capture probes are extended in the presence of polymerase and dNTPs. In some embodiments the extended capture probes are positively selected to generate a sample that is enriched for extended capture probes. In another embodiment extended capture probes are enriched by depleting non-extended products.

In another embodiment the capture probes comprise a first common sequence, a tag sequence, a target sequences and a recognition sequence for a Type IIs restriction enzyme (see, Figs. 4a and 4b, SEQ ID NOS: 4-12). The Type IIs recognition site is inserted within the target specific region so that there is target specific sequence on either side of the Type IIs recognition sequence and the tag sequence is 3' of the common sequence. In many embodiments there will be one or more mismatches between the probe and the target at the site of the Type IIs site. In some embodiments the Type IIs site is positioned so that when the fragment is digested the enzyme cuts between the polymorphic position and the base just 5' of the polymorphic position. The nucleic acid sample is fragmented and ligated to adapters comprising a second common sequence. The capture probes and adapter-ligated fragments are mixed under conditions that allow hybridization and the capture probes are extended. The extended capture probes are then made double stranded using a primer that is complementary to the adapter. The double stranded extended capture probes are amplified using primers to the common sequence in the capture probe and the common sequence in the adaptor.

To detect the allele or alleles present the amplified fragments are digested with a Type IIs restriction endonuclease and the fragments (Fig. 4b) are extended in the presence of labeled ddNTPs. The fragments will be extended by a single ddNTP which corresponds to the allele present at the polymorphic position. The extended fragments are hybridized to an array of tag probes and the labeled nucleotide or nucleotides present at each location are determined. In one embodiment the ddNTPs are all labeled with the same label, for example, biotin and the fragments are extended in four separate reactions, one for each of the four different ddNTPs. Each reaction is hybridized to a different array so four arrays are used. In another embodiment the ddNTPs are labeled with differentially detectable labels. In one embodiment there are four different labels and the extension reaction may be done in a single reaction and the hybridization may

be to a single array. In another embodiment there are two different labels and extension reaction may be done in two reactions and the hybridization may be to two different arrays.

In many embodiments of the present methods one or more enrichment step may be included to generate a sample that is enriched for extended capture probes prior to amplification with common sequence primers (*see*, Figures 5-7). In some embodiments it is desirable to separate extended capture probes from fragments from the starting nucleic acid sample, adapter-ligated fragments, adapter sequences or non-extended capture probes, for example. In one embodiment (Fig. 5) the capture probes are extended in the presence of a labeled dNTP, for example dNTPs labeled with biotin. The labeled nucleotides are incorporated into the extended capture probes and the labeled extended capture probes are then separated from non-extended material by affinity chromatography. When the label is biotin the labeled extended capture probes can be isolated based on the affinity of biotin for avidin, streptavidin or a monoclonal anti-biotin antibody. In one embodiment the antibody may be coupled to protein-A agarose, protein-A sepharose or any other suitable solid support known in the art. Those of skill in the art will appreciate that biotin is one label that may be used but any other suitable label or a combination of labels may also be used, such as fluorescein which may be incorporated in the extended capture probe and an anti-fluorescein antibody may be used for affinity purification of extended capture probes. Other labels such as, digoxigenin, Cyanine-3, Cyanine-5, Rhodamine, and Texas Red may also be used. Antibodies to these labeling compounds may be used for affinity purification. Also, other haptens conjugated to dNTPs may be used, such as, for example, dinitrophenol (DNP).

In another embodiment (Fig. 6) capture probes that have been extended through the adapter sequence (A2) on the adapter modified DNA are made double stranded by hybridizing and extending A2 primer. Only the fully extended capture probes will have the A2 priming site so partially extended capture probes will remain single-stranded. The sample is then digested with a nuclease that selectively digests single stranded

nucleic acid, such as E. Coli Exonuclease I. The sample is then amplified with primers A1 and A2.

In another embodiment (Fig. 7) extension products may be enriched by circularization followed by digestion with a nuclease such as Exonuclease VII or Exonuclease III. The extended capture probes may be circularized, for example, by hybridizing the ends of the extended capture probe to an oligonucleotide splint so that the ends are juxtaposed and ligating the ends together. The splint will hybridize to the A1 and A2 sequences in the extended capture probe and bring the 5' end of the capture probe next to the 3' end of the capture probe so that the ends may be ligated by a ligase, for example DNA Ligase or Ampligase Thermostable DNA. *See*, for example, U.S. Patent No. 5,871,921 which is incorporated herein by reference. The circularized product will be resistant to nucleases that require either a free 5' or 3' end.

A variety of nucleases may be used in one or more of the embodiments. Nucleases that are commercially available and may be useful in the present methods include: Mung Bean Nuclease, E. Coli Exonuclease I, Exonuclease III, Exonuclease VII, T7 Exonuclease, BAL-31 Exonuclease, Lambda Exonuclease, $RecJ_f$, and Exonuclease T. Different nucleases have specificities for different types of nucleic acids making them useful for different applications. Exonuclease I catalyzes the removal of nucleotides from single-stranded DNA in the 3′ to 5′ direction. Exonuclease I degrades excess single-stranded primer oligonucleotide from a reaction mixture containing double-stranded extension products. Exonuclease III catalyzes the stepwise removal of mononucleotides from 3′ -hydroxyl termini of duplex DNA. A limited number of nucleotides are removed during each binding event, resulting in coordinated progressive deletions within the population of DNA molecules. The preferred substrates are blunt or recessed 3′ -termini, although the enzyme also acts at nicks in duplex DNA to produce single-strand gaps. The enzyme is not active on single-stranded DNA, and thus 3′ -protruding termini are resistant to cleavage. The degree of resistance depends on the length of the extension, with extensions 4 bases or longer being essentially resistant to cleavage. This property can be exploited to produce unidirectional deletions from a linear molecule with one resistant (3′ -overhang) and

one susceptible (blunt or 5′ -overhang) terminus. Exonuclease VII is a single-strand directed enzyme with 5' to 3'- and 3' to 5'-exonuclease activities making it the only bi-directional *E. coli* exonuclease with single-strand specificity. The enzyme has no apparent requirement for divalent cation, and is fully active in the presence of EDTA. Initial reaction products are acid-insoluble oligonucleotides which are further hydrolyzed into acid-soluble form. The products of limit digests are small oligomers (dimers to dodecamers). For additional information about nucleases see catalogues from manufacturers such as New England Biolabs, Beverly, MA.

In some embodiments one of the primers added for PCR amplification is modified so that it is resistant to nuclease digestion, for example, by the inclusion of phosphorothioate. Prior to hybridization to an array one strand of the double stranded fragments may be digested by a 5' to 3' exonuclease such as T7 Gene 6 Exonuclease.

In some embodiments the nucleic acid sample, which may be, for example, genomic DNA, is fragmented, using for example, a restriction enzyme, DNase I or a non-specific fragmentation method such as that disclosed in U.S. patent application No. 09/358,664, which is incorporated herein by reference in its entirety. Adapters containing at least one priming site are ligated to the fragmented DNA. Locus-specific primers are synthesized which contain a different adapter sequence at the 5' end. The adapter-ligated genomic DNA is hybridized to the locus-specific primers and the locus specific primer is extended. This may be done for example, by the addition of DNA polymerase and dNTPs. Extension products may be amplified with primers that are specific for the adapter sequences. This allows amplification of a collection of many different sequences using a limited set of primers. For example, a single set of primers may be used for amplification. In another embodiment a second amplification step is carried out using the same or different primers.

In some embodiments the amplified products are analyzed by hybridization to an array of probes attached to a solid support. In some embodiments an array of probes is specifically designed to interrogate a collection of target sequences. The array of probes may interrogate, for example, from 1,000, 5,000, 10,000 or 100,000 to 2,000, 5,000, 10,000, 100,000, 1,000,000 or 3,000,000 different target sequences. In one

embodiment the target sequences contain SNPs and the array of probes is designed to interrogate the allele or alleles present at one or more polymorphic location. The array may comprise a collection of probes that hybridize specifically to one or more SNP containing sequences. The array may comprise probes that correspond to different alleles of the SNP. One probe or probe set may hybridize specifically to a first allele of a SNP, but not hybridize significantly to other alleles of the SNP and a second probe set may be designed to hybridize to a second allele of a SNP but not hybridize significantly to other alleles. A hybridization pattern from the array indicates which of the alleles are present in the sample. An array may contain probe sets to interrogate, for example, from 1,000, 5,000, 10,000 or 100,000 to 2,000, 5,000, 10,000, 100,000, 1,000,000 or 3,000,000 different SNPs.

In another embodiment an array of probes that are complementary to tag sequences present in the capture probes is used to interrogate the target sequences. In some embodiments the amplified targets are analyzed on an array of tag sequences, for example, the Affymetrix GenFlex® array (Affymetrix, Inc., Santa Clara, CA). In this embodiment the capture probes comprise a tag sequence that is unique for each species of capture probe. A detectable label that is indicative of the allele present at the polymorphic site of interest is associated with the tag. The labeled tags are hybridized to the one or more arrays and the hybridization pattern is analyzed to determine which alleles are present.

In another embodiment methods for generating a plurality of different oligonucleotides are disclosed. Oligonucleotides are synthesized in parallel on a solid support. The oligonucleotides are then released from the solid support and used for further analysis. The released probes may be used, for example, for multiplex PCR amplification of a collection of target sequences, for probes, for primers for reverse transcription or amplification or for any other use of oligonucleotides known in the art. In one embodiment the oligonucleotides on the solid support comprise a collection of capture probes.

In another embodiment kits that are useful for the present methods are disclosed. In one embodiment a kit for amplifying a collection of target sequences is disclosed.

The kit may comprise one or more of the following: a collection of capture probes as disclosed, one or more adapter, one or more generic primers for common sequences, one or more restriction enzymes, buffer, one or more polymerase, a ligase, buffer, dNTPs, ddNTPs, and one or more nucleases. The restriction enzyme of the kit may be a type-IIs enzyme. The capture probes may be attached to a solid support.

METHODS OF USE

The methods of the presently claimed invention can be used for a wide variety of applications. Any analysis of genomic DNA may be benefited by a reproducible method of complexity management. Furthermore, the methods and enriched fragments of the presently claimed invention are particularly well suited for study and characterization of extremely large regions of genomic DNA.

In a preferred embodiment, the methods of the presently claimed invention are used for SNP discovery and to genotype individuals. For example, any of the procedures described above, alone or in combination, could be used to isolate the SNPs present in one or more specific regions of genomic DNA. Selection probes could be designed and manufactured to be used in combination with the methods of the invention to amplify only those fragments containing regions of interest, for example a region known to contain a SNP. Arrays could be designed and manufactured on a large scale basis to interrogate only those fragments containing the regions of interest. Thereafter, a sample from one or more individuals would be obtained and prepared using the same techniques which were used to prepare the selection probes or to design the array. Each sample can then be hybridized to an array and the hybridization pattern can be analyzed to determine the genotype of each individual or a population of individuals. Methods of use for polymorphisms and SNP discovery can be found in, for example, in US Patent No. 6,361,947 and co-pending US application No. 08/813,159 which are herein incorporated by reference in their entirety for all purposes).

Correlation of Polymorphisms with Phenotypic Traits

Most human sequence variation is attributable to or correlated with SNPs, with the rest attributable to insertions or deletions of one or more bases, repeat length polymorphisms and rearrangements. On average, SNPs occur every 1,000-2,000 bases when two human chromosomes are compared. (*See*, The International SNP Map Working Group, *Science* 409: 928-933 (2001) incorporated herein by reference in its entirety for all purposes.) Human diversity is limited not only by the number of SNPs occurring in the genome but further by the observation that specific combinations of alleles are found at closely linked sites.

Correlation of individual polymorphisms or groups of polymorphisms with phenotypic characteristics is a valuable tool in the effort to identify DNA variation that contributes to population variation in phenotypic traits. Phenotypic traits include physical characteristics, risk for disease, and response to the environment. Polymorphisms that correlate with disease are particularly interesting because they represent mechanisms to accurately diagnose disease and targets for drug treatment. Hundreds of human diseases have already been correlated with individual polymorphisms but there are many diseases that are known to have an, as yet unidentified, genetic component and many diseases for which a component is or may be genetic.

Many diseases may correlate with multiple genetic changes making identification of the polymorphisms associated with a given disease more difficult. One approach to overcome this difficulty is to systematically explore the limited set of common gene variants for association with disease.

To identify correlation between one or more alleles and one or more phenotypic traits, individuals are tested for the presence or absence of polymorphic markers or marker sets and for the phenotypic trait or traits of interest. The presence or absence of a set of polymorphisms is compared for individuals who exhibit a particular trait and individuals who exhibit lack of the particular trait to determine if the presence or absence of a particular allele is associated with the trait of interest. For example, it might be found that the presence of allele A1 at polymorphism A correlates with heart

disease. As an example of a correlation between a phenotypic trait and more than one polymorphism, it might be found that allele A1 at polymorphism A and allele B1 at polymorphism B correlate with a phenotypic trait of interest.

Diagnosis of Disease and Predisposition to Disease

Markers or groups of markers that correlate with the symptoms or occurrence of disease can be used to diagnose disease or predisposition to disease without regard to phenotypic manifestation. To diagnose disease or predisposition to disease, individuals are tested for the presence or absence of polymorphic markers or marker sets that correlate with one or more diseases. If, for example, the presence of allele A1 at polymorphism A correlates with coronary artery disease then individuals with allele A1 at polymorphism A may be at an increased risk for the condition.

Individuals can be tested before symptoms of the disease develop. Infants, for example, can be tested for genetic diseases such as phenylketonuria at birth. Individuals of any age could be tested to determine risk profiles for the occurrence of future disease. Often early diagnosis can lead to more effective treatment and prevention of disease through dietary, behavior or pharmaceutical interventions. Individuals can also be tested to determine carrier status for genetic disorders. Potential parents can use this information to make family planning decisions.

Individuals who develop symptoms of disease that are consistent with more than one diagnosis can be tested to make a more accurate diagnosis. If, for example, symptom S is consistent with diseases X, Y or Z but allele A1 at polymorphism A correlates with disease X but not with diseases Y or Z an individual with symptom S is tested for the presence or absence of allele A1 at polymorphism A. Presence of allele A1 at polymorphism A is consistent with a diagnosis of disease X. Genetic expression information discovered through the use of arrays has been used to determine the specific type of cancer a particular patient has. (*See*, Golub et al. Science 286: 531-537 (2001) hereby incorporated by reference in its entirety for all purposes.)

Pharmacogenomics

Pharmacogenomics refers to the study of how genes affect response to drugs. There is great heterogeneity in the way individuals respond to medications, in terms of both host toxicity and treatment efficacy. There are many causes of this variability, including: severity of the disease being treated; drug interactions; and the individuals age and nutritional status. Despite the importance of these clinical variables, inherited differences in the form of genetic polymorphisms can have an even greater influence on the efficacy and toxicity of medications. Genetic polymorphisms in drug-metabolizing enzymes, transporters, receptors, and other drug targets have been linked to interindividual differences in the efficacy and toxicity of many medications. (*See*, Evans and Relling, *Science* 286: 487-491 (2001) which is herein incorporated by reference for all purposes).

An individual patient has an inherited ability to metabolize, eliminate and respond to specific drugs. Correlation of polymorphisms with pharmacogenomic traits identifies those polymorphisms that impact drug toxicity and treatment efficacy. This information can be used by doctors to determine what course of medicine is best for a particular patient and by pharmaceutical companies to develop new drugs that target a particular disease or particular individuals within the population, while decreasing the likelihood of adverse affects. Drugs can be targeted to groups of individuals who carry a specific allele or group of alleles. For example, individuals who carry allele A1 at polymorphism A may respond best to medication X while individuals who carry allele A2 respond best to medication Y. A trait may be the result of a single polymorphism but will often be determined by the interplay of several genes.

In addition some drugs that are highly effective for a large percentage of the population prove dangerous or even lethal for a very small percentage of the population. These drugs typically are not available to anyone. Pharmacogenomics can be used to correlate a specific genotype with an adverse drug response. If pharmaceutical companies and physicians can accurately identify those patients who would suffer adverse responses to a particular drug, the drug can be made available on a limited basis to those who would benefit from the drug.

Similarly, some medications may be highly effective for only a very small percentage of the population while proving only slightly effective or even ineffective to a large percentage of patients. Pharmacogenomics allows pharamaceutical companies to predict which patients would be the ideal candidate for a particular drug, thereby dramatically reducing failure rates and providing greater incentive to companies to continue to conduct research into those drugs.

Determination of Relatedness

There are many circumstances where relatedness between individuals is the subject of genotype analysis and the present invention can be applied to these procedures.

Paternity testing is commonly used to establish a biological relationship between a child and the putative father of that child. Genetic material from the child can be analyzed for occurrence of polymorphisms and compared to a similar analysis of the putative father's genetic material. Determination of relatedness is not limited to the relationship between father and child but can also be done to determine the relatedness between mother and child, (*see* e.g. Staub et al., U.S. Pat. No. 6,187,540) or more broadly, to determine how related one individual is to another, for example, between races or species or between individuals from geographically separated populations, (see for example H. Kaessmann, et al. *Nature Genet.* 22, 78 (1999)).

Forensics

The capacity to identify a distinguishing or unique set of forensic markers in an individual is useful for forensic analysis. For example, one can determine whether a blood sample from a suspect matches a blood or other tissue sample from a crime scene by determining whether the set of polymorphic forms occupying selected polymorphic sites is the same in the suspect and the sample. If the set of polymorphic markers does not match between a suspect and a sample, it can be concluded (barring experimental error) that the suspect was not the source of the sample. If the set of markers does match, one can conclude that the DNA from the suspect is consistent with that found at

the crime scene. If frequencies of the polymorphic forms at the loci tested have been determined (e.g., by analysis of a suitable population of individuals), one can perform a statistical analysis to determine the probability that a match of suspect and crime scene sample would occur by chance. A similar comparison of markers can be used to identify an individual's remains. For example the U.S. armed forces collect and archive a tissue sample for each service member. If unidentified human remains are suspected to be those of an individual a sample from the remains can be analyzed for markers and compared to the markers present in the tissue sample initially collected from that individual.

Marker Assisted Breeding

Genetic markers can assist breeders in the understanding, selecting and managing of the genetic complexity of animals and plants. Agriculture industry, for example, has a great deal of incentive to try to produce crops with desirable traits (high yield, disease resistance, taste, smell, color, texture, etc.) as consumer demand increases and expectations change. However, many traits, even when the molecular mechanisms are known, are too difficult or costly to monitor during production. Readily detectable polymorphisms which are in close physical proximity to the desired genes can be used as a proxy to determine whether the desired trait is present or not in a particular organism. This provides for an efficient screening tool which can accelerate the selective breeding process.

EXAMPLES

Example 1. Multiplexed Anchored Runoff Amplification

Genomic DNA was digested with *Mse*I and ligated to an adapter containing T7 promoter sequence as a priming site. The final concentration of the genomic DNA was 10 ng/µl in 1X T4 DNA Ligase Buffer. To generate extended capture probes 2.5 µl of adapter ligated DNA, 2.5 µl 10X Taq Gold Buffer, 2 µl 25 mM MgCl2, 2.5 µl 10X dNTPs, 5 µl of a 500 nM mixture of 150 different capture probes in TE buffer corresponding to 150 different forward primers from the HuSNP assay, 0.25 µl Perfect

Match Enhancer, 0.25 µl AmpliTaq Gold (Applied Biosystems, Foster City, CA) and 10 µl of water were mixed to give a final reaction volume of 25 µl. The reaction was incubated at 95°C for 6 min followed by 26 cycles of 95°C for 30 sec, 68°C for 2.5 min (decreasing 0.5°C on each subsequent cycle) and 72°C for 1 min, then to 4°C.

The extended capture probes were made double stranded by the addition of 0.25 µl of 1 µM T7 primer and incubation at 95°C for 2 min, 55°C for 2 min, 72°C for 6 min, then to 4°C. The reaction was passed over a G-25 Sephadex column and 5 µl of 10X Exonuclease I Buffer (NEB) and 2 µl of Exonuclease I (NEB) were added and the reaction was incubated at 37°C for 60 min, 80°C for 20 min, then to 4°C. The products were purified over a Qiagen (Valencia, CA) mini-elute column and eluted with 10 µl EB Buffer.

Generic PCR was done as follows: 65.5 µl water, 10 µl 10X Taq Gold Buffer, 8 µl 25 mM MgCl2, 10 µl 10X dNTPs, 1 µl 1 µM T3 primer, 1 µl 1 µM T7 primer 3 µl DNA, 0,5 µl Perfect Match Enhancer and 1 µl AmpliTaq Gold were mixed in a 100 µl final reaction volume and incubated at 95°C for 8 min, 40 cycles of 95°C for 30 sec, 55°C for 1 min, and 72 °C for 1 min, then 72°C for 6 min followed and finally to 4°C.

An aliquot of the reaction was analyzed on a 2% agarose gel. The products were concentrated using Qiagen QIAquick columns and eluted with 10 µl EB Buffer. The products were fragmented, labeled and hybridized to an array under standard conditions and hybridization patterns were analyzed.

Example 2.Multiplexed Anchored Runoff Amplification with Biotin Enrichment

Prepare adaptor ligated genomic DNA as above. To generate extended capture probes 2.5 µl of adapter ligated DNA, 2.5 µl 10X Taq Gold Buffer, 2 µl 25 mM $MgCl_2$, 0.5 µl 50X acGT (6 mM dATP, 6 mM dCTP, 10 mM dGTP, 10 mM dTTP), 5 µl of a 500 nM mixture of 150 different capture probes in TE buffer corresponding to 150 different forward primers from the HuSNP assay, 0.25 µl Perfect Match Enhancer, 0.25 µl Amplitaq Gold, 2 µl 1 mM Biotin-N6-dATP (Perkin Elmer, Boston, MA), 2 µl 1 mM Biotin-N4-dCTP (Perkin Elmer) and 8 µl of water were mixed to give a final reaction

volume of 25 µl. The reaction was incubated at 95°C for 6 min followed by 26 cycles of 95°C for 30 sec, 68°C for 2.5 min (decreasing 0.5°C on each subsequent cycle) and 72°C for 1 min, then to 4°C. Pass reaction over G-25 Sephadex column to remove unincorporated biotin-dNTPs.

Enrich for biotinylated extension products. Adjust the G-25 eluate to 1x PCR buffer and 2 mM $MgCl_2$. Add 15 µl monoclonal anti-biotin agarose (Clone BN-34, Sigma). Incubate at room temperature for 30 min with gentle agitation. Spin down agarose resin for 3 min at 5,000 rpm. Aspirate away supernatant and wash agarose resin with 250 µl 1x PCR buffer with 2 mM $MgCl_2$. Aliquot agarose resin into PCR tubes for generic PCR with T3 and T7 primers.

Generic PCR was done as follows: 65.5 µl water, 10 µl 10X Taq Gold Buffer, 8 µl 25 mM $MgCl_2$, 10 µl 10x dNTPs, 1 µl 1 µM T3 primer, 1 µl 1 µM T7 primer, 3 µl DNA, 0,5 µl Perfect Match Enhancer and 1 µl AmpliTaq Gold were mixed in a 100 µl final reaction volume and incubated at 95°C for 8 min, 40 cycles of 95°C for 30 sec, 55°C for 1 min, and 72 °C for 1 min, then 72°C for 6 min and finally to 4°C.
An aliquot of the reaction was analyzed on a 2% agarose gel. The products were concentrated using Qiagen QIAquick columns and eluted with 30 µl EB Buffer. The products were fragmented with DNase I, labeled with biotin-ddATP using TdT, and hybridized to an array under standard conditions. Hybridization patterns were analyzed.

Example 3. Multiplexed Anchored Runoff Amplification with Exo III Enrichment.

Prepare adaptor ligated genomic DNA as above. Kinase capture probes by incubating 12 µl of a 150-plex stock of either forward or reverse HuSNP® primers with 12.7 µl H_2O, 3 µl 10x T4 polynucleotide kinase buffer, 0.3 µl 100mM ATP, and 2 µl T4 Polynucleotide Kinase. Incubate the reaction at 37°C for 30 min. Adjust reaction volume to 50 µl and pass reaction over G-25 column to exchange buffer.

To generate extended capture probes 5 µl of adapter ligated DNA, 5 µl 10X Taq Gold Buffer, 4 µl 25 mM $MgCl_2$, 5 µl 10X dNTPs, 20 µl of the kinased mixture of 150 different capture probes, 1 µl Perfect Match Enhancer, 0.5 µl AmpliTaq Gold and 9.5 µl

of water were mixed to give a final reaction volume of 50 µl. The reaction was incubated at 95°C for 6 min followed by 26 cycles of 95°C for 30 sec, 68°C for 2.5 min (decreasing 0.5°C on each subsequent cycle) and 72°C for 1 min, then finally to 4°C. Pass the reaction over a G-25 column to exchange buffer.

Convert the single strand extension products to single strand circles using splint oligonucleotides and Ampligase Thermostable DNA Ligase (Epicenter, Madison, WI). The sequence of the T3-T7 splint oligo is (SEQ ID NO: 3) 5'TCTCCCTTTAGTGAGGGTTAATTTGTAATACGACTCACTATAGGGCA-3'. Mix 39.75 µl water, 7.5 µl 10x Ampligase Buffer, 1.25 µl 70 µM splint oligo, 25 µl 5' phosphorylated single strand extension products and 1.5 µl Ampligase Thermostable DNA Ligase 5 U/µl. Incubate the mixture at 95°C for 3 min, then 10 cycles of 95°C for 30 sec and 72°C for 3 min, then 10 cycles of 95°C for 30 sec and 70°C for 3 min, then 10 cycles of 95°C for 30 sec and 68°C for 3 min, then 10 cycles of 95°C for 30 sec and 66°C for 3 min, then 10 cycles of 95°C for 30 sec and 64°C for 3 min, then 10 cycles of 95°C for 30 sec and 62°C for 3 min. Hold at 4°C. Pass reaction over G-25 column to exchange buffer.

Digest uncircularized nucleic acids. Mix 13 µl water, 10 µl 10x Exo III Buffer, 75 µl Ampligase/splint reaction and 2 µl Exonuclease III 100 U/µl (NEB, Beverly, MA). Incubate at 37°C for 1 hour. Heat inactivate at 70°C for 20 min. Fragment, label and hybridize as above.

CONCLUSION

From the foregoing it can be seen that the present invention provides a flexible and scalable method for analyzing complex samples of DNA, such as genomic DNA. These methods are not limited to any particular type of nucleic acid sample: plant, bacterial, animal (including human) total genome DNA, RNA, cDNA and the like may be analyzed using some or all of the methods disclosed in this invention. This invention provides a powerful tool for analysis of complex nucleic acid samples. From experiment design to isolation of desired fragments and hybridization to an appropriate

array, the above invention provides for fast, efficient and inexpensive methods of complex nucleic acid analysis.

All publications and patent applications cited above are incorporated by reference in their entirety for all purposes to the same extent as if each individual publication or patent application were specifically and individually indicated to be so incorporated by reference. Although the present invention has been described in some detail by way of illustration and example for purposes of clarity and understanding, it will be apparent that certain changes and modifications may be practiced within the scope of the appended claims.

CLAIMS

What is claimed is:

1. A method of amplifying a collection of target sequences from a nucleic acid sample comprising:

generating a collection of capture probes comprising a plurality of different species of primers wherein each species comprises a first common sequence and a 3' variable region that is specific for a target sequence in said collection of target sequences, wherein said collection of capture probes is attached to a solid support and the 3' end of the capture probes is available for extension;

fragmenting the nucleic acid sample;

ligating an adapter to the fragments, said adapter comprising a second common sequence;

hybridizing the adapter-ligated fragments to the collection of capture probes;

extending the capture probes; and

amplifying the extended capture probes with first and second common sequence primers.

2. The method of claim 1 further comprising releasing the extended capture probes from the solid support prior to amplification.

3. The method of claim 2 wherein prior to releasing the extended capture probes from the solid support nucleic acids that are not covalently attached to the solid support are removed.

4. The method of claim 1 wherein said capture probes are attached to the solid support through a covalent interaction.

5. The method of claim 1 wherein the capture probes further comprise a tag sequence that is unique for each species of capture probe and the capture probes are

attached to the solid support by hybridization to a collection of tag probes that are covalently attached to the solid support.

6. The method of claim 1 wherein each species of capture probe is attached to the solid support in a discrete location.

7. The method of claim 1 wherein prior to amplification the extended capture probes are enriched in the sample to be amplified.

8. The method of claim 1 wherein labeled nucleotides are incorporated into the extended capture probes and extended capture probes are isolated by affinity chromatography.

9. The method of claim 8 wherein said labeled nucleotides are labeled with biotin and avidin, streptavidin or an anti-biotin antibody is used to isolate extended capture probes.

10. The method of claim 1 wherein prior to amplification the extended capture probes are made double stranded and single stranded nucleic acid in the sample is digested.

11. The method of claim 10 wherein the single stranded nucleic acid in the sample is digested with a nuclease.

12. The method of claim 11 wherein the nuclease is Exonuclease I.

13. The method of claim 1 wherein prior to amplification the extended capture probes are circularized and uncircularized nucleic acid in the sample is digested.

14. The method of claim 13 wherein extended capture probes are circularized by a method comprising:

hybridizing an oligonucleotide splint to the extended capture probes, wherein the oligonucleotide splint is complementary to the first and second common sequences, thereby juxtaposing the 5' and 3' ends of extended capture probes; and

ligating the ends of the extended capture probes to form circular extended capture probes.

15. The method of claim 13 wherein the uncircularized nucleic acid remaining in the sample is digested with a nuclease.

16. The method of claim 15 wherein the nuclease is Exonuclease III.

17. The method of claim 1 wherein the nucleic acid sample is fragmented by digestion with one or more restriction enzymes.

18. The method of claim 1 wherein said capture probes are synthesized on a solid support.

19. The method of claim 1 wherein there are 100 to 1500 different target sequences in the collection of target sequences.

20. The method of claim 1 wherein there are 1,000 to 5,000 different target sequences in the collection of target sequences.

21. The method of claim 1 wherein there are 2,000 to 10,000 different target sequences in the collection of target sequences.

22. The method of claim 1 wherein there are 10,000 to 1,000,000 different target sequences in the collection of target sequences.

23. A method of analyzing a nucleic acid sample comprising:

amplifying a collection of target sequences from said nucleic acid sample according to the method of claim 1;

hybridizing the amplified collection of target sequences to an array; and

analyzing the hybridization pattern to detect the presence or absence of target sequences from the collection of target sequences.

24. A method of genotyping one or more polymorphic locations in a sample comprising:

amplifying a collection of target sequences from the sample according to the method of claim 1;

hybridizing the amplified collection of target sequences to an array designed to interrogate at least one polymorphic location in the collection of target sequences; and

analyzing the hybridization pattern to determine the identify of the allele or alleles present at one or more polymorphic location in the collection of target sequences.

25. A method for analyzing sequence variations in a population of individuals comprising;

obtaining a nucleic acid sample from each individual;

amplifying a collection of target sequences from each nucleic acid sample according to the method of claim 1;

hybridizing each amplified collection of target sequences to an array designed to interrogate sequence variation in the collection of target sequences to generate a hybridization pattern for each sample; and

analyzing the hybridization patterns to determine the presence or absence of sequence variation in the population of individuals.

26. A method of amplifying a collection of target sequences from a nucleic acid sample said method comprising:

generating a collection of capture probes comprising a plurality of different species of primers wherein each species comprises a first common sequence and a 3' variable region that is specific for a target sequence in the collection of target sequences;

fragmenting the nucleic acid sample;

ligating an adapter to the fragments, wherein the adapter is ligated to the fragments so that the strand that is ligated to the 5' end of the fragment strands comprises a second common sequence and the strand that is ligated to the 3' end of the fragments lacks the complement of the second common sequence and is blocked from extension at the 3' end;

hybridizing the adapter-ligated fragments to the collection of capture probes;

extending the capture probes; and

amplifying the extended capture probes with first and second common sequence primers.

27. The method of claim 26 wherein an amino group is used to block extension at the 3' end of the adapter strand that is ligated to the 3' end of the fragments.

28. A method of analyzing a nucleic acid sample comprising:

amplifying a collection of target sequences from said nucleic acid sample according to the method of claim 26;

hybridizing the amplified collection of target sequences to an array; and

analyzing the hybridization pattern to detect the presence or absence of target sequences from the collection of target sequences.

29. A method of genotyping one or more polymorphic locations in a sample comprising:

preparing an amplified collection of target sequences from the sample according to the method of claim 26;

hybridizing the amplified collection of target sequences to an array designed to interrogate at least one polymorphic location in the collection of target sequences; and

analyzing the hybridization pattern to determine the identify of the allele or alleles present at one or more polymorphic location in the collection of target sequences.

30. A method for analyzing sequence variations in a population of individuals comprising;

obtaining a nucleic acid sample from each individual;

amplifying a collection of target sequences from each nucleic acid sample according to the method of claim 26;

hybridizing each amplified collection of target sequences to an array designed to interrogate sequence variation in the collection of target sequences to generate a hybridization pattern for each sample; and

analyzing the hybridization patterns to determine the presence or absence of sequence variation in the population of individuals.

31. The method of claim 26 wherein the nucleic acid sample is fragmented by digestion with one or more restriction enzymes.

32. The method of claim 26 wherein prior to amplification the extension products are enriched in the sample to be amplified.

33. The method of claim 26 wherein labeled nucleotides are incorporated into the extension products and the extension products are enriched by affinity chromatography.

34. The method of claim 33 wherein said labeled nucleotides are labeled with biotin and avidin, streptavidin or an anti-biotin antibody is used to isolate extension products.

35. The method of claim 26 wherein prior to amplification the extended capture probes are made double stranded and single stranded nucleic acid in the sample is digested.

36. The method of claim 35 wherein the single stranded nucleic acid in the sample is digested with a nuclease.

37. The method of claim 36 wherein the nuclease is Exonuclease I.

38. The method of claim 26 wherein prior to amplification the extended capture probes are circularized and uncircularized nucleic acid in the sample is digested.

39. The method of claim 38 wherein extended capture probes are circularized by a method comprising:

hybridizing an oligonucleotide splint to the extended capture probes, wherein the splint is complementary to the first and second common sequences, thereby juxtaposing the 5' and 3' ends of extended capture probes; and

ligating the ends of the extended capture probes to form circular extended capture probes.

40. The method of claim 38 wherein the uncircularized nucleic acid remaining in the sample is digested with a nuclease.

41. The method of claim 40 wherein the nuclease is Exonuclease III.

42. The method of claim 26 wherein there are 100 to 1,500 different target sequences in the collection of target sequences.

43. The method of claim 26 wherein there are 1,000 to 5,000 different target sequences in the collection of target sequences.

44. The method of claim 26 wherein there are 2,000 to 10,000 different target sequences in the collection of target sequences.

45. The method of claim 26 wherein there are 10,000 to 1,000,000 different target sequences in the collection of target sequences.

46. A method for genotyping one or more polymorphisms in a nucleic acid sample comprising:

fragmenting the nucleic acid sample to generate fragments;

ligating an adaptor to the fragments said adapter comprising a first common sequence;

hybridizing a collection of capture probes to the fragments, wherein said capture probes comprise a second common sequence, a tag sequence unique for each species of capture probe, a first target specific sequence, a Type IIs restriction enzyme recognition sequence, and a second target specific sequence wherein the Type IIs restriction enzyme recognition sequence is positioned so that the enzyme will cut on the 5' side of the polymorphic base;

extending said capture probes to generate extended capture probes;

amplifying the extended capture probes with first and second common sequence primers;

digesting the amplified product with a Type IIs restriction enzyme to generate amplified product fragments;

extending the amplified product fragments in at least one extension reaction;

hybridizing each extension reaction to an array comprising tag probes that hybridize to the tag sequences in the capture probes; and

analyzing the hybridization pattern on each of the arrays to determine at least one genotype.

47. The method of claim 46 wherein in each extension reaction there is at least one species of labeled ddNTP.

48. The method of claim 47 wherein one or more species of ddNTPs is labeled with biotin.

49. The method of claim 47 wherein there are four separate extension reactions wherein each extension reaction corresponds to a different species of labeled ddNTP and each extension reaction is hybridized to a different array.

50. The method of claim 47 wherein there are two separate extension reactions wherein two differentially labeled ddNTPs are present in each extension reaction and each extension reaction is hybridized to a different array.

51. The method of claim 47 wherein there is one extension reaction wherein four differentially labeled ddNTPs are present in the extension reaction and the extension reaction is hybridized to a single array.

52. The method of claim 46 wherein said capture probes are attached to a solid support.

53. The method of claim 52 wherein said capture probes are attached to the solid support through a covalent interaction.

54. The method of claim 52 wherein said capture probes are attached to said solid support by hybridization to a collection of tag probes that are attached to said solid support.

55. The method of claim 52 wherein each species of capture probe is attached to said solid support in a discrete location.

56. The method of claim 52 wherein said capture probes are synthesized on a solid support in a 5' to 3' direction.

57. The method of claim 46 wherein one of the common sequence primers is resistant to nuclease digestion and each extension reaction is digested with a 5' to 3' nuclease activity prior to hybridization to an array.

58. The method of claim 57 wherein the nuclease resistant common sequence primer comprises phosphorothioate linkages.

59. The method of claim 57 wherein said nuclease is T7 Gene 6 Exonuclease.

60. The method of claim 46 wherein prior to amplification the extended capture probes are enriched in the sample to be amplified.

61. The method of claim 46 wherein labeled nucleotides are incorporated into the extended capture probes and extended capture probes are isolated by affinity chromatography.

62. The method of claim 61 wherein said labeled nucleotides are labeled with biotin and avidin, streptavidin or an anti-biotin antibody is used to isolate extended capture probes.

63. The method of claim 46 wherein prior to amplification the extended capture probes are made double stranded and single stranded nucleic acid in the sample is digested.

64. The method of claim 63 wherein the single stranded nucleic acid in the sample is digested with a nuclease.

65. The method of claim 64 wherein the nuclease is Exonuclease I.

66. The method of claim 46 wherein prior to amplification the extended capture probes are circularized and uncircularized nucleic acid in the sample is digested.

67. The method of claim 66 wherein extended capture probes are circularized by a method comprising:

hybridizing an oligonucleotide splint to the extended capture probes, wherein the oligonucleotide splint is complementary to the first and second common sequences, thereby juxtaposing the 5' and 3' ends of extended capture probes; and

ligating the ends of the extended capture probes to form circular extended capture probes.

68. The method of claim 66 wherein the uncircularized nucleic acid remaining in the sample is digested with a nuclease.

69. The method of claim 68 wherein the nuclease is Exonuclease III.

70. The method of claim 46 wherein the nucleic acid sample is fragmented by digestion with one or more restriction enzymes.

71. The method of claim 46 wherein there are 100 to 1500 different target sequences in the collection of target sequences.

72. The method of claim 46 wherein there are 1,000 to 5,000 different target sequences in the collection of target sequences.

73. The method of claim 46 wherein there are 2,000 to 10,000 different target sequences in the collection of target sequences.

74. The method of claim 46 wherein there are 10,000 to 1,000,000 different target sequences in the collection of target sequences.

75. A method for screening for sequence variations in a population of individuals comprising:

providing a nucleic acid sample from each individual;

genotyping each sample according to the method of claim 46; and

comparing the genotypes from the samples to determine the presence or absence of sequence variation in the population of individuals.

76. A kit for amplifying a collection of target sequences said kit comprising:

a collection of capture probes, wherein each species of capture probe comprises a first common sequence and a 3' variable region that is specific for a target sequence in said collection of target sequences;

an adapter comprising a second common sequence; and

a pair of first and second common sequence primers.

77. The kit of claim 76 wherein said collection of capture probes is covalently attached to a solid support and the 3' end of the capture probes is available for extension.

78. The kit of claim 76 further comprising a restriction enzyme, buffer, DNA polymerase and dNTPs.

79. A kit for amplifying a collection of target sequences said kit comprising:

a collection of capture probes, wherein each species of capture probe comprises a first common sequence, a tag sequence unique for each species of capture probe, a first target specific sequence, a Type IIs restriction enzyme recognition sequence, and a second target specific sequence;

an adapter comprising a first strand comprising a second common sequence and a second strand that does not contain the complement of the second common sequence and is blocked from extension at the 3' end; and

a pair of first and second common sequence primers.

80. The kit of claim 79 further comprising a Type IIs restriction enzyme, a ligase, dNTPs, ddNTPs, buffer and DNA polymerase.

81. The kit of claim 79 wherein one of the common sequence primers is resistant to nuclease digestion.

82. A collection of capture probes attached to a solid support, wherein said solid support is selected from the group consisting of arrays, beads, microparticles, microtitre dishes and gels.

83. A collection of capture probes, wherein each species of capture probe comprises a first common sequence, a tag sequence unique for each species of capture probe, a first target specific sequence, a Type IIs restriction enzyme recognition sequence, and a second target specific sequence, wherein the target specific sequence are specific for targets in a collection of target sequences.

84. A method of generating a plurality of oligonucleotides comprising: synthesizing a plurality of oligonucleotides on a solid support and releasing the plurality of oligonucleotides from said solid support.

85. The method of claim 84 wherein said plurality of oligonucleotides comprises a collection of capture probes.

FIG 1

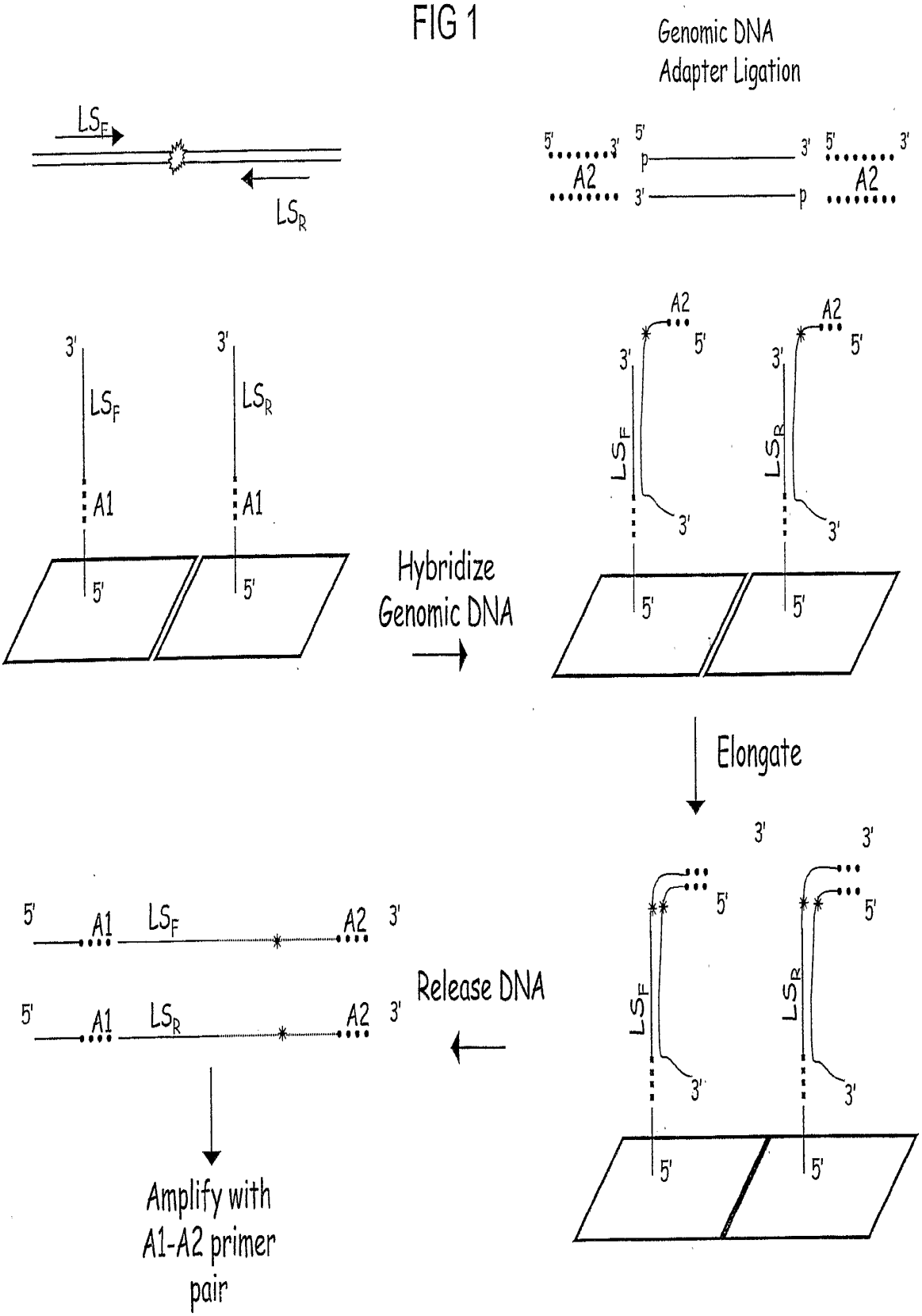

Genomic DNA
Adapter Ligation
LSF
LSR
A2
A1
Hybridize
Genomic DNA
Elongate
Release DNA
Amplify with
A1-A2 primer
pair

FIG 2

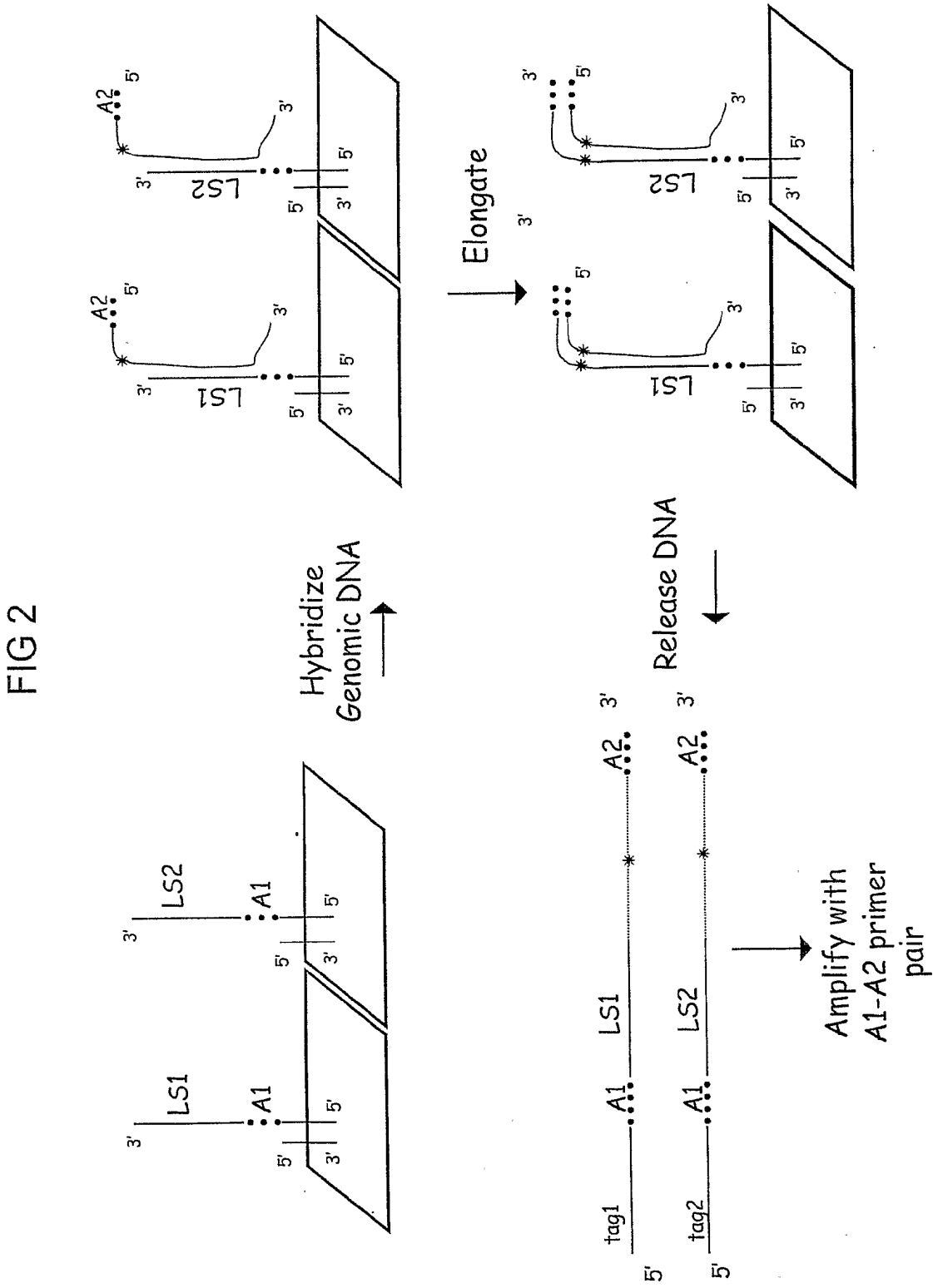

Hybridize Genomic DNA
Elongate
Release DNA
Amplify with A1-A2 primer pair
LS1
LS2
A1
A2
tag1
tag2
5'
3'

FIG 3

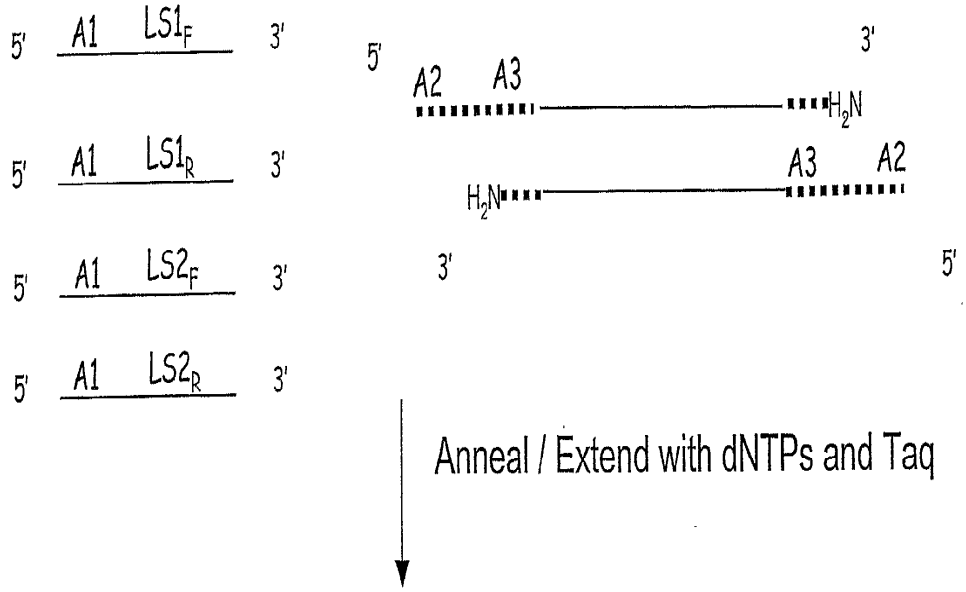

5' A1 LS1F 3'
5' A1 LS1R 3'
5' A1 LS2F 3'
5' A1 LS2R 3'
5'
3'
A2 A3
H2N
A3 A2
H2N
3'
5'
Anneal / Extend with dNTPs and Taq
Positively select extension products
or
Deplete non-extended products
Generic PCR with A1 / A2

FIG 4a

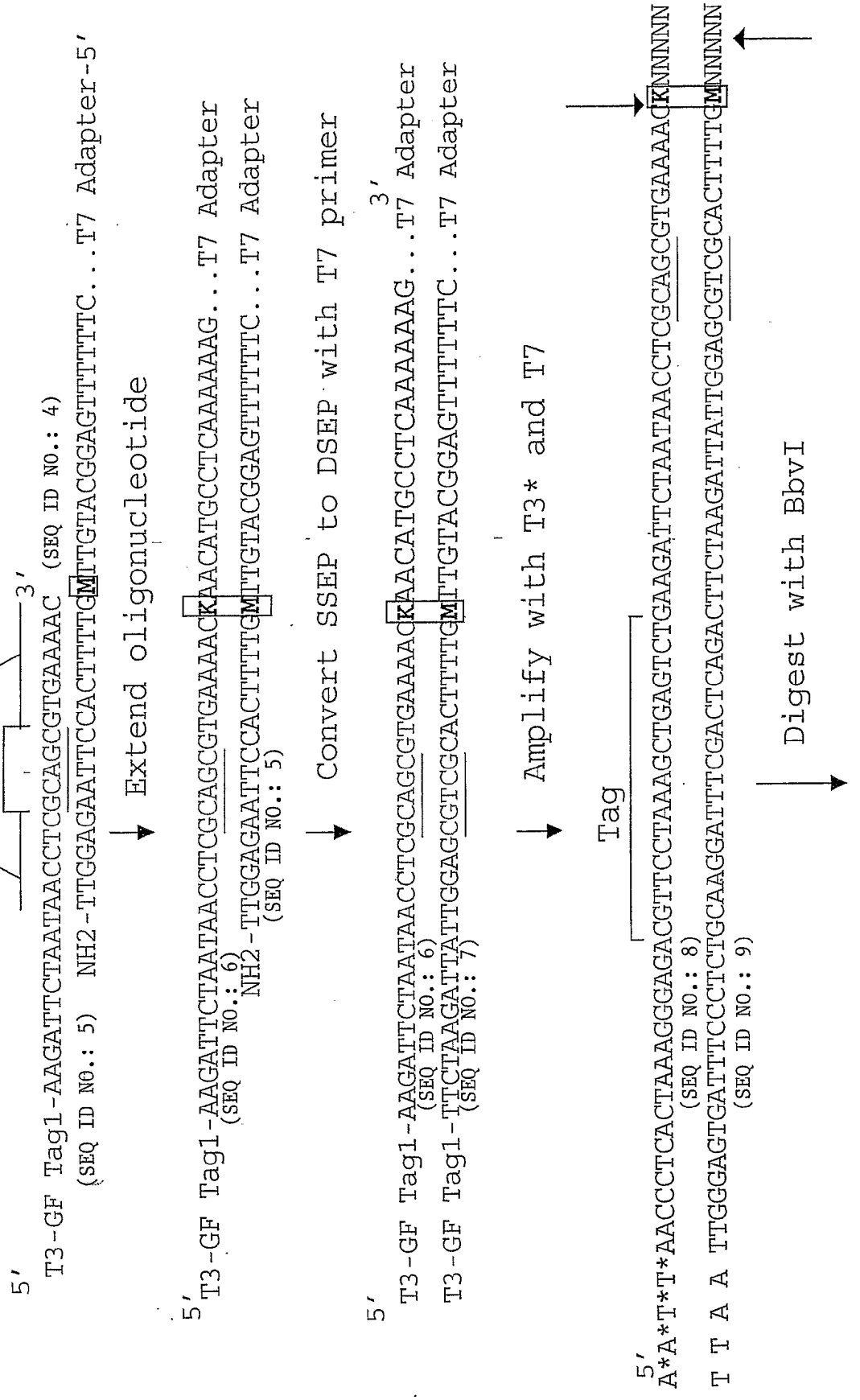
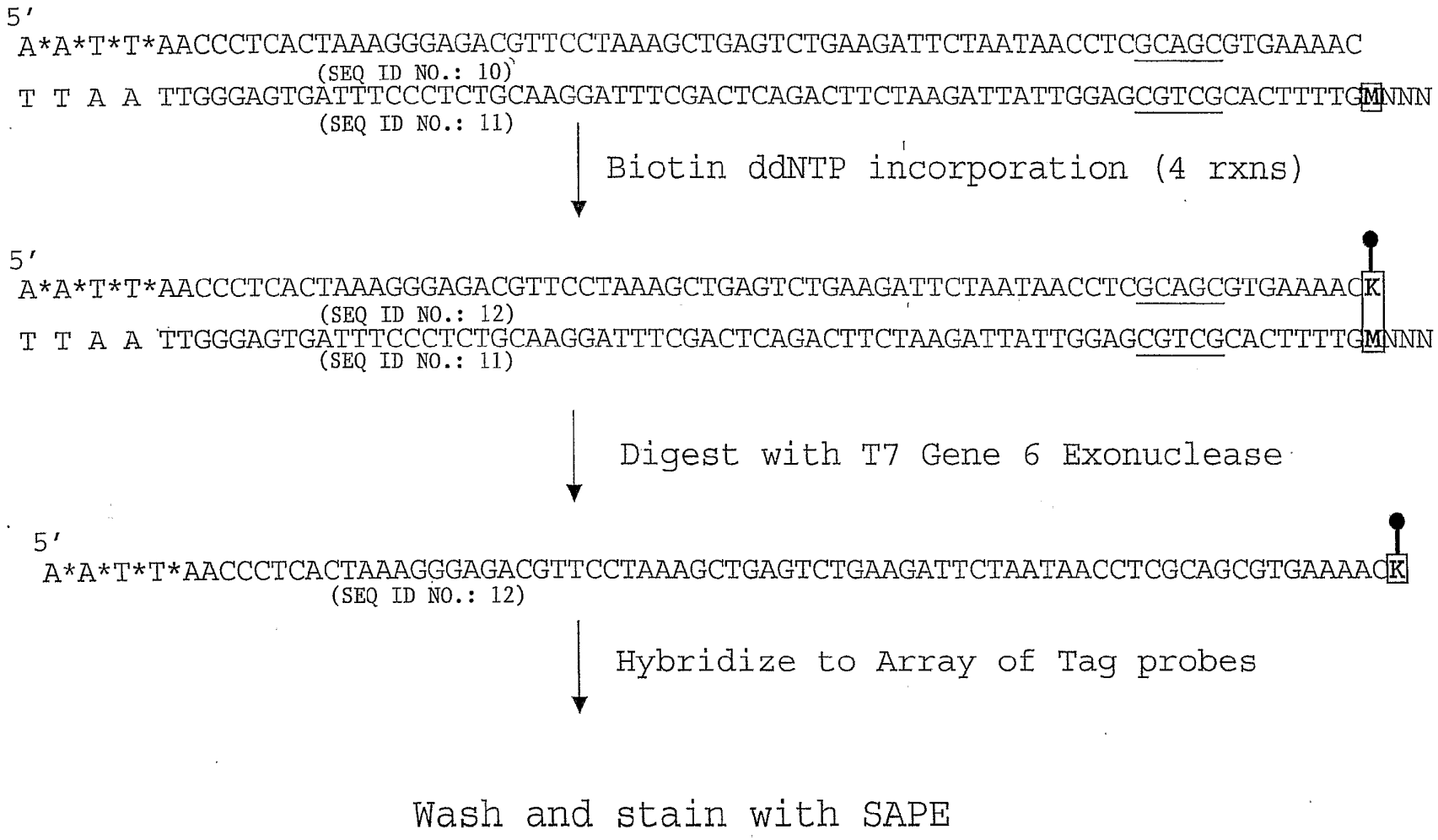

Target specific
BbvI
5′
3′
T3-GF Tag1-AAGATTCTAATACCTCGCAGCGTGAAAAC (SEQ ID NO.: 4)
(SEQ ID NO.: 5) NH2-TTGGAGAATTCCACTTTTTGMTTGTACGGAGTTTTTTTC...T7 Adapter-5′
Extend oligonucleotide
5′
T3-GF Tag1-AAGATTCTAATACCTCGCAGCGTGAAAACKAACATGCCTCAAAAAAG...T7 Adapter
(SEQ ID NO.: 6)
NH2-TTGGAGAATTCCACTTTTTGMTTGTACGGAGTTTTTTTC...T7 Adapter
(SEQ ID NO.: 5)
Convert SSEP to DSEP with T7 primer
5′
3′
T3-GF Tag1-AAGATTCTAATACCTCGCAGCGTGAAAACKAACATGCCTCAAAAAAG...T7 Adapter
(SEQ ID NO.: 6)
T3-GF Tag1-TTCTAAGATTATTGGAGCGTCGCACTTTTTGMTTGTACGGAGTTTTTTTC...T7 Adapter
(SEQ ID NO.: 7)
Amplify with T3* and T7
Tag
5′
A*A*T*T*AACCCTCACTAAAGGGAGACGTTCCTAAAGCTGAGTCTGAAGATTCTAATACCTCGCAGCGTGAAAACKNNNNNN
(SEQ ID NO.: 8)
T T A A TTGGGAGTGATTTCCCTCTGCAAGGATTTCGACTCAGACTTCTAAGATTATTGGAGCGTCGCACTTTTTGMNNNNNN
(SEQ ID NO.: 9)
Digest with BbvI

FIG 4b



IDT Exhibit 1020.9843

FIG 5

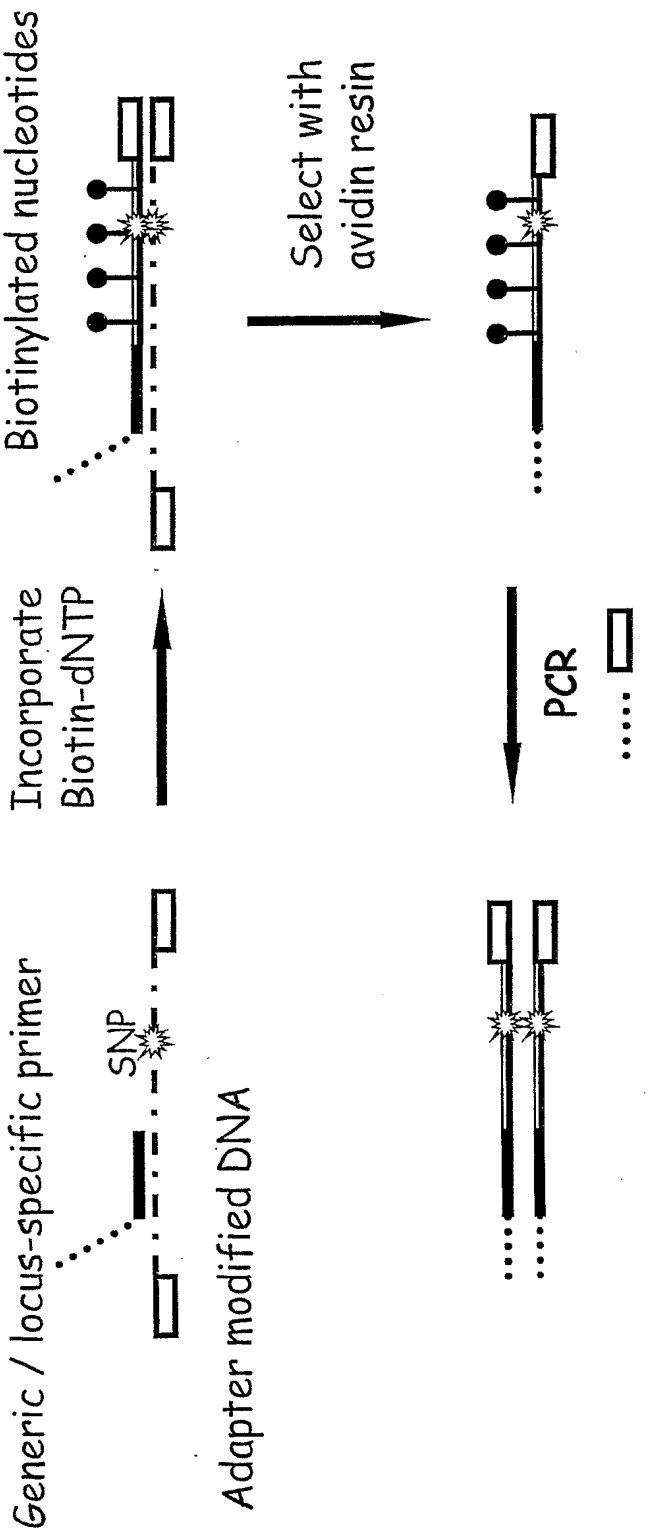

Biotinylated nucleotides
Incorporate Biotin-dNTP
Select with avidin resin
PCR
Generic / locus-specific primer
SNP
Adapter modified DNA

FIG 6

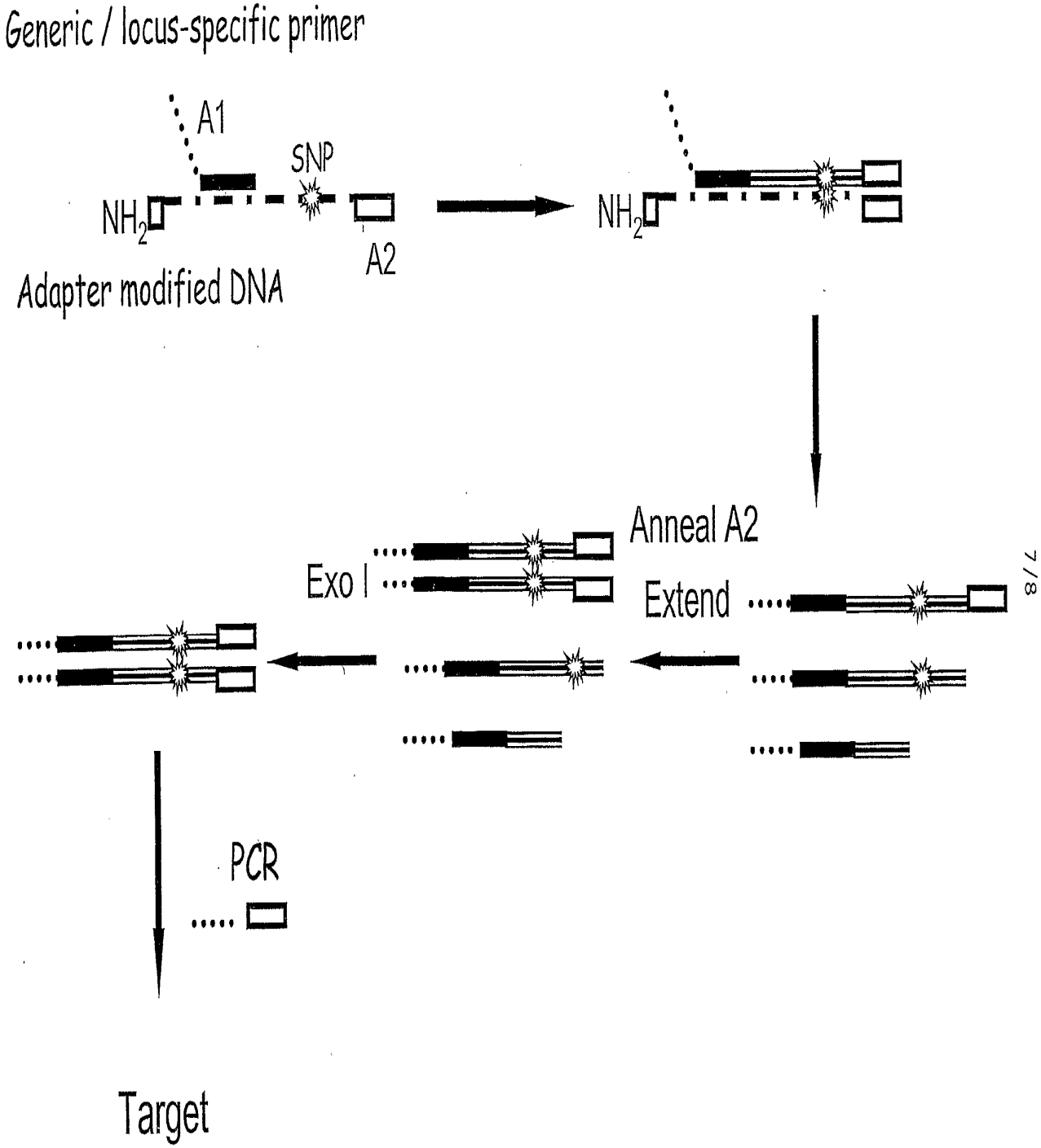
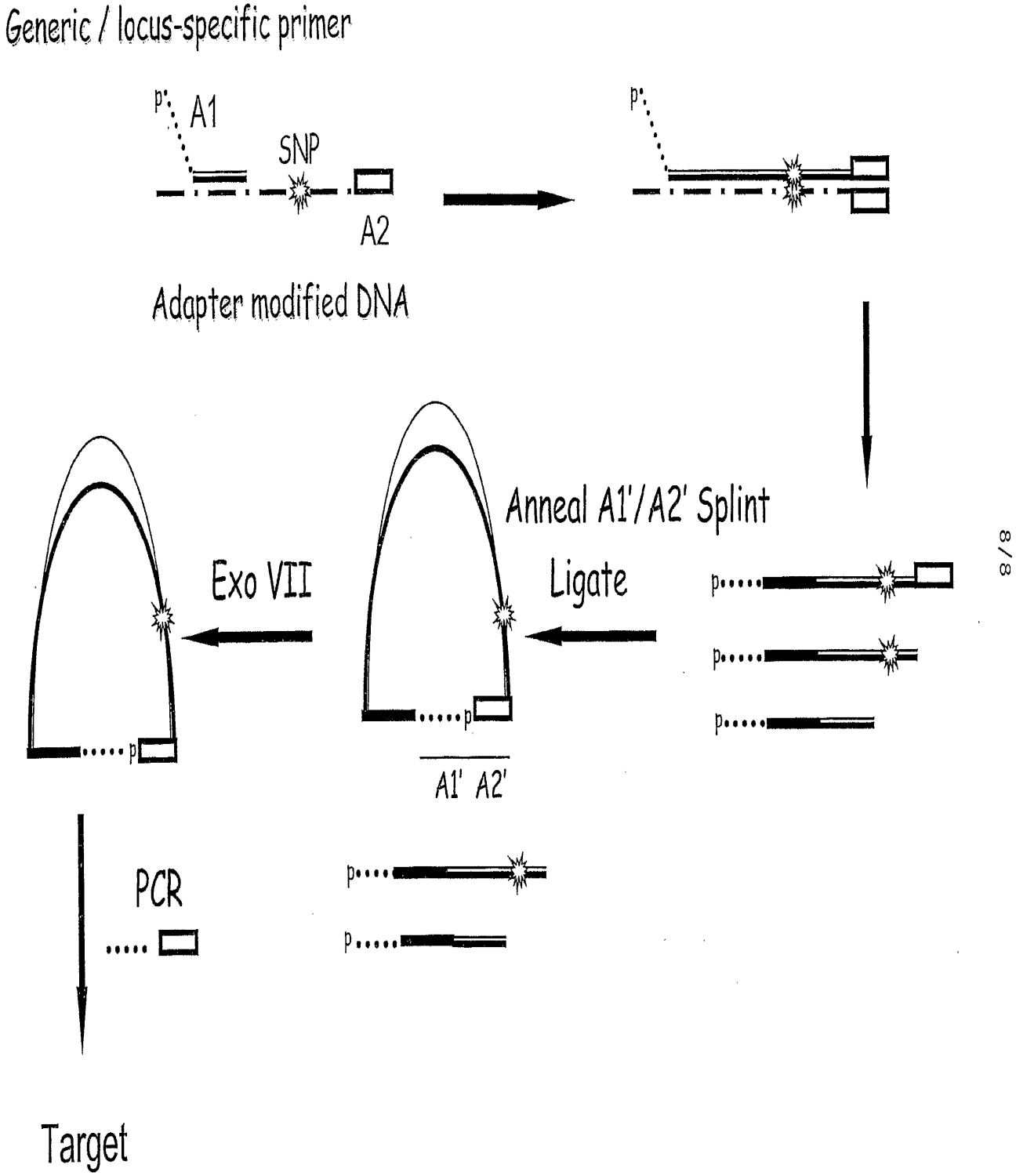

Generic / locus-specific primer
A1
SNP
NH2
A2
Adapter modified DNA
NH2
Anneal A2
Exo I
Extend
PCR
Target

FIG 7


Generic / locus-specific primer
p
A1
SNP
A2
Adapter modified DNA
Anneal A1'/A2' Splint
Ligate
Exo VII
PCR
A1' A2'
Target

SEQUENCE LISTING

<110> Affymetrix, Inc.
Jones, Keith W.
Shapero, Michael
Liu, Weiwei

<120> COMPLEXITY MANAGEMENT OF GENOMIC DNA BY LOCUS SPECIFIC AMPLIFICATION

<130> 2719.2029001

<150> US 10/272,155
<151> 2002-10-14

<150> US 60/389,747
<151> 2002-06-17

<160> 12

<170> FastSEQ for Windows Version 4.0

<210> 1
<211> 11
<212> DNA
<213> Artificial Sequence

<220>
<223> Type-II endonuclease EarI recognition sequence

<221> misc_feature
<222> 7, 8, 9, 10, 11
<223> n = A,T,C or G

<400> 1
ctcttcnnnn n 11

<210> 2
<211> 11
<212> DNA
<213> Artificial Sequence

<220>
<223> Type-II endonuclease EarI recognition sequence

<221> misc_feature
<222> 7, 8, 9, 10, 11
<223> n = A,T,C or G

<400> 2
nnnnngaaga g 11

<210> 3
<211> 47
<212> DNA
<213> Artificial Sequence

<220>
<223> T3-T7 splint oligonucleotide

<400> 3
tctcccttta gtgagggtta atttgtaata cgactcacta tagggca 47

<210> 4
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> Oligonucleotide

<400> 4
aagattctaa taacctcgca gcgtgaaaac 30

<210> 5
<211> 37
<212> DNA
<213> Artificial Sequence

<220>
<223> Oligonucleotide

<400> 5
ctttttgag gcatgttmgt tttcacctta agaggtt 37

<210> 6
<211> 48
<212> DNA
<213> Artificial Sequence

<220>
<223> Oligonucleotide

<400> 6
aagattctaa taacctcgca gcgtgaaaac kaacatgcct caaaaaag 48

<210> 7
<211> 48
<212> DNA
<213> Artificial Sequence

<220>
<223> Oligonucleotide

<400> 7
ctttttgag gcatgttmgt tttcacgctg cgaggttatt agaatctt 48

<210> 8
<211> 79
<212> DNA
<213> Artificial Sequence

<220>
<223> Oligonucleotide

<221> misc_feature
<222> 75, 76, 77, 78, 79
<223> n = A,T,C or G

<400> 8
aattaaccct cactaaaggg agacgttcct aaagctgagt ctgaagattc taataacctc 60
gcagcgtgaa aacknnnnn 79

<210> 9
<211> 79
<212> DNA
<213> Artificial Sequence

<220>
<223> Oligonucleotide

<221> misc_feature
<222> 1, 2, 3, 4, 5
<223> n = A,T,C or G

<400> 9
nnnnnmgttt tcacgctgcg aggttattag aatcttcaga ctcagcttta ggaacgtctc 60
cctttagtga gggttaatt 79

<210> 10
<211> 73
<212> DNA
<213> Artificial Sequence

<220>
<223> Oligonucleotide

<400> 10
aattaaccct cactaaaggg agacgttcct aaagctgagt ctgaagattc taataacctc 60
gcagcgtgaa aac 73

<210> 11
<211> 77
<212> DNA
<213> Artificial Sequence

<220>
<223> Oligonucleotide

<221> misc_feature
<222> 1, 2, 3
<223> n = A,T,C or G

<400> 11
nnnmgttttc acgctgcgag gttattagaa tcttcagact cagctttagg aacgtctccc 60
tttagtgagg gttaatt 77

<210> 12
<211> 74
<212> DNA
<213> Artificial Sequence

<220>
<223> Oligonucleotide

<400> 12
aattaaccct cactaaaggg agacgttcct aaagctgagt ctgaagattc taataacctc 60
gcagcgtgaa aack 74

[19] 中华人民共和国国家知识产权局

[51] Int. Cl.
C12Q 1/68 (2006.01)

[12] 发明专利申请公布说明书

[21] 申请号 200910026891.7

[43] 公开日 2009 年 10 月 28 日

[11] 公开号 CN 101565746A

[22] 申请日 2009.6.3
[21] 申请号 200910026891.7
[71] 申请人 东南大学
地址 214135 江苏省无锡市菱湖大道 99 号
[72] 发明人 涂 景 白云飞 杨 琦 孙蓓丽 陆祖宏

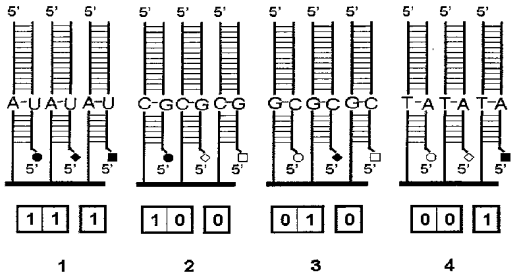
[74] 专利代理机构 南京君陶专利商标代理有限公司
代理人 奚胜元
权利要求书 2 页 说明书 14 页 附图 3 页

[54] 发明名称

带奇偶校验的信号组合编码 DNA 连接测序方法

[57] 摘要

本发明带奇偶校验的信号组合编码的 DNA 连接测序方法涉及一种信号组合编码的 DNA 测序方法，属于生物技术领域。 本发明的特征在于向信号组合编码的连接测序的编码中增加了奇偶校验，有效地对检测获得的编码结果进行奇偶校验，并成功分辨全空信号与未发生连接反应两种情况，提高测序的准确率。 本发明在信号组合编码标记物外增加一种奇偶校验标记物，利用奇偶校验标记物“有信号”和“无信号”两种状态与所获得的编码标记物状态的“有信号”的标记物的种类进行奇偶校验，校验成功后通过不同编码标记物在被检测时多个标记物“有信号”、“无信号”的排列组合进行信号组合编码的连接 DNA 测序反应，在不降低测序通量的基础上提高测序反应的准确率。



IDT Exhibit 1020.9850

1. 一种带奇偶校验的信号组合编码的 DNA 连接测序方法，其特征在于针对某一种特定的 DNA 测序探针，利用一组编码标记物和一种信号编码标记物状态的组合进行标记，对于一批 DNA 测序探针，采用一组编码标记物状态结合信号编码标记物的不同组合方案进行标记，制备一套带奇偶校验的信号组合编码的 DNA 测序探针，从而实现在检测时对不同 DNA 测序探针的区分和鉴别：

A 一套带奇偶校验的信号组合编码的 DNA 测序探针的制备：首先制备未标记的 DNA 测序探针，每一种未标记的 DNA 测序探针由一个或多个测序碱基、一个或多个简并碱基 N 或非严格配对碱基组成；测序碱基用于通过碱基互补配对法则测定待测 DNA 中相应位置的碱基信息，简并碱基 N 为 A、T、C、G 四种碱基中的任意一个；完成未标记的 DNA 测序探针的制备后，针对每一种未标记的 DNA 测序探针，采用多种编码标记物状态的一种组合进行标记，每种编码标记物在该 DNA 测序探针上标记与否由该编码标记物在状态的组合中所对应的状态决定，同时结合奇偶校验标记物，当标记的编码标记物个数为奇数个时，则不标记奇偶校验标记物，当标记的编码标记物为偶数个时，则标记奇偶校验标记物。采用不同的状态组合方案并结合奇偶校验标记物对不同的未标记 DNA 测序探针进行标记，完成一套带奇偶校验的信号组合编码的 DNA 测序探针的制备；

B 利用上述一套带奇偶校验的信号组合编码的 DNA 测序探针的测序流程如下：

a. 选取一条测序引物与待测的 DNA 模板依据碱基互补配对原理进行杂交；

b. 向反应体系中加入上述的一套带奇偶校验的信号组合编码的 DNA 测序探针和 DNA 连接酶及其反应体系，进行 DNA 连接反应；

c. 完成 DNA 连接反应后，检测全部参与状态组合的标记物及奇偶校验标记物的信号，判读编码标记物信号所处的状态，计算获得信号的编码标记物的数量，并利用奇偶校验标记物的信号对该数值进行奇偶校验，如校验通过，根据全部编码标记物状态的组合情况，确定发生连接反应的 DNA 测序探针的种类，从而确定该测序探针上测序碱基的类型和排布，并最终确定被测 DNA 模板上对应位置的碱基或碱基序列信息；

d. 完成对标记物信号的检测后，移除 DNA 测序探针上的标记物；

e. 重复上述步骤 b-d 直至 DNA 测序探针达到或超出待测 DNA 模板待测区域末端；

f. 该测序引物与待测 DNA 模板变性分离，选取第二条测序引物；

g. 重复上述步骤 a-f，直至全部未知 DNA 模板的待测区域的全部序列信息得以确定。

IDT Exhibit 1020.9851

2. 根据权利要求 1 所述的带奇偶校验的信号组合编码的 DNA 连接测序方法，其特征在于所述的多种标记物对未标记 DNA 测序探针进行组合标记，其方案是对同一 DNA 测序探针分子同时标记一种或多种标记物，或者分别用不同的标记物标记同一种 DNA 测序探针的不同分子，随后将上述不同测序探针分子混合。

3. 根据权利要求 1 所述的带奇偶校验的信号组合编码的 DNA 连接测序方法，其特征在于所述的测序引物是指一组只能和所有 DNA 模板的一段通用序列杂交的寡核苷酸片段，一组测序引物在 DNA 模板上杂交区域彼此相差一个或数个碱基，能够在数轮测序反应中完成对 DNA 模板全长的测序工作；测序模板上的通用序列，是在测序模板制备过程中通过连接反应加入，或者在扩增过程中通过引物引入。

4. 根据权利要求 1 所述的带奇偶校验的信号组合编码的 DNA 连接测序方法，其特征在于所述的 DNA 模板是通过对待测的 DNA 片段通过 DNA 扩增技术增加基因组中感兴趣的目的片段的量，扩增是单重的，即一次扩增一个目的片段，或者是多重的，即一次扩增多个目的片段；所述的待测的 DNA 模板中的固定是通过化学或者物理方法固定于平面片基上，或固定于“96 孔板”、“384 孔板”及各种修饰的珠子载体上。

5. 根据权利要求 1 所述的带奇偶校验的信号组合编码的 DNA 连接测序方法，其特征在于所述的连接反应，可以是自测序引物开始 5’ 末端向 3’ 末端的连接反应，也可以是 3’ 末端向 5’ 末端的连接反应。

6. 根据权利要求 1 所述的带奇偶校验的信号组合编码的 DNA 连接测序方法，其特征在于所述的标记物，是采用的荧光基团，或利用测序探针化学的、物理的性质的改变，如电阻变化、电流变化。

7. 根据权利要求 1 所述的带奇偶校验的信号组合编码的 DNA 连接测序方法，其特征在于所述的标记物的移除，是标记基团通过化学的、物理的方法与 DNA 测序探针分离，或者 DNA 测序探针化学的、物理的性质恢复原有状态；标记物的移除，是仅仅移除标记物本身，或者同时移除与标记物相连或不相连的一个或多个碱基及相关基团。

8. 根据权利要求 1 所述的带奇偶校验的信号组合编码的 DNA 连接测序方法，其特征在于所述的第二条测序引物是一组测序引物中除已经使用的测序引物中的任何一条，并不一定是与已使用的测序引物对应在测序模板上杂交位置最为接近的一条。

IDT Exhibit 1020.9852

带奇偶校验的信号组合编码 DNA 连接测序方法

技术领域

本发明带奇偶校验的信号组合编码DNA连接测序方法涉及一种采用信号组合编码的DNA连接测序方法，并结合奇偶校验方案，是一种实现DNA序列分析的方法，属于生物技术领域。

技术背景

随着基因组研究的深入，特别是人类基因组计划和各种模式生物基因组计划的完成，生物学研究和医学研究方式产生了巨大变革。从基因水平上认识生命的差异，疾病发生、发展的规律，药物与生命体的相互作用，不同物种之间的遗传差异以及同一物种内部不同个体间的遗传差异成为可能。尽管导致疾病发生的因素众多，但基因序列差异（包括单核苷酸多态性、DNA 甲基化等）被广泛认为是一个重要的内在因素。多数复杂疾病发生和发展，如癌症、糖尿病、心血管疾病、精神疾病等，是众多基因和环境共同作用的结果。通过对某一特定疾病大量基因组样本中基因突变的大规模检测及与非患病对照基因组样本的比对，即可获得与该疾病有关的基因型信息，通过随后对药物敏感基因突变位点的筛查，可以获得对临床治疗与用药具有指导性意义的信息。无论是亲缘关系较近的物种之间遗传差异，还是同一物种内部不同个体之间的遗传差异，都主要是以基因序列差异的形式体现出来。因此，如何快速的筛查基因组中的基因序列差异成为后基因组时代的主要课题之一。

现有的基因序列差异的检测方法主要为：传统 Sanger DNA 测序法、限制性酶切长度多态性、单链构象多态性和基于基因芯片的寡核苷酸探针杂交法等。在这些方法中，仅有传统的 Sanger DNA 测序法能完成对目标片段的全方位序列测定，其余方法均只能确定很少的一部分序列信息。但传统的 Sanger DNA 测序法存在通量低、成本高、耗时长等不足。第一个人类基因组序列测定的费用大约为 10 亿美元。尽管这一费用目前已经降低到 2000 万美元，但依然是制约功能基因组研究的瓶颈，大幅度降低 DNA 测序的成本将会大大推动生命科学的发展。为此，美国 Venter 基金会在 2003 年提出了 1000 美元人类全基因组测序的研究目标。2004 年初，美国国立卫生院投入 4 千多万美元支持 DNA 测序新技术的研究计划，累计已经超过 1 亿美元。其目标是发展 10 万美元的人类全基因组 DNA 测序技术，并最终减低为 1000 美元。

目前，除对现有的基于电泳的测序技术的改进外，正在研发的测序技术总体上可以分为四类。第一类是延伸测序法，将信号标记的碱基加入到正在发生聚合反应的 DNA 链中进行检

IDT Exhibit 1020.9853

测，第二类是连接测序法，将信号标记的寡核苷酸片段加入到正在发生连接反应的 DNA 链中进行检测；第三类是杂交测序法，通过制备一组高密度寡核苷酸微阵列芯片的杂交信号，进行目标基因的序列鉴定；第四类是分子影像等一系列可以在单分子的水平上进行测序的技术；最后一类技术是诱导 DNA 分子蜿蜒通过非常细微的小孔，在这个过程中借助电子学或者光学的方法对碱基进行读出，也成为纳米孔道测序技术。目前只有延伸测序方法和连接测序方法用于全基因组测序，并且开发出了商品化的仪器和试剂，大大提高了 DNA 测序的效率，大幅度降低了 DNA 测序的成本。然而，目前新一代的 DNA 测序在测序成本、通量和速度等方面仍然不能满足生命科学研究的需要。主要原因之一是在检测核酸的信号标记方法单一、效率不高。例如，在 DNA 连接测序方法中，受标记物（如荧光基团）种类的限制，每次连接反应一般仅能确定一个碱基的信息，如需在一次连接反应确定两个及两个以上碱基信息，则需要对标记物进行二维或多维编码。然而现有编码方法对多种标记物都“无信号”的状态和未成功发生连接反应的“空信号”之间无法有效分辨，测序反应的存在一定的错误率。

发明内容

本发明的目的就是针对现有 DNA 的连接测序技术在检测核酸的信号标记方法单一、效率不高，试图通过不同标记物状态的组合进行二维的编码，合成信号编码的 DNA 测序探针，并通过奇偶校验标记物的引入，分辨“无信号”和“空信号”，提高连接测序编码方案的准确率，实现使用相同数量的标记物检测较多碱基或碱基组合的目的，建立快速、准确、低成本和高通量的序列测定方法。

DNA 连接测序属于一种 DNA 合成测序方法。DNA 连接测序的基本步骤是：首先在待测的 DNA 模板上杂交一条与特定区域互补配对的测序引物；接着向含有待测模板和测序引物的反应器中加入一种或一组寡核苷酸测序探针，每个寡核苷酸探针由三部分组成：简并碱基或非严格配对碱基、标记物和测序碱基；简并碱基或非严格配对碱基是一组非特异性的可与测序模板杂交的碱基组合，标记物用于标记寡核苷酸探针，便于在发生 DNA 连接反应后进行检测；测序碱基是一个和多个确定的碱基，用于使得测序探针有选择性地与部分测序模板-测序引物复合物发生连接反应，发生反应的测序模板的相应碱基与测序探针的测序碱基互补配对，同时，测序碱基与标记物存在某种对应关系；在进行了 DNA 连接测序反应后，利用相应的检测手段对测序模板-引物-探针复合物进行检测，获得对应的信号信息，再利用预先设定的测序碱基和标记物的对应关系判读测序模板中相应位置的碱基信息。

当完成一次连接测序反应后，移除测序模板-引物-探针复合物上的标记物，以测序探针作为新的连接起点进行下一次的连接测序反应，连续进行连接反应直至完成所需测序长度。

IDT Exhibit 1020.9854

测序引物与待测模板变性分离，随后在待测模板上杂交与第一条引物序列平移一个和多个碱基的测序引物，进行新一轮多次连接反应；完成后重复变性、杂交、连接过程直至所有位置的碱基信息均完成测定。

本发明技术方案：带奇偶校验的信号组合编码的 DNA 连接测序方法，其特征在于针对某一种特定的 DNA 测序探针，利用一组编码标记物和一种信号编码标记物状态的组合进行标记，对于一批 DNA 测序探针，采用一组编码标记物状态结合信号编码标记物的不同组合方案进行标记，制备一套带奇偶校验的信号组合编码的 DNA 测序探针，从而实现在检测时对不同 DNA 测序探针的区分和鉴别：

A 一套带奇偶校验的信号组合编码的 DNA 测序探针的制备：首先制备未标记的 DNA 测序探针，每一种未标记的 DNA 测序探针由一个或多个测序碱基、一个或多个简并碱基 N 或非严格配对碱基组成。测序碱基用于通过碱基互补配对法则测定待测 DNA 中相应位置的碱基信息，简并碱基 N 为 A、T、C、G 四种碱基中的任意一个。完成未标记的 DNA 测序探针的制备后，针对每一种未标记的 DNA 测序探针，采用多种编码标记物状态的一种组合进行标记，每种编码标记物在该 DNA 测序探针上标记与否由该编码标记物在状态的组合中所对应的状态决定，同时结合奇偶校验标记物，当标记的编码标记物个数为奇数个时，则不标记奇偶校验标记物，当标记的编码标记物为偶数个时，则标记奇偶校验标记物。采用不同的状态组合方案并结合奇偶校验标记物对不同的未标记 DNA 测序探针进行标记，完成一套带奇偶校验的信号组合编码的 DNA 测序探针的制备。

B 利用上述一套带奇偶校验的信号组合编码的 DNA 测序探针的测序流程如下：

a. 选取一条测序引物与待测的 DNA 模板依据碱基互补配对原理进行杂交；

b. 向反应体系中加入上述的一套带奇偶校验的信号组合编码的 DNA 测序探针和 DNA 连接酶及其反应体系，进行 DNA 连接反应；

c. 完成 DNA 连接反应后，检测全部参与状态组合的标记物及奇偶校验标记物的信号，判读编码标记物信号所处的状态，计算获得信号的编码标记物的数量，并利用奇偶校验标记物的信号对该数值进行奇偶校验，如校验通过，根据全部编码标记物状态的组合情况，确定发生连接反应的 DNA 测序探针的种类，从而确定该测序探针上测序碱基的类型和排布，并最终确定被测 DNA 模板上对应位置的碱基或碱基序列信息；

d. 完成对标记物信号的检测后，移除 DNA 测序探针上的标记物；

e. 重复上述步骤 b-d 直至 DNA 测序探针达到或超出待测 DNA 模板待测区域末端；

f. 该测序引物与待测 DNA 模板变性分离，选取第二条测序引物；

IDT Exhibit 1020.9855

g. 重复上述步骤a-f，直至全部未知DNA模板的待测区域的全部序列信息得以确定。

所述的多种标记物对未标记DNA测序探针进行组合标记，其方案是对同一DNA测序探针分子同时标记一种或多种标记物，或者分别用不同的标记物标记同一种DNA测序探针的不同分子，随后将上述不同测序探针分子混合。

所述的测序引物是指一组只能和所有DNA模板的一段通用序列杂交的寡核苷酸片段，一组测序引物在DNA模板上杂交区域彼此相差一个或数个碱基，能够在数轮测序反应中完成对DNA模板全长的测序工作；测序模板上的通用序列，是在测序模板制备过程中通过连接反应加入，或者在扩增过程中通过引物引入。

所述的DNA模板是通过对待测的DNA片段通过DNA扩增技术增加基因组中感兴趣的目的片段的量，扩增是单重的，即一次扩增一个目的片段，或者是多重的，即一次扩增多个目的片段；所述的待测的DNA模板中的固定是通过化学或者物理方法固定于平面片基上，或固定于“96孔板”、“384孔板”及各种修饰的珠子载体上。

所述的连接反应，可以是自测序引物开始5’末端向3’末端的连接反应，也可以是3’末端向5’末端的连接反应。

所述的标记物，是采用的荧光基团，或利用测序探针化学的、物理的性质的改变，如电阻变化、电流变化。

所述的检测是指与所述的标记物性质相适应的，在激光激发的荧光标记基团时，激光激发强度可为95%、70%、45%等多个不同强度，光电倍增可以分别为95%、70%、45%等多个检测区间。

所述的标记物的移除，是标记基团通过化学的、物理的方法与DNA测序探针分离，或者DNA测序探针化学的、物理的性质恢复原有状态。标记物的移除，是仅仅移除标记物本身，或者同时移除与标记物相连或不相连的一个或多个碱基及相关基团。

所述的第二条测序引物是一组测序引物中除已经使用的测序引物中的任何一条，并不一定是与已使用的测序引物对应在测序模板上杂交位置最为接近的一条。

技术原理：本发明的目的是提供一种带奇偶验证信号编码的DNA连接测序方法，在一次连接反应中同时检测多个碱基，并通过奇偶校验标记物的引入，提高实验的准确率。该方法的原理是根据连接测序反应中信号编码标记物的奇偶数决定是否想连接测序反应中条件奇偶校验标记物，并通过在检测时奇偶校验标记物的状态校验信号编码标记物的实现对无信号状态和为成功发生连接反应的分辨，提高信号编码循环连接的DNA测序方法的可靠性。

本发明的目的是提供一种带背景验证信号组合编码的DNA连接测序方法，在一次连接反

IDT Exhibit 1020.9856

应中同时检测多个碱基，并通过背景标记物的引入，降低测序反应的错误率。该方法的原理是通过向连接测序反应中添加特定的背景验证标记物，有效分辨“无信号”未成功连接两种状态，提高信号组合编码的连接测序反应的准确率。

信号组合编码的连接测序方法利用不同标记物在被检测信号的有无（二维）和强度差异（多维）进行二进制编码或多进制编码。这里我们以信号的有无为例，当标记物为两种时，通过两种标记物“有信号”和“无信号”排列组合，共得到 2^2=4 种复合状态，即“标记物 1 有信号、标记物 2 有信号”、“标记物 1 有信号、标记物 2 无信号”、“标记物 1 无信号、标记物 2 有信号”、“标记物 1 无信号、标记物 2 无信号”。当标记物数量为 n 时（n 为大于等于 2 的整数），记录标记物复合状态为全部 n 位二进制数，二进制数的每一位记录 1 种标记物的被检状态，“1”为检测到信号，“0”为未检测到信号，全部 2^n 个 n 位二进制数与 n 个信号标记物的 2^n 种复合状态一一对应。由于构成 DNA 的常见碱基共有 4 种，分别为腺嘌呤（A）、鸟嘌呤（G）、胸腺嘧啶（T）、胞嘧啶（C），因此需要准确检测一个碱基的 4 种类型需要两种标记物“有信号”和“无信号”的 4 种复合状态；2 个碱基组成的碱基组合共有 4×4=16 种类型，需要 4 种标记物共 2^4=16 种复合状态来检测；依此类推，n 个（n 为大于等于 2 的整数）碱基组成的碱基组合共需要 2n 种标记物共 2^{2n} 种复合信号状态来检测，每种复合状态通过一个 2n 位二进制数来唯一表示，并与测序探针的类型一一对应。

然而，单纯的信号组合编码的连接测序方法存在如下问题：当待测碱基或碱基组合对应的互补测序探针没有与之对应的修饰标记物时，由于加入反应体系中的测序探针均不含修饰标记物，因此通过检测全部标记物后我们得到一组全为“无信号”的信号组合；同时，当加入反应体系的测序探针因为连接反应失败等或其他原因导致测序反应失败，通过检测我们同样可以获得一组全为“无信号”的信号组合。此时一组全为“无信号”的信号组合对应了两种可能，传统组合编码的测序方法无法有效分辨这两种可能，因而测序过程存在一定的错误率。同时，连接测序属于一种生物化学反应，其本身也存在一定的错误率。

本发明，在现有信号编码的基础上增加了一种标记物作为奇偶校验标记物（见表 1），参与测序探针的修饰。奇偶校验标记物和测序探针类型的组合按照如下规则进行：当一种测序探针类型同时由 0 种或偶数种编码标记物组合时，则增加奇偶校验标记物对该测序探针类型的修饰；当一种测序探针类型同时由奇数种编码标记物组合时，则不增加奇偶校验标记物对该测序探针类型的修饰；

奇偶校验标记物的主要作用是：1，验证测序反应后编码标记物“有信号”个数的奇偶数数量是否正确，以提高测序的准确性；2，分辨测序反应后全部编码标记物“无信号”的状态

IDT Exhibit 1020.9857

和连接反应失败的状态。

表 1 带奇偶校验的编码方法 2n 种标记物 2^{2n} 种复合状态检测 n 个碱基寡核苷酸配方列表

（表中“1”表示探针混合物含有该寡核苷酸，“0”表示探针混合物不含该寡核苷酸探针）

待测碱基组合类型	奇偶校验标记物	标记物 1	标记物 2	标记物 3	……	标记物 2n-2	标记物 2n-1	标记物 2n
测序探针类型 1	0	1	1	1	……	1	1	1
测序探针类型 2	1	1	1	1	……	1	1	0
测序探针类型 3	1	1	1	1	……	1	0	1
……	……	……	……	……	……	……	……	……
测序探针类型 $2^{2n}-2$	0	0	0	0		0	1	0
测序探针类型 $2^{2n}-1$	0	0	0	0	……	0	0	1
测序探针类型 2^{2n}	1	0	0	0	……	0	0	0

本发明的有益效果：

1. 本发明最大的优点通过奇偶校验标记物的引入，对信号编码标记物的状态进行奇偶校验，在有效分辨编码标记物全为“无信号”和未成功发生连接反应两种情况的同时，提高测序反应的准确率，使得信号编码的循环连接测序法更加具有实用性。通过的编码标记物的组合信号编码，实现在一次连接反应中同时检测较多碱基或碱基组合。
2. 由于在单次连接反应中检测多个碱基的类别所需的标记物种类与传统方法相比呈现指数型的下降，使得在单次连接反应中同时检测多个碱基的类别成为可能，成倍缩短了对全部序列测定所需的时间。
3. 通过信号组合编码，检测相同长度的序列所需的连接反应次数的减少，测序的成本也成倍降低。同时，通过信号组合编码，检测相同数量的碱基类型所需的标记物种类大幅降低，降低了对检测标记物设备的需求，从而降低成本。

附图说明

以下结合附图对本发明进行进一步说明。

图 1 是当标记物状态为二维时，即信号的“有”或“无”时，通过两种标记物的二维状态组合检测在一次连接测序反应中 A、T、C、G 四种碱基的原理图，图中实心图标表示“有信号”，空心图标表示“无信号”。1，当待测碱基为 A 时，由测序碱基为 U 的测序探针所对应的编码标记物的组合状态为“标记物 1 有信号，标记物 2 有信号”，即“1，1”，随后获得奇偶校验标记物（图中方块，实心为有信号，空心为无信号）信号为“有信号”，即“1”，根

IDT Exhibit 1020.9858

据预设就校验原理，编码标记物状态有两个1，为偶数，对应的奇偶校验标记物应为1，与实际奇偶校验标记物状态相符；当待测碱基为C时，由测序碱基为G的测序探针所对应的编码标记物的组合状态为“标记物1有信号，标记物2无信号”，即“1，0”，随后获得奇偶校验标记物（图中方块，实心为有信号，空心为无信号）信号为“无信号”，即“0”，根据预设就校验原理，编码标记物状态有1个1，为奇数，对应的奇偶校验标记物应为0，与实际奇偶校验标记物状态相符；当待测碱基为G时，由测序碱基为U的测序探针所对应的编码标记物的组合状态为“标记物1无信号，标记物2有信号”，即“0，1”，随后获得奇偶校验标记物（图中方块，实心为有信号，空心为无信号）信号为“无信号”，即“0”，根据预设就校验原理，编码标记物状态有1个1，为奇数，对应的奇偶校验标记物应为0，与实际奇偶校验标记物状态相符；当待测碱基为T时，由测序碱基为A的测序探针所对应的编码标记物的组合状态为“标记物1无信号，标记物2无信号”，即“0，0”，随后获得奇偶校验标记物（图中方块，实心为有信号，空心为无信号）信号为“有信号”，即“1”，根据预设就校验原理，编码标记物状态有0个1，为0，对应的奇偶校验标记物应为1，与实际奇偶校验标记物状态相符。

图2是实例2和实例3的工作流程图。1，将P1、P2引物扩增后的PCR产物固定于玻片表面；2，第一条测序引物I1与固定于玻片表面的PCR产物杂交；3，测序引物I1上连接了测序碱基与PCR待测位置互补的寡核苷酸探针；4，完成检测后移除荧光修饰基团；5，I1连接进行第二次测序探针的连接；6，完成检测后再次移除荧光修饰基团；7，I1完成全部4次连接反应；8，测序引物I1与PCR产物变性分离；9，第二条测序引物I2与PCR产物杂交开始新一轮测序反应。

图3是实例2测序引物I1进行前两次测序连接反应的详细图解。1，测序引物I1与固定于玻片表面的PCR产物杂交，I1的全部15个碱基与PCR引物P2互补链5’端15个碱基互补配对；2，进行第一次连接反应，40种寡核苷酸探针（32种编码寡核苷酸探针，8种奇偶校验寡核苷酸探针）组成的混合物中测序碱基为“5’-AA-3’”的寡核苷酸单体（寡核苷酸探针其余位置为简并碱基N）与杂交于PCR产物上的I1连接，通过检测全部标记物仅获得CY7的信号，编码标记物状态为全部“无信号”，奇偶校验标记物为“有信号”，符合预先设定的奇偶校验规则，奇偶校验通过。查寡核苷酸探针配方表得知当编码标记物均为“无信号”时待测DNA的3’末端第3、4号碱基为“5’-AA-3’”；3，移除寡核苷酸探针末端的荧光修饰基团；4，进行第二次连接，检测待测DNA的3’末端第11、12号碱基。

具体实施方案

本发明的实施方法分为两种：

IDT Exhibit 1020.9859

第一种是针对某一种测序探针，选用若干种编码标记物和一种奇偶校验标记物同时对同一测序探针分子进行标记。针对一次连接反应，各个标记物在被检测时分别呈现“有信号”、“无信号”的两种状态，由向测序连接反应体系中加入的特定测序探针决定，制备含有不同编码标记物和背景标记物的测序探针是本发明的关键。加入反应体系中的测序探针由一组标记了不同标记物的寡核苷酸探针组成。每个寡核苷酸探针由一个或多个测序碱基（通过碱基互补配对法则测定待测 DNA 中相应位置的碱基信息）、一个或多个简并碱基 N（N 为 A、T、C、G 四种碱基中的任意一个）或非严格配对碱基（如次黄嘌呤等）和一种具有被检测能力的标记物构成。测序碱基与标记物之间的组合按照表 1 所列状态制备，当表 1 中所确定的“测序探针类型 i”对应 k 种标记物的状态为“1”时，则在制备测序碱基为“碱基组合类型 i”的测序探针分子时，同时标记表格中“测序探针类型 i”所有状态为“1”的标记物；逐行合成表 1 所列全部 2^{2n} 种寡核苷酸探针，该探针混合物即为所需制备的寡核苷酸探针混合物。

第二种是将该种测序探针分为若干部分，每一部分分别标记一种标记物（标记物包含编码标记物和奇偶标记物）。针对一次连接反应，各个标记物在被检测时分别呈现“有信号”、“无信号”的两种状态，由向测序连接反应体系中加入的特定测序探针决定。因此制备含有不同编码标记物和背景标记物的测序探针是本发明的关键。加入反应体系中的测序探针由一组标记了不同标记物的寡核苷酸探针组成。每个寡核苷酸探针由一个或多个测序碱基（通过碱基互补配对法则测定待测 DNA 中相应位置的碱基信息）、一个或多个简并碱基 N（N 为 A、T、C、G 四种碱基中的任意一个）或非严格配对碱基（如次黄嘌呤等）和一种具有被检测能力的标记物构成。测序碱基与标记物之间的组合按照表 1 所列状态制备，当表 1 中所确定的“碱基组合类型 i”与“标记物 j”所列状态为“1”时，则制备测序碱基为“碱基组合类型 i”、标记物为“标记物 j”的寡核苷酸探针；如所列状态为“0”时则不合成，合成表 1 所列全部状态为“1”的寡核苷酸探针，该探针混合物即为所需制备的寡核苷酸探针混合物。

实施例 1：带奇偶校验的信号组合编码的 DNA 连接测序方法，其特征在于针对某一种特定的 DNA 测序探针，利用一组编码标记物和一种信号编码标记物状态的组合进行标记，对于一批 DNA 测序探针，采用一组编码标记物状态结合信号编码标记物的不同组合方案进行标记，制备一套带奇偶校验的信号组合编码的 DNA 测序探针，从而实现在检测时对不同 DNA 测序探针的区分和鉴别：

A 一套带奇偶校验的信号组合编码的 DNA 测序探针的制备：首先制备未标记的 DNA 测序探针，每一种未标记的 DNA 测序探针由一个或多个测序碱基、一个或多个简并碱基 N 或非严格配对碱基组成。测序碱基用于通过碱基互补配对法则测定待测 DNA 中相应位置的碱基信息，简并碱基 N 为 A、T、C、G 四种碱基中的任意一个。完成

IDT Exhibit 1020.9860

未标记的 DNA 测序探针的制备后，针对每一种未标记的 DNA 测序探针，采用多种编码标记物状态的一种组合进行标记，每种编码标记物在该 DNA 测序探针上标记与否由该编码标记物在状态的组合中所对应的状态决定，同时结合奇偶校验标记物，当标记的编码标记物个数为奇数个时，则不标记奇偶校验标记物，当标记的编码标记物为偶数个时，则标记奇偶校验标记物。采用不同的状态组合方案并结合奇偶校验标记物对不同的未标记 DNA 测序探针进行标记，完成一套带奇偶校验的信号组合编码的 DNA 测序探针的制备。

B 利用上述一套带奇偶校验的信号组合编码的 DNA 测序探针的测序流程如下：

a. 选取一条测序引物与待测的 DNA 模板依据碱基互补配对原理进行杂交；

b. 向反应体系中加入上述的一套带奇偶校验的信号组合编码的 DNA 测序探针和 DNA 连接酶及其反应体系，进行 DNA 连接反应；

c. 完成 DNA 连接反应后，检测全部参与状态组合的标记物及奇偶校验标记物的信号，判读编码标记物信号所处的状态，计算获得信号的编码标记物的数量，并利用奇偶校验标记物的信号对该数值进行奇偶校验，如校验通过，根据全部编码标记物状态的组合情况，确定发生连接反应的 DNA 测序探针的种类，从而确定该测序探针上测序碱基的类型和排布，并最终确定被测 DNA 模板上对应位置的碱基或碱基序列信息；

d. 完成对标记物信号的检测后，移除 DNA 测序探针上的标记物；

e. 重复上述步骤 b-d 直至 DNA 测序探针达到或超出待测 DNA 模板待测区域末端；

f. 该测序引物与待测 DNA 模板变性分离，选取第二条测序引物；

g. 重复上述步骤 a-f，直至全部未知 DNA 模板的待测区域的全部序列信息得以确定。

所述的多种标记物对未标记 DNA 测序探针进行组合标记，其方案是对同一 DNA 测序探针分子同时标记一种或多种标记物，或者分别用不同的标记物标记同一种 DNA 测序探针的不同分子，随后将上述不同测序探针分子混合。

所述的测序引物是指一组只能和所有 DNA 模板的一段通用序列杂交的寡核苷酸片段，一组测序引物在 DNA 模板上杂交区域彼此相差一个或数个碱基，能够在数轮测序反应中完成对 DNA 模板全长的测序工作；测序模板上的通用序列，是在测序模板制备过程中通过连接反应加入，或者在扩增过程中通过引物引入。

所述的 DNA 模板是通过对待测的 DNA 片段通过 DNA 扩增技术增加基因组中感兴趣的目的片段的量，扩增是单重的，即一次扩增一个目的片段，或者是多重的，即一次扩增多个目的片段；所述的待测的 DNA 模板中的固定是通过化学或者物理方法固定于平面片基上，

IDT Exhibit 1020.9861

或固定于“96 孔板”、“384 孔板”及各种修饰的珠子载体上。

所述的连接反应，可以是自测序引物开始 5’末端向 3’末端的连接反应，也可以是 3’末端向 5’末端的连接反应。

所述的标记物，是采用的荧光基团，或利用测序探针化学的、物理的性质的改变，如电阻变化、电流变化。

所述的检测是指与所述的标记物性质相适应的，在激光激发的荧光标记基团时，激光激发强度可为 95%、70%、45%等多个不同强度，光电倍增可以分别为 95%、70%、45%等多个检测区间。

所述的标记物的移除，是标记基团通过化学的、物理的方法与 DNA 测序探针分离，或者 DNA 测序探针化学的、物理的性质恢复原有状态。标记物的移除，是仅仅移除标记物本身，或者同时移除与标记物相连或不相连的一个或多个碱基及相关基团。

所述的第二条测序引物是一组测序引物中除已经使用的测序引物中的任何一条，并不一定是与已使用的测序引物对应在测序模板上杂交位置最为接近的一条。

实施例 2：带奇偶校验的四色荧光复合编码法（分别标记方案）检测人类 17 号染色体上，RP11-354P11 克隆上 45332-45363 共 32 个碱基对序列；

提取人类全基因组 DNA，利用 PCR 方法扩增目的片段，引物为 P1、P2（序列见表 2），引物 P1 的 5’末端采用羟基修饰。扩增后的 PCR 产物固定于醛基修饰的玻片表面后，加热玻片至 95℃使双链 PCR 产物的一条链与固定链脱离，随后开始测序反应。

表 2 实例 2 中相关寡核苷酸序列信息

序列名	序列信息
待测序列	5’-CACGGACCAGCTGCCCTGGACCAGCTGCAAGA-3’
P1	OH-5’-CGCTATACTACCTCATCTCCTCCTTCACG-3’
P2	5’-GCAGTTGCCAGTGTTCCAGGAGT-3’
I1	5’-CAGTGTTCCAGGAGT-3’
I2	5’-GTGTTCCAGGAGTNN-3’
I3	5’-GTTCCAGGAGTNNNN-3’
I4	5’-GCCAGTGTTCCAGGA-3’

向玻片表面加入杂交缓冲液和第一条测序引物 I1（序列见表 2），37℃杂交复性 20 分钟。向反应池中加入表 3 所列的全部 40 种测序探针组成的混合物，其中 CY7 为奇偶校验标记物，其余四种标记物为编码标记物，由标记物 CY7 组成的寡核苷酸探针仅出现在相应的测序碱基对应编码寡核苷酸探针种类为 0 种或偶数种时，而测序碱基对应的编码寡核苷酸探针为奇数

IDT Exhibit 1020.9862

种时，不合成由该测序碱基和 CY7 构成的寡核苷酸探针。采用 T4 核酸连接酶进行连接反应，反应条件为 25℃连接 30 分钟。连接完成后利用激光共聚焦显微镜对玻片进行扫描，扫描分别采用对应于 CY7、CY3、CY5、TXR 和 FTC 的波长进行，根据获得的荧光信号判读对应的碱基信息（附图 2）。

表 3 测序探针混合物详细列表

	测序探针类型（从 5’ 末端到 3’ 末端）				
测序碱基	CY7	CY3	CY5	TXR	FTC
AA	NNAANNNN-CY7	NNAANNNN-CY3	NNAANNNN-CY5	NNAANNNN-TXR	NNAANNNN-FTC
AC	无	NNACNNNN-CY3	NNACNNNN-CY5	NNACNNNN-TXR	无
AG	无	NNAGNNNN-CY3	NNAGNNNN-CY5	无	NNAGNNNN-FTC
AT	NNATNNNN-CY7	NNATNNNN-CY3	NNATNNNN-CY5	无	无
CA	无	NNCANNNN-CY3	无	NNCANNNN-TXR	NNCANNNN-FTC
CC	NNCCNNNN-CY7	NNCCNNNN-CY3	无	NNCCNNNN-TXR	无
CG	NNCGNNNN-CY7	NNCGNNNN-CY3	无	无	NNCGNNNN-FTC
CT	无	NNCTNNNN-CY3	无	无	无
GA	无	无	NNGANNNN-CY5	NNGANNNN-TXR	NNGANNNN-FTC
GC	NNGCNNNN-CY7	无	NNGCNNNN-CY5	NNGCNNNN-TXR	无
GG	NNGGNNNN-CY7	无	NNGGNNNN-CY5	无	NNGGNNNN-FTC
GT	无	无	NNGTNNNN-CY5	无	无
TA	NNTANNNN-CY7	无	无	NNTANNNN-TXR	NNTANNNN-FTC
TC	无	无	无	NNTCNNNN-TXR	无
TG	无	无	无	无	NNTGNNNN-FTC
TT	NNTTNNNN-CY7	无	无	无	无

在第一次连接反应中，I1 引物与 P2 互补链上的“5’-AAGAACTCCTGGAACACTG-3’”序列杂交，加入寡核苷酸探针混合物后，寡核苷酸探针混合物中测序碱基如果与待测 DNA 位于 I1 引物 3’ 端下游的第 3、4 号碱基（也是待测 DNA 的 3’ 末端第 3、4 号碱基）互补配对，则这些寡核苷酸探针与待测 DNA 杂交并与引物 I1 发生连接反应。待测 DNA 位于 I1 引物 3’ 端下游的第 3、4 个碱基为“5’-AA-3’”，其反向互补碱基为“5’-TT-3’”，则所有测序碱基为“5’-TT-3’”的寡核苷酸探针仅有一条，为“5’-NNTTNNNN-3’-CY7”，与待测 DNA 模板及引物 I1 发生连接反应（附图 3-2）。在进行信号检测时，首先检测了编码标记物的复合状态类型，为“CY3 无信号，CY5 无信号，TXR 无信号，FTC 无信号”，共有 0 种标记物有信号，同时检测到奇偶校验标记物状态为“有信号”，符合编码标记物为 0 种或偶数种标记物“有信号”时，奇偶校验标记物为“有信号”的奇偶校验规则，奇偶校验通过。随后通过查表得知当编

IDT Exhibit 1020.9863

码标记物状态均为“无信号”时，待测 DNA 的 3’末端第 3、4 号碱基的反向互补链信息为“5’-TT-3’”，从而可以获得待测 DNA 的 3’末端第 3、4 号碱基为“5’-AA-3’”。

完成次轮检测后，利用化学方法切除链接与引物 I1 上寡核苷酸探针 3’末端的荧光基团，在 3’末端获得一个游离的羟基进行再次连接反应（附图 3-3）。向反应池中再次加入表 3 所列全部苷酸探针混合物，进行第二轮连接反应，此次连接反应是检测待测 DNA 位于 I1 引物 3’端下游的第 11、12 号碱基（待测 DNA 的 3’末端第 11、12 号碱基）“5’-CC-3’”，此时三种测序碱基为“5’-GG-3’”的寡核苷酸探针“5’-NNGGNNNN-3’-CY7”、“5’-NNGGNNNN-3’-CY5”、“5’-NNGGNNNN-3’-FTC”发生连接反应，进行荧光检测是同时检测到 CY7、CY5、FTC，奇偶校验标记物验证成功，随后查表判读并进行反向互补获得待测 DNA 的 3’末端第 11、12 号碱基为“5’-CC-3’”（附图 3-4）。化学方法切除连接上的寡核苷酸末端荧光基团后再次进行 I1 引物的第 3、4 轮连接反应，分别测得待测 DNA 的 3’末端第 19、20 号碱基和第 27、28 号碱基信息。

完成四轮连接反应后，95℃使测序引物 I1 与待测 DNA 分离，向玻片表面加入杂交缓冲液和第二条测序引物 I2（序列见表 2），37℃杂交复性 20 分钟。随后开始四轮连接反应分别检测待测 DNA 的 3’末端第 5、6 号，13、14 号，21、22 号以及 29、30 号四组碱基。

同样的，利用测序引物 I3 检测待测 DNA 的 3’末端第 7、8 号，15、16 号，23、24 号以及 31、32 号四组碱基，利用测序引物 I4 检测待测 DNA 的 3’末端第 1、2 号，9、10 号，17、18 号以及 25、26 号四组碱基。

通过所获得序列信息进行拼接即可获得待测 DNA 全部 32 个碱基的序列信息。

实施例 3：带奇偶校验的四色荧光复合编码法（共同标记方案）检测人类 17 号染色体上，RP11-354P11 克隆上 45332-45363 共 32 个碱基对序列：

表 4 实例 3 中相关寡核苷酸序列信息

序列名	序列信息
待测序列	5’-CACGGACCAGCTGCCCTGGACCAGCTGCAAGA-3’
P1	OH-5’-CGCTATACTACCTCATCTCCTCCTTCACG-3’
P2	5’-GCAGTTGCCAGTGTTCCAGGAGT-3’
I1	5’-CAGTGTTCCAGGAGT-3’
I2	5’-GTGTTCCAGGAGTNN-3’
I3	5’-GTTCCAGGAGTNNNN-3’
I4	5’-GCCAGTGTTCCAGGA-3’

提取人类全基因组 DNA，利用 PCR 方法扩增目的片段，引物为 P1、P2（序列见表 4），引

IDT Exhibit 1020.9864

物 P1 的 5’ 末端采用羟基修饰。扩增后的 PCR 产物固定于醛基修饰的玻片表面后，加热玻片至 95℃使双链 PCR 产物的一条链与固定链脱离，随后开始测序反应。

向玻片表面加入杂交缓冲液和第一条测序引物 I1（序列见表 4），37℃杂交复性 20 分钟。向反应池中加入表 5 所列的全部 16 种寡核苷酸探针组成的混合物，表格中每行为一种寡核苷酸探针，其中 CY7 为奇偶校验标记物，其余四种标记物为编码标记物。每个寡核苷酸探针分子上同时标记了一种或多种的信号标记物，标记物 CY7 仅参与标记那些已经标记了 0 种或偶数种信号编码标记物的测序探针，而测序探针已标记的信号标记物为奇数种时，不标记 CY7 基团。采用 T4 核酸连接酶进行连接反应，反应条件为 25℃连接 30 分钟。连接完成后利用激光共聚焦显微镜对玻片进行扫描，扫描分别采用对应于 CY7、CY3、CY5、TXR 和 FTC 的波长进行，根据获得的荧光信号判读对应的碱基信息（附图 2）。

在第一次连接反应中，I1 引物与 P2 互补链上的“5’-AAGAACTCCTGGAACACTG-3’”序列杂交，加入寡核苷酸探针混合物后，寡核苷酸探针混合物中测序碱基如果与待测 DNA 位于 I1 引物 3’ 端下游的第 3、4 号碱基（也是待测 DNA 的 3’ 末端第 3、4 号碱基）互补配对，则这些寡核苷酸探针与待测 DNA 杂交并与引物 I1 发生连接反应。待测 DNA 位于 I1 引物 3’ 端下游的第 3、4 个碱基为“5’-AA-3’”，其反向互补碱基为“5’-TT-3’”，测序探针“5’-NNTTNNNN-3’”仅标记了 CY7 一种荧光基团，与待测 DNA 模板及引物 I1 发生连接反应。在进行信号检测时，首先检测了编码标记物的复合状态类型，为“CY3 无信号，CY5 无信号，TXR 无信号，FTC 无信号”，共有 0 种标记物有信号，同时检测到奇偶校验标记物状态为“有信号”，符合编码标记物为 0 种或偶数种标记物“有信号”时，奇偶校验标记物为“有信号”的奇偶校验规则，奇偶校验通过。随后通过查表得知当编码标记物状态均为“无信号”时，待测 DNA 的 3’ 末端第 3、4 号碱基的反向互补链信息为“5’-TT-3’”，从而可以获得待测 DNA 的 3’ 末端第 3、4 号碱基为“5’-AA-3’”。

完成次轮检测后，利用化学方法切除链接与引物 I1 上寡核苷酸探针 3’ 末端的荧光基团，在 3’ 末端获得一个游离的羟基进行再次连接反应。向反应池中再次加入表 5 所列全部苷酸探针混合物，进行第二轮连接反应，此次连接反应是检测待测 DNA 位于 I1 引物 3’ 端下游的第 11、12 号碱基（待测 DNA 的 3’ 末端第 11、12 号碱基）“5’-CC-3’”，此时测序探针“5’-NNGGNNNN-3’”上同时标记了 CY7、CY5 和 FTC 三种基团。发生连接反应后，进行荧光检测是同时检测到 CY7、CY5、FTC，奇偶校验标记物验证成功，随后查表判读并进行反向互补获得待测 DNA 的 3’ 末端第 11、12 号碱基为“5’-CC-3’”。化学方法切除连接上的寡核苷酸末端荧光基团后再次进行 I1 引物的第 3、4 轮连接反应，分别测得待测 DNA 的 3’ 末端第 19、20 号碱基和第 27、28 号碱基信息。

IDT Exhibit 1020.9865

完成四轮连接反应后，95℃使测序引物 I1 与待测 DNA 分离，向玻片表面加入杂交缓冲液和第二条测序引物 I2（序列见表 4），37℃杂交复性 20 分钟。随后开始四轮连接反应分别检测待测 DNA 的 3’末端第 5、6 号，13、14 号，21、22 号以及 29、30 号四组碱基。

表 5 寡核苷酸探针混合物详细列表

		共同修饰基团成分				
测序碱基	测序探针（从 5’末端到 3’末端）	CY7	CY3	CY5	TXR	FTC
AA	NNAANNNN	有	有	有	有	有
AC	NNACNNNN	无	有	有	有	无
AG	NNAGNNNN	无	有	有	无	有
AT	NNATNNN	有	有	有	无	无
CA	NNCANNNN	无	有	无	有	有
CC	NNCCNNNN	有	有	无	有	无
CG	NNCGNNNN	有	有	无	无	有
CT	NNCTNNNN	无	有	无	无	无
GA	NNGANNNN	无	无	有	有	有
GC	NNGCNNNN	有	无	有	有	无
GG	NNGGNNNN	有	无	有	无	有
GT	NNGTNNNN	无	无	有	无	无
TA	NNTANNNN	有	无	无	有	有
TC	NNTCNNNN	无	无	无	有	无
TG	NNTGNNNN	无	无	无	无	有
TT	NNTTNNNN	有	无	无	无	无

同样的，利用测序引物 I3 检测待测 DNA 的 3’末端第 7、8 号，15、16 号，23、24 号以及 31、32 号四组碱基，利用测序引物 I4 检测待测 DNA 的 3’末端第 1、2 号，9、10 号，17、18 号以及 25、26 号四组碱基。

通过所获得序列信息进行拼接即可获得待测 DNA 全部 32 个碱基的序列信息。

IDT Exhibit 1020.9866

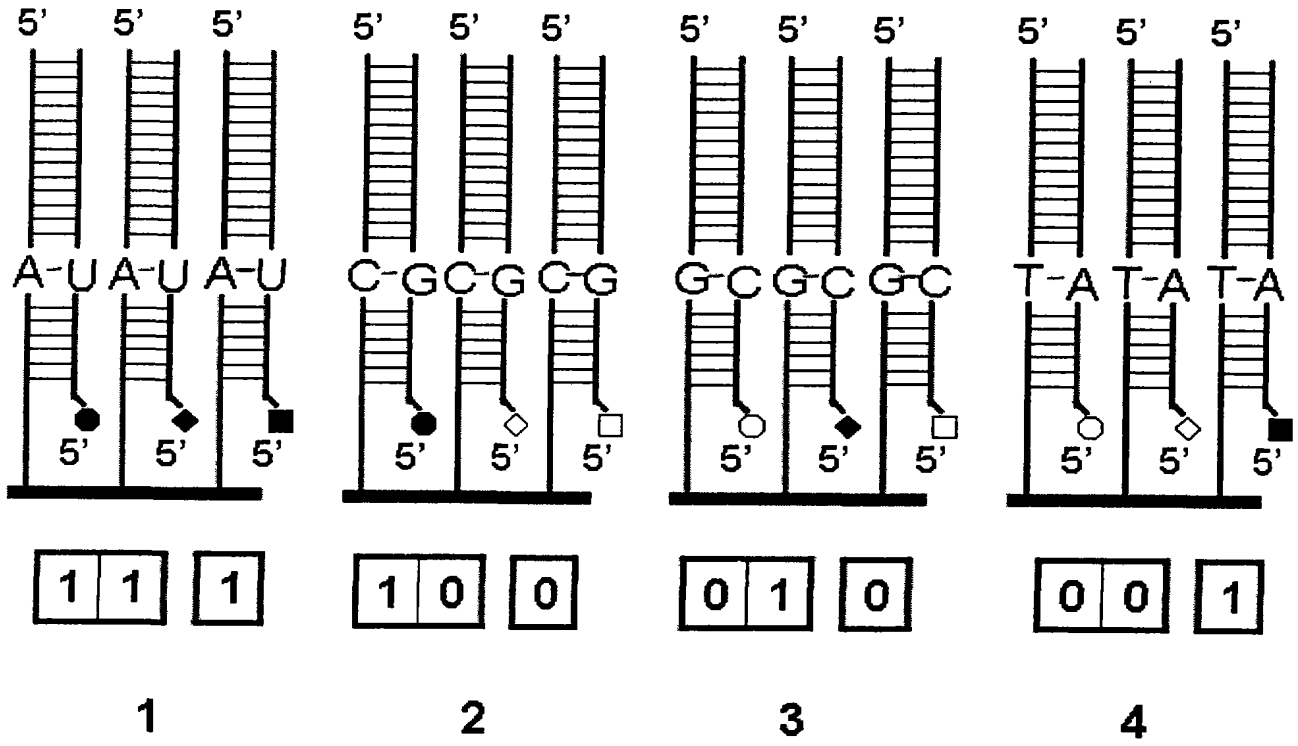


图 1

IDT Exhibit 1020.9867

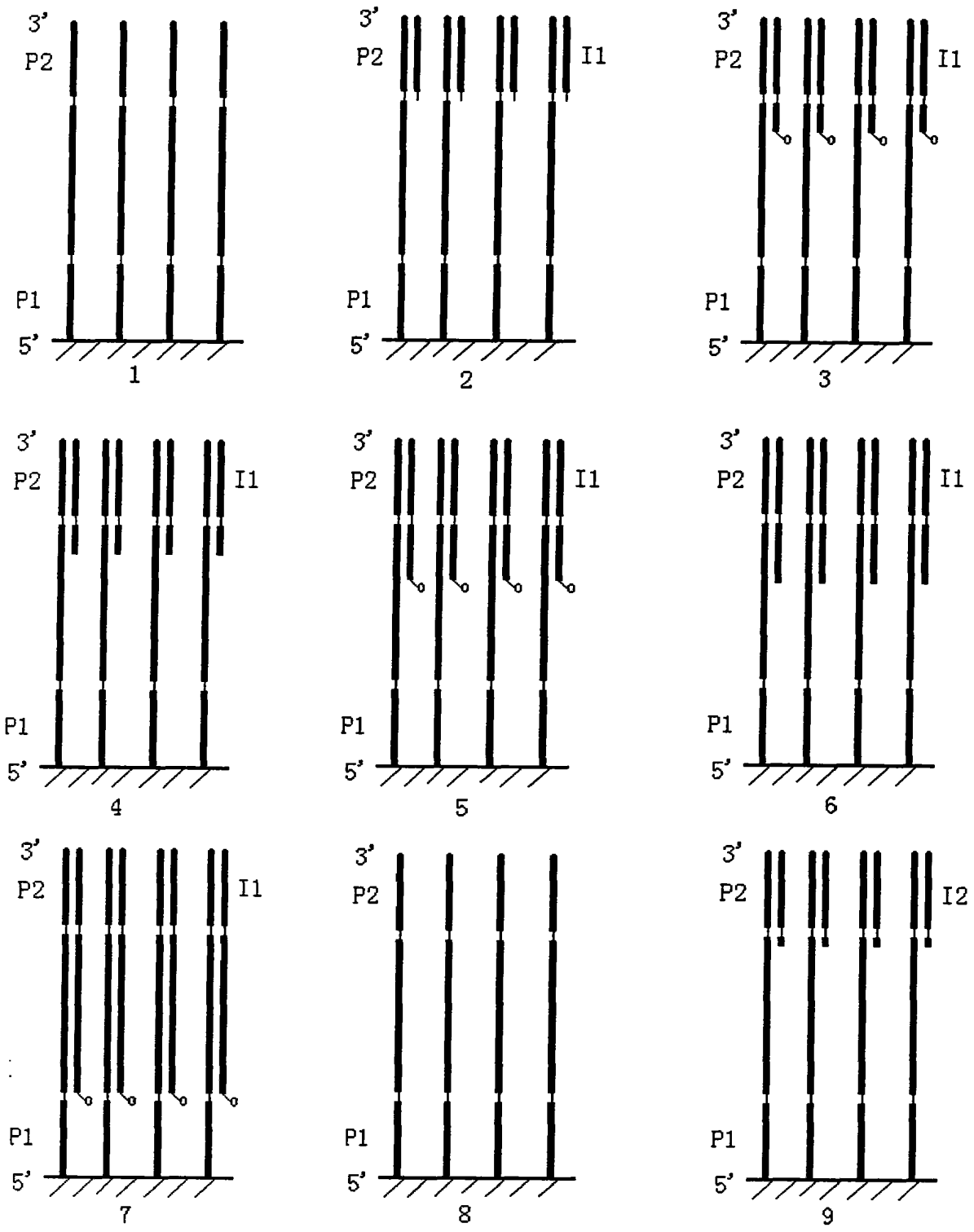


图 2

IDT Exhibit 1020.9868

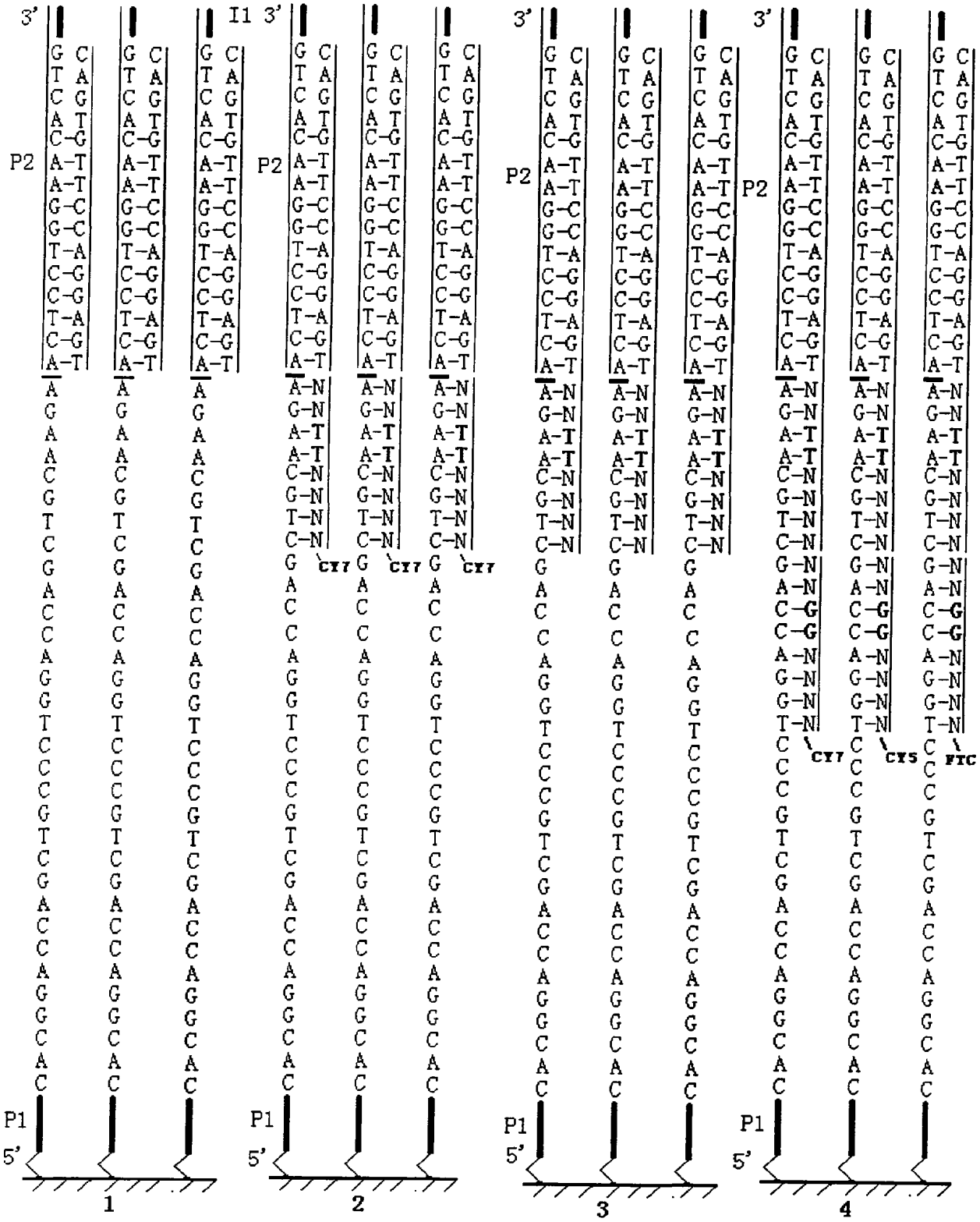


图 3

IDT Exhibit 1020.9869

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
19 August 2004 (19.08.2004)

PCT

(10) International Publication Number
WO 2004/070007 A2

(51) International Patent Classification[7]: C12N

(21) International Application Number:
PCT/US2004/002571

(22) International Filing Date: 28 January 2004 (28.01.2004)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

60/443,471	29 January 2003 (29.01.2003)	US
60/465,071	23 April 2003 (23.04.2003)	US
60/476,602	6 June 2003 (06.06.2003)	US
60/476,504	6 June 2003 (06.06.2003)	US
60/476,313	6 June 2003 (06.06.2003)	US
60/476,592	6 June 2003 (06.06.2003)	US
60/497,985	25 August 2003 (25.08.2003)	US

(63) Related by continuation (CON) or continuation-in-part (CIP) to earlier applications:

US	60/443,471 (CON)
Filed on	29 January 2003 (29.01.2003)
US	60/465,071 (CON)
Filed on	23 April 2003 (23.04.2003)

(71) Applicant *(for all designated States except US)*: **454 CORPORATION** [US/US]; 20 Commercial Street, Branford, CT 06405 (US).

(72) Inventors; and

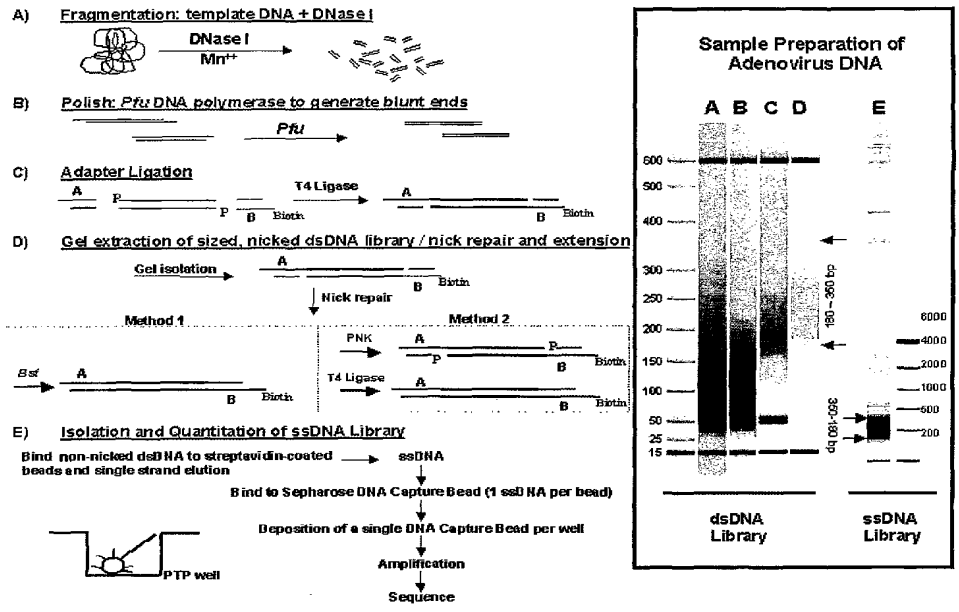
(75) Inventors/Applicants *(for US only)*: **COSTA, Gina, L.** [US/US]; 12 Pershing Avenue, Beverly, MA 01915 (US). **WEINER, Michael, P.** [US/US]; 52 Leighton Trail, Guilford, CT 06437 (US). **LEAMON, John** [US/US]; 251 Nortontown Road, Guilford, CT 06437 (US). **ROTHBERG, Jonathan** [US/US]; 1701 Moose Hill Road, Guilford, CT 06437 (US).

(74) Agent: **ELRIFI, Ivor, R.**; Mintz, Levin, Cohn, Ferris, Glovsky and Popeo, P.C, 666 Third Avenue, New York, NY 10017 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NI, NO, NZ, OM, PG,

[Continued on next page]

(54) Title: METHOD FOR PREPARING SINGLE-STRANDED DNA LIBRARIES



(57) Abstract: This invention relates to methods of generating single stranded DNA libraries for use in amplification and sequencing reactions. In various aspects, the disclosed methods include: fragmenting DNA; polishing the fragments' ends; lighting the fragments to universal adapters; performing strand displacement and extension of the nicked fragments; purifying the double-stranded libation products; capturing the double-stranded libation products onto a solid support; and isolating single stranded DNA library fragments, and binding these fragments to another solid support.

WO 2004/070007 A2

PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

METHOD FOR PREPARING SINGLE-STRANDED DNA LIBRARIES

RELATED APPLICATIONS

This application claims the benefit of priority to the following applications: USSN 60/443,471 filed January 29, 2003, USSN 60/465,071 filed April 23, 2003; USSN 60/476,504 filed June 6, 2003, USSN 60/476,313 filed June 6, 2003, USSN 60/476,592 filed June 6, 2003, USSN 60/476,602 filed June 6, 2003, USSN 60/476,592 filed June 6, 2003, and USSN 60/497,985 filed August 25, 2003. All patent and patent applications in this paragraph are hereby incorporated herein by reference in their entirety.

This application also incorporates by reference the following copending U.S. patent applications: "Bead Emulsion Nucleic Acid Amplification" filed January 29, 2004, "Bead Emulsion Nucleic Acid Amplification with Continuous Flow" filed January 29, 2004, "Double Ended Sequencing" filed January 29, 2004, and "Methods Of Amplifying And Sequencing Nucleic Acids" filed January 29, 2004.

FIELD OF THE INVENTION

The invention relates to protein chemistry, molecular biology, and methods of preparing single-stranded libraries for sequence analysis. More specifically, this invention includes methods of processing DNA for use in amplification and sequencing reactions.

BACKGROUND OF THE INVENTION

In amplification by polymerase chain reaction (PCR), two primers are designed to hybridize to the template DNA at positions complementary to respective primers that are separated on the DNA template molecule by some number of nucleotides. The base sequence of the template DNA between and including the primers is amplified by repetitive complementary strand extension reactions whereby the number of copies of the target DNA fragments is increased by several orders of magnitude. Amplification is exponential as 2^n, where n equals the number of amplification cycles. Following PCR, the amplified DNA may be sequenced through conventional sequencing methods (see, U.S. Patent No. 6,274,320).

Samples comprising large template DNA or whole DNA genomes comprising long nucleotide sequences are not conducive to efficient amplification by PCR. These long molecules do not naturally possess sequences useful for primer hybridization. In addition, if primer hybridization sequences are added to double stranded DNA molecules, it is

IDT Exhibit 1020.9872

difficult to ascertain the directionality of the amplified DNA molecules and this frustrates sequencing efforts.

Various methods have been designed to overcome some of these deficiencies. For example, United States Patent No. 5,508,169 describes that subsets of nucleic acid fragments may be indexed (i.e., selected or targeted) based upon the information contained in non-identical 5'-protruding or 3'-protruding cohesive ends. This includes fragments having 3, 4 or 5 base cohesive ends, such as those revealed by cleavage of DNA by Type II restriction endonucleases and interrupted palindrome recognizing type II restriction endonucleases. The patent describes nucleic acid molecules similar to adaptors (called indexing linkers) which contain protruding single strands complementary to the cohesive ends of cleavage sites of restriction endonucleases (rather than the recognition sequences). Various functional groups or specific nucleic acid sequences designed for particular applications may be selectively attached to the aforementioned subsets of fragments. Selective attachment of indexing linkers having known base sequences in their cohesive ends to a subset of fragments bearing the complementary cohesive ends can be used for the detection, identification, isolation, amplification, and manipulation of the subset of fragments.

United States Patent No. 6,468,748 describes a method of sorting genes and/or gene fragments comprising several steps. First, ds cDNA molecules are prepared from mRNA molecules by reverse transcription, using a poly-T primer optionally having a general primer-template sequence upstream from the poly-T sequence, yielding ds cDNA molecules having the poly-T sequence, optionally having the general primer-template sequence. Second, the ds cDNA molecules are digested with a restriction enzyme that produces digested cDNA molecules with cohesive ends having overhanging ssDNA sequences of a constant number of arbitrary nucleotides. Third, the digested cDNA molecules are ligated to a set of dsDNA oligonucleotide adaptors, each of which adaptors has at one of its ends a cohesive-end ssDNA adaptor sequence complementary to one of the possible overhanging ssDNA sequences of the digested cDNA, at the opposite end a specific primer-template sequence specific for the ssDNA adaptor complementary sequence, and in between the ends a constant sequence that is the same for all of the different adaptors of the set. Fourth, the ligated cDNA molecules are amplified by separate polymerase chain reactions, utilizing for each separate polymerase chain reaction a primer that anneals to the cDNA poly-T sequence optionally having the cDNA general primer-template, and a primer from a set of different specific primers that anneal to the cDNA specific primer-template sequences. Fifth, the amplified

IDT Exhibit 1020.9873

cDNA molecules are sorted into nonoverlapping groups by collecting the amplification products after each separate polymerase chain reaction, each group of amplified cDNA molecules determined by the specific primer that annealed to the specific primer-template sequence and primed the polymerase chain reaction.

United States Patent No. 5,863,722 describes a method and materials for sorting polynucleotides with oligonucleotide tags. The oligonucleotide tags are capable of hybridizing to complementary oligomeric compounds consisting of subunits having enhanced binding strength and specificity as compared to natural oligonucleotides. Such complementary oligomeric compounds are referred to as "tag complements." Subunits of tag complements may consist of monomers of non-natural nucleotide analogs, referred to as "antisense monomers" or they may comprise oligomers having lengths in the range of 3 to 6 nucleotides or analogs thereof, including antisense monomers, the oligomers being selected from a minimally cross-hybridizing set. In such a set, a duplex made up of an oligomer of the set and the complement of any other oligomer of the set contains at least two mismatches. In other words, an oligomer of a minimally cross-hybridizing set at best forms a duplex having at least two mismatches with the complement of any other oligomer of the same set. Tag complements attached to a solid phase support are used to sort polynucleotides from a mixture of polynucleotides each containing a tag. The surface of each support is derivatized by only one type of tag complement which has a particular sequence. Similarly, the polynucleotides to be sorted each comprise an oligonucleotide tag in the repertoire, such that identical polynucleotides have the same tag and different polynucleotides have different tags. Thus, when the populations of supports and polynucleotides are mixed under conditions which permit specific hybridization of the oligonucleotide tags with their respective complements, subpopulations of identical polynucleotides are sorted onto particular beads or regions. The subpopulations of polynucleotides can then be manipulated on the solid phase support by micro-biochemical techniques.

United States Patent No. 5,728,524 describes a process for the categorization of nucleic acid sequences in which these sequences are linked to a population of adaptor molecules, each exhibiting specificity for linking to a sequence including a predetermined nucleotide base. The resulting linked sequences are then categorized based upon selection for the particular base.

However, the art does not describe methods for generating libraries of unknown fragment sequences additionally comprising two known sequences, each different than the

IDT Exhibit 1020.9874

other, one being adjoined at each end. Thus, a need exists for a method which overcomes shortcomings of the prior art. Accordingly, the present invention is directed to describing such methods, materials, and kits as required to facilitate manipulation of multiple DNA sequences in a sample.

BRIEF SUMMARY OF THE INVENTION

This invention describes a novel method for preparing a library of multiple nucleic acid sequences from a sample where the library is suited to further quantitative and comparative analysis, particularly where the multiple nucleic acid sequences are unknown and derived from large template DNA or whole (or partial) genome DNA. In certain embodiments of the invention, sequences of single stranded DNA (ssDNA) are prepared from a sample of large template DNA or whole or partial DNA genomes through fragmentation, polishing, adaptor ligation, nick repair, and isolation of ssDNA.

Therefore, in one aspect, the present invention provides a method for clonally isolating a library comprising a plurality of ssDNAs, wherein each ssDNA comprises a first single stranded universal adaptor and a second single stranded universal adaptor, the method comprising:

(a) fragmenting large template DNA molecules to generate a plurality of fragmented DNA molecules;

(b) attaching a first or second universal double stranded adaptor to a first end of each fragmented DNA molecule and a first or second universal adaptor to a second end of each fragmented DNA molecule to form a mixture of adaptor ligated DNA molecules;

(c) isolating a plurality of single stranded DNA molecules each comprising a first single stranded universal adaptor and a second single stranded universal adaptor; and

(d) delivering the single stranded DNA molecules into reactors such that a plurality of the reactors include one DNA molecule, thereby clonally isolating the library.

In certain aspects, the single stranded DNA molecules are delivered into droplets in a water-in-oil emulsion (i.e., microreactors), or onto multiwell surfaces (e.g., PicoTiter plates). The single stranded DNA molecules may be delivered via attachment to a solid support (e.g., beads).

In other aspects, the adaptor ligated DNA molecules comprising a first double stranded universal adaptor and second double stranded universal adaptor is attached to a solid support via one strand of the double stranded universal adaptor (via the first or second

IDT Exhibit 1020.9875

universal adaptor). The adaptor ligated DNA molecules which have not attached to a solid support are washed away, and one strand of the adaptor ligated DNA molecules is released. This generates a mixture comprising a plurality of ssDNAs comprising a population of single stranded molecules with a first and second universal adaptor pair, thereby generating a library.

The sequence of the fragmented DNA may be known or unknown. In a preferred embodiment, the sequence of the fragmented DNA, particularly the sequence of the ends of the fragmented DNA, is unknown.

In another aspect, the present invention includes a method for generating a ssDNA library linked to solid supports comprising: (a) generating a library of ssDNA templates; (b) attaching the ssDNA templates to solid supports; and (c) isolating the solid supports on which one ssDNA template is attached. In still another aspect, the present invention includes a library of mobile solid supports made by the method disclosed herein.

BRIEF DESCRIPTION OF THE FIGURES

Figure 1 is a schematic representation of the entire process of library preparation including the steps of template DNA fragmentation (Figure 1A), end polishing (Figure 1B), adaptor ligation (Figure 1C), nick repair, strand extension and gel isolation (Figure 1D). Figure 1 also depicts a representative agarose gel containing a sample preparation of a 180-350 base pair adenovirus DNA library according to the methods of this invention.

Figure 2A is a schematic representation of the universal adaptor design according one embodiment of the present invention. Each universal adaptor is generated from two complementary ssDNA oligonucleotides that are designed to contain a 20 bp nucleotide sequence for PCR priming, a 20 bp nucleotide sequence for sequence priming and a unique 4 bp discriminating sequence comprised of a non-repeating nucleotide sequence (*i.e.*, ACGT, CAGT, etc.).

Figure 2B depicts a representative universal adaptor sequence pair for use with the invention. Adaptor A sense strand: SEQ ID NO:1; Adaptor A antisense strand: SEQ ID NO:2; Adaptor B sense strand: SEQ ID NO:3; Adaptor B antisense strand: SEQ ID NO:4.

Figure 2C is a schematic representation of universal adaptor design for use with the invention.

Figure 3 represents the strand displacement and extension of nicked double-stranded DNA fragments according to the present invention. Following the ligation of universal

IDT Exhibit 1020.9876

adaptors generated from synthetic oligonucleotides, double-stranded DNA fragments will be generated that contain two nicked regions following T4 DNA ligase treatment (Figure 3A). The addition of a strand displacing enzyme (i.e., *Bst* DNA polymerase I) will bind nicks (Figure 3B), strand displace the nicked strand and complete nucleotide extension of the strand (Figure 3C) to produce non-nicked double-stranded DNA fragments (Figure 3D).

Figure 4 represents the isolation of directionally-ligated single-stranded DNA according to the present invention using streptavidin-coated magnetic beads. Following ligation with universal adaptors A and B (the two different adaptors are sometimes referred to as a "first" and "second" universal adaptor), double-stranded DNA will contain adaptors in four possible combinations: AA, BB, AB, and BA. When universal adaptor B contains a 5'-biotin, magnetic streptavidin-coated solid supports are used to capture and isolate the AB, BA, and BB populations (population AA is washed away). The BB population is retained on the beads as each end of the double-stranded DNA is attached to a bead and is not released. However, upon washing in the presence of a low salt buffer, only populations AB and BA will release a single-stranded DNA fragment that is complementary to the bound strand. Single-stranded DNA fragments are isolated from the supernatant and used as template for subsequent applications. This method is described below in more detail.

Figure 5 represents an insert flanked by PCR primers and sequencing primers.

Figure 6 represents truncated product produced by PCR primer mismatch at cross-hybridization region (CHR).

Figures 7A-7D depict the assembly for the nebulizer used for the methods of the invention. A tube cap was placed over the top of the nebulizer (Figure 7A) and the cap was secured with a nebulizer clamp assembly (Figure 7B). The bottom of the nebulizer was attached to the nitrogen supply (Figure 7C) and the entire device was wrapped in parafilm (Figure 7D).

Figure 8 depicts representative BioAnalyzer output from analysis of a single stranded DNA library.

Figure 9A depicts representative results for LabChip analysis of a single stranded DNA library following nebulization and polishing.

Figure 9B depicts representative size distribution results for an adaptor-ligated single stranded DNA library following nebulization, polishing, and gel purification.

Figure 10 depicts the calculation for primer candidates based on melting temperature.

IDT Exhibit 1020.9877

DETAILED DESCRIPTION OF THE INVENTION

This invention relates to the preparation of sample DNA for amplification and sequencing reactions. The invention includes a method for preparing the sample DNA comprised of the following steps: (a) fragmenting large template DNA or whole genomic DNA samples to generate a plurality of digested DNA fragments; (b) creating compatible ends on the plurality of digested DNA samples; (c) ligating a set of universal adaptor sequences onto the ends of fragmented DNA molecules to make a plurality of adaptor-ligated DNA molecules, wherein each universal adaptor sequence comprises a PCR primer sequence, a sequencing primer sequence and a discriminating key sequence and wherein one adaptor is attached to biotin; (d) separating and isolating the plurality of ligated DNA fragments; (e) removing any portion of the plurality of ligated DNA fragments; (f) nick repair and strand extension of the plurality of ligated DNA fragments; (g) attaching each of the ligated DNA fragments to a solid support; and (h) isolating populations comprising single-stranded adaptor-ligated DNA fragments for which there is a unique adaptor at each end (i.e., providing directionality).

Unless otherwise defined, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. Methods and materials similar or equivalent to those described herein can be used in the practice of the present invention, and exemplified suitable methods and materials are described below. For example, methods may be described which comprise more than two steps. In such methods, not all steps may be required to achieve a defined goal and the invention envisions the use of isolated steps to achieve these discrete goals. The disclosures of all publications, patent applications, patents, and other references are incorporated *in toto* herein by reference. In addition, the materials, methods, and examples are illustrative only and not intended to be limiting.

As used herein, the term "universal adaptor" refers to two complementary and annealed oligonucleotides that are designed to contain a nucleotide sequence for PCR priming and a nucleotide sequence for sequence priming. Optionally, the universal adaptor may further include a unique discriminating key sequence comprised of a non-repeating nucleotide sequence (*i.e.*, ACGT, CAGT, etc.). A set of universal adaptors comprises two unique and distinct double-stranded sequences that can be ligated to the ends of double-stranded DNA. Therefore, the same universal adaptor or different universal adaptors can be ligated to either end of the DNA molecule. When comprised in a larger DNA molecule that

IDT Exhibit 1020.9878

is single stranded or when present as an oligonucleotide, the universal adaptor may be referred to as a single stranded universal adaptor.

As used herein, the term "discriminating key sequence" refers to a sequence including a combination of the four deoxyribonucleotides (i.e., A, C, G, T). The same discriminating sequence can be used for an entire library of DNA fragments. Alternatively, different discriminating key sequences can be used to track libraries of DNA fragments derived from different organisms. Longer discriminating key sequences can be used for a mixture of more than one library.

As used herein, the term "plurality of molecules" refers to DNA isolated from the same source, whereby different organisms may be prepared separately by the same method. In one embodiment, the plurality of DNA samples is derived from large segments of DNA, e.g., genomic DNA, cDNA, viral DNA, plasmid DNA, cosmid DNA, artificial chromosome DNA (e.g., BACs, YACs, MACs, PACs), synthetic DNA, phagemid DNA, phasemid DNA, or from reverse transcripts of viral RNA. This DNA may be derived from any source, including any mammal (i.e., human, nonhuman primate, rodent, or canine), plant, bird, reptile, fish, fungus, bacteria, or virus.

As used herein, the term "library" refers to a subset of smaller sized DNA species generated from a larger DNA template, e.g., a segmented or whole genome.

As used herein, the term "unique", as in "unique PCR priming regions" refers to a sequence that does not exist or exists at an extremely low copy level within the DNA molecules to be amplified or sequenced.

As used herein, the term "compatible" refers to an end of double stranded DNA to which an adaptor molecule may be attached (i.e., blunt end or cohesive end).

As used herein, the term "fragmenting" refers to a process by which a larger molecule of DNA is converted into smaller pieces of DNA.

As used herein, "large template DNA" would be DNA of more than 5 kb, 10 kb, or 25 kb, preferably more than 500 kb, more preferably more than 1 MB, and most preferably 5 MB or larger.

As used herein, the term "stringent hybridization conditions" refers to those conditions under which only fully complimentary sequences will hybridize to each other.

The following discussion summarizes the basic steps involved in the methods of the invention. The steps are recited in a specific order, however, as would be known by one of skill in the art, the order of the steps may be manipulated to achieve the same result. Such

IDT Exhibit 1020.9879

manipulations are contemplated by the inventors. Further, some steps may be minimized as would also be known by one of skill in the art.

Fragmentation

In the practice of the methods of the present invention, the fragmentation of the DNA sample can be done by any means known to those of ordinary skill in the art. Preferably, the fragmenting is performed by enzymatic, chemical, or mechanical means. The mechanical means may include sonication, French press, HPLC, HydroShear (GeneMachines, San Carlos, CA), and nebulization. The enzymatic means may be performed by digestion with Deoxyribonuclease I (DNase I), nonspecific nucleases, or single or multiple restriction endonucleases. In a preferred embodiment, the fragmentation results in ends for which the sequence adjacent to the end is not known. The sequence adjacent to the end may be at least 5 bases, 10 bases, 20 bases, 30 bases, or 50 bases.

Enzymatic Fragmentation

In a preferred embodiment, the enzymatic means is DNase I. DNase I is a versatile enzyme that nonspecifically cleaves double-stranded DNA (dsDNA) to release 5'-phosphorylated oligonucleotide products. DNase I has optimal activity in buffers containing Mn^{2+}, Mg^{2+} and Ca^{2+}. The purpose of the DNase I digestion step is to fragment a large DNA genome into smaller species comprising a library. The cleavage characteristics of DNase I will result in random digestion of template DNA (*i.e.*, minimal sequence bias) and in the predominance of blunt-ended dsDNA fragments when used in the presence of manganese-based buffers (Melgar, E. and D.A. Goldthwait. 1968. Deoxyribonucleic acid nucleases. II. The effects of metal on the mechanism of action of Deoxyribonuclease I. J. Biol. Chem. 243: 4409). The range of digestion products generated following DNase I treatment of genomic templates is dependent on three factors: i) amount of enzyme used (units); ii) temperature of digestion (°C); and iii) incubation time (minutes). The DNase I digestion conditions outlined below have been optimized to yield genomic libraries with a size range from 50-700 base pairs (bp).

In a preferred embodiment, DNase I is used to digest large template DNA or whole genome DNA for 1-2 minutes to generate a population of oligonucleotides that range from 50 to 500 bp, or 50 to 700 bp. In another preferred embodiment, the DNase I digestion is performed at a temperature of 10°C-37°C. In yet another preferred embodiment, the digested

IDT Exhibit 1020.9880

DNA fragments are 50 bp to 700 bp in length.

Mechanical Fragmentation

Another preferred method for nucleic acid fragmentation is mechanical fragmentation. Mechanical fragmentation methods include sonication and nebulization, and use of HydroShear, HPLC, and French Press devices. Sonication may be performed by a tube containing DNA in a suitable buffer (i.e., 10 mM Tris, 0.1 mM EDTA) and sonicating for a varying number of 10 second bursts using maximum output and continuous power Sonicators are commercially available from, e.g., Misonix Inc. (Farmingdale, NY), and can be used essentially as described by Bankier and Barrell (Bankier, A.T., Weston, K.M., and Barrell, B.G., "Random cloning and sequencing by the M13/dideoxynucleotide chain termination method", Meth. Enzymol. 155, 51-93 (1987). For sonication, it is preferred to maintain the nucleic acid at a uniform temperature by keeping the sample on ice. Constant temperature conditions, at 0° C for example, are preferred to maintain an even fragment distribution. The optimal conditions for sonication may be determined empirically for a given DNA sample before preparative sonication is performed. For example, aliquots of DNA can be treated for different times under sonication and the size and quality of DNA can be analyzed by PAGE. Once optimal sonication conditions are determined, the remaining DNA can be sonicated according to those pre-determined conditions.

Another preferred method for nucleic acid fragmentation is treatment by nebulizers (e.g., protocols, and hardware available from GeneMachines, San Carlos, California. Also see U.S. Patents 5,506,100 and 5,610,010). In nebulization, hydrodynamic shearing forces are used to fragment DNA strands. For example, DNA in a aqueous solution can be passed through a tube with an abrupt contraction. As the solution approaches the contraction, the fluid accelerates to maintain the volumetric flow rate through the smaller area of the contraction. During this acceleration, drag forces stretch the DNA until it snaps. Optionally, the DNA solution can be passed several times (e.g., 15 to 20 cycles) through the contraction until the fragments are too short for further shearing. By adjusting the contraction and the flow rate of the fluid, the size of the final DNA fragment may be determined. Software for controlling and monitoring reaction conditions is available to allow automation of the nebulizing process. As another advantage, there are no special buffer requirements for nebulization . For example, DNA may be suspended in various solutions including, but not limited to, water, Tris buffer, Tris-EDTA buffer, and Tris-EDTA with up to 0.5 M NaCl.

IDT Exhibit 1020.9881

Polishing

Polishing digestion of genomic DNA (gDNA) templates with DNase I in the presence of Mn^{2+} produces fragments of DNA that are either blunt-ended or have protruding termini with one or two nucleotides in length. Similarly, fragmentation of DNA by mechanical means provides a combination of fragments with blunt-ends or overhanging ends. These DNA fragments, whether generated enzymatically or mechanically, may be "polished" using the procedure described below.

Polishing (also called end repair) refers to the conversion of non-blunt ended DNA into blunt ended DNA. In one method, polishing may be performed by treatment with a single strand-specific exonuclease, such as BAL32 nuclease or Mung Bean nuclease. Generally, the nuclease should be calibrated prior to use.

In one embodiment, blunt ends are created with *Pfu* DNA polymerase. In other embodiments, blunt ends are created with other DNA polymerases such as T4 DNA polymerase or Klenow DNA polymerase. *Pfu* "polishing" or blunt ending can be used to increase the amount of blunt-ended species generated following genomic template digestion with DNase I. *Pfu* DNA polymerase fills in 5' overhangs. Additionally, *Pfu* DNA polymerase exhibits 3'→ 5' exonuclease activity. Thus, the enzyme can be used to remove single and double nucleotide extensions to further increase the amount of blunt-ended DNA fragments available for adaptor ligation (see, e.g., Costa, G.L. and M.P. Weiner. 1994a. Protocols for cloning and analysis of blunt-ended PCR-generated DNA fragments. *PCR Methods Appl* 3(5):S95; Costa, G.L., A. Grafsky and M.P. Weiner. 1994b. Cloning and analysis of PCR-generated DNA fragments. *PCR Methods Appl* 3(6):338; Costa, G.L. and M.P. Weiner. 1994c. Polishing with T4 or Pfu polymerase increases the efficiency of cloning of PCR products. *Nucleic Acids Res.* 22(12):2423).

Adaptor Ligation

Following fragmentation and blunt ending of the DNA library, universal adaptor sequences can be added to each DNA fragment. In various embodiments of the invention, the universal adaptors are designed to include: 1) a set of unique PCR priming regions that are typically 10-20 bp in length (any suitable size may be used) located adjacent to; 2) a set of unique sequencing priming regions that are typically 10-20 bp in length (any suitable size may be used) optionally followed by; 3) a unique discriminating key sequence (e.g., 1-12 bp

IDT Exhibit 1020.9882

in length) including a combination of at least one of each of the four deoxyribonucleotides (i.e., A, C, G, T). In a preferred embodiment, the discriminating key sequence is 4 bases in length. In another embodiment, the discriminating key sequence may be combinations of 1-4 bases. In another embodiment, the key sequence includes one of each of the four nucleotides. In certain embodiments, the key sequence includes one or more ribonucleotides, e.g., U.

In one embodiment of the invention, each unique universal adaptor is forty-four bp (44 bp) in length, although any suitable size may be used. In a preferred embodiment the universal adaptors are ligated, using T4 DNA ligase, onto each end of the DNA fragment to generate a total nucleotide addition of 88 bp to each DNA fragment, although any suitable size may be used. Different universal adaptors can be specifically designed for each DNA library preparation and therefore provide a unique identifier for each organism. For example, different library preparations can employ different key sequences. It is understood that the size and sequence of the universal adaptors may be modified as would be apparent to one of skill in the art. Thus, the adaptors for use with the invention are not limited to the size and sequence described herein.

For example, to prepare two distinct universal adaptors (i.e., "first" and "second"), single-stranded oligonucleotides may be ordered from a commercial vendor (e.g., Integrated DNA Technologies, IA or Operon Technologies, CA). In certain embodiments of the invention, all of the first adaptors for a library share one nucleotide sequence, including a PCR priming sequence, sequencing primer sequence, and discriminating key sequence, while all of the second adaptors share another nucleotide sequence. In another embodiment, the universal adaptor oligonucleotide sequences are modified during synthesis with one or more phosphorothioate linkages in place of phosphodiester linkages. For example, adaptor oligonucleotides can include two or three phosphorothioate linkages at both the 5' and 3' ends, or at one end. Unmodified oligonucleotides are typically subject to rapid degradation by contaminating nucleases that catalyze the hydrolytic cleavage of the phosphodiester linkage between nucleotide bases. One simple and widely used nuclease-resistant chemistry available for use in oligonucleotide applications is the phosphorothioate modification. In phosphorothioates, a sulfur atom replaces a non-bridging oxygen in the oligonucleotide backbone making it resistant to all forms of nuclease digestion (i.e. resistant to both endonuclease and exonuclease digestion). Each oligonucleotide is HPLC-purified to ensure there are no contaminating or spurious oligonucleotide sequences in the synthetic

IDT Exhibit 1020.9883

oligonucleotide preparation.

The universal adaptors are designed to allow directional ligation to the fragmented DNA. Each set of double-stranded universal adaptors is designed with a PCR priming region that includes non-complementary 5' four-base overhangs which are unable to ligate to each other or to the blunt-ended DNA fragment . Thus, binding occurs between the 3' end of the adaptor and the 5' end of the DNA fragment or between the 3' end of the DNA fragment and the 5' end of the adaptor. Double-stranded universal adaptor sequences are generated using single-stranded oligonucleotides that are designed with sequences that allow primarily complimentary oligonucleotides to anneal, and to prevent cross-hybridization between two non-complimentary oligonucleotides.

In one embodiment, 95% of the universal adaptors are formed from the annealing of complimentary oligonucleotides. In a preferred embodiment, 97% of the universal adaptors are formed from the annealing of complimentary oligonucleotides. In a more preferred embodiment, 99% of the universal adaptors are formed from the annealing of complimentary oligonucleotides. In a most preferred embodiment, 100% of the universal adaptors are formed from the annealing of complimentary oligonucleotides. Exemplary methods of primer design to minimize cross-hybridization between the primers and spurious target sequences are provided in Example 2.

In certain aspects of the invention, an overhanging nucleotide (e.g., T) is added to the blunt 3' ends of the first and second adaptors. In parallel, a polymerase is used to add an overhanging nucleotide (e.g., A) to the blunt 5' ends of the template DNA. The overhanging nucleotides of the adaptors and template are complementary to allow more efficient adaptor ligation.

In other aspects, a plasmid capture system is used in accordance with the disclosed methods. For example, the double stranded universal adaptors can be inserted into a plasmid. The adaptor region can include the following sequences, in order: Restriction Site(s), PCR Primer Sequence 1, Sequencing Primer Sequence 1, Key Sequence 1, Restriction Site(s), Key Sequence 2, Sequencing Primer Sequence 2, PCR Priming Sequence 2, and Restriction Site(s). In one approach, the plasmid is digested with one or more restriction enzymes that cut between Key Sequence 1 and Key Sequence 2. The fragmented template DNA is ligated between Key Sequence 1 and Key Sequence 2. The ligated construct is then digested to cut after PCR Priming Sequence 2. The digested end next to PCR Priming Sequence 2 is filled in with biotinylated nucleotides. The biotinylated construct is digested to cut prior to PCR

IDT Exhibit 1020.9884

Primer Sequence 1. The Adaptor-DNA fragment-Adaptor-biotin segment is excised and isolated, e.g., by binding to a streptavidin magnetic bead. Other embodiments of the plasmid capture system are also possible through application of known cloning techniques. These embodiments are also encompassed by the invention.

One of the two adaptors can be linked to a support binding moiety. In a preferred embodiment, a 5' biotin is added to the first universal adaptor to allow subsequent isolation of ssDNA template and noncovalent coupling of the universal adaptor to the surface of a solid support that is saturated with a biotin-binding protein (e.g., streptavidin, NeutrAvidinTM or avidin). Suitable supports include but are not limited to magnetic beads, affinity columns, membranes (e.g., PDVF membrane, nitrocellulose, etc.), which can be coated with streptavidin or another member of a binding pair. Other linkages are well known in the art and may be used in place of biotin-streptavidin. For example, antibody/antigen-epitope, receptor/ligand and oligonucleotide pairing or complimentarity linkages can be used. In one embodiment, the solid support is a bead, preferably a polystyrene bead. In one preferred embodiment, the bead has a diameter of about 2.8 μm, although any suitable size may be used. In another preferred embodiment, the bead is a paramagnetic bead (e.g., Dynal Biotech, Inc., Lake Success, NY). As used herein, this bead is referred to as a "sample prep bead".

Each universal adaptor may be prepared by combining and annealing two ssDNA oligonucleotides, one containing the sense sequence and the second containing the antisense (complementary) sequence. Schematic representation of the universal adaptor design is depicted in Figure 2.

Isolation of Ligation Products

The universal adaptor ligation results in the formation of fragmented DNAs with adaptors on each end, unbound single adaptors, and adaptor dimers. In a preferred embodiment, agarose gel electrophoresis is used as a method to separate and isolate the adapted DNA library population from the unligated single adaptors and adaptor dimer populations. In other embodiments, the fragments may be separated by size exclusion chromatography, filtration, sucrose sedimentation, or other nucleic acid separation techniques known to those skilled in the art. The procedure of DNase I digestion of DNA typically yields a library population that ranges from 50-700 bp. In a preferred embodiment, upon conducting agarose gel electrophoresis in the presence of a DNA marker, the addition of the

IDT Exhibit 1020.9885

88 bp universal adaptor set will shift the DNA library population to a larger size and will result in a migration profile in the size range of approximately 130-800 bp; adaptor dimers will migrate at 88 bp; and adaptors not ligated will migrate at 44 bp. Therefore, numerous double-stranded DNA libraries in sizes ranging from 200-800 bp can be physically isolated from the agarose gel and purified using standard gel extraction techniques. In one embodiment, gel isolation of the adapted ligated DNA library will result in the recovery of a library population ranging in size from 200-500 bp. Other methods of distinguishing adaptor-ligated fragments are known to one of skill in the art.

Nick Repair

Because the DNA oligonucleotides used for the universal adaptors are not 5' phosphorylated, gaps will be present at the 3' junctions of the fragmented DNAs following ligase treatment (see Figure 3A). These "gaps" or "nicks" can be filled in by using a DNA polymerase enzyme that can bind to, strand displace, and extend the nicked DNA fragments. DNA polymerases that lack 3'→ 5' exonuclease activity but exhibit 5' → 3' exonuclease activity have the ability to recognize nicks, displace the nicked strands, and extend the strand in a manner that results in the repair of the nicks and in the formation of non-nicked double-stranded DNA (see Figure 3B and 3C) (Hamilton, S.C., J.W. Farchaus and M.C. Davis. 2001. DNA polymerases as engines for biotechnology. *BioTechniques* 31:370).

Several modifying enzymes are utilized for the nick repair step, including but not limited to polymerases, ligases, and kinases. DNA polymerases that can be used in the methods of the invention include, for example, *E. coli* DNA polymerase I, *Thermoanaerobacter thermohydrosulfuricus* polymerase I, and bacteriophage phi 29. In a preferred embodiment, the strand displacing enzyme *Bacillus stearothermophilus* polymerase I (*Bst* DNA polymerase I) is used to repair the nicked dsDNA and results in non-nicked dsDNA (see Figure 3D). In another preferred embodiment, the ligase is T4 DNA ligase and the kinase is T4 polynucleotide kinase.

Isolation of Single-Stranded DNA

Following the generation of non-nicked dsDNA, ssDNAs comprising both the first and second adaptor molecules may be isolated. Double-stranded DNA libraries will have adaptors bound in the following configurations:

Universal Adaptor A – DNA fragment – Universal Adaptor A

IDT Exhibit 1020.9886

Universal Adaptor B – DNA fragment – Universal Adaptor A*

Universal Adaptor A – DNA fragment – Universal Adaptor B*

Universal Adaptor B – DNA fragment – Universal Adaptor B

"A" and "B" correspond to the first and second adaptors. The desired populations are designated with asterisks.

Preferably, the universal adaptors are designed such that only one universal adaptor has a 5' biotin moiety. For example, if universal adaptor B has a 5'biotin moiety, streptavidin-coated sample prep beads can be used to bind all double-stranded DNA library species with universal adaptor B. Genomic library populations that contain two universal adaptor A species will not contain a 5' biotin moiety and will not bind to streptavidin-containing sample prep beads and thus can be washed away. The only species that will remain attached to beads are those with universal adaptors A and B and those with two universal adaptor B sequences.

DNA species with two universal adaptor B sequences (i.e., biotin moieties at each 5'end) will be bound to streptavidin-coated sample prep beads at each end, as each strand comprised in the double strand will be bound. Double-stranded DNA species with a universal adaptor A and a universal adaptor B will contain a single 5'biotin moiety and thus will be bound to streptavidin-coated beads at only one end. Where the sample prep beads are magnetic, the beads will remain coupled to a solid support when magnetized. Accordingly, in the presence of a low-salt ("melt" or denaturing) solution, only those DNA fragments that contain a single universal adaptor A and a single universal adaptor B sequence will release the complementary unbound strand. This single-stranded DNA population attached to beads may be collected and quantitated by, for example, pyrophosphate sequencing, real-time quantitative PCR, agarose gel electrophoresis, fluorescent dye binding assay (PicoGreen®; Molecular Probes, Inc., Eugene, OR), or capillary gel electrophoresis.

In one embodiment, ssDNA libraries that are created according to the methods of the invention are quantitated to calculate the number of molecules per unit volume. For example, the molecules can be annealed to a solid support that includes oligonucleotide capture primers that are complementary to the PCR priming regions of the universal adaptor ends of the ssDNA species.

In certain embodiments, beads comprising capture primers annealed to ssDNA library molecules can be transferred to a thermocycler to allow PCR amplification. Clonal populations of single species of single stranded DNA captured on DNA beads may then

IDT Exhibit 1020.9887

sequenced. In one embodiment, the solid support is a bead, preferably a sepharose bead. As used herein, this bead is referred to as a "DNA capture bead".

The beads used herein may be of any convenient size and fabricated from any number of known materials. Example of such materials include: inorganics, natural polymers, and synthetic polymers. Specific examples of these materials include: cellulose, cellulose derivatives, acrylic resins, glass; silica gels, polystyrene, gelatin, polyvinyl pyrrolidone, co-polymers of vinyl and acrylamide, polystyrene cross-linked with divinylbenzene or the like (see, Merrifield Biochemistry 1964, 3, 1385-1390), polyacrylamides, latex gels, polystyrene, dextran, rubber, silicon, plastics, nitrocellulose, celluloses, natural sponges, silica gels, glass, metals plastic, cellulose, cross-linked dextrans (e.g., Sephadex^{TM}) and agarose gel (Sepharose^{TM}) and solid phase supports known to those of skill in the art. In one embodiment, the diameter of the DNA capture bead is 20-70 μm. In a preferred embodiment, the diameter of the DNA capture bead is 20-50 μm. In a more preferred embodiment, the diameter of the DNA capture bead is about 30 μm.

In one aspect, the invention includes a method for generating a library of solid supports comprising: (a) preparing a population of ssDNA templates according to the methods disclosed herein; (b) attaching each DNA template to a solid support such that there is one molecule of DNA per solid support; (c) amplifying the population of single-stranded templates such that the amplification generates a clonal population of each DNA fragment on each solid support; (d) sequencing clonal populations of the ssDNA templates.

In one embodiment, the solid support is a DNA capture bead. In another embodiment, the DNA is genomic DNA, cDNA, or reverse transcripts of RNA (e.g., viral RNA). The DNA may be attached to the solid support, for example, via a biotin-streptavidin linkage, a covalent linkage, or by complementary oligonucleotide hybridization. In one embodiment, each DNA template is ligated to a set of universal adaptors. In another embodiment, the universal adaptor pair comprises a PCR primer sequence, a sequencing primer sequence, and a discriminating key sequence. Single-stranded DNAs with unique ends are isolated and then attached to a solid support and exposed to amplification techniques for clonal amplification. The DNA may be amplified by PCR. In one aspect, the invention provides a library attached to solid supports made by the methods described herein.

The DNA prepared by this method may be used for many molecular biological procedures, such as linear extension, rolling circle amplification, PCR, and sequencing. The linkage reaction can be driven, for example, by using a high molar ratio of bead to DNA.

IDT Exhibit 1020.9888

The capture of single-stranded DNA molecules follows a Poisson distribution and results in subsets of beads having no DNA attached, one molecule of DNA attached, or more than one molecule of DNA attached. In a preferred embodiment, there is one molecule of DNA attached to each bead. In addition, it is possible to include additional modifications with the adaptors that may be useful for further manipulations of the isolated library.

Binding Nucleic Acid Template to Capture Beads

In certain embodiments of the invention, a single stranded nucleic acid template to be amplified is attached to a capture bead. The nucleic acid template may be attached to the solid support capture bead in any manner known in the art. Numerous methods exist in the art for attaching DNA to a solid support such as the preferred microscopic bead. According to the present invention, covalent chemical attachment of the DNA to the bead can be accomplished by using standard coupling agents, such as water-soluble carbodiimide, to link the 5'-phosphate on the DNA to amine-coated capture beads through a phosphoamidate bond. Another alternative is to first couple specific oligonucleotide linkers to the bead using similar chemistry, and to then use DNA ligase to link the DNA to the linker on the bead. Other linkage chemistries to join the oligonucleotide to the beads include the use of N-hydroxysuccinamide (NHS) and its derivatives. In such a method, one end of the oligonucleotide may contain a reactive group (such as an amide group) which forms a covalent bond with the solid support, while the other end of the linker contains a second reactive group that can bond with the oligonucleotide to be immobilized. In a preferred embodiment, the oligonucleotide is bound to the DNA capture bead by covalent linkage. However, non-covalent linkages, such as chelation or antigen-antibody complexes, may also be used to join the oligonucleotide to the bead.

Oligonucleotide linkers can be employed which specifically hybridize to unique sequences at the end of the DNA fragment, such as the overlapping end from a restriction enzyme site or the "sticky ends" of bacteriophage lambda based cloning vectors, but blunt-end ligations can also be used beneficially. These methods are described in detail in US 5,674,743. It is preferred that any method used to immobilize the beads will continue to bind the immobilized oligonucleotide throughout the steps in the methods of the invention.

In one embodiment, each capture bead is designed to have a plurality of nucleic acid primers that recognize (i.e., are complementary to) a portion of the nucleic template, and the nucleic acid template is thus hybridized to the capture bead. In the methods described herein,

IDT Exhibit 1020.9889

clonal amplification of the template species is desired, so it is preferred that only one unique nucleic acid template is attached to any one capture bead.

The beads used herein may be of any convenient size and fabricated from any number of known materials. Example of such materials include: inorganics, natural polymers, and synthetic polymers. Specific examples of these materials include: cellulose, cellulose derivatives, acrylic resins, glass, silica gels, polystyrene, gelatin, polyvinyl pyrrolidone, co-polymers of vinyl and acrylamide, polystyrene cross-linked with divinylbenzene or the like (as described, *e.g.,* in Merrifield, Biochemistry 1964, 3, 1385-1390), polyacrylamides, latex gels, polystyrene, dextran, rubber, silicon, plastics, nitrocellulose, natural sponges, silica gels, control pore glass, metals, cross-linked dextrans (e.g., Sephadex™) agarose gel (Sepharose™), and solid phase supports known to those of skill in the art. In a preferred embodiment, the capture beads are Sepharose beads approximately 25 to 40 μm in diameter.

Emulsification

For use with the present invention, capture beads with or without attached nucleic acid template may be suspended in a heat stable water-in-oil emulsion. It is contemplated that a plurality of the microreactors include only one template and one bead. There may be many droplets that do not contain a template or which do not contain a bead. Likewise there may be droplets that contain more than one copy of a template. The emulsion may be formed according to any suitable method known in the art. One method of creating emulsion is described below but any method for making an emulsion may be used. These methods are known in the art and include adjuvant methods, counter-flow methods, cross-current methods, rotating drum methods, and membrane methods. Furthermore, the size of the microcapsules may be adjusted by varying the flow rate and speed of the components. For example, in dropwise addition, the size of the drops and the total time of delivery may be varied. Preferably, the emulsion contains a density of about 3,000 beads encapsulated per microliter.

Various emulsions that are suitable for biologic reactions are referred to in Griffiths and Tawfik, EMBO, 22, pp. 24-35 (2003); Ghadessy et al., Proc. Natl. Acad. Sci. USA 98, pp. 4552-4557 (2001); United States Patent No. 6,489,103 and WO 02/22869, each fully incorporated herein by reference. It is noted that Griffiths et al., (U.S. Pat. No. 6,489,103 and WO 99/02671) refers to a method for *in vitro* sorting of one or more genetic elements encoding a gene products having a desired activity. This method involves

IDT Exhibit 1020.9890

compartmentalizing a gene, expressing the gene, and sorting the compartmentalized gene based on the expressed product. In contrast to the present invention, the microencapsulated sorting method of Griffith is not suitable for parallel analysis of multiple microcapsules because their nucleic acid product is not anchored and cannot be anchored. Since the nucleic acids of Griffiths are not anchored, they would be mixed together during demulsification.

The emulsion is preferably generated by adding beads to an amplification solution. As used herein, the term "amplification solution" means the sufficient mixture of reagents that is necessary to perform amplification of template DNA. One example of an amplification solution, a PCR amplification solution, is provided in the Examples below . It will be appreciated that various modifications may be made to the amplification solution based on the type of amplification being performed and whether the template DNA is attached to the beads or provided in solution. In one embodiment, the mixture of beads and amplification solution is added dropwise into a spinning mixture of biocompatible oil (e.g., light mineral oil, Sigma) and allowed to emulsify. In another embodiment, the beads and amplification solution are added dropwise into a cross-flow of biocompatible oil. The oil used may be supplemented with one or more biocompatible emulsion stabilizers. These emulsion stabilizers may include Atlox 4912, Span 80, and other recognized and commercially available suitable stabilizers. In preferred aspects, the emulsion is heat stable to allow thermal cycling, e.g., to at least 94°C, at least 95°C, or at least 96°C. Preferably, the droplets formed range in size from about 5 microns to about 500 microns, more preferably from about 10 microns to about 350 microns, even more preferably from about 50 to 250 microns, and most preferably from about 100 microns to about 200 microns. Advantageously, cross-flow fluid mixing allows for control of the droplet formation, and uniformity of droplet size. We note that smaller water droplets not containing beads may be present in the emulsion.

The microreactors should be sufficiently large to encompass sufficient amplification reagents for the degree of amplification required. However, the microreactors should be sufficiently small so that a population of microreactors, each containing a member of a DNA library, can be amplified by conventional laboratory equipment, e.g., PCR thermocycling equipment, test tubes, incubators and the like. Notably, the use of microreactors allows amplification of complex mixtures of templates (e.g., genomic DNA samples or whole cell RNA) without intermixing of sequences, or domination by one or more templates (e.g., PCR selection bias; see, Wagner et al., 1994, Suzuki and Giovannoni, 1996; Chandler et al., 1997,

IDT Exhibit 1020.9891

Polz and Cavanaugh, 1998).

With the limitations described above, the optimal size of a microreactor may be on average 100 to 200 microns in diameter. Microreactors of this size would allow amplification of a DNA library comprising about 600,000 members in a suspension of microreactors of less than 10 ml in volume. For example, if PCR is the chosen amplification method, 10 ml of microreactors would fit into 96 tubes of a regular thermocycler with 96 tube capacity. In a preferred embodiment, the suspension of 600,000 microreactors would have a volume of less than 1 ml. A suspension of less than 1 ml may be amplified in about 10 tubes of a conventional PCR thermocycler. In a most preferred embodiment, the suspension of 600,000 microreactors would have a volume of less than 0.5 ml.

Another embodiment of the invention is directed to a method of performing nucleic acid amplification with a template and a bead, but without attachment of the template to the bead. In one aspect, the bead may comprise a linker molecule that can bind the amplified nucleic acid after amplification. For example, the linker may be a linker that can be activated. Such linkers are well known and include temperature sensitive or salt sensitive binding pairs such as streptavidin/biotin and antibodies/antigen. The template nucleic acid may be encapsulated with a bead and amplified. Following amplification, the amplified nucleic acid may be linked to the beads, e.g., by adjustments in temperature or salt concentration.

Amplification

The template nucleic acid may be amplified, while attached or unattached to beads, by any suitable method of amplification including transcription-based amplification systems (Kwoh D. et al., Proc. Natl. Acad Sci. (U.S.A.) 86:1173 (1989); Gingeras T. R. et al., WO 88/10315; Davey, C. et al., EP Publication No. 329,822; Miller, H. I. et al., WO 89/06700), "RACE" (Frohman, M. A., In: PCR Protocols: A Guide to Methods and Applications, Academic Press, NY (1990)) and one-sided PCR (Ohara, O. et al., Proc. Natl. Acad. Sci. (U.S.A.) 86.5673-5677 (1989)). Still other methods such as di-oligonucleotide amplification, isothermal amplification (Walker, G. T. et al., Proc. Natl. Acad. Sci. (U.S.A.) 89:392-396 (1992)), Nucleic Acid Sequence Based Amplification (NASBA; see, e.g., Deiman B et al., 2002, Mol Biotechnol. 20(2):163-79), whole-genome amplification (see, e.g., Hawkins TL et al., 2002, Curr Opin Biotechnol. 13(1):65-7), strand-displacement amplification (see, e.g., Andras SC, 2001, Mol Biotechnol. 19(1):29-44), rolling circle amplification (reviewed in

IDT Exhibit 1020.9892

U.S. Pat. No. 5,714,320), and other well known techniques may be used in accordance with the present invention. In certain aspects, a nucleic acid template is amplified after encapsulation with a bead in a microreactor. Alternatively, a nucleic acid template is amplified after distribution onto a multiwell surface, e.g., a PicoTiter plate.

In a preferred embodiment, DNA amplification is performed by PCR. PCR according to the present invention may be performed by encapsulating the target nucleic acid with a PCR solution comprising all the necessary reagents for PCR. Then, PCR may be accomplished by exposing the emulsion to any suitable thermocycling regimen known in the art. In a preferred embodiment, 30 to 50 cycles, preferably about 40 cycles, of amplification are performed. It is desirable, but not necessary, that following the amplification procedure there be one or more hybridization and extension cycles following the cycles of amplification. In a preferred embodiment, 10 to 30 cycles, preferably about 25 cycles, of hybridization and extension are performed (e.g., as described in the examples). Routinely, the template DNA is amplified until typically at least 10,000 to 50,000,000 copies are immobilized on each bead. It is recognized that for nucleic acid detection applications, fewer copies of template are required. For nucleic acid sequencing applications we prefer that at least two million to fifty million copies, preferably about ten million to thirty million copies of the template DNA are immobilized on each bead. The skilled artisan will recognize that the size of bead (and capture site thereon) determines how many captive primers can be bound (and thus how many amplified templates may be captured onto each bead).

In one aspect, the invention encompasses a method for clonally isolating a library comprising a plurality of single stranded DNA molecules comprising: a) fragmenting large template DNA molecules to generate a plurality of fragmented DNA molecules; b) attaching a first or second universal double stranded adaptor to a first end of each fragmented DNA molecule and a first or second universal adaptor to a second end of each fragmented DNA molecule to form a mixture of adaptor ligated DNA molecules; c) isolating a plurality of single stranded DNA molecules each comprising a first single stranded universal adaptor and a second single stranded universal adaptor to obtain a library; and d) delivering the single stranded DNA molecules into reactors such that a plurality of the reactors include one DNA molecule, thereby clonally isolating the library.

In another aspect, the invention encompasses a method for generating a library comprising a plurality of single stranded DNA molecules, comprising: a) fragmenting large

IDT Exhibit 1020.9893

or whole genomic template DNA molecules to generate a plurality of fragmented DNA molecules; b) ligating a first universal double stranded adaptor or a second universal adaptor to a first end of each fragmented DNA molecule and a first universal adaptor or second universal adaptor to a second end of each fragmented DNA molecule to produce a mixture of adaptor ligated DNA molecules, wherein the first universal adaptor contains a moiety that binds to a solid support; c) attaching to a solid support those DNA molecules comprising a first double stranded universal adaptor; d) washing away adaptor ligated DNA molecules which have not attached to the solid support; e) strand separating those adaptor ligated DNA molecules that are attached to the solid support to release a plurality of single stranded DNA molecules comprising a first single stranded universal adaptor at one end and a second single stranded adaptor at the other end; and f) isolating the single stranded DNA molecules, thereby generating a library. In specific aspects, (f) can be accomplished by: i) delivering the single stranded DNA molecules onto a location on a reactor array; or ii) delivering the single stranded DNA molecules into droplets in a water-in-oil emulsion.

According to these methods, the first universal double stranded adaptor or the second universal adaptor can be attached to the fragmented DNA molecules by ligation. For example, DNA ligase may be used. These methods can further include a step of repairing single stranded nicks in the mixture of adaptor ligated DNA molecules using DNA repair and modifying enzymes, such as a polymerase, ligase, kinase, or combinations thereof. As particular examples, the enzymes can include *Bacillus stearothermophilus* polymerase I, T4 ligase, and T4 polynucleotide kinase. The template DNA for these methods can comprise genomic DNA, cDNA, plasmid DNA, cosmid DNA, artificial chromosome DNA, synthetic DNA, phasemid DNA, phagemid DNA, or reverse transcripts. Fragmenting may be performed by enzymatic, chemical, or mechanical means. For example, DNase I enzyme can be used in a digestion performed at a temperature of 10-37°C for 1-2 minutes. Alternatively, a restriction enzyme may be used. The mechanical means can be a nebulizer, French Press, sonicator, or a HydroShear.

For use with these methods, the fragmented DNA molecules can be 50 bp to 700 bp in length. The compatible ends can be blunt ends, or the compatible ends can include an A or T overhang. Blunt ends can be created with an enzyme such as *Pfu* polymerase, T4 DNA polymerase, and Klenow fragments. The first or second double stranded universal adaptor may comprise one or more phosphorothioate linkages, and may be attached to a biotin moiety. In addition, the first double stranded universal adaptors or the second double

IDT Exhibit 1020.9894

stranded universal adaptors or both double stranded universal adaptors may comprise a discriminating key sequence. For example, the discriminating key sequence is 3-12 nucleotides in length and comprise at least one nucleotide selected from the group consisting of A, G, C, U, and T. The first and second double stranded universal adaptor may comprise a PCR priming sequence and a sequencing primer sequence. In various aspects, the PCR priming sequence is 10-20 base pairs in length, and the sequencing primer sequence is 10-20 base pairs in length. In addition, the PCR priming sequence and the sequencing primer sequence may overlap.

In accordance with these methods, the mixture of adaptor ligated DNA molecules is separated by a method selected from the group consisting of gel electrophoresis, filtration, size exclusion chromatography, and sucrose sedimentation. The plurality of single stranded DNA molecules may be obtained by a treatment selected from the group consisting of low salt treatment, high pH treatment, and chemical denaturation treatment. In further aspects, the plurality of single stranded DNA molecules may be attached to a DNA capture bead. In addition, the DNA capture bead may comprise a component of a binding pair, such as avidin/biotin, ligand/receptor, antigen/antibody or complementary nucleotides. Preferably, the DNA capture bead is a paramagnetic bead.

The invention also encompasses a method for generating a single stranded DNA library attached to solid supports comprising: a) generating a plurality of single stranded DNA templates; b) attaching each of the plurality of ssDNA templates to a solid support; and c) isolating the solid supports on which the single stranded DNA templates are attached.

The invention further encompasses a method for generating a single stranded DNA library attached to solid supports comprising: a) fragmenting large template DNA molecules to generate a plurality of fragmented DNA molecules; b) attaching a first or second universal double stranded adaptor to a first end of each fragmented DNA molecule and a first or second universal adaptor to a second end of each fragmented DNA molecule to make a mixture of adaptor ligated DNA molecules; c) isolating those single stranded DNA molecules which comprise a first single stranded universal adaptor and a second single stranded universal adaptor; and d) attaching the isolated single stranded molecules from (c) to a solid support.

For use with this method, the solid support may be a DNA capture bead, and the DNA may be genomic DNA, cDNA, plasmid DNA, cosmid DNA, artificial chromosome DNA, synthetic DNA, phasemid DNA, or phagemid DNA. In certain aspects the DNA is attached to the solid support via a binding pair such as avidin/biotin, ligand/receptor, antigen/antibody

IDT Exhibit 1020.9895

and complementary nucleotides. Also encompassed is a library of mobile solid supports made by this method.

Additionally, the invention comprises a nucleic acid molecule comprising a first adaptor, a fragment of template DNA, and a second adaptor, wherein the first adaptor and second adaptor each comprise a sequencing primer, a PCR primer, and a discriminating key sequence, and wherein the first adaptor and second adaptor, when dissociated, do not cross-hybridize to each other under stringent hybridization conditions. In this method, the PCR primer may be 10-20 base pairs in length, the sequencing primer may be 10-20 base pairs in length, and the discriminating key sequence may be 3 to 12 base pairs in length. As examples, the template DNA can be genomic DNA, cDNA plasmid DNA, cosmid DNA, artificial chromosome DNA, synthetic DNA, phasemid DNA, or phagemid DNA. Preferably, the nucleic acid molecule, when dissociated, has minimal cross-hybridization to dissociated template DNA.

Also encompassed by the invention is a method for preparing single stranded DNA molecules, comprising: a) fragmenting large or whole genomic template DNA molecules to generate a plurality of fragmented DNA molecules; b) ligating a first universal double stranded adaptor or a second universal adaptor to a first end of each fragmented DNA molecule and a first universal adaptor or second universal adaptor to a second end of each fragmented DNA molecule to produce a mixture of adaptor ligated DNA molecules; c) attaching adaptor ligated DNA molecules comprising a first double stranded universal adaptor and a second double stranded adaptor to a solid support via one strand of the first double stranded universal adaptor; d) washing away adaptor ligated DNA molecules which have not attached to the solid support; e) strand separating those adaptor ligated DNA molecules that are attached to the solid support to release a plurality of single stranded DNA molecules comprising a first single stranded universal adaptor at one end and a second single stranded adaptor at the other end; and; and f) isolating the single stranded DNA molecules.

Further encompassed by the invention is a method for delivering nucleic acid templates to a plurality of reaction centers comprising the steps of: a) providing a population of nucleic acid templates; b) isolating each nucleic acid template from the population to a sequestering agent to form a population of sequestered nucleic acid templates; and c) delivering the population of sequestered nucleic acid templates to the plurality of reaction centers wherein each reaction center receives one sequestered nucleic acid. For this method, the isolating step may comprise attaching the nucleic acid templates to a bead, or

IDT Exhibit 1020.9896

encapsulating the nucleic acid template in an emulsion of a water-in-oil emulsion. The nucleic acid template may be encapsulated with a bead that can bind the nucleic acid. In this method, the delivering step may comprise delivering the sequestered nucleic acid to a plurality of reaction centers, wherein each reaction center is a well on a picotiter plate. The method can further comprising the step of attaching the isolated single stranded molecules each individually to a solid support.

Other features of the invention will become apparent in the course of the following description of exemplary embodiments which are given for illustration of the invention and are not intended to be limiting thereof. Throughout this specification, various patents, published patent applications and scientific references are cited to describe the state and content of the art. Those disclosures, in their entireties, are hereby incorporated into the present specification by reference.

EXAMPLES

Example 1: Sample Preparation

DNA Sample:

The DNA should be of high quality and free from contaminants such as proteins, nucleases, lipids, and other chemicals (such as residual EDTA from preparation) and salts. It is preferred that genomic DNA should have a 260/280 ratio of 1.8 or higher. If it is desired to sequence the genome of only one organism, then the DNA should be quality checked to ensure that there is no contaminating DNA. For example: a preparation of human DNA may be checked by PCR to ensure that it is not contaminated by bacterial DNA molecules. Another method of checking for contamination is by restriction digestion patterns and especially restriction digestion followed by Southern Blot using suitable probes known to be specific for an organism (e.g., human or mouse) and a second probe known to be specific for a possible contaminating organism (e.g., *E. coli*). If it is desired, the DNA should originate from a single clone of the organism (e.g., a colony if from bacteria).

Step 1: DNase I Digestion

The purpose of the DNase I digestion step is to fragment a large stretch of DNA such as a whole genome or a large portion of a genome into smaller species. This population of smaller-sized DNA species generated from a single DNA template is referred to as a

IDT Exhibit 1020.9897

"library". Deoxyribonuclease I (DNase I) is an endonuclease which cleaves double-stranded template DNA. The cleavage characteristics of DNase I allow random digestion of template DNA (i.e., minimal sequence bias) and will result in the predominance of blunt-ended, double-stranded DNA fragments when used in the presence of manganese-based buffers (Melgar and Goldthwait 1968). The digestion of genomic templates by DNase I is dependent on three factors: i) quantity of enzyme used (units); ii) temperature of digestion (°C); and iii) incubation time (minutes). The DNase I digestion conditions outlined below were optimized to yield DNA libraries in a size range from 50-700 base pairs (bp).

1. DNA was obtained and prepared to a concentration of 0.3 mg/ml in Tris-HCl (10mM, pH 7-8). A total of 134 µl of DNA (15 µg) was needed for this preparation. It is recommended to not use DNA preparations diluted with buffers containing EDTA (i.e., TE, Tris/EDTA). The presence of EDTA is inhibitory to enzyme digestion with DNase I. If the DNA preparation contains EDTA, it is important that the DNA be "salted" out of solution and reconstituted with the appropriate Tris-HCl buffer (10 mM, pH 7-8) or nanopure H_2O (pH 7-8).

2. In a 0.2 ml tube, DNase I Buffer, comprising 50 µl Tris pH 7.5 (1M), 10 µl $MnCl_2$ (1M), 1 µl BSA (100 mg/ml), and 39 µl water was prepared.

3. In a separate 0.2 ml tube, 15 µl of DNase I Buffer and 1.5 µl of DNase I (1U/ml) was added. The reaction tube was placed in a thermal cycler set to 15°C.

4. The 134 µl of DNA (0.3 mg/ml) was added to the DNase I reaction tube placed in the thermal cycler set at 15°C. The lid was closed and the sample was incubated for exactly 1 minute. Following incubation, 50 µl of 50 mM EDTA was added to stop the enzyme digestion.

5. The digested DNA was purified by using the QiaQuick PCR purification kit. The digestion reaction was then split into four aliquots, and four spin columns were used to purify each aliquot (37.5 µl per spin column). Each column was eluted with 30 µl elution buffer (EB) according to the manufacturer's protocol. The eluates were then combined to generate a final reaction volume of 120 µl.

6. One 3 µl aliquot of the digestion reaction was saved for analysis using a BioAnalzyer DNA 1000 LabChip.

Step 2: *Pfu* Polishing

Digestion of DNA templates with DNase I yields fragments of DNA that are primarily blunt-ended, however, some fragments will have ends that contain protruding

IDT Exhibit 1020.9898

termini that are one or two nucleotides in length. *Pfu* polishing is used to increase the amount of blunt-ended species by fill-in (i.e., "blunting") of 5' overhangs. Additionally, *Pfu* DNA polymerase has 3'→ 5' exonuclease activity that will result in the removal of single and double nucleotide extensions. *Pfu* polishing increases the amount of blunt-ended DNA fragments available for adaptor ligation (Costa 1994a, 1994b, 1994c). The following *Pfu* polishing protocol was used.

1. In a 0.2 ml tube, 115 µl purified, DNase I-digested DNA fragments, 15 µl 10X Cloned *Pfu* buffer, 5 µl dNTPs (10 mM), and 15 µl cloned *Pfu* DNA polymerase (2.5 U/µl) were added in order.

2. The polishing reaction components were mixed well and incubated at 72° C for 30 minutes.

3. Following incubation, the reaction tube was removed and placed on ice for 2 minutes.

4. The polishing reaction mixture was then split into four aliquots and purified using QiaQuick PCR purification columns (37.5 µL on each column). Each column was eluted with 30 µl buffer EB according to the manufacturer's protocol. The eluates were then combined to generate a final reaction volume of 120 µL.

5. One 3 µl aliquot of the final polishing reaction was saved for analysis using a BioAnalzyer DNA 1000 LabChip.

Step 3: Ligation of Universal Adaptors to Fragmented DNA Library

Following fragmentation and polishing of the genomic DNA library, primer sequences are added to the ends of each DNA fragment. These primer sequences are termed "Universal Adaptors" and are comprised of double-stranded oligonucleotides that contain specific priming regions that afford both PCR amplification and nucleotide sequencing. The Universal Adaptors are designed to include a set of unique PCR priming regions that are 20 base pairs in length located adjacent to a set of unique sequencing priming regions that are 20 base pairs in length, followed by a unique 4-base "key" consisting of one of each deoxyribonucleotide (i.e., A, C, G, T). Each unique Universal Adaptor (termed "Universal Adaptor A" and "Universal Adaptor B") is forty-four base pairs (44 bp) in length. Universal Adaptors are ligated, using T4 DNA ligase, onto each end of the DNA fragment to generate a total nucleotide addition of 88 bp to each DNA fragment. Different Universal Adaptors are designed specifically for each genomic DNA library preparation and will therefore provide a

IDT Exhibit 1020.9899

unique identifier for each organism.

To prepare a pair of Universal Adaptors, single-stranded oligonucleotides are designed in-house and are manufactured through a commercial vendor. Universal Adaptor DNA oligonucleotides are designed with two phosphorothioate linkages at each oligonucleotide end that serve to protect against nuclease activity (Samini, T.D., B. Jolles, and A. Laigle. 2001. Best minimally modified antisense oligonucleotides according to cell nuclease activity. *Antisense Nucleic Acid Drug Dev.* 11(3):129., the disclosure of which is incorporated *in toto* herein by reference.). Each oligonucleotide is HPLC-purified to ensure there are no contaminating or spurious DNA oligonucleotide sequences in the final prep.

The Universal Adaptors are designed to allow directional ligation to the blunt-ended, fragmented genomic DNA. For each Universal Adaptor pair, the PCR priming region contains a 5' four-base overhang and a blunt-ended 3' Key region. Directionality is achieved as the blunt-end side of the Universal Adaptor ligates to the blunt-ended DNA fragment while the 5' overhang of the adaptor cannot ligate to the blunt-ended DNA fragment. Additionally, a 5' biotin is added to the Universal Adaptor B to allow subsequent isolation of ssDNA template (Step 8). Each Universal Adaptor is prepared by annealing, in a single tube, the two single-stranded complementary DNA oligonucleotides (i.e., one oligo containing the sense sequence and the second oligo containing the antisense sequence). The following ligation protocol was used.

1. In a 0.2 ml tube, 39 µl nH_2O (molecular biology grade water), 25 µl digested, polished DNA Library, 100 µl 2X Quick Ligase Reaction Buffer, 20 µl MMP1 (10 pm/µl) adaptor set, 100:1 ratio, and 16 µl Quick Ligase were added in order. The ligation reaction was mixed well and incubated at RT for 20 minutes.

2. The ligation reaction was then removed and a 10-µl aliquot of the ligation reaction was purified for use on the BioAnalyzer. A single spin column from the Qiagen Min-Elute kit was used. The column was eluted with 10 µl EB according to the procedure per manufacturers' protocol. A 1-µl aliquot of the purified ligation reaction was loaded using a BioAnalyzer DNA 1000 LabChip. This purification step is recommended as the unpurified ligation reaction contains high amounts of salt and PEG that will inhibit the sample from running properly on the BioAnalyzer.

3. The remainder of the ligation reaction (190 µL) was used for gel isolation in Step 4.

IDT Exhibit 1020.9900

Step 3a: Microcon Filtration and Adaptor Construction. Total preparation time was approximately 25 min.

The Universal Adaptor ligation reaction requires a 100-fold excess of adaptors. To aid in the removal of these excess adaptors, the double-stranded gDNA library is filtered through a Microcon YM-100 filter device. Microcon YM-100 membranes can be used to remove double stranded DNA smaller than 125 bp. Therefore, unbound adaptors (44 bp), as well as adaptor dimers (88 bp) can be removed from the ligated gDNA library population. The following filtration protocol was used:

1. The 190 µL of the ligation reaction from Step 4 was applied into an assembled Microcon YM-100 device.
2. The device was placed in a centrifuge and spun at 5000 x g for approximately 6 minutes, or until membrane was almost dry.
3. To wash, 200 µl of 1X TE was added.
4. Sample was spun at 5000 x g for an additional 9 minutes, or until membrane was almost dry.
5. To recover, the reservoir was inserted into a new vial and spun at 3000 x g for 3 minutes. The reservoir was discarded. The recovered volume was approximately 10 µl. Next, 80 µl TE was added.

The Adaptors (A and B) were HPLC-purified and modified with phosphorothioate linkages prior to use. For Adaptor "A" (10 µM), 10 µl of 100 µM Adaptor A (44 bp, sense) was mixed with 10 µl of 100 µM Adaptor A (40 bp, antisense), and 30 µl of 1X Annealing Buffer (V_f = 50 µl) were mixed. The primers were annealed using the ANNEAL program on the Sample Prep Labthermal cycler (see below). For Adaptor "B" (10 µM), 10 µl of 100 µM Adaptor B (40 bp, sense) was mixed with 10 µl of 100 µM Adaptor B (44 bp, antisense), and 30 µl of 1X Annealing Buffer (V_f = 50 µl). The primers were annealed using the ANNEAL program on the Sample Prep Lab thermal cycler. Adaptor sets could be stored at -20°C until use.

ANNEAL-A program for primer annealing:

1.. Incubate at 95°C, 1 min;
2.. Decrease temperature to 15°C, at 0.1°C/sec; and
3.. Hold at 15°C.

IDT Exhibit 1020.9901

There was no orientation required for the genomic DNA insert fragment and the adaptors. Fragments could be ligated at either end. Four single-stranded DNA oligonucleotides were included in the Universal Adaptor set. Each single-stranded oligonucleotide was synthesized at 1 micromole scale and HPLC-purified. Each single-stranded oligonucleotide included four phosphorothioate linkages at each end.

Step 4: Gel Electrophoresis and Extraction of Adapted DNA Library

The Universal Adaptor ligation protocol produces the following: 1) fragmented DNAs with adaptors on either end; 2) unbound single adaptors; or 3) the formation of adaptor dimers. Agarose gel electrophoresis is used as a method to separate and isolate the adapted DNA library population from the unligated, single adaptors and adaptor dimer populations. The procedure of DNase I digestion of genomic DNA yields a library population that ranges from 50-700 bp (Step 1). The addition of the 88-bp Universal Adaptor set will shift the population to a larger size and will result in a migration profile in the size range of approximately 130-800 bp. Adaptor dimers will migrate at 88 bp and adaptors unligated will migrate at 44 bp. Therefore, genomic DNA libraries in size ranges > 200 bp can be physically isolated from the agarose gel and purified using standard gel extraction techniques. Gel isolation of the adapted DNA library will result in the recovery of a library population in a size range that is ≥200 bp (size range of library can be varied depending on application). The following electrophoresis and extraction protocol was used.

1. A 2% agarose gel was prepared.

2. 10 µl of 10X Ready-Load Dye was added to the remaining 90 µl of the DNA ligation mixture.

3. The dye/ligation reaction mixture was loaded into the gel using four adjacent lanes (25 µl per lane).

4. 10 µl of the100 bp ladder (0.1 µg/µl) was loaded two lanes away from ligation reaction lanes.

5. The gel was run at 100V for 3 hours.

6. When the gel run was complete, the gel was removed from the gel box and transferred to a flat surface covered with plastic wrap. DNA bands were visualized using a hand-held long-wave UV light. Using a sterile, single-use scalpel, the fragment sizes of 200 – 400 bp were cut out from the agarose gel. Using this approach, libraries with any size range can be isolated. It is also possible to isolate more than one size range. Where the

IDT Exhibit 1020.9902

library size range is 200-900 bp, it is possible to isolate several size ranges from a single well (i.e., 200-400 bp and 500-700 bp).

7. The DNA embedded in the agarose gel was isolated using a Qiagen MinElute Gel Extraction kit following the manufacturer's instructions. Briefly, Buffer QG was added to cover the agarose in the tube. The agarose was allowed to completely dissolve. The color of the Buffer QG was maintained by adjusting the pH according to the Qiagen instructions to minimize sample loss. Two MinElute spin columns (Qiagen) were used for purification. The large volume of dissolved agarose required each column to be loaded several times. The columns were eluded with 10 µl of Buffer EB which was pre-warmed at 55°C. The eluates were pooled to produce 20 µl of gDNA library.

8. One 1 µL aliquot of each isolated DNA library was analyzed using a BioAnalyzer DNA 1000 LabChip to assess the exact distribution of the DNA library population.

Step 5: Strand Displacement and Extension of Nicked Double Stranded DNA Library

Because the DNA oligonucleotides used for the Universal Adaptors are not phosphorylated, gaps are present at the 3' junctions of the fragmented gDNAs. These two "gaps" or "nicks" can be filled in by using a strand displacing DNA polymerase. The polymerase recognizes nicks, displaces the nicked strands, and extends the strand in a manner that results in repair of nicks and in the formation of non-nicked double-stranded DNA. The strand displacing enzyme used is the large fragment of *Bst* DNA polymerase.

1. In a 0.2 ml tube, 19 µl gel-extracted DNA library, 40 µl nH_2O, 8 µl 10X ThermoPol Reaction Buffer, 8 µl BSA (1 mg/ml), 2 µl dNTPs (10 mM), and 3 µl *Bst* I Polymerase (8 U/µl) were added in order.

2. The samples were mixed well and placed in a thermal cycler and incubated using the Strand Displacement incubation program: "BST". BST program for stand displacement and extension of nicked double-stranded DNA:

1. Incubate at 65° C, 30 minutes;
2. Incubate at 80° C, 10 minutes;
3. Incubate at 58° C, 10 minutes; and
4. Hold at 14° C.

3. One 1 µL aliquot of the *Bst*-treated DNA library was run using a BioAnalyzer DNA 1000 LabChip.

IDT Exhibit 1020.9903

Step 6: Preparation of Streptavidin Beads

Following the generation of unnicked double-stranded genomic DNA, it is necessary to isolate single-stranded genomic DNAs containing flanking Universal Adaptor sequences. This step outlines the binding of biotin-tagged double-stranded DNA to streptavidin beads. For preparing streptavidin beads, the following protocol was used.

1. 100 µl Dynal M-270 Streptavidin beads were washed two times with 200 µl of 1X Binding Buffer (1 M NaCl, 0.5 mM EDTA, 5 mM Tris, pH 7.5) by applying the magnetic beads to the MPC.

2. The beads were resuspended in 100 µl 2X Binding buffer, then the remaining 79 µl of the *Bst*-treated DNA sample (from Step 5) and 20 µl water was added.

3. The bead solution was mixed well and placed on a tube rotator at RT for 20 minutes. The bead mixtures were washed, using the MPC, two times with 100 µl of 1X Binding Buffer, then washed two times with nH_2O. Binding & Washing (B&W) Buffer (2X and 1X): 2X B&W buffer was prepared by mixing 10 mM Tris•HCl (pH 7.5), 1 mM EDTA, and 2 M NaCl. The reagents were combined as listed above and mixed thoroughly. The solution can be stored at RT for 6 months; 1X B&W buffer was prepared by mixing 2X B&W buffer with nH_2O, 1:1. The final concentrations were half the above, i.e., 5 mM Tris•HCl (pH 7.5), 0.5 mM EDTA, and 1 M NaCl.

Step 7: Isolation of single-stranded DNA Library using Streptavidin Beads

Following binding of the double-stranded gDNA library to streptavidin beads, it is preferred to isolate from the ligated pool only the single-stranded gDNAs containing Universal Adaptor A and Universal Adaptor B (desired populations are designated below with asterisks). Double-stranded genomic DNA fragment pools will have adaptors bound in the following possible configurations:

Universal Adaptor A- gDNA Fragment -Universal Adaptor A

Universal Adaptor B- gDNA Fragment -Universal Adaptor A*

Universal Adaptor A- gDNA Fragment -Universal Adaptor B*

Universal Adaptor B- gDNA Fragment -Universal Adaptor B

Because only the Universal Adaptor B has a 5' biotin moiety, magnetic streptavidin-containing beads can be used to bind all gDNA library species that possess the Universal Adaptor B. Genomic library populations that contain two Universal Adaptor A species (or

IDT Exhibit 1020.9904

nonligated species) do not bind to streptavidin-containing beads and are removed during the wash procedure. The species that remain bound to bead after washing include those with Universal Adaptors A and B or those with two Universal Adaptor B ends.

Genomic DNA species with two Universal Adaptor B sequences with two biotin molecules can bind to the streptavidin-containing beads at both ends. Species with A and B adaptors having only a single biotin molecule can bind to the beads only at the "B" end. To isolate the single-stranded population, the bead-bound double-stranded DNA is treated with a sodium hydroxide solution that serves to disrupt the hydrogen bonding between the complementary DNA strands. If the DNA fragment has biotin on each end (Universal Adaptor B ends), both resulting single strands remain bound to the beads. If the fragment has only a single biotin (Universal Adaptors A and B), then the complementary strand separates from the DNA-bead complex.

The resulting single-stranded genomic DNA library is collected from the solution phase and is quantitated, e.g., using pyrophosphate sequencing (PyroSequence) or by using a RNA Pico 6000 LabChip (Agilent, Palo Alto, CA). Single-stranded genomic DNA libraries are quantitated by calculating the number of molecules per unit volume. Single-stranded gDNA molecules are then annealed (at a half copy per bead to obtain one effective copy per bead) to 25-30 μm sepharose beads containing DNA capture primers (PCR primer B). The templates are then amplified using emulsion polymerase chain reaction protocols. Subsequent sequencing may be conducted using known techniques. For isolation of the single stranded library, the following protocol was used.

1. 250 μl Melt Solution (0.125 M NaOH, 0.1 M NaCl)was added to washed beads from Step 6 above.

2. The bead solution was mixed well and the bead mixture was incubated at room temperature for 10 minutes on a tube rotator.

3. A Dynal MPC (magnetic particle concentrator) was used, the pellet beads were carefully removed, and the supernatant was set aside. The 250-μl supernatant included the single-stranded DNA library.

4. In a separate tube, 1250 μl PB (from QiaQuick Purification kit) was added and the solution was neutralized by adding 9 μl of 20% acetic acid.

5. Using a Dynal MPC, beads from the 250-μl supernatant including the single-stranded gDNA library were pelleted and the supernatant was carefully removed and transferred to the freshly prepared PB/acetic acid solution.

IDT Exhibit 1020.9905

6. The 1500 µl solution was purified using a single QiaQuick purification spin column (load sample through same column two times at 750 µl per load). The single-stranded DNA library was eluted with 50 µl EB.

Step 8a: Single-stranded gDNA Quantitation using Pyrophosphate Sequencing. Total preparation time was approximately 1 hr.

1. In a 0.2 ml tube, the following reagents were added in order:

25 µl single-stranded gDNA
1 µl MMP2B sequencing primer
14 µl Library Annealing Buffer
40 µl total

2. The DNA was allowed to anneal using the ANNEAL-S Program (see Appendix, below).

3. The samples were run on PSQ (pyrophosphate sequencing jig) to determine the number of picomoles of template in each sample (see below). Methods of sequencing can be found in U.S. Patent 6,274,320; U.S. Patent 4,863,849; U.S. Patent 6,210,891; and U.S. Patent 6,258,568, the disclosures of which are incorporated *in toto* herein by reference. Calculations were performed to determine the number of single-stranded gDNA template molecules per microliter. The remaining 25 µL of prepared single-stranded gDNA library was used for amplification and subsequent sequencing (approximately 1 x 10^6 reactions).

Step 8b: Single-stranded gDNA Quantitation using RNA Pico 6000 LabChip. Total preparation time was approximately 30 minutes.

1. The mRNA Pico assay option was selected on the BioAnalyzer (Software version 2.12).

2. An RNA Pico 6000 LabChip was prepared on the BioAnalyzer according to the manufacturers' guidelines.

3. An RNA LabChip ladder (RNA 6000 ladder) was prepared according to manufacturer's (Ambion) directions. Briefly, the RNA LabChip ladder, in solution, was heated to 70°C for 2 minutes. The solution was chilled on ice for 5 minutes to snap cool the ladder. The solution was briefly centrifuged to clear any condensate from tube walls. The RNA LabChip Ladder was stored on ice and used within one day.

4. The ssDNA library to be analyzed was run in triplicate, in adjacent lanes,

IDT Exhibit 1020.9906

using three 1 µl aliquots.

5. The BioAnalyzer software was used to calculate the concentration of each ssDNA library lane (see the Table below and Figure 8. The average of all three lanes was used to calculate the DNA concentration of the library using the procedure outlined below.

a. The peak integration lower limit line (large dash in Figure 8) was moved immediately in front of the library peak (see below).

b. The peak integration upper limit line (large dash in the Figure 8) was moved immediately after the library peak. In this way, the peak integration line connecting the lower and upper integration lines followed the slope of the background.

c. The mouse arrow was used to determine the average size of the peak in bases (usually near the peaks highest point) or a defined peak was used as chosen by the software.

d. The integrated value was used for the amount of material in the peak. The value obtained for picograms recovered was converted into molecules recovered (see Table, below). The library concentration was then determined (molecules per microliter).

Table

1	2	3	4	5	6	7	8	9
	pg/µL (1)	pg/µL (2)	pg/µL (3)	Average pg/µL	Mean Size (bp) 1	Mean Size (bp) 2	Mean Size (bp) 3	Average Size (bp)
sample	1633	1639	1645	1639	435	435	432	434

10	11	12	13	14	15
Ave MW (g/mole) Ribonucleotide	Ave MW (g/mole)	Library g/µL	moles/g	moles/µL	molecules/µL
328.2	1.42E+05	1.64E-09	7.02E-06	1.15E-14	6.93E+09

As shown in the Table above, the concentration of Library 1 was calculated as 1639 pg/µl (Column 5) and the average fragment size was 434 nucleotides (Column 9). These values were obtained from the Agilent 2100 software as described in Steps (a)-(d), above. The average molecular weight (MW) of a ribonucleotide is 328.2 g/mole (Column 10). The MW of the average library fragment (1.42×10^5 g/mole, Column 11) was calculated by multiplying the average fragment length (434) by the average ribonucleotide (328.2). The

IDT Exhibit 1020.9907

quantitated library (1639 pg/μl) was converted to grams per microliter (1.64 x 10^{-9} g/μl, Column 12). The number of moles per microliter (1.15 x 10^{-14} moles/μl, Column 14) was calculated by dividing the grams per microliter (1.64 x 10^{-9} g/μl, Column 12) by the average molecular weight of the library fragments (1.42 x 10^{5}, Column 11). Finally, the number of molecules per microliter (6.93 x 10^{9} molecules/μl, Column 15) was derived by multiplying the number of moles per microliter (1.15 x 10^{-14} moles/μl, Column 14) by Avogadro's number (6.02 x 10^{23} molecules/mole).

The final library concentration was expected to be greater than 1 x 10^{8} molecules/μl. A more important factor for library quality was adaptor dimer concentration. In Figure 8, the height of the library peak was determined approximately 10 fold greater than the adaptor dimer peak (the first peak after the marker). A library of good quality is expected to have a peak height at least 2 fold greater than the dimer peak. It should be noted that the RNA Pico 6000 LabChip provided estimates within 500% accuracy of the single-stranded gDNA concentration. Thus, it was important to perform an initial sequencing run using a titration of template to determine the number of copies per bead (cpb) of input gDNA. The recommended input DNA is 2.5 cpb, 1 cpb, 0.5 cpb, and 0.1 cpb. This titration was easily checked using the 4slot bead loading chamber on a 14 x 43 PTP.

Step 9: Dilution and Storage of Single-Stranded gDNA library

The single-stranded gDNA library was eluted and quantitated in Buffer EB. To prevent degradation, the single-stranded gDNA library was stored frozen at -20°C in the presence of EDTA. After quantitation, an equal volume of 10 mM TE was added to the library stock. All subsequent dilutions was in TE. The yield was as follows:

Remaining final volume of ssDNA library following PSQ analysis = 25 μl.

Remaining final volume of ssDNA library following LabChip analysis = 47 μl.

For the initial stock dilution, single-stranded gDNA library was diluted to 100 million molecules/μl in 1X Library-Grade Elution Buffer. Aliquots of single-stranded gDNA library were prepared for common use. For this, 200,000 molecules/μl were diluted in 1X Library-Grade Elution Buffer and 20 μl aliquots were measured. Single-use library aliquots were stored at –20°C.

Step 10: Emulsion Polymerase Chain Reaction.

Where increased numbers of cpb were preferred, bead emulsion PCR was performed as

IDT Exhibit 1020.9908

described in U.S. Patent Application Serial No. 06/476,504 filed June 6, 2003, incorporated herein by reference in its entirety.

Reagent preparation

The Stop Solution (50 mM EDTA) included 100 µl of 0.5 M EDTA mixed with 900 µl of nH_2O to obtain 1.0 ml of 50 mM EDTA solution. For 10 mM dNTPs, (10 µl dCTP (100 mM), 10 µl dATP (100 mM), 10 µl dGTP (100 mM), and 10 µl dTTP (100 mM) were mixed with 60 µl molecular biology grade water. All four 100 mM nucleotide stocks were thawed on ice. Then, 10 µl of each nucleotide was combined with 60 µl of nH_2O to a final volume of 100 µl, and mixed thoroughly. Next, 1 ml aliquots were dispensed into 1.5 ml microcentrifuge tubes. The stock solutions could be stored at -20°C for one year.

The 10 X Annealing buffer included 200 mM Tris (pH 7.5) and 50 mM magnesium acetate. For this solution, 24.23 g Tris was added to 800 ml nH_2O and the mixture was adjusted to pH 7.5. To this solution, 10.72 g of magnesium acetate was added and dissolved completely. The solution was brought up to a final volume of 1000 ml and could be stored at 4°C for 1 month. The 10 X TE included 100 mM Tris•HCl (pH 7.5) and 50 mM EDTA. These reagents were added together and mixed thoroughly. The solution could be stored at room temperature for 6 months.

Example 2: Primer Design

As discussed above, the universal adaptors are designed to include: 1) a set of unique PCR priming regions that are typically 20 bp in length (located adjacent to (2)); 2) a set of unique sequencing priming regions that are typically 20 bp in length; and 3) optionally followed by a unique discriminating key sequence consisting of at least one of each of the four deoxyribonucleotides (i.e., A, C, G, T). The probability of cross-hybridization between primers and unintended regions of the genome of interest is increased as the genome size increases and length of a perfect match with the primer decreases. However, this potential interaction with a cross-hybridizing region (CHR) is not expected to produce problems for the reasons set forth below.

In a preferred embodiment of the present invention, the single-stranded DNA library is utilized for PCR amplification and subsequent sequencing. Sequencing methodology requires random digestion of a given genome into 150 to 500 base pair fragments, after which two unique bipartite primers (composed of both a PCR and sequencing region) are

IDT Exhibit 1020.9909

ligated onto the 5' and 3' ends of the fragments (Figure 5). Unlike typical PCR amplifications where an existing section of the genome is chosen as a priming site based on melting temperature (T_m), uniqueness of the priming sequence within the genome and proximity to the particular region or gene of interest, the disclosed process utilizes synthetic priming sites that necessitates careful *de novo* primer design.

Tetramer Selection:

Strategies for *de novo* primer design are found in the published literature regarding work conducted on molecular tags for hybridization experiments (see, Hensel, M. and D.W. Holden, Molecular genetic approaches for the study of virulence in both pathogenic bacteria and fungi. Microbiology, 1996. 142(Pt 5): p. 1049-58; Shoemaker, D.D., et al., Quantitative phenotypic analysis of yeast deletion mutants using a highly parallel molecular bar-coding strategy. Nat Genet, 1996. 14(4): p. 450-6) and PCR/LDR (polymerase chain reaction/ligation detection reaction) hybridization primers (see, Gerry, N.P., et al., Universal DNA microarray method for multiplex detection of low abundance point mutations. Journal of Molecular Biology, 1999. 292: p. 251-262; Witowski, N.E., et al., Microarray-based detection of select cardiovascular disease markers. BioTechniques, 2000. 29(5): p. 936-944.).

The PCR/LDR work was particularly relevant and focused on designing oligonucleotide "zipcodes", 24 base primers comprised of six specifically designed tetramers with a similar final T_m. (see, Gerry, N.P., et al., Universal DNA microarray method for multiplex detection of low abundance point mutations. Journal of Molecular Biology, 1999. 292: p. 251-262; U.S. Pat. No. 6,506,594). Tetrameric components were chosen based on the following criteria: each tetramer differed from the others by at least two bases, tetramers that induced self-pairing or hairpin formations were excluded, and palindromic (AGCT) or repetitive tetramers (TATA) were omitted as well. Thirty-six of the 256 (4^4) possible permutations met the necessary requirements and were then subjected to further restrictions required for acceptable PCR primer design (Table 1).

Table 1

Table 1 shows a matrix demonstrating tetrameric primer component selection based on criteria outlined by Gerry et al. 1999. *J. Mol. Bio.* 292: 251-262. Each tetramer was required to differ from all others by at least two bases. The tetramers could not be palindromic or complimentary with any other tetramer. Thirty-six tetramers were selected

IDT Exhibit 1020.9910

(bold, underlined); italicized sequences signal palindromic tetramers that were excluded from consideration.

Primer Design:

The PCR primers were designed to meet specifications common to general primer design (see, Rubin, E. and A.A. Levy, A mathematical model and a computerized simulation of PCR using complex templates. Nucleic Acids Res, 1996. 24(18): p. 3538-45; Buck, G.A., et al., Design strategies and performance of custom DNA sequencing primers. Biotechniques, 1999. 27(3): p. 528-36), and the actual selection was conducted by a computer program, MMP. Primers were limited to a length of 20 bases (5 tetramers) for efficient synthesis of the total bipartite PCR/sequencing primer. Each primer contained a two base GC clamp on the 5' end, and a single GC clamp on the 3' end (Table 2), and all primers shared similar T_m (+/- 2°C) (Figure 10). No hairpinning within the primer (internal hairpin stem $\Delta G > -1.9$ kcal/mol) was permitted. Dimerization was also controlled; a 3 base maximum acceptable dimer was allowed, but it could occur in final six 3' bases, and the maximum allowable ΔG for a 3' dimer was -2.0 kcal/mol. Additionally, a penalty was applied to primers in which the 3' ends were too similar to others in the group, thus preventing cross- hybridization between one primer and the reverse complement of another.

Table 2

	1-pos	2-pos	3-pos	4-pos	5-pos
1	CCAT	TGAT	TGAT	TGAT	ATAC
2	CCTA	CTCA	CTCA	CTCA	AAAG
3	CGAA	TACA	TACA	TACA	TTAG
4	CGTT	AGCC	AGCC	AGCC	AATC
5	GCAA	GACC	GACC	GACC	TGTC
6	GCTT	TCCC	TCCC	TCCC	AGTG
7	GGAC	ATCG	ATCG	ATCG	CTTG
8	GGTA	CACG	CACG	CACG	GATG
9		TGCG	TGCG	TGCG	TCTG
10		ACCT	ACCT	ACCT	
11		GTCT	GTCT	GTCT	
12		AGGA	AGGA	AGGA	
13		TTGA	TTGA	TTGA	
14		CAGC	CAGC	CAGC	
15		GTGC	GTGC	GTGC	
16		ACGG	ACGG	ACGG	
17		CTGT	CTGT	CTGT	
18		GAGT	GAGT	GAGT	
19		TCGT	TCGT	TCGT	

Table 2 shows possible permutations of the 36 selected tetrads providing two 5' and a single 3' G/C clamp. The internal positions are composed of remaining tetrads. This results in 8 x 19 x 19 x 19 x 9 permutations, or 493,848 possible combinations. Figure 10 shows

IDT Exhibit 1020.9911

first pass, T_m based selection of acceptable primers, reducing field of 493,848 primers to 56,246 candidates with T_m of 64 to 66°C.

Table 3

. The probability of perfect sequence matches for primers increases with decreasing match length requirements and increasing size of the genome of interest.

Match Length	Perfect match probability (1/(4^length))	% chance for match in Adeno ~ 35K bases	% chance for match in NCBI bacterial database ~ 488M bases	% chance for match in Human ~3B bases
20	9.1E-13	0.00%	0.04%	0.27%
19	7.3E-12	0.00%	0.65%	4.32%
18	4.4E-11	0.00%	5.76%	34.37%
17	2.3E-10	0.00%	35.69%	99.17%
16	1.2E-09	0.02%	97.52%	>100%
15	5.6E-09	0.12%	>100%	>100%
14	2.6E-08	0.64%	>100%	>100%
13	1.2E-07	3.29%	>100%	>100%
12	5.4E-07	15.68%	>100%	>100%
11	2.4E-06	58.16%	>100%	>100%
10	1.0E-05	99.35%	>100%	>100%
9	4.6E-05	99.77%	>100%	>100%
8	2.0E-04	>100%	>100%	>100%
7	8.5E-04	>100%	>100%	>100%
6	3.7E-03	>100%	>100%	>100%
5	1.6E-02	>100%	>100%	>100%
4	6.4E-02	>100%	>100%	>100%
3	2.5E-01	>100%	>100%	>100%
2	7.1E-01	>100%	>100%	>100%
1	1.0E+00	>100%	>100%	>100%

The possibility of complimentary regions occurring within the genome of interest was not a major concern in the primer design process despite the reported tolerance of PCR to mismatches in complex sample populations (see, e.g., Rubin, E. and A.A. Levy, A mathematical model and a computerized simulation of PCR using complex templates. Nucleic Acids Res, 1996. 24(18): p. 3538-45). Although the probability of finding a perfect match to a 20 base primer is extremely low (4^{20}) (Table 3), the probability of finding less non-consecutive matches increases significantly with the size of the genome of interest. As a result, the probability of finding a perfect match of at least 10 of 20 bases is 99.35% for an Adenovirus genome. The probability of finding a 16 base perfect match is 97% for the sequences in the NCBI database (approximately 100 times larger than the Adenovirus genome). The probability of finding a 17 base perfect match to a 20 base primer is 99% for the sequences in the human genome (3 billion bases).

The high probability of primer cross-hybridization to regions of the genome is less problematic than one might expect due to the random DNA digestion used to produce the template fragments. Thus, the effects of a cross-hybridizing region (CHR) are fairly benign.

IDT Exhibit 1020.9912

It is unlikely that a CHR would be able to successfully compete with the perfect match between the PCR primers in solution and the template. In addition, any primers that include mismatches at their 3' end would be at a significant competitive disadvantage. Even if a CHR should out compete the intended PCR primer, it would produce a truncated PCR product, without a downstream site for the sequencing primer. If the truncated product could be driven to the capture bead and immobilized, one of two situations would result. If the CHR out-competed the solution–phase primer, then the immobilized product would lack a sequencing primer binding site, and would result in an empty PicoTiter plate (PTP) well. If the CHR out-competed the bead-bound primer, the sequencing primer would still be present, and the only effect would be a shorter insert. Neither result would unduly compromise the sequencing quality. Given the large amount of genomic material used in the sample preparation process (currently 25 μg, containing 5.29×10^{16} copies of the 35 Kb Adenovirus genome), oversampling can be used to provide fragments that lack the complete CHR, and allow standard PCR amplification of the region in question.

Example 3: Sample Preparation by Nebulization

Preparation of DNA by Nebulization

The purpose of the Nebulization step is to fragment a large stretch of DNA such as a whole genome or a large portion of a genome into smaller molecular species that are amenable to DNA sequencing. This population of smaller-sized DNA species generated from a single DNA template is referred to as a library. Nebulization shears double-stranded template DNA into fragments ranging from 50 to 900 base pairs. The sheared library contains single-stranded ends that are end-repaired by a combination of T4 DNA polymerase, *E. coli* DNA polymerase I (Klenow fragment), and T4 polynucleotide kinase. Both T4 and Klenow DNA polymerases are used to "fill-in" 3' recessed ends (5' overhangs) of DNA via their 5'-3' polymerase activity. The single-stranded 3'-5' exonuclease activity of T4 and Klenow polymerases will remove 3' overhang ends and the kinase activity of T4 polynucleotide kinase will add phosphates to 5' hydroxyl termini.

The sample was prepared as follows:

1. 15 μg of gDNA (genomic DNA) was obtained and adjusted to a final volume of 100 μl in 10 mM TE (10 mM Tris, 0.1 mM EDTA, pH 7.6; see reagent list at the end of section). The DNA was analyzed for contamination by measuring the O.D. $_{260/280}$ ratio, which was 1.8 or higher. The final gDNA concentration was expected to be approximately

IDT Exhibit 1020.9913

300 µg/ml.

2. 1600 µl of ice-cold Nebulization Buffer (see end of section) was added to the gDNA.

3. The reaction mixture was placed in an ice-cold nebulizer (CIS-US, Bedford, MA).

4. The cap from a 15 ml snap cap falcon tube was placed over the top of the nebulizer (Figure 7A).

5. The cap was secured with a clean Nebulizer Clamp assembly, consisting of the fitted cover (for the falcon tube lid) and two rubber O-rings (Figure 7B).

6. The bottom of the nebulizer was attached to a nitrogen supply and the entire device was wrapped in parafilm (Figures 7C and 7D).

7. While maintaining nebulizer upright (as shown in Figure 7D), 50 psi (pounds per square inch) of nitrogen was applied for 5 minutes. The bottom of the nebulizer was tapped on a hard surface every few seconds to force condensed liquid to the bottom.

8. Nitrogen was turned off after 5 minutes. After the pressure had normalized (30 seconds), the nitrogen source was remove from the nebulizer.

9. The parafilm was removed and the nebulizer top was unscrewed. The sample was removed and transferred to a 1.5 ml microcentrifuge tube.

10. The nebulizer top was reinstalled and the nebulizer was centrifuged at 500 rpm for 5 minutes.

11. The remainder of the sample in the nebulizer was collected. Total recovery was about 700 µl.

12. The recovered sample was purified using a QIAquick column (Qiagen Inc., Valencia, CA) according to manufacturer's directions. The large volume required the column to be loaded several times. The sample was eluted with 30 µl of Buffer EB (10 mM Tris HCl, pH 8.5;supplied in Qiagen kit) which was pre-warmed at 55°C.

13. The sample was quantitated by UV spectroscopy (2 µl in 198 µl water for 1:100 dilution).

Enzymatic Polishing

Nebulization of DNA templates yields many fragments of DNA with frayed ends. These ends are made blunt and ready for ligation to adaptor fragments by using three enzymes, T4 DNA polymerase, *E. coli* DNA polymerase (Klenow fragment) and T4

IDT Exhibit 1020.9914

polynucleotide kinase.

The sample was prepared as follows:

1. In a 0.2 ml tube the following reagents were added in order:

28 µl purified, nebulized gDNA fragments
5 µl water
5 µl 10 X T4 DNA polymerase buffer
5 µl BSA (1mg/ml)
2 µl dNTPs (10 mM)
5 µl T4 DNA polymerase (3 units/µl)
50 µl final volume

2. The solution of step 1 was mixed well and incubated at 25°C for 10 minutes in a MJ thermocycler (any accurate incubator may be used).

3. 1.25 µl *E. coli* DNA polymerase (Klenow fragment) (5 units/ml) was added.

4. The reaction was mixed well and incubated in the MJ thermocycler for 10 minutes at 25°C and for an additional 2 hrs at 16°C.

5. The treated DNA was purified using a QiaQuick column and eluted with 30 µl of Buffer EB (10 mM Tris HCl, pH 8.5) which was pre-warmed at 55°C.

6. The following reagents were combined in a 0.2 ml tube:

30 µl Qiagen purified, polished, nebulized gDNA fragments
5 µl water
5 µl 10 X T4 PNK buffer
5 µl ATP (10 mM)
5 µl T4 PNK (10 units/ml)
50 µl final volume

7. The solution was mixed and placed in a MJ thermal cycler using the T4 PNK program for incubation at 37°C for 30 minutes, 65°C for 20 minutes, followed by storage at 14°C.

8. The sample was purified using a QiaQuick column and eluted in 30 µl of Buffer EB which was pre-warmed at 55°C.

10. A 2 µl aliquot of the final polishing reaction was held for analysis using a BioAnalyzer DNA 1000 LabChip (see below).

Ligation of Adaptors

IDT Exhibit 1020.9915

The procedure for ligating the adaptors was performed as follows:

1. In a 0.2 ml tube the following reagents were added in order:

20.6 µl molecular biology grade water
28 µl digested, polished gDNA Library
60 µl 2 X Quick Ligase Reaction Buffer
1.8 µl MMP (200 pmol/µl) Universal Adaptor set
9.6 µl Quick Ligase
120 µl total

The above reaction was designed for 5 µg and was scaled depending on the amount of gDNA used.

2. The reagents were mixed well and incubated at 25°C for 20 minutes. The tube was on ice until the gel was prepared for agarose gel electrophoresis.

Gel Electrophoresis and Extraction of Adapted gDNA Library

Nebulization of genomic DNA yields a library population that ranges from 50-900 bp. The addition of the 88-bp Universal Adaptor set will shift the population to a larger size and will result in a migration profile with a larger size range (approximately 130-980 bp). Adaptor dimers will migrate at 88 bp and adaptors not ligated will migrate at 44 bp. Therefore, genomic DNA libraries isolated in size ranges ≥250 bp can be physically isolated from the agarose gel and purified using standard gel extraction techniques. Gel isolation of the adapted gDNA library will result in the recovery of a library population in a size range that is ≥250 bp (size range of library can be varied depending on application). The library size range after ligation of adapters is 130 to 980 bp. It should be noted that the procedure may be adapted for isolation of any band size range, such as, for example, 130 to 200 bp, 200 to 400 bp, 250 to 500 bp, 300 to 600 bp, 500 to 700 bp and the like by cutting different regions of the gel. The procedure described below was used to isolated fragments of 250 bp to 500 bp.

A 150 ml agarose gel was prepared to include 2% agarose, 1X TBE, and 4.5 µl ethidium bromide (10 mg/ml stock). The ligated DNA was mixed with 10X Ready Load Dye and loaded onto the gel. In addition, 10 µl of a100-bp ladder (0.1 µg/µl) was loaded on two lanes away from the ligation reaction flanking the sample. The gel was electrophoresed at 100 V for 3 hours. When the gel run was complete, the gel was removed from the gel box, transferred to a GelDoc, and covered with plastic wrap. The DNA bands were visualized

IDT Exhibit 1020.9916

using the Prep UV light. A sterile, single-use scalpel, was used to cut out a library population from the agarose gel with fragment sizes of 250 – 500 bp. This process was done as quickly as possible to prevent nicking of DNA. The gel slices were placed in a 15 ml falcon tube. The agarose-embedded gDNA library was isolated using a Qiagen MinElute Gel Extraction kit. Aliquots of each isolated gDNA library were analyzed using a BioAnalyzer DNA 1000 LabChip to assess the exact distribution of the gDNA library population.

Strand Displacement and Extension of the gDNA Library and Isolation of the Single Stranded gDNA Library Using Streptavidin Beads

Strand displacement and extension of nicked double-stranded gDNA library was performed as described in Example 1, with the exception that the *Bst*-treated samples were incubated in the thermal cycler at 65° C for 30 minutes and placed on ice until needed. Streptavidin beads were prepared as described in Example 1, except that the final wash was performed using two washes with 200 µl 1X Binding buffer and two washes with 200 µl nH_2O. Single-stranded gDNA library was isolated using streptavidin beads as follows. Water from the washed beads was removed and 250 µl of Melt Solution (see below) was added. The bead suspension was mixed well and incubated at room temperature for 10 minutes on a tube rotator. In a separate tube, 1250 µl of PB (from the QiaQuick Purification kit) and 9 µl of 20% acetic acid were mixed. The beads in 250 µl Melt Solution were pelleted using a Dynal MPC and the supernatant was carefully removed and transferred to the freshly prepared PB/acetic acid solution. DNA from the 1500 µl solution was purified using a single MinElute purification spin column. This was performed by loading the sample through the same column twice at 750 µl per load. The single stranded gDNA library was eluted with 15 µl of Buffer EB which was pre-warmed at 55°C.

Single Strand gDNA Quantitation and Storage

Single-stranded gDNA was quantitated using RNA Pico 6000 LabChip as described in Example 1. In some cases, the single stranded library was quantitated by a second assay to ensure the initial Agilent 2100 quantitation was performed accurately. For this purpose, RiboGreen quantitation was performed as described (ssDNA Quantitation by Fluorometry) to confirm the Agilent 2100 quantitation. If the two estimates differed by more than 3 fold, each analysis was repeated. If the quantitation showed greater than a 3 fold difference between the two procedures, a broader range of template to bead was used.

IDT Exhibit 1020.9917

Dilution and storage of the single stranded gDNA library was performed as described in Example 1. The yield was as follows:

Remaining final volume of ssDNA library following LabChip analysis = 12 µl.

Remaining final volume of ssDNA library following RiboGreen analysis = 9 µl.

Final volume of ssDNA library after the addition of TE = 18 µl.

An equal volume of TE was added to single-stranded gDNA library stock. Single-stranded gDNA library to 1 x 10^8 molecules/µl in Buffer TE. Stock was diluted (1/500) to 200,000 molecules/µl in TE and 20 µl aliquots were prepared.

Library Fragment Size Distribution After Nebulization

Typical results from Agilent 2100 DNA 1000 LabChip analysis of 1 µl of the material following Nebulization and polishing are shown in Figure 9A. The size range distribution of the majority of the product was expected to fall around 50 to 900 base pairs. The mean size (top of peak) was expected to be approximately 450 bp. Typical results from gel purification of adaptor ligated library fragments are shown in Figure 9B.

Reagents

Unless otherwise specified, the reagents listed in the Examples represent standard reagents that are commercially available. For example, Klenow, T4 DNA polymerase, T4 DNA polymerase buffer, T4 PNK, T4 PNK buffer, Quick T4 DNA Ligase, Quick Ligation Buffer, *Bst* DNA polymerase (Large Fragment) and ThermoPol reaction buffer are available from New England Biolabs (Beverly, MA). dNTP mix is available from Pierce (Rockford, IL). Agarose, UltraPure TBE, BlueJuice gel loading buffer and Ready-Load 100bp DNA ladder may be purchased from Invitrogen (Carlsbad, CA). Ethidium Bromide and 2-Propanol may be purchased from Fisher (Hampton, NH). RNA Ladder may be purchased from Ambion (Austin, TX). Other reagents are either commonly known and/or are listed below:

Melt Solution:

Ingredient	Quantity Required	Vendor	Stock Number
NaCl (5 M)	200 µl	Invitrogen	24740-011
NaOH (10 N)	125 µl	Fisher	SS255-1
molecular biology grade water	9.675 ml	Eppendorf	0032-006-205

The Melt Solution included 100 mM NaCl, and 125 mM NaOH. The listed reagents were combined and mixed thoroughly. The solution could be stored at RT for six months.

IDT Exhibit 1020.9918

Binding & Washing (B&W) Buffer (2X and 1X):

Ingredient	Quantity Required	Vendor	Stock Number
UltraPure Tris-HCl (pH 7.5, 1 M)	250 µl	Invitrogen	15567-027
EDTA (0.5 M)	50 µl	Invitrogen	15575-020
NaCl (5 M)	10 ml	Invitrogen	24740-011
molecular biology grade water	14.7 ml	Eppendorf	0032-006-205

The 2X B&W buffer included final concentrations of 10 mM Tris-HCl (pH 7.5), 1 mM EDTA, and 2 M NaCl. The listed reagents were combined by combined and mixed thoroughly. The solution could be stored at RT for 6 months. The 1X B&W buffer was prepared by mixing 2X B&W buffer with picopure H_2O, 1:1. The final concentrations was half of that listed the above, i.e., 5 mM Tris-HCl (pH 7.5), 0.5 mM EDTA, and 1 M NaCl.

Other buffers included the following. 1X T4 DNA Polymerase Buffer: 50 mM NaCl, 10 mM Tris-HCl, 10 mM MgCl2, 1 mM dithiothreitol (pH 7.9 @ 25°C). TE: 10 mM Tris, 1 mM EDTA.

Special Reagent preparation:

TE (10 mM):

Ingredient	Quantity Required	Vendor	Stock Number
TE (1M)	1 ml	Fisher	BP1338-1
molecular biology grade water	99 ml	Eppendorf	0032-006-205

Reagents were mixed and the solution could be stored RT for six months.

Nebulization Buffer:

Ingredient	Quantity Required	Vendor	Stock Number
Glycerol	53.1 ml	Sigma	G5516
molecular biology grade water	42.1 ml	Eppendorf	0032-006-205
UltraPure Tris-HCl (pH 7.5, 1M)	3.7 ml	Invitrogen	15567-027
EDTA (0.5M)	1.1 ml	Sigma	M-10228

All reagents were added (glycerol was added last) to a Stericup and mixed well. The solution was labeled and could be stored at RT for six months.

ATP (10 mM):

Ingredient	Quantity Required	Vendor	Stock Number
ATP (100 mM)	10 µl	Roche	1140965
molecular biology grade water	90 µl	Eppendorf	0032-006-205

IDT Exhibit 1020.9919

The reagents were mixed and the solution could be stored at -20°C for six months.

BSA (1 mg/ml):

Ingredient	Quantity Required	Vendor	Stock Number
BSA (10 mg/ml)	10 µl	NEB	M0203 kit
Molecular Biology Grade water	90 µl	Eppendorf	0032-006-205

The reagents were mixed and the solution could be stored at 4°C for six months.

Library Annealing buffer, 10X:

Ingredient	Quantity Req.	Vendor	Stock No.
UltraPure Tris-HCl (pH 7.5, 1 M)	200 ml	Invitrogen	15567-027
Magnesium acetate, enzyme grade (1 M)	10.72 g	Fisher	BP-215-500
Molecular Biology Grade water	~ 1 L	Eppendorf	0032-006-205

The 10 X Annealing Buffer included 200 mM Tris (pH 7.5) and 50 mM magnesium acetate. For this buffer, 200 ml of Tris was added to 500 ml picopure H_2O. Next, 10.72 g of magnesium acetate was added to the solution and dissolved completely. The solution was adjusted to a final volume of 1000 ml. The solution could be stored at 4°C for six months. To avoid the potential for contamination of libraries, the buffer was aliquotted for single or short-term usage.

Adaptors:

Adaptor "A" (400 µM):

Ingredient	Quantity Req.	Vendor	Stock No.
Adaptor A (sense; HPLC-purified, phosphorothioate linkages, 44 bp, 1000 pmol/µl)	10.0 µl	IDT	custom
Adaptor A (antisense; HPLC-purified, Phosphorothioate linkages, 40 bp, 1000 pmol/µl)	10.0 µl	IDT	custom
Annealing buffer (10X)	2.5 µl	454 Corp.	previous table
molecular biology grade water	2.5 µl	Eppendorf	0032-006-205

For this solution, 10 µl of 1000 pmol/µl Adaptor A (44 bp, sense) was mixed with 10 µl of 1000 pmol/µl Adaptor A (40 bp, antisense), 2.5 µl of 10X Library Annealing Buffer, and 2.5 µl of water (V_f = 25 µl). The adaptors were annealed using the ANNEAL-A program (see Appendix, below) on the Sample Prep Lab thermal cycler. More details on adaptor design are provided in the Appendix.

IDT Exhibit 1020.9920

Adaptor "B" (400 μM):

Ingredient	Quantity Req.	Vendor	Stock No.
Adaptor B (sense; HPLC-purified, phosphorothioate linkages, 40 bp, 1000 pmol/μl))	10 μl	IDT	Custom
Adaptor B (anti; HPLC-purified, phosphorothioate linkages, 5'Biotinylated, 44 bp, 1000 pmol/μl)	10 μl	IDT	Custom
Annealing buffer (10X)	2.5 μl	454 Corp.	previous table
molecular biology grade water	2.5 μl	Eppendorf	0032-006-205

For this solution, 10 μl of 1000 pmol/μl Adaptor B (40 bp, sense) was mixed with 10 μl of 1000 pmol/μl Adaptor B (44 bp, anti), 2.5 μl of 10 X Library Annealing Buffer, and 2.5 μl of water (V_f = 25 μl). The adaptors were annealed using the ANNEAL-A program (see Appendix) on the Sample Prep Lab thermal cycler. After annealing, adaptor "A" and adaptor "B" (V_f = 50 μl) were combined. Adaptor sets could be stored at -20°C until use.

20% Acetic Acid:

Ingredient	Quantity Required	Vendor	Stock Number
acetic acid, glacial	2 ml	Fisher	A35-500
molecular biology grade water	8 ml	Eppendorf	0032-006-205

For this solution, glacial acetic acid was added to the water.. The solution could be stored at RT for six months.

Appendix

Adaptor Annealing Program:

ANNEAL-A program for primer annealing:

1. Incubate at 95°C, 1 min;
2. Reduce temperature to 15°C at 0.1°C/sec; and
3. Hold at 14°C.

T4 Polymerase/Klenow POLISH program for end repair:

1. Incubate at 25°C, 10 minutes;
2. Incubate at 16°C, 2 hours; and
3. Hold at 4°C.

T4 PNK Program for end repair:

1. Incubate at 37°C, 30 minutes;

IDT Exhibit 1020.9921

2. Incubate at 65°C, 20 minutes; and
3. Hold at 14°C.

BST program for stand displacement and extension of nicked double-stranded gDNA:

1. Incubate at 65° C, 30 minutes; and
2. Hold at 14° C.

References

Hamilton, S.C., J.W. Farchaus and M.C. Davis. 2001. DNA polymerases as engines for biotechnology. *BioTechniques* 31:370.

QiaQuick Spin Handbook (QIAGEN, 2001): hypertext transfer protocol://world wide web.qiagen.com/literature/handbooks/qqspin/1016893HBQQSpin_PCR_mc_prot.pdf.

Quick Ligation Kit (NEB): hypertext transfer protocol://world wide web.neb.com/neb/products/mod_enzymes/M2200.html.

MinElute kit (QIAGEN): hypertext transfer protocol://world wide web.qiagen.com/literature/handbooks/minelute/1016839_HBMinElute_Prot_Gel.pdf.

Biomagnetic Techniques in Molecular Biology, Technical Handbook, 3rd edition (Dynal, 1998): hypertext transfer protocol://world wide web.dynal.no/kunder/dynal/DynalPub36.nsf/cb927fbab127a0ad4125683b004b011c/4908f5b1a665858a41256adf005779f2/$FILE/Dynabeads M-280 Streptavidin.pdf.

Bio Analyzer User Manual (Agilent): hypertext transfer protocol://world wide web.chem.agilent.com/temp/rad31B29/00033620.pdf

BioAnalyzer DNA and RNA LabChip Usage (Agilent): hypertext transfer protocol://world wide web.agilent.com/chem/labonachip

BioAnalyzer RNA 6000 Ladder (Ambion): hypertext transfer protocol://world wide web.ambion.com/techlib/spec/sp_7152.pdf

Throughout this specification, various patents, published patent applications and scientific references are cited to describe the state and content of the art. Those disclosures, in their entireties, are hereby incorporated into the present specification by reference.

OTHER EMBODIMENTS

Although particular embodiments have been disclosed herein in detail, this has been done by way of example for purposes of illustration only, and is not intended to be

IDT Exhibit 1020.9922

limiting with respect to the scope of the appended claims, which follow. In particular, it is contemplated by the inventors that various substitutions, alterations, and modifications may be made to the invention without departing from the spirit and scope of the invention as defined by the claims.

IDT Exhibit 1020.9923

CLAIMS

We claim:

1. A method for clonally isolating a library comprising a plurality of single stranded DNA molecules comprising:

(a) fragmenting large template DNA molecules to generate a plurality of fragmented DNA molecules;

(b) attaching a first or second universal double stranded adaptor to a first end of each fragmented DNA molecule and a first or second universal adaptor to a second end of each fragmented DNA molecule to form a mixture of adaptor ligated DNA molecules;

(c) isolating a plurality of single stranded DNA molecules each comprising a first single stranded universal adaptor and a second single stranded universal adaptor to obtain a library; and

(d) delivering the single stranded DNA molecules into reactors such that a plurality of the reactors include one DNA molecule, thereby clonally isolating the library.

2. A method for generating a library comprising a plurality of single stranded DNA molecules, comprising:

(a) fragmenting large or whole genomic template DNA molecules to generate a plurality of fragmented DNA molecules;

(b) ligating a first universal double stranded adaptor or a second universal adaptor to a first end of each fragmented DNA molecule and a first universal adaptor or second universal adaptor to a second end of each fragmented DNA molecule to produce a mixture of adaptor ligated DNA molecules, wherein the first universal adaptor contains a moiety that binds to a solid support;

(c) attaching to a solid support those DNA molecules comprising a first double stranded universal adaptor;

(d) removing adaptor ligated DNA molecules which have not attached to a solid support;

(e) strand separating those adaptor ligated DNA molecules that are attached to a solid support at only one end to release a plurality of single stranded DNA molecules having a first single stranded universal adaptor at one end and a second single stranded adaptor at the other end; and

IDT Exhibit 1020.9924

(f) isolating a library of the released single stranded DNA molecules of step (e) away from those DNA molecules that remain attached to the solid support.

3. The method according to claim 1 or 2, wherein the first universal double stranded adaptor or the second universal adaptor is attached by ligation.

4. The method according to claim 1, wherein (f) is accomplished by: i) delivering the single stranded DNA molecules onto a location on a reactor array; or ii) delivering the single stranded DNA molecules into droplets in a water-in-oil emulsion.

5. The method according to claim 1 or 2, further comprising the step of repairing single stranded nicks in the mixture of adaptor ligated DNA molecules using DNA repair and modifying enzymes.

6. The method according to claim 5, wherein the DNA repair and modifying enzymes are selected from the group consisting of polymerase, ligase, kinase, and combinations thereof.

7. The method according to claim 6, wherein the polymerase is *Bacillus stearothermophilus* polymerase I, the ligase is T4 ligase, and the kinase is T4 polynucleotide kinase.

8. The method according to claim 1 or 2, wherein the template DNA is selected from the group consisting of genomic DNA, cDNA, plasmid DNA, cosmid DNA, artificial chromosome DNA, synthetic DNA, phasemid DNA, and phagemid DNA.

9. The method according to claim 1 or 2, wherein the template DNA comprises reverse transcripts.

10. The method according to claim 1 or 2, wherein the fragmenting is performed by a means selected from the group consisting of enzymatic, chemical, and mechanical means.

IDT Exhibit 1020.9925

11. The method according to claim 10, wherein the enzymatic means is DNase I.

12. The method according to claim 10, wherein the mechanical means is nebulization.

13. The method according to claim 11, wherein the DNase I digestion is performed at a temperature of 10-37°C for 1-2 minutes.

14. The method according to claim 10, wherein the enzymatic means is a restriction endonuclease.

15. The method according to claim 10, wherein the mechanical means is selected from the group consisting of a French Press, a sonicator, a HydroShear, and a nebulizer.

16. The method according to claim 1 or 2, wherein the fragmented DNA molecules are 50 bp to 700 bp in length.

17. The method according to claim 1 or 2, wherein the compatible ends are blunt ends.

18. The method according to claim 17 wherein the blunt ends are created with an enzyme selected from the group consisting of *Pfu* polymerase, T4 DNA polymerase and Klenow fragments.

19. The method according to claim 1 or 2, wherein the compatible ends include an A or T overhang.

20. The method according to claim 1 or 2, wherein the first or second double stranded universal adaptor comprises phosphorothioate linkages.

IDT Exhibit 1020.9926

21. The method according to claim 1 or 2, wherein a biotin moiety is attached to the first or second double stranded universal adaptor.

22. The method according to claim 1 or 2, wherein the first or second double stranded universal adaptor is ligated with T4 DNA ligase.

23. The method according to claim 1 or 2, wherein either the first double stranded universal adaptors or the second double stranded universal adaptors or both double stranded universal adaptors comprise a discriminating key sequence.

24. The method according to claim 23, wherein the discriminating key sequence is 3-12 nucleotides in length.

25. The method according to claim 23, wherein the discriminating key sequence comprises at least one nucleotide selected from the group consisting of A, G, C, U, and T.

26. The method according to claim 1 or 2, wherein first and second double stranded universal adaptor comprise a PCR priming sequence and a sequencing primer sequence.

27. The method according to claim 26, wherein the PCR priming sequence is 10-20 base pairs in length.

28. The method according to claim 26 wherein the sequencing primer sequence is 10-20 base pairs in length.

29. The method according to claim 26, wherein the PCR priming sequence and the sequencing primer sequence overlap.

30. The method according to claim 1 or 2, wherein the mixture of adaptor ligated DNA molecules is separated by a method selected from the group consisting of gel electrophoresis, filtration, size exclusion chromatography, and sucrose sedimentation.

IDT Exhibit 1020.9927

31. The method according to claim 1 or 2, wherein the plurality of single stranded DNA molecules is attached to a DNA capture bead.

32. The method according to claim 31, wherein the DNA capture bead comprises a component of a binding pair.

33. The method according to claim 31 wherein the binding pair is selected from the group consisting of avidin/biotin, ligand/receptor, antigen/antibody and complementary nucleotides.

34. The method according to claim 31, wherein the DNA capture bead is a paramagnetic bead.

35. The method according to claim 1 or 2, wherein the plurality of single stranded DNA molecules is obtained by a treatment selected from the group consisting of low salt treatment, high pH treatment, and chemical denaturation treatment.

36. A method for generating a single stranded DNA library attached to solid supports comprising:

(a) generating a plurality of single stranded DNA templates;

(b) attaching each of the plurality of ssDNA templates to a solid support; and

(c) isolating the solid supports on which the single stranded DNA templates are attached.

37. A method for generating a single stranded DNA library attached to solid supports comprising:

(a) fragmenting large template DNA molecules to generate a plurality of fragmented DNA molecules;

(b) attaching a first or second universal double stranded adaptor to a first end of each fragmented DNA molecule and a first or second universal adaptor to a second end of each fragmented DNA molecule to make a mixture of adaptor ligated DNA molecules;

IDT Exhibit 1020.9928

(c) isolating those single stranded DNA molecules which comprise a first single stranded universal adaptor and a second single stranded universal adaptor; and

(d) attaching the isolated single stranded molecules from (c) to a solid support.

38. The method according to claim 37, wherein the solid support is a DNA capture bead.

39. The method according to claim 37, wherein the DNA is selected from the group consisting of genomic DNA, cDNA, plasmid DNA, cosmid DNA, artificial chromosome DNA, synthetic DNA, phasemid DNA, and phagemid DNA.

40. The method according to claim 37, wherein the DNA is attached to the solid support via a binding pair.

41. The method according to claim 40, wherein the binding pair is selected from the group consisting of avidin/biotin, ligand/receptor, antigen/antibody and complementary nucleotides.

42. A library of mobile solid supports made by the method of claim 37.

43. A nucleic acid molecule comprising a first adaptor, a fragment of template DNA, and a second adaptor, wherein the first adaptor and second adaptor each comprise a sequencing primer, a PCR primer, and a discriminating key sequence, and wherein the first adaptor and second adaptor, when dissociated, do not cross-hybridize to each other under stringent hybridization conditions.

44. The nucleic acid molecule of claim 43, wherein the PCR primer is 10-20 base pairs in length.

45. The nucleic acid molecule of claim 43, wherein the sequencing primer is 10-20 base pairs in length.

IDT Exhibit 1020.9929

46. The nucleic acid molecule of claim 43, wherein the discriminating key sequence is 3 to 12 base pairs in length.

47. The nucleic acid molecule of claim 43, wherein the template DNA is selected from the group consisting of genomic DNA, cDNA plasmid DNA, cosmid DNA, artificial chromosome DNA, synthetic DNA, phasemid DNA, and phagemid DNA.

48. The nucleic acid molecule of claim 43, wherein the nucleic acid molecule, when dissociated, has minimal cross-hybridization to dissociated template DNA.

49. A method for preparing single stranded DNA molecules, comprising:

(a) fragmenting large or whole genomic template DNA molecules to generate a plurality of fragmented DNA molecules;

(b) ligating a first universal double stranded adaptor or a second universal adaptor to a first end of each fragmented DNA molecule and a first universal adaptor or second universal adaptor to a second end of each fragmented DNA molecule to produce a mixture of adaptor ligated DNA molecules;

(c) attaching adaptor ligated DNA molecules comprising a first double stranded universal adaptor and a second double stranded adaptor to a solid support via one strand of the first double stranded universal adaptor;

(d) washing away adaptor ligated DNA molecules which have not attached to a solid support;

(e) strand separating those adaptor ligated DNA molecules that are attached to a solid support at only one end to release a plurality of single stranded DNA molecules comprising a first single stranded universal adaptor at one end and a second single stranded adaptor at the other end; and; and

(f) isolating the single stranded DNA molecules.

50. A method for delivering nucleic acid templates to a plurality of reaction centers comprising the steps of:

(a) providing a population of nucleic acid templates;

(b) isolating each nucleic acid template from said population to a sequestering agent to form a population of sequestered nucleic acid templates;

IDT Exhibit 1020.9930

(c) delivering said population of sequestered nucleic acid templates to said plurality of reaction centers wherein each reaction center receives one sequestered nucleic acid.

51. The method of claim 50, wherein said isolating step comprises attaching said nucleic acid templates to a bead.

52. The method of claim 50, wherein said isolating step comprises encapsulating said nucleic acid template in an emulsion of a water-in-oil emulsion.

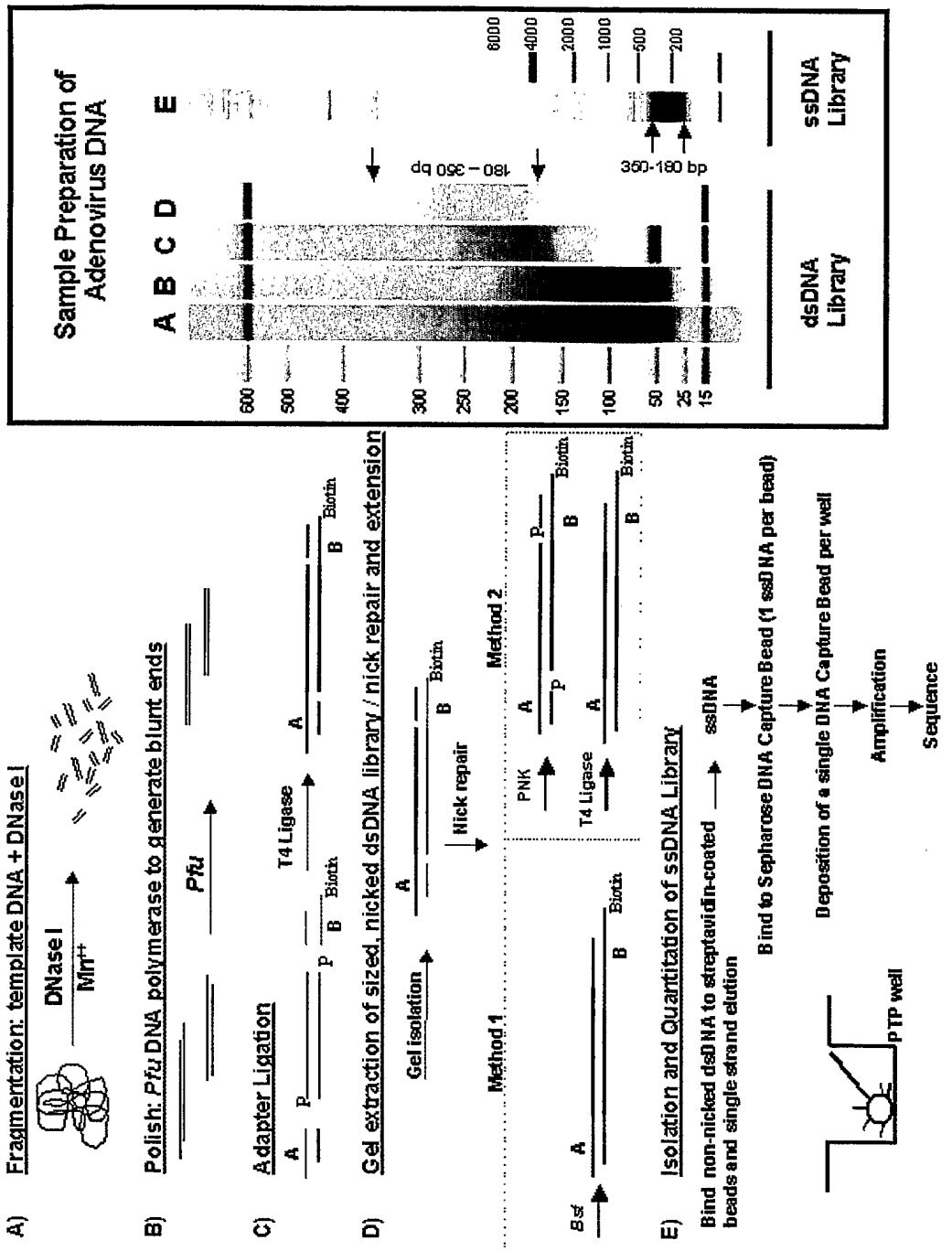
53. The method of claim 52, wherein said nucleic acid template is encapsulated with a bead and wherein the bead can bind said nucleic acid.

54. The method of claim 50, wherein said delivering step comprises delivering said sequestered nucleic acid to a plurality of reaction centers, wherein each reaction center is a well on a picotiter plate.

55. The method of claim 1, 2 or 49 further comprising the step of attaching the isolated single stranded molecules each individually to a solid support.

IDT Exhibit 1020.9931

Figure 1



IDT Exhibit 1020.9932

Figure 2A

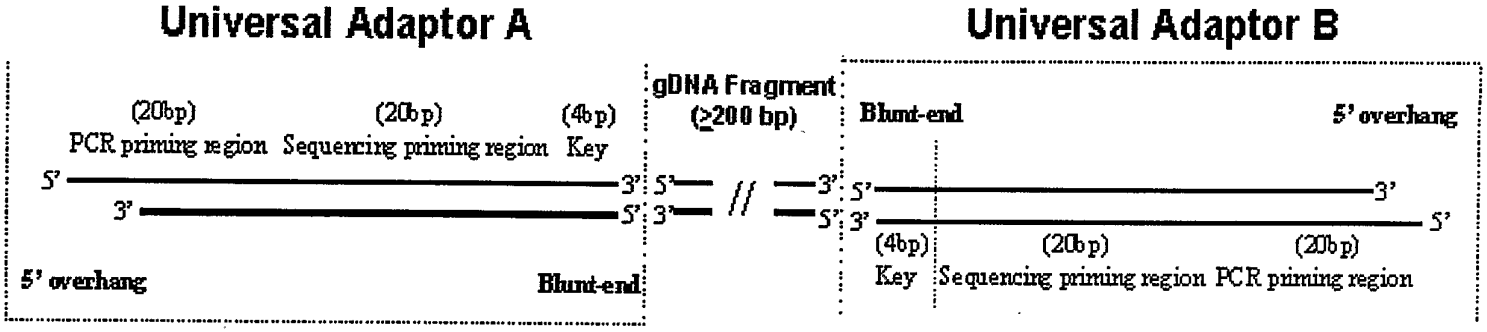


Figure 2B

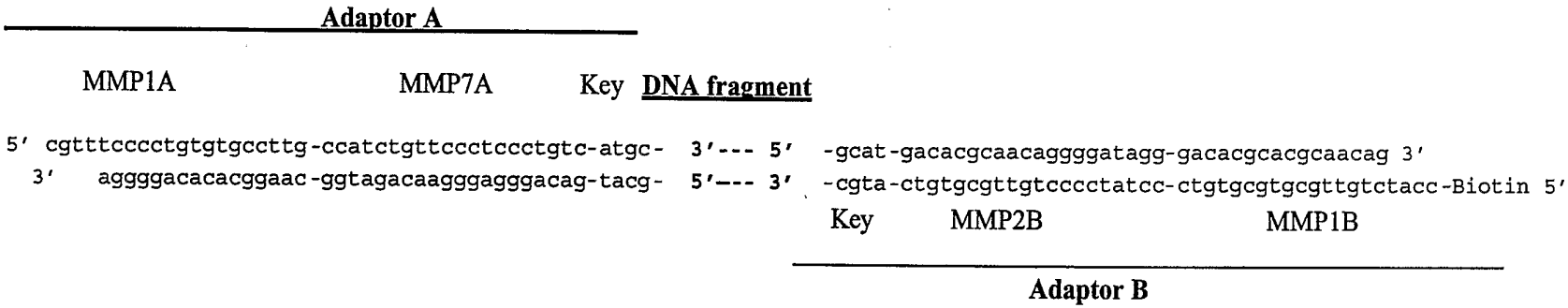


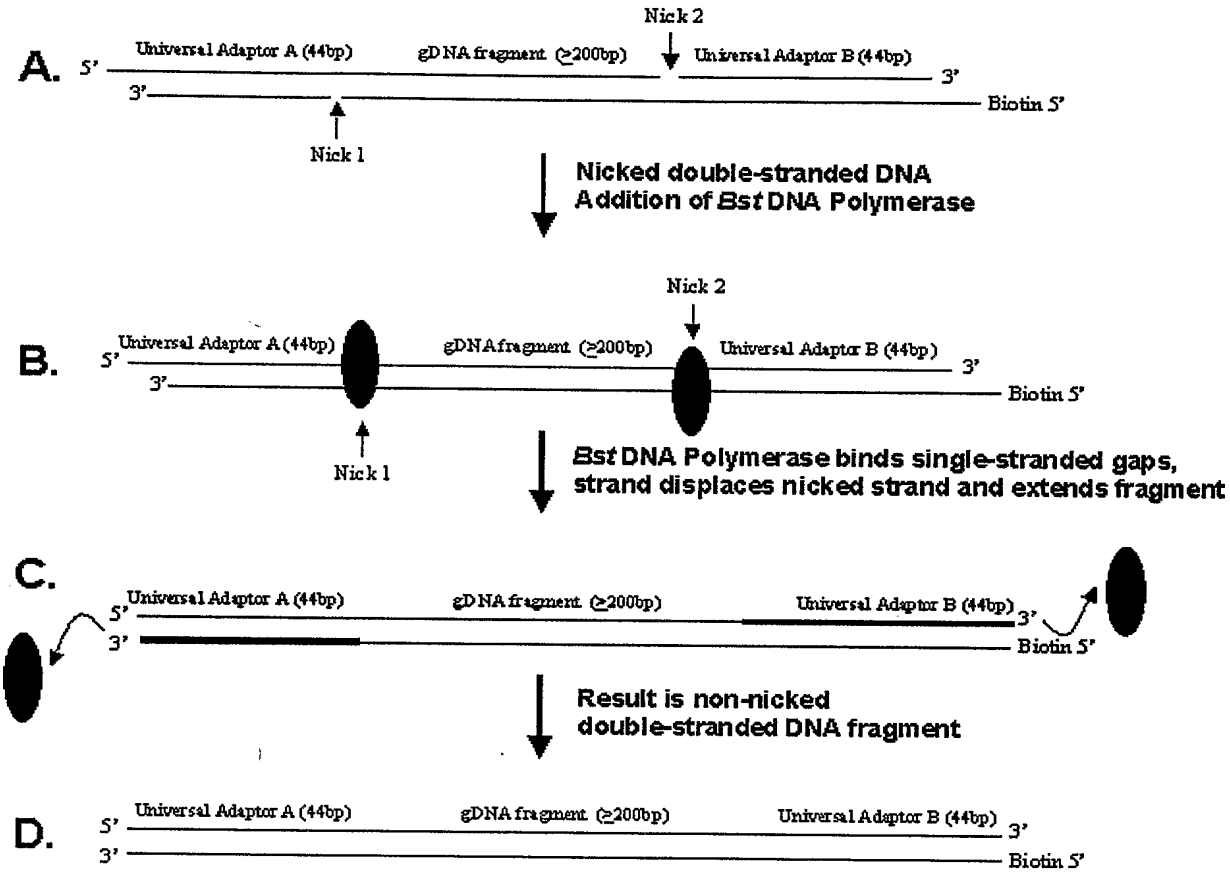
Figure 2C



Sense Strand 5' [PCR primer (20 bp)----- Seq Primer (20 bp)- Key (4 bp)] DNA Fragment [key (4 bp)-Seq Primer (20 bp)-PCR primer (20 bp)] 3'
Anti Strand 3' [PCR primer (20 bp)- Seq Primer (20 bp)- Key (4 bp)] DNA Fragment [key (4 bp)-Seq Primer (20 bp)-----PCR primer (20 bp)] 5' **(BEAD)**

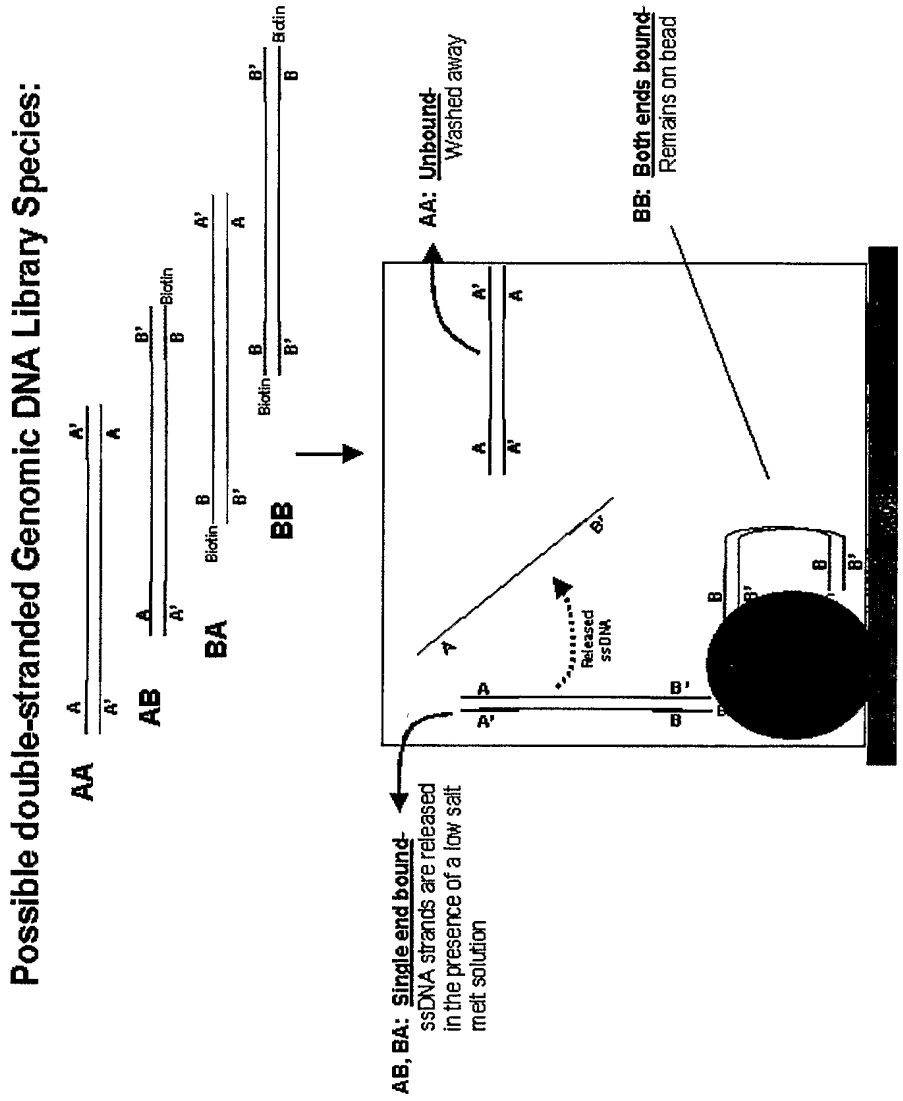
IDT Exhibit 1020.9933

Figure 3



IDT Exhibit 1020.9934

Figure 4

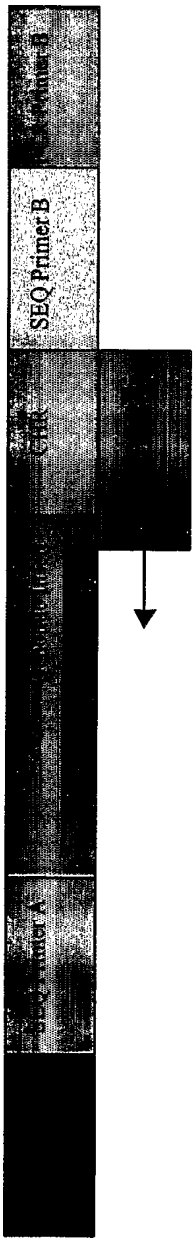


IDT Exhibit 1020.9935

Figure 5A

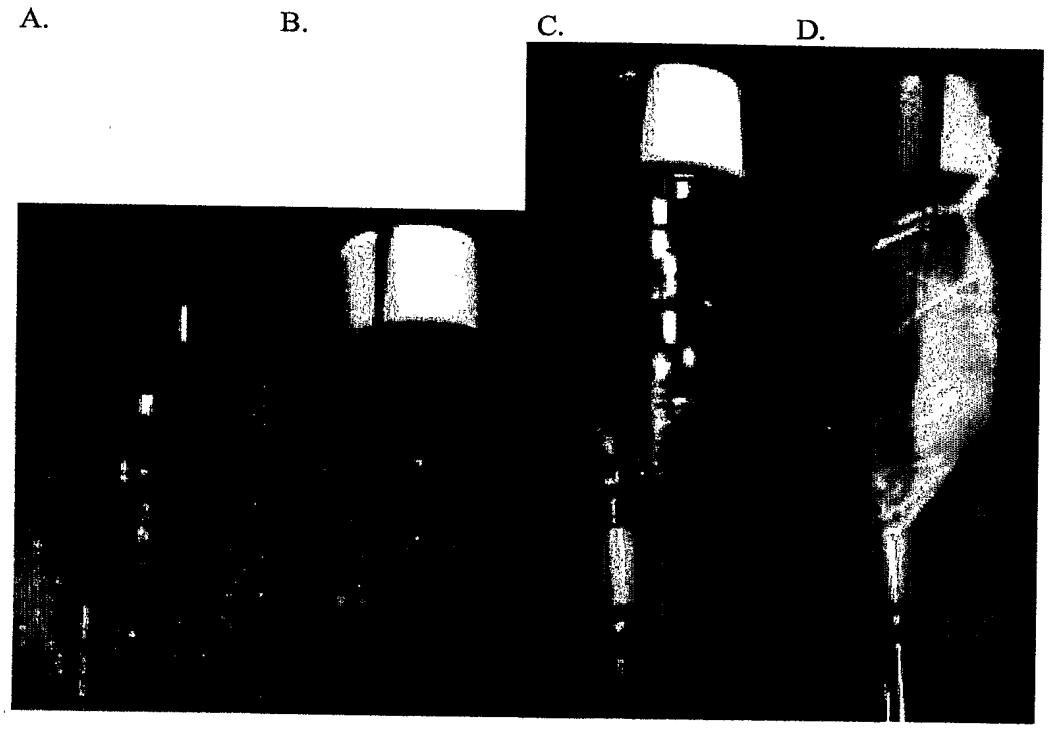


Figure 6



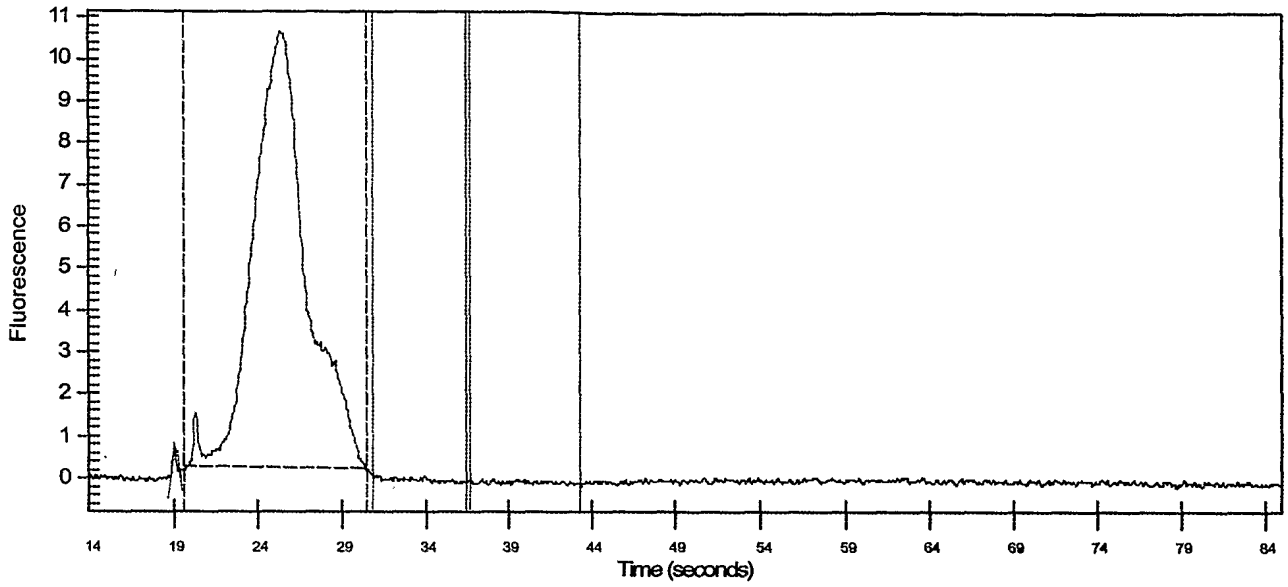
IDT Exhibit 1020.9936

Figure 7



IDT Exhibit 1020.9937

Figure 8



IDT Exhibit 1020.9938

Figure 9A

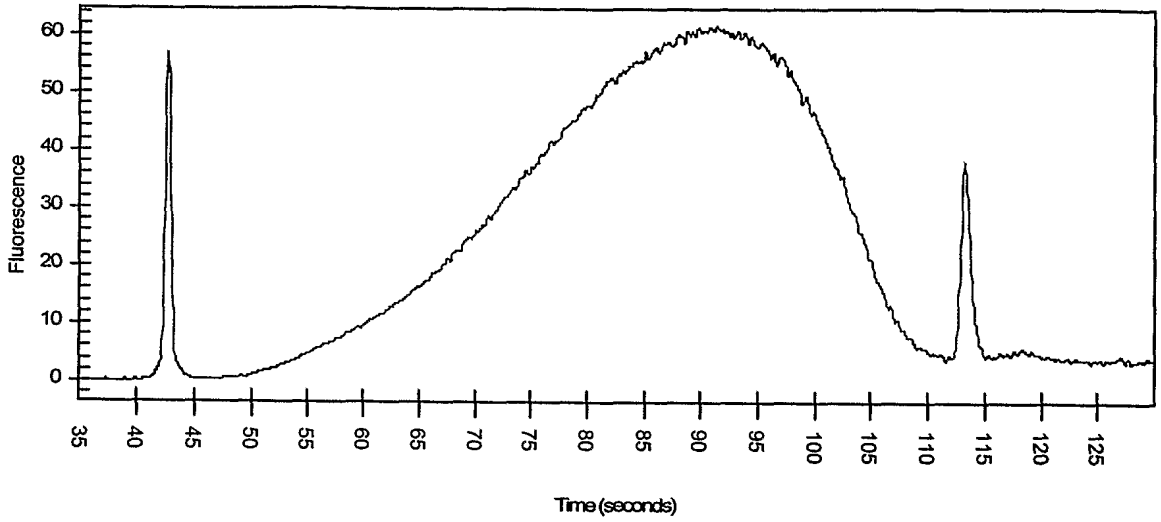
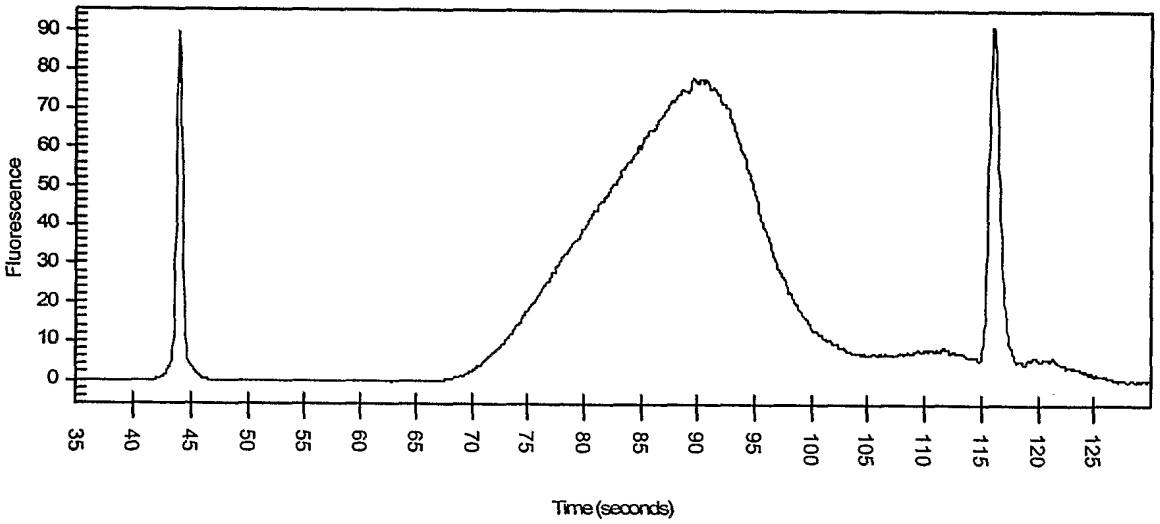
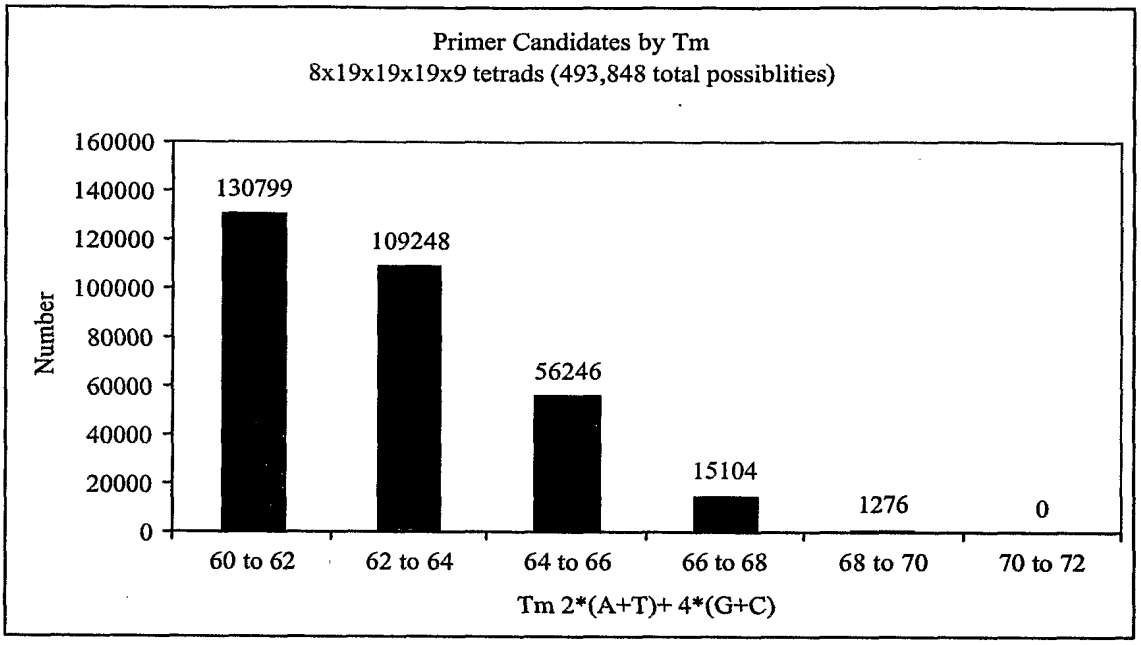


Figure 9B



IDT Exhibit 1020.9939

Figure 10



IDT Exhibit 1020.9940

SEQUENCE LISTING

<110> Costa, Gina L.
Weiner, Michael
Leamon, John
Rothberg, Jonathan

<120> Method For Preparing Single Stranded DNA Libraries

<130> 21465-507 UTIL

<150> US 60/443,471
<151> 2003-01-29

<150> US 60/465,071
<151> 2003-04-23

<150> US 60/476,592
<151> 2003-06-06

<150> US 60/476,504
<151> 2003-06-06

<150> US 60/476,592
<151> 2003-06-06

<150> US 60/476,602
<151> 2003-06-06

<150> US 60/497,985
<151> 2003-08-25

<160> 4

<170> PatentIn version 3.2

<210> 1
<211> 44
<212> DNA
<213> Artificial

<220>
<223> Oligonucleotide

<400> 1
cgtttcccct gtgtgccttg ccatctgttc cctccctgtc atgc 44

<210> 2
<211> 40
<212> DNA
<213> Artificial

<220>
<223> Oligonucleotide

<400> 2
gcatgacagg gagggaacag atggcaaggc acacagggga 40

<210> 3
<211> 40
<212> DNA
<213> Artificial

IDT Exhibit 1020.9941

<220>
<223> Oligonucleotide

<400> 3
gcatgacacg caacagggga tagggacacg cacgcaacag 40

<210> 4
<211> 44
<212> DNA
<213> Artificial

<220>
<223> Oligonucleotide

<400> 4
ccatctgttg cgtgcgtgtc cctatcccct gttgcgtgtc atgc 44

IDT Exhibit 1020.9942

Espacenet

Bibliographic data: WO03078645 (A2) — 2003-09-25

METHODS FOR GENERATING DOUBLE STRANDED DNA COMPRISING A 3' SINGLE STRANDED PORTION AND USES OF THESE COMPLEXES FOR RECOMBINATION

Inventor(s): KURN NURITH [US]; WANG MARTIN JUNHONG [US] ± (KURN, NURITH, ; WANG, MARTIN, JUNHONG)

Applicant(s): NUGEN TECHNOLOGIES INC [US]; KURN NURITH [US]; WANG MARTIN JUNHONG [US] ± (NUGEN TECHNOLOGIES, INC, ; KURN, NURITH, ; WANG, MARTIN, JUNHONG)

Classification: **- international:** ***C12N15/09; C12N15/10; C12P19/34; C12Q1/68;*** (IPC1-7): C12Q

- cooperative: **C12N15/1096; C12Q1/6853** more

Application number: WO2003US07425 20030311 Global Dossier

Priority number (s): US20020363729P 20020311

Also published as: WO03078645 (A3) AT405676 (T) AU2003222269 (A1) CA2478875 (A1) EP1488001 (A2) more

Abstract of WO03078645 (A2)

Methods of recombining nucleic acids are provided. In particular, methods for the production of partially double stranded nucleic acids comprising a 3' overhang from an RNA target and use in methods of recombining polynucleotides is described. These methods do not require thermocycling. The present invention also provides methods of recombining and selection which allow for identification of proteins comprising improved or desired characteristics.

IDT Exhibit 1020.9943

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
25 September 2003 (25.09.2003)

PCT

(10) International Publication Number
WO 03/078645 A2

(51) International Patent Classification[7]: C12Q

(21) International Application Number: PCT/US03/07425

(22) International Filing Date: 11 March 2003 (11.03.2003)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/363,729 11 March 2002 (11.03.2002) US

(71) Applicant *(for all designated States except US)*: **NUGEN TECHNOLOGIES, INC.** [US/US]; 821 Industrial Road, San Carlos, CA 94070 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **KURN, Nurith** [IL/US]; 2876 Ramona Street, Palo Alto, CA 94306 (US). **WANG, Martin, Junhong** [US/US]; 120 Boothbay Avenue, Foster City, CA 94404 (US).

(74) Agents: **POLIZZI, Catherine, M.** et al.; Morrison & Foerster LLP, 755 Page Mill Road, Palo Alto, CA 94304-1018 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NI, NO, NZ, OM, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, RO, SE, SI, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 03/078645 A2

(54) Title: METHODS FOR GENERATING DOUBLE STRANDED DNA COMPRISING A 3' SINGLE STRANDED PORTION AND USES OF THESE COMPLEXES FOR RECOMBINATION

(57) Abstract: Methods of recombining nucleic acids are provided. In particular, methods for the production of partially double stranded nucleic acids comprising a 3' overhang from an RNA target and use in methods of recombining polynucleotides is described. These methods do not require thermocycling. The present invention also provides methods of recombining and selection which allow for identification of proteins comprising improved or desired characteristics.

IDT Exhibit 1020.9944

METHODS FOR GENERATING DOUBLE STRANDED DNA COMPRISING A 3' SINGLE STRANDED PORTION AND USES OF THESE COMPLEXES FOR RECOMBINATION

CROSS REFERENCE TO RELATED APPLICATIONS

This application claims the priority benefit of provisional application U.S. Serial No. 60/363,729, filed March 11, 2002, the contents of which are incorporated by reference in their entirety.

TECHNICAL FIELD

The present invention relates to methods and compositions for generating hybridized first and second primer extension product comprising a 3' single stranded portion, from RNA template, and uses for these complexes in recombination methods. In one aspect, the invention provides methods, which do not require thermocycling, for generating recombined polynucleotides useful for screening and/or selection of a desired phenotype and/or encoding a protein having an advantageous predetermined property.

BACKGROUND ART

There has been a rapid and steady progress in the ability to clone and recombine DNA molecules in recent years. These advances began with the discovery of restriction enzymes that were capable of cleaving double-stranded DNA, so that DNA fragments were produced that could be recombined to generate new recombinant molecules. The revolution was extended by the discovery and development of the polymerase chain reaction (PCR), which allowed rapid amplification of particular DNA segments, producing large amounts of material that could be subsequently cleaved and recombined with other DNA molecules.

Despite the power of these digestion and amplification techniques, however, there remains substantial room for improvement. Restriction enzymes are expensive and sometimes inefficient or available in crude, contaminated preparations. Further, PCR amplification often yields products that are refractory to direct cloning, due to addition of a terminal 3'-dAMP residue by many thermophilic DNA polymerases, including *Taq*, the most commonly-used enzyme for PCR.

Particularly desirable systems would generate hybrid products suitable for nucleic acid recombination with minimal reliance on restriction enzymes, would provide for

IDT Exhibit 1020.9945

efficient recombination, and would be generally useful for recombination between nucleic acids having a wide variety of chemical structures.

Jarrell et al, *see* WO 00/40715 and U.S. Patent No. 6,358,712, teach methods for library assembly involving generation of double stranded products with 3' or 5' overhangs. However, this method uses polymerase chain reaction, thus requiring thermal cycling and exponential amplification. Furthermore, Jarrell discloses use of a double stranded DNA as template.

Methods using production of recombined products include molecular evolution and "DNA shuffling". Molecular evolution is a powerful tool for producing novel proteins with enhanced or unique selectable properties. Proteins with enhanced enzymatic activity, enhanced thermal stability, or enhanced stability in organic solvents or defined media such as alkaline or high salt media, novel enzymatic activity, or other desired features may be produced by recombining pieces of mutagenized variants of a single gene or pieces of different genes to form hybrid proteins. In general, *in vitro* directed evolution involves generation of hybrid polynucleotides, amplification of the polynucleotides, and screening to select a protein having the desired property or properties. "DNA shuffling" involves mutagenesis of a single gene rather than reassembly of multiple domains from related genes. Known methods for DNA shuffling include mutagenesis of a gene, selection of mutants, fragmentation by DNase I, and recombination of the fragments *in vitro* using PCR (Stemmer, *Proc. Natl. Acad. Sci.* 91:10747-41 (1994); Stemmer, *Nature* 370:389-91 (1994)).

Analyses have revealed that many proteins are composed of a number of discrete domains. The individual domains of such a protein are often involved in specific functions that contribute to the protein's overall activity. A number of domains have been found to be evolutionarily mobile. Mobile domains are characterized by their ability to fold independently. "Mosaic proteins" that include multiple mobile domains have been characterized and appear to have arisen through evolution by "exon shuffling." The natural process of exon shuffling can be mimicked *in vitro* by generating and screening libraries of exon shuffled genes (Kolkman and Stemmer, *Nature Biotechnology* 19:423-28 (2001)). The domain-encoding exons in such libraries may be from closely related or divergent genes. Further, domain-encoding exons may be derived from related genes from the same species or homologs from closely-related species.

IDT Exhibit 1020.9946

Thus, there is a serious need for methods of generating double stranded nucleic acid sequences comprising at least one defined single stranded DNA portion, where the double stranded nucleic acid sequences are generated from RNA using methods which do not require and thus avoid thermocycling, and where the double stranded nucleic acid sequences can be random, can be selected from a pool of all mRNA sequences, or can be selected from a related sequences (e.g., all sequences encoding a particular domain, sequences encoding members of a superfamily of genes, and the like).

All references cited herein, including patent applications and publications, are incorporated by reference in their entirety.

DISCLOSURE OF THE INVENTION

In one aspect, the invention provides methods for an aspect leading to generating a complex of first and second primer extension product, said complex comprising a 3' single stranded portion useful for recombining nucleic acids, said methods comprising: (a) preparing a complex of first and second primer extension products said complex comprising a 3' single stranded portion, wherein said complex of first and second primer extension products is prepared according to a method comprising: (i) extending a first primer hybridized to a target RNA with RNA-dependent DNA polymerase (an enzyme comprising RNA-dependent DNA polymerase activity), wherein the first primer is a composite primer comprising an RNA portion and a 3' DNA portion, whereby a complex comprising a first primer extension product and the target RNA is produced; (ii) cleaving RNA in the complex of step (i) (cleavage may be effected, for example, with an enzyme that cleaves RNA from an RNA/DNA hybrid); (iii) extending a second primer hybridized to the first primer extension product using a DNA dependent polymerase (which, as described herein, can comprise an RNA dependent DNA polymerase activity, or can comprise a DNA dependent DNA polymerase and an RNA dependent polymerase; thus, extension involves using an enzyme comprising DNA-dependent DNA polymerase activity and an enzyme comprising RNA-dependent DNA polymerase activity, which as described herein may be in same or separate enzymes), whereby a second primer extension product is produced to form a complex of first and second primer extension products; (iv) cleaving RNA from the composite primer in the complex of step (iii) with an enzyme that cleaves

IDT Exhibit 1020.9947

RNA from an RNA/DNA hybrid such that a complex comprising hybridized first and second primer extension product comprising a 3' single stranded DNA portion is generated. In some embodiments, cleavage of step (ii) is effected using an enzyme that cleaves RNA from an RNA/DNA hybrid. In some embodiments, the enzyme is RNaseH.

In one aspect, the invention provides methods for an aspect leading to recombining nucleic acids (making a hybrid), said methods comprising: hybridizing a 3' single stranded portion of a complex of first and second primer extension products with a recipient nucleic acid molecule; whereby a hybrid nucleic acid is generated.

In another aspect, the invention provides methods for an aspect leading to recombining nucleic acids (making a hybrid), said methods comprising: hybridizing a 3' single stranded portion of a complex of first and second primer extension products with a recipient nucleic acid molecule, whereby a hybrid nucleic acid is generated, wherein said complex of first and second primer extension products is prepared according to a method comprising: (i) extending a first primer hybridized to a target RNA with an enzyme comprising RNA-dependent DNA polymerase activity, wherein the first primer is a composite primer comprising an RNA portion and a 3' DNA portion, whereby a complex comprising a first primer extension product and the target RNA is produced; (ii) cleaving RNA in the complex of step (i) (cleavage may be effected, for example, with an enzyme that cleaves RNA from an RNA/DNA hybrid); (iii) extending a second primer hybridized to the first primer extension product with an enzyme comprising DNA-dependent DNA polymerase activity and an enzyme comprising RNA-dependent polymerase activity, whereby a second primer extension product is produced to form a complex of first and second primer extension products; (iv) cleaving RNA from the composite primer in the complex of step (iii) with an enzyme that cleaves RNA from an RNA/DNA hybrid such that a complex comprising hybridized first and second primer extension product comprising a 3' single stranded DNA portion is generated. In some embodiments, cleavage of step (ii) is effected using an enzyme that cleaves RNA from an RNA/DNA hybrid. In some embodiments, the enzyme is RNaseH.

In one aspect, the invention provides methods for recombining nucleic acids, said methods comprising: (a) preparing a complex of first and second primer extension products said complex comprising a 3' single stranded portion, wherein said complex of first and second primer extension products is prepared according to a method comprising: (i) extending a first primer hybridized to a target RNA with an enzyme comprising RNA-

IDT Exhibit 1020.9948

dependent DNA polymerase activity , wherein the first primer is a composite primer comprising an RNA portion and a 3' DNA portion, whereby a complex comprising a first primer extension product and the target RNA is produced; (ii) cleaving RNA in the complex of step (i) (cleavage may be effected, for example, with an enzyme that cleaves RNA from an RNA/DNA hybrid); (iii) extending a second primer hybridized to the first primer extension product with an enzyme comprising DNA-dependent DNA polymerase activity and RNA-dependent DNA polymerase activity, whereby a second primer extension product is produced to form a complex of first and second primer extension products; (iv) cleaving RNA from the composite primer in the complex of step (iii) with an enzyme that cleaves RNA from an RNA/DNA hybrid such that a complex comprising hybridized first and second primer extension product comprising a 3' single stranded DNA portion is generated; (b) hybridizing the 3' single stranded portion of the complex of first and second primer extension products with a recipient nucleic acid molecule; whereby a hybrid nucleic acid is generated; and (c) generating a recombined nucleic acid from the hybrid nucleic acid. In some embodiments, cleavage of step (ii) is effected using an enzyme that cleaves RNA from an RNA/DNA hybrid. In some embodiments, the enzyme is RNaseH.

In one aspect, the invention provides methods for an aspect leading to recombining nucleic acids (making a hybrid), said methods comprising: generating a recombined nucleic acid from a hybrid nucleic acid, said hybrid molecule prepared by hybridizing a 3' single stranded portion of a complex of first and second primer extension products with a recipient nucleic acid molecule; whereby the hybrid nucleic acid is generated; wherein said complex of first and second primer extension products is prepared according to a method comprising: (i) extending a first primer hybridized to a target RNA with an enzyme comprising RNA-dependent DNA polymerase activity, wherein the first primer is a composite primer comprising an RNA portion and a 3' DNA portion, whereby a complex comprising a first primer extension product and the target RNA is produced; (ii) cleaving RNA in the complex of step (i) (cleavage may be effected, for example, with an enzyme that cleaves RNA from an RNA/DNA hybrid); (iii) extending a second primer hybridized to the first primer extension product with an enzyme comprising DNA-dependent DNA polymerase activity and RNA-dependent polymerase activity, whereby a second primer extension product is produced to form a complex of first and second primer extension products; (iv) cleaving RNA from the composite primer in the complex of step (iii) with an enzyme that cleaves RNA from an RNA/DNA hybrid such that a complex comprising

IDT Exhibit 1020.9949

hybridized first and second primer extension product comprising a 3' single stranded DNA portion is generated; (b) hybridizing the 3' single stranded portion of the complex of first and second primer extension products with a recipient nucleic acid molecule; whereby a hybrid nucleic acid is generated. In some embodiments, cleavage of step (ii) is effected using an enzyme that cleaves RNA from an RNA/DNA hybrid. In some embodiments, the enzyme is RNaseH.

In one aspect, the invention provides methods for an aspect leading to generating a complex of first and second primer extension product said complex comprising a 3' single stranded portion useful for recombining nucleic acids, said methods comprising: preparing a complex of first and second primer extension products comprising two 3' single stranded portions, wherein said complex of first and second primer extension products is prepared according to a method comprising: (i) extending a first primer hybridized to a target RNA with an enzyme comprising RNA-dependent DNA polymerase activity, wherein the first primer is a composite primer comprising an RNA portion and a 3' DNA portion, whereby a complex comprising a first primer extension product and the target RNA is produced; (ii) cleaving RNA in the complex of step (i) (cleavage may be effected, for example, with an enzyme that cleaves RNA from an RNA/DNA hybrid); (iii) extending a second primer hybridized to the first primer extension product with an enzyme comprising DNA-dependent DNA polymerase activity and RNA-dependent polymerase activity, wherein the second primer is a composite primer comprising a RNA portion and a 3' DNA portion, whereby a second primer extension product is produced to form a complex of first and second primer extension products; (iv) cleaving RNA from the composite primers in the complex of step (iii) with an enzyme that cleaves RNA from an RNA/DNA hybrid such that a complex comprising hybridized first and second primer extension product comprising two 3' single stranded DNA portion is generated. In some embodiments, cleavage of step (ii) is effected using an enzyme that cleaves RNA from an RNA/DNA hybrid. In some embodiments, the enzyme is RNaseH.

In another aspect, the invention provides methods for an aspect leading to recombining nucleic acids (making a hybrid), said methods comprising: hybridizing a 3' single stranded portion of a complex of first and second primer extension products with a recipient nucleic acid molecule, whereby a hybrid nucleic acid is generated.

In one aspect, the invention provides a method for generating a hybrid nucleic acid, said method comprising: (a) preparing a complex of first and second primer extension

IDT Exhibit 1020.9950

products said complex comprising a 3' single stranded portion, wherein said complex comprising a 3' single stranded portion is prepared according to a method comprising: (i) extending a first primer hybridized to a target RNA with an enzyme comprising RNA-dependent DNA polymerase activity, wherein the first primer is a composite primer comprising an RNA portion and a 3' DNA portion, whereby a complex comprising a first primer extension product and the target RNA is produced; (ii) cleaving RNA in the complex of step (i) ; (iii) extending a second primer hybridized to the first primer extension product with an enzyme comprising DNA-dependent DNA polymerase activity and an enzyme comprising RNA-dependent DNA polymerase activity, whereby a second primer extension product is produced to form a complex of first and second primer extension products; (iv) cleaving RNA from the composite primer in the complex of step (iii) with an enzyme that cleaves RNA from an RNA/DNA hybrid such that a complex comprising hybridized first and second primer extension product comprising a 3' single stranded DNA portion is generated; and (b) hybridizing the 3' single stranded portion of the complex of first and second primer extension products with a recipient nucleic acid molecule; whereby a hybrid nucleic acid is generated. As described herein, a portion or the entire 3' singled stranded portion can hybridize, as long as annealing between complex and recipient molecule occurs. In some embodiments, the methods further comprise (c) generating a recombined nucleic acid from the hybrid nucleic acid. In some embodiments, cleavage of step (ii) is effected using an enzyme that cleaves RNA from an RNA/DNA hybrid. In some embodiments, the enzyme is RNaseH.

In some embodiments of the methods described herein, the second primer comprises a fragment of the target RNA hybridized to the primer extension product, said fragment generated by cleaving RNA in the complex of step (a). Cleavage may be effected by a number of ways, including with an enzyme that cleaves RNA from an RNA/DNA hybrid, heat treatment, or chemical treatment.

In some embodiments, the methods are practiced using a composite primer in which a 5'RNA portion is adjacent to a 3' DNA portion, and making second strand is effected by using a second primer that comprises a fragment of the target RNA hybridized to the primer extension product (which was generated by RNaseH cleavage of the RNA/DNA complex resulting from extension of the composite primer).

In one aspect, the invention provides methods for recombining nucleic acids, said methods comprising: (a) preparing a complex of first and second primer extension

IDT Exhibit 1020.9951

products, said complex comprising two 3' single stranded portions, wherein said complex of first and second primer extension products is prepared according to a method comprising: (i) extending a first primer hybridized to a target RNA with an enzyme comprising RNA-dependent DNA polymerase activity, wherein the first primer is a composite primer comprising an RNA portion and a 3' DNA portion, whereby a complex comprising a first primer extension product and the target RNA is produced; (ii) cleaving RNA in the complex of step (i) (cleavage may be effected, for example, with an enzyme that cleaves RNA from an RNA/DNA hybrid); (iii) extending a second primer hybridized to the first primer extension product with an enzyme comprising DNA-dependent DNA polymerase activity and RNA-dependent DNA polymerase activity, wherein the second primer is a composite primer comprising a RNA portion and a 3' DNA portion, whereby a second primer extension product is produced to form a complex of first and second primer extension products; (iv) cleaving RNA from the composite primers in the complex of step (iii) with an enzyme that cleaves RNA from an RNA/DNA hybrid such that a complex comprising hybridized first and second primer extension product comprising two 3' single stranded DNA portion is generated; (b) hybridizing at least one 3' single stranded portion of a complex of first and second primer extension products with a recipient nucleic acid molecule; whereby a hybrid nucleic acid is generated; and (c) generating a recombined nucleic acid from the hybrid nucleic acid. In some embodiments, cleavage of step (ii) is effected using an enzyme that cleaves RNA from an RNA/DNA hybrid. In some embodiments, the enzyme is RNaseH.

In another aspect, the invention provides methods for recombining nucleic acids, said methods comprising: (a) hybridizing a 3' single stranded portion of a complex of first and second primer extension products with a recipient nucleic acid molecule, whereby a hybrid nucleic acid is generated, wherein said complex of first and second primer extension products is prepared according to a method comprising: (a) preparing a complex of first and second primer extension products comprising two 3' single stranded portions, wherein said complex of first and second primer extension products is prepared according to the following steps: (i) extending a first primer hybridized to a target RNA with an enzyme comprising RNA-dependent DNA polymerase activity, wherein the first primer is a composite primer comprising an RNA portion and a 3' DNA portion, whereby a complex comprising a first primer extension product and the target RNA is produced; (ii) cleaving RNA in the complex of step (i) (cleavage may be effected, for example, with an enzyme

IDT Exhibit 1020.9952

that cleaves RNA from an RNA/DNA hybrid); (iii) extending a second primer hybridized to the first primer extension product with an enzyme comprising DNA-dependent DNA polymerase activity and an enzyme comprising RNA-dependent DNA polymerase activity, wherein the second primer is a composite primer comprising a RNA portion and a 3' DNA portion, whereby a second primer extension product is produced to form a complex of first and second primer extension products; (iv) cleaving RNA from the composite primers in the complex of step (iii) with an enzyme that cleaves RNA from an RNA/DNA hybrid such that a complex comprising hybridized first and second primer extension product comprising two 3' single stranded DNA portions is generated; and (b) generating a recombined nucleic acid from the hybrid nucleic acid. In some embodiments, cleavage of step (ii) is effected using an enzyme that cleaves RNA from an RNA/DNA hybrid. In some embodiments, the enzyme is RNaseH.

In one aspect, the invention provides methods for recombining nucleic acids, said methods comprising: generating a recombined nucleic acid from a hybrid nucleic acid, said hybrid molecule prepared by hybridizing a 3' single stranded portion of a complex of first and second primer extension products with a recipient nucleic acid molecule; whereby the hybrid nucleic acid is generated; wherein said complex of first and second primer extension products is prepared according to a method comprising: (i) extending a first primer hybridized to a target RNA with an enzyme comprising RNA-dependent DNA polymerase activity, wherein the first primer is a composite primer comprising an RNA portion and a 3' DNA portion, whereby a complex comprising a first primer extension product and the target RNA is produced; (ii) cleaving RNA in the complex of step (i) (cleavage may be effected, for example, with an enzyme that cleaves RNA from an RNA/DNA hybrid); (iii) extending a second primer hybridized to the first primer extension product with an enzyme comprising DNA-dependent DNA polymerase activity and an enzyme comprising RNA-dependent DNA polymerase activity, wherein the second primer is a composite primer comprising a RNA portion and a 3' DNA portion, whereby a second primer extension product is produced to form a complex of first and second primer extension products; (iv) cleaving RNA from the composite primers in the complex of step (iii) with an enzyme that cleaves RNA from an RNA/DNA hybrid such that a complex comprising hybridized first and second primer extension product comprising two 3' single stranded DNA portion is generated. In some embodiments, cleavage of step (ii) is effected using an enzyme that cleaves RNA from an RNA/DNA hybrid. In some embodiments, the enzyme is RNaseH.

IDT Exhibit 1020.9953

In another aspect, the invention provides methods for selection and/or screening recombined nucleic acids for, e.g., desired phenotype and/or encoding a protein having an advantageous predetermined property. Recombined products are useful for, e.g., methods for molecular evolution, DNA shuffling, and other methods utilizing recombined nucleic acids. Immobilized recombined products are useful, e.g., for generating microarrays.

In another aspect, the invention provides the products described herein (complexes with the 3' single stranded portion; hybrid nucleic acids; recombined nucleic acids) as made by the process(es) described herein. For example, the invention provides a hybrid nucleic acid made by a process comprising (a) preparing a complex of first and second primer extension products said complex comprising a 3' single stranded portion, wherein said complex comprising a 3' single stranded portion is prepared according to a method comprising: (i) extending a first primer hybridized to a target RNA with an enzyme comprising RNA-dependent DNA polymerase activity, wherein the first primer is a composite primer comprising an RNA portion and a 3' DNA portion, whereby a complex comprising a first primer extension product and the target RNA is produced; (ii) cleaving RNA in the complex of step (i) ; (iii) extending a second primer hybridized to the first primer extension product with an enzyme comprising DNA-dependent DNA polymerase activity and an enzyme comprising RNA-dependent DNA polymerase activity, whereby a second primer extension product is produced to form a complex of first and second primer extension products; (iv) cleaving RNA from the composite primer in the complex of step (iii) with an enzyme that cleaves RNA from an RNA/DNA hybrid such that a complex comprising hybridized first and second primer extension product comprising a 3' single stranded DNA portion is generated; and (b) hybridizing the 3' single stranded portion of the complex of first and second primer extension products with a recipient nucleic acid molecule; whereby a hybrid nucleic acid is generated. As described herein, a portion or the entire 3' singled stranded portion can hybridize, as long as annealing between complex and recipient molecule occurs. In some embodiments, cleavage of step (ii) is effected using an enzyme that cleaves RNA from an RNA/DNA hybrid. In some embodiments, the enzyme is RNaseH.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1A: shows a diagrammatic representation of an isothermal (avoiding thermocycling) process using a composite primer and a second primer to generate from

IDT Exhibit 1020.9954

target RNA a complex of first and second primer extension products comprising one 3' single stranded DNA portion (overhang).

Figure 1B: shows a diagrammatic representation of a process using a composite primer and a fragment of the RNA template to generate from target RNA a complex of first and second primer extension products comprising one 3' single stranded DNA portion (overhang). The figure illustrates using an enzyme that cleaves RNA from an RNA/DNA hybrid (here, RNaseH) to effect cleavage, although other means to effect cleavage as known in the art and described herein can be used.

Figures 2A and B: show a diagrammatic representation of an isothermal process using two composite primers to generate from target RNA a complex of first and second primer extension products comprising two 3' single stranded DNA portions (overhangs).

Figure 3: shows a diagrammatic representation of a method using a complex of first and second primer extension products comprising a 3' overhang prepared according to methods of the invention.

Figure 4: shows a diagrammatic representation of a method using complex of first and second primer extension products comprising a 3' overhang prepared according to methods of invention.

Figure 5: shows a diagrammatic representation of a method using complex of first and second primer extension products comprising a 3' overhang prepared according to methods of invention.

Figure 6: shows a diagrammatic representation of a method using an immobilized nucleic acid and complex of first and second primer extension products comprising a 3' overhang prepared according to methods of invention.

MODES FOR CARRYING OUT THE INVENTION

Methods of the invention

The methods of the invention are directed to generating hybrid nucleic acids and recombined nucleic acids using a complex with a 3' overhang (single stranded portion). The methods are convenient in that generating the complex (which arises from RNA) does not require thermocycling (e.g., PCR); as such, these methods are sometimes generally referred to being "isothermal", although the term does not indicate that any and all of the reaction temperatures must be the same (although, as described herein, for a given step, or

IDT Exhibit 1020.9955

even for a series of two or more steps, the reaction temperature may be (but need not be) the same). The methods of the invention are also directed to generating the complexes, which are useful in making hybrid and recombinant nucleic acid molecules. The methods involve generating a complex of first and second primer extension products comprising at least one 3' single stranded DNA portion (overhang), which is useful as a substrate for subsequent reactions including hybridization with a recipient molecule, whereby a hybrid nucleic acid is produced, and generation of a recombined nucleic acid (polynucleotide) from the hybrid nucleic acid.

The methods generally comprise using specially-designed primers, generally one or more RNA/DNA composite primers, to generate a partially double stranded complex of first and second strand cDNA comprising a 3' DNA single stranded end (overhang) of defined or random sequence, from RNA template. The 3' single stranded portion generally comprises the complement of an RNA portion of the composite primer. The 3' single stranded portion is the substrate for subsequent steps of the methods, such as recombination.

Generally, the complex is created as follows: a complex of first and second primer extension products comprising an RNA/DNA heteroduplex is generated from an RNA template as follows: (i) extending a first composite primer hybridized to a target RNA with an enzyme comprising RNA-dependent DNA polymerase activity, whereby a complex comprising a first primer extension product and the target RNA is produced; (ii) cleaving RNA in the complex of step (i) ; (iii) extending a second primer hybridized to the first primer extension product with an enzyme comprising DNA-dependent DNA polymerase activity and an enzyme comprising RNA-dependent DNA polymerase activity, whereby a second primer extension product is produced to form a complex of first and second primer extension products comprising an RNA/DNA heteroduplex at an end. An agent which cleaves RNA from an RNA/DNA hybrid (such as RNase H) cleaves RNA from the RNA/DNA hybrid, leaving a 3' single-stranded portion of the second primer extension product (an overhang) available as a substrate for further steps of the methods of the invention. In one embodiment, the second primer that primes second primer extension product synthesis is a fragment of the RNA template. As used herein, "double stranded cDNA" refers to the complex of first and second primer extension product, and "cDNA" is a polynucleotide as defined herein. Accordingly, the methods of the invention may be used for generation of a collection of partially double stranded complexes comprising sequences

IDT Exhibit 1020.9956

from a plurality of RNA sequences in the sample, each of said complexes comprising a 3' single stranded portion that is the complement of the 5' RNA portion of the composite primer.

The composite primer according to the methods of the invention comprises a 3'-DNA portion that generally is designed to be hybridizable to a target RNA(s). The sequence that is hybridizable to the RNA template defines the RNA template from which the partially double stranded complexes are generated. For example, the composite primer can comprise a poly-dT sequence, which would be expected to hybridize to the poly-A tails of all mRNA is a sample, or the composite primer can generally comprise at least a 3' portion that is hybridizable to random sequences such that the random primer portion would hybridize to random portion(s) of the template. In some embodiments, this primer further comprises a 5' portion that does not hybridize to the target mRNA under conditions which the composite primer hybridizes to the target mRNA. In another aspect, the methods of the invention may be used for generation of partially double stranded complexes from a specific RNA species or class of RNA species (e.g., a family or superfamily of RNA species). In this latter case, the composite primer generally comprises a 3'-portion which is complementary to a sequence of a specific RNA target (or family of RNA targets). In some embodiments, a portion(s) (such as a 5' RNA portion) of the composite primer can comprise a sequence that is not hybridizable (does not hybridize) to a target RNA, such as mRNA, (which would constitute a tail when the primer is bound to a target RNA) under conditions which the composit primer hbridizes to the target RNA . This non-hybridized tail is sometimes referred to as a "defined" sequence.

Target RNA can be any RNA, including translatable or untranslated RNA. RNA target may also be generated from any DNA source, including genomic DNA, using methods known in the art, including Kurn, U.S. Patent No. 6,251,639 B1. In some embodiments, target RNA is prepared by transcription of DNA template to create the RNA complement of an DNA template (including genomic DNA and cDNA). See Kurn, U.S. Patent No. 6,251,639 B1; Kurn, US patent publication no. 2002/0058270 A1, Thus, partially double stranded DNA comprising a 3' single stranded portion prepared according to the methods described herein can be generated from RNA copies of genomic DNA, and would thus include untranscribed sequences generally not found in mRNA, such as introns, regulatory and control elements, pseudo-genes, etc. As is evident, partially double stranded DNA comprising a 3' single stranded portion can be generated from untranslated RNA

IDT Exhibit 1020.9957

sequences, including tRNA, ribosomal RNA, small untranslated RNAs, untranslated RNA involved in disease mechanisms, and untranslated RNA that functions in gene regulation. Partially double stranded DNA comprising a 3' single stranded portion can also be generated from exogenous elements present within a genome, such as viral sequences.

As noted above, the 3' single stranded portion of the partially double stranded DNA, which is the complement of the RNA portion of the composite primer, serves as a template for further steps in the methods of the invention, for example, hybridizing at least one single stranded 3' portion of a double stranded DNA with a recipient molecule (which can be single or double stranded DNA, and can comprise a double stranded DNA comprising a single stranded 3' portion generated according to the methods of the invention described herein) whereby a hybrid nucleic acid is generated, and generation of a recombined nucleic acid from the hybrid nucleic acid. In some embodiments, the 3' single stranded portion may comprise a defined sequence (i.e., a known sequence, which is generally a sequence not hybridizable to a template nucleic acid). The hybrid molecule has a nick (created adjacent hybridized 3' end of a partially double stranded complex and 5' end of the recipient molecule, or the converse) or a gap, created as when, for example, the length of hybridizable sequence is less than the length of the 3' overhang, or when the nucleotides at the 3' end of duplex and/or 5' end of recipient molecule are not hybridized. Thus, in some embodiments, the entire 3' single stranded portion hybridizes to the recipient molecule. In other embodiments, a portion of the 3' single stranded portion (overhang) hybridizes to the recipient molecule. Thus, reference to hybridizing a singled stranded 3' portion of the complexes described herein refers to hybridization of all or a part (portion) of the single stranded 3' portion, as long as requisite annealing is accomplished. In some instances, where recombined product is produced, the extent of hybridization can be minimal, if, for example, other functions to effect recombination (joining) of nucleic acids are present (such as a polymerase and/or ligase).

The invention provides methods for recombination of hybrid polynucleotides. Recombination can be according to the methods of the invention described herein. It is also contemplated that any method known in the art can be used. Recombined polynucleotides are useful, for example, for screening and/or selecting for desirable properties, as further described herein. As used herein, and as known in this art, "recombined product" refers to a continuous (fully linked) polynucleotide comprising

IDT Exhibit 1020.9958

sequences corresponding to the partially double stranded DNA comprising a 3' single stranded portion and the recipient molecule.

Thus, one of the major advantages of the methods of the invention is the ability to create partially double stranded complexes comprising a 3' single stranded portion that include sequences from a pool of mRNAs (or other RNA targets) each connected to a defined or random 3' overhang. Depending on primer design, complexes may be generated from a specific RNA sequence of interest, or a multiplicity of RNAs, for example, members of a family of related RNAs. The methods of the invention also are suitable for generating complexes from a large multiplicity of RNAs, including all mRNA sequences in a sample. In some embodiments, double stranded complexes comprising sequences from a pool of mRNAs is generated using a single composite primer.

As noted above, the methods of the invention do not require thermocycling and all of the steps can be performed isothermally, although the various steps may be carried out at different temperatures. This feature provides numerous advantages, including facilitating automation and adaptation for high through-put procedures. The isothermal reaction is faster than that afforded by thermal cycling and is suitable for performing the methods of the invention in miniaturized devices.

As used herein, a "cDNA" or "double stranded DNA" generally is (i.e., encompasses) a polynucleotide(s). It is clear from the description herein that, although the term "polynucleotide" can refer to and include an RNA, the term "polynucleotide" is determined by the context to which it refers. For example, extension by a DNA dependent DNA polymerase generally refers to generation of a DNA molecule, but the term "DNA" is not meant (in this context, for example) to exclude molecules that contain other types of bases, analogs, linkages, etc., such as, any substrate that can be incorporated into a polymer by the polymerase Similarly, as used herein, "double stranded complexes", "DNA partial duplexes", "complex comprising a 3' single stranded portion", "double stranded DNA comprising a 3' single stranded region", "double stranded cDNA comprising a 3' overhang", and "hybridized first and second primer extension product comprising a 3' single stranded DNA portion" are used interchangeably, and refer to (encompass) polynucleotides.

General Techniques

The practice of the invention will employ, unless otherwise indicated, conventional techniques of molecular biology (including recombinant techniques), microbiology, cell

IDT Exhibit 1020.9959

biology, biochemistry, and immunology, which are within the skill of the art. Such techniques are explained fully in the literature, such as, "Molecular Cloning: A Laboratory Manual", second edition (Sambrook et al., 1989); "Oligonucleotide Synthesis" (M.J. Gait, ed., 1984); "Animal Cell Culture" (R.I. Freshney, ed., 1987); "Methods in Enzymology" (Academic Press, Inc.); "Current Protocols in Molecular Biology" (F.M. Ausubel et al., eds., 1987, and periodic updates); "PCR: The Polymerase Chain Reaction", (Mullis et al., eds., 1994).

Primers, oligonucleotides and polynucleotides employed in the invention can be generated using standard techniques known in the art.

Definitions

A "target sequence," "target nucleic acid," or "target RNA," as used herein, is a polynucleotide comprising a sequence of interest, for which generation of double stranded complex comprising a 3' single stranded portion is desired. The target sequence may be known or not known, in terms of its actual sequence. In some instances, the terms "target ," "template ," and variations thereof, are used interchangeably.

"Amplification," as used herein, generally refers to the process of producing multiple copies of a desired sequence. "Multiple copies" mean at least 2 copies. A "copy" does not necessarily mean perfect sequence complementarity or identity to the template sequence. For example, copies can include nucleotide analogs such as deoxyinosine, intentional sequence alterations (such as sequence alterations introduced through a primer comprising a sequence that is hybridizable, but not complementary, to the template, or a non-target sequence introduced through a primer), and/or sequence errors that occur during primer extension.

"Polynucleotide," or "nucleic acid," as used interchangeably herein, refer to polymers of nucleotides of any length, and include DNA and RNA. The nucleotides can be deoxyribonucleotides, ribonucleotides, modified nucleotides or bases, and/or their analogs, or any substrate that can be incorporated into a polymer by DNA or RNA polymerase. A polynucleotide may comprise modified nucleotides, such as methylated nucleotides and their analogs. If present, modification to the nucleotide structure may be imparted before or after assembly of the polymer. The sequence of nucleotides may be interrupted by non-nucleotide components. A polynucleotide may be further modified after polymerization,

IDT Exhibit 1020.9960

such as by conjugation with a labeling component. Other types of modifications include, for example, "caps", substitution of one or more of the naturally occurring nucleotides with an analog, internucleotide modifications such as, for example, those with uncharged linkages (e.g., methyl phosphonates, phosphotriesters, phosphoamidates, cabamates, etc.) and with charged linkages (e.g., phosphorothioates, phosphorodithioates, etc.), those containing pendant moieties, such as, for example, proteins (e.g., nucleases, toxins, antibodies, signal peptides, ply-L-lysine, etc.), those with intercalators (e.g., acridine, psoralen, etc.), those containing chelators (e.g., metals, radioactive metals, boron, oxidative metals, etc.), those containing alkylators, those with modified linkages (e.g., alpha anomeric nucleic acids, etc.), as well as unmodified forms of the polynucleotide(s). Further, any of the hydroxyl groups ordinarily present in the sugars may be replaced, for example, by phosphonate groups, phosphate groups, protected by standard protecting groups, or activated to prepare additional linkages to additional nucleotides, or may be conjugated to solid supports. The 5' and 3' terminal OH can be phosphorylated or substituted with amines or organic capping groups moieties of from 1 to 20 carbon atoms. Other hydroxyls may also be derivatized to standard protecting groups. Polynucleotides can also contain analogous forms of ribose or deoxyribose sugars that are generally known in the art, including, for example, 2'--O-methyl-, 2'-O-allyl, 2'-fluoro- or 2'-azido-ribose, carbocyclic sugar analogs, α-anomeric sugars, epimeric sugars such as arabinose, xyloses or lyxoses, pyranose sugars, furanose sugars, sedoheptuloses, acyclic analogs and abasic nucleoside analogs such as methyl riboside. One or more phosphodiester linkages may be replaced by alternative linking groups. These alternative linking groups include, but are not limited to, embodiments wherein phosphate is replaced by P(O)S("thioate"), P(S)S ("dithioate"), "(O)NR_2 ("amidate"), P(O)R, P(O)OR', CO or CH_2 ("formacetal"), in which each R or R' is independently H or substituted or unsubstituted alkyl (1-20 C) optionally containing an ether (-O-) linkage, aryl, alkenyl, cycloalkyl, cycloalkenyl or araldyl. Not all linkages in a polynucleotide need be identical. The preceding description applies to all polynucleotides referred to herein, including RNA and DNA.

A "labeled dNTP," or "labeled rNTP," as used herein, refers, respectively, to a dNTP or rNTP, or analogs thereof, that is directly or indirectly attached with a label. For example, a "labeled" dNTP or rNTP, may be directly labeled with, for example, a dye and/or a detectable moiety, such as a member of a specific binding pair (such as biotin-avidin). A "labeled" dNTP or rNTP, may also be indirectly labeled by its attachment to,

IDT Exhibit 1020.9961

for example, a moiety to which a label is/can be attached. A dNTP or rNTP, may comprise a moiety (for example, an amine group) to which a label may be attached following incorporation of the dNTP or rNTP into an extension product. Useful labels in the present invention include digoxigenin, fluorescent dyes (e.g., fluorescein isothiocyanate, Texas red, rhodamine, green fluorescent protein and the like), radioisotopes (e.g., ^{3}H, ^{35}S, ^{32}P, ^{33}P, ^{125}I, or ^{14}C), enzymes (e.g., LacZ, horseradish peroxidase, alkaline phosphatase) and colorimetric labels such as colloidal gold or colored glass or plastic (e.g., polystyrene, polypropylene, latex, etc.) beads. Various anti-ligands and ligands can be used (as labels themselves or as a means for attaching a label). In the case of a ligand that has a natural anti-ligand, such as biotin, thyroxine and cortisol, the ligand can be used in conjunction with labeled anti-ligands.

The "type" of dNTP or rNTP, as used herein, refers to the particular base of a nucleotide, namely adenine, cytosine, thymine, uridine, or guanine.

"Oligonucleotide," as used herein, generally refers to short, generally single stranded, generally synthetic polynucleotides that are generally, but not necessarily, less than about 200 nucleotides in length. Oligonucleotides in the invention include the composite primer. The terms "oligonucleotide" and "polynucleotide" are not mutually exclusive. The description above for polynucleotides is equally and fully applicable to oligonucleotides.

A "primer," as used herein, refers to a nucleotide sequence (a polynucleotide), generally with a free 3'-OH group, that hybridizes with a template sequence (such as a target RNA, or a primer extension product) and is capable of promoting polymerization of a polynucleotide complementary to the template. A "primer" can be, for example, an oligonucleotide. It can also be, for example, a sequence of the template (such as a primer extension product or a fragment of the template created following RNase cleavage of a template-DNA complex) that is hybridized to a sequence in the template itself (for example, as a hairpin loop), and that is capable of promoting nucleotide polymerization. Thus, a primer can be an exogenous (e.g., added) primer or an endogenous (e.g., template fragment) primer.

A "random primer," as used herein, is a primer that comprises a sequence that is designed not necessarily based on a particular or specific sequence in a sample, but rather is based on a statistical expectation (or an empirical observation) that the sequence of the random primer is hybridizable (under a given set of conditions) to one or more sequences

IDT Exhibit 1020.9962

in the sample. The sequence of a random primer (or its complement) may or may not be naturally-occurring, or may or may not be present in a pool of sequences in a sample of interest. The generation of double stranded complex from a plurality of RNA species in a single reaction mixture would generally, but not necessarily, employ a multiplicity, preferably a large multiplicity, of random primers. As is well understood in the art, a "random primer" can also refer to a primer that is a member of a population of primers (a plurality of random primers) which collectively are designed to hybridize to a desired and/or a significant number of target sequences. A random primer may hybridize at a plurality of sites on a nucleic acid sequence. The use of random primers provides a method for generating primer extension products complementary to a target polynucleotide which does not require prior knowledge of the exact sequence of the target.

A "complex" is an assembly of components. A complex may or may not be stable and may be directly or indirectly detected. For example, as is described herein, given certain components of a reaction, and the type of product(s) of the reaction, existence of a complex can be inferred. For purposes of this invention, a complex is generally an intermediate with respect to the final hybridized nucleic acid and/or recombination product(s). An example of a complex is a nucleic acid duplex comprising a first primer extension product and a second primer extension product.

A "portion" or "region," used interchangeably herein, of a polynucleotide or oligonucleotide is a contiguous sequence of 2 or more bases. In other embodiments, a region or portion is at least about any of 3, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 60, or more contiguous nucleotides.

A region, portion, or sequence which is "adjacent" to another sequence directly abuts that region, portion, or sequence. For example, an RNA portion which is adjacent to a 5' DNA portion of a composite primer directly abuts that region. For an illustration of this example, see Figures 1A and 2A.

A "reaction mixture" is an assemblage of components, which, under suitable conditions, react to form a complex (which may be an intermediate) and/or a product(s).

"A", "an" and "the", and the like, unless otherwise indicated include plural forms. "A" fragment means one or more fragments.

In accordance with the well-established principle patent law, "comprising" means including.

IDT Exhibit 1020.9963

Conditions that "allow" an event to occur or conditions that are "suitable" for an event to occur, such as hybridization, strand extension, and the like, or "suitable" conditions are conditions that do not prevent such events from occurring. Thus, these conditions permit, enhance, facilitate, and/or are conducive to the event. Such conditions, known in the art and described herein, depend upon, for example, the nature of the nucleotide sequence, temperature, and buffer conditions. These conditions also depend on what event is desired, such as hybridization, cleavage, or strand extension, hybridization or recombining..

Sequence "mutation," as used herein, refers to any sequence alteration in a sequence of interest in comparison to a reference sequence. A reference sequence can be a wild type sequence or a sequence to which one wishes to compare a sequence of interest. A sequence mutation includes single nucleotide changes, or alterations of more than one nucleotide in a sequence, due to mechanisms such as substitution, deletion or insertion. Single nucleotide polymorphism (SNP) is also a sequence mutation as used herein. A sequence mutation can be naturally occurring in the target sequence or incorporated by in vitro manipulation of the sequence, for example, producing a copy of the sequence using a polymerase with low fidelity and other methods known in the art.

"Microarray" and "array," as used interchangeably herein, comprise a surface with an array, preferably ordered array, of putative binding (e.g., by hybridization) sites for a biochemical sample (target) which often has undetermined characteristics. In a preferred embodiment, a microarray refers to an assembly of distinct polynucleotide or oligonucleotide probes immobilized at defined positions on a substrate. Arrays are formed on substrates fabricated with materials such as paper, glass, plastic (e.g., polypropylene, nylon, polystyrene), polyacrylamide, nitrocellulose, silicon, optical fiber or any other suitable solid or semi-solid support, and configured in a planar (e.g., glass plates, silicon chips) or three-dimensional (e.g., pins, fibers, beads, particles, microtiter wells, capillaries) configuration. Probes forming the arrays may be attached to the substrate by any number of ways including (i) *in situ* synthesis (e.g., high-density oligonucleotide arrays) using photolithographic techniques (see, Fodor et al., *Science* (1991), 251:767-773; Pease et al., *Proc. Natl. Acad. Sci. U.S.A.* (1994), 91:5022-5026; Lockhart et al., *Nature Biotechnology* (1996), 14:1675; U.S. Pat. Nos. 5,578,832; 5,556,752; and 5,510,270); (ii) spotting/printing at medium to low-density (e.g., cDNA probes) on glass, nylon or nitrocellulose (Schena et al, *Science* (1995), 270:467-470, DeRisi et al, *Nature Genetics* (1996), 14:457-460; Shalon

IDT Exhibit 1020.9964

et al., *Genome Res.* (1996), 6:639-645; and Schena et al., *Proc. Natl. Acad. Sci. U.S.A.* (1995), 93:10539-11286); (iii) by masking (Maskos and Southern, *Nuc. Acids. Res.* (1992), 20:1679-1684) and (iv) by dot-blotting on a nylon or nitrocellulose hybridization membrane (see, e.g., Sambrook et al., Eds., 1989, Molecular Cloning: A Laboratory Manual, 2nd ed., Vol. 1-3, Cold Spring Harbor Laboratory (Cold Spring Harbor, N.Y.)). Probes may also be noncovalently immobilized on the substrate by hybridization to anchors, by means of magnetic beads, or in a fluid phase such as in microtiter wells or capillaries. The probe molecules are generally nucleic acids such as DNA, RNA, PNA, and cDNA but may also include proteins, polypeptides, oligosaccharides, cells, tissues and any permutations thereof which can specifically bind the target molecules.

The term "3'" generally refers to a region or position in a polynucleotide or oligonucleotide 3' (downstream) from another region or position in the same polynucleotide or oligonucleotide.

The term "5'" generally refers to a region or position in a polynucleotide or oligonucleotide 5' (upstream) from another region or position in the same polynucleotide or oligonucleotide.

The term "3'-DNA portion," "3'-DNA region," "3'-RNA portion," and "3'-RNA region," refer to the portion or region of a polynucleotide or oligonucleotide located towards the 3' end of the polynucleotide or oligonucleotide, and may or may not include the 3' most nucleotide(s) or moieties attached to the 3' most nucleotide of the same polynucleotide or oligonucleotide. The 3' most nucleotide(s) can be preferably from about 1 to about 50, more preferably from about 10 to about 40, even more preferably from about 20 to about 30 nucleotides.

The term "5'-DNA portion," "5'-DNA region," "5'-RNA portion," and "5'-RNA region," refer to the portion or region of a polynucleotide or oligonucleotide located towards the 5' end of the polynucleotide or oligonucleotide, and may or may not include the 5' most nucleotide(s) or moieties attached to the 5' most nucleotide of the same polynucleotide or oligonucleotide. The 5' most nucleotide(s) can be preferably from about 1 to about 50, more preferably from about 10 to about 40, even more preferably from about 20 to about 30 nucleotides.

As used herein, "hybrid nucleic acid molecule" refers to hybridized partially-double stranded complex and recipient molecule (which may be a second partially double-stranded complex made using the methods of the invention). A "recipient molecule" as used here is

IDT Exhibit 1020.9965

a polynucleotide to which a complex of first and second primer extension product comprising at least one 3' overhang hybridizes.

As used herein "recombined product" refers to a continuous (fully linked) polynucleotide comprising sequences corresponding to the partially double stranded complex comprising a 3' single stranded portion and the recipient molecule. A recombined product can, but does not necessarily, comprise all sequences corresponding to the partially double stranded complex comprising a 3' single stranded portion and the recipient molecule, as some of the methods for generating recombined product described herein result in deletion and/or mutation of sequences corresponding to the original partially double stranded complex (comprising a 3' overhang) and recipient molecule. "Recombining nucleic acids" refers to generating a recombined product. This can be accomplished in any of a number of ways (including ligation) as described herein.

Methods of the invention using partially double stranded complexes comprising a 3' overhang

The following are examples of the methods of the invention. It is understood that various other embodiments may be practiced, given the general description provided above. For example, reference to using a composite primer means that any of the composite primers described herein may be used.

The methods of the invention involve production of a complex comprising a partially double stranded complex comprising a 3' DNA single stranded portion (overhang) of defined or random sequence, from an RNA template. The 3' single stranded end of the complex (all or part) is hybridizable to a 3' single stranded portion of a second nucleic acid, termed a recipient nucleic acid ,which may be single stranded or a second partially double stranded complex comprising a 3' single stranded portion (which can be a second partial double stranded complex comprising a 3' single stranded overhang according to the methods of the invention described herein). Hybridization of the 3' single stranded portions of the complex and the second nucleic acid generates a hybrid (annealed) nucleic acid. A recombined nucleic acid (which can be immobilized to a surface) is generated from the hybrid nucleic acid by any of a number of methods, as known in the art and described herein.

IDT Exhibit 1020.9966

Uses of recombined products are known in the art and include subsequent screening and/or selection of recombined products, and further uses described herein. These uses are encompassed by the invention.

For simplicity, the methods for production of the partially double stranded complex comprising a 3' single stranded portion (overhang) are described separately from the methods for production of hybrid nucleic acids and generation of recombined product. It is understood that each method for production of the complex is applicable to each method for production of a hybrid nucleic acid and/or generation of recombined product, and vice versa.

Formation of a complex comprising a partially double stranded complex comprising a 3' single stranded sequence

Figure 1A illustrates a schematic description of one embodiment of the methods for formation of partially double stranded DNA comprising a 3' single stranded DNA sequence from an RNA template. The methods involve the following steps: (a) formation of a double stranded cDNA comprising a RNA-DNA heteroduplex at one end as follows: (i) extending a first composite primer hybridized to a target RNA with an enzyme comprising RNA-dependent DNA polymerase activity, whereby a complex comprising a first primer extension product and the target RNA is produced; (ii) cleaving RNA in the complex of step (i) ; (iii) extending a second primer hybridized to the first primer extension product with an enzyme comprising DNA-dependent DNA polymerase activity and an enzyme comprising RNA-dependent DNA polymerase activity, whereby a second primer extension product is produced to form a complex of first and second primer extension products comprising an RNA/DNA heteroduplex at an end.; and (b) cleavage of RNA in the RNA-DNA heteroduplex to generate a partially double stranded DNA comprising a 3' single stranded DNA portion (interchangeably termed "overhang"). As illustrated, all steps are isothermal, although the temperatures for each of the steps may or may not be the same. The methods of the invention do not require thermal cycling (i.e., PCR), and the complex of template RNA and first primer extension product is not denatured or separated using thermal denaturation.

The embodiment illustrated in Figures 1A employs two oligonucleotides: a composite primer, (labeled 1); and a second primer (labeled 2), used for the formation of the second primer extension product (interchangeably called "second strand cDNA").

IDT Exhibit 1020.9967

As illustrated in these figures, the composite primer 1 comprises a DNA portion, A, at its 3' end, and an RNA portion, B, at its 5' end. As discussed herein, it is also possible to employ a composite primer in which the 3' DNA portion is followed, in the direction of its 5', by an RNA portion, which is followed by a portion which is DNA. The length of each of these sections is generally determined for maximum efficiency of the formation of the partially double stranded complexes comprising a 3' single stranded DNA sequence, and the subsequence recombination mediated by the 3' single stranded end. Only the two-portion (i.e., 3'-DNA-RNA-5') composite primer is shown in Figures 1A, 1B and Figures 2A-B.

The 5' portion of the composite primer may or may not be hybridizable to target sequence. In some embodiments, the 5' portion of composite primer 1 as illustrated in Figures 1A and B, and 2A comprises a sequence not hybridizable to the target sequence, e.g., a sequence forming a "tail" when the primer is hybridized to a target. The 5' portion generally is incorporated into the first primer extension product (first strand cDNA), and the 3' overhang (single stranded DNA portion) of the complex of double stranded DNA comprises the complement of the 5' portion of the composite primer. . As described herein, the single stranded DNA portion is the substrate for subsequence hybridization and recombination. In another embodiment, the 5' portion of the first primer is hybridizable to the target RNA.The 5' portion of the first primer can be designed in any of a number of ways (in terms of sequence) depending on which type, class, population, and/or species of recombination is desired between the complex and the second nucleic acid. As noted above, length and sequence of the overhang in the resultant partial double stranded DNA complex is determined by the length and sequence of the RNA portion of the composite primer that binds to target RNA. The length is generally at least two ribonucleotides, although as discussed below, longer RNA portions are contemplated by the invention.

The sequence of the 5' portion can be known (termed a "defined sequence"), and in some embodiments, comprises a restriction site. The defined sequence can be selected to be hybridizable to a known sequence or a family of known sequences (for example, a kinase domain). In some embodiments, the 5' defined sequence can be a unique sequence, in the sense that the sequence is not expected to be hybridizable to template RNA. In other embodiments, the 5' portion is a random sequence. In still other embodiments, the 5' portion comprises a defined sequence and a random sequence. The sequence of the RNA portion can be selected to, for example, comprise sequences corresponding to (the

IDT Exhibit 1020.9968

complement of which are hybridizable to) a 3' single stranded overhang generated by restriction digest of double stranded DNA, thus permitting hybridization of the complex to a recipient molecule cut with the restriction enzyme. As used herein, sequences that "correspond" to a 3' single stranded are sequences that are generally hybridizable to the complement of the 3' single stranded overhang sequence.

The 3' overhang can also be selected to hybridize with recipient molecules (interchangeably called "recipient nucleic acids") comprising a defined 3' overhang region. The defined region may be a unique sequence, such that hybridization occurs generally with recipient molecules comprising the same unique sequence. Alternatively, the defined sequence can correspond to a sequence hybridizable to a sequence present in 3' end of recipient molecule, or a family or related recipient molecules (e.g., gene family or superfamily). In another embodiment, the 3' overhang is random, so that the complexity of the hybridization step is greatly increased (in terms of number of potentially hybridizable recipient molecules, for example).

The 3' overhangs can also be selected to generate recombination products with insertions of varying lengths (corresponding to sequences contained in the 3' overhangs). For example, if a partially double stranded complex comprising a 3' overhang is hybridized to a recipient molecule comprising a 3' unique sequence (in the sense that it is not a sequence present in target RNA, but rather constitutes an appended 3' sequence), nucleotides corresponding to the unique sequence will be inserted into the recombination product. In another example, if the 3' overhang corresponds to a random sequence in the first composite primer, generally the recombination product comprises nucleotides hybridizable to the random sequence.

However, if either 3' end is generated from a sequence present in the target RNA, hybridization generates a double stranded portion corresponding to target RNA, and new nucleotides are generally not inserted in the final recombined product.

It is understood that the 3' ends of the partially double stranded complex and recipient need not hybridize with perfect complementary, depending on the level of stringency used during the hybridization of complex and recipient molecule. Selection of suitable hybridization conditions is described further herein. It is appreciated that use of reduced stringency increases the number of 3' overhangs which will tend to hybridize with one another.

IDT Exhibit 1020.9969

In addition, the 3' ends of the partially double stranded complex and recipient molecules need not be of the same length. One or more gaps can be present after hybridization, which may be filled using a variety of methods described herein.

3' ends may also be selected such that hybridization between partially double stranded complex and recipient molecule is directional. For example, a population of partially double stranded complex molecules and a second population of recipient molecules (which may also be partially double stranded complex molecules) may be generated such that hybridization is in a particular, specified orientation with respect to each other.

As noted above, the use of 3' ends corresponding to random 5' RNA portions would be expected to permit hybridization in all orientations (and thus to increase the complexity of recombination).

Further, the sequence of the 3' end may be selected such that hybridization is favored or disfavored, for example, increasing GC content of the 3' end, and using nucleotide analogs (including PNA) known to increase or decrease capacity to hybridize, as well as other factors known in the art.

The 3' portion of the first composite primer (which hybridizes to template RNA) can be designed in any of a number of ways (in terms of sequence), depending on which type, class, population, and/or species of RNA is desired to be incorporated into the partially double stranded DNA complex (that comprises a 3' single stranded region). In some embodiments, the 3' portion of composite primer 1 illustrated in Figures 1 and 2 comprises a sequence complementary to the poly-A tail of mRNA, and may further comprise additional random sequences (generally not complementary to a poly-A sequence) at the 3' end of the 3' portion. In other embodiments, the 3' portion of composite primer 1 is a random primer comprising sequences which are hybridizable to a multiplicity, a large multiplicity or a very large multiplicity of RNA species. Random primers are known in the art, for example, they have been used extensively in the preparation of cDNA libraries using PCR-based procedures. As is well understood in the art, a "random primer" can refer to a primer that is a member of a population of primers (a plurality of random primers) which collectively are designed to hybridize to a desired and/or a significant number of target sequences. A random primer may hybridize at a plurality of sites on a nucleic acid sequence. In other embodiments, the 3' portion of

IDT Exhibit 1020.9970

composite primer 1 can comprise a sequence hybridizable to a specific RNA or family of RNAs (or portions thereof).

In some embodiments, composite primer 1 is a mixture of different composite primers which comprise the same 5' RNA portion and a multiplicity of 3' DNA portions selected to amplify a multiplicity of RNA sequences of interest (which may range from 2 or more to many hundred or thousands or more).

The second primer is any sequence that is hybridizable to the first DNA strand such that it is capable of being extended by a DNA polymerase along a first primer extension product to create a second primer extension product. Primer 2 as illustrated in Figures 1A-B can be, but is not necessarily, composed of DNA and can comprise two sections (interchangeably called "portions" or "regions"). The 3' portion, F, of primer 2 can be selected for random priming of many, most and/or all possible mRNA sequences in a biological sample. The 5' portion, E, of primer 2 can be a sequence which is not complementary and not substantially hybridizable to a specific target sequence, i.e., it would not hybridize and would constitute a tail. The "tail" sequence would generally be incorporated into the second primer extension product. In other embodiments, the 5' end portion, E, of primer 2 can be hybridizable to the target sequence in the first primer extension product. In still other embodiments, it comprises a sequence of the first DNA strand (generally at the 3' end) that is hybridized to a sequence in the first DNA strand (for example, a hairpin or self-annealed structure).

In one embodiment, a fragment of the target RNA serves as the primer of the second primer extension product. This embodiment of the invention is illustrated in Figure 1B, and discussed further below. The target RNA in the initial complex comprising the target RNA and first primer extension product is cleaved such that at least one fragment of the template RNA remains hybridized to the first primer extension product. In some embodiments, the cleavage is effected using an agent such as an enzyme that effects cleavage of RNA from an RNA/DNA hybrid, such as RNaseH. In this aspect of the invention, one (or more) template RNA fragment(s) serves as a second "primer" in the manner described above, to generate a fragment extension product which has the same function as the second primer extension product in the methods described above. A suitable RNA fragment in the methods of the invention is long enough such that it does not dissociate from the first strand cDNA, preferably from about 3 to about 30, more preferably from about 5 to about 25, even more preferably from about 10 to about 20, and most

IDT Exhibit 1020.9971

preferably from about 12 to about 17, nucleotides in length. Template RNA can be cleaved using methods well-known in the art, including cleavage with an enzyme (such as RNAse H) that cleaves RNA from an RNA-DNA hybrid, cleavage resulting from heat treatment, and cleavage due to chemical treatment (e.g., treatment under alkaline conditions). As used herein, template cleavage does not encompass separation of template RNA and first primer extension using of heat that is sufficient to separate or denature template RNA from first primer extension product (for example, by using thermal cycler or other means of termperature controlled inclubation).

As illustrated in Figure 1A, in one embodiment, the process of the methods of the invention resulting in formation of partially double stranded complexes comprising a 3' single stranded DNA sequence from an RNA template is as follows:

1. Primer 1 binds to an RNA species in a sample by hybridization of the random sequence portion A (which can be based at least in part on the poly-A sequence of the mRNA), to form complex I (Figure 1A).
2. A reverse transcriptase, (indicated as "RT"), extends the hybridized primer 1 along the target RNA strand to which it is hybridized, to form an RNA/DNA duplex. An agent (such as RNase H) degrades the target RNA strand of the hybrid duplex to generate a single stranded first strand cDNA (labeled "II"). The 5' end of II is primer 1.
3. Primer 2 binds to the first strand cDNA, II, by hybridization of the random sequence F, to form complex III.
4. Primer 2 is extended along the first cDNA strand II by a DNA polymerase to form a double stranded product (labeled "IV"). Primer extension along the 5' RNA portion of II by an RNA-dependent DNA polymerase such as a reverse transcriptase results in formation of an RNA/DNA hybrid portion at one end of complex IV. DNA-dependent polymerase activity and RNA-dependent polymerase activity can be provided in the same enzyme.

RNase H degrades the RNA portion of the RNA/DNA hybrid at one end of complex IV, to create a partial double stranded complex (labeled "V") with a 3' DNA single stranded end, which has a sequence which is the complement of portion B of the composite primer 1. The RNase H activity may be supplied by the RNA-dependent DNA polymerase (such as reverse transcriptase) or may be provided in a separate enzyme. Reverse transcriptases useful for this method may or may not have RNase H activity. The DNA dependent DNA polymerase used for the second primer extension step may supply the

IDT Exhibit 1020.9972

reverse transcriptase activity for replication of the 5' RNA portion of the first primer extension product, or a second enzyme may provide it.

Figure 1B illustrates an embodiment of the invention wherein a fragment of the target RNA serves as the primer of the second primer extension product. As illustrated in Figure 1B, the single (composite) primer (labeled "1") can be composed of a 3'-DNA portion (labeled "A") which is complementary to a sequence on the target RNA, and a 5'-RNA portion (labeled "B") which comprises a non-target related sequence (i.e., it is not complementary or hybridizable under a given set of conditions to a sequence on the target RNA under which "A" does hybridize). The 3'-DNA portion of the composite primer may comprise poly-dT nucleotides, which would render it complementary/hybridizable (under a given set of conditions) to the poly-A 3' end of mRNA derived from a eukaryotic cell.

1. Composite primer 1 that is hybridized to a target RNA (complex "I" in Figure 1B) is extended by an RNA-dependent DNA polymerase, such as a reverse transcriptase, to form an RNA/DNA heteroduplex of the target RNA and a first strand cDNA (complex "II" in Figure 1B).
2. Target RNA is then degraded (fragmented), for example, by use of a ribonuclease such as RNase H, to form a complex of the first strand cDNA and one or more RNA fragments (oligonucleotides) (complex "III" in Figure 1B). The fragments are a result of incomplete degradation of the target RNA in the heteroduplex.
3. These fragments function as primers for a DNA-dependent DNA polymerase to form the second strand cDNA. Okayama & Berg, *Molecular and Cell Biology* (1982), 2:161; and Gubler & Hoffman, *Gene* (1983), 25:263. An RNA-dependent DNA polymerase (such as reverse transcriptase) then extends the 3'-end of the second strand cDNA in the duplex along the 5'-RNA sequence of the composite primer extension product (the first strand cDNA), to form an RNA/DNA heteroduplex at the end of the double stranded cDNA product (complex "IV" in Figure 1B).
4. The heteroduplex at the end of the double stranded cDNA is a substrate for RNase H. RNase H degrades the RNA portion of the RNA/DNA hybrid at one end of complex IV, to create a partial double stranded complex (labeled "V") with a 3' DNA single stranded end, which has a sequence which is the complement of portion B of the composite primer 1.

The methods of the invention may be used for generation of a "library" of partially double stranded complexes comprising sequences from a plurality of RNA sequences in the

IDT Exhibit 1020.9973

sample. For example, the composite primer can comprise a poly-dT sequence, which would be expected to hybridize to the poly-A tails of all mRNA is a sample, or the composite primer can generally comprise at least a 3' portion that is hybridizable to random sequences such that the random primer portion would hybridize to random portion(s) of the template. In another aspect, the methods of the invention may be used for generation of partially double stranded complexes from a specific RNA species or class of RNA species (e.g., a family or superfamily of RNA species). In this latter case, the composite primer generally comprises a 3'-portion which is complementary to a sequence of a specific RNA target (or family of RNA targets).

For convenience, only one first primer extension product (first strand cDNA) and second strand primer extension product (second strand cDNA) are described and illustrated in Figures 1A, 1B, and 2. It is to be understood that the methods of the invention are useful not only for generating partially double stranded complexes comprising double stranded copies of one RNA sequence, but also for generating complexes simultaneously from more than one different specific nucleic acid sequence located on the same or different targetRNA molecules.

As described herein, partially double stranded complexes can be generated from RNA target that is itself generated from a DNA source (such as genomic DNA), using methods known in the art, including Kurn, U.S. Patent No. 6,251,639. Thus, partially double stranded complexes prepared according to the methods described herein can correspond to RNA copies of genomic DNA, and would thus include untranscribed sequences generally not found in mRNA, such as introns, regulatory and control elements, etc. As is evident, partially double stranded DNA comprising a 3' single stranded portion can be generated from untranslated RNA sequences, including tRNA, ribosomal RNA, small untranslated RNAs, untranslated RNA involved in disease mechanisms, and untranslated RNA that functions in gene regulation.

Formation of a complex comprising a partially double stranded DNA comprising two 3' single stranded sequences

A schematic description of one embodiment of the formation of partially double stranded DNA comprising two 3' single stranded DNA sequences from an RNA template is given in Figure 2A and B. The methods involve (a) formation of a double stranded cDNA comprising a RNA-DNA heteroduplex at each end of the cDNA as follows: (i)

IDT Exhibit 1020.9974

extending a first composite primer hybridized to a target RNA with an enzyme comprising RNA-dependent DNA polymerase activity, whereby a complex comprising a first primer extension product and the target RNA is produced ; (ii) cleaving RNA in the complex of step (i) ; (iii) extending a second composite primer hybridized to the first primer extension product with an enzyme comprising DNA-dependent DNA polymerase activity and an enzyme comprising RNA-dependent DNA polymerase activity, whereby a second primer extension product is produced to form a complex of first and second primer extension products comprising an RNA/DNA heteroduplex at each end.; and (b) cleavage of RNA in the RNA-DNA heteroduplexes to generate a partially double stranded complex comprising two 3' single stranded DNA portions (or overhangs). As illustrated, all steps are isothermal, although the temperatures for each of the steps may or may not be the same. As is evident, in this aspect of the invention, production of the second strand cDNA is primed using another composite primer capable of hybridizing to the first strand cDNA. As used herein, the term "cDNA" refers to first and second primer extension product comprising polynucleotides.

The embodiment illustrated in Figures 2A-B employs two oligonucleotides: a first composite primer, (labeled 1) used for the formation of the first primer extension product; and a second composite primer (labeled 2), used for the formation of the second primer extension product. The 5' portion of the first composite primer is generally incorporated into the first primer extension product, and the 5' portion of the second (reverse) composite primer generally is incorporated into the second primer extension product (second strand cDNA). The complements of the 5' portion of the first and second composite primer comprise each of the two 3' single stranded portions of the complex of double stranded DNA.

In Figure 2A and B, two composite primers comprising "tail" sequences are used to generate a double stranded cDNA comprising a RNA-DNA heteroduplex at each end of the cDNA. Cleavage of RNA by an agent that cleaves RNA from an RNA/DNA heteroduplex creates a partially double stranded complex with two 3' DNA single stranded ends.

The 5' portion of composite primer 2 as illustrated in Figures 2A can comprise a sequence not hybridizable to the target sequence or to the first primer extension product, e.g., a sequence forming a "tail" when the primer is hybridized to the first primer extension product. In some embodiments, this 5' portion is a "defined" sequence or, in some

IDT Exhibit 1020.9975

embodiments, is a random sequence. The 5' portion of composite primer 2 can be the same as or different than the 5' portion of composite primer 1.

The process of the methods of the invention resulting in formation of partially double stranded complex comprising two 3' single stranded DNA portions from an RNA template is as follows:

1. Composite primer 1 binds to an RNA species in a sample by hybridization of sequence portion A (which can be based at least in part on the poly-A sequence of the mRNA), to form complex I. Composite primer 1 can, but does not necessarily, comprise a "tail" sequence (depicted as a thick black line) that is not hybridizable to the template mRNA.
2. A reverse transcriptase extends the hybridized primer 1 along the target RNA strand to which it is hybridized, to form an RNA/DNA duplex, labeled II. An agent (such as RNase H) degrades the target RNA strand of the hybrid duplex to generate a single stranded cDNA (labeled "III"). The 5' end of III is primer 1.
3. Composite primer 2 binds to the first strand cDNA, III, by hybridization of sequence F, to form complex IV. The reverse composite primer further comprises a "tail" sequence (depicted as a thick dotted line) that is not hybridizable to the template RNA or the first strand cDNA. The tail sequences of composite primer 1 and the second composite primer can be different sequences, although in some embodiments, the tail sequences can be the same sequence.
4. The composite primer 2 is extended along the first strand cDNA III by a DNA polymerase to form a double stranded product (labeled "V"). Primer extension along the 5' RNA portion of IV by an RNA-dependent DNA polymerase such as a reverse transcriptase results in formation of an RNA/DNA hybrid portion at one end of complex V. DNA-dependent polymerase activity and RNA-dependent polymerase activity can be provided in the same enzyme, as described herein.
5. An agent (such as RNase H) degrades the RNA portion of the RNA/DNA hybrid at one end of complex V, to create a partial double stranded complex (labeled "VI") with a 3' DNA single stranded end, which has a sequence which is the complement of portion B of the composite primer 1.
6. Composite primer 1 binds to complex VI by hybridization of the RNA portion to the single stranded DNA end, which is complementary to it, to form complex VII

IDT Exhibit 1020.9976

7. Primer extension of bound primer 1 in complex VII along the sense cDNA strand results in displacement of the previous cDNA product (VII), and replicates the tail sequence of the second composite primer, to form complex VIII.
8. Complex VIII has two RNA/DNA heteroduplexes comprised of composite primer 1 and the complement of composite primer 1 at one end, and composite primer 2 and the complement of the composite primer 2 at the other end. An agent (such as RNase H) cleaves the RNA portions of the RNA/DNA heteroduplexes, forming a partially double stranded DNA comprising two 3' single stranded DNA sequences. The RNase H activity may be supplied by the enzyme comprising RNA-dependent DNA polymerase activity (such as reverse transcriptase) or may be provided in a separate enzyme. Reverse transcriptases useful for this method may or may not have RNase H activity. One 3' DNA single stranded end has a sequence which is the complement of portion B of the composite primer 1, and one 3' DNA single stranded end has a sequence which is the complement of portion B of composite primer 2.

Design of the composite primers of these embodiments of the invention is generally governed by the same consideration described above for the methods of generating a partially double stranded complex with a single 3' end. However, it is understood that the composite primer that binds to target RNA and the composite primer that binds to first primer extension product generally have different 5' RNA portions. The 5' RNA portions can be selected such that the resultant 3' overhangs are mutually hybridizable (e.g., by designing hybridizable sequences or using random sequence in one or both 5' portion), or designed such that the 3' overhangs are hybridizable with one or more recipient molecules.

Methods using the partially double stranded complex comprising a 3' single stranded portion

The invention provides methods for aspects of recombining nucleic acids. In some embodiments, the methods comprise hybridizing a 3' single stranded portion of a partially double stranded DNA with a recipient nucleic acid molecule whereby a hybrid nucleic acid is generated; said partially double stranded complex comprising a 3' single stranded portion prepared according to any of the methods described herein. In another aspect, the invention provides methods for recombining nucleic acids comprising: hybridizing a 3' single stranded portion of a partially double stranded complex with a recipient nucleic acid

IDT Exhibit 1020.9977

molecule whereby a hybrid nucleic acid is generated (as noted above and herein, all or a part (portion) of the 3' single stranded portion can hybridize); said partially double stranded complex comprising a 3' single stranded portion prepared according to any of the methods described herein, and generating a recombined nucleic acid from the hybrid nucleic acid. As used herein "recombined product" refers to a continuous (fully linked) polynucleotide comprising sequences corresponding to the partially double stranded DNA comprising a 3' single stranded portion and the recipient molecule. A recombined product can, but does not necessarily, comprise all sequences corresponding to the partially double stranded DNA comprising a 3' single stranded portion and the recipient molecule, as some of the methods for generating recombined product described herein result in deletion and/or mutation of sequences corresponding to the original partially double stranded DNA (comprising a 3' overhang) and/or recipient molecule. As used herein, a "cDNA" or "partially double stranded DNA" or "double stranded DNA" is (i.e., encompasses) a polynucleotide(s). As noted above, tt is clear from the description herein that, although the term "polynucleotide" can refer to and include an RNA, the term "polynucleotide" is determined by the context to which it refers. For example, extension by a DNA dependent DNA polymerase generally refers to generation of a DNA molecule, but the term "DNA" is not meant (in this context, for example) to exclude molecules that contain other types of bases, analogs, linkages, etc., such as, any substrate that can be incorporated into a polymer by the polymerase

Methods for producing a hybrid nucleic acid The partially double stranded DNA comprising a 3' single stranded portion (interchangeably called "partially double stranded complex") is prepared according to any of the methods of the invention described herein. The partially double stranded complex can comprise a single 3' overhang or two 3' overhangs (one at each end). It is understood that partially double stranded complexes of the invention can be a multiplicity (from few to very many) of different partially double stranded complexes (corresponding to a multiplicity of target RNAs) generated according to any of the methods described herein. Methods and approaches for generating populations of different partially double stranded complexes are described herein. Such populations can be related in sequence (e.g., member of a gene family; comprising sequence corresponding to a certain motif or domain) or extremely diverse in sequence (e.g., generated from all mRNA using a poly-dT or random composite primer; generated from all RNA in a target population using a random primer). As described herein, partially

IDT Exhibit 1020.9978

double stranded complexes can be generated from RNA target that is itself generated from a DNA source (such as genomic DNA), using methods known in the art, including Kurn, U.S. Patent No. 6,251,639 B1. Thus, partially double stranded complexes prepared according to the methods described herein can be generated from RNA copies of genomic DNA, and would thus include untranscribed sequences generally not found in mRNA, such as introns, regulatory and control elements, etc.

A recipient molecule can (interchangeably called "recipient molecule" or "recipient") be partially double stranded comprising one or two 3' single stranded portions (3' overhangs). Double stranded recipient molecule can be a second partially double stranded polynucleotide (such as DNA) comprising one or two 3' single stranded portion(s) prepared according to any of the methods of the invention described herein. Other sources for partially double stranded recipient (comprising one or two 3' overhangs) include at least the following: double stranded cDNA, double stranded genomic DNA (from one or more genomic source), double stranded DNA prepared from a library, and double stranded DNA prepared using any method known in the art, including PCR (see, e.g., Jarrell WO 00/40715). Methods for generating partially double stranded DNA comprising one or two 3' (single stranded) overhangs are known in the art and include restriction digestion with a suitable restriction enzyme, Dnase digestion, mechanical shearing of double stranded DNA, and additional methods described in, for example, U.S. Patent Nos. 6,344,356; 6,319,714.

The recipient nucleic acid molecule can be single stranded. In some embodiments, the single stranded recipient is immobilized on a surface (as further described below). Single stranded recipient may be prepared from at least the following sources: single stranded DNA prepared using any method known in the art, including methods described in Kurn, U.S. Patent No. 6,251,639; Kurn, U.S. Patent Publication No. 2002/0058270; Kurn, co-pending U.S. patent application No. 10/100,321; first strand cDNA prepared using method known in the art; as well as from any double stranded DNA or DNA/RNA hybrid molecule simply by renderingthe DNA single stranded using methods well known in the art.

It is understood that a recipient molecule can be a multiplicity (from small to very large) of different recipient molecules. Such populations can be related in sequence (e.g., member of a gene family; comprising sequence corresponding to a certain motif or domain) or extremely diverse in sequence (e.g., generated from all mRNA, generated from a population of genomic DNA, etc.). Alternatively, recipient molecules can correspond to

IDT Exhibit 1020.9979

single sequence (which can be part or all of a known gene, for example a coding region, genomic portion, etc.)

Recipient molecules can be selected from one or more sources described herein. For example, recipient molecules can include recipient molecules that are partially double stranded DNA comprising a 3' overhang prepared according to methods described herein and other single stranded and/or double stranded recipient molecules.

Moreover, it is understood that partially double stranded complexes (comprising 3' overhang) and/or recipient molecules may correspond to RNA or DNA from one or a multiplicity (which can be small to very large) of organisms, and may be native or non-native (e.g., mutated, synthetic, etc.) DNA or RNA. Thus, the notion of a "gene family" (or related nucleic acid sequences) as used herein encompasses both related sequences from a particular organisms as well as sequences related across a multiplicity of organisms.

Partially double stranded complexes and/or recipient molecules can be selected in any of a number of ways depending on which type, class, population, and/or species of nucleic acid is desired to be incorporated into a hybrid molecule (i.e., following hybridization of 3' ends of a partially double stranded complex and a recipient molecule). For example, nucleic acid corresponding to a particular protein, protein domain, protein subdomain, protein motif, protein family or superfamily, etc. may be selected. In an additional example, nucleic acid corresponding to an exon, open reading frame, exon-intron region, and/or untranslated region (or any combination thereof) may also be selected. In another example, nucleic acids may correspond to randomly generated nucleic acids. In yet another example, nucleic acid can correspond to untranslated RNA sequences. These examples are further discussed herein.

Methods for selecting type, class, population, and/or species of partially double stranded DNA comprising a 3' overhang are described herein. Generally, the 3' portion of the first composite primer is designed in any of a number of ways (in terms of sequence), depending on which type, class, population, and/or species of RNA is desired to be incorporate into the partially double stranded complex (that comprises a 3' single stranded region), as discussed herein.

Methods for selecting type, class, population and/or species of recipient molecule are well known in the art, and include: amplification of desired sequence(s) using PCR; restriction endonuclease digestion of DNA, including genomic DNA; preparation of previously cloned DNA; and other well known methods in the art. It is understood that

IDT Exhibit 1020.9980

recipient molecules may be prepared from a variety of sources including: RNA, DNA, RNA/DNA hybrids; single stranded DNA; double stranded DNA, etc.

The following discussion describes certain exemplary nucleic acids which may be represented in a partially double stranded complex comprising a 3' overhang (prepared according to methods of the invention) and/or recipient molecule. It is understood that partially double stranded complexes and recipient molecules may comprise one or more of the features described below.

In some embodiments, partially double stranded complexes and/or recipient molecules correspond to open reading frames (or partial open reading frames). In other embodiments, partially double stranded complexes and/or recipient molecules correspond to genomic DNA (and thus may comprise exon, intron and/or untranslated regions, etc.). In other embodiments, partially double stranded complexes and/or recipient molecules correspond to randomly selected/generated portions.

In some embodiments, partially double stranded complexes and/or recipient molecules correspond to known sequences or groups of related sequences. For example, partially double stranded complexes and/or recipient molecules can correspond to one or more discrete full-length or partial functional domains of known biological activity. It is well known in the art that many proteins have discrete functional domains (see, for example, Traut, *Mol. Cell. Biochem.* 70:3 (1986); Go et al., *Adv. Biophys.* 19:91 (1985)). It is also well known that such domains may often be separated from one another and recombined with other discrete functional domains in a manner that preserves the activity of each individual functional domain.

Examplary functional protein domains include, for example, DNA binding domains (such as zinc fingers, homeodomains, helix-turn-helix motifs, etc.), ATP or GTP binding domains, transmembrane spanning domains, protein-protein interaction domains (such as leucine zippers, TPR repeats, WD repeats STYX domains (see, for example, Wishart et al., *Trends Biochem. Sci.* 23: 301, 1998), G-protein domains, tyrosine kinase domains (see, for example, Shokat, *Chem. Biol.* 2:509, 1995), SRC homology domains (see, for example, Sudol, *Oncogene* 17:1469, 1998), SH2 domains (see, for example, Schaffhausen, *Biochim. Biophys. Acta* 28:61, 1995), PTB domains (see, for example, van der Greer et al., *Trends Biochem. Sci.* 20:277, 1995), the PH domain (see, for example, Musacchio et al., *Trends Biochem. Sci.* 18:343, 1993), certain catalytic domains and cell surface receptor domains (see, for example, Campbell et al., *Immunol. Rev.* 163:11, 1998), carbohydrate recognition

IDT Exhibit 1020.9981

domains (see, for example, Kishore et al., *Matrix Biol.* 15:583, 1997), immunoglobulin domains (see, for example, Rapley, *Mol. Biotechnol.* 3:139, 1995).

Functional protein domains suitable for use in accordance with the present invention include those that have been reused through evolution to generate gene families. Examples of such gene families include, for example, the tissue plasminogen activator gene family (see, for example, WO 00/40715, Fig. 6), the family of voltage-gated sodium channels (see, for example, Marban et al., *J. Physiol.* 508:647, 1998), certain families of adhesion molecules (see, for example, Taylor et al., *Curr. Top. Microbiol. Immunol.* 228:135, 1998), various extracellular domain protein families (see, for example, Engle, *Matrix Biol.* 15:295, 1996), the protein kinase C family (see, for example, Dekker et al., *Curr. Op. Struct. Biol.* 5:396, 1995), the tumor necrosis factor receptor superfamily (see, for example, Naismith et al., *J. Inflamm.* 47:1, 1995), the lysin family (see, for example, Lopez et al., *Microb. Drug Resist.* 3:199, 1997), the nuclear hormone receptor gene superfamily (see, for example, Ribeiro et al., *Annu. Rev. Med.* 46:443, 1995; Carson-Jurica et al., *Endocr. Rev.* 11:201, 1990), the neurexin family (see, for example, Missler et al., *J. Neurochem.* 71:1339, 1998), the thioredoxin gene family (see, for example, Sahrawy et al., *J. Mol. Evol.* 42:422, 1996), the phosphoryl transfer protein family (see, for example, Reizer et al., *Curr. Op. Struct. Biol.* 7:407, 1997), the cell wall hydrolase family (see, for example, Hazelwood et al., *Prog. Nuc. Acid Res. Mol. Biol.* 61:211, 1998), and certain families of synthetic proteins (e.g., fatty acid synthases, polyketide synthases (see, for example, WO 98/01546, U.S. Pat. No. 5,252,474, U.S. Pat. No. 5,098,83), peptide synthetases (see, for example, Mootz et al., *Curr. Op. Chem. Biol.* 1:543, 1997; Stachelhaus et al., *FEMS Microbiol. Lett.* 125:3, 1995), and terpene synthases).

It is understood that, using the methods described herein, a multiplicity of partially double stranded complexes can be generated that comprise sequence corresponding to a specified protein domain (or part of a domain), wherein each partially double stranded complex further comprises sequences not necessarily related to the specified domain. For example, partially double stranded complex can be prepared that are comprise a kinase domain but are otherwise not necessarily related (e.g., receptor sequence, intracellular kinase, etc.).

Partially double stranded complexes and/or recipient molecules can also include nucleic acids corresponding to "linker" amino acids (generally, non-native nucleic acid

IDT Exhibit 1020.9982

sequence), which may reflect nucleic acids added during hybridization and recombining steps, as described herein.

Partially double stranded complexes and/or recipient molecules can also include nucleic acids corresponding to sequences capable of directly or indirectly regulating expression of a recombined molecule (or even of a second molecule or pathway of molecules), as further described herein. For example, partially double stranded complexes and/or recipient molecules may comprise sequences corresponding to an sequence element relating to or regulating transcription and/or expression, such as a promoter. Such elements may be known (in the sense that the sequence is known to be involved with transcription and/or expression) or may be novel. Methods for identifying novel elements (or novel arrangements of elements) are known in the art, and are briefly described below. In another example, partially double stranded complex and/or recipient molecules may comprise sequences corresponding to a small untranslated RNA, wherein such RNA, for example, is involved in regulation of gene expression.

Partially double stranded DNA comprising a 3' single stranded portion prepared according to the methods described herein can be generated from RNA copies of genomic DNA, and would thus include untranscribed sequences generally not found in mRNA, such as introns, regulatory and control elements, pseudo-genes, etc.

In some embodiments, recipient molecules (which may be single or double stranded) comprise sequence element relating to or regulating transcription and/or expression (*in vivo*, e.g. in a host cell, or *in vitro*), including at least the following: transcriptional control sequences, RNA splicing control sequences, other RNA modification control sequences, and/or translational control sequences. A wide variety of such transcription and/or translation and/or expression control sequences are well known in the art (see, for example, Sambrook et al., *Molecular Cloning: A Laboratory Manual*, 2nd ed., Cold Spring Harbor Press, Cold Spring Harbor, 1989).

Partially double stranded complex and/or recipient molecules can also include nucleic acids corresponding to untranslated RNA sequences, including tRNA, ribosomal RNA, and untranslated RNA sequences involved in, for example, disease processes and/or gene regulation as described above. Thus, partially double stranded DNA comprising a 3' single stranded portion prepared according to the methods described herein can be generated from RNA copies of genomic DNA, and would thus include untranscribed sequences generally not found in mRNA, such as introns, regulatory and control elements,

IDT Exhibit 1020.9983

pseudo-genes, etc. Partially double stranded DNA comprising a 3' single stranded portion can also be generated from exogenous elements present within a genome, such as viral sequences or any of RNA sequences which are the products of non coding sequences of either known or yet unknown functions.

Partially double stranded complexes and/or recipient molecules can also include nucleic acids corresponding to a detectable protein moiety (e.g., an enzyme moiety that catalyzes a detectable reaction such as color change or induction of fluorescence or luminescence; a moiety that interacts with a known monoclonal antibody, etc.), a moiety that allows ready purification of any polypeptide encoded by the recombined polynucleotide molecule (e.g., a tag such as GST domain, flag domain, six-cys domain, a copper chelate domain, etc.). In a preferred embodiment, the recipient molecule comprises such nucleic acids sequences.

Examples of further nucleic acid sequences that are useful in the present invention include: vector sequences (including expression vector sequences), vectors including a reporter gene or selectable marker, vectors including a "tag"; linear plasmid sequences, linker oligonucleotides, and the like. Suitable vector sequences are well known in the art. *See* Sambrook et al., *Molecular Cloning: A Laboratory Manual*, 2nd ed., Cold Spring Harbor Press, Cold Spring Harbor, 1989. Generally, such nucleic acids are included in recipient molecule, which is generally (but not necessarily) single stranded.

The methods of the invention further comprise hybridizing the 3' single stranded portion of a partially double stranded complex (prepared according to the methods of the invention) with a recipient nucleic acid molecule, whereby a hybrid nucleic acid is generated. (Stated alternatively, partially double-stranded complex and recipient are combined under conditions that allow hybridization.) As used herein, a hybrid nucleic acid molecule refers to hybridized partially double stranded complex and recipient molecule (which may be a second partially double stranded complex of the invention). The hybrid molecule has a nick (created adjacent hybridized 3' end of a partially double stranded complex and 5' end of the recipient molecule, or the converse) or a gap, created as when, for example, the length of hybridizable sequence is less than the length of the 3' overhang, or when the nucleotides at the 3' end of duplex and/or 5' end of recipient molecule are not hybridized. Accordingly, and as noted elsewhere, reference to hybridization of the single stranded 3' portion of the complex refers to hybridization of part of (a portion of) or the entire 3' portion.

IDT Exhibit 1020.9984

Hybridization of 3' ends depends on a number of factors, including length of hybridizable portion and the sequence of the 3' ends. The minimum length of overlap necessary for hybridization of two 3' ends is: about 2 nucleotides, and can be at least 3, at least 4, at least 5, at least 10, or more nucleotides in length. Overlap of hybridizing portion can be at least 2, 5, 10, 20, 30, 50, 70, or 100 or more nucleotides in length. Overlap as used herein can include hybridized 3' ends comprising hybridized portions interspersed (or adjacent to) portions not capable of hybridizing.

Design considerations for the sequence of the 3' end are discussed herein, and include at least the following: use of defined or relatively defined 3' ends (e.g., corresponding to a sequence or group of related sequences); use of random portions; use of nucleotides likely or less likely to hybridize (such as increased or decreased G:C content, use of nucleotide analogs and the like). Generally, the more related (in terms of complementarity) the 3' ends are, the smaller the number of potentially hybridizable 3' ends there are. Correspondingly, the use of 3' ends comprising random portions (in partially double stranded complex and/or recipient molecule) increases the number of potentially hybridizable 3' ends is created. Other factors for increasing (or decreasing) number of potentially hybridizable 3' ends are known in the art, and discussed herein. As is well known in the art, the degree of complementarity between two sequences contributes to hybridization.

Hybridization of 3' ends includes hybridization under conditions of different "stringency". Generally, the more stringent are the hybridization conditions, the higher the sequence identity between the hybridizing 3' ends. Thus, hybridization under conditions of reduced stringency (e.g., moderate stringency, low stringency) generally results in an increase of potentially hybridizable 3' ends, and an attendant increase in the number of possible combinations between partially double stranded complex(es) and recipient molecule(s).

Hybridization is well known and understood in the art, and depends on other factors such as, for example, ionic strength and temperature. Conditions that increase stringency of a hybridization reaction of widely known and published in the art. See, for example, Sambrook et al. (1989). Examples of relevant conditions include (in order of increasing stringency): incubation temperatures of 25°C, 37°C, 50°C and 68°C; buffer concentrations of 10 X SSC, 6 X SSC, 1 X SSC, 0.1 X SSC (where SSC is 0.15 M NaCl and 15 mM citrate buffer) and their equivalents using other buffer systems; formamide concentrations

IDT Exhibit 1020.9985

of 0%, 25%, 50%, and 75%; incubation times from 5 minutes to 24 hours; 1, 2, or more washing steps; wash incubation times of 1, 2, or 15 minutes; and wash solutions of 6 X SSC, 1 X SSC, 0.1 X SSC, or deionized water.

Generally, the sequence of hybridizable 3' ends (of a partially double stranded complex and recipient molecule) are at least about 45%, about 50%, about 55%, about 60%, about 65%, about 70%, about 75%, about 80 %, about 85%, about 95%, about 95% complementarity.

For convenience, hybridization of only one partially double stranded complex and one recipient molecule is described (and illustrated in Figures 3-6). It is understood that the methods of the invention are useful not only for producing hybrid molecules comprising a partially double stranded complex (according to the methods described herein) and one recipient molecule (which can be a partially double stranded complex as described herein), but also for producing hybrid molecules comprising more than partially double stranded complex and or recipient molecules. Accordingly, the invention provides for hybrid molecules comprising at least two partially double stranded complex and at least one recipient molecule, as described herein.

Recombining hybrid polynucleotide

In another aspect, the invention provides methods for generation of a recombined polynucleotide from hybrid polynucleotide(s) prepared according to the methods of the invention. Recombined polynucleotides are useful, for example, for screening and/or selecting for desirable properties, as further described herein. As used herein "recombined product" refers to a continuous (fully linked) polynucleotide comprising sequences corresponding to the partially double stranded complex comprising a 3' single stranded portion and the recipient molecule.

The invention provides methods for recombination of hybrid polynucleotides. Recombination can be according to the methods of the invention described herein. It is also contemplated that any method known in the art can be used. As described herein, hybrid nucleic acid molecule refers to hybridized partially double stranded complex and recipient molecule (which may be a second partially double stranded complex of the invention). The hybrid molecule has a nick (created between an adjacent hybridized 5' end of the recipient molecule (hybridized to 3' overhang of the partially double stranded complex) and a 3' end of partially double stranded complex), or a gap, created as when, for

IDT Exhibit 1020.9986

example, the length of hybridized 5' end of the recipient molecule is less than the length of the 3' overhang. Although a nick or gap is described on only one strand of a hybrid molecule, it is understood that a nick or gap may be created on both strands of a hybrid molecule (generally, when a recipient molecule is double stranded).

For convenience, the embodiments described herein generally related to a partially double stranded complex comprising a single 3' overhang hybridized to a 3' overhang of a recipient (which may be double or single stranded). Generally, in this configuration (and as illustratively diagrammed in Figures 3-6), the 3' overhang of recipient molecule is hybridized to the 3' single stranded portion of the partially double stranded complex, and as such, the 3' overhang of the recipient molecule can serve as a primer for primer extension. It is understood, however, that (unless specifically stated), the methods described herein apply to embodiments (e.g., partially double stranded complex with two 3' overhangs, recipient molecules comprising two 3' overhangs, and the like) in which a 3' overhang of partially double stranded complex can serve as a primer for primer extension.

In some embodiments, recombined product is generated by ligating the 3' end of the recipient molecule (which, as noted above, can also serve as a primer) with the 5' end of the non-overhanging strand of the partially double stranded complex using a DNA ligase. Suitable conditions and ligases (e.g., E. coli DNA ligase) are known in the art. If a 5' end ribonucleotide remains on the non-overhanging strand of the partially double stranded complex (i.e. at least one ribonucleotide of the composite primer remains following cleavage of RNA from a RNA/DNA heteroduplex), the ribonucleotide can be removed by addition of RNase H (for example, E. coli RNase H or T4 RNAse H) , which is known to cleave the ribonucleotides added by extension of Okazaki fragments in normal lagging strand replication. *See,* Bhagwat M and Nossal N (2001) *J. Bio. Chem.* 276: 28516-24.

In other embodiments, recombined product is generated by (a) primer extension from hybridized 3' end of recipient molecule, whereby the gap is filled, and (b) ligation as described herein. In other embodiments, the recombined product is generated by (a) primer extension from hybridized 3' end of the complex, whereby the gap is filled, and (b) ligation as described herein. In other embodiments, recombined product is generated by (a) primer extension from hybridized 3' end of recipient molele and 3' end of the complex, whereby the gap is filled, and (b) ligation as described herein. Generally, primer extension is by a DNA polymerase lacking strand displacement activity. As further described herein, a low fidelity DNA-dependent DNA polymerase or RNA-dependent DNA polymerase (e.g.,

IDT Exhibit 1020.9987

lacking proofreading capability) can be used if introduction of new mutations is desired during primer extension. Addition of new mutations results in greater diversity in the products.

In some embodiments, recombined product is generated by primer extension from hybridized 3' end of recipient molecule (and/or 3' end of hybridized complex) using a DNA polymerase comprising strand displacement activity, and displacement of the hybridized non-overlapping strand of partially double stranded complex. In some embodiments, a low fidelity polymerase comprising strand displacement activity is used, such that additional mutations are added during primer extension.

In some embodiments, recombined product is generated by nick translation (which can extend along part or all of a strand, generally a complementary strand). In some embodiments, the polymerase comprises a strand displacement activity with or without 3' and/or 5' exonuclease activity. When a polymerase with 3' exonuclease activity is used, it is desirable to add the polymerase after hybridization, to prevent degradation of the 3' single-stranded overhang.

In some embodiments, hybrid molecules are treated with agents (generally enzymes) comprising a 3' and or 5' exonuclease activity, whereby a free 3' end of recipient molecule and/or the 5' end of the non-overlapping strand of double stranded duplex are removed to varying extents, creating gaps of increased size. Exonuclease treatment of hybrid molecules, followed by gap-filling using low fidelity polymerases result in the introduction of mutations in a larger portion of the recombined product.

In another aspect of the invention, recombined molecules are generated from hybrid molecule comprising hybridized partially double stranded complex and single stranded recipient molecule (which may be immobilized, as further described herein). In some embodiments, the single stranded recipient molecule may be converted to a double stranded nucleic acid by extension with DNA polymerase (primed by hybridized 3' overhang portion of partially double stranded DNA complex). Other embodiments are described herein in the section describing immobilized recipient molecules, infra.

Figure 3 illustratively exemplifies one embodiment of the methods for generating recombined product as follows:

1. Partially double stranded complexes with 3' single-stranded overhang sequences are hybridized to a plurality of recipient molecules comprising partially double stranded DNA fragments generated by restriction digest, wherein the "sticky ends" produced

IDT Exhibit 1020.9988

from the restriction digest are capable of hybridizing with the 3' single-stranded overhang sequences of the first plurality of partially double stranded complexes;

2. Extension by DNA polymerase is used to fill gaps (if any);
3. Ligation of hybridized ends generates recombined double stranded product.

Figure 4 illustratively exemplifies one embodiment of the methods for generating recombined product as follows:

1. Partially double stranded complexes with 3' single-stranded overhang sequences are hybridized to a recipient molecules comprising single stranded nucleic acid sequences;
2. Extension by primer extension to fill gaps (if any) and ligation of hybridized ends to generate recombined single stranded product;

primer extension primed by 3' overhang of the partially double stranded complex whereby double stranded recombined product is generated.

Figure 5 illustratively exemplifies one embodiment of the methods for generating recombined product as follows:

1. Partially double stranded complexes comprising random 3' single-stranded overhangs are hybridized to a recipient molecules comprising random 3' overhang sequences;
2. Extension by primer extension to fill gaps (if any) and ligation of hybridized ends to generate recombined single stranded product.

Uses for recombined products of the invention

Recombined products are useful for screening and/or selecting products with useful properties. Methods for screening and selecting are well-known in the art. *See, e.g.*, U.S. Patent Nos. 5,830,721; 6,335,160; 6,344,356; 5,811,238; Zhang et al. (1997) PNAS 94:4505-4509; Gulik and Fahl (1995) PNAS 92:8140-8144; You and Arnold (1996) Protein Eng.9:77-83; Barbas (1994) PNAS 91:3809-3813; Crameria et al. (1997) Nature Biotech. 15:436-438; Heim and Tsien (1996) Curr. Biol. 6:;178-182.

Generally, recombined molecules may be expressed using methods well known in the art. For example, recombined molecules can be cloned into a plasmid or expression vector, amplified by PCR, or permitted to recombine *in vivo* (for example, by transfection into a suitable host). Recombined molecules can be translated, followed by screening or selection (in vivo or in vitro) for desired characteristics, using procedures that are known in

IDT Exhibit 1020.9989

the art. Desirable characteristics include: enhanced and/or altered and/or hybrid enzymatic activities; enhanced or improved folding of a protein domain; ability to regulate translation and/or expression of a reporter (e.g., promoter trapping) and/or a desired phenotype and/or encoding a protein having an advantageous predetermined property. It is understood that recombined molecules can be screened or selected independent of translation of recombined sequences, as when, for example, untranslated RNA or RNA interference is assessed or other RNA molecules which are the products of non-coding sequences and function in gene expression regulation. .

Generation of recombined nucleic acid attached to a solid (or semi-solid) support

The methods of the invention provide for generation of an immobilized recombined single or double stranded polynucleotide (such as DNA) to a support, which may be solid or semi-solid. Figure 6 illustrates one embodiment of the invention. Hybridization of the 3' single stranded portions of the partially double stranded complex and an immobilized single stranded recipient molecule results in generation of a hybrid nucleic acid comprising a complex of the partially double stranded complex and the immobilized single stranded recipient molecule. As used herein, a "cDNA" is (i.e., encompasses) a polynucleotide. Primer extension primed by the 3' hybridized end of the single stranded recipient molecule results in displacement of the previously hybridized non-overlapping strand of partially double stranded complex (corresponding to first strand cDNA) and formation of a complex comprising (a) immobilized chimeric single stranded DNA comprising (i) the first single stranded (DNA) nucleic acid and (ii) the first strand cDNA; and (b) second primer extension product. Denaturation or degradation of the hybridized second primer extension product results in immobilized single stranded chimeric DNA.

The immobilized recipient molecule is immobilized by using methods known in the art, which include: attachment to a solid or semisolid support, e.g., paper, glass, plastic (e.g., polypropylene, nylon, polystyrene), polyacrylamide, nitrocellulose, silicon or other metals, optical fiber. A solid or semi-solid support can be configured in any means known in the art, including in a planar (e.g., glass plates, silicon chips) or three-dimensional (e.g., pins, fibers, beads, particles, microtiter wells, capillaries) configuration.

Means for attaching and/or immobilizing nucleic acids to a solid support include (i) *in situ* synthesis (e.g., high-density oligonucleotide arrays) using photolithographic

IDT Exhibit 1020.9990

techniques (see, Fodor et al., *Science* (1991), 251:767-773; Pease et al., *Proc. Natl. Acad. Sci. U.S.A.* (1994), 91:5022-5026; Lockhart et al., *Nature Biotechnology* (1996), 14:1675; U.S. Pat. Nos. 5,578,832; 5,556,752; and 5,510,270); (ii) spotting/printing at medium to low-density on glass, nylon or nitrocellulose (Schena et al, *Science* (1995), 270:467-470, DeRisi et al, *Nature Genetics* (1996), 14:457-460; Shalon et al., *Genome Res.* (1996), 6:639-645; and Schena et al., *Proc. Natl. Acad. Sci. U.S.A.* (1995), 93:10539-11286); (iii) by masking (Maskos and Southern, *Nuc. Acids. Res.* (1992), 20:1679-1684) and (iv) by dot-blotting on a nylon or nitrocellulose hybridization membrane (see, e.g., Sambrook et al., Eds., 1989, Molecular Cloning: A Laboratory Manual, 2nd ed., Vol. 1-3, Cold Spring Harbor Laboratory (Cold Spring Harbor, N.Y.)). Oligonucleotide primers may also be noncovalently immobilized on the support by hybridization to anchors, by means of magnetic beads, or in a fluid phase such as in microtiter wells or capillaries.

In one embodiment of the method, recipient nucleic acids are immobilized as a microarray.

Optionally, labels may be attached to one or both strands of the partially double stranded complex, or to the immobilized recipient molecule, or to the solid support to which the recipient molecule is attached (e.g., beads with attached dye molecules), or any combination or permutation thereof. Labeling allows the identification and/or quantification of the immobilized recipient molecule which bind partially double stranded complex and/or which form further products. Labels include fluorochromes or radioisotopes. The label may also be a small molecule which is a member of a specific binding pair, and can be detected following binding of the other member of the specific binding pair, such as biotin and streptavidin, respectively, with the last member of the binding pair conjugated to an enzyme that catalyzes the generation of a detectable signal that could be detected by methods such as colorimetry, fluorometry or chemiluminescence. All of the above examples are well known in the art.

Generation of a hybrid molecule comprising immoblized double or single stranded recipient molecules is by any method described herein.

Generally, production of a recombined immobilized polynucleotide (which can be double or single stranded) is accomplished by any method described herein.

In other embodiments, immobilized recombined product is rendered double stranded by primer extension primed by the hybridized 3' overhang of partially double stranded complex.

IDT Exhibit 1020.9991

In other embodiments, immobilized partially double stranded recombined product is generated by primer extension primed by the hybridized 3' overhand of partially double stranded complex which is halted before the end of the 3' overhang strand is reached, in which case immobilized double stranded DNA with a 5' overhang is produced.

In other embodiments, 5' and/or 3' exonucleases can be used to remove part or all of the non-overlapping strand of double stranded heteroduplex (after, for example, primer extension and/or ligation. Such manipulations can result in production of single stranded recombined product that is shorter than the length of a combined partially double stranded complex and recipient molecule.

In other embodiments, following generation of a recombined strand that is immobilized (generally present in a complex comprising the 3' overhang strand of the partially double stranded complex), the 3' overhang strand of the original partially double stranded complex is removed to produce immobilized recombined product. Removal of the 3' overhang strand of partially double stranded complex may be achieved by denaturing the two strands through thermal denaturation or any other method known in the art, such as alkali treatment.

In other embodiment, the 3' overhang strand of the original partially double stranded complex can be partially or completely removed by degrading the 3' overhang strand using exonuclease; such as 5' exonuclease. Such 5' exonucleases and conditions for their use are known in the art. This procedure, if carried out from the 5' end of the 3' overhang strand and not allowed to go to completion, results in formation of an immobilized partially double stranded complex with a free 3' overhang.

Uses of immobilized recombined product

Uses of such immobilized polynucleotides (which can be double stranded or single stranded) will be readily apparent to those of skill in the art. In some embodiments, the methods of the invention may be used to provide an array containing a large number of different polynucleotide probes where the sequence and location of each probe is known. Such arrays are useful, for example, in expression analysis, either of a single sample or of multiple samples which may be compared. See e.g., U.S. Patent Nos. 6,344,316 and 6,303,301. In other embodiments, the solid support may be a bead that may be further dyed as a reporter group, offering means for identifying and/or quantifying levels of, e.g., mRNA. In yet other embodiments, single stranded DNA attached to the solid support may

IDT Exhibit 1020.9992

be used to sequence the single stranded DNA, by sequencing by synthesis (e.g. pyrosequencing). It is also possible to use the attached single stranded composition for selection of desired sequences from a mixture of sequences, by hybridization to the desired sequences, for example, if the attached single stranded sequences comprise sequences of expressed genes of one sample, it is possible to use the solid support with the attached single strand for selection of similar sequences from mRNA or cDNAs derived from another sequence. The attached single stranded polynucleotides may be one sequence or a plurality of sequences.

Kits

The invention also includes kits that are useful in methods of recombination of nucleic acids or directed evolution. Reagents that are useful for the practice of the present invention may desirably be provided together and assembled into a kit. Kits of the invention include partially double stranded complexes with 3' overhang sequences made by the method(s) described herein) and instructions for use in any of the methods described herein (and may further include components, such as appropriate enzyme(s), for achieving recombination). In some embodiments, the partially double stranded complexes comprise a cDNA library, full-length or partial domains for exon shuffling, or fragments of a coding sequence, for DNA shuffling. In some embodiments, the kits also comprise DNA polymerase. In some embodiments, the kits contain instructions for recombining the nucleic acids as described herein. In some embodiments, the kits comprise a solid support.

Components and reaction conditions used in the methods of the invention

Template nucleic acid

The RNA target includes RNAs from any source in purified or unpurified form, which can be RNA such as total RNA, tRNA, mRNA, rRNA, mitochondrial RNA, chloroplast RNA, DNA-RNA hybrids, or mixtures thereof, from any source and/or species, including human, animals, plants, and microorganisms such as bacteria, yeasts, viruses, viroids, molds, fungi, plants, and fragments thereof. It is understood that the RNA can be coding or noncoding RNA (such as untranslated small RNAs). RNAs can be obtained and purified using standard techniques in the art. Use of a DNA target (including genomic DNA target) would require initial transcription of the DNA target into RNA form, which can be achieved using methods disclosed in Kurn, U.S. Patent No. 6,251,639 B1, and by

IDT Exhibit 1020.9993

other techniques (such as expression systems) known in the art. Thus, partially double stranded complex can be generated from RNA target that is itself generated from a DNA source (such as genomic DNA), using methods known in the art, including Kurn, U.S. Patent No. 6,251,639. RNA copies of genomic DNA would generally include untranscribed sequences generally not found in mRNA, such as introns, regulatory and control elements, etc. RNA targets may also be generated from cloned genomic DNA sequences that can be subjected to in vitro transcription. Use of a DNA-RNA hybrid would require denaturation of the hybrid to obtain a single stranded RNA, denaturation followed by transcription of the DNA strand to obtain an RNA, or other methods known in the art such as digestion with an RNAse H to generate single stranded DNA. The target RNA can be only a minor fraction of a complex mixture such as a biological sample and can be obtained from various biological materials by procedures well known in the art. The target RNA can be known or unknown and may contain more than one desired specific nucleic acid sequence of interest, each of which may be the same or different from each other. Denaturation may also be carried out to remove secondary structure present in a RNA target molecule. If the target nucleic acid is double stranded (e.g., RNA/DNA hybrid) the initial step could be target denaturation. The denaturation step may be thermal denaturation or any other method known in the art. The removal of DNA strands from a complex material may be also achieved by treatment with DNase.

Composite primer

The methods of the invention employ a composite primer that is composed of RNA and DNA portions. The composite primer can be designed such that subsequent displacement of the primer extension product by binding of a new nucleic acid capable of priming primer extension.

Composite primers for use in the methods and compositions of the invention comprise a sequence capable of hybridizing to a target RNA. The sequence that is capable of hybridizing to the target RNA can be based on the particular sequence of a specific target RNA (for e.g., the mRNA of a particular gene), or be based on a more general sequence type known to be present in a plurality of RNA species, such as the poly-A tail sequence generally believed in the art to be present in all eukaryotic mRNA. In addition, the sequence that is capable of hybridizing to the target RNA may comprise a sequence complementary to the poly-A tail of mRNA, and may further comprise additional random

IDT Exhibit 1020.9994

sequences (generally not complementary to a poly-A sequence) at the 3' end of the 3' portion.

The sequence that is capable of hybridizing to the target RNA may also comprise a random sequence. Random primers are well known in the art, and include at least the following: primers hybridizable to two or more sequences in a sample; and primers comprising poly-dT sequences that are hybridizable to a multiplicity of RNAs in a sample (such as all mRNA). For convenience, a single random composite primer is discussed above. However, it is understood that the term "random primer" can refer to a primer that is a member of a population of primers which are collectively designed to a desired and/or significant population of target sequences.

It is also understood that generation of partially double stranded complexes from a plurality of mRNA species in a single reaction mixture may, but not necessarily, employ a multiplicity, or a large multiplicity of primers. Thus, the invention contemplates the use of a multiplicity of different composite primers (random or non-random) when partially double stranded complexes of the invention are generated from a plurality of mRNA species in a single reaction mixture.

In some embodiments, a first composite primer is used in the methods of the invention. In other embodiments, a multiplicity of different first composite primers are used in the methods of the invention. In other embodiments, a first and second composite primer are used in the methods of the invention. The second composite primer (used to primer second strand synthesis) may comprise some or all of the sequence of the first composite primer, and the first composite primer may comprise some or all of the sequence of the second composite primer. In some embodiments, the second composite primer comprises a different sequence than the first composite primer. In other embodiments, a multiplicity of different second composite primers are used in the methods of the invention.

The composite primer may comprise a sequence, preferably at the 5' end, that is not hybridizable (under a given set of conditions) to the target (for example, a non-hybridized 5' portion that would constitute a tail when the primer is bound to the target). Individual DNA and RNA portions of the composite primer may be completely or partially hybridizable to the target RNA.

Generation of primers suitable for extension by polymerization is well known in the art, such as described in PCT Pub. No. WO99/42618 (and references cited therein). The composite primer comprises a combination of RNA and DNA (see definition above), with

IDT Exhibit 1020.9995

the 3'-end nucleotide being a nucleotide suitable for nucleic acid extension. The 3'-end nucleotide can be any nucleotide or analog that when present in a primer, is extendable by a DNA polymerase. Generally, the 3'-end nucleotide has a 3'-OH. Suitable primers include those that comprise at least one portion of RNA and at least one portion of DNA. For example, composite primers can comprise a 5'-RNA portion and a 3'-DNA portion (in which the RNA portion is adjacent to the 3'-DNA portion); or 5'- and 3'-DNA portions with an intervening RNA portion. Accordingly, in one embodiment, the composite primer comprises a 5' RNA portion and a 3'-DNA portion, preferably wherein the RNA portion is adjacent to the 3'-DNA portion. In another embodiment, the composite primer comprises 5'- and 3'-DNA portions with at least one intervening RNA portion (i.e., an RNA portion between the two DNA portions). In yet another embodiment, the composite primer of the invention comprises a 3'-DNA portion and at least one intervening RNA portion (i.e., an RNA portion between DNA portions).

The length of an RNA portion in a composite primer comprising a 3'-DNA portion and an RNA portion can be from about 1 to about 50 or more nucleotides, from about 3 to about 20, from about 4 to about 15, and about 5 to about 10 nucleotides. In some embodiments of a composite primer comprising a 3'-DNA portion and an RNA portion, an RNA portion can be at least about any of 1, 3, 4, 5 nucleotides, with an upper limit of about any of 10, 15, 20, 25, 30, 50, 75, 100, 125, 150, or more nucleotides. In some embodiments, the RNA portion of the composite primer consists of about 10 to about 20 nucleotides. In some embodiments, the RNA portion of the composite primer consists of about 10 to about 20 nucleotides and the DNA portfion of the composite primer consists of about 5 to about 20 nucleotides.

The length of the 5'-RNA portion in a composite primer comprising a 5'-RNA portion and a 3'-DNA portion can be from about 2 to about 50 or more nucleotides, from about 5 to about 20 nucleotides, from about 7 to about 18 nucleotides, preferably from about 8 to about 17 nucleotides, and from about 10 to about 15 nucleotides. In other embodiments of a composite primer comprising a 5'-RNA portion and a 3'-DNA portion, the 5'-RNA portion can be at least about any of 3, 5, 7, 8, 10 nucleotides, with an upper limit of about any of 15, 17, 18, 20, 50, 75, 100, 125, 150, or more nucleotides.

In embodiments of a composite primer comprising a 5'-RNA portion and a 3'-DNA portion further comprising non-5'-RNA portion(s), a non-5'-RNA portion can be from about 1 to about 7 nucleotides, from about 2 to about 6 nucleotides, and from about 3 to

IDT Exhibit 1020.9996

about 5 nucleotides. In certain embodiments of a composite primer comprising a 5'-RNA portion and a 3'-DNA portion further comprising non-5'-RNA portion(s), a non-5'-RNA portion can be at least about any of 1, 2, 3, 5, with an upper limit of about any of 5, 6, 7, 10, 15, 20, 25, 30, or more nucleotides.

In embodiments of a composite primer comprising a 5'-RNA portion and a 3'-DNA portion, in which the 5'-RNA portion is adjacent to the 3'-DNA portion, the length of the 5'-RNA portion can be from about 3 to about 50 or more nucleotides, from about 5 to about 20 nucleotides, from about 7 to about 18 nucleotides, from about 8 to about 17 nucleotides, and from about 10 to about 15 nucleotides. In certain embodiments of a composite primer comprising a 5'-RNA portion and a 3'-DNA portion, in which the 5'-RNA portion is adjacent to the 3'-DNA portion, the 5'-RNA portion can be at least about any of 3, 5, 7, 8, 10 nucleotides, with an upper limit of about any of 15, 17, 18, 20, 50, 75, 100, 125, 150, or more nucleotides.

The length of an intervening RNA portion in a composite primer comprising 5'- and 3'-DNA portions with at least one intervening RNA portion can be from about 1 to about 7 nucleotides, from about 2 to about 6 nucleotides, and from about 3 to about 5 nucleotides. In some embodiments of a composite primer comprising 5'- and 3'-DNA portions with at least one intervening RNA portion, an intervening RNA portion can be at least about any of 1, 2, 3, 5 nucleotides, with an upper limit of about any of 5, 6, 7, 10, 15, 20, 25, 30, or more nucleotides. The length of an intervening RNA portion in a composite primer comprising a 3'-DNA portion and at least one intervening RNA portion can be from about 1 to about 7 nucleotides, from about 2 to about 6 nucleotides, and from about 3 to about 5 nucleotides. In some embodiments of a composite primer comprising a 3'-DNA portion and at least one intervening RNA portion, an intervening RNA portion can be at least about any of 1, 2, 3, 5 nucleotides, with an upper limit of about any of 5, 6, 7, 10, 15, 20, 25, 30 or more nucleotides. In a composite primer comprising a 3'-DNA portion and at least one intervening RNA portion, further comprising a 5'-RNA portion, the 5'-RNA portion can be from about 3 to about 25 nucleotides, from about 5 to about 20 nucleotides, from about 7 to about 18 nucleotides, about 8 to about 17 nucleotides, and from about 10 to about 15 nucleotides. In some embodiments of a composite primer comprising a 3'-DNA portion and at least one intervening RNA portion, further comprising a 5'-RNA portion, the 5'-RNA portion can be at least about any of 3, 5, 7, 8, 10 nucleotides, with an upper limit of about any of 15, 17, 18, 20, 25, 30, or more nucleotides.

IDT Exhibit 1020.9997

The length of the 3'-DNA portion in a composite primer comprising a 3'-DNA portion and an RNA portion can be from about 1 to about 20, from about 3 to about 18, from about 5 to about 15, and from about 7 to about 12 nucleotides. In some embodiments of a composite primer comprising a 3'-DNA portion and an RNA portion, the 3'-DNA portion can be at least about any of 1, 3, 5, 7, 10 nucleotides, with an upper limit of about any of 10, 12, 15, 18, 20, 22, 25, 30, 35, or more nucleotides.

The length of the 3'-DNA portion in a composite primer comprising a 5'-RNA portion and a 3'-DNA portion can be from about 1 to about 20 nucleotides, from about 3 to about 18 nucleotides, from about 5 to about 15 nucleotides, and from about 7 to about 12 nucleotides. In some embodiments of a composite primer comprising a 5'-RNA portion and a 3'-DNA portion, the 3' DNA portion can be at least about any of 1, 3, 5, 7, 10 nucleotides, with an upper limit of about any of 10, 12, 15, 18, 20, 22, 25, 30, 35, or more nucleotides.

In embodiments of a composite primer comprising a 5'-RNA portion and a 3'-DNA portion, further comprising non-3'-DNA portion(s), a non-3'-DNA portion can be from about 1 to about 10 nucleotides, from about 2 to about 8 nucleotides, and from about 3 to about 6 nucleotides. In some embodiments of a composite primer comprising a 5'-RNA portion and a 3'-DNA portion, further comprising non-3'-DNA portion(s), a non-3'-DNA portion can be at least about any of 1, 2, 3, 5 nucleotides, with an upper limit of about any of 6, 8, 10, 12, 15, 25, 30, 35 or more nucleotides.

In embodiments of a composite primer comprising a 5'-RNA portion and a 3'-DNA portion in which the 5'-RNA portion is adjacent to the 3'-DNA portion, the length of the 3'-DNA portion can be from about 1 to about 20 nucleotides, from about 3 to about 18 nucleotides, from about 5 to about 15 nucleotides, and from about 7 to about 12 nucleotides. In certain embodiments of the primer comprising a 5'-RNA portion and a 3'-DNA portion in which the 5'-RNA portion is adjacent to the 3'-DNA portion, the 3'-DNA portion can be at least about any of 1, 3, 5, 7, 10 nucleotides, with an upper limit of about any of 10, 12, 15, 18, 20, 22, 25, 30, 35, or more nucleotides.

The length of a non-3'-DNA portion in a composite primer comprising 5'- and 3'-DNA portions with at least one intervening RNA portion can be from about 1 to about 10 nucleotides, from about 2 to about 8 nucleotides, and from about 3 to about 6 nucleotides. In some embodiments of a primer comprising 5'- and 3'-DNA portions with at least one intervening RNA portion, a non-3'-DNA portion can be at least about any of 1, 2, 3, 5

IDT Exhibit 1020.9998

nucleotides, with an upper limit of about any of 6, 8, 10, 12, 15, 20, 25, 30, or more nucleotides.

The length of the 3'-DNA portion in a composite primer comprising 5'- and 3'-DNA portions with at least one intervening RNA portion can be from about 1 to about 20 nucleotides, from about 3 to about 18 nucleotides, from about 5 to about 15 nucleotides, and from about 7 to about 12 nucleotides. In some embodiments of a composite primer comprising 5'- and 3'-DNA portions with at least one intervening RNA portion, the 3'-DNA portion can be at least about any of 1, 3, 5, 7, 10 nucleotides, with an upper limit of about any of 10, 12, 15, 18, 20, 22, 25, 30 or more nucleotides.

The length of a non-3'-DNA portion (i.e., any DNA portion other than the 3'-DNA portion) in a composite primer comprising a 3'-DNA portion and at least one intervening RNA portion can be from about 1 to about 10 nucleotides, from about 2 to about 8 nucleotides, and from about 3 to about 6 nucleotides. In some embodiments of a composite primer comprising a 3'-DNA portion and at least one intervening RNA portion, a non-3'-DNA portion can be at least about any of 1, 3, 5, 7, 10 nucleotides, with an upper limit of about any of 6, 8, 10, 12, 15, 20, 25, 30 or more nucleotides. The length of the 3'-DNA portion in a composite primer comprising a 3'-DNA portion and at least one intervening RNA portion can be from about 1 to about 20 nucleotides, from about 3 to about 18 nucleotides, from about 5 to about 15 nucleotides, and from about 7 to about 12 nucleotides. In some embodiments of a composite primer comprising a 3'-DNA portion and at least one intervening RNA portion, the 3'-DNA portion can be at least about any of 1, 3, 5, 7, 10 nucleotides, with an upper limit of about any of 10, 12, 15, 18, 20, 22, 25, 30, or more nucleotides. It is understood that the lengths for the various portions can be greater or less, as appropriate under the reaction conditions of the methods of this invention.

In some embodiments, the 5'-DNA portion of a composite primer includes the 5'-most nucleotide of the primer. In some embodiments, the 5'-RNA portion of a composite primer includes the 5' most nucleotide of the primer. In other embodiments, the 3'-DNA portion of a composite primer includes the 3' most nucleotide of the primer. In other embodiments, the 3'-DNA portion is adjacent to the 5'-RNA portion and includes the 3' most nucleotide of the primer (and the 5'-RNA portion includes the 5' most nucleotide of the primer).

The total length of the composite primer can be from about 10 to about 50 nucleotides, from about 15 to about 30 nucleotides, and from about 20 to about 25

IDT Exhibit 1020.9999

nucleotides. In some embodiments, the length can be at least about any of 10, 15, 20, 25, 30 nucleotides, with an upper limit of about any of 25, 30, 50, 60, 70, 75, 85, 100 or more nucleotides. It is understood that the length can be greater or less, as appropriate under the reaction conditions of the methods of this invention.

To achieve hybridization to a target nucleic acid (which, as is well known and understood in the art, depends on other factors such as, for example, ionic strength and temperature), the portion of the primer that is hybridizable to the target RNA is preferably of at least about 60%, more preferably at least about 75%, even more preferably at least about 90%, and most preferably at least about 95% complementarity to the target nucleic acid.

In some embodiments, the RNA portion of the composite primer that hybridizes to target RNA is 5' with respect to the 3' DNA portion. In some embodiments, the 5' RNA portion is adjacent to the 3' DNA portion. In some embodiments, the target RNA is mRNA, and the composite primer that hybridizes to target RNA comprises a poly-dT sequence, and further comprises a 5' portion that does not hybridize to the target mRNA under conditions which the composite primer hybridizes to the target mRNA.. In some embodiments, the composite primer that hybridizes to target RNA comprises a random sequence. In some embodiments, a plurality of different composite primers are used for hybridizing to the target RNA.

Second primer

The second primer in the methods of the invention (which primes generation of the second strand cDNA) comprises a sequence (which may or may not be the whole of the primer) that is hybridizable (under a given set of conditions) to a first strand cDNA (interchangeably called first primer extension product) at a site on the first strand cDNA such that the second strand cDNA would include the RNA sequence of interest. In some embodiments, the hybridizable sequence of the second primer is designed based on a known sequence of the desired binding site on a first strand cDNA. In other embodiments, the hybridizable sequence is based on random sequences, for example, known in the art to be suitable for random priming of first strand cDNAs generated from a plurality of RNA species. In other embodiments, the second primer comprises a strand switch oligonucleotide, described in U.S. Patent Nos. 5,962,271 and 5,962,272, which is hybridizable to the Cap sequences present on mRNA and causes the reverse transcriptase to

IDT Exhibit 1020.10000

switch from the mRNA template to the switch oligonucleotide, permitting generation of a second strand cDNA primed by the "switch oligonucleotide". Alternatively, a homopolymeric tail is added to the 3' terminus of the first primer extension product, and the second primer comprises the complement of the homopolymeric tail.

In some embodiments, the second primer (which primes generation of the second strand cDNA) is a composite primer (as described above). In these embodiments, the method involves generation of a double stranded cDNA comprising two 3' single stranded DNA portions.

To achieve hybridization to a first strand cDNA (which, as is well known and understood in the art, depends on other factors such as, for example, ionic strength and temperature), the sequence of the second primer that is hybridizable to the first strand cDNA is preferably of at least about 60%, more preferably at least about 75%, even more preferably at least about 90%, and most preferably at least about 95% complementarity to the first strand cDNA.

In some embodiments, the second primer comprises DNA. In another embodiment, the second primer comprises RNA. In yet another embodiment, the second primer comprises DNA and RNA.

In some embodiments, the second primer is provided by self priming (for example, by a hairpin loop) at the 3' end of the composite primer extension product. In these embodiments, a sequence at the 3' end of the composite primer extension product hybridizes to another sequence in the composite primer extension product itself, for example as described in U.S. Pat. No. 6,132,997. In these embodiments, said sequence at the 3' of the composite primer extension product is generally cleaved (for example, with S1 nuclease) following its hybridization to the composite primer extension product and/or its extension along the composite primer extension product. U.S. Pat. No. 6,132,997.

In some embodiments, the second primer is provided by one or more target RNA fragment(s), and as such is not an exogenously added primer. Such a target RNA fragment can be generated using methods well known in the art, including incomplete degradation (fragmentation or cleavage) of a target RNA in a complex of target RNA and first primer extension product by an agent (such as an enzyme) that cleaves RNA in an RNA/DNA hybrid, such that one or more RNA fragments remain bound to the first primer extension product. As noted above in an earlier section, cleavage/fragmentation can also be effected by heat treatment (which is not sufficient to denature or separate template RNA from first

IDT Exhibit 1020.10001

primer extension product), or chemical treatment (such as treatment under alkaline conditions). In some embodiments, cleavage giving rise to target RNA fragment(s) which serve as primer(s) is effected by RNaseH.

DNA polymerase, an agent capable of cleaving an RNA-DNA hybrid, and RNA polymerase

The methods of the invention employ the following enzymes: an enzyme comprising RNA-dependent DNA polymerase activity, an enzyme comprising DNA-dependent DNA polymerase activity, an agent capable of cleaving an RNA strand of an RNA-DNA hybrid (for example, a ribonuclease such as RNase H), and, in some aspects a DNA-dependent RNA polymerase. One or more of these activities may be found and used in a single enzyme. For example, RNase H activity may be supplied by an RNA-dependent DNA polymerase (such as reverse transcriptase) or may be provided in a separate enzyme. Reverse transcriptases useful for this method may or may not have RNase H activity. DNA-dependent polymerase and RNA-dependent polymerase may be provided in a single enzyme, or in two separate enzymes.

One aspect of the invention is the formation of double stranded cDNA from a primer-RNA complex. This process generally utilizes the enzymatic activities of an RNA-dependent DNA polymerase, a DNA-dependent DNA polymerase and a ribonuclease activity.

RNA-dependent DNA polymerases for use in the methods and compositions of the invention are capable of effecting extension of a primer according to the methods of the invention. Accordingly, a preferred RNA-dependent DNA polymerase is one that is capable of extending a nucleic acid primer along a nucleic acid template that is comprised at least predominantly of ribonucleotides. Suitable RNA-dependent DNA polymerases for use in the methods and compositions of the invention include reverse transcriptase. Many reverse transcriptases, such as those from avian myeoloblastosis virus (AMV-RT), and Moloney murine leukemia virus (MMLV-RT) comprise more than one activity (for example, polymerase activity and ribonuclease activity) and can function in the formation of the double stranded cDNA molecules. However, in some instances, it is preferable to employ a reverse transcriptase which lacks the RNase H activity. Reverse transcriptase devoid of RNase H activity are known in the art, including those comprising a mutation of the wild type reverse transcriptase where the mutation eliminates the RNase H activity.

IDT Exhibit 1020.10002

The addition of an RNase H from other sources, such as that isolated from *E. coli*, can be employed for the formation of the double stranded cDNA.

DNA-dependent DNA polymerases for use in the methods and compositions of the invention are capable of effecting extension of the composite primer according to the methods of the invention. Accordingly, a preferred polymerase is one that is capable of extending a nucleic acid primer along a nucleic acid template that is comprised at least predominantly of deoxynucleotides. The formation of the double stranded cDNA can be carried out by reverse transcriptase which comprises both RNA-dependent DNA polymerase and DNA-dependent DNA polymerase activities. DNA polymerase may or may not possess strand displacement activity. Preferably, the DNA polymerase has high affinity for binding at the 3'-end of an oligonucleotide hybridized to a nucleic acid strand. DNA polymerase may or may not possess substantial nicking activity. DNA polymerase may or may not possess 5'->3' and/or 3'->5' exonuclease activity. Generally, this exonuclease activity is dependent on factors such as pH, salt concentration, whether the template is double stranded or single stranded, and so forth, all of which are familiar to one skilled in the art. Mutant DNA polymerases in which the 5'->3' exonuclease activity has been deleted. Mutant DNA polymerases which lack both 5' to 3' nuclease and 3' to 5' nuclease activities have also been described , for example, exo$^{-/-}$Klenow DNA polymerase. In some embodiments, the use of thermostable DNA polymerases with strand displacement activity is preferred. Such polymerases are known in the art, such as described in U.S. Pat. No. 5744312 (and references cited therein). The polymerases can further have proofreading activity or have little or no proofreading ability

Suitable DNA polymerases for use in the methods and compositions of the invention include those disclosed in U.S. Pat. Nos. 5648211 and 5744312, which include exo$^-$ Vent (New England Biolabs), exo$^-$ Deep Vent (New England Biolabs), Bst (BioRad), exo$^-$ Pfu (Stratagene), Bca (Panvera), sequencing grade Taq (Promega), exo$^{-/-}$Klenow DNA polymerase, and thermostable DNA polymerases from thermoanaerobacter thermohydrosulfuricus.

The agent (such as a ribonuclease) for use in the methods and compositions of the invention is capable of cleaving ribonucleotides in an RNA/DNA hybrid. Preferably, the ribonuclease cleaves ribonucleotides in an RNA/DNA hybrid regardless of the identity and type of nucleotides adjacent to the ribonucleotide to be cleaved. It is preferred that the ribonuclease cleaves independent of sequence identity. Examples of suitable ribonucleases

IDT Exhibit 1020.10003

for the methods and compositions of the invention are well known in the art, including ribonuclease H (RNase H), including Hybridase.

In general, the enzymes used in the methods and compositions of the invention should not produce substantial degradation of the nucleic acid components of said methods and compositions.

In some embodiments, the same enzyme comprises RNA-dependent DNA polymerase activity and cleaves RNA from an RNA/DNA hybrid. In some embodiments, the same enzyme comprises DNA-dependent DNA polymerase activity and cleaves RNA from an RNA/DNA hybrid. In some embodiments, the same enzyme comprises DNA-dependent DNA polymerase activity, RNA-dependent DNA polymerase activity and cleaves RNA from an RNA/DNA hybrid. In some embodiments, different enzymes comprise RNA -dependent DNA polymerase activity and DNA-dependent DNA polymerase activity. In some embodiments, different enzymes comprise RNA -dependent DNA polymerase activity and cleave RNA from an RNA/DNA hybrid. In some embodiments, different enzymes comprise DNA-dependent DNA polymerase activity and cleave RNA from an RNA/DNA hybrid.

Reaction conditions and detection

Appropriate reaction media and conditions for carrying out the methods of the invention are those that permit primer extension by polymerases, hybridization, and recombination according to the methods of the invention. Such media and conditions are known to persons of skill in the art, and are described in various publications, such as U.S. Pat. Nos. 5,554,516; 5,716,785; 5,130,238; 5,194,370; 6,090,591; 5,409,818; 5,554,517; 5,169,766; 5,480,784; 5,399,491; 5,679,512; and PCT Pub. No. WO99/42618. For example, a buffer may be Tris buffer, although other buffers can also be used as long as the buffer components are non-inhibitory to enzyme components of the methods of the invention. The pH is preferably from about 5 to about 11, more preferably from about 6 to about 10, even more preferably from about 7 to about 9, and most preferably from about 7.5 to about 8.5. The reaction medium can also include bivalent metal ions such as Mg^{2+} or Mn^{2+}, at a final concentration of free ions that is within the range of from about 0.01 to about 15 mM, and most preferably from about 1 to 10 mM. The reaction medium can also include other salts, such as KCl or NaCl, that contribute to the total ionic strength of the medium. For example, the range of a salt such as KCl is preferably from about 0 to about

IDT Exhibit 1020.10004

125 mM, more preferably from about 0 to about 100 mM, and most preferably from about 0 to about 75 mM. The reaction medium can further include additives that could affect performance of the reactions, but that are not integral to the activity of the enzyme components of the methods. Such additives include proteins such as BSA, single strand binding proteins (for e.g., T4 gene 32 protein), and non-ionic detergents such as NP40 or Triton. Reagents, such as DTT, that are capable of maintaining enzyme activities can also be included. Such reagents are known in the art. Where appropriate, an RNase inhibitor (such as Rnasin) that does not inhibit the activity of the RNase employed in the method can also be included. Any aspect of the methods of the invention can occur at the same or varying temperatures. Preferably, the reactions generating DNA double stranded complex (particularly, primer extension other than the first and second strand cDNA synthesis steps, and strand displacement) are performed isothermally, which avoids the cumbersome thermocycling process. The reactions are carried out at a temperature that permit primer extension, hybridization, and recombination as described herein, and that does not substantially inhibit the activity of the enzymes employed. The temperature can be in the range of any of about 15°C to about 85°C, about 25°C to about 85°C, about 30°C to about 80°C, and about 37°C to about 75°C. In some embodiments that include RNA transcription (for example, when RNA target is produced from DNA), the temperature for the transcription step is lower than the temperature(s) for the following steps (such as denaturation of the RNA target). In these embodiments, the temperature of the transcription steps can be in the range of about 25°C to about 85°C, about 30°C to about 75°C, and about 37°C to about 70°C. If cleavage of RNA is to be effected by heat, incubation occurs at a temperature sufficient to effect the desired cleavage without denaturation the first primer extension product (before cleavage), which can be about 65°C to about 80°C or about 70°C to about 75°C. Extension is performed at a temperature suitable to preserve sufficient enzyme activity.

Nucleotide and/or nucleotide analogs, such as deoxyribonucleoside triphosphates, that can be employed for synthesis of the primer extension products in the methods of the invention are provided in the amount of from preferably about 50 to about 2500 μM, more preferably about 100 to about 2000 μM, even more preferably about 200 to about 1700 μM, and most preferably about 250 to about 1500 μM. In some embodiments, a nucleotide or nucleotide analog whose presence in the primer extension strand enhances displacement of the strand (for example, by causing base pairing that is weaker than conventional AT, CG

IDT Exhibit 1020.10005

base pairing) is included. Such nucleotide or nucleotide analogs include deoxyinosine and other modified bases, all of which are known in the art. Nucleotides and/or analogs, such as ribonucleoside triphosphates, that can be employed for synthesis of the RNA transcripts in the methods of the invention are provided in the amount of from preferably about 0.25 to about 6 mM, more preferably about 0.5 to about 5 mM, even more preferably about 0.75 to about 4 mM, and most preferably about 1 to about 3 mM.

The oligonucleotide components of the reactions of the invention are generally in excess of the number of target nucleic acid sequence. They can be provided at about or at least about any of the following: 10, 10^2, 10^4, 10^6, 10^8, 10^{10}, 10^{12} times the amount of target nucleic acid. Composite primers can each be provided at about or at least about any of the following concentrations: 50 nM, 100 nM, 500 nM, 1000 nM, 2500 nM, 5000 nM.

In one embodiment, the foregoing components are added simultaneously at the initiation of the methods for generating recombined products. In another embodiment, components are added in any order prior to or after appropriate timepoints during the process, as required and/or permitted by the reaction. Such timepoints, some of which are noted below, can be readily identified by a person of skill in the art. The enzymes used for primer extension, and recombination according to the methods of the invention can be added to the reaction mixture either prior to the target nucleic acid denaturation step (denaturation in this context generally refers to subjecting target RNA to conditions which open sequence(s) to make them accessible), following the denaturation step, or following hybridization of the primer to the target RNA, following generation of double stranded complex comprising a 3' overhang, or following hybridization of double stranded complex comprising a 3' overhang to recipient molecule, as determined by their thermal stability and/or other considerations known to the person of skill in the art. The first strand cDNA (composite primer extension product) and the second strand cDNA (second primer extension product) synthesis reactions can be performed consecutively, followed by the hybridization and recombination steps. In these embodiments, the reaction conditions and components may be varied between the different reactions.

The recombination process (including generation of double stranded complex comprising a single stranded 3' overhang) can be stopped at various timepoints, and resumed at a later time. Said timepoints can be readily identified by a person of skill in the art. One timepoint is at the end of first strand cDNA synthesis. Another timepoint is at the end of second strand cDNA synthesis. Another timepoint is after generation of partially

IDT Exhibit 1020.10006

double stranded complex comprising a 3' single stranded overhang. Another timepoint is after hybridization of partially double stranded complex with recipient nucleic acids. Methods for stopping the reactions are known in the art, including, for example, cooling the reaction mixture to a temperature that inhibits enzyme activity or heating the reaction mixture to a temperature that destroys an enzyme. Methods for resuming the reactions are also known in the art, including, for example, raising the temperature of the reaction mixture to a temperature that permits enzyme activity or replenishing a destroyed (depleted) enzyme. In some embodiments, one or more of the components of the reactions is replenished prior to, at, or following the resumption of the reactions. Alternatively, the reaction can be allowed to proceed (i.e., from start to finish) without interruption.

The reaction can be allowed to proceed without purification of intermediate complexes, for example, to remove primer. Products can be purified at various timepoints, which can be readily identified by a person of skill in the art. One timepoint is at the end of first strand cDNA synthesis. Another timepoint is at the end of second strand cDNA synthesis. Another timepoint is after generation of partially double stranded complex comprising a 3' single stranded overhang. Another timepoint is after hybridization of partially double stranded complex with recipient nucleic acids.

The following Examples are provided to illustrate, but not limit, the invention.

EXAMPLES

Example 1: Creating a double stranded cDNA with a 3' overhang of a unique sequence appended to the second strand cDNA from poly-A mRNA using a single composite primer

Poly-A mRNA from MOLT-4 cell line (CLONTECH 6587-1) was used. The process was in two steps: 1) synthesis of first cDNA strand; 2) synthesis of second cDNA strand and cleavage to produce a double stranded cDNA (with a 3' overhang or single stranded portion) from the total mRNA of the sample. The double stranded cDNA product comprises at one end an RNA/DNA heteroduplex, which is a substrate for RNase H. The sequence of the two strands of this heteroduplex portion is not related to the target, and is incorporated through utilization of a composite (first) primer.

Primer sequences:

MTA1: *GACGGAUGCGGUCU*TTTTTTT

MTA2: *GACGGAUGCGGUCU*TTTTTTTN

IDT Exhibit 1020.10007

MTA3: *GACGGAUGCGGUCU*TTTTTTTNN

wherein italicized nucleotides denote ribonucleotides and "N" denotes a degenerate nucleotide (i.e., it can be A, T, C or G).

Step 1: Synthesis of the first strand cDNA from poly A mRNA

0.1 μg of total poly-A mRNA was mixed with the following reagents in a total volume of 10 ul:

0.2 μl primer MTA3 (100 μM)
0.5 μl dNTPs (25 mM)
0.1 μl Rnasin
0.1 μl DTT
2 μl 5X AMV reverse transcriptase reaction buffer
DEPC treated water to 10 μl total volume

The reaction mixture was incubated for 2 min at 75°C, and then cooled to 37°C. 1 μl AMV reverse transcriptase (USB 70041Y, 15U/μl) was added to each reaction and the reaction mixture further incubated at this temperature for 60 min.

Step 2: Second strand cDNA synthesis and cleavage to generate a 3' single stranded portion (overhang)

The first strand cDNA reaction mixture was mixed with 10 μl of the second strand cDNA synthesis mixture containing the following:

1 μl 10X Klenow reaction buffer
0.1 μl dNTPs (25 mM)
0.5 μl Klenow (USB 2141Y 5U/μl) DNA polymerase
0.1 μl of *E. coli* Ribonuclease H (BRL 18021-014, 4 U/μl
8.4 μl water

IDT Exhibit 1020.10008

The reaction mixture was incubated for 30 min at 37°C, followed by heating to 75°C for 5 min to stop the reactions by inactivating the enzymes.

Example 2: Characterization of products of step 2 reaction of Example 1

In the reaction of Example 1, a "unique" sequence (i.e., a sequence not hybridizable to the RNA template) is expected to be created at the 3'-end of the second strand cDNA due to the "unique" sequence of the 5' RNA portion of the composite primer used. This sequence (of the 3'-end of the second strand cDNA) is complementary to the 5'-RNA portion of the composite primer and is not related to sequences in the target RNA. To determine the presence of this sequence in the second strand cDNA that is obtained, PCR amplification of the reaction products (as found in reaction mix of step 2 of Example 1) was performed using a primer which is complementary to the expected sequence at the 3'-end of the second strand cDNA, as a forward primer, and a G3PDH-specific primer as a reverse PCR primer. This primer pair would be expected to amplify a specific product from a double stranded cDNA that has the "unique" sequence.

PCR reactions were carried out as follows:

Each 50 µl of PCR reaction contains:

0.4 µM of each primer (Biosource International)

100 µM of each dNTP (Epicenter)

2 mM Magnesium chloride (Epicenter)

1-2 units Polymerase (either MasterAmp taq or MasterAmp Tfl, both from Epicenter)

5 µl 10X buffer as supplied with the enzyme.

A 1:20 dilution of the cDNA generated in step 2 in Example 1.

The PCR amplification cycles were 94°C for 30 seconds, 51°C for 30 seconds, and 72°C for 30 seconds. Generally, the samples were cycled 20 or 25 times. There was a final 5-minute extension at 72°C before the samples were held at 4°C.

Similar experiments were carried out with primer specific to the T-cell receptor specific mRNA (TCR) which is expressed by the MOLT4 cell line.

Expected PCR product size (base pairs) using the G3PDH primers.

IDT Exhibit 1020.10009

REV PRIMER	FORWARD PRIMER G3PDH3	FORWARD PRIMER dMTA1
G3PDH5-2	18	62
G3PDH5-3	110	156
G3PDH5-4	157	203
G3PDH5	253	299
G3PDH5-6	309	354
G3PDH5-7	361	405

Primer sequences

G3PDH5: 5' TTT CCT GGT ATG ACA ACG AA

G3PDH5-4: 5' CCA GCA AGA GCA CAA GAG GA

G3PDH3: 5' GAT GGT ACA TGA CAA GGT

dMTA1: 5' GAC GGA TGC GGT CTT TTT TTT

Expected PCR product size (base pairs) using T-Cell receptor primers

	TCR3	DMTA1
TCR5-2	160	Approx. 440
TCR5	238	Approx. 500

Primer sequences

TCR5: 5' CCC GCA ACC ACT TCC GCT GTC

TCR5-2: 5' CAA ACC CGT CAC CCA GAT CGT

TCR3: 5' CAA CAC AAG GGC GCT GAC C

The results show that the unique sequence is incorporated into the second strand cDNA, as indicated by the presence of a product that was about 250 base pairs in length when the step 2 reaction mix was PCR amplified using primers DMTA1 and G3PDH5 (amplification of a sequence of G3PDH mRNA), and a product of about 400 base pairs in length when using primers DMTA1 and TCR5-2 (amplification of a sequence of TCR beta chain mRNA). Thus, the results demonstrated the incorporation of the "unique" sequence (of the RNA portion of the composite primer used in Example 1) into the partially double stranded cDNA products generated.

IDT Exhibit 1020.10010

Although the foregoing invention has been described in some detail by way of illustration and example for purposes of clarity of understanding, it will be apparent to those skilled in the art that certain changes and modifications may be practiced. Therefore, the descriptions and examples should not be construed as limiting the scope of the invention, which is delineated by the appended claims.

IDT Exhibit 1020.10011

CLAIMS

What is claimed is:

1. A method for generating a hybrid nucleic acid, said method comprising:

(a) preparing a complex of first and second primer extension products said complex comprising a 3' single stranded portion, wherein said complex comprising a 3' single stranded portion is prepared according to a method comprising:

(i) extending a first primer hybridized to a target RNA with an enzyme comprising RNA-dependent DNA polymerase activity, wherein the first primer is a composite primer comprising an RNA portion and a 3' DNA portion, whereby a complex comprising a first primer extension product and the target RNA is produced;

(ii) cleaving RNA in the complex of step (i);

(iii) extending a second primer hybridized to the first primer extension product with an enzyme comprising DNA-dependent DNA polymerase activity and an enzyme comprising RNA-dependent DNA polymerase activity, whereby a second primer extension product is produced to form a complex of first and second primer extension products;

(iv) cleaving RNA from the composite primer in the complex of step (iii) with an enzyme that cleaves RNA from an RNA/DNA hybrid such that a complex comprising hybridized first and second primer extension product comprising a 3' single stranded portion is generated; and

(b) hybridizing the 3' single stranded portion of the complex of first and second primer extension products with a recipient nucleic acid molecule; whereby a hybrid nucleic acid is generated.

2. The method of claim 1, further comprising: (c) generating a recombined nucleic acid from the hybrid nucleic acid.

3. The method of claim 1 or 2, wherein the second primer comprises a fragment of the target RNA hybridized to the primer extension product, said fragment generated by cleaving RNA in the complex of step (a).

4. The method of claim 1 or 2, wherein the second primer comprises DNA.

IDT Exhibit 1020.10012

5. The method of claim 1 or 2, wherein the RNA portion of the composite primer that hybridizes to target RNA is 5' with respect to the 3' DNA portion.

6. The method of claim 5, wherein the 5' RNA portion is adjacent to the 3' DNA portion.

7. The method of claim 1 or 2, wherein the target RNA is mRNA, and the composite primer that hybridizes to target RNA comprises a poly-dT sequence, and further comprises a 5' portion that does not hybridize to the target mRNA under conditions which the composite primer hybridizes to the target mRNA.

8. The method of claim 1 or 2, wherein the composite primer that hybridizes to target RNA comprises a random sequence.

9. The method of claim 1 or 2, wherein a plurality of different composite primers are used for hybridizing to the target RNA.

10. The method of claim 1 or 2, wherein the RNA portion of the composite primer consists of about 10 to about 20 nucleotides and the DNA portion of the composite primer consists of about 5 to about 20 nucleotides.

11. The method of claim 1 or 2, wherein the enzyme that cleaves RNA from an RNA/DNA hybrid is RNase H.

12. The method of claim 1 or 2, wherein the same enzyme comprises RNA-dependent DNA polymerase activity and cleaves RNA from an RNA/DNA hybrid.

13. The method of claim 1 or 2, wherein the same enzyme comprises DNA-dependent DNA polymerase activity and cleaves RNA from an RNA/DNA hybrid.

14. The method of claim 1 or 2, wherein the same enzyme comprises DNA-dependent DNA polymerase activity, RNA-dependent DNA polymerase activity and cleaves RNA from an RNA/DNA hybrid.

IDT Exhibit 1020.10013

15. The method of claim 1 or 2, wherein different enzymes comprise RNA - dependent DNA polymerase activity and DNA-dependent DNA polymerase activity.

16. The method of claim 1 or 2, wherein different enzymes comprise RNA - dependent DNA polymerase activity and cleave RNA from an RNA/DNA hybrid.

17. The method of claim 1 or 2, wherein different enzymes comprise DNA-dependent DNA polymerase activity and cleave RNA from an RNA/DNA hybrid.

18. The method of claim 1 or 2, wherein different enzymes comprise DNA-dependent DNA polymerase activity, RNA-dependent DNA polymerase activity and cleave RNA from an RNA/DNA hybrid.

19. The method of claim 1 or 2, wherein the entire 3' single stranded portion of the complex is hybridized to the recipient molecule.

20. The method of claim 1 or 2, wherein a portion of the 3' single stranded portion of the complex is hybridized to the recipient molecule.

21. The method of claim 1 or 2, wherein the cleaving RNA of step (ii) is by an enzyme that cleaves RNA from an RNA/DNA hybrid.

22. The method of claim 21, wherein the enzyme is RNaseH.

23. The method of claim 2, wherein the generating a recombined nucleic acid comprises ligation.

24. The method of claim 2, wherein the generating a recombined nucleic acid comprises extension of a 3' end by an enzyme having DNA polymerase activity.

25. The method of claim 24, wherein the enzyme having DNA polymerase activity has strand displacement acitivity.

IDT Exhibit 1020.10014

26. The method of of claim 2, wherein generating a recombined nucleic acid comprises gap filling and ligation.

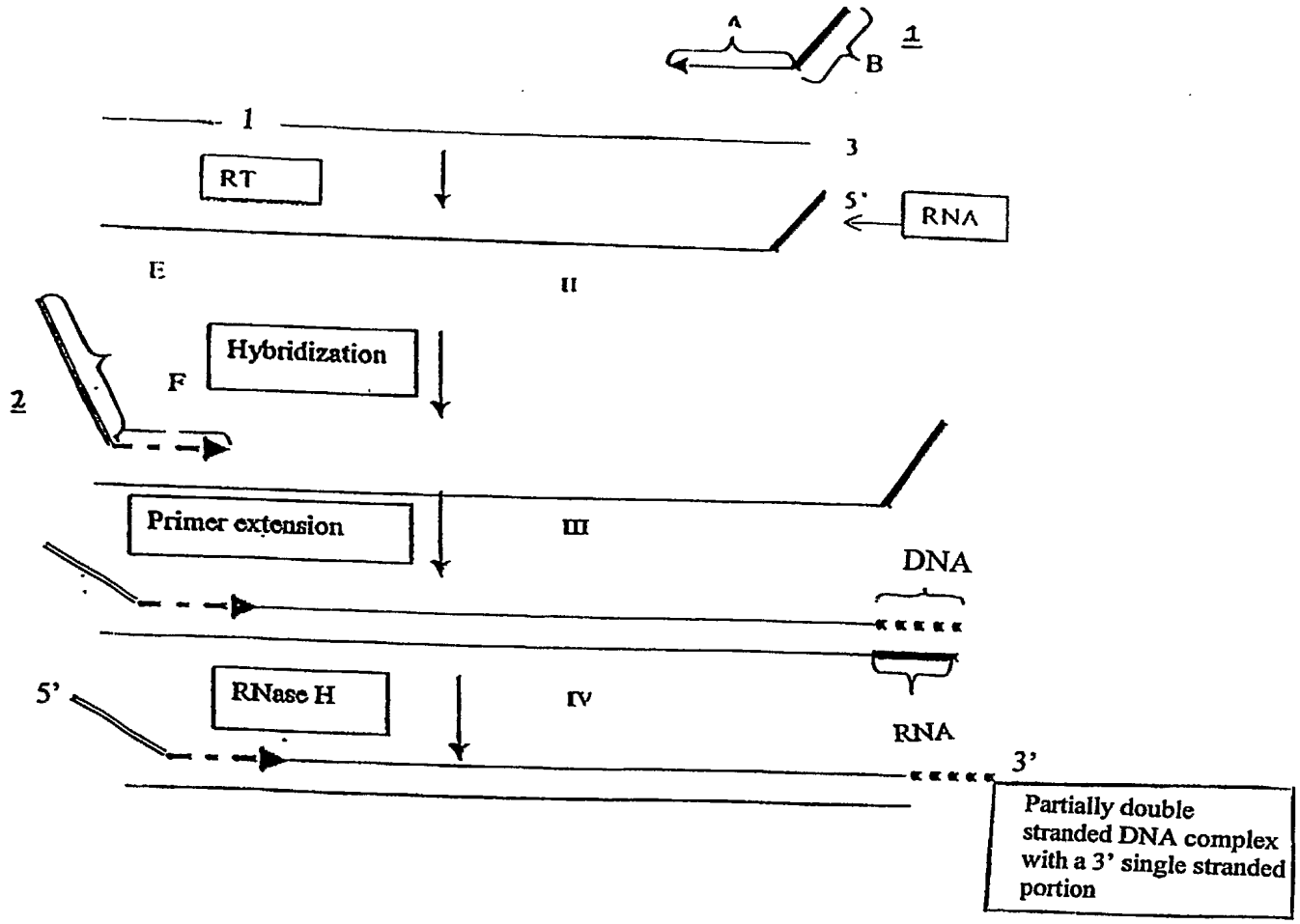
27. The method of claim 3, wherein cleavage is with an enzyme that cleaves RNA from an RNA/DNA hybrid.

28. The method of claim 27, wherein the enzyme is RNaseH.

29. A method of making a hybrid nucleic acid, said methods comprising: hybridizing a 3' single stranded portion of a complex of first and second primer extension products with a recipient nucleic acid molecule, whereby a hybrid nucleic acid is generated, wherein said complex of first and second primer extension products is prepared according to a method comprising: (i) extending a first primer hybridized to a target RNA with an enzyme comprising RNA-dependent DNA polymerase activity, wherein the first primer is a composite primer comprising an RNA portion and a 3' DNA portion, whereby a complex comprising a first primer extension product and the target RNA is produced; (ii) cleaving RNA in the complex of step (i); (iii) extending a second primer hybridized to the first primer extension product with an enzyme comprising DNA-dependent DNA polymerase activity and an enzyme comprising RNA-dependent polymerase activity, whereby a second primer extension product is produced to form a complex of first and second primer extension products; (iv) cleaving RNA from the composite primer in the complex of step (iii) with an enzyme that cleaves RNA from an RNA/DNA hybrid such that a complex comprising hybridized first and second primer extension product comprising a 3' single stranded portion is generated.

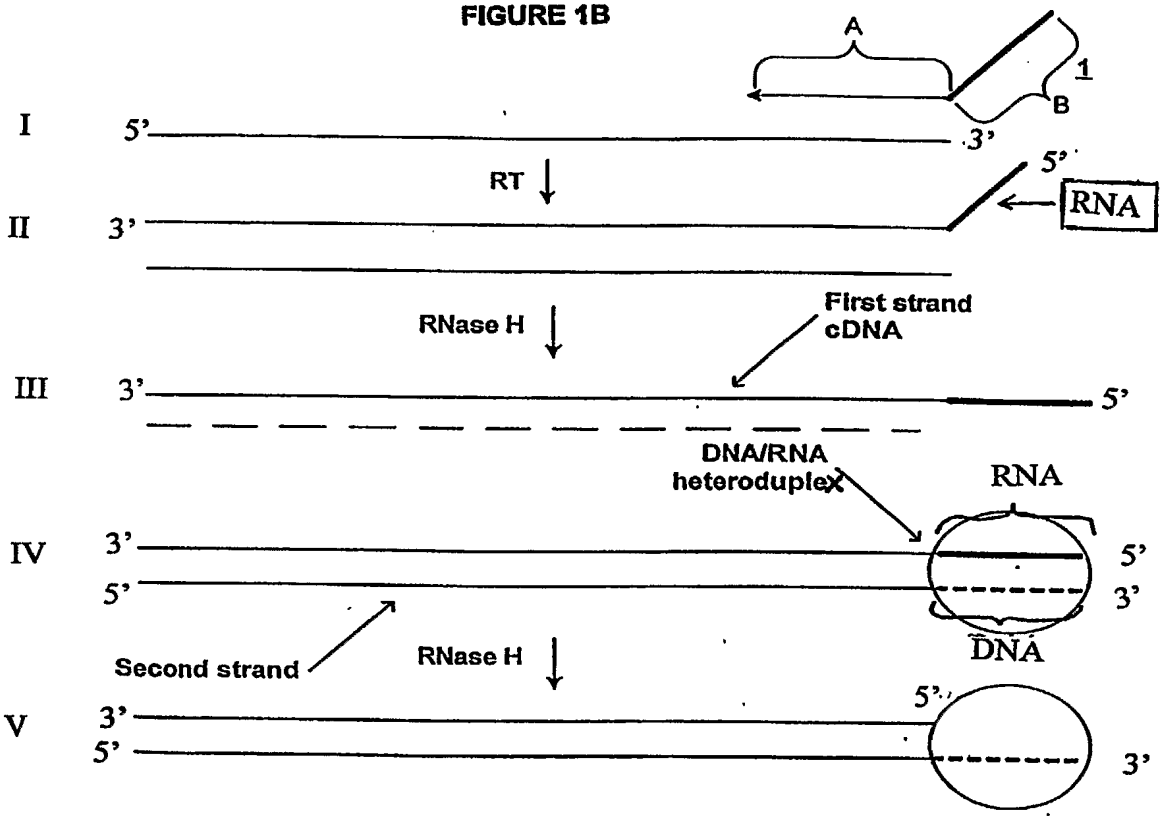
IDT Exhibit 1020.10015

FIGURE 1A



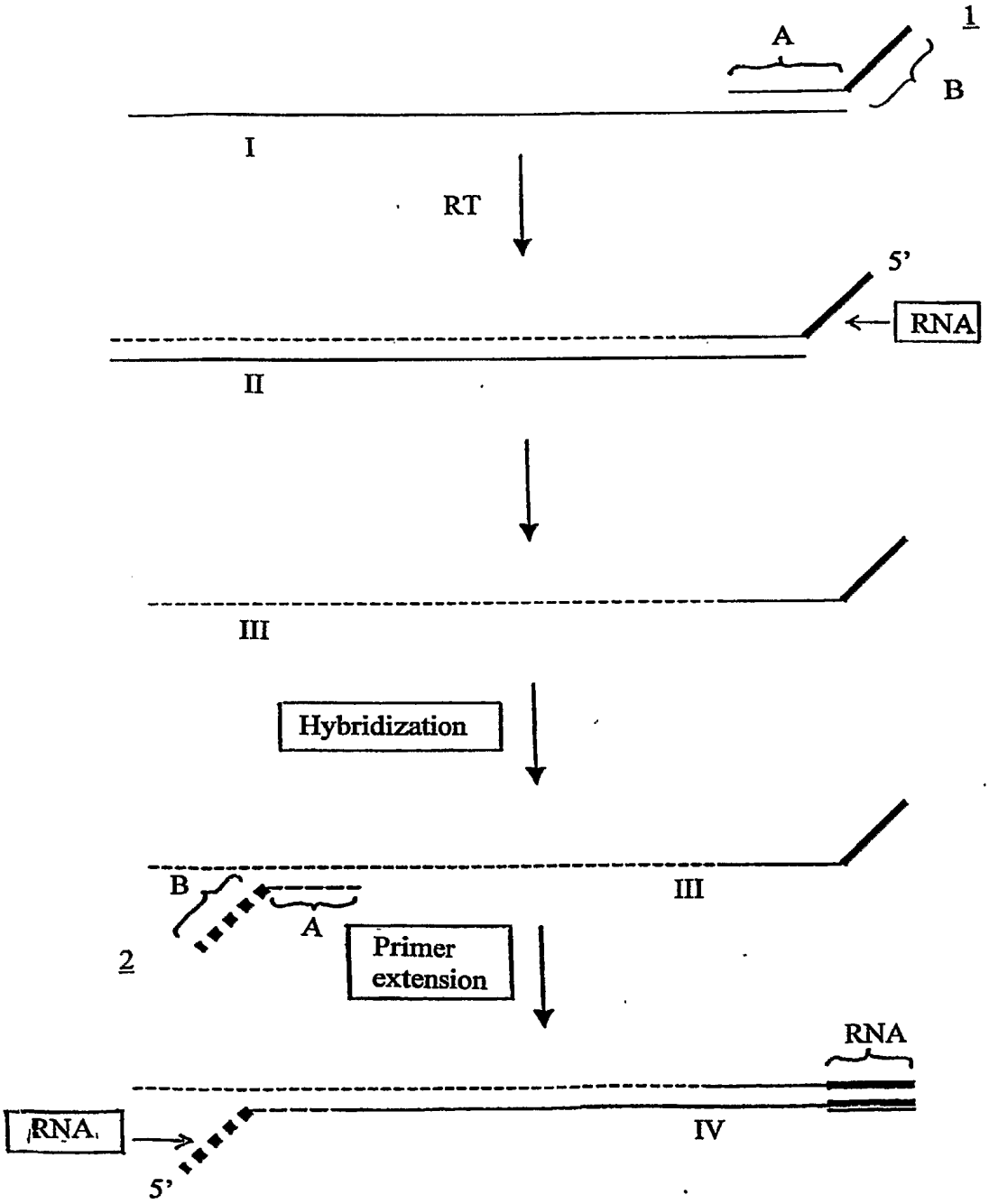
IDT Exhibit 1020.10016

FIGURE 1B



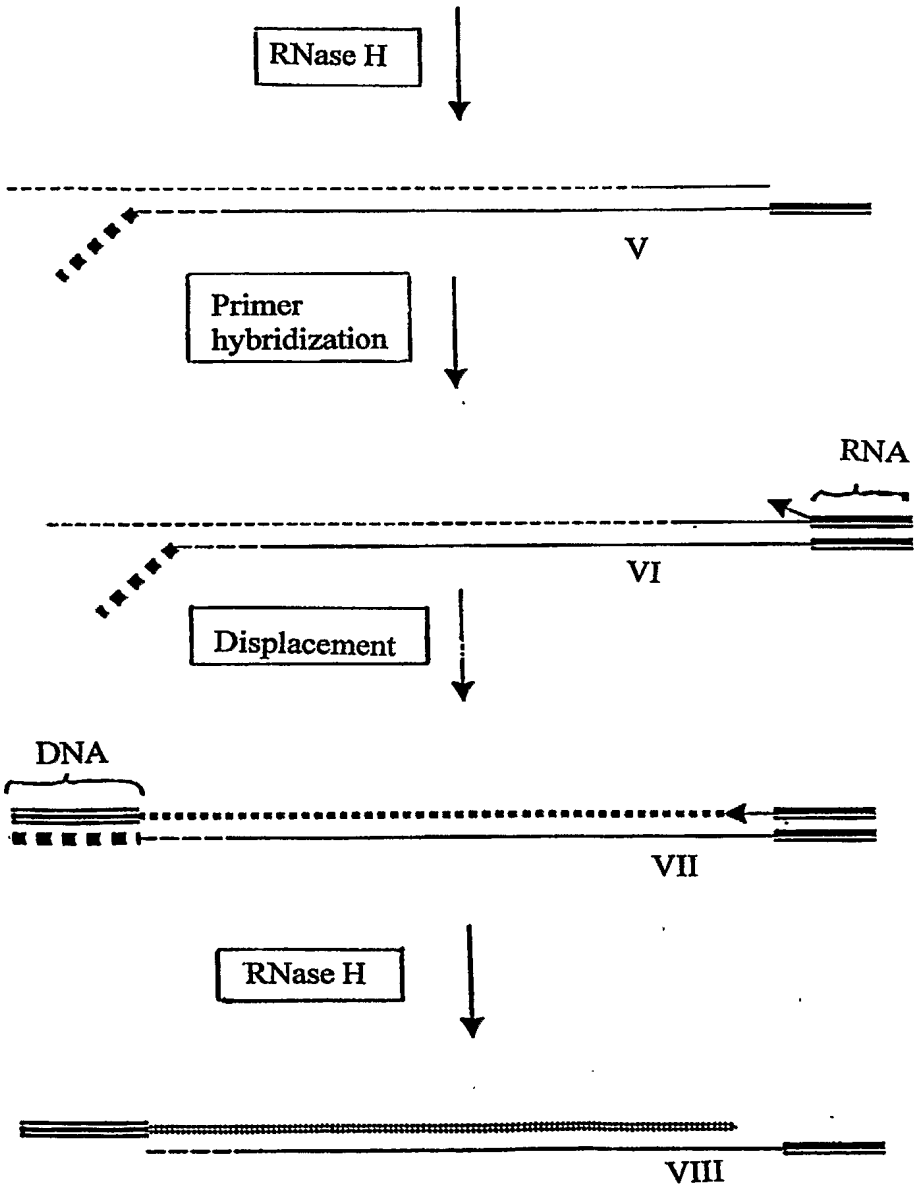
IDT Exhibit 1020.10017

FIGURE 2A



IDT Exhibit 1020.10018

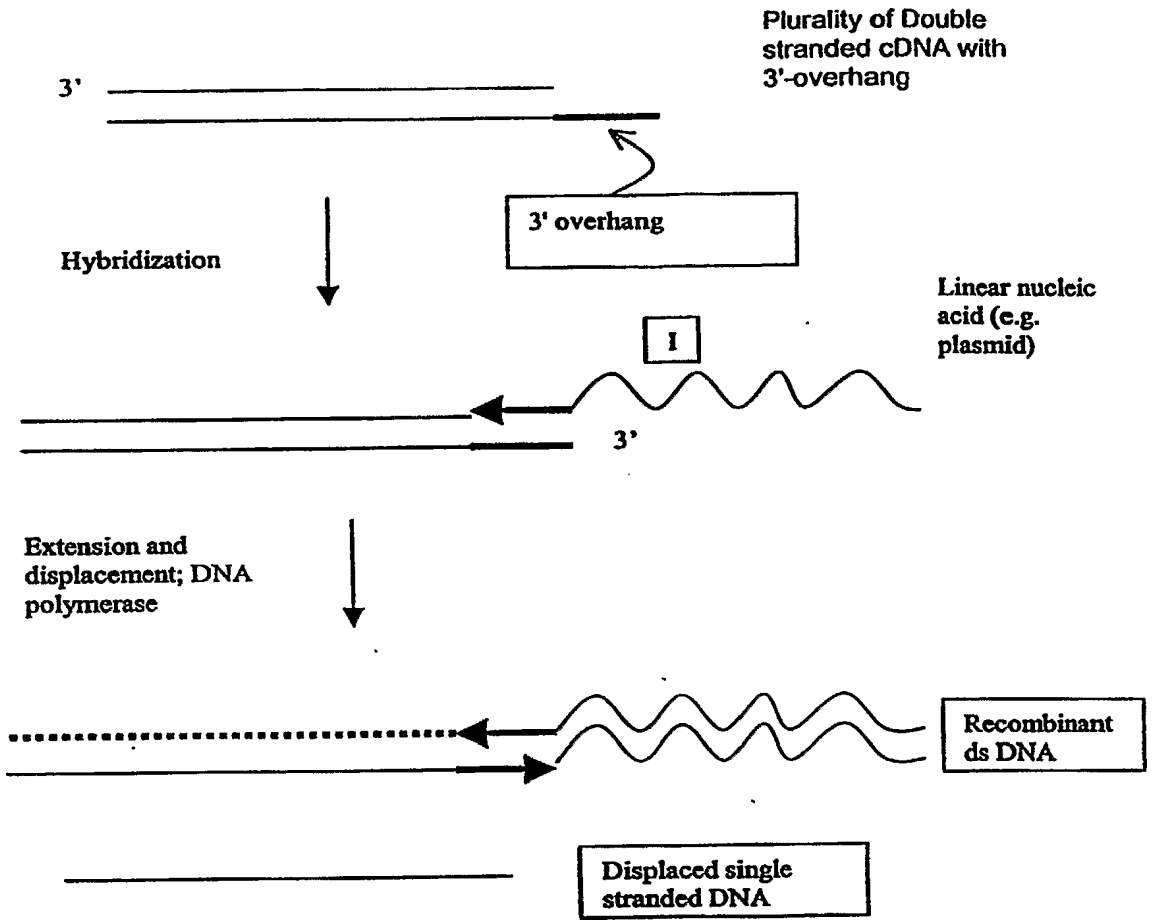
FIGURE 2B



IDT Exhibit 1020.10019

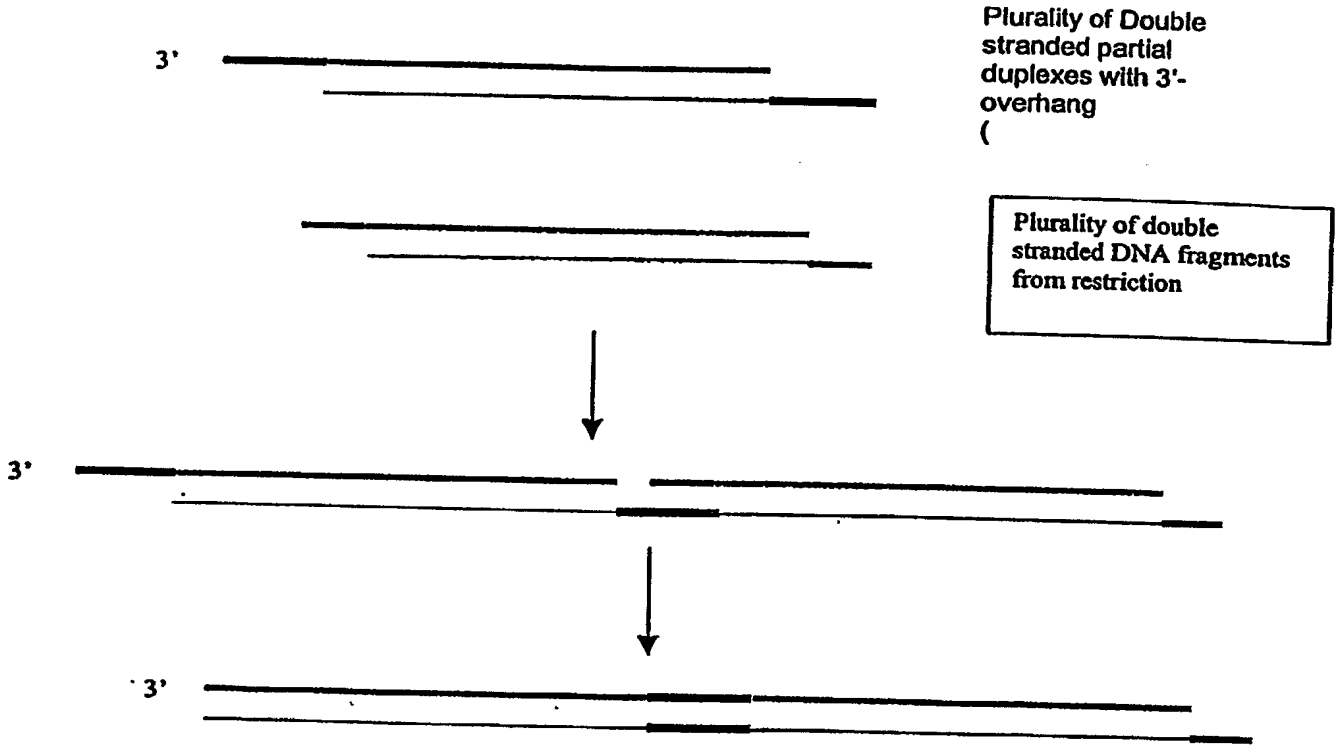
5/8

FIGURE 3



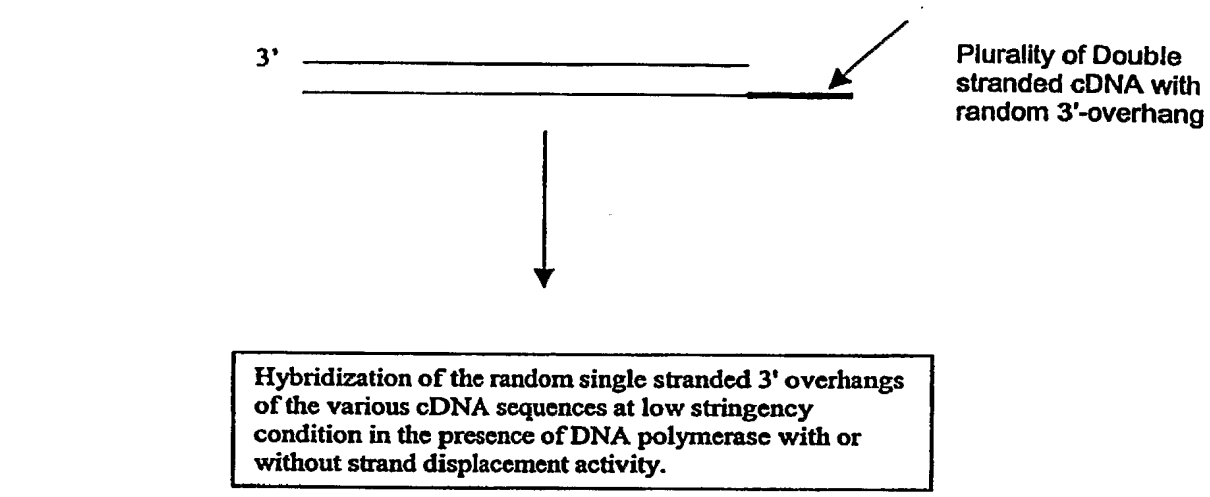
IDT Exhibit 1020.10020

FIGURE 4



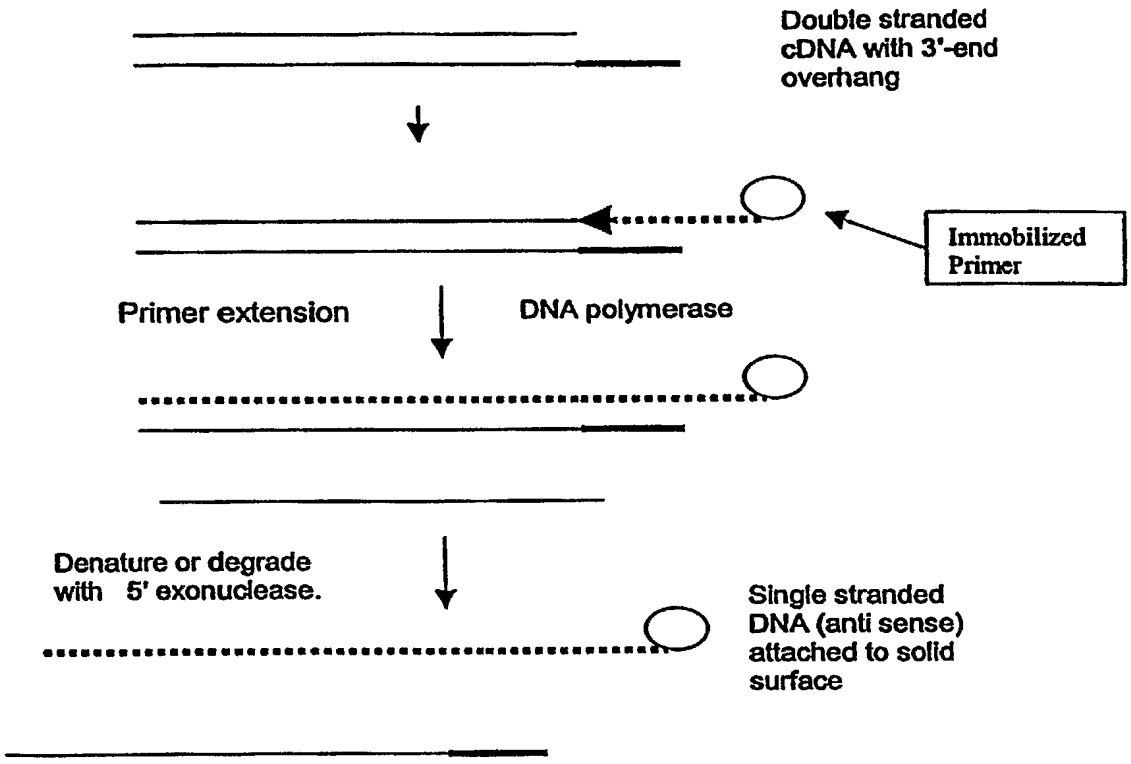
IDT Exhibit 1020.10021

FIGURE 5



IDT Exhibit 1020.10022

FIGURE 6



IDT Exhibit 1020.10023

Espacenet

Bibliographic data: CA2444926 (A1) — 2002-11-14

METHOD FOR LABELLING AND FRAGMENTING DNA

Inventor(s): LHOMME JEAN [FR]; TREVISIOL EMMANUELLE [FR]; TORA CHRISTELLE [FR]; SOTHIER ISABELLE [FR]; LAAYOUN ALI [FR]; BOURGET CECILE [FR]; KOTERA MITSUHARU [FR] ± (LHOMME, JEAN, ; TREVISIOL, EMMANUELLE, ; TORA, CHRISTELLE, ; SOTHIER, ISABELLE, ; LAAYOUN, ALI, ; BOURGET, CECILE, ; KOTERA, MITSUHARU)

Applicant(s): CENTRE NAT RECH SCIENT [FR]; BIO MERIEUX [FR]; UNIV GRENOBLE 1 [FR] ± (CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE, ; BIO MERIEUX, ; UNIVERSITE JOSEPH FOURIER (GRENOBLE 1))

Classification: **- international:** ***C12N15/09; C12Q1/68;*** (IPC1-7): C12Q1/68
- cooperative: **C12Q1/68 (EP)**

Application number: CA20022444926 20020503

Priority number(s): FR20010006039 20010504 ; WO2002FR01542 20020503

Also published as: CA2444926 (C) AT340269 (T) AU2002313026 (B2) CN1325659 (C) CN1639356 (A) more

Abstract not available for CA2444926 (A1)
Abstract of corresponding document: WO02090584 (A2)

The invention concerns a method for labelling and fragmenting a single-stranded or double-stranded deoxyribonucleic acid (DNA) comprising the following steps: chemically fragmenting the DNA by generating an abasic site on said DNA, attaching a marker on at least one of the fragments by means of a labelling reagent, said reagent being covalently and majority bonded on at least one phosphate of said fragment. The invention is applicable in the field of diagnosis.

IDT Exhibit 1020.10024

Patent Translate
Powered by EPO and Google

Notice

This translation is machine-generated. It cannot be guaranteed that it is intelligible, accurate, complete, reliable or fit for specific purposes. Critical decisions, such as commercially relevant or financial decisions, should not be based on machine-translation output.

DESCRIPTION CA2444926C

[0001]

CA 02444926 2003-10-17 WO 02/090584 PCT / FR02 /01542 The present invention relates to a method of DNA fragmentation and labeling as well as the applications of this method, in particular in the field of diagnostics. The state of the art shows that many methods exist for labeling nucleic acids. A first method consists in fixing the marker on the base, whether the latter is natural or modified. A second method involves attaching the marker to the sugar, again either natural or modified. A third method relates to the attachment of the marker to the phosphate. Labeling on the base has been used in particular in the approach for labeling nucleic acids by incorporation of directly labeled nucleotides. Labeling on the sugar is often used in the case of oligonucleotides prepared by chemical synthesis. Labeling on the phosphate was also used to introduce functionalized arms and markers during the chemical synthesis of oligonucleotides.

In fact the person skilled in the art, who must carry out the labeling of a nucleotide, or of a nucleotide analogue or of a nucleic acid, is inclined to carry out this fixing on the base or on the sugar which offers him more convenience and alternatives. This is what emerges from the study of numerous documents, such as EP-A-0.329.198, EP-A-0.302.175, EP- A- 0.097.373, EP-A-0.063.879 , US-A-5,449,767, US-A-5,328,824, WO-A-93/16094, DE-A-39 10 151, EP-A-0.567.841 for the base or EP-A-0.286.898 for the sugar . The attachment of the label to the phosphate is a more complex technique than the technique consisting in functionalizing the base or the sugar and has been used much less in particular because of the low reactivity of the phosphate (see for example Jencks W.P. et al J. Amer. Chem. Soc., 82, 1778-1785, 1960). Likewise in the review by O'Donnel and Mc Laughlin ("Reporter groups for the analysis of nucleic acid structure", p 216-243, in "Bioorganic Chemistry: Nucleic Acids", Ed Hecht SM, Oxford University Press, 1996) on on methods of introducing probes into oligonucleotide fragments,

[0002]

CA 02444926 2003-10-17 WO 02/090584 PCT / FR02 /01542

efficient alkylation of the internucleotide phosphodiester is considered to be impossible. A second problem arises for the labeling of nucleic acids especially for nucleic acids of large size, that is to say of more than one hundred (100) nucleotides, which must hybridize with nucleic acid probes. This problem is related to steric discomfort or to the lack of specificity between the nucleic acid and the nucleic acid probe. This results in a loss of detection sensitivity. 1o The steric hindrance can be the fact, not only of la length of the nucleic acid, but also of the existence or the conservation of secondary structures. Fragmentation makes it possible to destroy these structures and thus to optimize hybridization. This steric gene plays a particularly important role in the case of hybridization on surfaces containing high density capture probes, for example the DNA chips developed by the company Affymetrix ("Accessing Genetic Information with High-Density DNA arrays ", M. Shee et al., Science, 274, 610-614

"Light-generated oligonucleotide arrays for rapid DNA sequence analysis", A. Caviani Pease et al., Proc. Natl. Acad. Sci. USA, 1994, 91, 5022- 5026). As regards the fragmentation of nucleic acids, many methods are described in the state of the art. First, the fragmentation can be enzymatic, i.e. the fragmentation of nucleic acids can be carried out by nucleases (DNases). Second, the fragmentation can be chemical. For example in the case of DNA, the depurination or depyrimidination of the DNA can be carried out, which is then fragmented in the presence of a base by a so-called "~ i-elimination" mechanism. DNA fragmentation

IDT Exhibit 1020.10025

can be achieved by mechanisms of oxidation, alkylation, addition of free radicals among others. Finally, the fragmentation can be physical, for example by sonication or by photochemical 3o route.

[0003]

CA 02444926 2003-10-17 WO 02/090584 PCT / FR02 /01542

The difficulty consists in fact in combining these two stages of fragmentation and labeling. Patent application WO-A-99/65926 describes a method for labeling a synthetic or natural ribonucleic acid (RNA) which consists in fragmenting the RNA and in labeling at the terminal phosphate level. This document describes a certain number of reactive functions which can be used for labeling in connection with fragmentation. These functions make it possible to label the RNAs but a fragmentation step must be associated to have an effective labeling because this labeling occurs on the phosphate released during the fragmentation. 1o In addition, it is necessary to add a large excess of labeling reagent relative to the RNA to obtain effective labeling, which induces problems of background noise generated by the excess label. Finally, this method is not applicable to double-stranded DNA. There is therefore a need for a simple and efficient DNA labeling technique which allows good labeling yield and therefore good sensitivity, which is specific at the level of the labeling position, and in particular which does not affect the labeling position. properties of hybridization of the bases involved in the formation of the double helix via the hydrogen bonds, and finally which allows the fragmentation of the DNA with the aim of hybridizing these DNA fragments labeled on nucleic probes, and in particular on 2o nucleic acid probes fixed on a solid support.

'The present invention describes a method for labeling and fragmentation of a single or double stranded deoxyribonucleic acid (DNA) comprising the following steps - fragmenting the DNA by creating at least one abasic site on said DNA, - attaching a marker to the at least one of the fragments via a labeling reagent, said reagent coupling covalently and predominantly to at least one phosphate of said fragment. The fragmentation and the labeling is carried out in one step or in two steps and the 3o labeling can be carried out either before, after or simultaneously with the fragmentation.

[0004]

CA 02444926 2003-10-17 WO 02/090584 PCT / FR02 /01542

Preferably, the labeling and / or the fragmentation is carried out in a substantially aqueous homogeneous solution. Preferably, the labeling and the fragmentation are carried out simultaneously, that is to say that the reagents necessary for these two steps are put together in a substantially aqueous solution with the nucleic acid, for example. This is particularly the case for chemical or enzymatic fragmentation. In the case of mechanical fragmentation by physical means, labeling and fragmentation taking place simultaneously means that the physical means is applied to a solution containing at least the nucleic acids and the labeling reagent. By substantially aqueous solution is meant a solution containing at least 50% water. This solution preferably contains salts such as a buffer solution. By homogeneous solution is meant a single-phase solution such as a water / DMSO solution, as opposed to a two-phase solution such as a water chloroform solution.

DNA fragmentation through the creation of an abasic site occurs enzymatically, chemically or physically. The chemical fragmentation of DNA is carried out by bringing the nucleic acid into contact with a chemical means for creating an abasic site. Examples of conditions of chemical fragmentation via an abasic site of DNA are given in G. Pratviel et al., Angew. Chem. Int. Ed. Engl., 34, 20 p746-769, 1995; G. Pratviel et al., Adv. Inorg. Chem., 45, p251-312, 1998; D. S. Sigman et al. Chem. Rev., 93, p2295-2316, 1993; J. Lhomme et al., Biopolymers (Nucleic Acid Sciences), 52, p65-83, 1999. An abasic site is created when the N-glycosidic bond connecting the modified base to deoxyribose is hydrolyzed, leaving the phosphodiester bond intact. This phenomenon is called depurination (in the case of purines) or depyrimidination (in the case of pyrimidines). Depurination is much more common than depyrimidination.

Chemical agents, such as acidic pH, oxidizing agents or alkylating agents, can induce the phenomenon of depurination and therefore the formation of abasic sites. These alkylating agents consist of electrophilic species which react with the nucleophilic sites of a DNA fragment. The nitrogen atom in position 7 of guanine is the

[0005]

CA 02444926 2003-10-17 WO 02/090584 PCT / FR02 / 01542 main DNA alkylation site. In addition to the protonation of purines, alkylation is one of the chemical modifications that can make the N-glycosidic bond more labile.

0 0 OH O H 5 Abasic site DNA oxidation can also generate so-called "alkalilabile" abasic lesions, which lead to DNA fragmentation in the presence of a base. For example, the ribonolactone residue can be generated by reacting the hydroxyl

IDT Exhibit 1020.10026

radical (OH) with the C1 'carbon. o 0 -o O H Ribonolactone The abasic site is very unstable. In its aldehyde form, it can easily undergo a ~ i-elimination of the phosphodiester fixed in position 3 ', which leads to the fragmentation of the DNA strand. Depurination is spontaneous under physiological conditions (pH 7.4 at 37 ° C) 2o but the speed of the reaction is very low of the order of 3.10-11 depurination per second, that is to say unusable for a efficient fragmentation. To increase the reaction rate, alkylating agents which weaken the N-glycosidic bond or enzymes, such as DNA glycosylases, are used.

[0006]

A preferred embodiment of the fragmentation is obtained by using an acidic pH, that is to say a pH less than 5, preferably less than 4. Advantageously, the pH is around 3. A sodium formate buffer at pH 3 makes it possible to efficiently fragment the nucleic acids according to the present invention. This buffer is compatible with the one-step labeling conditions as will be demonstrated in the examples. Even more advantageously, an acidic medium (HCl, carbonate, H2SOø) is used. In a particular embodiment of the present invention and in order to further increase 1o the fragmentation, the deoxyribonucleic acid contains at least one modified base capable of generating an abasic site more easily. Various modified bases can be used, such as N7-alkylpurines, N3-alkylpurines, 06-alkylpurines, 8-bromopurines, 8-thiopurines, alkylthiopurines, 8 azidopurines or 8-alkylsulfonylpurines. .

In the case where the nucleic acid to be labeled is generated by an enzymatic amplification technique such as PCR, the use of an 8-bromopurine makes it possible to have an efficient incorporation of said nucleotide during the amplification, which facilitates both the fragmentation and labeling process according to the invention, while retaining excellent sensitivity for the enzymatic amplification step. An 8-methylthiopurine is also incorporated without hindering the efficiency of amplification and incorporation by PCR. Once incorporated, the thioether base is oxidized by a peracid or a monopersulfate derivative such as potassium peroxymonosulfate for example, to the corresponding sulphonyl which is very labile. It has been shown that the half-life corresponding to the hydrolysis of an 8-methylsulfonylguanosine is less than 2 minutes while the half-life for its natural counterpart is 1000 hours.

By label is meant at least one label capable of directly or indirectly generating a detectable signal. A non-limiting list of these markers follows ~ enzymes which produce a signal detectable for example by colorimetry, fluorescence, luminescence, such as horseradish peroxidase, alkaline phosphatase, ~ i-galactosidase, glucose-6-phosphate dehydrogenase,

[0007]

~ chromophores such as fluorescent, luminescent and dye compounds, ~ electron density groups detectable by electron microscopy or by their electrical property such as conductivity, amperometry, voltammetry, impedance, ~ detectable groups, for example including the molecules are of sufficient size to induce detectable modifications of their physical and / or chemical characteristics, this detection can be carried out by optical methods such as diffraction, surface plasmon resonance, surface variation, variation in angle of contact or physical methods such as atomic force spectroscopy, tunneling, and radioactive molecules such as 32P, 3sS or Iasi. Preferably, the label is not a radioactive label in order to avoid the security problems associated with these labels. In a particular embodiment of the present invention, the label is electrochemically detectable and in particular the label is a derivative of an iron complex such as a ferrocene.

Indirect systems can also be used, such as, for example, ligands capable of reacting with an anti-ligand. The ligand / anti-ligand pairs are well known to those skilled in the art, which is the case for example with the following pairs: biotin / streptavidin, hapten / antibody, antigen / antibody, peptide / antibody, sugar / lectin, polynucleotide / 2o complementary to the polynucleotide. In this case, it is the ligand which carries the reactive function. The anti-ligand can be detectable directly by the markers described in the preceding paragraph or be itself detectable by another ligand / anti-ligand pair. Another example of indirect systems uses a specific covalent bond between the ligand and the anti-ligand, for example methyl ketone and alkoxyamine. Examples of this system are described in patent applications WO-A-00/40590 and WO-A-9/05766. In the method of the present invention, the binding between the labeling reagent and the nucleic acid is covalent, but it is described above that non-covalent interactions can be used in particular in stacking systems or in the case where the 3o marker is indirectly detectable.

The term "attach" therefore covers these different possibilities.

[0008]

IDT Exhibit 1020.10027

These indirect detection systems can lead, under certain conditions, to an amplification of the signal and reference may be made to previous patent applications WO-A-00/07982, WO-A-01/92361 and WO-A-95. / 08000 from the Applicant for examples of chemical amplification using polymers or at the request WO-A-01/44506, also from the Applicant, for chemical amplification systems by stacking. In a particular mode of signal amplification, at least two markers are present on the labeling reagent. In a preferred embodiment of the invention, the tracer is a low fluorescent compound! o steric hindrance such as fluorescein, dansyl, IR type chromophores (LI-COR Inc, Lincoln NE, USA), cyanine derivatives such as Cy5 and Cy3 (Randolph JB et al, Nucleic Acids Res., 25 (14) , p2923-2929, 1997) and in particular the derivatives of Cy5 or else the tracer is a hapten of low steric hindrance such as biotin or a derivative of abietane (see application WO-A-00/07982).

By low steric hindrance is meant a molecular weight of less than 1000 g / mole. In the case of a fluorophore, it is preferable to work with fluorophores whose excitation wavelength is greater than 450 nm, preferably greater than 600 nm. In the case where the tracer is a hapten which does not produce a signal by itself, 2o such as for example biotin, the detection is carried out by the reaction of an anti-ligand as described above. In the case of biotin, use is preferably made of streptavidin or an anti-biotin antibody coupled to a fluorescent compound such as fluorescein, Cy5 or phycoerythrin. In the case of abietane, a monoclonal antibody is used as described in WO-A-00/07982. The term "deoxyribonucleic acid" or DNA means a sequence of at least two deoxyribonucleotides optionally comprising at least one modified nucleotide, for example at least one nucleotide comprising a modified base such as finosine, 5-methyl = deoxycytidine, 5-dimethylamino -deoxyuridine, deoxyuridine, diamino-2,6-purine, bromo-5-deoxyuridine or any other modified base allowing (hybridization.

[0009]

CA 02444926 2003-10-17 WO 02/090584 PCT / FR02 /01542

This DNA can also be modified at the level of the internucleotide bond, such as, for example, phosphorothioates, H-phosphonates, alkyl-phosphonates, at the level of the backbone such as, for example, alpha-oligonucleotides (FR-A-2,607,507) or PNA (M. Egholm et al., J. Am. Chem. Soc., 114, 1895-1897, 1992). DNA is natural or synthetic, and / or in fragment form. DNA is single or double stranded. In particular the DNA is obtained by an enzymatic amplification technique such as ~ PCR (Polymerase Chain Reaction), described in US-A-4,683,195, US-A-4,683,202 and US-A-4,800,159, and its derivative RT-PCR (Reverse Transcription-PCR), in particular in a one-step format as described in patent application EP-A- 0,569,272, ~ LCR (Ligase Chain Reaction), disclosed for example in the patent application EP-A- 0.201.184, ~ RCR (Repair Chain Reaction), described in patent application WO-A- 90/01069. We then speak of amplicons to designate the DNA generated by an enzymatic amplification technique.

The DNA can also comprise ribonucleotides in a small proportion, for example less than 10%. Each of these modifications can be taken in combination as long as at least one phosphate is present in the DNA. The labeling reagent comprises, as reactive function, a unit chosen from the following compounds: diazomethyl; alkyl halide; nitrosourea; spirocyclopropane; aziridine; epoxy; trifluorosulfonates. Responsive functions are described below

[0010]

CA 02444926 2003-10-17 WO 02/090584 PCT / FR02 / 01542 O diazomethyl ~ C = N2 nitrosourea -NC-NHRS I NO alkyl halide -CH2X spirocyclopropane HZC-CH2 O- CH2 epoxide ~~~ aziridine ~ - CH2 OII trifluorosulfonate -CH20-S-CF3 IIO For the reactive alkyl halide function, X signifies Br, Cl or I. Advantageously, the labeling reagent is 5- (bromomethyl) fluorescein.

For the reactive nitrosourea function, RS is alkyl or H. The diazomethyl function has already been used for the alkylation of phosphate groups, but a certain number of problems arise. On the one hand, the diazo derivatives in general are unstable themselves, which generates problems for the use of these labeling reagents in a labeling kit, and on the other hand, the coupling product is unstable this. which is prohibitive if the labeled product has the function of demonstrating the presence of a biological target molecule in any sample, or if it is the labeled target that it is desired to detect. Finally, the derivatives carrying the diazomethyl function are insoluble in water which leads to the use of two-phase conditions for coupling with nucleic acids, which are soluble and stable only in water or in aqueous buffers, but these conditions slow down the process. reaction rate and therefore adversely affect the efficiency of the coupling.

In a preferred embodiment of the process according to the invention, the labeling reagent is chosen from 2o among the compounds of formula (1)

[0011]

IDT Exhibit 1020.10028

CA 02444926 2003-10-17 WO 02/090584 PCT / FR02 / 01542 11 Ri z C-Na in which ~ R1 represents H or an alkyl, substituted alkyl, aryl or substituted aryl group, ~ Z comprises a detectable label.

Z and / or RI are chosen to stabilize the diazomethyl function, that is to say at least one of the two groups Z or RI has a phenyl ring. Preferably, the labeling reagent is chosen from the compounds of formula (2) RI R2, C = N2 - ~) n- Z 1o in which ~ RI represents H or an alkyl, aryl or substituted aryl group, ~ Ra is a detectable marker, ~ L is a linking arm comprising a linear sequence of at least two covalent bonds and n is equal to 0 or 1, and i5 ~ Z is chosen from YI R4 R4 XOO R3 or R3 or - C- NH- or - C- in which ~ R3 and R4 represent independently of one another: H, NO2, Cl, Br, F, I, OR, 2o SR, NR2, R, NHCOR, CONHR, COOR with R ~ alkyl or aryl, ~ -YX- represents -CONH-, -NHCO-, -CH2O-, -CH2S-. In a particular mode according to formula (1), Z has the following structure

[0012]

CA 02444926 2003-10-17 WO 02/090584 PCT / FR02 / 01542 12 In this case and if R1 is equal to H, the labeling reagent is 1-pyrénykliazomethane (PDAM). .

Although this marker is fluorescent, the excitation wavelength is too close to that of nucleic acids. Indirect detection by the use of a monoclonal antibody directed against the pyrene unit is preferred. The method of obtaining this antibody is well known to those skilled in the art (see for example patent application WO-A-00/07962). 1o In a preferred embodiment of the method, the labeling reagent is of formula (3) R3 R2 ~) ri ~ .'- X 4 ~ R1 R u N2 in which ~ RI represents H or an alkyl, aryl or aryl group substituted, ~ Ra represents a detectable label, ~ L is a linking arm comprising a linear sequence of at least two covalent bonds and n an integer equal to 0 or 1, ~ R3 and R4 independently represent one of the other: H, NOZ, Cl, Br, F, I, OR, SR, NR2, 2o R, NHCOR, CONHR, COOR with R = alkyl or aryl, ~ -YX- represents -CONH-, -NHCO-, -CHZO -, -CH2S-. Advantageously, the reagent is of formula (4)

[0013]

CA 02444926 2003-10-17 WO 02/090584 PCT / FR02 / 01542 13 R3 O Ra (L) ri C- NH R1 R4 N2 in which ~ RI represents H or an alkyl, aryl or substituted aryl group, ~ RZ represents a detectable marker, ~ L is a linking arm comprising a linear sequence of at least two covalent bonds and n an integer equal to 0 or 1, ~ R3 and R4 represent independently of one another: H, NO2, Cl, Br, F, I, OR, SR, NR2, R, NHCOR, CONHR, COOR with R = alkyl or aryl.

Advantageously the reagent is of formula (5) R3 O. R2 (L) ri NH - C R1 Rø N ~ in which ~ RI represents H or an alkyl, aryl or substituted aryl group, ~ R2 represents a detectable marker, ~ L is a linking arm comprising a linear sequence of at least two covalent bonds and n an integer equal to 0 or 1, ~ R3 and R4 represent independently of one another: H, NO2, Cl, Br, F, I, OR, SR, NRa, R, NHCOR, CONHR, COOR with R = alkyl or aryl. Advantageously, the reagent is of formula (6) R3 R2 ~) ri CH2-. O 'Rt R4 N2

[0014]

CA 02444926 2003-10-17 WO 02/090584 PCT / FR02 / 01542 14 in which ~ R1 represents -H or an alkyl, aryl or substituted aryl group, R2 represents a detectable marker, ~ L is a linker arm comprising a linkage linear of at least two covalent bonds and n an integer equal to 0 or 1, ~ R3 and R4 represent independently of each other: H, NO2, Cl, Br, F, I, OR, SR, NR2 , R, NHCOR, CONHR, COOR with R = alkyl or aryl.

In the above formulas (3) to (6), advantageously R3 and R4 represent 1o independently of one another: H, NO2, ÖCH3. Thus, a preferred compound according to formula (6) is of formula (6 ') N02 R2 (L) ri CH2- O ~ RI CH3O ~ ~ ~ N2 in which ~ RI represents H or an alkyl, aryl or substituted aryl group, ~ R2 represents a detectable marker, ~ L is a linking arm comprising a linear sequence of at least two covalent bonds and n an integer equal to 0 or 1. A preferred compound according to formula (4) is of formula (4 ') O R2 (L) ri C-NH N2 R1 in which ~ R1 represents H or an alkyl, aryl or substituted aryl group, ~ R2 represents a detectable label,

[0015]

CA 02444926 2003-10-17 WO 02/090584 PCT / FR02 / 01542 1S ~ L is a linking arm comprising a linear sequence of at least two covalent bonds and n an integer equal to 0 or 1.

IDT Exhibit 1020.10029

In a particular embodiment of the process where it is desired to amplify the signal, at least two markers are present on the labeling reagent. In particular a reagent which makes it possible to carry out the signal amplification according to the present invention has an Ra- (L) n structure of formula (7) below 2 O R2

[0016]

GOLD

[0017]

NH

[0018]

NH

[0019]

(CH2) P

[0020]

(CH ~) p

[0021]

R2- ~ NH

[0022]

f1 O m-1

[0023]

in which 1o ~ R2 represents a detectable marker, ~ m is an integer between 1 and 100, preferably between 1 and 20, ~ p is an integer between 1 and 10, advantageously 2 to 6, preferably 4 .

This structure of R2- (L) "applies equally to the preceding formulas (2) to (6).

Another preferred labeling reagent for signal amplification is the reagent of formula (8) N2 C iv '2- (i-) n 1' - X ~ - Y- (L) n- Ra R3 Rs, 2o in which ~ R2 represents a detectable marker,

[0024]

CA 02444926 2003-10-17 WO 02/090584 PCT / FR02 / 01542 16 ~ R3 represents H, NO2, Cl, Br, F, I, OR, SR, NR2, R, NHCOR, CONHR, COOR with R = alkyl or aryl, ~ L is a linking arm comprising a linear sequence of at least two covalent bonds and n an integer equal to 0 or 1, ~ -YX- represents -CONH-, -NHCO-, -CH20-, -CH2S -.

Advantageously, the reagent for signal amplification has the formula (9) N2 OCO RZ- (L) ri C- NH NH, C- (L) ri R2, R3 R3 in which 1o ~ RZ represents a detectable marker, ~ R3 represents H, NO2, Cl, Br, F, I, OR, SR, NR2, R, NHCOR, CONHR, COOR with R = alkyl or aryl, preferably R3 represents H, NOZ or OCH3., ~ L is an arm of bond comprising a linear sequence of at least two covalent bonds and n an integer equal to 0 or 1.

Certain advantageous reagents of the invention are a) of formula (10) O R2- (L) ri C-NH N2 R1 in which ~ R1 represents H or an alkyl or aryl or substituted aryl group, ~ R2 represents a detectable label, ~ L is a connecting arm comprising a linear sequence of at least two covalent bonds and n an integer equal to 0 or I, b) of formula (11)

[0025]

IDT Exhibit 1020.10030

CA 02444926 2003-10-17 WO 02/090584 PCT / FR02 / 01542 17 O R2- ~) n'-C- NH R1 N2 in which ~ R1 represents H or an alkyl or aryl or substituted aryl group, ~ R2 represents a detectable marker, ~ L is a linking arm comprising a linear sequence of at least two covalent bonds and n an integer equal to 0 or 1, c) of formula (12) Ri Il R2 (L) ri C-NH N2 in which 1o ~ Ri represents H or an alkyl or aryl or substituted aryl group, ~ R ~ represents a detectable marker, ~ L is a linking arm comprising a linear sequence of at least two covalent bonds and n an integer equal to 0 or 1.

15 Preferably, the labeling reagent a a) the structure O HN HN O H H C-NH S Ri N2 in which R1 represents a methyl group or a phenyl, or

[0026]

CA 02444926 2003-10-17 WO 02/090584 PCT / FR02 / 01542 18 b) the structure O HN HN OHH C-NH S Ri N2 in which R ~ represents a methyl or phenyl group, or c) the structure O HN HN O N2 HH C-NH .. ", ~ 1 S ~ R in which Ri represents a methyl or phenyl group.

Other preferred reagents according to the invention have the formula (13) H ~ C% Nz O h'2W) n'-C-NH N02 1o in which ~ R2 represents a detectable label, ~ L is a linker comprising a linear sequence of at least two covalent bonds and n an integer equal to 0 or 1.

15 In particular, the labeling reagents according to the method of the invention are soluble in polar solvents miscible with water such as DMF, DMSO, CH3CN, THF, DMA (dimethylacetamide), NMP (N-methylpyrrolidone), DME (dimethoxyethane)).

[0027]

CA 02444926 2003-10-17 WO 02/090584 PCT / FR02 / 01542 19 Preferably, the labeling reagents are soluble in DMSO or in water.

The term "water-miscible solvent" means a solvent which is miscible in a proportion of at least 5% by volume with water or an aqueous buffer containing salts.

Advantageously in the preceding formulas, the arm L comprises an ethylene glycol or polyethylene glycol unit to increase the solubility of the reagent in water.

These reagents can thus bind to the nucleic acids in a homogeneous phase, the homogeneous phase consisting of a substantially aqueous solution, that is to say containing at least 50% water. 1o An object of the present invention is to describe a method for detecting a target deoxyribonucleic acid (DNA) single or double stranded comprising the following steps a) fragmenting and labeling said DNA according to any one of the methods described above, b) hybridizing the labeled fragments on at least one sufficiently specific nucleic acid probe for the target nucleic acid, c) detecting the hybrid formed by means of the marker.

As will be demonstrated in the examples, the denaturation of the DNA before hybridization makes it possible to increase the sensitivity.

This denaturation step takes place after the 2o fragmentation and labeling. In a particular embodiment of the process, an enzymatic amplification step takes place before the fragmentation and the labeling. The enzymatic amplification step generates DNA amplicons from a DNA target nucleic acid, but also from an RNA target nucleic acid such as messenger RNA, transfer RNA or ribosomal RNA. The RT-PCR technique, for example, is a known means of amplifying RNA. Preferably, the stages of fragmentation, labeling and denaturation take place at the same time, both in the case of a natural target nucleic acid and in the case of a target nucleic acid obtained by an enzymatic amplification technique. Fragmentation by creating an abasic site involves the loss of a base. In the methods of detecting target nucleic acid and in particular in the genotyping methods that is to say in the methods where the target nucleic acid exhibits a

[0028]

CA 02444926 2003-10-17 WO 02/090584 PCT / FR02 / 01542 20 'polymorphism distributed over its sequence, it is necessary to identify a plurality of sequence modifications, some of which represent the modification of a single base between the target and nucleic acid probes whose function is to identify this modification (capture probes).

Specificity is therefore essential in this context. Surprisingly, ~ the present invention demonstrates that this loss of a base does not affect the specificity of the hybridization for a labeled and fragmented nucleic acid according to the invention. The present

IDT Exhibit 1020.10031

invention also relates to a method of demonstrating a polymorphism, distributed in predetermined positions or not of a target nucleic acid by the presence of a plurality of deletions and / or insertions and / or mutations in the sequence of said target nucleic acid relative to a so-called reference sequence comprising the following steps a) have a target DNA comprising all of the polymorphism to be studied, said DNA being optionally generated by an enzymatic amplification technique, b) fragmenting and labeling said DNA by one of the methods described above, c) hybridizing said fragments on a plurality of nucleic acid probes called capture probes, the plurality of capture probes being fixed on a solid support and the plurality of capture probes covering as a whole at least the polymorphism to be studied, d) detecting the hybrids formed between the labeled fragments and at least part of the nucleic acid probes by means of the marker and deduce therefrom the polymorphism of the target DNA.

As indicated above, a denaturation step after the fragmentation and labeling step makes it possible to improve the sensitivity of the process, and this step is preferably carried out simultaneously with the labeling and fragmentation step. The fragmentation and labeling process according to the invention is particularly useful in the case where the labeled and fragmented DNA must hybridize with a multitude of nucleic acids, in particular oligonucleotides, fixed on the solid support at a predetermined 3o position for form a DNA chip. By "DNA chip" is meant a solid support of reduced size where a multitude of capture probes are attached at positions

[0029]

CA 02444926 2003-10-17 WO 02/090584 PCT / FR02 / 01542 21 predetermined.

By multitude is meant a solid support comprising at least ten (10) nucleic acid probes of different sequences, advantageously at least four hundred (400), preferably at least one thousand (1000). In fact, the density of the nucleic acids fixed on the solid support imposes significant steric constraints during hybridization and the fragmentation makes it possible to improve this hybridization step. Examples of these DNA chips are given for example in the publications by G. Ramsay, Nature Biotechnology, 16, p40-44, 1998; F. Ginot, Human Mutation, 10, p1-10, 1997; J. Cheng et al, Molecular diagnostics, 1 (3), p183-200, 1996; T. Livache et al, Nucleic Acids Research, 22 (15), p2915-2921, 1994; J. Cheng et al, Nature Biotechnology, 16, p541-546, 1998. The term "solid support" as used herein includes all materials to which a nucleic acid can be attached. The solid support can be in the form of a microtiter plate, a membrane, a particle or a substantially planar plate.

The invention relates to the use of a labeled DNA, as defined above, as a probe for detecting a target nucleic acid. Lice allow the detection and / or quantification and / or purification of the target nucleic acid 2o, the labeled DNA is capable of forming a hybridization complex with a nucleic probe. By way of example, the labeled nucleic acid is sufficiently complementary to the target to hybridize specifically as a function of the reaction conditions and in particular of the temperature or the salinity of the reaction medium. The detection method is applicable for sequencing, expression profiling of messenger RNAs or screening of mutations for research purposes as well as drug screening in the pharmaceutical industry, diagnosis of infectious diseases (both in bacteriology , virology, parasitology) or genetics, food or industrial control. Numerous publications describe this type of application: Antoine de Saizieu et al, Nature Biotechnology, 16, p44-48, 1998; Thomas R. Gingeras et al Genome Research, 8, 3o p435-448, 1998; David J. Lockhart et al, Nature Biotechnology, 14, p1675-1680, 1996; Daniel D. Shoemaker et al, Nature Genetics, vol 14, p450-456, 1996; R.J. Lipshutz et al;

[0030]

CA 02444926 2003-10-17 WO 02/090584 PCT / FR02 / 01542 22 BioTechniques, 19 (3), p442-447,1995; David G. Wang et al, Science, 280, p1077-1082, 1998; Kevin L. Gunderson et al; Genome Research, 8, p112-1152, 1998; Joseph G. Hacia et al; Nature Genetics, vol 14, p441-447, 1996.

The tendency in terms of diagnosis and in particular for infectious diseases (AIDS or Tuberculosis for example) is to lower the level of sensitivity until the detection of a single molecule in a sample, which can represent several milliliters in the case of a liquid sample such as blood or urine or cerebrospinal fluid. This level of sensitivity can only be obtained if all the steps from taking the sample 1o to rendering the result are optimized. In particular in the case where an enzymatic amplification step is necessary to obtain the necessary sensitivity (viral or bacterial infection such as HIV, HCV or Tuberculosis), a labeling and / or fragmentation process as described in the present invention makes it possible not to affect the sensitivity of the amplification technique, either because it is not necessary to replace the deoxyribonucleotides used in the enzymatic amplification technique, or because the incorporated deoxyribonucleotides do not alter the sensitivity.

Additional information can be found in another patent application by the Applicant, filed on May 4, 2001 under the registration number FR01 / 06040, as well as in its international extension filed on the same day as the present invention. The accompanying figures and examples represent particular embodiments and cannot be considered as limiting the scope of the

IDT Exhibit 1020.10032

present invention. FIG. 1 represents the structural formulas of various reagents used in the present invention as well as the abbreviation designating them (o- means ortho, m-meta and p-paya). Figures 2A to 2I represent the profiles, as a function of time, analyzed by capillary electrophoresis 3o of the covalent coupling of different reagents carrying a function

[0031]

CA 02444926 2003-10-17 WO 02/090584 PCT / FR02 / 01542 23 diazomethyl on 3 'uridine monophosphate (3'-UMP) according to Example 6.1.

The molecules are as follows ~ PDAM in figure 2A, ~ DPDAM at 2 mM (millimole (s) per liter) in figure 2B, ~ DPDAM at 20 mM in figure 2C, PMDAM in figure 2D, ~ NPDAM in figure 2E, ~ BioDPDAM in Figure 2F, ~ meta-BioPMDAM in Figure 2G, 1o ~ paya-BioPMDAM in Figure 2H, and ~ o ~ tho-BioPMDAM in Figure 2I. FIGS. 3A to 3D represent the profiles as a function of time analyzed by capillary electrophoresis of the reaction of meta-BioPMDAM on four (4) 3 ' monophosphate nucleotides according to Example 6.2. The molecules are as follows ~ 3'-CMP in ribonucleotide series according to Figure 3A, ~ 3'-AMP in ribonucleotide series according to Figure 3B, ~ 3'-GMP in ribonucleotide series according to Figure 3C, and ~ 3'-TMP deoxyribonucleotide series according to Figure 3D. FIG. 4 represents the profiles as a function of time analyzed by capillary electrophoresis of the reaction of meta-BioPMDAM on a 5'-ApUp dinucleotide according to Example 6.3. Figures 5A to 5D represent the proton NMR spectrum in D20 of the various conjugates between the meta-BioPMDAM reagent and four (4) 3' -monophosphate ribonucleotides according to Example 6.4.

The molecules are as follows ~ 3'-GMP in figure 5A, ~ 3'-AMP in figure 5B, ~ 3 'CMP in figure 5C, and

[0032]

CA 02444926 2003-10-17 WO 02/090584 PCT / FR02 / 01542 24 ~ 3 'UMP in figure 5D. FIG. 6 represents a diagram of the synthesis of a labeling reagent carrying two biotins for the chemical amplification of the signal.

FIG. 7 represents the mechanism of fragmentation in an acidic medium by the formation of an abasic site. Figure 8 shows, according to Example 8.1, the degradation kinetics at acid pH for 1o different modified nucleosides (8-bromo-2'-deoxyadenosine (8-BrdA) and 5-bromo-2'-deoxycytidine (5- BrdC)) as well as the four natural nucleosides (dA, dC, dG and dT). The results are represented in the form of the percentage of hydrolysis of the starting nucleoside (on the ordinate) relative to the reaction time in minutes (on the abscissa). FIG. 9 represents, according to Example 11.2, the labeling kinetics as a function of time at a temperature of 60 ° C. with the PDAM reagent on a synthetic 5'-phosphate ODN. The results are represented in the form of percentage of labeling relative to the reaction time expressed in minutes (on the abscissa). FIG. 10 represents, according to Example 11.3, the percentage of labeling as a function of the reaction temperature. The results are presented in FIG. 10 with the percentage labeling on the ordinate and the reaction temperature in ° C on the abscissa. Figure 11 shows a synthetic route for a reagent according to formula 4 ', using the commercial reagent 5-Nitrovanillin.

The aldehyde function is the precursor of the diazomethyl function. Example 1: Synthesis of reagents with biotin 3o General synthesis diagram

[0033]

CA 02444926 2003-10-17 WO 02/090584 PCT / FR02 / 01542 R R N-methylmorpholine ~ p HN NH + i ~ 0 iBuOCOCl HN NH ~ H ~ H H OH NH DMF ~.

NH z OO D-Biotin 1a: R = CH3, meta substituent 1 b: R = CH3, para substituent 1 c: R = CH3, ortho substituent 1 d: R = Ph, meta substituent N2Ha_H20 R MnO2 ~ i ~ N ~ NHz HN NH H ~ H NH O 3a: R = CH3, mefa substituent 2a: R = CH3, meta substifuent 3b: R = CH3, para substituent 2b: R = CH3, para substituent 3c: R = CH3, ortho substituent 2c: R = CH3, ortho substituent 3d: R = Ph, meta substituent 2d: R = Ph, meta substituent. Example 1.1: Synthesis of naeta-SioPMDAM 5 ~ Com ose biotin meta-acetophenone la D-biotin (1.0 gram (g), 4.1 millimoles (mmol)) is solubilized in 45 milliliters (mL) of anhydrous DMF at hot. Cooled to 0 ° C under argon, then added successively N methylinorpholine (590 microliters (~ L), 5.33 mmol) and isobutyl chloroformate (840 ~ L, l0 6.60 mmol). The mixture is stirred for 30 minutes (min), then 3-aminoacetophenone (824 mg, 6.10 mmol) and N-methylmorpholine (480 ~ L, 4.35 mmol) in 10 mL of DMF are added. The solution is kept under stirring at 0 ° C. for 2 hours (h), then evaporated to dryness.

The residue is taken up in 3 mL of MeOH, then 50 mL of water is added. The precipitate obtained is filtered off, washed with water, CH2Cl2 and ether to give 1.2 g (80%) of crude product 1a. Recrystallization from MeOH-H2O gives 1a (1.01 g, 70%) in the form of a white powder.

IDT Exhibit 1020.10033

[0034]

CA 02444926 2003-10-17 WO 02/090584 PCT / FR02 / 01542 26 -1 F 145 ° C. - IR (KBr): 3280, 2931, 2857, 1691, 1590,1540, 1487, 1434, 1298, 1266 cm.

1H NMR (300 MHz, DMSO-d6) 8 = 1.3-1.7 (m, 6H); 2.33 (t, J = 8Hz, 2H); 2.55 (s, 3H); 2.58; (d, J = 12Hz, 1H); 2.83 (dd, J = 12 & 5Hz, 1H); 3.13 (m, 1H); 4.15 (m, 1H); 4.31 (m, 1H); 6.34 (s, 1H); 6.41 (s, 1H); 7.44 (t, J = 8Hz, 1H); 7.64 (d, J = 8Hz, 1H); 7.85 (d, J = 8Hz, 1H); 8.17 (s, 1H); 10.05 (s, 1H). - MS (FAB / glycerol), nz / z: 362 [M + H]. ~ Compound mâta-hydrazone 2a 1o A solution of Ia (500 mg, 1.38 mmol) and hydrazine monohydrate (200 ~, L, 4.15 mmol) in absolute ethanol (8 mL) is heated to reflux for 2 h. After cooling to room temperature, the white precipitate is filtered off, washed with water, then with ether and dried. 385 mg (74%) of product 2a are thus obtained in the form of a white powder. Mp 185 ° C. - IR (KBr): 3298, 2931, 2857, 1698, 1665, 1626, 1541, 1494, 1470, 1446, 1330, 1265 cm 1. - 1H NMR (300 MHz, DMSO-d6) 8 = 1.3-1 , 7 (m, 6H); 1.98 (s, 3H); 2.26 (t, J = 8Hz, 2H); 2.56; (d, J = 12Hz, 1H); 2.81 (dd, J = 12 & 5Hz, 1H): 3.11 (m, 1H): 4.13 (m, 1H); 4.29 (m, 1H); 6.39 (s, 3H); 6.42 (s, 1H); 7.22 (m, 2H); 7.50 (d, J = 8Hz, 1H); 7.84 (s, 1H); 9.82 (s, 1H). - MS (FAB / glycerol), m / z: 376 [M + H].

~ Mata-diazomethane compound 3a 2a (180 mg, 0.48 mmol) is dissolved in 2 mL of DMF. MnO 2 (340 mg, 3.9 mmol) is then added. After 30 minutes of stirring at room temperature, the mixture is filtered through a sintered funnel containing celite (thickness: 0.5 cm) and molecular sieves in powder 3 ~ (O, Som). The reaction mixture is concentrated to a volume of about 0.5 mL, then 5 mL of ether is added. The resulting precipitate is filtered off, washed with ether and then dried. Compound 3a (170 mg, 95%) is obtained in the form of a pink powder.

[0035]

CA 02444926 2003-10-17 WO 02/090584 PCT / FR02 / 01542 27 F 160 ° C. - IR (KBr): 3278, 2935, 2859, 2038, 1704, 1666, 1605, 1577, 1536, 1458, 1430, 1263 cm 1. - 1H NMR (300 MHz) 8 = 1.3-1.7 (m , 6H); 2.11 (s, 3H); 2.28 (t, J = 8Hz, 2H); 2.57; (d, J = 12Hz, 1H); 2.81 (dd, J = 12 & 5Hz, 1H); 3.11 (m, 1H); 4.13 (m, 1H); 4.29 (m, 1H); 6.33 (s, 1H); 6.41 (s, 1H); 6.60 (m, 1H); 7.25 (m, 3H); 9.84 (s, 1H).

). Example 1.2 Synthesis of dupara-BioPMDAM ~ Biotinpara-acetophenone Compound 1b D-biotin (1 g, 4.1 mmol) is solubilized in 45 mL of hot anhydrous DMF. Cooled to 0 ° C under argon and then successively added N methylmorpholine (590 ~. L, 5.33 mmol) and isobutyl chlorofonate (840 ~ L, 6.60 mmol). The mixture is left under stirring for 30 min, then 4-aminoacetophenone (824 mg, 6.10 mmol) is added. The solution is kept under stirring at 0 ° C. for 2 h then evaporated to dryness. The residue is taken up in 50 ml of water. The precipitate obtained is filtered, washed with water and then with SOmL of MeOH while hot. The white precipitate is dissolved in DMF with heating and then the solution obtained is filtered and washed with MeOH. The filtrate is collected and evaporated to dryness to give 868 mg of Ib (2.46 mmol, 60%). Mp 260 ° C. - IR (KBr): 3260, 2930, 2358, 1706, 1673, 1610, 1526, 1401, 1380, 1322, 1257, 1150 cm 1. - 1H NMR (200 MHz, DMSO-d6) 8 = 8.82 (s , 1H, NH-CO); 7.57 (d, 2H, J = 9Hz, Ar-H); 6.83 (d, 2H, J = 9Hz, Ar-H); 6.40 (br s, 1H, NH-CO-NH); 6.32 (br s, 1H, NH-CO-NH); 4.28 (m, 1H, CH2-CH-NH); 4.12 (m, 1H, CH-CH-NH); 3.11 (m, 1H, CH-S); 2.80 and 2.55 (ABX system, 2H, ZJ, ~ = S Hz, 3J, ex = 12 Hz, 3JBx = 0 Hz, CHZ-S); 2.35 (t, 2H, J = 8Hz, CH2-CO); 2.10 (s, 3H, CH3); 1.60-1.34 (m, 6H, (CH2) 3).

~ Composed ° a-h ~ drazone 2b Compound 1b X870 mg, 2.4 mmol) is dissolved hot in ethanol (99%, 8 mL) then hydrazine monohydrate (995 ~. L, 19.5 mmol) is added. The solution is heated at 3o reflux for 3 h. The white precipitate obtained is filtered off, washed with ice-cold water. In this way 820 mg (90%) of product 2b are obtained in the form of a white powder.

[0036]

CA 02444926 2003-10-17 WO 02/090584 PCT / FR02 / 01542 as F 305 ° C. - IR (KBr): 3281, 3183, 2930, 2857, 1698, 1658, 1593, 1521, 1459, 1401, 1325, 1263, 1187 cm 1. - 1H NMR (200 MHz, DMSO-d6) 8 = 9.68 (s, 1H, NH-CO); 7.52 (s, 4H, J = 9Hz, Ar-H); 6.43 (br s, 1H, NH-CO-NH); 6.35 (br s, 1H, NH-CO-NH); 6.21 (s, 2H, NHZ) 4.29 (m, 1H, CHZ-CH-NH); 4.12 (m, 1H, CH-CH-NH); 3.12 (m, 1H, CH-S); 2, 81 and 2.56 (ABX system, 2H, 2J ~ = 5Hz, 3J ~ = 12Hz, 3JBx = 0Hz, CHZ-S); 2.32 (t, 2H, J = 8Hz, CHZ-CO); 1.97 (s, 3H, CH3); 1.63-1.36 (m, 6H, (CHZ) 3).

~ Compepa ~ a-diazomethane 3b 2b (200 mg, 0.53 mmol) are dissolved in 10 mL of DMF. 800 mg of MnO2 are then added. After 10 minutes of stirring, the mixture is filtered through a mixed Celite (0.5 cm) - molecular sieve (0.5 cm powder) layer. The reaction mixture is evaporated to dryness then washed with ether and dried. Compound 3b (190 mg, 96%) is obtained in the form of a pink powder. F 180 ° C (dec). - IR (I ~ Br): 3257, 2930, 2857, 2032, 1698, 1597, 1524, 1510, 1455, 1404, 1307, 1259, 1180 cm 1. - 1H NMR (200 MHz, DMSO-d6) 8 = 10 , 18 (s, 1H, NH-CO); 7.88 (d, 2H, J = 6Hz, Ar-H); 7.7 (d, 2H, J = 6Hz, Ar-H); 6.41 (br s, 1H, NH-CO-NH); 6.34 (br s, 1H, NH-CO-NH); 4.28 (m, 1H, CH2-CH-NH); 4.12 (m, 1H, CH-CH-NH); 3.11 (m, 1H,

IDT Exhibit 1020.10034

CH-S); 2.80 and 2.55 (ABX system, 2H, aJ ~ = 5Hz, 3J ~ = 12Hz, 3JBx = 0Hz, CH2-S); 2.35 (t, 2H, J = 8Hz, CH2-CO); 2.10 (s, 3H, CH3); 1.60-1.34 (m, 6H, (CH2) 3). Example 1.3: Synthesis of ortho-BioPMDAM a5 ~ Biotin compound o ~ tho-acetophenone 1c D-biotin (1 g, 4.1 mmol) is solubilized in 45 mL of hot anhydrous DMF.

Cooled to 0 ° C under argon and then successively added N methylmorpholine (590 ~. L, 5.33 mmol) 3o and isobutyl chlorofonate (840 p, L, 6.60 mmol). The mixture is left under stirring for 30 min, then 2-aminoacetophenone (824 mg, 6.10 mmol) is added. The solution is kept under

[0037]

CA 02444926 2003-10-17 WO 02/090584 PCT / FR02 / 01542 29 stirring at room temperature for 3 h 30 then evaporated to dryness.

The residue is taken up in 50 ml of water. The precipitate obtained is filtered, washed with water and then with SOmL of MeOH while hot. The precipitate obtained is filtered off, washed with water. Recrystallization is carried out by dissolving the product in hot MeOH and reprecipitating by adding water. The precipitate is filtered off, washed with water and then with ether to give 1.1 g (2.95 mmol, 72%) of crude product 1c. F 150 ° C. - IR (KBr): 3248, 2930, 2857, 2359, 1691, 1669, 1651, 1582, 1528, 1448, 1354, 1310, 1245, 1161 cm 1. - 1H NMR (200 MHz, DMSO-d6) 8 = 11 , 24 (s, 1H, NH-CO); 8.33 (d, 1H, J = 8.5Hz, Ar-H); 7.97 (d, 2H, J = 8Hz, Ar-H); 7.57 (t, 1H, J = 7Hz, Ar- H); 7.18 (t, 1H, J = 7Hz, Ar-H); 6.44 (br s, 1H, NH-CO-NH); 6.35 (br s, 1H, NH-CO-NH); 4.30 (m, 1H, CH2-CH-NH); 4.14 (m, 1H, CH-CH-NH); 3.12 (m, 1H, CH-S); 2.80 and 2.55 (ABX system, 2H, aJ ~ = 5Hz, 3J, ~ = 12Hz, 3JBx = 0Hz, CH2-S); 2.61 (s, 3H, CH3); 2.37 (t, 2H, J = 8Hz, CH2-CO); 1.62-1.38 (m, 6H, (CH2) 3).

~ Compound o ~ tho-hydrazone 2c Compound 1c (500 mg, 1.38 mmol) is dissolved hot in ethanol (99%, 8 mL) then hydrazine monohydrate (572 ~ CL, 11.1 mmol) is added. The solution is heated at reflux for 50 minutes. The solution is evaporated to dryness. The white precipitate obtained is filtered, washed with water and then dried with ether. In this way 416 mg (11.1 mmol, 80%) of product 2c are obtained in the form of a white powder. Mp 161 ° C. - IR (KBr): 3412, 3240, 2930, 2857, 2351, 1706, 1677, 1604, 1590, 1531, 1463, 1444, 1372, 1303, 1270, 1169 cm i - 1H NMR (200 MHz, DMSO-d6) 8 = 11.97 (s, 1H, NH-CO); 8.35 (d, 1H, J = 8Hz, Ar-H); 7.45 (d; 1H, J = 7Hz, Ar-H); 7.19 (t, 1H, J = 7.5Hz, Ar- H); 7.04 (t, 1H, J = 7Hz, Ar-H); 6.61 (s, 2H, NH2); 6.42 (s lax, 1H, NH-CO-NH); 6.35 (br s, 1H, NH-CO-NH); 4.32 (m, 1H, CH2-CH-NH); 4.14 (m, 1H, CH-CH-NH); 3.12 (m, 1H, CH-S); 2.81 and 2.56 (ABX system, 2H, 2J ~ = 5Hz, 3J ~ = 12Hz, 3JBX = 0Hz, CHZ-S); 2.31 (t, 2H, J = 8Hz, CH2-CO); 2.09 (s, 3H, CH3); 1.63-1.36 (m, 6H, (CHa) 3).

[0038]

CA 02444926 2003-10-17 WO 02/090584 PCT / FR02 / 01542 ~ Ortho-diazomethane compound 3c 2c (200 mg, 0.53 mmol) are dissolved in 10 mL of DMF. 800 mg of MnO2 are then added.

After 15 minutes of stirring, the mixture is filtered through a mixed layer of Celite (O, Scm)-molecular sieve (O, Scm powder). The reaction mixture is evaporated to dryness then washed with ether and dried. Compound 3c (130 mg, 65%) is obtained in the form of a pink powder. F 110 ° C. - IR (KBr): 3248, 2930, 2857, 2367, 2342, 2038, 1699, 1521.1456 cm 1. - 1H 1o NMR (200 MHz, DMSO-d6) 8 = 9.37 (s, 1H, NH- CO); 7.26-7.00 (m, 4H, Ar-H); 6.43 (br s, 1H, NH-CO-NH); 6.35 (br s, 1H, NH-CO-NH); 4.30 (m, 1H, CH2-CH-NH); 4.15 (m, 1H, CH-CH-NH); 3.12 (m, 1H, CH-S); 2.82 and 2.54 (ABX system, 2H, 2J ~ = 5Hz, 3J ~ = 12Hz, 3JBx = 0Hz, CHZ-S); 2.24 (t, 2H, J = 8Hz, CHZ-CO); 2.12 (s, 3H, CH3); 1.63-1.37 (m, 6H, (CH2) 3). Example 1.4 Synthesis of Meta-BioDPDAM ~ Compound f ~ ieta-benzophenone 1d 2o D-biotin (500 mg, 2.05 mmol) is solubilized in 23 mL of hot anhydrous DMF. Cooled to 0 ° C under argon, then added successively 1a N methylmorpholine (295 ~ L, 2.67 mmol) and isobutyl chloroformate (420 ~, L, 3.28 mmol).

The mixture is left stirring for 30 min, then 3-aminobenzophenone (605 mg, 3.07 mmol) and N methylmorpholine (240 ~ L, 2.17 mmol) in 7 mL of DMF are added. The solution is kept under stirring at 0 ° C. for 2 h, then evaporated to dryness. The residue is taken up in 1 mL of MeOH, then 25 mL of water is added. The precipitate obtained is filtered, washed with water, then with ether to give 810 mg (93%) of crude product 1d. Recrystallization in the MeOH-HZO couple gives 1d (630 mg, 72%) in the form of a white powder.

[0039]

CA 02444926 2003-10-17 WO 02/090584 PCT / FR02 / 01542 31 1H NMR (200MHz, DMSO-d6) 8 = 10.10 (s, 1H, NH-CO); 8-7.39 (m, 9H, Ar-H); 6.43 (br s, 1H, NH-CO-NH); 6.35 (br s, 1H, NH-CO-NH); 4.27 (m, 1H, CHZ-CH-NH); 4.13 (m, 1H, CH-CH-

IDT Exhibit 1020.10035

NH); 3.12 (m, 1H, CH-S); 2.84 and 2.55 (ABX system, 2H, aJ ~ = 5Hz, 3J ~ = 12Hz, 3JBx = 0Hz, CH2-S); 2.31 (t, 2H, J = 8Hz, CHZ-CO); 1.59-1.36 (m, 6H, (CHZ) 3).

~ 2d meta-hydrazone compound 1d (350 mg, 0.83 mmol) is dissolved in 5.5 mL of absolute ethanol and then hydrazine monohydrate (140 ~ L, 2.48 mmol) is added. The solution is heated at reflux overnight. After evaporation, the product is taken up in 1 ml of ethanol and water. The white precipitate is recrystallized: it is dissolved in a minimum of hot ethanol and water is added until a slight cloudiness appears. After cooling, the precipitate obtained is washed with water and then dried with ether. In this way 264 mg (73%) of product 2d is obtained in the form of a white powder. 1H NMR (200 MHz, DMSO-d6) b = 9.99 (s, 1H, NH-CO); 9.80 (s, 2H, NH2); 7.54-6.88 (m, 9H, -Ar-H); 6.26 (br s, 1H, NH-CO-NH); 6.21 (br s, 1H, NH-CO-NH); 4.28 (m, 1H, CHZ-CH-NH); 4.13 (m, 1H, CH-CH-NH); 3.12 (m, 1H, CH-S); 2.78 and 2.59 (ABX system, 2H, ZJ, ~ = 5Hz, 3J, ex = 12Hz, 3JBx = 0Hz, CH2-S); 2.27 (t, 2H, J = 8Hz, CH2-CO); 1.57-1.36 (m, 6H, (CHZ) 3).

~ 3d meta-diazodiphenyl compound 3d (500 mg, 0.53 mmol) is dissolved in 1 mL of THF. 80 mg of activated MnO ~ are then added. After 5 minutes of stirring at room temperature, the mixture is filtered through a mixed Celite (0.5 cm) -molecular sieve (0.5 cm powder) layer. The reaction mixture is evaporated to dryness. Compound 3d (47 mg, 100%) is obtained in the form of a violet oil. 1H NMR (200 MHz, DMSO-d6) 8 = 9.95 (s, 1H, NH-CO); 7.60-6.9 (m, 9H, Ar-H); 6.42 (br s, 1H, -NH-CO-NH); 6.35 (br s, 1H, NH-CO-NH); 4.28 (m, 1H, CH2-CH-NH); 4, 14

[0040]

CA 02444926 2003-10-17 WO 02/090584 PCT / FR02 / 01542 32 (m, 1H, CH-CH-NH); 3.12 (m, 1H, CH-S); 2.83 and 2.59 (ABX system, 2H, 2J, ~ = 5Hz, 3J ~ = 12Hz, 3JBX = 0Hz, CH2-S); 2.27 (t, 2H, J = 8Hz, CH2-CO); 1.58-1.35 (m, 6H, (CH2) 3) ~ Example 2: Maraua ~ e reagent with Cy5: Cy5-PMDAM + A ~ Zo ~ / \ / e + PhHN ~ NHPh -> I + I ~ N ~ Ac ~ + C¿ I Ph 5 6 AczO \ / \ d N ~ ~ 'N I + I ~ OH O _ Cy5-COOH OH + N Br \ / N + Br ~ - ~ - ~ OH 10 O 7 $ 9 CHa H3 CH3 CHs ClC00r $ u ~ ~ O NHaNHz I ~ N-NH2 MnO2 / Nz + ~ ~ O -ai -ai 1 Cy5 NH _ Cy5 NH Cy5 ~ NH _ NHz CNJ ~ X Ö Cl O Cl Me 11 12 I2bis RR \ / N Ii, N \I -> \ I> I s, N \I

I ~ Nz ~ + + I w ~ ~ + _ I OH NHz X ~ NH OO 10 Cy5-COOH ~ Iodide compound of 2-j4- (N acet ~ -N phenylaminolbuta-1 3-dienyl] -1 2 3 3- tetramethyif 3H1 1o indolium 6 The mixture of malonaldehyde-bis (phenylimine) monohydrochloride 5 (18.3 g; 70.0 mmol), NaOAc (9.0 g; 110 mmol) and iodide of 1,2, 3,3-tetramethyl [3H] indolium 4 (4.25 g; 14.1 mmol) in (acetic anhydride (75 mL) is heated at 110 ° C for 20 min. After cooling, ether (350 mL) is added and the precipitated brown solid is filtered off, washed with (ether (3 x 100 mL). The solid is redissolved in 150 mL of CH2Cl2, filtered (removal of inorganic salts) then evaporated to give a brown solid (6.0 g, 90%).

[0041]

CA 02444926 2003-10-17 WO 02/090584 PCT / FR02 / 01542 33 1H NMR (CDCl3): S = 8.64 (d; 1H; J = 12 Hz; 1-H); 8.14 (t; 1H; J = 16; 12Hz; 3-H); 7.63-7.19 (m; 9H); 6.90 (d; 1H; J = 15Hz; 4-H); 5.82 (t; 1H; J = 12; 13 Hz; 2-H); 4.06 (s; 3H; NCH3); (2.16 (s; 3H; -COCH3); 1.74 (s; 6H; CH3).

~ Bromide compound of 1- (5-carboxypentyl) -2,3,3-trimethyl [3Hlindolium 9 We mix 2,3,3-trimethylindole 7 (10.0 g; 62.8 mmol) and (6-bromohexanoic acid 8 (12.3 g; 62.8 mmol) without solvent and heated at 110 ° C for 12 h under argon. The red-violet pasty reaction mixture is washed with (ethyl acetate (2 x 60 mL, the paste is triturated 1o with the spatula and the supernatant is decanted), then with (acetone (50 mL, the paste solidifies) . The pink solid is filtered off and then dried under vacuum (16.0 g; 73%). ~ Compound CySCOOH 10 The mixture of (iodkie 6 (6.0 g; 12.7 mmol), bromide 9 (4.5 g; 12.7 mmol) and NaOAc (2.6 g; 32 mmol) in (acetic anhydride (35 mL) is heated at 110 ° C for min. After cooling, (ether (150 mL) is added and the precipitate is filtered off and washed with (ether (3 x 50 mL). The solid is dissolved in 100 mL of CH2Cl2, filtered and purified by chromatography on a SiO2 column (eluent: 5-10% MeOH / CH2Cl2). 3.4 g (44%) are obtained.

1H NMR (CDCl3): b = 8.03 (t; 2H; J = 10; 11Hz, 2-H, 4-H); 7.38-6.91 (m; 9H; Ar-H, 3-H); 6.41 (d; 1H; J = 14Hz; 1-H); 6.31 (d; 1H; J = 13Hz; 5-H); 4.07 (t; 2H, J = 7; 7Hz; α, -CHZ); 3.68 (s; 3H; NCH3); 2.47 (t; 2H, J = 7; 7Hz; s-CHZ); 1.71 (m; 18H; CH3, ~ 3, y and ~ -CH2). ~ Coupling compound of 3-aminoacetophenone with CySCOOH 10 (product 11) A a solution of CySCOOH 10 (1.19 g; 1.9 mmol) in 12 mL of CHaCl2, N methylmorpholine (360 ~. L; 3.2 mmol). The solution is cooled with an ice bath, then isobutyl chloroformate (480 ~. L; 3.7 mmol). After 5 minutes of stirring, 3-aminoacetophenone (488 mg; 3.6 mmol) is added. The mixture is stirred at temperature

[0042]

IDT Exhibit 1020.10036

CA 02444926 2003-10-17 WO 02/090584 PCT / FR02 / 01542 34 ambient for 3 h.

By adding 250 mL of ether, a pasty solid is obtained. After stirring, the solid is allowed to stand at the bottom of the flask and the supernatant is decanted. Another 50 ml of ether is added and triturated with a spatula to obtain a solid. This is filtered, washed with water, ether, then dried under vacuum. The product (iodide) is then dissolved in ethanol, passed through a column of amberlite 1RA900 (Cl-; 15 g). The ethanolic solution collected is evaporated to dryness and then passed through a SiO 2 column. Obtained 0.93 g (77%) of blue solid. ~ Compound Cy5-hydrazone 12 is added to a solution of facetophenone 11 (0.93 g; 1.46 mmol) in 5 mL of absolute ethanol fhydrazine monohydrate (180 p. L; 3.1 mmol) which is stirred at room temperature for 7 h. 50 mL of ether are added and the precipitate is filtered off and washed with (ether. The crude product is dissolved in 50 mL of CH2Cl2, the solution is filtered and then concentrated to 10 mL. The product is precipitated by adding 100 mL of ether, filtered, washed with (ether and dried under vacuum.

540 mg of product 12 (57%) are obtained. ~ Compound CySPMDAM l2bis To a solution of 100 mg of hydrazone 12 in 2 mL of DMF, 300 mg of MnO2 are added and the mixture is stirred vigorously for 10 min. The suspension is filtered through a layer of celite and washed with DMF (3x500 ~, L). 50 ml of ether are added and the deposited oil is triturated with a spatula and the supernatant is decanted. It is repeated (washing operation three times with 25 ml of ether and the solid thus obtained is filtered off and dried. 65 mg (65%) of product 12bis are obtained. The purity of the product is about 80-85% (1H NMR). Example 3: Other reagents synthesized Example 3.1: Synthesis of para-nitrophenyldiazomethane (NPDAM)

[0043]

CA 02444926 2003-10-17 WO 02/090584 PCT / FR02 / 01542 4-nitrobenzaldehyde hydrazone is commercially available (reference 28,118-2, Aldrich, France).

We therefore work on 600 mg (3.64 mole) of this product which is dissolved in 9 mL of THF. The solution is left under stirring for 5 minutes then 1.26 g (4 equivalent, 14.56 mmol) of MnO2 are added with caution. The mixture is left under stirring for 10 minutes, then filtered. The filtrate recovered is evaporated to dryness. After washing with pentane, the paya-nitrophenyldiazomethane compound is obtained in the form of a bright orange powder with a yield of 79% (468 mg, 2.87 mmol). 1o F 80-82 ° C. - 1H NMR (300 MHz, DMSO-d6) 8 = 8.11 (d, 2H, J = 9 Hz, Ar- H); 7.18 (d, 2H, J = 9Hz, Ar-H 3; 6.06 (s, 1H, CH1-N2). Example 3.2: Synthesis of phenylmethyldiazomethane (PMDAM) Acetophenone hydrazone: Acetophenone (2.0 g, 16 mmol) is diluted in 16 mL of absolute ethanol and then hydrazine (2.3 mL, 48 mmole). We bring to reflux. After 2 h, the solvent is evaporated off and the residue is taken up in ether (150 mL). Washed with water (100mL).

After drying over Na2SO4, the ether is evaporated off. The pale yellow oil is obtained (1.5 g, 11 mmol, 69%). 20 Phenylmethyldiazomethane (PMDAM): The hydrazone (150 mg, 1.1 mmol) is dissolved in 3 mL of THF. MnO2 (480 mg, 5.5 mmol) is added. Stirred for 30 min at room temperature. The middle is colored red. It is filtered and the solvent is evaporated off. The red oil is obtained (145 mg, 100%). This reagent is used without purification. 25 Example 3.3: Synthesis of diphenyldiazomethane (DPDAM) Benzophenone hydrazone is commercial (reference B 960-2 Aldrich, France). We therefore work on 196 mg (1.0 mmol) which is dissolved in 5 mL of THF. 435 mg (5 eq, 5.0 mmol) of MnO 2 are added, the mixture is left under stirring for 10 3o minutes, then filtered. The filtrate recovered is evaporated to dryness. 193 mg (0.99 mmol) are obtained. This reagent is used without purification.

[0044]

CA 02444926 2003-10-17 WO 02/090584 PCT / FR02 / 01542 36 Example 3.4: Synthesis of NVDAM N02 NOZ ~ O 1) NH2-NH2, EtOH ~ ~ N

CH30 ~ 2) Mn02, THF CH30 OCH3 OCH3 NVDAM The synthesis is carried out according to the protocol described above from 6-nitrovératraldehyde (Aldrich, reference 27,960-9). Exem 1p e 4 ~ Synthesis of the biotinylated derivative from NPDA.M (according to ref.) 1. Pairing 1. HOCH2CH2OH, TsOH 2.hydrazine O 2. Na2S, reflux 3. Mn0 Bio N02 NH2 2 NH O2N ~ ~ CHO ....., O2N ~ ~ CHO -..... ~ O ~ N ~ ~ CHN2 M. E. Wall et al., J. Med. Chem. , 1993, 36, 2689. The 2-amino-4-nitro benzaldehyde derivative is prepared according to the method of ME Wall et al referenced above. The preparation of diazomethane NPDAM is identical to that described in Example 3.1 above. Example 5: Preparation of DNA nucleic acids The DNA amplicons are generated by PCR from Mycobacte ~ ium tuberculosis 16S genomic DNA targets (10 + 4 copies as starting targets) using the Fast kit

[0045]

IDT Exhibit 1020.10037

CA 02444926 2003-10-17 WO 02/090584 PCT / FR02 / 01542 37 Start from Roche, 0.2 mM of each deoxyribonucleotide (d-ATP, d-CTP, d-GTP, d-TTP), 0.3 ~.

M of primers and 0.4 ~, L of enzyme. The PCR parameters are as follows - 95 ° C: 4 min then 35 cycles (95 ° C: 30 sec; 55 ° C: 30 sec; 72 ° C: 30 sec) then 4 ° C. The amplicons are qualitatively analyzed by agarose gel electrophoresis (1.5%, TBE 0.5X). The deposited volume is 5 ~ L and the migration is carried out for 20 min at 100 volts (V). Visualization of the PCR products is carried out under a UV lamp after staining with ethidium bromide. The conditions for the culture, the extraction of the Mycobacteria as well as the amplification primers are given in patent application WO-A-99/65926. Example 6 Reactivity of the maraua ~ e reagents on model nucleotides The synthesis of the labeling reagents is described in Examples 1 to 4 above. The PDAM described in the present invention is commercially available. (reference P 1405, Molecular Probes, Eugene, OR) Example 6.1: Labeling of UMP 3'-phosphate monomers The reactivity of the labeling reagents bearing a diazomethyl function was studied in order to monitor the specificity of the reaction.

A protocol has been developed which consists in studying this reactivity by capillary electrophoresis on a model compound 3'-UMP (LTridine 3'monophosphate, reference U1126 Sigma) under the following standard conditions 3'-UMP 0.04 mM; 2.0 mM H3B03; 2.0 mM marker [added with an appropriate organic solvent (THF, AcOEt or DMSO)]; Solvents H20 - CH3CN -Organic solvent (ratio 1/3/1).

[0046]

CA 02444926 2003-10-17 WO 02/090584 PCT / FR02 / 01542 38 This solution is divided into ten fractions of 250 ~ L which are heated to 60 ° C. After a defined time, each fraction is treated with 250 ~, L of dichloromethane.

After stirring, the organic phase (lower phase) is removed. This operation is repeated two more times. After centrifugation (5 min, 5000 revolutions per minute (rpm)), the aqueous phase is analyzed by capillary electrophoresis. The conditions for capillary electrophoresis (CE) are as follows: EC analysis was performed by the Beckman P / ACE 5000 apparatus. An untreated fused silica capillary (75 ~, m x 50 cm) was used. The applied voltage was 30 kV (normal polarity) and the temperature of the capillary was maintained at 23 ° C. Electropherograms were recorded at 254 nm. Borate buffer solution (0.1M, pH 8.3) was prepared from boric acid by adjusting the pH with NaOH solution and filtered through a 0.2 ~, m filter. The samples were injected under pressure (0.5 psi, 5 seconds). Each analysis is preceded by regeneration of the capillary by successive passages of a solution of NaOH (0.1 M, 2 min), water (2 min) and borate buffer (2 min) under pressure (20 psi).

The reaction is carried out by varying the reaction time between 0 and 4 hours, as indicated in each figure, and the results are shown for each reagent tested in Figures 2A to 2I with the concentration of reagent used. The reaction time is indicated on each electropherogram. 2o With all the reagents tested, we obtain the exclusive formation of a monoalkylated product, which proves the specificity of the reaction. The reactivity, that is to say the reaction time of the reagent on the 3'-UMP, can thus be calculated, by comparison of the height of the peaks, and the results are presented in Table 2 below (standard conditions described above) HzC H Formulas C = NC = N

[0047]

z z H ~~ ~ ~ -Nz, N I 'ozN

[0048]

chemical z ~ and

[0049]

names PDAM DPDAM PMDAM NPDAM

[0050]

Reactivity 5 min 20 h 30 min 4 h

[0051]

CA 02444926 2003-10-17 WO 02/090584 PCT / FR02 / 01542 39 i HaC H C Ha Formulas ~ N " ~ H3 Nz ~ NH N

IDT Exhibit 1020.10038

[0052]

'x chemicals' Bio ~ NH II I

[0053]

o ~ p NH B Bio ~ NH Bto ~ II II O

[0054]

and ~

[0055]

names BioDPDAM m-BioPMDAM p-BioPMDAM o-BioPMDAM

[0056]

Reactivity 10 h 15 min 5 min 1 min

[0057]

* With o-BioPMDAM, the reactivity is estimated because under the conditions of concentration of 2 mM in labeling reagent, the reaction is complete in less than 5 min.

FIG. 2I therefore uses a concentration of 0.2 mM. Table I: Reactivity (half-reaction time) of reagents It can nevertheless be noted that, the reaction being very specific and not leading to by-products for all the reagents tested, it is possible to increase the concentration of the labeling reagent without consequence from the point of view of selectivity on the labeling.

Thus for the DPDAM reagent, if the concentration is increased to 20 mM (see FIG. 2C), the reactivity (half-reaction time) is 2 hours.

The same result is obtained with BioDPDAM (FIG. 2F) where the reactivity is 45 min at a concentration of 30 mM. Example 6.2: Test of different 3'-monophosphate nucleotides In order to avoid any misinterpretation, a complementary study on the mita-BioPMDAM marker, taken as a significant example, was carried out with the other 3'-monophosphate nucleotides .

The nucleotides tested are as follows: 3'-AMP (reference 85,194-9, Aldrich), 3'-GMP (reference 151214, ICN), 3'-CMP (reference C1133, Sigma), 3'-TMP (deoxyribo series) (reference T1008, Sigma).

The electropherograms obtained with the different nucleotides are represented in FIGS. 3A to 3D. The reaction times indicated in Figuxe 3A are identical for Figures 3B to 3D. Whatever the starting nucleotide (ribo or deoxyribo series), we observe the exclusive and complete formation of the alkylated product in 130 min at 60 ° C. It is important to note that in

[0058]

CA 02444926 2003-10-17 WO 02/090584 PCT / FR02 / 01542 the case of guanine (the most reactive base with the usual alkylating reagents), only the product alkylated with phosphate is observed, proving the very high selectivity of the reaction.

This study also makes it possible to verify that the speed of the reaction does not depend on the nature of the nucleotide as substrate. s Example 6.3: Study of a dinucleotide The alkylation of the dinucleotide ApUp (reference A4298, Sigma) was carried out with meta-BioPMDAM in order to verify the selectivity of the reaction towards the terminal phosphate with respect to the phosphate. internucleotide.

The monitoring of the reaction by electrophoresis is shown in Figure 4.

The conditions are the standard conditions of Example 6.1 already described.

The exclusive formation of a single product is observed showing a good selectivity of the meta-BioPMDAM reagent towards terminal phosphate over internucleotide phosphate.

IDT Exhibit 1020.10039

Example 6.4: Characterization of the adducts with the four (4) 3'-monophosphates nucleotides To ensure that the products obtained did indeed come from an alkylation with phosphate, the synthesis of the adducts of the 3'-UMP, 3'-CMP monophosphates, 3'-GMP and 3'-AMP with the meta-BioPMDAM reagent was performed.

The alkylation reaction is carried out at (preparative scale as indicated below.

The adducts, obtained with yields of the order of 70%, are purified and then studied by proton and phosphorus NMR. Preparation protocol The 3'-UMP (in the form of disodium salt; 9.3 mg; 21.1 mol) is dissolved in 2 mL of a 3o 0.1 M aqueous solution of H3B03, then 2 mL of H3B03 is successively added. CH3CN, 6 mL of MeOH then the meta-BioPMDAM reagent (75 mg; 0.20 mmol). The reaction is carried out for

[0059]

CA 02444926 2003-10-17 WO 02/090584 PCT / FR02 / 01542 41 2.5 h at room temperature.

It is followed by capillary electrophoresis. 3 ml of water are added, then (excess of the reagent is removed by extraction with CH2Cl2. The aqueous phase is evaporated. The residue is dissolved in a small amount of water and purified by passage through a column of reverse phase silica gel (Lichroprep RP-18, Merck; MeOH / H ~ O elution (20180)). 10 mg (69%) of 3'-UMP adduct are obtained. The 12MN spectra of the proton obtained for the 3'-NMP adducts (N = G, U, C, A) are presented in FIGS. 5A to 5D. The identification of the adducts was carried out by two-dimensional 1H / 1H NMR experiments (COZY). Two diastereoisomers for each of these 1o adducts in a 1/1 ratio are present. A single peak is present for the phosphorus NMR around 0 ppm (300 MHz, D2O). These experiments demonstrate that the reaction is very specific, that a single adduct is observed and that the labeling does indeed take place on the phosphorus. There is no side alkylation reaction on the bases.

The labeling products are therefore particularly suitable for a hybridization step. Example 7: Temperature stability study 2o All the diazomethane derivatives, described in Table 1 of Example 6.1 above, are stored in the solid state in the freezer at -20 ° C for at least three (3) months and no loss of reactivity is observed. The stability at room temperature on the bench was determined by 1 H NMR for the two reagents NPDAM and meta-BioPMDAM. We observed no decomposition by leaving the NPDAM for a month on the bench without special precautions. We observe approximately 50% decomposition by leaving meta-BioPMDAM on the bench for twenty-five (25) days. The temperature stability of the labeling reagent is an essential characteristic. Indeed, the final destination of such a reagent for industrial application is a 3o labeling kit. A reagent which is not stable for at least fifteen (15) days at -20 ° C, preferably one (1) month is unsaleable.

Even if there are means of storage and sending

[0060]

CA 02444926 2003-10-17 WO 02/090584 PCT / FR02 / 01542 42 down to -110 ° C, there is a relationship between the stability at - 110 ° C and at -20 ° C and that is why the value fifteen (15) days at -20 ° C, preferably one (1) month at -20 ° C is an industrial minimum.

Beyond -110 ° C, laboratories do not have the necessary equipment (freezer) to store these reagents from the point of view of the user and the manufacturer, and there is no simple dry ice-type way to ship them. from the manufacturer's point of view. With respect to stability at room temperature, a stability of a few hours, preferably one (1) day, is sufficient to allow the user to perform the labeling. 1o Example 8: Study of DNA fragmentation Example 8.1: Hydrolysis of different nucleosides in an acidic medium The aim of this study is to demonstrate the difference in terms of stability at acid pH between the 15 natural nucleosides, the modified nucleosides as well as purine and pyrimidine nucleosides. This study also makes it possible to better control DNA fragmentation by taking into account its base composition. NHz NHz 0 I Br H ~ N CH3 ONO HO NONO HO O HO O OH OH OH O NHz NHz H ~~ N ~ N ~ 'N ~ N ~ NHNN ~ Ni ~' Br z N ~ NN ~ N HO O HO O HO O OH OH OH

[0061]

CA 02444926 2003-10-17 WO 02/090584 PCT / FR02 / 01542 43 Two modified nucleosides, 8-bromo-2'-deoxyadenosine (8-BrdA) and 5- bromo-2'-deoxycytidine (5-BrdC) as well as the four (4) natural nucleosides (dA, dC, dG and dT) were used in this study. 50 nanomoles (nmol) of each nucleoside are incubated in 50 mM sodium formate pH 3 at 95 ° C. Incubation times vary from 0 to 45 min.

After drying under vacuum and taken up in 20 µl of pure water, the samples (10 nmol) are then analyzed by reverse phase HPLC. The results are given in the form of percentage of hydrolysis of the starting nucleoside (on the y-axis) relative to the

IDT Exhibit 1020.10040

incubation time in minutes (on the x-axis) see figure 8. 1o The curves of figure 8 show that the modification of the adenine in position 8 by a bromine atom makes this nucleoside less stable than the natural nucleoside. On the other hand, the results show that, under the conditions used, depurination is much more important than depyrimidination. This study shows that DNA fragmentation by depurination or depyrimidination can be controlled by optimizing hydrolysis conditions, or by incorporating either modified bases less stable than natural bases or bases that can be modified and hydrolyzed after their incorporation. Example 8.2: Fragmentation of double-stranded DNA whether or not incorporating a modified nucleotide Three PCR amplifications were carried out in parallel from the genomic DNA target Mycobacte ~ ium tuberculosis 16S (10 + 4 starting copies) using the Fast Start kit from Roche, 0.2 mM of each deoxyribonucleotide (d-ATP, d-CTP, d-GTP, d-TTP), 0.3 p.

M of primers and 0.4 ~, L of enzyme. The parameters of the PCR are those of Example 5. In the first case, the protocol is used as is: the case of so-called natural PCR. In the second case, the protocol is modified to obtain a PCR at 30% of 8Br-dATP. This is done by introducing 0.2 mM of d-CTP, d-GTP and d-TTP, 0.14 mM of d-ATP and 0.06 mM of 8-BrdATP are also introduced. (8-BrdATP is commercial (reference N 2005-1, TriLink Biotechnologies, San Diego CA).

[0062]

CA 02444926 2003-10-17 WO 02/090584 PCT / FR02 / 01542 44 In the third case, the protocol is modified to obtain a PCR at 50% of 8-BrdATP. This is done by introducing 0.2 mM of d-CTP, d-GTP and d-TTP, 0.1 mM d-ATP and 0.1 mM 8-BrdATP are also introduced. The study of the only fragmentation of these amplicons was carried out under the conditions described above: 50 mM of sodium formate pH 3 at 95 ° C. Analysis was performed on denaturing polyacrylamide gel (8% polyacrylamide, 7 M urea, 1 X TBE) using ethidium bromide staining. 1o After 15 min of incubation at 95 ° C in 50 mM sodium formate pH 3, no difference is visible between the three (3) targets.

In all cases, we observed a total fragmentation of the PCR amplicons. The depurination of the PCR amplicons was also carried out at different pH and at different temperatures and incubation times. Analysis on gel, under the above conditions, shows that at pH 3, fragmentation is complete after only 10 min of incubation at 95 ° C. At this pH, fragmentation is also complete at 60 ° C. after 30 min of incubation. At pH 4, 30 min of incubation are necessary to obtain a total fragmentation of the DNA amplifiers even at 95 ° C. This result is very interesting and shows that the 2o abasic site generated at acidic pH is unstable and therefore leads to the fragmentation of the DNA chain without any special treatment. Example 9: Maraua ~ e and DNA fragmentation uar the naeta-bioPMDAM derivative (3a) in two steps The meta-bioPMDAM derivative (3a) was obtained according to the reaction scheme described in Example 1.1.

The DNA amplicons were prepared by PCR amplification according to the protocol described in Example 5.

[0063]

CA 02444926 2003-10-17 WO 02/090584 PCT / FR02 / 01542 Mar ua ~ e A 10 ~, l of PCR, 38 ~, L of pure water (sigma), 50 ~.

L of sodium formate at pH 3 (100 mM in pure water) are added and the mixture is incubated for 30 minutes at 60 ° C. Then 2 ~. L of meta-bioPMDAM (100 mM in DMSO) are then added. The solution was mixed by vortexing, then incubated an additional 15 minutes at 60 ° C. The tests are carried out in duplicate in order to be able to analyze the fragmentation of the DNA on gel and the efficiency of labeling by hybridization and reading of the DNA chip. Purification Purification is performed on QIAQUICKT "" columns (Nucleotide Removal kit, Qiagen, Hilden, Germany). The purification protocol used is that recommended by the supplier. After purification, the eluate is transferred to a clean tube containing 400 wL of hybridization buffer (1.75 mL 20X SSPE; 2.9 mL 5M Betaine; 280 ~, L 0.1M DTAB; 10 ~ L Antifoam 30%) are added. The references of these substances are, for ~ the betaine reference B-2754 Sigma, and ~ the DTAB reference D-5047 Sigma.

The solution is mixed by vortexing and incubated 10 min at 95 ° C, in order to separate the DNA strands which are not separated during the labeling step during fragmentation (denaturation step). The tube is then immersed in a water-ice mixture at 0 ° C. before hybridization on a DNA chip. Hybridization on a DNA chip 25 After the labeling step during the fragmentation, the fragments obtained are hybridized on the DNA chips designed for the analysis of the region 213-415 of the sequence M20940 "Genbank" of the 16S RNA of Mycobacterium tuberculosis. This DNA chip is described in A. Troesch et al, J. Clin Microbiol., 37 (1), p49-S5, 1999. The hybridization steps were performed on the fluidic stations (Affymetrix, Santa Clara, CA) using the hybridization protocol and buffers described in A. Troesch et al, J. Clin

[0064]

CA 02444926 2003-10-17 WO 02/090584 PCT / FR02 / 01542 46 Microbiol., 37 (1), p49-55, 1999.

IDT Exhibit 1020.10041

An additional step is necessary to reveal the biotin (indirect detection). Hybridization is revealed by the coupling of streptavidin labeled with phycoerythrin (PE) which interacts with the biotin of meta-BioPMDAM under the following conditions 300 ~. L of pure water; 300 ~ L of 100 mM Tris buffer pH 7 / 1M NaCl / 0.05% tween / 0.005% defoamer; 6 ~. L of BSA (50 mg / mL); 6 ~ L of streptavidin-PE (300 ~ g / mL). The references of these substances are, for ~ Streptavidin-Phycoerythrin: reference 80438, Dako, Denmark, ~ Streptavidin-CYS: reference 00050, Dako, Denmark, antifoam reference MS-575, Ultra Additives Inc., and ~ Tween reference P-7949, Sigma. Reading of the DNA chip The reading of the fluorescence emitted on the surface of the DNA chip after labeling and hybridization as well as the generation of the data in terms of signal intensity and the percentage of homology are carried out by the reading systems. and software provided by Affymetrix (GeneChip ~ Instrument System and GeneChip ~ Information System, Santa Clara CA).

The reading system supplies signal and background noise intensities expressed in rfu ("relative fluorescence unit"). The percentage of homology is given relative to a reference sequence which in this case is the sequence of Mycobacterium tuberculosis. The results in terms of mean signal intensity (I), background noise (B) and percentage homology (%) are given in Table 2 below. In general, it is considered that a percentage of Homology greater than 90% is a satisfactory result although one generally seeks a result greater than 95%. Above 95%, the values are no longer indicated because they are not significant in the case of the Mycobacteria DNA chip. A high intensity with a low background noise is the second result sought in the examples which follow. In all the results, the background noise B is 3o deduced from the average intensity I.

[0065]

CA 02444926 2003-10-17 WO 02/090584 PCT / FR02 / 01542 47 Analysis on polyacr ~ lamide gel The samples intended to be analyzed on gel are dried under vacuum, taken up in 10 p, L of pure water and 10 ~ .

L of 2X formamide blue. The migration is carried out on 8% acrylamide gel, in 1X TBE, one (1) hour at 150 V. The acidic pH was used for the fragmentation of the DNA. Indeed at this pH, the phenomenon of depurination generates very unstable abasic sites leading to an almost immediate fragmentation of the DNA sequences at high temperature. This type of fragmentation produces DNA-5 'phosphate fragments. The gel analysis shows that the incubation of the PCR amplicons at 60 ° C. for 30 min in solution in formate buffer (50 mM, pH 3) leads to total fragmentation of these amplicons. This allowed us to evaluate the labeling during the fragmentation of DNA amplicons in the presence of the meta-bioPMDAM marker. The results of labeling during the fragmentation of DNA amplicons in terms of percentage homology, signal intensity and background noise are given in Table 2 below. Labeling conditions for Homology I (rfu) B (rfu) I / B PCR amplicons (%)

[0066]

Buffer: formate pH3, 50 mM

[0067]

Marker: 2 mM mta-bioPMDAM> 95 4456 593 7,5

[0068]

Incubation: 30 min 60C

[0069]

Table 2: Labeling and fragmentation of DNA amplicons in terms of homology, signal intensity (i) and background noise (B) This example shows that the derivatives of the invention can be used for the labeling of DNA fragments produced by enzymatic amplification in a two-step protocol.

They can also be used for the labeling of unamplified natural DNA.

[0070]

CA 02444926 2003-10-17 WO 02/090584 PCT / FR02 / 01542 48 Example IO: Maraua ~ e of the DNA uar the derivative Cy5-PMDAM (iZbis) The labeling of the DNA by this new marker carrying the diazomethyl function was evaluated using a synthetic DNA fragment.

IDT Exhibit 1020.10042

~ / I 'N2 ~~ / ~' ~ Nue 'NH O Cy5-PMDAM (I2bis) The labeling reagent Cy5-PMDAM (I2bis) is prepared according to the protocol described in Example 2.

A twenty (20) seas oligodeoxyribonucleotide (ODN) is prepared according to the so-called phosphoramidite method.

A phosphate is introduced at the 5 'end by a standard phosphorylation reagent compatible with phosphoramidite chemistry.

The sequence of this ODN consists of all the natural bases of DNA (sequence of the ODN: 5'-CTGAACGGTAGCATCTTGAC-3 '). This sequence is complementary to the capture sequences of a so-called "model" DNA chip, synthesized according to Affymetrix technology. This DNA chip contains identical capture probes in sequence and which are distributed in a checkerboard pattern on its surface. Reading this DNA chip gives information as to the 2o performance of the labeling in terms of intensity but no homology result. Brand: At 50 picomoles (pmoles) of this ODN, 10 ~. L of Cy5-PMDAM (100 mM in DMSO) are added. The final volume is 100 ~. L. After homogenization, incubation is carried out at 60 ° C for 30 minutes. Purification and reading: The purification to remove the excess labeling reagent is carried out according to Example 9. The DNA chip reading is carried out according to Example 17.

[0071]

CA 02444926 2003-10-17 WO 02/090584 PCT / FR02 / 01542 49 Results The average of the labeling intensities (I) read on the DNA chip is 16,644 rfu for a background noise (B) of 450 rfu.

This level of intensity is very high and shows that the Cy5-PMDAM labeling reagent (12 bis) is completely compatible with the labeling of DNA fragments on the phosphate group. 1o Example 11: Maraua ~ e and DNA fragmentation by the PDAM reagent: Example 11.1: Labeling with 1-Pyrényldiazométhane (PDAM) CH ~ N ~ 1-Pyrényldiazométhane The PDAM is obtained from Molecular Probes (Eugene, OR) and solubilized in anhydrous DMSO. 2o Two twenty (20) mer ODNs were used as DNA templates: a 20mer 5'-hydroxyl ODN and the same 20mer ODN carrying a phosphate at the 5 'end. The sequence of the ODN is described in Example 10. The labeling reaction is carried out in a mixture of 50% DMSO and 1.5 mM 1-Pyrényldiazomethane (PDAM) at 60 ° C. for 30 minutes or one hour. The efficiency of the labeling was evaluated by thin layer chromatography (in normal phase) in an eluent isopropanol / ammonia / water 60110/30.

After 30 minutes, the coupling is complete on the ODN 5'-phosphate. It takes one hour to obtain partial coupling on the 5'-hydroxyl ODN, ie about 50%.

[0072]

CA 02444926 2003-10-17 WO 02/090584 PCT / FR02 / 01542 The results of Example 6 are confirmed on a model sequence of 20 bases as regards the preferential labeling of the reagents carrying a diazomethyl function on the terminal phosphate.

Labeling on an intranucleotide phosphate is not a crippling drawback since it can lead to an increase in sensitivity by introducing more than one label on the nucleic acid fragment. This allows the nucleic acid to hybridize with good sensitivity to the complementary target while retaining good hybridization specificity. Those skilled in the art by optimizing reactions can thus control the specificity of the labeling, by playing for example on the labeling reagent, the reaction time and the temperature, in order to have an exclusive labeling on the terminal phosphate. Example 11.2 Kinetic study of the PDAM labeling reaction This study was carried out using the ODN 20mer 5'-phosphate under the above conditions by varying the reaction time. The labeling yields were evaluated by reverse phase high pressure liquid chromatography (HPLC) analysis under the following conditions Reverse phase Spheri-5 RI'-18 5 ~, m, 220x 4.6 mm column (Perkin Elmer).

The buffers and the gradient used are 20 ~ Buffer A: 0.1 M TEAA; Buffer B = 50% Buffer A + 50% CH3CN, and ~ Gradient of 10 to 50% B in 30 min at 1 ml / min at room temperature. The results are shown in FIG. 9 with the reaction time on the abscissa, expressed in minutes, and the percentage labeling on the ordinate. 25 The yield is close to 90% after only 10 min of incubation at 60 ° C. Example 11.3: Effect of temperature on PDAM labeling Labeling was carried out using the ODN 20mer 5'-phosphate under the above 3o conditions by varying the incubation temperature and with an incubation time of 10 minutes in each case.

[0073]

CA 02444926 2003-10-17 WO 02/090584 PCT / FR02 / 01542 51 The labeling yields were evaluated by reverse phase HPLC analysis.

The results are presented in figure 10 with the percentage of labeling on the y-axis and the reaction temperature in ° C on the x-axis. It is very important to note that even at room temperature (25 ° C), we observe a labeling of the ODN. . After 10 minutes

IDT Exhibit 1020.10043

of incubation at 25 ° C., the labeling yield is approximately 25%. At temperatures above 50 ° C, yields above 80% are obtained. This shows the efficiency and flexibility of this DNA labeling chemistry with reagents carrying a diazomethyl function. Example 12: Marcrua ~ e and fragmentation of DNA amulicons obtained by PCR amplification with the maraua ~ e meta-bioPMDAM reagent (3a) The meta-bioPMDAM derivative (3a) was obtained according to the reaction scheme described in Example 1.1. The DNA amplicons were prepared by PCR amplification according to the protocol described in Example 5. Example 12.1: Comparison of labeling with and without fragmentation a. Ma ~ guage dais the co ~ zditions of fi ~ a ~ merztatioia A 10 ~ CL of PCR are added 50 ~.

L of sodium formate pH 3 (50 mM) and 2 p. L of m ~ ta BioPMDAM 100 mM in DMSO). The volume is adjusted to 100 ~, L. The solution is incubated for 30 min at 60 ° C. b. Brand without, fr ~~ mentation A 10 ~. L of PCR are added 2 ~. L of meta-BioPMDAM (100 mM in DM80). The volume is adjusted to 100 ~. L. The solution is incubated for 30 min at 60 ° C. The rest of the protocol is identical to that of Example 9.

[0074]

CA 02444926 2003-10-17 WO 02/090584 PCT / FR02 / 01542 52 Results Protocol used Homology I (rfu) B (rfu) I / B (%)

[0075]

at. Marking under the conditions of

[0076]

>95 3995 569 7,0

[0077]

fragmentation

[0078]

b. Marking without fragmentation 94.1 500 542 0,9

[0079]

Table 3: Comparison of labeling with and without fragmentation The results in Table 3 above show that without fragmentation, the mean of the intensities obtained is at the same level as the background noise (500 rfu).

Labeling during fragmentation gives a much higher level of intensity (about 4000 rfu) and a very good percentage of homology.

The combination of the two steps therefore clearly represents a significant improvement for the detection of a nucleic acid longer than one hundred nucleotides.

Example 12.2 Effect of denaturation before hybridization on a DNA chip Two labeling reactions were carried out in parallel in two separate tubes according to the following protocol: To 10 wL of PCR are added 50 ~ L of sodium formate buffer pH 3 (50 mM) and 2 p.

L of meta-bioPMDAM (100mM in DMSO).

The total volume is adjusted to 100 ~ L and incubated for 30 min at 60 ° C. After purification on a column (example 9), the solution from the first tube is incubated for 10 min at 95 ° C (in order to mismatch the DNA double strand), then the tube is immersed in a water-ice mixture at 0 ° C. until hybridization on the DNA chip. The solution from the second tube is hybridized on the DNA chip without prior denaturation.

[0080]

CA 02444926 2003-10-17 WO 02/090584 PCT / FR02 / 01542 53 The biotinylated fragments hybridized on the capture probes on the surface of the DNA chip are revealed by the introduction of a streptavidin labeled with phycoerythrin using the conditions described in Example 9.

IDT Exhibit 1020.10044

Results Conditions used Homology I (rfu) B (rfu) I / B (%)

[0081]

With denaturation> 95 22812 570 40,1

[0082]

Without denaturation 93.5 4795 661 7,0

[0083]

Table 4: Effect of denaturation before hybridization on a DNA chip 1o The results obtained, presented in Table 4, with the prehybridization denaturation are superior to those obtained without the denaturation step.

This shows that DNA denaturation is necessary to obtain a good level of intensity.

Fragmentation via the abasic sites is a means of facilitating the denaturation of a double-stranded DNA and of reinforcing the hybridization on the capture probes.

To test other labeling reagents and taking into account the results obtained with the various conditions above, we have defined a reference protocol using fragmentation in sodium formate buffer (50 mM, pH 3) and a pre-denaturation step. - hybridization.

Example 13: Margua ~ e and fragmentation of the PCR amplicons by the biotinylated reagents in a one-step protocol The mita-, o ~ tho-, and para-bioPMDAM derivatives were prepared according to the protocol described in Examples 1.1, 1.2 and 1.3 . They were solubilized in anhydrous DMSO at a concentration of 100 mM.

[0084]

CA 02444926 2003-10-17 WO 02/090584 PCT / FR02 / 01542 54 The protocol is identical to that of Example 12.2 above (labeling and fragmentation in one step then a prehybridization denaturation step).

Results Marker Homology I (rfu) B (rfu) I / B (%)

[0085]

ortho-bioPMDAM> 95 25 951 820 31,8

[0086]

mta-bioPMDAM> 95 22 960 581 '39,5

[0087]

para-bioPMDAM 94.1 43785 1205 36,3

[0088]

Table 5: Labeling and fragmentation of PCR amplicons by biotinylated reagents in a one-step protocol 1o The optimized protocol with fragmentation and one-step labeling gives excellent results with different labeling reagents with the reactive diazomethyl function, as shown by the results shown in Table 5.

Example 14: Maraua ~ e and DNA fragmentation by BioDPDAM The DNA amplicons were prepared by PCR amplification according to the protocol described in Example 5.

The synthesis of the labeling reagent is described in Example 1.

The protocol is identical to that of Example 22.2, including denaturation at 95 ° C. before the hybridization step.

[0089]

IDT Exhibit 1020.10045

CA 02444926 2003-10-17 WO 02/090584 PCT / FR02 / 01542 55.

Results Marker used Homology I (rfu) B (rfu) I / B (%)

[0090]

BioDPDAM 93.0 32359 3610 9,1

[0091]

Table 6: Labeling and Fragmentation of DNA by BioDPDAM This result, described in Table 6, shows that substitutions as large as the phenyl group can be used to optimize the reactivity of the labeling reagents bearing a diazomethyl function.

Example 15 Maraua ~ e and DNA fragmentation by 5- (bromomethyl) fluorescein The DNA amplicons were prepared by PCR amplification according to the protocol described in Example 5.

At 10 ~, L of PCR are added 50 ~.

L of sodium formate pH3 (100 mM) and 2 ~. L of 5- (bromomethyl) fluorescein (Molecular probes, Eugen, OR) (100 mM in DMSO). The volume is adjusted to 100%. L. The solution is incubated for 30 min at 60 ° C. The purification conditions are in accordance with those of Example 9. A denaturation step is carried out as described in Example 12.2. The other hybridization and reading conditions are identical to those described in the article by A. Troesch et al. J. Clin. Microbiol., 37 (1), p 49-55 1999. Fluorescein is directly detectable on the reader.

[0092]

CA 02444926 2003-10-17 WO 02/090584 PCT / FR02 / 01542 56 Protocol used Homology T (rfu) B (rfu) I / B (%)

[0093]

Marking and fragmentation of

[0094]

PCR amplicons with 5- (bromo-> 95 855 183 4,7

[0095]

methyl) fluorescine

[0096]

Table 7: Labeling and fragmentation of DNA by 5- (bromornethyl) fluorescein_ This result of Table 7 shows that the fragmentation of DNA by the creation of abasic sites is completely compatible with a labeling reagent carrying a function reactive alkyl halide.

This protocol is carried out in one step (fragmentation and labeling) but with lower intensities than with the markers carrying a diazomethyl function. 10 Exemule 16: Maraua ~ e and DNA ampficons fragmentation in uresence of another fragmentation chemipue derived from phenanthroline DNA amplicons were prepared by PCR amplification according to the protocol described in Example 5.

Two types of conditions are used Conditions A To 10 ~ L of PCR are added 20 ~ L of phenanthroline-FeSO4 (2S mM) and 2 ~, L of meta-BioPMDAM (100 mM in DMSO).

The total volume is adjusted to 100 pl, L. The mixture is incubated 60 min at 95 ° C. Conditions b To 10 ~ L of PCR are added 50 ~ L of sodium formate buffer pH3 (100 mM in pure water) and 2 ~ L of meta-BioPMDAM (100 mM in DMSO).

The total volume is adjusted to 100 ~ L. The mixture is incubated 60 min at 95 ° C. The other conditions of the protocol are identical to those of Example 9.

[0097]

IDT Exhibit 1020.10046

CA 02444926 2003-10-17 WO 02/090584 PCT / FR02 / 01542 57 Homology conditions I (rfu) B (rfu) I / B (%)

[0098]

Phnanthroline FeS04 (5 mM)

[0099]

mta-bioPMDAM (2 mM)> 95 2236 500 4,5

[0100]

60 min 95C

[0101]

Formate buffer (SOmM) pH3

[0102]

mta-bioPMDAM (2 mM)> 95 6786 565 12,0

[0103]

60 min 95C

[0104]

Table 8: Labeling and Fragmentation of DNA Amplicons in the Presence of Phenanthroline The two fragmentation conditions allow a satisfactory result as demonstrated in Table 8.

The best result is obtained with conditions (b) using fragmentation at acidic pH.

Example 17 Fragmentation of the PCR amulicons labeled by incorporation of d-UTP-1o fluorescein Incorporation of the labeled nucleotide PCR amplification was carried out according to the following protocol in order to generate PCR amplicons labeled with fluorescein (labeling on the bases).

From target genomic DNA Mycobacte ~ ium tuberculosis 16S (10 + 4 copies) using the Fast Staf t kit from Roche, 0.2 mM of the deoxyribonucleotides d-ATP, d-CTP and d-GTP as well as 0.14 mM of d-TTP and 0.06 mM of d-UTP-12-fluorescein, 0.3 ~.

M of primers and 0.4 ~ L of enzyme.

The percentage of the labeled nucleotide relative to its natural d-UTP homolog is 30%.

It is this ratio which is generally used in reactions for labeling amplicons by incorporation of labeled nucleotides.

D-UTP-12-fluorescein is commercially available from Roche Diagnostics reference 1373242, Mannheim, Germany).

The parameters of the PCR are those of Example 5.

[0105]

CA 02444926 2003-10-17 WO 02/090584 PCT / FR02 / 01542 58 a.

Fr ~ a ~ mentation of PCR amplicons ma ~ e ~ ués 30r6 no UTP- ~ rescein A 10 ~ L of PCR are added 50 ~ L of sodium formate buffer pH 3 (50 mM).

The volume is adjusted to 100 ~, L. The solution is then incubated for 30 min at 60 ° C. b. Man ~ ua ~ e, ~ endant the f ~ 'a ~ mentation of PCR amplicons containing 30% by UTP- ~ luorescein A 10 ~. L of PCR are added 50 ~. L of sodium formate buffer pH 3 (50 mM) and 2 ~. L of meta-BioPMDAM (100 mM in DMSO). The volume is adjusted to 100 ~, L. The solution is

IDT Exhibit 1020.10047

then incubated 1o for 30 min at 60 ° C. This test corresponds to a reference protocol which makes it possible to compare the different labeling strategies while freeing themselves from the variability due to the amplification step. A column purification step and a denaturation step at 95 ° C. are carried out in all cases as in Example 9. Dal protocol The nucleic acids obtained by fragmentation of the amplicons labeled with d-UTP-fluorescein (conditions a)) are hybridized on a DNA chip and initially detected by direct reading of the fluorescent signals emitted by fluorescein as described in l example 15.

Protocol, a2 ~ A signal amplification step was used to improve labeling sensitivity. The signal was amplified by the introduction, during the hybridization step, of a biotinylated anti-fluorescein antibody (reference 216-065-0 ~ 4, Jackson ImmunoResearch) then of a streptavidin labeled with phycoerythrin using the following successive conditions ~ 300 p, L of pure water,

[0106]

CA 02444926 2003-10-17 WO 02/090584 PCT / FR02 / 01542 59 ~ 300 ~, L of 100 mM Tris buffer pH 7 / 1M NaCl / 0.05% tween / 0.005% defoamer, 2.4 ~.

L of BSA (50 mg / mL), 1.2 ~, L of biotinylated anti-fluorescein antibody. (1 mg 1 mL), ~ 300 qL of pure water, and ~ 300 ~, L of 100 mM Tris buffer pH 7 / 1M NaCl / 0.05% tween / 0.005% antifoam; 6 ~, L BSA (50 mg / mL); 6 ~. L of streptavidin-PE. (300 ~, g / mL). In this protocol, fluorescein acts as a hapten (tracer detectable indirectly by a labeled antibody) and not as a fluorophore. Protocol ~, b) 1o The biotinylated fragments (condition b) hybridized on a DNA chip are revealed by the introduction of a streptavidin labeled with la phycoerythrin using the following conditions: ~ 300 ~, L of pure water, and ~ 300 ~, L of 100 mM Tris buffer pH 7 / 1M NaCl / 0.05% tween / 0.005% defoamer; 6 ~. L of BSA (50 mg / mL); 6 ~ L of labeled streptavidin. (300 ~ g / mL). The reading of the fluorescence emitted on the surface of the DNA chip after labeling and hybridization as well as the generation of the data in terms of signal intensity and the percentage of homology are carried out by the reading systems and the software provided by 2o Affymetrix.

In this regard, it is important to note that the reading system used contains two filters making it possible to directly detect ~ fluorescein in the case where the amplicons are labeled with d-UTP-fluorescein only, according to the al protocol, or else ~ la phycoerythrin in the case where the amplicons are labeled - with d-UTP-fluorescein with signal amplification, according to protocol a2, or - by pieta-bioPMDAM during their fragmentation, according to protocol b. In the two cases where the revelation is carried out by phycoerythrin, the use of a filter makes it possible to dispense with the signal generated by the fluorescein and it is indeed the phycoerythrin signal which is detected.

[0107]

CA 02444926 2003-10-17 WO 02/090584 PCT / FR02 / 01542 Results Marker Homology

[0108]

Protocol used I B (rfu) I / B (rfu)

[0109]

detected (%)

[0110]

al: Fragmentation of PCR amplicons

[0111]

Flu * 81.6 595 342 1,7

[0112]

fluorescin-labeled UTP

[0113]

IDT Exhibit 1020.10048

a2.

Fragmentation of PCR amplicons

[0114]

d-labeled UTP fluorescin pg *> 95 221073461 6.4 with

[0115]

signal amplification

[0116]

b. Fragmentation and marking on the same

[0117]

PCR amplicons modified by d-UTP-pE *> 95 217001503 14,4

[0118]

fluorescin and labeled with mta-

[0119]

BioPMDAM 1 I

[0120]

* Flu = Fluorescein and PE = Phycoerythrin 5 Table 9: Fragmentation of labeled PCR amplicons by incorporation of d-UTP-fluorescein The results, from Table 9 above, show that chemical fragmentation using the creation of an abasic site is compatible with the enzymatic labeling of DNA amplicons and that labeling can take place before fragmentation.

The intensity levels as well as the percentage of homology obtained with this enzymatic fluorophore incorporation protocol are low compared to those obtained with labeling during fragmentation using the labeling reagent with a diazomethyl function such as meta-bioPMDAM (conditions (b)).

15 To reach the level of intensity obtained with the meta-bioPMDAM derivative, a signal amplification step is necessary (conditions a2).

This clearly shows the efficiency of the diazomethyl reactive function compared to the traditional incorporation of a modified base such as fluorescein d-UTP (reference protocol (b)).

[0121]

CA 02444926 2003-10-17 WO 02/090584 PCT / FR02 / 01542 61 Example 18: Fragmentation of double-stranded DNA by sonication DNA amplicons were obtained using the protocol described in Example 5.

These amplicons were fragmented by sonication in the presence and absence of the marker. at.

Marc, ~ ua eg PCR amplicons during sonication A 10 ~, L of PCR reaction are added 2 ~, L of meta-BioPMDAM (100 mM in DMSO).

The volume is adjusted to 100 ~ L with pure water, the pH is adjusted to 6.5.

The mixture is incubated for 30 minutes at 60 ° C. in a bath of an ultrasonic tank (frequency 35 kHz, model T460-H, Bioblock, France). b.

Marauding during the ~ chemical rementation of PCR amplicons (one-step reference protocol A 10 ~.

IDT Exhibit 1020.10049

L of PCR reaction are added 50 ~ L of sodium formate pH 3 (50 mM) and 2 ~, L of meta-BioPMDAM (100 mM in DMSO).

The volume is adjusted to 100 ~ L. The solution is incubated for 30 min at 60 ° C. The tests are carried out in duplicate in order to be able to analyze the fragmentation of the DNA on gel and the efficiency of labeling by hybridization and reading of the DNA chip as described above (example 9 on the detection of phycoerythrin).

Gel Analysis Analysis was performed on denaturing polyacrylamide gel (8% polyacrylamide, 7 M urea, 1 X TBE) using ethidium bromide staining.

Gel analysis shows that DNA amplicons are fragmented by sonication at 60 ° C.

[0122]

CA 02444926 2003-10-17 WO 02/090584 PCT / FR02 / 01542 62 Results Conditions Homology I (rfu) B (rfu) I / B (%)

[0123]

at. Marking during fragmentation

[0124]

93,~ 2271 631 3,6

[0125]

by sonication

[0126]

b. Marking during fragmentation

[0127]

>95 19639 1459 13,5

[0128]

chemical (reference conditions)

[0129]

Table I O: Fragmentation of double stranded DNA by sonication The labeling results of Table 10 during sonication (conditions a) are satisfactory.

This shows that the physical fragmentation by sonication of the DNA targets is compatible with the labeling chemistry by labeling reagents carrying a diazomethyl function. 1o The poor labeling results in this case are certainly due to the fact that the marker degrades under the effect of ultrasound.

The results with fragmentation by creation of an abasic site by the action of acidic pH are better.

Example 19: Marcrua ~ e, fragmentation and denaturation of DNA in one step The DNA amplicons were prepared by PCR amplification according to the protocol described in Example 5.

Two labeiing reactions were carried out a.

Labeling fra entation and denaturation at 95 ° C in a single step, ~ e A 10 ~ L of PCR are added 50 ~, L of sodium formate buffer pH3 (50 mM) and 2 p, L of the meta-BioPMDAM marker (100 mM in anhydrous DMSO).

The final volume is adjusted to 100 ~ L. The solution is then incubated for 30 min at 95 ° C. In this case, the reaction mixture was hybridized on a DNA chip without any prior purification.

IDT Exhibit 1020.10050

[0130]

CA 02444926 2003-10-17 WO 02/090584 PCT / FR02 / 01542 63 b.

Margin and f ~ mentatioh at 60 ° C at 10 ~.

L of PCR are added 50 ~ L of sodium formate buffer pH3 (50 mM) and 2 ~, L of the meta-BioDPDAM marker (100 mM in anhydrous DMSO).

The final volume is adjusted to 100 ~ L. The solution is then incubated for 30 min at 60 ° C. The reaction mixture was then purified according to the protocol described previously.

In this protocol and before hybridization on a DNA chip, the fragments were denatured according to the protocol described in Example 12.2.

Results Protocol used Homology I (rfu) B (rfu) I / B (%)

[0131]

at. Marking, fragmentation and

[0132]

>95 5426 506 10,7

[0133]

dhaturatiou 95C

[0134]

b. Masking and fragmentation

[0135]

>95 7015 818 6,6

[0136]

60Cpuis dnaturatiov ~ 95C

[0137]

Table 11: Labeling, Fragmentation and Denaturation of DNA in One Step Example 12.2 demonstrated the importance of denaturing double-stranded DNA for the sensitivity of detection.

These results of Table 11 show that with the fragmentation approach by creation of an abasic site, the labeling, the fragmentation and the denaturation of DNA can be carried out in a single step, which represents a notable improvement from the point of view of simplicity and time for the user without affecting the detection sensitivity.

[0138]

CA 02444926 2003-10-17 WO 02/090584 PCT / FR02 / 01542 64 Example 20: Synthesis of paya-Bio-EG3-PDAM O ~ ~ Ol MeOH ~ I CH (OMe) 3 / TMSCI ~ \ DowexWXB-100 OOH OOMe OOMe HzN-OHZ) 3-OO-CHz-CHz) 3-CHZ NHZ 4-carboxybenzaldehyde 24 25 ar I o ~ H2N ~ O ~ O ~ O ~ N 26 U 1) Biotin, CDI, in DMF purification O ~ 2) Deprotection: HCl 2N O HN- _NH HH ~ H NH ~.

O ~ O ~ O ~ N S o 28 O NH2-NH2, EtOH abs N ~ NHZ HN- 'NH H H H r H NH ~ O ~ O ~ O ~ N S,.

IDT Exhibit 1020.10051

U 29 O Mn02, DMF HN '-NH N2 H H H NH ~ O ~~ O ~ N S O 30 ~ Protection of 4-carboxylbenzaldehyde 4-carboxybenzaldehyde is commercial.

It is dissolved (3 g; 20 mmol) in a 1o solution of trimethylsilyl chloride (10 g; 92 mmol) in 100 mL of MeOH. The mixture is left under stirring for 40 h at room temperature.

After evaporation, a white solid

[0139]

CA 02444926 2003-10-17 WO 02/090584 PCT / FR02 / 01542 corresponding to 4-methoxycarbonylbenzaldehyde 24 is isolated, characterized by NMR and used as it is for the following reaction.

1H NMR (200MHz, CDCl3) 8 = 10.07 (s, 1H, -CHO); 8.17 (d, 2H, J = 8Hz, Ar-H2.6); 7.92 (d, 2H, J = 8Hz, Ar-H3.5); 3.93 (s, 3H, CO-O-CH3).

~ Protection of 4'-methox c ~ ylbenzaldehyde 4-methoxycarbonylbenzaldehyde (3.35 g; 20 mmol) is dissolved with 1o trimethyl orthoformate (4.8 g; 40 mmol) in the presence of Dowex SOWXB- 100 (1 g).

The mixture is heated at reflux for 2 h, then filtered and evaporated.

After a recrystallization test, an NMR analysis shows that the reaction is not complete and it is restarted in 30 mL of MeOH, 30 mL of CH (OMe) 3 and 1 g of Dowex SOWXB-100 at room temperature . Filtration is carried out followed by evaporation, in order to obtain 3.55 g (16.89 mmol, 84%) of product 25. 1H NMR (200MHz, CDCl3) b = 8.01 (d, 2H, J = 8Hz, Ar-H2.6); 7.50 (d, 2H, J = 8Hz, Ar-H3.5); 5.41 (s, 1H, CH); 3.93 (s, 3H, -CO-O-CH3); 3.29 (s, 6H, -O-CH3). 2o t Compound 26 Compound 25 (3.1 g; 14.8 mmol) is dissolved in 16 mL (73 mmol) of 4,7,10-Trioxa-1,13-tridecanediamine. The resulting solution is heated at 140-150 ° C for 2 h. The mixture is then dissolved in 100 mL of DCM (dichloromethane or CH2C12) and washed 6 times with 10 mL of water. The organic phase is dried with Mg S04 then evaporated until an oil is obtained. This oil is washed with pentane 3 times in a row by decantation, then a new extraction with DCM and H20 is carried out.

After drying over MgSO4 and evaporation, product 26 is isolated with a yield of 63% (9.27 mmol). 30 1H NMR (200MHz, CDCl3) ô = 7.78 (d, 2H, J = 8Hz, Ar-Hz, 6); 7.46 (d, 2H, J = 8Hz, Ar-H3.5); 5.39 (s, 1H, CH); 3.62-3.47 (m, 14H, H7 ~, 8 ~, 10 °, m and HS ~, i3 ~ and H3 ~); 3.29 (s, 6H, -O-CH3); 2.72 (m, 2H, His ~) I, 87 (m, 2H, H4>); 1.64 (m, 2H, H14); 1.30 (br s, 2H, NH2).

[0140]

CA 02444926 2003-10-17 WO 02/090584 PCT / FR02 / 01542 66 1 Biotinylated compound 27 Biotin (500 mg; 2.05 mmol) is suspended in 10 mL of DMF then 365 mg (2.25 mmol) CDIs are added.

This solution is left under stirring for 30 min at room temperature. Compound 26 (900 mg; 2.26 mmol) is dissolved in 1 mL of DMF, then added little by little to the previous solution. The mixture thus obtained is left under stirring for 1 h at room temperature. After evaporation, purification on a column by flash chromatography (20 mm diameter column) is carried out with 250 mL of 1o MeOH-DCM 6%, then with 200 mL of 7% MeOH-DCM and finally 200 mL of MeOH-DCM 8%. The fractions corresponding to product 27 are combined and then evaporated to dryness to give 1.00 g of oil with an estimated yield of 50%. 1H NMR (200MHz, CDCl3) ~ = 9.50 (br s, 1H, NH "niaa ~ goose); 7.80 (d, 2H, J = 8Hz, Ar- H2.6); 7.64 (s , 1H, H ~ "; d ~ ole); 7.46 (d, 2H, J = 8Hz, Ar-H3, s and 1H, NHZ -); 7.05 (s, 2H, Hi ~ a ~ ole); 6.76 (t, 1H, NHI6-), 6.20 (br s, 1H, NHB1); 5.44 (br s, 1H, NHB3); 5.37 (s, 1H, CH); 4.42 (m, 1H, HB6a); 4.24 (m, 1H, H ~ 3a); 3.59-3.44 (m, 14H, H7 ~, 8>, 10 °, m and Hs ~, 13 ~ and H3 ~); 3.29 (m, 8H, HIS ~ and 2-O-CH3); 3.07 (m, 1H, HB4); 2.84 and 2.66 (ABX system, 2H, 2J ~ = 5Hz, 3J, ~, X = 12Hz, 3JBx = 0Hz, HB6); 2.13 (t, 2H, J = 8Hz, HB10); 1.85 (m, 2H, H4 ~); 1.66 (m, 2H, H14>); 1.40-1.37 (m, 6H, HB7, Bg, B9).

~ Aldehyde compound 28 The acetal 27 is dissolved in 50 mL of chloroform, then 20 mL of 2N HCl are added. The two-phase mixture is vigorously stirred for 15 min. The organic phase is recovered and dried over anhydrous NaHCO3. It is filtered, evaporated and compound 28 is obtained in the form of a paste (495 mg; 0.855 mmol) with an overall yield of 42% from biotin. 30 1H NMR (300MHz, CDCl3) = 10.05 (s, 1H, CHO); 7.98 (d, 2H, J = 8Hz, Ar-H2.6); 7.92 (d, 2H, J = 8Hz, Ar-H3, s); 7.58 (t, 1H, NH2 -); 6.46 (t, 1H, NH16 3), 6.02 (br s, 1H, NHm);

[0141]

CA 02444926 2003-10-17 WO 02/090584 PCT / FR02 / 01542 67 5.19 (br s, 1H, NHB3); 4.46 (m, 1H, HB6a); 4.27 (m, 1H, HB3a); 3.66-3.56 (m, 10H, H7 ~, s °, 10 °, u> and HS ~); 3.50-3.29 (m, 4H, H3>, 13 °); 3.28 (m, 2H, H15 ~); 2.95 (m, 1H, HB4);

IDT Exhibit 1020.10052

2.84 and 2.71 (ABX system, 2H, 2J ~ = 5Hz, 3J ~ = 12Hz, 3JBx = 0Hz, HB6); 2.15 (t, 2H, J = 8Hz, HB10); 1.89 (m, 2H, H4 ~); 1.72-1.63 (m, 6H, H14>, HB7, B9); 1.23 (m, 2H, HBS).

~ Hydrazone compound 29 The aldehyde 28 (495 mg; 0.855 mmol) is dissolved in 10 mL of absolute ethanol. Hydrazine (350 p, L; 7.20 mmol) is added, then the reaction mixture is heated at reflux for 1 h. 1o The oil obtained after evaporation is dissolved in EtOH abs, to be evaporated again. A foam is then obtained which is triturated with pentane. The paste corresponding to product 29 (511 mg; 0.862 mmol) is obtained with a yield of 100%. 1H NMR (300MHz, CDCl3) 8 = 7.76 (d, 2H, J = 8Hz, Ar-H2.6); 7.72 (s, 1H, CH); 7.56 (d, 2H, J = 8Hz, Ar-H3.5); 7.34 (t, 1H, NHa>); 6.45 (t, 1H, NH16 -), 5.98 (br s, 1H, NHB1); 5.78 (br s, 2H, NH2); 5.18 (br s, 1H, NHB3); 4.44 (m, 1H, HB6a); 4.26 (m, 1H, HB3a); 3.62-3.56 (m, 10H, H ~~, s>, lo ~, m and HS>); 3.48-3.45 (m, 4H, H3>, 13 '); 3.27 (m, 2H, Hls.); 3.07 (m, 1H, HB4); 2.64 and 2.68 (ABX system, 2H, ZJ ~ = 5Hz, 3J ~ = 12Hz, 3JB ~ = 0Hz, HB6); 2.11 (t, 2H, J = 8Hz, HB10); 1.86 (m, 2H, H4 ~); 1.72-1.59 (m, 6H, H14>, HB7, B9); 1.21 (m, 2H, Has). t Diazo compound 30 Hydrazone 29 (357 mg; 0.602 mmol) is dissolved in 17.5 mL of DMF. MnO 2 (700 mg; 7.7 mmol) is then added.

After 12 min of stirring at room temperature, the mixture is filtered through Millipore containing velite (thickness: 2 cm) and molecular sieve powder 3 å (0.5 cm). The reaction mixture is evaporated to dryness. The residual oil obtained is washed with ether three times in a row. Compound 30 (290 mg, 0.491 mmol) is obtained in the form of a slightly pink solid with a yield of 82%.

[0142]

CA 02444926 2003-10-17 WO 02/090584 PCT / FR02 / 01542 68 1H NMR (300MHz, DMSO-d6) 8 = 8.28 (t, 1H, NH2>); 7.77 (d, 2H, J = 8Hz, Ar-H2.6); 7.74 (t, 1H, NH16 -); 7.00 (d, 2H, J = 8Hz, Ar-H3, s); 6.38 (br s, 1H, NHBi); 6.32 (br s, 1H, NH63); 5.80 (s, 1H, CH-N2); 4.27 (m, 1H, HB6a); 4.11 (m, 1H, Hssa); 3.51- 3.44 (m, 10H, H7>, s °, 10 °, n ° and Hs ~); 3.37 (m, 2H, H1s °); 3.32 (m, 4H, H3 ~, 13 '); 3.05 (m, 1H, HB4); 2.79 and 2.58 (ABX system, 2H, 2J ~ = 5Hz, 3J, ~ x = 12Hz, 3JBx = 0Hz, HB6); 2.02 (t, 2H, J = 8Hz, Halo); 1.69 (m, 2H, H4 ~); 1.59-1.48 (m, 6H, H14 ~, HB7, B9); 1.25 (m, 2H, HB8).

The reactivity of compound 30 was tested on 3'-Uridine-monophosphate and followed by capillary electrophoresis. The analysis conditions are those of Example 6.1. The 1o results show a half-life of 45 minutes. The reagent is stable at -20 ° C for at least 1 month. Example 21: Mara ~ ua ~ e and fragmentation of DNA amplicons with the reagent of marg-ua ~ e para-Bio-EG3-PDAM The main interests of this kind of molecules, that is to say derivatives of PDAM carrying a polyethylene glycol-based linker arm, are to allow the separation of the diazo function from biotin, and to increase the solubility and, ultimately, the reactivity of these 2o molecules. The paya-Bio-EG3-PDAM derivative was obtained according to the reaction scheme described in Example 20. The DNA amplicons were prepared by PCR amplification according to the protocol described in Example 5. Two labeling reactions were carried out. at. Labeling with the reagentpara-Bio-EG3-PDAM A 10 ~, L of PCR are added 10 ~.

L of para-Bio-EG3-PDAM (100 mM in DMSO) and 77 ~, L of water Dnase / Rnase free. The solution is homogeneous and does not present any precipitates. 3o This solution is incubated for 10 min at 95 ° C., then 3 wL of 0.1 M HCl are added and the solution is incubated for 10 min at 95 ° C.

[0143]

CA 02444926 2003-10-17 WO 02/090584 PCT / FR02 / 01542 69 The rest of the protocol is identical to that of example 8. b.

Labeling with the meta-BioPMDAM reagent A 10 p, L of PCR are added 10 p. L of meta-BioPMDAM (100 mM in DMSO) and 77 p. L of Dnase / Rnase free water. The synthesis of this product is mentioned in Example 1.1. The solution shows a slight precipitate. This solution is incubated for 10 min at 95 ° C. Then 3 ~ L of 0.1 M HCl are added and the solution is incubated for 10 min at 95 ° C. The rest of the protocol is identical to that of Example 8. l0 Results Protocol used Homology (%) I (rfu) B (rfu) I! S

[0144]

at. Reagent marking

[0145]

> 95% 15151 621 24,4

[0146]

IDT Exhibit 1020.10053

para-Bio-EG3-PDAM

[0147]

b. Reagent marking

[0148]

> 95% 11226 515 21,8

[0149]

mta-BioPMDAM

[0150]

Table 12: Comparative study of the labeling and fragmentation of DNA amplicons with lepara-Bio-EG3-PDAM and meta-BioPMDAM The signal intensities obtained in this table 12 are very satisfactory and the percentage of homology is high .

This result shows that the introduction of a polyethylene glycol arm on the diazo labeling molecule makes it possible to increase the solubility in the aqueous phase of the reagent.

The test is therefore homogeneous.

In addition, the increase in solubility makes it possible to increase the reactivity of the label.

Example 22 Synthesis of other PDAM derivatives comprising a linker arm based on polyethylene 1 ~ yçol

[0151]

CA 02444926 2003-10-17 WO 02/090584 PCT / FR02 / 01542 Example 22.1: Synthesis of meta-Bio-EG3-PMDAM Synthesis diagram OOO vo 0 O NH2 NH Succinic anhydride HO 3-aminoacetophenone O 68 CDI EG3 OO NH H2NI ~ O ~ O ~ O ~ N 69 CDI D-biotin OO NH HNI ~ O ~ O ~ O ~ NOO 7U H2N-NH2 Mn02 N2 O NH HN ~ O ~ O ~ O ~ NO meta-Bio-EG3-PMDAM 71

[0152]

CA 02444926 2003-10-17 WO 02/090584 PCT / FR02 / 01542 71 Compound 68 The 3-aminoacetophenone (14.5 g, 107 mmol) is dissolved in 50 ml of anhydrous DMF.

(Succinic anhydride (10.7 g, 107 mmol) is added and the mixture is left under stirring, under argon and at room temperature.

After 6 h, the solution is concentrated in vacuo and 50 ml of methanol are added.

The precipitate obtained is filtered off and washed with methanol and ether.

In this way 19.4 g (81%) of product 68 are obtained in the form of an off-white powder.

1H NMR (200 MHz, DMSO-d6): d = 2.5-2.6 (m, 7H); 7.45 (t, 1H); 7.64 (d, 1H); 7.83 (d, 1H); 10 8.19 (s, 1H); 10.16 (s, 1H); 12.12 (s, 1H). Compound 69 5.07 g (22 mmol) of compound 68 is dissolved in 10 ml of anhydrous DMF, under argon. Put on ice and add 5.00 g (32 mmol) of carbonyldümidazole. After 20 min, 20 ml (94.6 mmol) of 4, 7, 10-trioxatridecanediamine (EG3) are slowly added. After reacting for 3 hours at room temperature, the DMF is evaporated off and the residue is taken up in 100 ml of CH2CLa. Extractions were made with saturated NaHCO3 and H2O, after which the organic phase was dried with anhydrous Na2SO4 and the solvent evaporated. 4.34 g (46%) of product 69 are thus obtained. 2o 1H NMR (200 MHz, DMSO-d6): d = 1.59 (m, 2H); 1.87 (m, 2H); 2.16 (s, 3H); 2.40 (m, 2H); 2.55 (m, 2H); 3.08 (m, 2H); 3.45 (m, 16H); 7.30 (t, 1H); 7.42 (d, 1H); 7.70 (d, 1H); 7.83 (t, 1H); 7.97 (s, 1H); 10.00 (s, 1H). Biotinylated Compound 70 D-biotin (1.0 g, 4.1 mmol) is solubilized in 10 ml of anhydrous DMF, under argon.

Cool on ice and add the carbonyldümidazole (CDI) (0.665 g, 4.1 mmol) in 10 ml of anhydrous DMF. After 15 min, compound 69 (1.8 g, 4.1 mmol) in 2 ml of anhydrous DMF is added. The mixture is left to react for 3 h at 35 ° C., then the DMF is evaporated off and the residue is taken up in 100 ml of CH2Cl. Extractions were made with saturated NaHCO3 and H2O, after which the organic phase was dried with anhydrous Na2SO4 and the solvent evaporated. Characterization by

IDT Exhibit 1020.10054

[0153]

CA 02444926 2003-10-17 WO 02/090584 PCT / FR02 / 01542 72 NMR of the product thus obtained shows that a mixture of product 70 and free EG3 is obtained.

Another purification step is carried out before continuing the synthesis. The final compound, meta-Bio-EG3-PMDAM, is obtained after two synthesis steps according to the scheme described in Example 1. The interest of this synthesis is twofold. On the one hand, product 69 is obtained in just two steps; this product can be used as a precursor of diazo with the possibility of attaching detectable molecules of a different nature to it, thanks to the terminal amine group. This group also makes it possible to graft compound 69 onto solid supports, with (1o objective of immobilizing nucleic acids. On the other hand, compound 71 has the same reactive center as meta-Bio-PMDAM (our reference molecule), which facilitates (analysis of the advantages linked to (inclusion of the ethylene glycol arm (EG3). The reagent is stable at -20 ° C for at least 1 month. Example 22.2: Synthesis of meta-Bio-EG4-PMDAM Synthesis scheme

[0154]

CA 02444926 2003-10-17 WO 02/090584 PCT / FR02 / 01542 73 O ~ 1 / -j ° O ° O w Ethylene glycol, APTE '~ \ Hz ~' dC '~ \ Br Br Toluene ~ \ EtOH ~ ~ \ O NOZ, NOZ NHZ Et \ N 3-nitro-acetophenone 13 14 3 a ° Ol w ~ NH 1S O- (CHz-CHz-O-CHz-CHz-OH) z, NaH, THF r ~ 'B purification ° O ° HO Q ~ O ~ O ~ O ~ / O ~ NH

TsCl, pyridine purification ° I Ts " ° ~ p ~ 0 ~ ° ~ O ~ NH Phthalimide, DBU purification O ° r N ~ '~ O ~ / O ~ O ~ / O ~ NH' NHz-NHz EtOH abs r ° H2 ~ O ~ ° ~ O ~ ° ~ NH 1 1) Biotin, CDI, DMF purification O 2) Deprotection: HCl 2N O HN- _NH HHHO 'IN g ~ O ~ ° ~ O ~ ° ~ NH O 1) NHz-NHz EtOH abs O N ~ 2) MnOz, DMF HN '_NH H H H H ~ O N' ~ (g ~ O ~ ° ~ O ~ Ov 'NH U 23

[0155]

CA 02444926 2003-10-17 WO 02/090584 PCT / FR02 / 01542 74 t Protection of 3-vitro-aceto ~ henone 13 33 g (0.20 mol) of 3-vitro-acetophenone are dissolved in 400 mL of toluene , then 40 mL (0.717 mol) of ethylene glycol and 600 mg (3.15 mmol) of para-toluene sulfonic acid (APTS) are added.

We're building a Dean Stark system. The solution is heated for 3 h at 130 ° C. After allowing the solution to come to room temperature, 400 mL of ethyl acetate is added, then the latter is washed with 8 mL of saturated NaHCO 3 solution. The organic phase is dried over MgSO4. After evaporation, a pale yellow solid 13 (39.72 g; 0.190 mol) is obtained with a yield of 95%. 1H NMR (200MHz, CDCl3) 8 = 7.11 (t, 1H, J = 8Hz, Ar-H); 6.87-6.78 (m, 2H, Ar- H); 6.59 (dd, 1H, J = 6.5Hz, Ar-H); 4.00 (m, 2H, H Caceta ~ e); 3.79 (m, 2H, H acetate); 1.61 (s, 3H, CH3). t Preparation of Amine 14 Compound 13 (39.7 g; 0.190 mol) is dissolved in 500 mL of EtOH, then 1 g of 10% palladium on charcoal is added. It is heated to dissolve everything, then the solution is allowed to come to room temperature. After having created a vacuum and placed the solution under H2, the mixture is left under vigorous stirring for 5 h. The solution is then filtered hot and then evaporated.

The product 14 is washed with pentane, and isolated in the form of a solid (34 g; 0.189 mol) with a yield of 99%. 1H NMR (200MHz, CDCl3) 8 = 7.14 (t, 1H, J = 8Hz, Ar-H); 6.85 (m, 2H, J = 7.5Hz, Ar-H); 6.79 (s, 1H, Ar-H); 6.59 (dd, 1H, J = 6.5Hz, Ar-H); 4.00 (m, 2H, H Cacetal); 3.77 (m, ZH, H Cacetal); 1.61 (s, 3H, CH3). ~ Brominated compound 15 3o Amine 14 (12.3 g; 68.7 mmol) and triethylamine (7 g; 69 mmol) are dissolved in 150 mL of DCM under argon. A solution of 13.8 g (60 mmol) of bromoacetyl bromide dissolved in 150 mL of DCM is added dropwise at -5 ° C.. At the end of the addition,

[0156]

CA 02444926 2003-10-17 WO 02/090584 PCT / FR02 / 01542 100 mL of 1N aqueous NaHCO3 are added. The organic phase is washed twice in succession with aqueous NaHCO3 and dried over MgSO4.

After evaporation to dryness, 22.6 g of brown oil is obtained corresponding to compound 15 and used as it is for the following reaction. 5 1H NMR (200MHz, CDCl3) 8 = 8.29 (br s, 1H, NH); 7.62 (dt, 1H4, J = 5Hz, Ar-H); 7.47 (s, 1H, Ar-H2); 7.38-7.19 (m, 2H, Ar-H5.6); 4.00 (m, 2H, Br-CH); 3.75 (m, 4H, H C- H Kacetal); 1.61 (s, 3H, CH3). ~ Alcohol compound 16 1o Sodium hydride (3.3 g; 82.5 mmol) is washed three times with pentane and then suspended in 150 mL of THF, tetraethylene glycol (50 mL; 0.29 mol) is then added at room temperature. The reaction is allowed to stir for 15 min, then the solution is cooled to -5 ° C. The addition of compound 15 previously diluted in 25 mL of THF is done dropwise. It is left stirring for 30 min to allow it to return to room temperature. The solution is concentrated to 100 mL, then it is diluted with 500 mL of CHCl3. This organic phase is washed three times in a row with 250 mL of 1N aqueous NaHCO3, then dried over MgSO4 before evaporating.

IDT Exhibit 1020.10055

The product is purified on a silica column by flash chromatography (6S mm diameter column) with 1.5 L of 5% MeOH-DCM, then with 500 mL of 7% MeOH-DCM, and finally 500 mL of MeOH -DCM 10%. The fractions corresponding to compound 16 are combined, then evaporated to dryness to give 17.4 g (42.1 mmol) of product, with a yield of 61%. 25 1 H NMR (200MHz, CDCl3) 8 = 8.86 (broad s, 1H, NH); 7.71 (d, 1H, J = 7.5Hz, Ar-H4); 7.51 (s, 1H, Ar-HZ); 7.29-7.24 (m, 2H, Ar-H5,6); 4.09 (m, 2H, CO-CH -O); 3.99 (m, 2H, H Cacetal); 3.72-3.53 (m, 20H, O-CH -CH -O, H2Cacetal and HO-CH); 1.61 (s, 3H, CH3). ~ Compound tosylate 17 Alcohol 16 (4.13 g; 10.0 mmol) is dissolved in 5 mL of pyridine. 2.0 g (10.5 mmol) of tosyl chloride are then added at room temperature. We put under stirring under

[0157]

CA 02444926 2003-10-17 WO 02/090584 PCT / FR02 / 01542 76 argon for 10 h.

Dilute with 100 mL of DCM, wash the organic phase three times with 20 mL of 1N aqueous NaHCO3, then dry over MgSO4 before coevaporating with toluene. Purification on a column by flash chromatography (50 mm diameter column) is carried out with 500 mL of 2% MeOH-DCM, then 500 mL of 3% MeOH-DCM, and finally 500 mL of 4% MeOH-DCM. The fractions corresponding to product 17 are combined and then evaporated to dryness to give 3.68 g (6.48 mmol) of oil with an estimated yield of 65%. 1H NMR (300MHz, CDCl3) 8 = 8.86 (broad s, 1H, NH); 7.76 (d, 4H, J = 5.5Hz, Ar-Ht ° sy); 7.60 (d, 1H, J = 7.5Hz, Ar-H4); 7.50 (s, 1H, Ar-H2); 7.32-7.22 (m, 2H, Ar-H5,6); 4.10 (m, 2H, CO-CH -O); 4.00 (m, 2H, H Cacetal); 3.73-3.54 (m, 20H, O-CH -CH -O, H Cacetal and HO-CH); 2.42 (s, 3H, Ar-CH3); 1.61 (s, 3H, CH3). ~ Compound phthalimide 18 The tosylate 17 (3.68 g; 6.48 mmol) is dissolved with 1.52 g (10.0 mmol) of DBU (1,8-diazobicyclo- [5.4.0] -undecene) then the phthalimide (1.47 g; 10 mmol) is added.

The solution thus obtained is heated at 85-90 ° C for 17 h and then evaporated. The product is purified on a silica column by flash chromatography (50 mm diameter column) with 1 L of 15% acetone-DCM, then with 1 L of 20% acetone-DCM. The fractions corresponding to compound 18 are combined, then evaporated to dryness to give 3.15 g (5.8 mmol) of product, with a yield of 90%. 1H NMR (200MHz, CDCl3) 8 = 8.73 (broad s, 1H, NH); 7.79 (m, 2H, Ar-Hphta); 7.99 (m, 2H, Ax-Hphta and Ar-H4); 7.49 (s, 1H, Ar-HZ); 7.27-7.18 (m, 2H, Ar-H5,6); 4.10 (m, 2H, CO-CH - O); 4.00 (m, 2H, H Caceta ~ e); 3.69-3.56 (m, 20H, O-CH -CH -O, H Ca-étale and Nphta CH); 1.61 (s, 3H, CH3).

[0158]

CA 02444926 2003-10-17 WO 02/090584 PCT / FR02 / 01542 77 ~ Amine compound 19 The product 18 is dissolved in 20 mL of absolute EtOH while heating under reflux at 75-80 ° C. Hydrazine (1.07 mL; 22.1 mmol) is then added, and the reaction is left under stirring for 1 hour 15 minutes.

The precipitate obtained is filtered off on a frit and the ethanolic phase evaporated. The white precipitate is then washed with DCM, and the DCM phase is evaporated. The yellow oil obtained (2.3 g; 5.57 mmol) is used directly for the following reaction, even if it contains imidazole which can be eliminated subsequently, during the deprotection step of the i. 'acetal. 1H NMR (300MHz, CDCl3) 8 = 8.83 (br s, 1H, NH); 7.69 (d, 1H, J = 7.5Hz, Ar-H4); 7.51 (s, 1H, Ar-H2); 7.30-7.19 (m, 2H, Ar-H5 g); 4.10 (m, 2H, CO-CH -O); 4.00 (m, 2H, H Ca ~ etate); 3.69-3.56 (m, 20H, O-CH2-CH2-O, H2Ca-etate and HaN-CH2); 1.61 (s, 3H, CH3). Compound biotin ~ Lé 20 D-Biotin (1.05 g; 4.32 mmol) is solubilized in 10 mL of anhydrous DMF. 790 mg (4.87 mmol) of carbonyl dümidazole (CDT) are added under argon. After 10 min of stirring, the amine 19 diluted in 5 ml of DMF is added. The solution is left for 40 min with stirring, then evaporated before being purified on a column by flash chromatography. 2o For this we use a column 50 mm in diameter with as eluent 500 mL of MeOH-DCM 5%, then S00 mL of MeOH-DCM IO%, and finally 500 mL of MeOH-DCM 15%.

The fractions corresponding to product 20 are combined and then evaporated to dryness to give a yellow oil (1.66 g; 2.6 mmol). The yellow oil obtained (2.4 g) containing, according to the NMR spectrum, approximately 30% by weight of imidazole. It is therefore deduced that the yield of the reaction resulting in product 20 is about 60% relative to the starting biotin. 1H NMR (300MHz, CDCl3) 8 = 8.80 (broad s, 1H, NH); 7.66 (m, 3H, Ar-H4 and Himidazole) 7.54 (s, 1H, Ar-HZ); 7.28-7.24 (m, 2H, Ar-Hs, 6); 7.07 (s, 2H, H:, ~ dazole), 6.59 (t, 1H, NHIS>); 6.06 (br s, 1H, NHB1); 5.19 (br s, 1H, NH ~ 3); 4.45 (m, 1H, HBga); 4.27 (m, 1H, HB3a); 4.10 (s, 2H ,. H3>); 4.00 (m, 2H, H Ca ° etate); 3.75-3.49 (m, 18H, O-CH -CH -O and H Cacetate,);

[0159]

CA 02444926 2003-10-17 WO 02/090584 PCT / FR02 / 01542 78. 3.36 (m, 2H, H14>); 3.09 (m, 1H, HB4); 2.85 and 2.66 (ABX system, 2H, 2J ~ = 5Hz, 3J ~ = 12Hz, 3J $ x = 0Hz, H ~ 6); 2.16 (t, 2H, J = 8Hz, HB10); 1.61 (s, 3H, CH3); 1.59-1.3 (m, 6H, HB9, Ba, s7).

IDT Exhibit 1020.10056

~ Compound ketone 21 The acetal 20 is dissolved in 80 mL of chloroform, then 30 mL of 2N HCl are added. The mixture is left under vigorous stirring for 45 min. The organic phase is recovered and then dried over anhydrous NaHCO3. After filtration, the solution is evaporated and the oil obtained is washed with 10 pentane to give the product 21 (1.48 g; 2.48 mmol) with a yield of 99%. 1H NMR (300MHz, CDCl3) 8 = 8.99 (broad s, 1H, NH); 8.07 (s, 1H, Ar-H2); 7.98 (d, 2H, J = 8Hz, Ar-H4); 7.66 (d, 2H, J = 8Hz, Ar-H6); 7.42 (t, 2H, J = 8Hz, Ar-HS); 6.38 (t, 1H, NHIS.); 5.76 (br s, 1H, NHBI); 4.96 (br s, 1H, NHB3); 4.47 (m, 1H, H ~ 6a); 4.29 (m, 1H, HB3a); 4.13 (s, 2H, H3.); 3.76-3.37 (m, 16H, O-CH -CH -O); 3.32 (m, 2H, H14>); 3.11 (m, 1H, HB4); 2.89 and 2.75 (ABX system, 2H, 2J ~ = 5Hz, 3J ~ = 12Hz, 3JBx = 0Hz, HBS); 2.59 (s, 3H, CH3); 2.16 (t, 2H, J = 8Hz, HB10); 1.64-1.40 (m, 6H, H $ 9, ss, B7).

1

Hydrazone compound 22 Ketone 21 is dissolved in 20 mL of absolute EtOH. The mixture is heated to reflux at 75-80 ° C. Hydrazine (816 p. L; 16.81 mmol) and the mixture is left under stirring for 3 h. After filtration, evaporated to dryness, redissolved in ethanol until a sticky white foam is obtained. In a second step, this foam is dissolved in 50 mL of chloroform and then 20 mL of a saturated NaHCO 3 solution is added. It is washed well and then the organic phase is recovered. It is dried over anhydrous Na2CO3 and after filtration, evaporated to dryness in order to obtain a new sticky foam. This corresponds to product 22 (842 mg g; 1.38 mmol) and is obtained with a yield of 86%. 1H NMR (300MHz, CDCl3) 8 = 8.81 (br s, 1H, NH); 8.82 (s, 1H, Ar-H2); 7.64 (d, 2H, J = 8Hz, Ar-H4); 7.32 (m, 4H, Ar-H5.6); 6.43 (t, 1H, NH ~S>); 5.89 (br s, 1H, NHB1); 5.46 (s

[0160]

CA 02444926 2003-10-17 WO 02/090584 PCT / FR02 / 01542 79 br, 2H, NHa); 4.99 (br s, 1H, NHB3); 4.44 (m, 1H, HB6a); 4.27 (m, 1H, HB3a); 4.11 (s, 2H, H3.); 3.70-3.37 (m, 16H, O-CHz-CH -O); 3.32 (m, 2H, H14.); 3.08 (m, 1H, HB4); 2.87 and 2.67 (ABX system, 2H, ZJ ~ = 5Hz, 3JAx = 12Hz, 3JBx = 0Hz, HB6); 2.11 (m, SH, CH3 and HB10); 1.64-1.40 (m, 6H, HB9, Ba, B7).

~ Diazo compound 23 Hydrazone 22 (100 mg; 0.164 mmol) is dissolved in 1 mL of DMF under argon. 80 mg of activated MnO 2 are added and the mixture is left under vigorous stirring for 30 min. The mixture is filtered through a mixed Celite (3 cm) - powdered molecular sieve (1 cm) layer. The solution is then evaporated to dryness. The oil obtained at the end of evaporation is triturated until a pink powder corresponding to compound 23 (78 mg; 0.128 mmol; 78%) is obtained. 1H NMR (300MHz, DMSO-d6) = 9.60 (broad s, 1H, NH); 7.89 (s, 1H, Ar-HZ); 7.76 (t, 1H, NHIS ~); 7.35-7.25 (m, 4H, Ar-H5.6); 6.64 (d, 2H, J = 8Hz, Ar-H4); 6.36 (br s, 1H, NHsl); 6.32 (br s, 1H, NI-iB3); 4.28 (m, 1H, HB6a); 4.08 (m, 1H, HB3a); 4.06 (s, 2H, H3 ~); 3.55-3.31 (m, 16H, O-CH -CH -O); 3.17 (m, 2H, H14 ~); 3.08 (m, 1H, HB4); 2.80 and 2.59 (ABX system, 2H, 2J ~ = 5Hz, 3J, ~ x = 12Hz, 3JBx = 0Hz, HB6); 2.13 (m, 5H, CH3); 2.13 (t, 2H, J = 8Hz, HB10); 1.99-1.30 (m, 6H, HB9, Bs, s7).

The reactivity of compound 23 was tested on 3'-LTridine-monophosphate and followed by capillary electrophoresis. The analysis conditions are those of Example 6.1. The results show a half-life of 30 minutes. Reagent stability is greater than 1 month at -20 ° C. Example 22.3: Synthesis of para-Bio-EG3-PMDAM Synthesis scheme

[0161]

CA 02444926 2003-10-17 WO 02/090584 PCT / FR02 / 01542 OO \ O \ O MeOH ~ I CH (OMe) 3 / \> w> I TMSCI DowexWX8-100 ~ \ HZN- (CHZ) 3- (O -CHz-CHZ) 3-CHZ-NHZ COOH COOMe COOM \ e 4-Acetylbenzoic acid 32 a! L_o H2N ~ O ~ O ~ O ~ N \ I 33 O 1) Biotin, CDI, in DMF purification O 2) Deprotection: HCl 2N HN ~~ ~ NH OH ~ H ~. . r ~~ ~ 2 ~ 2 ~ EtOH abs ~~ NH2 r NH ~, O ~ O ~ O ~ N \ 0 36 O Mn02, DMF HN ~ NH N2 HH NH ~ O ~ O ~ O ~ N \ ISO 0 37 ~ Protection of 4-Acetylbenzoic Acid 4-Acetylbenzoic acid (1 g; 6.1 mmol) is dissolved in a solution of trimethylsilyl chloride (TMSCI, 10 g; 92 mmol) in 5 mL of MeOH. The mixture is heated to 90 ° C overnight.

After evaporation, a white solid corresponding to compound 31 (1.21 g; 5.75 mmol) is isolated, characterized by NMR and used as it is for the following reaction.

[0162]

CA 02444926 2003-10-17 WO 02/090584 PCT / FR02 / 01542 81 1H NMR (200MHz, CDCl3) ~ = 8.08 (d, 2H, J = 8Hz, Ar-H2.6); 7.59 (d, 2H, J = 8Hz, Ar-H3.5); 3.18 (s, 6H, -O-CH3); 1.53 (s, 3H, CH3).

~ Compound 32 Compound 31 (1.21 g; 5.75 mmol) is dissolved in 5 ml of trimethyl orthoformate in the presence of Dowex SOWX8-100 (0.3 g). The mixture is heated to 60 ° C overnight, then filtered and evaporated to give compound 32 (1.19 g; 5.3 mmol) with a yield of 87%. 1H NMR (200MHz, CDCl3) 8 = 8.00 (d, 2H, J = 8Hz, Ar-H2.6); 7.54 (d, 2H, J = 8Hz, Ar-1o H3.5);

IDT Exhibit 1020.10057

3.89 (s, 1H, CO-O-CH3); 3.16 (s, 6H, -O-CH3); 1.51 (s, 3H, CH3). Compound 33 Compound 32 (1.17 g; 5.22 mmol) is dissolved in 5 mL (22.7 mmol) of 4,7,10-trioxa-1s 1,13-tridecanediamine. The solution obtained is heated at 140 ° C. for 4 h. The mixture is then dissolved in 30 mL of DCM and washed 3 times with 10 mL of water. The organic phase is dried with Mg 504, then evaporated until an oil corresponding to product 33 (1.44 g; 3.49 mmol) is obtained with a yield of 67%. 20 1H NMR (200MHz, CDCl3) 8 = 7.76 (d, 2H, J = 8Hz, Ar-H2.6); 7.51 (d, 2H, J = 8Hz, Ar-H3.5); 3.62-3.47 (m, 14H, H7 ~, 8 ~, io ~, m and H5 ~, 13 ~ and H3 ~); 3.15 (s, 6H, -O-CH3); 2.73 (m, 2H, His>) 1.88 (m, 2H, H4 ~); I. 65 (m, 2H, H14 ~); I, 38 (br s, 2H, NH2).

~ Biotinylated compound 34 Biotin (780 mg; 3.19 mmol) is suspended in 13 mL of DMF. Then 590 mg (3.60 mmol) of CDI are added. This solution is left under stirring for 30 min at room temperature. Compound 33 is dissolved in 1 mL of DMF, then added little by little to the previous solution. The mixture thus obtained is left under stirring for 1 h at room temperature. After evaporation of the DMF, purification on a column by flash chromatography (35 mm diameter column) is carried out with 500 mL of MeOH-

[0163]

CA 02444926 2003-10-17 WO 02/090584 PCT / FR02 / 01542 82 DCM 6%, then with 250 mL of 8% MeOH-DCM, and finally 250 mL of 8% MeOH-DCM.

The fractions corresponding to product 34 are combined, then evaporated to dryness to give 1.05 g of oil with an estimated yield of 30%. 1H NMR (300MHz, CDCl3) = = 8.49 (br s, 1H, NH;, ~,; daZOl); 7.79 (d, 2H, J = 8Hz, Ar-H2.6); 7.66 (s, 1H, H; n, ~ d ~ ole); 7.50 (d, 2H, J = 8Hz, Ar-H3.5); 7.38 (t, 1H, NH2 -); 7.11 (s, 2H, H; ",; a ~ ole); 6.67 (t, 1H, NH16>), 5.99 (br s, 1H, NHB1); 5.15 (br s, 1H , NH83); 4.46 (m, 1H, HB6a); 4.27 (m, 1H, H $ 3a); 3.61-3.45 (m, 14H, H ~~, s ~, io ~, II and HS ~, i3 ~ and H3 ~); 3.28 (m, 2H, His ~); 3.15 (s, 6H, -OCH3); 2.85 (m, 1H, HB4); 2.85 and 2.69 (ABX system, 2H, 2J, 8.8 = 5Hz, 10 3JAx = 12Hz, 3JBx = 0Hz, HB6); 2.14 (t, 2H, J = 8Hz, HBIO); 1 , 86 (m, 2H, H4>); 1.69 (m, 2H, Hr4 ~); 1.49 (s, 3H, CH3); 1.42-1.39 (m, 6H, HB7, ss, B9). ~ Compound 35 The acetal 34 is dissolved in 45 mL of chloroform, then 10 mL of 2N HCl are added. The two-phase mixture is vigorously stirred for 5 min. The organic phase is recovered and dried over anhydrous NaHCO3.

Filtered, evaporated, and compound 35 is obtained as a light yellow solid (504 mg; 0.87 mmol) with an overall yield of 27% from biotin. 1H NMR (300MHz, CDCl3) 8 = 7.97 (d, 2H, J = 8Hz, Ar-H2.6); 7.91 (d, 2H, J = 8Hz, Ar-H3.5); 7.51 (t, 1H, NHa -); 6.50 (t, 1H, NH16), 6.05 (br s, 1H, NHBI); 5.23 (br s, 1H, NH83); 4.45 (m, 1H, HB6a); 4.27 (m, 1H, HB3a); 3.62-3.56 (m, 10H, H ~~ a °, 10 °, u> and HS>); 3.48- 3.46 (m, 4H, H3 °, 13 '); 3.27 (m, 2H, His ~); 3.10 (m, 1H, HB4); 2.85 and 2.71 (ABX system, 2H, 2J, e, 8 = 5Hz, 3Jfx = 12Hz, 3JBx = 0Hz, HB6); 2.60 (s, 3H, CH3); 2.14 (t, 2H, J = 8Hz, HB10); 1.89 (m, 2H, H4 ~); 1.72-1.61 (m, 6H, H14., HB7 B9); 1.40 (m, 2H, HBS). ~ Hydrazone compound 36 3o The ketone 35 (500 mg; 0.864 mmol) is dissolved in 11 mL of absolute EtOH. Hydrazine (335 ~ L; 6.911 mmol) is added, then the reaction mixture is heated at reflux for 1

[0164]

CA 02444926 2003-10-17 WO 02/090584 PCT / FR02 / 01542 83 h.

The oil obtained after evaporation is dissolved in EtOH abs, to be evaporated again. A sticky foam corresponding to product 36 (488 mg; 0.823 mmol) is then obtained with a yield of 95%. S 1H NMR (300MHz, CDCl3) 8 = 7.76 (d, 2H, J = 8Hz, Ar-H2.6); 7.67 (d, 2H, J = 8Hz, Ar-H3, s); 7.29 (t, 1H, NH2 -); 6.46 (t, 1H, NH16 3), 5.98 (br s, 1H, NHB1); 5.55 (br s, 2H, NHa); 5.14 (br s, 1H, NHB3); 4.45 (m, 1H, HB6a); 4.24 (m, 1H, HB3a); 3.62-3.51 (m, 10H, H7 °, s °, 10 °, n ° and Hs>); 3.47-3.45 (m, 4H, H3>, 13 '); 3.27 (m, 2H, HIS ~); 3.07 (m, 1H, HB4); 2.84 and 2.69 (ABX system, 2H, ZJ, ~ = 5 Hz, 3J ~ °x = 12 Hz, 3JBx = 0 Hz, HB6); 2.11 (t, 2H, J = 8 Hz, 1o HBIO and s, 3H, CH3); 1.86 (m, 2H, Hø ~); 1.72-1.59 (m, 6H, H14., HB7, B9); 1.21 (m, 2H, HB8). ~ Diazo compound 37 Hydrazone 36 (200 mg; 0.337 mmol) is dissolved in 5 mL of DMF. Then 15 MnO2 (450 mg; 5.17 mmol) is added. After 15 minutes of stirring at room temperature, the mixture is filtered through a millipore containing celite (thickness: 2 cm) and molecular sieve powder 3.

E ~ (0.5 cm). The reaction mixture is evaporated to dryness. The residual oil obtained is washed with ether three times in a row, until a powder is obtained. Compound 37 (290 mg, 0.491 mmol) is obtained in the form of a pink solid with a yield of 93%. 1H NMR (300MHz, DMSO-d6) b = 8.33 (t, 1H, NH2 ~); 7.83 (d, 2H, J = 8Hz, Ar-H2.6); 7.73 (t, 1H, NH16>); 6.98 (d, 2H, J = 8Hz, Ar-H3, s); 6.39 (br s, 1H, NHB1); 6.33 (br s, 1H, NHB3); 4.30 (m, 1H, HB6a); 4.12 (m, 1H, HBSa); 3.51-3.45 (m, 16H, H7 ~, 8>, 1o., 11 and Hs ~ and His ° and H3 ~, 13 ~); 3.07 (m, 1H, HB4); 2.79 and 2.58 (ABX system, 2H, 2J ~ = 5Hz, 3JAx = 12Hz, 3JBx = 0Hz, HB6); 2.14 (s, 3H, CH3); 2.04 (t, 2H, J = 8Hz, HB10); 1.77 (m, 2H, H4>); 1.62-1.48 (m, 6H, H14 °, HB7, s9); 1.31 (m, 2H, HB8). The reactivity of compound 37 was tested on 3'-Uridine-monophosphate and followed by capillary electrophoresis. The analysis conditions are those of Example 6.1.

The results show a half-life of 60 minutes. The reagent is stable at -20 ° C for at least 1 month.

IDT Exhibit 1020.10058

[0165]

CA 02444926 2003-10-17 WO 02/090584 PCT / FR02 / 01542 84 Example 22.4: Synthesis of ~ neta-Bio-EG3-PDAM Synthesis scheme O ~ \ O \ O / CH (OMe) 3 COOMe OowexWX8-100 ~ I g2N_ (CHZ) 3- (O-CHZ-CHZ) 3-CHZ-NHz \ COOM 'Methyl-3-formylbenzoate 39 38 HI) Biotin, CDI, in DMF 2) Deprotection: HCl 2N O HN ~ NH, HH _ _ NH ~ O ~ O ~~ NSOO 42 NHa-NH2, EtOF I H 43 Mn02, DMF HN NH N2 ~ / IHH NH ~ O ~ O ~ O ~ NSOO 44 ~ Protection of methyl-3-formylbenzoate 38

[0166]

CA 02444926 2003-10-17 WO 02/090584 PCT / FR02 / 01542 The Dowex SOWX8-100 resin (2 g) is dissolved in 25 mL of MeOH and 25 mL of trimethyl orthoformate and then stirred for IS min .

After decantation, the resin is washed twice in succession with 20 ml of MeOH. This resin is then placed in 100 mL of MeOH, 50 mL of CH (OMe) 3 and 7.12 g (43.4 mmol) of methyl-3-formylbenzoate are added.

The solution is left stirring for 15 min, then filtered through pleated paper before evaporation. Product 39 (9 g; 43.1 mmol) is isolated in the form of a light yellow liquid with a yield of 99%. 1H NMR (200MHz, CDCl3) 8 = 8.10 (s, H, Ar-HZ); 7.9 (d, H, J = 8Hz, Ar-H4); 7.63 (d, H, J 10 = 8Hz, Ar-H6); 7.42 (t, H, J = 8Hz, Ar-HS); 5.40 (s, 1H, CH); 3.90 (s, 3H, -CO-O-CH3); 3.31 (s, 6H, -O-CH3). ~ Compound 40 15 Compound 39 (2 g; 9.5 mmol) is solubilized in 10.4 mL (47.6 mmol) of 4,7,10-Trioxa-1,13-tridecanediamine. The solution obtained is heated at 165 ° C. for 2 h. The mixture is then dissolved in 80 mL of DCM and washed 4 times with 20 mL of water. After drying over MgSO4 and evaporation, the product 40 is isolated with a yield of 60% (2.27 g; 5.69 mmol). 1H NMR (200MHz, CDCl3) 8 = 7.84 (s, H, Ar-H2); 7.75 (d, H, J = 8Hz, Ar-H4); 7.53 (d, H, J = 8Hz, Ar-H6); 7.39 (t, H, J = 8Hz, Ar-HS); 5.38 (s, 1H, CH); 3.64-3.43 (m, 14H, H7 °, a °, to °, m and HS ~, i3 ~ and H3 ~); 3.29 (s, 6H, -O-CH3); 2.72 (m, 2H, Hls ~) 1.87 (m, 2H, H4 ~); 1.64 (m, 2H, Hl4 ~); 1.30 (br s, 2H, NH2).

~ Biotin compound ~ Lé 41 D-biotin (344 mg; 1.40 mmol) is suspended in 4 mL of DMF and then 250 mg (1.54 mmol) of CDI are added. This solution is left under stirring for 30 min at room temperature. Compound 40 (616 mg; 1.54 mmol) is dissolved in 2 mL of DMF, then added little by little to the previous solution. The mixture thus obtained is left under

[0167]

CA 02444926 2003-10-17 WO 02/090584 PCT / FR02 / 01542 86 stirring for 50 min at room temperature.

After evaporation, purification on a column by flash chromatography (column 30 mm in diameter) is carried out with 750 ml of 10% MeOH-DCM, then with 20 ml of 1S% MeOH-DCM. The fractions corresponding to product 41 are combined and then evaporated to dryness to give 740 mg of oil with an estimated yield of SO%. 1H NMR (200MHz, CDCl3) 8 = 7.87 (s, H, Ar-Ha); 7.76 (d, H, J = 8Hz, Ar-H4); 7.65 (s, 1H, Himictaz ° 1e); 7.53 (d, H, J = 8Hz, Ar-H6); 7.39 (t, H, J = 8Hz, Ar-HS); 7.07 (s, 2H, H; i ,,; daZ ° 1e) 6.65 (t, 1H, NH16 ~), S, 9S (br s, 1H, NHBl); 5.38 (s, 1H, CH); 5.15 (br s, 1H, NHB3); 10 4.43 (m, 1H, HB6a); 4.27 (m, 1H, HB3a); 3.59-3.44 (m, 14H, H7 ~, ~~, lo ~, II and H5 ~, 13 ~ and H3>); 3.29 (m, 6H, His ~ and 2-O-CH3); 3.07 (m, 1H, HB4); 2.84 and 2.66 (ABX system, 2H, 2J ~ = 5Hz, 3J ~ = 12Hz, 3JBX = 0Hz, HB6); 2.13 (t, 2H, J = 8Hz, HB10); 1.85 (m, 2H, H4 ~); 1.66 (m, 2H, H14 ~); 1.40-1.37 (m, 6H, HB7, Bs, B9). t Aldehyde compound 42 The acetal 41 is dissolved in 20 mL of chloroform, then S mL of 2N HCl are added.

The two-phase mixture is vigorously stirred for 15 min. The organic phase is recovered and dried over anhydrous NaHCO3. It is filtered, evaporated and compound 42 is obtained in the form of a yellow oil (S93 mg; 1.02 mmol) with an overall yield of 87% from biotin. 1H NMR (300MHz, CDCl3) S = 10.04 (S, 1H, CHO); 8.34 (s, H, Ar-H2); 8.16 (d, H, J = 8Hz, Ar-H4); 7.96 (d, H, J = 8Hz, Ar-H6); 7.72 (t, 1H, NH2 -); 7.39 (t, H, J = 8Hz, Ar-HS); 6.51 (t, 1H, NH16>); 6.00 (br s, 1H, NHBt); 5.30 (br s, 1H, NHB3); 4.46 (m, 1H, Ha6a); 4.27 (m, 1H, HB3a); 3.66-3.56 (m, 10H, H7 ~, 8 ~, io ~, 1 and H5 ~); 3.50-3.29 (m, 4H, H3 ~, 13 '); 3.28 (m, 2H, Hls.); 2.95 (m, 1H, HB4); 2.84 and 2.71 (ABX system, 2H, 2J, ~ = SHz, 3J, ~ = 12Hz, 3JBx = 0Hz, HB6); 2.15 (t, 2H, J = 8Hz, HB10); 1.69 (m, 2H, H4 ~); 1.72-1.63 (m, 6H, Hla °, H87, B9); 1.23 (m, 2H, HB8). 1 Cornuose hydrazone 43

[0168]

CA 02444926 2003-10-17 WO 02/090584 PCT / FR02 / 01542 87 The aldehyde 42 (593 mg; 1.02 mmol) is dissolved in 10 mL of absolute ethanol.

Hydrazine (400 p. L; 8.19 mmol) is added, then the reaction mixture is heated at reflux for 50 min. The yellow oil obtained after evaporation is triturated with ether until a beige powder corresponding to product 43 (404 mg; 0.68 mmol) is obtained with a

IDT Exhibit 1020.10059

yield of 66%. Purification on a column by flash chromatography (15 mm diameter column) is then carried out on a sample of 150 mg (0.253 mmol) with 200 mL of 20% MeOH-DCM. The fractions are combined and then evaporated to dryness to give 144 mg of product 43 with a yield of 76%. 1H NMR (300MHz, CDCl3) = 7.95 (s, H, Ar-H2); 8.16 (d, H, J = 8Hz, Ar-H ~); 7.76 (s, 1H, CH); 7.96 (d, H, J = 8Hz, Ar-H6); 7.38 (t, H, J = 8Hz, Ar-HS); 6.45 (t, 1H, NH16 ~); 5.98 (br s, 1H, NHB1); 5.72 (br s, 2H, NH2); 5.18 (br s, 1H, NHB3); 4.44 (m, 1H, HBSa); 4.26 (m, 1H, HB3a); 3.62-3.56 (m, 10H, H7>, s ~, io ~, m and HS>); 3.48-3.45 (m, 4H, H3 ~, 13 '); 3.27 (m, 2H, His ~); 3.07 (m, 1H, Hue); 2.84 and 2.68 (ABX system, 2H, 2J, ~ = 5Hz, 3J ~ = 12Hz, 3JBx = 0Hz, HB6); 2.11 (t, 2H, J = 8Hz, HB10); 1.86 (m, 2H, H4 ~); 1.72-1.59 (m, 6H, H14 °, Hs7, s9); 1.21 (m, 2H, HB8). t Diazo compound 44 Hydrazone 43 (100 mg; 0.187 mmol) is dissolved in 4 mL of DMF. MnO 2 (200 mg; 2.3 mmol) is then added.

After 13 min of stirring at room temperature, the mixture is filtered through a millipore containing velite (thickness: 2 cm) and 3 ä molecular sieve powder (0.5 cxn). The reaction mixture is evaporated to dryness. The residual oil obtained is washed with ether three times in a row. Compound 44 (290 mg, 0.491 mmol) is obtained in the form of an orange solid with a yield of 83%. 1H NMR (300MHz, DMSO-d6) 8 = 8.39 (t, 1H, NH2>); 7.78 (t, 1H, NH16>); 7.39-7.34 (m, Ar- H); 7.09 (d, Ar-H); 6.38 (br s, 1H, NHB1); 6.32 (br s, 1H, NHB3); 5.78 (s, 1H, CH-Na); 4.27 (m, 1H, HB6a); 4.11 (m, 1H, HB3a); 3.51-3.44 (m, 10H, H7 ~, 8 ~, io ~, 11 and H5 ~); 3.37 (m, 2H, His ~); 3.32 (m, 4H, H3 ~, 13 '); 3.05 (m, 1H, HB4); 2.79 and 2.58 (ABX system, 2H,

[0169]

CA 02444926 2003-10-17 WO 02/090584 PCT / FR02 / 01542 88 2J, s, 8 = 5 Hz, 3J, ~ = 12 Hz, 3JBx = 0 Hz, HBS); 2.02 (t, 2H, J = 8Hz, HB10); 1.69 (m, 2H, H4.); 1.59-1.48 (m, 6H, H14 ~, HB7, B9); 1.25 (m, 2H, HB8).

The stabilliy of the product is greater than 1 month at -20 ° C. Example 23: Synthesis of para-Cy5-EG3-PDAM As has already been mentioned in Example 2, the biotin can be replaced by another marker such as CyS. This example shows that it is also possible to link the diazo function, 1o carried by PDAM, to this Cy5 label by the intermediary of a polyethylene glycol linker. Synthetic scheme: The counterion I- is not represented in the formulas 46, 47, 50 'and 51. J N + Ac20 46 '+ AcÖ ~ / ~ N' " ~ 0 4'7 Ph PhHN ~ NHPh Ac20 I \ / / d 120 ° C ~ + 20 min 50 'I ~ OAc AcOH 0 120 ° C ~ ri ~ \ / V _ 4s '~~ ~ \ / + _ ~ OH ~' N ~ / / ~ N OH B ~ + gp O ~ g1 OH 49 O ~ 2- [4- (N acetyl-N henylamino) buta-1 iodide , 3-dien ~] -1,2,3,3-tetramethyl [3H] 'indolium 47

[0170]

CA 02444926 2003-10-17 WO 02/090584 PCT / FR02 / 01542 89 The mixture of malonaldehyde-bis (phenylimine) monohydrochloride 45 (13 g; 50.2 mmol), of NaOAc (6.0 g; 69.7 mmol) and (1,2,3,3-tetramethyl [3H] indolium iodide 46 (3.01 g; mmol) in acetic anhydride (50 mL) is heated at 100 ° C for precisely 20 minutes.

After cooling, (ether (350 mL) is added and the precipitated brown solid is filtered off, washed with (ether (3 x 100 mL). The solid is redissolved in 150 mL of CH2Cl2, filtered (elimination, of inorganic salts) then precipitated with 350 mL of ether to give a brown solid (3.54 g, 54%). 1H NMR (CDCl3): 8 = 8.64 (d; 1H; J = 12Hz; 1-H); 8.14 (t; 1H; J = 16; 12Hz; 3-H); 1o 7.63-7.19 (m; 9H); 6.90 (d; 1H; J = 15Hz; 4-H); 5.82 (t; 1H; J = 12; 13 Hz; 2-H); 4.06 (s; 3H; NCH3); (2.16 (s; 3H; -COCH3); 1.74 (s; 6H; CH3). T 1 ~ 5-carboxypentyl) -2,3,3-trimethyl [3Hlindolium bromide 50 Mix the 2,3 , 3-trimethylindole 48 (10.0 g; 62.8 mmol) and 6-bromohexanoic acid 49 (12.3 g; 62.8 mmol) without solvent and heat at 110 ° C for 12 h under argon. The red-violet paste reaction mixture is washed with (ethyl acetate (2 x 60 mL, the paste is triturated with the spatula and the supernatant is decanted), then with acetone (50 mL, the paste solidifies).

The pink solid is filtered off and then dried under vacuum (16.0 g; 73%). ~ Compound C ~ SCOOH 51 The mixture of (Iodide 47 (2.5 g; 5.3 mmol), bromide 50 (1.87 g; 5.3 mmol) and NaOAc (1.08 g; 12.1 mmol) in (acetic anhydride (11 mL) is heated at 120 ° C for 25 min. After cooling, ether (200 mL) is added and the precipitate is filtered off and washed with ether (3 x 50 mL). The solid corresponding to product 50 ' is dissolved in 100 mL of CHaCIz and then evaporated. It is then dissolved in 15 mL of acetic acid and stirred for 30 min at 120 ° C. The precipitate corresponding to product 51 is thus obtained after addition of 200 ml of ether and filtration through a frit, with a yield of 84% (2.71 g; 4.4430 mmol).

[0171]

CA 02444926 2003-10-17 WO 02/090584 PCT / FR02 / 01542 1H NMR (CDCl3): 8 = 8.03 (t; 2H; J = 10; 11 Hz, 2-H, 4-H); 7.38-6.91 (m; 9H; Ar-H, 3-H); 6.41 (d; 1H; J = 14Hz; 1-H); 6.31 (d; 1H; J = 13Hz; 5-H); 4.07 (t; 2H, J = 7; 7Hz; a-CHZ); 3.68 (s; 3H; NCH3); 2.47 (t; 2H, J = 7; 7Hz; s-CH2); 1.71 (m; 18H; CH3, (3, y and 8-CH2). Bbbb 145 ° C -1- HzN ~~ NHz> 4, 7, 10-trioxa-1, 13- tridecanediamine N ~ p ~~ / O ~ O ~ NH2 H 4-methylformylbenzoate 26 dimethyl acetal bb 1) iBuOCOCl, NMM, DCM, ta, 51 2) purification ~ o i_ OH ~ ~ O ~ p ~ / ~ H ~ CS CY 52 1) exchange resin d 'Cl- ions ~ ~~ ~ o ~~ 2) Deprotection

IDT Exhibit 1020.10060

HCl 2 N, DCM N'V1'O N"CCy H ~% 3 H + 53 ~~ NH ~ -NH2, EtOH abs z ta I 0 H ~~ H ~ GSCy 54 N2 Mn02, DMF 0 or H ~~ H ~~ ~ ~ Couplag ~, e of compound 26 with CySCOOH 51 (product 52)

[0172]

CA 02444926 2003-10-17 WO 02/090584 PCT / FR02 / 01542 91 Å a solution of CySCOOH 51 (1.5 g; 2.46 mmol) in 15 mL of CH2C12, N methylmorpholine (NMM, 405 p, L; 3.68 mmol).

The solution is cooled with an ice bath and placed under argon, then isobutyl chloroformate (494 ~, L; 3.81 mmol) is added.

After 10 min of stirring, the amine 26 (1.86 mole; 4.67 mole) diluted in 8 mL of CH 2 Cl 2 is added. The mixture is left under stirring at room temperature for 1 hour 30 minutes. 20 mL of CH2Cl2 is added and washed with 25 mL of NaHCO3 (1N) three times in a row. After drying over Na2CO3, the solution is filtered to recover the dichloromethane phase which is evaporated. 1o A purification on a column by flash chromatography (column 45 mm in diameter, fractions of 20 mL)) is carried out with 10% MeOH-DCM as eluent. The fractions corresponding to product 52 are combined and then evaporated to dryness to give a blue solid which is dissolved in CH2Cl2. The product 52 is then precipitated and washed with ether to give a blue product with a yield of 72% (I, 45 g; 1.77 mmol). The product 52 (iodide) is then dissolved in 54 ml of methanol and then passed through a column of amberlite IRA900 (Cl-; 15 g). The methanolic solution collected is evaporated to dryness to give a sticky oil which is redissolved in CH2C12.

Evaporation allows the product 52 'to be obtained with a yield of 87%. ~ Aldeh d The acetal 52 'is dissolved in 10 mL of DCM, then 10 mL of 2N HCL are added. The solution is left with vigorous stirring for 3 h 30 min. After adding 20 mL of DCM, the dichloromethane phase is recovered and then dried over NaHCO3. The product obtained after evaporation is washed with ether to give aldehyde 53 with a yield of 90 (1.188; 1.46 mmol). ~ Hydrazone 54 Aldehyde 53 (200 mg; 0.247 mmol) is dissolved in 1 mL of absolute ethanol and thydrazine monohydrate (15.6 ~ L; 0.321 mmol) is added. The solution is stirred at temperature

[0173]

CA 02444926 2003-10-17 WO 02/090584 PCT / FR02 / 01542 92 ambient for 30 min.

8 ml of ether are added; washed with ether by decantation, three times in a row and then dried under vacuum. 172 mg of hydrazine 54 (0.209 mmol; 85% yield) are obtained, which are stored in the freezer. ~ Diazo 55 Å a solution of 20 mg (0.243 mmol) of hydrazone 54 in 2 mL of DMF, 100 mg of MnO2 are added and stirred vigorously for 5 min under argon at room temperature. The suspension is filtered through a layer of celite (thickness: 2 cm) and molecular sieve in powder 3 1 ~ (0.5 cm) and washed with DMF. The solution is evaporated and then triturated with ether. The solid thus obtained is dried. 18 mg (0.185 mmol; 76%) of diazo 55 are obtained. Reagent stability is greater than 1 month at -20 ° C. Example 24: Synthesis of ~ rzeta-Fluo-EG3-PMDAM As has already been mentioned in Example 23, the biotin can be replaced by another marker. This example shows that it is also possible to link the diazo function, carried by PMDAM, to this fluorescein label by means of a polyethylene glycol linker arm.

[0174]

CA 02444926 2003-10-17 WO 02/090584 PCT / FR02 / 01542 93 OO ~, / ~ ~ O ~ ~ NH HZN OON 69 ° 0 so / O ~ ~ NH NH ~ HN OON n II O 72 H2N-NH- Tosyl N2 7S KOH / EtOH HO SO NH NH ~ HN ~ O ~ O ~ O ~ N Ö 'mets-Fluo-EG3-PMDAM 74 Compound 72 Fluorescein- isothiocyanate (250 mg, 0.64 mmol) is dissolved in 1, 6 ml of anhydrous DMF, with 2% pyridine, under argon.

Product 69 (0.356 g, 0.81 rrunol) dissolved in 1.6 ml of anhydrous DMF is added. The mixture is left to react for 3.5 h at room temperature, then the DMF is evaporated off and the residue is taken up in 25 ml of H20. Three extractions are made with 50 ml of 1o CHaCl2, and the aqueous phase is evaporated. 255 mg (48%) of product 72 are obtained. Synthesis diagram

[0175]

CA 02444926 2003-10-17 WO 02/090584 PCT / FR02 / 01542 94 Meta-Fluo-EG3-Hydrazone-Tosyle compound 75 Compound 72 (255 mg, 0.31 mmol) is dissolved in 1.5 ml of ethanol at reflux.

P-toluen-sulfonyl-hydrazine (69.2 mg, 0.37 mmol) in 1.5 ml of ethanol is added and the mixture is left to react for 6 h. Evaporated to dryness, and the solid washed with CH2Cl2, H2O and ether. 18.5 mg (74%) of product 75 are obtained in the form of an orange powder. 1H NMR (200 MHz, DMSO-d6): d = 1.6-1.8 (m, 4H); 2.13 (s, 1H); 2.28 (s, 1H); 2.36 (s, 1H); 2.80 (m,

IDT Exhibit 1020.10061

1H); 3.07 (m, 2H); 3.46 (m, 12H); 6.5-6.7 (m, 6H); 7.1-8.3 (m, 9H). Çom ~ osé meta-Fluo-EG3-PMDAM 74 Fhydrazone 75 (176 mg, 0.18 mmol) is dissolved in 720 ~, 1 of a 10% KOH solution in anhydrous methanol. The solution is left at reflux for 3 h. The solution is allowed to cool and a precipitate appears. The solution is filtered and evaporated to dryness. The residue is washed with ether and dried. NMR analysis shows the disappearance of the signals at 2.36 and 2.13 ppm (corresponding to the methyls of tosyl and of hydrazone) and (appearance of a peak at 1.96 ppm (corresponding to the methyl of diazo). Example 25: Diazomethyl intermediate allowing a subsequent maraua ~ e It may be advantageous not to have a direct labeling with the diazomethyl labeling reagent also carrying the RZ label but to proceed in two steps with an indirect labeling.

In this case, the labeling reagent comprising the diazomethyl function is said to be prefunctionalized, ie it also comprises a chemical function capable of subsequently reacting with direct or indirect labels. The pre-functionalization can take place by introducing a covalent reactive function within the labeling reagent which can react with an anti-covalently reactive function of the direct or indirect label. These functions can be constituted by an electrophilic organic chemical function 3o and a nucleophilic organic chemical function, or vice versa.

[0176]

CA 02444926 2003-10-17 WO 02/090584 PCT / FR02 / 01542 An example of such a labeling strategy is illustrated by the diagram below R1 RI N2-C \ + Wa_Ra ~ N2-Co Z_Wi ~ Z_Wi-W2 -Ra In which the labeling reagent comprises, in addition to a diazomethyl function, an electrophilic or nucleophilic Wi function capable of reacting with an R2 label comprising a W2 function complementary to Wi. For example if W1 is a methyl ketone or aldehyde function, W ~ is an alkoxyamine function.

In a method of labeling a biological molecule such as a nucleic acid, the nucleic acid is brought into contact with the labeling reagent coprenant the diazomethyl function and 1o in a subsequent step, the marker W2- R2 reacts with the nucleic acid via the Wi function. One of the uses is for example constituted by a method for amplifying a sequence of a nucleic acid or by a method for amplifying the signal. Additional information on this type of marking can be found in patent application WO-A-98/05766, under priority of August 2, 1996. and in patent application WO-A-00/40590 under priority of OS January 1999. Example 25.1: Synthesis of MeCO-PMDAM Synthesis scheme

[0177]

CA 02444926 2003-10-17 WO 02/090584 PCT / FR02 / 01542 96 OO NH HzN ~ O ~ O ~ O ~ 69 0 0 Fmoc-Cl 0 NH FmocHN ~ 0 ~ 0 ~ 0 ~ O $ I NNH-Tos H2N -NH-Tos 0 NH FmocHN ~ O ~ O ~ O '~ $ ~ O NH-Tos piperidine 0 NH H2N ~ O ~ O ~ O ~ $ 3 O NH- Tos OO Rome o ~ 00 0 NH ~ HN ~ O' ~ O ~ O ~ $ Q. O N2 MnO ~ o o 0 ~ ~ NH ~ HN ~ O ~ O ~ O ~ O MeCO-PMDAM $ 5

[0178]

CA 02444926 2003-10-17 WO 02/090584 PCT / FR02 / 01542 97 The product 85, the synthesis of which is described in this example, makes it possible to carry out the labeling of natural nucleic acids, thanks to. the reactivity of the diazomethyl function with the phosphate groups, and thus to introduce a methyl ketone function, which can be used subsequently to introduce a detectable molecule (fluorescent, biotin) having an alkoxyamine group.

This synthesis is based on known methods and routinely used in chemistry.

The starting product is the same as for the synthesis of markers 71 and 74. The first step consists in the protection of terminal starvation by Fluorenylmethyl-formate (Fmoc, 99). The choice of this protective group is based on its stability and cleavage conditions. 0 99 / \ After formation of fhydrazone protected 82 by the method used previously (example meta-Fluo-EG3-PMDAM,), terminal starvation is deprotected under mild basic conditions which guarantee the stability of fhydrazone. Methyl acetoacetate is used to create the methyl ketone function by a terminal starvation acylation reaction (see formation of compounds 26 and 36). The formation of the diazomethyl is then carried out by means of the methods described above. Example 25.2: Synthesis of HZNO-PMDAM Synthesis scheme

[0179]

CA 02444926 2003-10-17 WO 02/090584 PCT / FR02 / 01542 9 ~ OO NH 69 H2N ~~ o ~ ° ~ o ~ 0 Fmoc-Cl \ 81 0 NH FmocHN ~ O ~ 'O ~ O ~ U NH- Tos N ° 82 FmocHN " ~ O ~ NH-Tos ri piperidine 83 0 NH HzN ~ 0 ~ 0 ~ 0 ~ NH-Tos ri NMM COOH 'BuClC00 F ~ c-HN 0 0 86 FmoaHN0.

IDT Exhibit 1020.10062

J ~ O NH ~ H ~ O ~ ~ O ~ I Nz Mn02 0 0 NH g7 FmocHNO ~ H ~~ O ~~ Nz piperidine o 0 NH H2N ~ H ~ 0 ~ 0 ~ 0 ~ H2N0-PMDAM 88

[0180]

CA 02444926 2003-10-17 WO 02/090584 PCT / FR02 / 01542 99 The product 88, the synthesis of which is described in this example, makes it possible to carry out the labeling of natural nucleic acids, thanks to the reactivity of the diazomethyl function with the phosphate groups, and thus introduce an alkoxyamine function, which can be used subsequently to introduce a detectable molecule (fluorescent, biotin) having a methyl ketone group.

This synthesis is based on the model of the previous one, that is to say the use of precursor 69, of Fmoc for the protection of starvation and of tosyl for the protection of thydrazone. 1o The introduction of the alkoxyamine function (compound 86) is done by (use of carboxymethoxylamine (commercial) protected by the Fmoc function (Thesis E. Trévisiol, LEDSS Grenoble, 1999). Given the mild conditions of the final deprotection (compound 86), this is carried out immediately after the formation of diazomethyl. Example 26: Preparation of PDAM derivatives allowing signal amplification Example 26.1: Synthesis of bis-biotinylated markers such as [Bio-EG3] 2-PDAM 2o Synthesis diagram

[0181]

CA 02444926 2003-10-17 WO 02/090584 PCT / FR02 / 01542 100 COOMe H LiBHa ~ ~ MnOz ~ Me00 ~ COOMe I ~ e Me00 I ~ COOMe Me00 ~ ~ COOM 56 57 58 MeOH / TMSCI

[0182]

M OMe Me0 OMe KOH!

[0183]

HOO I ~ COOH MepO I ~ COOMe

[0184]

60 59

[0185]

H N ~ NHz CFaCOOEt ~ HzN ~~~~ Fa ~ 7 " ~ ~ ~ / J3 3 p 61 62 = D-Biotin, CDI, Oll HNnNH ~~~ Fa I liS O NHs aq. conc.

O O II 63 HN ~ NH ~~~ NHz p 3 64 a HCI ~ S ~ 1.

NH2NHz 66 2.

MnOz 67

[0186]

CA 02444926 2003-10-17 WO 02/090584 PCT / FR02 / 01542 law ~ Reduction of trimethyl benzene-1,3,5-tricarboxylate 56 to alcohol 57 The triester 56 is dissolved (12.6 g; 50.0 mmol) in 100 mL of THF then add 1.1 g (50.5 mmol) of LiBH4 at room temperature.

The red solution is heated to 40-45 ° C with stirring under argon for 1 h.

After cooling (ice), the excess hydride is carefully destroyed (evolution of H2) by adding water (200 mL) then 2N HCl (30 mL).

We observe the color change to light yellow. This solution is extracted with CH2Cl2 (100mL then 3 times 50 mL), the organic phase is washed with anhydrous NaHCO3, dried over MgSO4 then the solvent is evaporated until an oil is obtained (11.1 g). By flash chromatography i0 on a silica column (diameter = 40 mm, eluent: ethyl acetate / cyclohexane = 1/1), the alcohol 57 (6.38 g, 57%) is obtained. 1H NMR (200 MHz, CDCl3): 8 = 8.53 (t, 1H, J = 2Hz); 8.18 (d, 2H, J = 2Hz); 4.76 (s, 2H); 3.91 (s, 6H); 2.30 (s, 1H). 1s ~ Oxidation of Alcohol 57 to Aldehyde 58 (Alcohol 57 (5.86 g; 26.1 mmol) is dissolved in 100 mL of THF and then 40.0 g of MnOa is added little by little over 5 min at room temperature. The solution is left overnight with stirring 2o under argon.

IDT Exhibit 1020.10063

The solution is filtered through a Büchner funnel fitted with a layer of Celite 545, washed with CH2Cl2 and then the solvents evaporated. The crude solid (4.4 g) is purified by flash chromatography on a silica column (diameter = 50 mm, eluent: ethyl acetate / cyclohexane 3/7).

3.44 g (59%) of (aldehyde 58 are obtained. 25 1H NMR (200 MHz, CDCl3): = 10.11 (s, 1H); 8.89 (t, 1H, J = 1Hz); 8.59 (d, 2H, J = 1Hz); 3.98 (s, 6H). ~ Formation of acetal 59 3o (aldehyde 58 (3.21 g; 14.4 mmol) is dissolved in 30 mL of methanol and then 6.0 mL of TMSCl is added. The solution is left at room temperature with stirring under argon for 1 hr.

[0187]

CA 02444926 2003-10-17 WO 02/090584 PCT / FR02 / 01542 102 The solution is diluted with 200 mL of CH2C12, and stirred with 1 M NaHCO3 (100 mL) (attention release of CO2).

The two phases are separated, the aqueous phase extracted three times with CH2Cl2 (25 mL), the organic phases are combined, dried over MgSO4 then the solvent is evaporated off. 3.55 g (92%) of (acetal 59. 1H NMR (200 MHz, CDCl3): ~ = 8.63 (t, 1H, J = 2Hz); 8.29 (d, 2H, J = 2Hz); 5.45 (s, 2H); 3.93 (s, 6H); 3.32 (s, 6H). t Hydro ~ se of the diester 59 in diacid 60 1o The diester 59 (3.18 g; 11.9 mmol) is dissolved in 10 mL of THF then a solution of l ~ OH (2.0 g, 85% pellet) is added in 10 mL of methanol. After 15 min at room temperature, the solvents are evaporated off. The residue is dissolved in H2O (50 mL). H3PO4 (about 2.5 mL, 85%) is added until pH 3 and the white precipitate is filtered through the frit (# 3), washed with water and dried in vacuo. 2.59 g (91%) of the diacid 60 are obtained. 1H NMR (200 MHz, DMSO-db): ~ = 8.43 (t, 1H, J = 1Hz); 8.15 (d, 2H, J = 1Hz); 5.53 (s, 1H); 3.27 (s, 6H). ~ Trifluoroacetamide 62 The diamine 61 (66 g; 0.30 mol) is dissolved in 250 mL of CHaCl2 then the ethyl trifluoroacetate (11.8 mL, 0.10 mol) is added dropwise over 5 min at 10 ° C. with stirring under argon.

After 15 min at room temperature, the solution is transferred to a separatory funnel, washed with H2O (3 x 100 mL), dried over MgSO4 and the solvent evaporated. 22.4 g (71%) of monoamide 62 with a purity of about 85% (determined by F19 NMR) are obtained. This compound is stored at -20 ° C and used without purification. 1H NMR (200 MHz, CDCl3): b = 3.5-3.6 (m, 12H); 3.42 (t, 2H, J = 6Hz); 2.75 (t, 2H, J = 6Hz); 1.81 (quantiplet, 2H, J = 6Hz); 1.67 (quantiplet, 2H, J = 6Hz); 1.30 (br s, 2H). 30 1gF NMR (190 MHz, CDCl3): 8 = -76.3.

[0188]

CA 02444926 2003-10-17 WO 02/090584 PCT / FR02 / 01542 103 ~ Compound 63 To a suspension of D-biotin (6.39 g; 26.2 mmol) in 50 mL of DMF, carbonyldiimidazole (CDI) is added , 6.32 g, 80%, 31.2 mmol).

The mixture is heated to 55-60 ° C with stirring under argon for 30 min. One initially observes a complete dissolution of material then a solidification with a precipitation of white solid (COa evolution). Starvation (oil) is added to (using 5 mL of CHZCl2 to rinse and the mixture is heated to 55-60 ° C for 3 h. The DMF is evaporated in vacuo (<1 mmHg) and the residue is stirred with 10 CH2Clz (700 mL) and 2N HCl (100 mL). After filtration of the two phases through a layer of Celite 545, the phases are separated, the aqueous phase is extracted with CHaCla (15 x 100 mL), the organic phases are combined, dried over anhydrous NaHCO3 and MgSO4, then the solvent is obtained. evaporated. The oily residue is triturated with 150 mL of ether to obtain a suspension. The pasty solid is difficult to filter. The supernatant is decanted and the ether wash is repeated. After drying under vacuum, 9.13 g (64%) of compound 63 are obtained. 1H NMR (200 MHz, CDCl3): = 3.5-3.6 (m, 12H); 3.42 (t, 2H, J = 6Hz); 2.75 (t, 2H, J = 6Hz); 1.8l (quantiplet, 2H, J = 6l ~ z); 1.67 (quantiplet, 2H, J = 6Hz); 1.30 (br s, 2H).

1 Compound 64 A solution of compound 63 in ammonia (100 mL, 22% aqueous) is heated in a 250 mL flask stoppered with a septum at 55-60 ° C for 2 h. After cooling, the solvent is evaporated off. The residue is dissolved in methanol (20 mL) and passed through a Dowex 21K anion exchange resin column [height 12 cm ~ e diameter 35 mm, OH- form obtained by prior washing with 1N NaOH (1.5 L) then H2O (1.5 L) then methanol (1.5 L)]. Compound 64 free of trifluoroacetate flon passes into the first fractions with 200 mL of methanol. After evaporation, the residue is triturated with 50 ml of ether and then decanted. The ether wash is repeated five times in a row. After drying, compound 64 is obtained (6.43 g, 86%).

[0189]

CA 02444926 2003-10-17 WO 02/090584 PCT / FR02 / 01542 104 1H NMR (300 MHz, CDCl3): b = 6.77 (t, 1H, J = 4 Hz); 6.32 (s, 1H); 5.52 (s, 1H): 4.45 (m, 1H); 4.28 (m, 1H), 3.50-3.68 (m, 12H); 3.30 (m, 2H), 3.11 (m, 1H); 2.86 (dd, 1H, J = 13 & 5Hz), 2.75 (t, 2H, J = 13Hz), 2.68 (d, 1H, J = 13Hz); 2.16 (t, 2H, J = 7Hz); 1.60-1.85 (m, 8H); 1.41 (m, 2H).

1 Bio-EGO- acetal 65 To a suspension of the diacid 60 (120 mg; O, S00 mmol) in dichloroethane (5 mL), add the carbonyldiimidazole (225 mg, 90%, 1.25 mmol) and heat to SS -60 ° C for 10 30 min with stirring under argon. Starvation 64

IDT Exhibit 1020.10064

(SSO mg; 1.23 mmol) is added and the solution is heated at 55-60 ° C for 6 h. After evaporation, one passes through a silica column (diameter: 25 mm, eluent: methanol 1 S-30% in CH2Cl2). 413 mg (75%) of compound 65 are obtained. 1H NMR (300 MHz, CDCl3): = 8.34 (s, 1H); 8.06 (s, 2H); 7.87 (m, 2H); 6.85 (m, 2H); 6.60 (s, 2H); 5.93 (s, 2H): 5.40 (s, 1H); 4.45 (m, 2H); 4.27 (m, 2H), 3.43- 3.68 (m, 24H); 3.31 (s, 6H); 3.25 (m, 4H), 3.08 (m, 2H); 2.83 (dd, 2H, J = 13 & 5Hz), 2.70 (t, 2H, J = 13Hz); 2.13 (t, 4H, J = 7Hz); 1.89 (quintuplet, 4H, J = 7Hz); 1.55-1.70 (m, 12H); 1.37 (m, 4H). Bio-EG3 ~ - aldehyde 66 Acetal 65 (413 mg; 0.376 mmol) in solution in methanol is treated with 2N HCl (0.5 mL). After evaporation and washing with (ether, aldehyde 66 is obtained (0.37 g, 90%).

1H NMR (300 MHz, DMSO-d6): ~ = 10.11 (s, 1H); 8.82 (t, 2H, J = 6Hz); 8.62 (s, 1H); 8.47 (s, 2H); 7.73 (t, 2H, J = 5Hz); 4.30 (m, 2H); 4.11 (m, 2H), 3.30-3.60 (m, 24H); 3.06 (m, 6H); 2.60 (dd, 2H, J = 12 & 5Hz), 2.56 (t, 2H, J = 12Hz); 2.03 (t, 4H, J = 7Hz); 1.78 (quintuplet, 4H, J = 7Hz); 1.35-1.60 (m, 12H); 1.28 (m, 4H). ~ Bio-EG3] ~ - PDAM 67

[0190]

CA 02444926 2003-10-17 WO 02/090584 PCT / FR02 / 01542 105 The aldehyde 66 is converted into diazomethane 67 according to the method used for the preparation of diazomethanes (Example 1).

Reagent stability is greater than 1 month at -20 ° C. Example 26.2: Synthesis of meta-Bio7-EG3-PMDAM Synthesis scheme

[0191]

CA 02444926 2003-10-17 WO 02/090584 PCT / FR02 / 01542 106 OO 0 O ~ " ~ CDI 0 HO ~ O ~ O ~ D ~ OH HOOC-EGO - IC H2N ~ 3, 6, 9-trioxa -1,11- HN 3-aminoacetophenone undecanedioic acid D- (NH2) 8 CDI O 0 N-NH2 HAN-NH2 EtOH reflux ~ D-Biotin ~ 0 O --EG ~ NH-EG3-D-Biot ~ NH-EG3 -D-Biot ~ H ~ NH 79 ~ 8 ~~ D- (NH2h Mn02 r 'meta-Bio7-EG3-PMD.AM 80

[0192]

CA 02444926 2003-10-17 WO 02/090584 PCT / FR02 / 01542 107 EG3-acetophenone compound 76 3, 6, 9-trioxa-1, 11-undecanedioic acid (EG3, 12.64 ml, 74 mmol) is dissolved in 80 ml of anhydrous DMF under argon and cooled in an ice bath.

The dicyclohexylcaxbodiimide (DCC, 11.45 g, 55.5 mmol)) is then dissolved in 20 ml of anhydrous DMF and added slowly.

After 30 min, 3-aminoacetophenone (5.0 g, 37 mmol) is added and the mixture is left to react for 1 h at room temperature under argon. The DMF is then evaporated in vacuo and 70 ml of CH2CL ~ are added. The solution is filtered and extracted with 3x25 ml of 1% acetic acid. The aqueous phases are combined and washed with 25 ml of CH2Cl ~. The organic phases are mixed 1o, dried with anhydrous sodium sulfate and evaporated to dryness. The product is recrystaillized from the MeOH: H2O couple. 8.74 g (70%) of product 76 are thus obtained. 1H NMR (200 MHz, DMSO-d6): d = 2.55 (s, 3H); 3.5-3.7 (m, 8H); 4.0 (s, 2H); 4.1 (s, 2H); 7.45 (t, 1H); 7.65 (d, 1H); 7.90 (d, 2H); 8.2 (s, 1H); 9.8 (s, 1H). Compound (NHa_) ~ -EG3-acetophenone 77 The product 76 (120 mg, 0.35 mmol) is dissolved in 15 ml of anhydrous DMF under argon, cooled on ice, and the DCC (110 mg, 0.53 mmol) is then added. After 30 min, this solution is added 2o to a solution of the commercial dendrimer "Starburst PAMAM Dendrimer, Generation 1" (Aldrich, St Quentin Fallavier) (1 g, 0.71 mmol, in 5 ml of methanol), slowly and under strong restlessness.

Allowed to react for 1 h at room temperature and evaporated. The residue is taken up in 10 ml of CHZCl ~ and extracted twice with 30 ml of 1% acetic acid. Compound Biot ~ -EG3- acetophenone 78 D-Biotin (1.73 g, 7.08 mmol) is dissolved in 80 ml of anhydrous DMF under argon, and the solution is cooled on ice. N-methylmorpholine (NMM, 856 p1, 7.7 mmol) and isobutyl chloroformate (1022 ~, 1, 7.7 mmol) are successively added. After 30 min, the 3o product 77 (1.13 g, 0.7 mmol, in 5 ml of methanol) is added and the mixture is left to react for 3 h on ice and under argon. Concentrate in vacuo to 50 ml and add 100 ml of CHZCl2. A precipitate

[0193]

CA 02444926 2003-10-17 WO 02/090584 PCT / FR02 / 01542 108 form, which is filtered, washed with (ether, and dried under vacuum.

1.3 g of 78 are obtained in the form of a white powder. Compound Biot7-EG; -Hydrazone 79 s Compound 78 (300 mg, 0.09 mmol) is dissolved in 10 ml of absolute ethanol at reflux. Hydrazine monohydrate (20 ml, 0.40 mmol) is added and the mixture is left to react for 3 h at reflux. After cooling, a precipitate forms, which is filtered, washed with (ether and dried in vacuo. 109 mg (36%) of product 79 are thus obtained in the form of a white powder. i0 Compound Biot ~ -EG3-PMDAM 80 Fhydrazone 79 (100 mg, 0.03 mmol) is dissolved in 5 ml of anhydrous DMF at 70 ° C. Allowed to warm to room temperature and the MnO ~ (31 mg,

IDT Exhibit 1020.10065

0.36 mmol) is added. The mixture is left to react for 10 min and the manganese oxide is removed by filtration through a fruct with celite (0.5 cm and powdered molecular sieve (0.5 cm). The filtrate is evaporated to dryness, washed with (ether and dried under vacuum. 78 mg (78%) of product 80 are thus obtained. Dendrimers are tree molecules having at the ends several 2o reactive groups such as amines, carboxyis, hydroxyis or others (for review, see Newcome et al, (1996) Dehdritie Molecules: G'oncept, Syntheses, Perspectives.

VCH Ed., Weinheim, Germany). The synthesis of these molecules is now well controlled, and many dendrimers are marketed by (chemical industry. The choice of PAMAM (Sigma-Aldrich) was made on the basis of its stability, solubility and flexibility, since several versions of this molecule, with different number and type of terminations, are available. The "PAMAM Generation 1" makes it possible to add seven molecules of the marker (in a single synthesis step) for each diazomethyl group.

[0194]

CA 02444926 2003-10-17 WO 02/090584 PCT / FR02 / 01542 109 Exemulates 27 'Maraua ~ e and fragmentation in two vices of DNA amlicons with meta- BioPMDAM The DNA axnplicons were prepared by PCR amplification according to the protocol described in example 5.

Two labeling reactions were carried out. at. Marking and fra ~ nentation in two steps A 10 ~. L of PCR are added 10 ~, L of meta-BioPMDAM (100 mM in DMSO) and 77 ~ L 1o of water Dnase / Rnase free. The solution is incubated for 10 min at 95 ° C. Then 3 ~, L of 0.1M HCl are added and the solution is incubated for 10 min at 95 ° C. b. Labeling and fragmentation in one step A 10 wL of PCR are added 10 wL of meta-BioPMDAM (100 mM in DMSO), 5 ~. L of 0.1 M HCl and 75 ~. L of Dnase / Rnase free water. The solution is incubated for 30 min at 60 ° C. The rest of the protocol is identical to that of Example 9. Results Protocol used - Homology I (rfu) B (rfu) I / B (%)

[0195]

at. Marking and fragmentation in 99.5 14129 624 22.7 two-step

[0196]

b. Marking and fragmentation in ~ 98.9 ~ 4431 ~ 667 6.6 a step ~

[0197]

Table 13: Comparative study of labeling and fragmentation in two distinct steps and in a single step As shown in Table 13, the results obtained with the one-step protocol are satisfactory.

Those obtained with two-step labeling and fragmentation are even better.

This example shows that the labeling and cleavage steps can be dissociated to improve the labeling depending on the target used.

[0198]

CA 02444926 2003-10-17 WO 02/090584 PCT / FR02 / 01542 110 Example 28: Maraua ~ e and fragmentation of DNA amlicons in different reaction formats The DNA amplicons were prepared by PCR amplification according to the protocol described in Example S. Three labeling reactions were carried out. at.

Margua e ~ and fragmentation in a format of 2S0 ~ L 1o A SO p, L of PCR are added 7S wL of mâta-BioPMDAM (100 mM in DMSO) and 102.5 ~.

L of Dnase / Rnase free water. The solution is incubated for 2S min at 9 ° C. Then 22.5 ~. L of 0.1 M HCl are added and the solution is incubated for S min at 9 ° C. I5 b. Marc ~ uag_e and fra ~ nentation in a 200 ~ format. L A SO p. L of PCR are added 7S ~. L of mâta-BioPMDAM (100 mM in DMSO) and S2, S ~. L of Dnase / Rnase free water. The solution is incubated for 2S min at 9 ° C. Then 22.5 ~. L of 0.1 M HCl are added and the solution is incubated for S min at 9 ° C. vs. Labeling and fragmentation in a format of 1S0 wL A SO ~ L of PCR are added 7S ~, L of mâta-BioPMDAM (100 mM in DMSO) and 2, S p, L of water Dnase / Rnase free. The solution is incubated for 2S min at 9 ° C. 2S Then 22.5 ~. L of 0.1M HCl are added and the solution is incubated for S min at 9 ° C. The rest of the protocol is identical to that of Example 9.

IDT Exhibit 1020.10066

[0199]

CA 02444926 2003-10-17 WO 02/090584 PCT / FR02 / 01542 111 Results Protocol used Homology (%) I (rfu) B (rfu) I / B

[0200]

at. 250 format l00.0 5606 549 10.2, uL

[0201]

b. Formatde200, uL 99.4 5886 557 10.6

[0202]

vs. Format of l50 99.4 6800 537 12.7, uL

[0203]

Table I4: Labeling and Fragmentation According to Different Formats The results obtained in terms of signal and percentage of homology are very satisfactory in all cases.

In addition, although the format of the reaction varies from 150 to 250 p.

L, the results have similar values.

This example shows a flexibility of the reaction format of the labeling protocol which can 1o accept different volumes and in particular different volumes of amplification products.

Example 29: Comparison between a protocol using a purification step before fragmentation and a protocol using a purification step after fragmentation 1s The DNA amplicons were prepared by PCR amplification according to the protocol described in Example 5.

Two labeling reactions were carried out. at. Labeling, purification then fragmentation of the DNA amplicons A 10 wL of PCR are added 10 ~, L of meta-BioPMDAM (100 mM in DMSO) and 80 p. L of Dnase / Rnase free water. The solution is incubated for 10 min at 95 ° C. The purification is then carried out according to the protocol described in Example 9. To the solution of purified labeled amplicons 6 ~, L of 0.1 M HCl are added. The solution is incubated for 10 min at 95 ° C. Four hundred (400) wL of hybridization buffer preheated to 95 ° C for 10 min is added. The composition of the hybridization buffer as well as the rest of the protocol are identical to that of Example 9.

[0204]

CA 02444926 2003-10-17 WO 02/090584 PCT / FR02 / 01542 112 b.

Marqua ~, e fra ~ nentation then purification of DNA amplicons A 10 ~. L of PCR are added 10 wL of meta-BioPMDAM (100 mM in DMSO) and 77 ~. L of Dnase / Rnase free water. The solution is incubated for 10 min at 95 ° C. Then 3 ~, L of 0.1 M HCl are added and the solution is incubated again for 10 min at 95 ° C. The rest of the protocol is identical to that of Example 9. Results Protocol used homology (%) I (rfu) B (rfu) I / B

[0205]

at. Purification before fragmentation96,9 6256 473 13

[0206]

b. Fragmentation before purification96,1 6066 556 11

[0207]

IDT Exhibit 1020.10067

Table 15: Comparison between protocol using a purification step before fragmentation and protocol using a purification step after fragmentation This result, presented in Table 15, shows that the purification step can be introduced between the labeling and fragmentation steps.

In addition, the introduction of purification between the labeling and fragmentation steps makes it possible to carry out the denaturation during the cleavage and to hybridize on the chip la all of the fragments of labeled amplicons.

Example 30: Synthesis of 2-vitro-para-BioPDAM Synthesis scheme

[0208]

CA 02444926 2003-10-17 WO 02/090584 PCT / FR02 / 01542 113 1 CHO 02 \ Ethylene glycol o2N \ NaZS-H20 ~ o2N \ / pTsOH, toluene ~ / EtOH Dean-Stark NO2 NO2 101 NH2 2,4-dinitrobenzaldehyde 100 Biotin NMM iBuOCOCl DMF r O ~ O CHO 02N \ O O2 \ HN ~ NH ~ / H2S04 aq HNxNH H ~ HH ~ H, ~~~~~ NH S ,, ~~~~ NH 102 NHz NHa ~ 103 i H2 104 N j 2 O2N \ HN ~ NH DMF HNxNH ~ / H ~ H ~ H ~ H NH., ~ Mn02 ~ S., ~~ 105 SO 2-nitro para-BioPDAM ~ Protection of the aldehyde 5 g (25.5 mmol) of 2,4-dinitrobenzaldehyde in 250 mL of toluene, and 20 mL of ethylene glycol and 150 mg of para-toluenesulfonic acid are added.

The mixture is heated to reflux while recovering the water in a Dean-Stark system for 6 h.

Treat with 150 mL of EtOAc and 100 mL of H2O. The solution is extracted twice with ethyl acetate, the organic phase is dried with MgSO4 and then evaporated. The oil obtained corresponding to product 100 is used for the following reaction.

[0209]

CA 02444926 2003-10-17 WO 02/090584 PCT / FR02 / 01542 114 1H NMR (200 MHz, CDCl3): 8 = 8.70 (d,!

Haro, J = 2Hz, H3); 8.44 (dd,! Haro, J = 2Hz, J = 6Hz, HS); 8.02 (d,! Haro, J = 8Hz, H6); 6.49 (s, 1H, CH); 4.12-4.06 (m, 4H, CH2-CH2). ~ Reduction of the dinitro derivative 100 The protected 2-4-dinitrobenzaldehyde (5.4 g; 25.5 mmol) is dissolved in an ethanol-water mixture (6l1), then 2 equivalents of Na2S nonahydrate (12.3 g; 51, 1 mmol) are added. The reaction mixture is then heated for 30 min. Evaporation then extraction with dichloromethane are carried out. After drying and filtration, the reaction medium is evaporated to obtain an oil, which is directly purified on a silica column (cyclohexane-ethyl acetate 60! 40). Compound 101 is isolated with a yield of 45%. Compound 101: F 58-60 ° C. - 1H NMR (200 MHz, CDCl3): 7.49 (d,! Haro, J = 2Hz, H3); 7.09 (d, 1H ~, J = 2Hz, H6); 6.80 (dd,! Haro, J = 2Hz, J = 6Hz, HS); 6.27 (s, 1H, CH); 3.99-3.97 i5 (m, 4H, CH2-CHZ). . ~ Couple with biotin D-biotin (1.0 g; 4.1 mmol) is solubilized in 20 mL of anhydrous DMF and 600 ~ L of N-methylmorpholine.

Isobutyl chloroformate (700 ~, L; 5.5 mmol) is added under argon while cooling in an ice bath. The mixture is stirred for 5 min, then 1 g (4.75 mmol) of compound 101 and 500 ~ L of N-methylmorpholine are added. The solution is kept under stirring at room temperature for 4 h, then evaporated to dryness. The oil obtained is passed directly through a silica column with 7% MeOH-DCM as elution solvent then 10%. Product 102 (1.1 g; 2.52 mmol) is obtained with a yield of 62%. 1H NMR (200MHz, DMSO-d6) 8 = 10.40 (s, 1H, NH-CO); 8.31 (d,! Haro, J = 2Hz, H3); 7.77 (dd,! Haro, J = 2Hz, J = 6Hz, HS); 7.68 (d,! Haro, J = 2Hz, H6); 6.43 (br s, 1H, NH-CO-NH); 6.36 (br s, 1H, NH-CO-NH); 6.23 (s, 1H, CH); 4.28 (m, 1H, CH2-CH-NH); 4.14 (m, 1H, CH-CH-NH); 3.92 (s, 4H, CH2-CH2); 3.12 (m, 1H, CH-S); 2.85 and 2.76 (ABX system, 2H, 2J ~ = 5Hz, 3J ~ = 12Hz, 3JBx = 0Hz, CH2-S); 2.29 (t, 2H, J = 8Hz, CH2-CO); 1.61-1.39 (m, 6H, (CH2) 3).

[0210]

CA 02444926 2003-10-17 WO 02/090584 PCT / FR02 / 01542 115 ~ Deprotection of the acetal The product 102 (768 mg; 1.76 mmol) is suspended in 25 mL of THF. Everything dissolves after adding 4 mL of 2N H2SO4. The mixture is left under stirring for 2 h.

Evaporated then rinsed and washed with water on a frit. Compound 103 (694 mg) is obtained in the form of a yellow powder with a yield of 90%. F 165 ° C. - 1H NMR (200MHz, DMSO-d6) 8 = 10.69 (s, 1H, NH-CO); 10.09 (s, H, CHO); 10 8.43 (d, 1 Har ° a J = 2Hz, H3); 7.91 (s, 2H -. °, HS and H6); 6.42 (br s, 1H, NH-CO-NH); 6.35 (br s, 1H, NH-CO-. N ~ H; 8 = 6.23 (s, 1H, CH); 4.29 (m, 1H, CH2-CH-NH); 4.13 (m, 1H, CH-CH-NH); 3.12 (m, 1H, CH-S); 2.84 and 2.78 (ABX system, 2H, 2J, ~ = 5Hz, 3J ~ = 12Hz, 3JBX = 0Hz, CH2-S); 2.29 (t, 2H, J = 8Hz, CH2-CO); 1.61-1.39 (m, 6H, (CH2) 3). ~ Formation of hydrazone l04 Aldehyde 103 is suspended in ethanol and heated to 80 ° C. When hydrazine is added, everything dissolves and the solution immediately turns orange. A precipitate forms after 5 min. The mixture is heated with stirring for 1 h. It is filtered through a frit and then the precipitate is dried. The product 104 (700 mg; 690 mmol) is obtained with a yield of 98%.

IDT Exhibit 1020.10068

Mp 169 ° C. - 1H NMR (200MHz, DMSO-d6) ~ = 10.31 (s, 1H, NH-CO); 8.31 (d, 1Har °, J = 2Hz, H3); 7.96 (s, H, CHO); 7.87 (d, Haro, J = 2Hz, H6); 7.68 (dd, 1H ~ °, J = 2Hz, J = 6Hz, HS); 7.31 (s, 2H, NH2); 6.42 (br s, 1H, NH-CO-NH); 6.34 (br s, 1H, NH-CO-NH); 4.29 (m, 1H, CH2-CH-NH); 4.13 (m, 1H, CH-CH-NH); 3.12 (m, 1H, CH-S); 2.84 and 2.78 (ABX system, 2H, ZJ, ~ = 5Hz, 3J ~ = 12Hz, 3JB ~ = 0Hz, CH2-S); 2.29 (t, 2H, J = 8Hz, CH2-CO); 1.61-1.39 (m, 6H, (CHZ) 3). ~ Formation of the diazo 105

[0211]

CA 02444926 2003-10-17 WO 02/090584 PCT / FR02 / 01542 116 Compound 104 (200 mg; 0.492 mmol) is dissolved in 8 mL of DMF. 400 mg of MnO2 are added.

Vigorously stirred for 10 min. It is filtered through Millipore containing velite (thickness: 2 cm) and molecular sieve powder 3 ~ (0.5 cm). Evaporated to dryness then washed with ether. We filter again through millipore. Compound I05 (180 mg; 0.445 mmol) is obtained in the form of an orange powder with a yield of 96%. Mp 155 ° C. - 1H NMR (200MHz, DMSO-d6) 8 = 10.21 (s, 1H, NH-CO); 8.60 (d, 1Ha, .o, J = 2Hz, H3); 7.77 (d, Haro, J = 6Hz, HS); 7.22 (d, 1Ha, .o, J = 6Hz, H6); 6.60 (s, H, CH-N); 6.41 (br s, 1H, NH-CO-NH); 6.33 (br s, 1H, NH-CO-NH); 4.29 (m, 1H, CHa-CH-NH); 10 4.13 (m, 1H, CH-CH-NH); 3.12 (m, 1H, CH-S); 2.84 and 2.78 (ABX system, 2H, 2J ~ = 5Hz, 3JAx = 12Hz, 3JBx = 0Hz, CH2-S); 2.29 (t, 2H, J = 8Hz, CHZ-CO); 1.61-1.39 (m, 6H, (CHz) 3). The reactivity of compound 105 was tested on 3'-Uridine-monophosphate followed by capillary electrophoresis. The analysis conditions are those of Example 6.1. The results show a half-reaction time of 45 minutes.

The stability of the reagent is greater than I month at -20 ° C Example 31: Marpua ~ e and fragmentation of DNA amplicons with the reagent of maraua ~ e 2-vitro-yara-BioPDAM The 2-vitro para-BioPDAM derivative has was obtained according to the reaction scheme described in Example 30. The DNA amplicons were prepared by PCR amplification according to the protocol described in Example 5. Two labeling reactions were carried out. at. Mar ~ uag-e by the reagent 2-vitro para-BioPDAM A 10 p, L of PCR are added 2 ~ L of 2-vitro paf ~ a-BioPDAM (100 mM in DMSO), S ~, L of HCl at 0.1 M and 83 ~, L of water Dnase / Rnase free. This solution is incubated for 30 min at 60 ° C.

[0212]

CA 02444926 2003-10-17 WO 02/090584 PCT / FR02 / 01542 117 b.

Labeling with the meta-bioPMDAM reagent A 10 ~, L of PCR are added 2 ~, L of meta-BioPMDAM (100 mM in DMSO), S ~, L of 0.1 M HCl and 83 ~, L d Dnase / Rnase free water. This solution is incubated for 30 min at 60 ° C. The rest of the protocol is identical to that of Example 9. Result Protocol used Homology I (rfu) B (rfu) IIE (%)

[0213]

at. Labeling with 2-nitro- reagent

[0214]

100,0 24392 899 27,1

[0215]

paya-BioPDAM

[0216]

b. Masking by mta- reagent

[0217]

98,9 21883 774 28,3

[0218]

IDT Exhibit 1020.10069

BioPMI ~ AM

[0219]

Table 16: Comparative study of DNA labeling with the 2-vitro para-BioPDAM derivative with respect to the meta-BioPMDAM derivative The 2-vitro para-BioPDAM reagent used for DNA labeling gives interesting results in terms of term of labeling intensity and percentage of homology.

Example 32: Insertion of a double bond between the diazomethyl function and the phenyl ring, removal of the DAM and synthesis of a molecule uarticularly suited to this 2o removal The objective is the removal of the diazomethyl function (DAM) of the aromatic structure to minimize the effect of steric hindrance during the alkylation of phosphates and also during the hybridization of the labeled nucleic acid with its complementary sequence.

[0220]

CA 02444926 2003-10-17 WO 02/090584 PCT / FR02 / 01542 118 Synthesis scheme / OO OEt KOH / EtOH O 89 O 'OMe O OMe -formylbenzoate of Acetoacetate O methyl ethyl H2N-EG3-NH2 H2N ~ O ~ O ~ O ~ NH O -Biotin MM UOCCIO O Mn02 HN '_NH SN!

~ / ~ O ~ O ~ O 90 For the aldol reaction for the formation of (paya-methoxycarbonyl) - styrylmethylketone 89, ethyl facetoacetate is used for the high acidity of the methylene protons, which facilitates (attack of the formyl group, with subsequent elimination of H20 (favored by the conjugation of the double bond with (aromatic ring) and decarboxylation by hydrolysis 1o due to the basic medium.

The rest of the synthesis is similar to what has been shown in the other examples.

The final product paya-Bio-EG3-SMDAM 90 has two more carbons between diazomethyl and (aromatic ring, which limits possible steric problems, while retaining the stabilization of diazomethyl by the aromatic system by conjugation.

D N H2N-NH2 ~ iB

[0221]

CA 02444926 2003-10-17 WO 02/090584 PCT / FR02 / 01542 119 Example 33 'Capture and detection of a nucleic acid on a solid support bearing diazomethyl groups The reactivity of a resin bearing diazomethyl groups was studied for determine its ability to bind nucleic acids.

4- (Diazomethyl) phenoxymethyl-polystyrene (reference 17338, Fluka) is a resin described for its ability to bind carboxylic groups, in particular those present in proteins (G. Bhalay, AR Dunstan, Tetrahedron Lett. 39, 7803 - 1998), but it is not described for its ability to bind DNA molecules. We tested the possibility of capturing nucleic acids with this reagent, and revealing them by a colorimetric test.

The experiment is carried out with part of the reagents present in the HLA-DR oligo-detection kit (reference 33 202, bioMérieux, France, basic principle described in patent EP 549 776-B1),. allowing the detection of nucleic acids amplified by PCR in microplates, by colorimetric reading.

In the context of the experiment described, nucleic acids produced by PCR are simultaneously reacted with the resin tested, as well as with a paya-Bio-EG3-PDAM molecule, the synthesis of which is mentioned in Example 20.

If the DNA reacts with the diazomethyl functions present on the two compounds, it will be possible to reveal it, after washing and elimination of the molecules fixed in a non-covalent manner, thanks to a colorimetric reaction involving an enzyme coupled to streptavidin. Streptavidin is associated with a horseradish peroxidase, this enzyme capable of breaking down an OrthoPhenylene Diamine molecule (Color 1 reagent from the kit) into a compound detectable at a wavelength of 492 nm. Example 33.1: Capture and detection of DNA 10 mg of resin are incubated for 30 minutes at 60 ° C with 50 ~ L of PCR carried out as described in Example 5, in 400 ~, L of pure water (Sigma) plus 5%. L of pa ~~ a-Bio-EG3-PDAM. This resin is then washed with 500 ~, L of PBS tween buffer (Color reagent 0

[0222]

IDT Exhibit 1020.10070

CA 02444926 2003-10-17 WO 02/090584 PCT / FR02 / 01542 120 HLA of the kit, PBS pH 7.0; Tween 1%, BND 0.2 g / L; Ciprofiaxacin 0.01 g / L).

The resin is then resuspended in 100 ~ L of PBS Tween, and 250 ~, L of streptavidin hybridization buffer (BPS pH 7.0 TWEEN 0.5%) supplemented with streptavidin HRP (S-911, MOLECULAR PROBES, EUGENE, OR, USA) diluted to 1/10000. The reaction mixture is incubated for 30 min at room temperature. The resin is then washed three (3) times with 500 ~ L of PBS tween buffer, and incubated at room temperature in the presence of chromogenic reagent (1 Color tablet, orthophenylene diamine hydrochloride, diluted in 5 mL of buffer Color 2, 100mM Sodium Phosphate, 50mM Citric Acid, 0.03% HaO2). After an incubation of 20 min in the dark, Ia reaction is then blocked by 50 ~, L of 1o H2SO4 (I, ~ N reagent Color 3). The supernatant is then pipetted and placed in a microplate to read the absorbent of the reaction medium at 492 nm. Example 33.2: Control without nucleic acid 10 mg of resin are incubated for 30 minutes at 60 ° C in 425 ~. L of pure water (Sigma) supplemented with 5 ~, L of para-Bio-EG3-PDAM. This resin is then washed with 500 ~ L of PBS buffer. The sample is then treated identically to the process described in Example 33.1. 2o Example 33.3: Control with PCR carried out without targets 10 mg of resin are incubated in 400 p.

L of pure water supplemented with 5 ~, L of paya- Bio-EG3- PDAM for 30 minutes at 60 ° C, with 50 pL of PCR carried out with a volume of 25 ~ L of pure water instead of the volume i Genomic DNA described. This resin is then washed with 500 ~. L of PBS buffer. The sample is then treated identically to the process described in Example 33.1. Example 33.4: Control with PCR without revealing molecule 3o 10 mg of resin are incubated with 50 ~, L of PCR in 400 ~, L of pure water for 30 minutes at 60 ° C. This resin is then washed with 500 ~, L of PBS buffer. The sample is then treated identically to the process described in Example 33.1.

[0223]

CA 02444926 2003-10-17 WO 02/090584 PCT / FR02 / 01542 121 Example 33.5: Control with non-captured nucleic acid 10 mg of resin are incubated for 30 minutes at 60 ° C with 400 wL of pure water added with 5 ~ , L depa ~ a-Bio-EG3-PDAM. This resin is then washed with 500 ~.

L of PBS tween buffer (Color 0 HLA reagent from the kit, PBS pH 7.0; Tween 1%, BND 0.2 g / L; Ciprofiaxacin 0.01 g / L). The resin is then resuspended in 100 ~. L of PBS Tween, and 250 ~ L of streptavidin hybridization buffer supplemented with streptavidin HRP diluted to 1/10000. To this preparation is added 50 ~. L of a DNA preparation prepared as follows 1o We add 5 wL of para-Bio-EG3-PDAM and 70 ~, L of pure water at 25 ~, L of DNA from a PCR prepared as described in example 5. This mixture is incubated for 30 min at 60 ° C, then the excess label is removed by subjecting the preparation to purification on a QIAquick column (Nucleotide Removal Kit, Qiagen, Hilden, Germany) according to the protocol recommended by the supplier, by performing a final elution in a volume of 50 p. L. The reaction mixture is incubated for 30 min at room temperature, then it is treated according to the procedure described in Example 33.1. 2o Results Conditions Absorbance 492 nm

[0224]

Ex. 33.1: DNA captured on Rsin527

[0225]

Ex. 33.2: Without nucIic acid249

[0226]

Ex. 33.3: PCR wilhout target 261

[0227]

Ex. 33.4: Witness without marker264

[0228]

Ex. 33.5: Nuclear acid not captured 249

[0229]

IDT Exhibit 1020.10071

Table 17: Study of the reactivity of a resin bearing diazomethyl groups

[0230]

CA 02444926 2003-10-17 WO 02/090584 PCT / FR02 / 01542 122 In table 17, a high colorimetric value indicates a high concentration of enzymes in the reaction medium, corresponding to a significant presence of nucleic acid carrying derivatives biotins.

Controls indicate that the signal is not due to non-specific adsorption of DNA on the bead, a reaction of paya-Bio-EG3-PDAM on the resin, or adsorption of streptavidin HRP on the resin. , but in the presence of DNA captured covalently and labeled with iepa ~ a-Bio-EG3-PDAM. Example 34: Maraua ~ e of a PCR product allowing its cauture and its detection in a microplate It is shown in this example the possibility of labeling a DNA molecule with a single type of molecule carrying a diazomethyl function, so as to capture and detecting this nucleic acid in a single step, on a microplate.

The experiment is carried out with part of the reagents present in the HLA-DR oligo-detection kit (reference 33202, bioMérieux), allowing the detection of nucleic acids amplified by PCR in microplates, by colorimetric reading.

In the context of the experiment described, para-Bio-EG3-PDAM, the synthesis of which is described in Example 20, is reacted with nucleic acids produced by PCR. The DNA reacts with the diazomethyl groups of the molecule, and is thus endowed with biotins grafted on its phosphate.

It will then be possible to capture the nucleic acid by incubation on a microplate where streptavidin molecules are adsorbed, and to reveal it, thanks to a colorimetric reaction.

A detection reagent is used which is also a streptavidin molecule, associated with a horseradish peroxidase (Horse Radish Peroxidase, HRP).

Under the conditions used, the peroxidase can decompose an OrthoPhenylene Diamine molecule (Color 1 reagent from the kit) into a compound detectable at a wavelength of 492 nm.

[0231]

CA 02444926 2003-10-17 WO 02/090584 PCT / FR02 / 01542 123 Example 34.1: Capture and detection of the DNA obtained from PCR, on a microplate 10 ~, L of DNA obtained by PCR amplification are marked as described in l Example 5 by incubating it for 30 min at 80 ° C in 80 ~, L of pure water (Sigma) supplemented with 10 ~ L of pa ~ a-Bio-EG3-PDAM. After labeling, the DNA is purified on a QIAquick column (Nucleotide Removal I ~ it, Qiagen, Hilden, Germany) according to the protocol recommended by the supplier, and the final eluate is collected in 50 ~.

L of EB buffer (10 mM Tris EDTA, pH 8.5).

Twenty (20) ~. L of this eluate is diluted in 180 ~, L of PEG buffer (0.1 M Na P03; 0.5 M NaCl; 0.65% Tween 20; 0.14 mg / mL DNA of salmon sperm (Gibco); 2% PEG 4000) supplemented with streptavidin HRP (S-911, Molecular Probes, Eugene, OR, USA) diluted to 1: 10,000. One hundred (100) ~. L of this preparation are then incubated for 1 hour at 37 ° C., either in a well of Combiplate 8, streptavidin coated, (reference 95029263, Labsystem, Helsinki, Finland) or in a control well, obtained from a Maxisorb strip. (Nunc, Denmark). Example 34.2: Production of witnesses Witnesses are simultaneously produced as follows 2o A - Labeling control without DNA Ninety (90) ~ L of pure water added with 10 ~, L of paya-Bio-EG3- PDAM are incubated for 30 min at 60 ° C. The reaction mixture is then treated in a manner similar to the process previously described in Example 34.1. B - Labeling control without markers Ten (10) ~ L of DNA obtained by PCR amplification, as described in Example 5, are incubated for 30 min at 60 ° C. in 90 ~ L of pure water.

The reaction mixture is then treated similarly to the process previously described in Example 34.1. 3o All the strips are then washed with three times 100 ~ .l of PBS Tween buffer (Color 0 HLA reagent from the kit, PBS pH 7.0; Tween 1%, BND 0.2 g / L; Ciproflaxacin 0, 01

[0232]

CA 02444926 2003-10-17 WO 02/090584 PCT / FR02 / 01542 124 g / L) then the presence of streptavidin HRP is revealed by adding 100 ~.

IDT Exhibit 1020.10072

L of chromogenic reagent (1 Color 1 tablet, Orthophenylenediamine hydrochloride, diluted in 5 mL of Color 2 buffer, 100 mM sodium phosphate, 50 mM citric acid, 0.03% Ha02), incubated 20 min in the dark , the reaction then being blocked by 50 ~. L of H2S04 (1.8 N Reagent Color3). The absorbance of the reaction medium is then measured at 492 nm. Results Conditions Combiplate Tmoin maxisorp

[0233]

A - DNA marked 1382 152

[0234]

B - unlabeled DNA 178,136

[0235]

C - Without DNA 140 192

[0236]

Table 16: Detection of DNA captured and labeled with dupara-Bio-EG3-PDAM The experiment, set out in Table 16, therefore shows that DNA labeled with para-Bio-EG3-PDAM can be captured and detected by a single step in a microplate well.

As indicated by the reaction controls, the signal generated is due solely to DNA and does not result from non-specific adsorption of the nucleic acid to the wall of the microplate, or to sireptavidin, or to a nonspecific reaction of streptavidin HRP on unlabeled DNA, or on the plastic of the microplate.

Example 35: Double maraua ~ e of a PCR product uermettant its capture and detection on a solid support of the microplague type It is shown in this example the possibility of labeling with two molecules, carrying diazomethyl functions, and in a single step, a DNA molecule, so as to capture and detect it on a microplate.

[0237]

CA 02444926 2003-10-17 WO 02/090584 PCT / FR02 / 01542 125 The experiment is carried out with part of the reagents present in the HLA-DR oligo-detection kit, allowing the detection of nucleic acids amplified by PCR in microplates, by simple colorimetric reading.

Within the framework of the described experiment, one reacts simultaneously on nucleic acids produced by PCR of ~ 1-Pyrényldiazomethane (PDAM), and ~ para-Bio-EG3-PDAM. If the DNA reacts with the diazomethyl functions present on the two compounds, it will be able to bind to a support carrying anti-pyrene antibodies, and it will be possible to reveal it with a streptavidin molecule associated with a horseradish peroxidase.

This 1o enzyme can decompose an OrfhoPhénylène Diamine molecule, acting as a revelation reagent, into a compound detectable at a wavelength of 492 nm.

Example 35.1: Double labeling of the DNA and detection on a microplate Anti-pyrene antibodies were adsorbed onto eight (8) wells Maxisorp strips, by incubating overnight at room temperature 100 ~ L of a solution of 1. , I ~. L of anti-pyrene antibodies diluted in 100 ~, L of bicarbonate buffer (O, OSM pH 9.6). Such antibodies, called Rabbit anti-pyrene (Ref .: YSRT-AHP236) are available from Accurate Chemical & Scientific (Westbury, New York, USA). Of course, this could have been achieved with other commercially available antibodies, without there being any divergent results compared to those we obtained in this example. Was then labeled 10 ~ L of DNA obtained by PCR amplification, as described in Example 5, incubating 30 min at 60 ° C in 40 ~. L of pure water (Sigma), 10 tCL of para-Bio-EG3-PDAM, 2 ~. L of PDAM (P-1405, 1-pyrenyldiazomethane, Molecular Probes, Eugene, OR, USA) and 38 ~, L of DMSO. After labeling, the DNA was purified using a QiAquick kit (QLAGEN) and the final eluate was collected in 50 p.

L of EB buffer (10 mM Tris EDTA, pH 8.5). Twenty (20) ~, L of this eluate was diluted in 180 ~, L of PEG buffer (0.1 M Na P03; 0.5 M 3o NaCl; 0.65% Tween 20; 0.14 mg / mL DNA of Salmon semen (Gibco); 2% PEG 4000) supplemented with streptavidin HRP (S-911, MOLECULAR PROBES, EUGENE, OR,

[0238]

IDT Exhibit 1020.10073

CA 02444926 2003-10-17 WO 02/090584 PCT / FR02 / 01542 126 USA) diluted to 1/10 000.

One hundred (100) ~, L of this preparation were then incubated for 1 hour at 37 ° C., either in a well of adsorbed Maxisorp strip, or in a control well, not adsorbed. Example 35.2 Production of controls Witnesses were simultaneously produced in the following manner A - Control of marguage with., Para-Bio-EG3-PDAM only 1o Ten (10) ~ L of DNA obtained by PCR amplification, as described in Example 5, are labeled by incubation for 30 min at 60 ° C in 90 ~, L of pure water added with 10 p. L of paya-Bio-EG3-PDAM. After labeling, the DNA is purified using a QIAquick kit and the final eluate is collected in 50 ~ L of EB buffer. Twenty (20) wL of this eluate is diluted in 180 ~, L of PEG buffer supplemented with streptavidin HRP diluted to 1: 10,000. 1s One hundred (100) ~ L of this preparation are then incubated for 1 hour at 37 ° C, i.e. in an adsorbed Maxisorp strip well, or in a non-adsorbed well. B - Branded control with PDAM alone Ten (10) ~, L of DNA obtained by PCR amplification, as described in Example 5, are 2o labeled by incubation for 30 min at 60 ° C in 90 wL of pure water added with 2 ~.

L of PDAM and 38 ~ L of DMSO. After labeling, the DNA is purified using a QIAquick kit and the final eluate is collected in 50 ~ L of EB buffer. Twenty (20) ~. L of this eluate is diluted in 180 ~, L of PEG buffer supplemented with streptavidin HRP diluted to 1/10 000. One hundred (100) wL of this preparation is then incubated for 1 hour at 37 ° C., either in a well of adsorbed Maxisorp strip or in a non-adsorbed well. C - Witness without eg- ~ mark. Ten (10) ~ L of DNA obtained by PCR amplification, as described in Example 5, are incubated for 30 min at 60 ° C in 100 ~. L of pure water. After labeling, the DNA is purified by 3o use of a QIAquick kit and the final eluate is collected in 50 ~. L of buffer EB. Twenty (20) ~, L of this eluate is diluted in 180 ~, L of PEG buffer supplemented with streptavidin

[0239]

CA 02444926 2003-10-17 WO 02/090584 PCT / FR02 / 01542 127 HRP diluted to 1/10 000.

One hundred (100) ~, L of this preparation are then incubated for 1 hour at 37 ° C., either in a well of adsorbed Maxisorp strip, or in a non-adsorbed well. The strips are then washed with three times 100 µl of PBS Tween buffer (color 0) then the presence of streptavidin HRP is revealed by adding 100 ~. L of chromogenic reagent (Color 2), incubated 20 min in the dark, the reaction then being blocked by 50 wL of H ~ SO4 (Color3). The absorbance of the reaction medium is then measured at 492 nm. 1o Results Antibody anti-pyrene plaque indicator

[0240]

Conditions adsorbs on plaquenon adsorbs

[0241]

(rfu) (rfu)

[0242]

Ex. 35.1: DNA labeled PDAM + para-Bio-

[0243]

348 16

[0244]

EG3-PDAM

[0245]

Ex. 35.2.

A: DNA taggedpaYa-Bio-EG3-

[0246]

IDT Exhibit 1020.10074

44 19

[0247]

PDAM

[0248]

Ex. 35.2

B: DNA labeled PDAM 68 12

[0249]

Ex. 35.2.

C: unlabeled DNA 75 19

[0250]

Table 19: Double labeling of DNA by PDAM and lepa ~ a-Bio-EG3-PDAM The result of Table 19 clearly shows a significant signal resulting from the capture of the DNA in the wells by the anti-pyrene antibodies, as well as the simultaneous labeling thereof with streptavidin HRP, which is fixed.

As shown by the absence of signal at the level of controls, this detection is specific for labeled DNA, and is not due to non-specific adsorption of DNA or streptavidin HRP to the plastic, or to non-specific binding of the enzyme to the captured DNA.

This example therefore shows that it is possible to

[0251]

CA 02444926 2003-10-17 WO 02/090584 PCT / FR02 / 01542 128 carrying out a double labeling of DNA in a single step, this double labeling being able to be used to capture and detect it simultaneously.

Example 36: Maraua ~ e of a PCR product simultaneously allowing its capture and its detection by complementary nucleic acid probes This experiment makes it possible to demonstrate that it is possible to specifically detect a DNA, said DNA being captured on a solid surface using the reactivity of the diazomethyl function vis-à-vis a phosphate group of DNA. The experiment is carried out with part of the reagents present in the HLA-DR oligo-detection kit (reference 33 202, bioMérieux, France), allowing the detection of nucleic acids amplified by PCR in microplates, by colorimetric reading.

In the context of the experiment described, para-Bio-EG3-PDAM is reacted with nucleic acids produced by PCR. DNA reacts with the diazomethyl functions of the molecule, and is thus endowed with biotins grafted onto its phosphates.

It will then be possible to capture the nucleic acid by incubation on a microplate where streptavidin molecules are adsorbed, and to reveal it, using a probe consisting of an oligonucleotide complementary to the captured sequence, associated with a horseradish peroxidase, this enzyme capable of breaking down colorless OrthoPhenylene Diamine molecules (Color 1 reagent from the kit) into a compound detectable at a wavelength of 492 nm.

Example 36.1 Capture and specific detection of DNA on a microplate The labeling of 10 ~ is carried out in duplicate.

L of DNA obtained by PCR amplification, incubating it for 30 min at 60 ° C with 20 ~.

L of para-Bio-EG3-PDAM. After labeling, the DNA is purified on a QIAquick column (Nucleotide Removal Kit, Qiagen, Hilden, Germany) 3o according to the protocol recommended by the supplier and the final eluate is collected in 50 ~.

L of EB buffer (10 mM Tris-HCl, pH 8.5). Eighty-five (65) pL of the mixture of these

[0252]

IDT Exhibit 1020.10075

CA 02444926 2003-10-17 WO 02/090584 PCT / FR02 / 01542 129 eluates are denatured with 8.5 ~, L of reagent R4 (2N NaOH) for 5 min at room temperature, then the solution is then neutralized with 8, 5 p.

L of RS reagent (2N acetic acid). 850 ~, L of hybridization buffer (R6 - 10 mM Tris-HCL, pH 7.0 BND 0.2 g / L, Ciproflaxacin 0.01 g / L.) And 85 µL of oligonucleotide are added to the mixture. detection (R7 - 4 mM sodium phosphate, 1 mM potassium phosphate, pH 7.0, 0.1% bovine serum albumin, 0.5% phenol). One hundred (100) ~, L of this preparation are deposited either on the positive control of an R1 strip supplied with the kit, (capture by a consensus sequence of the amplified gene), or on a Combiplate 8 streptavidin coated plate (reference 95029263, Labsystem, Helsinki, Finland) or on a Maxisorp witness plate (Nunc, 10 Denmark). At the same time, the same hybridization reaction is carried out on a tenth and one hundredth dilution, in EB buffer, of the DNA preparation, in order to test the sensitivity of the technique. Example 36.2 Production of the controls Witnesses were simultaneously produced in the following manner 2o A - Comparison with the HLA-DR kit Ten (10) ~, L of DNA obtained by PCR amplification, as described in Example 5, are incubated 30 min at 60 ° C with 20 ~, L depa ~ a-Bio-EG3-PDAM. After labeling, the DNA is purified on a QIAquick column, and the final eluate is collected in a volume of 50 ~, L of EB buffer. The 45 ~ L of eluate are denatured by 4.5 ~.

L of reagent R4 for 5 min at room temperature, then the solution is neutralized with 4.5 ~, L of reagent R5. 450 ~ L of hybridization buffer R6 and 45 ~, L of detection oligonucleotide are added to the mixture. One hundred (100) ~ L of this preparation are deposited either on the positive control of an R1 strip supplied with the kit (hybridization with a consensus sequence of the amplified gene), or on 3o a Combiplate 8 streptavidin plate, or on a control plate Maxisorp.

[0253]

CA 02444926 2003-10-17 WO 02/090584 PCT / FR02 / 01542 130 B - Hybridization carried out on a DNA which does not hybridize to the specific probe Dix (10) ~, L of DNA obtained by PCR amplification , as described in Example 5, are incubated for 30 min at 60 ° C with 20 ~ L of step ° a-Bio-EG3-PDAM. The sample is then treated identically to the process described above in Example A. C - Control without DNA Ten (10) ~ L of reagent R6 (hybridization buffer) and 100 ~ L of reagent R7 (oligonucleotide 1o of detection) are deposited either on the positive control of an R1 strip supplied with the kit, on a streptavidin plate, or on a Maxisorp type control plate.

All the strips of the above protocols are incubated for an hour and a half at 37 ° C, then washed with three times 100 ~, L of PBS Tween buffer (color 0 HLA reagent) then the presence of the specific detection probe is revealed by addition of 100 ~ L of the chromogenic reagent (Color Reagent 2,, PBS pH 7.0; Tween 1%, BND 0.2 g / L; Ciproffaxacin 0.01 g / L), incubated 20 min in the dark, the reaction then being blocked by 50 ~ L of HZS04 (1.8 N reagent Color 3). The absorbent of the reaction medium is then measured at 492 nm. Results Conditions Barette R1 Kit Combiplate Maxisorp HLA DR (Streptavidin) (Control) (Specific capture)

[0254]

A - labeled DNA - Ex 36.1 1215 2160 16

[0255]

A - labeled DNA (diluted to 1110th) NA 900 NA

[0256]

A - labeled DNA (diluted to 11100) NA 53 NA

[0257]

B - Unmarked DNA witness - 1153 40 17 Ex 36.2 A

[0258]

C- Bacterial DNA witness - Ex 24 15 NA 36.2 B

[0259]

IDT Exhibit 1020.10076

D - Witness without DNA - Ex 36.2 13 12 17 C

[0260]

Table 20: Specific detection on microplate of DNA obtained from PCR

[0261]

CA 02444926 2003-10-17 WO 02/090584 PCT / FR02 / 01542 131 The results of Table 20 indicate an excellent amplification of the target making it possible to envisage use in a diagnostic setting.

This example shows that the labeling at the level of the phosphate groups allows the capture of the DNA and does not prevent specific hybridization thereon.

Example 37: Capture and amplification of DNA from a bacterial lysate and marauded by para-Bio-EG3-PDAM 1o This example shows that it is possible to capture and amplify a bacterial DNA using a capture based on the reactivity of the diazomethyl function on the phosphate of the nucleic acid.

In the present case, the nucleic acids contained in a bacterial lysate and labeled with para-Bio-EG3-PDAM, are captured on magnetic beads coated with streptavidin.

The use of magnetic beads makes it possible to immobilize them by magnetization during the successive washings aimed at eliminating the cellular residues, present in the reaction medium, these residues having to be eliminated because they can inhibit the amplification by PCR which is practiced by the following. 2o The amplification products could be analyzed by passage through DNA chips. Bacterial DNA is obtained by lysis of cells contained in a culture of Illycobacterium tubeYCUiosis.

Lysis is carried out by mechanical lysis.

More precisely, it is carried out by sonication, the treated liquid sample containing glass beads.

Such a process is already well described by the Applicant in its patent application WO-A-99/15621, as regards the beads, and in its patent application WO-A-00/60049, as regards the sonication. .

Sonication can also be carried out using a liquid bath. However, other techniques, known to those skilled in the art, can be used, such as those described in patent US-A-5,902,746 and patent applications WO-A-98/54306 and WO-A-00105338. All these Industrial Property titles belong to the Applicant.

[0262]

CA 02444926 2003-10-17 WO 02/090584 PCT / FR02 / 01542 132 The bacterial DNA was quantified by Picogreen (P-7589; Molecular Probes, Eugene, OR, USA) according to the protocol described by the supplier, at a concentration of 107 copies per ~.

L. Ten (10) p, L of lysate are incubated in the presence of 20 ~ L of para-Bio-EG3-PDAM for 30 minutes at 60 ° C. At the same time, 10 ~. L of lysate are incubated in 20 µl of pure water (Sigma) under the same conditions. The reaction medium is then purified on a QIAquick column (Nucleotide Removal Kit, Qiagen, Hilden, Germany). The purification protocol used is that recommended by the supplier. The final elution volume is 50 wL. The labeled DNA fragments are then captured on Dynal magnetic beads (Dynabeads M-280 streptavidin; reference 112.05; Dynal Biotech ASA, Oslo, Norway), which are prepared according to the following protocol Ninety (90) wL of Dynal beads are washed twice with 200 µL of free pure water (Sigma), then are taken up in 200 ~ L of PEG buffer (0.1 M Na P04, pH 7, 0.5 M NaCl; 0.65 Tween 20 ; 0.14 mL Herring sperm DNA (Reference 15634-017, GibcoBRL); 2 PEG 4000) and incubated for 30 min at 37 ° C. They are then washed twice with 200 ~, L of PBS 1X Tween 20 0.5% buffer, then finally taken up with 90 µL of the same buffer.

Ten (10) wL of the labeled or unlabeled DNA eluates are incubated for 5 min at room temperature with 40 µL of PEG buffer and 2.5 ~ L of the preparation of magnetic beads described above.The beads are then washed three times. times with 200 ~, L of 1 X 0.5% tween PBS buffer, taken up in 200 ~, L of water and incubated for 20 min at 60 ° C, then washed again four times with 200 µl of PBS tween. The beads are finally taken up in 25 µl of water and PCR is carried out following the protocol described in Example 5. Two reaction controls are carried out, one with 25 p. L of pure water and the other with 2.5 p, L of beads prepared and washed under the same 3o conditions as the biological samples, and taken up in 25 ~, L of water.

[0263]

IDT Exhibit 1020.10077

CA 02444926 2003-10-17 WO 02/090584 PCT / FR02 / 01542 133 Result The PCR products are then quantified by Picogreen (P-7589: Molecular Probes, Eugene, OR, USA) according to the protocol described by the supplier.

This method is based on the use of a molecule (Picogreen) having the particularity of becoming fluorescent only when it positions itself inside a DNA molecule (by inserting itself between the bases). Due to the very specific nature of this intercalation, and the fact that the fluorescent signal produced is directly proportional to the amount of DNA present in the medium, it is possible to assay in this way, very precisely, the concentration of 1o nucleic acid present in a sample. The signal is then expressed in rfu (relative fluorescent unit). Analysis of the gel PCR results shows the presence of a single specific band at the expected size in the samples produced from genomic DNA labeled with para-Bio-EG3-PDAM. The bands are not detected when PCR was performed from unlabeled genomic DNA. Quantification of DNA by Picogreen makes it possible to confirm the production of DNA from genomic DNA captured on beads.

Rfu conditions

[0264]

Unmarked witnesses 51

[0265]

DNA mar us 170

[0266]

Background noise 20

[0267]

Table 21: Quantification of the DNA produced by PCR, from a bacterial lysate, captured and purified by lepara-Bio-EG3-PDAM An analysis on a DNA chip, following the protocol described in Example 8 makes it possible to confirm the specificity of the amplification as shown in table 34 below.

Homology I (rfu) B (rfu) I / B%

[0268]

Sample98 11 531 723 16

[0269]

Table 22: Specific detection of the target captured and purified by the a-Bio-EG3-PDAM step

[0270]

CA 02444926 2003-10-17 WO 02/090584 PCT / FR02 / 01542 134 This example shows that it is possible to prepare a biological sample so as to amplify the nucleic acid that it contains, using a capture technique based on on the reactivity of the diazomethyl function on the phosphate groups thereof.

Example 38: Successive amplification of two ~ enes from a bacterial DNA captured on a solid support 1o This example shows that it is possible to amplify several times and on different targets a DNA captured thanks to the reactivity of the diazomethyl function on the phosphate groups thereof.

In the present case, the nucleic acids, contained in a bacterial lysate and labeled with para-Bio-EG3-PDAM, are captured on magnetic beads coated with streptavidin.

The use of magnetic beads makes it possible to preserve them by magnetization during the successive washings aimed at eliminating the cellular residues present in the reaction medium, these residues having to be eliminated because they can inhibit the amplifications by PCR, which will be carried out subsequently. .

IDT Exhibit 1020.10078

These amplifications will take place on two different genes present 2o in the genomic DNA, respectively designated under the names of 16S and rpoB. These two genes are then analyzed using DNA microarrays. The bacterial DNA is obtained by lysis of the cells contained in a culture of Mycobacterium tuberculosis, according to the protocol already described in Example 37.

Bacterial DNA was quantified by Picogreen following the protocol described by the supplier, at a concentration of 107 copies per ~, L. Ten (10) ~, L of lysate are incubated in the presence of 20 ~, L of para-Bio-EG3-PDAM for 30 minutes at 60 ° C. At the same time, 10 ~, L of lysate are incubated in 20 ~, L of pure water (Sigma) 3o under the same conditions.

[0271]

CA 02444926 2003-10-17 WO 02/090584 PCT / FR02 / 01542 135 The reaction medium is then purified on a QIAquick column (Nucleotide Removal Kit, Qiagen, Hilden, Germany).

The purification protocol used is that recommended by the supplier.

The final elution volume is 50 ~, L. The labeled DNA fragments are then captured on Dynal magnetic beads, which are prepared according to the following protocol. Ninety (90) ~, L of Dynal beads are washed twice with 200 ~. L of free pure water (Sigma), then are taken up by 200 ~. L of PEG buffer (0.1 M Na P04, pH 7, 0.5 M NaCl; 0.65% Tween 20; 0.14 mL salmon sperm DNA (Gibco); 2% PEG 4000) and incubated for 30 min at 37 ° C. They are then washed twice with 200 ~, L of PBS 1X Tween 20 0.5% buffer, then finally taken up with 90 ~. L of the same buffer. Ten (10) ~ L of the labeled or unlabelled DNA eluates are incubated for 5 min at room temperature with 40 ~, L of PEG buffer and 2.5 ~. L of the preparation of magnetic beads described above 1 S. The beads are then washed three times with 200 ~, L of PBS buffer 1 - tween O, S%, taken up in 200 ~. L of water and incubated for 20 min at 60 ° C, then washed again four times with 200 ~, L of PBS tween.

The beads are finally taken up in 2S ~, L of water and a PCR is carried out following the protocol described in Example 5. Two reaction controls are carried out, one with 25 ~. L of pure water (Sigma) and the other with 2, S ~, L of beads prepared and washed under the same conditions as the biological samples, and taken up in 25 ~, L of water. After amplification, the reaction medium is collected, and the beads are separated and washed with 150 ~. L of PBS 1 X Tween 0.5%, then resuspended in 25 ~. L of pure water (Sigma). A new amplification is carried out on the beads, but in the presence of primers intended to amplify the rpoB gene. Control amplifications carried out from uncaptured genomic DNA are carried out in parallel on the two amplification systems (rpoB and 16S).

[0272]

CA 02444926 2003-10-17 WO 02/090584 PCT / FR02 / 01542 136 Result The PCR products obtained are then analyzed by DNA chips according to the protocol described in Example 8.

Homology conditions I (rfu) S (rfu) I / B (%)

[0273]

Control PCR 16S 16S 96 4662 367 12

[0274]

Results on RpoB 23 237 358 1 rpoB sequence

[0275]

PCR witness rpoB RpoB 98 5438 397 14

[0276]

Results on sequence 16S 17 183 391 1 16S

[0277]

16S PCR on 16S beads 98 2726 534 5

IDT Exhibit 1020.10079

[0278]

Results on RpoB sequence 8 161 480 <1 rpoB

[0279]

RpoB PCR on lavaRpoB beads 97 3205 349 9

[0280]

Results on 16S sequence 14 84 358 <1 16S

[0281]

Table 23: Analysis on DNA chips of the DNA amplicons resulting from the PCR amplified successively or not 1o The analysis of the PCR results on gel shows the presence of a single band specific to the expected size in the samples produced from DNA para-Bio-EG3-PDAM labeled genomics. The bands are not detected when PCR has been performed from unlabeled genomic DNA.

A DNA chip analysis, as presented in Table 23 above, makes it possible to confirm the specificity of the two successive amplifications, and therefore the ability to carry out a successive amplification of several genes from DNA immobilized on a solid support. , this making it possible to avoid the development of multiplex systems, which often significantly reduce the sensitivity and efficiency of nucleic acid amplifications.

[0282]

CA 02444926 2003-10-17 WO 02/090584 PCT / FR02 / 01542 137 Example 39: Capture and Amplification of DNA on a Nylon Membrane Carrying Diazomethyl Groups A nylon membrane activated so as to carry diazomethyl groups was used in order to capture bacterial DNA, in order to amplify it by PCR. Example 39.1: Modification of the Biodyne C filter Synthesis diagram OOO w vo -I- OO Cr ~ NH NH2 H '~ 1II ~ CDI EG3 3 ~ _ ~ inoacetophenone 68 0 succinic anhydride HZN ~ 69 Nz OO ° ~ NH O ~ _ ~ ~ NH H H2N-NHa o ° Nx ~ ---- 0 0 '---- / ~ 0 ~ ° ~ o ~, /, - NH NH // N 0 92 N ~ ° ~ "x ~ ° ~ ° ° ~ ° ° ~ J ~ '91 N NH O Compound 68 COOH COOH CDI C ~ H (Biodyne C filter)

[0283]

CA 02444926 2003-10-17 WO 02/090584 PCT / FR02 / 01542 138 The 3'-aminoacetophenone (14.5 g, 107 mmol) is dissolved in 50 ml of anhydrous DMF.

(Succinic anhydride (10.7 g, 107 mmol) is added and the mixture is left under stirring, under argon and at room temperature.

After 6 h, the solution is concentrated in vacuo and 50 ml of methanol are added.

The precipitate obtained is filtered off and washed with methanol and ether.

In this way 19.4 g (61%) of product 68 are obtained in the form of an off-white powder.

1H NMR (200 MHz, DMSO-d6): d - 2.5-2.6 (m, 7H); 7.45 (t, 1H); 7.64 (d, 1H); 7.83 (d, 1H); 8.19 (s, 1H); 10.16 (s, 1H); 12.12 (s, 1H).

Compound 69 5.07 g (22 mmol) of compound 68 is dissolved in 10 ml of anhydrous DMF, under argon.

Put on ice and add 5.00 g (32 mmol) of carbonyldiimidazole.

After 20 min, 20 ml (94.6 mmol) of 4,7,10-trioxatridecanediamine (EG3) are slowly added.

After reacting for 3 hours at room temperature, the DMF is evaporated off and the residue is taken up in 100 ml of CH2Cl2.

Extractions were made with saturated NaHCO3 and H2O, after which the organic phase was dried with anhydrous Na2SO4 and the solvent evaporated. 4.34 g (46%) of product 69 are thus obtained. 1H NMR (200 MHz, DMSO-d6): d-- I, 59 (m, 2H); 1.87 (m, 2H); 2.16 (s, 3H); 2.40 (m, 2H); 2.55 (m, 2H); 3.06 (m, 2H); 3.45 (m, 16H); 7 30 (t, 1H); 7.42 (d, 1H); 7.70 (d, 1H); 7.83 (t, 1H); 7.97 (s, 1H); 10.00 (s, 1H). Compound 91 A rectangle of 4 cm2 is cut from a sheet of Biodyne C filter (reference 60314; Pall

IDT Exhibit 1020.10080

Gelman Laboratory; Ann Arbor; Michigan; USA), introduced into a flask and brought into contact with 0.97 g (6 mmol) of carbonyldiimidazole (CDn in 3 ml of anhydrous DMF, on ice, under argon and with vigorous stirring. After 20 min, the solution is removed and the filter washed with DMF. An amount of 0.53 g of product 68 (1 mmol) in 3 ml of anhydrous DMF is then added, and the reaction is carried out overnight at room temperature. The solution is then removed and the filter is rinsed with ethanol, dried in vacuo and kept under argon.

[0284]

CA 02444926 2003-10-17 WO 02/090584 PCT / FR02 / 01542 139 Compound 92 The modified filter 91 is placed in a solution of 97 ml of hydrazine hydrate (2 mmol) in 4 ml of absolute ethanol.

The solution is refluxed for 5 h. After allowing to cool, the filter is washed with H2O, ethanol and ether, dried under vacuum and placed under argon. Then, 4 ml of anhydrous DMF and 86 mg of MnO 2 (1 mmol) are added, and the mixture is left to react with vigorous stirring. After 20 min, the solution is discarded, and the filter is rinsed with DMF and ether. The modified-diazomethyl 92 filter is stored under argon at a temperature of -19 to -31 ° C. Example 39.2: Biological tests The activated membrane is cut into small fragments of 2 mma which will be incubated for 30 minutes at room temperature in 25 ~, 1 of bacterial lysate of Mycobacterium tuberculosis, prepared by mechanical lysis, according to the same technique and the same final concentration as in Example 37, and 375 ~ .1 pure water (Sigma). The membrane is then placed at 65 ° C for 60 min in 100 mL of washing buffer (5 Formamide (Sigma), 1X SSPE (Perbio), 0.01% triton X-100) in order to remove the non-specific nucleic acids adsorbed. on the membrane, then the latter is stored in 1 ml of pure water before amplification.

PCR is carried out as described in paragraph 5.1, the reaction volume being completed with a sufficient quantity of pure water. At the same time, checks are carried out following the same process with membranes that cannot covalently bind the nucleic acids ~ Unmodified Biodyne C membrane (A membrane), ~ Biodyne membrane chemically modified according to the process described, but not treated with 3o Anhydrous DMF and MnO2; this control makes it possible to check the behavior of the membrane in the absence of diazomethyl groups (membrane B), and

[0285]

CA 02444926 2003-10-17 WO 02/090584 PCT / FR02 / 01542 140 ~ Biodyne C membrane not modified, but treated for 20 min with vigorous stirring with anhydrous DMF and Mn02, this control making it possible to verify that this last step does not modify not adsorption of DNA to the membrane (C membrane).

In order to check the absence of PCR inhibition caused by the treatment of the membranes, another fragment of the membranes A, B, C is amplified simultaneously with 25 w1 of bacterial lysate (tubes A ', B', C '). The PCR products are then quantified by Picogreen according to the protocol described by the supplier. Result Tests carried out Si nal rfu

[0286]

Membrane modifies 111

[0287]

Membrane not modified A 18

[0288]

Unmodified membrane, co-amplified with 25 260 ~, L of bacterial lysate (A ')

[0289]

Modified membrane not active (B) 26

[0290]

Membrane modifies inactive, co-amplifies 264 with 25 ~.

L of bacterial lysate (B ')

IDT Exhibit 1020.10081

[0291]

Unmodified membrane having undergone activation21 (C)

[0292]

Unmodified membrane activated, 268 co-amplified with 25 ~, L of bacterial lysate (C ')

[0293]

Table 24: Quantification of the DNA obtained by PCR from DNA of bacterial lysate captured on a solid support These results of Table 24 indicate that it is possible to covalently capture on a solid support, nucleic acids, obtained from lysate, using diazomethyl chemistry.

The amplification observed is not due to non-specific adsorption of DNA to the membrane.

On the other hand, it is observed with the controls carried out with the PCRs that the membranes do not cause inhibition of the amplification reaction.

[0294]

CA 02444926 2003-10-17 WO 02/090584 PCT / FR02 / 01542 141 In order to check the nature of the product amplified on the membrane, the amplification product was analyzed by passage on a DNA chip, according to the protocol described above.

IDT Exhibit 1020.10082

Patent Translate
Powered by EPO and Google

Notice

This translation is machine-generated. It cannot be guaranteed that it is intelligible, accurate, complete, reliable or fit for specific purposes. Critical decisions, such as commercially relevant or financial decisions, should not be based on machine-translation output.

CLAIMS CA2444926A1

1.

Process for labeling and fragmentation of a single or double stranded deoxyribonucleic acid (DNA) comprising the following steps: - fragmenting the DNA by creating at least one abasic site on said DNA, - attaching a marker to at least one of the fragments by the intermediary of a labeling reagent, said reagent coupling covalently and predominantly to at least one phosphate of said fragment.

2.

Process according to Claim 1, characterized in that the fragmentation and the labeling are carried out in two stages.

3.

Process according to Claim 1, characterized in that the fragmentation and the labeling are carried out in one step.

4.

Process according to any one of Claims 1 to 3, characterized in that the abasic site is generated by the action of an acidic aqueous medium.

5.

Process according to Claim 4, characterized in that the pH of the acid medium is less than 5, preferably less than 4.

6.

Process according to Claim 5, characterized in that the acidic aqueous medium is a sodium formate buffer at a pH of approximately 3.

7.

Process according to any one of Claims 1 to 6, characterized in that the DNA contains at least one modified base capable of generating an abasic site.

8.

IDT Exhibit 1020.10083

Process according to Claim 7, characterized in that the modified base capable of generating an abasic site is chosen from 8bromo purine derivatives.

9.

Process, according to any one of claims 1 to 3, characterized in that the abasic site is generated by an alkylating agent or an oxidizing agent.

10.

Labeling and fragmentation process according to any one of Claims 1 to 9, characterized in that the labeling reagent comprises, as reactive function, a unit chosen from the following compounds: diazomethyl; alkyl halide; nitrosourea; spirocyclopropane; aziridine; epoxy; trifluorosulfonates.

11.

Process according to Claim 10, characterized in that the said labeling reagent is 5- (bromomethyl) fluorescein.

12.

Process according to Claim 10, characterized in that the labeling reagent is chosen from the compounds of formula (1): [image] in which ~ R1 represents H or an alkyl, substituted alkyl, aryl or substituted aryl group, and ~ Z includes a detectable marker.

13.

Method according to Claim 12, characterized by the fact that Z is [image]

14.

Process according to Claim 12, characterized in that the labeling reagent is chosen from the compounds of formula (2) [image] in which. R1 represents H or an alkyl, aryl or substituted aryl group,. R2 is a detectable marker,. L is a linking arm comprising a linear sequence of at least two covalent bonds and n is equal to 0 or 1, and. Z is chosen from: [images] in which: ~ R3 and R4 represent independently of one another: H, No2, Cl, Br, F, I, OR, SR, NR2, R, NHCOR, CONHR, COOR with R = alkyl or aryl, and ~ -YX- represents -CONH-, -NHCO-, -CH2o-, -CH2S-.

15.

Process according to Claim 14, characterized in that the labeling reagent has the formula (3): [image] in which:. R1 represents H or an alkyl, aryl or substituted aryl group,. R2 represents a detectable marker,. L is a linking arm comprising a linear sequence of at least 2 covalent bonds and n an integer equal to 0 or 1,. R3 and R4 represent independently of each other: H, NO2, Cl, Br, F, I, OR, SR, NR2, R, NHCOR, CONHR, COOR with R = alkyl or aryl, and. -Y-X- represents -CONH-, -NHCO-, -CH2O-, -CH2S-.

16.

A method according to any one of claims 14 or 15, characterized in that R2 is a fluorescent compound or a hapten.

17.

Process according to any one of Claims 10 to 16, characterized in that the said labeling reagent is soluble in a solvent miscible with water.

18.

Method for detecting a single or double stranded target deoxyribonucleic acid (DNA) comprising the following steps:. fragment and label said DNA, according to any one of the methods according to claims 1 to 17,. hybridize the labeled fragments on at

IDT Exhibit 1020.10084

least one sufficiently specific nucleic acid probe for the target nucleic acid, and ~ detect the hybrid formed by means of the marker.

19.

A method of detecting a double-stranded target deoxyribonucleic acid (DNA), according to claim 18, further comprising a denaturing step after fragmentation and labeling.

20.

Process according to Claim 19, characterized in that the fragmentation, the labeling and the denaturation are carried out in a single step.

21.

A method of detecting a target nucleic acid comprising the following steps. enzymatically amplifying the target nucleic acid to generate a multitude of double-stranded DNA amplicons, fragmenting and labeling said DNA amplicons, according to any one of the methods according to claims 1 to 17,. hybridizing the labeled fragments on at least one sufficiently specific nucleic acid probe for the target nucleic acid, and. detecting the hybrid formed through the label.

22.

A method according to claim 21 further comprising a denaturing step after the fragmentation and labeling step.

23.

Process according to Claim 22, characterized in that the fragmentation, the labeling and the denaturation are carried out in a single step.

24.

Method of demonstrating a polymorphism distributed in predetermined positions or not of a target nucleic acid by the presence of a plurality of deletions and / or insertions and / or mutations in the sequence of said target nucleic acid with respect to a so-called reference sequence comprising the following steps:. have a target DNA comprising all of the polymorphism to be studied, said DNA optionally being generated by an enzymatic amplification technique. . fragmenting and labeling said DNA by a method according to any one of claims 1 to 17.. hybridizing said fragments on a plurality of nucleic acid probes called capture probes, the plurality of capture probes being fixed on a solid support and the plurality of capture probes covering as a whole at least the polymorphism to be studied. . detecting the hybrids formed between the labeled fragments and at least part of the nucleic acid probes by means of the label and deducing therefrom the polymorphism of the target DNA.

25.

A method according to claim 24, further comprising a denaturing step after the fragmentation and labeling step.

26.

Process according to Claim 25, characterized in that the fragmentation, the labeling and the denaturation are carried out in a single step.

27.

Process, according to any one of Claims 24 to 26, characterized in that the solid support comprises at least ten (10) nucleic acid probes of different sequences, advantageously at least four hundred (400), preferably at least one thousand (1000).

28.

IDT Exhibit 1020.10085

Use of a labeled nucleic acid fragment, according to the method of any one of claims 1 to 17, as a probe for detecting a target nucleic acid.

IDT Exhibit 1020.10086

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
28 December 2006 (28.12.2006)

PCT

(10) International Publication Number
WO 2006/137733 A1

(51) International Patent Classification:
C12Q 1/68 (2006.01)

(21) International Application Number:
PCT/NL2006/000311

(22) International Filing Date: 23 June 2006 (23.06.2006)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

60/693,053	23 June 2005 (23.06.2005)	US
06075104.7	16 January 2006 (16.01.2006)	EP
60/759,034	17 January 2006 (17.01.2006)	US

(71) Applicant *(for all designated States except US)*: **KEYGENE N.V.** [NL/NL]; 90, Agro Business Park, NL-6708 PW Wageningen (NL).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: **VAN EIJK, Michael, Josephus, Theresia** [NL/NL]; 12, Pastoor Strijboschstraat, NL-5373 EJ Herpen (NL). **VAN DER POEL, Henricus, Johannes, Adam** [NL/NL]; 21A, Beukenlaan, NL-6871 CK Renkum (NL).

(74) Agent: **DE LANG, R.-J.**; Exter Polak & Charlouis B.V., P.O. Box 3241, NL-2280 GE Rijswijk (NL).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LV, LY, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LT, LU, LV, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *with international search report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: STRATEGIES FOR HIGH THROUGHPUT IDENTIFICATION AND DETECTION OF POLYMORPHISMS

(57) Abstract: The invention relates to a method for the high throughput identification of single nucleotide polymorphisms by performing a complexity reduction on two or more samples to yield two or more libraries, sequencing at least part of the libraries, aligning the identified sequences and determining any putative single nucleotide polymorphisms, confirming any putative single nucleotide polymorphism, generating detection probes for the confirmed single nucleotide polymorphisms, subjection a test sample to the same complexity reduction to provide a test library and screen the test library for the presence or absence of the single nucleotide polymorphisms using the detection probe.

WO 2006/137733 A1

Title: Strategies for high throughput identification and detection of polymorphisms

Technical field

The present invention relates to the fields of molecular biology and genetics. The invention relates to rapid identification of multiple polymorphisms in a nucleic acid sample. The identified polymorphisms may be used for development of high-throughput screening systems for polymorphisms in test samples.

Background of the invention

Exploration of genomic DNA has long been desired by the scientific, in particular medical, community. Genomic DNA holds the key to identification, diagnosis and treatment of diseases such as cancer and Alzheimer's disease. In addition to disease identification and treatment, exploration of genomic DNA may provide significant advantages in plant and animal breeding efforts, which may provide answers to food and nutrition problems in the world.

Many diseases are known to be associated with specific genetic components, in particular with polymorphisms in specific genes. The identification of polymorphisms in large samples such as genomes is at present a laborious and time-consuming task. However, such identification is of great value to areas such as biomedical research, developing pharmacy products, tissue typing, genotyping and population studies.

Summary of the invention

The present invention provides for a method of efficiently identifying and reliably detecting polymorphisms in a complex, e.g. very large, nucleic acid sample (e.g. DNA or RNA) in a rapid and economical manner using a combination of high-throughput methods.

This integration of high-throughput methods together provide a platform that is particularly suited for the rapid and reliable identification and detection of polymorphisms in highly complex nucleic acid samples wherein conventional identification and mapping of polymorphisms would be laborious and time-consuming.

One of the things the present inventors have found is a solution for the identification of polymorphisms, preferably Single Nucleotide Polymorphisms, but likewise for (micro)satellites and/or indels, in particular in large genomes. The method is unique in its applicability to large and small genomes alike, but provides particular advantages for large genomes, in particular polyploidal species.

To identify SNPs (and subsequently detect the identified SNPs) there are several possibilities available in the art. In a first option the whole genome can be sequenced, and this can be done so for several individuals. This is mostly a theoretical exercise as this is cumbersome and expensive and, despite the rapid development of technology simply not feasible to do for every organism, especially the ones with larger genomes. Second option is to use available (fragmented) sequence information, such as EST libraries. This allows the generation of PCR primers, resequencing and comparison between individuals. Again, this requires initial sequence information that is not available or only in a limited amount. Furthermore separate PCR-assays have to be developed for each region which adds enormously to costs and development time.

The third option is to limit one self to part of the genome for each individual. The difficulty resides that the provided part of the genome must be the same for different individuals in order to provide comparable result for successful SNP identification. The present inventors now have solved this dilemma by integration of highly reproducible methods for selecting part of genome with high throughput sequencing for the identification of polymorphisms integrated with sample preparation and high throughput identification platforms. The present invention accelerates the process of polymorphism discovery and uses the same elements in the subsequent process for the exploitation of the discovered polymorphisms to allow for effective and reliable high throughput genotyping.

Further envisaged applications of the method of the present invention include screening enriched microsatellite libraries, performing transcript profiling cDNA-AFLP (digital Northern), sequencing of complex genomes, EST library sequencing (on whole cDNA or cDNA-AFLP), microRNA discovery (sequencing of small insert libraries), Bacterial Artificial Chromosome (BAC)(contig) sequencing, Bulked Segregant analysis approach AFLP/cDNA-AFLP, routine detection

of AFLP fragments, e.g. for marker-assisted back-crosses (MABC), etcetera.

Definitions

In the following description and examples a number of terms are used. In order to provide a clear and consistent understanding of the specification and claims, including the scope to be given such terms, the following definitions are provided. Unless otherwise defined herein, all technical and scientific terms used have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. The disclosures of all publications, patent applications, patents and other references are incorporated herein in their entirety by reference.

Polymorphism: polymorphism refers to the presence of two or more variants of a nucleotide sequence in a population. A polymorphism may comprise one or more base changes, an insertion, a repeat, or a deletion. A polymorphism includes e.g. a simple sequence repeat (SSR) and a single nucleotide polymorphism (SNP), which is a variation, occurring when a single nucleotide: adenine (A), thymine (T), cytosine (C) or guanine (G) - is altered. A variation must generally occur in at least 1% of the population to be considered a SNP. SNPs make up e.g. 90% of all human genetic variations, and occur every 100 to 300 bases along the human genome. Two of every three SNPs substitute Cytosine (C) with Thymine (T). Variations in the DNA sequences of e.g. humans or plants can affect how they handle diseases, bacteria, viruses, chemicals, drugs, etc.

Nucleic acid: a nucleic acid according to the present invention may include any polymer or oligomer of pyrimidine and purine bases, preferably cytosine, thymine, and uracil, and adenine and guanine, respectively (*See* Albert L. Lehninger, *Principles of Biochemistry*, at 793-800 (Worth Pub. 1982) which is herein incorporated by reference in its entirety for all purposes). The present invention contemplates any deoxyribonucleotide, ribonucleotide or peptide nucleic acid component, and any chemical variants thereof, such as methylated, hydroxymethylated or glycosylated forms of these bases, and the like. The polymers or oligomers may be heterogenous or homogenous in composition, and may be isolated from naturally occurring sources or may be artificially or synthetically produced. In addition, the nucleic acids may be DNA or RNA, or a mixture thereof, and may exist

permanently or transitionally in single-stranded or double-stranded form, including homoduplex, heteroduplex, and hybrid states.

Complexity reduction: the term complexity reduction is used to denote a method wherein the complexity of a nucleic acid sample, such as genomic DNA, is reduced by the generation of a subset of the sample. This subset can be representative for the whole (i.e. complex) sample and is preferably a reproducible subset. Reproducible means in this context that when the same sample is reduced in complexity using the same method, the same, or at least comparable, subset is obtained. The method used for complexity reduction may be any method for complexity reduction known in the art. Examples of methods for complexity reduction include for example AFLP® (Keygene N.V., the Netherlands; see e.g. EP 0 534 858), the methods described by Dong (see e.g. WO 03/012118, WO 00/24939), indexed linking (Unrau et al., vide infra), etc. The complexity reduction methods used in the present invention have in common that they are reproducible. Reproducible in the sense that when the same sample is reduced in complexity in the same manner, the same subset of the sample is obtained, as opposed to more random complexity reduction such as microdissection or the use of mRNA (cDNA) which represents a portion of the genome transcribed in a selected tissue and for its reproducibilty is depending on the selection of tissue, time of isolation etc..

Tagging: the term tagging refers to the addition of a tag to a nucleic acid sample in order to be able to distinguish it from a second or further nucleic acid sample. Tagging can e.g. be performed by the addition of a sequence identifier during complexity reduction or by any other means known in the art. Such sequence identifier can e.g. be a unique base sequence of varying but defined length uniquely used for identifying a specific nucleic acid sample. Typical examples thereof are for instance ZIP sequences. Using such tag, the origin of a sample can be determined upon further processing. In case of combining processed products originating from different nucleic acid samples, the different nucleic acid samples should be identified using different tags.

Tagged library: the term tagged library refers to a library of tagged nucleic acid.

Sequencing: The term sequencing refers to determining the order of nucleotides (base sequences) in a nucleic acid sample, e.g. DNA or RNA.

Aligning and alignment: With the term "aligning" and "alignment" is meant the comparison of two or more nucleotide sequence based on the presence of short or long stretches of identical or similar nucleotides. Several methods for alignment of nucleotide sequences are known in the art, as will be further explained below.

Detection probes: The term "detection probes" is used to denote probes designed for detecting a specific nucleotide sequence, in particular sequences containing one or more polymorphisms.

High-throughput screening: High-throughput screening, often abbreviated as HTS, is a method for scientific experimentation especially relevant to the fields of biology and chemistry. Through a combination of modern robotics and other specialised laboratory hardware, it allows a researcher to effectively screen large amounts of samples simultaneously.

Test sample nucleic acid: The term "test sample nucleic acid" is used to indicate a nucleic acid sample that is investigated for polymorphisms using the method of the present invention.

Restriction endonuclease: a restriction endonuclease or restriction enzyme is an enzyme that recognizes a specific nucleotide sequence (target site) in a double-stranded DNA molecule, and will cleave both strands of the DNA molecule at every target site.

Restriction fragments: the DNA molecules produced by digestion with a restriction endonuclease are referred to as restriction fragments. Any given genome (or nucleic acid, regardless of its origin) will be digested by a particular restriction endonuclease into a discrete set of restriction fragments. The DNA fragments that result from restriction endonuclease cleavage can be further used in a variety of techniques and can for instance be detected by gel electrophoresis.

Gel electrophoresis: in order to detect restriction fragments, an analytical method for fractionating double-stranded DNA molecules on the basis of size can be required. The most commonly used technique for achieving such fractionation is (capillary) gel electrophoresis. The rate at which DNA fragments move in such gels depends on their molecular weight; thus, the distances travelled decrease as the fragment lengths increase. The DNA fragments fractionated by gel electrophoresis can be visualized directly by a staining procedure e.g. silver staining or staining using ethidium bromide, if the number of fragments included in the pattern is

sufficiently small. Alternatively further treatment of the DNA fragments may incorporate detectable labels in the fragments, such as fluorophores or radioactive labels.

Ligation: the enzymatic reaction catalyzed by a ligase enzyme in which two double-stranded DNA molecules are covalently joined together is referred to as ligation. In general, both DNA strands are covalently joined together, but it is also possible to prevent the ligation of one of the two strands through chemical or enzymatic modification of one of the ends of the strands. In that case the covalent joining will occur in only one of the two DNA strands.

Synthetic oligonucleotide: single-stranded DNA molecules having preferably from about 10 to about 50 bases, which can be synthesized chemically are referred to as synthetic oligonucleotides. In general, these synthetic DNA molecules are designed to have a unique or desired nucleotide sequence, although it is possible to synthesize families of molecules having related sequences and which have different nucleotide compositions at specific positions within the nucleotide sequence. The term synthetic oligonucleotide will be used to refer to DNA molecules having a designed or desired nucleotide sequence.

Adaptors: short double-stranded DNA molecules with a limited number of base pairs, e.g. about 10 to about 30 base pairs in length, which are designed such that they can be ligated to the ends of restriction fragments. Adaptors are generally composed of two synthetic oligonucleotides which have nucleotide sequences which are partially complementary to each other. When mixing the two synthetic oligonucleotides in solution under appropriate conditions, they will anneal to each other forming a double-stranded structure. After annealing, one end of the adaptor molecule is designed such that it is compatible with the end of a restriction fragment and can be ligated thereto; the other end of the adaptor can be designed so that it cannot be ligated, but this need not be the case (double ligated adaptors).

Adaptor-ligated restriction fragments: restriction fragments that have been capped by adaptors.

Primers: in general, the term primers refers to a DNA strand which can prime the synthesis of DNA. DNA polymerase cannot synthesize DNA *de novo* without primers: it can only extend an existing DNA strand in a reaction in which the complementary strand is used as a template to direct the order of nucleotides to be

assembled. We will refer to the synthetic oligonucleotide molecules which are used in a polymerase chain reaction (PCR) as primers.

DNA amplification: the term DNA amplification will be typically used to denote the *in vitro* synthesis of double-stranded DNA molecules using PCR. It is noted that other amplification methods exist and they may be used in the present invention without departing from the gist.

Detailed description of the invention

The present invention provides for a method for identifying one or more polymorphisms, said method comprising the steps of:

a) providing a first nucleic acid sample of interest;
b) performing a complexity reduction on the first nucleic acid sample of interest to provide a first library of the first nucleic acid sample;
c) consecutively or simultaneously performing steps a) and b) with a second or further nucleic acid sample of interest to obtain a second or further library of the second or further nucleic acid sample of interest;
d) sequencing of at least a portion of the first library and the second or further library;
e) aligning the sequences obtained in step d);
f) determining one or more polymorphisms between the first nucleic acid sample and second or further nucleic acid sample in the alignment of step e);
g) using the one or more polymorphisms determined in step f) to design one or more detection probes;
h) providing a test sample nucleic acid of interest;
i) performing the complexity reduction of step b) on the test sample nucleic acid of interest to provide a test library of the test sample nucleic acid;
j) subjecting the test library to high-throughput screening to identify the presence, absence or amount of the polymorphisms determined in step f) using the one or more detection probes designed in step g).

In step a), a first nucleic acid sample of interest is provided. Said first nucleic acid sample of interest is preferably a complex nucleic acid sample such as total genomic DNA or a cDNA library. It is preferred that the complex nucleic acid sample is total genomic DNA.

In step b), a complexity reduction is performed on the first nucleic acid sample of interest to provide a first library of the first nucleic acid sample.

In one embodiment of the invention, the step of complexity reduction of the nucleic acid sample comprises enzymatically cutting the nucleic acid sample in restriction fragments, separating the restriction fragments and selecting a particular pool of restriction fragments. Optionally, the selected fragments are then ligated to adaptor sequences containing PCR primer templates/binding sequences.

In one embodiment of complexity reduction, a type IIs endonuclease is used to digest the nucleic acid sample and the restriction fragments are selectively ligated to adaptor sequences. The adaptor sequences can contain various nucleotides in the overhang that is to be ligated and only the adaptor with the matching set of nucleotides in the overhang is ligated to the fragment and subsequently amplified. This technology is depicted in the art as 'indexing linkers'. Examples of this principle can be found inter alia in Unrau P. and Deugau K.V. (1994) Gene 145:163-169.

In another embodiment, the method of complexity reduction utilizes two restriction endonucleases having different target sites and frequencies and two different adaptor sequences.

In another embodiment of the invention, the step of complexity reduction comprises performing an Arbitrarily Primed PCR upon the sample.

In yet another embodiment of the invention, the step of complexity reduction comprises removing repeated sequences by denaturing and reannealing the DNA and then removing double-stranded duplexes.

In another embodiment of the invention, the step of complexity reduction comprises hybridizing the nucleic acid sample to a magnetic bead which is bound to an oligonucleotide probe containing a desired sequence. This embodiment may further comprise exposing the hybridized sample to a single strand DNA nuclease to remove the single-stranded DNA, ligating an adaptor sequence containing a Class IIs restriction enzyme to release the magnetic bead. This embodiment may or may not comprise amplification of the isolated DNA sequence. Furthermore, the adaptor sequence may or may not be used as a template for the PCR oligonucleotide primer. In this embodiment, the adaptor sequence may or may not contain a sequence identifier or tag.

In another embodiment, the method of complexity reduction comprises exposing the DNA sample to a mismatch binding protein and digesting the sample with a 3′ to 5′ exonuclease and then a single strand nuclease. This embodiment may or may not include the use of a magnetic bead attached to the mismatch binding protein.

In another embodiment of the present invention, complexity reduction comprises the CHIP method as described herein elsewhere or the design of PCR primers directed against conserved motifs such as SSRs, NBS regions (nucleotide biding regions), promoter/enhancer sequences, telomer consensus sequences, MADS box genes, ATP-ase gene families and other gene families.

In step c), steps a) and b) are consecutively or simultaneously performed with a second or further nucleic acid sample of interest to obtain a second or further library of the second or further nucleic acid sample of interest. Said second or further nucleic acid sample of interest is preferably also a complex nucleic acid sample such as total genomic DNA. It is preferred that the complex nucleic acid sample is total genomic DNA. It is also preferred that said second or further nucleic acid sample is related to the first nucleic acid sample. The first nucleic acid sample and the second or further nucleic acid may for example be different lines of a plant, such as different pepper lines, or different varieties. Steps a) en b) may be performed for merely a second nucleic acid sample of interest, but may also additionally be performed for a third, fourth, fifth, etc. nucleic acid sample of interest.

It is to be noted that the method according to the present invention will be most useful when complexity reduction is performed using the same method and under substantially the same, preferably identical, conditions for the first nucleic acid sample and the second or further nucleic acid sample. Under such conditions, similar (comparable) fractions of the (complex) nucleic acid samples will be obtained.

In step d), at least a portion of the first library and of the second or further library is sequenced. It is preferred that the amount of overlap of sequenced fragments from the first library and second or further library is at least 50%, more preferably at least 60%, yet more preferably at least 70%, even more preferably at least 80%, yet more preferably at least 90%, and most preferably at least 95%.

The sequencing may in principle be conducted by any means known in the art, such as the dideoxy chain termination method. It is however preferred that the sequencing is performed using high-throughput sequencing methods, such as the methods disclosed in WO 03/004690, WO 03/054142, WO 2004/069849, WO 2004/070005, WO 2004/070007, and WO 2005/003375 (all in the name of 454 Corporation), by Seo et al. (2004) Proc. Natl. Acad. Sci. USA 101:5488-93, and technologies of Helios, Solexa, US Genomics, etcetera, which are herein incorporated by reference. It is most preferred that sequencing is performed using the apparatus and/or method disclosed in WO 03/004690, WO 03/054142, WO 2004/069849, WO 2004/070005, WO 2004/070007, and WO 2005/003375 (all in the name of 454 Corporation), which are herein incorporated by reference. The technology described allows sequencing of 40 million bases in a single run and is 100 times faster and cheaper than competing technology. The sequencing technology roughly consists of 4 steps: 1) fragmentation of DNA and ligation of specific adaptor to a library of single-stranded DNA (ssDNA); 2) annealing of ssDNA to beads and emulsification of the beads in water-in-oil microreactors; 3) deposition of DNA carrying beads in a PicoTiterPlate®; and 4) simultaneous sequencing in 100,000 wells by generation of a pyrophosphate light signal. The method will be explained in more detail below.

In step e), the sequences obtained in step d) are aligned to provide an alignment. Methods of alignment of sequences for comparison purposes are well known in the art. Various programs and alignment algorithms are described in: Smith and Waterman (1981) Adv. Appl. Math. 2:482; Needleman and Wunsch (1970) J. Mol. Biol. 48:443; Pearson and Lipman (1988) Proc. Natl. Acad. Sci. USA 85:2444; Higgins and Sharp (1988) Gene 73:237-244; Higgins and Sharp (1989) CABIOS 5:151-153; Corpet et al. (1988) Nucl. Acids Res. 16:10881-90; Huang et al. (1992) Computer Appl. in the Biosci. 8:155-65; and Pearson et al. (1994) Meth. Mol. Biol. 24:307-31, which are herein incorporated by reference . Altschul et al. (1994) Nature Genet. 6:119-29 (herein incorporated by reference) present a detailed consideration of sequence alignment methods and homology calculations.

The NCBI Basic Local Alignment Search Tool (BLAST) (Altschul et al., 1990) is available from several sources, including the National Center for Biological Information (NCBI, Bethesda, Md.) and on the Internet, for use in connection with the sequence analysis programs blastp, blastn, blastx, tblastn and tblastx. It can be accessed at

<http://www.ncbi.nlm.nih.gov/BLAST/>. A description of how to determine sequence identity using this program is available at <http://www.ncbi.nlm.nih.gov/BLAST/blast_help.html>. A further application can be in microsatellite mining (see Varshney et al. (2005) Trends in Biotechn. 23(1):48-55.

Typically, the alignment is performed on the sequence data that have been trimmed for the adaptors/primer and/or identifiers, i.e. using only the sequence data from the fragments that originate from the nucleic acid sample. Typically, the sequence data obtained are used for identifying the origin of the fragment (i.e. from which sample), the sequences derived from the adaptor and/or identifier are removed from the data and alignment is performed on this trimmed set.

In step f), one or more polymorphisms are determined between the first nucleic acid sample and second or further nucleic acid sample in the alignment. The alignment can be made such that the sequences derived from the first nucleic acid sample and the second or further nucleic acid sample can be compared. Differences can then be identified reflecting polymorphisms.

In step g) the one or more polymorphisms determined in step g) are used to design detection probes, for example for detection by hybridization on DNA chips or a beads-based detection platform. The detection probes are designed such that a polymorphism is reflected therein. In case of single nucleotide polymorphisms (SNPs) the detection probes typically contain the variant SNP alleles at the central position such as to maximize allele discrimination. Such probes can advantageously be used to screen test samples having a certain polymorphism. The probes can be synthesized using any method known in the art. The probes are typically designed such that they are suitable for high throughput screening methods.

In step h), a test sample nucleic acid of interest is provided. The test sample nucleic acid may be any sample, but is preferably another line or variety to be mapped for polymorphisms. Commonly, a collection of test samples representing the germplasm of the organisms studied is used to validate experimentally that the (SN) polymorphism is genuine and detectable and to calculate allele frequencies of the observed alleles. Optionally, samples of a genetic mapping population are included in the validation step in order to determine the genetic map position of the polymorphism too.

In step i), the complexity reduction of step b) is performed on the test sample nucleic acid of interest to provide a test library of

the test sample nucleic acid. It is highly preferred that throughout the method according to the present invention the same method for complexity reduction is used using substantially the same, preferably identical, conditions, thus covering a similar fraction of the sample. It is however not required that a tagged test library is obtained, although a tag may be present on the fragments in the test library.

In step j) the test library is subjected to high-throughput screening to identify the presence, absence or amount of the polymorphisms determined in step f) using the detection probes designed in step g). One skilled in the art knows several methods for high-throughput screening using probes. It is preferred that one or more probes designed using the information obtained in step g) are immobilized on an array, such as a DNA chip, and that such array is subsequently contacted with the test library under hybridizing conditions. DNA fragments within the test library complementary to one or more probes on the array will under such conditions hybridize to such probes, and can thus be detected. Other methods of high-throughput screening are also envisaged within the scope of the present invention, such as immobilization of the test library obtained in step j) and contacting of said immobilized test library with the probes designed in step h) under hybridizing conditions.

Another high-throughput sequencing screening technique is provided *inter alia* by Affymetrix using chip-based detection of SNPs and bead technology provided by Illumina.

In an advantageous embodiment, step b) in the method according to the present invention further comprises the step of tagging of the library to obtain a tagged library, and said method further comprising step c1) of combining the first tagged library and second or further tagged library to obtain a combined library.

It is preferred that the tagging is performed during the complexity reduction step as to reduce the amount of steps required to obtain the first tagged library of the first nucleic acid sample. Such simultaneous tagging can e.g. be achieved by AFLP, using adaptors that comprise a unique (nucleotide) identifier for each sample.

The tagging is intended to distinguish between samples of different origin, e.g. obtained from different plant lines, when the libraries of two or more nucleic acid samples are combined to obtain a combination library. Thus, preferably different tags are used for

preparing the tagged libraries of the first nucleic acid sample and the second or further nucleic acid sample. When for example five nucleic acid samples are used, it is intended to obtain five differently tagged libraries, the five different tags denoting the respective original samples.

The tag may be any tag known in the art for distinguishing nucleic acid samples, but is preferably a short identifier sequence. Such identifier sequence can e.g. be a unique base sequence of varying length used to indicate the origin of the library obtained by complexity reduction.

In a preferred embodiment, the tagging of the first library and the second or further library is performed using different tags. As discussed above, it is preferred that each library of a nucleic acid sample is identified by its own tag. The test sample nucleic acid does not need to be tagged.

In a preferred embodiment of the invention, the complexity reduction is performed by means of AFLP® (Keygene N.V., the Netherlands; *see* e.g. EP 0 534 858 and Vos et al. (1995). AFLP: a new technique for DNA fingerprinting, *Nucleic Acids Research*, vol. 23, no. 21, 4407-4414, which are herein incorporated in their entirety by reference).

AFLP is a method for selective restriction fragment amplification. AFLP does not any prior sequence information and can be performed on any starting DNA. In general, AFLP comprises the steps of:

(a) digesting a nucleic acid, in particular a DNA or cDNA, with one or more specific restriction endonucleases, to fragment the DNA into a corresponding series of restriction fragments;

(b) ligating the restriction fragments thus obtained with a double-stranded synthetic oligonucleotide adaptor, one end of which is compatible with one or both of the ends of the restriction fragments, to thereby produce adapter-ligated, preferably tagged, restriction fragments of the starting DNA;

(c) contacting the adapter-ligated, preferably tagged, restriction fragments under hybridizing conditions with at least one oligonucleotide primer that contains at least one selective nucleotide at its 3'-end;

(d) amplifying the adapter-ligated, preferably tagged, restriction fragment hybridised with the primers by PCR or a similar technique so as to cause further elongation of the

hybridised primers along the restriction fragments of the starting DNA to which the primers hybridised; and

(e) detecting, identifying or recovering the amplified or elongated DNA fragment thus obtained.

AFLP thus provides a reproducible subset of adaptor-ligated fragments. Other suitable methods for complexity reduction are Chromatine Immuno Precipitation (ChiP). This means that nuclear DNA is isolated, whilst proteins such as transcription factors are linked to the DNA. With ChiP first an antibody is used against the protein, resulting in Ab-protein-DNA complex. By purifying this complex and precipitating it, DNA to which this protein binds is selected. Subsequently, the DNA can be used for library construction and sequencing. I.e., this is a method to perform a complexity reduction in a non-random fashion directed to specific functional areas; in the present example specific transcription factors.

One useful variant of the AFLP technology uses no selective nucleotides (i.c. +0/+0 primers) and is sometimes called linker PCR. This also provides for a very suitable complexity reduction.

For a further description of AFLP, its advantages, its embodiments, as well as the techniques, enzymes, adaptors, primers and further compounds and tools used therein, reference is made to US 6,045,994, EP-B-0 534 858, EP 976835 and EP 974672, WO01/88189 and Vos *et al.* Nucleic Acids Research, 1995, 23, 4407-4414, which are hereby incorporated in their entirety.

Thus, in a preferred embodiment of the method of the present invention, the complexity reduction is performed by

- digesting the nucleic acid sample with at least one restriction endonuclease to fragment it into restriction fragments;
- ligating the restriction fragments obtained with at least one double-stranded synthetic oligonucleotide adaptor having one end compatible with one or both ends of the restriction fragments to produce adapter-ligated restriction fragments;
- contacting said adapter-ligated restriction fragments with one or more oligonucleotide primers under hybridizing conditions; and
- amplifying said adapter-ligated restriction fragments by elongation of the one or more oligonucleotide primers,

wherein at least one of the one or more oligonucleotide primers include a nucleotide sequence having the same nucleotide sequence as the terminal parts of the strands at the ends of said adapter-ligated restriction fragments, including the nucleotides involved in the

formation of the target sequence for said restriction endonuclease and including at least part of the nucleotides present in the adaptors, wherein, optionally, at least one of said primers includes at its 3' end a selected sequence comprising at least one nucleotide located immediately adjacent to the nucleotides involved in the formation of the target sequence for said restriction endonuclease.

AFLP is a highly reproducible method for complexity reduction and is therefore particularly suited for the method according to the present invention.

In a preferred embodiment of the method according to the present invention, the adaptor or the primer comprises a tag. This is particularly the case for the actual identification of the polymorphisms, when it is important to distinguish between sequences derived from separate libraries. Incorporating an oligonucleotide tag in an adaptor or primer is very convenient as no additional steps are required to tag a library.

In another embodiment, the tag is an identifier sequence. As discussed above, such identifier sequence may be of varying length depending on the amount of nucleic acid samples to be compared. A length of about 4 bases (4^4 = 256 different tag sequences possible) is sufficient to distinguish between the origin of a limited number of samples (up to 256), although it is preferred that the tag sequences differ by more than one base between the samples to be distinguished. As needed, the length of the tag sequences can be adjusted accordingly.

In an embodiment, the sequencing is performed on a solid support such as a bead (*see* e.g. WO 03/004690, WO 03/054142, WO 2004/069849, WO 2004/070005, WO 2004/070007, and WO 2005/003375 (all in the name of 454 Corporation), which are herein incorporated by reference). Such sequencing method is particularly suitable for cheap and efficient sequencing of many samples simultaneously.

In a preferred embodiment, the sequencing comprises the steps of:

- annealing adapter-ligated fragments to beads, each bead being annealed with a single adapter-ligated fragment;
- emulsifying the beads in water-in-oil microreactors, each water-in-oil microreactor comprising a single bead;
- loading the beads in wells, each well comprising a single bead; and
- generating a pyrophosphate signal.

In the first step, sequencing adaptors are ligated to fragments within the combination library. Said sequencing adaptor includes at least a "key" region for annealing to a bead, a sequencing primer region and a PCR primer region. Thus, adapter-ligated fragments are obtained.

In a further step, adapter-ligated fragments are annealed to beads, each bead annealing with a single adapter-ligated fragment. To the pool of adapter-ligated fragments, beads are added in excess as to ensure annealing of one single adapter-ligated fragment per bead for the majority of the beads (Poisson distribution).

In a next step, the beads are emulsified in water-in-oil microreactors, each water-in-oil microreactor comprising a single bead. PCR reagents are present in the water-in-oil microreactors allowing a PCR reaction to take place within the microreactors. Subsequently, the microreactors are broken, and the beads comprising DNA (DNA positive beads) are enriched.

In a following step, the beads are loaded in wells, each well comprising a single bead. The wells are preferably part of a PicoTiter™Plate allowing for simultaneous sequencing of a large amount of fragments.

After addition of enzyme-carrying beads, the sequence of the fragments is determined using pyrosequencing. In successive steps, the picotiterplate and the beads as well as the enzyme beads therein are subjected to different deoxyribonucleotides in the presence of conventional sequencing reagents, and upon incorporation of a deoxyribonucleotide a light signal is generated which is recorded. Incorporation of the correct nucleotide will generate a pyrosequencing signal which can be detected.

Pyrosequencing itself is known in the art and described inter alia on www.biotagebio.com; www.pyrosequencing.com / tab technology The technology is further applied in e.g. WO 03/004690, WO 03/054142, WO 2004/069849, WO 2004/070005, WO 2004/070007, and WO 2005/003375 (all in the name of 454 Corporation), which are herein incorporated by reference.

The high-throughput screening of step k) is preferably performed by immobilization of the probes designed in step h) onto an array, followed by contacting of the array comprising the probes with a test library under hybridizing conditions. Preferably, the contacting step is performed under stringent hybridizing conditions (see Kennedy et al. (2003) Nat. Biotech.; published online 7

September 2003: 1-5). One skilled in the art is aware of suitable methods for immobilization of probes onto an array and of methods of contacting under hybridizing conditions. Typical technology that is suitable for this purpose is reviewed in Kennedy et al. (2003) Nat. Biotech.; published online 7 September 2003: 1-5.

One particular advantageous application is found in the breeding of polyploidal crops. By sequencing polyploidal crops with a high coverage, identifying SNPs and the various alleles and developing probes for allele-specific amplification significant progress can be made the breeding of polyploidal crops.

As part of the invention, it has been found that the combination of generating random selected subsets using selective amplification for a plurality of samples and high throughput sequencing technology presents certain complex problems that had to be solved for the further improvement of the herein described method for the efficient and high throughput identification of polymorphisms. More in detail, it has been found that when multiple (i.e. the first and the second or further) samples are combined in a pool after performing a complexity reduction, a problem occurs that many fragments appear to be derived from two samples, or, put differently, many fragments were identified that could not be allocated uniquely to one sample and thus could not be used in the process of identifying polymorphisms. This lead to a reduced reliability of the method and to less polymorphisms (SNPs, indels, SSRs) that could be adequately identified.

After careful and detailed analysis of the entire nucleotide sequence of the fragments that could not be allocated, it was found that those fragments contained two different tag-comprising adaptors and were probably formed between the generation of the complexity reduced samples and the ligation of the sequencing adaptors. The phenomenon is depicted as 'mixed tagging'. The phenomenon depicted as 'mixed tagging', as used herein, thus refers to fragments that contain a tag relating the fragment to one sample on side, whereas the opposite side of the fragment contains a tag relating the fragment to another sample. Thus one fragment appears to be derived from two samples (quod non). This leads to erroneous identification of polymorphisms and is hence undesirable.

It has been theorised that the formation of heteroduplex fragments between two samples lies afoot to this anomaly.

The solution to this problem has been found in a redesigning of the strategy for the conversion of samples of which the complexity is reduced to bead-annealed fragments that can be amplified prior to high throughput sequencing. In this embodiment, each sample is subjected to complexity reduction and optional purification. After that, each sample is rendered blunt (end-polishing) followed by ligation of the sequencing adapter that is capable of annealing to the bead. The sequencing adapter-ligated fragments of the samples are then combined and ligated to the beads for emulsion polymerisation and subsequent high-throughput sequencing.

As a further part of this invention, it was found that the formation of concatamers hampered the proper identification of polymorphisms. Concatamers have been identified as fragments that are formed after complexity reduction products have been 'blunted' or 'polished', for instance by T4 DNA polymerase, and instead of ligating to the adapters that allow annealing to the beads, ligate to each other, hereby creating concatamers, i.e. a concatamer is the result of the dimerisation of blunted fragments.

The solution to his problem was found in the use of certain specifically modified adapters. The amplified fragments obtained from the complexity reduction typically contain a 3'-A overhang due to the characteristics of certain preferred polymerases that do not have 3'-5' exonuclease proof-reading activity. The presence of such a 3'-A overhang is also the reason why fragments are blunted prior to adapter ligation. By providing an adapter that can anneal to a bead wherein the adapter contains a 3'-T overhang, it was found that both the problem of the 'mixed tags' and of the concatamers can be solved in one step. A further advantage of using these modified adapters is that the conventional 'blunting' step and the subsequent phosporylation step can be omitted.

Thus, in a further preferred embodiment, after the complexity reduction step of each sample, a step is performed on the amplified adapter-ligated restriction fragments obtained from the complexity reduction step, whereby to these fragments sequencing adapters are ligated, which sequencing adapters contain a 3'-T overhang and are capable of annealing to the beads.

It has further been found that, when the primers used in the complexity reduction step are phosporylated, the end-polishing (blunting) step and intermediate phosporylation prior to ligation can be avoided.

Thus, in a highly preferred embodiment of the invention, the invention relates to a method for identifying one or more polymorphisms, said method comprising the steps of:

a) providing a plurality of nucleic acid samples of interest;
b) performing a complexity reduction on each of the samples to provide a plurality of libraries of the nucleic acid samples, wherein the complexity reduction is performed by
- digesting each nucleic acid sample with at least one restriction endonuclease to fragment it into restriction fragments;
- ligating the restriction fragments obtained with at least one double-stranded, synthetic oligonucleotide adaptor having one end compatible with one or both ends of the restriction fragments to produce adapter-ligated restriction fragments;
- contacting said adapter-ligated restriction fragments with one or more phosporylated oligonucleotide primers under hybridising conditions; and
- amplifying said adapter-ligated restriction fragments by elongation of the one or more oligonucleotide primers, wherein at least one of the one or more oligonucleotide primers include a nucleotide sequence having the same nucleotide sequence as the terminal parts of the strands at the ends of said adapter-ligated restriction fragments, including the nucleotides involved in the formation of the target sequence for said restriction endonuclease and including at least part of the nucleotides present in the adaptors, wherein, optionally, at least one of said primers includes at its 3' end a selected sequence comprising at least one nucleotide located immediately adjacent to the nucleotides involved in the formation of the target sequence for said restriction endonuclease and wherein the adaptor and/or the primer contain a tag;

c) combining said libraries to a combined library;
d) ligating sequencing adapters capable of annealing to beads to the amplified adapter-capped fragments in the combined library, using an sequencing adapter carrying a 3'-T overhang and subjecting the bead-annealed fragments to emulsion polymerisation;

e) sequencing of at least a portion of the combined library;
f) aligning the sequences from each sample obtained in step e);
g) determining one or more polymorphisms between the plurality of nucleic acid samples in the alignment of step f);
h) using the one or more polymorphisms determined in step g) to design detection probes;
i) providing a test sample nucleic acid of interest;
j) performing the complexity reduction of step b) on the test sample nucleic acid of interest to provide a test library of the test sample nucleic acid;
k) subjecting the test library to high-throughput screening to identify the presence, absence or amount of the polymorphisms determined in step g) using the detection probes designed in step h).

Brief description of the drawings

Figure 1A shows a fragment according to the present invention annealed onto a bead ('454 bead') and the sequence of primer used for pre-amplification of the two pepper lines. 'DNA fragment' denotes the fragment obtained after digestion with a restriction endonuclease, 'keygene adaptor' denotes an adaptor providing an annealing site for the (phosphorylated) oligonucleotide primers used to generate a library, 'KRS' denotes an identifier sequence (tag), '454 SEQ. Adaptor' denotes a sequencing adaptor, and '454 PCR adaptor' denotes an adaptor to allow for emulsion amplification of the DNA fragment. The PCR adaptor allows for annealing to the bead and for amplification and may contain a 3'-T overhang.

Figure 1B shows a schematic primer used in the complexity reduction step. Such a primer generally comprises a recognition site region indicated as (2), a constant region that may include a tag section indicated as (1) and one or more selective nucleotides in a selective region indicated as (3) at the 3'-end thereof.).

Figure 2 shows DNA concentration estimation using 2% agarose gel-electrophoresis. S1 denotes PSP11; S2 denote PI201234. 50, 100, 250 and 500 ng denotes respectively 50 ng, 100 ng, 250 ng and 500 ng to estimate DNA amounts of S1 and S2. Fig 2C and 2D show DNA concentration determination using Nanodrop spectrophotometry.

Figure 3 shows the results of intermediate quality assessments of example 3.

Figure 4 shows flow charts of the sequence data processing pipeline, i.e. the steps taken from the generation of the sequencing data to the identification of putative SNPs, SSRs and indels, via steps of the removal of known sequence information in Trimming & Tagging resulting in trimmed sequence data which are clustered and assembled to yield contigs and singletons (fragments that cannot be assembled in a contig) after which putative polymorphisms can be identified and assessed. Figure 4B further elaborates on the process of polymorphisms mining.

Figure 5 addresses the problem of mixed tags and provides in panel 1 an example of a mixed tag, carrying tags associated with sample 1 (MS1) and sample 2 (MS2). Panel 2 provides a schematic explanation of the phenomenon. AFLP Restriction fragments derived from sample 1 (S1) and from sample 2 (S2) are ligated with adaptors ("Keygene adaptor") on both sides carrying sample specific tags S1 and S2. After amplification and sequencing, e Expected fragments are those with S1-S1 tags and S2-S2 tags. What additionally and unexpectedly is observed are also fragments that carry tags S1-S2 or S2-S1. Panel 3 explains the hypothesized cause of the generation of mixed tags whereby heteroduplex products are formed from fragments from samples 1 and 2. The heteroduplexes are subsequently, due to the 3'-5' exonuclease activity of T4 DNA polymerase or Klenow, rendered free from the 3'-protruding ends. During polymerization, the gaps are filled with nucleotides and the incorrect tag is introduced. This works for heteroduplexes of about the same length (top panel) but also for heteroduplexes of more varying length. Panel 4 provides on the right the conventional protocol leading to the formation of mixed tags and on the right the modified protocol.

Figure 6 addresses the problem of concatamer formation, whereby in panel 1 a typical example of a concatamer is given, whereby the various adapter and tag sections are underlined and with their origin (i.e. MS1, MS2, ES1 and ES2 corresponding respectively to a MseI restriction site- adapter from sample 1, MseI restriction site- adapter from sample 2, EcoRI restriction site- adapter from sample 1, EcoRI restriction site- adapter from sample 2). Panel 2 demonstrates the expected fragments carrying S1-S1 tags and S2-S2 tags and the observed but unexpected S1-S1-S2-S2, being a concatamer of a fragments from sample 1 and from sample 2. Panel 3 solution to avoid the generation of concatamers as well as mixed tags by introducing an overhang in the AFLP adaptors, modified sequencing adaptors and

omission of the end-polishing step when ligating sequencing adaptors. No concatamer formation is found because the ALP fragments can not ligate to each other and no mixed fragments occur as the end-polishing step is omitted. Panel 4 provides the modified protocol using modified adaptors to avoid concatamer formation as well as mixed tags.

Figure 7. Multiple alignment "10037_CL989contig2" of pepper AFLP fragment sequences, containing a putative single nucleotide polymorphism (SNP). Note that the SNP (indicated by an the black arrow) is defined by an A allele present in both reads of sample 1 (PSP11), denoted by the presence of the MS1 tag in the name of the top two reads, and a G allele present in sample 2 (PI201234), denoted by the presence of the MS2 tag in the name of the bottom two reads. Read names are shown on the left. The consensus sequence of this multiple alignment is (5'- 3'):
TAACACGACTTTGAACAAACCCAAACTCCCCCAATCGATTTCAAACCTAGAACA[A/G]TGTTGGTTTT
GGTGCTAACTTCAACCCCACTACTGTTTTGCTCTATTTTTG.

Figure 8A. Schematic representation of enrichment strategy for targeting simple sequence repeats (SSRs) in combination with high throughput sequencing for *de novo* SSR discovery.

Figure 8B: Validation of a G/A SNP in pepper using SNPWave detection. P1 = PSP11; P2 = PI201234. Eight RIL offspring are indicated by numbers 1-8.

Examples

Example 1

EcoRI/MseI restriction ligation mixture (1) was generated from genomic DNA of the pepper lines PSP-11 and PI20234. The restriction ligation mixture was 10 times diluted and 5 microliter of each sample was pre-amplified (2) with EcoRI +1(A) and MseI +1(C) primers (set I). After amplification the quality of the pre-amplification product of the two pepper samples was checked on a 1% agarose gel. The preamplification products were 20 times diluted, followed by a KRSEcoRI +1(A) and KRSMseI +2(CA) AFLP pre-amplification. The KRS (identifier)sections are underlined and the selective nucleotides are in bold at the 3'-end in the primersequence SEQ ID 1-4 below. After amplification the quality of the pre-amplification product of the two pepper samples was checked on a 1% agarose gel and by an EcoRI +3(A) and MseI +3(C) (3) AFLP fingerprint (4). The pre-amplification

products of the two pepper lines were separately purified on a QiagenPCR column (5). The concentration of the samples was measured on the nanodrop. A total of 5006.4 ng PSP-11 and 5006,4 ng PI20234 was mixed and sequenced.

Primer set I used for preamplification of PSP-11

E01LKRS1 5'-CGTCAGACTGCGTACCAATTC**A**-3' [SEQ ID 1]

M15KKRS1 5'-TGGTGATGAGTCCTGAGTAA**CA**-3' [SEQ ID 2]

Primer set II used for preamplification of PI20234

E01LKRS2 5'-CAAGAGACTGCGTACCAATTC**A**-3' [SEQ ID 3]

M15KKRS2 5'-AGCCGATGAGTCCTGAGTAA**CA**-3' [SEQ ID 4]

(1) EcoRI/MseI restriction ligation mixture

Restriction mix (40ul/sample)

DNA	6µl (±300ng)
ECoRI (5U)	0.1µl
MseI(2U)	0.05µl
5xRL	8µl
MQ	25.85µl
Totaal	40µl

Incubation during 1h. at 37°C

Addition of:

Ligation mix (10µl/sample)

10mM ATP	1µl
T4 DNA ligase	1µl
ECoRI adapt. (5pmol/µl)	1µl
MseI adapt. . (50pmol/µl)	1µl
5xRL	2µl
MQ	4µl
Totaal	10µl

Incubation during 3h. at 37°C

EcoRI-adaptor

91M35/91M36: *-CTCGTAGACTGCGTACC :91M35 [SEQ ID 5]

± bio CATCTGACGCATGGTTAA :91M36 [SEQ ID 6]

MseI-adaptor

92A18/92A19: 5-GACGATGAGTCCTGAG-3 :92A18 [SEQ ID 7]

3-TACTCAGGACTCAT-5 :92A19 [SEQ ID 8]

(2) Pre-amplification

Preamplification (A/C):

RL-mix (10x)	5µl
EcoRI-pr E01L(50ng/ul)	0.6µl
MseI-pr M02K(50ng/ul)	0.6µl
dNTPs (25mM)	0.16µl
Taq.pol.(5U)	0.08µl
10XPCR	2.0µl
MQ	11.56µl
Total	20µl/reaction

Pre-amplification thermal profile

Selective pre amplification was done in a reaction volume of 50µl. The PCR was performed in a PE GeneAmp PCR System 9700 and a **20** cycle profile was started with a 94°C denaturation step for 30 seconds, followed by an annealing step of 56°C for 60 seconds and an extension step of 72°C for 60 seconds.

EcoRI +1(A)[1]

E01 L 92R11: 5-AGACTGCGTACCAATTCA-3 [SEQ ID 9]

MseI +1(C)[1]

M02k 93E42: 5-GATGAGTCCTGAGTAAC-3 [SEQ ID 10]

Preamplification A/CA:

PA+1/+1-mix (20x)	:5µl
EcoRI-pr	:1.5µl
MseI-pr.	:1.5µl
dNTPs (25mM)	:0.4µl
Taq.pol.(5U)	:0.2µl
10XPCR	:5µl
MQ	:36.3µl
Total	:50µl

Selective pre amplification was done in a reaction volume of 50µl. The PCR was performed in a PE GeneAmp PCR System 9700 and a **30** cycle profile was started with a 94°C denaturation step for 30 seconds, followed by an annealing step of 56°C for 60 seconds and an extension step of 72°C for 60 seconds.

(3) KRSEcoRI +1(A)and KRSMseI +2(CA)[2]

05F212	E01LKRS1	CGTCAGACTGCGTACCAATTC**A**	-3' [SEQ ID 11]
05F213	E01LKRS2	CAAGAGACTGCGTACCAATTC**A**	-3' [SEQ ID 12]

05F214 M15KKRS1 TGGTGATGAGTCCTGAGTAA**CA** -3' [SEQ ID 13]

05F215 M15KKRS2 AGCCGATGAGTCCTGAGTAA**CA** -3' [SEQ ID 14]

selective nucleotides in bold and tags (KRS) underlined

Sample PSP11 : E01LKR**S1**/M15KKR**S1**

Sample PI120234 : E01LKR**S2**/M15KKR**S2**

(4) AFLP protocol

Selective amplification was done in a reaction volume of 20µl. The PCR was performed in a PE GeneAmp PCR System 9700. A **13** cycle profile was started with a 94°C denaturation step for 30 seconds, followed by an annealing step of 65°C for 30 seconds, with a touchdown phase in witch the annealing temperature was lowered 0.7°C in each cycle, and an extension step of 72°C for 60 seconds. This profile was followed by a 23 cycle profile with a 94°C denaturation step for 30 seconds, followed by an annealing step of 56°C for 30 seconds and an extension step of 72°C for 60 seconds.

EcoRI +3(**AAC**) and MseI +3(**CAG**)

E32 92S02: 5-GACTGCGTACCAATTC**AAC**-3 [SEQ ID 15]

M49 92G23: 5-GATGAGTCCTGAGTAA**CAG**-3 [SEQ ID 16]

(5) Qiagen column

Qiagen purification was performed according to the manufacturer's instruction: QIAquick® Spin Handbook (http://www1.qiagen.com/literature/handbooks/PDF/DNACleanupAndConcentration/QQ_Spin/1021422_HBQQSpin_072002WW.pdf)

Example 2: PEPPER

DNA from the *Pepper* lines PSP-11 and PI20234 was used to generate AFLP product by use of AFLP *Keygene Recognition Site* specific primers. (These AFLP primers are essentially the same as conventional AFLP primers, e.g. described in EP 0 534 858, and will generally contain a *recognition site region*, a constant region and one or more selective nucleotides in a selective region.

From the pepper lines PSP-11 or PI20234 150ng of DNA was digested with the restriction endonucleases *EcoRI* (5U/reaction) and *MseI* (2U/reaction) for 1 hour at 37°C following by inactivation for 10 minutes at 80°C. The obtained restriction fragments were ligated with double-stranded synthetic oligonucleotide adapter, one end of witch is compatible with one or both of the ends of the EcoRI and/or MseI restriction fragments. AFLP preamplification reactions (20µl/reaction) with +1/+1 AFLP primers were performed on 10 times

diluted restriction-ligation mixture. PCR profile:20*(30 s at 94°C + 60 s at 56°C + 120 s at 72°C). Additional AFLP reactions (50µl/reaction) with different +1 *EcoRI* and +2 *MseI* AFLP Keygene Recognition Site specific primers (Table below, tags are in bold, selective nucleotides are underlined.) were performed on 20 times diluted +1/+1 EcoRI/MseI AFLP preamplification product. PCR profile: 30*(30 s at 94°C + 60 s at 56°C + 120 s at 72°C). The AFLP product was purified by using the QIAquick PCR Purification Kit (QIAGEN) following the QIAquick® Spin Handbook 07/2002 page 18 and the concentration was measured with a NanoDrop® ND-1000 Spectrophotometer. A total of 5µg of +1/+2 PSP-11 AFLP product and 5µg of +1/+2 PI20234 AFLP product was put together and solved in 23.3µl TE. Finally a mixture with a concentration of 430ng/µl +1/+2 AFLP product was obtained.

Table

SEQ ID	PCR primer	Primer -3'	Pepper	AFLP reaction
[SEQ ID 17]	05F21	**CGTC**AGACTGCGTACCAATTCA	PSP-	1
[SEQ ID 18]	05F21	**TGGT**GATGAGTCCTGAGTAACA	PSP-	1
[SEQ ID 19]	05F21	**CAAG**AGACTGCGTACCAATTCA	PI2023	2
[SEQ ID 20]	05F21	**AGCC**GATGAGTCCTGAGTAACA	PI2023	2

Example 3 : Maize

DNA from the *Maize* lines B73 and M017 was used to generate AFLP product by use of AFLP *Keygene Recognition Site* specific primers. (These AFLP primers are essentially the same as conventional AFLP primers, e.g. described in EP 0 534 858, and will generally contain a *recognition site region*, a constant region and one or more selective nucleotides at the 3'-end thereof.).

DNA from the pepper lines B73 or M017 was digested with the restriction endonucleases *TaqI* (5U/reaction) for 1 hour at 65°C and *MseI* (2U/reaction) for 1 hour at 37°C following by inactivation for 10 minutes at 80°C. The obtained restriction fragments were ligated with double-stranded synthetic oligonucleotide adapter, one end of witch is compatible with one or both of the ends of the TaqI and/or MseI restriction fragments.

AFLP preamplification reactions (20µl/reaction) with +1/+1 AFLP primers were performed on 10 times diluted restriction-ligation

mixture. PCR profile:20*(30 s at 94°C + 60 s at 56°C + 120 s at 72°C). Additional AFLP reactions (50µl/reaction) with different +2 *TaqI* and *MseI* AFLP Keygene Recognition Site primers (Table below, tags are in bold, selective nucleotides are underlined.) were performed on 20 times diluted +1/+1 TaqI/MseI AFLP preamplification product. PCR profile: 30*(30 s at 94°C + 60 s at 56°C + 120 s at 72°C). The AFLP product was purified by using the QIAquick PCR Purification Kit (QIAGEN) following the QIAquick® Spin Handbook 07/2002 page 18 and the concentration was measured with a NanoDrop® ND-1000 Spectrophotometer. A total of 1.25µg of each different B73 +2/+2 AFLP product and 1.25µg of each different M017 +2/+2 AFLP product was put together and solved in 30µl TE. Finally a mixture with a concentration of 333ng/µl +2/+2 AFLP product was obtained.

Table

SEQ ID	PCR Primer	Primer sequence	Maize	AFLP Reaction
[SEQ ID 21]	05G360	**ACGT**GTAGACTGCGTACCGAAA	B73	1
[SEQ ID 22]	05G368	**ACGT**GATGAGTCCTGAGTAACA	B73	1
[SEQ ID 23]	05G362	**CGTA**GTAGACTGCGTACCGAAC	B73	2
[SEQ ID 24]	05G370	**CGTA**GATGAGTCCTGAGTAACA	B73	2
[SEQ ID 25]	05G364	**GTAC**GTAGACTGCGTACCGAAG	B73	3
[SEQ ID 26]	05G372	**GTAC**GATGAGTCCTGAGTAACA	B73	3
[SEQ ID 27]	05G366	**TACG**GTAGACTGCGTACCGAAT	B73	4
[SEQ ID 28]	05G374	**TACG**GATGAGTCCTGAGTAACA	B73	4
[SEQ ID 29]	05G361	**AGTC**GTAGACTGCGTACCGAAA	M017	5
[SEQ ID 30]	05G369	**AGTC**GATGAGTCCTGAGTAACA	M017	5
[SEQ ID 31]	05G363	**CATG**GTAGACTGCGTACCGAAC	M017	6
[SEQ ID 32]	05G371	**CATG**GATGAGTCCTGAGTAACA	M017	6
[SEQ ID 33]	05G365	**GAGC**GTAGACTGCGTACCGAAG	M017	7
[SEQ ID 34]	05G373	**GAGC**GATGAGTCCTGAGTAACA	M017	7
[SEQ ID 35]	05G367	**TGAT**GTAGACTGCGTACCGAAT	M017	8
[SEQ ID 36]	05G375	**TGAT**GATGAGTCCTGAGTAACA	M017	8

Finally the 4 P1-samples and the 4 P2-samples were pooled and concentrated. A total amount of 25µl of DNA product and a final concentration of 400ng/ul (total of 10µg) was obtained. Intermediate quality assessments are given in figure 3.

SEQUENCING BY 454

Pepper and maize AFLP fragment samples as prepared as described hereinbefore were processed by 454 Life Sciences as described (Margulies et al., 2005. Genome sequencing in microfabricated high-density picolitre reactors. Nature 437 (7057):376-80. Epub July 31, 2005).

DATA PROCESSING

Processing Pipeline:

Input data

raw sequence data were received for each run:

- 200,000 – 400,000 reads
- base calling quality scores

Trimming and tagging

These sequence data are analyzed for the presence of Keygene Recognition Sites (KRS) at the beginning and end of the read. These KRS sequences consist of both AFLP-adaptor and sample label sequence and are specific for a certain AFLP primer combination on a certain sample. The KRS sequences are identified by BLAST and trimmed and the restriction sites are restored. Reads are marked with a tag for identification of the KRS origin. Trimmed sequences are selected on length (minimum of 33 nt) to participate in further processing.

Clustering and assembly

A *MegaBlast* analysis is performed on all size-selected, trimmed reads to obtain clusters of homologous sequences. Consecutively all clusters are assembled with *CAP3* to result in assembled contigs. From both steps unique sequence reads are identified that do not match any other reads. These reads are marked as singletons.

The processing pipeline carrying out the steps described herein before is shown in Figure 4A

Polymorphism mining and quality assessment

The resulting contigs from the assembly analysis form the basis of polymorphism detection. Each 'mismatch' in the alignment of each cluster is a potential polymorphism. Selection criteria are defined to obtain a quality score:

- number of reads per contig
- frequency of 'alleles' per sample
- occurrence of homopolymer sequence
- occurrence of neighbouring polymorphisms