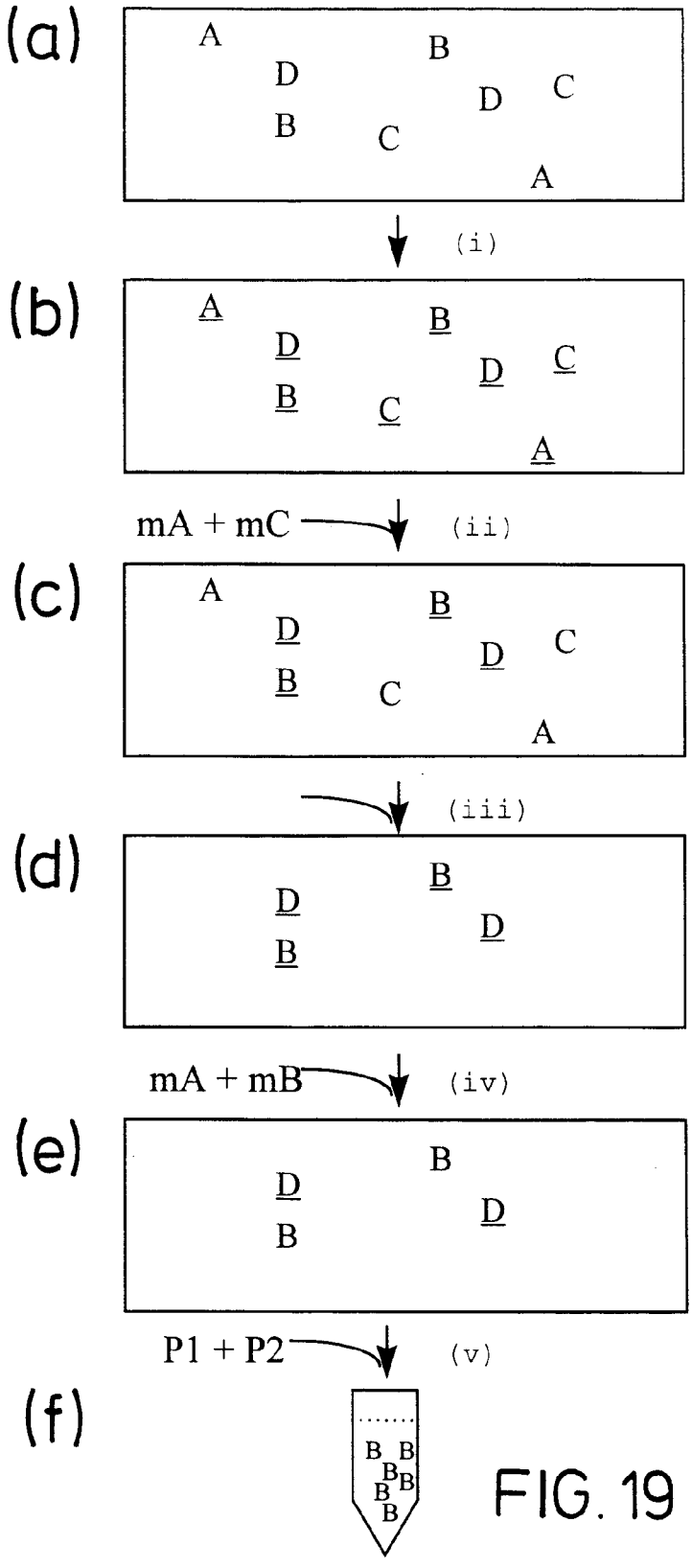


FIG. 19

IDT Exhibit 1020.0207

INTERNATIONAL SEARCH REPORT

Int tional Application No
PCT/GB 98/00961

A. CLASSIFICATION OF SUBJECT MATTER
IPC 6 C12Q1/68

According to International Patent Classification(IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 6 C12Q

Documentation searched other than minimumdocumentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X Y	WO 96 04404 A (MOSAIC TECHNOLOGIES INC ;WHITEHEAD FOR BIOMEDICAL RESEA (US); ADAM) 15 February 1996 see the whole document	1-29, 48-56, 63,64 30-47, 57-62
Y	WO 96 32504 A (UNIV BOSTON) 17 October 1996 see the whole document	30-47
Y	OROSKAR ET AL.: "DETECTION OF IMMOBILIZED AMPLICONS BY ELISA-LIKE TECHNIQUES" CLIN.CHEM., vol. 42, no. 9, 1996, pages 1547-1555, XP002074418 cited in the application see the whole document	57-59

-/--

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

° Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publicationdate of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of theinternational search	Date of mailing of the international search report
18 August 1998	01/09/1998

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Hagenmaier, S

2

Form PCT/ISA/210 (second sheet) (July 1992)

IDT Exhibit 1020.0208

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication,where appropriate, of the relevant passages	Relevant to claim No.
Y	YERSHOV ET AL.: "DNA ANALYSIS AND DIAGNOSTIC ON OLIGONUCLEOTIDE MICROCHIPS" PNAS, vol. 93, May 1996, pages 4913-4918, XP002074419 cited in the application see the whole document	60-62
A	LOCKHART ET AL.: "EXPRESSION MONITORING BY HYBRIDIZATION TO HIGH-DENSITY OLIGONUCLEOTIDE ARRAYS" NATURE BIOTECHNOLOGY, vol. 14, 14 December 1996, pages 1675-1680, XP002074420 cited in the application see the whole document	1-64
A	EP 0 665 293 A (HAMAMATSU PHOTONICS KK) 2 August 1995 see abstract	1-64
A	FU D -J ET AL: "SEQUENCING DOUBLE-STRANDED DNA BY STRAND DISPLACEMENT" NUCLEIC ACIDS RESEARCH, vol. 25, no. 3, 1997, pages 677-679, XP000196977 see the whole document	1-64
A	WO 93 04199 A (SCIENT GENERICS LTD) 4 March 1993 see the whole document	1-64
A	WO 91 06678 A (STANFORD RES INST INT ;TSIEN ROGER Y (US)) 16 May 1991 see the whole document	1-64
A	US 5 589 332 A (SHIH ANDY ET AL) 31 December 1996 see the whole document	19,20

2

IDT Exhibit 1020.0209

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No
PCT/GB 98/00961

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9604404 A	15-02-1996	US 5641658 A	24-06-1997
		CA 2196604 A	15-02-1996
		EP 0784701 A	23-07-1997
		JP 10505492 T	02-06-1998
WO 9632504 A	17-10-1996	AU 5544696 A	30-10-1996
		CA 2218188 A	17-10-1996
		EP 0830460 A	25-03-1998
EP 0665293 A	02-08-1995	JP 7203998 A	08-08-1995
WO 9304199 A	04-03-1993	NONE	
WO 9106678 A	16-05-1991	CA 2044616 A	27-04-1991
		EP 0450060 A	09-10-1991
US 5589332 A	31-12-1996	AU 675482 B	06-02-1997
		AU 5739694 A	04-07-1994
		EP 0681613 A	15-11-1995
		JP 8507202 T	06-08-1996
		WO 9413833 A	23-06-1994

IDT Exhibit 1020.0210

Espacenet

Bibliographic data: WO9942618 (A1) — 1999-08-26

QUANTITATIVE DETERMINATION OF NUCLEIC ACID AMPLIFICATION PRODUCTS

Inventor(s): PATEL RAJESH; KURN NURITH ± (PATEL, RAJESH, ; KURN, NURITH)

Applicant(s): DADE BEHRING INC [US] ± (DADE BEHRING INC)

Classification: **- international:** ***C12N15/09; C12Q1/68;*** (IPC1-7): C12Q1/68
- cooperative: **C12Q1/6851**; C12Q1/6818 more

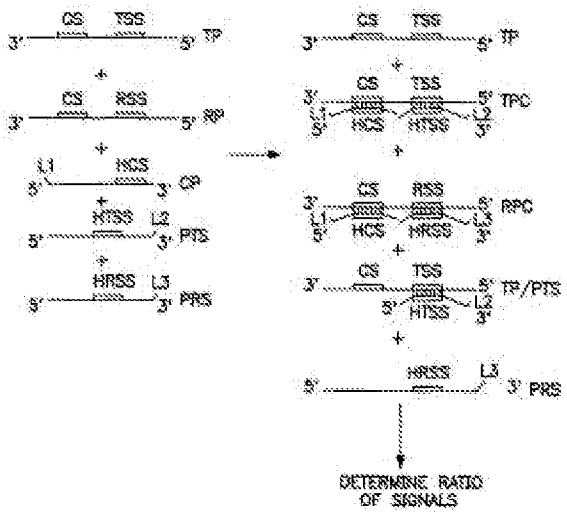
Application number: WO1999US03426 19990217 Global Dossier

Priority number (s): US19980025639 19980218

Also published as: DE69935314 (T2) EP1055005 (A1) EP1055005 (B1) ES2283124 (T3) JP2002504350 (A) more

Abstract of WO9942618 (A1)

The present invention relates to a method for detecting the amount of a target polynucleotide in a sample. A combination is provided in a medium. The combination comprises: (i) a sample suspected of containing the target polynucleotide, the target polynucleotide being in single stranded form; (ii) a reference polynucleotide comprising a sequence that is common with a sequence of the target polynucleotide; and (iii) a predetermined amount of an oligonucleotide probe that has a sequence that hybridizes with the sequence that is common. The combination is subjected to conditions for amplifying the target polynucleotide and the reference polynucleotide. The conditions permit formation of substantially non-dissociative complexes of the target polynucleotide and the reference polynucleotide,



IDT Exhibit 1020.0211

respectively, with the oligonucleotide probe. Furthermore, the predetermined amount of the oligonucleotide probe is less than the expected amount of the amplified target polynucleotide. The ratio of the amount of the complex of the target polynucleotide with the oligonucleotide probe to the amount of the complex of the reference polynucleotide with the oligonucleotide probe is determined. Determination of the ratio is facilitated by employing second and third oligonucleotide probes. The second oligonucleotide probe has a sequence that hybridizes only with the second sequence of the target polynucleotide. The third oligonucleotide probe has a sequence that hybridizes only with a respective second sequence of the reference polynucleotide. The ratio is related to the known amount of the reference polynucleotide to determine the amount of the target polynucleotide in the sample. One or more reference polynucleotides may be employed with a corresponding third oligonucleotide probe for each reference probe. Kits for carrying out the above methods are also disclosed. The method is particularly applicable to the amplification and detection of RNA.

IDT Exhibit 1020.0212

PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : C12Q 1/68

A1

(11) International Publication Number: **WO 99/42618**

(43) International Publication Date: 26 August 1999 (26.08.99)

(21) International Application Number: PCT/US99/03426

(22) International Filing Date: 17 February 1999 (17.02.99)

(30) Priority Data:
09/025,639 18 February 1998 (18.02.98) US

(71) Applicant: DADE BEHRING INC. [US/US]; 1717 Deerfield Road, Deerfield, IL 60015 (US).

(72) Inventors: PATEL, Rajesh; 34270 Northwind Terrace, Fremont, CA 94555 (US). KURN, Nurith; 2876 Ramona Street, Palo Alto, CA 94306 (US).

(74) Agents: RUSZALA, Lois, K. et al.; Dade Behring Inc., 1717 Deerfield Road, Deerfield, IL 60015 (US).

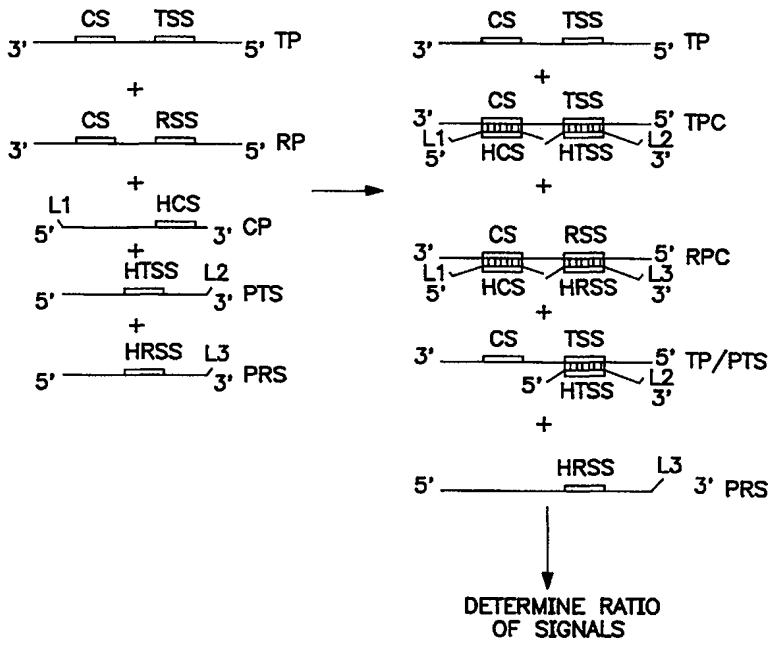
(81) Designated States: CA, JP, European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: QUANTITATIVE DETERMINATION OF NUCLEIC ACID AMPLIFICATION PRODUCTS

(57) Abstract

The present invention relates to a method for detecting the amount of a target polynucleotide in a sample. A combination is provided in a medium. The combination comprises: (i) a sample suspected of containing the target polynucleotide, the target polynucleotide being in single stranded form; (ii) a reference polynucleotide comprising a sequence that is common with a sequence of the target polynucleotide; and (iii) a predetermined amount of an oligonucleotide probe that has a sequence that hybridizes with the sequence that is common. The combination is subjected to conditions for amplifying the target polynucleotide and the reference polynucleotide. The conditions permit formation of substantially non–dissociative complexes of the target polynucleotide and the reference polynucleotide, respectively, with the oligonucleotide probe. Furthermore, the predetermined amount of the oligonucleotide probe is less than the expected amount of the amplified target polynucleotide. The ratio of the amount of the complex of the target polynucleotide with the oligonucleotide probe to the amount of the complex of the reference polynucleotide with the oligonucleotide probe is determined. Determination of the ratio is facilitated by employing second and third oligonucleotide probes. The second oligonucleotide probe has a sequence that hybridizes only with the second sequence of the target polynucleotide. The third oligonucleotide probe has a sequence that hybridizes only with a respective second sequence of the reference polynucleotide. The ratio is related to the known amount of the reference polynucleotide to determine the amount of the target polynucleotide in the sample. One or more reference polynucleotides may be employed with a corresponding third oligonucleotide probe for each reference probe. Kits for carrying out the above methods are also disclosed. The method is particularly applicable to the amplification and detection of RNA.



IDT Exhibit 1020.0213

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	**ES**	Spain	**LS**	Lesotho	**SI**	Slovenia
AM	Armenia	**FI**	Finland	**LT**	Lithuania	**SK**	Slovakia
AT	Austria	**FR**	France	**LU**	Luxembourg	**SN**	Senegal
AU	Australia	**GA**	Gabon	**LV**	Latvia	**SZ**	Swaziland
AZ	Azerbaijan	**GB**	United Kingdom	**MC**	Monaco	**TD**	Chad
BA	Bosnia and Herzegovina	**GE**	Georgia	**MD**	Republic of Moldova	**TG**	Togo
BB	Barbados	**GH**	Ghana	**MG**	Madagascar	**TJ**	Tajikistan
BE	Belgium	**GN**	Guinea	**MK**	The former Yugoslav Republic of Macedonia	**TM**	Turkmenistan
BF	Burkina Faso	**GR**	Greece			**TR**	Turkey
BG	Bulgaria	**HU**	Hungary	**ML**	Mali	**TT**	Trinidad and Tobago
BJ	Benin	**IE**	Ireland	**MN**	Mongolia	**UA**	Ukraine
BR	Brazil	**IL**	Israel	**MR**	Mauritania	**UG**	Uganda
BY	Belarus	**IS**	Iceland	**MW**	Malawi	**US**	United States of America
CA	Canada	**IT**	Italy	**MX**	Mexico	**UZ**	Uzbekistan
CF	Central African Republic	**JP**	Japan	**NE**	Niger	**VN**	Viet Nam
CG	Congo	**KE**	Kenya	**NL**	Netherlands	**YU**	Yugoslavia
CH	Switzerland	**KG**	Kyrgyzstan	**NO**	Norway	**ZW**	Zimbabwe
CI	Côte d'Ivoire	**KP**	Democratic People's Republic of Korea	**NZ**	New Zealand		
CM	Cameroon			**PL**	Poland		
CN	China	**KR**	Republic of Korea	**PT**	Portugal		
CU	Cuba	**KZ**	Kazakstan	**RO**	Romania		
CZ	Czech Republic	**LC**	Saint Lucia	**RU**	Russian Federation		
DE	Germany	**LI**	Liechtenstein	**SD**	Sudan		
DK	Denmark	**LK**	Sri Lanka	**SE**	Sweden		
EE	Estonia	**LR**	Liberia	**SG**	Singapore		

IDT Exhibit 1020.0214

QUANTITATIVE DETERMINATION OF NUCLEIC ACID AMPLIFICATION PRODUCTS

BACKGROUND OF THE INVENTION

1. Field of the Invention.

Significant morbidity and mortality are associated with infectious diseases. More rapid and accurate diagnostic methods are required for better monitoring and treatment of disease. Molecular methods using DNA probes, nucleic acid hybridizations and in vitro amplification techniques are promising methods offering advantages to conventional methods used for patient diagnoses.

Nucleic acid hybridization has been employed for investigating the identity and establishing the presence of nucleic acids. Hybridization is based on complementary base pairing. When complementary single stranded nucleic acids are incubated together, the complementary base sequences pair to form double stranded hybrid molecules. The ability of single stranded deoxyribonucleic acid (ssDNA) or ribonucleic acid (RNA) to form a hydrogen bonded structure with a complementary nucleic acid sequence has been employed as an analytical tool in molecular biology research. The availability of radioactive nucleoside triphosphates of high specific activity and the ^{32}P labeling of DNA with T4 polynucleotide kinase has made it possible to identify, isolate, and characterize various nucleic acid sequences of biological interest. Nucleic acid hybridization has great potential in diagnosing disease states associated with unique nucleic acid sequences. These unique nucleic acid sequences may result from genetic or environmental change in DNA by insertions, deletions, point mutations, or by acquiring foreign DNA or RNA by means of infection by bacteria, molds, fungi, and viruses. Nucleic acid hybridization has, until now, been employed primarily in academic and industrial molecular biology laboratories. The application of nucleic acid hybridization as a diagnostic tool in clinical medicine is limited because of the frequently very low concentrations of disease related DNA or RNA present in a patient's body fluid and the unavailability of a sufficiently sensitive method of nucleic acid hybridization analysis.

One method for detecting specific nucleic acid sequences generally involves immobilization of the target nucleic acid on a solid support such as nitrocellulose paper, cellulose paper, diazotized paper, or a nylon membrane. After the target nucleic acid is fixed on the support, the support is contacted with a suitably labeled probe nucleic acid for about two to forty-eight hours. After the above time period, the solid support is washed several times at a controlled temperature to remove unhybridized probe. The support is then dried and the hybridized material is detected by autoradiography or by spectrometric methods.

When very low concentrations must be detected, the above method is slow and labor intensive, and nonisotopic labels that are less readily detected than radiolabels are frequently not suitable.

Recently, a method for the enzymatic amplification of specific segments of DNA known as the polymerase chain reaction (PCR) method has been described. This in vitro amplification procedure is based on repeated cycles of denaturation, oligonucleotide primer annealing, and primer extension by thermophilic polymerase, resulting in the exponential increase in copies of the region flanked by the primers. The PCR primers, which anneal to opposite strands of the DNA, are positioned so that the polymerase catalyzed extension product of one primer can serve as a template strand for the other, leading to the accumulation of a discrete fragment whose length is defined by the distance between the 5' ends of the oligonucleotide primers.

Other methods for amplifying nucleic acids have also been developed. These methods include single primer amplification, ligase chain reaction (LCR), nucleic acid sequence based amplification (NASBA) and the Q-beta-replicase method. Regardless of the amplification used, the amplified product must be detected.

After amplification of a particular nucleic acid, a separate step is carried out prior to detecting amplified material. The various nucleic acid amplification procedures developed over the past few years greatly enhanced the sensitivity of detecting defined nucleic acid species in a test sample. The frequent formation of non-specific amplification products requires selective detection of the specific

amplification product, which is often carried out using multiple probes complementary to regions within the specific amplified sequence.

One method for detecting nucleic acids is to employ nucleic acid probes that have sequences complementary to sequences in the amplified nucleic acid. One method utilizing such probes is described in U.S. Patent No. 4,868,104. A nucleic acid probe may be, or may be capable of being, labeled with a reporter group or may be, or may be capable of becoming, bound to a support. Detection of signal depends upon the nature of the label or reporter group. If the label or reporter group is an enzyme, additional members of the signal producing system include enzyme substrates and so forth.

Usually, the probe is comprised of natural nucleotides such as ribonucleotides and deoxyribonucleotides and their derivatives although unnatural nucleotide mimetics such as peptide nucleic acids (PNA) and oligomeric nucleoside phosphonates are also used. Commonly, binding of the probes to the target is detected by means of a label incorporated into the probe. Binding can be detected by separating the bound probe from the free probe and detecting the label. For this purpose it is usually necessary to form a sandwich comprised of the labeled probe, the target and a probe that is or can become bound to a surface. Alternatively, binding can be detected by a change in the signal-producing properties of the label upon binding, such as a change in the emission efficiency of a fluorescent or chemiluminescent label. This permits detection to be carried out without a separation step.

Homogeneous methods that have been utilized include the Taqman method used by Roche Molecular Diagnostics. In this approach a probe is used that is labeled with a fluorescer and a quencher. The polymerase used in PCR is capable of cutting the probe when it is bound to the target DNA and causing separation of these labels. Changes in the polarization of fluorescence upon binding of a fluorescer-labeled probe to target DNA are used by Becton Dickenson to detect the formation of DNA in Strand Displacement Amplification (SDA). Binding of two probes, one with a chemiluminescer bead and one with a sensitizer bead has been used by Behring Diagnostics Inc. for detection of DNA produced by PCR and single

primer amplification. Binding of an electroluminescent ruthenium labeled probe to a biotinylated target RNA and capture of the complex on magnetic beads has been used by Organon Teknika for detection of RNA produced in NASBA. GenProbe has carried out detection of RNA by means of an acridinium labeled probe that changes chemiluminescence efficiency when the probe is bound to target RNA.

Each of the above methods has limitations. Where two or more probes are required for detection and quantitation of products of specific nucleic acid amplifications, increasing the amount of target increases the signal up to a point and then the signal fall off (the "prozone" phenomenon). In general, loss of signal is realized under the prozone phenomenon when the analyte concentration exceeds probe concentration. Methods that require a capturable ligand in the target cannot be used on non-amplified nucleic acids nor are all amplification methods capable of introducing a ligand into the amplified product. Fluorescence polarization changes on binding are small and the sensitivity is therefore limited. Taqman is subject to problems with emission form the quencher, which limits sensitivity; GenProbe's chemiluminescent probe requires reagent additions prior to detection.

It is desirable to have a sensitive, simple method for amplifying and detecting nucleic acids preferably, in a homogeneous format. The method should minimize the number and complexity of steps and reagents and avoid the prozone phenomenon.

Also, it is desirable to know the concentration of the amplified product in the reaction medium. Quantitation of nucleic acid amplification products has become an important molecular diagnostic tool.

2. Description of the Related Art.

Rapid, non-separation electrochemiluminescent DNA hybridization assays for PCR products using 3'-labeled oligonucleotide probes is described by Gudibande, *et al.*, (1992) Molecular and Cellular Probes, 6: 495-503. A related disclosure is found in international patent application WO 9508644 A1 (950330).

Marmaro, *et al.*, (Meeting of the American Association of Clinical Chemists, San Diego, California, November 1994, Poster No. 54) discusses the design and use of fluorogenic probes in TaqMan, a homogeneous PCR assay.

German patent application DE 4234086-A1 (92.02.05) (Henco, *et al.*) discusses the determination of nucleic acid sequences amplified *in vitro* in enclosed reaction zone where probe(s) capable of interacting with target sequence is present during or after amplification and spectroscopically measurable parameters of probe undergo change thereby generating signal.

A process for the determination of nucleic acid molecules at low concentrations by amplification using labeled primers is discussed in patent application WO 96-EP5472 961206 (Eigen, *et al.*).

A self-sustained sequence replication electrochemiluminescent nucleic acid assay is disclosed in patent application WO 94-US10732 940921 (Kenten, *et al.*).

A hybridization protection assay is discussed by Matsuoka in Rinsho Kensa (1991) 35(6):627-631.

U.S. Patent No. 5,593,867 (Walker, *et al.*) discloses a fluorescence polarization detection of nucleic acid amplification using fluorescently labeled oligonucleotide probes and detector-probe extension products.

Molecular beacons: probes that fluoresce upon hybridization are discussed by Tyagi, *et al.*) in Nat. Biotechnol. (1996) 14(3):303-308.

U.S. Patent No. 5,656,207 (Woodhead, *et al.*) discloses a method for detecting or quantifying multiple analytes using labelling techniques.

U.S. Patent No. 5,340,716 (Ullman, *et al.*) describes an assay method utilizing photoactivated chemiluminescent labels.

Photoactivatable chemiluminescent matrices are described in patent application WO 94/03812 (Pease, *et al.*).

European Patent Application No. 0 515 194 A2 (Ullman, *et al.*) discloses assay methods utilizing induced luminescence. The references cited therein are incorporated herein by reference including without limitation U.S. Patent No. 5,017,473 (Wagner), which discloses a homogeneous chemiluminescence immunoassay using a light absorbing material, European Patent Application No.

0,345,776 (McCapra), which discloses specific binding assays that utilize a sensitizer as a label, U.S. Patent No. 4,193,983 (Ullman, et al.), which discloses labeled liposome particle compositions and immunoassays therewith, U.S. Patent No. 4,891,324 (Pease, et al.), which describes a particle with luminescer for assays.

SUMMARY OF THE INVENTION

One embodiment of the present invention is a method for detecting the amount of a target polynucleotide in a sample. A combination is provided in a medium. The combination comprises (i) a sample suspected of containing the target polynucleotide, the target polynucleotide being in single stranded form, (ii) a reference polynucleotide comprising a sequence that is common with a sequence of the target polynucleotide, and (iii) a predetermined amount of an oligonucleotide probe that has a sequence that hybridizes with the sequence that is common. The combination is subjected to conditions for amplifying the target polynucleotide and the reference polynucleotide. The conditions permit formation of substantially non-dissociative complexes of the target polynucleotide and the reference polynucleotide, respectively, with the oligonucleotide probe. Furthermore, the predetermined amount of the oligonucleotide probe is less than the expected amount of the amplified target polynucleotide. The ratio of the amount of the complex of the target polynucleotide with the oligonucleotide probe to the amount of the complex of the reference polynucleotide with the oligonucleotide probe is determined. The ration is related to the known amount of the reference polynucleotide to determine the amount of the target polynucleotide in the sample.

Another aspect of the present invention concerns a method for detecting the amount of a target polynucleotide in a sample. In this aspect a combination is provided in a medium. The combination comprises (i) a sample suspected of containing the target polynucleotide, the target polynucleotide being in single stranded form, (ii) predetermined amounts of one or more reference polynucleotides, each of the reference polynucleotides comprising a first sequence that is common with a first sequence of the target polynucleotide and a second sequence that is different from a second sequence of the target polynucleotide, (iii)

a predetermined amount of a first oligonucleotide probe that has a sequence that hybridizes with the sequence that is common, (iv) a second oligonucleotide probe that has a sequence that hybridizes only with the second sequence of the target polynucleotide, and (v) one or more third oligonucleotide probes. Each of the third oligonucleotide probes has a sequence that hybridizes only with a respective second sequence of one of the reference polynucleotide. The combination is subjected to isothermal conditions for amplifying with equal efficiency the target polynucleotide and the one or more reference polynucleotides. The conditions permit formation of a substantially non-dissociative first termolecular complex of the target polynucleotide, the first oligonucleotide probe and the second oligonucleotide probe and a substantially non-dissociative second termolecular complex of each of the reference polynucleotide with the first oligonucleotide probe and a respective third oligonucleotide probe. The predetermined amount of the first oligonucleotide probe is less than the expected amount of the amplified target polynucleotide. A determination is made of the ratio of the amount of the first termolecular complex to the amount of each of the second termolecular complexes. Each of the ratios is related to the predetermined amount of each of the reference polynucleotides to determine the amount of the target polynucleotide in the sample.

Another embodiment of the present invention is a method for detecting the amount of a target polynucleotide in a sample. In this embodiment a combination is provided in a medium. The combination comprises (i) a sample suspected of containing the target polynucleotide, the target polynucleotide being in single stranded form, (ii) predetermined amounts of one or more reference polynucleotides, each of the reference polynucleotides comprising a first sequence that is common with a first sequence of the target polynucleotide and a second sequence that is different from a second sequence of the target polynucleotide, (iii) a predetermined amount of a first oligonucleotide probe that has a sequence that hybridizes with the sequence that is common wherein the first oligonucleotide probe has, or is capable of having, a sensitizer attached thereto, (iv) a second oligonucleotide probe that has a sequence that hybridizes only with the second sequence of the target polynucleotide wherein the second oligonucleotide probe

has, or is capable of having, a first chemiluminescent compound attached thereto, and (v) one or more third oligonucleotide probes. Each of the one or more third oligonucleotide probes has a sequence that hybridizes only with a respective second sequence of one of the reference polynucleotide. In addition, each of the third oligonucleotide probes has, or is capable of having, a second chemiluminescent compound attached thereto. The first and the second chemiluminescent compounds differ in signal produced when activated by the photosensitizer. The second chemiluminescent compound is different for each of the third oligonucleotide probes. The combination is subjected to isothermal conditions for amplifying with equal efficiency the target polynucleotide and each of the reference polynucleotide. The conditions permit formation of a substantially non-dissociative first termolecular complex of the target polynucleotide, the first oligonucleotide probe and the second oligonucleotide probe and a substantially non-dissociative second termolecular complex of each of the reference polynucleotides with the first oligonucleotide probe and a respective third oligonucleotide probe. The predetermined amount of the first oligonucleotide probe is less than the expected amount of the amplified target polynucleotide. The ratio of the amount of the signal produced by the first termolecular complex to the amount of signal produced by each of the second termolecular complexes is determined. Each of the ratios is related to the amount of each respective reference polynucleotide to determine the amount of the target polynucleotide in the sample.

Another embodiment of the present invention is a kit for use in amplification and detection of a target polynucleotide. The kit is a packaged combination and comprises reagents for conducting an amplification of the target polynucleotide and predetermined amounts of one or more reference polynucleotides. Each of the reference polynucleotides comprises a first sequence that is common with a first sequence of the target polynucleotide and a second sequence that is different from a second sequence of the target polynucleotide. Also in the kit are a predetermined amount of a first oligonucleotide probe that has a sequence that hybridizes with the sequence that is common and a second oligonucleotide probe that has a sequence that hybridizes only with the second sequence of the target polynucleotide. The kit

further comprises one or more third oligonucleotide probes, each of the third oligonucleotide probes having a sequence that hybridizes only with a respective second sequence of one of the reference polynucleotide. The kit may also comprise reagents for conducting an isothermal amplification.

Another aspect of the present invention is a kit for use in an amplification and quantitation of a specific RNA. The kit comprises in packaged combination one or more reference RNA's, a promoter, an enzyme and a predetermined amounts of one or more reference polynucleotides. Each of the reference polynucleotides comprising a first sequence that is common with a first sequence of the target polynucleotide and a second sequence that is different from a second sequence of the target polynucleotide. The kit also comprises a predetermined amount of a first oligonucleotide probe that has a sequence that hybridizes with the sequence that is common. The first oligonucleotide probe is labeled with a sensitizer. A second oligonucleotide probe is included that has a sequence that hybridizes only with the second sequence of the target polynucleotide. The second oligonucleotide probe is labeled with a chemiluminescer. The kit further comprises one or more third oligonucleotide probes. Each of the third oligonucleotide probes has a sequence that hybridizes only with a respective second sequence of one of the reference polynucleotide. In addition, each of the third oligonucleotide probes is labeled with a chemiluminescer where the chemiluminescer is different from that of the second oligonucleotide probe and different for each of the third oligonucleotide probes.

BRIEF DESCRIPTION OF THE DRAWINGS

Fig. 1 is a schematic diagram depicting an embodiment in accordance with the present invention.

Fig. 2 is a schematic diagram depicting an alternate embodiment in accordance with the present invention.

DESCRIPTION OF THE SPECIFIC EMBODIMENTS

The present invention provides for detection of nucleic acid sequences, particularly, the products of nucleic acid amplification reactions. The invention is

especially useful for monitoring the formation of a target nucleic acid produced during amplification under isothermal conditions such as that found in NASBA, 3SR, SDA or amplifications using Q-β-replicase. The method is dependent on the formation of the detectable complexes during the isothermal amplification process. A characteristic feature of the isothermal amplification process is the absence of a high temperature step of nucleic acid denaturation. This feature prevents dissociation of the detectable complexes of the single strand amplification product and the probes. The complexes that are formed simultaneously with the amplification process do not undergo dissociation thus avoiding the formation of single probe/product complexes, which would in turn lead to the prozone effect.

In the present invention the test nucleic acid sequence is amplified in the presence of a control sequence designed to be amplified as efficiently as the test sequence. The quantification of the specific amplification product is calculated as a ratio of signal obtained form the reference and test products. Two probes are used for the generation of signal from each of the products. The signal is generated by the association of the two probes, each carrying a reporter group, through hybridization to the single stranded amplification product. The signals from the test and reference products are distinct. The reference sequence is designed to differ from the test sequence by substituting a short sequence of the test with a unique sequence of similar size. This portion of the test and reference sequence is used for binding of the specific probes. The common probe is complementary to a sequence common for the two nucleic acid targets, the test and reference.

Binding of the oligonucleotide probes to the respective single stranded amplification products occurs simultaneously with the formation of the respective products. Limiting the amount of the common probe used results in a binding of the probe to a defined total amount of the single stranded amplification products, the sum of products of amplification of the test and reference targets. Thus, product formed in excess of the amount of the common probe is not considered in the quantitative analysis. In so far as the test and reference nucleic acid targets are amplified at equal efficiency, the amount of common probe binding to the two specific amplification products is dependent of the ratio of input of the two targets.

Binding of the specific probes results in formation of detectable complexes that can be distinguished. An additional important feature of the disclosed method is limiting the quantification of the initial stages of the amplification process, which are expected to be more directly correlated with the initial amount of target than at later stages.

Before proceeding further with a description of the specific embodiments of the present invention, a number of terms will be defined.

Polynucleotide analyte -- a compound or composition to be measured that is a polymeric nucleotide, which in the intact natural state can have about 30 to 5,000,000 or more nucleotides and in an isolated state can have about 20 to 50,000 or more nucleotides, usually about 100 to 20,000 nucleotides, more frequently 500 to 10,000 nucleotides. It is thus obvious that isolation of the analyte from the natural state often results in fragmentation. The polynucleotide analytes include nucleic acids, and fragments thereof, from any source in purified or unpurified form including DNA (dsDNA and ssDNA) and RNA, including t-RNA, m-RNA, r-RNA, mitochondrial DNA and RNA, chloroplast DNA and RNA, DNA-RNA hybrids, or mixtures thereof, genes, chromosomes, plasmids, the genomes of biological material such as microorganisms, e.g., bacteria, yeasts, viruses, viroids, molds, fungi, plants, animals, humans, and the like. The polynucleotide analyte can be only a minor fraction of a complex mixture such as a biological sample. The analyte can be obtained from various biological materials by procedures well known in the art. Some examples of such biological material by way of illustration and not limitation are disclosed in the following Table 1:

Table 1

Microorganisms of interest include:

Corynebacteria
Corynebacterium diphtheria
Pneumococci
Diplococcus pneumoniae
Streptococci
Streptococcus pyrogenes
Streptococcus salivarus
Staphylococci

Staphylococcus aureus
Staphylococcus albus
Neisseria
Neisseria meningitidis
Neisseria gonorrhea
Enterobacteriaciae
Escherichia coli
Aerobacter aerogenes
Klebsiella pneumoniae
Salmonella typhosa
Salmonella choleraesuis
Salmonella typhimurium
Shigella dysenteria
Shigella schmitzii
Shigella arabinotarda
Shigella flexneri
Shigella boydii
Shigella sonnei
Other enteric bacilli
Proteus vulgaris
Proteus mirabilis
Proteus morgani
Pseudomonas aeruginosa
Alcaligenes faecalis
Vibrio cholerae
Hemophilus-Bordetella group
Hemophilus influenza, H. ducryi
Hemophilus hemophilus
Hemophilus aegypticus
Hemophilus parainfluenza
Bordetella pertussis
Pasteurellae
Pasteurella pestis
Pasteurella tulareusis
Brucellae
Brucella melitensis
Brucella abortus
Brucella suis
Aerobic Spore-forming Bacilli
Bacillus anthracis
Bacillus subtilis
Bacillus megaterium
Bacillus cereus
Anaerobic Spore-forming Bacilli
Clostridium botulinum
Clostridium tetani

The colliform bacteria

The *Salmonellae*

The *Shigellae*

Proteus species

Rhizopus oryzae
Rhizopus arrhizua Phycomycetes
Rhizopus nigricans
Sporotrichum schenkii
Flonsecaea pedrosoi
Fonsecacea compact
Fonsecacea dermatidis
Cladosporium carrionii
Phialophora verrucosa
Aspergillus nidulans
Madurella mycetomi
Madurella grisea
Allescheria boydii
Phialophora jeanselmei
Microsporum gypseum
Trichophyton mentagrophytes
Keratinomyces ajelloi
Microsporum canis
Trichophyton rubrum
Microsporum adouini
Viruses

Clostridium perfringens
Clostridium novyi
Clostridium septicum
Clostridium histolyticum
Clostridium tertium
Clostridium bifermentans
Clostridium sporogenes
Mycobacteria Pox Viruses
Mycobacterium tuberculosis *Variola (smallpox)*
Hominis
Mycobacterium bovis *Vaccinia*
Mycobacterium avium *Poxvirus bovis*
Mycobacterium leprae *Paravaccinia*
Mycobacterium paratuberculosis *Molluscum contagiosum*
Actinomycetes (fungus-like bacteria) Picornaviruses
Actinomyces Isaeli Poliovirus
Actinomyces bovis Coxsackievirus
Actinomyces naeslundii Echoviruses
Nocardia asteroides Rhinoviruses
Nocardia brasiliensis Myxoviruses
The Spirochetes Influenza (A, B, and C)
Treponema pallidum *Spirillum minus* Parainfluenza (1-4)
Treponema pertenue *Streptobacillus monoiliformis* Mumps Virus
Newcastle Disease Virus
Treponema carateum Measles Virus
Borrelia recurrentis Rinderpest Virus
Leptospira icterohemorrhagiae Canine Distemper Virus
Leptospira canicola Respiratory Syncytial Virus
Trypanasomes Rubella Virus
Mycoplasmas Arboviruses
Mycoplasma pneumoniae
Other pathogens Eastern Equine Eucephalitis Virus
Listeria monocytogenes Western Equine Eucephalitis Virus
Erysipelothrix rhusiopathiae Sindbis Virus
Streptobacillus moniliformis Chikugunya Virus
Donvania granulomatis Semliki Forest Virus
Bartonella bacilliformis Mayora Virus
Rickettsiae (bacteria-like parasites) St. Louis Encephalitis Virus
Rickettsia prowazekii California Encephalitis Virus
Rickettsia mooseri Colorado Tick Fever Virus
Rickettsia rickettsii Yellow Fever Virus
Rickettsia conori Dengue Virus
Rickettsia australis Reoviruses
Rickettsia sibiricus Reovirus Types 1-3
Retroviruses

Rickettsia akari
Rickettsia tsutsugamushi

Rickettsia burnetti
Rickettsia quintana
Chlamydia (unclassifiable parasites bacterial/viral)
Chlamydia agents (naming uncertain)
Fungi
Cryptococcus neoformans
Blastomyces dermatidis
Hisoplasma capsulatum
Coccidioides immitis
Paracoccidioides brasiliensis
Candida albicans
Aspergillus fumigatus
Mucor corymbifer (Absidia corymbifera)

Human Immunodeficiency Viruses (HIV)
Human T-cell Lymphotrophic Virus I & II (HTLV)
Hepatitis
Hepatitis A Virus
Hepatitis B Virus
Hepatitis nonA-nonB Virus
Tumor Viruses
Rauscher Leukemia Virus
Gross Virus
Maloney Leukemia Virus

Human Papilloma Virus

Also included are genes, such as hemoglobin gene for sickle-cell anemia, cystic fibrosis gene, oncogenes, cDNA, and the like.

The polynucleotide analyte, where appropriate, may be cleaved to obtain a fragment that contains a target polynucleotide sequence, for example, by shearing or by treatment with a restriction endonuclease or other site specific chemical cleavage method.

For purposes of this invention, the polynucleotide analyte, or a cleaved fragment obtained from the polynucleotide analyte, will usually be at least partially denatured or single stranded or treated to render it denatured or single stranded. Such treatments are well-known in the art and include, for instance, heat or alkali treatment. For example, double stranded DNA can be heated at 90-100° C. for a period of about 1 to 10 minutes to produce denatured material.

Amplification of nucleic acids or polynucleotides -- any method that results in the formation of one or more copies of a nucleic acid or polynucleotide molecule or in the formation of one or more copies of the complement of a nucleic acid or polynucleotide molecule.

Exponential amplification of nucleic acids or polynucleotides -- any method that depends on the product catalyzed formation of multiple copies of a nucleic acid or polynucleotide molecule or its complement. The amplification products are

sometimes referred to as "amplicons." One such method for the enzymatic amplification of specific double stranded sequences of DNA is known as the polymerase chain reaction (PCR), as described above. This in vitro amplification procedure is based on repeated cycles of denaturation, oligonucleotide primer annealing, and primer extension by thermophilic template dependent polynucleotide polymerase, resulting in the exponential increase in copies of the desired sequence of the polynucleotide analyte flanked by the primers. The two different PCR primers, which anneal to opposite strands of the DNA, are positioned so that the polymerase catalyzed extension product of one primer can serve as a template strand for the other, leading to the accumulation of a discrete double stranded fragment whose length is defined by the distance between the 5' ends of the oligonucleotide primers.

Another method for amplification is mentioned above and involves amplification of a single stranded polynucleotide using a single oligonucleotide primer. The single stranded polynucleotide that is to be amplified contains two non-contiguous sequences that are complementary to one another and, thus, are capable of hybridizing together to form a stem-loop structure. This single stranded polynucleotide already may be part of a polynucleotide analyte or may be created as the result of the presence of a polynucleotide analyte.

Another method for achieving the result of an amplification of nucleic acids is known as the ligase chain reaction (LCR). This method uses a ligase enzyme to join pairs of preformed nucleic acid probes. The probes hybridize with each complementary strand of the nucleic acid analyte, if present, and ligase is employed to bind each pair of probes together resulting in two templates that can serve in the next cycle to reiterate the particular nucleic acid sequence.

Another method for achieving a nucleic acid amplification is the nucleic acid sequence based amplification (NASBA). This method is a promoter-directed, enzymatic process that induces *in vitro* continuous, homogeneous and isothermal amplification of a specific nucleic acid to provide RNA copies of the nucleic acid. The reagents for conducting NASBA include a first DNA primer with a 5' tail comprising a promoter, a second DNA primer, reverse transcriptase, RNAse-H, T7 RNA polymerase, NTP's and dNTP's.

Another method for amplifying a specific group of nucleic acids is the Q-beta-replicase method, which relies on the ability of Q-beta-replicase to amplify its RNA substrate exponentially. The reagents for conducting such an amplification include "midi-variant RNA" (amplifiable hybridization probe), NTP's, and Q-beta-replicase.

Another method for amplifying nucleic acids is known as 3SR and is similar to NASBA except that the RNAse-H activity is present in the reverse transcriptase. Amplification by 3SR is an RNA specific target method whereby RNA is amplified in an isothermal process combining promoter directed RNA polymerase, reverse transcriptase and RNase H with target RNA.

Another method for amplifying nucleic acids is the Transcription Mediated Amplification (TMA) used by Gen-Probe. The method is similar to NASBA in utilizing two enzymes in a self-sustained sequence replication. See Fahy, *et al.*, patent application WO 91-US8488 911113.

Linear amplification of nucleic acids or polynucleotides -- any method that depends on the self catalyzed formation of one or more copies of the complement of only one strand of a nucleic acid or polynucleotide molecule, usually a nucleic acid or polynucleotide analyte. Thus, the primary difference between linear amplification and exponential amplification is that the latter is autocatalyzed, that is, the product serves to catalyze the formation of more product, whereas in the former process the starting sequence catalyzes the formation of product but is not itself replicated. In linear amplification the amount of product formed increases as a linear function of time as opposed to exponential amplification where the amount of product formed is an exponential function of time.

Target polynucleotide -- a sequence of nucleotides to be identified, usually existing within a portion or all of a polynucleotide analyte, the identity of which is known to an extent sufficient to allow preparation of various oligonucleotides, such as probes and primers, and other molecules necessary for conducting an amplification of the target polynucleotide.

In general, in primer extension amplification primers hybridize to, and are extended along (chain extended), at least the target sequence within the target polynucleotide and, thus, the target sequence acts as a template. The extended primers are chain "extension products." The target sequence usually lies between two defined sequences but need not. In general, the primers hybridize with the defined sequences or with at least a portion of such target polynucleotide, usually at least a ten-nucleotide segment at the 3'-end thereof and preferably at least 15, frequently 20 to 50 nucleotide segment thereof. The target sequence usually contains from about 30 to 5,000 or more nucleotides, preferably 50 to 1,000 nucleotides. The target polynucleotide is generally a fraction of a larger molecule or it may be substantially the entire molecule (polynucleotide analyte). The minimum number of nucleotides in the target polynucleotide sequence is selected to assure that the presence of target polynucleotide in a sample is a specific indicator of the presence of polynucleotide analyte in a sample. Very roughly, the sequence length is usually greater than about 1.6 log L nucleotides where L is the number of base pairs in the genome of the biologic source of the sample. The maximum number of nucleotides in the target polynucleotide is normally governed by the length of the polynucleotide analyte and its tendency to be broken by shearing, or other processes during isolation and any procedures required to prepare the sample for assay and the efficiency of detection and/or amplification of the sequence.

Oligonucleotide -- a polynucleotide, usually single stranded, usually a synthetic polynucleotide but may be a naturally occurring polynucleotide. The oligonucleotide(s) are usually comprised of a sequence of at least 5 nucleotides, preferably, 10 to 100 nucleotides, more preferably, 20 to 50 nucleotides, and usually 10 to 30 nucleotides in length.

Various techniques can be employed for preparing an oligonucleotide utilized in the present invention. Such oligonucleotide can be obtained by biological synthesis or by chemical synthesis. For short sequences (up to about 100 nucleotides) chemical synthesis will frequently be more economical as compared to the biological synthesis. In addition to economy, chemical synthesis provides a convenient way of incorporating low molecular weight compounds and/or modified bases during the synthesis step. Furthermore, chemical synthesis is very flexible in the choice of length and region of the target polynucleotide binding sequence. The oligonucleotide can be synthesized by standard methods such as those used in commercial automated nucleic acid synthesizers. Chemical synthesis of DNA on a suitably modified glass or resin can result in DNA covalently attached to the surface. This may offer advantages in washing and sample handling. For longer sequences standard replication methods employed in molecular biology can be used such as the use of M13 for single stranded DNA as described by J. Messing (1983) Methods Enzymol, 101, 20-78.

Other methods of oligonucleotide synthesis include phosphotriester and phosphodiester methods (Narang, *et al.* (1979) Meth. Enzymol 68: 90) and synthesis on a support (Beaucage, *et al.* (1981) Tetrahedron Letters 22: 1859-1862) as well as phosphoramidate technique, Caruthers, M. H., *et al.*, "Methods in Enzymology," Vol. 154, pp. 287-314 (1988), and others described in "Synthesis and Applications of DNA and RNA," S.A. Narang, editor, Academic Press, New York, 1987, and the references contained therein.

Oligonucleotide probe -- an oligonucleotide employed in the present invention to bind to a portion of a polynucleotide such as an oligonucleotide probe or a target polynucleotide. The design and preparation of the oligonucleotide probes are important in performing the methods of this invention. A more detailed description of oligonucleotide probes in accordance with the present invention is found hereinbelow.

Oligonucleotide primer(s) -- an oligonucleotide that is usually employed in a chain extension on a polynucleotide template such as in, for example, an amplification of a nucleic acid. The oligonucleotide primer is usually a synthetic

nucleotide that is single stranded, containing a sequence at its 3'-end that is capable of hybridizing with a defined sequence of the target polynucleotide. Normally, an oligonucleotide primer has at least 80%, preferably 90%, more preferably 95%, most preferably 100%, complementarity to a defined sequence or primer binding site. The number of nucleotides in the hybridizable sequence of an oligonucleotide primer should be such that stringency conditions used to hybridize the oligonucleotide primer will prevent excessive random non-specific hybridization. Usually, the number of nucleotides in the oligonucleotide primer will be at least as great as the defined sequence of the target polynucleotide, namely, at least ten nucleotides, preferably at least 15 nucleotides and generally from about 10 to 200, preferably 20 to 50, nucleotides.

Nucleoside triphosphates -- nucleosides having a 5'-triphosphate substituent. The nucleosides are pentose sugar derivatives of nitrogenous bases of either purine or pyrimidine derivation, covalently bonded to the 1'-carbon of the pentose sugar, which is usually a deoxyribose or a ribose. The purine bases include adenine(A), guanine (G), inosine (I), and derivatives and analogs thereof. The pyrimidine bases include cytosine (C), thymine (T), uracil (U), and derivatives and analogs thereof. Nucleoside triphosphates include deoxyribonucleoside triphosphates such as the four common triphosphates dATP, dCTP, dGTP and dTTP and ribonucleoside triphosphates such as the four common triphosphates rATP, rCTP, rGTP and rUTP.

The term "nucleoside triphosphates" also includes derivatives and analogs thereof, which are exemplified by those derivatives that are recognized in a similar manner to the underivatized nucleoside triphosphates. Examples of such derivatives or analogs, by way of illustration and not limitation, are those which are biotinylated, amine modified, alkylated, and the like and also include phosphorothioate, phosphite, ring atom modified derivatives, and the like.

Nucleotide -- a base-sugar-phosphate combination that is the monomeric unit of nucleic acid polymers, i.e., DNA and RNA.

Modified nucleotide -- is the unit in a nucleic acid polymer that results from the incorporation of a modified nucleoside triphosphate during an amplification reaction and therefore becoming part of the nucleic acid polymer.

Nucleoside -- is a base-sugar combination or a nucleotide lacking a phosphate moiety.

Nucleotide polymerase -- a catalyst, usually an enzyme, for forming an extension of a polynucleotide along a DNA or RNA template where the extension is complementary thereto. The nucleotide polymerase is a template dependent polynucleotide polymerase and utilizes nucleoside triphosphates as building blocks for extending the 3'-end of a polynucleotide to provide a sequence complementary with the polynucleotide template. Usually, the catalysts are enzymes, such as DNA polymerases, for example, prokaryotic DNA polymerase (I, II, or III), T4 DNA polymerase, T7 DNA polymerase, Klenow fragment, reverse transcriptase, Vent DNA polymerase, Pfu DNA polymerase, Taq DNA polymerase, and the like, derived from any source such as cells, bacteria, such as E. coli, plants, animals, virus, thermophilic bacteria, and so forth. RNA polymerases include T7 RNA polymerase, AMV polymerase, Q-beta-replicase, and so forth.

Wholly or partially sequentially -- when the sample and various agents utilized in the present invention are combined other than concomitantly (simultaneously), one or more may be combined with one or more of the remaining agents to form a subcombination. Subcombination and remaining agents can then be combined and can be subjected to the present method.

Hybridization (hybridizing) and binding -- in the context of nucleotide sequences these terms are used interchangeably herein. The ability of two nucleotide sequences to hybridize with each other is based on the degree of complementarity of the two nucleotide sequences, which in turn is based on the fraction of matched complementary nucleotide pairs. The more nucleotides in a given sequence that are complementary to another sequence, the more stringent the conditions can be for hybridization and the more specific will be the binding of the two sequences. Increased stringency is achieved by elevating the temperature, increasing the ratio of cosolvents, lowering the salt concentration, and the like.

Homologous or substantially identical polynucleotides -- In general, two polynucleotide sequences that are identical or can each hybridize to the same polynucleotide sequence are homologous. The two sequences are homologous or

substantially identical where the sequences each have at least 90%, preferably 100%, of the same or analogous base sequence where thymine (T) and uracil (U) are considered the same. Thus, the ribonucleotides A, U, C and G are taken as analogous to the deoxynucleotides dA, dT, dC, and dG, respectively. Homologous sequences can comprise DNA, RNA or modified polynucleotides and may be homoduplexes, e.g., RNA:RNA and DNA:DNA or heteroduplexes, e.g., RNA:DNA.

Complementary -- Two sequences are complementary when the sequence of one can bind to the sequence of the other in an anti-parallel sense wherein the 3'-end of each sequence binds to the 5'-end of the other sequence and each A, T(U), G, and C of one sequence is then aligned with a T(U), A, C, and G, respectively, of the other sequence.

Non-contiguous -- two sequences within a first single polynucleotide sequence are non-contiguous when the 5' end of one sequence is joined to the 3' end of the other sequence by more than a bond, usually by a chain of one or more nucleotides that are not hybridized to a second single polynucleotide strand to which the two sequences of the first single strand are hybridized to form a duplex.

Contiguous -- two sequences within a first single polynucleotide strand are contiguous when the 5' end of one sequence is joined by a covalent bond directly to the 3' end of the other sequence without any intervening atoms or chain of nucleotides that are not hybridized to a second single polynucleotide strand to which the two sequences of the first single strand are hybridized to form a duplex.

Copy of a sequence -- a sequence that was copied from, and has the same base sequence as, a single stranded polynucleotide sequence as differentiated from a sequence that is copied from and has a complementary base sequence to the sequence of such single stranded polynucleotide.

Means for extending a primer -- a nucleotide polymerase or a single stranded template polynucleotide having a sequence other than at its 3'-end that can hybridize to at least the 3'-end of the primer or both. Means for extending a primer also includes nucleoside triphosphates or analogs thereof capable of acting as substrates for the enzyme and other materials and conditions required for enzyme

activity such as a divalent metal ion (usually magnesium), pH, ionic strength, organic solvent (such as formamide), and the like.

Member of a specific binding pair ("sbp member") -- one of two different molecules, having an area on the surface or in a cavity which specifically binds to and is thereby defined as complementary with a particular spatial and polar organization of the other molecule. The members of the specific binding pair are referred to as ligand and receptor (antiligand). These may be members of an immunological pair such as antigen-antibody, or may be operator-repressor, nuclease-nucleotide, biotin-avidin, hormones-hormone receptors, nucleic acid duplexes, IgG-protein A, DNA-DNA, DNA-RNA, and the like.

Ligand -- any compound for which a receptor naturally exists or can be prepared.

Receptor ("antiligand") -- any compound or composition capable of recognizing a particular spatial and polar organization of a molecule, e.g., epitopic or determinant site. Illustrative receptors include naturally occurring receptors, e.g., thyroxine binding globulin, antibodies, enzymes, Fab fragments, lectins, nucleic acids, repressors, protection enzymes, protein A, complement component C1q, DNA binding proteins or ligands and the like.

Small organic molecule -- a compound of molecular weight less than 1500, preferably 100 to 1000, more preferably 300 to 600 such as biotin, fluorescein, rhodamine and other dyes, tetracycline and other protein binding molecules, and haptens, etc. The small organic molecule can provide a means for attachment of a nucleotide sequence to a label or to a support.

Support or surface -- a porous or non-porous water insoluble material. The support can be hydrophilic or capable of being rendered hydrophilic and includes inorganic powders such as silica, magnesium sulfate, and alumina; natural polymeric materials, particularly cellulosic materials and materials derived from cellulose, such as fiber containing papers, e.g., filter paper, chromatographic paper, etc.; synthetic or modified naturally occurring polymers, such as nitrocellulose, cellulose acetate, poly (vinyl chloride), polyacrylamide, cross linked dextran, agarose, polyacrylate, polyethylene, polypropylene, poly(4-methylbutene),

polystyrene, polymethacrylate, poly(ethylene terephthalate), nylon, poly(vinyl butyrate), etc.; either used by themselves or in conjunction with other materials; glass available as Bioglass, ceramics, metals, and the like. Natural or synthetic assemblies such as liposomes, phospholipid vesicles, and cells can also be employed.

Binding of sbp members to a support or surface may be accomplished by well-known techniques, commonly available in the literature. See, for example, "Immobilized Enzymes," Ichiro Chibata, Halsted Press, New York (1978) and Cuatrecasas, J. Biol. Chem., 245:3059 (1970). The surface can have any one of a number of shapes, such as strip, rod, particle, including bead, and the like.

Label -- a member of a signal producing system. Usually the label is part of an oligonucleotide probe either being conjugated thereto or otherwise bound thereto or associated therewith and is capable of being detected directly or indirectly. The label may be part of the oligonucleotide primer. Labels include reporter molecules that can be detected directly by virtue of generating a signal, specific binding pair members that may be detected indirectly by subsequent binding to a cognate that contains a reporter molecule, oligonucleotide primers that can provide a template for amplification or ligation or a specific polynucleotide sequence or recognition sequence that can act as a ligand such as for a repressor protein, wherein in the latter two instances the oligonucleotide primer or repressor protein will have, or be capable of having, a reporter molecule. In general, any reporter molecule that is detectable can be used. The reporter molecule can be isotopic or nonisotopic, usually non-isotopic, and can be a catalyst, such as an enzyme, a polynucleotide coding for a catalyst, promoter, dye, fluorescent molecule, chemiluminescer, coenzyme, enzyme substrate, radioactive group, a small organic molecule, amplifiable polynucleotide sequence, a particle such as latex or carbon particle, metal sol, crystallite, liposome, cell, etc., which may or may not be further labeled with a dye, catalyst or other detectable group, and the like. The reporter group can be a fluorescent group such as fluorescein, a chemiluminescent group such as luminol, a terbium chelator such as N-(hydroxyethyl) ethylenediaminetriacetic acid that is capable of detection by delayed fluorescence, and the like.

The label is a member of a signal producing system and can generate a detectable signal either alone or together with other members of the signal producing system. As mentioned above, a reporter molecule can be bound directly to a nucleotide sequence or can become bound thereto by being bound to an sbp member complementary to an sbp member that is bound to a nucleotide sequence. Examples of particular labels or reporter molecules and their detection can be found in U.S. Patent Application Serial No. 07/555,323 filed July 19, 1990, the relevant disclosure of which is incorporated herein by reference.

Signal Producing System -- the signal producing system may have one or more components, at least one component being a label. The signal producing system generates a signal that relates to the presence or amount of target polynucleotide in a sample. The signal producing system includes all of the reagents required to produce a measurable signal. The labels and other reagents of the signal producing system must be stable at the temperatures used in an amplification of a target polynucleotide. Detection of the signal will depend upon the nature of the signal producing system utilized.

Preferably, the signal producing system is characterized in that the members of the system are chosen such that the binding of the second and third oligonucleotide probes to their respective duplexes to form termolecular complexes alters the signal generated by the signal producing system of the respective termolecular complexes. For example, in one approach the second oligonucleotide probe and the third oligonucleotide probe comprise a member of a different signal producing system and the signals are measured from the ternary complexes and a ratio of signals is determined. The first oligonucleotide probe may comprise a member of both signal producing systems. In another approach the first oligonucleotide probe in combination with each of the second oligonucleotide probe and the third oligonucleotide probe comprise members of different signal producing systems and the signals measured from the ternary complexes are used to determine a ratio of signals. In another approach the first oligonucleotide probe comprises a first member of a each of two signal producing systems and the second

oligonucleotide probe and the third oligonucleotide probe each respectively comprise a second member of each of the signal producing systems. Preferably, when the first member is brought into close proximity with the second members of the signal producing system, a signal is produced.

A number of signal producing systems in accordance with the above may be employed. The following discussion is by way of illustration and not limitation. In one such system the first member is a catalyst such as an enzyme and the second members are catalysts such as enzymes that are different from the first enzyme and from each other and the products of the reaction of the enzyme comprising the first member are the substrates for the other of the enzymes. By employing different second enzymes signals are produced that can be differentiated and used to determine a ratio of signals that is related to concentration of the target polynucleotide.

A list of enzymes is found in U.S. Patent No. 4,299,916 at column 30 to column 33. As mentioned above, of particular interest in the subject invention is the use of coupled catalysts, usually two or more enzymes, where the product of one enzyme serves as the substrate of the other enzyme. One of the enzymes is used as the label in the first oligonucleotide probe. Different second enzymes are used in the second and third oligonucleotide probes. The solute will be the substrate of any one of the enzymes, but preferably of an enzyme bound to the first oligonucleotide probe. The enzymatic reaction may involve modifying the solute to a product that is the substrate of another enzyme or production of compound that does not include a substantial portion of the solute, which serves as an enzyme substrate. The first situation may be illustrated by glucose-6-phosphate being catalytically hydrolyzed by alkaline phosphatase to glucose, wherein glucose is a substrate for glucose oxidase. The second situation may be illustrated by glucose being oxidized by glucose oxidase to provide hydrogen peroxide which would enzymatically react with the signal generator precursor to produce a signal generator. Coupled catalysts can also include an enzyme with a non-enzymatic catalyst. The enzyme can produce a reactant that undergoes a reaction catalyzed by the non-enzymatic catalyst or the non-enzymatic catalyst may produce a substrate (includes coenzymes) for the

enzyme. For example, Medola blue can catalyze the conversion of NAD and hydroquinones to NADH, which reacts with FMN oxidoreductase and bacterial luciferase in the presence of long chain aldehydes to produce light. Examples of particular catalytic systems that may be utilized in the present invention are found in U.S. Patent No. 4,299,916 at column 33, line 34, to column 38, line 32, the disclosure of which is incorporated herein by reference. For enzyme labels, additional members of the signal producing system include enzyme substrates and so forth. The product of the enzyme reaction is preferably a luminescent product or a fluorescent or non-fluorescent dye, any of which can be detected spectrophotometrically, or a product that can be detected by other spectrometric or electrometric means.

In another approach the first member of the signal producing system is a quencher and the second members are fluorescent compounds that emit at different wavelengths or with different decay rates. Fluorescers of interest generally emit light at a wavelength above 350 nm, usually above 400 nm and preferably above 450 nm. Desirably, the fluorescers have a high quantum efficiency, a large Stokes shift and are chemically stable under the conditions of their conjugation and use. The term fluorescer is intended to include substances that emit light upon activation and include fluorescent and phosphorescent substances, scintillators and chemiluminescent substances. In this approach the medium is irradiated with light and the fluorescence is determined. As will be appreciated, when the quencher is brought into close proximity to the fluorescent molecule by the formation of a termolecular complex, the fluorescence of the medium is decreased because of the absorption by the quencher of the light emitted by the fluorescer.

Fluorescers of interest fall into a variety of categories having certain primary functionalities. These primary functionalities include 1- and 2-aminonaphthalene, p,p-diaminostilbenes, pyrenes, quaternary phenanthridine salts, 9-aminoacridines, imines, anthracenes, oxacarboxyamine, merocyanine, 3-aminoequilenin, perylene, bis-benzoxazole, bis-p-oxazolyl benzene, 1,2-benzophenazine, retinal, bis-3-aminopyridinium salts, hellebrigenin, tetracycline, sterophenol, benzimidazolylphenylamine, 2-oxo-3-chromen, indole, xanthene, 7-

hydroxycoumarin, 4,5-benzimidazoles, phenoxazine, salicylate, strophanthidin, porphyrins, triarylmethanes, flavin and rare earth chelates oxides and salts. Exemplary fluorescers are enumerated in U.S. Patent No. 4,318,707 at columns 7 and 8, the disclosure of which is incorporated herein by reference.

A diverse number of energy absorbers or quenchers may be employed. The quencher must be able to quench the fluorescence of the fluorescer when brought into proximity with the fluorescer by virtue of the binding of the probes. Quenchers are chromophores having substantial absorption higher than 310 nm, normally higher than 350 nm, and preferably higher than about 400 nm. Generally, the quencher is a fluorescent compound or fluorescer but energy acceptors that have weak or no fluorescence are also useful. For example, one group of quenchers is the xanthene dyes, which include the fluoresceins derived from 3,6-dihydroxy-o-phenyl-xanthhydrol and rhodamines, derived form 3,6-diamino-9-phenylxanthhydrol. Another group of compounds are the naphthylamines such as, e.g., 1-anilino-8-naphthalene sulfonate, 1-dimethylaminonaphthyl-5-sulfonate and the like. Other examples of quenchers that may be employed are those fluorescers of interest mentioned above wherein one fluorescer can absorb the energy of another fluorescer and quench its fluorescence.

Energy acceptors that are non-fluorescent can include any of a wide variety of azo dyes, cyanine dyes, 4,5-dimethoxyfluorescein, formazans, indophenols and the like.

Another example of quenchers is energy absorbent or quenching particles. Examples of such particles are carbon particles, such as charcoal, lampblack, graphite, colloidal carbon and the like. Besides carbon particles metal sols may also find use, particularly of the noble metals, gold, silver, and platinum. Other metal derived particles may include metal sulfides, such as lead, silver or copper sulfides or metal oxides, such as iron or copper oxide.

As mentioned above, Heller (U.S. Patent No. 5,565,322) discloses donor and acceptor chromophores at column 9, line 37, to column 14, line 7, the disclosure of which is incorporated herein by reference. A further discussion of fluorescers and quenchers may also be found in U.S. Patent Nos. 4,261,968, 4,174,384, 4,199,983

and 3,996,345, the relevant disclosures of which are incorporated herein by reference.

In another approach the first member of the signal producing system is a sensitizer and the second members are chemiluminescent compounds that emit at different wavelengths or with different decay rates. Alternatively, the first member is a chemiluminescent compound and the second members are sensitizers that can be independently excited by different wavelengths of light. Examples of chemiluminescent compounds and sensitizers are set forth in U.S. Serial No. 07/923,069 filed July 31, 1992, the disclosure of which is incorporated herein by reference. Particularly preferred are photosensitizers and photoactivatable chemiluminescent compounds such as described in U.S. Patent No. 5,340,716 at column 19, line 30, to column 20, line 45, and column 22, line 58, to column 30, line 10, the disclosure of which is incorporated herein by reference. The sensitizers are those compounds that generate singlet oxygen usually by excitation with light. The sensitizer can be photoactivatable (e.g., dyes and aromatic compounds) or chemi-activated (e.g., enzymes and metal salts). Typical photosensitizers include acetone, benzophenone, 9-thioxanthone, eosin, 9,10-dibromoanthracene, methylene blue, metallo-porphyrins, such as hematoporphyrin, phthalocyanines, chlorophylls, rose bengal, buckminsterfullerene, etc., and derivatives thereof. Photoactivatable chemiluminescent compounds are substances that undergo a chemical reaction upon direct of sensitized excitation by light of upon reaction with singlet oxygen to form a metastable reaction product that is capable of decomposition with the simultaneous or subsequent emission of light, usually with the wavelength range of 250 to 1200 nm. Preferably, these compounds react with singlet oxygen to form dioxetanes or dioxetanones. The latter are usually electron rich olefins. Exemplary of such olefins are enol ethers, enamines, 9-alkylidene-N-alkylacridans, arylvinylethers, dioxenes, arylimidazoles, 9-alkylidene-xanthanes and lucigenin. Other compounds include luminol and other phthalhydrazides and chemiluminescent compounds that are protected form undergoing a reaction such as firefly luciferin, aquaphorin, luminol, etc.

Other components of the signal producing system may include substrates, enhancers, activators, chemiluminescent compounds, cofactors, inhibitors, scavengers, metal ions, specific binding substances required for binding of signal generating substances, and the like. Other components of the signal producing system may be coenzymes, substances that react with enzymic products, other enzymes and catalysts, and the like.

Termolecular complex -- a complex formed in accordance with the present methods upon the binding of the two oligonucleotide probes with a target polynucleotide or a reference polynucleotide. Such complex is termolecular in that it involves three molecules, namely, the two oligonucleotide probes and the single strand of such target polynucleotide or reference polynucleotide. In the present invention the termolecular complex is relatively stable under the isothermal conditions employed.

Sample -- any solution, synthetic or natural, containing a polynucleotide analyte or target polynucleotide including body fluids such as, for example, whole blood, blood fractions such as serum and plasma, synovial fluid, cerebro-spinal fluid, amniotic fluid, semen, cervical mucus, sputum, saliva, gingival fluid, urine, and the like. The amount of the sample depends on the nature of the sample and the analyte contained therein. For fluid samples such as whole blood, saliva, urine and the like the amount of the sample is usually about 1 to 1000 nanoliters, more usually, about 10 to 100 nanoliters. The sample can be pretreated and can be prepared in any convenient medium, which does not interfere with the reactions conducted as part of the present methods.

Ancillary Materials -- Various ancillary materials will frequently be employed in the methods carried out in accordance with the present invention. For example, buffers will normally be present in the assay medium, as well as stabilizers for the assay medium and the assay components. Frequently, in addition to these additives, proteins may be included, such as albumins, organic solvents such as formamide, quaternary ammonium salts, polycations such as dextran sulfate, surfactants, particularly non-ionic surfactants, binding enhancers, e.g., polyalkylene glycols, or the like.

In one embodiment the present invention is directed to a method for detecting the amount of a target polynucleotide in a sample. The target polynucleotide is single stranded either in the natural state or rendered single stranded as described above. A sample suspected of containing the target polynucleotide is combined in a medium with one or more of a reference polynucleotide and a predetermined amount of an oligonucleotide probe. The reference polynucleotide comprises a sequence that is common with a sequence of the target polynucleotide. The oligonucleotide probe has a sequence that hybridizes with the sequence that is common between the target polynucleotide and the reference polynucleotide. The combination is subjected to conditions for amplifying the target polynucleotide and the reference polynucleotide with substantially equal efficiency. The conditions permit formation of substantially non-dissociative complexes of the target polynucleotide and the reference polynucleotide, respectively, with the oligonucleotide probe. Furthermore, the predetermined amount of the oligonucleotide probe is less than the expected amount of the amplified target polynucleotide. The ratio of the amount of the complex of the target polynucleotide with the oligonucleotide probe to the amount of the complex of the reference polynucleotide with the oligonucleotide probe is determined. The ratio is related to the known amount of the reference polynucleotide to determine the amount of the target polynucleotide in the sample. Because the reference polynucleotide is amplified at substantially the same efficiency as the target polynucleotide and because the common oligonucleotide is employed in a predetermined amount less than the expected amount of the amplified target polynucleotide, the ratio of the amounts of the complexes at the end of amplification is directly related to the ratio of the amounts of the target and reference polynucleotides prior to amplification. Since the amount of the reference polynucleotide is known, the ratio of the complexes in conjunction with this known amount may be used to determine the amount of the target polynucleotide in the original sample.

The reference polynucleotide is specifically selected in light of the target polynucleotide. The reference polynucleotide may be naturally occurring or synthetic. As mentioned above, the reference polynucleotide has a sequence that is

common with the target polynucleotide and a sequence that is different from that of the target polynucleotide. The common sequence is usually about 5 to about 100 nucleotides, preferably, about 10 to about 70 nucleotides, more preferably, about 20 to about 50 nucleotides, in length. It is preferred also that the common sequence of the reference polynucleotide be greater than about 90% homologous, preferably, greater than about 95% homologous, most preferably, fully homologous, to the common sequence of the target polynucleotide. A sequence that contains a greater number of G, C nucleotides has a stronger degree of binding to its complementary sequence and, thus, a greater G, C content in the common sequence lessens the need for full homology between the two common sequences.

The number of reference polynucleotides employed in the present method is determined by the range of quantification desired. In other words the greater the range, the more reference polynucleotides can be used.

As mentioned above, the reference polynucleotide should be amplified with substantially the same efficiency as the target polynucleotide. By the term "efficiency" is meant that the parameters of amplification such as rate of reaction and the like are substantially similar.

By the term "substantially the same efficiency" is meant that amplification of the reference polynucleotide proceed with efficiency that is about 90% similar to the amplification of the target polynucleotide, more preferably, about 95% similar to the efficiency of amplification of the target polynucleotide. In general, the efficiency of the amplification of the reference polynucleotide should be as close to that for the amplification of the target polynucleotide as possible so that an accurate determination of the amount of the target polynucleotide in a sample may be made. Most preferably, 100% similarity between the efficiencies of amplification of the target polynucleotide and the reference polynucleotide is desired.

The length and nucleotide content of the reference polynucleotide are chosen to achieve this efficiency. In general, the length of the reference polynucleotide should be within about 100 to about 1000 nucleotides of he sequence of the target polynucleotide to be amplified. Likewise, the nucleotide content of the amplification product of the reference polynucleotide should be within about at least 90% of the

amplification product of the target polynucleotide. Other considerations are that the primers used for amplification of both target polynucleotide and reference oligonucleotide hybridize to the target polynucleotide with equal efficiency and that polymerase activity with both be substantially equally efficient.

The oligonucleotide probe has a sequence that is hybridizable with the common sequence. Usually, this sequence of the oligonucleotide probe is about 5 to about 100 nucleotides, preferably, about 10 to about 60 nucleotides, more preferably, about 20 to about 40 nucleotides, in length. The degree of complementarity of the sequence of the oligonucleotide probe and the common sequence of the target polynucleotide and the reference polynucleotide should be sufficient to achieve one of the objects of the present invention, namely, the ability to form complexes, respectively, with the target polynucleotide and the reference polynucleotide where the complexes are non-dissociative under the isothermal conditions. It is important that the complexes, once formed, not dissociate to any significant degree under the isothermal. By the term "not dissociate to any significant degree" is meant that the complex is stable to dissociation so that less than 1% of the oligonucleotide probe:target polynucleotide complex dissociates. Consequently, it is preferred that the sequence of the oligonucleotide probe be 70-100% complementary, preferably, 90-100% complementary, to the common sequence of the target polynucleotide and the reference polynucleotide. However, the degree of complementary depends on the relative lengths and nucleotide composition of the sequence of the oligonucleotide probes and the common sequences. It is preferred that the sequence of the oligonucleotide probe be greater than about 90% complementary, preferably, greater than about 95% complementary, most preferably, fully complementary, to the common sequence.

The sequence of the reference polynucleotide that is different from that in the target polynucleotide is usually about 10 to about 100 nucleotides; preferably, about 10 to about 50 nucleotides, more preferably, about 20 to about 40 nucleotides, in length. It is preferred also that the different sequence of the reference polynucleotide be greater than about 90% different, preferably, greater than about

95% different, most preferably, completely different, from the common sequence of the target polynucleotide.

In carrying out the present method, an aqueous medium is usually employed. In general, an aqueous medium is employed for the entire method in accordance with the present invention. Other polar cosolvents may also be employed, usually oxygenated organic solvents of from 1-6, more usually from 1-4, carbon atoms, including alcohols, ethers and the like. Usually these cosolvents, if used, are present in less than about 70 weight percent, more usually in less than about 30 weight percent.

The pH for the medium is usually in the range of about 4.5 to 9.5, more usually in the range of about 5.5 to 8.5, and preferably in the range of about 6 to 8. Various buffers may be used to achieve the desired pH and maintain the pH during the determination. Illustrative buffers include borate, phosphate, carbonate, Tris, barbital and the like. The particular buffer employed is not critical to this invention but in individual methods one buffer may be preferred over another. In general for amplification, the pH and temperature are chosen based on the particular method of amplification employed.

As mentioned above, the present method is conducted under isothermal conditions. In general, the temperature for the method is chosen so that the above complexes of the oligonucleotide probe with the target polynucleotide and reference polynucleotide, respectively, do not dissociate to any significant degree after formation. Thus, the temperature for the method is below the melting temperature of the above complexes. The particular temperature chosen depends on the salt concentration, pH, solvents used, the length of and composition of the common sequence and the sequence in the oligonucleotide probe and any other probes used in the method. For amplification by NASBA and 3SR, the reaction is usually conducted at isothermal temperature, which is usually about 38 to about 44°C, preferably about 41°C. For amplification by Q-beta replicase and SDA, the reaction is usually conducted at isothermal temperature, which is usually about 30 to about 45°C, preferably about 38°C.

The isothermal conditions include a uniform or constant temperature at which the present method is conducted. As mentioned above, the temperature for the method is chosen so that the above complexes of the oligonucleotide probe with the target polynucleotide and reference polynucleotide, respectively, do not dissociate to any significant degree after formation. By this is meant that the complexes do not dissociate more than about 5 %, preferably no more than about 1%, most preferably, no more than about 0.1%. Although constant temperature is preferred, some fluctuation of the temperature may occur and still achieve the benefits of the present invention. Accordingly, the term ""isothermal conditions" includes the use of a fluctuating temperature, particularly random or uncontrolled fluctuations in temperature. However, the fluctuation in temperature should not be more than about 5%, preferably no more than about 3%, most preferably, no more than about 1%. The fluctuation may be tolerated provided the temperature falls below, rather than above, the temperature of dissociation of the complexes formed in the present method. Generally, the particular isothermal temperature chosen is arrived at empirically by carrying out the present method at different temperatures and determining the optimum temperature resulting in the greatest amplification in accordance with the present invention. Computer models may also be used to select the appropriate temperature. A combination of the above may also be used. The temperature is generally determined by the amplification efficiency as determined by the method employed. This in turn is used to design the probes, since the melting temperature of the probe is a direct function of nucleotide content and length and can be calculated.

The amplification is conducted for a time sufficient so that all of the reference polynucleotide becomes part of a complex with the oligonucleotide probe. During this time period the target polynucleotide becomes complexed with free oligonucleotide probe, i.e., oligonucleotide probe not complexed with the reference polynucleotide. As mentioned above, the reference polynucleotide is employed in a predetermined amount. As a matter of convenience, it will usually be desirable to minimize the time period. The time period for amplification involving isothermal temperatures is usually about 10 to 90 minutes.

In the next step of the method, the ratio of the amount of the complex of the target polynucleotide with the oligonucleotide probe to the amount of the complex of the reference polynucleotide with the oligonucleotide probe is determined. To this end there must be some means for identifying the two complexes. In one approach a signal is produced by the complex of the target polynucleotide with the oligonucleotide probe. Such signal is different from a signal produced by the complex of the reference polynucleotide with the oligonucleotide probe. The ratio of the signals is related to the amount of the reference polynucleotide to determine the amount of the target polynucleotide in the sample. The signals produced may differ because a member of a signal producing system is used that recognizes only the complex of the oligonucleotide probe with the target polynucleotide and another member of a signal producing system is used that recognizes only the complex of the oligonucleotide probe with the reference polynucleotide. The corresponding complexes contain the other member of the respective signal producing system so that appropriate interactions between the respective members may occur to generate a signal. In another approach the respective complexes are identified and quantified directly.

In a preferred approach in accordance with the present invention, other probes are employed in addition to the oligonucleotide probe that has a sequence that hybridizes with the sequence that is common. In this embodiment a second oligonucleotide probe is employed that has a sequence that hybridizes only with a second sequence of the target polynucleotide. Also employed are one or more third oligonucleotide probes. Each of the third oligonucleotide probes has a sequence that hybridizes only with a respective second sequence of one of the reference polynucleotides.

The sequence of the second oligonucleotide probe that hybridizes only with a second sequence of the target polynucleotide is usually about 5 to about 80 nucleotides, preferably, about 10 to about 50 nucleotides, more preferably, about 15 to about 30 nucleotides, in length. The degree of complementarity of the sequence of the second oligonucleotide probe and the second sequence of the target polynucleotide should be sufficient to achieve one of the objects of the present

invention, namely, the ability to form complexes, respectively, with the target polynucleotide and the reference polynucleotide where the complexes are substantially non-dissociative or do not dissociate to any substantial degree under the isothermal conditions employed in the method. Consequently, it is preferred that the sequence of the second oligonucleotide probe be 70-100% complementary, preferably, 90-100% complementary, to the second sequence of the target polynucleotide. However, as mentioned above, the degree of complementary depends on the relative lengths and nucleotide composition of the sequence of the oligonucleotide probes and the common sequences. It is preferred that the sequence of the second oligonucleotide probe be greater than about 90% complementary, preferably, greater than about 95% complementary, most preferably, fully complementary, to the second sequence of the target polynucleotide.

The second sequence of the target polynucleotide may be separated from the common sequence of the target polynucleotide by about 1 to about 20 nucleotides, usually, about 1 to about 10 nucleotides. In a preferred embodiment the second sequence and the common sequence of the target polynucleotide are contiguous.

The sequence of the third oligonucleotide probe that hybridizes only with a second sequence of the reference polynucleotide is 5 to about 80 nucleotides, preferably, about 10 to about 50 nucleotides, more preferably, about 15 to about 30 nucleotides, in length. The degree of complementarity of the sequence of the third oligonucleotide probe and the second sequence of the reference polynucleotide should be sufficient to achieve one of the objects of the present invention, namely, the ability to form complexes, respectively, with the target polynucleotide and the reference polynucleotide where the complexes are substantially non-dissociative or do not dissociate to any substantial degree under the isothermal conditions employed in the method. Consequently, it is preferred that the sequence of the third oligonucleotide probe be 70-100% complementary, preferably, 90-100% complementary, to the second sequence of the reference polynucleotide. However, as mentioned above, the degree of complementary depends on the relative lengths and nucleotide composition of the sequence of the oligonucleotide probes and the

common sequences. It is preferred that the sequence of the third oligonucleotide probe be greater than about 90% complementary, preferably, greater than about 95% complementary, most preferably, fully complementary, to the second sequence of the reference polynucleotide.

The second sequence of the reference polynucleotide is separated from the common sequence of the reference polynucleotide by about 1 to about 20 nucleotides, usually, about 1 to about 10 nucleotides. In a preferred embodiment the second sequence and the common sequence of the target polynucleotide are contiguous. In a most preferred embodiment the number of nucleotides between the second sequence and the common sequence in the reference polynucleotide is the same as that in the target polynucleotide.

A different signal producing system is chosen for each of the respective second oligonucleotide probes and one or more third oligonucleotide probes corresponding to one or more of the reference polynucleotides. In this way the level of the different signals produced by the complexes formed are detected and a ratio is determined. The ratio of the respective signals is used to determine the ratio of the amount of the complex of the common (or first) oligonucleotide probe with the target polynucleotide to the amount of the complex of the common (or first) oligonucleotide probe with the reference oligonucleotide probe. To this end signal from the termolecular complex of the target polynucleotide and the first oligonucleotide probe and the second oligonucleotide probe is measured. Also measured are signals from the one or more termolecular complexes, each comprising a common (or first) oligonucleotide probe and a respective third oligonucleotide probe and its corresponding reference polynucleotide.

The ratio of the amount of the complex of the common (or first) oligonucleotide probe with the target polynucleotide to the amount of the complex of the common (or first) oligonucleotide probe with the reference oligonucleotide probe is then related to the known amount of the reference polynucleotide to determine the amount of the target polynucleotide in the sample. This is carried out for each of the reference polynucleotides employed.

The number of reference polynucleotides employed is dependent on a number of factors such as the desired range of quantification and the experimental accuracy of the determination of the differences between the target polynucleotide and the reference polynucleotide.

A third oligonucleotide probe is employed for each of the reference polynucleotides. Each of the third oligonucleotide probes has a sequence that hybridizes only with a respective second sequence of one of the reference polynucleotides.

An embodiment of the present method is depicted in Fig. 1 by way of illustration and not limitation. Target polynucleotide TP has a sequence CS that is common with sequence CS of reference polynucleotide RP. A common oligonucleotide probe CP has a sequence HCS that is hybridizable with sequence CS of TP and RP. Also included is second oligonucleotide probe PTS that comprises a sequence HTSS that is hybridizable with sequence TSS of TP. Also included is a third oligonucleotide probe PRS that has a sequence HRSS that is hybridizable with a sequence RSS in the RP. The above polynucleotides and oligonucleotide probes are combined in an appropriate medium and subjected to isothermal conditions. As the single stranded oligonucleotide amplification product is formed, complexes TRC and RPC are formed. As a result, substantially non-dissociative termolecular complexes TPC and RPC are formed. The reference polynucleotide RP and the first oligonucleotide probe CP are employed in predetermined amounts. Accordingly, the reaction medium contains unreacted second oligonucleotide probe PTS and unreacted third oligonucleotide probe PRS. Since the predetermined amount of the first oligonucleotide probe is less than the expected amount of amplified target polynucleotide, the reaction mixture also contains target polynucleotide that is hybridized only to the second oligonucleotide probe PTS. Depending on the amount of the amplified target polynucleotide formed and the amount of PTS, the reaction medium may also contain uncomplexed target polynucleotide or uncomplexed second oligonucleotide PTS. In the situation depicted in Fig. 1, an excess amount of amplified target polynucleotide is produced, so that the reaction medium contains uncomplexed target polynucleotide as well as

target polynucleotide complexed with the second oligonucleotide probe (TP/PTS). The predetermined amount of the first oligonucleotide probe may be about 30 to 160-fold, preferably, about 50-fold, less than the expected amount of the amplified target polynucleotide.

After an appropriate incubation period for complex formation, signals are determined from the reaction medium. One signal results from the interaction of label L1 on CP and label L2 on PTS. The labels can be part of the common oligonucleotide probe and the second oligonucleotide probe initially. On the other hand, the labels can be included as separate reagents either in the initial reaction mixture or added subsequent to the incubation period. In the latter situation the common probe CP and the second oligonucleotide probe PTS may each respectively be capable of having the respective labels, L1 and L2. In this regard an sbp member may be attached to CP and the corresponding sbp member attached to label L1. The two sbp members bind to each other and L1 becomes part of the common probe CP. This situation is discussed above. The two sbp members may be, for example, hapten-antibody, nucleic acid-nucleic acid, and the like. Label L2 may be incorporated with the second oligonucleotide probe PTS in a similar manner. This is also true for label L3 and the third oligonucleotide probe RPS.

Labels L1 and L2 are part of a signal producing system as discussed above. When the complex TPC is formed, L1 and L2 can interact to produce a signal that is then measured. The signal produced by the interaction of L1 and L2 is different from the signal produced by the interaction of L1 and L3. This difference permits the determination of the ratio of the signals from complex TPC and RPC, which is then related to the amount of the target polynucleotide in the sample.

The amount of the target polynucleotide to be amplified can be as low as one or two molecules in a sample but generally may vary from about 10 to 10^{10}, more usually from about 10^3 to 10^8 molecules in a sample preferably at least 10^{-21}M in the sample and may be 10^{-10} to 10^{-19}M, more usually 10^{-14} to 10^{-19}M.

As mentioned above, the first oligonucleotide probe or common oligonucleotide probe is present in a predetermined amount. The predetermined amount of the first oligonucleotide probe is generally less than the expected amount

of the amplified target polynucleotide. Usually, the amount is about 1 %, more usually, about 0.1%, and preferably, about 0.01% less than that of the expected amount of the amplified target polynucleotide. The amount of the first oligonucleotide probe is usually about 5 to about 100 nM, more usually, about 10 to about 50 nM.

The concentrations of the second and third oligonucleotide probes will usually be similar, preferably identical, and may be an excess amount. Usually, the concentration of the second and third oligonucleotide probes is consistent with the concentration of the other reagents in the reaction. For examjple, the concentration may be about 5 to about 100 M, more usually, about 10 to about 50 M.

As mentioned above, the reference polynucleotide (or reference polynucleotides) is present in a predetermined amount so that the ratio of the signals determined may be related to the predetermined amount of the reference polynucleotide to determine the amount of target polynucleotide that was originally in the sample. The predetermined amount of the reference polynucleotide is usually about 10 to about 10^6, more usually, 10^2 to about 10^5, molecules per assay.

The concentration of an enzyme that may be used in an amplification is usually determined empirically. Preferably, a concentration is used that is sufficient such that further increase in the concentration does not decrease the time for the amplification by over 5-fold, preferably 2-fold. The primary limiting factor generally is the cost of the reagent.

For amplification under isothermal conditions that may involve primer extension, the amount of the oligonucleotide primer(s) will be at least as great as the number of copies desired and will usually be 10^{-13} to 10^{-8} moles per sample, where the sample is 1-1,000 μL. Usually, the primer(s) are present in at least 10^{-9} M, preferably 10^{-7} M, and more preferably at least about 10^{-6} M. Preferably, the concentration of the oligonucleotide primer(s) is substantially in excess over, preferably at least 100 times greater than, more preferably, at least 1000 times greater than, the concentration of the target polynucleotide sequence.

For amplifications involving primer extension, the concentration of the nucleoside triphosphates in the medium can vary widely; preferably, these reagents

are present in an excess amount. The nucleoside triphosphates are usually present in 10^{-6} to 10^{-2}M, preferably 10^{-5} to 10^{-3}M.

As mentioned above one aspect of the present invention provides for quantitation of a target polynucleotide in a sample suspected of containing the target polynucleotide. The present invention has particular application to amplification reactions conducted at isothermal temperature. All of the necessary reagents for amplification and detection can be included in the reaction mixture prior to amplification and it is not necessary to open the reaction vessel after amplification and prior to detection. Thus, contamination is avoided. At the very least, complexes containing labels and any remaining members of the signal producing system can be added after amplification but without a separation step prior to detection. In some amplification procedures a preincubation may be desirable for the purpose of denaturation of double stranded structure and secondary intra strand structure. This preincubation is usually carried out at a temperature of about 50°C to about 90°C , preferably, about 60°C to about 70°C , for a period of about 1 to about 30 minutes, preferably, about 5 to about 10 minutes. In this regard all of the reagents may be combined except for the enzyme necessary for a particular amplification. Once the preincubation is carried out, the enzyme is added and the amplification is allowed to proceed.

It should be noted that the length of the various oligonucleotide probes and the relevant sequences contained therein may be greater or less than that indicated above. The length depends on a number of factors such as the possibility of strand invasion, whether the nucleotides are natural or modified, the temperature, the pH, the salt concentration of the medium, and so forth. In general, the length of the above sequences should be sufficient to ensure substantially non-dissociative complexes following hybridization to the amplification product at the amplification conditions of temperature, buffer and oligonucleotide probe concentration.

Detection of the signal, and the conditions therefor, depend upon the nature of the signal producing system utilized. Such conditions are well known in the art. If the reporter molecule is an enzyme, additional members of the signal producing system would include enzyme substrates and so forth. The product of the enzyme

reaction is preferably a luminescent product, or a fluorescent or non-fluorescent dye, any of which can be detected spectrophotometrically, or a product that can be detected by other spectrometric or electrometric means. If the reporter molecule is a fluorescent molecule, the medium can be irradiated and the fluorescence determined. Where the label is a radioactive group, the first probe can be separated from the duplex of the first and second probes and one of the fractions can be counted to determine the radioactive count.

The association of the labels within the termolecular complex may also be determined by using labels that provide a signal only if the labels become part of, or dissociate from, the complex. The binding of the single stranded target polynucleotide, if present, to the second probe causes displacement of the first probe from the second probe and thereby alters a signal generated by the signal producing system. This approach is particularly attractive when it is desired to conduct the present invention in a homogeneous manner. Such systems include enzyme channeling immunoassay, fluorescence energy transfer immunoassay, electrochemiluminescence assay, induced luminescence assay, latex agglutination and the like.

In one aspect of the present invention detection of the complex is accomplished by employing at least one suspendable particle as a support, which may be bound directly to a nucleic acid strand or may be bound to an sbp member that is complementary to an sbp member attached to a nucleic acid strand, either first or second oligonucleotide probe. Such a particle serves as a means of segregating the bound target polynucleotide sequence from the bulk solution. A second label, which is attached to the other of the first or second oligonucleotide probes, becomes part of the termolecular complex. Typical labels that may be used in this particular embodiment are fluorescent labels, particles containing a sensitizer and a chemiluminescent olefin (see U.S. Serial No. 07/923,069 filed July 31, 1992, the disclosure of which is incorporated herein by reference), chemiluminescent and electroluminescent labels.

Preferably, the particle itself can serve as part of a signal producing system that can function without separation or segregation. The second label is also part of

the signal producing system and can produce a signal in concert with the particle to provide a homogeneous assay detection method. A variety of combinations of labels can be used for this purpose. When all the reagents are added at the beginning of the reaction, the labels are limited to those that are stable to the temperatures used for amplification.

The particles, for example, may be simple latex particles or may be particles comprising a sensitizer, chemiluminescer, fluorescer, dye, and the like. Typical particle/reporter molecule pairs include a dye crystallite and a fluorescent label where binding causes fluorescence quenching or a tritiated reporter molecule and a particle containing a scintillator. Typical reporter molecule pairs include a fluorescent energy donor and a fluorescent acceptor dye. Typical particle pairs include (1) two latex particles, the association of which is detected by light scattering or turbidimetry, (2) one particle capable of absorbing light and a second label particle which fluoresces upon accepting energy from the first, and (3) one particle incorporating a sensitizer and a second particle incorporating a chemiluminescer as described for the induced luminescence immunoassay referred to in U.S. Serial No. 07/704,569, filed May 22, 1991, entitled "Assay Method Utilizing Induced Luminescence", which disclosure is incorporated herein by reference.

Briefly, detection of the termolecular complex using the induced luminescence assay as applied in the present invention involves employing a photosensitizer as part of one label and a chemiluminescent compound as part of the other label. If the complex is present the photosensitizer and the chemiluminescent compound come into close proximity. The photosensitizer generates singlet oxygen and activates the chemiluminescent compound when the two labels are in close proximity. The activated chemiluminescent compound subsequently produces light. The amount of light produced is related to the amount of the complex formed.

If the amplification uses a DNA polymerase, preferably, the oligonucleotide probes are blocked at the 3'-end to avoid any potential interference with and during amplification. Blocking may be accomplished, for example, by employing a group that cannot undergo chain extension, such as, for example, an unnatural group such as a 3'-phosphate, a 3'-terminal dideoxy, an abasic ribophosphate, a polymer or

surface, or other means for inhibiting chain extension. Alternatively, a polynucleotide that does not hybridize to the amplicon is attached to the 3'-end. Such an end group can be introduced at the 3' end during solid phase synthesis or a group can be introduced that can subsequently be modified. For example, in order to introduce dextran at the 3'-end a ribonucleotide can be introduced at the 3'-end and then oxidized with periodate followed by reductive amination of the resulting dialdehyde with borohydride and aminodextran. The details for carrying out the above modifications are well known in the art and will not be repeated here.

An embodiment of the above is depicted in Fig. 2 by way of illustration and not limitation. Target polynucleotide TP has a sequence CS that is common with sequence CS of reference polynucleotide RP1 and reference polynucleotide RP2. In addition to the common sequence, RP1 has a sequence RSS1 that is different from sequence TSS of the target polynucleotide. In addition to the common sequence, RP2 has a sequence RSS2 that is different from sequence TSS of the target polynucleotide and from the sequence RSS1 of RP1. A common oligonucleotide probe CP has a sequence HCS that is hybridizable with sequence CS of TP and RP1 and RP2. CP also comprises sequence LCS, which is part of one of the signal producing systems employed in the method depicted in Fig. 2. LCS is hybridizable with sequence HLCS in oligonucleotide OL5, which also comprises a bead having a sensitizer S associated therewith. Also included is second oligonucleotide probe PTS that comprises a sequence HTSS that is hybridizable with sequence TSS of TP. PTS also comprises a sequence LTS that is part of one of the signal producing systems employed in the method depicted in Fig. 2. LTS is hybridizable with sequence HLTS in oligonucleotide OL6, which also comprises a bead having a chemiluminescer, CL1, associated therewith. Also included is a third oligonucleotide probe PRS1 that has a sequence HRSS1 that is hybridizable with a sequence RSS1 in the RP1. PRS1 also comprises sequence LRS1, which is part of another of the signal producing systems employed in the method depicted in Fig. 2. LRS1 is hybridizable with sequence HLRS1 in oligonucleotide OL7, which also comprises a bead having a chemiluminescer, CL2, associated therewith. Also included is a fourth oligonucleotide probe PRS2 that has

a sequence HRSS2 that is hybridizable with a sequence RSS2 in the RP2. PRS2 also comprises sequence LRS2, which is part of another of the signal producing systems employed in the method depicted in Fig. 2. LRS2 is hybridizable with sequence HLRS2 in oligonucleotide OL8, which also comprises a bead having a chemiluminescer, CL3, associated therewith.

The above polynucleotides and oligonucleotide probes are combined in an appropriate medium and subjected to isothermal conditions. As a result, substantially non-dissociative termolecular complexes TPC, RPC1 and RPC2 are formed. Complex TPC includes OL5 and OL6, by virtue of the presence of sequences that are hybridizable to sequences in the respective oligonucleotide probes CP and PTS. Complex RPC1 includes OL5 and OL7, by virtue of the presence of sequences that are hybridizable to sequences in the respective oligonucleotide probes CP and PRS1. Complex RPC2 includes OL5 and OL8, by virtue of the presence of sequences that are hybridizable to sequences in the respective oligonucleotide probes CP and PRS2.

As mentioned above, sensitizer S forms part of each of three signal producing systems corresponding, respectively, to chemiluminescers CL1, CL2 and CL3. As a result of the formation of the termolecular complexes TPC, RPC1 and RPC2, the sensitizer is brought into close proximity with CL1, CL2 and CL3, respectively. The chemiluminescers differ by virtue of emitting light of different wavelength, thus permitting discrimination of signals produced by the respective termolecular complexes.

The reference polynucleotides RP1 and RP2 and the first oligonucleotide probe CP are employed in predetermined amounts. Accordingly, the reaction medium contains unreacted second oligonucleotide probe PTS and unreacted third and fourth oligonucleotide probes, PRS1 and PRS2, respectively. Since the predetermined amount of the first oligonucleotide probe CP is less than the expected amount of amplified target polynucleotide, the reaction mixture also contains target polynucleotide that is hybridized only to the second oligonucleotide probe PTS. Depending on the amount of the amplified target polynucleotide formed and the amount of PTS, the reaction medium may also contain uncomplexed target

polynucleotide or uncomplexed second oligonucleotide PTS as well as excess amounts of OL5, OL6, OL7 and OL8. In the situation depicted in Fig. 2, an excess amount of amplified target polynucleotide is produced, so that the reaction medium contains uncomplexed target polynucleotide as well as target polynucleotide complexed with the second oligonucleotide probe (TP/PTS).

After an appropriate incubation period for complex formation, signals are determined from the reaction medium. The difference is signals from the respective termolecular complexes permits the determination of the ratio of the signals from complex TPC and RPC1 and RPC2, which is then related to the amount of the target polynucleotide in the sample.

The present method may also be used to monitor the progress of an amplification conducted under isothermal conditions. A signal is measured at the outset of the amplification. For example, the signal may be produced as a result of the induced luminescence resulting from the interaction of the sensitizer with oxygen to form singlet oxygen, which reacts with the chemiluminescent compound to produce a product that spontaneously produces luminescence. The level of signal is measured periodically during the amplification. During the amplification the signal measured initially is changed as a result of the increase in concentration of the complex of the target polynucleotide with the first oligonucleotide probe and the second oligonucleotide probe. Since more sensitizer is in close proximity of the chemiluminescent compound, signal increases. This increase in signal can be monitored as an indication of the progress of the amplification of the target polynucleotide. When the increase in signal is no longer observed, the amplification is complete.

As mentioned above, the nature of the signal from each of the reference polynucleotides and that for the target polynucleotide must be separately detectable. This entails different signal producing system members for each of the above. For example, in the example depicted in Fig. 2, each of the respective chemiluminescent particles corresponding to the different reference polynucleotides and the target polynucleotide may emit at different wavelengths of light as a result of the sensitization by the sensitizer.

As mentioned above, a particular advantage of the present invention is that excess target polynucleotide does not interfere with the determination. This results from the fact that the first oligonucleotide probe is present in an amount that is less than the expected amount of amplified target polynucleotide.

Another particular example of the present method as applied to the detection of an RNA target polynucleotide in a sample suspected of containing the target polynucleotide is next described by way of illustration and not limitation. The amplification method chosen is NASBA. Accordingly, a measured amount of the sample suspected of containing the RNA target polynucleotide is combined with a predetermined amount of reference polynucleotides H and J. The predetermined amount of H is 100 copies and the predetermined amount of J is 1000 copies. The RNA is isolated from the sample and combined in a reaction medium with all of the reagents for conducting a NASBA amplification, namely, RNA polymerase, a DNA primer with a promoter sequence, a second DNA primer, T7 polymerase reverse transcriptase, RNAse-H, nucleotide triphosphates and deoxynucleoside triphosphates. These reagents are provided in amounts to achieve a concentration of the RNA target polynucleotide, if present, in the reaction medium after NASBA amplification of about 10^{-7} to 10^{-5} M. The reaction medium further contains a first oligonucleotide probe K, which is a universal probe and has the energy donor 7-dimethylaminocoumarin-3-carboxylic acid attached to a terminal nucleotide thereof at the 5' end. The medium also includes a second oligonucleotide probe K that has a sequence complementary to a sequence in the target polynucleotide, a third oligonucleotide probe L that has a sequence complementary to a sequence in reference polynucleotide M, and a fourth oligonucleotide probe N that has a sequence complementary to a sequence in reference polynucleotide H. Universal probe K has a sequence that is complementary to a sequence in the target polynucleotide as well as oligonucleotides L, M and N. Oligonucleotide probe L has fluorescein attached to a terminal nucleotide at the 3' end thereof, oligonucleotide probe M has the fluorophore, phycoerythrin, attached to a terminal nucleotide at the 3' end thereof and oligonucleotide N has phycoerythrin conjugated to allophycocyanine attached to a terminal nucleotide at the 3' end thereof. The

reaction medium is subjected to conditions for carrying out the NASBA amplification, namely, a temperature of 41°C for a time of 90 minutes.

As indicated above, the detection and quantitation of the target polynucleotide can be carried out at the end of the amplification or during the amplification. In the latter situation, the amplification is conducted and a determination of the signal is determined at a variety of points in time. This is achieved by irradiating the reaction medium at a wavelength of 415 nm and determining the ratios of the light emitted at three different wavelengths corresponding to the emission maxima of the three fluorophors fluorescein 520 nm, phycoerythrin 576 nm and phycoerythrin/allophycocyanine 660 nm. The target polynucleotide, if present, hybridizes with K and with L to give a termolecular complex TKL. Likewise, reference polynucleotide H hybridizes with K and M to give termolecular complex HKM, and reference polynucleotide J hybridizes with K and N to give termolecular complex JKN.

Briefly, detection using the induced luminescence assay as applied in the present invention involves an energy donor as part of one label and a fluorescent energy acceptor as part of the other label in a particular signal producing system. If the target polynucleotide is present the donor and the acceptor come into close proximity and the fluorescent acceptor is activated and emits light in relation to the amount of the complex formed.

As a matter of convenience, predetermined amounts of reagents employed in the present invention can be provided in a kit in packaged combination. The kit comprises in packaged combination one or more reagents for conducting an amplification and detection of the amplified product as well as the oligonucleotide probes and the reference polynucleotides. An example of a kit in accordance with the present invention is a kit comprising reagents for conducting an amplification of the target polynucleotide together with one or more reference polynucleotides, first and second oligonucleotide probes and one or more third oligonucleotide probes. Each reference polynucleotide is present in a known predetermined amount and each comprises a sequence that is common with a sequence of the target polynucleotide. The first oligonucleotide probe is employed in a predetermined

amount and hybridizes with the sequence that is common between the target and the reference polynucleotides. The second oligonucleotide probe has a sequence that hybridizes only with the second sequence of the target polynucleotide . The third oligonucleotide probe has a sequence that hybridizes only with a respective second sequence of the reference polynucleotide.

Another example of a kit in accordance with the present invention is a kit for RNA amplification. Such a kit comprises reference RNA's, a promoter, an enzyme, one or more reference polynucleotides, first and second oligonucleotide probes and one or more third oligonucleotide probes, each corresponding to a reference polynucleotide. One of the first and second oligonucleotide probes is labeled with a sensitizer and the other thereof is labeled with a chemiluminescer. One of the first and third oligonucleotide probes is labeled with a sensitizer and the other thereof is labeled with a chemiluminescer. Usually, the first oligonucleotide probe is labeled with a sensitizer and the second and third oligonucleotide probes are labeled with a different chemiluminescer. In this way the signals from the respective labels are differentially detectable.

The kit can further include any additional members of a signal producing system and also various buffered media, some of which may contain one or more of the above reagents. The kits above can further include in the packaged combination reagents for conducting an amplification of the target polynucleotide. For example, in the case of NASBA, the kit may include a first DNA primer with a 5' tail comprising a promoter, a second DNA primer, reverse transcriptase, RNAse-H, T7 RNA polymerase, NTP's and dNTP's.

The relative amounts of the various reagents in the kits can be varied widely to provide for concentrations of the reagents necessary to achieve the objects of the present invention. Under appropriate circumstances one or more of the reagents in the kit can be provided as a dry powder, usually lyophilized, including excipients, which on dissolution will provide for a reagent solution having the appropriate concentrations for performing a method or assay in accordance with the present invention. Each reagent can be packaged in separate containers or some reagents can be combined in one container where cross-reactivity and shelf life permit. The

kits may also include a written description of a method in accordance with the present invention as described above.

EXAMPLES

The invention is demonstrated further by the following illustrative examples. Parts and percentages recited herein are by weight unless otherwise specified. Temperatures are in degrees centigrade (°C).

Melting points were determined on a Hoover capillary apparatus and are uncorrected. 'HNMR spectra were recorded on a Brucker WP-250 MHz or Brucker WP-300 MHz NMR spectrometer. Chemical shifts were reported in parts per million (δ 0.0). NMR multiplicities are recorded by use of the following abbreviations: s, singlet; d, doublet; t, triplet; m, multiplet; Hz, hertz. Infrared spectra were recorded on a Perkin-Elmer 297IR spectrometer. Desorption chemical ionization (C.I.) and electron ionization (E.I.) were done on a Varian-MAT 311A, double focusing high-resolution mass spectrometer. A Finnigan TSQ-70 or MAT-8230 was used for fast atom bombardment mass spectra (FAB/LSIMS). UV-visible absorption spectra were done on a HP 8452A diode array spectrophotometer. Fluorescence and chemiluminescence measurements were done on a Spex fluorolog spectrophotometer or a Perkin Elmer 650-40 spectrophotometer. Chemiluminescence measurements were also performed on an in-house chemiluminometer (Oriel box).

Toluene was distilled from sodium over argon. Unless mentioned otherwise, all solvents were used without purification, and most reactions were carried out under argon. Silica gel used for flash chromatography was 230-400mesh ASTM, purchased from Scientific Products while preparative plates (1000μ) and analytical plates were purchased from Analtech.

C-28 thioxene, substituted N-phenyl oxazine and thioxene attached to 9,10-bis(phenylethynyl) anthracene (BPEA) were prepared as described below. 2-Chloro 9,10- bis(phenylethynyl) anthracene (1-Cl-BPEA) and rubrene (5,6,11,12-tetraphenyl naphthacene) were purchased from Aldrich Chemical Co. Rubrene was recrystallized from methylene chloride and stored at 4°C in a brown bottle prior to

use. Silicon phthalocyanine was prepared as described below and phthalocyanine tetrasulfonates was obtained from Ultra Diagnostics, Inc. Carboxylate-modified polystyrene (latex) particles were purchased from Seradyn, Inc. The particles were 203 ± 4.0 nM. The carboxyl parking area was 49.5 angstroms squared (0.09 milliequivalents/g). Solids were 10% (100 mg/ml).

2-ethoxyethanol was from Aldrich Chemical Co. and was redistilled under vacuum. Sodium hydroxide was 0.1 N. Isopropanol was from Aldrich Chemical Co..

Unless otherwise indicated, oligonucleotides used in the following examples were prepared by synthesis using an automated synthesizer and were purified by gel electrophoresis or HPLC.

The following abbreviations have the meanings set forth below:

Tris HCl - Tris(hydroxymethyl)aminomethane-HCl (a 10X solution) from BioWhittaker, Walkersville, MD.

DTT - dithiothreitol from Sigma Chemical Company, St. Louis, MO.

HPLC - high performance liquid chromatography.

DPP - 4,7-diphenylphenanthroline from Aldrich Chemical Company, Milwaukee WI.

BSA - bovine serum albumin from Sigma Chemical Company, St. Louis MO

ELISA - enzyme linked immunosorbent assay as described in "Enzyme-Immunoassay," Edward T. Maggio, CRC Press, Inc., Boca Raton, Florida (1980)

bp - base pairs

ddc - dideoxycytidine

g - grams

mmol - millimolar

DMF - dimethyl formamide

THF - tetrahydrofuran

LSIMS - fast ion bombardment mass spectroscopy

NMR - nuclear magnetic resonance spectroscopy

TMSCl - tetramethylsilylchloride

EDAC - 1-ethyl-3-(3-dimethylaminopropyl)-carbodiimide hydrochloride.

MES - 2-(N-morpholino)ethane sulfonic acid.

SPDP - N-succinimidyl 3-(2-pyridylthio)-propionate.

Sulfo-SMCC – N-sulfosuccinimidyl-4-(N-maleimidomethyl)cyclohexane-1-carboxylate.

TCEP - tris-carboxyethyl phosphine.

PREPARATION OF REAGENTS

C-28 thioxene:

To a solution of 4-bromoaniline (30g, 174mmol) in dry DMF (200mL) was added 1-bromotetradecane (89.3mL, 366mmol) and N,N-diisopropylethylamine (62.2mL, 357mmol). The reaction solution was heated at 90°C for 16 hr under argon before being cooled to room temperature. To this reaction solution was again added 1-bromotetradecane (45mL, 184mmol) and N,N-diisopropylethylamine (31mL, 178mmol) and the reaction mixture was heated at 90°C for another 15 hr. After cooling, the reaction solution was concentrated in vacuo and the residue was diluted with CH_2Cl_2 (400mL). The CH_2Cl_2 solution was washed with 1N aqueous NaOH (2x), H_2O, and brine, was dried over Na_2SO_4 and was concentrated in vacuo to yield a dark brown oil (about 110g). Preparative column chromatography on silica gel by a Waters 500 Prep LC system eluting with hexane afforded a yellow oil that contained mainly the product (4-bromo-N,N-di-($C_{14}H_{29}$)-aniline) along with a minor component 1-bromotetradecane. The latter compound was removed from the mixture by vacuum distillation (bp 105-110°C, 0.6mm) to leave 50.2g (51%) of the product as a brown oil. To a mixture of magnesium turnings (9.60g, 395mmol) in dry THF (30mL) under argon was added dropwise a solution of the above substituted aniline product (44.7g, 79mmol) in THF (250mL). A few crystals of iodine were added to initiate the formation of the Grignard reagent. When the reaction mixture became warm and began to reflux, the addition rate was regulated to maintain a gentle reflux. After addition was complete, the mixture was heated at reflux for an additional hour. The cooled supernatant solution was transferred via cannula to an addition funnel and added dropwise (over 2.5 hr) to a solution of phenylglyoxal (11.7g, 87mmol) in THF (300mL) at -30°C under argon. The reaction mixture was gradually warmed to 0°C over 1 hr and stirred for another 30 min. The resulting

mixture was poured into a mixture of ice water (800mL) and ethyl acetate (250mL). The organic phase was separated and the aqueous phase was extracted with ethyl acetate (3x). The combined organic phases were washed with H_2O (2x), brine and were dried over $MgSO_4$. Evaporation of the solvent gave 48.8g of the crude product as a dark green oily liquid. Flash column chromatography of this liquid (gradient elution with hexane, 1.5:98.5, 3:97, 5:95 ethyl acetate:hexane) afforded 24.7g (50%) of the benzoin product (LSIMS ($C_{42}H_{69}NO_2$): $[M-H]^+$ 618.6, ^1H NMR (250 MHz, $CDCl_3$) was consistent with the expected benzoin product. To a solution of the benzoin product from above (24.7g, 40mmol) in dry toluene (500mL) was added sequentially 2-mercaptoethanol (25g, 320mmol) and TMSCl (100mL, 788mmol). The reaction solution was heated at reflux for 23 hr under argon before being cooled to room temperature. To this was added additional TMSCl (50mL, 394mmol); and the reaction solution was heated at reflux for another 3 hr. The resulting solution was cooled, was made basic with cold 2.5N aqueous NaOH and was extracted with CH_2Cl_2 (3x). The combined organic layers were washed with saturated aqueous $NaHCO_3$ (2x) and brine, was dried over Na_2SO_4 and was concentrated in vacuo to give a brown oily liquid. Preparative column chromatography on silica gel by using a Waters 500 Prep LC system (gradient elution with hexane, 1:99, 2:98 ethyl acetate:hexane) provided 15.5g (60%) of the C-28 thioxene as a yellow oil (LSIMS ($C_{44}H_{71}NOS$): $[M-H]^+$ 661.6, ^1H NMR (250 MHz, $CDCl_3$) was consistent with the expected C-28 thioxene product 2-(4-(N,N-di-($C_{14}H_{29}$)-anilino)-3-phenyl thioxene.

Silicon tetra-t-butyl phthalocyanine:

Sodium metal, freshly cut (5.0g, 208mmol), was added to 300mL of anhydrous methanol in a two-liter, 3-necked flask equipped with a magnetic stirrer, reflux condenser, a drying tube and a gas bubbler. After the sodium was completely dissolved, 4-t-butyl-1,2-dicyanobenzene (38.64g, 210mmol, from TCI Chemicals, Portland OR) was added using a funnel. The mixture became clear and the temperature increased to about 50°C. At this point a continuous stream of anhydrous ammonia gas was introduced through the glass bubbler into the reaction mixture for 1 hr. The reaction mixture was then heated under reflux for 4 hr. while

the stream of ammonia gas continued during the course of the reaction, as solid started to precipitate. The resulting suspension was evaporated to dryness (house vacuum) and the residue was suspended in water (400mL) and filtered. The solid was dried (60°C, house vacuum, P_2O_5). The yield of the product (1,3-diiminoisoindoline, 42.2g) was almost quantitative. This material was used for the next step without further purification. To a one-liter, three-necked flask equipped with a condenser and a drying tube was added the above product (18g, 89mmol) and quinoline (200mL, Aldrich Chemical Company, St. Louis MO). Silicon tetrachloride (11mL, 95mmol, Aldrich Chemical Company) was added with a syringe to the stirred solution over a period of 10 minutes. After the addition was completed, the reaction mixture was heated to 180-185°C in an oil bath for 1 hr. The reaction was allowed to cool to room temperature and concentrated HCl was carefully added to acidify the reaction mixture (pH 5-6). The dark brown reaction mixture was cooled and filtered. The solid was washed with 100mL of water and dried (house vacuum, 60°C, P_2O_5). The solid material was placed in a 1-liter, round bottom flask and concentrated sulfuric acid (500mL) was added with stirring. The mixture was stirred for 4 hr. at 60°C and was then carefully diluted with crushed ice (2000g). The resulting mixture was filtered and the solid was washed with 100mL of water and dried. The dark blue solid was transferred to a 1-liter, round bottom flask, concentrated ammonia (500mL) was added, and the mixture was heated and stirred under reflux for 2 hr., was cooled to room temperature and was filtered. The solid was washed with 50mL of water and dried under vacuum (house vacuum, 60°C, P_2O_5) to give 12g of product silicon tetra-t-butyl phthalocyanine as a dark blue solid. 3-picoline (12g, from Aldrich Chemical Company), tri-n-butyl amine (anhydrous, 40mL) and tri-n-hexyl chlorosilane (11.5g) were added to 12g of the above product in a one-liter, three-necked flask, equipped with a magnetic stirrer and a reflux condenser. The mixture was heated under reflux for 1.5 hr. and then cooled to room temperature. The picoline was distilled off under high vacuum (oil pump at about 1mm Hg) to dryness. The residue was dissolved in CH_2Cl_2 and purified using a silica gel column (hexane) to give 10g of pure product di-(tri-n-hexylsilyl)-silicon tetra-

t-butyl phthalocyanine as a dark blue solid. (LSIMS: $[M-H]^+$ 1364.2, absorption spectra: methanol: 674nm (ε 180,000): toluene 678nm, ^1H NMR (250 MHz, $CDCl_3$): δ: -2.4(m,12H), -1.3(m, 12H), 0.2-0.9 (m, 54H), 1.8(s, 36H), 8.3(d, 4H) and 9.6 (m, 8H) was consistent with the above expected product.

Hydroxypropylaminodextran: ($1NH_2$/ 7 glucose) was prepared by dissolving Dextran T-500 (Pharmacia, Uppsala, Sweden) (50g) in 150 mL of H_2O in a 3-neck round-bottom flask equipped with mechanical stirrer and dropping funnel. To the above solution was added 18.8g of $Zn(BF_4)_2$ and the temperature was brought to 87°C with a hot water bath. Epichlorohydrin (350mL) was added dropwise with stirring over about 30 min while the temperature was maintained at 87-88°C. The mixture was stirred for 4 hr while the temperature was maintained between 80°C and 95°C, then the mixture was cooled to room temperature. Chlorodextran product was precipitated by pouring slowly into 3L of methanol with vigorous stirring, recovered by filtration and dried overnight in a vacuum oven.

The chlorodextran product was dissolved in 200mL of water and added to 2L of concentrated aqueous ammonia (36%). This solution was stirred for 4 days at room temperature, then concentrated to about 190mL on a rotary evaporator. The concentrate was divided into two equal batches, and each batch was precipitated by pouring slowly into 2L of rapidly stirring methanol. The final product was recovered by filtration and dried under vacuum.

Hydroxypropylaminodextran ($1NH_2$/ 7 glucose), prepared above, was dissolved in 50mM MOPS, pH 7.2, at 12.5 mg/mL. The solution was stirred for 8 hr at room temperature, stored under refrigeration and centrifuged for 45 min at 15,000 rpm in a Sorvall RC-5B centrifuge immediately before use to remove a trace of solid material. To 10mL of this solution was added 23.1mg of Sulfo-SMCC in 1mL of water. This mixture was incubated for 1 hr at room temperature and used without further purification.

N-phenyl oxazine (NPhe):

The N-phenyl oxazine was prepared by a procedure similar to that described in U.S. Patent No. 5,578,498 (Singh, *et al.*) for the preparation of compound 16. The relevant disclosure of the above patent is incorporated herein by reference.

Particle Dyeing

Preparation of N-phenyl oxazine (N- Phe) particles

Microgon setup: A Microgon was assembled as described in the MiniKros lab system manual page 8 - 10. A 0.1 micron module was used operated between about 5 - 10 psi. The Microgon apparatus was with ethanol (200 proof).

Setup of apparatus:

1. An oil bath was heated to 95° C±1.O°C. A three-necked 1 liter roundbottom flask (rbf) equipped with a digital Cafeama mechanical stirrer from the middle neck was immersed into the oil bath.
2. Addition of particles: 200 ml ± 5.0 ml of latex particles were added to the rbf by means of a measuring cylinder, which was washed with 2x30 ml of ethoxyethanol (ee) and the contents were transferred to the rbf. To the flask was added 20 ± 5 ml of 0.1 N sodium hydroxide. The particles were stirred at 330 rpm at 95° C for 20 minutes.
3. Addition of N-Phe: N-Phe (1.93 gram, 5.4 mmol) was dissolved in 85 ml of ethoxyethanol and the resulting solution was added to the particles dropwise over 85 - 100 minutes at a constant addition rate of approximately 1.0 ml per minute. The particles were stirred for 5 minutes, and 6.0 ml of 0.1 N sodium hydroxide and 30 ml of deionized water were added over 10 minutes. The particles were stirred for 5 minutes.
4. Addition of Rubrene: Rubrene (480 mg, 0.9 mmol) was dissolved in 200 ml of 1,2 - dimethoxyethane and was then added with stirring to the above particles over 70 - 90 minutes. Rate of addition was 3.0 ml per minute. Next, 30 ml of 0.1 N sodium hydroxide and 120 ml of 0.1 N sodium hydroxide and 120 ml of deionized water were added to the particles over 30 minutes and the medium was stirred for

10 minutes. The medium was cooled to 40° C over 1 hour with stirring. The particles were subjected to filtration on Microgon apparatus using a 43 micron (Tetko) filter (from Tetko Inc. Briarcliff Manor, NY).

Preparation of Thioxene - diphenylethynyl anthracene (DPA) particle

Microgon setup: A Microgon was assembled as described in the MiniKros lab system manual page 8 - 10. A 0.1 micron module was used operated between about 5 - 10 psi. The Microgon apparatus was with ethanol (200 proof).

Setup of apparatus:

1. An oil bath was heated to 95° C±1.O°C. A three-necked 1 liter roundbottom flask (rbf) equipped with a digital Cafeama mechanical stirrer from the middle neck was immersed into the oil bath.

2. Addition of particles: 200 ml ± 5.0 ml of latex particles were added to the rbf by means of a measuring cylinder, which was washed with 2x30 ml of ethoxyethanol (ee) and the contents were transferred to the rbf. To the flask was added 20 ± 5 ml of 0.1 N sodium hydroxide. The particles were stirred at 330 rpm at 95° C for 20 minutes.

3. Addition of C - 28 thioxene: 3.6 grams of C - 28 thioxene was dissolved in 85 ml of ethoxyethanol and the resulting solution was added to the particles dropwise over 85 - 100 minutes at a constant addition rate of approximately 1.0 ml per minute. The particles were stirred for 5 minutes and 6.0 ml of 0.1 N sodium hydroxide and 30 ml of deionized water were added over 10 minutes. The particles were stirred for 5 minutes.

4. Addition of DPA: DPA (341 mg, 0.9 mmol) was dissolved in 200 ml of 1,2 - dimethoxyethane and was then added with stirring to the above particles over 70 - 90 minutes. Rate of addition was 3.0 ml per minute. Next, 30 ml of 0.1 N sodium hydroxide and 120 ml of 0.1 N sodium hydroxide and 120 ml of deionized water were added to the particles over 30 minutes and the medium was stirred for 10 minutes. The medium was cooled to 40° C over 1 hour with stirring. The particles were subjected to filtration on Microgon apparatus using a 43 micron (Tetko) filter (from Tetko Inc. Briarcliff Manor, NY).

Oligonucleotide bound Sensitizer Particles:

The oligonucleotide was immobilized on the surface of the above particles in the following manner. Aminodextran (500 mg) was partially maleimidated by reacting it with sulfo-SMCC (157 mg, 10 mL H_2O). The sulfo-SMCC was added to a solution of the aminodextran (in 40 mL, 0.05 M Na_2HPO_4, pH 7.5) and the resulting mixture was incubated for 1.5 hr. The reaction mixture was then dialyzed against MES/NaCl (2x2L, 10 mM MES, 10 mM NaCl, pH 6.0, 4°C). The maleimidated dextran was centrifuged at 15,000 rpm for 15 minutes and the supernatant collected. The supernatant dextran solution (54 mL) was then treated with imidazole (7 mL of 1.0 M solution) in MES buffer (pH 6.0) and into this stirred solution was added the stained photosensitizer particles (10 mL of 10mg/mL). After stirring for 10 minutes the suspension was treated with EDAC (7 mmol in 10 mM pH 6.0 MES) and the suspension stirred for 30 minutes. After this time, SurfactAmps® (Pierce Chemical Company) Tween-20 (10%, 0.780 mL) was added to the reaction mixture for a final concentration of 0.1%. The particles were then centrifuged at 15,000 rpm for 45 minutes and the supernatant discarded. The pellet was resuspended in MES/NaCl (pH 6.0, 10 mM, 100 mL) by sonication. Centrifugation at 15,000 rpm for 45 minutes, followed by pellet resuspension after discarding the supernatant, was performed twice. The maleimidated dextran photosensitizer particles were stored in water as a 10 mg/mL suspension.

Thiolated oligonucleotide (oligonucleotide bearing a 5'-bis(6-hydroxyethyldisulfide) group) (Oligos Etc.) was dissolved in water at a concentration of 0.49 mM. To 116 μL of this solution was added 8.3 μL of 3.5 M sodium acetate, pH 5.3 and 8.9 μL of tris(carboxyethyl)phosphine (20 mM). After 30 minutes incubation at room temperature, 548 μL of cold ethanol. Was added and the mixture was maintained at about 20°C for 1.5 hour. The precipitated oligonucleotide was recovered by centrifugation for 2 min. at 15,000 rpm in an Eppendorf centrifuge, then dissolved in 37.5 μL of 5mM sodium phosphate, 2 mM EDTA, pH 6.

An aliquot of the maleimidated beads prepared above containing 22 mg beads was centrifuged for 30 min. at about 37,000 g, and the pellet was

resuspended in 96 μL of 0.26 M NaCl, 0.05% Tween-20, 95 mM sodium phosphate, and 0.95 mM EDTA, pH7. The thiolated oligonucleotide was added and the mixture was maintained at 37°C for 64 hours under argon. A 10 μL aliquot of sodium thioglycolate was added and incubation was continued for 2 hours at 37°C. Water was added to a total volume of 1 mL, and the beads were recovered by centrifugation, then resuspended in 5 mL of 0.1 M NaCl, 0.17 M glycine, 10 mg/mL BSA, 1 mM EDTA, 0.1% Tween-20, and 0.5 mg/mL Calf thymus DNA (Sigma Molecular Biology grade), pH 9.2. After three hours, the beads were recovered and washed three times by centrifugation, twice in buffer A and once in standard PCR buffer. The product was stored refrigerated in PCR buffer. Buffer A contained 0.1 M Tris base (J.T. Baker Chemical Co.), 0.3 M NaCl (Mallinckrodt), 25 mM EDTA Na_2 H_2O (Sigma Chemical Co.), 0.1% BSA (Sigma Chemical Co.), 0.1% dextran (Pharmacia), HBR-1 (Scantibodies), 0.05% Kathon and 0.01% gentamycin sulfate (GIBCO) prepared by dissolving and adjusting pH to 8.20 with concentrated HCL and made up to 10 L with distilled water.

A similar procedure was used for the preparation of oligonucleotide bound chemiluminescer particles

The above procedure may be modified in a manner similar to that described by Ullman, *et al.*, Proc. Natl. Acad. Sci. USA (1994) 91:5426-5427 at column 1 of page 5427.

Example 1

Target HIV RNA (obtained from Organon Teknika, Boxtel, Netherlands) was amplified by the isothermal NASBA amplification procedure. The NASBA Amplification Kit of Organon Teknika was employed. Target RNA, referred to below as WT) at various initial input (number of molecules per reaction) was amplified in the presence of reference RNA, at known number of molecules. The reference RNA molecules, referred to below as Q_a) were engineered (by Organon Teknika) to be homologous to the target RNA except for an internal sequence of 21 nucleotides. This sequence in the reference RNA was 63 nucleotides from the 3'-end thereof and was complementary to the sequence in the third oligonucleotide probe as set forth

below. The corresponding sequence in the target RNA was complementary to the sequence in the second oligonucleotide probe as set forth below. A sequence in both the target and reference RNA that was 29 nucleotides from the 3' end was complementary to a sequence in the first oligonucleotide probe or common probe.

The above design of the reference RNA ensures amplification of the reference and target RNA by the same NASBA primers and enzyme at equal amplification efficiency for target and reference RNA. The sequence of the target and reference RNA was as described by the manufacturer in the J. of Virological Methods (1993) 43:177-188.

A homogenous chemiluminescence detection method as described above was used to generate signals for determination. Chemiluminescence signal was produced when a pair of probes became bound to the target RNA analyte, one of which was bound to a singlet oxygen-producing particle, the sensitizer particle, and the other of which was bound to a particle dyed with a specific acceptor dye. These probes are referred to herein as particle detection probes. The particle detection probes included a first oligonucleotide detection probe that comprised a sequence that was complementary to a sequence in the first oligonucleotide probe indicated by underlining in the sequence below. A particle with which a sensitizer was associated was attached at the 3'-end of the first oligonucleotide detection probe. A second oligonucleotide detection probe comprised a sequence that was complementary to a sequence in the second oligonucleotide probe indicated by underlining in the sequence below.

DPA dyed particles were attached at the 3'-end of the second oligonucleotide detection probe. A third oligonucleotide detection probe comprised a sequence that was complementary to a sequence in the third oligonucleotide probe indicated by underlining in the sequence below. N-PHE dyed particles were attached at the 3'-end of the third oligonucleotide detection probe. The particle detection probes became bound to the respective first, second or third oligonucleotide probes used in this example, where the binding was non-covalent and based on the hybridization of a sequence of the particle detection probes, either a 3'- or 5'-oligonucleotide tail, to a complementary sequence in the respective first, second or third oligonucleotide

probes. Chemiluminescence signals produced by complexes of the common sensitizer particle, with each of the specific chemiluminescer particles, were specifically detected.

First oligonucleotide probe:

5'$(dT)_{20}$TGTTAAAAGAGACCATCAATGAGGA 3' (SEQ ID NO:1)

Second oligonucleotide probe:

5'$(TACT)_5$GCTGCAGAATGGGATAGA3' (SEQ ID NO:2)

Third oligonucleotide probe:

5'GATGACAGTCGCATGCAG$(CTAT)_5$ 3' (SEQ ID NO:3)

All probes were 3'- amino blocked and gel-purified by the manufacturer, namely, Oligos Etc. The underlined portions of the probes represent the sequences of 20 nucleotides that are complementary to the particle detection probes.

The sequence of the particle detection probe bound to sensitizer particles had the sequence dA_{24}.

The sequence of the particle detection probe bound to chemiluminescent particles dyed with N-PHE: 5'$(ATAG)_6$.

The sequence of the particle probe bound to the chemiluminescent dyed with DPA: 5'$(ATGA)_6$.

RNA amplification of the test and reference RNA was carried out in mixtures containing all probes, detection particles and amplification reagents. The amplification reagents used were provided by the manufacturer (Organon Teknika). The lyophilized NASBA reagent Accusphere™ was reconstituted as directed. The reagent mixture includes all primers, NTPs, $MgCl_2$ and buffer components. The target and reference RNA, the first, second and third oligonucleotide probes and corresponding particle detection probes were added to the reconstituted reaction mixture. The concentration of the first oligonucleotide probe was 25 nM and the concentration of the second and third oligonucleotide probes were at final concentration of 25 nM. The concentration of the particle detection probes were at

final concentration of 1μg per reaction. Target and reference RNA molecules, at known number of molecules, were added (2μl) to the corresponding amplification tubes. The total volume of the initial reaction mixtures, including the target and reference RNA, was 15μl. The mixtures were overlaid with 20 μl white, light mineral oil (Aldrich Chemical Co.) and incubated at 65°C for 5 min. Following incubation at 41°C for 10 min., the mixture of enzymes (5μl from Organon Teknika) was added.

Following addition of the enzyme mixture, the amplification reactions were carried out by incubation at 41°C for 60 min. The signals were read using the following program: Illumination for 0.1 sec. and read for 2.0 sec. (380-440 nm filter), followed by illumination for 0.5 sec., delay for 30 sec. and read for 10 sec. (550-660 nm filter). All measurements were carried out using a manual reader built in-house. This reader is similar to conventional readers except that it had a mechanism for fast filter change. The results are summarized in Table 2.

Table 2

Target WT/Qa	Corrected signal DPA (WT)	Ratio N-Phe* (Qa)	Input	Signal
0/0			/1.89**	
100/0	34471	-30		
50/0	23321	-14		
0/100	47	22540		
50/100	19240	13559	0.5	0.72
100/100	24875	12445	1	1.00
500/100	26787	2095	5	5.00
1000/10 0	36528	292	10	14.82

* Corrected N-Phe signal = [S-$B_{N\text{-}Phe}$ minus S-B_{DPA}] x 0.028 wherein S-B is Signal-Background. The correction factor 0.028 reflects 2.8% cross-over of DPA signal into the N-Phe signal channel.

** The ratio of corrected signal from test (WT) and reference (Qa) was divided by 1.89, to reflect the difference of chemiluminescence signal detection.

In summary, the above results demonstrate that the ratio of concentrations of the target RNA sequences can be measured by measuring the corrected ratio of the chemiluminescent signals.

Example 2

NASBA amplification and quantification of an HIV target in sample was carried out as described in Example 1. Known input amount of the test HIV target (WT) was mixed with known input amount of reference target, Qa. 2 μl of the mixture of targets was added to 13 μl of reaction mixture containing the primers, dNTPs and NTPs, the NASBA buffer, the three probes and the chemiluminescer beads and the photosensitizer beads. The oligonucleotide chemiluminescer beads were dyed with DPA, for the WT specific probe, and N-PHE, for the Qa specific probe. 1 μg of each of the chemiluminescer beads was used per assay. The reaction mixture contained 0.25 μg (per assay) of the sensitizer beads (specific for binding of the common probe). The following probes were used in the example: Oligonucleotide SEQ ID NO: 1(OB-1 common probe) at 12.5 nM; SEQ ID NO:4 (EF-4; specific for WT) at 25 nM; and SEQ ID NO 5 (EF-7; specific for Qa) at 25 nM.

The oligonucleotides EF-4 and EF-7 were as follows:

5' GCTGCAGAATGGGATAGA(TACT)$_5$ 3' SEQ ID NO:4 (EF-4)

5' GCTGCAGACAGTGTAGATA(CTAT)$_5$ 3' SEQ ID NO:5 (EF-7)

All sequences were blocked at the 3'-end by the manufacturer and gel purified. The 20 nucleotide tails on the oligonucleotide probes were complementary to the 24 nucleotide sequence present on the oligonucleotide bound chemiluminescer particles.

20 μl of mineral oil was added to each reaction tube and the tubes were incubated at 65°C for 5 minutes followed by incubation at 41°C for 10 minutes. 5 μl of a mixture of the amplification enzymes was added to each reaction tube and amplification was carried out for 90 minutes at 41°C. The signals were read as follows: Three cycles of 1 sec illumination and 0.5 sec read, followed by 30 sec delay and 10 sec read.

The obtained signals for the DPA and N-PHE chemiluminescence were corrected as follows: The background signal was subtracted and the obtained signal

was corrected using a correction factor calculated from signal obtained with the individual targets. This correction was required due to cross over of the signal from one chemiluminescence into the other. The results are shown in Table 3 below.

Table 3

Input Number of target molecules WT/Qa	Corrected Signal WT	Corrected Signal Qa	Ratio Input	Ratio Calc. Signal WT/Qa/1.5
1e5/5e3	47404	1348	**20**	**23.44**
5e4/5e3	49879	3562	10	9.34
5e4/5e3	51865	5609	10	6.16
5e3/5e3	40364	26688	**1**	**1.01**
5e3/5e3	36936	24148	**1**	**1.02**
5e2/5e3	14846	56322	0.1	0.18
5e2/5e3	13187	55392	0.1	0.16
2.5e2/5e	962	67457	**0.05**	**0.01**
0/0	0	0		
5e3/0	49174	274		
0/5e3	1	63622		

The above discussion includes certain theories as to mechanisms involved in the present invention. These theories should not be construed to limit the present invention in any way, since it has been demonstrated that the present invention achieves the results described.

The above description and examples fully disclose the invention including preferred embodiments thereof. Modifications of the methods described that are obvious to those of ordinary skill in the art such as molecular biology and related sciences are intended to be within the scope of the following claims.

WHAT IS CLAIMED IS:

1. A method for detecting the amount of a target polynucleotide in a sample, said method comprising:

(a) providing in combination in a medium (i) a sample suspected of containing said target polynucleotide, said target polynucleotide being in single stranded form, (ii) a reference polynucleotide comprising a sequence that is common with a sequence of said target polynucleotide, and (iii) a predetermined amount of an oligonucleotide probe that has a sequence that hybridizes with said sequence that is common;

(b) subjecting said combination to conditions for amplifying said target polynucleotide and said reference polynucleotide wherein said conditions permit formation of substantially non-dissociative complexes of said target polynucleotide and said reference polynucleotide, respectively, with said oligonucleotide probe and wherein said predetermined amount of said oligonucleotide probe is less than the expected amount of said amplified target polynucleotide,

(c) determining the ratio of the amount of said complex of said target polynucleotide with said oligonucleotide probe to the amount of said complex of said reference polynucleotide with said oligonucleotide probe, and

(d) relating said ratio to said known amount of said reference polynucleotide to determine the amount of said target polynucleotide in said sample.

2. The method of claim 1 wherein a signal is produced by the complex of said target polynucleotide with said oligonucleotide probe wherein said signal is different from a signal produced by the complex of said reference polynucleotide with said oligonucleotide probe and the ratio of said signals is related to the amount of said reference polynucleotide to determine the amount of said target polynucleotide in said sample.

3. The method of Claim 1 wherein said amplification is selected from the group consisting of NASBA, 3SR, SDA and amplifications utilizing Qβ-replicase.

4. The method of Claim 1 wherein said polynucleotide is DNA.

5. The method of Claim 1 wherein said polynucleotide is RNA.

6. A method for detecting the amount of a target polynucleotide in a sample, said method comprising:

(a) providing in combination in a medium (i) a sample suspected of containing said target polynucleotide, said target polynucleotide being in single stranded form, (ii) predetermined amounts of one or more reference polynucleotides, each of said reference polynucleotides comprising a first sequence that is common with a first sequence of said target polynucleotide and a second sequence that is different from a second sequence of said target polynucleotide, (iii) a predetermined amount of a first oligonucleotide probe that has a sequence that hybridizes with said sequence that is common, (iv) a second oligonucleotide probe that has a sequence that hybridizes only with said second sequence of said target polynucleotide, and (v) one or more third oligonucleotide probes, each of said third oligonucleotide probes having a sequence that hybridizes only with a respective second sequence of one of said reference polynucleotide;

(b) subjecting said combination to isothermal conditions for amplifying with equal efficiency said target polynucleotide and said one or more reference polynucleotides wherein said conditions permit formation of a substantially non-dissociative first termolecular complex of said target polynucleotide, said first oligonucleotide probe and said second oligonucleotide probe and a substantially non-dissociative second termolecular complex of each of said reference polynucleotide with said first oligonucleotide probe and a respective third oligonucleotide probe and wherein said predetermined amount of said first oligonucleotide probe is less than the expected amount of said amplified target polynucleotide,

(c) determining the ratio of the amount of said first termolecular complex to the amount of each of said second termolecular complexes, and

(d) relating each of said ratios to the predetermined amount of each of said reference polynucleotides to determine the amount of said target polynucleotide in said sample.

7. The method of claim 6 wherein said first termolecular complex has a first signal producing system and each of said second termolecular complexes has a second signal producing system wherein a signal produced by said first signal producing system is different from a signal produced by each of said second signal producing systems.

8. The method of claim 7 wherein the ratio of said signals is determined and related to the predetermined amount of said reference polynucleotide to determine the amount of said target polynucleotide in said sample.

9. The method of claim 7 wherein said signal producing systems comprise labels selected from the group consisting of a luminescent energy donor and acceptor pair, a singlet oxygen generator and chemiluminescent reactant pair, and an enzyme pair wherein a product of the first enzyme serves as a substrate for the second enzyme.

10. The method of Claim 6 wherein said amplification is selected from the group consisting of NASBA, 3SR, SDA and amplifications utilizing Qβ-replicase.

11. The method of Claim 6 wherein said polynucleotide is DNA.

12. The method of Claim 6 wherein said polynucleotide is RNA.

13. The method of Claim 6 wherein one of said first and second oligonucleotide probes is labeled with a sensitizer.

14. The method of Claim 6 wherein one of said first and second oligonucleotide probes is labeled with a chemiluminescent compound.

15. The method of Claim 6 wherein said predetermined amount of said first oligonucleotide probe is about 50-fold less than the expected amount of said amplified target polynucleotide.

16. A method for detecting the amount of a target polynucleotide in a sample, said method comprising:

(a) providing in combination in a medium (i) a sample suspected of containing said target polynucleotide, said target polynucleotide being in single stranded form, (ii) predetermined amounts of one or more reference polynucleotides, each of said reference polynucleotides comprising a first sequence that is common with a first sequence of said target polynucleotide and a second sequence that is different from a second sequence of said target polynucleotide, (iii) a predetermined amount of a first oligonucleotide probe that has a sequence that hybridizes with said sequence that is common wherein said first oligonucleotide probe has, or is capable of having, a sensitizer attached thereto, (iv) a second oligonucleotide probe that has a sequence that hybridizes only with said second sequence of said target polynucleotide wherein said second oligonucleotide probe has, or is capable of having, a first chemiluminescent compound attached thereto, and (v) one or more third oligonucleotide probes, each having a sequence that hybridizes only with a respective second sequence of one of said reference polynucleotide wherein each of said third oligonucleotide probes has, or is capable of having, a second chemiluminescent compound attached thereto, said first and said second chemiluminescent compounds differ in signal produced when activated by said photosensitizer; said second chemiluminescent compound being different for each of said third oligonucleotide probes,

(b) subjecting said combination to isothermal conditions for amplifying with equal efficiency said target polynucleotide and each of said reference polynucleotide wherein said conditions permit formation of a substantially non-dissociative first

termolecular complex of said target polynucleotide, said first oligonucleotide probe and said second oligonucleotide probe and a substantially non-dissociative second termolecular complex of each of said reference polynucleotides with said first oligonucleotide probe and a respective third oligonucleotide probe and wherein said predetermined amount of said first oligonucleotide probe is less than the expected amount of said amplified target polynucleotide,

(c) determining the ratio of the amount of said signal produced by said first termolecular complex to the amount of signal produced by each of said second termolecular complexes, and

(d) relating each of said ratios to the amount of each respective reference polynucleotide to determine the amount of said target polynucleotide in said sample.

17. The method of Claim 16 wherein said chemiluminescent compounds are each independently selected from the group consisting of enol ethers, enamines, 9-alkylidene-N-alkylacridans, arylvinylethers, dioxenes, arylimidazoles, 9-alkylidene-xanthanes and lucigenin.

18. The method of Claim 16 wherein said sensitizer is a photosensitizer.

19. The method of Claim 18 wherein said photosensitizer is selected from the group consisting of methylene blue, rose bengal, porphyrins and phthalocyanines.

20. The method of Claim 16 wherein said amplification is selected from the group consisting of NASBA, 3SR, SDA and amplifications utilizing Qβ-replicase.

21. The method of Claim 16 wherein said polynucleotide is DNA.

22. The method of Claim 16 wherein said polynucleotide is RNA.

23. A kit for use in the determination of the amount of a target polynucleotide in a sample, said kit comprising in packaged combination:

(a) reagents for conducting an amplification of said target polynucleotide,

(b) predetermined amounts of one or more reference polynucleotides, each of said reference polynucleotides comprising a first sequence that is common with a first sequence of said target polynucleotide and a second sequence that is different from a second sequence of said target polynucleotide,

(c) a predetermined amount of a first oligonucleotide probe that has a sequence that hybridizes with said sequence that is common,

(d) a second oligonucleotide probe that has a sequence that hybridizes only with said second sequence of said target polynucleotide, and

(e) one or more third oligonucleotide probes, each of said third oligonucleotide probes having a sequence that hybridizes only with a respective second sequence of one of said reference polynucleotide;

24. The kit of Claim 23 further comprising a signal producing system comprising a label on one of said first or said second or said third oligonucleotide probes wherein said label is a particle having a chemiluminescent compound associated therewith.

25. The kit of Claim 23 further comprising a signal producing system comprising a label on one of said first oligonucleotide probes wherein said label is a sensitizer.

26. The kit of Claim 23 wherein said reagents for conducting an amplification comprise a promoter and an enzyme.

27. The kit of Claim 26 wherein said amplification is selected from the group consisting of NASBA, 3SR, SDA and amplifications utilizing Qβ-replicase.

28. The kit of Claim 23 wherein said target polynucleotide is DNA.

29. The kit of Claim 23 wherein said target polynucleotide is RNA.

30. A kit for use in a determination of the amount of a specific RNA in a sample, said kit comprising in packaged combination:

(a) one or more reference RNA's,

(b) a promoter,

(c) an enzyme and

(d) predetermined amounts of one or more reference polynucleotides, each of said reference polynucleotides comprising a first sequence that is common with a first sequence of said target polynucleotide and a second sequence that is different from a second sequence of said target polynucleotide,

(e) a predetermined amount of a first oligonucleotide probe that has a sequence that hybridizes with said sequence that is common, said first oligonucleotide probe being labeled with a sensitizer,

(f) a second oligonucleotide probe that has a sequence that hybridizes only with said second sequence of said target polynucleotide, said second oligonucleotide probe being labeled with a chemiluminescer, and

(g) one or more third oligonucleotide probes, each of said third oligonucleotide probes having a sequence that hybridizes only with a respective second sequence of one of said reference polynucleotide; each of said third oligonucleotide probes being labeled with a chemiluminescer, said chemiluminescer being different from that of said second oligonucleotide probe and different for each of said third oligonucleotide probes.

31. The kit of Claim 30 wherein said sensitizer is a photosensitizer.

32. The kit of Claim 30 where said chemiluminescer is a singlet oxygen reactive olefin.

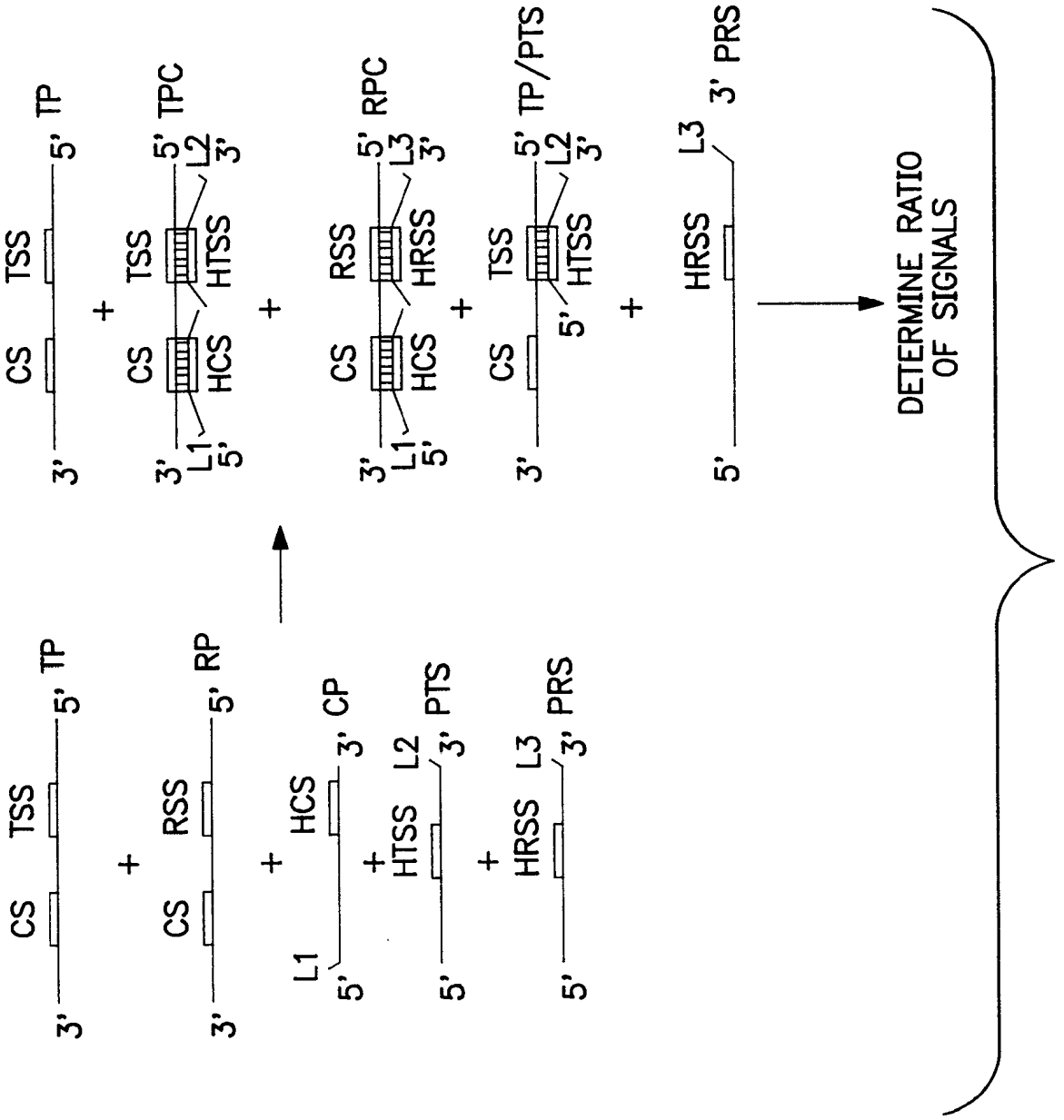


FIG. 1

IDT Exhibit 1020.0286

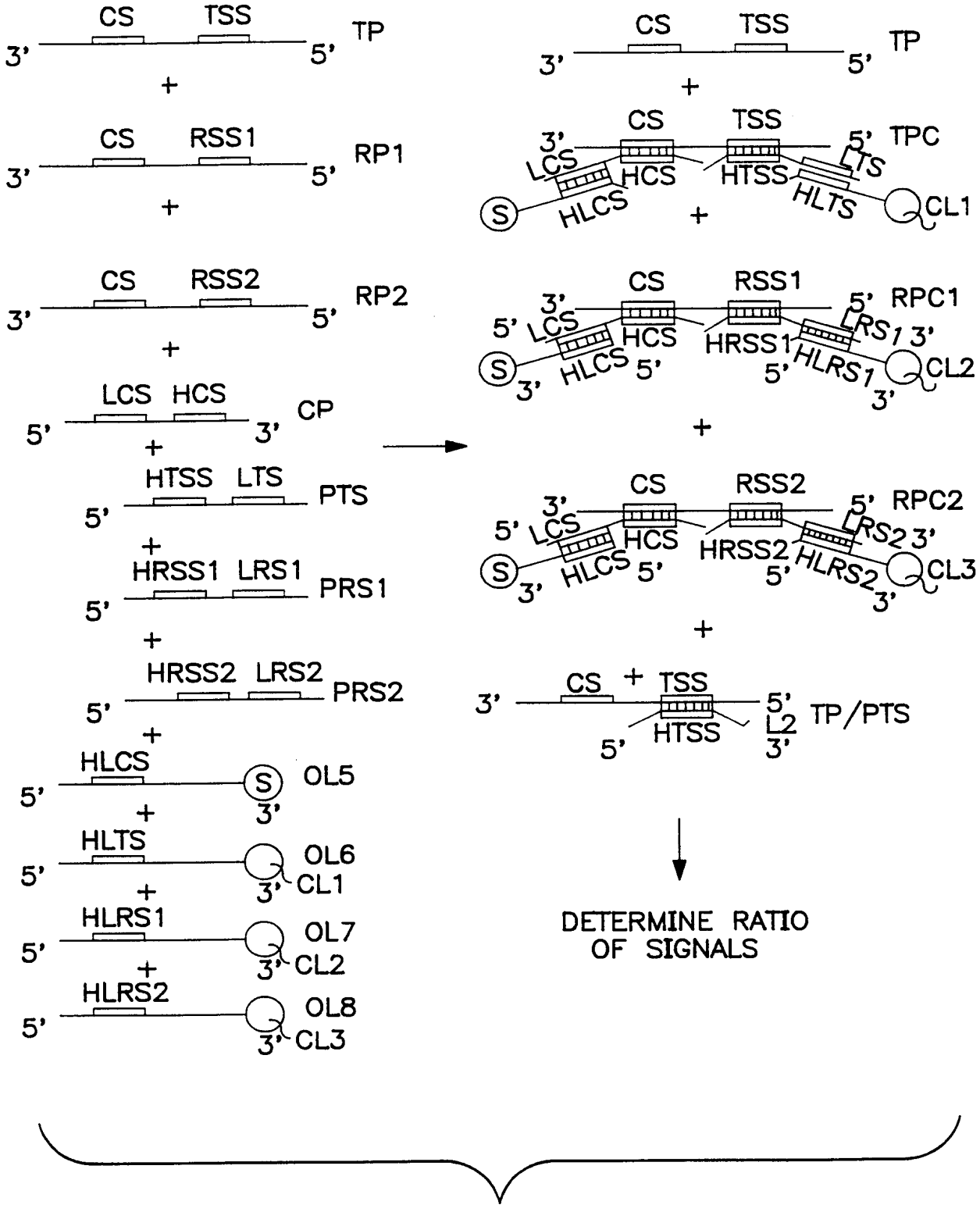


FIG. 2

IDT Exhibit 1020.0287

INTERNATIONAL SEARCH REPORT

Inte .ational Application No
PCT/US 99/03426

A. CLASSIFICATION OF SUBJECT MATTER
IPC 6 C12Q1/68

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 6 C12Q

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 94 04706 A (AKZO N.V.) 3 March 1994 see page 2, line 4 - page 6, line 12 see page 8, line 23 - page 9, line 15; claims ---	1-8, 10-12, 15,23, 26-30
X	EP 0 525 882 A (AKZO N.V.) 3 February 1993 see page 2, line 24 - line 42; claims; example 5 --- -/--	1,3-6, 10-12, 23,26-30

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

° Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
19 July 1999	30/07/1999

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Luzzatto, E

1

Form PCT/ISA/210 (second sheet) (July 1992)

IDT Exhibit 1020.0288

INTERNATIONAL SEARCH REPORT

International Application No
PCT/US 99/03426

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication,where appropriate, of the relevant passages	Relevant to claim No.
X	VAN GEMEN B. ET AL.: "A one-tube quantitative HIV-1 RNA NASBA nucleic acid amplification assay using electrochemiluminescent (ECL) labelled probes" JOURNAL OF VIROLOGICAL METHODS, vol. 49, 1994, pages 157-168, XP000600186 ISSN: 0166-0934 see abstract see p. 160, chapter 2.5, see discussion ---	1,3-5
A	EP 0 623 682 A (BECTON DICKINSON AND CO.) 9 November 1994 see page 4, line 35 - page 6, line 14; claims ---	1-22
A	ULLMAN E.F. ET AL.: "Luminescent oxygen channeling assay (LOCI (tm)): sensitive, broadly applicable homogeneous immunoassay method" CLINICAL CHEMISTRY, vol. 42, no. 9, 1996, pages 1518-1526, XP002109518 US see abstract see page 1519, column 2, line 45 - page 1520, column 2, line 21 discussion ---	9,13,14, 16-19
A	WO 94 03812 A (SYNTEX INC.) 17 February 1994 cited in the application see page 4, line 30 - page 7, line 3 see page 10, line 15 - line 19 see page 33, line 9 - page 44, line 34; claims see page 51, line 18 - line 20 ---	9,13,14, 16-22, 24,25, 31,32
A	EP 0 345 776 A (LONDON DIAGNOSTICS) 13 December 1989 cited in the application see page 2, line 25 - page 3, line 50; claims -----	9,13,32

1

Form PCT/ISA/210 (continuation of second sheet) (July 1992)

IDT Exhibit 1020.0289

INTERNATIONAL SEARCH REPORT

Information on patent family members

Inte..ational Application No
PCT/US 99/03426

Patent document cited in search report	Publication date	Patent family member(s)		Publication date
WO 9404706 A	03-03-1994	AT	162226 T	15-01-1998
		AU	685396 B	22-01-1998
		AU	4951393 A	15-03-1994
		CA	2143202 A	03-03-1994
		DE	69316372 D	19-02-1998
		DE	69316372 T	27-08-1998
		EP	0656955 A	14-06-1995
		ES	2114066 T	16-05-1998
		FI	950836 A	23-02-1995
		GR	3026631 T	31-07-1998
		JP	8500732 T	30-01-1996
		US	5770360 A	23-06-1998
		ZA	9306015 A	10-03-1994
EP 525882 A	03-02-1993	AU	670535 B	25-07-1996
		AU	2071892 A	11-03-1993
		CA	2075147 A	03-02-1993
		FI	923447 A	03-02-1993
		JP	5219999 A	31-08-1993
		US	5834255 A	10-11-1998
EP 623682 A	09-11-1994	US	5457027 A	10-10-1995
		AU	675503 B	06-02-1997
		AU	6058094 A	10-11-1994
		CA	2121658 A,C	06-11-1994
		JP	6343497 A	20-12-1994
		SG	50707 A	20-07-1998
		US	5470723 A	28-11-1995
		US	5561044 A	01-10-1996
		US	5736365 A	07-04-1998
		US	5840487 A	24-11-1998
WO 9403812 A	17-02-1994	AT	162892 T	15-02-1998
		CA	2141451 A	17-02-1994
		DE	69316757 D	05-03-1998
		DE	69316757 T	07-05-1998
		EP	0653066 A	17-05-1995
		ES	2113547 T	01-05-1998
		JP	9502253 T	04-03-1997
		US	5618732 A	08-04-1997
		US	5709994 A	20-01-1998
EP 345776 A	13-12-1989	AT	114373 T	15-12-1994
		AU	636110 B	22-04-1993
		AU	3771889 A	05-01-1990
		CA	1338927 A	25-02-1997
		DE	68919440 D	05-01-1995
		DE	68919440 T	08-06-1995
		DK	291690 A	25-01-1991
		ES	2064380 T	01-02-1995
		GR	3014969 T	31-05-1995
		IE	66734 B	24-01-1996
		IL	90542 A	11-11-1994
		JP	2681061 B	19-11-1997
		JP	3505373 T	21-11-1991
		WO	8912232 A	14-12-1989
		US	5705622 A	06-01-1998
		US	5516636 A	14-05-1996

IDT Exhibit 1020.0290

Espacenet

Bibliographic data: WO9712061 (A1) — 1997-04-03

METHOD FOR CHARACTERIZING NUCLEIC ACID MOLECULES

Inventor(s): JENDRISAK JEROME J; HOFFMAN LESLIE M; SMITH ROBERT E ± (JENDRISAK, JEROME, J, ; HOFFMAN, LESLIE, M, ; SMITH, ROBERT, E)

Applicant(s): EPICENTRE TECHNOLOGIES CORP [US] ± (EPICENTRE TECHNOLOGIES CORPORATION)

Classification: - international: ***C12N15/00; C12N15/09; C12Q1/68; G01N33/53;*** (IPC1-7): C12Q1/68
- cooperative: **C12Q1/6848; C12Q1/6869** more

Application number: WO1996US15355 19960925 Global Dossier

Priority number (s): US19950534799 19950927

Also published as: AU704625 (B2) AU7118396 (A) CA2233079 (A1) CA2233079 (C) EP0854936 (A1) more

Abstract of WO9712061 (A1)

A method of characterizing a nucleic acid molecule is disclosed. The method comprises synthesizing DNA in the presence of a reaction mixture comprising a nucleic acid template, a primer molecule, an enzyme that extends the primer so that a DNA molecule may be synthesized, four canonical deoxynucleoside triphosphates and at least one non-canonical deoxynucleoside triphosphate. The non-canonical deoxynucleoside triphosphate is incorporated into the synthesized DNA in place of a portion of only one canonical deoxynucleoside triphosphate. The synthesized DNA is treated with an N-glycosylase that excises a base portion of the non-canonical deoxynucleoside triphosphate from the synthesized DNA. The DNA is then treated in such a manner that the phosphodiester backbone of the DNA is broken at the abasic site, thus creating at least two DNA fragments. The fragments are separated according to size.

IDT Exhibit 1020.0291

PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : C12Q 1/68	A1	(11) International Publication Number: WO 97/12061
		(43) International Publication Date: 3 April 1997 (03.04.97)

(21) International Application Number: PCT/US96/15355

(22) International Filing Date: 25 September 1996 (25.09.96)

(30) Priority Data:
08/534,799 27 September 1995 (27.09.95) US

(71) Applicant: EPICENTRE TECHNOLOGIES CORPORATION [US/US]; 1202 Ann Street, Madison, WI 53713 (US).

(72) Inventors: JENDRISAK, Jerome, J.; 2721 Mason Street, Madison, WI 53705 (US). HOFFMAN, Leslie, M.; 1510 Comanche Glen, Madison, WI 53704 (US). SMITH, Robert, E.; 2504 South Whitney Way, Madison, WI 53719 (US).

(74) Agent: BAKER, Jean; Quarles & Brady, Suite 2550, 411 East Wisconsin Avenue, Milwaukee, WI 53202-4497 (US).

(81) Designated States: AU, CA, JP, PL, European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published
With international search report.

(54) Title: METHOD FOR CHARACTERIZING NUCLEIC ACID MOLECULES

(57) Abstract

A method of characterizing a nucleic acid molecule is disclosed. The method comprises synthesizing DNA in the presence of a reaction mixture comprising a nucleic acid template, a primer molecule, an enzyme that extends the primer so that a DNA molecule may be synthesized, four canonical deoxynucleoside triphosphates and at least one non-canonical deoxynucleoside triphosphate. The non-canonical deoxynucleoside triphosphate is incorporated into the synthesized DNA in place of a portion of only one canonical deoxynucleoside triphosphate. The synthesized DNA is treated with an N-glycosylase that excises a base portion of the non-canonical deoxynucleoside triphosphate from the synthesized DNA. The DNA is then treated in such a manner that the phosphodiester backbone of the DNA is broken at the abasic site, thus creating at least two DNA fragments. The fragments are separated according to size.

IDT Exhibit 1020.0292

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AM	Armenia	**GB**	United Kingdom	**MW**	Malawi
AT	Austria	**GE**	Georgia	**MX**	Mexico
AU	Australia	**GN**	Guinea	**NE**	Niger
BB	Barbados	**GR**	Greece	**NL**	Netherlands
BE	Belgium	**HU**	Hungary	**NO**	Norway
BF	Burkina Faso	**IE**	Ireland	**NZ**	New Zealand
BG	Bulgaria	**IT**	Italy	**PL**	Poland
BJ	Benin	**JP**	Japan	**PT**	Portugal
BR	Brazil	**KE**	Kenya	**RO**	Romania
BY	Belarus	**KG**	Kyrgystan	**RU**	Russian Federation
CA	Canada	**KP**	Democratic People's Republic of Korea	**SD**	Sudan
CF	Central African Republic			**SE**	Sweden
CG	Congo	**KR**	Republic of Korea	**SG**	Singapore
CH	Switzerland	**KZ**	Kazakhstan	**SI**	Slovenia
CI	Côte d'Ivoire	**LI**	Liechtenstein	**SK**	Slovakia
CM	Cameroon	**LK**	Sri Lanka	**SN**	Senegal
CN	China	**LR**	Liberia	**SZ**	Swaziland
CS	Czechoslovakia	**LT**	Lithuania	**TD**	Chad
CZ	Czech Republic	**LU**	Luxembourg	**TG**	Togo
DE	Germany	**LV**	Latvia	**TJ**	Tajikistan
DK	Denmark	**MC**	Monaco	**TT**	Trinidad and Tobago
EE	Estonia	**MD**	Republic of Moldova	**UA**	Ukraine
ES	Spain	**MG**	Madagascar	**UG**	Uganda
FI	Finland	**ML**	Mali	**US**	United States of America
FR	France	**MN**	Mongolia	**UZ**	Uzbekistan
GA	Gabon	**MR**	Mauritania	**VN**	Viet Nam

IDT Exhibit 1020.0293

METHOD FOR CHARACTERIZING NUCLEIC ACID MOLECULES

FIELD OF THE INVENTION

The field of the present invention is methods of characterizing nucleic acid molecules. Specifically, the present invention concerns characterizing nucleic acid molecules by synthesizing DNA in the presence of a non-canonical deoxynucleoside triphosphate, excising the base portion of the non-canonical deoxynucleoside triphosphate, breaking the phosphodiester backbone of the DNA at the abasic sites, and analyzing the resulting DNA fragments.

BACKGROUND OF THE INVENTION

Methods of Characterizing Nucleic Acid Molecules

There are many reasons for characterizing nucleic acid molecules. For example, genes are rapidly being identified and characterized which are causative or related to many human, animal and plant diseases. Even within any particular gene, numerous mutations are being identified that are responsible for particular pathological conditions. Thus, although many methods for detection of both known and unknown mutations have been developed (e.g., see Cotton, 1993), our growing knowledge of human and other genomes makes it increasingly important to develop new, better, and faster methods for characterizing nucleic acids. Besides diagnostic uses, improved methods for rapidly characterizing nucleic acids will also be useful in many other areas, including human forensics, paternity testing, animal and plant breeding, tissue typing, screening for smuggling of endangered species, and biological research.

A variety of methods for characterizing DNA molecules are known in the art. For example, one can characterize DNA molecules by size based on their electrophoretic migration through an agarose or polyacrylamide gel. In these methods, the negatively charged DNA molecules move through a gel in the direction of the positively charged electrode. Provided

that the percentage of agarose or polyacrylamide in the gel is appropriate for the size range of the DNA molecules being electrophoresed, smaller DNA molecules move through the pores of the gel more readily than larger DNA molecules. Because DNA molecules move in the gel at size-dependent rates, molecule sizes can be determined by staining and visualizing the DNA and then comparing the migration of the sample DNA molecules in the gel with the migration of marker DNA molecules of known size. Under the appropriate conditions, single-stranded DNA molecules differing in length by even a single nucleotide can be distinguished by denaturing polyacrylamide gel electrophoresis.

Another way to characterize DNA molecules is to treat each DNA molecule with one or more restriction endonucleases and then to determine the sizes of the various DNA fragments resulting from this treatment by agarose gel electrophoresis. Restriction endonucleases are enzymes that recognize specific sequences of bases in DNA (often 4, 5, 6 or sometimes 8 bases on each DNA strand) and then cut the phosphodiester bonds of the polynucleotide chains of DNA within to the recognition sequence. Because many restriction endonucleases with different recognition sequences are available, one can obtain a restriction map of an entire DNA molecule showing locations of restriction enzyme recognition sites and distances between them by determining which other restriction enzymes will cut each DNA fragment generated by any given restriction enzyme and what are the sizes of all of the resulting fragments. Such a restriction map is characteristic for a particular DNA molecule and can be used to obtain a rough identification of a particular sequence. Additionally, changes such as those caused by mutations in DNA may result in a loss or gain of a restriction site - a so-called "restriction fragment length polymorphism" (RFLP) (Kazazian, et al., 1989). An example of a diagnostically significant RFLP is a single base mutation in the beta-hemoglobin gene, the change from A to T which eliminates a *Dde* I restriction site, which results in sickle cell anemia (Kazazian, et al., 1989).

One of the most informative ways to characterize a DNA

molecule is to determine its nucleotide sequence. One method for sequencing DNA (Maxam and Gilbert, 1977) is accomplished by treating each of four aliquots of one strand of a 5'- or 3'-end-labelled DNA molecule to be sequenced with one of four different chemical reagents. One chemical specifically modifies only the guanine base in the DNA, another modifies only cytosine, another modifies either guanine or adenine bases, and the last chemical modifies either thymine or cytosine bases. The chemical treatments are carried out under conditions so that only a small proportion of the total susceptible bases will actually be modified.

It is important that the chemical reactions are limited in order to generate a nested set of fragments differing by one base of a specific type. If all G residues, for example, were modified, the residues would all be susceptible to phosphodiester bond cleavage. Therefore, a collection of partially modified nucleic acids is required for sequencing. Subsequent treatment with piperidine results in cleavage of the phosphodiester bonds of the DNA molecule at the abasic sites, generating a mixture of all sizes of DNA molecules that are possible following chemical modification and loss of each one of the corresponding susceptible bases. The DNA molecules in each of the four reactions are then resolved by electrophoresis in adjacent lanes of a polyacrylamide gel and the pattern of bands is revealed by exposing the gel to X-ray film if the DNA molecules are labelled with a radioisotope. The sequence of the DNA is revealed by analyzing the exposed X-ray film. Alternatively, if DNA molecules are labelled with a fluorescent, chemiluminescent or some other non-radioactive moiety, the sequence is revealed by an appropriate method known in the art.

The most commonly used method for sequencing DNA at this time (Sanger, et al., 1977) uses a DNA polymerase to produce differently sized fragments depending on the positions (sequence) of the four canonical bases (A = Adenine; C = Cytidine; G = Guanine; and T = Thymine) within the DNA to be sequenced. In this method, the DNA to be sequenced is used as a template for *in vitro* DNA synthesis.

In addition to all four of the canonical deoxynucleotides (dATP, dCTP, dGTP and dTTP), a 2',3'-dideoxynucleotide is also included in each *in vitro* DNA synthesis reaction at a concentration that will result in random substitution of a small percentage of a canonical nucleotide by the corresponding dideoxynucleotide. Thus, each DNA synthesis reaction yields a mixture of DNA fragments of different lengths corresponding to chain termination wherever the dideoxynucleotide was incorporated in place of the normal deoxynucleotide. The DNA fragments are labelled, either radioactively or non-radioactively, by one of several methods and the label(s) may be incorporated into the DNA by extension of a labelled primer, or by incorporation of a labelled deoxy or dideoxy nucleotide. By carrying out DNA synthesis reactions for each of the four dideoxynucleotides (ddATP, ddCTP, ddGTP or ddTTP), then separating the products of each reaction in adjacent lanes of a denaturing polyacrylamide gel, and detecting those products by one of several methods, the sequence of the DNA template can be read directly.

Cycle Sequencing is a variation of Sanger sequencing that achieves a linear amplification of the sequencing signal by using a thermostable DNA polymerase and repeating chain terminating DNA synthesis during each of multiple rounds of denaturation of a template DNA (e.g., at 95°C), annealing of a single primer oligonucleotide (e.g., at 55°C), and extension of the primer (e.g., at 70°C).

Nucleic acid sequencing provides the highest degree of certainty as to the identity of a particular nucleic acid. Also, nucleic acid sequencing permits one to detect mutations in a gene even if the site of the mutation is unknown. Sequencing data may even provide enough information to permit an estimation of the clinical significance of a particular mutation or of a variation in the sequence.

In order to characterize a nucleic acid by sequencing, the nucleic acid must be isolated in sufficient quantity to be used for the particular method. Although it may be possible to obtain sufficient quantities of a nucleic acid

for sequencing by first cloning it into a plasmid or other vector, this procedure is time-consuming and is often not practical for routine analysis of samples for clinical diagnostics or other purposes. When the amount of nucleic acid in a sample is less than optimal for a given method, it may be advantageous to use one of several methods which have been developed for amplifying parts of nucleic acid molecules. The polymerase chain reaction (PCR), nucleic acid sequence-based amplification (NASBA), self-sustained sequence replication (3SR), transcription-mediated amplification (TMA) and strand displacement amplification (SDA) are examples of some of the methods which have been developed for amplifying nucleic acid molecules *in vitro*.

By way of example, a specific portion of a DNA molecule may be amplified using PCR by temperature cycling of a sample DNA in a buffer containing two primers (one primer complementary to each of the DNA strands and which, together, flank the DNA sequence of interest), a thermostable DNA polymerase, and all four canonical 2'-deoxynucleoside-5'-triphosphates (dATP, dCTP, dGTP and dTTP). The specific nucleic acid sequence is geometrically amplified during each of about 30 cycles of denaturation (e.g., at 95°C), annealing of the two primers (e.g., at 55°C), and extension of the primers by the DNA polymerase (e.g., at 70°C), so that up to about a billion copies of the nucleic acid sequence are obtained. RNA may be similarly amplified using one of several protocols for RT-PCR, such as, for example, by carrying out the reaction using a thermostable DNA polymerase which also has reverse transcriptase activity (Myers and Gelfand, 1991).

The polymerase chain reaction (PCR), as discussed above, is the subject of numerous publications, including Mullis, KB, et al., U.S. Pat. No. 4,683,202 and U.S. Pat. No. 4,683,195; Mullis, KB, EP 201,184; Ehrlich, H., EP 50,424, EP 84,796, EP 258,017, & EP 237,362; Ehrlich, H., U.S. Pat. No. 4,582,788; Saiki, R., et al., U.S. Pat. No. 4,683,202; Mullis, KB, et al. (1986) in Cold Spring Harbor Symp. Quant. Biol. 51:263; Saiki, R., et al. (1985) Science 230:1350;

Saiki, et al. (1985) Science 231:487; and Loh, EY, et al. (1988) Nature 335:141.

By way of a second example, all or a specific portion of an RNA molecule may be amplified using NASBA (Fahy, et al., 1991) by isothermal incubation of a sample RNA in a buffer containing two primers (a first primer complementary to the RNA molecule and encoding a promoter sequence for an RNA polymerase and a second primer complementary to the 3'-end of the first cDNA strand resulting from reverse transcription of the RNA molecule), an RNA- and DNA-dependent DNA polymerase which also has RNase H activity (or a separate RNase H enzyme), all four canonical 2'-deoxynucleoside-5'-triphosphates (dATP, dCTP, dGTP and dTTP), an RNA polymerase that recognizes the promoter sequence of the first primer, and all four ribonucleoside-5'-triphosphates (rATP, rCTP, rGTP and rUTP).

A first cDNA strand is synthesized by extension of the first primer by reverse transcription. Then, the RNase H digests the RNA of the resulting DNA:RNA hybrid, and the second primer primes synthesis of the second cDNA strand. The RNA polymerase then transcribes the resultant double-stranded DNA (ds-DNA) molecule from the RNA polymerase promoter sequence, making many more copies of RNA, which in turn, are reverse-transcribed into cDNA and the process begins all over again. This series of reactions, from ds-DNA through RNA intermediates to more ds-DNA, continues in a self-sustained way until reaction components are exhausted or the enzymes are inactivated. DNA samples can also be amplified by other variations of NASBA or 3SR.

Strand Displacement Amplification (SDA) is another isothermal nucleic acid amplification technique (Walker, 1994). SDA is a method of nucleic acid amplification in which extension of primers, displacement of single stranded extension products, annealing of primers to the extension products (or the original target sequence) and subsequent extension of the primers occurs concurrently in the reaction mix. This is in contrast to the PCR, in which the steps of the reaction occur in discrete phases or cycles as a result

of the temperature constraints of the reaction. SDA is based upon 1) the ability of a restriction endonuclease to nick the unmodified strand of a hemiphosphorothioate form of its double-stranded recognition site and 2) the ability of certain polymerases to initiate replication at the nick and displace the downstream non-template strand. After an initial incubation at increased temperature (about 95°C) to denature double-stranded target sequences for annealing of the primers, subsequent polymerization and displacement of newly synthesized strands takes place at a constant temperature (usually about 37°C). Production of each new copy of the target sequence consists of five steps: 1) binding of amplification primers to an original target sequence or a displaced single-stranded extension product previously polymerized, 2) extension of the primers by exonuclease deficient (exo$^-$) klenow polymerase incorporating an α-thio deoxynucleoside triphosphate, 3) nicking of a hemiphosphorothioate double-stranded restriction site, 4) dissociation of the restriction enzyme from the nick site, and 5) extension from the 3'-end of the nick by exo$^-$ klenow with displacement of the downstream non-template strand. Nicking, polymerization and displacement occur concurrently and continuously at a constant temperature because extension from the nick regenerates another nickable restriction site. When primers which hybridize to both strands of a double-stranded target sequence are used, amplification is exponential, as the sense and antisense strands serve as templates for the opposite primer in subsequent rounds of amplification.

PCR, NASBA and the other methods of nucleic acid amplification can be very useful for obtaining greater quantities of a nucleic acid for additional characterization. However, in general, the amplified nucleic acid molecules must be purified away from primers, nucleotides, incomplete amplification products and other impurities prior to being used for sequencing. Otherwise, for example, the PCR primers may compete with labelled sequencing primers and the PCR nucleotides may compete with the sequencing nucleotide mixes

that are used for Sanger dideoxy sequencing. Also, Sanger sequencing clearly can not be done at the same time as PCR or another amplification method, at least not efficiently, because the dideoxynucleotides used for sequencing will result in termination of the DNA amplification reactions as well as the sequencing reactions.

One group has developed a method that attempts to decrease the number of steps required for sequencing nucleic acids (Shaw and Porter, PCT WO 95/06752). According to this method, 5'-alpha-borano-deoxynucleoside triphosphates, which were found to be resistant to exonuclease III (exo III) digestion, were incorporated into DNA during *in vitro* DNA synthesis in lieu of one of the canonical nucleotides (dATP, dCTP, dGTP, dTTP) in one of four primer extension reactions. Treatment with exo III will digest the synthesized DNA up to the point of alpha-borano deoxynucleoside incorporation. After digestion with exo III and resolution of the labeled fragments on a polyacrylamide gel, the sequence of the nucleic acid can determined.

An advantage of the alpha-borano/exo III method is that it can be integrated into PCR amplification. However, there are also disadvantages. A key disadvantage is a lower degree of accuracy of the sequence data compared to Maxam-Gilbert or Sanger sequencing because the alpha-borano/exo III method gives both extra bands and missing bands on sequencing gels. Although the mechanism by which these sequencing artifacts are generated is still uncertain, the extra bands may be due to incomplete digestion of non-boronated regions of DNA by exo III, while missing bands may be due to preferential digestion by exo III through some sequences containing alpha-borano-nucleotides. Another serious disadvantage of the alpha-borano/exo III method is related to the substrate requirements of exo III. Because exo III digests only double-stranded DNA, beginning at the 3'-end of each strand, the sequence can only be determined for the 3'-half of each strand of a PCR product, so it is not possible to obtain the sequence near to the primer. Also, because exo III digestion yields only fragments that are between 50% and 100% of the

length of the full-size PCR product, the size range of the DNA which can be sequenced by the method is somewhat limited. For example, if a PCR product of 1000 base pairs in length is sequenced according to the alpha-borano/exo III method, the fragments to be electrophoresed would be approximately 500-1000 nucleotides long. Fragments of such a length are more difficult to resolve in DNA sequencing gels.

Uracil N-Glycosylase

"Uracil N-glycosylase" or "uracil-DNA glycosylase" (UNG or UDG) is an enzyme that catalyzes the cleavage of the N-glycosidic bond between the base uracil and the sugar deoxyribose in DNA into which the non-canonical nucleotide 2'-deoxyuridine-5'-triphosphate (dUTP) has been incorporated in place of the canonical nucleotide dTTP (Lindahl, 1979). UDG does not catalyze cleavage of uracil from free dUTP, free deoxyuridine or RNA (Duncan, 1981).

U.S. patent 5,035,996 describes a process in which UDG is used for controlling contamination of nucleic acid amplification reactions (Hartley, U.S. Pat. No. 5,035,996).

The purpose of this invention was to pretreat reaction mixtures containing a new sample on which an amplification is to be carried out with UDG to assure that any uracil-containing DNA from a prior amplification of another sample had been destroyed and would not contaminate the second amplification reaction. Digesting the second amplification mixture with UDG prior to carrying out the second amplification reaction destroys the ability of any residual products of the first amplification from serving as a template for further amplifications.

Somewhat similarly, U.S. patent 5,418,149 discloses the use of glycosylases to reduce non-specific amplification of nucleic acids.

A method for introducing site-specific mutations into DNA has also been described that relies upon replacement of thymine with uracil in DNA and subsequent treatment with uracil-DNA glycosylase. See U.S. patent 4,873,192 and Kunkel, 1985. Also, uracil-containing phage were suggested

as a part of a biological containment system that would transfer genetic information only to uracil-N glycosylase deficient cells and not to naturally occurring bacteria. See Warner, et al., 1979.

Another use for UDG in molecular biology has been described by Nisson, et al., 1991. In Nisson, et al. UDG is used to facilitate directional cloning of PCR products. Thus, primers to be used for PCR amplification are made to contain a specific 12-base 5' sequence that contains dUMP in place of dTMP. Then, after the PCR amplification (without dUTP in the reaction mixture), since the amplification products contain dUMP residues at each 5' terminus, treatment of these PCR products with UDG removes the uracil residues from the 5' termini. Subsequent treatment with heat results in cleavage of the phosphodiester bonds at the abasic sites where uracil has been lost and thereby, generates 12-base cohesive termini which can be easily cloned into vectors with complementary termini.

What is needed in the art is a method for characterizing nucleic acids that is as accurate and as specific as Maxam-Gilbert or Sanger sequencing in detecting and identifying nucleic acids and differences between nucleic acids, but that is easier, faster, more sensitive and/or requires less sample DNA. The new method should also be capable of being used for relatively impure nucleic acid samples, such as amplification products, without needing to purify the sample away from primers, nucleotides or other impurities. The method also should be capable of being integrated into an amplification method such as PCR in a way similar to the alpha-borano/exo III method, but without having the disadvantages of the latter method.

SUMMARY OF THE INVENTION

The present invention provides a simple, reliable method for characterizing nucleic acid polymers. In one embodiment, the invention permits the detection and localization of mutations in nucleic acid polymers. In another embodiment, the present invention allows one to determine whether nucleic

acid molecules are similar or different and the degree of similarity or difference. In another embodiment, the invention permits one to determine the sequence of a nucleic acid molecule. Because the sequence-delimiting non-canonical nucleotides used for the method are not chain terminators for DNA synthesis, the method may also be part of or coupled to an amplification method such as PCR, NASBA, or SDA.

In the present invention, a DNA molecule is synthesized in the presence of a nucleic acid template, a primer, a polymerase and a non-canonical deoxynucleoside triphosphate as well as all four canonical deoxynucleoside triphosphates. The *in vitro* synthesized DNA is then treated with an enzyme that excises the non-canonical base from the DNA, whereby an abasic site is created. One then breaks the phosphodiester bonds of the synthesized DNA at the abasic sites and a series of molecules of different lengths is generated.

Each size class of molecule corresponds to the site where a non-canonical nucleotide substituted for the canonical nucleotide within the original sequence. Size resolution of the nested sets of cleaved DNA molecules by gel electrophoresis, capillary electrophoresis, or another method shows the position of individual substitutions by non-canonical nucleotides.

In one embodiment, the non-canonical deoxynucleoside triphosphate is dUTP. The enzyme that excises the non-canonical residue is preferably an N-glycosylase, such as uracil N-glycosylase.

In one embodiment of the present invention, the nucleotide sequence of a DNA molecule is deduced by creating four different reactions in which four different non-canonical deoxynucleoside triphosphates substitute for a portion of each of the canonical deoxynucleoside triphosphates and comparing the products of the reactions.

It is an object of the present invention to compare nucleic acid molecules at the nucleotide level.

It is another object of the present invention to provide a method to detect mutations.

It is another object of the present invention to provide

a method of incorporating a non-canonical deoxynucleoside triphosphate that may be compatible with amplification procedures.

It is another object of the present invention to provide a method of determining the sequence of a nucleic acid molecule.

It is another object of the present invention to provide a method to directly sequence amplification products.

It is an advantage of the present invention that the method is accurate and reproducible.

It is another advantage of the present invention that the method does not result in artificial results, such as missing or duplicated bands.

Other advantages, features and objects of the present invention will become apparent after examination of the specification, claims and drawings.

DESCRIPTION OF THE DRAWINGS

Figure 1 is a schematic diagram of the BESS method.

Figure 2 is a schematic diagram showing the use of BESS to evaluate specific mutation events.

Figure 3 is a schematic diagram showing coupled PCR and BESS.

Figure 4 is a schematic diagram of DNA sequencing using BESS.

Figure 5 is a schematic diagram showing the use of BESS to determine if and to what extent nucleic acid molecules are similar or different.

DETAILED DESCRIPTION OF THE INVENTION

Definitions:

The term "mutation", as used herein, refers to alterations in the nucleotide composition of a nucleic acid, and may consist of deletions, insertions, single and multiple residue changes in nucleic acid sequence. These changes normally consist of the substitution of a nucleotide residue by another species of commonly-occurring (canonical) nucleotide.

"Nucleotide" refers to a base-sugarphosphate compound. Nucleotides are the monomeric subunits of both types of nucleic acid polymers, RNA and DNA. "Nucleotide" refers to ribonucleoside triphosphates, rATP, rGTP, rUTP and rCTP, and deoxyribonucleoside triphosphates, such as dATP, dGTP, dTTP, dCTP. "Nucleoside" refers to a base-sugar combination without a phosphate group. "Base" refers to the nitrogen-containing base, for example adenine (A), cytidine (C), guanine (G) and thymine (T).

"Incorporation" refers to becoming a part of a nucleic acid polymer. There is a known flexibility in the terminology about incorporation of nucleic acid precursors. For example, the nucleotide dGTP is a deoxyribonucleoside triphosphate. Upon incorporation into DNA, it becomes a dGMP, or deoxyguanosine monophosphate moiety. Although there is no dGTP molecule in DNA, one may say that one incorporates dGTP into DNA.

The term "canonical" is used to refer to the four common nucleic acid bases adenine, cytosine, guanine and thymine that are commonly found in DNA or to the respective deoxyribonucleosides, deoxyribonucleotides or 2'-deoxyribonucleoside-5'-triphosphates that contain a canonical base. The term "non-canonical" is used to refer to nucleic acid bases in DNA other than the four canonical bases, or to the respective deoxyribonucleosides, deoxyribonucleotides, or 2'-deoxyribonucleoside-5'-triphosphates that contain a non-canonical base. Although uracil is a common nucleic acid base in RNA, uracil is a non-canonical base in DNA. "Non-canonical bases" are found in nucleic acids as a result of incorporation of non-canonical nucleotides or as a result of modification of existing bases (canonical or non-canonical).

The term "amplicon" has come to describe the product of amplification by the PCR or by other nucleic acid amplification techniques known to those skilled in the art.

All temperatures described herein are expressed in degrees centrigrade, unless otherwise specified.

In General

We disclose a new method, referred to as "Base Excision Sequence Screening" or "BESS™", for characterizing nucleic acids. Figure 1 describes an embodiment of BESS.

The BESS method is based on the incorporation of non-canonical nucleotides into DNA during *in vitro* DNA synthesis using the nucleic acid to be characterized as a template. Referring to Figure 1, a labelled primer is annealed to a nucleic acid template in the presence of a polymerase capable of extending the primer, all four canonical deoxynucleoside triphosphates and a non-canonical nucleotide. The specific non-canonical nucleotide demonstrated as an example in Figure 1 is dUTP. Following DNA synthesis, each non-canonical base is excised from DNA by an enzyme specific for that base. In the Figure 1 example, this enzyme is uracil DNA glycosylase.

Still referring to Figure 1, the DNA chain is then broken at the abasic sites corresponding to the positions of each non-canonical nucleotide. In Figure 1, the example demonstrated was the use of an AP endonuclease, Endonuclease IV, to break the DNA at abasic sites. Then, separation of the resulting DNA fragments yields patterns of fragments that end at the positions where non-canonical bases were removed. The positions of each non-canonical nucleotide within the *in vitro*-synthesized DNA correspond to the positions of a canonical nucleotide within the nucleic acid being characterized.

BESS, even when applied to a single non-canonical nucleotide, may be used to obtain a sequence-related pattern of DNA fragments in order to characterize either the presence or absence of a particular nucleic acid sequence or its relatedness to another nucleic acid sequence (see Figs. 2 and 5). Because non-canonical nucleotides used for BESS do not terminate the DNA chain, BESS may be part of an amplification process such as PCR, RT-PCR, NASBA, or SDA, or any other method involving *in vitro* DNA synthesis (see Figure 3). BESS may also be used to determine the complete nucleotide sequence of a nucleic acid molecule (see Figure 4). A preferred method for DNA synthesis during the BESS reaction is described in PCR Technology, Griffin and Griffin, eds.,

CRC Press, Boca Raton, FL (1994). The preferred minimal components for synthesis of DNA are: a template DNA (double- or single-stranded), a primer oligodeoxynucleotide of a sequence which is the reverse complement of a portion of the template, the four canonical deoxynucleoside triphosphates at a concentration of 50 - 250 micromolar, a buffer of pH 6.5 - 9.0, magnesium ions at 1 - 10 mM, and a DNA polymerase enzyme. The concentration of the non-canonical deoxynucleoside triphosphate and its proportion in the reaction mixture with respect to the corresponding canonical nucleotide are determined empirically for each non-canonical nucleotide and DNA polymerase combination. The present invention requires the incorporation of a quantity of non-canonical nucleotide such that, for any given site along the polynucleotide, only a fraction of all molecules will be cleaved at that site by an N-glycosylase enzyme. The non-canonical nucleotide may be a specific substitute for any of the four common nucleotide components for synthesizing DNA: deoxythymidine triphosphate, deoxyadenosine triphosphate, deoxycytosine triphosphate, or deoxyguanosine triphosphate. Some preferred methods are described in the Examples below.

In the BESS method, the incorporation of the non-canonical nucleotide occurs during a standard DNA synthesis reaction. Either a DNA or RNA molecule may serve as the template. Those of skill in the art know of many modifications of this reaction. Preferably, the reaction is performed with a labelled primer containing a nucleotide sequence at least part of which is sufficiently complementary to a portion of the template molecule to hybridize with the template. If the template is a DNA molecule, preferably a DNA-dependent DNA polymerase extends the primer molecule. Most preferably, this DNA-dependent DNA polymerase is a thermostable DNA-dependent DNA polymerase. If an RNA molecule is the template, an RNA-dependent DNA polymerase (reverse transcriptase) is used. Preferably, this enzyme is thermostable.

A radioactively-labelled primer enables one to visualize the reaction product by exposing the separated reaction

products to an X-ray film. Alternatively, one may label the primer and detect the reaction products by non-radioactive methods known to one of skill in the art.

dUTP as the Non-Canonical Nucleotide

In one embodiment of the invention, the non-canonical nucleotide dUTP is incorporated into a DNA sequence during DNA synthesis. The BESS method and some of its uses are illustrated in the figures using dUTP as an exemplary non-canonical nucleotide and uracil-DNA glycosylase as an exemplary N-glycosylase. For example, a specific mutation involving the loss or gain of a deoxythymidine (dT) residue may be detected after the incorporation of deoxyuridine (dU) into the DNA. Digestion of the dU-substituted sample with UDG and hydrolysis of the phosphodiester bonds at the abasic sites (points of dU removal) leads to a nested set of fragments ending at each of the abasic sites. In this embodiment, the loss or gain of a dT residue by mutation must occur on the same strand of the nucleic acid which will be detected following gel electrophoresis or another separation method.

The incorporation may be during a PCR or another amplification involving DNA synthesis known to the art. A mixture of deoxynucleotides dGTP, dCTP, dATP, dTTP and dUTP is used as the precursor pool for incorporation into newly synthesized DNA. Although the nucleotide incorporated opposite a dA on the template strand is dT for the majority of the synthesized DNA molecules, dU constitutes the nucleotide at any such position for a fraction of the DNA molecules synthesized.

The synthesized nucleic acid is then treated with an N-glycosylase, such as UDG, which removes only uracil from such sites where it has been incorporated as dU. This removal creates an "abasic" site. Because uracil is present in only a fraction of the possible incorporation sites, the nucleic acid is cleaved by UDG at that position in a fraction of the molecules. If dU is randomly incorporated into the nucleic acid at sites normally containing dT, UDG digestion and

phosphodiester bond scission generates a collection of molecules differing in length by the distance between dT residues in the original nucleic acid sequence. Such a nested set of molecules may then be resolved by electrophoretic separation means, and the precise location of the dT residues in the original nucleic acid may be inferred.

Mutation Analysis

Figure 2 describes the use of the BESS technique to examine specific mutations. Specifically, Figure 2 shows an example in which a mutation results from substitution of a single T residue in wild-type DNA by a C residue in mutated DNA. This change is detected using BESS as the loss of band on a PAGE gel for the sample with mutated DNA compared to the sample with wild-type DNA.

Referring to Figure 2 both the wild-type DNA and the mutated DNA are denatured and annealed with labelled primers. The strand complementary to the one containing the T-to-C transition serves as the template for DNA synthesis. An appropriate non-canonical nucleotide is incorporated during the DNA synthesis reaction in place of a small portion of a canonical nucleotide. Figure 2 describes the products of the wild-type template and a template with an A-to-G mutation.

Still referring to Figure 2, the products are exposed to an appropriate N-glycosylase. The N-glycosylase will excise the non-canonical base at all sites in which it is incorporated. The DNA is then treated so as to cleave it at the abasic sites and the fragments are separated. Figure 2 illustrates the result of comparing the labelled reaction products from the wild-type samples and the labelled reaction products from the mutated samples. Multiple fragments will be obtained as a result of cleavage at all abasic sites where a non-canonical base was removed by an N-glycosylase. The example shown in Figure 2, depicts only the portion of the electrophoretic gel where a mutation is observed.

If the above process is carried out using either (1) two PCR primers with different detection labels or (2) two separate PCR mixtures, each carried out with a different one

of the two PCR primers labelled, the position of dT in both strands of DNA may be determined.

As seen in Table 1, ten out of twelve possible base substitution mutations of DNA involve the loss or gain of a T on one or the other strand. Therefore, an embodiment of the invention using dUTP as the non-canonical nucleotide can be used to locate most point mutations. Table 1 compares various mutations and detection of the mutation after UDG digestion.

Table 1

Detection of single-base substitution mutations using two differently-labelled PCR primers

MUTATION IN SENSE STRAND	SENSE STRAND PRIMER	ANTI-SENSE PRIMER	MUTATION DETECTED after UDG
A---> G	--	lose T	YES
A---> C	--	lose T	YES
A---> T (U)	gain T	lose T	YES
G---> C	--	--	NO
G---> T (U)	gain T	--	YES
G---> A	--	gain T	YES
C---> T (U)	gain T	--	YES
C---> A	--	gain T	YES
C---> G	--	--	NO
(U) T---> A	lose T	gain T	YES
(U) T---> G	lose T	--	YES
(U) T---> C	lose T	--	YES
	6/12 detected with one primer	**6/12 detected with one primer**	**10/12 detected with two primers**

Thus, ten out of twelve possible point mutations may be determined with great sensitivity by coupling an embodiment of the present invention using dUTP incorporation and UDG digestion and PCR with two different labelled primers. With one primer labeled, one-half (6 out of 12) of the possible mutations could be determined using this embodiment of the invention. The detection of the other two mutations, dG to dC and dC to dG, requires the use of non-canonical nucleotides that specifically substitute for dG or for dC, and N-glycosylases which remove each respective non-canonical base. Both of these mutations could be detected using only one non-canonical nucleotide, substituting for dGTP or dCTP, and one corresponding N-glycosylase if two different labelled primers

were used (one for priming each strand of a DNA template).

Use of Other Non-Canonical Nucleotides

For embodiments of the invention involving non-canonical nucleotides other than dUTP, it is necessary that one select N-glycosylases which will selectively cleave the respective non-canonical base from the synthesized DNA strand. There are other N-glycosylases known to the art (Demple and Harrison, 1994 and Lindahl, 1979) which may be utilized to remove specific non-canonical nucleotides from DNA. Table 2 lists some of these enzymes.

Table 2

CANONICAL BASE in DNA	NON-CANONICAL NUCLEOTIDE	NON-CANONICAL NUCLEOTIDE REFERENCE	GLYCOSYLASE	SOURCE OF GLYCO-SYLASE	GLYCO-SYLASE REFERENCE
T (thymine)	dUTP (deoxyuridine triphosphate)	Bessmans, et al., 1958	UDG or UNG	*E. coli*	Lindahl, 1974
G (guanine)	dITP (deoxyinosine triphosphate)	Thomas, et al., 1978	HXNG (hypoxanthine-N-glycosylase)	a) calf thymus b) *E. coli*	Karran and Lindahl, 1980 Karran and Lindahl, 1978
C (cytosine)	5-OHMe-dCTP (5-hydroxymethyl deoxycytidine triphosphate)	Stahl and Chamberlin, 1976	hydroxy-methyl cytosine-N-glycosylase	calf thymus	Cannon, et al., 1988

A non-canonical deoxynucleotide of the present invention must be incorporated into DNA with fidelity and must be compatible as a component of product DNA during subsequent rounds of DNA synthesis. That is, the non-canonical deoxynucleotide must not either terminate the extension of the product DNA strand or cause the polymerase to falter or result in a mutation on the complementary strand. In some cases, the presence of an unusual non-canonical nucleotide may alter the DNA synthesis reaction or the physical properties of the DNA into which it is incorporated so that DNA synthesis parameters need to be optimized.

There are several other general requirements envisioned for the invention. The non-canonical nucleotide must be obtainable by some means, such as by chemical synthesis, by degradation of polynucleotides, by enzymatic modification of

nucleotides, or by a combination of these processes, and the product of said process or processes must then be converted into a deoxynucleoside triphosphate. Additionally, there must be a specific N-glycosylase to excise the non-canonical base from a nucleic acid into which it is incorporated.

There are variations of the invention envisioned which allow for the position of canonical nucleotides to be specifically determined without direct incorporation during DNA synthesis of the non-canonical nucleotide recognized by the N-glycosylase. Thus, canonical or non-canonical nucleotides in DNA are first specifically converted to non-canonical nucleotides which are specific substrates for appropriate glycosylase enzymes. Such a conversion to a non-canonical nucleotide recognized specifically by an N-glycosylase may be accomplished by chemical or enzymatic means. The phosphodiester bonds at the abasic sites generated by glycosylase activity are then broken as described herein. For example, DNA with abasic sites may be prepared by deamidation of cytosine followed by treatment with UNG (Sagher and Strauss, 1985).

Use of N-Glycosylases

Other specific N-glycosylases will become available and known to those of skill in the art. In order to determine whether or not an N-glycosylase is suitable for the present invention, one would first incorporate the non-canonical nucleotide into the synthesizing DNA and determine whether or not the non-canonical base can be specifically removed by the candidate N-glycosylase. The example above, describing the use of UDG, would serve as a comparative control.

By "N-glycosylase" or "DNA-glycosylase" we mean an enzyme with N-glycosylase activity, whether or not the enzyme is formally called a glycosylase or has a glycosylase activity combined with other enzymatic activities. Glycosylases are sometimes referred to as "glycosidases," and we therefore mean the definition of N-glycosylase to cover N-glycosidases.

As defined herein, an N-glycosylase or DNA glycosylase is an enzyme that catalyzes hydrolysis of the bond between a non-

canonical nucleic acid base and a sugar in DNA to generate an abasic (AP) site. Such enzymes are present in many species. An example from *Escherichia coli* is uracil-DNA glycosylase (UDG), also called uracil N-glycosylase (UNG). Other examples are described by Demple and Harison (1994) and by Duncan (1981).

Cleavage at the Abasic Site

Once one has created an abasic site, the method of the present invention requires that the synthesized DNA strand be cleaved at this site. There are various methods known to those of skill in the art to cleave an abasic site. Heat and/or basic conditions may be used to break the DNA molecule at the abasic sites. For example, the following protocol may be used: Nucleic acids containing abasic (AP) sites following removal of non-canonical bases are heated in a buffer solution containing an amine, for example, 25 mM Tris-HCl and 1 to 5 mM magnesium ions, for a period of 10 to 30 minutes at 70°C - 95°C. Alternatively, the following treatment may be used to break the DNA at abasic sites:

1.0 M piperidine, a base, is added to DNA which has been precipitated with ethanol and vacuum dried. The solution is then heated for 30 minutes at 90°C and lyophilized to remove the piperidine.

Preferably, enzymatic treatment is used to break the DNA polymer at the abasic site. For example, apurinic/ apyrimidinic endonucleases (AP endonucleases) have been described which are able to cleave the phosphodiester bonds of DNA at apurinic or apyrimidinic (AP) sites (Lindahl, 1979; Demple and Harrison, 1994).

The employment of AP endonucleases such as Endonuclease IV from *E. coli* appears to be the preferred method of phosphodiester bond cleavage. As defined herein, an AP endonuclease is any enzyme that catalyzes cleavage of DNA at abasic (AP) sites. Such enzymes are present in many species. Examples of AP endonucleases from *Escherichia coli* include, but are not limited by, endonuclease III and endonuclease IV. Also, *E. coli* exonuclease III in the presence of calcium ions

is an AP endonuclease. Enzymes useful in the present invention include any enzyme with AP endonuclease-like activity, whether it is called by that name or by some other name.

Endonuclease IV cleaves between the 5'-phosphate group of the AP residue and the deoxyribose ring of the adjacent nucleotide, generating a free 3'-hydroxyl group. In contrast to AP endonuclease-catalyzed cleavage, pyrolysis of the phosphodiester bond may occur on either side of the AP residue, producing a mixture of 3'-phosphorylated ends on the cleavage products. Thus, the result of AP DNA cleavage with an AP endonuclease, such as Endonuclease IV, is a set of sharper bands after electrophoresis than is obtainable with pyrolysis.

Examination of BESS Products

Once cleavage has occurred at the abasic sites, one would wish to resolve the BESS products. Preferably, this resolution will be by gel electrophoresis or capillary electrophoresis. Then, the sizes of the fragments will be visualized either by staining the DNA molecules or by exposing radioactively labelled DNA molecules to X-ray film. The fragments may likewise be visualized using non-radioactive detection methods known to those of skill in the art. The reaction products from 1, 2, 3, 4 or more BESS reactions may be visualized in the same lane of an electrophoresis gel or a single capillary if different distinguishable non-radioactive labels are used for each reaction.

A suggested protocol for electrophoresis of the BESS fragments is: a 0.2 - 0.4 mm gel of 6 or 8% polyacrylamide, 89 mM tris-borate, pH 8.3, 1 mM EDTA (TBE buffer) and 7 M urea is used to separate DNA fragments. Denaturing loading buffer (95% formamide, 0.1% bromophenol blue, 0.1% xylene cyanol and 10 mM EDTA, pH 7.6) is added to samples which were heated at 85 - 95° for 5 minutes. Then, the samples are electrophoresed at 1200 - 2500 volts in TBE buffer.

Advantages of BESS Protocol

The BESS method has several advantages. The principal advantage is the ability to characterize nucleic acid molecules at the nucleotide level. This method allows one to select specific regions of DNA or RNA for analysis at the nucleotide level.

Figure 3 demonstrates the coupling of the BESS protocol with an amplification reaction. Note that Figure 3 includes the step of incorporation of the non-canonical nucleotide in place of a small portion of a canonical nucleotide during PCR amplification. As described above for non-amplification products, the PCR product is exposed to an N-glycosylase that excises the non-canonical base at all sites in which it is incorporated. The samples are then treated to cleave the strands at the abasic sites. Figure 3 demonstrates the expected products that would occur after strand cleavage and demonstrates the analysis of the separated fragments.

Another advantage of the invention over current technologies could be the generation of a pattern of bands which may be used to confirm the identity of an amplification product (see Figure 3, for example). In some cases in the art, a PCR product of the correct molecular size is presumed to be derived from the desired template sequence. The DNA moiety may in fact be the amplification product of another nucleic acid template, and the result is then a "false positive". If the PCR product fragments generated by this invention are compared with a standard set of fragments derived from the sequence in question, they may be positively identified, even if there are mutations present. A mutation is recognizable as a deviation from the normal pattern of bands. The invention may avoid the occurrence of false positive results because the identity of the amplicon may be checked from its characteristic banding pattern after N-glycosylase digestion and phosphodiester bond breakage.

Another advantage of the invention is that the amplification process itself can incorporate non-canonical nucleotides which are then detected in the product.

An especially advantageous feature of the invention is that the amplification product does not require further

purification before the positions of the non-canonical nucleotide are determined. Many current methods of sequencing PCR products involve labor-intensive purification steps, followed by an additional single-strand amplification which incorporates label and chain terminating nucleotides. Purification of the amplicon is usually necessary to remove components of the original amplification, such as nucleotides, primers, truncated extension products and unwanted non-specific products of amplification. The present invention eliminates the need for purification steps prior to determining a nucleotide's position in the sequence.

The process described in PCT WO 95/06752 wherein exonuclease III digestion can be used to determine the presence of alpha-borano nucleotide incorporation sites, also enables the direct sequencing of PCR products. However, the process generates extra and missing bands in sequencing gels. Although the reasons for these artifacts are still uncertain, extra bands may be caused by incomplete digestion by exonuclease III (exo III). The present invention utilizes an N-glycosylase to remove specific bases from DNA, which in turn allows chain cleavage at AP sites. Sequencing gel band artifacts common with the WO 95/06752 process should not occur using the current invention.

Use of BESS to Determine Nucleotide Sequence

By incorporating non-canonical deoxynucleoside triphosphates in place of a portion of the four canonical deoxynucleoside triphosphates and by using appropriate N-glycosylase enzymes, the positions of all four canonical nucleoside triphosphates may be determined using the invention. Therefore, the complete sequence of the nucleic acid may be determined using the present invention. In this embodiment, there are separate N-glycosylases to remove each specific non-canonical nucleotide incorporated into a nucleic acid (see Figure 4). In such a case, there are distinct non-canonical nucleotides incorporated into each of four versions of the nucleic acid. The four versions of the nucleic acid are then separately digested by N-glycosylases specific for

each non-canonical nucleotide. After the glycosylase digestion and phosphodiester bond cleavage, the fragments are separated by gel electrophoresis, capillary electrophoresis or other means. Cleavage of the phosphodiester bonds of the sugar-phosphate moiety is carried out by non-enzymatic beta-elimination or by enzymatic means. If each non-canonical nucleotide is a specific substitute for each of the four common nucleotides, the complete sequence of said nucleic acid may be determined.

Figure 4 describes a method of DNA sequencing in which a labelled primer is first annealed to a denatured sample of DNA. Separate A, T, G and C-specific reactions are created to incorporate an appropriate non-canonical nucleotide in place of a small portion of the canonical nucleotide during DNA synthesis. We note that differently-labelled non-radioactive primers may be used for each reaction to permit single lane detection. Referring to Figure 4, the symbol Am, Tm, Cm and Gm refer to "modified A," "modified T," "modified C" and "modified G." These "modified bases" are the non-canonical nucleotides.

Still referring to Figure 4, the products of each nucleotide-specific reaction are then contacted with an N-glycosylase that specifically excises the appropriate non-canonical base at all sites in which it is incorporated. After non-canonical base excision, the products are treated to cleave at the abasic sites, preferably using an AP endonuclease such as *E. coli* endonuclease IV. These fragments may then be separated, preferably by denaturing polyacrylamide gel electrophoresis, by capillary electrophoresis, or by other methods, and analyzed. The sequence of the template molecule may be determined from this comparison, as demonstrated in Figure 4. Figure 4 also demonstrates how the sequence can be determined from one-lane electrophoresis if differently labelled non-radioactive primers are used for each reaction.

Alternatively, one could perform all the nucleotide-specific reactions in one large reaction providing that the incorporated non-canonical nucleotides do not cause mutagenic events. This reaction can be aliquoted into separate

reactions for treatment with the appropriate N-glycosylase. The products can then be separated and analyzed as before.

Non-canonical nucleotides may be incorporated using cycle sequencing reactions in which each of four parallel reactions contains one non-canonical nucleotide along with the four common nucleotides. Each reaction is then digested by an N-glycosylase specific for the non-canonical nucleotide incorporated. The products obtained after phosphodiester bond breakage may then be resolved using denaturing gel electrophoresis or other methods as previously described.

Use of BESS to Determine Similarities or Differences between Nucleic Acids

Just as the electrophoretic pattern of DNA fragments obtained by digesting an entire DNA molecule with different restriction enzymes can be used to characterize and roughly identify the DNA molecule (by a restriction map), the electrophoretic pattern of DNA fragments generated using the BESS method can be used to characterize and identify nucleic acid molecules at the more precise nucleotide level. Thus, the BESS method, even using a single non-canonical nucleotide and a single N-glycosylase, may be used to determine if and to what extent different nucleic acid samples are similar or different at the nucleotide level.

Figure 5 demonstrates the use of the BESS method to determine similarities and differences between nucleic acids. The method begins with annealing a labelled primer 1 and an unlabelled or differently labelled primer 2 to a denatured DNA template. (The reaction may also be performed using only a single primer.)

Non-canonical nucleotides are then incorporated in place of a small portion of a canonical nucleotide using *in vitro* DNA synthesis. Figure 5 demonstrates examples of synthesis products from four different DNA template samples to be evaluated. The products of each reaction are treated with an N-glycosylase that specifically excises the incorporated non-canonical base at all sites in which it is incorporated. The DNA is then treated so as to cleave at the abasic sites,

preferably using an AP endonuclease such as *E. coli* endonuclease IV.

Still referring to Figure 5, the cleavage products of all four samples are illustrated. These cleavage samples are separated, preferably by denaturing polyacrylamide gel electrophoresis or another method, as described herein, and the labelled fragments are detected. Many bands would be generated from most samples and the Figure 5 example shows only a few bands for illustrative purposes.

Kit for Characterizing Nucleic Acids

The present invention is also a kit for characterizing nucleic acids. In one embodiment this kit will comprise an enzyme capable of extending a primer which is hybridized to a nucleic acid template, four canonical deoxynucleoside triphosphates, at least one non-canonical deoxynucleoside triphosphate, an N-glycosylase corresponding to each non-canonical nucleotide, and an AP endonuclease. Preferably, the AP endonuclease is Endonuclease IV.

Another form of the kit would be useful for one who was performing characterization on a specific DNA template. In this embodiment of the kit, at least one primer designed to hybridize to a short section of the specific DNA template would be provided in the kit. The kit may also contain other components required to couple BESS with an amplification method. For example, a kit for coupling BESS with PCR amplification of a specific nucleic acid sequence may contain two different primers, labelled and/or unlabelled, which flank the sequence of interest.

In a most preferred embodiment of the kit, the enzyme would be a thermostable DNA polymerase.

Another embodiment of the kit would be designed for DNA sequencing applications. In this embodiment, the kit would contain four different non-canonical nucleotides designed to replace a portion of each of the four different canonical nucleotides and four different N-glycosylases which catalyze specific removal of each of the non-canonical bases following their incorporation into DNA. Alternatively, the kit could

contain only two or three non-canonical nucleotides, if both strands of a specific DNA are to be sequenced.

EXAMPLES

Example 1

This example demonstrates the detection of a single-base mutation of clinical significance using the BESS method. The example also demonstrates coupling of BESS with PCR. Coupling BESS with amplification permits one to obtain results with smaller amounts of sample nucleic acid, which may be limited in some situations.

A polymerase chain reaction (PCR) was performed to amplify a region of the human ß-globin gene contained within the plasmids pHB4 or pSic2. The plasmids were constructed by the insertion of a 290 bp segment of the human ß-globin gene into the vector pT7Blue (Novagen, Madison, WI). The two plasmids differ only by the substitution of an adenosine by a thymidine in the sense strand of the globin gene portion of pSic2. Such a mutation in human DNA is the molecular cause of the $ß^S$ or sickle globin defect (Kazazian, 1989). The primers used to amplify the globin gene were the standard forward and reverse M13 sequencing 24-mer oligonucleotides, 5'-CGCCAGGGTTTTCCCAGTCACGAC-3' and 5'-AGCGGATAACAATTTCACACAGGA-3' (SEQ ID NOs:5 and 6), respectively.

100 picograms of plasmid pHB4 or pSic2 DNA was added to PCR reactions of 50 microliters, containing 50 mM Tris-HCl pH 9.0, 20 mM ammonium sulfate, 1.5 mM magnesium chloride, 0.2 mM each of dATP, dCTP, and dGTP, 0.16 mM dTTP, 0.04 mM dUTP, 0.1 microgram of unlabeled forward primer, 0.1 microgram of ^{32}P-end-labeled reverse primer, and 2.5 units of Tth DNA polymerase or Taq DNA polymerase (Perkin-Elmer Corporation, Branchburg, NJ). The reactions were heated at 94° for 2 minutes, cycled at 95° for 30 seconds, 55° for 30 seconds, and 72° for 30 seconds for 30 repetitions, and finally held at 4°.

Following the amplification, the sample was treated with one unit of UDG (Epicentre Technologies, Madison, WI) at 37° for 20 minutes, and the phosphodiester backbone of the amplicon was cleaved at sites of uridine removal by heating at

70° for 10 minutes. The sample was diluted with 95% formamide, 10 mM EDTA, pH 9.5, 10 mM NaOH and 0.1% each of bromophenol blue and xylene cyanol.

No purification of the reaction products was performed. Rather, the products of the reaction were heat denatured and directly electrophoresed in an 8% polyacrylamide, 8 M urea gel. The gel was fixed, dried and exposed to X-ray film. Parallel samples containing sequencing reactions for the same plasmids were generated using the SequiTherm™ Cycle Sequencing Kit (Epicentre Technologies, Madison, WI) according to the manufacturer's directions for a ^{32}P-end-labeled reverse primer.

The results show that the UDG-digested products of the pHB4 amplicon containing deoxyuridine displays an additional band at the site of the A to T mutation which is characteristic of the sickle globin gene. The corresponding sequences obtained by cycle sequencing were electrophoresed in adjacent lanes for comparison. The UDG-generated lanes have the same pattern of bands as the "T" lane of the cycle sequencing products, except, as expected, the bands from the UDG-treated material migrate slightly faster. The mutation of the sickle gene is also clearly visible in the cycle sequencing lanes. The same results were obtained with either Tth DNA polymerase or Taq DNA polymerase.

Example 2

The following example demonstrates that BESS, coupled with an amplification method, is capable of detecting a clinically-significant single-base mutation, even when the mutation is present in only one allele of a heterozygous individual. The DNA samples used were obtained from humans who were either homozygous for the normal allele or heterozygous for the normal and mutant alleles. The results obtained using BESS were confirmed by sequencing both normal and mutant-containing DNA samples using standard methods.

The alpha$_1$-antitrypsin gene was amplified from human DNA using a protocol similar to that of Example 1. The Z gene mutation of the alpha$_1$-antitrypsin gene is a G-to-A transition

in exon five, which leads to an amino acid change of glutamate to lysine at position 342. An embodiment of the present invention was used to detect the Z gene mutation in one allele of a heterozygous MZ individual.

The "M" form of the alpha$_1$-antitrypsin gene is the normal or wild-type condition. Amplification reactions contained one-half microgram of human genomic DNA from apparently normal or MZ alpha$_1$-antitrypsin heterozygous persons (Coriell Institute for Medical Research, Camden, NJ) as the template in place of plasmid DNA. The primers, 5'-GAGGAGGAGCAAGGCCTATGTG-3' and 5'-GGGATTCACCACTTTTCCCATG-3', (SEQ ID NOs:1 and 2) were chosen to amplify parts of intron four and exon five of the alpha$_1$-antitrypsin gene (Cox, 1990). The reactions contained 50 mM Tris-HCl pH 9.0, 20 mM ammonium sulfate, 1.5 mM magnesium chloride, 0.2 mM each of dATP, dCTP, and dGTP, 0.16 mM dTTP, 0.04 mM dUTP, 0.1 microgram of each of the primers described above, and 2.5 units of Taq DNA polymerase (Perkin-Elmer, Branchburg, NJ).

The reactions were submitted to 35 cycles consisting of 91° for 1.5 minutes, 55° for 30 seconds, and 68° for 1 minute, then were held at 4°. The amplicons resulting from amplification in the presence of dUTP were treated with UDG and pyrolyzed as described in Example 1.

Without further purification steps, both the UDG-treated amplicons and the cycle sequence reaction samples were electrophoresed in a polyacrylamide-urea gel as described in Example 1. The MZ heterozygous DNA yielded an additional "T" band at the position of the mutation which is one-half the intensity of the other T bands, because only one amplified allele contains the Z mutation. Using BESS, the AAT Z mutation was detectable in the amplification products obtained from a mixture of normal and mutant alleles in chromosomal DNA.

As a control display of the Z mutation, the PCR products of both normal and MZ heterozygote genes were sequenced. Amplicons synthesized without dUTP and containing only the canonical four deoxynucleotides were electrophoresed in a discontinuous 13% polyacrylamide gel containing sodium dodecyl

sulfate detergent. The 310 bp amplicon was localized by staining with ethidium bromide, and cut from the gel. The PCR products were purified by crushing and soaking the gel fragment in 0.5 M ammonium acetate, 1 mM EDTA, and 0.1% SDS using methodology known to the art. This DNA was used as the template to generate cycle sequencing data using ^{32}P end-labeled primers as described in the product information for SequiTherm™ Cycle Sequencing Kits (Epicentre Technologies). Standard cycle sequencing performed as described in Example 1 confirmed the sequence of the normal and mutant DNA samples that were obtained using BESS coupled with PCR.

Example 3

The effectiveness of endonuclease IV digestion of AP DNA generated using BESS is demonstrated in Example 3. As a standard of reference, the T lane of the SequiTherm-sequenced AAT PCR product (Example 2) was compared with the results of BESS in several of its embodiments. The first embodiment was UNG digestion of dU-substituted PCR product (37° for 10 minutes) followed by pyrolysis (90° for 20 minutes) to cleave the phosphodiester bonds at AP sites. The second is the same as the first, except that endonuclease IV was added concurrently with UNG, and no 90° heat step followed. The third is the same as the second, except that EDTA was added to 10 mM final concentration prior to UNG and endonuclease IV addition. Several conclusions may be reached from the experiment:

(1) The products of the pyrolysis of AP DNA result in electrophoretic bands that are less sharp than those resulting from endonuclease IV cleavage of the same amplification product.

(2) The endonuclease IV reaction allows the polymerase to extend, in some cases, the cleavage product 3'-hydroxyl ends, resulting in spurious bands. The addition of a chelating agent such as EDTA before or concurrently with UNG and endonuclease IV prevents the polymerase from extending the cleavage product ends and eliminates the spurious bands.

(3) As expected, the BESS products migrate more rapidly

during electrophoresis than the corresponding products from Sanger cycle sequencing reactions using SequiTherm™. The migration difference is more apparent at the bottom of the electrophoresis gel.

(4) Artifactual bands obtained in cycle sequencing of the PCR products were eliminated using the invention. More accurate sequence information may be obtained using the invention in its preferred embodiments than using methods now known to the art.

Example 4

The second exon of the DQB1 gene of the human HLA class II complex is a highly polymorphic region (Bugawan and Ehrlich, 1991; Thorsby and Ronningen, 1993). Analysis of the HLA regions is useful for tissue typing prior to tissue or organ transplantation or for identity testing in forensics or other fields. To demonstrate the usefulness of BESS to distinguish QB1 alleles, several genotypes of the HLA gene were cloned into a pUC19 plasmid vector and subjected to analysis using BESS. In this example, incorporation of a non-canonical nucleotide into DNA according to the BESS procedure was carried out using only a single primer and multiple rounds of DNA synthesis in a thermocycler. Thus, only a linear amplification of only one strand of the nucleic acid template was obtained, in a manner similar to cycle sequencing. Therefore, the BESS procedure was used to linearly amplify single-stranded copies of DQB1 genes from the plasmids. As described previously, dU was partially substituted for dT in the single-stranded amplification products. Following digestion with UNG and endonuclease IV, the products were electrophoresed in all lanes adjacent to "T" reaction products obtained from standard cycle sequencing of the same plasmids.

Primers DB130 (5'-AGGGATCCCCGCAGAGGATTTCGTGTACC-3') and GH29 (5'-GAGCTGCAGGTAGTTGTGTCTGCACAC-3') (SEQ ID NOs:3 and 4) (Bugawan and Ehrlich, 1991) were used to amplify exon 2 of the DQB1 gene from human peripheral blood DNA (Novagen, Madison, WI). The blood samples from five individuals were used for the DNA preparation, which should contain many alleles of the

HLA gene target. Fifty microliter reactions containing 50 mM Tris-HCl pH 9.0, 20 mM ammonium sulfate, 1.5 mM magnesium chloride, 0.2 mM each of dATP, dGTP, dTTP and dCTP, 0.1 microgram each of DB130 and GH29 primers, 0.5 microgram human DNA, and 1 unit of Taq DNA polymerase (Perkin-Elmer, Branchburg, NJ) were amplified in an MJ Research DNA Engine thermal cycler. The samples were denatured at 94° for 3 minutes, then subjected to thirty cycles of 93° for 10 seconds and 60° for 40 seconds, followed by 75° for 3 minutes. The thermal cycler was programmed to begin timing when the sample reaches the designated temperature.

The resulting 250 bp PCR product was digested with restriction endonucleases *Bam* HI and *Pst* I, which cleave the product within its primer-encoded regions. The PCR product was purified by polyacrylamide gel electrophoresis and elution from gel fragments. Plasmid pUC19 was digested with *Bam* HI and *Pst* I, treated with HK Phosphatase (Epicentre Technologies, Madison, WI), and ligated with the PCR product. Ampicillin-resistant colonies were screened for the presence of the PCR product in their plasmids by PCR using standard forward (5'-CGCCAGGGTTTTCCCAGTCACGAC-3') and reverse (5'-AGCGGATAACAATTTCACACAGGA-3') M13 primers (SEQ ID NOs:5 and 6). Candidate bacterial colonies were swabbed with a polypropylene pipette tip and added to PCR mixtures which were cycled as above in a DNA Engine thermal cycler. PCR products encoding HLA genes were mapped by restriction endonuclease digestion (Mitsunaga, et al., 1995) Three unique clones were selected for further analysis using BESS.

For BESS analysis, DNA synthesis was primed from the three HLA clones in a manner analogous to cycle sequencing, but with a mixture of dGTP, dCTP, dTTP, dATP and dUTP as previously described in place of a dideoxy-termination mixture. The single-stranded products of thirty cycles of linear amplification were subjected to UNG and endonuclease IV digestion in the presence of 10 mM EDTA as previously described, and were electrophoresed in a denaturing polyacrylamide gel. For comparison, the nucleotide sequences of the same three clones were determined using the SequiTherm

Cycle Sequencing Kit (Epicentre Technologies, Madison, WI).

The BESS technique gave a unique set of bands for each HLA genotype, well correlated with the T sequencing lanes obtained by cycle sequencing each clone, except that, as expected and as in the previous examples, the BESS-derived bands migrated slightly more rapidly in the gel than the corresponding dideoxy-sequencing bands. The sequence-related data obtained using BESS was even more accurate than that obtained using standard cycle sequencing. Thus, the presence of dideoxy-sequencing artifacts in one region of the gel that were not present in the BESS products lanes aided the correct interpretation of the DNA sequence data. From this example, it is clear that the invention is not limited to sequence screening of PCR products or of double-stranded DNA molecules.

A kit may be made for determining the HLA genotype of human DNA samples using BESS. A kit containing the reagents listed below could be used in conjunction with a thermostable DNA polymerase such as Tth or Taq DNA polymerase (Perkin-Elmer Corp.) in order to perform PCR. The kit might contain: HLA-specific primers at 0.1 micrograms per microliter; a dNTP mixture containing 2.5 mM each of dGTP, dCTP and dATP, 2 mM dTTP and 0.5 mM dUTP; a DNA polymerase buffer containing 1.0 M Tris-HCl, pH 9.0, and 400 mM ammonium sulfate; 1 unit per microliter of uracil-DNA glycosylase; 50 mM EDTA; 25 mM magnesium chloride; 1 unit per microliter of *E. coli* Endonuclease IV; and a Stop Buffer containing 95% formamide, 10 mM EDTA, pH 9.5, 10mM NaOH and 0.1% each of bromophenol blue and xylene cyanol.

The kit could be used in a similar manner to methods described in Examples 1 through 4 and Figs. 1 through 5. Specifically, the DQB1 group of HLA genes could be amplified from human DNA obtained from peripheral blood or cheek (buccal) swabs in a 50 microliter mixture containing 50 mM Tris-HCl, pH 9.0, 20 mM ammonium sulfate, 1.5 mM magnesium chloride, 0.2 mM each of dATP, dGTP and dCTP, 0.16 mM dTTP, 0.04 mM dUTP, 0.1 micrograms each of primers DB130 and GH29 (SEQ ID NOs:3 and 4), approximately 0.5 micrograms of human DNA, and 1 unit of Taq DNA polymerase (Perkin-Elmer). The

mixture could be denatured at 94°C for three minutes, and then cycled thirty times at 93°C for 10 seconds and 60°C for 40 seconds and then ten minutes at 75°C. Then, after addition of EDTA to a final concentration of 10 mM, the amplification products could be cleaved by treatment with 1 unit each of uracil-DNA glycosylase and Endonuclease IV for ten minutes at room temperature. Following addition of the Stop Buffer, the samples could be denatured by incubating for five minutes at 95°C and then analyzed by electrophoresis in an 8% polyacrylamide-8 M urea gel. The migration of the bands generated from samples using the BESS procedure could be compared to BESS reaction products from cloned DQB1 genes, perhaps included as a control in the kit, or with DNA size standards, which could also be included in the kit. The primers in the kit may be labelled by some means, preferably by a non-radioactive means, and an automated scanner and/or image analysis system could be used to compare the band patterns of samples with the controls.

REFERENCES

U.S. PATENT DOCUMENTS

U.S. patent 4,683,202;
U.S. patent 4,873,192;
U.S. patent 5,035,996;
U.S. patent 5,418,149.

INTERNATIONAL PATENT APPLICATIONS

WO 95/06752; Shaw and Porter; 3/1995.

OTHER PUBLICATIONS

Bessmans, et al., "Enzymatic Synthesis of Deoxyribonucleic Acid. III. The Incorporation of Pyrimidine and Purine Analogs into Deoxyribonucleic Acid", Proc. Natl. Acad. Sci. USA 44:633-640, 1958.

Bugawan and Ehrlich, "Rapid Typing of HLA-DQB1 DNA Polymorphism Using Nonradioactive Oligonucleotide Probes and Amplified DNA", Immunogenetics 33:163-170, 1991.

Cannon, et al., "5-hydroxymethyl Cytosine DNA Glycosylase Activity in Mammalian Tissue", Biochem. Biophys. Res. Commun. 151:1173-1179, 1988.

Cook, et al., "Enzyme-Labeled Oligonucleotides for the Detection of Alpha$_1$-Antitrypsin Deficiency: Optimization of Enzyme Activity for Single Point Mutation Detection", Annals of Clinical Biochemistry 32:91-93, 1995.

Cotton, "Current Methods of Mutation Detection", Mutation Research 285:125-144, 1993.

Cox, "Alpha$_1$-Antitrypsin Deficiency", in The Metabolic Basis of Inherited Disease, 6th Edition, Scriver, et al. eds., McGraw-Hill, New York, pp. 2409-2437, 1990.

Demple and Harrison, "Repair of Oxidative Damage to DNA", Annual Rev. Biochemistry 63:915-948, 1994.

Duncan, "DNA Glycosylases", (1981), in The Enzymes, Boyer ed., 14:565-586, 1981.

Fahy, et al., "Self-Sustained Sequence Replication (3SR): An Isothermal Transcription-based Amplification System Alternative to PCR", PCR Methods and Applications 1:25-33, 1991.

Kwoh, et al., "Transcription-based amplification system and detection of amplified human immunodeficiency virus type I with a bead-based sandwich hybridization format", Proc. Natl. Acad. Sci. USA 86:1173-1177, 1989.

Lindahl, "An N-Glycosylase from Escherichia coli That Releases Free Uracil from DNA Containing Deaminated Cytosine Residues", Proc. Natl. Acad. Sci. USA 71:3649-3653, 1974.

Lindahl, "DNA Glycosylases, Endonucleases for Apurinic/Apyrimidinic Sites, and Base Excision-Repair", Progress in Nucleic Acid Research 22:135-192, 1979.

Karran and Lindahl, "Enzymatic Excision of Free Hypoxanthine from Polydeoxy-nucleotides and DNA Containing Deoxyinosine Monophosphate Residues", J. Biol. Chem. 253:5877-5879, 1978.

Karran and Lindahl, "Hypoxanthine in Deoxyribonucleic Acid: Generation by Heat-Induced Hydrolysis of Adenine residues and Release in Free Form by a Deoxyribonucleic Acid Glycosylase from Calf Thymus", Biochemistry 19:6005-6011, 1980.

Kazazian, et al., "Prenatal Diagnosis of Sickle Cell Anemia-1988", Annals of the New York Academy of Sciences 565:44-47, 1989.

Kunkel, 1985, Proc. Natl. Acad. Sci. USA 82:488-492.

Maxam and Gilbert, "A New Method for Sequencing DNA", Proc. Natl. Acad. Sci. USA 74:560-564, 1977.

Merajver, et al., "Somatic Mutations in the *BRCA1* Gene in Sporadic Ovarian Tumors", Nature Genetics 9:439-443, 1995.

Mitsunaga, et al., "High Resolution HLA-DQB1 Typing by Combination of Group-Specific Amplification and Restriction Fragment Length Polymorphism", Human Immunology 42:307-314, 1995.

Myers and Gelfand, "Reverse Transcription and DNA Amplification by a *Thermus thermophilus* DNA Polymerase", Biochemistry 30:7661-7666, 1991.

Nisson, et al., "Rapid and Efficient Cloning of Alu-PCR Products Using Uracil DNA Glycosylase", PCR Methods and Applications", 1:120-123, 1991.

Sagher and Strauss, "Abasic Sites From Cytosine as Termination Signals for DNA Synthesis", Nucl. Acids Res. 13(12):4285-4298, 1985.

Sanger, et al., "DNA Sequencing with chain-terminating inhibitors", Proc. Natl. Acad. Sci. USA 74:5463-5468, 1977.

Stahl and Chamberlin, "Groups on the Outside of the DNA Helix Effect Promoter Utilization by T7 RNA Polymerase", in RNA Polymerase, Losich and Chamberlin eds., Cold Spring Harbor Press, Cold Spring Harbor, N.Y., pp. 429-440, 1976.

Thorsby and Ronningen, "Particular HLA-DQ Molecules Play a Dominant Role in Determining Susceptibility to Resistance to Type 1 (Insulin-Dependent) Diabetes Mellitus", Diabetologia 36:371-377, 1993.

Walker, "Strand Displacement Amplification", In: Novel Amplification Technologies for DNA/RNA-Based Diagnostics, a book presented at a meeting of that name on April 20-22, 1994 in San Francisco, CA, organized by International Business Communications, Southborough, MA, 1772-1749.

Warner, et al., 1979, J. Biol. Chem. 245(16):4734-7539.

Yap and McGee, "Detection of Mutations by PCR", in PCR Technology, Griffin and Griffin, eds., CRC Press, Boca Raton, FL, pp. 107-120, 1994.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT: Jendrisak, Jerome J.
Hoffman, Leslie M.
Smith, Robert E.

(ii) TITLE OF INVENTION: METHOD FOR CHARACTERIZING NUCLEIC ACID MOLECULES

(iii) NUMBER OF SEQUENCES: 6

(iv) CORRESPONDENCE ADDRESS:
(A) ADDRESSEE: Quarles & Brady
(B) STREET: 411 East Wisconsin Avenue
(C) CITY: Milwaukee
(D) STATE: Wisconsin
(E) COUNTRY: U.S.A.
(F) ZIP: 53202-4497

(v) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.25

(vi) CURRENT APPLICATION DATA:
(A) APPLICATION NUMBER:
(B) FILING DATE:
(C) CLASSIFICATION:

(viii) ATTORNEY/AGENT INFORMATION:
(A) NAME: Baker, Jean C.
(B) REGISTRATION NUMBER: 35,433
(C) REFERENCE/DOCKET NUMBER: 310307.90053

(ix) TELECOMMUNICATION INFORMATION:
(A) TELEPHONE: (414) 277-5709
(B) TELEFAX: (414) 271-3552

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 22 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (oglionucleotide)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

GAGGAGGAGC AAGGCCTATG TG 22

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 22 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (oglionucleotide)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

GGGATTCACC ACTTTTCCCA TG 22

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 29 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (oligonucleotide)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

AGGGATCCCC GCAGAGGATT TCGTGTACC 29

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 27 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (oligonucleotide)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

GAGCTGCAGG TAGTTGTGTC TGCACAC 27

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 24 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (oligonucleotide)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

CGCCAGGGTT TTCCCAGTCA CGAC 24

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 24 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (oligonucleotide)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

AGCGGATAAC AATTTCACAC AGGA 24

CLAIMS

We claim:

1. A method for characterizing a nucleic acid molecule comprising:

a) synthesizing DNA in the presence of a reaction mixture comprising

(i) a nucleic acid template,

(ii) a primer molecule that contains a nucleotide sequence at least part of which is sufficiently complementary to a portion of the template to hybridize therewith,

(iii) an enzyme that extends the primer so that a DNA molecule may be synthesized,

(iv) four canonical deoxynucleoside triphosphates, and

(v) at least one non-canonical deoxynucleoside triphosphate, wherein the non-canonical deoxynucleoside triphosphate is incorporated into the synthesized DNA in place of only one canonical deoxynucleoside triphosphate;

b) contacting the synthesized DNA with an N-glycosylase that excises a base portion of the non-canonical deoxynucleoside triphosphate from said synthesized DNA, whereby an abasic site is created;

c) treating the DNA in such manner that said treatment results in breakage of a phosphodiester backbone at all abasic sites, whereby at least two DNA fragments are created; and

d) separating the fragments according to size.

2. The method of claim 1 wherein the fragments are detected by transferring said size-separated fragments to a solid support and then revealing the presence of said fragments by hybridization thereto of a labelled complementary probe.

3. The method of claim 1 wherein the synthesized DNA is labelled with a detectable probe.

4. The method of claim 1 wherein the template is a DNA molecule and the enzyme of step (a) is a DNA-dependent DNA polymerase.

5. The method of claim 4 wherein the enzyme is thermostable.

6. The method of claim 1 wherein the template is an RNA molecule and the enzyme of step (a) is an RNA-dependent DNA polymerase.

7. The method of claim 6 wherein the enzyme is thermostable.

8. The method of claim 1 wherein the non-canonical deoxynucleoside triphosphate is 2'-deoxyuridine-5'-triphosphate and the N-glycosylase is uracil-N-glycosylase.

9. The method of claim 1 wherein the non-canonical deoxynucleoside triphosphate is 2'-deoxyinosine-5'-triphosphate and the N-glycosylase is hypoxanthine-N-glycosylase.

10. The method of claim 1 wherein a non-canonical base is modified after incorporation into the synthesized DNA by treating the synthesized DNA in order to make the modified base a non-canonical base which is_susceptible to loss upon treatment of the synthesized DNA with an N-glycosylase.

11. The method of claim 1 additionally comprising the step of providing three additional DNA synthesis reactions, to provide a total of four DNA synthesis reactions, wherein a different non-canonical deoxynucleoside triphosphate is used in each reaction.

12. A method for determining the complete nucleotide sequence of a nucleic acid template by analyzing the nucleic acid fragments of claim 11.

13. The method of claim 1 wherein step (c) comprises heating the DNA to break the phosphodiester backbone at all abasic sites.

14. The method of claim 1 wherein step (c) comprises using a basic solution to effect the breakage of the phosphodiester backbone at all abasic sites.

15. The method of claim 1 wherein step (c) comprises using an enzyme to effect the breakage of the phosphodiester backbone at all abasic sites.

16. The method of claim 15 wherein step (c) comprises using an AP endonuclease to effect the breakage of the phosphodiester backbone at all abasic sites.

17. The method of claim 1 additionally comprising the step of providing at least two different non-canonical nucleotides in the synthesizing step, wherein the reaction is divided into separate aliquots in the contacting step so that a different N-glycosylase is provided for each aliquot.

18. The method of claim 1 wherein DNA synthesis is a part of a method for amplification of the template nucleic acid.

19. The method of claim 1 wherein DNA synthesis is a part of amplification of a template nucleic acid using the Polymerase Chain Reaction.

20. The method of claim 1 wherein DNA synthesis is a part of a NASBA reaction.

21. The method of claim 1 wherein DNA synthesis is a part of 3SR.

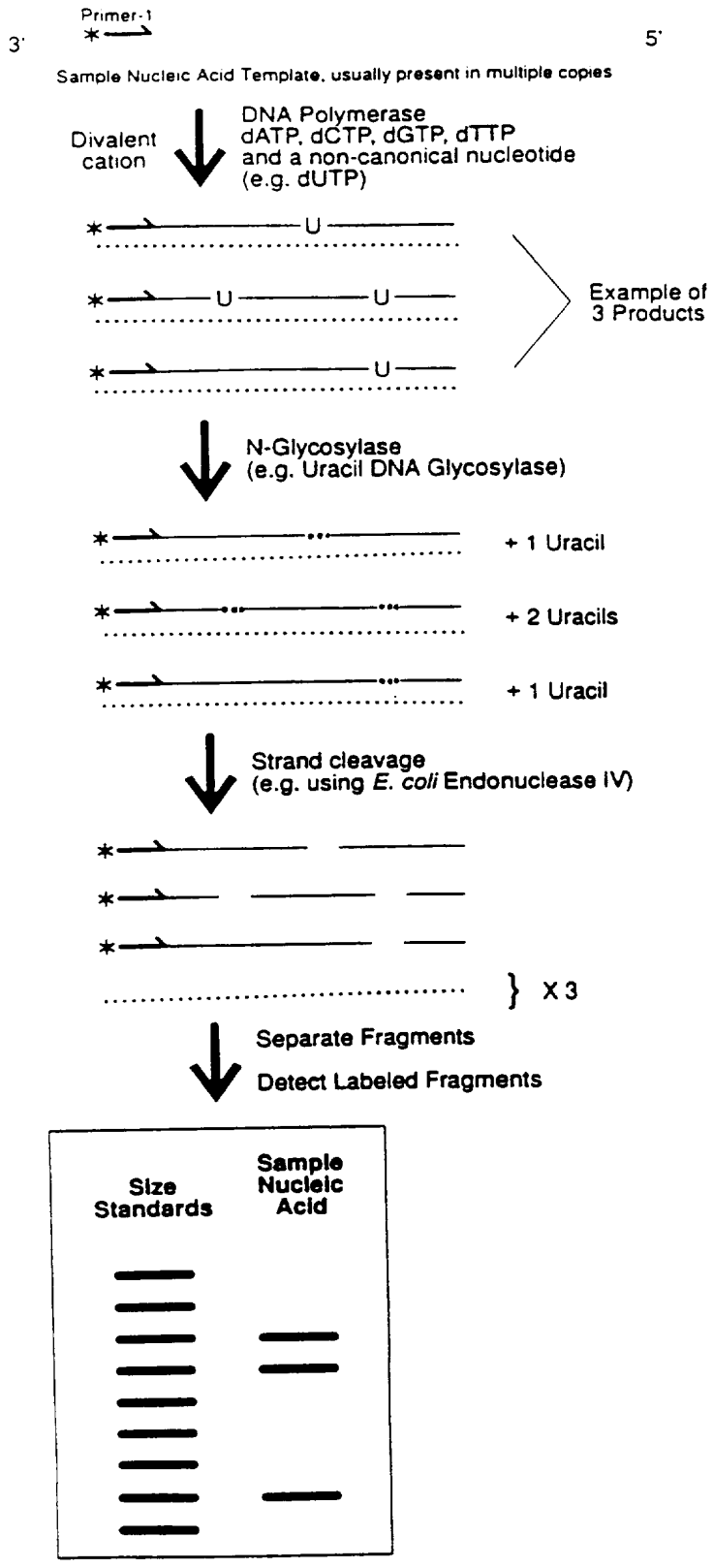
22. The method of claim 1 wherein DNA synthesis is a part of Strand Displacement Amplification.

23. The method of claim 1 in which electrophoresis is used to separate the DNA fragments.

24. A kit for characterizing a nucleic acid template, comprising

an enzyme capable of extending a primer bound to a nucleic acid template so that a DNA molecule may be synthesized, four canonical deoxynucleoside triphosphates, at least one non-canonical deoxynucleoside triphosphate, an N-glycosylase capable of excising the base portion of each respective non-canonical deoxynucleoside triphosphate from said synthesized DNA, and an AP endonuclease.

25. The kit of claim 24 additionally comprising at least one primer molecule.



① Anneal labeled Primer to DNA template.

② Incorporate the appropriate non-canonical nucleotide in place of a small portion of a canonical nucleotide during *in vitro* DNA synthesis.

③ Contact the *in vitro*-synthesized DNA with an N-glycosylase that excises the non-canonical base at all sites in which it is incorporated.

④ After non-canonical base excision, treat the DNA so as to cleave it at abasic sites.

⑤ Separate the fragments (e.g. by denaturing PAGE).

⑥ Detect the labeled fragments (e.g.s X-ray film for a radioactively-labeled primer or a fluorescent detector for a fluorescently-labeled primer).

FIG. 1

IDT Exhibit 1020.0339

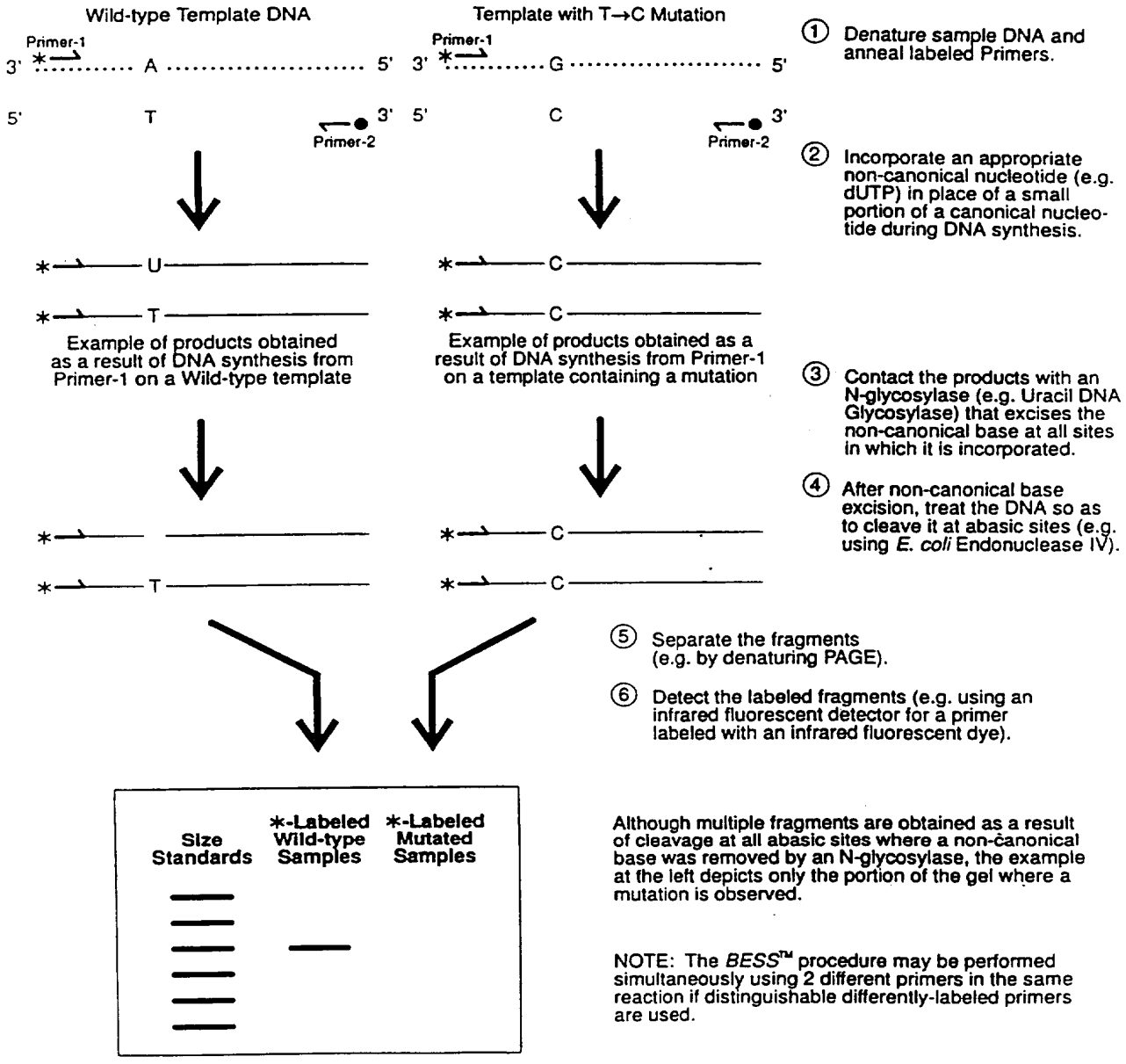
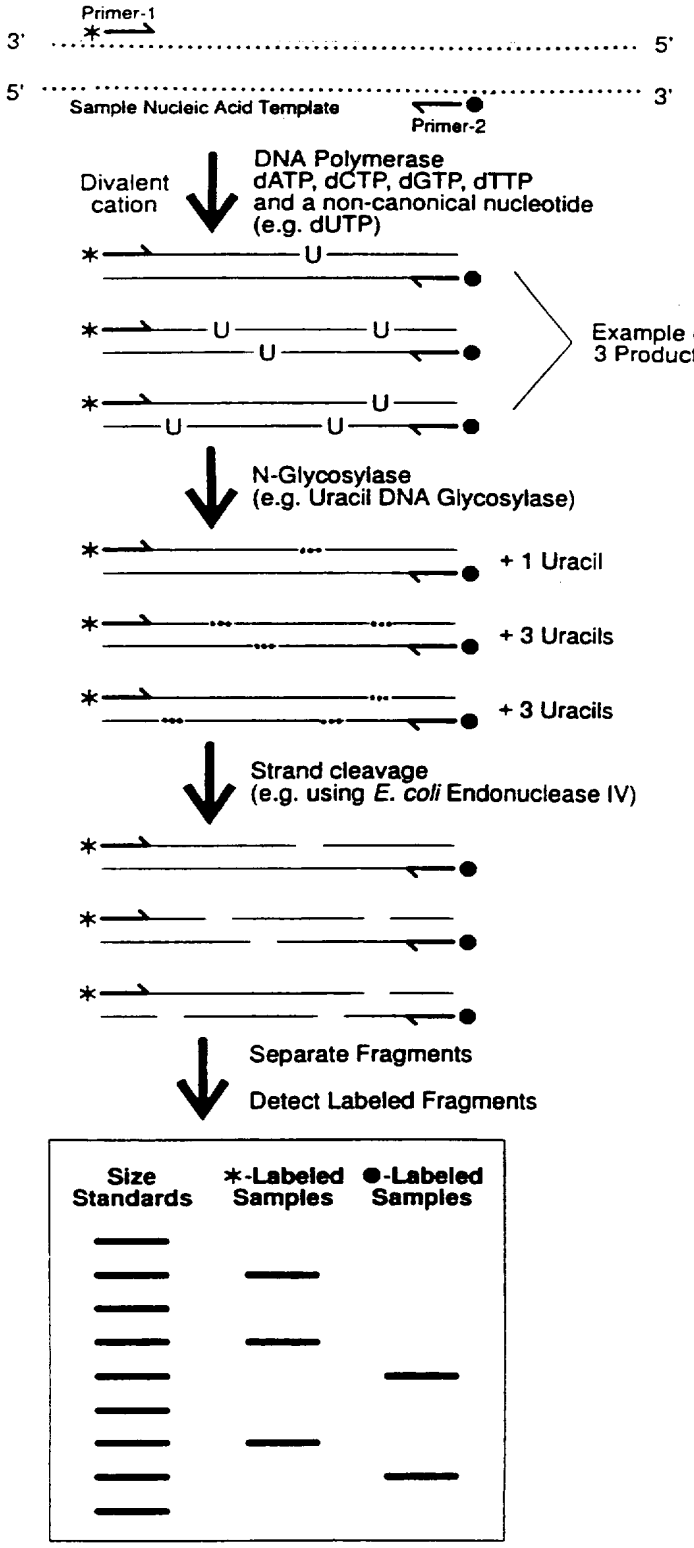


FIG. 2

IDT Exhibit 1020.0340



① Denature sample DNA and anneal labeled Primer-1 and an unlabeled or differently-labeled Primer-2 to DNA template.

② Incorporate the appropriate non-canonical nucleotide in place of a small portion of a canonical nucleotide during PCR amplification.

③ Contact the PCR product with an N-glycosylase that excises the non-canonical base at all sites in which it is incorporated.

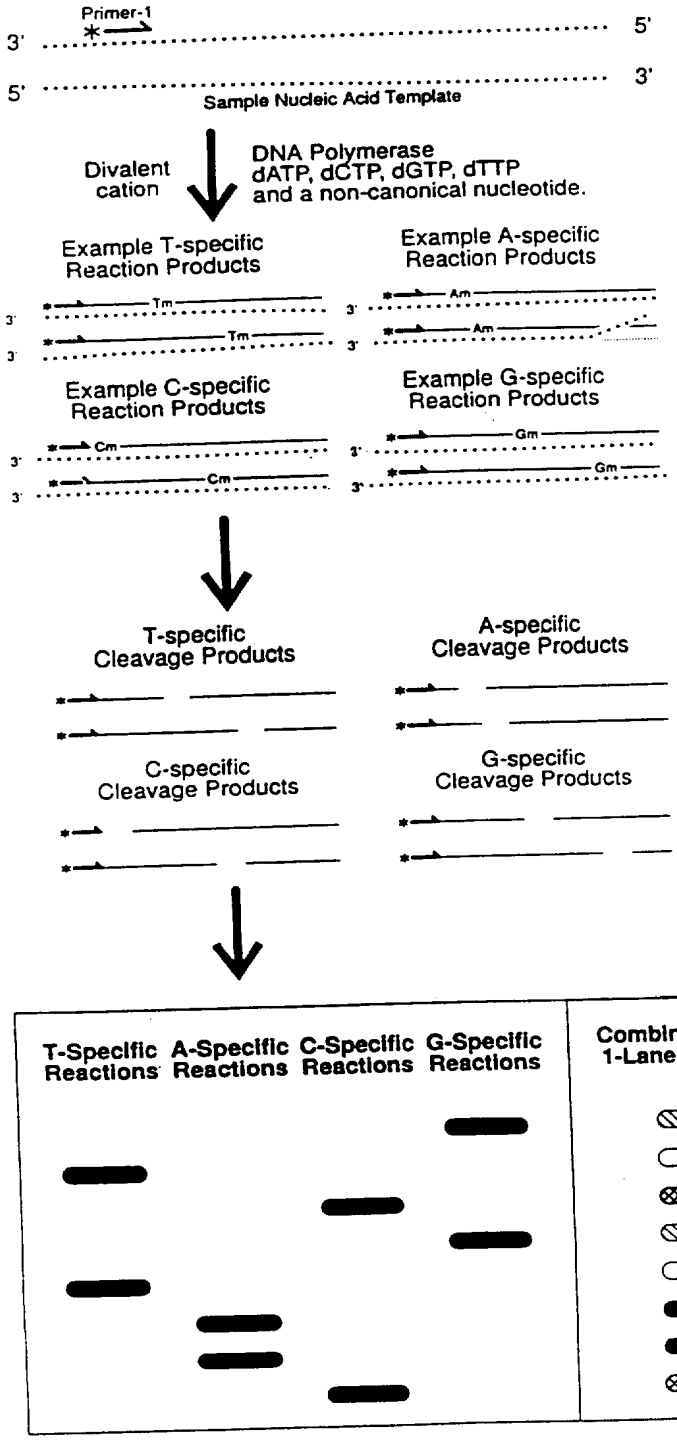
④ After non-canonical base excision, treat the DNA so as to cleave it at abasic sites.

⑤ Separate the fragments (e.g. by denaturing PAGE).

⑥ Detect the labeled fragments (e.g.s X-ray film for a radioactively-labeled primer or a fluorescent detector for a fluorescently-labeled primer). Multiple differently-labeled primers may be detected in the same lane if they can be detected differentially.

FIG. 3

IDT Exhibit 1020.0341



① Denature sample DNA and anneal labeled Primer.

② In separate A, T, G and C-specific reactions, incorporate the appropriate non-canonical nucleotide in place of a small portion of a canonical nucleotide during DNA synthesis (e.g. "modified A" (Am), "modified T" (Tm), "modified C" (Cm) and "modified G" (Gm)). Differently-labeled non-radioactive primers may be used for each reaction to permit single-lane detection.

③ Contact the products of each reaction set with an N-glycosylase that specifically excises the appropriate non-canonical base at all sites in which it is incorporated.

④ After non-canonical base excision, treat the DNA so as to cleave it at abasic sites (e.g. using *E. coli* Endonuclease IV).

⑤ Separate the fragments (e.g. by denaturing PAGE).

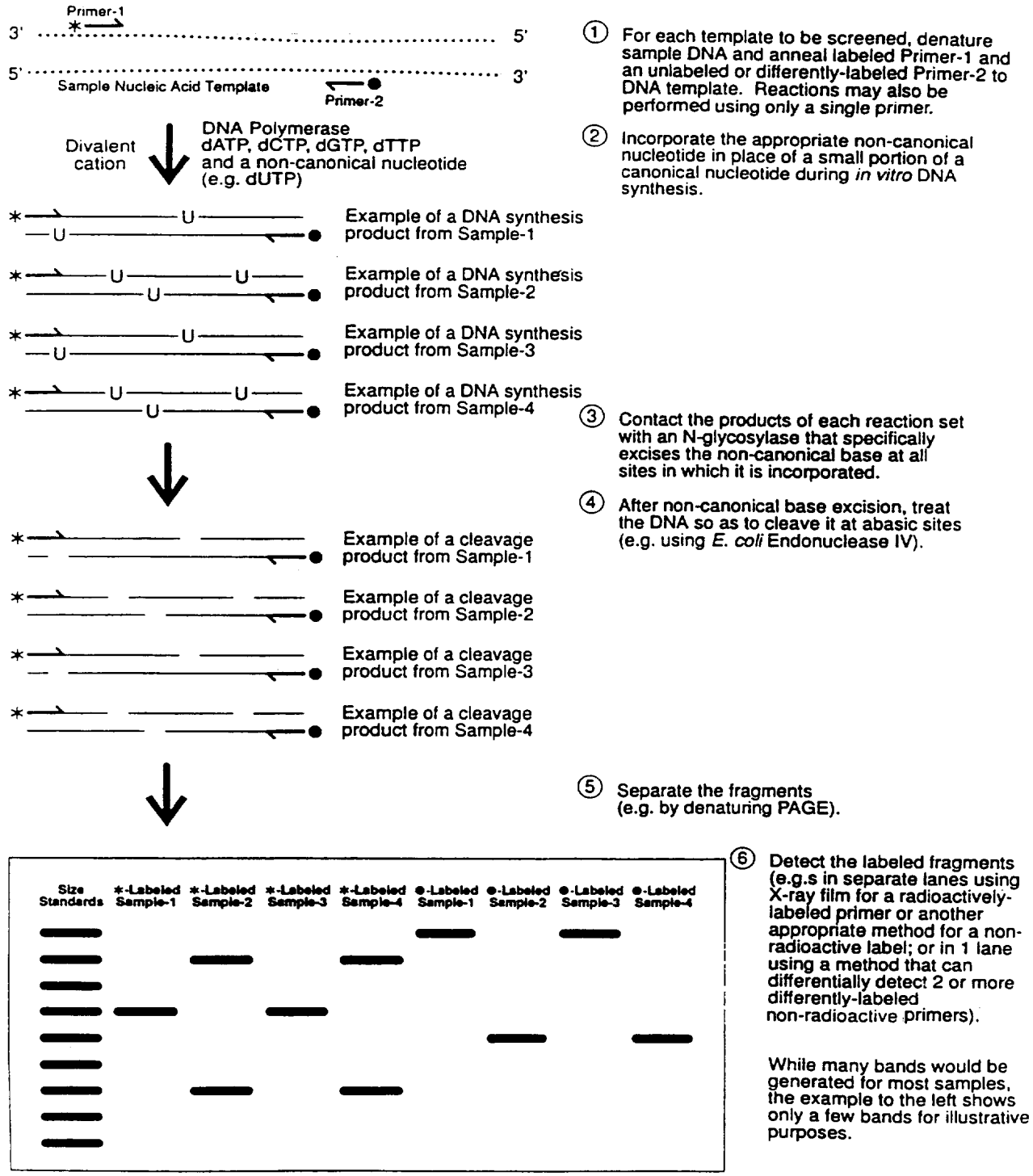
⑥ Detect the labeled fragments (e.g.s in 4 separate lanes using X-ray film for a radioactively-labeled primer or another appropriate method for a non-radioactive label; or in 1 lane using a method that can differentially detect 4 differently-labeled non-radioactive primers).

Sequence would be read from bottom to top of gel as:

5'- C A A T G C T G -3'

FIG. 4

SUBSTITUTE SHEET (RULE 26)



① For each template to be screened, denature sample DNA and anneal labeled Primer-1 and an unlabeled or differently-labeled Primer-2 to DNA template. Reactions may also be performed using only a single primer.

② Incorporate the appropriate non-canonical nucleotide in place of a small portion of a canonical nucleotide during *in vitro* DNA synthesis.

③ Contact the products of each reaction set with an N-glycosylase that specifically excises the non-canonical base at all sites in which it is incorporated.

④ After non-canonical base excision, treat the DNA so as to cleave it at abasic sites (e.g. using *E. coli* Endonuclease IV).

⑤ Separate the fragments (e.g. by denaturing PAGE).

⑥ Detect the labeled fragments (e.g.s in separate lanes using X-ray film for a radioactively-labeled primer or another appropriate method for a non-radioactive label; or in 1 lane using a method that can differentially detect 2 or more differently-labeled non-radioactive primers).

While many bands would be generated for most samples, the example to the left shows only a few bands for illustrative purposes.

FIG. 5

SUBSTITUTE SHEET (RULE 26)

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US96/15355

A. CLASSIFICATION OF SUBJECT MATTER

IPC(6) :C12Q 1/68
US CL :435/6

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 435/6

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

aps, cas, biosis

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	PCR Methods and Applications, Volume 1, issued in 1991, Fahy et al, "Self-sustained Sequence Replication (3SR): An isothermal Transcription-based Amplification System Alternative to PCR", pages 25-33, see entire document.	1-25
Y	US, A, 5,035,996 (HARTLEY) 30 July 1991, see entire document.	1-25
Y	Proc. Natl. Acad. Sci. Volume 86, issued February 1989, Kwoh et al, "Transcription-based amplification system and detection of amplified human immonodeficiency virus type I with a bead-based sandwich hybridization format", pages 1173-1177, see entire document.	1-25

[X] Further documents are listed in the continuation of Box C. [ ] See patent family annex.

*	Special categories of cited documents:	"T"	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A"	document defining the general state of the art which is not considered to be of particular relevance	"X"	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"E"	earlier document published on or after the international filing date	"Y"	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"L"	document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"&"	document member of the same patent family
"O"	document referring to an oral disclosure, use, exhibition or other means		
"P"	document published prior to the international filing date but later than the priority date claimed		

Date of the actual completion of the international search 26 NOVEMBER 1996	Date of mailing of the international search report 26 DEC 1996
Name and mailing address of the ISA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231 Facsimile No. (703) 305-3230	Authorized officer EGGERTON CAMPBELL Telephone No. (703) 308-0196

Form PCT/ISA/210 (second sheet)(July 1992)*

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US96/15355

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	Mutation Research, Volume 285, issued 1993, Cotton, R.G.H. "Current methods of mutation detection", pages 125-144, see entire document.	1-25

Form PCT/ISA/210 (continuation of second sheet)(July 1992)*

IDT Exhibit 1020.0345

Espacenet

Bibliographic data: WO9838296 (A1) — 1998-09-03

METHOD FOR SIMULTANEOUS LIGATION OF MULTIPLE DNA FRAGMENTS

Inventor(s): SHARON GIL [IL] ± (SHARON, GIL)

Applicant(s): GESHER ISRAEL ADVANCED BIOTECS [IL]; SHARON GIL [IL] ± (GESHER-ISRAEL ADVANCED BIOTECS (1996) LTD, ; SHARON, GIL)

Classification: - international: ***C12N15/09; C12N15/10; C12N15/66; C12Q1/68;*** (IPC1-7): C12N15/10; C12N15/66; C12Q1/68
- cooperative: **C12N15/10; C12N15/66; C12Q1/686; C12Q1/6862** more

Application number: WO1998IL00094 19980226 Global Dossier

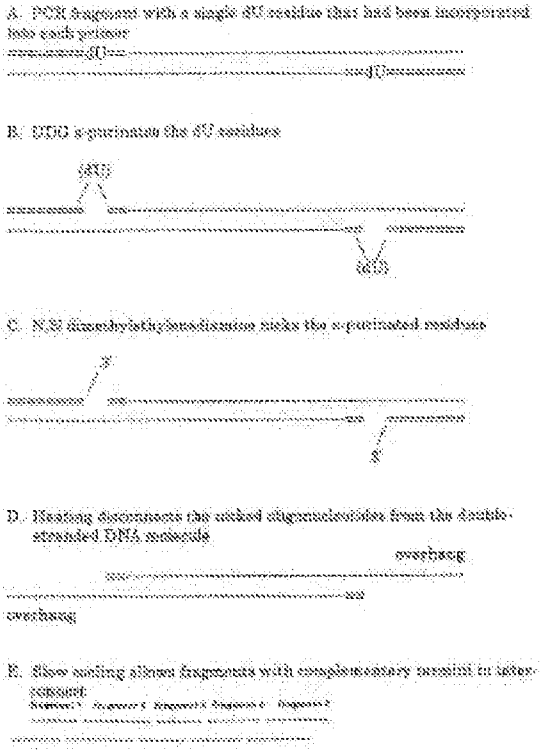
Priority number (s): IL19970120337 19970227

Also published as: AU6228698 (A) EP1000141 (A1) JP2001514498 (A)

Abstract of WO9838296 (A1)

A method is described for the simultaneous assembly of two or more primer extension generated, double stranded DNA fragments, into a construct, comprising the steps of: 1) designing the primers such that the fragments subsequently produced by PCR, have 3' termini of 15 nucleotides or more being complementary to a second and third fragment, both having 3' termini of 15 nucleotides or more, complementary to each terminal of first said fragment; 2) providing in each of said primers and subsequently in each of said fragments, at least one dU residue in place of at least one dT residue, being positioned at least 15 nucleotides from the 5' end of said primer and subsequent fragments; 3) adding to the mixture of fragments the enzyme UDG (Uracil DNA Glycosylase) and a compound such as N,N,dimethylethylenedimethylamine, to a-purinate and generate single strand nicks at said dU residues, respectively, thereby to generate a 3' single strand overhang on each fragment to be joined thereto, being complementary to the other 3' single strand overhangs of second, third or more aforementioned fragments, and 4) allowing for all fragments thus generated to anneal into a single and stable DNA construct.

IDT Exhibit 1020.0346



IDT Exhibit 1020.0347

PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 :
C12N 15/10, C12Q 1/68, C12N 15/66

A1

(11) International Publication Number: **WO 98/38296**

(43) International Publication Date: 3 September 1998 (03.09.98)

(21) International Application Number: PCT/IL98/00094

(22) International Filing Date: 26 February 1998 (26.02.98)

(30) Priority Data:
120337 27 February 1997 (27.02.97) IL

(71) Applicant *(for all designated States except US)*: GESHER–ISRAEL ADVANCED BIOTECS (1996) LTD. [IL/IL]; P.O. Box 98, 90830 Bet Neqofa (IL).

(72) Inventor; and
(75) Inventor/Applicant *(for US only)*: SHARON, Gil [IL/IL]; Cramim Street 48, 90805 Mevasseret Zion (IL).

(74) Agents: LUZZATTO, Kfir et al.; Luzzatto & Luzzatto, P.O. Box 5352, 84152 Beer–Sheva (IL).

(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, GW, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).

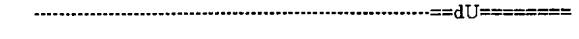
Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: METHOD FOR SIMULTANEOUS LIGATION OF MULTIPLE DNA FRAGMENTS

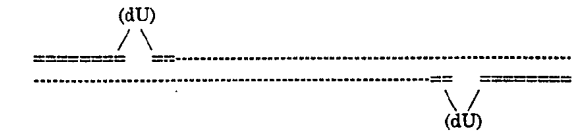
(57) Abstract

A method is described for the simultaneous assembly of two or more primer extension generated, double stranded DNA fragments, into a construct, comprising the steps of: 1) designing the primers such that the fragments subsequently produced by PCR, have 3' termini of 15 nucleotides or more being complementary to a second and third fragment, both having 3' termini of 15 nucleotides or more, complementary to each terminal of first said fragment; 2) providing in each of said primers and subsequently in each of said fragments, at least one dU residue in place of at least one dT residue, being positioned at least 15 nucleotides from the 5' end of said primer and subsequent fragments; 3) adding to the mixture of fragments the enzyme UDG (Uracil DNA Glycosylase) and a compound such as N,N,dimethylethylenedimethylamine, to a–purinate and generate single strand nicks at said dU residues, respectively, thereby to generate a 3' single strand overhang on each fragment to be joined thereto, being complementary to the other 3' single strand overhangs of second, third or more aforementioned fragments, and 4) allowing for all fragments thus generated to anneal into a single and stable DNA construct.

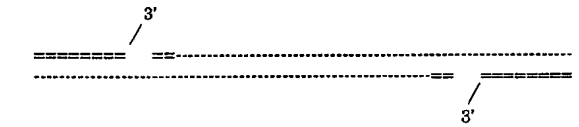
A. PCR fragment with a single dU residue that had been incorporated into each primer



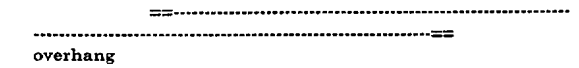
B. UDG a-purinates the dU residues



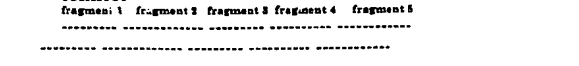
C. N,N dimethylethylenediamine nicks the a-purinated residues



D. Heating disconnects the nicked oligonucleotides from the double-stranded DNA molecule



E. Slow cooling allows fragments with complementary termini to interconnect



IDT Exhibit 1020.0348

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	**ES**	Spain	**LS**	Lesotho	**SI**	Slovenia
AM	Armenia	**FI**	Finland	**LT**	Lithuania	**SK**	Slovakia
AT	Austria	**FR**	France	**LU**	Luxembourg	**SN**	Senegal
AU	Australia	**GA**	Gabon	**LV**	Latvia	**SZ**	Swaziland
AZ	Azerbaijan	**GB**	United Kingdom	**MC**	Monaco	**TD**	Chad
BA	Bosnia and Herzegovina	**GE**	Georgia	**MD**	Republic of Moldova	**TG**	Togo
BB	Barbados	**GH**	Ghana	**MG**	Madagascar	**TJ**	Tajikistan
BE	Belgium	**GN**	Guinea	**MK**	The former Yugoslav Republic of Macedonia	**TM**	Turkmenistan
BF	Burkina Faso	**GR**	Greece			**TR**	Turkey
BG	Bulgaria	**HU**	Hungary	**ML**	Mali	**TT**	Trinidad and Tobago
BJ	Benin	**IE**	Ireland	**MN**	Mongolia	**UA**	Ukraine
BR	Brazil	**IL**	Israel	**MR**	Mauritania	**UG**	Uganda
BY	Belarus	**IS**	Iceland	**MW**	Malawi	**US**	United States of America
CA	Canada	**IT**	Italy	**MX**	Mexico	**UZ**	Uzbekistan
CF	Central African Republic	**JP**	Japan	**NE**	Niger	**VN**	Viet Nam
CG	Congo	**KE**	Kenya	**NL**	Netherlands	**YU**	Yugoslavia
CH	Switzerland	**KG**	Kyrgyzstan	**NO**	Norway	**ZW**	Zimbabwe
CI	Côte d'Ivoire	**KP**	Democratic People's Republic of Korea	**NZ**	New Zealand		
CM	Cameroon			**PL**	Poland		
CN	China	**KR**	Republic of Korea	**PT**	Portugal		
CU	Cuba	**KZ**	Kazakstan	**RO**	Romania		
CZ	Czech Republic	**LC**	Saint Lucia	**RU**	Russian Federation		
DE	Germany	**LI**	Liechtenstein	**SD**	Sudan		
DK	Denmark	**LK**	Sri Lanka	**SE**	Sweden		
EE	Estonia	**LR**	Liberia	**SG**	Singapore		

METHOD FOR SIMULTANEOUS LIGATION OF MULTIPLE DNA FRAGMENTS

Field of the Invention

The present invention concerns an improved method for joining a number of DNA fragments into a single, complex DNA construct in a desired predetermined directional manner in which method the actual joining of the DNA fragments is essentially a single step. More specifically, the present invention concerns an improved method for combining DNA fragments having regulatory function, such as, for example, promoters and the like, and DNA fragments encoding various proteins, such as, for example, enzymes (such as polyketide synthases), cytokines, hormones and the like, into a single construct for the purposes of cloning and expression of such a construct in a simple and rapid procedure, by way of which the various DNA fragments are connected in a predetermined direction and after joining of the fragments, there is no unwanted linker DNA between the joined fragments. Accordingly, the method of the present invention provides also for a seamless joining of DNA fragments.

Background of the Invention

Building DNA constructs is the core of genetic engineering. Building complex constructs requires time, money and highly skilled personnel. The construction is performed by fusing together specific DNA fragments in a desired way. The state of the art, as concerns cloning DNA fragments, has been published in a very large number of books, articles, patent applications, patents, and the like, and is usually readily available and known to all of skill in the art. For example, a comprehensive account of DNA cloning procedures is provided in the three volume text by Sambrook et al. (1989) entitled "Molecular Cloning - a Laboratory Manual", 2nd Edition, Cold Spring Harbor Laboratory Press. This extensive account of the prior art techniques for the combination of DNA fragments, cloning and expression thereof, is included herein by reference, in its entirety.

In accordance with the prior art, each of the DNA fragments is initially cleaved from larger DNA entities (DNA from plasmids, cDNA, genomic DNA) with enzymes called restriction enzymes. The desired fragments are then covalently connected together by use of an enzyme called ligase.

A given restriction enzyme is able to cleave DNA at a specific short sequence known as a restriction site. With more than two hundred different restriction enzymes which are commercially available, restriction sites situated at random up and downstream of a desired gene (or any other sequence of choice on a given DNA entity), can be identified and then cleaved by the respective enzymes. The DNA sequences which constitute restriction sites are mostly palindromic, between four to eight base pair long. Most enzymes cleave the DNA within the restriction site, leaving either "blunt" or "staggered" ends, depending on the specific enzyme. DNA with a "staggered" end has a short stretch (also known as "overhang") of single-stranded DNA between two to four bases long.

The enzyme ligase can connect, or ligate, two blunt ends of two DNA fragments (each fragment is a separate molecule) and form one longer fragment (longer molecule). Such a ligation is extremely inefficient. ligase can also ligate two molecules with staggered ends if the overhangs of these two molecules are complementary to each other. In fact, matching overhangs greatly improve ligation between molecules: the single-stranded DNA on one molecule has an affinity to the complementary single-stranded DNA on the other molecule. The staggered ends, also known as "sticky ends", or overhangs, form non-covalent connection with one another via hydrogen bonds. Since the overhangs produced by restriction enzymes are short, these connections are weak and unstable. Nevertheless, they rightly align the molecules long enough as to assist ligase in performing its task.

Sticky-end overhangs produced by the majority of the restriction enzymes consist of an overhang of two or four nucleotides. In theory, only fragments containing matching complementary overhangs can be connected to one another by ligase. In practice, illegitimate connections are a common occurrence. Due to the low affinity between such short single strand overhangs and consequently, due to the unstable nature of the connection between them, legitimate ligation is an inefficient process yielding a low amount of desired product. Furthermore, the palindromic nature of the sticky ends always results in undesired by-products such as "head-to-head" connections between identical molecules which further reduce the amount of the desired product.

Connecting DNA fragments with non-matching ends can be carried out by either blunting the ends by special enzymes or by adding very short, artificial DNA molecules called "linkers". These molecules are specifically designed to have an overhang that would match one fragment on one of their sides and another overhang, on the other side, that would match the other fragment. The addition of linkers further reduces the amount of the desired product.

Because of the inefficiency of the process, connecting more than two fragments at once is avoided if possible, and the building of DNA constructs is done one step at a time. Each step consists of several stages: first, desired DNA fragments are cut by restriction enzymes from larger molecules. Next, two DNA fragments are ligated to each other. Since the amount of the desired fragment is low, it has to be amplified, usually by transfecting cells of choice, such as, for example, bacterial cells. In order to do so, the product has to be circular DNA and has to contain certain components that will allow its amplification in bacteria. The third stage is therefore transformation of the DNA product into bacterial cells. Because of the high background of undesired product, a fourth, verification, stage has to be carried out. In this stage, the DNA from various bacterial clones

is purified and tested in order to distinguish between the desired product and all the others. Only afterwards can one proceed to the next step. The construction of a complex DNA molecule requires numerous such steps. In terms of time, the construction of sophisticated molecules may take anywhere between several weeks to several months, and sometimes it is not achieved at all.

Although the affinity between complementary overhangs of two or four nucleotides is low, both the affinity between complementary overhangs and the stability of the hydrogen bonds, once formed, greatly increase when the overhangs are longer. Several methods for connecting fragments employing overhangs of longer than 4 bases have been described.

One prior art method uses single strand extensions created by adding nucleotides at the 3' end of a DNA strand in a template-independent fashion (Roychoudhury, R. Gene Amplif Anal. 2:41-83, 1981). The enzyme used in this method, terminal transferase, will incorporate nucleotides at the 3' hydroxyl terminus of a double-stranded DNA fragment, thus creating a single-stranded tail. Since the enzyme uses the nucleotides randomly, the only way to ensure that the single-stranded tail will be complementary to a corresponding overhang created on a second DNA molecule, is to provide for each extension only one of the four nucleotides. The overhangs created with this method must therefore be homopolymeric, so that only four types of overhangs can be used, corresponding to the residues dA, dC, dG or dT. Since the overhangs created on both termini of a DNA fragment must be identical, cloning with this method is directionless and can only involve two fragments that are connected to each other at both ends, forming a circular molecule. Furthermore, the length of the overhangs cannot be specifically controlled. Finally, the method necessarily introduces an unwanted stretch of nucleotides into the final construct, the length of which cannot be

determined exactly, making the method unsuitable for the purpose of cloning into vectors where the reading frame must be preserved.

Various cloning kits sold in the market are based on the fact that the hydrogen bonds between overhangs 12 nucleotides long are stable enough to make the addition of ligase prior to transformation into bacteria unnecessary. The hydrogen bound fragments remain attached to one another during the transformation procedure and then become covalently bound by the bacterial ligation machinery. In fact, in a number of recent issued U.S. patents and a published PCT patent application: U.S. 5,137,814; U.S. 5,229,283; U.S. 5,334,515; and WO/18175 (PCT/US93/01965), all assigned to Life Technologies, Inc., U.S., there have been described methods for inserting DNA fragments of choice into a DNA vector of choice by way of generation of such 12 nucleotide long complementary overhangs, as well as ways of altering the nucleotide sequence of such DNA fragments at the time of their preparation and insertion into the vectors.

The above-mentioned U.S. patents and PCT application assigned to Life Technologies, Inc., also describe the use of so-called exo-sample nucleotides such as, for example, dU for the purposes of generating the desired complementary overhangs on the vector and the fragment to be inserted therein. Hence, these patents and application are also incorporated herein in their entirety by reference for all matters concerning the use of exo-sample nucleotides and the various associated polymerase chain reaction (PCR) procedures and the construction of single-stranded DNA primers, including such exo-sample nucleotides, to generate the fragment of choice to be inserted into the vector.

However, the above-mentioned prior art as a whole, and the specific disclosures according to the above-mentioned U.S. patents and PCT application, all have drawbacks when it is desired to use such known

technology to combine two or more DNA fragments together. Of these drawbacks, one of the main ones is the fact that for the generation of the complementary overhangs, usually repeats of three nucleotides are employed, each repeat containing a single EXO-sample nucleotide, in the primer sequence to generate the desired fragment. As a result, the generated fragment will have at its terminal ends an additional segment of DNA corresponding to these repeats. This additional segment of DNA, also called "linker" DNA, has the length of the repeats. Subsequently, when the fragment is inserted into the vector, the two junction sites between the fragment and the vector will contain this extra linker DNA. Hence, if it was desired, for the purpose of expression of the fragment, to directly fuse the original fragment DNA to a desired sequence within the vector, by these prior art techniques such a direct fusion is not possible, because of the presence of the linker segment. Such a linker segment may therefore be detrimental to the proper expression of the fragment within the vector, and hence may have to be removed after the initial construction of the fragment within the vector, a procedure requiring an additional input of time and resources.

Similar methods have been described by Rashtchian et al. (Anal. Biochem. 206, p. 91-97, 1992), who described 12 nucleotide overhangs generated by uracil DNA glycosylase (UDG) to achieve high-efficiency cloning of single inserts into a vector. Using such complementary overhangs of only 12 nucleotides in length represents a further drawback of the above earlier methods, namely, such short overhangs often lead to a relatively low efficiency of joining between two fragments and hence, when three or more of such fragments are desired to be joined, these methods do not provide a useful yield of the desired products. Thus, using such 12 nucleotide overhangs with the prior art methods, it is likely that when it is desired to join three or more DNA molecules together, this would have to be done in a step-by-step process in which, at first, two molecules would be joined, the so-joined molecule then isolated and purified, and afterwards, the

third fragment would be added thereto, and so on for any additional fragments to be joined. As a result, these prior art methods are also time-consuming and require a significant input of resources.

One of the aims of the present invention is to provide a method by which two or more fragments may be joined together in a specific predetermined directional manner, very efficiently, in which the joining step is essentially a single step and in which the junction sites between the various fragments are seamless, namely, do not contain any linker DNA segments to provide a fully in-phase joining of one fragment to the next.

In a copending application of the same applicant herein, filed together with the present application and identified as Attorney Docket 4190/96, (Israeli Patent Application No. 120338) there is described a method for joining DNA fragments in an efficient manner, by way of a rapid procedure, and this by providing long terminal overhangs of at least 15 nucleotides on each of the fragments to be joined. It is another purpose of the present invention to provide a method for joining three or more DNA fragments in an efficient manner, and at the same time ensuring that there is only complementarity between the terminal overhangs of any two fragments which are to be joined together, ensuring that only the predetermined desired order of joining between the fragments is achieved.

Other aims and aspects of the present invention will be readily apparent from the foregoing disclosure.

It should be noted that all of the various terms, procedures and the like, used herein throughout, unless otherwise indicated, are all well known terms in the art, known by all of average skill in the art. Thus, for example, terms such as nucleotides, primers, PCR, and the like, are readily known to all of skill in the art and are well defined in all standard texts and publications, for example, the above-noted Sambrook et al. The

above-noted U.S. patents and PCT application, incorporated herein by reference, are less known. The term "exo-sample nucleotides", of which dU is an example, as used herein throughout, is a relatively new term of the art, but is well defined and described in the above-noted reference U.S. patents and PCT application.

Summary of the Invention

In accordance with the present invention, there is provided an improved method for combining two or more DNA fragments together into a multi-DNA fragment assembly, which overcomes the above-mentioned drawbacks of the prior art. By the method of the present invention, the DNA fragments are joined in an essentially single-step joining reaction in a desired, predetermined order and in a seamless fashion.

Accordingly, the present invention provides a method for the simultaneous multi-DNA fragment assembly of two or more double-stranded DNA fragments produced by a primer extension reaction, particularly the polymerase chain reaction (PCR), comprising the steps of:

(a) providing for each DNA fragment to be joined to a second DNA fragment and to a third DNA fragment, a pair of primers for the primer extension procedure thereof, wherein one of said two primers has a portion which is complementary to one terminal portion of the strand of said fragment, and the second of said two primers has a portion which is complementary to the other terminal portion of the strand of said fragment, and wherein the second portion of at least one of the said primers is complementary to the terminal portion of a different fragment to be joined to said first fragment in a specific positioned relationship;

(b) providing, in each of said primers, at least one dU residue or the like in place of at least one dT residue of the original primer sequence, the first of said at least one dU residue(s) being positioned at least 15 nucleotides from the 5' end of said primer;

(c) producing each said DNA fragment by the primer extension reaction procedure;

(d) carrying out in any suitable order the steps of: 1. mixing the produced DNA fragments 2. adding the enzyme UDG to apurinate said dU residue and a second enzyme or a compound such as N,N-dimethyl ethylenediamine to generate single-strand nicks at said apurinated dU residue, said UDG and second enzyme or compound being in amounts sufficient thereby to generate at the end of substantially all fragments a 3' single strand overhang which is complementary to another 3' single strand overhang or another fragment to be joined thereto; and

(e) providing suitable conditions and allowing sufficient time for the specific joining of said fragments to generate a multi-DNA fragment assembled product.

Of course, if two fragments are to be joined to a circular construct, then the above termini of said second and third DNA fragments are termini of the same fragment. Furthermore, it will be readily appreciated by all of skill in the art that it is possible, by the above method of the invention, to join a fragment at one end only, i.e., it is not necessary to join both ends of any or all of the above fragments.

The invention also provides the above method wherein the DNA mixing step is carried out before the UDG enzyme addition step. The invention further provides the above method wherein the UDG enzyme addition step is carried out before the DNA mixing step.

Still further, the invention provides the above method wherein step (e) includes the removal of the nicked oligonucleotides from the reaction.

One embodiment of the above method of the invention is a method wherein the said 3' complementary overhangs on said DNA fragments are

generated by the addition to said fragments, the enzyme UDG to apurinate said dU residue and a second enzyme selected from endonuclease III (Endo III) or endonuclease IV (Endo IV) to generate single-stranded nicks of said apurinated dU residue to provide said 3' overhangs.

Another embodiment of the above method of the present invention is a method wherein said 3' complementary overhangs of said DNA fragments are generated by the addition to said fragments, the enzyme UDG to apurinate said dU residue and the compound N,N-dimethyl ethylenediamine to generate single-stranded nicks at said a-purinated dU residue to provide said 3' overhangs (for the role of this, and other chemical reagents in forming nicks, see McHugh and Knowland. (1995) Nucleic Acid Res., 23, 1664-1670).

The joining of the fragments via the 3' complementary overhangs is facilitated by:

(a) heating the reaction mixture to a temperature suitable to allow dissociation of the oligonucleotides at the termini of the fragment after said apurination of the dU residue and generation of nicks at the apurinated dU residue, and incubating said mixture;

(b) keeping the mixture of (a) at a temperature suitable to maximize correct connections between complementary overhangs; and

(c) optionally, adding ligase to said mixture to facilitate covalent joining of the DNA strands.

In accordance with any of the above methods and embodiments according to the present invention, the DNA fragments to be joined are selected from two or more DNA fragments having regulatory function such as, for example, promoters, enhancers, terminators, ribosome binding sites and the like; and DNA fragments encoding proteins such as, for example,

enzymes (such as citrate synthases, polyketide synthases, and succinyl-CoA-synthetase), cytokines, hormones and the like.

Furthermore, in accordance with any of the methods or embodiments of the present invention as noted above, there is also provided a method wherein one or more of the DNA fragments to be joined is a mutant fragment having been subjected to site directed mutagenesis during its preparation. Thus, for example, if it is desired to join a promoter carried by one DNA fragment to another DNA fragment encoding the protein of choice, it is possible to produce one or more site directed mutations in the promoter sequence and this by introducing into the primers made to synthesize this promoter sequence one or more specific site-directed mutations to provide a mutated promoter sequence, or similarly, by introducing mutations into the primers made to synthesize the protein sequence, it is possible to generate a site-directed mutated protein sequence. This is particularly useful if, for example, it is desired to provide a promoter from a particular source but in which its activity is either increased or decreased with respect to the naturally occurring promoter or, likewise, if it is desired to alter the biological activity of the protein as compared to its naturally occurring form. Thus, by virtue of the present invention, it is possible not only to join fragments of choice together, but at the time of preparing such fragments, it is possible to also mutate the sequences of the fragments of choice. In addition, it is also possible to generate random mutations in defined sequences by employing mutagenic PCR condictions on at least one of the fragments which are produced.

According to the present invneiton, there is also provided the above method, wherein at least some of said primers for the primer extension production of said DNA fragments are biotinylated at their 5' ends, and wherein when said enzyme UDG and said second enzyme or compound are added to said mixture of fragments, there is also added streptavidin, either

free or bound to beads, whereby single-stranded oligonucleotides containing at their 5' end a biotin moiety generated following the formation of said single-strand nicks at said apurinated dU residues are bound via a biotin-streptavidin connection and are effectively removed from the joining reaction.

The present invention also provides a primer for use in any of the above-mentioned methods or embodiments thereof of the present invention, comprising at least one dU residue in place of at least one dT residue, the first of said at least one dU residue(s) being positioned at least 15 nucleotides from the 5' end of the primer. In view of the above-mentioned, there is also provided, as an embodiment of the primer of the invention, a primer wherein the primer is biotinylated at its 5' end.

The present invention also provides an assembled DNA construct whenever prepared by any one of the methods or embodiments thereof as set forth hereinabove, and wherein said construct has been assembled from the joining together of a plurality of DNA fragments according to the invention.

Still further, the invention provides a DNA fragment comprising an overhang of at least 15 nucleotides or an end portion suitable to be converted into such an overhang. The invention also provides said DNA fragment, for use in the above method according to the invention.

Illustrative, but non-limitative, examples of the above assembled DNA construct of the invention include constructs which have been assembled by the joining together of three DNA fragments; constructs which have been assembled by the joining together of four DNA fragments; constructs which have been assembled by the joining together of five DNA fragments, and constructs which have been assembled by the joining together of eight and more DNA fragments. Furthermore, other examples of the assembled DNA constructs of the invention include constructs as indicated above,

which are in the form of a linear DNA molecule or which are in the form of a closed circular DNA molecule.

Other aspects and embodiments of the present invention will be readily apparent from the following detailed description of the invention.

Brief Description of the Drawings

- Fig. 1 (A-E) is a schematic flow-diagram depicting one embodiment of the preparation of the fragments, their treatment and their joining, in accordance with the method of the invention, as detailed in the Examples. === depicts the primers, ---- depicts a single strand of the DNA fragment and /__\ depicts the a-purinated site.

- Fig. 2 is a schematic illustration of a plasmid prepared in accordance with the method of the present invention, as detailed in Example 1 and 2, wherein the open bars depict the Tet fragment which may be synthesized as a single fragment from primers SEQ. ID NO. 1 (also designated 31162) and SEQ. ID NO. 5 (also designated 30402), or in the form of two subfragments, TetA and TetB from primers SEQ. ID NO. 1 and SEQ. ID NO. 4 (also designated 27341), and SEQ. ID NO. 3 (also designated 25595) and SEQ. ID NO. 5, respectively; the dark bar depicts the Amp + ColE1-ORI fragment, which is synthesized as a single fragment from primers SEQ. ID NO. 8 (also designated 27342) and SEQ. ID NO. 6 (also designated 3885); and wherein the hatched bar represents the Cm fragment, which may be produced as a single fragment from primers SEQ. ID NO. 2 (also designated 3597) and SEQ. ID NO. 7 (also designated 4144), or as two separate subfragments, CmA and CmB from primers SEQ. ID NO. 2 and SEQ. ID NO. 9 (also designated 27343), and SEQ. ID NO. 10 (also designated 25596) and SEQ. ID NO. 7, respectively, and wherein the primers are indicated in the figure with arrows denoting their direction of synthesis, the arrowheads being the 3' end of the primer.

Fig. 3 is a schematic illustration of a plasmid constructed out of eight fragments in accordance with the method of the present invention, as detailed in example 3.

Detailed Description of the Invention

The present invention concerns an improved method for combining two or more DNA fragments together into a single DNA construct, by which method the fragments are joined in an essentially single-step joining reaction, in a desired, predetermined order and in a seamless fashion, namely, no linker DNA is inserted between the joined fragments.

In the present invention, each primer is synthesized by standard, automated single-stranded (oligonucleotide) DNA synthesis and has two parts. A 3' part being complementary to the fragment to be produced, and a 5' part complementary to the terminal portion of the strand of another fragment to be specifically joined to this first fragment. Thus, the primers, in accordance with the present invention, essentially correspond to the desired predetermined junction region at those ends of the two fragments to be joined. In this way, for each fragment generated, a unique pair of primers is used.

In the above primers, during their synthesis there is incorporated at least one dU residue in place of at least one dT residue of the original sequence of the fragment, this dU residue being placed at least 15 nucleotides from the 5' end of the primer. In this way, as is fully detailed below, when the so-generated fragment is to be joined to the other so-generated fragments, upon addition of the enzyme UDG and the compound N,N-dimethyl ethylenediamine, or another enzyme which is the functional equivalent of this compound, at the ends of each fragment there is generated a 3' single strand overhang which is complementary to only one other such overhang present only on the fragment to be joined, to ensure that only the right fragments will join.

According to the present invention, the use of dU and the associated UDG and N,N-dimethyl ethylenediamine for the generation of the specific overhangs, is the preferred choice, as all of these reagents are readily available at relatively low costs. However, any other suitable exo-sample nucleotides, and corresponding reagents to remove them and generate single-stranded overhangs as are known in the art may also be used in accordance with the present invention. See, for example, the above-referenced patents assigned to Life Technologies, Inc., in which there is described dU, UDG and various other exo-sample nucleotides and reagents for their removal.

One such example of an alternate procedure is the use of UDG to cause apurination of the dU residues in the PCR - generated DNA fragments followed by the use of a second enzyme, which may be either endonuclease III (Endo III) or endonuclease IV (Endo IV), instead of N,N-dimethyl ethylenediamine, to nick the DNA fragments at the apurinated dU residue and thereby to generate the desired 3' overhangs on the DNA fragments.

In accordance with the present invention, overhangs of at least 15 nucleotides long are generated to ensure stable and efficient joining between the various fragments. In fact, in a series of experiments in which overhangs of only 12 nucleotides was used, disclosed in a copending patent application of the same applicant filed on the same day as this application (Attorney Docket 4191/96, Israeli Patent application No. 120339), the specification of which is incorporated herein by reference, it was found that when using fragments with overhangs of 12 bases, only low efficiency joining between two fragments could be achieved, and no joining between three or more fragments into a single construct could be achieved. Hence, in accordance with the present invention, it has been found that such overhangs must be at least 15 nucleotides long to ensure the joining of two or more fragments together in an efficient manner.

In accordance with the present invention, it is of course also possible to generate one or more site-specific mutations within the one or more fragments to be joined together and this by standard procedures, in which the various above-noted primers, when generated, are generated having a specific nucleotide substitution, deletion or addition at a selected site within the primer. Subsequently, the fragment generated from such a primer will have the pre-selected site-specific mutation. Likewise, it is also possible to generate regions of either high mutation rates or very low mutation rates by amplifying fragments using mutagenic PCR protocols or high-fidelity enzymes, respectively.

As regards the various DNA fragments to be joined in accordance with the method of the present invention, these fragments may encode any DNA molecule of choice. For example, the fragments may encode various regulatory sequences such as, for example, promoters, enhancers, terminators or the like. The fragments may also encode various proteins having various biological activity of pharmaceutical or veterinary importance, for example, various metabolic enzymes (such as polyketide synthase), hormones, cytokines, and the like. Hence, it is possible to generate, in accordance with the present invention, new chimeric promoters having improved or decreased activity and this by joining two fragments encoding parts of promoters from various sources. Likewise, it is possible to generate a wide range of chimeric structural and regulatory proteins, for example, chimeric cytokine molecules, receptors, enzymes and the like, of improved or other desired biological activity, by combining fragments encoding different domains of such molecules from different sources. Likewise, a multi-fragment assembly may be devised in which a new desired promoter is directly connected to a new desired DNA molecule encoding a biologically active protein in a single new construct for the purposes of enhanced expression of this new desired protein, then the construct is used to transfect/transform suitable cells of any organism of

choice, for example, prokaryotic or eukaryotic cells such as bacterial or yeast cells, respectively, or mammalian, insect or any other eukaryotic cells. Likewise, such a multifragment assembly can also be in the form of a modified bacterial or animal virus carrying one or more genes of choice for the purposes of infecting prokaryotic or eukaryotic cells of choice, and thereby introducing into these cells the gene(s) of choice.

In accordance with the present invention, the PCR procedure is essentially the well known, now standard, procedure, which may be augmented by using recently described new high-fidelity DNA Polymerase enzymes, as well as newly developed, improved automated machinery for this purpose. Hence, any PCR procedure and reagents for use therewith may be utilized in accordance with the present invention to generate the very specific fragments which are to be combined together.

Hence, in accordance with the present invention, one may readily prepare a DNA construct of choice, namely, a "custom-made" DNA construct in which any desired DNA fragment encoding any desired structural or regulatory function, can be joined in a seamless fashion to other such DNA fragments. As the procedure in accordance with the present invention is both rapid and simple to perform, it is therefore possible to prepare any DNA construct of choice, be it a linear DNA molecule for insertion into cells directly by known techniques, or a circular DNA molecule to be used as a vector for transfecting/transforming cells of choice, or a linear construct for insertion into another vector of choice, and any other such purpose readily apparent to any of skill in the art. The list of possible constructs which may be prepared in accordance with the present invention is essentially endless, as the only limitation on the preparation of constructs in accordance with the present invention is the availability of sequence information for the ends of each of the fragments to be joined.

The present invention will now be described in more detail in the following

non-limiting examples and the accompanying drawings:

The General Procedure

The fragments which are to be attached to each other in a directional fashion are preferably prepared by utilizing the well-established Polymerase Chain Reaction (PCR) procedure, this being a standard procedure of the art. This provides for suitable amounts of the fragments.

In this way, the desired fragments of the original DNA are greatly amplified and by virtue of the use of pre-selected specific primers in the PCR procedure, the 5' and 3' ends (termini) of the PCR-prepared fragments will have the desired pre-selected sequences which will ultimately provide for the directional attachment of the fragments to yield the DNA molecule of choice, in which all of the fragments have been attached to each other in the pre-selected order.

Furthermore, by virtue of the necessity to use pre-selected primers in the PCR procedure, this necessity is also convenient for the purposes of the present invention, namely, into these primers, one or more dU residues may be incorporated in place of one or more dT residues, these one or more dU residues being the sites at which further specific treatment will provide for each of the various fragments having the desired single-stranded DNA overhangs (cohesive ends) that will be complementary only to the overhangs on the other fragments to which each fragment is to be connected, thereby ensuring the directionality of the connection.

In accordance with the present invention, it is preferred to use more than one such dU insertion into each primer in those sites of the primer that it is desired to treat further following the PCR procedure. However, it must be noted that the number of dU insertions is entirely dependent upon the chosen (predetermined) primer sequence, not all primer sequences of

choice will necessarily have many dT residues for replacement by dU residues. Further, it is also desirable to have such dU replacements that are spread out along the length of the primer, but here too, the number and spread of dU replacements is dependent upon the primer sequences of choice. In any event, at least one dU insertion in place of one dT residue is essential in each primer sequence to ensure the successful operation of the method of the invention. According to one embidiment of the present invention, following the PCR procedure, all of the fragments are mixed together and subjected to the preferred treatment with UDG and N,N-dimethyl ethylenediamine, which results in the apurination of the dU residues, followed by cleavage (nick) at the 3' side of each apurinated residue to create a single strand nick. After this treatment, the temperature is raised to 75°C, resulting in the denaturation and disconnection of the double-stranded structures in between dU residues and up to the dU residue most distal from the 5' end of the primer (the most 3' dU residue), or when only one dU residue was inserted, then the denaturation and disconnection is up to the position of this dU residue, thereby yielding fragments having long specific 3' overhangs. These 3' overhangs are specific both as regards their sequence, and as regards their length, the length being determined by the original placement of the most distal (most 3') dU residue in the primer. In the present invention, overhangs of about 15 nucleotides and longer are used.

The above predetermined annealing can be done by a slow cooling, (starting from 75°C and down to about 20-30°C), which facilitates specific and directional annealing of the different fragments to each other, by virtue of specific connections between the complementary 3' overhangs of the neighboring fragments. Alternatively, the annealing of the fragment can be done at 37°C or at any temperature desired when more than one dU residue is inserted into the primer and the inserted dU residues are spaced along the primer, their subsequent removal results in the generation of very short released oligonucleotides which compete less

successfully for annealing to the aforesaid strand (to which they were originally annealed) with the complementary overhangs of the fragments to be joined. Further, at the temperatures at which the various fragments join each other by annealing of their complementary 3' overhanging portions, the various released oligonucleotides are less efficient in reannealing, the shorter the oligonucleotide, the less the annealing efficiency. Most preferably, the oligonucleotides that are generated 3' to the apurinated dU residue or residues are removed from the mixture. Two examples of how to do it are included herein, one using biotin (Example 2), the other using a different procedure (Example 3).

As a final step in the above procedure, ligase (e.g., T_4 ligase; of course any ligase, other than T4 ligase, may be used) may be added to covalently connect all of the fragments and thereby complete the formation of a single DNA molecule without any nicks. This ligase step is optional in view of the fact that by the preferred use of long overhangs at the termini of the fragments, the connection between the matching complementary overhangs, i.e., the hydrogen bonding therebetween, is of such a nature that even without the final covalent connection with a ligase, the hydrogen bonding is strong enough to hold all of the fragments together. As will be described below, fragments combined in this fashion remain intact in a linear form even when subjected to gel electrophoresis, or when fragments are so connected to form a circularized DNA molecule, such a circularized molecule is structurally stable enough to be used to transfect cells to provide transformed cells containing this circularized molecule, which in these cells become ligated by the endogenous cellular ligases.

For the purpose of simplicity, it should be noted that in the above general procedure, when, for example, it is desired to connect five fragments together to form a linear molecule, then, the specific primers used in the PCR step are designed in such a way that the ends that are to be connected to each other are homologous (i.e., have perfectly

complementary 3' overhangs, once these overhangs have been generated as noted above and below). For example, if a linear DNA construct is to be made from five fragments numbered 1-5 and for each fragment, for the purposes of illustration, we designate the sense direction to be "left to right" such that when joined fragment number 1 will be the extreme left-hand side and fragment 5 will be the extreme right-hand side of the combined molecule (or by convention, the (+) sense strand will begin with its 5' terminus at the extreme 5' end of fragment 1 and end with its 3' terminus being at the extreme end of fragment number 5; and the (-) non-sense strand will have its extreme 5' end at the extreme end of fragment 5 and its 3' end at the extreme end of fragment number 1).

Hence, to make this construct of five fragments, the right terminus of fragment 1 should be homologous to the left terminus of fragment 2, the right terminus of fragment 2 should be homologous to the left terminus of fragment 3, the right terminus of fragment 3 should be homologous to the left terminus of fragment 4, and the right terminus of fragment 4 should be homologous to the left terminus of fragment 5. With such homology, once the 3' overhangs at the ends of each fragment are generated (as noted above and below), there will thus be perfect complementarity between the 3' overhang at the right terminus of fragment 1 and the 3' overhang at the left terminus of fragment 2 to provide for a specific directional joining of these two fragments by complementary interaction or joining between the two 3' overhangs, and likewise for the joining of the above-mentioned other termini of the various fragments to be joined. In such a construct, therefore, the left terminus of fragment 1 and the right terminus of fragment 5, i.e., the extreme ends of the molecule, should not have any homology to each other or to any of the other left or right termini of all of the various fragments, in this way ensuring that the so-produced molecule will be linear and will not be capable of circularization or undergoing any other inter-fragment interactions which would disturb the 1-2-3-4-5 desired configuration to be formed.

However, when it is desired to make such a molecule which is circular, then of course the right terminus of fragment 5 should be homologous to the left terminus of fragment 1, thereby ensuring that circularization occurs only by connection between fragment 5 and fragment 1. As also arises from the above-mentioned, all of the left and right termini of each of the fragments which are not designed to be connected must be non-homologous, this condition being relatively easy to meet in view of the standard technology available for the automated synthesis of primers, the sequence of which is easy to determine.

Thus, it is apparent from the above general procedure that the method of the present invention is essentially a single-step procedure in which all of the reaction components are present in a single reaction vessel in which all of the various reactions and treatments are carried out, thereby greatly simplifying the overall process and providing an end-product that is essentially only the desired product which can be readily obtained and further utilized.

It should of course be noted that the initial PCR amplification of the fragments which it is desired to combine should preferably be carried out in separate reaction vessels for each specific fragment to ensure the fidelity of the PCR products.

However, this requirement for separate PCR amplifications of the selected sequences is now also a simple laboratory procedure in view of the widely available automated apparatuses for carrying out PCR procedures, which can handle a large number of separate samples at the same time. Also, as mentioned above, the production of the desired PCR primers which will carry the preferred more than one dU residue, or at least a single dU residue is also carried out by automated machinery which allows for the generation of large numbers of primers of high fidelity and purity

simultaneously in a very short period of time (for example, there are machines which can simultaneously produce 96 different primers at a cost of only about 10-30 cents per base per primer). Hence, it is possible to readily produce almost any desired primer carrying the preferred more than one or at least single dU residue in the preferred positions or position within the primer so as to ultimately result in an at least 15 nucleotide overhang and preferably about 20 or more nucleotide overhang at the ends of the fragments to be joined.

A schematic representation of the above-mentioned general procedure is set forth in Fig. 1. In parts A-D, there is illustrated how a PCR fragment having only a single dU residue (for simplicity of illustration) may be treated with UDG and N,N-dimethyl ethylenediamine to yield a PCR fragment with 3' overhangs at its two termini. In parts A-B of Fig. 1 there is shown schematically the preferred position of the dU residue in the end-product of the PCR procedure, i.e., the desired PCR-produced fragment which has incorporated at its terminal ends the sequence defined by the two primers used in its preparation, and in which the dU residue is inserted distal from the terminal ends of the fragment so that once it is removed by the treatment with UDG and N,N-dimethyl ethylenediamine, schematically shown in parts C and D of Fig. 1, large single-stranded terminal overhangs will be formed, these preferably being at least 15 nucleotides in length, and more preferably about 20 or more nucleotides in length. When the primers are prepared it is of importance to make the replacement of the dT residue by the dU residue at a site in the primer which will be about 20 base pairs (bp) distal to the 5' end of the sequence of the desired fragment, so that when incorporated, this dU residue will be about 20 bp distal to the 5' terminus of the PCR end-product.

It should be noted that for the purposes of illustration in Fig. 1, parts A-D, there has been shown only the general outline of the UDG and

N,N-dimethyl ethylenediamine treatment of dU-containing PCR fragments for a single exemplary fragment, all the other fragments that are to be joined together essentially having the general overall structural configurations with the dU inserted distal to the terminal ends of the fragments. In part E of Fig. 1, there is shown schematically the linear alignment of five fragments which are to be connected to each other in the predetermined order of fragments 1-5, this being facilitated by the homology at the ends of the fragments which are to be specifically joined to each other, namely, the so-called right-hand terminus of fragment 1 is homologous to the left-hand terminus of fragment 2, the right-hand terminus of fragment 2 is homologous to the left-hand terminus of fragment. 3, and so on, so that only connections in the specific order of fragments 1-5 can be obtained by the specific complementary interactions between the complementary 3' overhangs. The length of the overhangs is playing a critical role in determining the strength of the hydrogen bonding between the complementary overhangs which will connect the various fragments and hence the overall high specificity of the order in which these fragments will be connected to each other.

Materials and Methods

All of the fragments which were joined together, as set forth in the following examples, were first individually prepared by the PCR procedure using a commercially available PCR apparatus (Robocycler Gradient 96™, Stratagene, U.S.A.) according to the manufacturer's instructions.

The concentrations of the reagents used in all of the PCR procedures, were as follows:

Reagent	Concentration	Volume in Reaction Vessel
DNA*(see below)	0.1 µg/µl	0.5 µl
Primer 1	10 pmol/µl	10 µl
Primer 2	10 pmol/µl	10 µl
dNTP**(see below)	2.5 mM (each)	16 µl
10X Buffer	--	20 µl
H_2O	--	144 µl
TaqDNA Polymerase***(see below)	5 U/µl	0.8 µl
		200 µl - Final Volume

The temperature regime for all of the PCR procedures was as follows, in accordance with the manufacturer's instructions:

94°C for 60 sec., 40°C for 80 sec., 72°C for 240 sec., this temperature regime being carried out for 30 cycles and then followed by a final 72°C for 600 sec. After the last 72°C incubation, the reaction vessel was cooled to 6°C.

*The DNA which was used as the template in all of the following examples was either the plasmid pBR322 or the plasmid pACYC184, both commercially available and for both the full sequence and restriction maps are available. This will be described in more detail in Example 1 below.

**dNTP is a mixture of dATP, dTTP, dCTP and dGTP, in equimolar amounts, all obtained from commercial suppliers and used in accordance with the manufacturer's instructions (Boehringer-Mannheim, Germany). It should be noted, as mentioned above, that during the PCR amplification of the fragments, only these normal nucleotides are employed and that dUTP is not inserted at all, this having been already inserted into the above-noted primer 1 and primer 2 prior to commencement of the PCR procedure.

***The Taq DNA polymerase was also obtained from a commercial supplier and used in accordance with the manufacturer's instructions (Boehringer-Mannheim, Germany).

Further, it should be noted that the above 10X Buffer is the usual buffer employed in the PCR procedure and is also purchased from a commercial supplier and used in accordance with the manufacturer's instructions (Boehringer-Mannheim, Germany).

Example 1

Preparation of a circularized plasmid by the directional connection of 3-5 fragments constituting the plasmid

Overview

The plasmid to be constructed was designed to have three or more different regions, each to be prepared separately by PCR amplification and then joined in a specific directional fashion to provide a circularized plasmid as the end-product. This required the initial preparation of specific primers, the sequence of which was to provide the basis on which the various PCR fragments, constituting the various regions of the plasmid, would be able to combine with each other to ultimately connect all the regions in the specific directional manner. The sites within the primer sequences which would provide for the interconnection between the various PCR fragments were determined irrespective of any restriction enzyme sites or any other DNA sequence features at these connection sites in order to demonstrate that, in accordance with the method of the present invention, any DNA sequence at any particular site within a DNA molecule can be utilized as a connection site by preparing the necessary specific primers to provide for this connection site.

In Fig. 2, there is shown schematically the plasmid that was designed and produced by the method of the present invention. This plasmid carries

three independent antibiotic resistance genes, for resistance to ampicillin (Amp gene, or hereinafter Amp); tetracycline (Tet^r gene, or hereinafter Tet); and chloramphenicol (Cm^r gene, or hereinafter Cm). The plasmid also carries the ColE1 origin of replication (ColE1-ORI), which in this specific instance is situated next to the Amp gene, thus these two entities constituting a single region of the plasmid. The Tet and Cm genes constitute additional separate regions. Such a plasmid is capable of being replicated in a host cell and will endow the host cell, successfully transformed therewith, with resistance to all three types of antibiotic. Accordingly, it is also possible to readily select for those host cells transformed by this plasmid by growing the transformed cells in the presence of one of the antibiotics and then to screen for the resistance to the other antibiotics. This therefore provides for a genetic analysis of the end-products of the construction procedure at the *in vivo* level. Further verification step was done by testing the plasmids in the transformed colonies by PCR, as detailed below.

To prepare the above plasmids, a number of constructions were made in which 3, 4 and 5 individual fragments were produced by PCR and then combined to yield the plasmids. Thus, when the plasmid was to be constructed from three fragments, the above-noted three regions of the plasmid were PCR synthesized using appropriate primers so that three PCR fragments together having all three regions would be obtained. When the plasmid was constructed from four fragments, other appropriate primers were used to yield four PCR fragments encompassing the three regions of the plasmid. Likewise when the plasmid was constructed from five fragments, other appropriate primers were used to yield five PCR fragments encompassing the three regions of the plasmid. The results (not shown) of these constructions using three and four fragments showed that the so-constructed plasmids were fully active *in vivo* when used to transfect host cells, there being obtained a large number of resulting transformed cells having resistance to all three antibiotics, this large

number being indicative that the transformation efficiency of the so-constructed plasmid was high, and further, that the so-transformed cells actively expressed the antibiotic resistance genes carried by the plasmid.

For the purposes of exemplifying the method of the present invention, it has been chosen not to show the building of the simpler constructs and to demonstrate only the more difficult procedure for successfully combining five PCR fragments which together encompass the above three regions of the plasmid that was constructed and used to transform host cells with high efficiency.

For this five fragment construction, the following was carried out:

(i) Preparation of the specific primers for the PCR amplification of the desired five fragments of the plasmid.

As shown in Fig. 2, and as outlined above, it was desired to construct a circularized plasmid having three regions:

(a) a Tet region;

(b)an Amp + ColE1-ORI region; and

(c) a Cm region.

These three regions were to be connected in a highly specific manner, namely, the Tet region was to be connected between two ends of Cm and Amp + ColE1-ORI regions, while at the other two ends, the Cm and Amp+ColE1-ORI regions were to be connected to each other. Thus, it was necessary to connect one end (upstream) of the Cm region to one end (upstream) of the Tet region, to connect the other end of the Tet region to one end (downstream) of the Amp + ColE1-ORI region, and to connect the other ends (downstream of Cm and the upstream end of Amp + ColE1-ORI region) to each other.

To achieve this task, five separate fragments were designed and produced by PCR:

(1) A large fragment of around 1700 bp which carries the Amp gene and the ColE1-ORI sequence;

(2) Two fragments together constituting the Tet region being designated "Tet A" fragment of around 850 bp containing the upstream end of the Tet gene; and "Tet B" fragment of 600 bp containing the downstream end of the Tet gene; and

(3) Two fragments together constituting the Cm region, designated "Cm A" fragment of around 430 bp containing the upstream end of the Cm gene; and "Cm B" fragment of around 270 bp containing the downstream end of the Cm gene.

For the production of these fragments, it was necessary to prepare specific primers. The above three regions of the plasmid to be constructed are well known, as they are derived from two commercially available plasmids, namely pBR322 and pACYC184 (New England Biolabs, U.S.A.). The full sequence and maps of the various regions of these plasmids are known. The full sequence of these plasmids can be accessed from GenBank database under accession Nos. J01749 and X06403, respectively. For the preparation of PCR fragments containing the Tet and Amp + ColE1-ORI regions the pBR322 sequence and plasmid (as a template) was utilized; and for the Cm region, the pACYC184 sequence and plasmid (as a template) was used.

In this way, using standard automated procedures to produce polynucleotide oligomers (Applied Biosystems, U.S.A.), the following primers were synthesized:

(1) Upon a study of the sequences of the Cm gene provided with the commercially obtained pACYC184 plasmid and of the Tet gene provided with the commercially obtained pBR322 plasmid, it was

decided that the desired region containing the junction between the Cm and Tet genes should be of the following sequence:

junction point
5' CCATTTTAGC TTCCTT AGCTCCTGA↓TTCTCAT GTTT GACAGCTTATCATC
3' GGTAAAATCGAAGGAATCGAGGACT↑AAGAGTACAAACTGTCGAATAGTAG
<---------- | ---------->
Cm upstream (5' region) junction point Tet upstream..(5') region

Hence, the primers which will provide the overlap between the Cm and Tet genes and allow their connection to each other were designed as follows:

(a) Primer designated SEQ. ID NO. 1, also designated 31162, as shown schematically in Fig. 2, having the sequence:

5' A G C T C C T G A↑T T C T C A T G T T U G A C A G C T T A T C
sequence complementary ↑ beginning of Tet 5' (sense) sequence
to the Cm end junction point

(b) Primer designated SEQ. ID NO. 2, also designated 3597, as shown schematically in Fig. 2, having the sequence:

5'A A A C A T G A G A A↑T C A G G A G C U A A G G A A G C T A A A A T G
sequence complementary ↑ beginning of Cm 5' (sense) sequence
to the Tet 5' (sense) end junction point

As depicted schematically in Fig. 2, the above primer SEQ. ID NO. 1 will enable synthesis of the Tet gene from the upstream side, inclusive of the Tet A fragment, this synthesis being in the clockwise direction with respect to the completed plasmid shown in Fig. 2. At the same time, primer SEQ. ID NO. 1 has also an extension at its 5' end, as noted above, that is complementary to the Cm upstream region, to facilitate joining of the Tet and Cm genes.

Likewise, the above primer SEQ. ID NO. 2 will enable synthesis of the Cm gene from the upstream side, inclusive of the CmA fragment, this synthesis being in the counterclockwise direction with respect to the completed plasmid shown in Fig. 2. At the same time, primer SEQ. ID NO. 2 also has an extension at its 5' end, as noted above, that is complementary to the Tet upstream region, to facilitate joining of the Cm and Tet genes.

It should also be noted that, by design, the two genes, Cm and Tet are to be joined at their 5' ends, i.e., when expressed they will be transcribed in opposite directions, this being apparent from the description above.

Furthermore, as arises from the positioning of the dU in the above primers SEQ. ID NO. 1 and SEQ. ID NO. 2, this was chosen so as to produce a large overlap (preferably greater than 15 nucleotides), namely, in the resulting PCR fragments produced with these primers, the dU is twenty nucleotides from the 5' end, and following UDG, N,N-dimethyl ethylenediamine treatment and heating, this entire twenty nucleotide single-stranded 5' fragment will be removed from both PCR fragments, leaving single-stranded 3' overhangs on the complementary fragments. This complementarity is unique (all the other primers for the other fragments having different sequences) to facilitate specific directional joining of the two fragments. As will be apparent to one of skill in the art from the above description and knowledge of the now standard PCR technology, use of the above primers will yield PCR fragments of the following 5'-end sequences containing the dU residues:

5' A G C T C C T G A T T C T C A T G T T U G A C A G C T T A T C...
3' T C G A G G A C T A A G A G T A C A A A C T G T C G A A T A G...
(from Primer SEQ. ID NO. 1)

5' A A A C A T G A G A A T C A G G A G C U A A G G A A G C T A A A A T G...
3' T T T G T A C T C T T A G T C C T C G A T T C C T T C G A T T T T A C...
(from Primer SEQ. ID NO. 2)

Following UDG, N,N-dimethyl ethylenediamine treatment and heating, the 3' overlaps on each of these 5' end sequences will be:

```
                                        5' G A C A G C T T A T C...
3' T C G A G G A C T A A G A G T A C A A A C T G T C G A A T A G...
                    (from Primer SEQ. ID NO. 1)

                                                  5' A A G G A A G C T A A A A T G...
3' T T T G T A C T C T T A G T C C T C G A T T C C T T C G A T T T T A C...
          (from Primer SEQ. ID NO. 2)
```

from which it is apparent that when these overhangs are brought into proximity with each other, they are completely complementary:

```
                                                    | (from primer SEQ. ID NO. 1)
                                                 3' | 5'
          TCC TT AGCTCCTGATTCT CATGTTT | GACAGCTTA...
          AGGAA | TCGAGGACTAAGAGTACAAA CTGTCGAAT...
                5' | 3'
 (from primer    |
 SEQ. ID NO. 2)
```

Therefore, a long overlap of 20 nucleotide bases generated on each of the PCR fragments to be joined by virtue of the specific placement of the dU residue in the primers to prepare the PCR fragments, provides for a highly specific directional joining of these fragments, which is also a stable joint due to the length of those overlaps or overhangs. In fact, it is not necessary to perform a ligase reaction to covalently link the respective single-strand overlapping segments to the rest of the PCR fragments at the joint sites because the strength of hydrogen bonding interactions between the two complementary 3' overhangs is great enough to maintain the joint between the two PCR fragments and prevent them from becoming detached.

As will also be apparent from the above sequences, there are a large number of dT residues which may have been chosen to be replaced by a dU residue during the preparation of the specific primers. Some of the dT residues are nearer the 5' end of the primer sequence and hence would

have been less favorable as they would have led to the generation of a 3' overhang having less than 20 nucleotide bases. Some of the dT residues are more distal from the 5' end of the primer sequence and hence might be even more favorable than the one exemplified above as they will lead to generation of an even longer 3' overhang.

(2) In a similar fashion to that mentioned above concerning the specific directional joining of the upstream end of the Cm gene to the upstream end of the Tet gene, it was chosen to prepare an internal joint within the Tet gene so that the Tet gene may be generated from two separate fragments, one to be joined as above to the upstream end of the Cm gene, and one as noted below to be joined to the first Tet fragment, in a specific directional manner on one side and to the downstream end fragment on the other side. The primers for these other PCR fragments were specifically designed to provide for these joints. Furthermore, it was chosen to prepare the Cm gene in the form of two separate PCR fragments to be joined to each other (internal joint) and to the Tet gene as above, as well as to the Amp ColEI-ORI region (as below). For all of these connections it was also necessary to prepare all of the necessary primers for synthesis of the various PCR fragments so that each will eventually provide for the desired specific directional joining of all the fragments.

Thus, for the internal Tet gene fusion the following sequence, as obtained from the sequence provided with the commercially obtained pBR322 plasmid, was chosen as the desired linkage region for the two Tet fragments to provide the desired internal linkage.

internal Tet gene region

5' C G A T G A T C G G C C T G T C G C T T G C G G T A T T C G 3'
3' G C T A C T A G C C G G A C A G C G A A C G C C A T A A G C 5'

To produce the two Tet gene fragments (Tet A and Tet B in Fig. 2) by PCR procedure and to ensure the above internal connection in a specific

directional way, the following primers containing the noted dU residues were prepared:

a) Primer SEQ. ID NO. 3, also designated 25595:

Tet gene internal region - sense strand

5' A T C G G C C T G T C G C T T G C G G T A U T C G 3'

b) Primer SEQ. ID NO. 4, also designated 27341:

Tet gene internal region - anti sense strand.

5' A T A C C G C A A G C G A C A G G C C G A U C A T C G 3'

For the relative positions of these primers with respect to the completed constructed plasmid and the direction of synthesis from these primers during the PCR procedure, there is shown in Fig. 2 the two primers with arrows indicating their directions.

In an analogous manner to that described above with respect to the Cm upstream - Tet upstream junction, the above primers were used to generate the required PCR fragments having terminal sequences inclusive of the above primer sequences. Following UDG, N,N-dimethylethylenediamine treatment and heating of these PCR fragments, 3' single-stranded overhangs of 22 nucleotide bases are generated on each PCR fragment. As is apparent from the above primer sequences and their directionality, these 3' single-stranded overhangs are completely complementary and hence will readily combine with each other to form the desired Tet gene internal region linkage. Here, too, the nature of the complementary interaction of the 22 base pairs is strong enough to link tightly the two fragments without the necessity for a subsequent ligase reaction.

(3) For the next desired linkage in the plasmid to be constructed, namely, the linkage between the downstream end of the Tet gene and the

downstream end of the Amp-ColE1-ORI region, both from plasmid pBR322, the following sequence was chosen as the desired linkage region for the PCR fragments to be produced containing the above gene regions:

downstream region of Amp | Tet gene downstream region
junction point
5' GAAATGTTGAATACTCATACTCTT↓TGGCCAGGACCCAACGCTGCCCGAG 3'
3' CTTTACAACTTATGAGTATGAGAA↑ACCGGTCCTGGGTTGCGACGGGCTC 5'

To produce the PCR fragments which will contain the above linkage region to enable specific directional joining between these fragments, the following primers containing the noted dU residues were prepared:

a) Primer SEQ. ID NO. 5, also designated as 30402:

(Amp region) | (Tet region)
junction point
5' A T A C T C T T↑T G G C C A G G A C C C A A C G C U G C C C 3'

b) Primer SEQ. ID NO. 6, also designated as 3885:

(Tet region) junction point (Amp region)
5' AGCGTTGGGTCCTGGCCA↑AAGAGTAUGAGTATTCAACA 3'

For the relative positions of these primers with respect to the completed constructed plasmid and the direction of synthesis from these primers during the PCR procedures, there is shown in Fig. 2 the two primers with arrows indicating their directions.

As noted above for the other fragments, these primers were used to generate the required PCR fragments having terminal sequences inclusive of the above primer sequences. Following UDG, N,N-dimethylethylenediamine treatment and heating of these PCR fragments, 3' single-stranded overhangs of 26 nucleotide bases are generated on each PCR fragment. As is apparent from the above primer sequences and their directionality, these 3' single-stranded overhangs are completely complementary and hence will readily combine with each other to form the desired 3' Amp-ColE1-ORI - 3' Tet gene linkage. There

too, the nature of the complementary interaction of the 26 base pairs is strong enough to tightly combine the two fragments without the necessity for a subsequent ligase reaction.

(4) The next desired linkage in the plasmid to be constructed was the linkage between the Amp -ColE1-ORI fragment, upstream of the Amp gene and the downstream of the Cm gene. A study of the respective sequences revealed the following as the desired sequence for the linkage region for the PCR fragments to be produced containing the above gene regions:

junction point

Amp **upstream** region ↓ Cm **downstream** region

...5' TGCTCACAT GTTCTTT CCTGC GTT↓CAGGCGTTTAAGGGCACCAATAAC 3'...

...3' ACGAGTGTACAAGAAAGGACGCAA↑GTCC GCAAATTCCCGTGGTTATTG 5'...

To produce the PCR fragments which will contain the above linkage region to enable specific directional linkage between these fragments, the following primers containing the noted dU residues were prepared:

a) Primer SEQ. ID NO. 7, also designated as 4144:

junction point

Amp region Cm region

5'ATGTTCTTTCCTGCGTT↑CAGGCGUTTAAGGGCACCAATAAC 3'

b) Primer SEQ. ID NO. 8, also designated as 27342:

junction point

Cm region Amp region

5'ACGCCTG↑AACGCAGGAAAGAACAUGTG 3'

The relative positions of these primers with respect to the completed constructed plasmid and the direction of synthesis from these primers during the PCR procedures are shown in Fig. 2, with arrows indicating their directions.

As mentioned above, with regard to the other primers, these primers were used to generate the required PCR fragments having terminal

sequences inclusive of the above primer sequences. Following UDG, N,N-dimethyl ethylene diamine treatment and heating, of these PCR fragments, 3' single-stranded overhangs of 24 nucleotide bases are generated on each PCR fragment. As is apparent from the above primer sequences and their directionality, these 3' single-stranded overhangs are completely complementary and hence will readily combine with each other to form the desired Amp-ColE1-ORI Cm gene linkage. Again, the nature of the complementary interaction of the 24 base pairs is strong enough to facilitate tight binding between the two fragments without the necessity for a subsequent ligase reaction.

(5) The next desired linkage in the plasmid to be constructed was the internal linkage between the two PCR fragments each having part of the Cm gene (Cm A and Cm B fragments in Fig. 2). As noted above, all the sequence information of the internal part of the Cm gene is available with the purchased pACYC184 plasmid. Hence, upon a study of the Cm gene sequence, the following internal Cm sequence was chosen as the desired sequence for the internal joining of the PCR fragments to be produced containing the two Cm gene portions:

internal region of Cm gene (bases 4018-4046 of the pACYC184 map)

5' G G A T T G G C T G A G A C G A A A A A C A T A T T C T C 3'
3' C C T A A C C G A C T C T G C T T T T T G T A T A A G A G 5'

To produce the PCR fragments which will contain the above linkage region to enable specific directional linkage between these fragments, the following primers containing the noted dU residues were prepared:

a) Primer SEQ. ID NO. 9, also designated as 27343:

Cm internal region - non-coding strand (anti-sense)
5' A T T G G C T G A G A C G A A A A A C A T A U T C T C 3'

b) Primer SEQ. ID NO. 10, also designated as 25596:

Cm internal region - coding strand (sense)
5' A T A T G T T T T T C G T C T C A G C C A A U C C 3'

The relative positions of these primers with respect to the completed constructed plasmid and the direction of synthesis from these primers during the PCR procedures are shown in Fig. 2 with arrows indicating their directions.

As mentioned above, with regard to the other primers, these primers were used to generate the required PCR fragments having terminal sequences inclusive of the above primer sequences. Following UDG, N,N-dimethyl ethylendiamine treatment and heating of these PCR fragments, 3' single-stranded overhangs of 23 nucleotide bases are generated on each PCR fragment. As is apparent from the above primer sequences and their directionality, these 3' single-stranded overhangs are completely complementary, and hence will readily combine with each other to form the desired internal Cm gene linkage. Again, the nature of the complementary interaction of the 23 bases pairs is strong enough to facilitate tight binding between the two fragments without the necessity for a subsequent ligase reaction.

It should be mentioned that the selection of the above primer sequences, in particular, those for the joining of two fragments making up a single gene region (internal junction), was not a rigid selection, as any suitable sequence could have been chosen, as long as it provided the desired junction between two specific fragments to be joined. In other words, each of the above primer sequences represents but one possibility from among many which could have been equally suitable. The only constraints in the selection of the primer sequences are that they: (a) will have at least one dU insertion at a position providing for a subsequent 3' overlap on the resulting PCR fragments of at least 15 nucleotides and preferably about 20 nucleotides; (b) they are chosen so as to include in their sequence portions of the sequence of both of the DNA fragments it is desired to join, thereby ensuring a "seamless" joining of the two

fragments, without any undesired "linker" sequence being inserted into the junction region between the two joined DNA fragments; and (c) that the primer sequences are so chosen to provide for complete complementarity only between the subsequently generated 3' overhangs of two fragments (the specific ends of these fragments) that it is desired to combine, such that when a number of fragments are to be joined in a desired pre-selected order, only the desired junctions will be possible.

For example, in the above primers SEQ. ID NO. 9 and SEQ. ID NO. 10 for the generation of PCR fragments encoding the internal region of the Cm gene to be specifically joined together, the primer sequences may have been chosen from among many possibilities, each having part of the Cm A fragment sequence and part of the Cm B fragment sequence, i.e., each primer covering the desired junction region between the Cm A and Cm B fragments. In this instance, because the junction is an internal one within a single gene region, there is even more flexibility of choice of primer sequence, as the junction point may have been pre-selected anywhere within the Cm gene. Further, the so-chosen primers SEQ. ID NO. 9 and SEQ. ID NO. 10 also have in their sequence numerous dT residues, a number of which (in addition to the exemplified one) may also have been replaced by a dU residue to provide subsequently the desired long 3' overhangs for the purposes of joining the PCR fragments generated from these primers. As the primer sequences were chosen specifically from the Cm gene sequence, there is ensured that once the generated PCR fragments are joined via their 3' overhangs, the joined fragments will contain only the Cm gene sequence, i.e., there is a "seamless" junction.

(ii) <u>Preparation of the PCR fragments</u>

Taking all of the above into consideration, the following is a brief description of how each fragment of the plasmid of Fig. 2 was prepared

using all of the various primers for each fragment, and how all the so-produced fragments were linked together:

a) The Amp-ColE1-ORI region fragment was prepared by standard PCR conditions using the above-noted primers SEQ. ID NO. 8 and SEQ. ID NO. 6 (see Fig. 2 for their relative positions and directionality with respect to the completed plasmid), as well as the purchased pBR322 plasmid as the template DNA. The amounts of primer DNA, template DNA and other conditions of the PCR production were as noted hereinabove.

b) The upstream portion of the Tet gene (the Tet A fragment shown in Fig. 2) was prepared by standard PCR conditions using the above-noted primers SEQ. ID NO. 1 and SEQ. ID NO. 4 (see Fig. 2 for their relative positions and directionality with respect to the completed plasmid), as well as the purchased pBR322 plasmid as the template DNA. The amounts of primer DNA, template DNA and other conditions of the PCR procedures were as noted hereinabove.

c) The downstream portion of the Tet gene (the Tet B fragment shown in Fig. 2) was prepared by standard PCR conditions using the above-noted primers SEQ. ID NO. 3 and SEQ. ID NO. 5 (see Fig. 2 for their relative positions and directionality, with respect to the completed plasmid), as well as the purchased pBR322 plasmid as the template DNA. The amounts of primers DNA, template DNA and other conditions of the PCR procedure were as noted hereinabove.

d) The upstream portion of the Cm gene (the Cm A fragment shown in Fig. 2) was prepared by standard PCR conditions using the above-noted primers SEQ. ID NO. 2 and SEQ. ID NO. 9 (see Fig. 2 for their relative positions and directionality, with respect to the completed plasmid), as well as the purchased pACYC184 plasmid as the template

DNA. The amounts of primer DNA, template DNA and other conditions of the PCR procedures were as noted hereinabove.

e) The downstream portion of the Cm gene (the Cm B fragment shown in Fig. 2) was prepared by standard PCR conditions using the above-noted primers SEQ. ID NO. 10 and SEQ. ID NO. 7 (see Fig 2 for their relative positions and directionality, with respect to the completed plasmid), as well as the purchased pACYC184 plasmid as the template DNA. The amounts of primer DNA, template DNA and other conditions of the PCR procedures were as noted hereinabove.

For PCR, the following conditions and conentrations were used: 10µl of a 0.1 mg/ml solution of primer #1 and primer #2, respectively, were mixed and 180µl of PCR mix as given below added. Then, 1µl of 40ng/µl of DNA template was added, and the PCR carried out at the temperature regime indicated below.

PCR mix:

dNTPs (2.5mM each)	128µl
Buff. x10	160µl
H_2O	1088µl
Taq DNA Polymerase 5u/µl	6.4µl

Temperature regime

94°C 40sec

40°C 2min

72°C 4min

30 cycles

72°C 5min

6°C infinitely

Following PCR synthesis of the individual fragments separately, each fragment was then purified by standard agarose-gel purification techniques using the commercially available Bio-Rad "Prep-A-Gene™" DNA purification kit and adhering to the manufacturer's instructions. Following purification, the concentration of the purified fragment DNA was determined by standard procedures using the Pharmacia "Gene-Quant™ RNA/DNA Calculator" and adhering to the manufacture's instructions.

(iii) Connection of the PCR - produced fragments

The 5 PCR fragments as produced and purified according to the above-mentioned procedure, were connected to each other in a one-step reaction mixture in a single reaction vessel. This was achieved by mixing the fragments together. Each DNA fragment was in an amount of 0.15 pmol. The volume of the DNA fragments mixture was 9 µl. 10 µl "dU nicking mixture" was added tc the DNA fragments mixture. The dU nicking mixture containing 40mM Tris-HCI pH 8.4, 100 mM KCl, 3mM $MgCl_2$, 200mM N,N-dimethyl ethylenediamine. 1 µl UDG (in a concentration of 5 U/µl purchased from GibcoBRL), was then added to the reaction, and the mixture was incubated at 37°C for 30 minutes.

Following the above reaction to apurinate the uracil bases from the fragments with UDG and to nick the DNA 3' to the apurinated dU residues with N,N-dimethyl ethylenediamine, thereby preparing the way for dissociation of short single-strands (between 21-26 bases in length, as noted above), the reaction mixture was subjected to conditions to facilitate dissociation of the short nicked single-stranded DNA (oligonucleotides) from the 5' ends of the fragments in order to expose the desired 3' overhangs. These conditions included adding 20 µl Mineral Oil to the reaction mixture (to prevent evaporation) and then heating the reaction mixture to 75°C, at which temperature the mixture was

incubated for 30 min. At 75°C, the oligonucleotides (short single-strands) upstream of the dU residue at each end of the fragments dissociate from the remaining major portion of the fragments, leaving exposed 3' single-stranded overhangs on each end of each fragment.

The next stage was to facilitate the completion of the specific directional joining of the fragments by cooling the above reaction mixture to 37°C (cooling from 75°C at a rate of about 2°C per hour).

(iv) **analysis of the products**

A 1 µl sample of the above reaction mixture containing the newly constructed plasmid made from joining the 5 separate fragments was used to transform suitable bacterial cells by standard procedures. The mode of transformation was by the preferred standard method of electroporation of *E. coli* DH10B cells. In the transformation procedure 20 µl of electrocompetent "ElectroMax" cells (purchased from Gibco BRL) were mixed with the above 1 µl DNA sample and subjected to electroporation in a commercially available apparatus (BioRad "*E. coli* Pulser Apparatus" set at 1.8 kV and operated according to the manufacturer's instructions).

Following electroporation (transformation) the cells were plated on a standard LB Agar plate containing 100 mM ampicillin (to select for transformants having ampicillin resistance by virtue of having being transformed with a DNA carrying the Amp gene)

The Agar plates were incubated under standard conditions. Over a hundred colonies were obtained and tested further for resistance to chloramphenicol and tetracycline. Thus, colonies found to be resistant to all three antibiotics are indicative of those having received a fully intact constructed plasmid which carries all three resistance genes, namely, the

desired constructed plasmid (having the desired structure as depicted in Fig. 2).

The results showed that about 10% of the originally obtained ampicillin-resistant colonies were also tetracycline and choramphenicol resistant. Additional PCR verification tests were carried out on these colonies in order to verify that they carried the required plasmids. Pairs of primers originally used to produce two different fragments were used in a PCR test procedure in which the template DNA was the construct DNA obtained from the transformants. If the fragments had been connected in the correct manner, the PCR product would be equal in size to the sum of the original fragments. The PCR procedure and conditions were the same as noted above under "Materials and Methods". The PCR products were subjected to standard Agarose gel electrophoresis and the resultant bands separated on the gel were analyzed against standard molecular weight markers to determine their approximate size and to determine whether these sizes are the expected ones.

All of the colonies that were subjected to the above PCR test, had constructs that yielded PCR bands having the expected sizes (results not shown).

The above result is a very significant one, as it provides evidence for the first time that it is possible to join correctly 5 separate DNA fragments in a specific directional manner, in a single reaction mixture by an essentially one-step procedure.

Example 2

Improvements in the procedure for the construction of a plasmid from 5 separate fragments.

In view of the results set forth in Example 1 above, it was desired to improve the procedure of plasmid construction to obtain a higher yield of fully constructed plasmids from 5 original separate fragments.

In this improved procedure, it was chosen to utilize recently developed technology whereby DNA molecules can be biotinylated and subsequently attached to streptavidin beads to effectively immobilize such biotinylated DNA molecules and remove them from the reaction mixture. Such biotinylation was performed at the terminal adenosine residue for ease of operation. The advantage of this approach is that it removes almost all the oligonucleotides which may interfere with the connection between fragments. Furthermore, it also removes fragments that have two, or even one unexposed terminus, further improving the efficiency of the assembly of fragments.

Using this technology, it was chosen to biotinylate all of the above noted primers SEQ. ID NO. 1, SEQ. ID NO. 2, SEQ. ID NO. 3, SEQ. ID NO. 4, SEQ. ID NO. 5, SEQ. ID NO. 6, SEQ. ID NO. 7, SEQ. ID NO. 8, SEQ. ID NO. 9 and SEQ. ID NO. 10 at their 5' terminal adenosine residues. Biotinylation of the primers is carried out at the time of their synthesis, namely, a biotinylated dA residue is prepared by standard procedures or purchased from a commercial supplier, and is used in the primer synthesis reaction as the last nucleotide, i.e., the 5' terminal nucleotide (primer synthesis using automated apparatus and procedures has the synthesis in the 3'-5' direction). All the other dA residues added during primer synthesis will be normal, non-biotinylated ones. In this way, each primer so synthesized will be biotinylated only at its 5' terminal adenosine residue. As a result, the PCR fragments to be produced with these primers will be 5' biotinylated (two biotinylated 5' adenosine

residues per fragment produced as each fragment is generated with two primers - see above procedure for the fragment preparation). These biotinylated short strands may be attached to streptavidin-coated beads.

The entire procedure of PCR fragment preparation, was carried out using biotinylated primers which were otherwise identical to those described in Example 1.

In addition, it was also chosen to perform a phosphorothiolation of the 3' ends of each of the primers before use in the PCR procedure, to ensure that the 3' ends of each of the various primers would not be digested by "proof-reading" functions of the high-fidelity DNA polymerases to be used in the PCR procedure. In this way, all of the above primers were also 3' phosphorothiolated during their synthesis before being used in the PCR reactions to generate the various DNA fragments. This phosphorothiolation of the primers being another improvement of the basic method of the invention described in Example 1.

It should, however, be noted that for the purposes of employing high-fidelity DNA polymerases, as noted above, in the PCR reactions, different buffers and other reaction conditions are necessary. Hence, for the purposes of the present PCR reactions, it was chosen to use the same DNA polymerase and PCR reaction conditions as noted above in Example 1. In any event, such modified primers are also capable of being employed with high-fidelity DNA polymerases in accordingly modified PCR reactions.

Thus, once the primers were prepared as above with 5' biotinylation and 3' phosphorothiolation, the PCR fragments were produced by the same procedure as in Example 1, up to and including the step of treating the fragments with UDG and N,N-dimethylethylenediamine and incubating the reaction mixture at 37°C for 30 min. For the subsequent joining of the

fragments, the reaction was divided at this stage into two separate vessels (one serving as a control).

To the first vessel of 100µl mixture, magnetic streptavidin-coated beads were added (100 µl, purchased from MPG) as was 20 µl of Mineral Oil to prevent evaporation, and the vessel was then incubated with occasional shaking at 75°C for 30 min. As noted above, at 75°C, the short 5' oligonucleotides upstream of the dU residue at each end of the fragments dissociate from the main portion of the fragments, leaving 3' overhangs. In the presence of the streptavidin-coated beads, these 5' oligonucleotides which are biotinylated are "captured" (i.e., bind tightly,) by the streptavidin molecules and are effectively removed from the "construction mixture", preventing them from competing with the fragment-fragment connections. Following the 30 min incubation at 75°C, the streptavidin beads, which are magnetic, were removed from the reaction vessel by the simple standard procedure of collecting and removing them with a magnet, under suitable conditions so as not to lead to any damage of the DNA fragments in the mixture. Therefore, the short biotinylated 5' oligonucleotides were physically removed from the reaction mixture before the cooling stage, in which complete stabilization of the interacting fragments takes place, to facilitate the final stage of the specific directional connections between the various fragments.

In the second reaction vessel, 100 µl H_2O was added instead of streptavidin-coated beads. This was thus a control experiment, being essentially the same as that described in Example 1 above, in which the 5' oligonucleotides remain in the reaction mixture and can interfere with the fragment-fragment interconnections. All other reaction conditions were as before.

For both of the above reactions, following the 30 min. incubation at 75°C, the slow cooling procedure was carried out, as described in Example 1.

Likewise, 1 µl DNA samples from each of the above two reaction vessels (with and without streptavidin beads treatment) were used to transform the same strain of *E. coli* cells by the same procedure as detailed in Example 1.

The results of the *E. coli* transformations revealed that for the streptavidin-treated reaction mixture, 480 ampicillin-resistant colonies were obtained, of which 35% were chloramphenicol- and tetracycline-resistant. In contrast, for the non-streptavidin-treated reaction mixture, only 160 ampicillin-resistant colonies were obtained, of which only about 10% were also resistant to chloramphenicol and ampicillin (this being essentially the same result as obtained for the basic procedure set forth in Example 1).

The above significantly greater number of completely constructed desired plasmids obtained by biotinylation-streptavidin treatment indicates that the free released 5' oligonucleotides do interfere with the connections between the various fragments, and their removal by the biotinylation-streptavidin treatment is desirable.

The use of the Biotin-Streptavidin system, for removing of the oligonucleotide generated by nicking at the apurinated site, is for illustrative purposes only. It can be done by several different methods, as for example the use of commercially available kits for removing oligonucleotides (see Example 3 hereinbelow).

Example 3

Preparation of a circularized plasmid by the directional connection of 8 fragments constituting the plasmid

Overview

The plasmid to be constructed was designed to consist of eight fragments, each to be prepared separately by PCR amplification and then joined in a specific directional fashion to provide a circularized plasmid as the end-product. This example is similar to example 1, the difference being the number of fragments the plasmid was constructed of.

In Fig. 3, there is shown schematically the plasmid that was designed and produced by the method of the present invention. This plasmid carries four independent antibiotic resistance genes, for resistance to ampicillin (Amp^r gene, or hereinafter Amp) ; tetracycline (Tet^r gene, or hereinafter Tet); chloramphenicol (Cm^r gene, or hereinafter Cm) and kanamycin (Kn^r gene, or hereinafter Kn). The plasmid also carries the ColEl origin of replication (ColEl-ORI), which in this specific instance is situated next to the Amp gene. Hence, such a plasmid is capable of being replicated in a host cell and will endow the host cell with resistance to all four types of antibiotic.

Accordingly, it is possible to readily select for those host cells transformed by this plasmid by growing the transfected cells in the presence of one of the antibiotics and then to screen for the resistance to the other antibiotics. In order to verify that mis-connections did not occur, additional verification tests may be carried out. One such test carried out included testing transformed colonies by restriction enzyme analysis. DNA was prepared from a number of transformed colonies and checked by restriction enzyme analysis.

This constitutes the first known disclosure of a successful, specific directional connection of eight independent PCR fragments to form a single active plasmid construct, and this by an essentially one-step procedure in accordance with the present invention.

For this construction of eight fragments, the following was carried out:

(i) Preparation of the specific primers for the PCR amplification.

As shown in Fig. 3, and as outlined above, it was desired to construct a circularized plasmid having four regions:

(a) a Tet region.

(b) an Amp + ColEl-ORI region.

(c) a Cm region.

(d) a Kn region.

The order of the connection is depicted in figure 3.

To achieve this task, eight separate fragments were designed:

Name of fragment	Size(bp)	plasmid	Location
AmpA 241365 — 3885	574	pBR322	3603-4159
AmpB 27342 — 241366	1171	pBR322	2460-3624
CmA 40122 — 27343	481	PACYC184	4021-240
CmB 25596 — 4144	293	PACYC184	3768-4043
TetA 27341 — 36176	862	pBR322	1-848
TetB 30402 — 25595	631	pBR322	827-1449
KnA 31254 — 25953	569	PACYC177	1809-2368
KnB 25952 — 31253	652	PACYC177	2349-2991

Written are the names of the fragments, their sizes, the plasmid from which they were PCR amplified, the exact location of the site on the original plasmid and the numbers of the primers that were used to amplify the fragments. The full sequence and maps of the various regions of these plasmids are known and can be accessed from GenBank database under accession Nos. J01749 (pBR322), X06403 (pACYCI84) and X06402 (pACYC177).

The **AmpB** fragment includes the 5' part of the Amp fragment and the ColEl-ORI sequence. The **AmpA** fragment includes the 3' part of the Amp fragment. The **CmA** fragment includes the 5' part of the Cm fragment and the **CmB** fragment includes the 3' part of the Cm fragment. The **TetB** fragment includes the 5' part of the Tet fragment and the **TetA** fragment includes the 3' part of the Tet fragment. The **KnA** fragment includes the 5' part of the Kn fragment and the **KnB** fragment includes the 3' part of the Kn fragment.

As in example 1, each primer consists of two regions: a 3' region complementary to the DNA to be amplified, and a 5' region complementary to the fragment it should be connect to.

Using standard automated procedures to produce polynucleotide oligomers (Applied Biosystems, U.S.A.), the following primers were synthesized:

In all the cases below, the (-) in the middle of the sequence represents the junction between the regions of the original DNA.

Primers for the amplification of the **AmpA** fragments

Primer SEQ. ID NO. 11, also designated as 241365:

internal Amp region
5'ATTGCTGCAGGCATCGTGGTGUCA 3'

Primer SEQ. ID NO. 6, also designated as 3885:

Cm region Amp region
5'AGCGTTGGGTCCTGGCCA - AAGAGTAUGAGTATTCAACA 3'

Primers for the amplification of the **AmpB** fragments

Primer SEQ. ID NO. 8, also designated as 27342:

Cm region Amp region
5'ACGCCTG - AACGCAGGAAAGAACAUGTG 3'

Primer SEQ. ID NO. 12, also designated as 241366:

internal Amp region
5'ACACCACGATGCCTGCAGCAAUGG 3'

Primers for the amplification of the **CmA** fragments

Primer SEQ. ID NO. 13, also designated as 40122:

Kn region Cm region
AGG CCT GGT ATG AGT C - TCA GGA GCU AAG GAA GCT AAA ATG

Primer SEQ. ID NO 9, also designated as 27343:

Cm region
5'ATTGGCTGA GACGAAAAACATAUTCTC 3-

Primers for the amplification of the **CmB** fragment

Primer SEQ. ID NO. 10, also designated as 25596:

Cm region
5' ATATGTTTTTCGTCTCAGCCAAUCC 3'

Primer SEQ. ID NO. 7, also designated as 4144:

Amp region Cm region
5'ATGTTCTTTCCTGCGTT - CAGGCGUTTAAGGGCACCAATAAC 3-

Primers for the amplification of the **TetA** fragment:

Primer SEQ. ID NO. 4, also designated 27341:

Tet region
5'ATACCGCAAGCGACAGGCCGAUCATCG 3'

Primer SEQ. ID NO. 14, also designated as 36176:

Kn antisense Tet sense
5'ACGTGGCTTTGTTG - TTCTCATGUTTGACAGCTTATC 3'

Primers for the amplification of the **TetB** fragment:

Primer SEQ. ID NO. 5, also designated as 30402:

Amp region Tet region
5'ATACTCTT - TGGCCAGGACCCAACGCUGCCC 3-

Primer SEQ. ID NO. 3, also designated 25535:

Tet region
5'ATCGGCCTGTCGCTTGCGGTAUTCG 3'

Primers for the amplification of the **KnA** fragment:

Primer SEQ. ID NO. 15, also designated 31254:

Tet region Kn region
5'ACATGAGAA - CAACAAAGCCACGUTGTGTCTC 3'

Primer SEQ. ID NO. 16, also designated as 25953:

Kn region
5'AGACGAAATACGCGATCGCUGTTAA 3'

Primers for the amplification of the **KnB** fragment:

Primer SEQ. ID NO. 17, also designated as 25952:

Kn region
5'AGCGATCGCGTATTTCGTCUCGCTC 3'

Primer SEQ. ID NO. 18, also designated as 31253:

Cm region Kn region
5'AGCTCCTGA - GACTCATACCAGGCCUGAATCG 3'

(ii) Preparation of the PCR fragments

The PCR was carried out as in Examples 1 and 2. Following PCR synthesis of the individual fragments, each fragment was purified by standard agarose-gel purification techniques using the commercially available Bio-Rad "Prep-A-Gene™" DNA purification kit and adhering to the manufacturer's instructions. Following purification, the

concentration of the purified fragment DNA was determined by standard procedures.

(iii) Connection of the PCR-produced fragments

The eight PCR fragments were connected to each other in a one-step reaction mixture in a single reaction vessel. This was achieved by mixing the fragments together in a 25µl reaction mixture that included: 0.15 pmol of each fragment, 2.5µl buffer (200mM Tris-HCl pH 8.4, 500 mM KCl, 15mM $MgCl_2$), 2.5 µl of 1M N,N-dimethyl ethylenediamine and 6.25 units of UDG (GibcoBRL). The mixture was incubated at 37°C for 4 hours and then transferred to 70°C for 5 minutes to facilitate dissociation of the short nicked single-stranded DNA from the 5' ends of the fragments (as explained in example 1). After dissociation of the short nicked single-stranded DNA, the DNA was cleaned from them with "QIAquick PCR purification kit" (QIAGEN) adhering to the manufacturer's instructions. Before adding the first buffer of the kit, 200 µl of hot (70°C) buffer (20mM Tris-HCl pH 8.4, 50 mM KCl, 1.5mM $MgCl_2$) was added in order to minimize the reannealing of the short nicked single-stranded DNA. This was carried out instead of the biotin-method desribed in Example 2. The DNA was eluted in 30µl of sterilized water. 27µl of the DNA mixture was incubated with 3µl of buffer (200mM Tris-HCl pH 8.4, 500 mM KCl, 15mM $MgCl_2$) in a water-bath at 70°C. The bath was shut down and the temperature was slowly decreased to 37°C. This allows for the joining between the complementary overhangs, while minimizing illegitimate connections.

(iv) analysis of the construct

A 1 µl sample of the above DNA containing the newly constructed plasmid, made by joining the 8 separate fragments, was used to

transform *E. coli* DH10B cells by electroporation (as detailed in example 1).

Following electroporation (transformation) the cells were plated on LB Agar plate containing 100 mM ampicillin (to select for transformants having ampicillin resistance by virtue of having been transformed with a DNA carrying the Amp^r gene). On the day after, three colonies were picked and checked for resistance to chloramphenicol, tetracycline and kanamycin by plating them on Agar plates containing the appropriate antibiotics.

The results showed that the three colonies were resistant to all the four antibiotics. The colonies were further checked by restriction enzyme analysis and proved to be correct (data not shown).

The above result is very significant, since it proves that it is possible to correctly join 8 separate DNA fragments in a specific directional manner, in a single reaction mixture by an essentially one-step procedure.

The examples above demonstrate that in accordance with the present invention, it is possible for the first time to construct a fully active plasmid molecule from numerous (e.g., 5, 8) separate fragments in a simple, essentially one-step process, and even under less favorable conditions (without biotinylation-streptavidin treatment), this process is feasible, producing fully constructed desired plasmid molecules.

The above examples have been provided for the purpose of illustration only. As will be appreciated by the skilled person, the invention is not

limited to any particular conditions, except as defined in the appended claims.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT:

(A) NAME: Gil Sharon and Gesher Israel Advanced Biotecs (1996) Ltd.

(B) STREET: POB 98

(C) CITY: Beit Neqofa

(E) COUNTRY: Israel

(F) POSTAL CODE (ZIP): 90830

(G) TELEPHONE: +972 2 5700 381

(H) TELEFAX: +972 2 5700 586

(ii) TITLE OF INVENTION: Method for joining DNA fragments

(iii) NUMBER OF SEQUENCES: 18

(iv) COMPUTER READABLE FORM:

(A) MEDIUM TYPE: Floppy disk

(B) COMPUTER: IBM PC compatible

(C) OPERATING SYSTEM: PC-DOS/MS-DOS

(D) SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)

(2) INFORMATION FOR SEQ ID NO: 1:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 31 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: unknown

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

AGCTCCTGAT TCTCATGTTU GACAGCTTAT C 31

(2) INFORMATION FOR SEQ ID NO: 2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 35 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: unknown
(D) TOPOLOGY: unknown

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

AAACATGAGA ATCAGGAGCU AAGGAAGCTA AAATG 35

(2) INFORMATION FOR SEQ ID NO: 3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: unknown

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

ATCGGCCTGT CGCTTGCGGT AUTCG 25

(2) INFORMATION FOR SEQ ID NO: 4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 27 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: unknown

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

ATACCGCAAG CGACAGGCCG AUCATCG 27

(2) INFORMATION FOR SEQ ID NO: 5:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: unknown

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

ATACTCTTTG GCCAGGACCC AACGCUGCCC 30

(2) INFORMATION FOR SEQ ID NO: 6:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 38 base pairs
(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: unknown

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:

AGCGTTGGGT CCTGGCCAAA GAGTAUGAGT ATTCAACA 38

(2) INFORMATION FOR SEQ ID NO: 7:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 41 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: unknown

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 7:

ATGTTCTTTC CTGCGTTCAG GCGUTTAAGG GCACCAATAA C 41

(2) INFORMATION FOR SEQ ID NO: 8:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 27 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: unknown

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 8:

ACGCCTGAAC GCAGGAAAGA ACAUGTG 27

(2) INFORMATION FOR SEQ ID NO: 9:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 27 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: unknown

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 9:

ATTGGCTGAG ACGAAAAACA TAUTCTC 27

(2) INFORMATION FOR SEQ ID NO: 10:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 10:

ATATGTTTTT CGTCTCAGCC AAUCC 25

(2) INFORMATION FOR SEQ ID NO: 11:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: unknown

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 11:

ATTGCTGCAG GCATCGTGGT GUCA 24

(2) INFORMATION FOR SEQ ID NO: 12:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: unknown

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 12:

ACACCACGAT GCCTGCAGCA AUGG 24

(2) INFORMATION FOR SEQ ID NO: 13:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 40 base pairs
(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: unknown

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 13:

AGGCCTGGTA TGAGTCTCAG GAGCUAAGGA AGCTAAAATG 40

(2) INFORMATION FOR SEQ ID NO: 14:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 36 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: unknown

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 14:

ACGTGGCTTT GTTGTTCTCA TGUTTGACAG CTTATC 36

(2) INFORMATION FOR SEQ ID NO: 15:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 31 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: unknown

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 15:

ACATGAGAAC AACAAAGCCA CGUTGTGTCT C 31

(2) INFORMATION FOR SEQ ID NO: 16:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: unknown

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 16:

AGACGAAATA CGCGATCGCU GTTAA 25

(2) INFORMATION FOR SEQ ID NO: 17:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: unknown

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 17:

AGCGATCGCG TATTTCGTCU CGCTC 25

(2) INFORMATION FOR SEQ ID NO: 18:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 31 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: unknown

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 18:

AGCTCCTGAG ACTCATACCA GGCCUGAATC G 31

Claims

1. A method for the simultaneous multi-DNA fragment assembly of two or more double-stranded DNA fragments produced by a primer extension reaction, particularly the polymerase chain reaction (PCR), comprising the steps of:

(a) providing for each DNA fragment to be joined to a second DNA fragment and optionally to a third DNA fragment, a pair of primers for the primer extension procedure thereof, wherein one of said two primers has a portion which is complementary to one terminal portion of the strand of said fragment, and the second of said two primers has a portion which is complementary to the other terminal portion of the strand of said fragment, and wherein the second portion of at least one of the said primers is complementary to the terminal portion of a different fragment to be joined to said first fragment in a specific positioned relationship;

(b) providing, in each of said primers, at least one dU residue or the like in place of at least one dT residue of the original primer sequence, the first of said at least one dU residue(s) being positioned at least 15 nucleotides from the 5' end of said primer;

(c) producing each said DNA fragment by the primer extension reaction procedure;

(d) carrying out in any suitable order the steps of: 1. mixing the produced DNA fragments 2. adding the enzyme UDG to apurinate said dU residue and a second enzyme or a compound such as N,N-dimethyl ethylenediamine to generate single-strand nicks at said apurinated dU residue, said UDG and second enzyme or compound being in amounts sufficient thereby to generate a nick at the end of substantially all fragments, and dissociating the nicked oligonucleotides which are formed at the 5' end or ends of the fragments, and

(e) providing suitable conditions and allowing sufficient time for the specific joining of said fragments to generate a multi-DNA fragment assembled product.

2. A method according to claim 1, wherein the DNA mixing step is carried out before the UDG enzyme addition step.

3. A method according to claim 1, wherein the UDG enzyme addition step is carried out before the DNA mixing step.

4. A method according to claim 1-3, wherein step (e) includes the removal of the nicked oligonucleotides from the reaction.

5. A method according to claim 4, wherein at least some of said primers for the primer extension production of said DNA fragments are biotinylated at their 5' ends, and wherein when said enzyme UDG and said second enzyme or compound are added to said mixture of fragments, there is also added streptavidin, either free or bound to beads, whereby single-stranded oligonucleotides containing at their 5' end a biotin moiety generated following the formation of said single-strand nicks at said apurinated dU residues are bound via a biotin-streptavidin connection and are effectively removed from the joining reaction.

6. A method according to claims 1-5, wherein said 3' complementary overhangs on said DNA fragments are generated by the addition to said fragments, the enzyme UDG to apurinate said dU residue and a second enzyme selected from endonuclease III (Endo III) or endonuclease IV (Endo IV) to generate single-stranded nicks at said apurinated dU residue to provide said 3' overhangs.

7. A method according to claims 1-5, wherein said 3' complementary overhangs on said DNA fragments are generated by the addition to said fragments of the enzyme UDG to apurinate said dU residue and of a

compound suitable to generate single-strand nicks at said apurinated dU residue to provide said 3' overhangs.

8. A method according to claim 7, wherein the compound suitable to generate single-strand nicks at the apurinated dU residue to provide the 3' overhangs is N,N-dimethyl ethylenediamine.

9. A method according to any one of claims 1-8, wherein the joining of the fragments via the 3' complementary overhangs is facilitated *in vitro* by:

(a) heating the reaction mixture to a temperature suitable to allow dissociation of the oligonucleotides at the termini of the fragment after said apurination of the dU residue and generation of nicks at the apurinated dU residue, and incubating said mixture;

(b) keeping the mixture of (a) at a temperature suitable to maximize correct connections between complementary overhangs; and

(c) optionally, adding ligase to said mixture to facilitate covalent joining of the DNA strands.

10. A method according to any one of claims 1-9, wherein after the joining of the fragments via the 3' complementary overhangs is facilitated *in vitro*, the DNA is used to transform and/or transfect host cells.

11. A method according to any one of claims 1-9, wherein the covalent connection of the fragments via the 3' complementary overhangs is facilitated *in vivo*, by the host cells that are being transformed and/or transfected by the mixture of fragments.

12. A method according to any one of claims 1-11, wherein one or more of said DNA fragments to be joined is a mutant fragment having been subjected to site-directed mutagenesis or mutagenic PCR during its preparation.

13. A primer for use in a method according to any one of claims 1-12 comprising at least one dU residue in place of at least one dT residue, the first of said at least one dU residue(s) being positioned at least 15 nucleotides from the 5' end of the primer.

14. A primer according to claim 13, wherein said primer is biotinylated at its 5' end.

15. An assembled DNA construct whenever prepared by a method according to any one of claims 1-12, and wherein said construct has been assembled from the joining together of two or more DNA fragments.

16. An assembled DNA construct according to claim 15, wherein said construct is in the form of a linear DNA molecule.

17. An assembled DNA construct according to claim 15, wherein said construct is in the form of a closed circular DNA molecule.

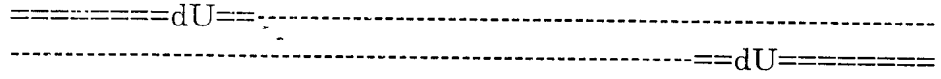
18. A method according to claim 1 wherein one or more of the DNA fragments to be joined comprise sequences coding for one or more proteins selected from the group including cytokines, hormones, and enzymes.

19. A method according to claim 18 wherein the enzymes are selected from the group consisting essentially of citrate synthases, polyketide synthases, and succinyl-CoA-synthetase.

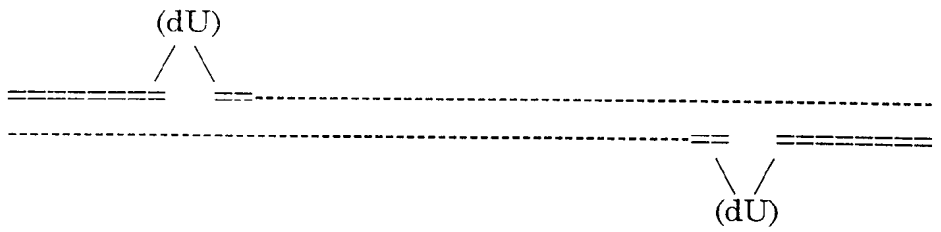
20. A DNA fragment comprising an overhang of at least 15 nucleotides or an end portion suitable to be converted into such an overhang.

21. A DNA fragment as claimed in claim 20, for use in the method of claim 1.

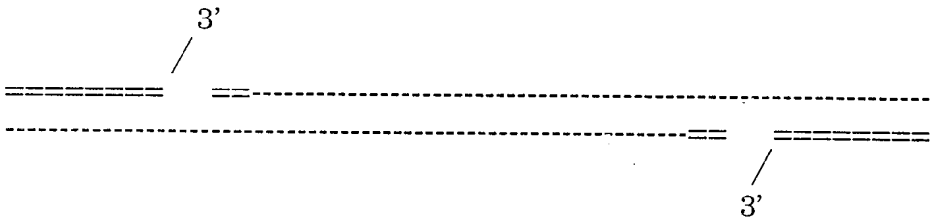
A. PCR fragment with a single dU residue that had been incorporated into each primer



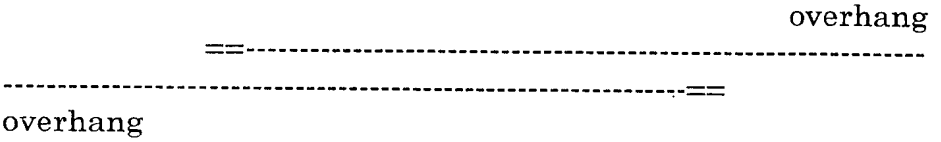
B. UDG a-purinates the dU residues



C. N,N dimethylethylenediamine nicks the a-purinated residues



D. Heating disconnects the nicked oligonucleotides from the double-stranded DNA molecule



E. Slow cooling allows fragments with complementary termini to inter-connect

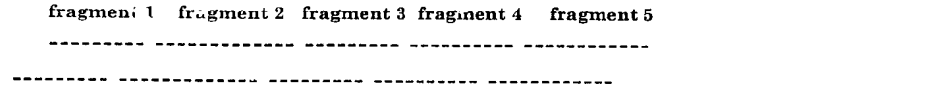


Fig. 1

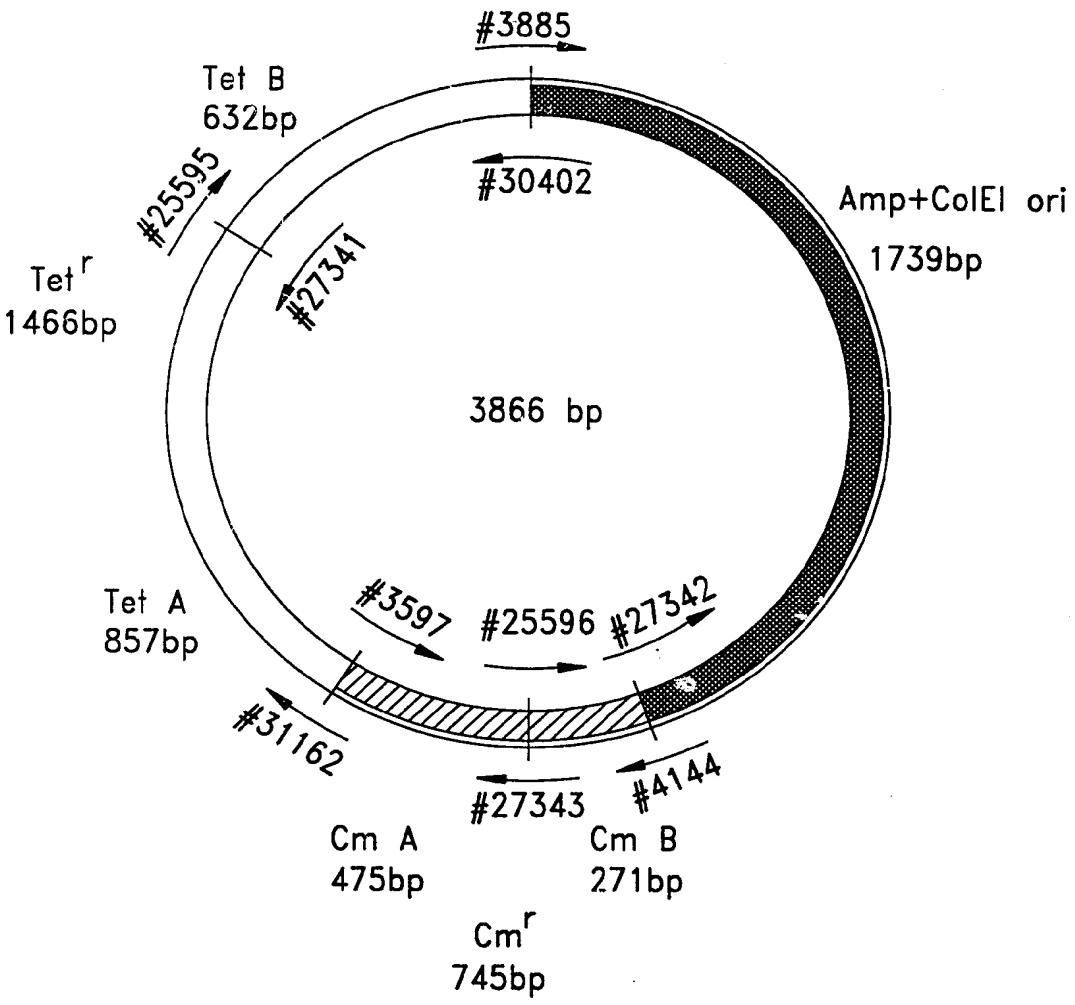

#3885
Tet B
632bp
#25595
#30402
Amp+ColEI ori
1739bp
Tet r
1466bp
#27341
3866 bp
Tet A
857bp
#3597
#25596
#27342
#31162
#27343
#4144
Cm A
475bp
Cm B
271bp
Cm r
745bp

Fig. 2

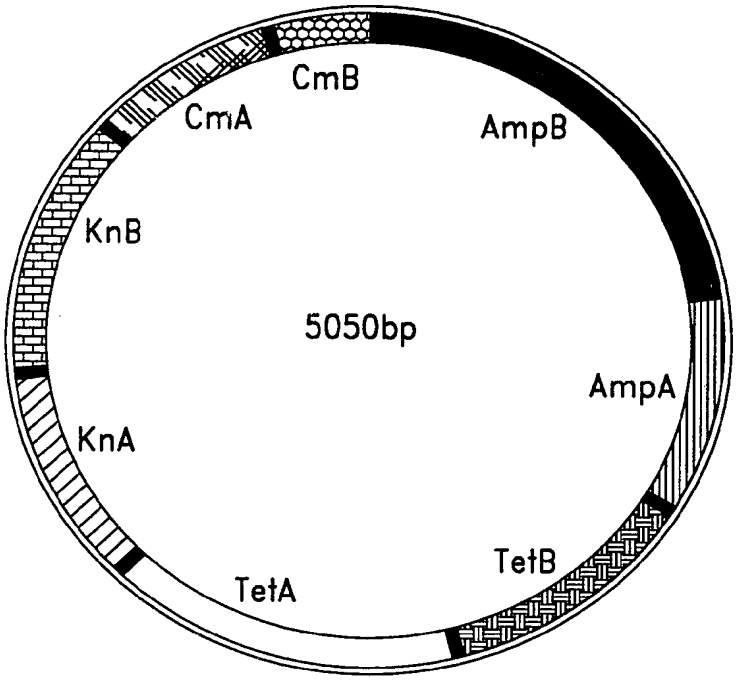


Fig. 3

INTERNATIONAL SEARCH REPORT

Int ational Application No
PCT/IL 98/00094

A. CLASSIFICATION OF SUBJECT MATTER
IPC 6 C12N15/10 C12Q1/68 C12N15/66

According to International Patent Classification(IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 6 C12N C12Q

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	BOOTH P.M. ET AL.: "Assembly and cloning of coding sequences for neurotrophic factors directly from genomic DNA using polymerase chain reaction and uracil DNA glycosylase" GENE., vol. 146, no. 2, 1994, AMSTERDAM NL, pages 303-308, XP002071998 see the whole document --- -/--	1-21

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

° Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
4 August 1998	17/08/1998

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Panzica, G

2

Form PCT/ISA/210 (second sheet) (July 1992)

IDT Exhibit 1020.0422

INTERNATIONAL SEARCH REPORT

International Application No
PCT/IL 98/00094

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	SMITH C. ET AL.: "Generation of cohesive ends on PCR products by UDG-mediated excision of dU, and application for cloning into restriction digest-linearized vectors" PCR METHODS & APPLICATIONS., vol. 2, no. 4, May 1993, ING HARBOR LABORATORY PRESS US, pages 328-332, XP002071999	20
Y	see the whole document	1-3,6,7, 9,13, 15-18,21

X	EP 0 385 410 A (CANON KK) 5 September 1990	20
Y	see abstract see figure 1 see page 7, line 45 - page 13, line 55 see examples 1-4,8-13	1-3,6,7, 9,13, 15-18,21

X	RASHTCHIAN A. ET AL.: "URACIL DNA GLYCOSYLASE-MEDIATED CLONING OF POLYMERASE CHAIN REACTION-AMPLIFIED DNA: APPLICATION TO GENOMIC AND cDNA CLONING" ANALYTICAL BIOCHEMISTRY, vol. 206, no. 1, 1 October 1992, pages 91-97, XP000311343 see the whole document	20

X	NEWTON C R ET AL: "PCR" 1994 , BIOS SCIENTIFIC PUBLISHERS , OXFORD GB XP002072707 see page 59 - page 70	20
A		1,12

A	WO 93 18175 A (LIFE TECHNOLOGIES INC) 16 September 1993 cited in the application see the whole document	1,12

A	EP 0 456 304 A (EASTMAN KODAK COMPANY) 13 November 1991 see the whole document	1,5,14

X,P	WATSON D.E., BENNETT G.N.: "Cloning and assemby of PCR products using modified primers and DNA repair enzymes" BIOTECHNIQUES., vol. 23, no. 5, 1997, NATICK US, pages 858-860, 862, 864, XP002071997 see the whole document	1,4,6,7, 9,10,13, 15-18, 20,21

2

Form PCT/ISA/210 (continuation of second sheet) (July 1992)

IDT Exhibit 1020.0423

INTERNATIONAL SEARCH REPORT

Information on patent family members

Int ational Application No
PCT/IL 98/00094

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
EP 0385410 A	05-09-1990	JP 3210195 A	13-09-1991
		JP 3232489 A	16-10-1991
		JP 3232490 A	16-10-1991
		JP 3127999 A	31-05-1991
		JP 3007583 A	14-01-1991
		AT 143696 T	15-10-1996
		DE 69028725 D	07-11-1996
		DE 69028725 T	13-03-1997
WO 9318175 A	16-09-1993	NONE	
EP 456304 A	13-11-1991	US 5387505 A	07-02-1995
		AT 130048 T	15-11-1995
		CA 2039222 A	05-11-1991
		DE 69114353 D	14-12-1995
		DE 69114353 T	18-04-1996
		FI 912141 A	05-11-1991
		HK 38296 A	15-03-1996
		JP 4228076 A	18-08-1992

IDT Exhibit 1020.0424

Espacenet

Bibliographic data: WO9725416 (A2) — 1997-07-17

NOVEL DNA GLYCOSYLASES AND THEIR USE

Inventor(s): KROKAN HANS E [NO]; SLUPPHAUG GEIR [NO]; KAVLI BODIL [NO]; TAINER JOHN A [US]; MOL CLIFFORD D [US] ± (KROKAN, HANS, E, ; SLUPPHAUG, GEIR, ; KAVLI, BODIL, ; TAINER, JOHN, A, ; MOL, CLIFFORD, D)

Applicant(s): NYFOTEK AS [NO]; DZIEGLEWSKA HANNA EVA [GB]; KROKAN HANS E [NO]; SLUPPHAUG GEIR [NO]; KAVLI BODIL [NO]; TAINER JOHN A [US]; MOL CLIFFORD D [US] ± (NYFOTEK A/S, ; DZIEGLEWSKA, HANNA, EVA, ; KROKAN, HANS, E, ; SLUPPHAUG, GEIR, ; KAVLI, BODIL, ; TAINER, JOHN, A, ; MOL, CLIFFORD, D)

Classification: - international: ***C12N15/56; C12N9/24; C12Q1/68;***
(IPC1-7): C12N15/56; C12N9/24; C12Q1/68
- cooperative: **C12N9/2497; C12Q1/6848; C12Q1/6869; C12Y302/02027** more

Application number: WO1997GB00057 19970109 Global Dossier

Priority number(s): GB19960000384 19960109

Also published as: WO9725416 (A3) AU1389597 (A) EP0873403 (A2) US2006134631 (A1) US6713294 (B1) more

Abstract of WO9725416 (A2)

Novel cytosine-, thymine- and uracil-DNA glycosylases, subcellular localization peptides, nucleic acid molecules containing the same, methods of identifying such enzymes and their use in various methods including mutagenesis, cell killing and DNA sequencing and modification, are described.

IDT Exhibit 1020.0425

PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : C12N 9/24, 15/56, C12Q 1/68

A2

(11) International Publication Number: **WO 97/25416**

(43) International Publication Date: 17 July 1997 (17.07.97)

(21) International Application Number: PCT/GB97/00057

(22) International Filing Date: 9 January 1997 (09.01.97)

(30) Priority Data:
9600384.3 9 January 1996 (09.01.96) GB

(71) Applicant *(for all designated States except US)*: NYFOTEK A/S [NO/NO]; Strindveien 4, N-7034 Trondheim (NO).

(71) Applicant *(for GB only)*: DZIEGLEWSKA, Hanna, Eva [GB/GB]; Frank B. Dehn & Co., 179 Queen Victoria Street, London EC4V 4EL (GB).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: KROKAN, Hans, E. [NO/NO]; Unigen Center for Molecular Biology, University of Trondheim, N-7005 Trondheim (NO). SLUPPHAUG, Geir [NO/NO]; Unigen Center for Molecular Biology, University of Trondheim, N-7005 Trondheim (NO). KAVLI, Bodil [NO/NO]; Unigen Center for Molecular Biology, University of Trondheim, N-7005 Trondheim (NO). TAINER, John, A. [US/US]; Department of Molecular Biology, The Scripps Research Institute, 10666 North Torrey Pines Road, La Jolla, CA 92037 (US). MOL, Clifford, D. [US/US]; Department of Molecular Biology, The Scripps Research Institute, 10666 North Torrey Pines Road, La Jolla, CA 92037 (US).

(74) Agents: DZIEGLEWSKA, Hanna, Eva et al.; Frank B. Dehn & Co., 179 Queen Victoria Street, London EC4V 4EL (GB).

(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, HU, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, TJ, TM, TR, TT, UA, UG, US, UZ, VN, ARIPO patent (KE, LS, MW, SD, SZ, UG), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).

Published
Without international search report and to be republished upon receipt of that report.

(54) Title: NOVEL DNA GLYCOSYLASES AND THEIR USE

(57) Abstract

Novel cytosine-, thymine- and uracil-DNA glycosylases, subcellular localization peptides, nucleic acid molecules containing the same, methods of identifying such enzymes and their use in various methods including mutagenesis, cell killing and DNA sequencing and modification, are described.

IDT Exhibit 1020.0426

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AM	Armenia	**GB**	United Kingdom	**MW**	Malawi
AT	Austria	**GE**	Georgia	**MX**	Mexico
AU	Australia	**GN**	Guinea	**NE**	Niger
BB	Barbados	**GR**	Greece	**NL**	Netherlands
BE	Belgium	**HU**	Hungary	**NO**	Norway
BF	Burkina Faso	**IE**	Ireland	**NZ**	New Zealand
BG	Bulgaria	**IT**	Italy	**PL**	Poland
BJ	Benin	**JP**	Japan	**PT**	Portugal
BR	Brazil	**KE**	Kenya	**RO**	Romania
BY	Belarus	**KG**	Kyrgystan	**RU**	Russian Federation
CA	Canada	**KP**	Democratic People's Republic of Korea	**SD**	Sudan
CF	Central African Republic			**SE**	Sweden
CG	Congo	**KR**	Republic of Korea	**SG**	Singapore
CH	Switzerland	**KZ**	Kazakhstan	**SI**	Slovenia
CI	Côte d'Ivoire	**LI**	Liechtenstein	**SK**	Slovakia
CM	Cameroon	**LK**	Sri Lanka	**SN**	Senegal
CN	China	**LR**	Liberia	**SZ**	Swaziland
CS	Czechoslovakia	**LT**	Lithuania	**TD**	Chad
CZ	Czech Republic	**LU**	Luxembourg	**TG**	Togo
DE	Germany	**LV**	Latvia	**TJ**	Tajikistan
DK	Denmark	**MC**	Monaco	**TT**	Trinidad and Tobago
EE	Estonia	**MD**	Republic of Moldova	**UA**	Ukraine
ES	Spain	**MG**	Madagascar	**UG**	Uganda
FI	Finland	**ML**	Mali	**US**	United States of America
FR	France	**MN**	Mongolia	**UZ**	Uzbekistan
GA	Gabon	**MR**	Mauritania	**VN**	Viet Nam

IDT Exhibit 1020.0427

NOVEL DNA GLYCOSYLASES AND THEIR USE

This invention relates to new DNA-glycosylases, in particular new cytosine-, thymine- and uracil-DNA glycosylases, and their use for mutagenesis, for DNA modification and cell killing.

Damage to DNA arises continually throughout the cell cycle and must be recognised and repaired prior to the next round of replication to maintain the genomic integrity of the cell. DNA base damage can be recognised and excised by the ATP-dependent nucleoside excision repair systems or by base excision repair systems exemplified by the DNA glycosylases.

DNA glycosylases are enzymes that occur normally in cells. They release bases from DNA by cleaving the bond between deoxyribose and the base in DNA. Naturally occurring glycosylases remove damaged or incorrectly placed bases. This base excision repair pathway is the major cellular defence mechanism against spontaneous DNA damage.

DNA glycosylases which have been identified are directed to specific bases or modified bases. An example of a DNA glycosylase which recognizes an unmodified base is uracil DNA glycosylase (UDG), which specifically recognises uracil in DNA and initiates base excision repair by hydrolysing the N-C1'glycosylic bond linking the uracil base to the deoxyribose sugar. This creates an abasic site that is removed by a 5'-acting apurinic/apyrimidic (AP) endonuclease and a deoxyribophosphodiesterase, leaving a gap which is filled by DNA polymerase and closed by DNA ligase.

The activity of UDG serves to remove uracil which arises in DNA as a result of incorporation of dUMP instead of dTMP during replication or from the spontaneous deamination of cytosine. Deamination of cytosine to uracil creates a premutagenic U:G mismatch

that, unless repaired, will cause a GC → AT transition mutation.

In vivo, UDGs specifically recognise and remove uracil from within DNA and cleave the glycosylic bond to initiate the uracil excision pathway. In vitro, UDG's can recognise and remove uracil from both single stranded DNA (ssDNA) and double-stranded DNA (dsDNA) substrates.

UDGs are ubiquitous enzymes and have been isolated from a number of sources. Amino acid sequencing reveals that the enzymes are conserved throughout evolution with greater than about 55% amino acid identity between human and bacterial proteins. A cDNA for human UDG has been cloned and the corresponding gene has been named *UNG* (Olsen *et al.* (1989) EMBO J., 8: 3121-3125).

The crystal structures of the human enzyme (Mol et al., (1995) Cell, 80: 869-878) and the herpes simplex virus enzymes (Savva *et al.* (1995) Nature, 373: 487-493) have recently been determined and reveal that uracil binds in a rigid pocket at the base of the DNA binding groove of human UDG. The absolute specificity of the enzyme for uracil over the structurally related DNA bases thymine and cytosine is conferred by shape complementarity, as well as main chain and side chain hydrogen bonds.

Although UDG's do not have activity against other bases as a result of the afore-mentioned specific spatial and charge characteristics of the active site, other glycosylases with different activities have been identified, which may or may not be restricted to single substrates.

A naturally-occurring thymine-DNA glycosylase has been identified which in addition to releasing thymine also releases uracil (Nedderman & Jiricny (1993) J. Biol. Chem., 268: 21218-21114 ; Nedderman & Jiricny (1994) J. Proc. Natl. Acad. Sci. U.S.A., 91: 1642-1646). This thymine-DNA glycosylase however has activity in

respect of only certain substrates and has an absolute requirement for a mismatched U or T opposite of a G in a double-stranded substrate and will not recognise T or U from T(U):A matches or a single-stranded substrate. DNA glycosylases which recognize and release unmodified bases other than uracil and thymine (in certain substrates, as mentioned above) have not been identified.

A DNA glycosylase recognizing unmodified cytosine has not been reported, although a 5-hydroxymethylcytosine-DNA glycosylase activity was detected in mammalian cells (Cannon *et al.* (1988) Biochem. Biophys. Res. Comm., 151: 1173-1179). The sequences of the afore-mentioned thymine and 5-hydroxymethylcytosine DNA glycosylases have not yet been reported and it is unknown whether their active site may be structurally related to UDG.

It has now surprisingly been found that the substitution of certain of the UDG amino acids has a profound effect on the substrate specificity of the glycosylase. In particular, the replacement of Asn204 by Asp204 results in the production of a mutant enzyme which has acquired cytosine-DNA glycosylase (CDG) activity, while retaining some UDG-activity. Alternatively, replacing Tyr147 with Ala147 allows for binding of thymine, resulting in an enzyme that has acquired thymine-DNA glycosylase (TDG) activity.

These new DNA glycosylases are not product-inhibited by added uracil, in contrast to UDG and other UDG-mutants. Compared with the efficiency of wild type UDG in removal of uracil, the activity of the new DNA glycosylases that remove normal pyrimidines in DNA is low, but distinct and easily detectable. However, it should be noted that the very high turnover of UDG appears to be unique among DNA glycosylases and turnover numbers of other DNA glycosylases may be as low, or even lower than those of the engineered glycosylases CDG and

TDG. This may result from the narrow substrate specificity of UDG.

Furthermore, an additional new UDG has been identified. The complete sequence of the *UNG* gene was recently published (Haug et al., 1996, Genomics, 36, p408-416). As mentioned previously, cDNA to this *UNG* gene has been identified by Olsen et al., 1989, supra (hereinafter referred to as *UNG1* cDNA and the expressed protein referred to as UNG1). It has now surprisingly been found that alternative splicing of the genomic DNA (*UNG*) with an exon located 5' of exon 1 which was not previously recognized results in a new distinct cDNA with an open reading frame of 313 amino acids. The new *UNG* cDNA is referred to hereinafter as *UNG2* cDNA, and the product which it encodes, UNG2. The latter protein has a predicted size of 36kDa.

UNG2 differs from the previously known form (UNG1, ORF 304 amino acid residues) in the 44 amino acids of the N-terminal presequence, which is not necessary for catalytic activity. The rest of the presequence and the catalytic domain, altogether 269 amino acids, are identical. The alternative presequence in UNG2 arises by splicing of a previously unrecognized exon (exon 1A) into a consensus splice site after codon 35 in exon 1B (previously designated exon 1). The UNG1 presequence starts at codon 1 in exon 1B and thus has 35 amino acids not present in UNG2. Coupled transcription/translation in rabbit reticulocyte lysates demonstrated that both proteins are catalytically active. Similar forms of UNG1 and UNG2 are expressed in mouse which has an identical organization of the homologous gene. Furthermore, the presequence of a putative *Xiphophorus* UNG2 protein predicted from the gene structure is homologous to mammalian UNG2, but much shorter, suggesting a very high degree of conservation from fish to man.

The invention therefore provides a DNA glycosylase

capable of releasing cytosine bases from single stranded (ss) DNA and/or double stranded (ds) DNA or thymine bases from both single stranded (ss) DNA and double stranded (ds) DNA or from single stranded (ss) DNA or uracil bases from single stranded (ss) DNA and/or double stranded (ds) DNA, wherein said uracil-DNA glycosylase is encoded by a nucleic acid molecule comprising the sequence (SEQUENCE I.D. Nos 1 and 2):

```
1   CACAGCCACA GCCAGGGCTA GCCTCGCCGG TTCCCGGGTG GCGCGCGTTC GCTGCCTCCT

61  CAGCTCCAGG ATGATCGGCC AGAAGACGCT CTACTCCTTT TTCTCCCCCA GCCCCGCCAG
                M  I  G  Q  K  T  L  Y  S  F  F  S  P  S  P  A

121 GAAGCGACAC GCCCCCAGCC CCGAGCCGGC CGTCCAGGGG ACCGGCGTGG CTGGGGTGCC
    R  K  R  H  A  P  S  P  E  P  A  V  Q  G  T  G  V  A  G  V

181 TGAGGAAAGC GGAGATGCGG CGGCCATCCC AGCCAAGAAG GCCCCGGCTG GGCAGGAGGA
    P  E  E  S  G  D  A  A  A  I  P  A  K  K  A  P  A  G  Q  E

241 GCCTGGGACG CCGCCCTCCT CGCCGCTGAG TGCCGAGCAG TTGGACCGGA TCCAGAGGAA
    E  P  G  T  P  P  S  S  P  L  S  A  E  Q  L  D  R  I  Q  R

301 CAAGGCCGCG GCCCTGCTCA GACTCGCGGC CCGCAACGTG CCCGTGGGCT TTGGAGAGAG
    N  K  A  A  A  L  L  R  L  A  A  R  N  V  P  V  G  F  G  E

361 CTGGAAGAAG CACCTCAGCG GGGAGTTCGG GAAACCGTAT TTTATCAAGC TAATGGGATT
    S  W  K  K  H  L  S  G  E  F  G  K  P  Y  F  I  K  L  M  G

421 TGTTGCAGAA GAAAGAAAGC ATTACACTGT TTATCCACCC CCACACCAAG TCTTCACCTG
    F  V  A  E  E  R  K  H  Y  T  V  Y  P  P  P  H  Q  V  F  T

481 GACCCAGATG TGTGACATAA AAGATGTGAA GGTTGTCATC CTGGGACAGG ATCCATATCA
    W  T  Q  M  C  D  I  K  D  V  K  V  V  I  L  G  Q  D  P  Y

541 TGGACCTAAT CAAGCTCACG GGCTCTGCTT TAGTGTTCAA AGGCCTGTTC CGCCTCCGCC
    H  G  P  N  Q  A  H  G  L  C  F  S  V  Q  R  P  V  P  P  P

601 CAGTTTGGAG AACATTTATA AAGAGTTGTC TACAGACATA GAGGATTTTG TTCATCCTGG
    P  S  L  E  N  I  Y  K  E  L  S  T  D  I  E  D  F  V  H  P

661 CCATGGAGAT TTATCTGGGT GGGCCAAGCA AGGTGTTCTC CTTCTCAACG CTGTCCTCAC
    G  H  G  D  L  S  G  W  A  K  Q  G  V  L  L  L  N  A  V  L

721 GGTTCGTGCC CATCAAGCCA ACTCTCATAA GGAGCGAGGC TGGGAGCAGT TCACTGATGC
    T  V  R  A  H  Q  A  N  S  H  K  E  R  G  W  E  Q  F  T  D

781 AGTTGTGTCC TGGCTAAATC AGAACTCGAA TGGCCTTGTT TTCTTGCTCT GGGGCTCTTA
    A  V  V  S  W  L  N  Q  N  S  N  G  L  V  F  L  L  W  G  S
```

```
841  TGCTCAGAAG AAGGGCAGTG CCATTGATAG GAAGCGGCAC CATGTACTAC AGACGGCTCA
     Y  A  Q  K  K  G  S  A  I  D  R  K  R  H  H  V  L  Q  T  A

901  TCCCTCCCCT TTGTCAGTGT ATAGAGGGTT CTTTGGATGT AGACACTTTT CAAAGACCAA
     H  P  S  P  L  S  V  Y  R  G  F  F  G  C  R  H  F  S  K  T

961  TGAGCTGCTG CAGAAGTCTG GCAAGAAGCC CATTGACTGG AAGGAGCTGT GATCATCAGC
     N  E  L  L  Q  K  S  G  K  K  P  I  D  W  K  E  L

1021 TGAGGGGTGG CCTTTGAGAA GCTGCTGTTA ACGTATTTGC CAGTTACGAA GTTCCACTGA

1081 AAATTTTCCT ATTAATTCTT AAGTACTCTG CATAAGGGGG AAAAGCTTCC AGAAAGCAGC

1141 CATGAACCAG GCTGTCCAGG AATGGCAGCT GTATCCAACC ACAAACAACA AAGGCTACCC

1201 TTTGACCAAA TGTCTTTCTC TGCAACATGG CTTCGGCCTA AAATATGCAG AAGACAGATG

1261 AGGTCAAATA CTCAGTTGGC TCTCTTTATC TCCCTTGCCT TTATGGTGAA ACAGGGGAGA

1321 TGTGCACCTT TCAGGCACAG CCCTAGTTTG GCGCCTGCTG CTCCTTGGTT TTGCCTGGTT

1381 AGACTTTCAG TGACAGATGT TGGGGTGTTT TTGCTTAGAA AGGTCCCCTT GTCTCAGCCT

1441 TGCAGGGCAG GCATGCCAGT CTCTGCCAGT TCCACTGCCC CCTTGATCTT TGAAGGAGTC

1501 CTCAGGCCCC TCGCAGCATA AGGATGTTTT GCAACTTTCC AGAATCTGGC CCAGAAATTA

1561 GGGCTCAATT TCCTGATTGT AGTAGAGGTT AAGATTGCTG TGAGCTTTAT CAGATAAGAG

1621 ACCGAGAGAA GTAAGCTGGG TCTTGTTATT CCTTGGGTGT TGGTGGAATA AGCAGTGGAA

1681 TTTGAACAAG GAAGAGGAGA AAAGGGAATT TTGTCTTTAT GGGGTGGGGT GATTTTCTCC

1741 TAGGGTTATG TCCAGTTGGG GTTTTTAAGG CAGCACAGAC TGCCAAGTAC TGTTTTTTTT

1801 AACCGACTGA AATCACTTTG GGATATTTTT TCCTGCAACA CTGGAAAGTT TTAGTTTTTT

1861 AAGAAGTACT CATGCAGATA TATATATATA TATTTTTCCC AGTCCTTTTT TTAAGAGACG

1921 GTCTTTATTG GGTCTGCACC TCCATCCTTG ATCTTGTTAG CAATGCTGTT TTTGCTGTTA

1981 GTCGGGTTAG AGTTGGCTCT ACGCGAGGTT TGTTAATAAA AGTTTGTTAA AAGTTCAAAA

2041 AAAAAAAAAA AAA
```

or a fragment thereof encoding a catalytically active product comprising at least nucleotides 121 to 130, preferably 71 to 202 in addition to the catalytic domain, or a sequence which is degenerate, substantially homologous with or which hybridizes with at least nucleotides 121 to 130, preferably 71 to 202 of any such

aforesaid sequence.

In particular, viewed from one aspect, the invention can be seen as providing a cytosine-DNA glycosylase (CDG) capable of releasing cytosine bases from ssDNA and/or dsDNA.

A further aspect of the invention provides a cytosine-DNA glycosylase (CDG) capable of releasing both cytosine and uracil bases from ssDNA and/or dsDNA.

Preferably, the cytosine-DNA glycosylase is one derived from a UDG and especially from the human UDG protein which has Asn at amino acid position 204. In particular, the novel CDG of the invention is preferably derived from human UDG and has an amino acid substitution or modification at position 204. Modification of UDG from other species at an equivalent residue is similarly preferred. Especially preferably, the glycosylase is human UDG having an aspartic acid residue (Asp) at position 204.

Another aspect of the invention provides a thymine-DNA glycosylase (TDG) capable of releasing thymine bases from both ssDNA and dsDNA.

A further aspect of the invention provides a thymine-DNA glycosylase (TDG) capable of releasing both thymine and uracil bases from both ssDNA and dsDNA.

Yet further aspects of the invention provide a thymine-DNA glycosylase (TDG) capable of releasing thymine bases from A:T DNA pairs and a thymine-DNA glycosylase (TDG) capable of releasing thymine bases from single stranded DNA.

Preferably, the thymine-DNA glycosylase is one derived from a UDG, and especially from the human UDG protein which has Tyr at amino acid position 147. In particular, the novel CDG of the invention is preferably derived from human UDG and has an amino acid substitution or modification at position 147. Modification of UDG from other species at an equivalent residue is similarly preferred. Especially preferably,

the glycosylase is human UDG having a alanine residue (Ala) at position 147.

A yet further aspect of the invention provides a uracil-DNA glycosylase encoded by a nucleic acid molecule comprising the sequence (SEQUENCE I.D Nos 1 and 2):

```
1   CACAGCCACA GCCAGGGCTA GCCTCGCCGG TTCCCGGGTG GCGCGCGTTC GCTGCCTCCT

61  CAGCTCCAGG ATGATCGGCC AGAAGACGCT CTACTCCTTT TTCTCCCCCA GCCCCGCCAG
                M  I  G  Q  K  T  L  Y  S  F  F  S  P  S  P  A

121 GAAGCGACAC GCCCCCAGCC CCGAGCCGGC CGTCCAGGGG ACCGGCGTGG CTGGGGTGCC
    R  K  R  H  A  P  S  P  E  P  A  V  Q  G  T  G  V  A  G  V

181 TGAGGAAAGC GGAGATGCGG CGGCCATCCC AGCCAAGAAG GCCCCGGCTG GGCAGGAGGA
    P  E  E  S  G  D  A  A  A  I  P  A  K  K  A  P  A  G  Q  E

241 GCCTGGGACG CCGCCCTCCT CGCCGCTGAG TGCCGAGCAG TTGGACCGGA TCCAGAGGAA
    E  P  G  T  P  P  S  S  P  L  S  A  E  Q  L  D  R  I  Q  R

301 CAAGGCCGCG GCCCTGCTCA GACTCGCGGC CCGCAACGTG CCCGTGGGCT TTGGAGAGAG
    N  K  A  A  A  L  L  R  L  A  A  R  N  V  P  V  G  F  G  E

361 CTGGAAGAAG CACCTCAGCG GGGAGTTCGG GAAACCGTAT TTTATCAAGC TAATGGGATT
    S  W  K  K  H  L  S  G  E  F  G  K  P  Y  F  I  K  L  M  G

421 TGTTGCAGAA GAAAGAAAGC ATTACACTGT TTATCCACCC CCACACCAAG TCTTCACCTG
    F  V  A  E  E  R  K  H  Y  T  V  Y  P  P  P  H  Q  V  F  T

481 GACCCAGATG TGTGACATAA AAGATGTGAA GGTTGTCATC CTGGGACAGG ATCCATATCA
    W  T  Q  M  C  D  I  K  D  V  K  V  V  I  L  G  Q  D  P  Y

541 TGGACCTAAT CAAGCTCACG GGCTCTGCTT TAGTGTTCAA AGGCCTGTTC CGCCTCCGCC
    H  G  P  N  Q  A  H  G  L  C  F  S  V  Q  R  P  V  P  P  P

601 CAGTTTGGAG AACATTTATA AAGAGTTGTC TACAGACATA GAGGATTTTG TTCATCCTGG
    P  S  L  E  N  I  Y  K  E  L  S  T  D  I  E  D  F  V  H  P

661 CCATGGAGAT TTATCTGGGT GGGCCAAGCA AGGTGTTCTC CTTCTCAACG CTGTCCTCAC
    G  H  G  D  L  S  G  W  A  K  Q  G  V  L  L  L  N  A  V  L

721 GGTTCGTGCC CATCAAGCCA ACTCTCATAA GGAGCGAGGC TGGGAGCAGT TCACTGATGC
    T  V  R  A  H  Q  A  N  S  H  K  E  R  G  W  E  Q  F  T  D

781 AGTTGTGTCC TGGCTAAATC AGAACTCGAA TGGCCTTGTT TTCTTGCTCT GGGGCTCTTA
    A  V  V  S  W  L  N  Q  N  S  N  G  L  V  F  L  L  W  G  S

841 TGCTCAGAAG AAGGGCAGTG CCATTGATAG GAAGCGGCAC CATGTACTAC AGACGGCTCA
    Y  A  Q  K  K  G  S  A  I  D  R  K  R  H  H  V  L  Q  T  A
```