

Yeast two-hybrid assay. Liquid mating of yeast was performed essentially as described in Buckholz et al. (1999). To validate this method, 5 μ l from each well of the prey yeast library was transferred into a fresh 96-well V-bottom plate. Bait cultures were spun down and the cells resuspended in 45 mL YEP galactose +
5 raffinose broth with antibiotics. A 25 μ l aliquot of bait culture was added to the 5 μ l of prey culture in each well. Following a 48 hour incubation at 30°C, 200 μ l of minimal selective dropout media minus uracil, histidine, tryptophan and leucine, plus 2% galactose and 1% raffinose (SGR-UHWL) was added diluting the rich YPD media 1:10. After incubating an additional 48 hours, 5 μ l of samples were
10 transferred to a new microtiter plate and diluted 1:40 using SGR-UHWL to a final volume of 205 μ l. The diluted matings were incubated for an additional 3-5 days at 30 °C. The mating mixture (25 μ l) was transferred to 96-well assay plates and β -galactosidase assay was performed as described.

For high-throughput analysis 175 μ l of each prey culture was transferred to a
15 50 ml conical centrifuge tube and spun down, the supernatant removed, and the cells resuspended in 45 mL YNB -Trp + glucose with antibiotics. A 5 μ l aliquot of each bait culture was transferred into a fresh 96-well V-bottom plate and 25 μ l of the pooled prey culture was added to each well.

20 *Decoding assay.* Positive-interactors detected via the β -gal assay served as template for PCR in which the chloramphenicol cassettes of the interacting prey (pMAR101.x) were amplified using biotinylated primers, Watson and nCrick, and standard PCR conditions. Products were hybridized against a pool of microspheres containing 250 beads per μ l in 1.5 M NaCl. The pool was populated with 20
25 different types of microspheres; where each type was coupled to the complement of one of the ZipCode sequences used in pMAR101.x construction. During the hybridization, samples were denatured at 96 °C for 2 minutes and incubated at 45 °C for greater than one hour. Following hybridization, samples were washed in 1X SSC containing 0.2% Tween-20, resuspended in a 1.5M NaCl solution containing
30 streptavidin-phycoerythrin and incubated for 20 minutes at room temperature. The

reactions were diluted with 60 μ l of 1X SSC containing 0.2% Tween-20 prior to analysis on the LX100 (Luminex, Austin, TX).

Validating the vectors. We synthesized 96 “library” vectors, each vector
5 identifiable by a unique pair of 2 of 20 possible 25-base ZipCodes. The ZipCodes we used were DNA sequences derived from a sequence of the *M. tuberculosis* genome. Chen et al. (2000); Iannone (2000). The particular ZipCode sequences were chosen to: 1) be absent from the known human genome sequence (by BLAST analysis), 2) have no discernable secondary structure, and 3) not hybridize to any of
10 the other ZipCodes under the conditions (determined empirically) used for the genotyping and decoding analysis.

Twelve of the 20 ZipCodes (ZipCodes₁₋₁₂) have been placed 5’ upstream of a Cm^r gene while the other 8 (ZipCodes_{A-H}) have been located at the 3’ end of the Cm^r
15 gene. Each 5’ ZipCode was bracketed by a similar DNA sequence (5’-TGGGCGACTTCTCCAAAC-3’ (SEQ ID NO:2), a translation of the amino acid sequence WATSON which we have labeled the “Watson” sequence); each 3’ sequence was bracketed by a second DNA sequence (5’-CTTGCAGATTCGGCAGTT-3’ (SEQ ID NO:3), a translation of the amino acid
20 sequence NCRICK which we have labeled the “nCrick” sequence). PCR amplification was used to generate 96 different cassettes containing the Cm^r gene; each fragment with the following order: 5’-Watson-ZipCode₍₁₋₁₂₎-Cm^r-ZipCode_(A-H)-nCrick-3’. The 96 different cassettes were cloned into a unique *Swa* I site of the pMAR101 vector to synthesize the vectors pMAR101.1 ... pMAR101.96.

25 The set of vectors was purified and biotin-labeled Watson and nCrick primers were used to generate a PCR fragment. This PCR fragment was then hybridized against the set of 20 microspheres (Figure 3). Figure 4 illustrates the results from a subset of the analysis and demonstrates clear discrimination of the appropriate pair of the beads to the labeled fragment.

30

Identifying interactors using a non-random array. Prey clones were constructed

in groups of (for this example) 96 where each novel cDNA fragment of the group was cloned into a unique vector of the 96 library vectors. A Y2H analysis was performed using a bait protein against the multiplexed prey. Positive clones were isolated and the cassettes containing the Cm^r genes were amplified using biotin-dye-
5 labeled 'Watson' and 'nCrick' primers. The PCR product was then used in a bead-based decoding assay and both of the ZipCodes on either side of the Cm^r gene were identified by hybridization to a set of 20 different beads. The data points were used to decode which member of the 96 vector set contained the interactor cDNA fragment which interacted with our bait protein. To further validate this result, we
10 used vector specific primers to amplify the region containing the cloned cDNA fragment and submitted this fragment for DNA sequencing. Results were analyzed by BLAST and compared with the results of the bead-based assay, shown in Table 3.

The data show that this method can be used for high-throughput yeast two-
15 hybrid analysis where the proteins of interest are encoded by known or predicted cDNA sequences. The method is easily automated and can be scaled to accommodate projects of varying complexity. In this analysis of 96 clones per well, 20 ZipCodes in two "dimensions," the dimensions containing 12 and 8 elements (i.e., a 96 well plate), respectively, were used. In practice, a reduced set of 13
20 elements in 6 dimensions ($2 \times 2 \times 2 \times 2 \times 3$) can be used to encode the 96 different beads. In fact, the complexity of any number of analyses per well can be reduced to the sum of the prime factors defining the maximum number of clones one desires to analyze. For example, a 10-fold increase in well complexity, i.e., 960 clones per well, ($12 \times 8 \times 10$) could be encoded by either 30 beads or by the sum of the prime
25 factors making up 960 ($2 \times 2 \times 2 \times 2 \times 3 \times 5$), i.e., with just 20 beads.

To access the complete set of human genes with just a set of the currently existing 100 bead set, it is reasonable to expect that a multiplex of 1000 clones per well could be decoded using a $10 \times 12 \times 8$ matrix or array. A single master library plate could be used to represent close to the predicted number of genes.

Table 3. Example of results from an assay of approximately 1500 bait clones against 350 prey clones.¹

Sample ²	Avg. β -gal Score	Bead Set Values ³				Hit Identity	Sequence Verification ⁴
		1	2	3	4	Plasmid	
B707.P677.F05	8603	79 C 2964.922	58 10 2155.084	71 11 67.3227	8 E 52.8565	C10 pMAR101.75	(+)
B707.P675.E05	2470	58 10 1811.079	77 B 703.7803	53 H 16.65197	8 E 14.02155	B10 pMAR101.74	(+)
B691.P679.G06	2224	56 9 1436.165	8 E 761.8906	34 2 1.283019	81 D -2.06527	E09 pMAR101.69	(+)
B674.P675.G06	6480	52 6 2102.083	53 H 1564.84	50 5 18.56156	33 G 16.50951	H06 pMAR101.48	(+)
B674P675.A11	4556	38 4 1008.487	75 A 351.1794	79 C 36.33744	53 H 34.78622	A04 pMAR101.25	(-) ⁵
B678P679.D01	2784	50 5 2203.989	53 H 1553.246	73 12 36.82699	52 6 26.41147	D01 pMAR101.04	(+)
B680P679.H04	5444	36 3 1957.287	53 H 1955.586	33 G 333.8939	50 5 236.088	H03 pMAR101.24	(+)
B674P675.G05	12924	50 5 2120.277	33 G 1602.547	73 12 20.97098	79 C 16.38077	G05 pMAR101.39	(+)
B674P675.G06	6480	52 6 2019.826	53 H 1536.088	33 G 42.05269	50 5 26.03052	H06 pMAR101.47	(+)

¹ All samples have been assayed in duplicate. Samples in which both assays were successful have been included in the data.

² Sample names reflect ("bait" plate, "prey" pool, "bait" well)

³ Luminex bead values for 4 highest scoring beads (out of 20-bead set) after background subtraction.

⁴ Sequence results of inserts cloned by homologous recombination into pMAR101.x were analyzed by BLAST.

⁵ Top BLAST hit reveals a protein family member closely related to cloned gene.

Example 2: Phage display-, F pilli- and lacI-fusion display systems.

Using the protocol set forth in Example 1, 96 or more differentially-labeled M13 clones are made. These are used in either gpIII or gpVIII vectors in an arrayed format as in the Y2H example. Alternatively, the ZipCodes are generated in a 'wild-card' (i.e.,
5 random) synthesis and then cloned into the vectors as pools of up to several thousands to millions. These libraries are used in a typical phage display experiment and the ZipCodes identified after panning.

LacI and F pilli are other display-systems and can be used in a manner identical to the phage display methods.

10 Example 3: Baculovirus (in insect and eukaryotic) adenovirus, adeno-associated virus (AAV), and retrovirus.

Similar to that described for phage display and yeast two-hybrid, these
15 eukaryotic viruses are used to express foreign protein in eukaryotic cells. Baculovirus can be used to infect both insect and mammalian cells and also as a fusion vector (in a manner similar to the M13 gp-fusion vectors).

Example 4: Cell-based assay, cell-surface marker hapten or cell-surface protein.

20 The transcriptional readout used in the Y2H system is engineered such that the reporters are either small haptens or peptides that are transcribed and that eventually appear on the surface of the cell as *in vivo* fusions (for example, in yeast the Mat alpha gene product, in *E.coli* the *F pilli* gene product, and in mammalian cell lines
25 the CD40 gene product). These cell lines are used in panning experiments against a set of antibodies or other specific-interactors. The interactors are fused to beads or labeled with a reporter molecule. In a 2-dimensional analysis of 20 different fusions, 96 (using 8 x 12) different types of cells are analyzed at once. In a manner similar to that described for yeast two-hybrid, the number of dimensions or factors
30 within a dimension are increased to increase the number of cell lines that could be

simultaneously examined.

Example 5: Cell-based assay, other markers

5 In place of haptens as in the previous example, the reporters are external protein fusions or nucleic acid molecules (for example, DNA and/or RNA). The reporter molecules can be one of the many forms of green fluorescent proteins (GFPs) available.

10 Example 6: Protein tags with antibodies

Proteins are synthesized with genetic fusion tags. In a specific example, a set of twenty tags is designed that do not cross-react. Sets of tags are used in a manner similar to the dimensions used in the yeast two-hybrid experiment described.

15 Throughout this application, various publications are referenced. The disclosures of these publications in their entireties are hereby incorporated by reference into this application in order to more fully describe the state of the art to which this invention pertains.

It will be apparent to those skilled in the art that various modifications and variations can be made in the present invention without departing from the scope or spirit of the invention. Other embodiments of the invention will be apparent to those skilled in the art from consideration of the specification and practice of the invention disclosed herein. It is intended that the specification and examples be considered as exemplary only, with a true scope and spirit of the invention being indicated by the following claims.

References

1. Buckholz, R.G., et al. (1999) Automation of yeast two-hybrid screening. *J. Molec. Microbiol. Biotechnol.* 1: 34-38.
2. Chen, J., et al. (2000) A Microsphere-Based Assay for Multiplexed Single

- Nucleotide Polymorphism Analysis Using Single Base Chain Extension. *Genome Research* 10:549-557.
3. Finley, R.L., Brent, R. (1994) Interaction mating reveals binary and ternary connections between *Drosophila* cell cycle regulators. *Proc. Natl. Acad. Sci. USA* 91: 12980-12984.
 4. Gyuris, J., et al. (1993) Cdi1, a human G1 and S phase protein phosphatase that associates with Cdk2. *Cell* 75: 791-803.
 5. Iannone, M.A., et al. (2000) Multiplexed Single Nucleotide Polymorphism Genotyping by Oligonucleotide Ligation and Flow Cytometry. *Cytometry* 39:131-140.
 6. Oldenburg, K., et al. (1997) Recombination-mediated PCR-directed plasmid construction in vivo in yeast. *Nucleic Acids Research* 25:451-452.
 7. Shoemaker, D.D., et al. (1996) Quantitative phenotypic analysis of yeast deletion mutants using a highly parallel molecular bar-coding strategy. *Nat. Genet.* 14:450-456.
 8. Watson, M.A., et al. (1996) Vectors encoding Alternative Antibiotic Resistance for Use in the Yeast Two-Hybrid System. *BioTechniques* 21:255-259.

What is claimed is:

1. A method of encoding a complex mixture of assay constituents comprising using combinations of detectable tags and a total number of detectable tags that is less than the total number of constituents to be encoded, to encode the complex mixture of assay constituents.
2. The method of claim 1, further comprising:
 - (a) determining the total number of constituents to be encoded;
 - (b) determining the number of detectable tags in each combination, wherein the number of detectable tags in each combination is more than one and less than or equal to the number of prime numbers in the number of constituents to be encoded; and
 - (c) determining the total number of detectable tags, wherein the total number of detectable tags equals a sum of a set of factors of the total number of constituents, wherein the number of factors equals the number of detectable tags in each combination.
3. The method of claim 2, wherein the total number of detectable tags is minimized by selecting factors of the total number of constituents that are equal or approximate.
4. The method of claim 1, wherein the constituents are selected from the group consisting of proteins, peptides, amino acids, small molecules, nucleotides, fatty acids, sugars, cofactors, receptors, receptor ligands, protein domains, oligonucleotides, transcription factors, nucleic acids, and small compounds.
5. The method of claim 1, wherein the detectable tags are directly or indirectly coupled to the constituents.

6. The method of claim 1, wherein the detectable tags are selected from the group consisting of radiolabels, dyes, fluorescent labels, and combinations thereof.
7. The method of claim 5, wherein the detectable tags are contained in or coupled to microspheres.
8. The method of claim 7, wherein the microspheres are coupled to a means of specifically binding the constituents.
9. The method of claim 8, wherein the coupled means is a nucleic acid complementary to a nucleic acid in or bound to the constituent to be bound.
10. A method of performing a multiplexed assay using complex mixtures of assay constituents encoded according to the method of claim 2, comprising:
 - (a) Performing an assay to produce assay products using an array of the complex mixtures, wherein each constituent in a single complex mixture is detectably tagged with a unique combination of detectable tags;
 - (b) Detecting which complex mixtures of assay products in the array have a positive response; and
 - (c) Decoding the constituents in the complex mixtures having the positive response to determine which specific constituent or constituents are positive.
11. The method of claim 10, wherein the assay is selected from the group consisting of a chemical assay, protein assay, pharmacologic assay, antibody binding assay, hybrid assay, display assay, and genomic assay readout assay.
12. The method of claim 10, wherein the constituents are selected from the group

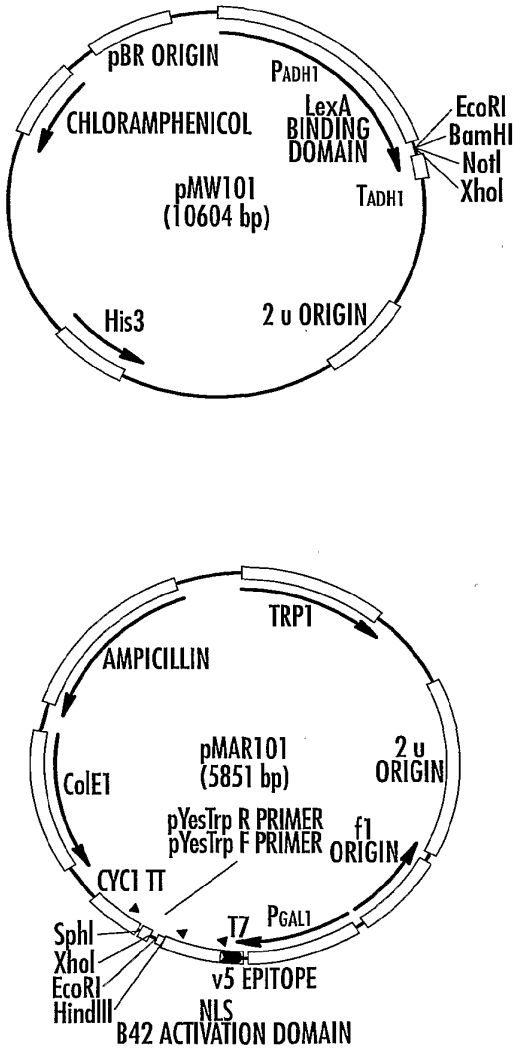
consisting of nucleic acids, proteins, peptides, amino acids, small molecules, nucleotides, fatty acids, sugars, cofactors, receptors, receptor ligands, protein domains, oligonucleotides, transcription factors, and small compounds.

13. The method of claim 10, wherein the detectable tags are directly or indirectly coupled to the constituents.
14. The method of claim 10, wherein the detectable tags are selected from the group consisting of radiolabels, dyes, fluorescent labels, and combinations thereof.
15. The method of claim 10, wherein the detectable tags are contained in or coupled to microspheres.
16. The method of claim 15, wherein the microspheres are coupled to a means of specifically binding the constituents.
17. The method of claim 16, wherein the coupled means is a nucleic acid complementary to a nucleic acid in or bound to the constituent to be bound to the microspheres.
18. The method of claim 10, wherein the decoding is performed by detecting and distinguishing with a detection device the detectable tags in each complex mixture of constituents.
19. A kit for performing a multiplexed assay using complex mixtures of encoded assay constituents, comprising a means of detectably tagging assay constituents encoded according to method of claim 2 and an arraying means for a plurality of complex mixtures.

1/4

BAIT AND PREY VECTORS

FIG. 1.



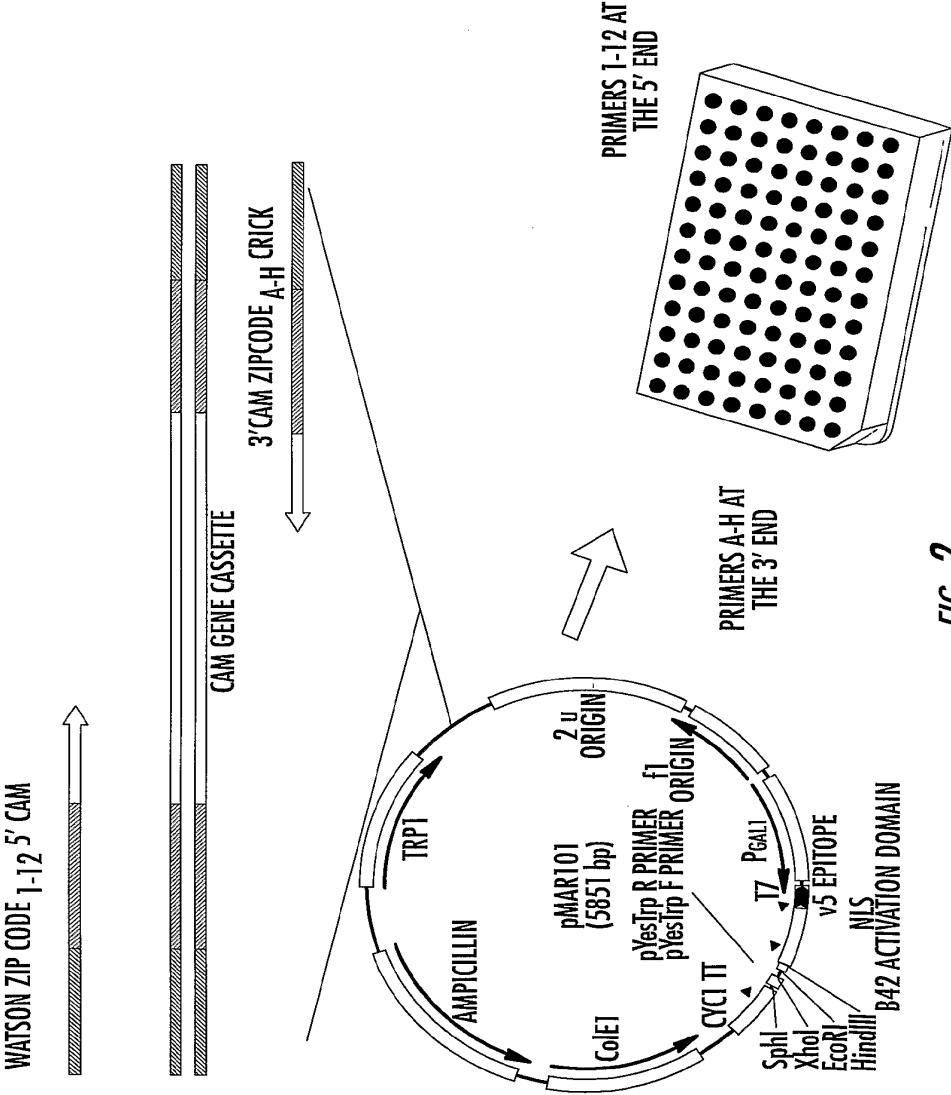


FIG. 2.

3/4

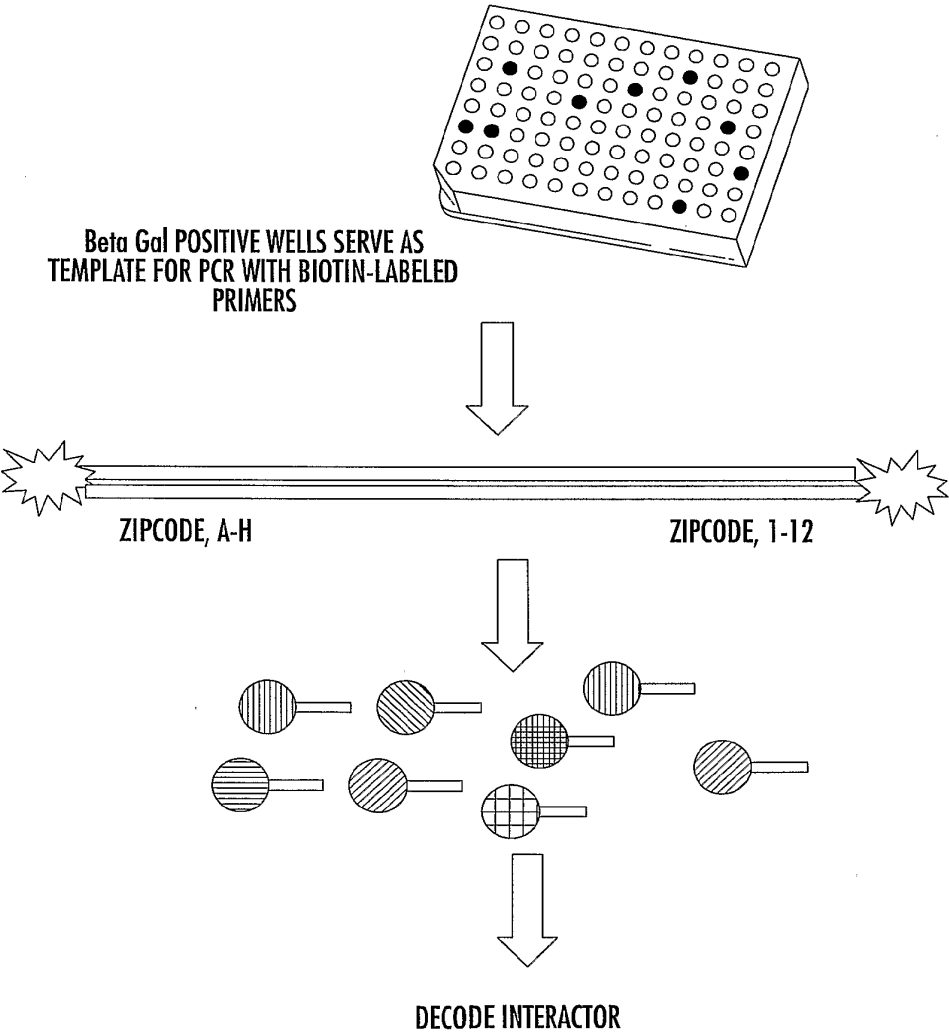


FIG. 3.

4/4

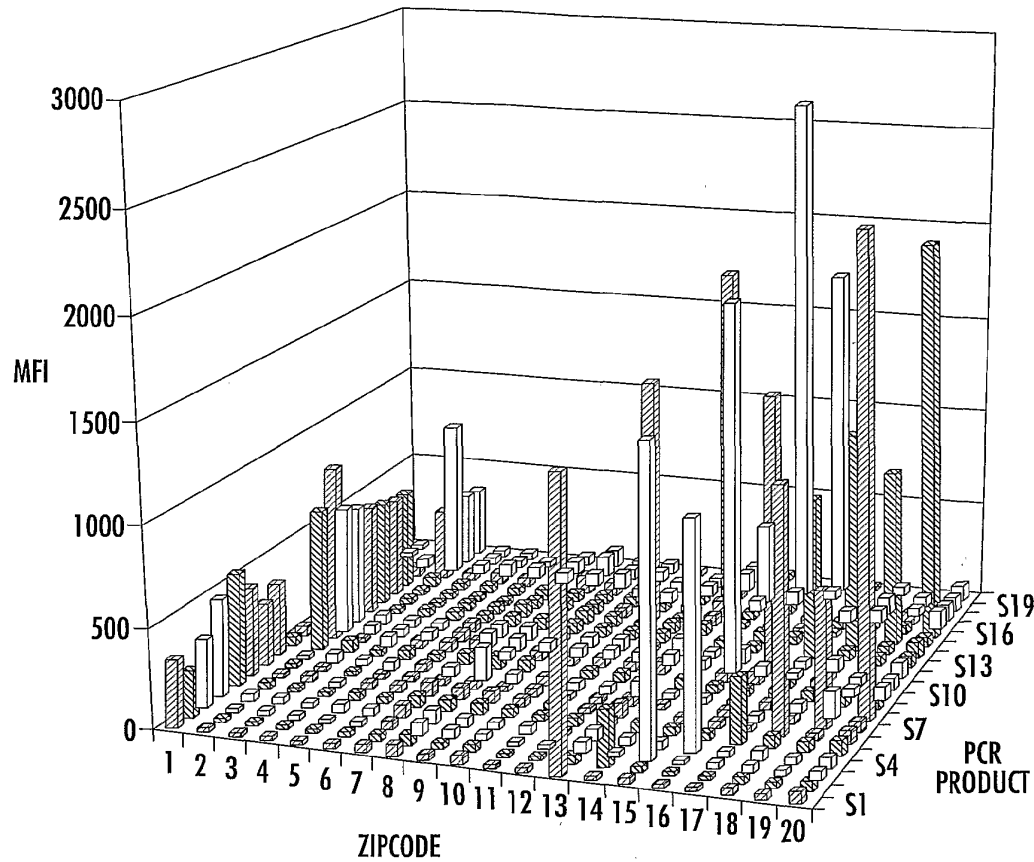


FIG. 4.

SEQUENCE LISTING

<110> Glaxo Group Limited
Michael P. Weiner
Lottie L. Peppers
Anita J. Nelsen

<120> METHODS FOR ENCODING AND DECODING
COMPLEX MIXTURES IN ARRAYED ASSAYS

<130> 07083.0010P1

<160> 29

<170> FastSEQ for Windows Version 4.0

<210> 1
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:/note =
synthetic construct

<400> 1
caggccaagt aacttcttcg 20

<210> 2
<211> 18
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:/note =
synthetic construct

<400> 2
tgggcgactt ctccaaac 18

<210> 3
<211> 18
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:/note =
synthetic construct

<400> 3
cttgagatt cggcagtt 18

<210> 4
<211> 65
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:/note =
synthetic construct

<400> 4
tgggcgactt ctccaaacga tgatcgacga gacactctcg ccactgtgac ggaagatcac 60
ttcgc 65

<210> 5
<211> 65
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:/note =
synthetic construct

<400> 5
tgggcgactt ctccaaaccg gtgcgacgagc tgccgcgcaa gatctgtgac ggaagatcac 60
ttcgc 65

<210> 6
<211> 65
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:/note =
synthetic construct

<400> 6
tgggcgactt ctccaaacga cattcgcgat cgccgcccgc tttctgtgac ggaagatcac 60
ttcgc 65

<210> 7
<211> 65
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:/note =
synthetic construct

<400> 7
tgggcgactt ctccaaaccg gtatcgcgac cgcattccaa tctctgtgac ggaagatcac 60
ttcgc 65

<210> 8
<211> 65
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:/note =
synthetic construct

<400> 8
tgggcgactt ctccaaacgc tcgaagaggc gctacagatc ctctgtgac ggaagatcac 60
ttcgc 65

<210> 9
<211> 65
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:/note =
synthetic construct

<400> 9
tgggcgactt ctccaaacca cgcagctc ggcttcgagt tcgctgtgac ggaagatcac 60
ttcgc 65

<210> 10
<211> 65
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:/note =
synthetic construct

<400> 10
tgggcgactt ctccaaacgt aaatctccag cggaagggtta cggctgtgac ggaagatcac 60
ttcgc 65

<210> 11
<211> 65
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:/note =
synthetic construct

<400> 11
tgggcgactt ctccaaacct ttcccgctcc gtcacgctc aagctgtgac ggaagatcac 60
ttcgc 65

<210> 12
<211> 65
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:/note =
synthetic construct

<400> 12
tgggcgactt ctccaaacgg ctgggtctac agatcccaa cttctgtgac ggaagatcac 60
ttcgc 65

<210> 13
<211> 65
<212> DNA
<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:/note =
synthetic construct

<400> 13
tgggcgactt ctccaaacga acctttcgct tcaccggccg atcctgtgac ggaagatcac 60
ttcgc 65

<210> 14
<211> 65
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:/note =
synthetic construct

<400> 14
tgggcgactt ctccaaactt tcggcacgcg cgggatcacc atcctgtgac ggaagatcac 60
ttcgc 65

<210> 15
<211> 65
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:/note =
synthetic construct

<400> 15
tgggcgactt ctccaaacct cgggtggtgct gacggtgcaa tccctgtgac ggaagatcac 60
ttcgc 65

<210> 16
<211> 65
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:/note =
synthetic construct

<400> 16
cttgcagatt cggcagtttc aacgtgccag cgccgtcctg ggactccacg gggagagcct 60
gagca 65

<210> 17
<211> 65
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:/note =
synthetic construct

<400> 17
cttgcagatt cggcagttgc gaaggaactc gacgtggacg ccgctccacg gggagagcct 60
gagca 65

<210> 18
<211> 65
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:/note =
synthetic construct

<400> 18
cttgcagatt cggcagttcg gggataccga tctcgggcgc aactccacg gggagagcct 60
gagca 65

<210> 19
<211> 65
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:/note =
synthetic construct

<400> 19
cttgcagatt cggcagttgg agcttacgcc atcacgatgc gatctccacg gggagagcct 60
gagca 65

<210> 20
<211> 65
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:/note =
synthetic construct

<400> 20
cttgcagatt cggcagttcg tggcgggtgcg gagtttcccc gaactccacg gggagagcct 60
gagca 65

<210> 21
<211> 65
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:/note =
synthetic construct

<400> 21
cttgcagatt cggcagttcg atccaacgca ctggccaaac ctactccacg gggagagcct 60
gagca 65

<210> 22
<211> 65
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:/note =

synthetic construct

<400> 22
 cttgcagatt cggcagttct gaatcctcca accgggttgt cgactccacg gggagagcct 60
 gagca 65

<210> 23
 <211> 65
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Description of Artificial Sequence:/note =
 synthetic construct

<400> 23
 cttgcagatt cggcagtttt cggcgctggc gtaaagcttt tggctccacg gggagagcct 60
 gagca 65

<210> 24
 <211> 21
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Description of Artificial Sequence:/note =
 synthetic construct

<400> 24
 tccacgggga gagcctgagc a 21

<210> 25
 <211> 22
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Description of Artificial Sequence:/note =
 synthetic construct

<400> 25
 ctgtgacgga agatcacttc gc 22

<210> 26
 <211> 23
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Description of Artificial Sequence:/note =
 synthetic construct

<400> 26
 gcaacggcga ctggctggaa ttc 23

<210> 27
 <211> 23
 <212> DNA
 <213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:/note =
synthetic construct

<400> 27

gcttggtgc aggtcgactc gag

23

<210> 28

<211> 39

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:/note =
synthetic construct

<400> 28

gccgaagctt gcggttgggg tattcgcaac ggcgactgg

39

<210> 29

<211> 40

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:/note =
synthetic construct

<400> 29

atacgcatgc aattcgcccg gaattagctt ggctgcaggt

40

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

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
18 July 2002 (18.07.2002)

PCT

(10) International Publication Number
WO 02/056014 A3

(51) International Patent Classification⁷: **G01N 33/53**,
C12Q 1/00, G01N 33/543

(21) International Application Number: PCT/US01/49132

(22) International Filing Date:
18 December 2001 (18.12.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
09/747,003 22 December 2000 (22.12.2000) US

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

(72) Inventors; and

(75) Inventors/Applicants (for US only): **NELSEN, Anita, J.** [US/US]; GlaxoSmithKline, Five Moore Drive, P.O. Box 13398, Research Triangle Park, NC 27709 (US). **PEPPERS, Lottie, L.** [US/US]; GlaxoSmithKline, Five Moore Drive, P.O. Box 13398, Research Triangle Park, NC 27709 (US). **WEINER, Michael, Phillip** [US/US]; c/o 454 Corporation, 20 Commercial Street, Branford, CT 06405 (US).

(74) Agent: **GRASSLER, Frank, P.**; GlaxoSmithKline, Five Moore Drive, P.O. Box 1398, Research Triangle Park, NC 27709 (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, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, 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, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

- with international search report
- before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments

(88) Date of publication of the international search report:
9 October 2003

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 02/056014 A3

(54) Title: METHODS FOR ENCODING AND DECODING COMPLEX MIXTURES IN ARRAYED ASSAYS

(57) Abstract: The present invention is a method of encoding a complex mixture of assay constituents comprising using combinations of detectable tags and a total number of detectable tags less than the total number of constituents to be encoded. The method comprises determining the total number of constituents to be encoded; determining the number of detectable tags in each combination, wherein the number of detectable tags in each combination is more than one and less than or equal to the number of prime numbers in the number of constituents to be encoded; and determining the total number of detectable tags, wherein the total number of detectable tags equals a sum of a set of factors of the total number of constituents, wherein the number of factors equals the number of detectable tags in each combination. The encoding methods are useful in a multiplexed assay using complex mixtures of assay constituents.

INTERNATIONAL SEARCH REPORT

International Application No
PCT/US 01/49132

A. CLASSIFICATION OF SUBJECT MATTER IPC 7 G01N33/53 C12Q1/00 G01N33/543		
According to International Patent Classification (IPC) or to both national classification and IPC		
B. FIELDS SEARCHED Minimum documentation searched (classification system followed by classification symbols) IPC 7 G01N		
Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched		
Electronic data base consulted during the international search (name of data base and, where practical, search terms used) EPO-Internal, WPI Data, PAJ, BIOSIS, MEDLINE, EMBASE		
C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	FULTON R J ET AL: "Advanced multiplexed analysis with the FlowMetrix(TM) system" CLINICAL CHEMISTRY, AMERICAN ASSOCIATION FOR CLINICAL CHEMISTRY. WINSTON, US, vol. 43, no. 9, September 1997 (1997-09), pages 1749-1756, XP002142645 ISSN: 0009-9147	1,4-6
Y	the whole document	1-19
X	WO 99 52708 A (LUMINEX CORPORATION) 21 October 1999 (1999-10-21) page 17, line 27 -page 19; claims 1-17	1-19
A	WO 99 64867 A (AMERSHAM PHARMACIA BIOTECH UK LIMITED) 16 December 1999 (1999-12-16) the whole document	1-19
-/--		
☒ Further documents are listed in the continuation of box C. ☒ Patent family members are listed in annex.		
* Special categories of cited documents : *A* document defining the general state of the art which is not considered to be of particular relevance *E* earlier document but published on or after the international filing date *L* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) *O* document referring to an oral disclosure, use, exhibition or other means *P* document published prior to the international filing date but later than the priority date claimed *T* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention *X* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone *Y* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art. *&* document member of the same patent family		
Date of the actual completion of the international search 22 July 2003		Date of mailing of the international search report 06/08/2003
Name and mailing address of the ISA European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016		Authorized officer Moreno de Vega, C

INTERNATIONAL SEARCH REPORT

International Application No
PCT/US 01/49132

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	CHEN J ET AL: "A MICROSPHERE-BASED ASSAY FOR MULTIPLEXED SINGLE NUCLEOTIDE POLYMORPHISM ANALYSIS USING SINGLE BASE CHAIN EXTENSION" GENOME RESEARCH, COLD SPRING HARBOR LABORATORY PRESS, US, vol. 10, no. 4, April 2000 (2000-04), pages 549-557, XP000927257 ISSN: 1088-9051 the whole document	1-19
Y	WO 00 68434 A (YALE UNIVERSITY) 16 November 2000 (2000-11-16) the whole document	1-19

Form PCT/ISA/210 (continuation of second sheet) (July 1992)

page 2 of 2

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No

PCT/US 01/49132

Patent document cited in search report		Publication date	Patent family member(s)	Publication date
WO 9952708	A	21-10-1999	AU 3555599 A EP 1079967 A1 WO 9952708 A1	01-11-1999 07-03-2001 21-10-1999
WO 9964867	A	16-12-1999	CA 2313047 A1 EP 1036332 A1 WO 9964867 A1 JP 2002517759 T	16-12-1999 20-09-2000 16-12-1999 18-06-2002
WO 0068434	A	16-11-2000	AU 4824800 A CA 2371843 A1 EP 1196632 A2 JP 2002543813 T WO 0068434 A2	21-11-2000 16-11-2000 17-04-2002 24-12-2002 16-11-2000

Form PCT/ISA/210 (patent family annex) (July 1992)

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
9 January 2003 (09.01.2003)

PCT

(10) International Publication Number
WO 03/002979 A2

(51) International Patent Classification⁷: **G01N**

(21) International Application Number: PCT/US02/22095

(22) International Filing Date: 28 June 2002 (28.06.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/302,213 28 June 2001 (28.06.2001) US

(63) Related by continuation (CON) or continuation-in-part (CIP) to earlier application:
US 60/302,213 (CIP)
Filed on 28 June 2001 (28.06.2001)

(71) Applicant (for all designated States except US): **ILLUMINA, INC.** [US/US]; 9885 Towne Centre Drive, San Diego, CA 92121 (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **CHEE, Mark, S.** [AU/US]; 155 15th Street, No. 22, Del Mar, CA 92014

(US). **STUELPNAGEL, John, R.** [US/US]; 38 Briggs Avenue, Encinitas, CA 92024 (US). **CZARNIK, Anthony, W.** [US/US]; 11017 Camino Abrojo, San Diego, CA 92127 (US).

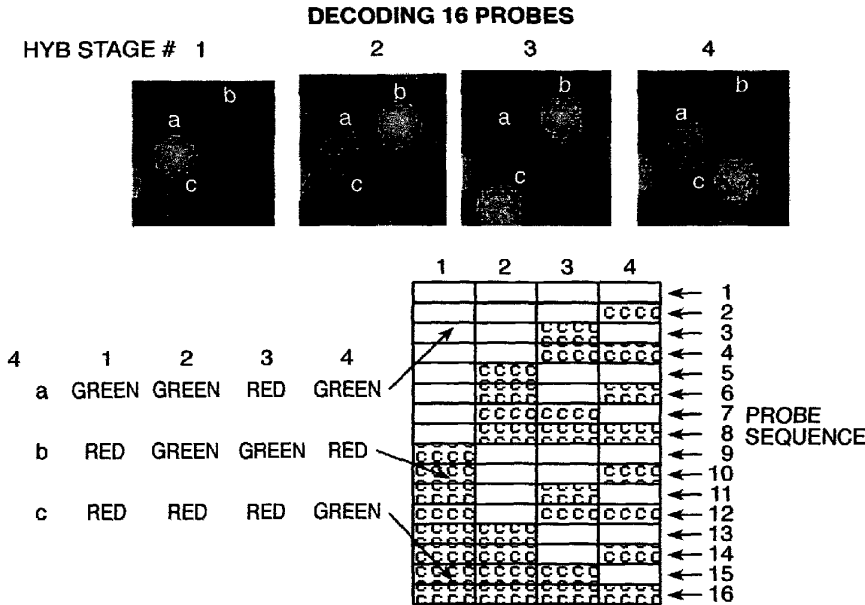
(74) Agents: **SILVA, Robin, M.** et al.; DORSEY & WHITNEY LLP, 4 Embarcadero Center, Suite 3400, San Francisco, CA 94111 (US).

(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, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, 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, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

[Continued on next page]

(54) Title: MULTIPLEX DECODING OF ARRAY SENSORS WITH MICROSPHERES



(57) Abstract: The invention relates to compositions and methods for multiplex decoding of microsphere array sensors.

WO 03/002979 A2



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.

MULTIPLEX DECODING OF ARRAY SENSORS WITH MICROSPHERES

This application claims the benefit of U.S.S.N. 60/302, 213, filed June 28, 2001.

FIELD OF THE INVENTION

The invention relates to compositions and methods for decoding microsphere array sensors.

BACKGROUND OF THE INVENTION

There are a number of assays and sensors for the detection of the presence and/or concentration of specific substances in fluids and gases. Many of these rely on specific ligand/antiligand reactions as the mechanism of detection. That is, pairs of substances (i.e. the binding pairs or ligand/antiligands) are known to bind to each other, while binding little or not at all to other substances. This has been the focus of a number of techniques that utilize these binding pairs for the detection of the complexes. These generally are done by labeling one component of the complex in some way, so as to make the entire complex detectable, using, for example, radioisotopes, fluorescent and other optically active molecules, enzymes, etc.

Of particular use in these sensors are detection mechanisms utilizing luminescence. Recently, the use of optical fibers and optical fiber strands in combination with light absorbing dyes for chemical analytical determinations has undergone rapid development, particularly within the last decade. The use of optical fibers for such purposes and techniques is described by Milanovich et al., "Novel Optical Fiber Techniques For Medical Application", Proceedings of the SPIE 28th Annual International Technical Symposium On Optics and Electro-Optics, Volume 494, 1980; Seitz, W.R., "Chemical

Sensors Based On Immobilized Indicators and Fiber Optics" in *C.R.C. Critical Reviews In Analytical Chemistry*, Vol. 19, 1988, pp. 135-173; Wolfbeis, O.S., "Fiber Optical Fluorosensors In Analytical Chemistry" in *Molecular Luminescence Spectroscopy, Methods and Applications* (S. G. Schulman, editor), Wiley & Sons, New York (1988); Angel, S.M., *Spectroscopy* 2 (4):38 (1987); Walt, et al., "Chemical Sensors and Microinstrumentation", *ACS Symposium Series*, Vol. 403, 1989, p. 252, and Wolfbeis, O.S., *Fiber Optic Chemical Sensors*, Ed. CRC Press, Boca Raton, FL, 1991, 2nd Volume.

When using an optical fiber in an *in vitro/in vivo* sensor, one or more light absorbing dyes are located near its distal end. Typically, light from an appropriate source is used to illuminate the dyes through the fiber's proximal end. The light propagates along the length of the optical fiber; and a portion of this propagated light exits the distal end and is absorbed by the dyes. The light absorbing dye may or may not be immobilized; may or may not be directly attached to the optical fiber itself; may or may not be suspended in a fluid sample containing one or more analytes of interest; and may or may not be retainable for subsequent use in a second optical determination.

Once the light has been absorbed by the dye, some light of varying wavelength and intensity returns, conveyed through either the same fiber or collection fiber(s) to a detection system where it is observed and measured. The interactions between the light conveyed by the optical fiber and the properties of the light absorbing dye provide an optical basis for both qualitative and quantitative determinations.

Of the many different classes of light absorbing dyes which conventionally are employed with bundles of fiber strands and optical fibers for different analytical purposes are those more common compositions that emit light after absorption termed "fluorophores" and those which absorb light and internally convert the absorbed light to heat, rather than emit it as light, termed "chromophores."

Fluorescence is a physical phenomenon based upon the ability of some molecules to absorb light (photons) at specified wavelengths and then emit light of a longer wavelength and at a lower energy. Substances able to fluoresce share a number of common characteristics: the ability to absorb light energy at one wavelength λ_{ab} ; reach an excited energy state; and subsequently emit light at another light wavelength, λ_{em} . The absorption and fluorescence emission spectra are individual for each fluorophore and are often graphically represented as two separate curves that are slightly overlapping. The same fluorescence emission spectrum is generally observed irrespective of the wavelength of the exciting light and, accordingly, the wavelength and energy of the exciting light may be varied within limits; but the light emitted by the fluorophore will always provide the same emission spectrum. Finally, the strength of the fluorescence signal may be measured as the quantum yield of light emitted. The fluorescence quantum yield is the ratio of the number of photons emitted in comparison to the number of photons initially absorbed by the fluorophore. For more detailed information regarding each of these characteristics, the following references are recommended: Lakowicz, J. R., *Principles of Fluorescence Spectroscopy*, Plenum Press, New York, 1983; Freifelder, D., *Physical Biochemistry*, second edition, W. H. Freeman and Company, New York, 1982; "Molecular Luminescence

Spectroscopy Methods and Applications: Part I" (S.G. Schulman, editor) in *Chemical Analysis*, vol. 77, Wiley & Sons, Inc., 1985; *The Theory of Luminescence*, Stepanov and Gribkovskii, Iliffe Books, Ltd., London, 1968.

In comparison, substances which absorb light and do not fluoresce usually convert the light into heat or kinetic energy. The ability to internally convert the absorbed light identifies the dye as a "chromophore." Dyes which absorb light energy as chromophores do so at individual wavelengths of energy and are characterized by a distinctive molar absorption coefficient at that wavelength. Chemical analysis employing fiber optic strands and absorption spectroscopy using visible and ultraviolet light wavelengths in combination with the absorption coefficient allow for the determination of concentration for specific analyses of interest by spectral measurement. The most common use of absorbance measurement via optical fibers is to determine concentration which is calculated in accordance with Beers' law; accordingly, at a single absorbance wavelength, the greater the quantity of the composition which absorbs light energy at a given wavelength, the greater the optical density for the sample. In this way, the total quantity of light absorbed directly correlates with the quantity of the composition in the sample.

Many of the recent improvements employing optical fiber sensors in both qualitative and quantitative analytical determinations concern the desirability of depositing and/or immobilizing various light absorbing dyes at the distal end of the optical fiber. In this manner, a variety of different optical fiber chemical sensors and methods have been reported for specific analytical determinations and applications such as pH measurement, oxygen detection, and carbon dioxide analyses. These developments are exemplified by the following publications: Freeman, et al., *Anal. Chem.* 53:98 (1983); Lippitsch et al., *Anal. Chem. Acta.* 205:1, (1988); Wolfbeis et al., *Anal. Chem.* 60:2028 (1988); Jordan, et al., *Anal. Chem.* 59:437 (1987); Lubbers et al., *Sens. Actuators* 1983; Munkholm et al., *Talanta* 35:109 (1988); Munkholm et al., *Anal. Chem.* 58:1427 (1986); Seitz, W. R., *Anal. Chem.* 56:16A-34A (1984); Peterson, et al., *Anal. Chem.* 52:864 (1980); Saari, et al., *Anal. Chem.* 54:821 (1982); Saari, et al., *Anal. Chem.* 55:667 (1983); Zhujun et al., *Anal. Chem. Acta.* 160:47 (1984); Schwab, et al., *Anal. Chem.* 56:2199 (1984); Wolfbeis, O.S., "Fiber Optic Chemical Sensors", Ed. CRC Press, Boca Raton, FL, 1991, 2nd Volume; and Pantano, P., Walt, D.R., *Anal. Chem.*, 481A-487A, Vol. 67, (1995).

More recently, fiber optic sensors have been constructed that permit the use of multiple dyes with a single, discrete fiber optic bundle. U.S. Pat. Nos. 5,244,636 and 5,250,264 to Walt, et al. disclose systems for affixing multiple, different dyes on the distal end of the bundle, the teachings of each of these patents being incorporated herein by this reference. The disclosed configurations enable separate optical fibers of the bundle to optically access individual dyes. This avoids the problem of deconvolving the separate signals in the returning light from each dye, which arises when the signals from two or more dyes are combined, each dye being sensitive to a different analyte, and there is significant overlap in the dyes' emission spectra.

U.S.S.N.s 08/818,199 and 09/151,877 describe array compositions that utilize microspheres or beads on a surface of a substrate, for example on a terminal end of a fiber optic bundle, with each individual fiber comprising a bead containing an optical signature. Since the beads go down randomly, a unique optical signature is needed to "decode" the array; i.e. after the array is made, a correlation of the location of an individual site on the array with the bead or bioactive agent at that particular site can be made. This means that the beads may be randomly distributed on the array, a fast and inexpensive process as compared to either the *in situ* synthesis or spotting techniques of the prior art. Once the array is loaded with the beads, the array can be decoded, or can be used, with full or partial decoding occurring after testing, as is more fully outlined below.

One drawback with the previous system is that it requires a set of unique optical signatures. While large sets of such signatures are available, for example by using different ratios of different dyes, it would be preferable to use decoding systems that do not rely on the use of sets of optical signatures. Accordingly, it is an object of the invention to provide methods to allow decoding of bead arrays without relying solely on unique optical signatures.

SUMMARY OF THE INVENTION

In accordance with the above objects, the present invention provides array compositions comprising a substrate with a surface comprising discrete sites. The composition further comprises a population of microspheres comprising at least a first and a second subpopulation; each subpopulation comprises a bioactive agent; and an identifier binding ligand that will bind a decoder binding ligand such that the identity of the bioactive agent can be elucidated. The microspheres are distributed on the surface.

In a further aspect, the present invention provides array compositions comprising a substrate with a surface comprising discrete sites, and a population of microspheres comprising at least a first and a second subpopulation. Each subpopulation comprises a bioactive agent and does not comprise an optical signature.

In an additional aspect, the present invention provides methods of making an array composition as outlined above. The methods comprise forming a surface comprising individual sites on a substrate and distributing microspheres on said surface such that said individual sites contain microspheres. The microspheres comprise at least a first and a second subpopulation each comprising a bioactive agent and do not comprise an optical signature.

In a further aspect, the invention provides methods of making a composition comprising forming a surface comprising individual sites on a substrate and distributing microspheres on the surface such that the individual sites contain microspheres. The microspheres comprise at least a first and a second subpopulation each comprising a bioactive agent and an identifier binding ligand that will bind a decoder binding ligand such that the identification of the bioactive agent can be elucidated.

In an additional aspect, the invention provides methods of decoding an array composition comprising providing an array composition as outlined above, and adding a plurality of decoding binding ligands to the array composition to identify the location of at least a plurality of the bioactive agents.

In a further aspect, the invention provides methods of determining the presence of a target analyte in a sample. The methods comprise contacting the sample with an array composition as outlined herein, and determining the presence or absence of the target analyte.

In a further aspect, the invention provides a method comprising providing an array composition comprising a population of microspheres comprising at least a first and a second subpopulation, wherein each subpopulation comprises a bioactive agent and at least a first and a second decoding attribute, and detecting each of said first and second decoding attributes to identify each of said bioactive agents.

In a further aspect the invention provides a method of increasing the information obtained in a decoding step. The method includes the use of degenerate probes as DBL-IBL combinations. In addition the invention provides the use of multiple decoding attributes on a bead.

In further aspect, the invention provides a method of increasing the confidence of decoding. The method includes using the decoding as a quality control measure. In addition, the invention provides quality control fibers. In addition, the invention provides for parity analysis of decoding data.

In a further aspect, the invention provides a method of decoding an array composition comprising providing an array composition comprising a population of microspheres comprising at least 50 subpopulations, wherein each subpopulation comprises a bioactive agent adding a plurality of decoding binding ligands to said population of microspheres to identify at least 50 of the bioactive agents.

In a further aspect, the invention provides a method of determining the presence of a target analyte in a sample comprising contacting said sample with a composition comprising a population of microspheres comprising at least 50 subpopulations, wherein each subpopulation comprises a bioactive agent adding a plurality of decoding binding ligands to said population of microspheres to identify at least 50 of the bioactive agents and determining the presence or absence of said target analyte.

FIGURES

Fig. 1 Illustrates a two color decoding process wherein either FAM-labeled or Cy3-labeled oligo complements are use to "paint" (label) the different bead types on the array.

Fig. 2 Depicts the decoding 128 different bead types with four colors and four decode stages. (inset shows a single decode stage using four different dyes to decode 16 bead types.)

Fig. 3. Grey scale decoding of 16 different bead types. (A) Combinatorial pooling scheme for complementary decoding oligos. A (B) Two independent normalizing images were acquired, and the resulting bead intensities compared. (C) The alpha values (ratio of bead intensity in indicated decode stage to intensity in normalization image) are plotted for three decodes stage described in (A).

Fig. 4 Depicts the use of fluorescence resonance energy transfer as a code for a probe on a random array. A. Depicts FRET oligos with linkers of different lengths. B. Depicts a plot of the ration of emission of the fluors from the oligos in Fig. 4A.

Figure 5 Use of fluorescence resonance energy transfer to discriminate between linkers of varying length.

Figure 6 depicts clustering in hyperspectral alpha space ($\Delta_1 = I_1/6I_1$, $\Delta_2 = I_2/6I_1$, $\Delta_3 = I_3/6I_1$, etc.). A set of 128 different bead types present on a fiber bundle were decoded with by hybridizing set of complementary oligonucleotides labeled with four dyes: Bodipy-493, Bodipy-R6G, Bodipy-TXR, and Bod-564 (only one dye per oligonucleotide). Shown is the second stage of a four stage decode in which 4013 beads were decoded. Ovals are drawn around zones of hue clusters.

Figure 7 depicts an error checking scheme.

Figure 8 depicts an example of primer extension labeling of decoder oligonucleotides.

Figure 9 depicts two color decoding of 16 beads with 16 distinct IBLs.

Figure 10 depicts primer extension labeling of degenerate IBL-DBL probes.

Figure 11 depicts a scatter plot of signals obtained from three distinct labels recorded in two channels.

Figure 12 depicts decoding by sequencing. A polymerase (e.g. Sequenase, Thermosequenase) and dye-labeled chain-terminating nucleotides are used to sequence the encoding region of a DNA strand present on the bead. The four nucleotides are each labeled with a discrete label. Because only 4 colors are detected at once, the colors are removed after each extension step. In subsequent extension steps, the primers are indexed or extended by one base so that the primer can query the next base in the encoding region.

Figure 13 Depicts construction of probes on bead containing encoding sequences, zipcodes, and a gene-specific sequence. Two different encoding cassettes are employed to facilitate the primer extension reaction using primers with universal or degenerate bases. Using four color sequencing and a single base code, 6 bases generates $4^6 = 4096$ codes, likewise four color hybridization using single hybridization colors also generates 4096 codes. The grand total number of codes is $4096 \times 4096 = >16$ million. If only a single color/two state scheme is employed for hybridization, than $2^6 = 64$ codes are generated. The grand total is $4096 \times 64 = 262,144$ codes. The zipcode sequences can also be constructed so as to be overlapping to reduce the length of the overall sequence (i.e. cZip#1 vs. cZip#2).

Figure 14 Depicts incorporation of a cleavable linker between encoding sequences and gene-specific sequence. A disulfide cleavable linker is shown for illustration purposes. If the linker is not present, the terminal portion of the gene-specific sequence can be included as part of the encoding sequence to reduce overall length of the probe.

Figure 15 depicts decoding with exonuclease.

Figure 16 depicts a graph showing decoding of about 1500 bead subpopulation on each of several different fiber bundles.

DETAILED DESCRIPTION OF THE INVENTION

The present invention is generally based on previous work comprising a bead-based analytic chemistry system in which beads, also termed microspheres, carrying different chemical functionalities are distributed on a substrate comprising a patterned surface of discrete sites that can bind the individual microspheres. Since the beads are generally put onto the substrate randomly, the previous work relied on the incorporation of unique optical signatures, generally fluorescent dyes, that could be used to identify the chemical functionality on any particular bead. This allows the synthesis of the candidate agents (i.e. compounds such as nucleic acids and antibodies) to be divorced from their placement on an array, i.e. the candidate agents may be synthesized on the beads, and then the beads are randomly distributed on a patterned surface. Since the beads are first coded with an optical signature, this means that the array can later be "decoded", i.e. after the array is made, a correlation of the location of an individual site on the array with the bead or candidate agent at that particular site can be made. This means that the beads may be randomly distributed on the array, a fast and inexpensive process as compared to either the in situ synthesis or spotting techniques of the prior art.

However, the drawback to these methods is that for a large array, the system requires a large number of different optical signatures, which may be difficult or time-consuming to utilize. Accordingly, the present invention provides several improvements over these methods, generally directed to methods of coding and decoding the arrays. That is, as will be appreciated by those in the art, the placement of

the bioactive agents is generally random, and thus a coding/decoding system is required to identify the bioactive agent at each location in the array. This may be done in a variety of ways, as is more fully outlined below, and generally includes: a) the use of decoding binding ligands (DBLs), generally directly labeled, that binds to either the bioactive agent or to identifier binding ligands (IBLs) attached to the beads; b) positional decoding, for example by either targeting the placement of beads (for example by using photoactivatable or photocleavable moieties to allow the selective addition of beads to particular locations), or by using either sub-bundles or selective loading of the sites, as are more fully outlined below; c) selective decoding, wherein only those beads that bind to a target are decoded; or d) combinations of any of these. In some cases, as is more fully outlined below, this decoding may occur for all the beads, or only for those that bind a particular target analyte. Similarly, this may occur either prior to or after addition of a target analyte.

Once the identity of the bioactive agent and its location in the array has been fixed, the array is exposed to samples containing the target analytes, although as outlined below, this can be done prior to or during the analysis as well. The target analytes will bind to the bioactive agents as is more fully outlined below, and result in a change in the optical signal of a particular bead.

In the present invention, "decoding" does not rely on the use of optical signatures, but rather on the use of decoding binding ligands that are added during a decoding step. The decoding binding ligands will bind either to a distinct identifier binding ligand partner that is placed on the beads, or to the bioactive agent itself, for example when the beads comprise single-stranded nucleic acids as the bioactive agents. The decoding binding ligands are either directly or indirectly labeled, and thus decoding occurs by detecting the presence of the label. By using pools of decoding binding ligands in a sequential fashion, it is possible to greatly minimize the number of required decoding steps.

In one embodiment, the invention provides methods for multiplex decoding of bead-arrays. That is, microspheres with bioactive agents are decoded to elucidate the identity of the bioactive agent on the bead. In a preferred embodiment at least 50 different bead types or subpopulations are simultaneously decoded, or decoded in a single decoding experiment. More preferably at least 100; more preferably at least 500; more preferably at least 1000; or preferably at least 1500 bead types are simultaneously decoded or decoded in a single decoding experiment by the methods of the invention. Preferably at least 5000 bead types are simultaneously decoded or decoded in a single decoding experiment.

Accordingly, the present invention provides array compositions comprising at least a first substrate with a surface comprising individual sites. By "array" herein is meant a plurality of candidate agents in an array format; the size of the array will depend on the composition and end use of the array. Arrays containing from about 2 different bioactive agents (i.e. different beads) to many millions can be made, with very large fiber optic arrays being possible. Generally, the array will comprise from two to as many as a billion or more, depending on the size of the beads and the substrate, as well as the end use of the array, thus very high density, high density, moderate density, low density and very low

density arrays may be made. Preferred ranges for very high density arrays are from about 10,000,000 to about 2,000,000,000 (all numbers are per square cm), with from about 100,000,000 to about 1,000,000,000 being preferred. High density arrays range about 100,000 to about 10,000,000, with from about 1,000,000 to about 5,000,000 being particularly preferred. Moderate density arrays range from about 10,000 to about 100,000 being particularly preferred, and from about 20,000 to about 50,000 being especially preferred. Low density arrays are generally less than 10,000, with from about 1,000 to about 5,000 being preferred. Very low density arrays are less than 1,000, with from about 100 to about 1,000 being preferred, and from about 100 to about 500 being particularly preferred. In some embodiments, the compositions of the invention may not be in array format; that is, for some embodiments, compositions comprising a single bioactive agent may be made as well. In addition, in some arrays, multiple substrates may be used, either of different or identical compositions. Thus for example, large arrays may comprise a plurality of smaller substrates.

In addition, one advantage of the present compositions is that particularly through the use of fiber optic technology, extremely high density arrays can be made. Thus for example, because beads of 200 nm or less (with beads of 200 nm possible) can be used, and very small fibers are known, it is possible to have as many as 40,000 or more (in some instances, 1 million) different fibers and beads in a 1 mm² fiber optic bundle, with densities of greater than 15,000,000 individual beads and fibers (again, in some instances as many as 25-50 million) per 0.5 cm² obtainable.

By "substrate" or "solid support" or other grammatical equivalents herein is meant any material that can be modified to contain discrete individual sites appropriate for the attachment or association of beads and is amenable to at least one detection method. As will be appreciated by those in the art, the number of possible substrates is very large. Possible substrates include, but are not limited to, glass and modified or functionalized glass, plastics (including acrylics, polystyrene and copolymers of styrene and other materials, polypropylene, polyethylene, polybutylene, polyurethanes, Teflon, etc.), polysaccharides, nylon or nitrocellulose, resins, silica or silica-based materials including silicon and modified silicon, carbon, metals, inorganic glasses, plastics, optical fiber bundles, and a variety of other polymers. In general, the substrates allow optical detection and do not themselves appreciably fluoresce.

Generally the substrate is flat (planar), although as will be appreciated by those in the art, other configurations of substrates may be used as well; for example, three dimensional configurations can be used, for example by embedding the beads in a porous block of plastic that allows sample access to the beads and using a confocal microscope for detection. Similarly, the beads may be placed on the inside surface of a tube, for flow-through sample analysis to minimize sample volume. Preferred substrates include optical fiber bundles as discussed below, and flat planar substrates such as glass, polystyrene and other plastics and acrylics.

In a preferred embodiment, the substrate is an optical fiber bundle or array, as is generally described in U.S.S.N.s 08/944,850 and 08/519,062, PCT US98/05025, and PCT US98/09163, all of which are expressly incorporated herein by reference. Preferred embodiments utilize preformed unitary fiber optic arrays. By "preformed unitary fiber optic array" herein is meant an array of discrete individual fiber optic strands that are co-axially disposed and joined along their lengths. The fiber strands are generally individually clad. However, one thing that distinguished a preformed unitary array from other fiber optic formats is that the fibers are not individually physically manipulatable; that is, one strand generally cannot be physically separated at any point along its length from another fiber strand.

Generally, the array of array compositions of the invention can be configured in several ways; see for example U.S.S.N. 09/473,904, and WO 00/39587, both of which are hereby expressly incorporated by reference. In a preferred embodiment, as is more fully outlined below, a "one component" system is used. That is, a first substrate comprising a plurality of assay locations (sometimes also referred to herein as "assay wells"), such as a microtiter plate, is configured such that each assay location contains an individual array. That is, the assay location and the array location are the same. For example, the plastic material of the microtiter plate can be formed to contain a plurality of "bead wells" in the bottom of each of the assay wells. Beads containing the capture probes of the invention can then be loaded into the bead wells in each assay location as is more fully described below.

Alternatively, a "two component" system can be used. In this embodiment, the individual arrays are formed on a second substrate, which then can be fitted or "dipped" into the first microtiter plate substrate. A preferred embodiment utilizes fiber optic bundles as the individual arrays, generally with "bead wells" etched into one surface of each individual fiber, such that the beads containing the capture probes are loaded onto the end of the fiber optic bundle. The composite array thus comprises a number of individual arrays that are configured to fit within the wells of a microtiter plate.

By "composite array" or "combination array" or grammatical equivalents herein is meant a plurality of individual arrays, as outlined above. Generally the number of individual arrays is set by the size of the microtiter plate used; thus, 96 well, 384 well and 1536 well microtiter plates utilize composite arrays comprising 96, 384 and 1536 individual arrays, although as will be appreciated by those in the art, not each microtiter well need contain an individual array. It should be noted that the composite arrays can comprise individual arrays that are identical, similar or different. That is, in some embodiments, it may be desirable to do the same 2,000 assays on 96 different samples; alternatively, doing 192,000 experiments on the same sample (i.e. the same sample in each of the 96 wells) may be desirable. Alternatively, each row or column of the composite array could be the same, for redundancy/quality control. As will be appreciated by those in the art, there are a variety of ways to configure the system. In addition, the random nature of the arrays may mean that the same population of beads may be added to two different surfaces, resulting in substantially similar but perhaps not identical arrays.

At least one surface of the substrate is modified to contain discrete, individual sites for later association of microspheres. These sites may comprise physically altered sites, i.e. physical configurations such as wells or small depressions in the substrate that can retain the beads, such that a microsphere can rest in the well, or the use of other forces (magnetic or compressive), or chemically altered or active sites, such as chemically functionalized sites, electrostatically altered sites, hydrophobically/ hydrophilically functionalized sites, spots of adhesive, etc.

The sites may be a pattern, i.e. a regular design or configuration, or randomly distributed. A preferred embodiment utilizes a regular pattern of sites such that the sites may be addressed in the X-Y coordinate plane. "Pattern" in this sense includes a repeating unit cell, preferably one that allows a high density of beads on the substrate. However, it should be noted that these sites may not be discrete sites. That is, it is possible to use a uniform surface of adhesive or chemical functionalities, for example, that allows the association of beads at any position. That is, the surface of the substrate is modified to allow association of the microspheres at individual sites, whether or not those sites are contiguous or non-contiguous with other sites. Thus, the surface of the substrate may be modified such that discrete sites are formed that can only have a single associated bead, or alternatively, the surface of the substrate is modified and beads may go down anywhere, but they end up at discrete sites.

In a preferred embodiment, the surface of the substrate is modified to contain wells, i.e. depressions in the surface of the substrate. This may be done as is generally known in the art using a variety of techniques, including, but not limited to, photolithography, stamping techniques, molding techniques and microetching techniques. As will be appreciated by those in the art, the technique used will depend on the composition and shape of the substrate.

In a preferred embodiment, physical alterations are made in a surface of the substrate to produce the sites. In a preferred embodiment, the substrate is a fiber optic bundle and the surface of the substrate is a terminal end of the fiber bundle, as is generally described in 08/818,199 and 09/151,877, both of which are hereby expressly incorporated by reference. In this embodiment, wells are made in a terminal or distal end of a fiber optic bundle comprising individual fibers. In this embodiment, the cores of the individual fibers are etched, with respect to the cladding, such that small wells or depressions are formed at one end of the fibers. The required depth of the wells will depend on the size of the beads to be added to the wells.

Generally in this embodiment, the microspheres are non-covalently associated in the wells, although the wells may additionally be chemically functionalized as is generally described below, cross-linking agents may be used, or a physical barrier may be used, i.e. a film or membrane over the beads.

In a preferred embodiment, the surface of the substrate is modified to contain chemically modified sites, that can be used to associate, either covalently or non-covalently, the microspheres of the invention to the discrete sites or locations on the substrate. "Chemically modified sites" in this context includes, but is not limited to, the addition of a pattern of chemical functional groups including amino groups, carboxy groups, oxo groups and thiol groups, that can be used to covalently attach microspheres, which generally also contain corresponding reactive functional groups; the addition of a pattern of adhesive that can be used to bind the microspheres (either by prior chemical functionalization for the addition of the adhesive or direct addition of the adhesive); the addition of a pattern of charged groups (similar to the chemical functionalities) for the electrostatic association of the microspheres, i.e. when the microspheres comprise charged groups opposite to the sites; the addition of a pattern of chemical functional groups that renders the sites differentially hydrophobic or hydrophilic, such that the addition of similarly hydrophobic or hydrophilic microspheres under suitable experimental conditions will result in association of the microspheres to the sites on the basis of hydroaffinity. For example, the use of hydrophobic sites with hydrophobic beads, in an aqueous system, drives the association of the beads preferentially onto the sites. As outlined above, "pattern" in this sense includes the use of a uniform treatment of the surface to allow association of the beads at discrete sites, as well as treatment of the surface resulting in discrete sites. As will be appreciated by those in the art, this may be accomplished in a variety of ways.

The compositions of the invention further comprise a population of microspheres. By "population" herein is meant a plurality of beads as outlined above for arrays. Within the population are separate subpopulations, which can be a single microsphere or multiple identical microspheres. That is, in some embodiments, as is more fully outlined below, the array may contain only a single bead for each bioactive agent; preferred embodiments utilize a plurality of beads of each type.

By "microspheres" or "beads" or "particles" or grammatical equivalents herein is meant small discrete particles. The composition of the beads will vary, depending on the class of bioactive agent and the method of synthesis. Suitable bead compositions include those used in peptide, nucleic acid and organic moiety synthesis, including, but not limited to, plastics, ceramics, glass, polystyrene, methylstyrene, acrylic polymers, paramagnetic materials, thoria sol, carbon graphite, titanium dioxide, latex or cross-linked dextrans such as Sepharose, cellulose, nylon, cross-linked micelles and Teflon may all be used. *"Microsphere Detection Guide"* from Bangs Laboratories, Fishers IN is a helpful guide.

The beads need not be spherical; irregular particles may be used. In addition, the beads may be porous, thus increasing the surface area of the bead available for either bioactive agent attachment or IBL attachment. The bead sizes range from nanometers, i.e. 100 nm, to millimeters, i.e. 1 mm, with

beads from about 0.2 micron to about 200 microns being preferred, and from about 0.5 to about 5 micron being particularly preferred, although in some embodiments smaller beads may be used.

It should be noted that a key component of the invention is the use of a substrate/bead pairing that allows the association or attachment of the beads at discrete sites on the surface of the substrate, such that the beads do not move during the course of the assay.

Each microsphere comprises a bioactive agent, although as will be appreciated by those in the art, there may be some microspheres which do not contain a bioactive agent, depending on the synthetic methods. By "candidate bioactive agent" or "bioactive agent" or "chemical functionality" or "binding ligand" herein is meant as used herein describes any molecule, *e.g.*, protein, oligopeptide, small organic molecule, coordination complex, polysaccharide, polynucleotide, etc. which can be attached to the microspheres of the invention. It should be understood that the compositions of the invention have two primary uses. In a preferred embodiment, as is more fully outlined below, the compositions are used to detect the presence of a particular target analyte; for example, the presence or absence of a particular nucleotide sequence or a particular protein, such as an enzyme, an antibody or an antigen. In an alternate preferred embodiment, the compositions are used to screen bioactive agents, *i.e.* drug candidates, for binding to a particular target analyte.

Bioactive agents encompass numerous chemical classes, though typically they are organic molecules, preferably small organic compounds having a molecular weight of more than 100 and less than about 2,500 Daltons. Bioactive agents comprise functional groups necessary for structural interaction with proteins, particularly hydrogen bonding, and typically include at least an amine, carbonyl, hydroxyl or carboxyl group, preferably at least two of the functional chemical groups. The bioactive agents often comprise cyclical carbon or heterocyclic structures and/or aromatic or polyaromatic structures substituted with one or more of the above functional groups. Bioactive agents are also found among biomolecules including peptides, nucleic acids, saccharides, fatty acids, steroids, purines, pyrimidines, derivatives, structural analogs or combinations thereof. Particularly preferred are nucleic acids and proteins.

Bioactive agents can be obtained from a wide variety of sources including libraries of synthetic or natural compounds. For example, numerous means are available for random and directed synthesis of a wide variety of organic compounds and biomolecules, including expression of randomized oligonucleotides. Alternatively, libraries of natural compounds in the form of bacterial, fungal, plant and animal extracts are available or readily produced. Additionally, natural or synthetically produced libraries and compounds are readily modified through conventional chemical, physical and biochemical means. Known pharmacological agents may be subjected to directed or random chemical

modifications, such as acylation, alkylation, esterification and/or amidification to produce structural analogs.

In a preferred embodiment, the bioactive agents are proteins. By "protein" herein is meant at least two covalently attached amino acids, which includes proteins, polypeptides, oligopeptides and peptides. The protein may be made up of naturally occurring amino acids and peptide bonds, or synthetic peptidomimetic structures. Thus "amino acid", or "peptide residue", as used herein means both naturally occurring and synthetic amino acids. For example, homo-phenylalanine, citrulline and norleucine are considered amino acids for the purposes of the invention. The side chains may be in either the (R) or the (S) configuration. In the preferred embodiment, the amino acids are in the (S) or L-configuration. If non-naturally occurring side chains are used, non-amino acid substituents may be used, for example to prevent or retard *in vivo* degradations.

In one preferred embodiment, the bioactive agents are naturally occurring proteins or fragments of naturally occurring proteins. Thus, for example, cellular extracts containing proteins, or random or directed digests of proteinaceous cellular extracts, may be used. In this way libraries of procaryotic and eukaryotic proteins may be made for screening in the systems described herein. Particularly preferred in this embodiment are libraries of bacterial, fungal, viral, and mammalian proteins, with the latter being preferred, and human proteins being especially preferred.

In a preferred embodiment, the bioactive agents are peptides of from about 5 to about 30 amino acids, with from about 5 to about 20 amino acids being preferred, and from about 7 to about 15 being particularly preferred. The peptides may be digests of naturally occurring proteins as is outlined above, random peptides, or "biased" random peptides. By "randomized" or grammatical equivalents herein is meant that each nucleic acid and peptide consists of essentially random nucleotides and amino acids, respectively. Since generally these random peptides (or nucleic acids, discussed below) are chemically synthesized, they may incorporate any nucleotide or amino acid at any position. The synthetic process can be designed to generate randomized proteins or nucleic acids, to allow the formation of all or most of the possible combinations over the length of the sequence, thus forming a library of randomized bioactive proteinaceous agents.

In a preferred embodiment, a library of bioactive agents are used. The library should provide a sufficiently structurally diverse population of bioactive agents to effect a probabilistically sufficient range of binding to target analytes. Accordingly, an interaction library must be large enough so that at least one of its members will have a structure that gives it affinity for the target analyte. Although it is difficult to gauge the required absolute size of an interaction library, nature provides a hint with the immune response: a diversity of 10^7 - 10^8 different antibodies provides at least one combination with sufficient affinity to interact with most potential antigens faced by an organism. Published *in vitro*

selection techniques have also shown that a library size of 10^7 to 10^8 is sufficient to find structures with affinity for the target. Thus, in a preferred embodiment, at least 10^6 , preferably at least 10^7 , more preferably at least 10^8 and most preferably at least 10^9 different bioactive agents are simultaneously analyzed in the subject methods. Preferred methods maximize library size and diversity.

In a preferred embodiment, the library is fully randomized, with no sequence preferences or constants at any position. In a preferred embodiment, the library is biased. That is, some positions within the sequence are either held constant, or are selected from a limited number of possibilities. For example, in a preferred embodiment, the nucleotides or amino acid residues are randomized within a defined class, for example, of hydrophobic amino acids, hydrophilic residues, sterically biased (either small or large) residues, towards the creation of cysteines, for cross-linking, prolines for SH-3 domains, serines, threonines, tyrosines or histidines for phosphorylation sites, etc., or to purines, etc.

In a preferred embodiment, the bioactive agents are nucleic acids (generally called "probe nucleic acids" or "candidate probes" herein). By "nucleic acid" or "oligonucleotide" or grammatical equivalents herein means at least two nucleotides covalently linked together. A nucleic acid of the present invention will generally contain phosphodiester bonds, although in some cases, as outlined below, nucleic acid analogs are included that may have alternate backbones, comprising, for example, phosphoramidate (Beaucage, *et al.*, *Tetrahedron*, **49**(10):1925 (1993) and references therein; Letsinger, *J. Org. Chem.*, **35**:3800 (1970); Sprinzl, *et al.*, *Eur. J. Biochem.*, **81**:579 (1977); Letsinger, *et al.*, *Nucl. Acids Res.*, **14**:3487 (1986); Sawai, *et al.*, *Chem. Lett.*, 805 (1984), Letsinger, *et al.*, *J. Am. Chem. Soc.*, **110**:4470 (1988); and Pauwels, *et al.*, *Chemica Scripta*, **26**:141 (1986)), phosphorothioate (Mag, *et al.*, *Nucleic Acids Res.*, **19**:1437 (1991); and U.S. Patent No. 5,644,048), phosphorodithioate (Briu, *et al.*, *J. Am. Chem. Soc.*, **111**:2321 (1989)), O-methylphosphoroamidite linkages (see Eckstein, *Oligonucleotides and Analogues: A Practical Approach*, Oxford University Press), and peptide nucleic acid backbones and linkages (see Egholm, *J. Am. Chem. Soc.*, **114**:1895 (1992); Meier, *et al.*, *Chem. Int. Ed. Engl.*, **31**:1008 (1992); Nielsen, *Nature*, **365**:566 (1993); Carlsson, *et al.*, *Nature*, **380**:207 (1996), all of which are incorporated by reference)). Other analog nucleic acids include those with positive backbones (Denpcy, *et al.*, *Proc. Natl. Acad. Sci. USA*, **92**:6097 (1995)); non-ionic backbones (U.S. Patent Nos. 5,386,023; 5,637,684; 5,602,240; 5,216,141; and 4,469,863; Kiedrowski, *et al.*, *Angew. Chem. Intl. Ed. English*, **30**:423 (1991); Letsinger, *et al.*, *J. Am. Chem. Soc.*, **110**:4470 (1988); Letsinger, *et al.*, *Nucleosides & Nucleotides*, **13**:1597 (1994); Chapters 2 and 3, ASC Symposium Series 580, "Carbohydrate Modifications in Antisense Research", Ed. Y.S. Sanghui and P. Dan Cook; Mesmaeker, *et al.*, *Bioorganic & Medicinal Chem. Lett.*, **4**:395 (1994); Jeffs, *et al.*, *J. Biomolecular NMR*, **34**:17 (1994); *Tetrahedron Lett.*, **37**:743 (1996)) and non-ribose backbones, including those described in U.S. Patent Nos. 5,235,033 and 5,034,506, and Chapters 6 and 7, ASC Symposium Series 580, "Carbohydrate Modifications in Antisense Research", Ed. Y.S. Sanghui and P. Dan Cook. Nucleic acids containing one or more carbocyclic sugars are also included within the definition of

nucleic acids (see Jenkins, *et al.*, Chem. Soc. Rev., (1995) pp. 169-176). Several nucleic acid analogs are described in Rawls, C & E News, June 2, 1997, page 35. All of these references are hereby expressly incorporated by reference. These modifications of the ribose-phosphate backbone may be done to facilitate the addition of additional moieties such as labels, or to increase the stability and half-life of such molecules in physiological environments; for example, PNA is particularly preferred. In addition, mixtures of naturally occurring nucleic acids and analogs can be made. Alternatively, mixtures of different nucleic acid analogs, and mixtures of naturally occurring nucleic acids and analogs may be made. The nucleic acids may be single stranded or double stranded, as specified, or contain portions of both double stranded or single stranded sequence. The nucleic acid may be DNA, both genomic and cDNA, RNA or a hybrid, where the nucleic acid contains any combination of deoxyribo- and ribo-nucleotides, and any combination of bases, including uracil, adenine, thymine, cytosine, guanine, inosine, xanthine, hypoxanthine, isocytosine, isoguanine, and base analogs such as nitropyrrole and nitroindole, etc.

In a preferred embodiment, the bioactive agents are libraries of clonal nucleic acids, including DNA and RNA. In this embodiment, individual nucleic acids are prepared, generally using conventional methods (including, but not limited to, propagation in plasmid or phage vectors, amplification techniques including PCR, etc.). The nucleic acids are preferably arrayed in some format, such as a microtiter plate format, and beads added for attachment of the libraries.

Attachment of the clonal libraries (or any of the nucleic acids outlined herein) may be done in a variety of ways, as will be appreciated by those in the art, including, but not limited to, chemical or affinity capture (for example, including the incorporation of derivatized nucleotides such as AminoLink or biotinylated nucleotides that can then be used to attach the nucleic acid to a surface, as well as affinity capture by hybridization), cross-linking, and electrostatic attachment, etc.

In a preferred embodiment, affinity capture is used to attach the clonal nucleic acids to the beads. For example, cloned nucleic acids can be derivatized, for example with one member of a binding pair, and the beads derivatized with the other member of a binding pair. Suitable binding pairs are as described herein for IBL/DBL pairs. For example, the cloned nucleic acids may be biotinylated (for example using enzymatic incorporation of biotinylated nucleotides, or by photoactivated cross-linking of biotin). Biotinylated nucleic acids can then be captured on streptavidin-coated beads, as is known in the art. Similarly, other hapten-receptor combinations can be used, such as digoxigenin and anti-digoxigenin antibodies. Alternatively, chemical groups can be added in the form of derivatized nucleotides, that can then be used to add the nucleic acid to the surface.

Preferred attachments are covalent, although even relatively weak interactions (i.e. non-covalent) can be sufficient to attach a nucleic acid to a surface, if there are multiple sites of attachment per each

nucleic acid. Thus, for example, electrostatic interactions can be used for attachment, for example by having beads carrying the opposite charge to the bioactive agent.

Similarly, affinity capture utilizing hybridization can be used to attach cloned nucleic acids to beads. For example, as is known in the art, polyA+RNA is routinely captured by hybridization to oligo-dT beads; this may include oligo-dT capture followed by a cross-linking step, such as psoralen crosslinking). If the nucleic acids of interest do not contain a polyA tract, one can be attached by polymerization with terminal transferase, or via ligation of an oligoA linker, as is known in the art.

Alternatively, chemical crosslinking may be done, for example by photoactivated crosslinking of thymidine to reactive groups, as is known in the art.

In general, special methods are required to decode clonal arrays, as is more fully outlined below.

As described above generally for proteins, nucleic acid bioactive agents may be naturally occurring nucleic acids, random nucleic acids, or "biased" random nucleic acids. For example, digests of procaryotic or eukaryotic genomes may be used as is outlined above for proteins.

In general, probes of the present invention are designed to be complementary to a target sequence (either the target analyte sequence of the sample or to other probe sequences, as is described herein), such that hybridization of the target and the probes of the present invention occurs. This complementarity need not be perfect; there may be any number of base pair mismatches that will interfere with hybridization between the target sequence and the single stranded nucleic acids of the present invention. However, if the number of mutations is so great that no hybridization can occur under even the least stringent of hybridization conditions, the sequence is not a complementary target sequence. Thus, by "substantially complementary" herein is meant that the probes are sufficiently complementary to the target sequences to hybridize under the selected reaction conditions. High stringency conditions are known in the art; see for example Maniatis et al., *Molecular Cloning: A Laboratory Manual*, 2d Edition, 1989, and *Short Protocols in Molecular Biology*, ed. Ausubel, et al., both of which are hereby incorporated by reference. Stringent conditions are sequence-dependent and will be different in different circumstances. Longer sequences hybridize specifically at higher temperatures. An extensive guide to the hybridization of nucleic acids is found in Tijssen, *Techniques in Biochemistry and Molecular Biology--Hybridization with Nucleic Acid Probes*, "Overview of principles of hybridization and the strategy of nucleic acid assays" (1993). Generally, stringent conditions are selected to be about 5-10 C lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength pH. The T_m is the temperature (under defined ionic strength, pH and nucleic acid concentration) at which 50% of the probes complementary to the target hybridize to the target sequence at equilibrium (as the target sequences are present in excess, at T_m , 50% of the probes are

occupied at equilibrium). Stringent conditions will be those in which the salt concentration is less than about 1.0 M sodium ion, typically about 0.01 to 1.0 M sodium ion concentration (or other salts) at pH 7.0 to 8.3 and the temperature is at least about 30 °C for short probes (e.g. 10 to 50 nucleotides) and at least about 60 °C for long probes (e.g. greater than 50 nucleotides). Stringent conditions may also be achieved with the addition of destabilizing agents such as formamide. In another embodiment, less stringent hybridization conditions are used; for example, moderate or low stringency conditions may be used, as are known in the art; see Maniatis and Ausubel, *supra*, and Tijssen, *supra*.

The term "target sequence" or grammatical equivalents herein means a nucleic acid sequence on a single strand of nucleic acid. The target sequence may be a portion of a gene, a regulatory sequence, genomic DNA, cDNA, RNA including mRNA and rRNA, or others. It may be any length, with the understanding that longer sequences are more specific. As will be appreciated by those in the art, the complementary target sequence may take many forms. For example, it may be contained within a larger nucleic acid sequence, i.e. all or part of a gene or mRNA, a restriction fragment of a plasmid or genomic DNA, among others. As is outlined more fully below, probes are made to hybridize to target sequences to determine the presence or absence of the target sequence in a sample. Generally speaking, this term will be understood by those skilled in the art.

In a preferred embodiment, the bioactive agents are organic chemical moieties, a wide variety of which are available in the literature.

In a preferred embodiment, each bead comprises a single type of bioactive agent, although a plurality of individual bioactive agents are preferably attached to each bead. Similarly, preferred embodiments utilize more than one microsphere containing a unique bioactive agent; that is, there is redundancy built into the system by the use of subpopulations of microspheres, each microsphere in the subpopulation containing the same bioactive agent.

As will be appreciated by those in the art, the bioactive agents may either be synthesized directly on the beads, or they may be made and then attached after synthesis. In a preferred embodiment, linkers are used to attach the bioactive agents to the beads, to allow both good attachment, sufficient flexibility to allow good interaction with the target molecule, and to avoid undesirable binding reactions.

In a preferred embodiment, the bioactive agents are synthesized directly on the beads. As is known in the art, many classes of chemical compounds are currently synthesized on solid supports, including beads, such as peptides, organic moieties, and nucleic acids.

In a preferred embodiment, the bioactive agents are synthesized first, and then covalently attached to the beads. As will be appreciated by those in the art, this will be done depending on the composition

of the bioactive agents and the beads. The functionalization of solid support surfaces such as certain polymers with chemically reactive groups such as thiols, amines, carboxyls, etc. is generally known in the art. Accordingly, "blank" microspheres may be used that have surface chemistries that facilitate the attachment of the desired functionality by the user. Some examples of these surface chemistries for blank microspheres include, but are not limited to, amino groups including aliphatic and aromatic amines, carboxylic acids, aldehydes, amides, chloromethyl groups, hydrazide, hydroxyl groups, sulfonates and sulfates.

These functional groups can be used to add any number of different candidate agents to the beads, generally using known chemistries. For example, candidate agents containing carbohydrates may be attached to an amino-functionalized support; the aldehyde of the carbohydrate is made using standard techniques, and then the aldehyde is reacted with an amino group on the surface. In an alternative embodiment, a sulfhydryl linker may be used. There are a number of sulfhydryl reactive linkers known in the art such as SPDP, maleimides, -haloacetyls, and pyridyl disulfides (see for example the 1994 Pierce Chemical Company catalog, technical section on cross-linkers, pages 155-200, incorporated herein by reference) which can be used to attach cysteine containing proteinaceous agents to the support. Alternatively, an amino group on the candidate agent may be used for attachment to an amino group on the surface. For example, a large number of stable bifunctional groups are well known in the art, including homobifunctional and heterobifunctional linkers (see Pierce Catalog and Handbook, pages 155-200). In an additional embodiment, carboxyl groups (either from the surface or from the candidate agent) may be derivatized using well known linkers (see the Pierce catalog). For example, carbodiimides activate carboxyl groups for attack by good nucleophiles such as amines (see Torchilin et al., Critical Rev. Therapeutic Drug Carrier Systems, 7(4):275-308 (1991), expressly incorporated herein). Proteinaceous candidate agents may also be attached using other techniques known in the art, for example for the attachment of antibodies to polymers; see Slinkin et al., Bioconj. Chem. 2:342-348 (1991); Torchilin et al., *supra*; Trubetskoy et al., Bioconj. Chem. 3:323-327 (1992); King et al., Cancer Res. 54:6176-6185 (1994); and Wilbur et al., Bioconjugate Chem. 5:220-235 (1994), all of which are hereby expressly incorporated by reference). It should be understood that the candidate agents may be attached in a variety of ways, including those listed above. Preferably, the manner of attachment does not significantly alter the functionality of the candidate agent; that is, the candidate agent should be attached in such a flexible manner as to allow its interaction with a target.

Specific techniques for immobilizing enzymes on microspheres are known in the prior art. In one case, NH₂ surface chemistry microspheres are used. Surface activation is achieved with a 2.5% glutaraldehyde in phosphate buffered saline (10 mM) providing a pH of 6.9. (138 mM NaCl, 2.7 mM, KCl). This is stirred on a stir bed for approximately 2 hours at room temperature. The microspheres are then rinsed with ultrapure water plus 0.01% tween 20 (surfactant) -0.02%, and rinsed again with a

pH 7.7 PBS plus 0.01% tween 20. Finally, the enzyme is added to the solution, preferably after being prefiltered using a 0.45 μ m amicon micropure filter.

In some embodiments, the microspheres may additionally comprise identifier binding ligands for use in certain decoding systems. By "identifier binding ligands" or "IBLs" herein is meant a compound that will specifically bind a corresponding decoder binding ligand (DBL) to facilitate the elucidation of the identity of the bioactive agent attached to the bead. That is, the IBL and the corresponding DBL form a binding partner pair. By "specifically bind" herein is meant that the IBL binds its DBL with specificity sufficient to differentiate between the corresponding DBL and other DBLs (that is, DBLs for other IBLs), or other components or contaminants of the system. The binding should be sufficient to remain bound under the conditions of the decoding step, including wash steps to remove non-specific binding. In some embodiments, for example when the IBLs and corresponding DBLs are proteins or nucleic acids, the dissociation constants of the IBL to its DBL will be less than about 10^{-4} - 10^{-6} M $^{-1}$, with less than about 10^{-5} to 10^{-9} M $^{-1}$ being preferred and less than about 10^{-7} - 10^{-9} M $^{-1}$ being particularly preferred.

IBL-DBL binding pairs are known or can be readily found using known techniques. For example, when the IBL is a protein, the DBLs include proteins (particularly including antibodies or fragments thereof (FABs, etc.)) or small molecules, or vice versa (the IBL is an antibody and the DBL is a protein). Metal ion- metal ion ligands or chelators pairs are also useful. Antigen-antibody pairs, enzymes and substrates or inhibitors, other protein-protein interacting pairs, receptor-ligands, complementary nucleic acids (including nucleic acid molecules that form triple helices), and carbohydrates and their binding partners are also suitable binding pairs. Nucleic acid - nucleic acid binding proteins pairs are also useful, including single-stranded or double-stranded nucleic acid binding proteins, and small molecule nucleic acid binding agents. Similarly, as is generally described in U.S. Patents 5,270,163, 5,475,096, 5,567,588, 5,595,877, 5,637,459, 5,683,867, 5,705,337, and related patents, hereby incorporated by reference, nucleic acid "aptamers" can be developed for binding to virtually any target; such an aptamer-target pair can be used as the IBL-DBL pair. Similarly, there is a wide body of literature relating to the development of binding pairs based on combinatorial chemistry methods.

In a preferred embodiment, the IBL is a protein, particularly an enzyme, that is conjugated to a bead. One advantage of using proteins/enzymes as IBLs is that many proteins and enzymes are commercially available in relatively pure preparations. Proteins contain many suitable functional groups for attachment to beads. Moreover, the identity of many small molecules including substrates that bind the proteins/enzymes are known. Such small molecules serve as the DBL. In addition, the small molecule can serve as the IBL and the enzyme as the DBL. As will be appreciated by those in the art, combinations of enzymes and small molecules can be used as well, or multiple enzymes and small molecules. Another potential advantage relates to the purification of the labeled proteins. If one

conjugates a fluorophore to a recrystallizable protein, then the unlabeled protein can be removed from labeled protein by use of crystallization. By filtering off the crystalline solid, the labeled protein will be concentrated in the filtrate.

In an alternative embodiment, antigen-antibody pairs are used as DBL-IBL combinations. In one embodiment antibodies are conjugated to different labels, defined below, that are then used to decode antigen-labeled beads.

In a preferred embodiment, the IBL is a molecule whose color or luminescence properties change in the presence of a selectively-binding DBL.

In one embodiment, the DBL may be attached to a bead, i.e. a "decoder bead", that may carry a label such as a fluorophore.

In a preferred embodiment, the IBL-DBL pair comprise substantially complementary single-stranded nucleic acids. In this embodiment, the binding ligands can be referred to as "identifier probes" and "decoder probes". Generally, the identifier and decoder probes range from about 4 basepairs in length to about 1000, with from about 6 to about 100 being preferred, and from about 8 to about 40 being particularly preferred. What is important is that the probes are long enough to be specific, i.e. to distinguish between different IBL-DBL pairs, yet short enough to allow both a) dissociation, if necessary, under suitable experimental conditions, and b) efficient hybridization.

In a preferred embodiment, as is more fully outlined below, the IBLs do not bind to DBLs. Rather, the IBLs are used as identifier moieties ("IMs") that are identified directly, for example through the use of mass spectroscopy.

Alternatively, in a preferred embodiment, the IBL and the bioactive agent are the same moiety; thus, for example, as outlined herein, particularly when no optical signatures are used, the bioactive agent can serve as both the identifier and the agent. For example, in the case of nucleic acids, the bead-bound probe (which serves as the bioactive agent) can also bind decoder probes, to identify the sequence of the probe on the bead. Thus, in this embodiment, the DBLs bind to the bioactive agents. This is particularly useful as this embodiment can give information about the array or the assay in addition to decoding. For example, as is more fully described below, the use of the DBLs allows array calibration and assay development. This may be done even if the DBLs are not used as such; for example in non-random arrays, the use of these probe sets can allow array calibration and assay development even if decoding is not required.

In a preferred embodiment, the microspheres do not contain an optical signature. That is, as outlined in U.S.S.N.s 08/818,199 and 09/151,877, previous work had each subpopulation of microspheres comprising a unique optical signature or optical tag that is used to identify the unique bioactive agent of that subpopulation of microspheres; that is, decoding utilizes optical properties of the beads such that a bead comprising the unique optical signature may be distinguished from beads at other locations with different optical signatures. Thus the previous work assigned each bioactive agent a unique optical signature such that any microspheres comprising that bioactive agent are identifiable on the basis of the signature. These optical signatures comprised dyes, usually chromophores or fluorophores, that were entrapped or attached to the beads themselves. Diversity of optical signatures utilized different fluorochromes, different ratios of mixtures of fluorochromes, and different concentrations (intensities) of fluorochromes.

Thus, the present invention does not rely solely on the use of optical properties to decode the arrays. However, as will be appreciated by those in the art, it is possible in some embodiments to utilize optical signatures as an additional coding method, in conjunction with the present system. Thus, for example, as is more fully outlined below, the size of the array may be effectively increased while using a single set of decoding moieties in several ways, one of which is the use in combination with optical signatures on beads. Thus, for example, using one "set" of decoding molecules, the use of two populations of beads, one with an optical signature and one without, allows the effective doubling of the array size. The use of multiple optical signatures similarly increases the possible size of the array.

In a preferred embodiment, each subpopulation of beads comprises a plurality of different IBLs. By using a plurality of different IBLs to encode each bioactive agent, the number of possible unique codes is substantially increased. That is, by using one unique IBL per bioactive agent, the size of the array will be the number of unique IBLs (assuming no "reuse" occurs, as outlined below). However, by using a plurality of different IBLs per bead, n , the size of the array can be increased to 2^n , when the presence or absence of each IBL is used as the indicator. For example, the assignment of 10 IBLs per bead generates a 10 bit binary code, where each bit can be designated as "1" (IBL is present) or "0" (IBL is absent). A 10 bit binary code has 2^{10} possible variants. However, as is more fully discussed below, the size of the array may be further increased if another parameter is included such as concentration or intensity; thus for example, if two different concentrations of the IBL are used, then the array size increases as 3^n . Thus, in this embodiment, each individual bioactive agent in the array is assigned a combination of IBLs, which can be added to the beads prior to the addition of the bioactive agent, after, or during the synthesis of the bioactive agent, i.e. simultaneous addition of IBLs and bioactive agent components.

Alternatively, when the bioactive agent is a polymer of different residues, i.e. when the bioactive agent is a protein or nucleic acid, the combination of different IBLs can be used to elucidate the sequence of the protein or nucleic acid.

Thus, for example, using two different IBLs (IBL1 and IBL2), the first position of a nucleic acid can be elucidated: for example, adenosine can be represented by the presence of both IBL1 and IBL2; thymidine can be represented by the presence of IBL1 but not IBL2, cytosine can be represented by the presence of IBL2 but not IBL1, and guanosine can be represented by the absence of both. The second position of the nucleic acid can be done in a similar manner using IBL3 and IBL4; thus, the presence of IBL1, IBL2, IBL3 and IBL4 gives a sequence of AA; IBL1, IBL2, and IBL3 shows the sequence AT; IBL1, IBL3 and IBL4 gives the sequence TA, etc. The third position utilizes IBL5 and IBL6, etc. In this way, the use of 20 different identifiers can yield a unique code for every possible 10-mer.

The system is similar for proteins but requires a larger number of different IBLs to identify each position, depending on the allowed diversity at each position. Thus for example, if every amino acid is allowed at every position, five different IBLs are required for each position. However, as outlined above, for example when using random peptides as the bioactive agents, there may be bias built into the system; not all amino acids may be present at all positions, and some positions may be preset; accordingly, it may be possible to utilize four different IBLs for each amino acid.

In this way, a sort of "bar code" for each sequence can be constructed; the presence or absence of each distinct IBL will allow the identification of each bioactive agent.

In addition, the use of different concentrations or densities of IBLs allows a "reuse" of sorts. If, for example, the bead comprising a first agent has a 1X concentration of IBL, and a second bead comprising a second agent has a 10X concentration of IBL, using saturating concentrations of the corresponding labelled DBL allows the user to distinguish between the two beads.

In addition, the use of different ratios of labeled to unlabeled IBLs allows for a differentiation of the intensity rather than the concentration of IBLs. For example, if a bead comprising a first agent has a ratio of 1:1 (labeled: unlabeled IBL), while a second bead comprising a second agent has a ratio of 1:10 (labeled: unlabeled IBL), the user will distinguish the different intensities between the two beads. This procedure offers the advantage that a broad spectrum of intensities of labels can be created using only a single label.

Once the microspheres comprising the candidate agents and the unique IBLs are generated, they are added to the substrate to form an array. It should be noted that while most of the methods described

herein add the beads to the substrate prior to the assay, the order of making, using and decoding the array can vary. For example, the array can be made, decoded, and then the assay done. Alternatively, the array can be made, used in an assay, and then decoded; this may find particular use when only a few beads need be decoded. Alternatively, the beads can be added to the assay mixture, i.e. the sample containing the target analytes, prior to the addition of the beads to the substrate; after addition and assay, the array may be decoded. This is particularly preferred when the sample comprising the beads is agitated or mixed; this can increase the amount of target analyte bound to the beads per unit time, and thus (in the case of nucleic acid assays) increase the hybridization kinetics. This may find particular use in cases where the concentration of target analyte in the sample is low; generally, for low concentrations, long binding times must be used.

In addition, adding the beads to the assay mixture can allow sorting or selection. For example, a large library of beads may be added to a sample, and only those beads that bind the sample may be added to the substrate. For example, if the target analyte is fluorescently labeled (either directly (for example by the incorporation of labels into nucleic acid amplification reactions) or indirectly (for example via the use of sandwich assays)), beads that exhibit fluorescence as a result of target analyte binding can be sorted via Fluorescence Activated Cell Sorting (FACS) and only these beads added to an array and subsequently decoded. Similarly, the sorting may be accomplished through affinity techniques; affinity columns comprising the target analytes can be made, and only those beads which bind are used on the array. Similarly, two bead systems can be used; for example, magnetic beads comprising the target analytes can be used to "pull out" those beads that will bind to the targets, followed by subsequent release of the magnetic beads (for example via temperature elevation) and addition to an array.

In general, the methods of making the arrays and of decoding the arrays is done to maximize the number of different candidate agents that can be uniquely encoded. The compositions of the invention may be made in a variety of ways. In general, the arrays are made by adding a solution or slurry comprising the beads to a surface containing the sites for association of the beads. This may be done in a variety of buffers, including aqueous and organic solvents, and mixtures. The solvent can evaporate, and excess beads removed.

In a preferred embodiment, when non-covalent methods are used to associate the beads to the array, a novel method of loading the beads onto the array is used. This method comprises exposing the array to a solution of particles (including microspheres and cells) and then applying energy, e.g. agitating or vibrating the mixture. This results in an array comprising more tightly associated particles, as the agitation is done with sufficient energy to cause weakly-associated beads to fall off (or out, in the case of wells). These sites are then available to bind a different bead. In this way, beads that exhibit a high affinity for the sites are selected. Arrays made in this way have two main advantages as

compared to a more static loading: first of all, a higher percentage of the sites can be filled easily, and secondly, the arrays thus loaded show a substantial decrease in bead loss during assays. Thus, in a preferred embodiment, these methods are used to generate arrays that have at least about 50% of the sites filled, with at least about 75% being preferred, and at least about 90% being particularly preferred. Similarly, arrays generated in this manner preferably lose less than about 20% of the beads during an assay, with less than about 10% being preferred and less than about 5% being particularly preferred.

In this embodiment, the substrate comprising the surface with the discrete sites is immersed into a solution comprising the particles (beads, cells, etc.). The surface may comprise wells, as is described herein, or other types of sites on a patterned surface such that there is a differential affinity for the sites. This differential affinity results in a competitive process, such that particles that will associate more tightly are selected. Preferably, the entire surface to be "loaded" with beads is in fluid contact with the solution. This solution is generally a slurry ranging from about 10,000:1 beads:solution (vol:vol) to 1:1. Generally, the solution can comprise any number of reagents, including aqueous buffers, organic solvents, salts, other reagent components, etc. In addition, the solution preferably comprises an excess of beads; that is, there are more beads than sites on the array. Preferred embodiments utilize two-fold to billion-fold excess of beads.

The immersion can mimic the assay conditions; for example, if the array is to be "dipped" from above into a microtiter plate comprising samples, this configuration can be repeated for the loading, thus minimizing the beads that are likely to fall out due to gravity.

Once the surface has been immersed, the substrate, the solution, or both are subjected to a competitive process, whereby the particles with lower affinity can be disassociated from the substrate and replaced by particles exhibiting a higher affinity to the site. This competitive process is done by the introduction of energy, in the form of heat, sonication, stirring or mixing, vibrating or agitating the solution or substrate, or both.

A preferred embodiment utilizes agitation or vibration. In general, the amount of manipulation of the substrate is minimized to prevent damage to the array; thus, preferred embodiments utilize the agitation of the solution rather than the array, although either will work. As will be appreciated by those in the art, this agitation can take on any number of forms, with a preferred embodiment utilizing microtiter plates comprising bead solutions being agitated using microtiter plate shakers.

The agitation proceeds for a period of time sufficient to load the array to a desired fill. Depending on the size and concentration of the beads and the size of the array, this time may range from about 1 second to days, with from about 1 minute to about 24 hours being preferred.

In a preferred embodiment, the substrate is pressed into a mixture of dry beads (although slurries may be used as well), and then tapped to remove excess beads.

It should be noted that not all sites of an array may comprise a bead; that is, there may be some sites on the substrate surface which are empty. In addition, there may be some sites that contain more than one bead, although this is not preferred.

In some embodiments, for example when chemical attachment is done, it is possible to associate the beads in a non-random or ordered way. For example, using photoactivatable attachment linkers or photoactivatable adhesives or masks, selected sites on the array may be sequentially rendered suitable for attachment, such that defined populations of beads are laid down.

The arrays of the present invention are constructed such that information about the identity of the candidate agent is built into the array, such that the random deposition of the beads in the fiber wells can be "decoded" to allow identification of the candidate agent at all positions. This may be done in a variety of ways, and either before, during or after the use of the array to detect target molecules.

Thus, after the array is made, it is "decoded" in order to identify the location of one or more of the bioactive agents, i.e. each subpopulation of beads, on the substrate surface.

In a preferred embodiment, a selective decoding system is used. In this case, only those microspheres exhibiting a change in the optical signal as a result of the binding of a target analyte are decoded. This is commonly done when the number of "hits", i.e. the number of sites to decode, is generally low. That is, the array is first scanned under experimental conditions in the absence of the target analytes. The sample containing the target analytes is added, and only those locations exhibiting a change in the optical signal are decoded. For example, the beads at either the positive or negative signal locations may be either selectively tagged or released from the array (for example through the use of photocleavable linkers), and subsequently sorted or enriched in a fluorescence-activated cell sorter (FACS). That is, either all the negative beads are released, and then the positive beads are either released or analyzed in situ, or alternatively all the positives are released and analyzed. Alternatively, the labels may comprise halogenated aromatic compounds, and detection of the label is done using for example gas chromatography, chemical tags, isotopic tags, or mass spectral tags.

As will be appreciated by those in the art, this may also be done in systems where the array is not decoded; i.e. there need not ever be a correlation of bead composition with location. In this embodiment, the beads are loaded on the array, and the assay is run. The "positives", i.e. those beads displaying a change in the optical signal as is more fully outlined below, are then "marked" to

distinguish or separate them from the “negative” beads. This can be done in several ways, preferably using fiber optic arrays. In a preferred embodiment, each bead contains a fluorescent dye. After the assay and the identification of the “positives” or “active beads”, light is shown down either only the positive fibers or only the negative fibers, generally in the presence of a light-activated reagent (typically dissolved oxygen). In the former case, all the active beads are photobleached. Thus, upon non-selective release of all the beads with subsequent sorting, for example using a fluorescence activated cell sorter (FACS) machine, the non-fluorescent active beads can be sorted from the fluorescent negative beads. Alternatively, when light is shown down the negative fibers, all the negatives are non-fluorescent and the the positives are fluorescent, and sorting can proceed. The characterization of the attached bioactive agent may be done directly, for example using mass spectroscopy.

Alternatively, the identification may occur through the use of identifier moieties (“IMs”), which are similar to IBLs but need not necessarily bind to DBLs. That is, rather than elucidate the structure of the bioactive agent directly, the composition of the IMs may serve as the identifier. Thus, for example, a specific combination of IMs can serve to code the bead, and be used to identify the agent on the bead upon release from the bead followed by subsequent analysis, for example using a gas chromatograph or mass spectroscope.

Alternatively, rather than having each bead contain a fluorescent dye, each bead comprises a non-fluorescent precursor to a fluorescent dye. For example, using photocleavable protecting groups, such as certain ortho-nitrobenzyl groups, on a fluorescent molecule, photoactivation of the fluorochrome can be done. After the assay, light is shown down again either the “positive” or the “negative” fibers, to distinguish these populations. The illuminated precursors are then chemically converted to a fluorescent dye. All the beads are then released from the array, with sorting, to form populations of fluorescent and non-fluorescent beads (either the positives and the negatives or vice versa).

In an alternate preferred embodiment, the sites of association of the beads (for example the wells) include a photopolymerizable reagent, or the photopolymerizable agent is added to the assembled array. After the test assay is run, light is shown down again either the “positive” or the “negative” fibers, to distinguish these populations. As a result of the irradiation, either all the positives or all the negatives are polymerized and trapped or bound to the sites, while the other population of beads can be released from the array.

In a preferred embodiment, the location of every bioactive agent is determined using decoder binding ligands (DBLs). As outlined above, DBLs are binding ligands that will either bind to identifier binding ligands, if present, or to the bioactive agents themselves, preferably when the bioactive agent is a nucleic acid or protein.

In a preferred embodiment, as outlined above, the DBL binds to the IBL.

In a preferred embodiment, the bioactive agents are single-stranded nucleic acids and the DBL is a substantially complementary single-stranded nucleic acid that binds (hybridizes) to the bioactive agent, termed a decoder probe herein. A decoder probe that is substantially complementary to each candidate probe is made and used to decode the array. In this embodiment, the candidate probes and the decoder probes should be of sufficient length (and the decoding step run under suitable conditions) to allow specificity; i.e. each candidate probe binds to its corresponding decoder probe with sufficient specificity to allow the distinction of each candidate probe.

In a preferred embodiment, the DBLs are either directly or indirectly labeled. By "labeled" herein is meant that a compound has at least one element, isotope or chemical compound attached to enable the detection of the compound. In general, labels fall into three classes: a) isotopic labels, which may be radioactive or heavy isotopes; b) magnetic, electrical, thermal; and c) colored or luminescent dyes; although labels include enzymes and particles such as magnetic particles as well. Examples of such labels include: peroxidase; alkaline phosphatase; biotin; urease; E. alactosidase; BSA/KLH; gold particles; quantum dots; redox indicators; pH indicators; E-lactamase; luciferase; TSA; SPA; chemiluminescence; sonoluminescence; fluorophores; phosphors; and, fluorescent metal ion sensors. Preferred labels include luminescent labels. In a preferred embodiment, the DBL is directly labeled, that is, the DBL comprises a label. In an alternate embodiment, the DBL is indirectly labeled; that is, a labeling binding ligand (LBL) that will bind to the DBL is used. In this embodiment, the labeling binding ligand-DBL pair can be as described above for IBL-DBL pairs. Suitable labels include, but are not limited to, fluorescent lanthanide complexes, including those of Europium and Terbium, fluorescein, rhodamine, tetramethylrhodamine, eosin, erythrosin, coumarin, methyl-coumarins, pyrene, Malachite green, stilbene, Lucifer Yellow, Cascade Blue™, Texas Red, FITC, PE, cy3, cy5 and others described in the 6th Edition of the Molecular Probes Handbook by Richard P. Haugland, hereby expressly incorporated by reference. In addition, ion-sensitive dyes such as bis-BITC, Newport green, fluo 3, fura red find use in the invention.

In one embodiment, the label is a molecule whose color or luminescence properties change in the presence of the IBL, due to a change in the local environment. For example, the label may be: (1) a fluorescent pH indicator whose emission intensity changes with pH; (2) a fluorescent ion indicator, whose emission properties change with ion concentration; or (3) a fluorescent molecule such as an ethidium salt whose fluorescence intensity increases in hydrophobic environments.

In one embodiment ion sensitive dyes are used for decoding. That is, dyes that fluoresce in the presence of certain ions are used. In a preferred embodiment the ion sensitive dyes are attached to the beads. An advantage of attaching dyes directly to microspheres is that the cost associated with oligonucleotide synthesis is avoided.

Accordingly, when such dyes are attached to the beads they can be used to encode a specific nucleotide sequence, or the identity of the bioactive agent. Once the beads are distributed on the array, the array is contacted with buffers that include certain ions. In the presence of appropriate ion, certain dyes will fluoresce and fluorescence is detected. In a preferred embodiment, the array is contacted with different buffers sequentially. Each buffer contains a different ion. Preferably the array is washed between buffer solutions so that any signal produced as a result of the contact with the ion solution is diminished.

In this embodiment multiple different ion sensitive dyes can be used. That is, preferably, more than 2 with more than 3 or 4 dyes are attached to the microspheres. In some embodiments, more than 10 dyes are attached to the microsphere.

One advantage to using ion sensitive dyes is that fluorescence bleed-through is diminished. That is, because the response of the dyes is either positive or negative, i.e. they interact or they don't, bleed-through is avoided.

In a preferred embodiment, decoding of self-assembled random arrays is done on the basis of pH titration. In this embodiment, in addition to bioactive agents, the beads comprise optical signatures, wherein the optical signatures are generated by the use of pH-responsive dyes (sometimes referred to herein as "pH dyes") such as fluorophores. This embodiment is similar to that outlined in PCT US98/05025 and U.S.S.N. 09/151,877, both of which are expressly incorporated by reference, except that the dyes used in the present invention exhibits changes in fluorescence intensity (or other properties) when the solution pH is adjusted from below the pKa to above the pKa (or vice versa). In a preferred embodiment, a set of pH dyes is used, each with a different pKa, preferably separated by at least 0.5 pH units. Preferred embodiments utilize a pH dye set of pKa's of 2.0, 2.5, 3.0, 3.5, 4.0, 4.5, 5.0, 5.5, 6.0, 6.5, 7.0, 7.5, 8.0, 8.5, 9.0, 9.5, 10.0, 10.5, 11, and 11.5. Each bead can contain any subset of the pH dyes, and in this way a unique code for the bioactive agent is generated. Thus, the decoding of an array is achieved by titrating the array from pH 1 to pH 13, and measuring the fluorescence signal from each bead as a function of solution pH.

In one embodiment, the decoder probes are labeled in situ; that is, they need not be labeled prior to the decoding reaction. In this embodiment, the incoming decoder probe is shorter than the candidate probe, creating a 5' "overhang" on the decoding probe. The addition of labeled ddNTPs (each labeled with a unique tag) and a polymerase will allow the addition of the tags in a sequence specific manner, thus creating a sequence-specific pattern of signals. Similarly, other modifications can be done, including ligation, etc.

In one embodiment, a series of sequential extension and denaturing stages allows for positional decoding of the beads. That is, following one round of extension, the DBL is removed or stripped from

the IBL and a second round of extension is performed. The contents of any given primer extension reaction is shown in Figure 8.

In the primer extension method, up to four fluorescently-labeled dideoxynucleotide terminators each with a different colored label (e.g. using ABI's dRhodamine® terminators such as dR110, dR6G, dTAMRA, dROX or Big Dye® Terminators) could be used for 'painting'. The number of unique beads that can be decoded is given by $N = y^x$, where n = the number of unique codes, y = the number of colors and x = the number of stages or primer extension reactions. This formula yields 4096 unique codes using 6 primer extension stages and four colors (i.e. $4096 = 4^6$).

In a preferred embodiment the bioactive agent is synthesized with encoding sequences that allows decoding to be accomplished independent of the useable portion of the bioactive agent, i.e. nucleic acid sequence. By "useable" portion is meant the adapter sequence or target sequence. In a preferred embodiment the bioactive agent also is synthesized with a primer region to facilitate sequencing or primer extension analysis of the encoding sequence (see Figure 12). As shown in Figure 12, a primer complementary to a portion of the ssDNA on the bead is extended by single base polymerase extension using dye-labeled chain terminating nucleotides. After incorporation of a labeled, i.e. fluorescent, chain terminator and reading of the array, i.e. detecting the label, the fluorescence is removed. Removal of the fluorescence is accomplished directly when reversible chain terminators are used (see Canard and Sarfati, Gene, 148(1): 1-6, 1994, which is expressly incorporated herein by reference). However, in a preferred embodiment the primer is dissociated from the bead and a subsequent primer is annealed. Preferably the subsequent primer is 1 nucleotide longer than the former primer so that an additional nucleotide in the encoding sequence can be interrogated. In a preferred embodiment the subsequent primer is lengthened by the addition of a degenerate base. This can be accomplished either by incorporating all 4 nucleotides into the additional position of the primer, or by the incorporation of a universal base into the primer. The new primer is then subject to extension as described above. The primer extension can be repeated as many times as are necessary to decode the microsphere.

Decoding by sequencing has the advantage that encoding a vast number of distinct codes by using only a small number of nucleotides. That is, the information content of the encoding region depends on the number of combinations of the bases used, i.e. 4 when A, C, G and T are used, that are present at each position and the length of the encoding region. When 4 nucleotides are used, there are up to fifteen possible combinations of nucleotides at each position (A, C, G, T, A/C, A/G, A/T, C/G, C/T, G/T, A/C/G, A/C/T, A/G/T, C/G/G and A/C/G/T). As such, the information content is 15^n , where n is equal to the length of the encoding sequence. As such, for example, with just 3 base positions, 3375 different codes can be generated.

In addition an additional preferred embodiment includes the combination of decoding by sequencing with decoding by hybridization. In one application of decoding by sequencing, universal primers, each one base longer than the previous, are sequentially hybridized to the array. The base adjacent to the 3' end of the universal primer is determined by a single base extension reaction employing labeled-chain terminators, preferably using four different color labels one for each base. Since a hybridization event must be performed for each base read, one can improve the decoding power by using the hybridization event to also decode. However, rather than using universal primers, zipcode sequences would be used instead (see Figure 13)

The placement of the encoding sequences can be either on the "bead-side" of the gene-specific sequence or on the terminal side. Placement of the encoding sequences on the terminal side offers that advantage of incorporating a cleavable linker between the encoding sequences and the gene-specific sequence (Fig. 14). After completion of decoding, the encoding sequences can be removed to prevent possible interference in the analytical assay.

In one embodiment following decoding, the label is extinguished. By "extinguished" is meant that the signal from the label is eliminated. This is accomplished by a variety of techniques including cleaving the label or destroying the signal of the label. That is, in one embodiment, the label is attached to the DBL via a cleavable linker. Following a decoding step, the linker is cleaved thereby releasing the label. In one embodiment cleavage of the label results in deprotection of the bioactive agent, for example when the bioactive agent contains a protecting group. Alternatively, cleavage of the label preserves the nature of the bioactive agent. That is, what is important is that cleavage of the label does not interfere with the bioactive agent.

In an alternative embodiment, the label is not cleaved, but rather is eliminated by destroying the signal of the label. That is, following decoding, the label is contacted with an agent that degrades or destroys the signal. For example, organic fluorescent dyes are subject to chemical degradation under appropriate conditions. Thus, following a decoding step where the fluorescent signal of the dyes are detected, the fluorophore is destroyed by an appropriate chemical solution that degrades the fluorescent molecule sufficient leaving no trace of fluorescence. In this way, beads are tagged with fluorescent labels, i.e. barcodes of sorts, their location or identity deciphered and recorded. The beads are then exposed to the "fluorescent destroying solution" and the fluorescence is eliminated, allowing an analytical assay to be conducted. Suitable fluorescent dyes include but are not limited to various bodipy, or Alexa dyes from Molecular Probes. The "fluorescent destroying solution" is a reducing or oxidizing solution, such as 0.01 M sodium periodate.

In a preferred embodiment not only is the label cleaved, but the encoding nucleic acid is cleaved. This method finds particular use in removing residual signal following a decoding step. Accordingly, in a preferred embodiment a cleavable linker can be used to bridge the adapter and beads. After

decoding, the adapter oligonucleotide(s) can be easily removed through cleavable linkers. There are many commercially available cleavable linkers. They are routinely used in the combinatorial solid-phase synthesis field. The linkers can be cleaved under acid, base or photo radiation conditions.

In addition, restriction enzymes can be used to cleave a linker. Alternatively, because of the sequence specificity of restriction enzymes, they also are used to cleave encoding nucleic acids. A change in signal following incubation with a sequence specific enzyme decodes the bead. An example is illustrative.

Consider the decoding of 6 bead types with three zip codes or adapters. In general, the strategy will have $B = (2^n) - 2$. Where B is the number of types encoded and n is the number of zip codes. The bead types are z1, z2, z3, z1z2, z1z3, and z2z3. Here zizj indicates that both codes are on a bead. Every code is already dye labeled.

Generally, each code is designed with a site that can be specifically cut with some agent. This includes, but is not limited to, restriction enzymes. For ease of description, the process using restriction enzymes and zip codes that contain specific restriction sites is described. When restriction enzymes are used, the zip codes are double stranded. Each stage would consist of sequentially adding one restriction enzyme. An image is taken prior to stage 1, and the intensity of every bead is measured. Restriction enzyme 1 (Re1) is added. New intensity measurements are taken. The information of which beads decrease in intensity, and which remain the same. Detecting the pattern of beads for which the signal "decreased" and "stayed the same" uniquely determines the bead type. If we denote "decreased" by 1 and "stayed the same" by 0, then the decoding would give the following codes:

z1 : 1,0,0

z2 : 0,1,0

z3 : 0,0,1

z1z2 : 1,1,0

z1z3 : 1,0,1

z2z3 : 0,1,1.

The general approach uses n codes, which would give $(2^n) - 2$ types. Subtracting 2 leaves off the all off state, and the all zip codes used state.

In another embodiment the array is decoded by using exonuclease enzyme. In this embodiment the microspheres are encoded with dye-labeled nucleotides. Preferably each nucleotide is labeled with a discrete label. In this particular embodiment it is desirable to protect the bioactive agent from cleavage. This can be accomplished by a variety of known methods including hybridizing the bioactive agent to a complementary sequence or blocking the terminus of the bioactive agent. Alternatively, if

the exonuclease enzyme operates from 5' to 3', one can design zipcodes from 5' to 3' and SNP-gene specific oligos from 3' to 5'. This way zipcode sequences will be operated on by the exonuclease, but the SNP/gene-specific sequences will remain intact.

The microsphere is then contacted with an exonuclease that successively removes the labeled nucleotides of the encoding sequence. Monitoring of the signal of the microspheres should begin upon addition of the exonuclease and monitored in all channels that correspond to the signal emitted by the labels in use. For example, assume the microsphere has all the zipcodes available (with the above specifications), and is in a solution. At time $T=0$, the exonuclease enzyme is applied to the solution. At the same time, the acquisition system (CCD) starts taking images every few seconds (Figure 15). The assumption is that the reaction of the exonuclease is slow enough, so one would have ample time to take several images, as the reaction is proceeding. If this is not the case, the reaction can be slowed down by increasing the stringency conditions.

The exonuclease removes the nucleotides, one at a time. Therefore, ideally, one would notice a step reduction in the signal of the appropriate channel, once the nucleotide is removed (and moved away from the focal zone). The complete set of waveforms (i.e., the four color channels, and from T_0 to T_{end}) will provide a unique signature for every zipcode.

In practice, the removal of the nucleotides will not be completely synchronized on different oligos (as shown in Figure 15). For instance, for two oligos on the same bead (same zipcode), the exonuclease may have removed one nucleotide from one, and two nucleotides from the other. However, since one monitors many oligos from each bead, the overall trend should be preserved, i.e., in general the nucleotides closer to the end should be removed first. This means that in general, although the transitions on the waveforms may not be sharp, the signatures could still be unique.

Accordingly, the identification of the location of the individual beads (or subpopulations of beads) is done using one or more decoding steps comprising a binding between the labeled DBL and either the IBL or the bioactive agent (i.e. a hybridization between the candidate probe (sometimes referred to herein as a "decoding sequence") and the decoder probe when the bioactive agent is a nucleic acid). After decoding, the DBLs can be removed and the array can be used; however, in some circumstances, for example when the DBL binds to an IBL and not to the bioactive agent, the removal of the DBL is not required (although it may be desirable in some circumstances). In addition, as outlined herein, decoding may be done either before the array is used in an assay, during the assay, or after the assay.

In one embodiment, a single decoding step is done. In this embodiment, each DBL is labeled with a unique label, such that the the number of unique labels is equal to or greater than the number of bioactive agents (although in some cases, "reuse" of the unique labels can be done, as described herein; similarly, minor variants of candidate probes can share the same decoder, if the variants are

encoded in another dimension, i.e. in the bead size or label). For each bioactive agent or IBL, a DBL is made that will specifically bind to it and contains a unique label, for example one or more fluorochromes. Thus, the identity of each DBL, both its composition (i.e. its sequence when it is a nucleic acid) and its label, is known. Then, by adding the DBLs to the array containing the bioactive agents under conditions which allow the formation of complexes (termed hybridization complexes when the components are nucleic acids) between the DBLs and either the bioactive agents or the IBLs, the location of each DBL can be elucidated. This allows the identification of the location of each bioactive agent; the random array has been decoded. The DBLs can then be removed, if necessary, and the target sample applied.

In a preferred embodiment, the number of unique labels is less than the number of unique bioactive agents, and thus a sequential series of decoding steps are used. To facilitate the discussion, this embodiment is explained for nucleic acids, although other types of bioactive agents and DBLs are useful as well. In this embodiment, decoder probes are divided into n sets for decoding. The number of sets corresponds to the number of unique tags. Each decoder probe is labeled in n separate reactions with n distinct tags. All the decoder probes share the same n tags. Each pool of decoders contains only one of the n tag versions of each decoder, and no two decoder probes have the same sequence of tags across all the pools. The number of pools required for this to be true is determined by the number of decoder probes and the n . Hybridization of each pool to the array generates a signal at every address comprising an IBL. The sequential hybridization of each pool in turn will generate a unique, sequence-specific code for each candidate probe. This identifies the candidate probe at each address in the array. For example, if four tags are used, then $4 \times n$ sequential hybridizations can ideally distinguish 4^n sequences, although in some cases more steps may be required. After the hybridization of each pool, the hybrids are denatured and the decoder probes removed, so that the probes are rendered single-stranded for the next hybridization (although it is also possible to hybridize limiting amounts of target so that the available probe is not saturated. Sequential hybridizations can be carried out and analyzed by subtracting pre-existing signal from the previous hybridization).

An example is illustrative. Assuming an array of 16 probe nucleic acids (numbers 1-16), and four unique tags (four different fluors, for example; labels A-D). Decoder probes 1-16 are made that correspond to the probes on the beads. The first step is to label decoder probes 1-4 with tag A, decoder probes 5-8 with tag B, decoder probes 9-12 with tag C, and decoder probes 13-16 with tag D. The probes are mixed and the pool is contacted with the array containing the beads with the attached candidate probes. The location of each tag (and thus each decoder and candidate probe pair) is then determined. The first set of decoder probes are then removed. A second set is added, but this time, decoder probes 1, 5, 9 and 13 are labeled with tag A, decoder probes 2, 6, 10 and 14 are labeled with tag B, decoder probes 3, 7, 11 and 15 are labeled with tag C, and decoder probes 4, 8, 12 and 16 are labeled with tag D. Thus, those beads that contained tag A in both decoding steps contain candidate probe 1; tag A in the first decoding step and tag B in the second decoding step contain candidate probe 2; tag A in the first decoding step and tag C in the second step contain candidate probe 3; etc.

As will be appreciated by those in the art, the decoder probes can be made in any order and added in any order.

In addition, since the size of the array will be set by the number of unique decoding binding ligands, there are a number of techniques that allow new combinations and the "reuse" of DBLs to allow for a greater number of test sites. This may be done in a wide variety of ways, as outlined below.

In a preferred embodiment, a spatial or positional coding system is done. In this embodiment, there are sub-bundles or subarrays (i.e. portions of the total array) that are utilized. By analogy with the telephone system, each subarray is an "area code", that can have the same labels (i.e. telephone numbers) of other subarrays, that are separated by virtue of the location of the subarray. Thus, for example, the same unique labels can be reused from bundle to bundle. Thus, the use of 50 unique labels in combination with 100 different subarrays can form an array of 5000 different bioactive agents. In this embodiment, it becomes important to be able to identify one bundle from another; in general, this is done either manually or through the use of marker beads; these can be beads containing unique tags for each subarray, or the use of the same marker bead in differing amounts, or the use of two or more marker beads in different ratios.

In a preferred embodiment, the coding and decoding is accomplished by sequential loading of the microspheres into the array. As outlined above for spatial coding, in this embodiment, the optical signatures can be "reused". In this embodiment, the library of microspheres each comprising a different bioactive agent (or the subpopulations each comprise a different bioactive agent), is divided into a plurality of sublibraries; for example, depending on the size of the desired array and the number of unique tags, 10 sublibraries each comprising roughly 10% of the total library may be made, with each sublibrary comprising roughly the same unique tags. Then, the first sublibrary is added to the fiber optic bundle comprising the wells, and the location of each bioactive agent is determined, generally through the use of DBLs. The second sublibrary is then added, and the location of each bioactive agent is again determined. The signal in this case will comprise the signal from the "first" DBL and the "second" DBL; by comparing the two matrices the location of each bead in each sublibrary can be determined. Similarly, adding the third, fourth, etc. sublibraries sequentially will allow the array to be filled.

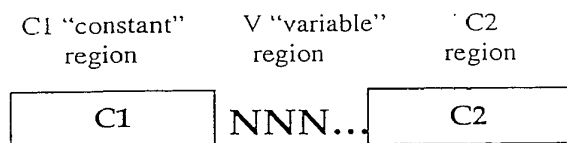
In a preferred embodiment, codes can be "shared" in several ways. In a first embodiment, a single code (i.e. IBL/DBL pair) can be assigned to two or more agents if the target analytes differ sufficiently in their binding strengths. For example, two nucleic acid probes used in an mRNA quantitation assay can share the same code if the ranges of their hybridization signal intensities do not overlap. This can occur, for example, when one of the target sequences is always present at a much higher concentration than the other. Alternatively, the two target sequences might always be present at a similar concentration, but differ in hybridization efficiency.

Alternatively, a single code can be assigned to multiple agents if the agents are functionally equivalent. For example, if a set of oligonucleotide probes are designed with the common purpose of detecting the presence of a particular gene, then the probes are functionally equivalent, even though they may differ in sequence. Similarly, if classes or "families" of analytes are desired, all probes for different members of a class such as kinases or G-protein coupled receptors could share a code. Similarly, an array of this type could be used to detect homologs of known genes. In this embodiment, each gene is represented by a heterologous set of probes, hybridizing to different regions of the gene (and therefore differing in sequence). The set of probes share a common code. If a homolog is present, it might hybridize to some but not all of the probes. The level of homology might be indicated by the fraction of probes hybridizing, as well as the average hybridization intensity. Similarly, multiple antibodies to the same protein could all share the same code.

In a preferred embodiment, the set of codes can be increased using a kind of degeneracy on the IBLs on the beads. For most of the systems described herein, particularly when the IBL is a nucleic acid (including the case where the nucleic acid IBL is also the candidate agent), any single bead contains a single IBL sequence (decoding sequence). However, in this embodiment, it is possible to have a single bead comprise two or more decoding sequences, that differ at a particular nucleotide position. For a degeneracy of two at position X, there are 10 different possible combinations: the decoding sequence has an A at position X, a T, a C, a G, or a mixture of decoding sequences with either A/T, A/G, A/C, C/G, C/T or G/T. By using decoder probes that have different labels for each different nucleotide at the position, the number of codes can be increased, since beads containing the two different sequences will exhibit a mixture of signals. In addition, as will be appreciated by those in the art, rather than a "degeneracy" of two, degeneracies of three or four can be done as well. Thus, a bead containing three different decoding sequences will exhibit a signal comprising three different colors, etc.

Decoding the degenerate probes is done by a series of hybridizations to labeled oligonucleotides. Each hybridization decodes one position in the probe so the total number of hybridizations needed to decode a library is equal to the length of the variable region. Each hybridization is done with four oligonucleotides, each designed to hybridize specifically to a sequence containing a specific base at the interrogated position and each tagged with a distinctly colored fluorescent dye.

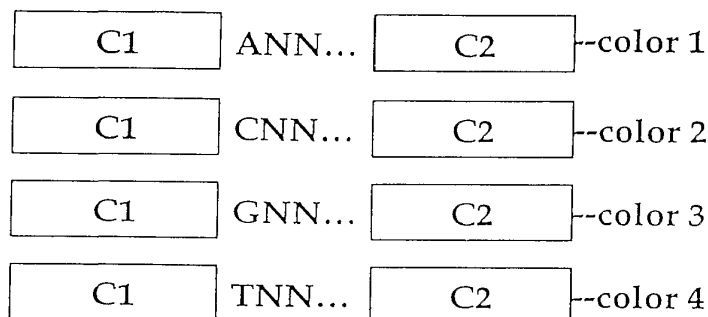
In a preferred embodiment, a degenerate probe sequence is attached to one or more constant probe regions. By "constant" probe region is meant a portion of a probe that is invariant among a population of decoder probes. In a particularly preferred embodiment, the degenerate or variable region is flanked on either side by constant regions. As such, each oligonucleotide has the following structure:



Although the C1 and C2 regions may be any length, preferably each is from 5 to 50 nucleotides in length. More preferably each is from 7 to 30 nucleotides in length and most preferably, each is about 10 bp in length. In one embodiment, the constant regions do not vary in sequence among the probes designed to decode a particular set of beads. As such, by using constant and variable regions (i.e. "CV" or "CVC" when two constant regions are used) in a set of probes, an array specific set or library of probes is generated.

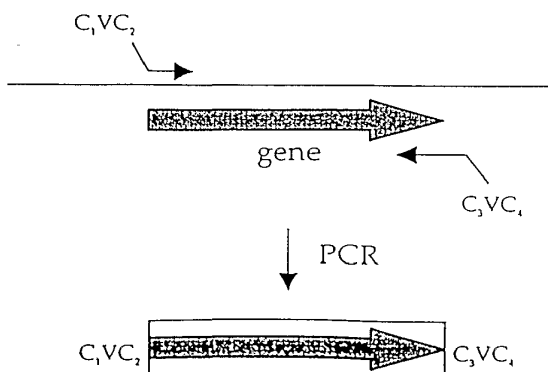
The variable region contains a sequence that is unique to each address. Each "N" or variable position within the variable region can contain one of ten possibilities as described above. As such, the number of unique addresses in a library is determined by the length of the variable region: if L is length of the variable region, the number of unique addresses, A, is $A=10^L$.

Accordingly, decoding the first base of the variable region of a decoding sequence is illustrated below:



Any address with an A at the first position would hybridize with the fourth oligonucleotide and give the corresponding color. An address with an A/C at the first position would hybridize with both the third and fourth decoders and give the corresponding combination of colors. By serial hybridizations, each using four decoders and each interrogating a different position of the V region, the entire sequence of each address can be determined. The "N" positions in the decoders are synthesized with degeneracies of the four standard bases.

In one embodiment, CVC probes are incorporated³⁸ into the probes of an oligonucleotide array during synthesis as is described for the addition of adapters; see U.S.S.Ns 09/553,993, filed April 20, 2000, 09/556,463, filed April 21, 2000, 60/160,917, filed October 22, 1999, and 60/135,123, filed May 20, 1999 and WO 00/63437, all of which are expressly incorporated herein by reference. In a similar fashion, the CVC probes could be synthesized in the 5-prime ends of PCR primers used to generate a PCR-product-based array. By incorporating different probes into the two primers of each PCR reaction, a combinatorial expansion of addresses is possible:



The constant regions in the C_1VC_2 addresses would be different in sequence from the C_3VC_4 addresses. Each would therefore be decoded with its own set of CVC complement decoders. If C_1VC_2 has a 3-base V region (1000 addresses) and C_3VC_4 has a 3-base V region (1,000 addresses), then the combination results in 1,000 x 1,000, or one million addresses.

In a preferred embodiment, there are additional ways to increase the number of unique or distinct tags. That is, the use of distinct attributes on each bead can be used to increase the number of codes. In addition, sequential decoding allows a reuse of codes in new ways. These attributes are independent of each other, thus allowing the number of codes to grow exponentially as a function of the number of decoding steps and the number of attributes (e.g. distinct codes). However, by increasing the amount of decoding information obtained in a single decoding step, the number of decoding steps is markedly reduced. By increasing the number of attributes per decoding step, fewer decoding steps are required for a given number of codes. Thus, in a preferred embodiment, a variety of methods are used to generate a number of codes for use in the process of decoding the arrays, while minimizing the necessary decoding steps. For example, a variety of different coding strategies can be combined: thus, different "colors", combinations of colors ("hues"), different intensities of colors or hues or both, different colors and different bead sizes, etc. can all be combined. Alternatively, the number of distinct codes is markedly increased.

In a preferred embodiment DBLs rely on attaching or embedding a quantitative or discrete set of physical attributes to the bead, i.e. labeling the bead. Preferred physical attributes of a bead include but are not limited to: surface "smoothness" or "roughness", color (fluorescent and otherwise), color

intensity, size, detectable chemical moieties, chemical reactivity, magnetization, pH sensitivity, energy transfer efficiency between dyes present, hydrophobicity, hydrophilicity, absorptivity, charge, pH sensitivity, etc. In general, physical attributes that can be measured optically are preferred, including color, size, surface smoothness etc.

In a preferred embodiment, the DBLs may be reused by having some subpopulations of beads comprise optical signatures. In a preferred embodiment, the optical signature is generally a mixture of reporter dyes, preferably fluorescent. By varying both the composition of the mixture (i.e. the ratio of one dye to another) and the concentration of the dye (leading to differences in signal intensity), matrices of unique optical signatures may be generated. This may be done by covalently attaching the dyes to the surface of the beads, or alternatively, by entrapping the dye within the bead. The dyes may be chromophores or phosphors but are preferably fluorescent dyes, which due to their strong signals provide a good signal-to-noise ratio for decoding. Suitable dyes for use in the invention include those listed for labeling DBLs, above.

In a preferred embodiment, the encoding can be accomplished in a ratio of at least two dyes, although more encoding dimensions may be added in the size of the beads, for example. In addition, the labels are distinguishable from one another; thus two different labels may comprise different molecules (i.e. two different fluorors) or, alternatively, one label at two different concentrations or intensity.

In a preferred embodiment, the dyes are covalently attached to the surface of the beads. This may be done as is generally outlined for the attachment of the bioactive agents, using functional groups on the surface of the beads. As will be appreciated by those in the art, these attachments are done to minimize the effect on the dye.

In a preferred embodiment, the dyes are non-covalently associated with the beads, generally by entrapping the dyes in the pores of the beads.

Additionally, encoding in the ratios of the two or more dyes, rather than single dye concentrations, is preferred since it provides insensitivity to the intensity of light used to interrogate the reporter dye's signature and detector sensitivity.

In alternative embodiments, additional encoding parameters can be added, such as microsphere size. For example, the use of different size beads may also allow the reuse of sets of DBLs; that is, it is possible to use microspheres of different sizes to expand the encoding dimensions of the microspheres. Optical fiber arrays can be fabricated containing pixels with different fiber diameters or cross-sections; alternatively, two or more fiber optic bundles, each with different cross-sections of the individual fibers, can be added together to form a larger bundle; or, fiber optic bundles with fiber of the same size cross-sections can be used, but just with different sized beads. With different diameters, the largest wells can be filled with the largest microspheres and then moving onto progressively

smaller microspheres in the smaller wells until all size wells are then filled. In this manner, the same dye ratio could be used to encode microspheres of different sizes thereby expanding the number of different oligonucleotide sequences or chemical functionalities present in the array. Although outlined for fiber optic substrates, this as well as the other methods outlined herein can be used with other substrates and with other attachment modalities as well.

A bead decoding scheme includes assigning/imbuing a single quantifiable attribute to each bead type wherein each bead type differs in the quantifiable value of that attribute. For instance, one can attach a given number of fluorophores to a bead and quantitate the number of attached fluorophores in the decoding process; however, in practice, attaching a "given amount" of an attribute to a bead and accurately measuring the attribute may be problematic. Typically, the coefficient of variation (CV) seen in associating an attribute and the resultant measurement of this attribute is quite large ($> 20\%$). By coefficient of variation is meant the variability in labeling a bead in successive labelings. This CV can be determined by labeling beads with a defined given number of label (fluorophore, for example) in multiple tests and measuring the resulting signal emitted by the bead. This large CV limits the number of useable and resolvable "levels" for any given attribute.

A more robust decoding scheme employs ratiometric rather than absolute measurements for segmenting a quantitative attribute into codes. By ratiometric decoding is meant labeling a bead with a ratio of labels or attributes (i.e. 1:10, 1:1, and 10:1). In theory any number of ratios can be used so long as the difference in signals between the ratios is detectable. This process produced smaller CVs and allowing more attribute segmentation within a given dynamic range. Thus, in a preferred embodiment, the use of ratiometric decoding reduces the coefficient of variability.

In addition, as will be appreciated by those in the art, ratiometric decoding can be accomplished in a different way. In this embodiment, rather than add a given number of beads with a first dye (or dye combination) intensity in the first decoding reaction and a second number with a second dye intensity in the sequential second decoding reaction, this ratiometric analysis may be done by using a ratio of labelled:unlabelled beads. That is, given a set saturating concentration of decoding beads, for example 100,000 beads/reaction, the first intensity decoding step may be done by adding 100,000 labelled beads and the second step can be done by adding 10,000 labelled beads and 90,000 unlabeled beads. Equilibrium dictates that the second step will give one tenth the signal intensity.

Because of the spread in values of a quantitatively measured attribute value, the number of distinct codes is practically limited to less than a dozen or so codes. However, by serially "painting" (i.e. temporarily attaching an attribute level to a bead) and "stripping" (removing the attribute level) a bead with different attribute values, the number of possible codes grows exponentially with the number of serial stages in the decoding process.

An example is illustrative. For instance, 9 different bead types and three distinguishable attribute distributions (Table 1). "Painting" (labeling) the beads with different attribute values in a combinatorially distinct pattern in the two different stages, generates a unique code for each bead type, i.e. nine distinct codes are generated. Thus, in a preferred embodiment beads are labeled with different attributes in a combinatorially distinct pattern in a plurality of stages. This generates unique codes for each bead type. Examples of different attributes are described above. Labeling of beads with different attributes is performed by methods known in the art.

	stage 1	stage 2	
Bead Type	attribute value	attribute value	Code
1	L	L	(L, L)
2	L	M	(L, M)
3	L	H	(L, H)
4	M	L	(M, L)
5	M	M	(M, M)
6	M	H	(M, H)
7	H	L	(H, L)
8	H	M	(H, M)
9	H	H	(H, H)

Number of unique codes =
Number of attributes^{Number of stages}

Table 1 Serial decode generates unique codes using a small number of attribute levels.

As outlined herein, fluorescent colors are a particularly convenient and preferable attribute to use in a decoding scheme. Fluorescent colors can be attached to any agent that recognizes an IBL to form a labeled DBL. A fluorescently labeled oligonucleotide is a particularly useful DBL since it can specifically and reversibly "paint" (label) any desired subset of beads with a particular color simply by the process of hybridization and dehybridization (i.e. to the DBL with a complementary sequence). Moreover, fluorescence is easily imaged and quantitated using standard optical hardware and software. In order to "paint" a given bead type with a particular color, the bead type must be labeled

with a unique hybridizable DNA sequence (IBL) and the decoding solution must contain the color-labeled complement of that sequence.

One consideration in implementing a decoding scheme is to minimize the number of images collected. In a color-based scheme, the number of images collected is the product of the number of colors and the number of stages. The number of images can be reduced by "painting" a bead with multiple colors for each given stage. By assigning multiple colors to a bead, the number of effective codes is increased. As an example, in a 24 bit three color scheme (e.g. red, green, blue) coloring process used by computers, a total of $256 \times 256 \times 256 = 16.7$ million different "hues" can be generated from just three colors (red, green, blue).

Thus, in a preferred embodiment DBLs are labeled with a combination of colored fluorophores. As such, this method finds use in increasing the number of available codes for labeling DBLs using only a handful of different dyes (colors). Increasing the number of codes available at each decoding step will greatly decrease the number of decoding steps required in a given decoding process.

In one embodiment a population of oligonucleotides encoding a single DBL is labeled with a defined ratio of colors such that each bead to which the DBL binds is identified based on a characteristic "hue" formulated from the combination of the colored fluorophores. In a preferred embodiment two distinct colors are used. In a preferred embodiment, three or more distinct dyes (colors) are available for use. In this instance the number of differentiable codes generated by labeling a population of oligonucleotides encoding a single DBL with any given color is three. However by allowing combinations of colors and color levels in the labeling, many more codes are generated.

For decoding by hybridization, a preferred number of distinguishable color shades is from 2 to 2000; a more preferred number of distinguishable color shades is from 2 to 200 and a most preferred number of distinguishable color shades is from 2 to 20. Utilizing three different color shades (intensities) and three colors, the number of different hues will be $3^4 = 81$. Combining a hue with sequential decoding allows a virtually limitless number of codes to be generated.

As previously described, the DBL can be any agent that binds to the IBL. In a preferred embodiment, a single DBL is labeled with a pre-determined ratio of colors. This ratio is varied for each DBL thus allowing for a unique "hue" for each DBL labeled as such. Following treatment of the beads with the DBL, the bead is analyzed to determine the "hue" associated with each bead, thereby identifying the bead with its associated bioactive agent.

For instance, with four primary colors and two intensity levels (color is present or absent), fifteen different hues/stage are possible. If four dyes and three different intensity levels are used (absent, half-present, fully present), then 73 different hues/stage are possible. In this case, acquisition of only 4 color images is sufficient to obtain information on 73 different coding hues.

In a preferred embodiment the DBL relies on a Fluorescence Resonance Energy Transfer (FRET) identifier. Fluorescence resonance energy transfer (FRET) is well known in the art and has been employed for various biotechnological applications. FRET requires that the energy of the donor excited state overlap with the ground state of the acceptor, implying that the donor emission spectrum overlaps with the acceptor absorbance spectrum. In a FRET experiment, the donor is excited with light; some of the energy of the excited state is released as light emission while some of the energy is transferred to the acceptor, which becomes excited emitting light at a characteristic wavelength. The distance over which energy transfer may occur can be varied by the choice of energy donor and acceptor, wherein the use of various different donor and acceptor molecules is contemplated. The formula for the efficiency of energy transfer is:

$$T = R_0^6 / (R_0^6 + R^6)$$

Where R is the distance between the donor and acceptor, and R_0 is a constant related to the characteristic overlap of the absorption and emission spectra of the donor-acceptor pair.

A FRET identifier comprises a linker region that is labeled at each terminus with a fluorophore. The first fluorophore serves as a donor fluorophore, whose signal is detectable; the other acts as an acceptor fluorophore. The emission of the acceptor fluorophore also is a detectable signal, the intensity of which varies proportionally with the distance between the two fluorophores. That is, by varying the distance between the two fluorophores different codes can be generated.

Identification of the decoding oligonucleotide, therefore, is accomplished by identifying the "efficiency" or ratio of intensities of the acceptor fluorophore and the donor fluorophore. This ratio should be invariant with regard to the number of decoding oligonucleotides bound to a particle (bead); rather, without being bound by any theory, the distance between the fluorophores determines the ratio. Thus, by varying the distance between the fluorophores (i.e. increasing or decreasing the length of the linker), unique "codes" can be assigned to the DBLs. Each code is characterized by the ratio of emission intensities between the two fluorophores.

An even greater advantage is realized when different bead attributes can be collected in a single image rather than from multiple images. For instance, if the colors of red, green and blue are used as attributes, three separate images need to be collected. However, if different shades or intensities of each color are used in the coding process, a much larger number of attributes can be realized in a single image. The difficulty with using mixtures of colors, is that the color ratios can vary somewhat between bead types since individual components are mixed together to determine the color ratio. If on the other hand, the color ratio is built into the decoding label, as with the above described FRET decoding oligonucleotides, the decoding process should be more reliable and more efficient.

In preferred embodiments of the present invention, the first component may be a lanthanide such as a chelate of europium(III) (Eu(III)) or terbium(III) (Tb(III)) which serves as a fluorescent donor and the second component may be an activated ester of cyanine 5 (Cy5) or tetramethylrhodamine (TMR) which serve as energy acceptors of Eu(III) and Tb(III). Other first and second component pairs that will find use for FRET include, for example, fluorescein and rhodamine; FITC and rhodamine; and fluorescein and trinitrophenyl, and other components as are well known in the art.

The linker is comprised of any spacer element including alkyl linkers, polymer linkers, etc. For ease of synthesis, preferably the linker comprises nucleic acids (including nucleic acid analogs).

The FRET decoding oligonucleotide comprises a spacer or linker region that is labeled with fluorophores (dyes). In general, this may be used in two general ways. In a preferred embodiment, as generally depicted in Figure 4, the linker is attached to the terminus of a decoding nucleotide that hybridizes with (is complementary to) the IBL. In this embodiment, the linker length can be varied to adjust signal intensity. Thus, in a preferred embodiment a FRET oligonucleotide is adjoined to an oligonucleotide DBL. In one embodiment the FRET oligonucleotide is synthesized independent of the DBL oligo. Subsequently the FRET oligonucleotide is ligated to the DBL oligo according to methods well known in the art. Alternatively, the DBL oligo and FRET oligo are synthesized as a single oligonucleotide comprising the FRET fluorophores at the desired positions.

Alternatively, in a second embodiment the linker is itself the oligonucleotide that hybridizes with (is complementary to) the IBL. That is, fluorophores are attached to the decoding oligonucleotide at different basepair lengths; some may be 10 nucleotides apart, some 20, etc. In a preferred embodiment the oligonucleotide DBL is itself labeled with fluors; as such the oligonucleotide serves as both DBL and FRET oligonucleotide.

The linker region of a FRET oligonucleotide comprises any set of nucleotides; in a preferred embodiment, for example when the linker does not hybridize to anything, the linker comprises a single repeated nucleotide. A single repeated nucleotide is less likely to hybridize with the IBL or bioactive agent on the bead because of the reduced likelihood of the presence of a complementary nucleotide sequence in either the IBL or bioagent on the bead. In a preferred embodiment, the linker comprises a polynucleotide linker such as poly T, poly A, poly G or poly C, or combinations.

The linker can be of any length so long as the emission of the first fluorophore is absorbed by the second fluorophore. In a preferred embodiment the linker is formed of from 1 to at least 100 nucleotides. More preferably the linker is formed from 1 to 70 nucleotides. Most preferably the linker is formed from 1 to 40 nucleotides, with from 5 to 20 being especially preferred.

In a preferred embodiment, for a given DBL sequence, several codes can be generated by varying the distance of separation of the two fluorophores and hence generating several energy transfer "efficiency" ratios.

In another preferred embodiment, for a given DBL sequence, a single code can be generated by varying the distance of separation of the two fluorophores and hence generating a single energy transfer ratio for a single DBL. However, numerous DBLs can be labeled with distinct FRET oligos thus allowing for the simultaneous determination or identification of multiple DBLs in a single imaging session as a result of the plurality of energy transfer ratios.

For example, given a 40-mer decoding oligo, the donor and acceptor fluorophores can be located as close as one or as far apart as 40 bases. In principle this should allow 39 codes to be generated for a single decoding sequence. Empirically, the number of codes generated depends on dispersion in the distributions for the ratiometric measurements.

Also, given a 40-mer decoding oligo, 39 unique codes can be generated which will identify 39 different oligos.

The present invention provides array compositions comprising a substrate with a surface comprising discrete sites. A population of microspheres is distributed on the sites, and the population comprises at least a first and a second subpopulation. Each subpopulation comprises a bioactive agent, and, in addition, at least one optical dye with a given pKa. The pKas of the different optical dyes are different.

In a preferred embodiment, when for example the array comprises cloned nucleic acids, there are several methods that can be used to decode the arrays. In a preferred embodiment, when some sequence information about the cloned nucleic acids is known, specific decoding probes can be made as is generally outlined herein.

In a preferred embodiment, "random" decoding probes can be made. By sequential hybridizations or the use of multiple labels, as is outlined above, a unique hybridization pattern can be generated for each sensor element. This allows all the beads representing a given clone to be identified as belonging to the same group. In general, this is done by using random or partially degenerate decoding probes, that bind in a sequence-dependent but not highly sequence-specific manner. The process can be repeated a number of times, each time using a different labeling entity, to generate a different pattern of signals based on quasi-specific interactions. In this way, a unique optical signature is eventually built up for each sensor element. By applying pattern recognition or clustering algorithms to the optical signatures, the beads can be grouped into sets that share the same signature (i.e. carry the same probes).

In order to identify the actual sequence of the clone itself, additional procedures are required; for example, direct sequencing can be done. By using an ordered array containing the clones, such as a spotted cDNA array, a "key" can be generated that links a hybridization pattern to a specific clone whose position in the set is known. In this way the clone can be recovered and further characterized.

Alternatively, clonal arrays can be decoded using binary decoding with vector tags. For example, partially randomized oligos are cloned into a nucleic acid vector (e.g. plasmid, phage, etc.). Each oligonucleotide sequence consists of a subset of a limited set of sequences. For example, if the limited set comprises 10 sequences, each oligonucleotide may have some subset (or all of the 10) sequences. Thus each of the 10 sequences can be present or absent in the oligonucleotide. Therefore, there are 2^{10} or 1,024 possible combinations. The sequences may overlap, and minor variants can also be represented (e.g. A, C, T and G substitutions) to increase the number of possible combinations. A nucleic acid library is cloned into a vector containing the random code sequences. Alternatively, other methods such as PCR can be used to add the tags. In this way it is possible to use a small number of oligo decoding probes to decode an array of clones.

There are a variety of methods used to evaluate decoding. In a preferred embodiment, discriminant analysis and cluster algorithms and computer apparatus are used to analyze the decoding data from the arrays of the invention. The potentially large number of codes utilized in the invention, coupled with the use of different intensities and "hues" of fluorophores in multi-step decoding processes, requires good classification of the data. The data, particularly intensity data, is acquired in a multi-step process during which beads are "painted" (for example by hybridizing dye-labeled complementary decoding oligonucleotides to the IBL probes on the beads, or the formation of binding ligand pairs for non-nucleic acid IBL-DBL pairs) with different colors or mixtures of colors ("hues") at each stage. The challenge is to accurately classify a bead as to which color with which it was painted at each step. The more closely related the "paints" (colors) are to one another (as determined by the optical imaging system), the more difficult the classification.

The proximity of the dyes as seen by the imaging system is determined by the spectral properties of the decoding dyes and the spectral channel separation of the imaging system. Better color separation is achieved by employing fluorescent dyes with narrow emission spectra, and by employing an optical system with narrow band pass excitation and emission filters which are designed to excite the dye "on peak" and measure its emission "on peak". The process of optically imaging the dyes on the beads is similar to the human vision process in which our brain sees color by measuring the ratio of excitation in the three different cone types within our eye. However, with an optical imaging system, the number of practical color channels is much greater than the three present in the human eye. CCD based imaging systems can "see" color from 350 nm up to 850 nm whereas the cones in the eye are tuned to the visible spectrum from 500 - 600 nm.

The problem of decoding bead arrays is essentially a discriminant analysis classification problem. Thus, in a preferred embodiment, an analysis of variance in hyperspectral alpha space is performed on a known set of bead colors or hues. The center of the bead clusters in alpha space are termed the centroids of the clusters, and the scatter of the points within a cluster determines the spread of the cluster. A robust classification scheme requires that the distance between the centroids of the different bead classes (hues) is much greater than the spread of any cluster class. Moreover, the location of the centroids should remain invariant from substrate to substrate and from experiment to experiment.

Thus, in a preferred embodiment, a hue "zone" is defined as a region in alpha space surrounding the hue centroid and extending out to the spread radius of the cluster. Given a reference set of hue centroids and spread radii, as determined empirically, the classification of a new set of data can be accomplished by asking whether a given bead point falls closest to or within the "zone" of a hue cluster. This is accomplished by calculating the Mahalanobis distance (in this case, it is simply a Euclidean distance metric) of the bead point from the centroids of the different hue classes. For the data shown in Fig. 6, the location of the centroids and their distances from one another are indicated in Table 2.

Table 2 dye/channel	Centroid position				Distance between centroids			
	Blue	Green	Yellow	Red	Bod-493	Bod-R6G	Bod-564	Bod-TXR
Bod-493	0.63	0.22	0.11	0.03	0.00			
Bod-R6G	0.03	0.51	0.37	0.09	0.72	0.00		
Bod-564	0.06	0.04	0.57	0.32	0.81	0.55	0.00	
Bod-TXR	0.09	0.05	0.04	0.82	0.99	0.93	0.73	0.00

For classifying the different beads into a particular hue class, a Euclidean distance cutoff of 0.3 was chosen. The closest two centroids, the Bod-R6G and Bod-564 (dist = 0.55), have a slight overlap in their decoding zones when using a Euclidean or Mahalanobis distance of 0.3. An improvement in classification can be achieved by decreasing this distance, and by weighting the different coordinate axes appropriately.

Accordingly, the present invention provides computer methods for analyzing and classifying the color of a bead. The classification of the color of the bead is done by viewing the bead in hyperspectral "alpha" space ($\Delta_1 = I_1/6I_t$, $\Delta_2 = I_2/6I_t$, $\Delta_3 = I_3/6I_t$, etc.) in which each coordinate axis represents the fraction of the bead intensity within a given imaging channel. For instance, if four imaging channels are used to image the beads, the color or hue of a bead can be represented by a point in 3-D alpha space (the fourth dimension is not necessary since $S_{a_i} = 1$). Given a set of different primary dyes by which to "paint" the beads, the number of hues that can be generated from these dyes is unlimited since the dyes can be combined in varying ratios and in varying combinatorial patterns. The

number of practical hues is experimentally determined by the separation of the different hue clusters in hyperspectral alpha space.

Fig. 6 shows a hyperspectral alpha plot of beads "painted" with four different hues imaged in four separate imaging channels. Note that the beads form four distinct clusters. The fact that these four clusters are well separated allows a robust decode classification scheme to be implemented.

In a preferred embodiment, a quality control analysis of the decoding process is done. This is achieved by performing a cluster analysis of alpha space for each decoding stage. The number of clusters determined will be fixed by the expected number of hues. The positions of the cluster centroids will be monitored and any deviations from the expected position will be noted.

Thus the invention provides an apparatus for decoding the arrays of the invention. In addition to the compositions outlined herein, the apparatus includes a central processing unit which communicates with a memory and a set of input/output devices (e.g., keyboard, mouse, monitor, printer, etc.) through a bus. The general interaction between a central processing unit, a memory, input/output devices, and a bus is known in the art. One aspect of the present invention is directed toward the hyperspectral "alpha" space classification system stored in the memory.

The classification system program includes a data acquisition module that receives data from the optical reader or confocal microscope (or other imaging system). In general, the classification program also includes an analysis module, that can analyze the variance in hyperspectral alpha space, calculate the centroids of the clusters, calculate the scatter of the cluster (the spread) and define the hue zone and distance cutoff. In general, the analysis module will further determine whether a data point falls within the hue zone by calculating the Mahalanobis distance.

Finally, the analysis module will analyze the different sequential decoding information to finally assign a bioactive agent to a bead location.

In this way, sequential decoding steps are run, with each step utilizing the discriminant analysis calculations to assign each bead in the array to a hue cluster at each step. The buildup of the sequential decoding information allows the correlation of the location of a bead and the chemistry contained on it.

In one embodiment, the invention includes modifying the number of detection channels available for analysis relative to the number of labels to be detected. That is, in one embodiment, the number of detection channels is decreased relative to the number of labels used. In an alternative embodiment the number of detection channels is increased relative to the number of labels used.

When fewer detection channels are used, fewer images are required during decoding. As such the invention provides a method of reducing the number of images obtained during decoding. Fluorescent dyes have broad band absorption and emission spectra. Accordingly, the emission of certain dyes is detected in multiple channels, generally by choosing channels to maximize the signal. However by choosing the dyes and the channels carefully, fewer channels than dyes may be used. As such, decoding with fewer channels than dyes, allows for detection of signals that are maximally detected in a particular channel, and overlapping in several channels. Thus, for example, beads with dye-1 are read in channel-1, beads with dye-2 are read in channel-2 and beads with dye-3 show signals in both channel-1 and channel-2. Example 7 is illustrative and the results of the experiment are found in Figure 11. It is noted, however, that three dyes are clearly detected and distinguished following detection in only two channels. Accordingly, using fewer detection channels than the number of dyes to be analyzed, allows for a reduction in the number of images and steps required for decoding.

In an alternative embodiment, the invention provides for using additional channels to increase the number of detectable codes. That is, additional readout channels are used to increase the coding/decoding capacity of arrays by analyzing attributes of the bead in addition to an optical signal such as fluorescence. For example, bead size, shape, density and surface characteristics can alter the light-scattering properties of the beads and can be detected optically. Additional channels that provide information independent of fluorescence are forward-scatter and side-scatter. Low-angle forward light scatter provides some information on the relative size of beads, whereas side light scatter provides some information on the relative granularity or texture of individual beads. By analogy these channels are often used in combination to distinguish the different major categories of white cells in unseparated mammalian blood, but are useful in a wide variety of other assays as well.

With respect to particle size, particles may simply be varied in diameter. Alternatively, particles could be assembled as doublets, triplets, or other higher-order multimers. Doublets and triplets often occur spontaneously in a population, and can be enriched for by fractionation.

With respect to surface properties, in one embodiment, the invention provides the use of metallization. This can dramatically alter the scattering properties of a surface. One way to accomplish this is to attach gold particles (e.g. colloidal gold) to an affinity molecule, such as streptavidin or DNA, and then to bind the molecule at defined loadings to specific sites on the bead. For example, ten different loadings could give ten new codes. Together with 100 fluorescence codes, this would allow the recognition of ~ 1000 bead types in an array. Alternatively, metallization could be used for the assay readout, and fluorescence for the encoding.

With respect to magnetism, superparamagnetic beads could be deflected according to the amount of magnetism, and directed to different readout channels -i.e. this mechanism could be used as a

"pre-sorter" prior to fluorescence readout. In one version, bead subpopulations could be localized to different regions of a reaction vessel (e.g. microtiter well) by the use of small magnets. Each sub-population could be released sequentially for analysis.

Although the previously described method involving additional detection channels finds use in planar bead arrays, it also finds particular use in liquid arrays. By "liquid arrays" is meant a collection of beads that are not spatially organized, but are typically encoded by fluorescence labeling and "decoded" by readout on a flow- cytometer or equivalent instrument. Assays are carried out in bulk in solution. Generally, liquid arrays are designed such that the assay result and decoding result are obtained essentially simultaneously. However, because only optical images are detected simultaneously, the number of codes that can be analyzed is limited. That is, there is a limit to the number of codes that can be used since the number of assay signals and encoding signals must be determined at different wavelengths at essentially the same time. However, by using additional detection channels as described above, the number of codes that can be analyzed is markedly increased.

This method finds particular use in analyzing the results of genomics based assays such as gene expression profiling, genotyping and SNP analysis. In such genomics based assays, it is frequently desirable to make hundreds to millions of measurements simultaneously. Thus, by increasing the number of codes that can be simultaneously detected, analysis of such assays is enhanced.

This invention provides methods of maximizing the information that is obtained from a single bead. That is, by providing ways of multiplexing assays on a single bead, multiple parameters can be simultaneously detected. That is, after a cycle of performing a particular assay, decoding and reading the assay results, the bead population is recovered. The detection molecules are removed from the bead and the cycle is repeated.

In a preferred embodiment, parity analysis is used during decoding to increase the robustness and accuracy of the system. By "parity analysis" is meant a decoding step wherein the signal of a particular element is analyzed across a plurality of decoding stages. That is, following at least one decoding step, but most preferably following multiple decoding steps, the signal of an array element across the decoding stages is analyzed. That is, the signal from a particular bead is evaluated across multiple stages. Although the analysis includes any parameter that can be obtained from the signals such as evaluating the total signal obtained across the stages, in a preferred embodiment, the parity of the signals across the stages is analyzed.

By "parity" is meant the digital or modular readout of signals, i.e. odd or even, when binary signals are used. As depicted in Figure 7, the digit sum of the signals across a plurality of stages is translated into a parity determination. The parity determination is then useful in evaluating the decoding process. An

example is illustrative. Codes are designed to have an odd number of a particular signal, for example a red signal, when viewed across all stages or decoding steps, or a pre-determined plurality of stages or decoding steps. The detection of an even number of red stages, provides an indication that an error has occurred at some point in decoding. When this result is obtained, the faulty code can either be discarded, or the analysis repeated.

In a preferred embodiment the invention includes introducing a "redundant stage" into the decoding system. By "redundant stage" is meant a stage that serves as a parity check. That is, following the decoding stages, an additional stage is included to analyze the parity. This analysis provides an indication of the competence or validity of the decoding. That is, when codes are designed with a pre-determined parity, the redundant stage is used to detect the parity of the signals obtained from the decoding step. That is, the redundant stage detects errors in parity because if there has been an error in decoding, the parity detected following the redundant stage will be different from the parity designed into the codes.

An example is illustrative. In the following illustration, array elements are decoded in two stages with two colors. This results in the detection of 4 (2^2) codes. The codes are designed to result in odd parity following the analysis with the redundant stage. That is, the code sequences can be made such that each code will contain an odd number of "red" colors (when viewed across all stages, or when viewed across the pre-determined number of stages). In the illustration below, only the stage where the "red" signal is detected is indicated.

Code	Stage 1	Stage 2	Redundant Stage	Parity (number of "red" stages)
1	0	0	red	1 (odd)
2	0	red	0	1 (odd)
3	red	0	0	1 (odd)
4	red	red	red	3 (odd)

However, if there is an error during decoding, this will be reflected as detecting a parity that is different from that designed into the codes. In the following illustration the same codes as above are used however, an error is detected. In this case if one detects an even number of "red" stages it can be inferred that there has been a decoding error.

Code	Stage 1	Stage 2	Redundant Stage	Parity (number of "red" stages)
1	0	0	red	1 (odd)
2	0	red	0	1 (odd)

3	red	0	0	1 (odd)
4	red	0 (error)	red	2 (even)

Although in some embodiments, the parity is binary, i.e. odd or even, in other embodiments the parity is modular or based on numbers other than 2. That is, the parity can be based on detection of 3 or more signals. In a preferred embodiment, the signals are multiple colors. In a particularly preferred embodiment the colors are red, green and blue.

An example is illustrative. In the following illustration, codes are decoded in two steps with three colors. This results in the detection of 3^2 (9) signals. In the illustration, red is indicated by 1, green is indicated by 2 and blue is indicated by 3. The parity is based on a modulus 3. By including a third stage as a redundant stage the parity can be checked as described above.

Code	Stage 1	Stage 2	Redundant stage	Parity
1	1	1	1	3
2	1	2	3	6
3	1	3	2	6
4	2	1	3	6
5	2	2	2	6
6	2	3	1	6
7	3	1	2	6
8	3	2	1	6
9	3	3	3	9

However, when there is a decoding error, the parity will be incorrect; in the illustration the parity readout is not divisible by 3.

Code	Stage 1	Stage 2	Redundant stage	Parity
1	1	1	1	3
2	1	2	3	6
3	1	3	2	6
4	2	1	3	6
5	2	2	2	6

6	2	3	1	6
7	3	1	2	6
8	3	2	1	6
9	3	2 (error)	3	8

Thus, a redundant stage allows for the detection of errors on the array or in the signal obtained from an array.

Although described above for detecting errors on arrays, the method of using redundant stages is not limited to arrays. That is, the method finds use in detecting errors in a variety of systems that send, receive and/or interpret or analyze data such as communications systems including modems etc. In particular, the method of using the modular redundant stages allows for processing of signals of vastly increased complexity as compared to sending bits of data. That is, when using modular redundant stages, characters as opposed to bits, can be sent across modem lines in the communications field. The modular redundant stage is used to detect errors in the communication or processing of the signals.

In an additional embodiment a redundant channel is used during the decoding. By "redundant channel" is meant either the use of multiple filters to analyze a particular signal or repeated use of the same filter to analyze the signal, i.e. multiple signals from the same channel. Redundant channels find use as a source of a "back-up" signal of sorts for a particular array element. That is, when multiple images are obtained as a result of redundant channels, the likelihood of obtaining the correct signal increases. That is, when it is realized that one of the results for an element is compromised for some reason, the image obtained from the redundant channel is used instead. When the initial image is not compromised, the result from the redundant channel need not be used, although it may be used, for example as a way to confirm the initial signal.

An example is illustrative. During decoding, when it is apparent that one stage fails, for example as a result of mis-registration of a channel, the use of a redundant channel serves to rescue the decoding at that particular stage. Mis-registration can also be addressed as outlined in USSN 09/636,387, filed August 9, 2000, which is expressly incorporated herein by reference. As noted above, the redundant channel can either be an image separate filter to analyze the signal, or alternatively, it can be multiple images taken from the same filter.

In one embodiments, the number of redundant channels increases as the number of decoding stages increases. Because the investment in decoding increases as each decoding stage or step proceeds, it is important in this embodiment to ensure the competence of decoding at later decoding stages.

In a preferred embodiment, redundant systems are used to avoid the loss of information. It is appreciated that when "S" decoding steps (sometimes referred to herein as "stages") and "C" colors are used, the maximum number of distinct signals is C^S (C raised to the "S"th power). In some instances, the corruption of a stage results in the loss of $C^S - (C-1) \times C^{(S-1)}$ codes. Depending on the values of C and S, the loss can be substantial. Accordingly, the present invention provides for methods of improving the robustness of the decoding system. That is, the invention provides for a method of confirming the results of a decoding analysis or detecting errors in a decoding analysis during or after the decoding procedure. The method also decreases the number of false positive or false negative signals; similarly, the number of true positives and/or true negatives increases.

In a preferred embodiment, several levels of redundancy are built into the arrays of the invention. Building redundancy into an array gives several significant advantages, including the ability to make quantitative estimates of confidence about the data and significant increases in sensitivity. Thus, preferred embodiments utilize array redundancy. As will be appreciated by those in the art, there are at least two types of redundancy that can be built into an array: the use of multiple identical sensor elements (termed herein "sensor redundancy"), and the use of multiple sensor elements directed to the same target analyte, but comprising different chemical functionalities (termed herein "target redundancy"). For example, for the detection of nucleic acids, sensor redundancy utilizes a plurality of sensor elements such as beads comprising identical binding ligands such as probes. Target redundancy utilizes sensor elements with different probes to the same target: one probe may span the first 25 bases of the target, a second probe may span the second 25 bases of the target, etc. By building in either or both of these types of redundancy into an array, significant benefits are obtained. For example, a variety of statistical mathematical analyses may be done.

In addition, while this is generally described herein for bead arrays, as will be appreciated by those in the art, this techniques can be used for any type of arrays designed to detect target analytes.

In a preferred embodiment, sensor redundancy is used. In this embodiment, a plurality of sensor elements, e.g. beads, comprising identical bioactive agents are used. That is, each subpopulation comprises a plurality of beads comprising identical bioactive agents (e.g. binding ligands). By using a number of identical sensor elements for a given array, the optical signal from each sensor element can be combined and any number of statistical analyses run, as outlined below. This can be done for a variety of reasons. For example, in time varying measurements, redundancy can significantly reduce the noise in the system. For non-time based measurements, redundancy can significantly increase the confidence of the data.

In a preferred embodiment, a plurality of identical sensor elements are used. As will be appreciated by those in the art, the number of identical sensor elements will vary with the application and use of the sensor array. In general, anywhere from 2 to thousands may be used, with from 2 to 100 being preferred, 2 to 50 being particularly preferred and from 5 to 20 being especially preferred. In general,

preliminary results indicate that roughly 10 beads gives a sufficient advantage, although for some applications, more identical sensor elements can be used.

Once obtained, the optical response signals from a plurality of sensor beads within each bead subpopulation can be manipulated and analyzed in a wide variety of ways, including baseline adjustment, averaging, standard deviation analysis, distribution and cluster analysis, confidence interval analysis, mean testing, etc.

In a preferred embodiment, the first manipulation of the optical response signals is an optional baseline adjustment. In a typical procedure, the standardized optical responses are adjusted to start at a value of 0.0 by subtracting the integer 1.0 from all data points. Doing this allows the baseline-loop data to remain at zero even when summed together and the random response signal noise is canceled out. When the sample is a fluid, the fluid pulse-loop temporal region, however, frequently exhibits a characteristic change in response, either positive, negative or neutral, prior to the sample pulse and often requires a baseline adjustment to overcome noise associated with drift in the first few data points due to charge buildup in the CCD camera. If no drift is present, typically the baseline from the first data point for each bead sensor is subtracted from all the response data for the same bead. If drift is observed, the average baseline from the first ten data points for each bead sensor is subtracted from all the response data for the same bead. By applying this baseline adjustment, when multiple bead responses are added together they can be amplified while the baseline remains at zero. Since all beads respond at the same time to the sample (e.g. the sample pulse), they all see the pulse at the exact same time and there is no registering or adjusting needed for overlaying their responses. In addition, other types of baseline adjustment may be done, depending on the requirements and output of the system used.

Once the baseline has been adjusted, a number of possible statistical analyses may be run to generate known statistical parameters. Analyses based on redundancy are known and generally described in texts such as Freund and Walpole, *Mathematical Statistics*, Prentice Hall, Inc. New Jersey, 1980, hereby incorporated by reference in its entirety.

In a preferred embodiment, signal summing is done by simply adding the intensity values of all responses at each time point, generating a new temporal response comprised of the sum of all bead responses. These values can be baseline-adjusted or raw. As for all the analyses described herein, signal summing can be performed in real time or during post-data acquisition data reduction and analysis. In one embodiment, signal summing is performed with a commercial spreadsheet program (Excel, Microsoft, Redmond, WA) after optical response data is collected.

In a preferred embodiment, cumulative response data is generated by simply adding all data points in successive time intervals. This final column, comprised of the sum of all data points at a particular

time interval, may then be compared or plotted with the individual bead responses to determine the extent of signal enhancement or improved signal-to-noise ratios.

In a preferred embodiment, the mean of the subpopulation (i.e. the plurality of identical beads) is determined, using the well known Equation 1:

Equation 1

$$\mu = \sum \frac{x_i}{n}$$

In some embodiments, the subpopulation may be redefined to exclude some beads if necessary (for example for obvious outliers, as discussed below).

In a preferred embodiment, the standard deviation of the subpopulation can be determined, generally using Equation 2 (for the entire subpopulation) and Equation 3 (for less than the entire subpopulation):

Equation 2

$$\sigma = \sqrt{\frac{\sum (x_i - \mu)^2}{n}}$$

Equation 3

$$s = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n - 1}}$$

As for the mean, the subpopulation may be redefined to exclude some beads if necessary (for example for obvious outliers, as discussed below).

In a preferred embodiment, statistical analyses are done to evaluate whether a particular data point has statistical validity within a subpopulation by using techniques including, but not limited to, t distribution and cluster analysis. This may be done to statistically discard outliers that may otherwise skew the result and increase the signal-to-noise ratio of any particular experiment. This may be done using Equation 4:

Equation 4

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

In a preferred embodiment, the quality of the data is evaluated using confidence intervals, as is known in the art. Confidence intervals can be used to facilitate more comprehensive data processing to measure the statistical validity of a result.

In a preferred embodiment, statistical parameters of a subpopulation of beads are used to do hypothesis testing. One application is tests concerning means, also called mean testing. In this application, statistical evaluation is done to determine whether two subpopulations are different. For example, one sample could be compared with another sample for each subpopulation within an array to determine if the variation is statistically significant.

In addition, mean testing can also be used to differentiate two different assays that share the same code. If the two assays give results that are statistically distinct from each other, then the subpopulations that share a common code can be distinguished from each other on the basis of the assay and the mean test, shown below in Equation 5:

Equation 5

$$Z = \frac{\overline{x_1} - \overline{x_2}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

Furthermore, analyzing the distribution of individual members of a subpopulation of sensor elements may be done. For example, a subpopulation distribution can be evaluated to determine whether the distribution is binomial, Poisson, hypergeometric, etc.

In addition to the sensor redundancy, a preferred embodiment utilizes a plurality of sensor elements that are directed to a single target analyte but yet are not identical. For example, a single target nucleic acid analyte may have two or more sensor elements each comprising a different probe. This adds a level of confidence as non-specific binding interactions can be statistically minimized. When nucleic acid target analytes are to be evaluated, the redundant nucleic acid probes may be overlapping, adjacent, or spatially separated. However, it is preferred that two probes do not compete for a single binding site, so adjacent or separated probes are preferred. Similarly, when proteinaceous target analytes are to be evaluated, preferred embodiments utilize bioactive agent binding agents that bind to different parts of the target. For example, when antibodies (or antibody fragments) are used as bioactive agents for the binding of target proteins, preferred embodiments utilize antibodies to different epitopes.

In this embodiment, a plurality of different sensor elements may be used, with from about 2 to about 20 being preferred, and from about 2 to about 10 being especially preferred, and from 2 to about 5

being particularly preferred, including 2, 3, 4 or 5. However, as above, more may also be used, depending on the application.

As above, any number of statistical analyses may be run on the data from target redundant sensors.

One benefit of the sensor element summing (referred to herein as "bead summing" when beads are used), is the increase in sensitivity that can occur.

Once made, the compositions of the invention find use in a number of applications. In a preferred embodiment, the compositions are used to probe a sample solution for the presence or absence of a target analyte, including the quantification of the amount of target analyte present. By "target analyte" or "analyte" or grammatical equivalents herein is meant any atom, molecule, ion, molecular ion, compound or particle to be either detected or evaluated for binding partners. As will be appreciated by those in the art, a large number of analytes may be used in the present invention; basically, any target analyte can be used which binds a bioactive agent or for which a binding partner (i.e. drug candidate) is sought.

Suitable analytes include organic and inorganic molecules, including biomolecules. When detection of a target analyte is done, suitable target analytes include, but are not limited to, an environmental pollutant (including pesticides, insecticides, toxins, etc.); a chemical (including solvents, polymers, organic materials, etc.); therapeutic molecules (including therapeutic and abused drugs, antibiotics, etc.); biomolecules (including hormones, cytokines, proteins, nucleic acids, lipids, carbohydrates, cellular membrane antigens and receptors (neural, hormonal, nutrient, and cell surface receptors) or their ligands, etc); whole cells (including procaryotic (such as pathogenic bacteria) and eukaryotic cells, including mammalian tumor cells); viruses (including retroviruses, herpesviruses, adenoviruses, lentiviruses, etc.); and spores; etc. Particularly preferred analytes are nucleic acids and proteins.

In a preferred embodiment, the target analyte is a protein. As will be appreciated by those in the art, there are a large number of possible proteinaceous target analytes that may be detected or evaluated for binding partners using the present invention. Suitable protein target analytes include, but are not limited to, (1) immunoglobulins; (2) enzymes (and other proteins); (3) hormones and cytokines (many of which serve as ligands for cellular receptors); and (4) other proteins.

In a preferred embodiment, the target analyte is a nucleic acid. These assays find use in a wide variety of applications. Such applications are described in more detail in USSNs 60/130,089, filed April 20, 1999; 60/160,927, filed October 22, 1999, 09/513,362, filed February 25, 2000, 60/135,051 filed May 20, 1999, 60/161,148, filed October 22, 1999, 09/517, 945, filed March 3, 2000, 09/425,633, filed October 22, 1999, 09/535,854, filed March 27, 2000, 09/535,993, filed April 20, 2000, 09/556,463, filed April 21, 2000, 60/224,119, filed October 26, 2000, all of which are expressly incorporated herein by reference.

In a preferred embodiment, the probes are used in genetic diagnosis. For example, probes can be made using the techniques disclosed herein to detect target sequences such as the gene for nonpolyposis colon cancer, the BRCA1 breast cancer gene, P53, which is a gene associated with a variety of cancers, the Apo E4 gene that indicates a greater risk of Alzheimer's disease, allowing for easy presymptomatic screening of patients, mutations in the cystic fibrosis gene, cytochrome p450s or any of the others well known in the art.

In an additional embodiment, viral and bacterial detection is done using the complexes of the invention. In this embodiment, probes are designed to detect target sequences from a variety of bacteria and viruses. For example, current blood-screening techniques rely on the detection of anti-HIV antibodies. The methods disclosed herein allow for direct screening of clinical samples to detect HIV nucleic acid sequences, particularly highly conserved HIV sequences. In addition, this allows direct monitoring of circulating virus within a patient as an improved method of assessing the efficacy of anti-viral therapies. Similarly, viruses associated with leukemia, HTLV-I and HTLV-II, may be detected in this way. Bacterial infections such as tuberculosis, chlamydia and other sexually transmitted diseases, may also be detected.

In a preferred embodiment, the nucleic acids of the invention find use as probes for toxic bacteria in the screening of water and food samples. For example, samples may be treated to lyse the bacteria to release its nucleic acid, and then probes designed to recognize bacterial strains, including, but not limited to, such pathogenic strains as, *Salmonella*, *Campylobacter*, *Vibrio cholerae*, *Leishmania*, enterotoxigenic strains of *E. coli*, and Legionnaire's disease bacteria. Similarly, bioremediation strategies may be evaluated using the compositions of the invention.

In a further embodiment, the probes are used for forensic "DNA fingerprinting" to match crime-scene DNA against samples taken from victims and suspects.

In an additional embodiment, the probes in an array are used for sequencing by hybridization.

The present invention also finds use as a methodology for the detection of mutations or mismatches in target nucleic acid sequences, as is described in more detail in USSNs 09/425,633, filed October 22, 1999, 09/535,854, filed March 27, 2000, both of which are expressly incorporated herein by reference. For example, recent focus has been on the analysis of the relationship between genetic variation and phenotype by making use of polymorphic DNA markers. Previous work utilized short tandem repeats (STRs) as polymorphic positional markers; however, recent focus is on the use of single nucleotide polymorphisms (SNPs). Common SNPs occur at an average frequency of more than 1 per kilobase in human genomic DNA. Some SNPs, particularly those in and around coding sequences, are likely to be the direct cause of therapeutically relevant phenotypic variants. There are a number of well known polymorphisms that cause clinically important phenotypes; for example, the apoE2/3/4 variants are associated with different relative risk of Alzheimer's and other diseases (see Cordor et al., Science

261 (1993). Multiplex PCR amplification of SNP loci with subsequent hybridization to oligonucleotide arrays has been shown to be an accurate and reliable method of simultaneously genotyping at least hundreds of SNPs; see Wang et al., Science, 280:1077 (1998); see also Schafer et al., Nature Biotechnology 16:33-39 (1998). The compositions of the present invention may easily be substituted for the arrays of the prior art.

In a preferred embodiment, the compositions of the invention are used to screen bioactive agents to find an agent that will bind, and preferably modify the function of, a target molecule. As above, a wide variety of different assay formats may be run, as will be appreciated by those in the art. Generally, the target analyte for which a binding partner is desired is labeled; binding of the target analyte by the bioactive agent results in the recruitment of the label to the bead, with subsequent detection.

In a preferred embodiment, the binding of the bioactive agent and the target analyte is specific; that is, the bioactive agent specifically binds to the target analyte. By "specifically bind" herein is meant that the agent binds the analyte, with specificity sufficient to differentiate between the analyte and other components or contaminants of the test sample. However, as will be appreciated by those in the art, it will be possible to detect analytes using binding which is not highly specific; for example, the systems may use different binding ligands, for example an array of different ligands, and detection of any particular analyte is via its "signature" of binding to a panel of binding ligands, similar to the manner in which "electronic noses" work. This finds particular utility in the detection of chemical analytes. The binding should be sufficient to remain bound under the conditions of the assay, including wash steps to remove non-specific binding, although in some embodiments, wash steps are not desired; i.e. for detecting low affinity binding partners. In some embodiments, for example in the detection of certain biomolecules, the dissociation constants of the analyte to the binding ligand will be less than about 10^{-4} - 10^{-6} M⁻¹, with less than about 10^{-5} to 10^{-9} M⁻¹ being preferred and less than about 10^{-7} - 10^{-9} M⁻¹ being particularly preferred.

Generally, a sample containing a target analyte (whether for detection of the target analyte or screening for binding partners of the target analyte) is added to the array, under conditions suitable for binding of the target analyte to at least one of the bioactive agents, i.e. generally physiological conditions. The presence or absence of the target analyte is then detected. As will be appreciated by those in the art, this may be done in a variety of ways, generally through the use of a change in an optical signal. This change can occur via many different mechanisms. A few examples include the binding of a dye-tagged analyte to the bead, the production of a dye species on or near the beads, the destruction of an existing dye species, a change in the optical signature upon analyte interaction with dye on bead, or any other optically interrogatable event.

In a preferred embodiment, the change in optical signal occurs as a result of the binding of a target analyte that is labeled, either directly or indirectly, with a detectable label, preferably an optical label such as a fluorochrome. Thus, for example, when a proteinaceous target analyte is used, it may be

either directly labeled with a fluor, or indirectly, for example through the use of a labeled antibody. Similarly, nucleic acids are easily labeled with fluorochromes, for example during PCR amplification as is known in the art. Alternatively, upon binding of the target sequences, a hybridization indicator may be used as the label. Hybridization indicators preferentially associate with double stranded nucleic acid, usually reversibly. Hybridization indicators include intercalators and minor and/or major groove binding moieties. In a preferred embodiment, intercalators may be used; since intercalation generally only occurs in the presence of double stranded nucleic acid, only in the presence of target hybridization will the label light up. Thus, upon binding of the target analyte to a bioactive agent, there is a new optical signal generated at that site, which then may be detected.

Alternatively, in some cases, as discussed above, the target analyte such as an enzyme generates a species that is either directly or indirectly optically detectable.

Furthermore, in some embodiments, a change in the optical signature may be the basis of the optical signal. For example, the interaction of some chemical target analytes with some fluorescent dyes on the beads may alter the optical signature, thus generating a different optical signal.

As will be appreciated by those in the art, in some embodiments, the presence or absence of the target analyte may be done using changes in other optical or non-optical signals, including, but not limited to, surface enhanced Raman spectroscopy, surface plasmon resonance, radioactivity, etc.

The assays may be run under a variety of experimental conditions, as will be appreciated by those in the art. A variety of other reagents may be included in the screening assays. These include reagents like salts, neutral proteins, e.g. albumin, detergents, etc which may be used to facilitate optimal protein-protein binding and/or reduce non-specific or background interactions. Also reagents that otherwise improve the efficiency of the assay, such as protease inhibitors, nuclease inhibitors, anti-microbial agents, etc., may be used. The mixture of components may be added in any order that provides for the requisite binding. Various blocking and washing steps may be utilized as is known in the art.

In a preferred embodiment, two-color competitive hybridization assays are run. These assays can be based on traditional sandwich assays. The beads contain a capture sequence located on one side (upstream or downstream) of the SNP, to capture the target sequence. Two SNP allele-specific probes, each labeled with a different fluorophor, are hybridized to the target sequence. The genotype can be obtained from a ratio of the two signals, with the correct sequence generally exhibiting better binding. This has an advantage in that the target sequence itself need not be labeled. In addition, since the probes are competing, this means that the conditions for binding need not be optimized. Under conditions where a mismatched probe would be stably bound, a matched probe can still displace it. Therefore the competitive assay can provide better discrimination under those conditions. Because many assays are carried out in parallel, conditions cannot be optimized for every probe

simultaneously. Therefore, a competitive assay system can be used to help compensate for non-optimal conditions for mismatch discrimination.

In a preferred embodiment, dideoxynucleotide chain-termination sequencing is done using the compositions of the invention. In this embodiment, a DNA polymerase is used to extend a primer using fluorescently labeled ddNTPs. The 3' end of the primer is located adjacent to the SNP site. In this way, the single base extension is complementary to the sequence at the SNP site. By using four different fluorophors, one for each base, the sequence of the SNP can be deduced by comparing the four base-specific signals. This may be done in several ways. In a first embodiment, the capture probe can be extended; in this approach, the probe must either be synthesized 5'-3' on the bead, or attached at the 5' end, to provide a free 3' end for polymerase extension. Alternatively, a sandwich type assay can be used; in this embodiment, the target is captured on the bead by a probe, then a primer is annealed and extended. Again, in the latter case, the target sequence need not be labeled. In addition, since sandwich assays require two specific interactions, this provides increased specificity which is particularly helpful for the analysis of complex samples.

In addition, when the target analyte and the DBL both bind to the agent, it is also possible to do detection of non-labelled target analytes via competition of decoding.

In a preferred embodiment, the methods of the invention are useful in array quality control. Prior to this invention, no methods have been described that provide a positive test of the performance of every probe on every array. Decoding of the array not only provides this test, it also does so by making use of the data generated during the decoding process itself. Therefore, no additional experimental work is required. The invention requires only a set of data analysis algorithms that can be encoded in software.

The quality control procedure can identify a wide variety of systematic and random problems in an array. For example, random specks of dust or other contaminants might cause some sensors to give an incorrect signal-this can be detected during decoding. The omission of one or more agents from multiple arrays can also be detected; that is, since the arrays formulated herein are randomly self-assembled, arrays comprising different statistical distributions of each sensor elements may be made; the invention allows actual QC on the arrays, for "sorting" arrays with good distributions from those with less than ideal distributions. These procedures also allow the determination of sites that do not comprise beads. This quality control procedure can also be used in non-random arrays.

An advantage of this quality control procedure is that it can be implemented immediately prior to the assay itself (or after, in some instances), and is a true functional test of each individual sensor. Therefore any problems that might occur between array assembly and actual use can be detected. In applications where a very high level of confidence is required, and/or there is a significant chance of

sensor failure during the experimental procedure, decoding and quality control can be conducted both before and after the actual sample analysis.

In an alternative method of quality control, a single fiber bundle of an array is dedicated to quality control. That is a dedicated bundle or subset of fibers within an array is independently analyzed for quality control. In one embodiment, the control fiber(s) are targeted or labeled with a unique code or label. At each state of hybridization, in addition to the experimental probes, the quality control probes are hybridized with the array. Upon taking the image from the first hybridization of the first stage, the intensity of the control fiber(s) is analyzed. An aberrant reading or signal from the control fiber indicates a failed or at least non-optimal hybridization step.

In a preferred embodiment, the arrays can be used to do reagent quality control. In many instances, biological macromolecules are used as reagents and must be quality controlled. For example, large sets of oligonucleotide probes may be provided as reagents. It is typically difficult to perform quality control on large numbers of different biological macromolecules. For example, when large populations of different random oligonucleotides are synthesized, it may be desirable to check that each population is represented. The approach described here can be used to do this by treating the reagents (formulated as the DBLs) as variable instead of the arrays.

In a preferred embodiment, the methods outlined herein are used in array calibration. For many applications, such as mRNA quantitation, it is desirable to have a signal that is a linear response to the concentration of the target analyte, or, alternatively, if non-linear, to determine a relationship between concentration and signal, so that the concentration of the target analyte can be estimated. For example, different probes can show differences in hybridization efficiencies, etc., and it may be desirable to choose appropriate probes under experimental conditions. Thus, for example, when expression profiling is done, and quantitation is desirable, calibration curves can be done to see how individual probes react, and then the probes that give the best response (i.e. linearity at the concentrations and conditions of interest) can be chosen for further assays. Accordingly, the present invention provides methods of creating calibration curves in parallel for multiple beads in an array. The calibration curves can be created under conditions that simulate the complexity of the sample to be analyzed. Each curve can be constructed independently of the others (e.g. for a different range of concentrations), but at the same time as all the other curves for the array.

Thus, in these embodiments, different types of experiments can be done. For example, the sequential decoding scheme can be implemented with different concentrations being used as the code "labels", rather than different fluorophores. In this way, signal as a response to concentration can be measured for each bead. This calibration can be carried out just prior to array use, so that every probe on every array is individually calibrated as needed. Alternatively, different concentrations of the decoding probe can have different labels.

It should be noted that the assay calibration methods find use in non-random arrays, as well; that is, other types of support-bound nucleic acid arrays can be calibrated using these methods as well. Thus, for example, the sequential addition of different pools of probes, wherein the concentration of the probes is varied, to biochips can allow calibration of any assay system. This type of analysis can also be done on non-random arrays for quality control, to verify the integrity and sequence of the support bound probes, and in assay development to identify good probes.

In a preferred embodiment, the methods of the invention can be used in assay development as well. Thus, for example, the methods allow the identification of good and bad probes; as is understood by those in the art, some probes do not function well because they do not hybridize well, or because they cross-hybridize with more than one sequence. These problems are easily detected during decoding. The ability to rapidly assess probe performance has the potential to greatly reduce the time and expense of assay development. Thus, probes that respond linearly with concentration, show low non-specific binding, or give signals in a particular range, can be chosen for addition to a new array for assays.

Similarly, in a preferred embodiment, the methods of the invention are useful in quantitation in assay development. Major challenges of many assays is the ability to detect differences in analyte concentrations between samples, the ability to quantitate these differences, and to measure absolute concentrations of analytes, all in the presence of a complex mixture of related analytes. An example of this problem is the quantitation of a specific mRNA in the presence of total cellular mRNA. One approach that has been developed as a basis of mRNA quantitation makes use of a multiple match and mismatch probe pairs (Lockhart et al., 1996), hereby incorporated by reference in its entirety. While this approach is simple, it requires relatively large numbers of probes. In this approach, a quantitative response to concentration is obtained by averaging the signals from a set of different probes to the gene or sequence of interest. This is necessary because only some probes respond quantitatively, and it is not possible to predict these probes with certainty. In the absence of prior knowledge, only the average response of an appropriately chosen collection of probes is quantitative. However, in the present invention, that can be applied generally to nucleic acid based assays as well as other assays. In essence, the approach is to identify the probes that respond quantitatively in a particular assay, rather than average them with other probes. This is done using the array calibration scheme outlined above, in which concentration-based codes are used. Advantages of this approach include: fewer probes are needed; the accuracy of the measurement is less dependent on the number of probes used; and that the response of the sensors is known with a high level of certainty, since each and every sequence can be tested in an efficient manner. It is important to note that probes that perform well are selected empirically, which avoids the difficulties and uncertainties of predicting probe performance, particularly in complex sequence mixtures. In contrast, in experiments described to date with ordered arrays, relatively small numbers of sequences are checked by performing quantitative spiking experiments, in which a known mRNA is added to a mixture.

In a preferred embodiment, cDNA arrays are made for RNA expression profiling. In this embodiment, individual cDNA clones are amplified (for example, using PCR) from cDNA libraries propagated in a host-vector system. Each amplified DNA is attached to a population of beads. Different populations are mixed together, to create a collection of beads representing the cDNA library. The beads are arrayed, decoded as outlined above, and used in an assay (although as outlined herein, decoding may occur after assay as well). The assay is done using RNA samples (whole cell or mRNA) that are extracted, labeled if necessary, and hybridized to the array. Comparative analysis allows the detection of differences in the expression levels of individual RNAs. Comparison to an appropriate set of calibration standards allows quantification of absolute amounts of RNA.

The cDNA array can also be used for mapping, e.g. to map deletions/insertions or copy number changes in the genome, for example from tumors or other tissue samples. This can be done by hybridizing genomic DNA. Instead of cDNAs (or ESTs, etc.), other STS (sequence tagged sites), including random genomic fragments, can also be arrayed for this purpose.

In one embodiment, the invention provides a method of single nucleotide polymorphism (SNP) genotyping. As described herein and in U.S.S.N.s 09/553,993 filed April 20, 2000, 09/556,463 filed April 21, 2000 and 60/244,119, filed October 26, 2000, all of which are expressly incorporated herein by reference, a target nucleic acid is contacted with a probe that is specific for a particular allele of a single nucleotide polymorphism. The target nucleic acid can be immobilized prior to contacting with the probe, or alternatively, the probe: target hybridization can occur in solution following by immobilization of the probe: target nucleic acid complex. Alternatively, a capture probe on a bead may be specific for a particular SNP. What is important in this invention is the nature of the probe for detecting a SNP.

There are four possibilities for the bases at a polymorphic site: -A, T, G, or C. However, it is known that most of the human SNP's are dimorphic, i.e., they only include the combinations of two of the four bases. As such, the prior art generally relied on distinguishing between two alleles that are present within a SNP. and characterized them as either "AA", "AB, or "BB".

However, according to the present invention, the probes are designed such that each of the four possible nucleotides that could be present at a particular position are included in the SNP probes. In addition, the probes are labeled with a specific color that corresponds with the nucleotide that hybridizes to the nucleotide to be interrogated. For instance, all the alleles that correspond to an A-polymorphic site are labeled with red. Similarly, all the alleles that correspond to a G-polymorphic site are labeled green, etc. As such the true identity of the polymorphism is determined. That is, if the polymorphism is of the type GT, one would not classify is as "AB" genotype. Instead, the GT code will come up directly from the genotyping routine.

An advantage of this method over the prior art is the capability of independent error checking. That is, knowing that polymorphisms are typically dimorphic, by using probes that correspond to all possibilities of polymorphisms at a sites, one obtains a check against false results, for example of a signal is detected that does not correspond with the known polymorphisms. That is, in the prior art, the resultant genotyping would be one of the following three possibilities: AA, AB, or BB. However, by using this invention, the number of possible results is increased from 3 to 10: AA, AG, AC, AT, GG, GC, GT, CC, CT, TT . Of these ten possibilities, only 3 are the possible combinations for each SNP, e.g., a T/G polymorphism can only render one of the following three possibilities: TT, TG, or GG. Therefore, in this case, if the computed genotype falls into any of the remaining $10-3=7$ combinations (e.g., either of AA, AG, AC, AT, GC, CC, CT), it would be a clear indication that an error has occurred.

All references cited herein are incorporated by reference in their entirety.

The following examples serve to more fully describe the manner of using the above-described invention, as well as to set forth the best modes contemplated for carrying out various aspects of the invention. It is understood that these examples in no way serve to limit the true scope of this invention, but rather are presented for illustrative purposes. All references cited herein are incorporated by reference in their entirety.

Examples

Example 1

Sixteen microspheres (beads) were labeled combinatorially with two different fluorophores (FAM and Cy3). In a first round of labeling, either FAM or Cy3 labeled oligonucleotides that were complementary to the oligonucleotide (IBL) on the microsphere, were hybridized with the microsphere. Labeling of oligonucleotides was performed as is well known in the art. Hybridization conditions are known in the art.

Following a first round of hybridization, the two pools of beads were divided into two pools each and each labeled either with the FAM or Cy3 labeled oligonucleotide. This process was repeated two additional times. Thus, following four successive rounds of labeling, each microsphere was labeled with a unique code (see Figure 1). The identity of each microsphere was elucidated by determining the identity of each fluorophore in succession; the terminal fluorophore was determined and then removed to allow for the identification of the next fluorophore. In this fashion, with as few as 4 decoding steps, the identity of 16 microspheres is determined.

Example 2

A decoding scheme similar to that described in Example 1 was implemented for four color decoding. In this example, beads were labeled as described in Example 1 with the exception that 4 labels were used at each stage. 4013 beads were labeled using Bod493, BodR6G, Bod564 and BodTXR labeled oligonucleotides. 128 different bead types were identified based on the successive decoding of the four colors.

Example 3

An alternative method to using multiple colors is to use ratiometric intensities as a coding scheme. A normalizing image is acquired in which every bead exhibits its "full" intensity. Subsequent decode stages generate intensity codes by hybridizing mixtures of "labeled": "unlabeled" complementary oligonucleotides. For instance, Fig. 1 depicts three different intensity shades (low, medium, and high) which can be ratioed to a stage with all complements present at a "high" shading value. An experiment using grey scale decoding on 16 different bead types is shown in Fig. 3.

Figure 3A depicts the combinatorial pooling scheme for labeling beads with different ratios of labeled oligonucleotides. A particular oligo is present at either 100% Cy3-labeled, 40% Cy3-labeled (60% unlabeled), or 10% Cy3-labeled (90% unlabeled) fraction. Decode oligos were hybridized to the array

for 2 min. at a 50 nM concentration. Subsequently, two independent normalizing images (all oligo complements are present as 100% Cy3-labeled species) were acquired, and the resulting bead intensities compared. This is depicted in Figure 3B as the normalized values are plotted against each other. Finally, to identify or decode the beads, the alpha values (ratio of bead intensity in indicated decode stage to intensity in normalization image) are plotted for three decode stages described in (A). In stage 1, only two peaks are observed in the alpha value histogram since only 16 bead types are present on the array. Three distinguishable peaks are observed in the second and third decode stages indicating the feasibility of grey scale decoding.

Physical attributes and different "levels" of the attributes can be used as codes by which to distinguish bead types from another. Thus, for an attribute to act as a robust code, it should be possible to imbue a bead with different "levels" of a particular attribute. Each "level" of an attribute should be quantitatively well separated from other "levels". The important point is to maximize the dynamic range of the attribute measurement, and minimize the spread of the measurement.

Example 4

For a given oligo sequence, several codes can be generated by varying the distance of separation of the two fluorophores. As shown in Figure 4, a complementary sequence containing FRET dyes separated by a linker of varying lengths allows multiple codes to be generated (Figure 4A). As an example, Probe A and B are hybridized to a complementary sequence immobilized on a solid support. Probe A contains dyes separated by an 8 poly T linker, whereas probe B is hybridized to a linker with a 4 poly T linker. The efficiency of energy transfer (ratio of intensities of dye2:dye1) is greater for probe B than for probe A (Figure 4B).

Example 5

The following FRET oligonucleotides (probes ET1, ET2, ET5 and ET7) were synthesized as is known in the art and labeled with Cy3 and fluorscein separated by linkers of varying length:

ET1 T**G*ACGAGAATGGAGGTATCT

ET2 C**TGTCGC*ACGAGAATGGAGGTATCT

ET5 C**TGTCGGGGCACTCATTTGTGC*ACGAGAATGGAGGTATCT

ET7 C**TGTCGGGGCACTCATTTGTCTGTCTGGGGCGC*ACGAGAATGGAGGTATCT.

**is Cy3 * is fluorescein

An aliquot of each FRET oligonucleotide was assayed as is known in the art. Emission spectra were normalized relative to emission at 530nm. As can be seen in Figure 5 emission of fluoresein increases as the linker is shortened. The ratio of intensities of the two fluorophores, therefore, is characteristic of the distance between the fluorophores.

Example 6

To illustrate the primer extension decoding, a two-color model is illustrated in Figure 9. To decode 16 different bead addresses using only two colors. 16 different beads are labeled by attaching 16 different unique oligos (IBLs) to each of 16 batches of beads and pooling the bead batches together. Along with the beads 16 oligos (DBLs) that are complementary to these IBLs are synthesized. Each unique IBL used in this method contains two domains. The domains are composed of a binding domain and an extension domain. The pooled bead populations are then self-assembled onto a fiber optic tip end as described herein. After array self-assembly a combinatorial color decoding process is initiated as follows: A primer extension reaction is initiated at the end of the fiber bundle tip containing the arrayed beads by simultaneously hybridizing 16 complementary oligos (DBLs) to their cognate IBL. In the first extension stage, the DBLs would have exactly the same length (n) as the binding domain of the IBL. The first stage primer extension reaction would contain 16 primer oligos, a DNA polymerase, appropriate buffer, and two fluorescently-labeled dideoxynucleotide terminators. During the extension reaction, hybridized primers are extended by a single nucleotide complementary to the first nucleotide in the extension domain of the immobilized IBL. The described extension reaction 'paints' or labels any given bead with one of two possible fluorescent terminator colors. After the reaction is over the array tip is imaged to capture the color of the beads at this stage. Beads are then immediately stripped of color or label by denaturing and washing off the first extended primer at each zip code loci. This process can be repeated through multiple stages. In subsequent stages, the primer extension reactions contain the same reagents as the previous extension reaction except that the binding complementary oligos are one base longer. These ($n+1$) primers hybridize with their cognate binding domain and one base into the 'extension domain'. Figure 9 illustrates a four-stage extension process (2-color) whereby each stage allows for the assignment of a unique combinatorial color-coded 'address' to each bead type.

In addition, if this primer-based extension method were to be combined with degenerate DBL-BL combinations for decoding large combinatorial libraries, for example using the CVC (constant-variable-constant domain idea), as described herein, the extension reaction would extend into one of 10 different possible nucleotide combinations in a 'variable extension region'. This would require only three extension reactions to decodify 1000 unique addresses (i.e. $A = \text{lot}$ where L is the length of the variable region). The combination of the degenerate DBL-IBL combinations such as CVC with the primer extension method is illustrated in Figure 10.

Example 7

This method relates to fluorescent based encoding and decoding process based on the fact that fluorescent dyes have broad band absorption and emission spectra. Fewer detection channels are used than the number of encoding dyes used in the process. That is, for n detection channels, a particular dye, x , can be represented by a vector I_x in the n -dimensional space.

$$I_x = \left\{ \frac{I}{n} \mid I_1, I_2, I_3, \dots, I_n \right\}$$

where I_i is the intensity detected at channel i . Theoretically, an infinite number of vectors are possible in the n -dimensional space when n is larger or equal to 2. With an appropriate dye selection and optical filter design, for any given number of detection channels, a larger number of dyes can be used for decoding. For example, to distinguish three species, it is possible to tag the species with three different dyes, x , y , z and use only two detection channels, such that

$$\begin{aligned} I_x &= \{1 \ 0\} \\ I_y &= \{a \ 1-a\} \quad 0 < a < 1 \\ I_z &= \{0 \ 1\} \end{aligned}$$

Figure 11 illustrates the experimental results for this example scheme used in decoding. Three classes of beads were tagged with Bodipy-493, Bodipy-R6G and Bodipy564. Detection channel 1 consisted of a 410nm/20 excitation filter and a 540nm/20 emission filter. Detection channel 2 consisted of a 535nm/20 excitation filter and a 585nm/nm emission filter. Excited with a xenon arc lamp, the fluorescence emission of Bodipy-493 can only be detected in channel 1. Bodipy-564 emission can be detected in channel 2. Bodipy-R6G can be detected in both channels. As expected, the scatter plot (Figure 11) of the intensity in the two channels of the beads clearly indicates three classes. Within each class, the beads have the same relative intensity ratio (corresponding to the same unit vector) with a certain uncertainties.

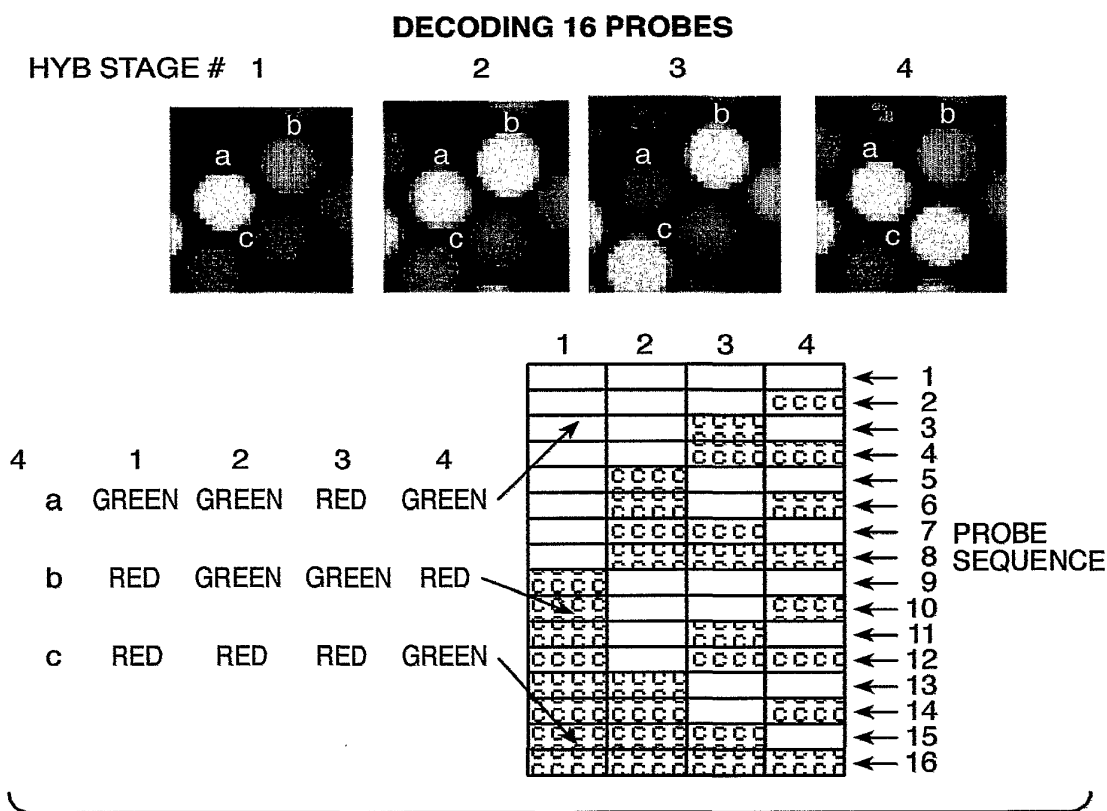
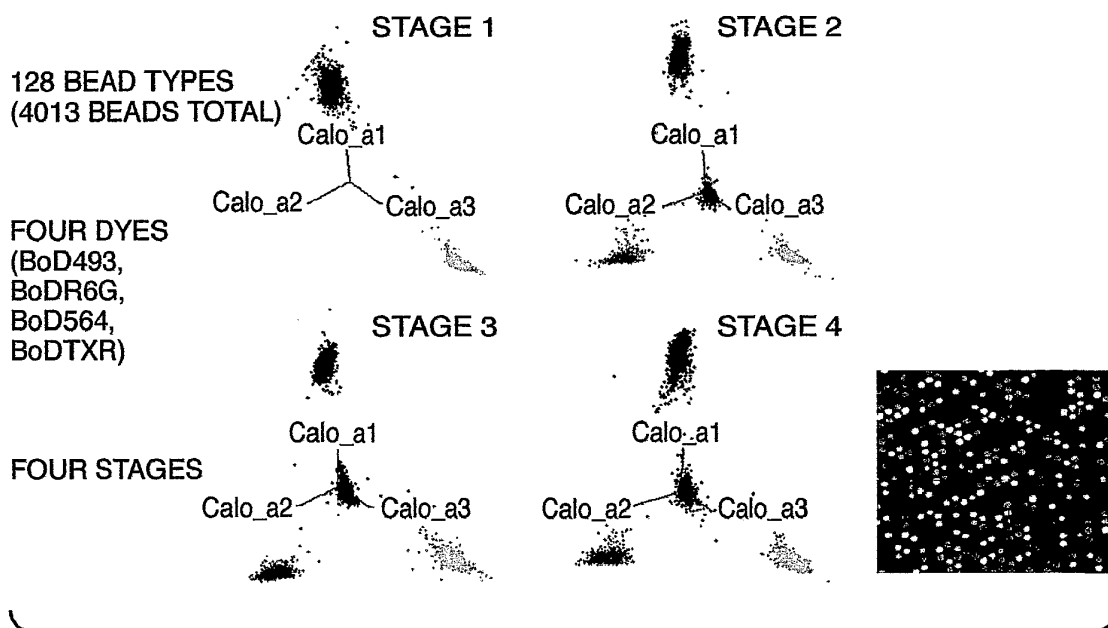
Similar design and data analysis procedure can be used for other dyes and detection channels.

CLAIMS

We claim:

1. A method of decoding an array composition comprising
 - a) providing an array composition comprising a population of microspheres comprising at least 500 subpopulations, wherein each subpopulation comprises a bioactive agent;
 - b) adding a plurality of decoding binding ligands to said population of microspheres to identify at least 500 of the bioactive agents.
2. A method according to claim 1 wherein at least one subpopulation of microspheres comprises an identifier binding ligand to which a decoding binding ligand can bind.
3. A method according to claim 1 or 2 wherein said decoding binding ligands bind to said bioactive agents.
4. A method according to claim 1, 2 or 3, wherein said decoding binding ligands are labeled.
5. A method according to claim 1, 2, 3 or 4, wherein the bioactive agent of each subpopulation is determined.
6. A method according to claim 1, 2, 3, 4 or 5, wherein said array composition further comprises a substrate with a surface comprising discrete sites, wherein said microspheres are randomly distributed on said substrate.
7. A method according to claim 1, 2, 3, 4 or 5, wherein said array composition is a liquid array.
8. A method according to claim 1, 2, 3, 4 or 5, wherein each subpopulation comprises a bioactive agent and does not comprise an optical signature

9. A method according to claim 1, 2, 3, 4, 5, 6, 7 or 8. wherein said bioactive agents are nucleic acids.
10. A method according to claim 1, 2, 3, 4, 5, 6, 7 or 8 wherein said bioactive agents are proteins.
11. A method according to claim 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10, wherein said population comprises at least 1000 subpopulations.
12. A method according to claim 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, or 11, wherein said population comprises at least 1500 subpopulations.
13. A method of determining the presence of a target analyte in a sample comprising:
 - a) contacting said sample with a composition comprising a population of microspheres comprising at least 500 subpopulations, wherein each subpopulation comprises a bioactive agent;
 - b) adding a plurality of decoding binding ligands to said population of microspheres to identify at least 500 of the bioactive agents; and
 - c) determining the presence or absence of said target analyte.
14. A method according to claim 13, wherein said subpopulations are distributed on a substrate.
15. A method according to claim 14 wherein said array is a liquid array.
16. A method according to claim 15, wherein said determining is by FACS.

**FIG. 1****FIG. 2**

SUBSTITUTE SHEET (RULE 26)

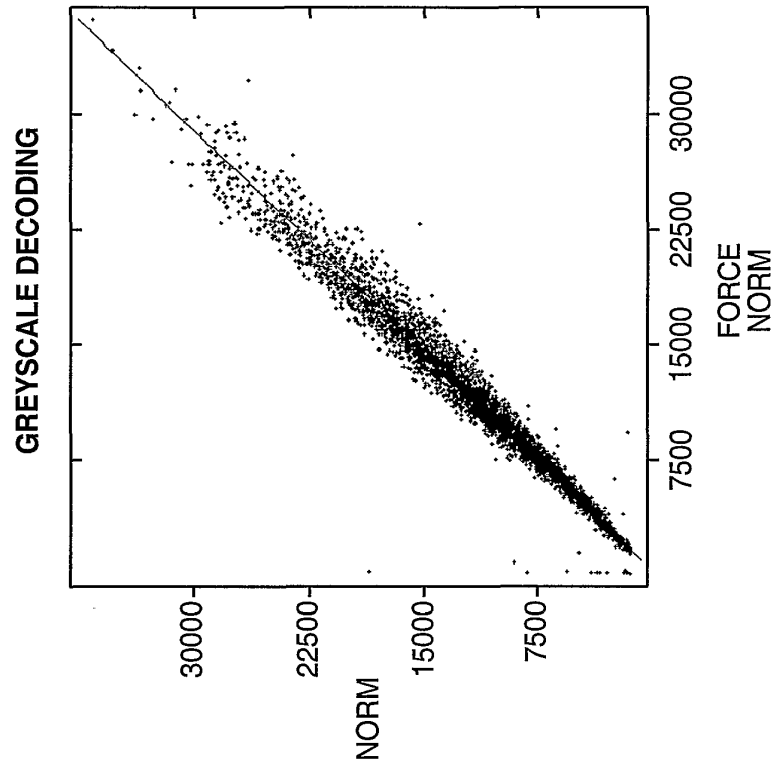
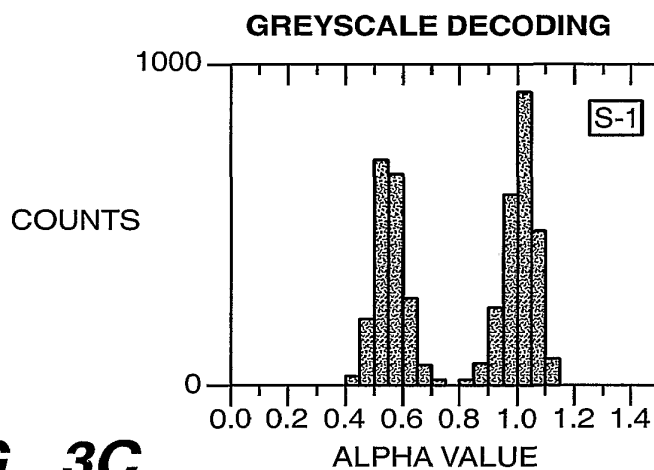
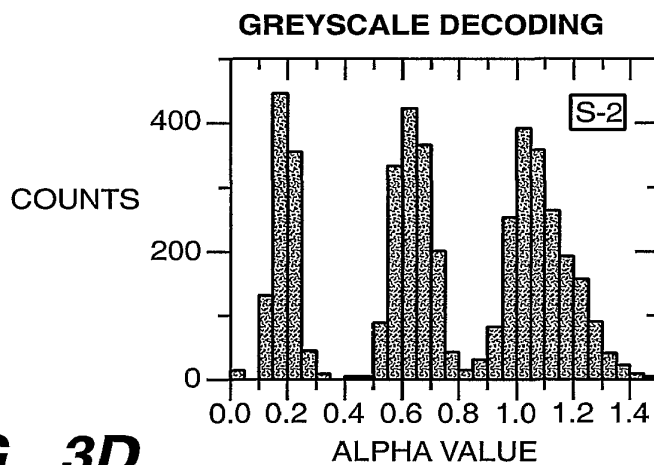
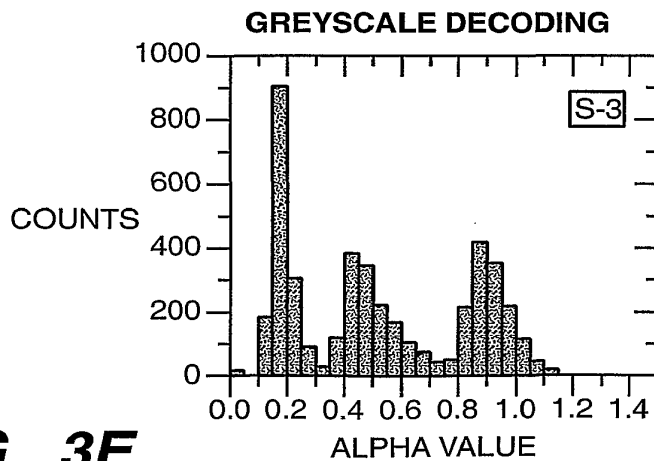


FIG._3B

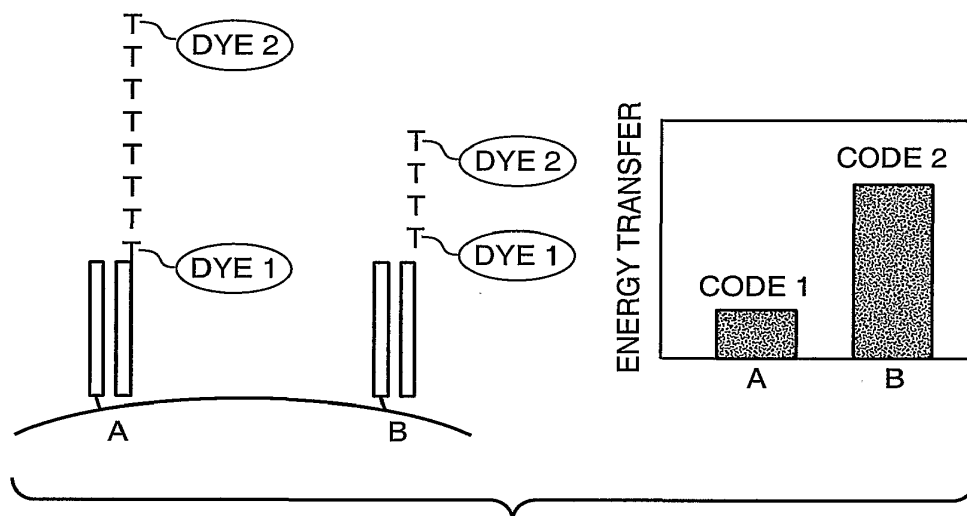
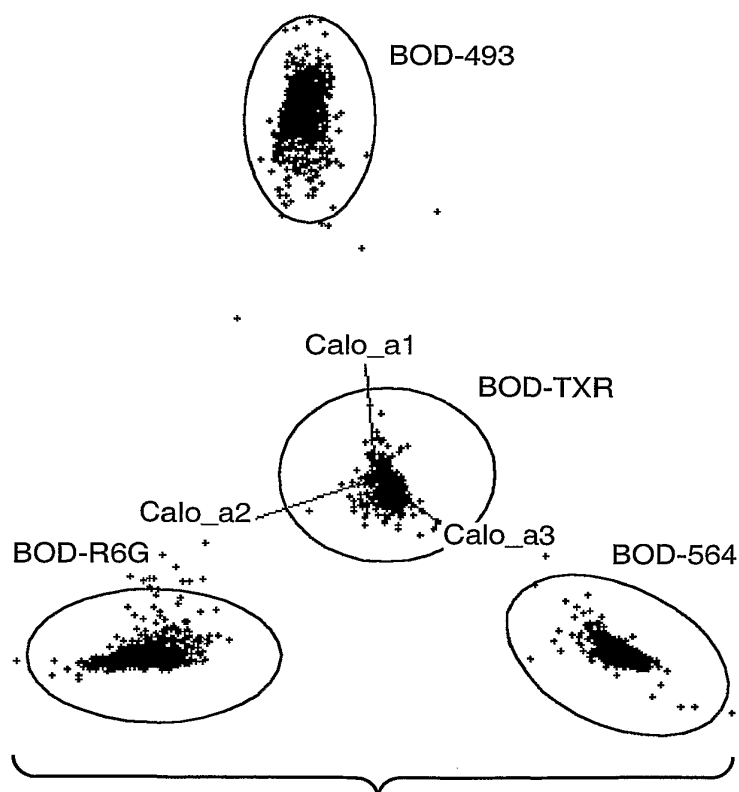
GREYSCALE DECODING

CODE	S1	S2	S3
1	100	100	100
2	100	100	40
3	100	100	10
4	100	40	100
5	100	40	40
6	100	40	10
7	100	10	100
8	100	10	40
9	100	10	10
10	40	100	100
11	40	100	40
12	40	100	10
13	40	40	100
14	40	40	40
15	40	40	10
16	40	10	100
17	40	10	40
18	40	10	10
19	10	100	100
20	10	100	40
21	10	100	10
22	10	40	100
23	10	40	40
24	10	40	10
25	10	10	100
26	10	10	40
27	10	10	10

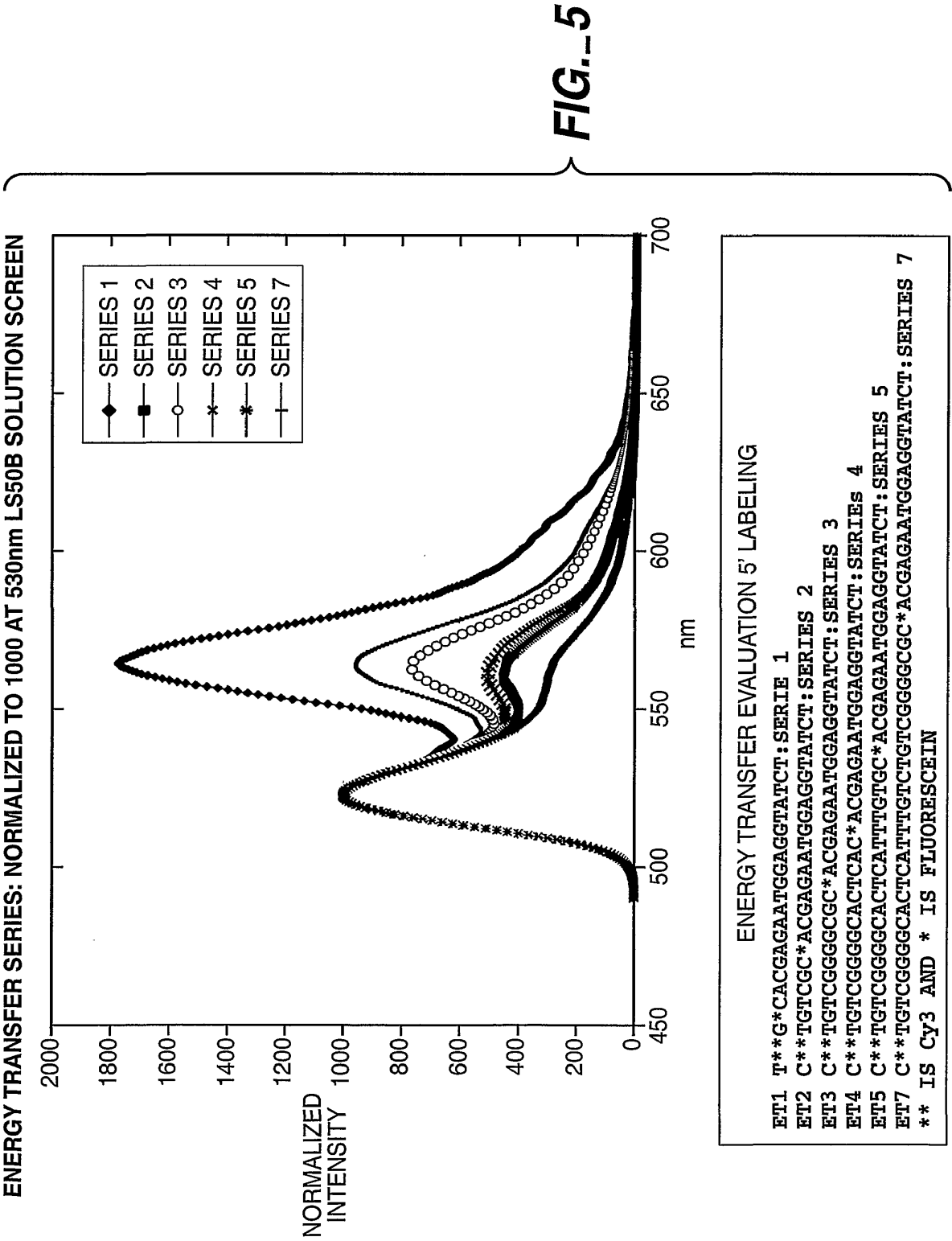
FIG._3A

**FIG._3C****FIG._3D****FIG._3E**

SUBSTITUTE SHEET (RULE 26)

**FIG._4****FIG._6**

SUBSTITUTE SHEET (RULE 26)



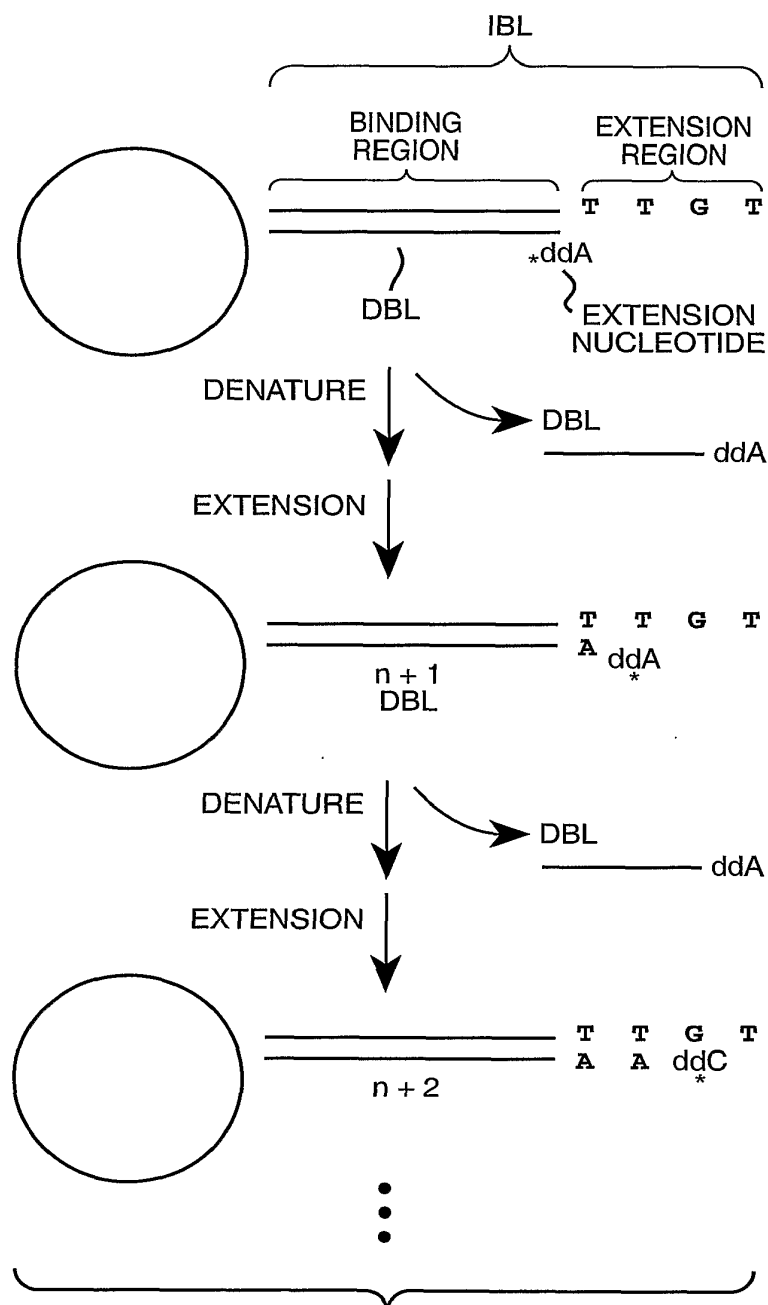
	1 "COLOR" DECODING					
BEAD	STAGE 1	STAGE 2	STAGE 3	STAGE 4	CODE	PARITY
1	1	1	1	1	1111	EVEN
	1	1	1		1110	ODD
	1	1		1	1101	ODD
2	1	1			1100	EVEN
	1		1	1	1011	ODD
3	1		1		1010	EVEN
4	1			1	1001	EVEN
	1				1000	ODD
		1	1	1	0111	ODD
5		1	1		0110	EVEN
6		1		1	0101	EVEN
		1			0100	ODD
7			1	1	0011	EVEN
			1		0010	ODD
				1	0001	ODD
8					0000	EVEN

1001 DIGIT SUM = 2, EVEN PARITY



1011 DIGIT SUM = 3, ODD PARITY

FIG. 7



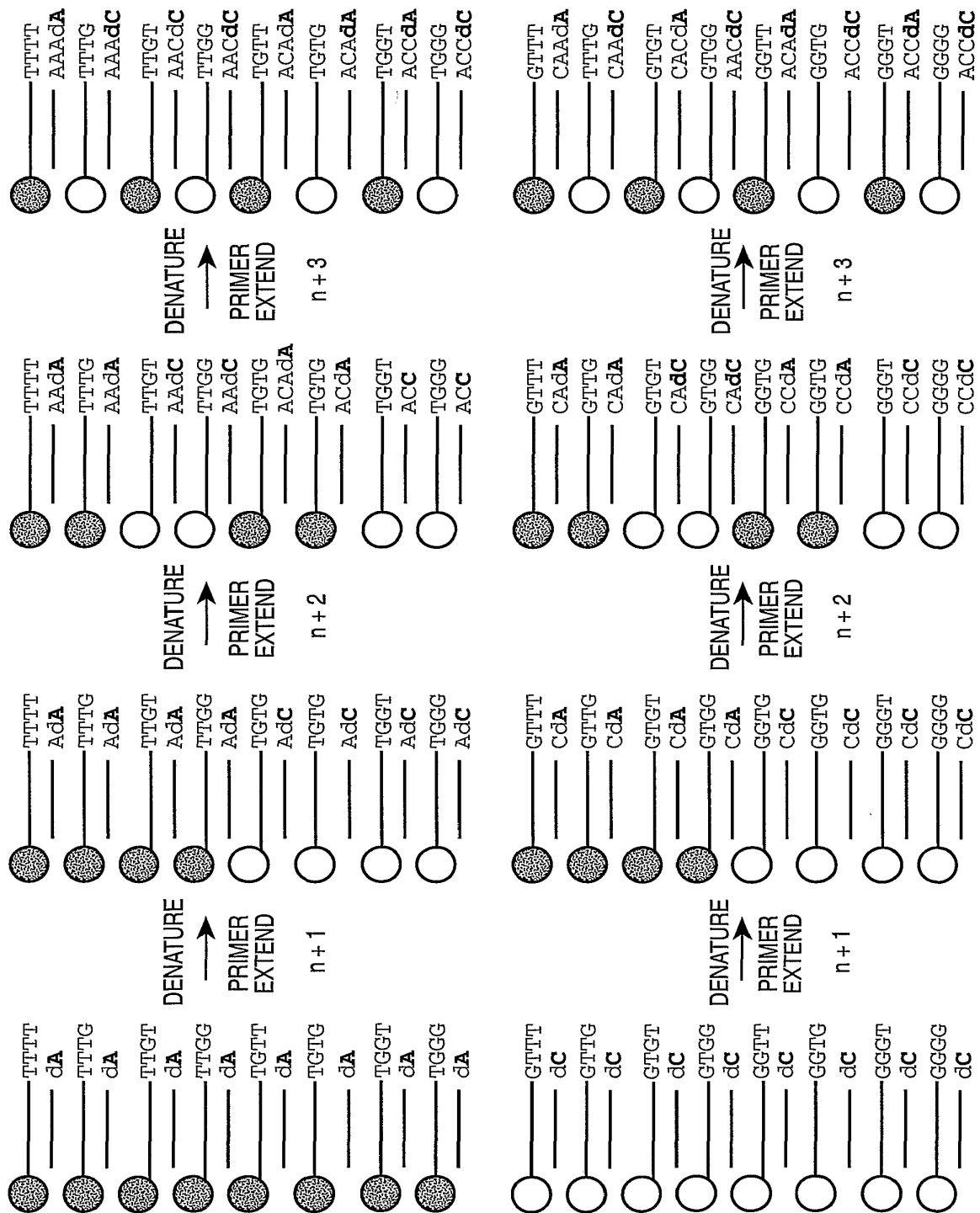
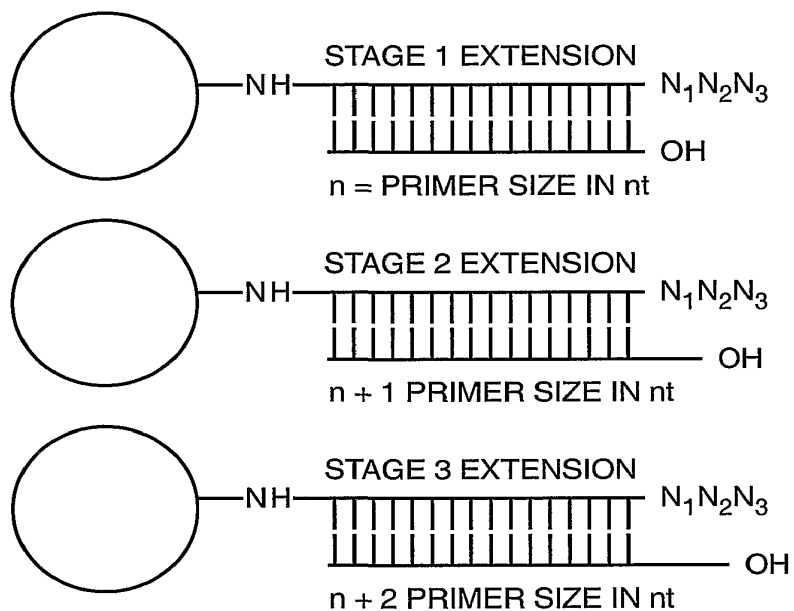


FIG. 9

SUBSTITUTE SHEET (RULE 26)



EACH 'N' WITHIN THE VARIABLE REGION
CAN CONTAIN ONE OF 10 POSSIBILITIES:

$N = A, G, C, T, A/C, A/G, A/T, C/G, C/T, G/T$

FIG._10

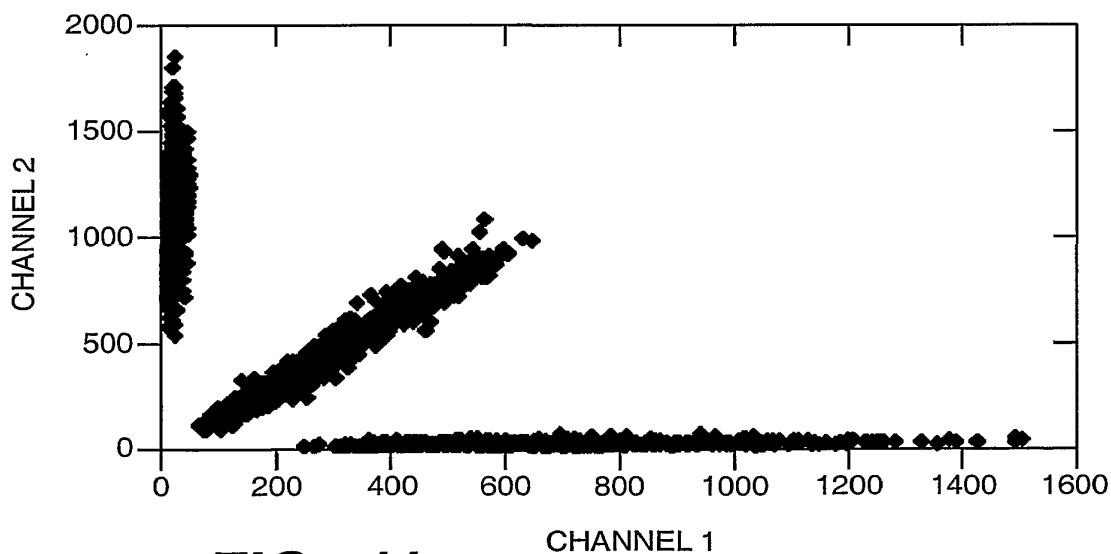


FIG._11

SUBSTITUTE SHEET (RULE 26)

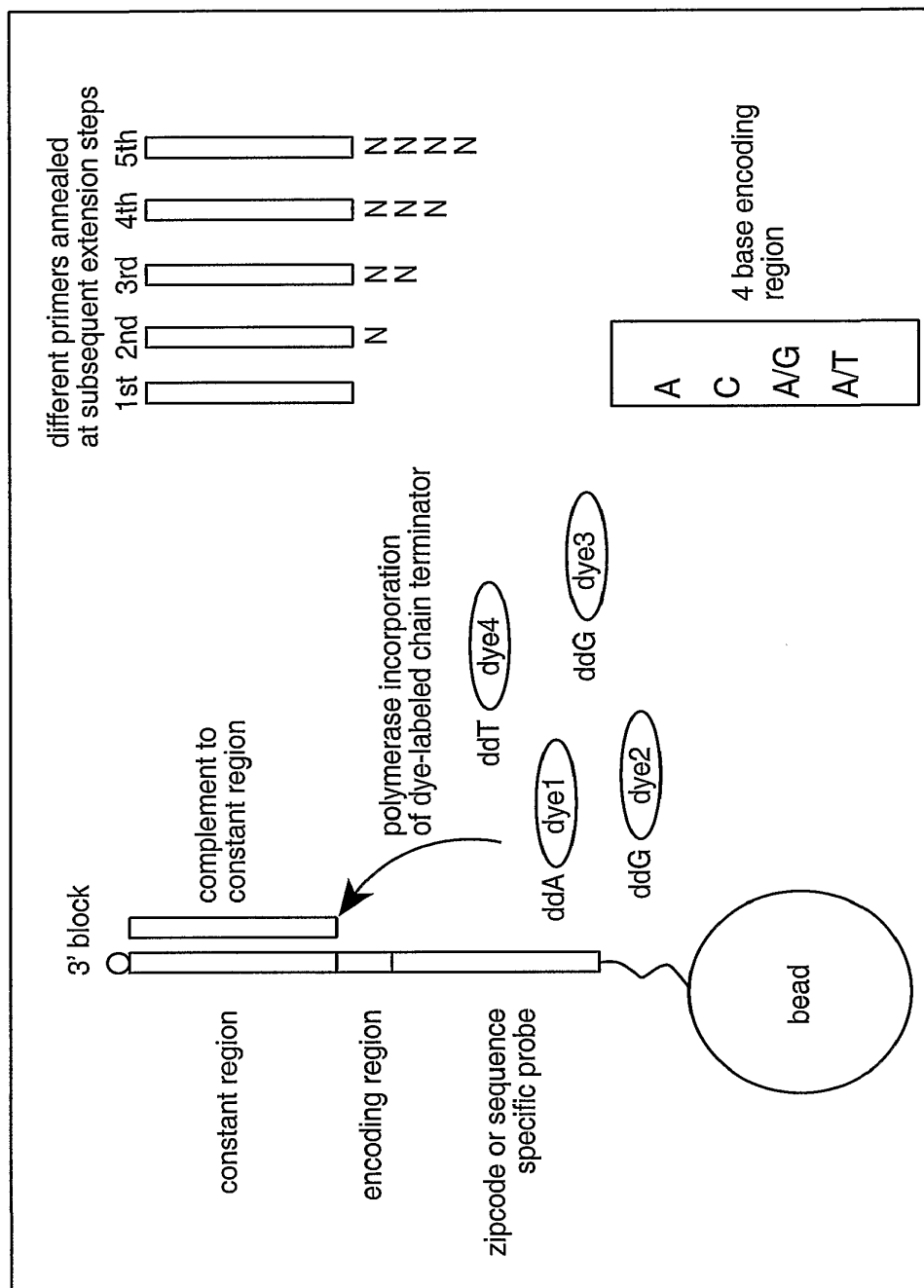
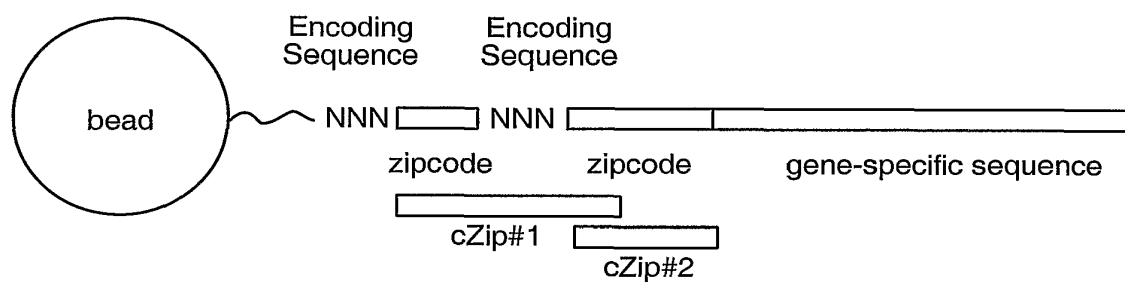
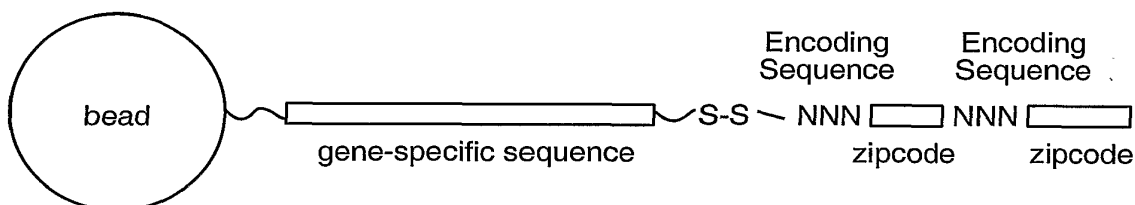
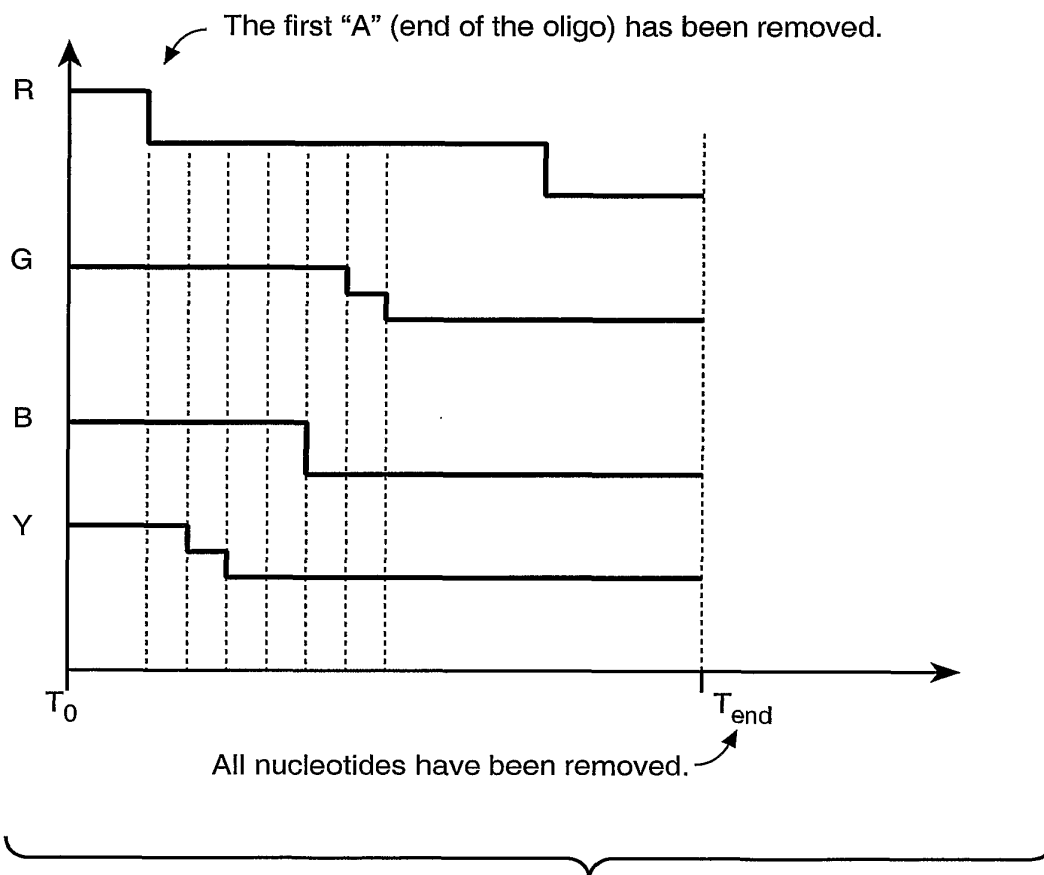
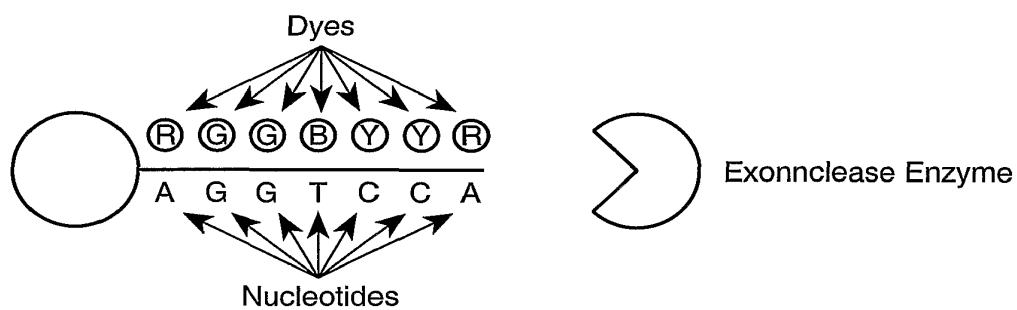
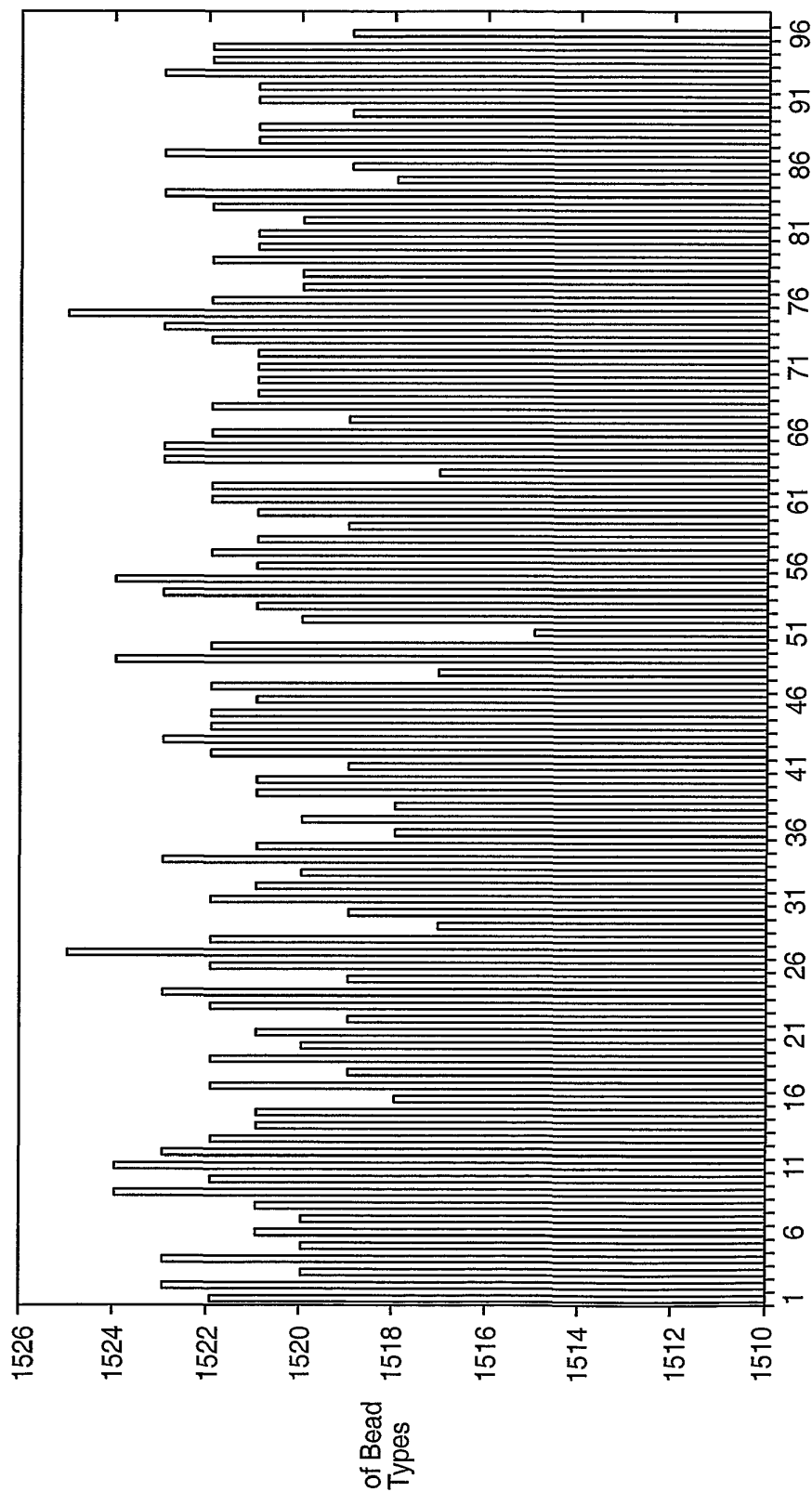


FIG. 12

**FIG. 13****FIG. 14**

**FIG. 15**

Decoding of >1500 Bead Types



Bundle Number

FIG. 16

SUBSTITUTE SHEET (RULE 26)

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

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
13 March 2003 (13.03.2003)

PCT

(10) International Publication Number
WO 03/020967 A1

(51) International Patent Classification⁷: **C12Q 1/68**

(21) International Application Number: PCT/EP02/09496

(22) International Filing Date: 24 August 2002 (24.08.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
01120466.6 28 August 2001 (28.08.2001) EP

(71) Applicant (for DE only): **ROCHE DIAGNOSTICS GMBH** [DE/DE]; Sandhofer Strasse 116, 68305 Mannheim (DE).

(71) Applicant (for all designated States except DE, US): **F.HOFFMANN-LA ROCHE AG** [CH/CH]; Grenzacherstrasse 124, CH-4070 Basel (CH).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **WEINDEL, Kurt**

[DE/DE]; Fichtenstrasse 3 b, 82407 Wielenbach-Hardt (DE). **KRAISS, Stefan** [DE/US]; 4283 Lucero Ct., Pleasanton, CA 94566 (US). **BERGMANN, Frank** [DE/DE]; Faltergatter 5, 82393 Iffeldorf (DE). **JOSEL, Hans-Peter** [DE/DE]; Ulmenstrasse 28, 82362 Weilheim (DE). **HEINDL, Dieter** [DE/DE]; Waldschmidtstrasse 1, 82327 Tutzing (DE).

(81) Designated States (national): AU, BR, CN, CZ, JP, KR, MX, PL, SG, US.

(84) Designated States (regional): European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, SK, TR).

Declaration under Rule 4.17:

— of inventorship (Rule 4.17(iv)) for US only

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.



WO 03/020967 A1

(54) Title: A METHOD FOR THE DETERMINATION OF MULTIPLE ANALYTES

(57) Abstract: The present invention is directed to a method for the determination of at least 3 analytes using specific binding partners. Combinations of detectable labels are coupled to the binding partners assuring that each binding partner molecule is coupled with only one detectable label. This allows the determination of more different analytes than different labels used by not limiting the range of useable labels. The present invention is also related to kits, composition of matters and their use.

A method for the determination of multiple analytes

5 This invention is directed to methods for the determination of multiple analytes using analyte specific binding partners wherein the binding partners are labeled with less different labels than different binding partners used. Additionally, this invention is directed to compositions of matter containing these differently labeled binding partners and the use of such compositions for the determination of multiple analytes as well as suitable kits.

10 Background of the invention

The determination of an analyte in a sample has acquired especial importance particularly in the field of health care, nutrition and ecology. Depending on the analyte different methods for determination can be used. Small molecules like metal ions, sugar monomers, amino acids often are determined by their chemical or physical properties.

15 Analytes having a higher molecular weight like proteins and nucleic acid polymers can also be determined using binding partners having a specific affinity to them. Useful pairs of binding partners are antibody - antigen, substrate - enzyme, nucleic acid - complementary nucleic acid, sugar - lectin. In case it is intended to determine a specific protein for which one knows that it exhibits antigenic properties a specific antibody can be used for the determination. For the determination of a specific nucleic acid sequence a nucleic acid probe having a complementary sequence can be used. Such binding assays and suitable protocols are well known in the art and are fully explained in the literature, see for example Sambrook et al., 1985, Molecular Cloning - A Laboratory Manual, Coldspring Harbor Laboratory, Cold Spring Harbor, New York; Nucleic Acid

20 Hybridization (B.D. Hames and S.J. Higgins, eds. 1984) and a series Methods in Enzymology (Academic Press, Inc.).

In order to allow the detection of the specific binding partner - analyte complexes it is very common to label the binding partners, whereby these labels can be detected directly or indirectly using additional reagents which bind to these labels (for example

30 using biotin as a label, which can be detected using a avidin - horseradish peroxidase

conjugate and a suitable enzyme – substrate reaction for detection). Suitable labels are known in the art.

For many cases, especially in the diagnostic field it is necessary to determine two or more analytes in order to get a complete picture about a distinct situation. For example
5 regarding a patient who might have a Chlamydia trachomatis infection, a doctor normally will not only check for a Chlamydia trachomatis but also for a Neisseria gonorrhoeae infection. Also when establishing the HLA pattern of a patient one need to determine the alleles of several different HLA loci.

When conducting a multiple analyte binding assay one has to pay attention to the labels
10 used. In a first case the same label is used for all different binding partners. In this case two alternative formats have to be differentiated. If all binding reactions are conducted together in one reaction a positive result does only show that there is at least one analyte present in the sample, but not which one it is. As an alternative one can separate each binding reaction by conducting sequential reactions or using parallel reaction formats.

15 Suitable formats are for example microtiterplate- or dot-blot-assays. Methods using such formats are known in the art. Such formats are time consuming and require a lot of handling steps, which could result in increased costs and contamination risks. More sophisticated methods using biochips are very expensive and are difficult to handle with regard to contaminations especially in the field of nucleic acid amplification.

20 In order to allow the discrete detection of several analytes within only one binding and/or detection reaction the different specific binding partners can be labeled using different labels, which can be separately detected. Such labels and useable detection methods are well known in the art. For example one can use labels which can be detected by its optical emission spectrum, whereby each of the label has a different
25 emission spectrum. Although, it has to be noted that such assays are in reality limited to the determination of only a few analytes, because there is not an unlimited number of suitable labels which are detectable separately with high sensitivity, and sufficiently stable. In case of suitable rare earth labels, expensive and highly sophisticated detectors for time-resolved fluorimetry are necessary . These limitations are of special importance
30 with regard to homogeneous detection methods. Such formats normally suffer from a higher signal background due to missing washing steps to eliminate binding partners which are not specifically bound to an analyte, which makes it difficult to interpret the multiple detection signals properly.

Multilabelling in conjunction with detectors featuring a plurality of optical channels has been used by many investigators. For instance, Vet et al. (PNAS 96, 6394-6399 (1999)) use methodology, in which 4 specific probes (Molecular Beacons) complementary to 4 different targets are labelled with 4 different reporter dyes, 1 dye each. The entire
5 fluorescence spectrum of each dye is stored in the computer and used to interpret complex multiplex results. This entails need for quite sophisticated software in order to cope with the demand to be able to resolve co-infections via spectral contribution analysis. This includes normalisation of fluorescent signals from different dyes individually for each particular dye (optical channel) on maximum gain, i.e.
10 normalisation of channel-specific growth curves pertinent to the corresponding plateau, not across different optical channels on a reference dye in order to be able to analyse signal distribution on a common basis. Other examples include Josephsson et al. (J. Clin. Microbiol. 37/3, 490-496 (1999); up to 3 differently labelled Molecular Beacons for up to 3 targets per assay) and Mercier et al. (J. Virol. Methods 77, 1-9 (1999); 2
15 differently labelled probes for 2 targets per assay).

In order to increase the number of analytes detectable with a distinct number of labels also combinations of labels for labeling of binding partners can be used. In addition to binding partners having one distinct label attached also binding partners are used which are coupled with two or more labels using the same kind of labels also coupled to the
20 single – labeled binding partners. Using for example 3 different labels 7 combinations of labels are possible by which also 7 different analytes can be detected. An in-situ-hybridization method using a similar principle is described by T. Ried et al., Proc. Natl. Acad. Sci USA, Vol. 89, pp. 1388 – 1392 (1992). A similar method for determination of 7 different Human Papilloma Virus types in a sample using PCR and a subsequent
25 probe hybridization assay are described by Samiotaki et al., Analytical Biochemistry Vol. 253, p. 156 – 161 (1997). For this purpose, oligonucleotides are 5'-terminally modified by a plurality of chelator moieties. The individual chelate units are subsequently filled with either one and the same sort of rare earth ion, or a precisely adjusted mixture of up to 3 rare earth ions. Thus, in case of > 1 sort of ion being fixed by complexation,
30 multiple fluorescent rare earth chelates (“dye mixtures”) are coupled to one and the same probe molecule. Such probes are not easy to synthesise, and poorly applicable (if at all) in homogeneous amplification/detection formats.

There are a number of other technological approaches to combinatorial labelling. Ballard et al. (US 5,759,781), Speicher et al. (Nat. Genetics 12, 368-375 (1996)) and Ried

et al. (PNAS 89, 1388-1392 (1992)) have developed assay formats, in which each specific probe is known by its labelling signature, as all copies of a given probe carry 1 or more distinct fluorophors in a combinatorial fashion. Labelling is done by means of nick translations, which is less precise in terms of number of labels incorporated per probe molecule than chemical oligonucleotide synthesis, i.e. the methods doesn't lend itself well to accurately analysing discrete distributions. Therefore an expert would not use such probes in nucleic acid amplification methods, especially not in homogeneous amplification methods.

The so-called CFET probes described by Tong et al. (Nat. Biotechnology 19, 756-759 (2001)) have a principally different design and mode of application. 1 to 3 dyes are coupled 5'-terminal per probe molecule, with spacer units in precisely adjusted number interspersed. The oligonucleotide moiety follows after the last dye or spacer towards the 3'-end. Thus, there is a mixture of different probe types as with Ballard et al. The spacers function as tuners for electrophoretic mobility, and for resonance intensity as well. The latter affects the fluorescent signature of the particular probe. Probes are excited at a common wavelength, and the recorded signal is resolved by the position in the capillary electropherogram and by overlay of the complete emission spectra at the particular position and calculation of relative contributions from components of the label set directly as digital ratios. Again, there is no signal distribution analysis of corrected & normalised signals *per se*, nor dynamically along the reaction coordinate, no "same sequence / different label" oligonucleotide design, and no signal generation as a consequence of a biochemical reaction altering probe configuration or ~ constitution.

The labeling of binding partners with more than one label per binding partner molecule bears some disadvantages. Such multiple labeled binding partners are rather demanding (and costly) in synthesis and steric hindering of the labels can occur which might decrease the efficiency of the binding of the binding partner to its analyte or the detection of the attached labels. Also interactions between the labels attached to the binding partner might occur, which could lead to wrong results. This could be relevant for example when using optical labels. It is certainly of especial importance when conducting homogeneous detection methods using fluorescence energy transfer labels as for example used in TaqMan assays (US 5210015, EP 0543942).

Therefore, it is an object of the present invention to improve the methods for determination of multiple analytes avoiding all or a part of the disadvantages of the known methods.

Summary of the invention

The main aspect of the present invention is related to a method for determination of multiple analytes using less labels than analytes determined. This is achieved by labelling the whole ensemble of specific probe molecules for some targets with a particular dye
5 each, while for other targets one part of the ensemble of specific probes is labelled with a given dye, while another part of the ensemble is labelled with a different dye (same sequence / different label oligonucleotides in admixture). Preferably by performing homogeneous (solution-phase) real-time amplification assays (for example by PCR or TMA), correcting the resulting signal for all kind of relevant noise, and normalising the
10 signal from different labels on a reference dye, targets can be deduced from the distribution of processed signal collected in a multi-channel detector. For this purpose combinations of labels are used.

For example when determining 3 analytes only two different detectable labels are necessary. A first label is attached to a first binding partner specific for the first analyte.
15 A second label is attached to a second binding partner specific for the second analyte. For labeling of the third binding partner specific for the third analyte the same kind of labels as has been coupled to the first and second binding partners are used. One amount of the third binding partner is coupled with the first label. In case of three analytes preferably one half of the third binding partner is labeled this way. Another
20 part of the third binding partner is coupled to the second label, which would be in case of three analytes preferably the remaining half of the third binding partner. Therefore each binding partner specific for this analyte contains a first label or a second label. This results in binding partner molecules each labeled with only one label, preferably enabling multi-colour analysis for resolving multiplex results based on signal generation
25 which is derived from a biochemical reaction and thus provides means for kinetic response recording, while maintaining simple probe designs and state-of-the art instrumentation for detection. In contrast thereto probes are described in the literature that have bound multiple labels to one probe molecule (Samiotaki, M. et al., Analytical Biochemistry 253, 156 – 161 (1997)). Such probes are more difficult to synthesize
30 compared with the binding partners according to the present invention.

Different detection efficiencies which may occur for the different labels attached to the third binding partner, can be compensated for in 2 ways. In case a first label has a higher signal output compared to a second label, one can adjust this difference either chemically by mixing the third binding partner coupled to the first label with the third

binding partner coupled to the second label in a defined non-1:1 ratio, or mathematically via normalization of signal output to a given label selected as standard. Such different detection efficiencies may be due to the intrinsic properties of the labels, such as absorptivity or quantum yield, but can also result from different coupling efficiencies, or susceptibility to solvent effects. In the latter case the mixing of the different amounts of the third binding partner following the coupling reaction provides a good possibility to adapt such discrepancies.

In addition possible interference which may occur when two or more detectable label are attached to one binding partner molecule is avoided. Therefore, binding partners according to the present invention are compatible with broad spectrum of labels, which also encompass fluorophores and combinations of fluorophores, like classical Förster type labels (Styer and Haugland, Proc. Natl. Acad. Sci. USA 98, 719 (1967)) which can be used for example in homogeneous detection assays like the TaqMan-method for determination of low concentrated nucleic acid analytes.

It should be noted that also an increased number of analytes may be determined using three or even more labels which can be combined in the same way as shown for the two labels. When using three labels up to seven analytes may be determined. Using four different labels even 15 analytes can be detected if applied singular as well as in dual, triplicate and quadruplicate combinations. If applied singular or in dual combinations only (which might be advantageous from a practical point of view), up to 10 different analytes can be differentiated with a set of 4 labels.

Therefore the present invention is related to a method for the determination of at least 3 analytes comprising the steps:

- a) Providing a mixture of a sample containing said at least 3 analytes or suspected of containing one or more of said analytes and at least 3 different binding partners under conditions allowing the specific binding of said binding partners to said analytes, wherein
 - a first of said at least 3 binding partners is specific for a first of said at least 3 analytes and is coupled to a first label,
 - a second of said at least 3 binding partners is specific for a second of said at least 3 analytes and is coupled to a second label, which label is separately detectable from the label coupled to said first binding partner, and

- 5 - a third of said at least 3 binding partners specific for a third of said at least 3 analytes whereby a first amount of said third binding partner is coupled to the same label as coupled to said first binding partner and a second amount of said third binding partner is coupled to the same label as the second binding partner and said first amount of said third binding partner is not coupled to the same label as coupled to the second binding partner and said second amount of said third binding partner is not coupled to the same label as coupled to the first binding partner,
- 10 b) Detecting the signal intensities indicative for said first and second label,
- c) Determining the analytes present in said sample using said signal intensities detected in step b).

Preferred analytes are nucleic acid analytes which can be determined using sequence specific probes. Such nucleic acid analytes can also be amplified nucleic acids using one of several nucleic acid amplification method known in the art, like LCR (U.S. Patent 15 Nos. 5,185,243, 5,679,524 and 5,573,907; EP 0 320 308 B1; WO 90/01069; WO 89/12696; and WO 89/09835), cycling probe technology (U.S. Patent Nos. 5,011,769, 5,403,711, 5,660,988, and 4,876,187, and PCT published applications WO 95/05480, WO 95/1416, and WO 95/00667), Invader TM technology (U.S. Patent Nos. 5,846,717; 5,614, 402; 20 5,719,028; 5,541,311; and 5,843,669), Q-Beta replicase technology (U.S. Patent No. 4,786,600) , NASBA (U.S. Patent No. 5,409,818; EP-0 329 822), TMA (U.S. Patent Nos. 5,399,491, 5,888,779, 5,705,365, 5,710,029), SDA (U.S. Patent Nos. 5, 455,166 and 5,130,238) and PCR (US-A-4,683,202), whereas the PCR method is most preferred.

- The invention is also related to a composition of matter comprising
- 25 - a first binding partner specific for a first analyte coupled to a first label,
- a second binding partner specific for a second analyte coupled to a second label, which label is separately detectable from the label coupled to said first binding partner, and
- 30 - a third binding partner specific for a third analyte whereby a first amount of said third binding partner is coupled to the same label as coupled to said first binding partner and a second amount of said third binding partner is coupled to the same label as the second binding partner and said first amount of said third binding partner is not coupled to the same label as

coupled to the second binding partner and said second amount of said third binding partner is not coupled to the same label as coupled to the first binding partner.

and its use for the determination of at least three analytes in a sample.

- 5 Furthermore the invention is related to kits for determination of at least three analytes.

Brief description of the drawings

Fig. 1 shows a possible data processing scheme for an automated multiple analyte determination assay using the determination methods described. See also example 2.

Detailed description of the invention

- 10 Analytes according to the present invention are analytes which can be determined by binding assays known in the art. These are preferably components of samples for medical diagnostics or other biological analytics, i.e. in particular ingredients of body components such as antigens, antibodies, cells or nucleic acids. Such assays can be used for example for determination of infectious agents like bacteria e.g. chlamydia, neisseria
15 and mycobacteria and viruses like HBV, HCV and HIV. Depending on the aim of an assay one can determine the amount of an analyte present in a sample as well as the configuration of an analyte, for example a nucleic acid analyte can be analysed for its allelic form or whether a patient carries a mutated form.

A sample according to the present invention contains the analyte to be determined.

- 20 With regard to medical diagnostics samples derived from human or animal are preferred, e.g. whole blood, tissue sections, urine, sputum, serum, plasma, buffy coat and smears can be used. Depending on the analyte and the sample it might be necessary to pre-process the sample in order to allow determination of the analyte, for example for most samples nucleic acids need to be extracted in a first step. Such pre-processed
25 samples are also samples according to the present invention.

- According to the present invention at least three analytes are determined. This could be for example a pattern of infectious agents in a patient, like the viruses HIV, HBV and HCV, which have to be tested for each blood sample in blood banks. Another example could be the determination of the histocompatibility locus antigene pattern, which
30 consists of several loci and distinct allelic forms.

The analytes are bound by specific binding partners. Depending on the analyte one has to choose specific binding partners. Analytes having antigenic properties can be bound

by using specific antibodies. Antibodies in a sample can be determined by using the specific antigens as binding partners. If a nucleic acid analyte should be determined one can use a nucleic acid sequence being complementary to the analyte as specific probe. Further specific binding pairs like substrate – enzyme or sugar – lectin are known in the art, which might also be useful in the described methods.

A nucleic acid analyte is usually brought into available form by processing the original sample with one of various methods. This comprises for example change of pH (alkaline), heating, cyclic changes of temperature (freezing/thawing), change of the physiological growing conditions, use of detergents, chaotropic salts or enzymes (for example proteases or lipases), alone or in combination. For example one can use magnetic glass particles which are able to bind nucleic acid under specific conditions. Suitable particles and protocols are described in WO 96/41811 and WO 01/37291.

It is important to choose reaction conditions which promote the specific binding of the binding partners to the analytes which allows that the specific binding complexes occur but minimize the binding of the binding partners to unrelated reaction and sample components leading to an increased background signal. Such reaction conditions and suitable protocols are known in the art.

A specific nucleic acid binding partner, also called probe, for determination of a nucleic acid analyte or an amplified nucleic acid analyte preferably is an oligonucleotide, but also analogues having for example a peptide-backbone instead of the natural phosphate-sugar backbone (PNA, WO 92/20702) can be used. In order to allow a specific binding of the probe to the analyte the probes are preferably longer than 10 nucleotides, even more preferred have a length of 10 to 40 nucleotides. It is further required that the probe is sufficiently complementary to the analyte sequence. Therefore probes are preferably at least 80 % complementary, more preferred are at least 90 % complementary. In the most preferred case the probes are fully complementary to the analyte. The exact determination of complementarity and homology can be determined by using computer programs such as FastA (Pearson and Lipman, Proc. Natl. Acad. Sci. USA Vol. 85, pp. 2444 – 2448 (88)).

Depending on the initial concentration of a nucleic acid analyte it might be necessary to amplify the analyte in order to allow its determination. For this purpose several amplification methods are known in the art (as referenced above). Especially when using primer based amplification methods a specific amplification of a nucleic acid analyte is possible. In combination with a probe binding assay an increased specificity of

an assay can be achieved. Also methods are known in the art, like LCR, which use labeled primers in order to allow the detection of an analyte. Also such methods can be modified according to the present invention.

5 A primer according to the present invention is a molecule capable of being extended or modified preferably by enzymes, more preferably by a polymerase of pro-
caryotic origin, when hybridized to a nucleic acid template. When using PCR
methodology thermostable polymerases like *T. aquaticus* or *T. thermophilus* DNA-
polymerase are preferred. These extend primers by adding mononucleotide units from
monodesoxyribonucleosidetriphosphates to the 3'-OH-terminal end of said primers.
10 The overall length and base sequence of a primer is dictated by the required specificity of the amplification reaction. Preferred primer lengths for performing PCR are from 10 to 40, most preferred from 15 to 30 base containing subunits, selected from
mononucleotides and/or nucleic acid analog monomers. In general primers of that
length are also useful for other amplification methods. If more than one primer is used
15 for amplification, for example when using PCR or amplifying multiple target nucleic acids in one reaction, preferably primers are used which cannot hybridize to each other, because they do not contain any stretch of more than 5 consecutive complementary bases.

Labels are generally known to those skilled in the art as being a group which is
20 detectable or can be made detectable for determination the presence of an analyte. Well-known labels are fluorescent labels, like fluoresceine and lanthanide chelates, electro-
chemiluminescent labels, like ruthenium complexes, or moieties that can be recognized by another molecular entity, like haptens which can be recognized by an antibody raised
against this hapten or moieties that can be immobilized, like biotin which can be bound
25 for example to streptavidin coated solid phases, like beads or tubes. Most preferred labels especially with regard to homogenous formats are chromophores. Such
chromophores can be used alone or in combination with another chromophore or for
example with a non-fluorescent quencher.

In case homogeneous detection formats are used, especially for determination of nucleic
30 acid analytes several formats are known in the art. Method like the TaqMan assay (US Nos. 5,210,015 and 5,487,972) and the kissing probe- assay are based on fluorescent
energy transfer of fluorescent dyes (FET, Styer and Haugland Proc. Natl. Acad. Sci. USA Vol. 98, pp. 719- (67)). When brought in close proximity with each other a first
fluorophor can interact with a second fluorophor or a non-fluorescent quencher. For

example, the electromagnetic emission or vibrational excitation of a first fluorescent dye can induce resonance in a second fluorescent dye. Depending on the format used one can for example detect a decreasing light emission of the first fluorescent dye or an increasing light emission of the second fluorescent dye as a mean for the presence of an
5 analyte. Combinations of fluorescent dyes, which can be used for this purpose, are known in the art. Examples are classical dyes suitable for Förster-type Resonance Energy Transfer like Pentamethin-indodicarbocyanin (Cy5) combined with 6-Carboxyfluorescein (6-FAM).

Label in this context means a signal generating entity, which is actually detected in the
10 method of the present invention. With regard to the homogeneous formats described, pairs of fluorochromes are used, which get into resonance with each other and which are detected in these methods as a single detection signal. Although two distinct molecules, for example two fluorochromes, are involved, they function as one single label. In addition, the term label has to be understood that it could also mean that more
15 than one label molecule of a distinct label is coupled to a binding partner molecule.

Depending on the labels used different detectors have to be used for measuring of the signals of the labels. Beside others, especially fluorimeters are widely used for this purpose. Fluorescent labelling is most advantageous in conjunction with homogeneous PCR, as no chemical trigger reagent has to be added for signal generation, in contrast to
20 enzymatic or chemiluminescent labelling techniques. This allows for a closed tube procedure, the most effective way to avoid cross-contamination with amplified material. When determining multiple label signals it is important that the signals are distinguishable from each other. In order to avoid such cross-talk it is preferred when using optical labels that each label has a different emission and/ or extinction-
25 spectrum. Although when using most of the commercially available labels in multiple analyte detection assay at least some cross-talk cannot be avoided. When determining only a few analytes this cross-talk can be compensated by processing the measured signals with computer programs or using more expensive spectral fluorimeters instead of filter based fluorimeters. With regard to homogenous formats especially when using
30 fluorescence energy transfer-based methods the need of additional chromophores even increases the possibilities of interferences. Therefore in reality the number of suitable labels which can be used together in a multiple assay is limited.

According to the present invention at least one of the binding partners used is coupled with a combination of detectable labels, whereby each of these binding partner

molecules is coupled to only one detectable label. In contrast probes as described by Samiotaki et al (as referenced above) are labeled with multiple labels (e.g. 10-20 coupled 5'-terminal per oligonucleotide). Due to the proximity of the detectable labels coupled to the same probe molecule the risk of interference between the labels is very high and
5 not all types of labels especially fluorescent dyes can be used.

The methods according to the present invention can be used for determination of an increased number of analytes in a multiple analyte assay and are compatible with a broad range of labels.

By decreasing the number of labels necessary in a multiplex assay the need for exotic
10 labels is avoided and also cheaper detectors like filter based fluorimeters can be used instead of spectral fluorimeters which are much more expensive. This further allows for homogeneous formats, which are of outstanding importance with regard to commercial diagnostic assays due to the limited contamination risks and fewer handling steps.

As indicated above, three labels combined in different combinations can be coupled
15 with up to seven different specific binding partners allowing the distinct determination of six different analytes. This is visualized in the table below. P stands for Parameter = analyte, a channel corresponds to an optical tract as part of a detector optimised to capture the signal of a particular detectable label. The table shows the expected percentage distribution of signal over all three channels, being normalized either chemically or
20 mathematically. Considered is the case that of the analytes P1 to P7 only one at a time would be present in a sample. The labeling of the different binding partners are: type P1 – Label1 (signal in channel 1, P2 – Label 2 (signal in channel 2), P3 – label 3(signal in channel 3), P4 – half amount label 1, half amount label 2, P5 - half amount label 1, half amount label 3, P6 - half amount label 2, half amount label 3 and P7 – 1/3 amount label
25 1, 1/3 amount label 2, 1/3 amount label 3.

Parameter	Channel		
	1	2	3
P1	100		
P2		100	
P3			100
P4	50	50	
P5	50		50
P6		50	50
P7	33	33	33

If a signal can be only measured in one channel a clear indication for the presence of one of the analytes of parameter 1, 2 or 3 is given (see table). If signals occur in two channels in a 50: 50 ratio there is a strong indication for an analyte of parameter 4, 5 or 6 as indicated in the table. A signal ratio of 33:33:33 would indicate that a P7 analyte is present in the sample. However this reflects an idealized situation. Real signals may deviate and often need to be normalized in order to compensate for example different detection efficiencies of the labels, background- and cross talk-signals.

The different labels coupled to the binding partners should not change the binding efficiency of a binding partner, which might also be adjustable by varying the ratios of binding partner bound to the first label: binding partner bound to the second label. Also different detection efficiencies can be adjusted by this way.

A 50: 50 signal ratio might also occur in case of the presence of two analytes. However this would probably be extremely rare, since their occurrence would depend on either a 1:1 co-infection in terms of number of starting molecules and identical extraction / amplification / detection efficiencies yielding a 50:50 signal distribution into 2 channels, or on different numbers of starting molecules in conjunction with extraction / amplification / detection efficiencies that precisely counter-balance the differences in titer and again yield a 50:50 distribution. Thus, co-infections of 2 parameters indicated by single labeled probe, respectively, e.g. P1 and P2, should result in an uneven distribution of >>50: <<50: 0 or <<50: >>50: 0 and so be distinguishable from e.g. P4

indicated by an even distribution of 50: 50: 0. However such results could also derived from samples co-infected with two agents for example P4 and P1.

In order to further decrease the risk of ambiguous results and possible errors in interpretation of the measured signals it is possible to arrange the parameters in a way that signals occurring in two or more channel can be clearly assigned to one possible analyte distribution in the sample as explained in example 1. In addition one can run an additional assay especially testing for the remaining possible analyte distributions which may be present in the sample. Such a sequential method is very economical as it allows to determine a distinct analyte distribution in most samples in a first run, whereas ambiguous samples can be re-evaluated in a second run using for example a multiple analyte assay with a reduced number of analytes (for which an unambiguous one label – one binding partner combination should be possible). Furthermore regarding nucleic acid analytes and amplified nucleic acid analytes one could for example determine the melting curve of the analyte-probe hybridization complexes or of the amplified nucleic acid, which allows distinguishing the different analytes and coming to an unambiguous result.

Also kits for determination of at least three analytes are subject of the present invention. Such kits containing in one or more containers

- a first binding partner specific for a first analyte coupled to a first label,
- a second binding partner specific for a second analyte coupled to a second label, which label is separately detectable from the label coupled to said first binding partner, and
- a third binding partner specific for a third analyte, whereby a first amount of said third binding partner is coupled to the same label as coupled to said first binding partner and a second amount of said third binding partner coupled to the same label as said second binding partner and said first amount of said third binding partner is not coupled to the same label as coupled to the second binding partner and said second amount of said third binding partner is not coupled to the same label as coupled to the first binding partner.

In addition such kits can also contain means for specific binding of the binding partners to their analytes, which may include buffers, blocking reagents and other reaction components known in the art. Kits for determination of nucleic acid analytes may also

contain reaction components for nucleic acid amplification methods as described above. A PCR reaction kit could for example also contain a polymerase, buffers, nucleotide triphosphates and/ or primers.

The present invention is exemplified by the following examples:

Examples

Example 1

Multi-color analysis according to the present invention can be done using the Roche Cobas Taqman amplification/ detection technology and 5'-nuclease assay technology (EP 0543942 and US 5210015) building on "classical" fluorochrome compounds and "classical" resonance energy transfer principles. A Roche Cobas Taqman instrument is equipped with up to 4 filter pairs (see also EP 0953379, EP 0953837, US 6134000 and US 6084669). A putative indicator set could consist of 4 reporters and preferentially 1 or 2, optionally non-fluorescent, quenchers. Candidate compounds that span the spectral range set by the Cobas Taqman detector, i.e. ca. 420 nm – ca. 710 nm, may include:

- for quencher (Q) dyes with broad-banded absorption range in the red, for instance polymethine-cyanine dyes (e.g. $n = 5$, i.e. 4 conjugated olefinic entities building the mesomeric structure), or rhodamin derivatives at the flanks symmetrically fused to unsaturated N-heterocyclic moieties that participate in the conjugated mesomeric π - e^{-} -system, both preferentially as non-fluorescent nitrophenyl-derivatives
- for reporter (R) dyes fluorescent molecules like coumarine dyes, fluorescein-type dyes like FAM or chlorinated derivatives, rhodamin-type dyes, oxazines or bodipy-type compounds. Three of these would be assigned to target parameters, one for monitoring IC.

General requirements pertinent to candidate dyes include: good chemical stability in stock solution and also at use level as ingredients in mastermix reagents; stable when exposed to ambient light during synthesis, purification and kit manufacture; stable when exposed to multiple heating /cooling cycles (e.g. $n = 60$) in PCR; sufficient spectral overlap and efficient energy transfer between corresponding RQ resonance pairs; high extinction coefficients (i.e. absorptivity); and especially for reporters high quantum yield in aqueous solutions and in the pH range employed for Taqman-PCR (e.g. pH 7-8,5).

The new concept allows the measuring of the distribution of crosstalk- and background-corrected normalized signals of 3 reporter dyes and dual combinations thereof across the corresponding 3 optical channels of Cobas Taqman instead of having a fixed assignment of a particular dye, representing a particular assay parameter, and a corresponding channel. The principle would provide for a complete resolution (via

reporter R1, R2, R3, R1+2, R1+3, R2+3) of a positive result obtained with a 6-parameter-multiplex assay, the maximum complexity taken into consideration.

Theoretically, in cases of single-label use, 100% of the respective processed signal is found in a given channel; in cases of dual label use, processed signals would be distributed evenly across 2 corresponding channels (50% + 50%).

Discrimination of positive / negative pools (donations) could be done based on an optimized algorithm for determination of threshold cycles (ct values, i.e. the point on the reaction coordinate, where the fluorescent signal intensity rises significantly above background level), sensing the slope profile of signal-over-time curves.

Resolution of the kind of infection (assay parameter, i.e. kind of pathogen, yielding detectable product) could be done via analysis of channel-specific contributions to the total area under the signal/time curves (AUC), or to the total normalized plateau signal intensity. There is no firm dose-response correlation for criteria like these which, however, is not mandatory for qualitative assays as needed for blood screening.

A suitable 6-parameter (P1 to P6) multiplex assay could comprise the following probes (Q= quencher, R= reporter and N_i= one of the four nucleotides):

parameter 1 - probe = 5'-Q-(N_i)_g-R1-(N_i)_h-3'-PO₄ + 5'-Q-(N_i)_g-R2-(N_i)_h-3'-PO₄
5 pmol each / 100 µl reaction mix

parameter 2 - probe = 5'-Q-(N_i)_a-R1-(N_i)_b-3'-PO₄
10pmol / 100 µl reaction mix

parameter 3 - probe = 5'-Q-(N_i)_c-R2-(N_i)_d-3'-PO₄
10pmol / 100 µl reaction mix

parameter 4 - probe = 5'-Q-(N_i)_e-R3-(N_i)_f-3'-PO₄
10pmol / 100 µl reaction mix

parameter 5 - probe = 5'-Q-(N_i)_k-R1-(N_i)_l-3'-PO₄ + 5'-Q-(N_i)_k-R3-(N_i)_l-3'-PO₄
5 pmol each / 100 µl reaction mix

parameter 6 - probe = 5'-Q-(N_i)_n-R2-(N_i)_m-3'-PO₄ + 5'-Q-(N_i)_n-R3-(N_i)_m-3'-PO₄
5 pmol each / 100 µl reaction mix

In other words:

in cases 2 - 4, ca. $1 \cdot 10^{12}$ molecules of uniformly labelled probe (same oligonucleotide / same reporter) would be introduced into PCR per parameter;

in cases 1, 5 and 6, again $ca. 1 \times 6^{12}$ molecules of labeled probe would be introduced into PCR per parameter, $ca. \frac{1}{2} \times 6^{12}$ molecules with one sort of tag, and $\frac{1}{2} \times 6^{12}$ molecules with the other sort of tag (same oligonucleotide / different reporters), which would be consumed statistically balanced;

5 in all case, 10 pmol of parameter-specific probe would be available.

Regarding the risk of ambiguous results, and focusing on the use of singular labels and dual combinations only, one may argue that ambiguous results would probably be extremely rare in the case of co-infections of 2 parameters indicated by single labeled probe, respectively, e.g. $P2 \Leftrightarrow R1 + P4 \Leftrightarrow R3$. The probability of either a 1:1 co-
 10 infection in terms of number of starting molecules and identical extraction / amplification / detection efficiencies yielding a 50:50 signal distribution into 2 channels, or of different numbers of starting molecules in conjunction with extraction / amplification / detection efficiencies that precisely counter-balance the differences in titre and again yield a 50:50 distribution, is considered negligibly low. Thus, an uneven
 15 distribution of $>(>)50 : <(<)50 : 0$ or $<(<)50 : >(>)50 : 0$ would result (*case I*), and so be distinguishable from e.g. $P5 \Leftrightarrow R1R3$ indicated by an even distribution of $50 : 0 : 50$.

Co-infection by 2 parameters indicated by 1:1 mixtures of 2 dual labeled probes (*case II*), e.g. $P1 \Leftrightarrow R1R2$ and $P5 \Leftrightarrow R1R3$, should always be evident due to signal distribution into 3 channels (e.g. $ca. 50 : 25 : 25$, comparable extraction / amplification / detection
 20 efficiencies assumed).

Co-infection by 2 parameters, one represented by 1 reporter and the other by two, should be recognized due to signal distribution into 3 channels, $\approx 33 : 33 : 33$ (*case III*, e.g. $R1R2 + R3$), or yield a $\approx 75 : \approx 25 : 0$ pattern (*case IV*, e.g. $R1R2 + R1$), again assuming comparable extraction / amplification / detection efficiencies. For case II and
 25 III, the common feature would that there is no optical channel related to target with essentially "zero signal". In other words, the "zero+x" threshold to discriminate against "true zero" will be important. In *case IV*, probably the most critical one, resolution does depend on the overall efficiency of the assay, which determines whether the resulting distribution is substantially different from a 50 : 50 situation, or not. Most probably,
 30 additional mathematical criteria have to be taken into view in order to resolve ambiguities, especially of the latter kind. Advantageously, real-time Taqman-PCR offers such "helper arithmetics". This topic is dealt with below in some more detail.

The rare occurrences of ambiguity, i.e. the event of a 50 : 50 signal - distribution occurring from co-infection can further minimized by assigning extreme rarely found, and less efficiently amplified parameters to single labeled probes, which might be complemented by considering geographical distributions.

- 5 A putative assignment of reporters (R = Label) and assays for the determination of HIV-1-M, HIV-1-O, HIV-2, HCV, HBV and HAV could be as follows:

	P1	↔	HIV-1-M	↔	R1 / R2	Amplification and	high
	P2	↔	HIV-1-O	↔	R1	detection efficiency:	low
10	P3	↔	HIV-2	↔	R2		low
	P4	↔	HCV	↔	R3		high
	P5	↔	HBV	↔	R1 / R3		high
	P6	↔	HAV	↔	R2 / R3		medium

- 15 IC (internal control) ↔ R4.

For safety reasons and especially with regard to a commercial assay one can also add an internal control, which for example may be detected by a probe coupled to a fourth label (R4).

- 20 Here, co-infection of HIV-1-O and HIV-2 should be very unlikely, both being very rare and geographically well separated (which, however, may change over the years), and the overall result would be "HIV positive" even in case of such an unexpected event.

- On the other hand, co-infections of HIV-1-O and HCV or HIV-2 and HCV should not result in an even distribution of 50 : 0 : 50 or 0 : 50 : 50, given the current amplification efficiencies (low for HIV-1-O and HIV-2), and not be mistaken for HBV or HAV single infections, respectively. Thus it is possible to use a standard four channel fluorimeter for conducting a six-parameter assay including an internal control being separately detected, which is for example used in a Roche Cobas TaqMan-instrument.

- Thus, this concept for complete resolution of a 6 parameter multiplex assay in terms of assay parameter would necessitate moderate additional efforts compared to the present chemical system and instrument platform. Likewise, the requirements in terms of a

more sophisticated analytical software are not very demanding, as shown by the considerations depicted below. In some detail, this will include:

- double logistic tracking for 3 probes (same oligonucleotide, different fluorescent tags)
- 5 - standardized procedure to establish signal outputs for individual reporters from single tagged probes and elaboration of normalization factors (taking into account impact from the sequence environment in the respective oligonucleotides, degree of purification as indicated by e.g. Em(R/Q) ratios, wavelength dependent energy contents and spectral sensitivity of CTM ASICS)
- 10 - elaboration of the most reliable output criterion or set of criteria to serve as 100% count, determination of the variance of distribution patterns, and setting of acceptance range borders for expected values
- development of a software algorithm for analysis of distribution of corrected normalized signals into 3 target channels that incorporates all the parameters
- 15 mentioned above

Example 2

With regard to an assay protocol for automated test devices like the Roche Cobas TaqMan™ instrument a suitable process scheme is given below (see also Figure 1).

- 20 The availability of new reagents given, the principal data collecting and data processing scheme of an automated assay could be as follows:

- data collection:			status:
25	1. dark mean (= instrument noise)	DM	already done
	2. dark drift (= drift correction factor as $f(t)$)	DD	already done
	3. cross-talk and preset correction factors	XT	already done
	4. signal output normalization factors for reporters (= preset factors to correct for quantum yield, micro-environmental effects, spectral sensitivity, ...)	NF _{Ri}	additional
30	5. background intensity (= chemical noise, accumulated from for all probes contained in the mix)	BG	already done
	6. measured total signal over time	TFI	already done
35	Preferably all steps are done separately for the optical channels 1, 2, 3, and 4 on the Cobas Taqman- instrument.		

- data processing:

- 5 1) for each channel, and as a signal-over-time graph,

$$\{TFI_{f(t)} - \{ (DM * DD_{f(t)} + BG) \} * XT * NF_{Ri} = \underline{NFI}$$

is calculated, i.e.
channel-specific normalized corrected fluorescence intensities
- 10 2) if channels 1, 2, 3 correspond to reporter dyes, and channel 4 to IC dye, then

$$NFI_4 = 100\% \text{ IC response intensity}$$
- 15 $NFI_{1+2+3} = 100\% \text{ target response intensity}$,
3) which is to be analyzed in terms of distribution across these channels, i.e. x % in channel 1, y % in channel 2, and z % in channel 3.
- 20 The resulting signal distribution numbers are then compared to the pre-calculated values given above, and the identity of the pathogen detected is deduced from the correlation of optical channel-specific gain and corresponding reporter dyes, plus auxiliary factors.
- 25 Thus, in other words, the assay is not calibrated with respect to target titre (i.e. it is still a qualitative screening assay, yielding a pos/neg result as appropriate e.g. for screening applications in blood banks), but with respect to signal output of parameter-specific probes at a given concentration (to identify the cause for the positive result of a multiplex amplification).
- 30 The individual sample analysis could proceed as indicated in Figure 1.
To comprehend the logic of this processing algorithm, let's consider the underlying assay principles.
In 5'-nuclease technology (Taqman-PCR, the preferred embodiment for the invention set forth here) amplification and detection reactions, respectively, are closely
35 interwoven. To this end, detection probes with 2 particular chemical modification are added to the PCR mastermix. One of these modifications is a fluorogenic reporter group (R, for instance a derivative of 6-carboxy-fluorescein) covalently attached to the backbone of the probe, the other one is a dye (for instance a polymethine-cyanine derivative) capable of absorbing the fluorescent light of the reporter and to quench it
40 (quencher, Q). The quencher is typically attached to the probe backbone at the 5'-end,

whereas the reporter is located within the oligo sequence, spaced from the quencher by a number of nucleotide building blocks. Probes bind to target nucleic acids (sense or anti-sense strand) close to the 3'-end of a primer (reverse or forward). As soon as primer has annealed to the target and DNA-polymerase gets bound to the primer :
5 target hybrid, elongation starts. Due to the 5'-nuclease activity of the enzyme, simultaneously with copy strand synthesis the probe is cleaved as soon as the polymerase reaches the probe binding site, reporter and quencher get separated, and the fluorescent signal becomes measurable. This process is repeated with every cycle, and more and more fluorescent reporter is accumulated in solution until reagent depletion
10 at the end of the reaction. So, in a signal-over-time plot, sigmoidal growth curves are generated. The point on the time axis, where the arbitrary fluorescent intensity (AFI, also called relative light units, RLU) can be significantly distinguished from background signal is called threshold cycle (ct). ct is a measure of analyte titre: the smaller the ct-value, the higher the number of starting molecules and/or the better the extraction or
15 amplification/detection efficiency. ct values can be calculated by means of different mathematical operations. For example, cut-off approaches (average background signal intensity multiplied by a constant factor yields a cut-off signal intensity to discern negative from positive) can be used, or approaches where the location of the maximum of the first or second derivative of the signal-over-time curve (i.e. the steepness or slope
20 profile) is sensed after curve fitting. The slope profile approach is, in contrast to the signal threshold approach, essentially independent of the background intensity level and, thus highly attractive for multiplex assays with cumulated background from all probes present in the reaction mixture.

Based on these principles, and according to the processing scheme depicted above, one
25 would employ different means to generate the primary result (i.e. positive or negative?), and the secondary result (i.e. which kind of infection renders the sample positive?).

Compared to performing a multiplex assay with one common label, there is potential loss in sensitivity due to dilution of signal into several optical channels and, thus, reduced specific signal growth curves in particular channels. Therefore, in order to
30 generate the primary result, one would optionally work with composite signal-over-time curves. By this means, the entire specific signal yield is collected into one signal-over-time curve which then is used to determine the threshold cycle (ct) and the principal status of the sample (pos. / neg.). Using the slope profile approach, the addition of background intensity would not be detrimental, and one could exploit the

potential advantage of sensing the more pronounced growth characteristics of the specific signal gain.

On the other hand, for generation of secondary results, one would refer to the channel-specific distribution of normalized signal (NFI intensities), and deduce the kind of infection by comparing the resulting distribution patterns to the pre-calculated matrix of patterns given above. In addition, and in order to enhance the discriminatory power, also non-normalized *specific* signal intensities may be considered for analysis of ca. 50 : ca. 50 or ca. 75 : ca. 25 distributions of normalized signal, i.e. the really critical ones.

- In case of a *mono-infection* represented by a 1:1 mixture of probes, the resulting ratio of non-normalized, yet background-corrected signals R1 / R2 would be constant throughout the entire amplification/detection process, for instance 40 / 50. This must be precisely equivalent to the value of the normalization factor (0,80) obtained with equal amounts of purified dye, because both sorts of labelled probe (same oligonucleotide sequence!) would be cleaved at identical rates. Background (BG)-correction is necessary, since with a resolving 6-parameter-multiplex assay according to the concept presented here, each channel-specific background is made up from the signal output of 3 different probes tagged with the same label. Therefore, BG intensity is not only a function of the optical features of the label, but also of the structural features of the probe, i.e. linear or hairpin, or (more generally) by the degree of self-complementarity, and of the temperature during measurements which affects conformational equilibria of secondary structures.. So, secondary structures of unhybridized probe must be taken into view with respect to BG intensity, and auxiliary calculations have to be based on *specific* fluorescence intensities, SFI, with

$$SFI = TFI_{f(t)} - [(DM * DD_{f(t)}) + BG] * XT_i = AFI - BG.$$

Specific signal, by contrast, is generated by way of 5'-nuclease activity of the polymerase depending on essentially perfect hybridization of probe to target, i.e. linearization of the probe oligonucleotide and, thus, essentially independent of probe structure. This pertains also to the truncated probe after having been cleaved and dissociated from the target strand due to the concomitant drop in T_m (melting temperature). Short internal hybrids in the residual probe oligomer should not be stable at the elevated temperatures used for annealing / elongation in PCR. After normalization, which compensates for differences in quantum yield or susceptibility to solvent effects, signal distribution in

channel 1 and 2 would be 50 : 50. Finding NFI and SFI ratios as expected is double proof for mono-infection.

- In case of a **double infection** indicated via 2 specific probes labelled R1 and R2 (*case I*), respectively, it is very likely that the signal dynamics (i.e. growth as a function of time) is different for the 2 different assays, as either initial analyte titre, or overall assay efficiency, or both will probably be different. Thus, in addition to the probability of a > 50 : < 50 situation on the NFI level, also the SFI ratio of R1 / R2 must be different for rather early, central, and late phases of the reaction, i.e. the SFI ratio will be variable. The ratios will have a certain trend, and be different from the respective normalization factor obtained with equivalent mixtures of purified dye, preferably coupled to a model probe (to account for inevitable microenvironmental factors, e.g. linker, local charges), e.g. a T₃-R-T₁₆-oligonucleotide, which does not modulate signal output by its structural configuration.
- In case of a **double infection** represented by R1 + R1R2 (*case IV*) and substantially different overall recoveries for the 2 assays, a normalized signal distribution not very different from 50 : 50, e.g. 55 : 45, might be observed in channel 1 and 2. In terms of non-normalized signal intensities, however, the ratio R1 / R2 would be e.g. 44 / 50 (0,88), i.e. greater than the normalization factor. Thus, the double bias of "greater than expected" on both the normalized and non-normalized level, would indicate double infection. What is more, the ratio R1 / R2, e.g. traced e.g. via AUC, would most probably not be constant throughout the entire assay due to differing signal dynamics resulting from different amplification efficiencies, which will enhance discrimination.
- If, in case of double infection, both assays have ca. equal efficiency and start from either substantially different or ca. equal analyte titres, a 75 : 25 : 0 NFI pattern would be approximated, or even exceeded, with either a R1 + R2 or a R1 + R1R2 labelling (*case I OR case IV?*). Here, discrimination between a *case IV* and a strongly *biased case I*-type co-infection may be achieved by means of absolute *specific* signal intensities (R1 intensity should be much lower in a *case I* situation) or absolute differences of channel-specific, optionally rectified, SFI values. Results might not completely unambiguous in some cases, yet.

In addition to this, composite AFI/t curves may be analysed concerning curve shape characteristics. In case of a mono-infection, one does expect a mono-sigmoidal curve

with 1 point of inflection. In case of a double infection, however, it is most likely that the 2 growth curves generated from each of the 2 kind of nucleic acid targets extracted from the 2 infectious agents are not identical in terms of ct (i.e. point on the time axis where the curve bends upward and rises above background and drift level), in terms of shape, and specific signal intensity. Thus, a bi-sigmoidal curve shape may be generated, with 2 points of inflection. So, mono- or bi-sigmoidal curve shapes might be indicative of mono- or bi-infections present in the sample under inspection. In unfavourable cases, however, superposition of 2 or more curves may lead to irregular shapes with no clearcut maximum of slope, and with noise (i.e. the roughness of the curve) being increased. Still another possibility is a composite curve with 1 point of inflection, when 2 nearly identical curves are superposed.

To sum up, there are a number of mathematical means available to achieve resolution of a positive multiplex assay result in terms of the type in infection. These include, but are not limited to:

- 15 Basic means – reference NFI distribution patterns, and
 - type of deviation, either in situations with significant signal intensity in 2, or in 3, optical channels related to target. The accepted range of deviations from the reference pattern can be set by absolute numbers, % bands relative to the respective NFI value, based on precision data, or a combination of these.
- 20 Auxiliary means –
 - analysis of NFI and SFI ratios pertinent to particular optical channels for common bias;
 - analysis of NFI or SFI ratios as a function of cycle number , i.e. dynamically along the reaction coordinate (constant or variable);
 - 25 relationship of NFI ratios and corresponding absolute SFI values, or absolute SFI differences between any 2 optical channels (optionally rectified by relating to a standard dose difference), taking into account known assay-specific overall efficiencies (extraction / amplification / detection) for the different infectious agents under inspection;
 - 30 analysis of composite AFI-over-time curve shapes (if applicable)

A putative application mode for this set of criteria is illustrated in Fig. 1 + 2.

What is more, in order to monitor the whole process, an artificial nucleic acid construct may be added to all samples, preferably packaged (armoured) in a modified virus particle, which is co-extracted and co-amplified with the natural target. This internal control (IC) features a unique probe binding region for an IC detection probe, which
5 differs from target-specific probes by a different reporter group with distinguishable emission characteristics. Thus, IC signal can be discerned from target signal, and as the IC is known to be present in the sample, it functions as a monitoring agent. If there is no IC response, the respective reaction has to be considered invalid, and repeated.

10

Additional to the co-extracted/-amplified/-detected internal control (IC), added to each and every specimen to be processed (i.e. both unknowns and external controls or calibrators) and indicated by R4, external positive controls will be part of the assay. These are „knowns“ which must be found positive, as opposed to „unknowns“, and may
15 be designed as mixed control samples that give rise to signal in all target related optical channels. Combined with the IC, utility of the entire optical system would be monitored then in addition to the functionality of the complete extraction/amplification/detection process. For nucleic acid-based screening applications in blood banks, for instance, a highly sensitive, yet qualitative assay is required to identify each and every positive
20 donation, as far as technically feasible. Therefore, titres of external positive controls would be set in the lower range of concentration

Example 3

Based on the parameter list given in example 1 the signal distribution shown below is
25 analyzed. For each of the 6 parameters under inspection, there is a reference pattern of signal distribution across the 3 optical channels assigned to target detection. Each reference pattern is complemented by a concomitant accepted range, for instance

30

	Ch1	Ch2	Ch3
P1	50	50	0
P2	100	0	0
P3	0	100	0
P4	0	0	100
P5	50	0	50
P6	0	50	50

NFI reference pattern

	Ch1	Ch2	Ch3
P1	10	10	9,17
P2	10	1,67	6,25
P3	7,08	10	7,92
P4	4,17	1,67	10
P5	10	4,17	10
P6	5,83	10	10

Putative 10% / 20% off-set bands pertinent to 100% and 50% mean values, and 5 X multiples of near zero values, which translates into the following accepted ranges for the reference pattern

	Ch1	Ch2	Ch3
P1	40-60	40-60	0-9
P2	90-100	0-2	0-6
P3	0-7	90-100	0-8
P4	0-4	0-2	90-100
P5	40-60	0-8	40-60
P6	0-8	40-60	40-60

- 5 Accepted ranges of deviation may be calculated, for example, as percent values (%), absolute numbers, multiples of simple standard deviation, or combinations thereof.

For each sample (parameter) the signal distribution in Channel 1 (reporter R1), Channel 2 (reporter R2) and Channel 3 (reporter R3) is given by 6 parallel simulated data sets based on 6-fold determinations each, as shown on the table (right side). On the
 10 left side the subsequent steps for evaluating the signal data are shown.

Simulated distribution				Simulated data-set:				
	Ch1	Ch2	Ch3		Ch1	Ch2	Ch3	Total
P1	49,33	48,83	1,83	P1	52	48	0	100
P2	98,42	0,33	1,25		49	48	3	100
P3	1,42	97,00	1,58		47	52	1	100
P4	0,83	0,33	98,83		50	46	4	100
P5	48,17	1,67	50,17		48	49	3	100
P6	1,17	50,33	48,50		50	50	0	100
mean values (arithmetic)				P2	98	0	2	100
According to reference pattern, <u>deviated</u>					97,5	0	2,5	100
					99	0	1	100
					100	0	0	100
					96	2	2	100
P1	5,25	6,12	5,17		100	0	0	100
P2	4,69	2,45	3,25		96	2	2	100
P3	2,75	5,02	3,34	P3	100	0	0	100
P4	3,51	1,55	4,81		1	96	3	100
P5	7,92	6,75	7,92		2	96	2	100
P6	3,51	6,48	7,29		2	97	1	100
threefold standard deviation					1	99	0	100
according to reference pattern, <u>deviated</u>					0	99	0	100
					2,5	95	2,5	100
				P4	0	0	100	100
					0	0	100	100
					3	1	96	100
P1	50	50	0		1	0	99	100
P2	100	0	0		0	0	100	100
P3	0	100	0		1	1	98	100
P4	0	0	100	P5	46	6	48	100
P5	50	0	50		50	1	49	100
P6	0	50	50		45	0	55	100
NFI reference pattern					52	0	48	100
					49	1	50	100
					47	2	51	100
				P6	0	50	50	100
P1	10,00	10,00	9,17		1	49	50	100
P2	10,00	1,67	6,25		1	53	46	100
P3	7,08	10,00	7,92		2	47	51	100
P4	4,17	1,67	10,00		3	52	45	100
P5	10,00	8,33	10,00		0	51	49	100
P6	7,50	10,00	10,00					

10% / 20%-off-set bands pertinent to 100% and 50%-mean values, and factor 5 pertinent to 0% values i.e. accepted ranges for reference pattern

With regard to P5 a distribution of Ch1:Ch2:Ch3 of approximately 50:0:50 is found, with both mean values and scatter being within the accepted ranges. This would lead to the result "P5-positiv", which means monoinfection with HBV.

5 *Example 4*

As outlined above in the course of the principal considerations given in Examples 1 and 2, 50 : 50 distributions may not always be a valid indicator of mono-infections. Although it has to be mentioned that there are several applications where such ambiguous results do not occur, in many other cases especially for diagnostic applications a further analysis often is desired. In order to allow a further analysis for example additional curve characteristics obtained in homogeneous amplification methods can be used. This is reflected in the scheme shown in Fig. 5, where additional curve characteristics like absolute SFI levels of the Ry dye and NFI trace features (NFI ratios constant / variable; early / late approximation of final distribution; bias <1 / > 1; etc.) are employed to resolve the infecting parameter(s).

It is evident from these examples that resolution should be significantly enhanced by analysing the real-time assay data dynamically along the reaction coordinate, instead of considering end-point conditions only. Especially for the samples S4 – S6, pure end-point analysis only would engender the risk of diagnosing mono-infections, while the comprehensive set of analytical tools allows for more precise assignments. In case of S5, for instance, the end-point NFI distribution mimicks HBV mono-infection as in Example 3, whereas by way of analysis of NFI distributions dynamically along the reaction coordinate, the false assignment „HBV mono-infection“ can be rejected easily, and replaced by the correct assignment „Co-Infection, HIV-1-O + HCV“, as HCV being amplified efficiently and represented by R3, gives rise to specific fluorescence intensity (SFI) much earlier than HIV-1-O, being amplified less efficiently, and represented by R1. Thus, the R1/R3 graph falls below the 50 : 50 baseline quickly, continues with positive bias, as R1 is produced with some delay, approximating the final distribution.

30

*Example 5*Experimental data with HBV mono-infections

Genomic target (HBV subtype A), varying in dose from 25 – 5000 copies /ml, and
5 internal control constructs were extracted from plasma samples using magnetic glass
particle technology as described in PCT/EP00/11459. Eluted nucleic acids were
specifically amplified and detected in homogeneous real-time Taqman-PCR on a
COBAS TaqMan™ instrument operating under AmpliLink® version 2.1 (Roche
Molecular Systems). For the basic studies reported here, HBV specific probe (code
10 JW144) was applied @ 7,5 pmol/100 µl labelled with FAM ($\lambda_{EX} \approx 494$ nm; $\lambda_{EM} \approx 518$
nm), and @ 7,5 pmol/100 µl labelled with JA274 ($\lambda_{EX} \approx 580$ nm; $\lambda_{EM} \approx 609$ nm).
Results were exported to MS Excel® Version 7.0 for further data analysis according to
the present invention (e.g. calculation of valid cycle range [SFI >> AFI-BG noise], of
specific signal gain, normalised signal intensities, NFI ratios and NFI traces along the
15 reaction coordinate, classification of NFI trace graphs into constant / offset / variable
with positive slope / variable with negative slope, and for graphical representation of the
results as well).

Homogeneous multiplex PCR reactions were performed under standard PCR conditions
20 using primer pairs and probes specific for HIV-1M, HIV-1O, HIV2, HCV and HBV.

As shown in Fig. 3, a-d, FAM and JA274 signals generated from one and the same
biochemical reaction (i.e. extraction and amplification of HBV-DNA) behave in a
coupled manner according to the concept of the present invention, i.e. swing in to the
25 50:50 distribution line as soon as SFI levels rise significantly above background. The
higher the HBV dose, the earlier this is achieved, and remains constant until the end-
point of the reaction. This is most noteworthy, since signal output as measured on the
COBAS TaqMan™ is very different for FAM and JA274, resulting in a normalisation
factor $NF_{(JA274/FAM)}$ of ca. 11,5 : 1. This in turn translates into enhanced analytical
30 sensitivity of reactions indicated by JA274 compared with similar reaction indicated by
FAM, i.e. at the lower end of detectable HBV doses, attaining to a 50 : 50 NFI
distribution is critical using this dye pair („worst case model“)! Yet, upon translation of

SFI into NFI using the experimentally established $NF_{(JA274/FAM)}$, these dyes – coupled by way of being generated from a common biochemical reaction - yield 50:50 distributions with remarkable precision (see also Fig. 4, b).

5 *Example 6*

Experimental data with different co-infections (type I & type IV)

Procedures and reagent sets were as described for Example 5, except that here, dependent on the targets extracted from co-infected samples, the probes specific for
 10 HBV, HCV, or HIV-1-O were consumed in the amplification / detection reaction. HBV probe was employed as a mixture of equal amounts of FAM- or JA274 labelled oligonucleotides, HCV probe was FAM labelled in total, and HIV-1-O probe was JA274 labelled in total. HBV + HCV positive plasma samples and HBV + HIV-1-O positive plasma samples are type IV co-infections, HCV + HIV-1-O positive is a type I co-
 15 infection. Note: for HIV-1-O, titres are given in X-fold dilutions of a culture supernatant; for all other parameters, precisely quantitated [cp/ml] and standardised material was available.

As can be seen from the plots in Fig. 4, a-d, NFI distribution *per se*, and NFI traces
 20 along the reaction coordinate are very different from those resulting from mono-infections. For instance,, in Fig. 4, a, the reactions # 1-12 feature low and approximately equivalent titres of HBV (amplifying very well; no reverse transcription step involved) and HCV (amplifying well, yet somewhat slower than HBV, via reverse transcription-PCR; HCV targets being not easy to access due to pronounced secondary structures).
 25 Consequently, soon after the on-set of the specific reaction $SFI_{(FAM)}$ and $SFI_{(JA274)}$ are produced via HBV reaction in considerable amounts, and $NFI_{(JA274/FAM)}$ ratios tend to approach the 50.50 distribution line (=1). As cycling progresses, however, HCV reacts with little delay, and additional $SFI_{(FAM)}$ is generated, pushing the $NFI_{(JA274/FAM)}$ ratio below the equity line with the graph having a negative slope. Reactions # 13-24 in Fig.
 30 4, a feature low dose HBV infection plus a 100-fold excess of HCV dose, resulting in a shift of the valid cycle range towards earlier kick-off, an $NFI_{(JA274/FAM)}$ ratio graph that rushes down far below the equity line and then rises slowly with a positive slope as

SFI_(FAM) is generated from the high dose HCV reaction in advance of the „coupled“ generation of SFI_(FAM) and SFI_(JA274) from the low dose HBV reaction, which gradually increases the ratio of NFI_(JA274/FAM).

- 5 Thus, applying the criteria set forth in Example 2 (p. 25-26) in an ordered sequential manner as depicted in Fig. 1+2, the analysis would yield „Co-infection HBV + HCV“. This holds true even for most reactions with a combination of high dose HBV and low dose HCV (high levels of SFI_(FAM) and SFI_(JA274) plus little additional SFI_(FAM) alone, see Fig. 4, d) or from of high dose HBV and low dose HIV-1-O (high levels of SFI_(FAM) and
- 10 SFI_(JA274) plus very little additional SFI_(JA274) alone, see Fig. 4, c; reactions # 7-24). In a few cases, especially with the latter setting, both the shapes of NFI curves inspected visually and preliminary mathematical descriptors (NFI end-point ratios, NFI trace characteristics, absolute SFI levels, etc.) are quite close to those indicating mono-
- 15 similar emission intensities, i.e. NFI ratios more stable independent of target dose and concomitant level of absolute signal generation), mathematical (more sophisticated statistical tools applied to the analysis of primary growth and secondary NFI curves, respectively), as well as technical means (TC block and operations software configuration designed specifically for rapid cycling as with the TC profile used here; this helps
- 20 to smooth the primary growth curves which in turn stabilises NFI ratio trace graphs).

- What is more, Fig. 4, b (reactions # 13-24) and Fig. 4, c (reactions # 1-6) show examples of type-I co-infection, i.e. HCV plus HIV-1-O in this case, which results in a „non-coupled“ generation of SFI_(FAM) and SFI_(JA274). Consequently, the NFI ratio trace
- 25 graph runs, and remains, far below the equity line since HCV is amplified much better than HIV-1-O (a consequence, inter alia, of polymorphisms affecting the primer / probe binding sites). Moreover, also JA274 specific signal intensity generated from HIV-1-O alone are considerably lower than those generated from HBV mono-infections, which again is according to theory. This helps to further discriminate co-infection from mono-
- 30 infection.

Generally speaking, type IV co-infections are more difficult to discriminate against mono-infections, especially when high doses of parameter 5 (generating high SFI levels for both R1 [Ry] and R3 [Rx]) is combined with very low doses of parameter 2 (generating only little additional R1), but in most cases, this is possible even using
5 rather simple mathematical criteria as in this preliminary test system, and a „worst case model“ dye pair. Combining 2 well amplifying targets in a type IV co-infection, e.g. parameter 5 [indicated by R1 + R3] and parameter 4 [indicated by R3] poses much less ambiguity, as substantial amounts of R3 intensity is added asynchronously and asymmetrically. Yet, if the relative abundance of parameter 4 versus parameter 5
10 becomes negligible, similar limits of resolution are observed. In contrast to type IV co-infection involving parameter 2 (and as expected), type I co-infections are discriminated quite easily against mono-infections indicated by the same, yet biochemically coupled dye pair. This becomes evident on the level of end-point NFI ratios and the NFI traces along the reaction coordinate as well, and also the significantly lower signal intensities
15 of the R1 [Ry] dye. This is due to the fact that substantial R3 intensity is generated by parameter 4 even at low doses, while R1 intensity produced from cleavage of parameter 2-specific probes is principally low and generated rather late during the course of 60 cycles. So, due to almost inevitable differences in either initial titre, or extraction / elution / amplification efficiency, it is most unlikely to yield a 50:50 distribution of
20 signal throughout the reaction.

Claims

- 1) A method for the determination of at least 3 analytes comprising the steps:
- 5 a) Providing a mixture of a sample containing said at least 3 analytes or suspected of containing one or more of said analytes and at least 3 different binding partners under conditions allowing the specific binding of said binding partners to said analytes, wherein
- a first of said at least 3 binding partners is specific for a first of said at least 3 analytes and is coupled to a first label,
- 10 - a second of said at least 3 binding partners is specific for a second of said at least 3 analytes and is coupled to a second label, which label is separately detectable from the label coupled to said first binding partner, and
- a third of said at least 3 binding partners specific for a third of said at least 3 analytes whereby a first amount of said third binding partner is coupled to
- 15 the same label as coupled to said first binding partner and a second amount of said third binding partner is coupled to the same label as the second binding partner and
- said first amount of said third binding partner is not coupled to the same label as coupled to the second binding partner and said second amount of
- 20 said third binding partner is not coupled to the same label as coupled to the first binding partner,
- b) Detecting the signal intensities indicative for said first and second label,
- c) Determining the analytes present in said sample using said signal intensities detected in step b),
- 25 characterized in that said analytes are amplified nucleic acids.
- 2) Method of claim 1 characterized in that the different binding partners are nucleic acid sequence specific probes.
- 3) Method of claim 1 or 2 characterized in that said nucleic acids were amplified using PCR.
- 30 4) Method of any of claims 1 to 3 characterized in that said labels were detected homogeneously.

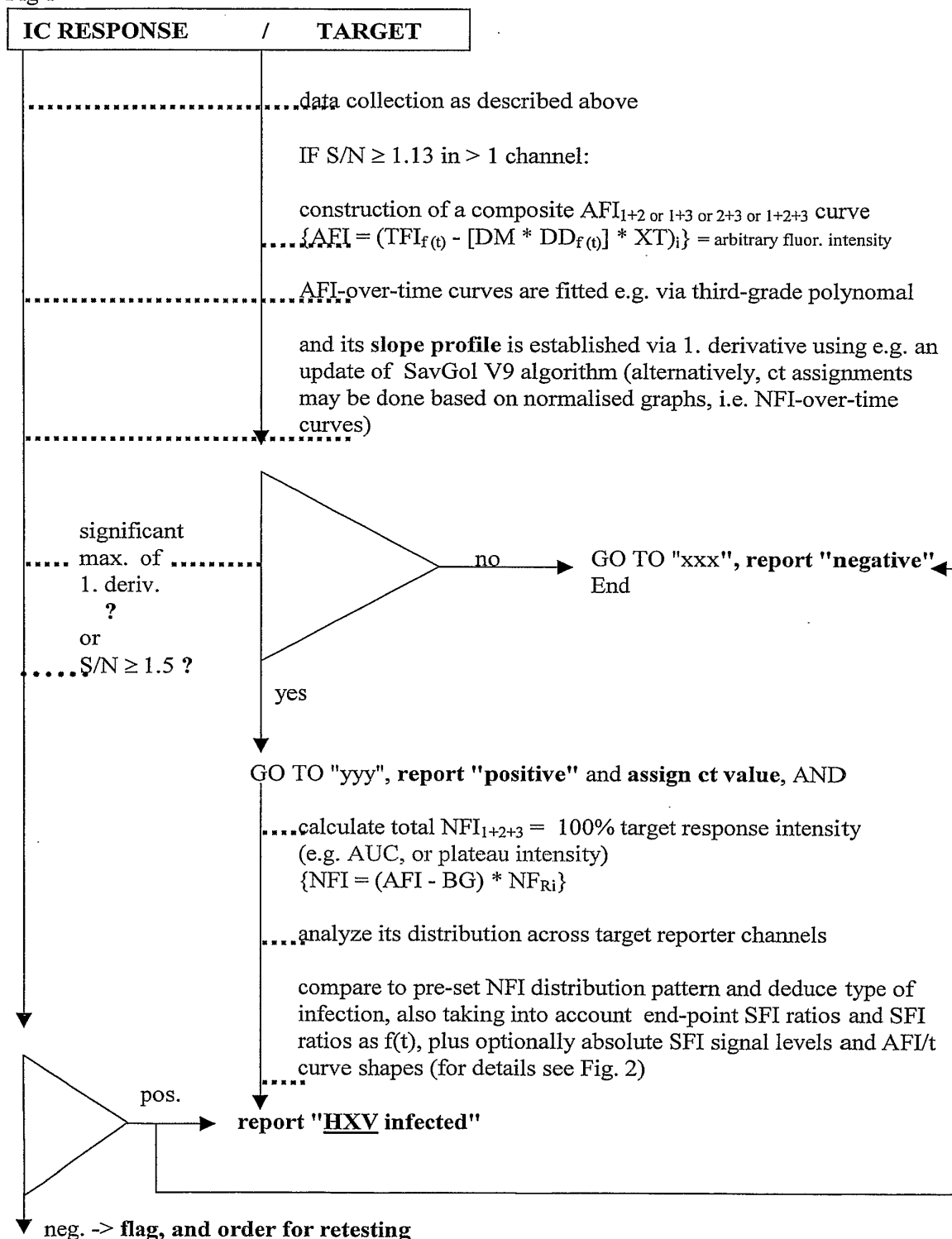
- 5) Method of claim 4 characterized in that said labels were detected using fluorescence energy transfer.
- 6) Method of any of claims 1 to 5 characterized in that a first of said first, second and third binding partners is specific for HIV, a second binding partner is specific for HCV and a third binding partner is specific for HBV.
- 7) Method of any of claims 1 to 5 characterized in that a first binding partners is specific for HIV-1-M, a second binding partner is specific for HIV-1-O, a third binding partner is specific for HIV-2, a fourth binding partner is specific for HCV, a fifth binding partner is specific for HBV and a sixth binding partner is specific for HAV.
- 8) A composition of matter comprising
- a first binding partner specific for a first analyte coupled to a first label,
 - a second binding partner specific for a second analyte coupled to a second label, which label is separately detectable from the label coupled to said first binding partner, and
 - a third binding partner specific for a third analyte whereby a first amount of said third binding partner is coupled to the same label as coupled to said first binding partner and a second amount of said third binding partner is coupled to the same label as the second binding partner and said first amount of said third binding partner is not coupled to the same label as coupled to the second binding partner and said second amount of said third binding partner is not coupled to the same label as coupled to the first binding partner,
- characterized in that said analytes are amplified nucleic acids.
- 9) A kit for the determination of at least 3 analytes containing in one or more containers
- a first binding partner specific for a first analyte coupled to a first label,
 - a second binding partner specific for a second analyte coupled to a second label, which label is separately detectable from the label coupled to said first binding partner, and
 - a third binding partner specific for a third analyte, whereby a first amount of said third binding partner is coupled to the same label as coupled to said first

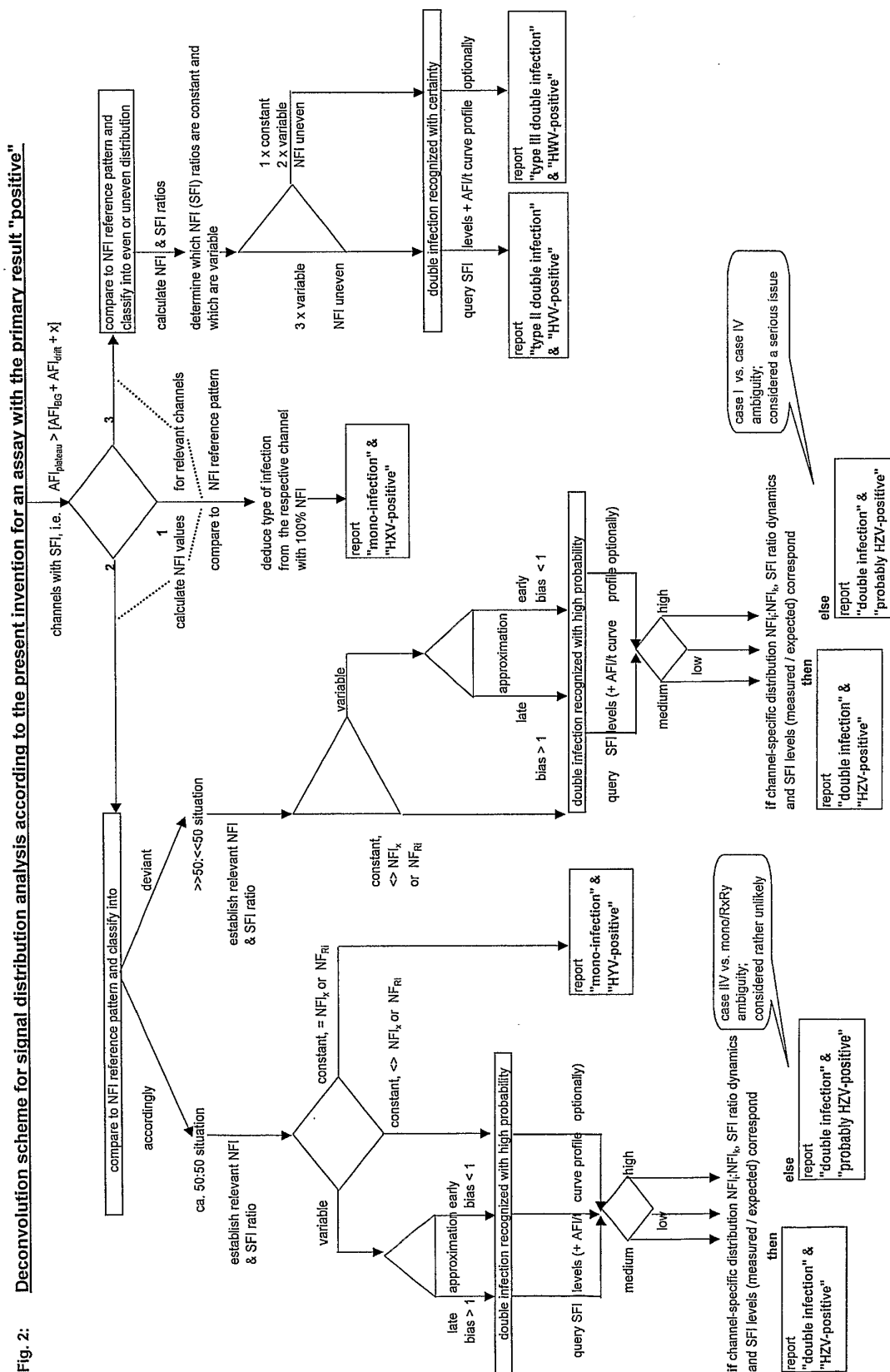
binding partner and a second amount of said third binding partner coupled to the same label as said second binding partner,

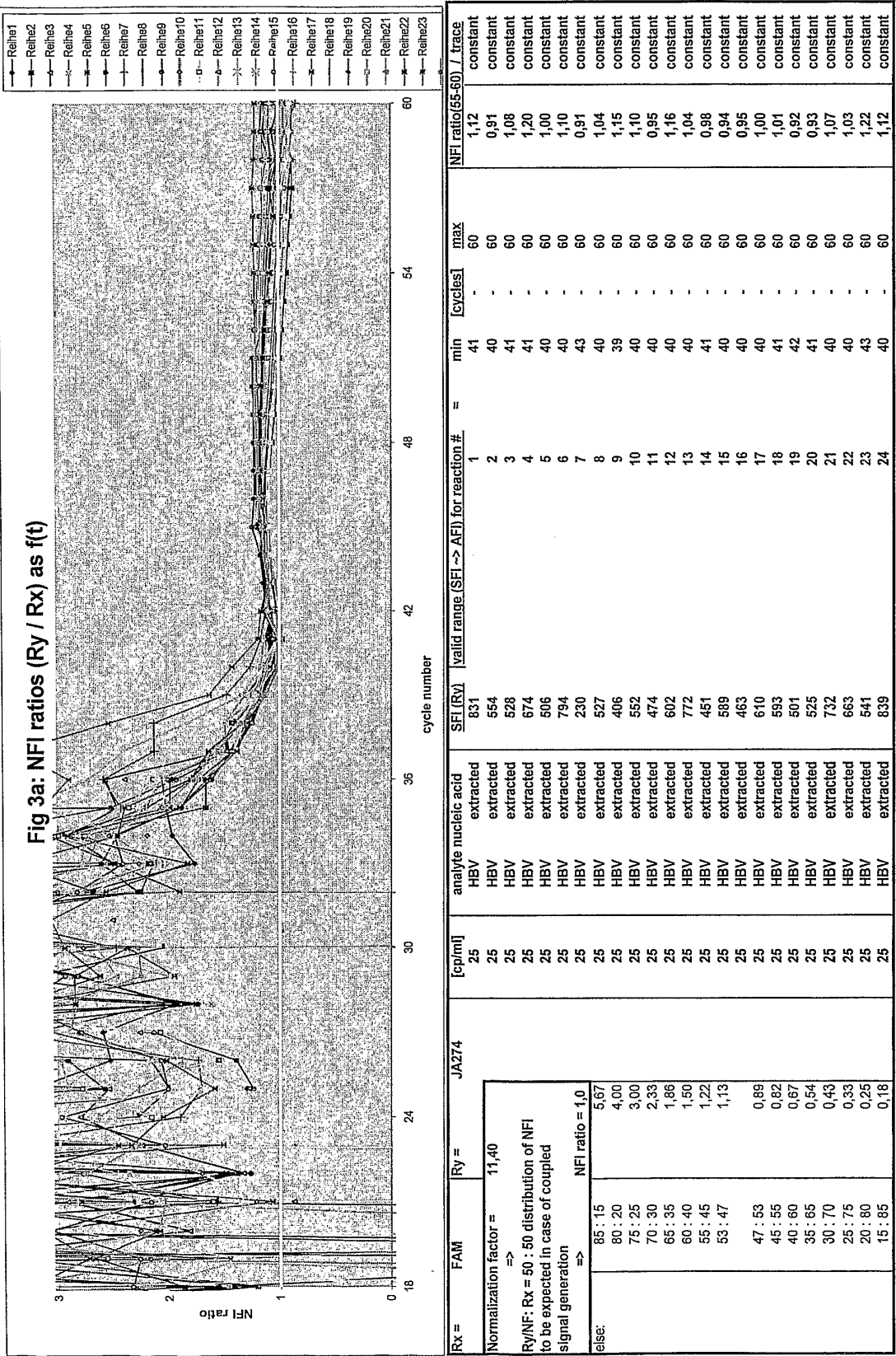
characterized in that said analytes are amplified nucleic acids.

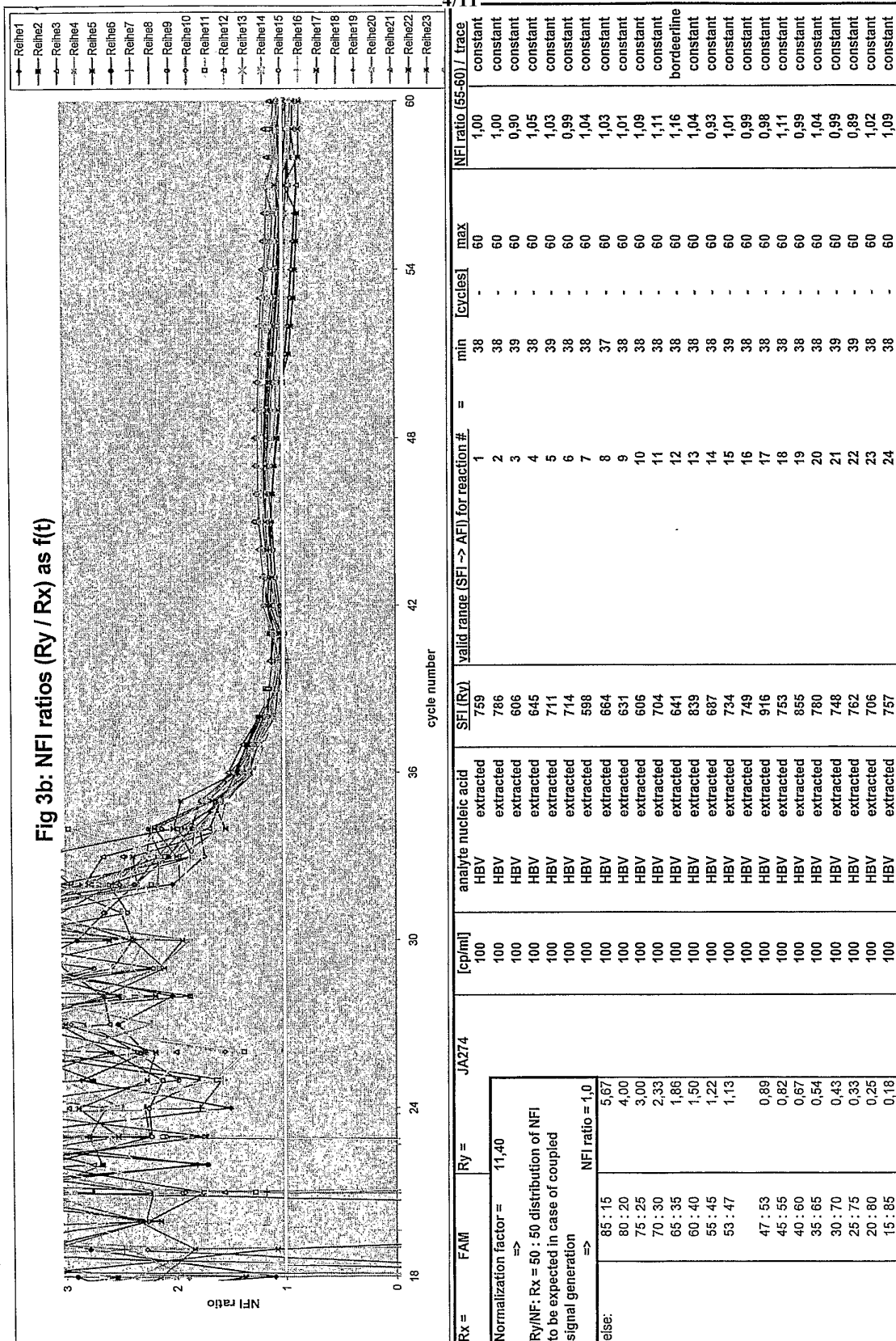
- 5 10) Kit of claim 9 further containing means for specific binding of said binding partners to said analytes in said one container or in one or more further containers.
- 11) Kit of claim 9 or 10 further containing reagents for nucleic acid amplification in said one container or in one or more further containers.
- 10 12) Use of at least 3 binding partners for the determination of at least 3 analytes in a sample, wherein
- a first binding partner specific for a first analyte is coupled to a first label,
 - a second binding partner specific for a second analyte is coupled to a second label, which label is separately detectable from the label coupled to said first binding partner, and
 - 15 - a third binding partner specific for a third analyte, whereby a first amount of said third binding partner is coupled to the same label as coupled to said first binding partner and a second amount of said third binding partner is coupled to the same label as said second binding partner and
 - 20 said first amount of said third binding partner is not coupled to the same label as coupled to the second binding partner and said second amount of said third binding partner is not coupled to the same label as coupled to the first binding partner,
- characterized in that said analytes are amplified nucleic acids.

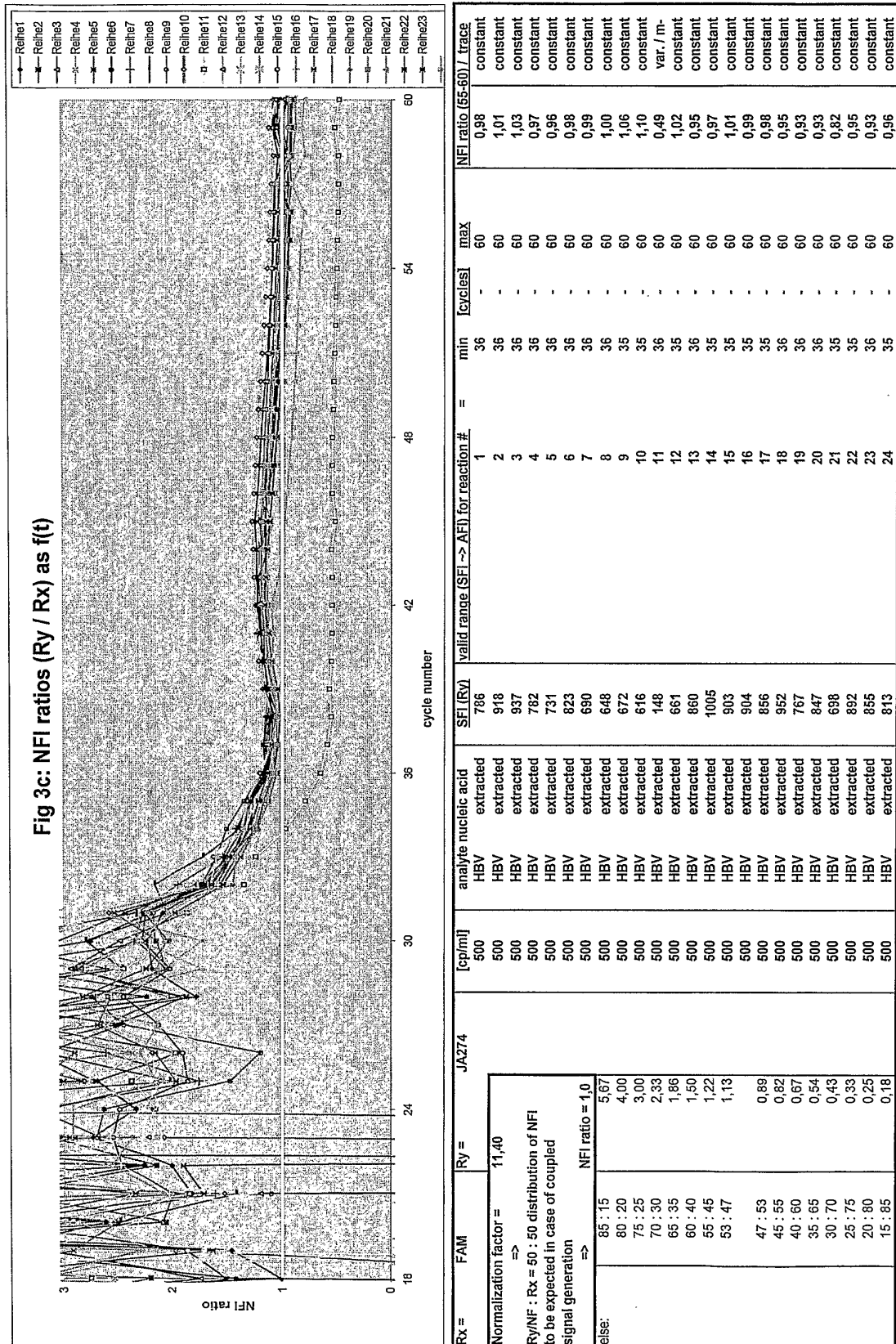
Fig 1

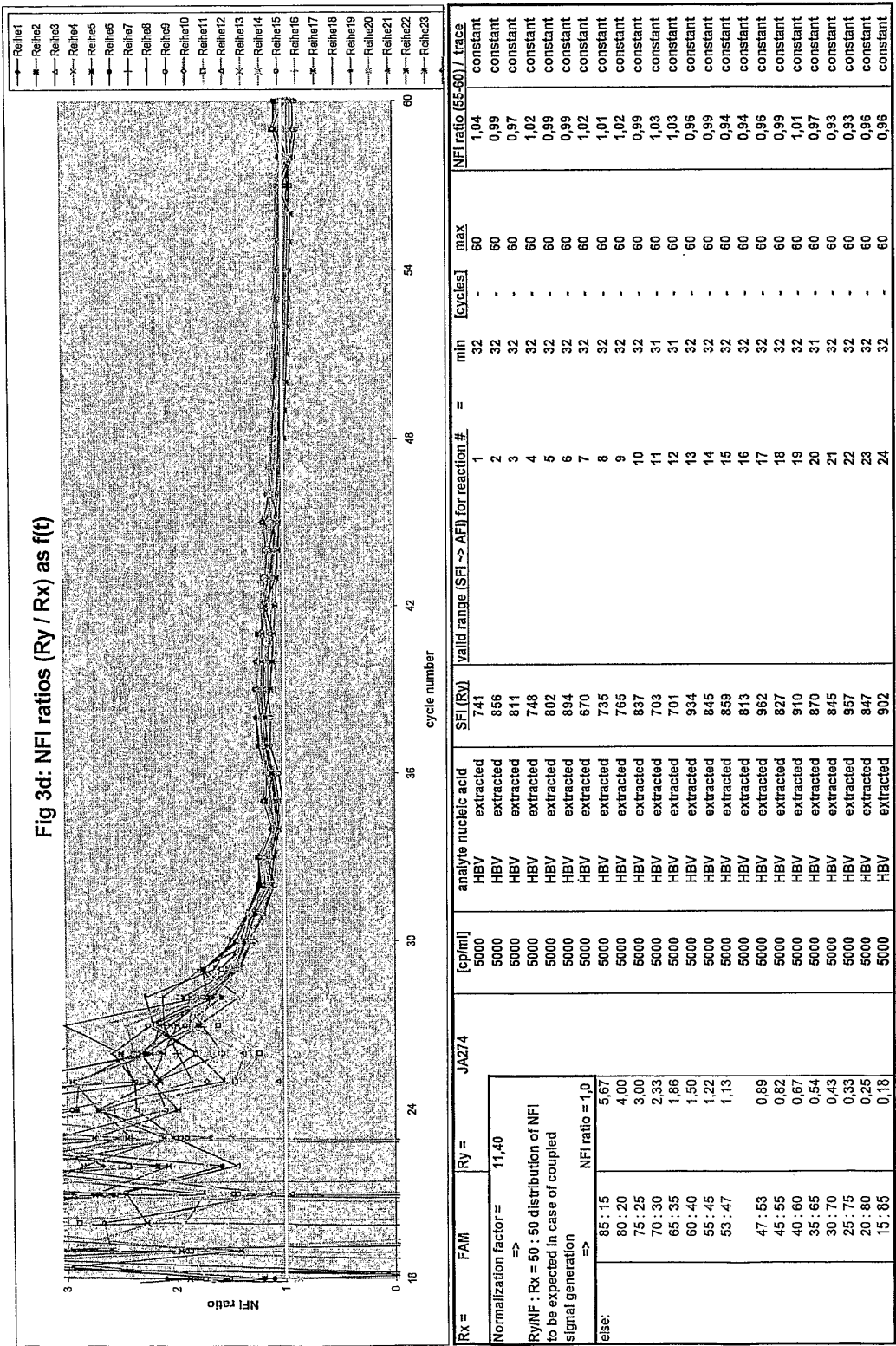


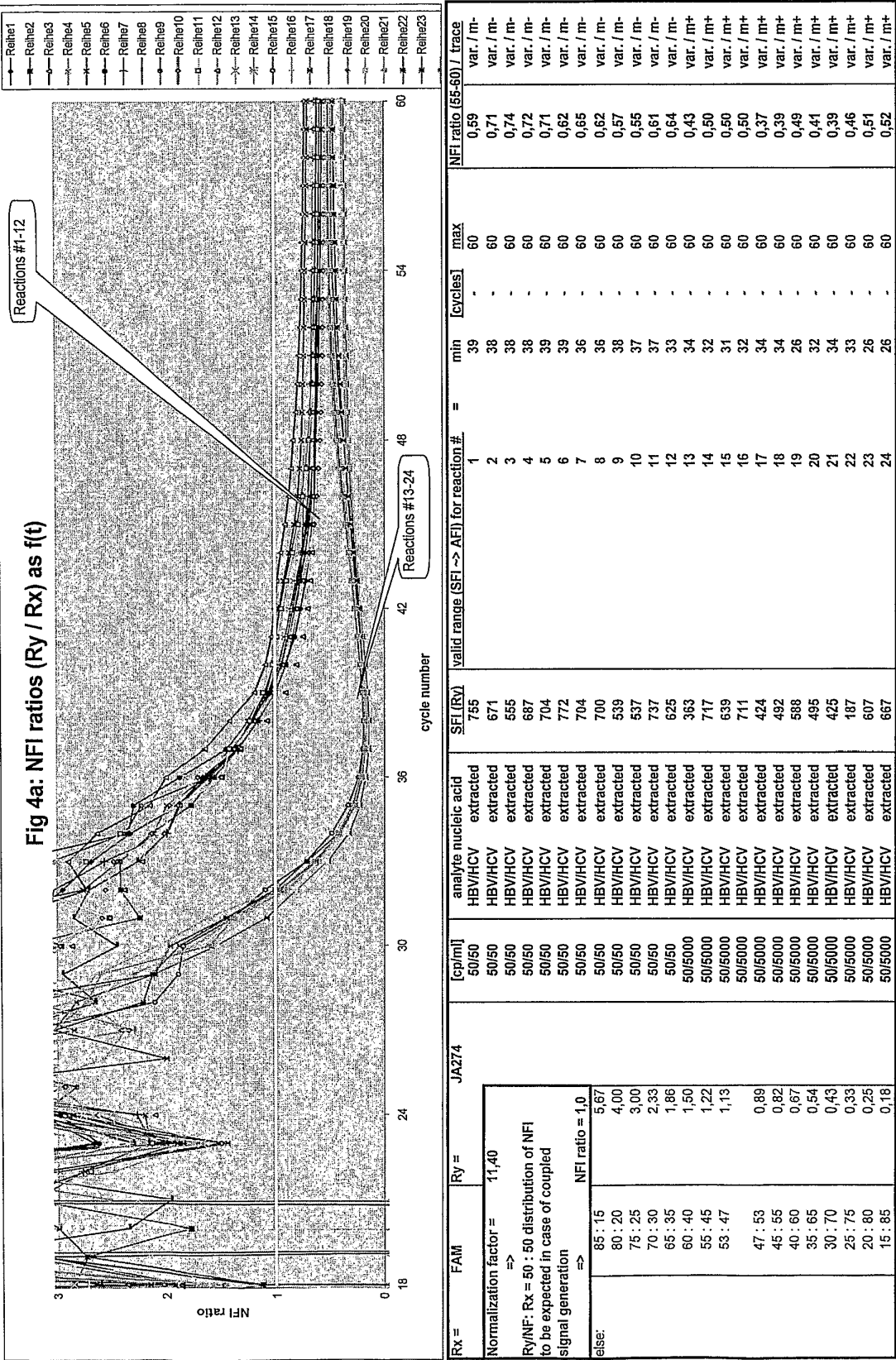


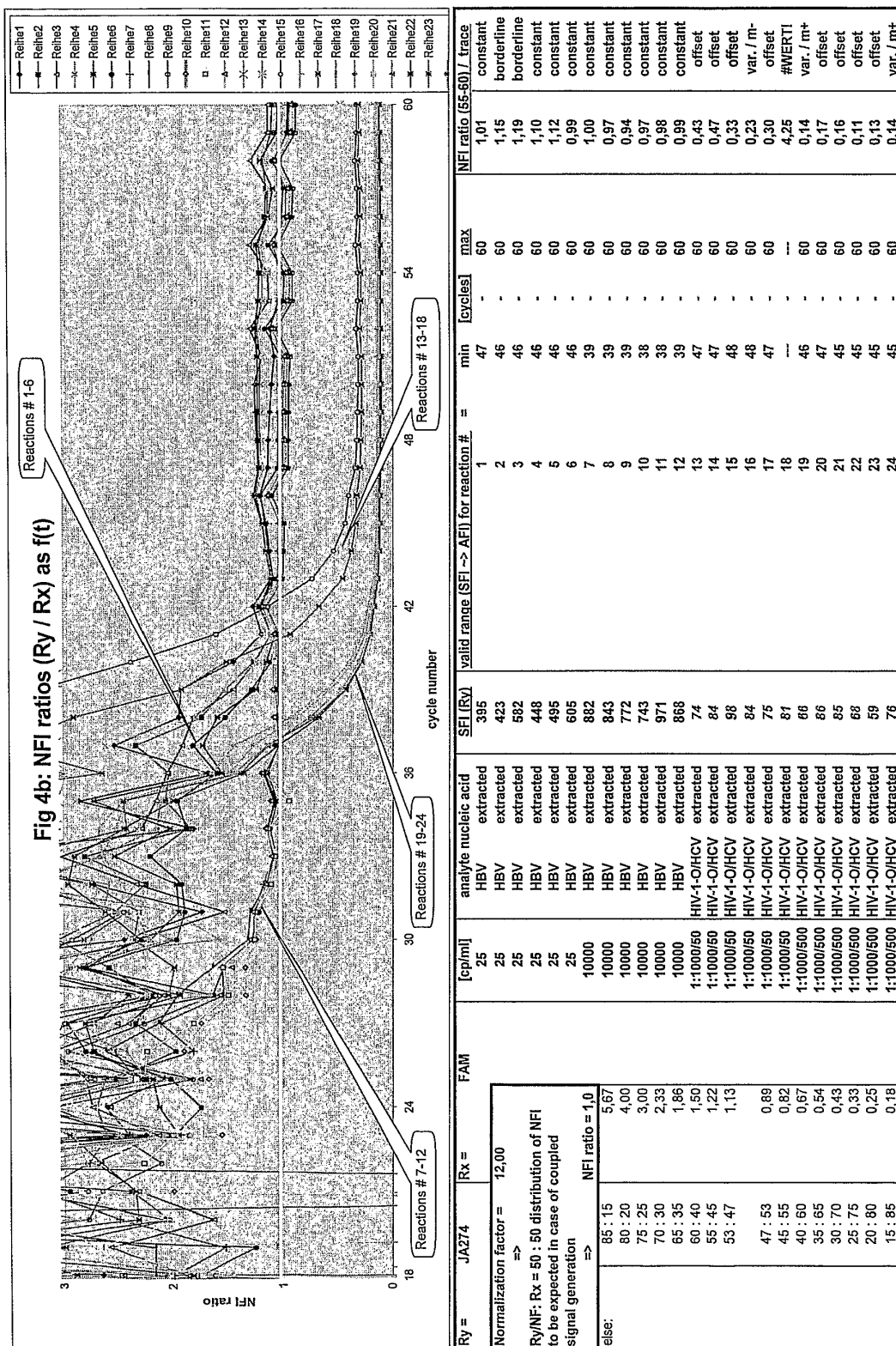


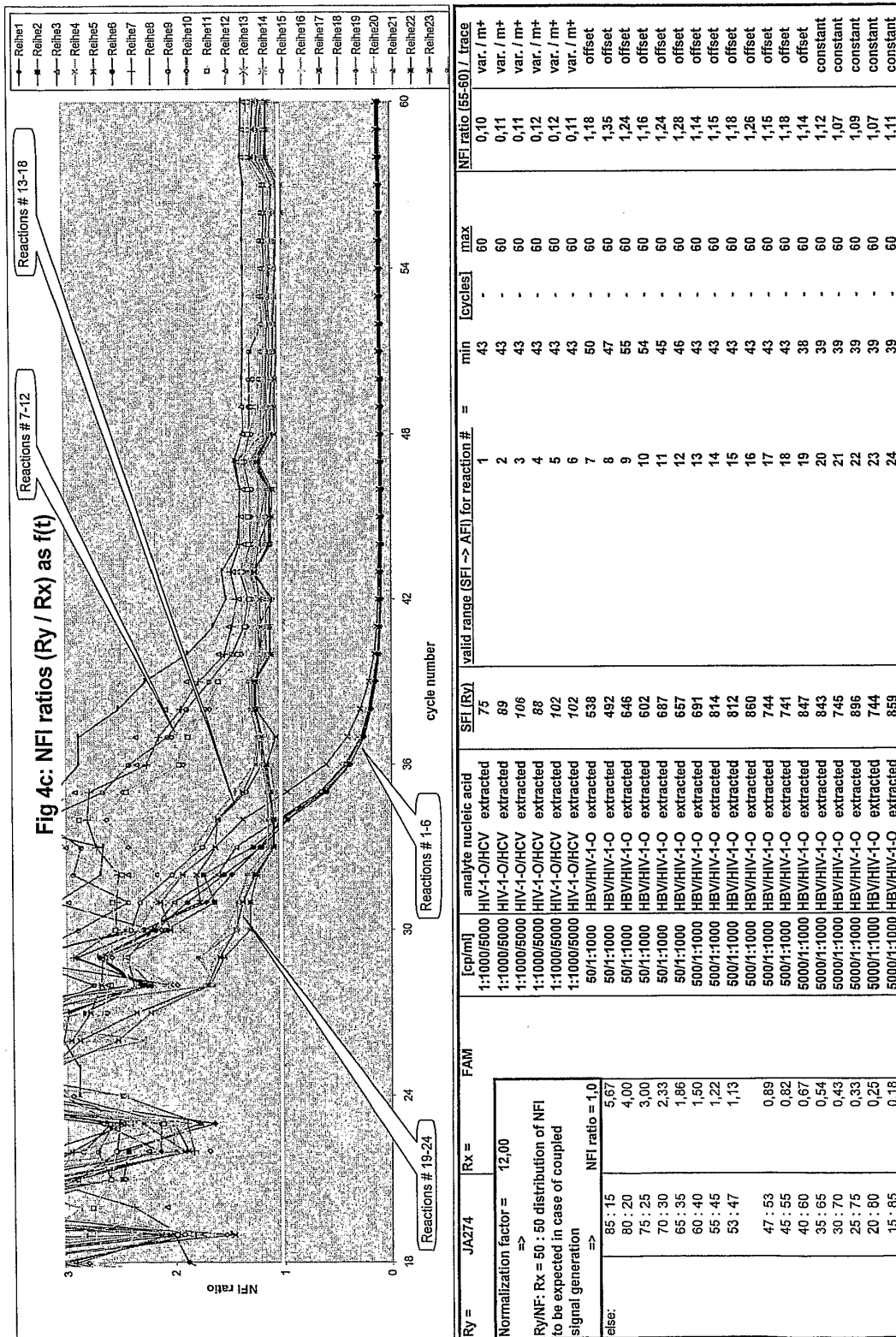












10/11

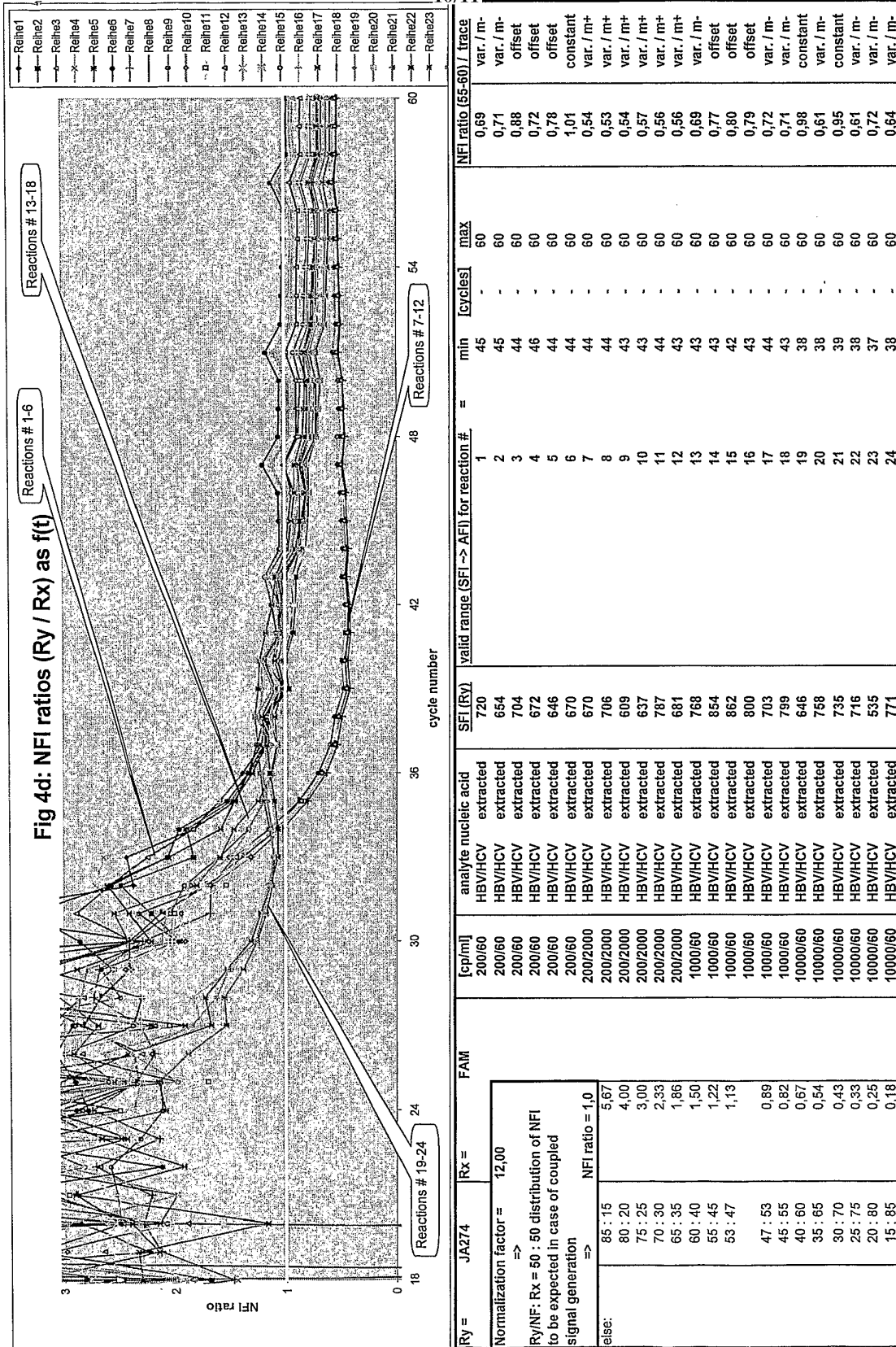


Fig. 5:

Fig. 5: we consider a panel of potentially ambiguous samples

model data set 3:

NFI values

CH1	CH2	CH3	total
31	30	38	99
34	33	33	100
32	35	33	100
33	32	35	100
32	30	38	100
30	37	100	100
3	15	82	100
2	20	78	100
1	24	75	100
0	27	73	100
1	20	79	100
3	21	76	100
24	49	27	100
21	49	30	100
28	46	26	100
27	44	29	100
25	48	27	100
22	53	25	100
52	48	0	100
50	49	1	100
47	52	1	100
50	48	2	100
49	50	1	100
50	50	0	100
46	1	52	99
48	2	50	100
45	0	55	100
49	0	51	100
46	1	53	100
47	2	51	100
40	0	60	100
41	1	58	100
43	1	56	100
37	2	61	100
42	3	55	100
38	0	62	100

NFI distribution; NFI ratios

CH1	CH2	CH3
31	32.90	35.67
32	33.00	37.17
33	34.30	37.33
34	35.57	38.50
35	36.83	39.68
36	38.10	40.85
37	39.37	42.02
38	40.64	43.19
39	41.91	44.36
40	43.18	45.53
41	44.45	46.70
42	45.72	47.87
43	46.99	49.04
44	48.26	50.21
45	49.53	51.38
46	50.80	52.55
47	52.07	53.72
48	53.34	54.89
49	54.61	56.06
50	55.88	57.23
51	57.15	58.40
52	58.42	59.57
53	59.69	60.74
54	60.96	61.91
55	62.23	63.08
56	63.50	64.25
57	64.77	65.42
58	66.04	66.59
59	67.31	67.76
60	68.58	68.93
61	69.85	70.10
62	71.12	71.27
63	72.39	72.44
64	73.66	73.61
65	74.93	74.78
66	76.20	75.95
67	77.47	77.12
68	78.74	78.29
69	79.01	79.46
70	80.28	80.63
71	81.55	81.80
72	82.82	82.97
73	84.09	84.14
74	85.36	85.31
75	86.63	86.48
76	87.90	87.65
77	89.17	88.82
78	90.44	89.99
79	91.71	91.16
80	92.98	92.33
81	94.25	93.50
82	95.52	94.67
83	96.79	95.84
84	98.06	97.01
85	99.33	98.18
86	100.60	99.35
87	101.87	100.52
88	103.14	101.69
89	104.41	102.86
90	105.68	104.03
91	106.95	105.20
92	108.22	106.37
93	109.49	107.54
94	110.76	108.71
95	112.03	109.88
96	113.30	111.05
97	114.57	112.22
98	115.84	113.39
99	117.11	114.56
100	118.38	115.73

mean values (arithmetic)
according to reference; deviating

threshold standard deviation

NFI reference pattern

CH1	CH2	CH3
31	32.90	35.67
32	33.00	37.17
33	34.30	37.33
34	35.57	38.50
35	36.83	39.68
36	38.10	40.85
37	39.37	42.02
38	40.64	43.19
39	41.91	44.36
40	43.18	45.53
41	44.45	46.70
42	45.72	47.87
43	46.99	49.04
44	48.26	50.21
45	49.53	51.38
46	50.80	52.55
47	52.07	53.72
48	53.34	54.89
49	54.61	56.06
50	55.88	57.23
51	57.15	58.40
52	58.42	59.57
53	59.69	60.74
54	60.96	61.91
55	62.23	63.08
56	63.50	64.25
57	64.77	65.42
58	66.04	66.59
59	67.31	67.76
60	68.58	68.93
61	69.85	70.10
62	71.12	71.27
63	72.39	72.44
64	73.66	73.61
65	74.93	74.78
66	76.20	75.95
67	77.47	77.12
68	78.74	78.29
69	79.01	79.46
70	80.28	80.63
71	81.55	81.80
72	82.82	82.97
73	84.09	84.14
74	85.36	85.31
75	86.63	86.48
76	87.90	87.65
77	89.17	88.82
78	90.44	89.99
79	91.71	91.16
80	92.98	92.33
81	94.25	93.50
82	95.52	94.67
83	96.79	95.84
84	98.06	97.01
85	99.33	98.18
86	100.60	99.35
87	101.87	100.52
88	103.14	101.69
89	104.41	102.86
90	105.68	104.03
91	106.95	105.20
92	108.22	106.37
93	109.49	107.54
94	110.76	108.71
95	112.03	109.88
96	113.30	111.05
97	114.57	112.22
98	115.84	113.39
99	117.11	114.56
100	118.38	115.73

10% / 20% offset bands pertinent to 100% ± 50%
mean values, and factor 5 pertinent to 0% values
i.e. accepted ranges for reference pattern

NFI trace as f(t)

R1/R2	R1/R3	R2/R3
signal in all target-related channels		
constant	variable	variable
variable	bias < 1	bias < 1
bias < 1	/ early	/ early
variable	/ late	variable
late	/ late	bias > 1
bias > 1	/	/
variable	variable	variable
late	bias < 1	bias > 1
/	/	/
late	late	late
variable	variable	variable
/	late	/
bias > 1	bias > 1	bias > 1
/	/	/
late	late	late
variable	variable	variable
/	late	/
bias > 1	bias > 1	bias > 1
variable	variable	variable
/	late	/
bias > 1	bias > 1	bias > 1
variable	variable	variable
/	early	/
bias < 1	bias < 1	bias < 1

legend: - variable / constant pertains to the
NFI ratio along the reaction coordinate
- early / late denotes the approximation
of the final value to case of var. ratios

assumptions:

Assumptions	HIV-1-M	R1/R2	efficiency:
P1	HIV-1-M	R1/R2	high
P2	HIV-1-O	R1	low
P3	HIV-2	R2	low
P4	HCV	R3	high
P5	HBV	R1/R3	high
P6	CMV	R2/R3	medium
P7	IC (internal control)	R4	

Roughly similar litres are assumed for all infections considered here.

basic analytical means^a

auxiliary analytical means^a

INTERNATIONAL SEARCH REPORT

International Application No
PCT/EP 02/09496

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 C12Q1/68

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 7 C12Q

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

EPO-Internal, WPI Data, PAJ, BIOSIS, MEDLINE

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 5 759 781 A (BALLARD STEPHEN GWYNN ET AL) 2 June 1998 (1998-06-02)	1,2,4,8-10,12
Y	column 2, line 1 -column 8, line 39; example 1	3,5-7,11
X	----- SPEICHER M R ET AL: "KARYOTYPING HUMAN CHROMOSOMES BY COMBINATORIAL MULTI-FLUOR FISH" NATURE GENETICS, NEW YORK, NY, US, vol. 12, April 1996 (1996-04), pages 368-375, XP000930084 ISSN: 1061-4036	1,2,4,8-10,12
Y	* Methods * page 368, column 1 -page 370, column 2; table 2 ----- -/--	3,5-7,11

☒ Further documents are listed in the continuation of box C.

☒ Patent family members are listed in annex.

° Special categories of cited documents :

A document defining the general state of the art which is not considered to be of particular relevance

E earlier document but published on or after the international filing date

L document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

O document referring to an oral disclosure, use, exhibition or other means

P document published prior to the international filing date but later than the priority date claimed

T later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

X document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

Y document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

* & * document member of the same patent family

Date of the actual completion of the international search

30 October 2002

Date of mailing of the international search report

07/11/2002

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,
Fax: (+31-70) 340-3016

Authorized officer

Aguilera, M

INTERNATIONAL SEARCH REPORT

International Application No
PCT/EP 02/09496

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT		
Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	VET J A M ET AL: "Multiplex detection of four pathogenic retroviruses using molecular beacons" PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF USA, NATIONAL ACADEMY OF SCIENCE. WASHINGTON, US, vol. 96, 1999, pages 6394-6399, XP002145609 ISSN: 0027-8424 * Abstract; Materials and Methods *	5-7,11
Y	MERCIER BERNARD ET AL: "Simultaneous screening for HBV DNA and HCV RNA genomes in blood donations using a novel TaqMan PCR assay." JOURNAL OF VIROLOGICAL METHODS, vol. 77, no. 1, January 1999 (1999-01), pages 1-9, XP002193782 ISSN: 0166-0934 * whole document *	5-7,11
Y	TONG ANTHONY K ET AL: "Combinatorial fluorescence energy transfer tags for multiplex biological assays." NATURE BIOTECHNOLOGY, vol. 19, no. 8, August 2001 (2001-08), pages 756-759, XP002193783 August, 2001 ISSN: 1087-0156 page 756, column 1, paragraph 2 -column 2, paragraph 1; figure 1	5
A	MENG Q ET AL: "Automated multiplex assay system for simultaneous detection of hepatitis B virus DNA, hepatitis C virus RNA, and human immunodeficiency virus type 1 RNA." JOURNAL OF CLINICAL MICROBIOLOGY, vol. 39, no. 8, August 2001 (2001-08), pages 2937-2945, XP002193784 ISSN: 0095-1137 page 2938, column 2, paragraph 1	1-12
A	SAMIOTAKI M ET AL: "SEVEN-COLOR TIME-RESOLVED FLUORESCENCE HYBRIDIZATION ANALYSIS OF HUMAN PAPILLOMA VIRUS TYPES" ANALYTICAL BIOCHEMISTRY, ACADEMIC PRESS, SAN DIEGO, CA, US, vol. 253, no. 2, 15 November 1997 (1997-11-15), pages 156-161, XP000721248 ISSN: 0003-2697 cited in the application * whole document *	1-12
	-/--	

Form PCT/ISA/210 (continuation of second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT

International Application No
PCT/EP 02/09496

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	JOSEFSSON AGNETHA ET AL: "Detection and quantitation of human papillomavirus by using the fluorescent 5' exonuclease assay." JOURNAL OF CLINICAL MICROBIOLOGY, vol. 37, no. 3, March 1999 (1999-03), pages 490-496, XP002193785 ISSN: 0095-1137 * whole document * -----	1-12
A	RIED T ET AL: "SIMULTANEOUS VISUALIZATION OF SEVEN DIFFERENT DNA PROBES BY IN SITU HYBRIDIZATION USING COMBINATORIAL FLUORESCENCE AND DIGITAL IMAGING MICROSCOPY" PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF USA, NATIONAL ACADEMY OF SCIENCE. WASHINGTON, US, vol. 89, 1 February 1992 (1992-02-01), pages 1388-1392, XP002073155 ISSN: 0027-8424 cited in the application * whole document * -----	1-12

Form PCT/ISA/210 (continuation of second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No

PCT/EP 02/09496

Patent document cited in search report		Publication date	Patent family member(s)	Publication date
US 5759781	A	02-06-1998	AU 724912 B2	05-10-2000
			AU 1423597 A	17-07-1997
			CA 2241148 A1	03-07-1997
			EP 0874912 A1	04-11-1998
			JP 2000507809 T	27-06-2000
			US 6007994 A	28-12-1999
			WO 9723648 A1	03-07-1997
			CA 2241364 A1	03-07-1997
			EP 0879297 A1	25-11-1998
			JP 11510707 T	21-09-1999
			WO 9723649 A1	03-07-1997

Form PCT/ISA/210 (patent family annex) (July 1992)

IPR2024-01451

(19) World Intellectual Property
Organization
International Bureau



(43) International Publication Date
14 October 2004 (14.10.2004)

PCT

(10) International Publication Number
WO 2004/087950 A2

(51) International Patent Classification⁷: **C12Q 1/68**

(21) International Application Number:
PCT/EP2004/003457

(22) International Filing Date: 1 April 2004 (01.04.2004)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
03007458.7 4 April 2003 (04.04.2003) EP
03014929.8 1 July 2003 (01.07.2003) EP
03017561.6 7 August 2003 (07.08.2003) EP

(71) Applicant (for DE only): **ROCHE DIAGNOSTICS GMBH** [DE/DE]; Sandhofer Strasse 116, 68305 Mannheim (DE).

(71) Applicant (for all designated States except DE): **F. HOFFMANN-LA ROCHE AG** [CH/CH]; Grenzacherstrasse 124, CH-4070 Basel (CH).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **SAGNER, Gregor**

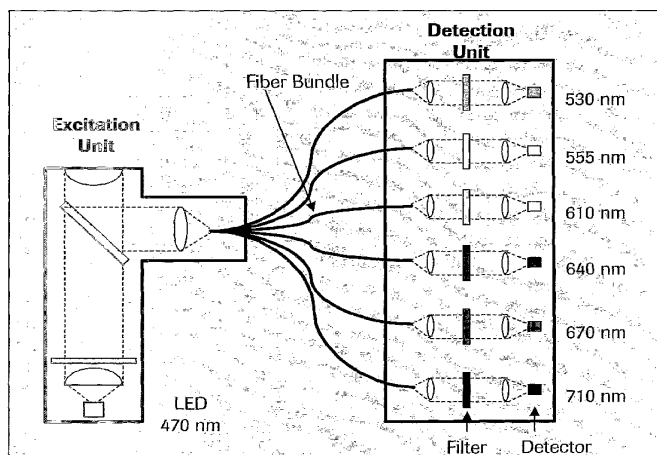
[DE/DE]; Meichelbeckstrasse 16, 82377 Penzberg (DE). **BECHLER, Ingrid** [DE/DE]; Wolfratshauser Strasse 34 a, 82538 Gelting (DE). **BOLTE, Joachim** [DE/DE]; Martin-Luther Strasse 2c, 68642 Buerstadt (DE). **HEINDL, Dieter** [DE/DE]; Sternstrasse 4, 82396 Paehl (DE). **JOSEL, Hans-Peter** [DE/DE]; Ulmenstrasse 28, 82362 Weilheim (DE). **GUTEKUNST, Martin** [DE/DE]; Escherstrasse 12, 82390 Eberfing (DE). **SEIBL, Rudolf** [DE/DE]; Kapellenwiese 29, 82377 Penzberg (DE). **MUELLER, Christoph** [DE/DE]; Ambacher Strasse 19, 81476 Muenchen (DE).

(74) Common Representative: **ROCHE DIAGNOSTICS GMBH**; C/O Heiko Hillebrandt, 82377 Penzberg (DE).

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

[Continued on next page]

(54) Title: IMPROVED SYSTEM FOR MULTI COLOR REAL TIME PCR



(57) Abstract: The invention is directed to a system for performing multi-color real time PCT, comprising a flexible real time PCT instrument and a specific composition or reaction mixture for performing multiplex PCR: In particular, the present invention is directed to a composition or reaction mixtuer which comprises at least 3, preferably 4-5 and most preferably exactly 4 pairs of FRET hybridization probes. Each pair of said hybridization probes consists of a FRET donor probe carrying a FRET donor moiety and a FRET acceptor probe carrying a FRET acceptor moiety having an emission maximum between 550 and 710 nm.

WO 2004/087950 A2



(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, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Declarations under Rule 4.17:

- as to the applicant's entitlement to claim the priority of the earlier application (Rule 4.17(iii)) for the following designations AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, 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, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW, ARIPO patent (BW, 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, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG)
- as to the applicant's entitlement to claim the priority of the earlier application (Rule 4.17(iii)) for the following designations AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, 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, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY,

TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW, ARIPO patent (BW, 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, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG)

- as to the applicant's entitlement to claim the priority of the earlier application (Rule 4.17(iii)) for the following designations AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, 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, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW, ARIPO patent (BW, 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, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG)
- of inventorship (Rule 4.17(iv)) for US only

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.

Improved system for multi color real time PCR

Field of invention

The present invention relates to the field of Real time PCR. In particular, the present invention is directed to a system for performing multiplex real time PCR.

5 Prior art background

Amplification of DNA by polymerase chain reaction (PCR) is a technique fundamental to molecular biology. Nucleic acid analysis by PCR requires sample preparation, amplification, and product analysis. Although these steps are usually performed sequentially, amplification and analysis can occur simultaneously. DNA
10 dyes or fluorescent probes can be added to the PCR mixture before amplification and used to analyze PCR products during amplification. Sample analysis occurs concurrently with amplification in the same tube within the same instrument. This combined approach decreases sample handling, saves time, and greatly reduces the risk of product contamination for subsequent reactions, as there is no need to
15 remove the samples from their closed containers for further analysis. The concept of combining amplification with product analysis has become known as "real time" PCR. See, for example, U.S. Patent No. 6,174,670.

Monitoring fluorescence during each cycle of PCR initially involved the use of ethidium bromide. (Higuchi R, G Dollinger, PS Walsh and R. Griffith, Simultaneous amplification and detection of specific DNA sequences, Bio/Technology 10 (1992) 413-417; Higuchi R, C Fockler G Dollinger and R Watson, Kinetic PCR analysis: real time monitoring of DNA amplification reactions, Bio/Technology 11 (1993) 1026-1030). In that system fluorescence is measured once per cycle as a relative measure of product concentration. Ethidium
20 bromide detects double stranded DNA; if template is present fluorescence intensity increases with temperature cycling. Furthermore, the cycle number where an increase in fluorescence is first detected increases inversely proportionally to the log of the initial template concentration. Other fluorescent systems have been developed that are capable of providing additional data concerning the nucleic acid
25 concentration and sequence.
30

- 2 -

In kinetic real time PCR, the formation of PCR products is monitored in each cycle of the PCR. The amplification is usually measured in thermocyclers which have additional devices for measuring fluorescence signals during the amplification reaction.

5 Real time PCR instrumentation

Several types of real time detection thermocyclers are known in the art.

EP 0 640 828, for example discloses an apparatus for monitoring multiple nucleic acid amplification simultaneously. It is characterized in that it comprises a metal block thermal cycler, including a heat conducting member having multiple recesses formed therein in order to take up a multiwell plate such as a microtiter plate. Detection is obtained by means of a CCD camera, arranged for detecting light emitted (simultaneously) from all of said recesses. Alternatively, the use of fiber optics is suggested. Depending on the fluorescent dyes which are used and – more important-, depending on the presence of a filter wheel close to the CCD camera, the system has the capability of analyzing multiplex amplification reactions, wherein in one reaction chamber, one or more different amplicons are detected by 2 or more differently labeled hybridization probes. Yet, EP 0 640 828 does neither anticipate nor suggest, which labels or detection formats could be used for such a multiplex/multicolor approach.

20 US 6,015,674 discloses an apparatus and a system for real time PCR detection and quantification, characterized in that it is capable of detecting first and second fluorescent indicators, which may be used as labels for different hybridization probes in order to detect different amplicons in the same reaction vessel. Yet, US 6,015,674 does not disclose a system for performing multiplex experiments with a higher degree of complexity.

Another typical example is the Roche Diagnostics LightCycler (Cat. No. 2 0110468). It is a fast PCR system enabling kinetic on-line PCR quantification and subsequent analysis of PCR-product melting curves. The optical system of the current LightCycler version 1.2 being commercially available contains one light source, a blue light emitting diode (470 nm LED) and three detection channels. The amplification products are detected by means of fluorescent labeled hybridization

- 3 -

probes which only emit fluorescence signals when they are bound to the target nucleic acid or in certain cases also by means of fluorescent dyes that bind to double-stranded DNA. A defined signal threshold is determined for all reactions to be analysed and the number of cycles C_p required to reach this threshold value is determined for the target nucleic acid as well as for the reference nucleic acids such as the standard or housekeeping gene. The absolute or relative copy numbers of the target molecule can be determined on the basis of the C_p values obtained for the target nucleic acid and the reference nucleic acid.

The fluorescence emitted by a sample is separated by a set of dichroic mirrors and filters into different wavelengths that can be recorded in one of the three detection channels (530 / 640 / 710 nm). This allows detection of the double-stranded DNA-binding dye SybrGreenI, mono color detection of the TaqMan Probe format and dual color detection of the Hybridization Probe(HybProbe) format. Details of the Lightcycler system are disclosed in WO 97/46707, WO 97/46712 and WO 98/46714, The complete contents of these applications are herewith incorporated by reference.

A very important feature of the LightCycler instrument is the color compensation software. In principle this software allows for accurate quantification and melting curve analysis by means of correcting spectral overlap of monitored fluorescent radiation in a temperature dependent manner. Technical details are disclosed in US 6,197,520.

Similar to the LightCycler system, the Corbett Rotor-Gene Real time PCR Thermocycler (www.corbettresearch.com) is a 4 channel multiplexing system comprising 4 different LEDs as excitation sources and corresponding photodiodes as fluorescent detection units. Thus, although this instrument hardware at least theoretically has the capacity of performing multiplex experiments with up to four differently labeled hybridization probes within one reaction vessel, no respective successful application protocol has been published so far.

Another real time PCR instrument ist the Biorad iQ Multi-color Real time PCR detection system (Cat. No: 170-8740), which allows for a fluorophore excitation and emission from 400 nm to 700 nm. The system is based on a conventional multiwell heating block for thermocycling, a tungsten lamp as an excitation source, a filter wheel for providing appropriate excitation wavelengths, a second filter wheel

- 4 -

for selecting appropriate emission wavelengths and a CCD camera as a detection unit. The instrument has successfully been used in a multiplex assay for the detection of 4 different amplicons generated from targets with more or less equimolar concentrations, using four differently labeled TaqMan probes in the same reaction vessel (Pedersen, S., Bioradiations 107 (2001) 10-11).

In a further approach to increase multiplexing capacities, US 6,369,893 discloses a real time PCR thermocycling instrument comprising a first optics assembly with at least two light sources and a second optics assembly comprising at least two detectors for detecting and discriminating light of different emission wavelengths. In particular, a specific embodiment of different 4 LEDs as light sources and 4 different photodiodes as detectors is disclosed. Thus, the instrument disclosed in 6,369,893 in principle can be used for real time PCR detection with a broad selection of different fluorescent dyes which are known in the art. Yet, US 6,369,893 does neither anticipate nor suggest any approach on how a multiplex experiment comprising multiple different probes each labeled with a different fluorescent entity needs to be designed.

Real time PCR detection formats

In general, there exist different formats for real time detection of amplified DNA, of which the following are well known and commonly used in the art:

a) DNA binding dye format

Since the amount of double stranded amplification product usually exceeds the amount of nucleic acid originally present in the sample to be analyzed, double-stranded DNA specific dyes may be used, which upon excitation with an appropriate wavelength show enhanced fluorescence only if they are bound to double-stranded DNA. Preferably, only those dyes may be used which like SybrGreen I, for example, do not affect the efficiency of the PCR reaction.

All other formats known in the art require the design of a fluorescent labeled Hybridization Probe which only emits fluorescence upon binding to its target nucleic acid.

- 5 -

b) TaqMan probe

A single-stranded Hybridization Probe is labeled with two components. When the first component is excited with light of a suitable wavelength, the absorbed energy is transferred to the second component, the so-called quencher, according to the principle of fluorescence resonance energy transfer. During the annealing step of the PCR reaction, the hybridization probe binds to the target DNA and is degraded by the 5'-3' exonuclease activity of the Taq Polymerase during the subsequent elongation phase. As a result the excited fluorescent component and the quencher are spatially separated from one another and thus a fluorescence emission of the first component can be measured (US5,538,848).

c) Molecular Beacons

These hybridization probes are also labeled with a first component and with a quencher, the labels preferably being located at both ends of the probe. As a result of the secondary structure of the probe, both components are in spatial vicinity in solution. After hybridization to the target nucleic acids both components are separated from one another such that after excitation with light of a suitable wavelength the fluorescence emission of the first component can be measured (US 5,118,801).

d) Single Label Probe (SLP) Format

This detection format consists of a single oligonucleotide labeled with a single fluorescent dye at either the 5'- or 3'-end (WO 02/14555). Two different designs can be used for oligo labeling: G-Quenching Probes and Nitroindole-Dequenching probes.

In the G-Quenching embodiment, the fluorescent dye is attached to a C at oligo 5'- or 3'-end. Fluorescence decreases significantly when the probe is hybridized to the target, in case two G's are located on the target strand opposite to C and in position 1 aside of complementary oligonucleotide probe.

In the Nitroindole Dequenching embodiment, the fluorescent dye is attached to Nitroindole at the 5'- or 3'-end of the oligonucleotide. Nitroindole somehow

- 6 -

decreases the fluorescent signaling of the free probe. Fluorescence increases when the probe is hybridized to the target DNA due to a dequenching effect.

ε) FRET hybridization probes

5 The FRET Hybridization Probe test format is especially useful for all kinds of homogenous hybridization assays (Matthews, J.A., and Kricka, L.J. , Analytical Biochemistry 169 (1988) 1-25. It is characterized by a pair of two single-stranded hybridization probes which are used simultaneously and are complementary to adjacent sites of the same strand of the amplified target nucleic acid. Both probes are labeled with different fluorescent components. When excited with light of a
10 suitable wavelength, a first component transfers the absorbed energy to the second component according to the principle of fluorescence resonance energy transfer such that a fluorescence emission of the second component can be measured when both hybridization probes bind to adjacent positions of the target molecule to be detected.

15 When annealed to the target sequence, the hybridization probes must sit very close to each other, in a head to tail arrangement. Usually, the gap between the labeled 3' end of the first probe and the labeled 5' end of the second probe is as small as possible, i.e. 1-5 bases. This allows for a close vicinity of the FRET donor compound and the FRET acceptor compound, which is typically 10-100 Angstrom.

20 Alternatively to monitoring the increase in fluorescence of the FRET acceptor component, it is also possible to monitor fluorescence decrease of the FRET donor component as a quantitative measurement of hybridization event.

25 In particular, the FRET Hybridization Probe format may be used in real time PCR, in order to detect the amplified target DNA. Among all detection formats known in the art of real time PCR, the FRET-Hybridization Probe format has been proven to be highly sensitive, exact and reliable (WO 97/46707; WO 97/46712; WO 97/46714). Yet, the design of appropriate FRET Hybridization Probe sequences may sometimes be limited by the special characteristics of the target nucleic acid sequence to be detected.

As an alternative to the usage of two FRET hybridization probes, it is also possible to use a fluorescent-labeled primer and only one labeled oligonucleotide probe (Bernard, P. S., et al., Analytical Biochemistry 255 (1998) 101-107). In this regard, it may be chosen arbitrarily, whether the primer is labeled with the FRET donor or the FRET acceptor compound.

Besides PCR and real time PCR, FRET hybridization probes are used for melting curve analysis. In such an assay, the target nucleic acid is amplified first in a typical PCR reaction with suitable amplification primers. The hybridization probes may already be present during the amplification reaction or added subsequently. After completion of the PCR-reaction, the temperature of the sample is constitutively increased, and fluorescence is detected as long as the hybridization probe was bound to the target DNA. At melting temperature, the hybridization probes are released from their target, and the fluorescent signal is decreasing immediately down to the background level. This decrease is monitored with an appropriate fluorescence versus temperature-time plot such that a first derivative value can be determined, at which the maximum of fluorescence decrease is observed.

There exist many different pairs of fluorescent dyes known in the art which according to the invention are principally capable of acting together as a FRET donor/FRET acceptor pair. Yet, prior to the present invention, no functional example has been disclosed, characterized in that 4 different FRET pairs have successfully been used in a multiplex detection assay. Among other reasons, this may be due to lack of appropriate instrumentation and, moreover, due to fact that the functionality of the FRET process of a specific FRET pair is interfered by other fluorescent compounds which are present in the same reaction mixture.

As discussed above, there exist different real time detection thermocycler instruments having a maximum of 4 detector channels for multiplex/multicolor detection. Yet, the utility of all of these instruments for multiplex detection up to now has been very limited due to the fact that attempts to establish real time multicolor multiplex assays with several (at least more than two) differently labeled probes with sufficient sensitivity and specificity have not been successful so far.

Thus, it was the object of the present invention to provide an improved system which allows for an optimized and at the same time flexible design of

- 8 -

multiplex/multicolor detection experiments. In one aspect the problem to be solved relates to improvements in the design of appropriate hybridization probes. In another aspect, the problem to be solved relates to an improvement of instrumentation.

5 Brief description of the invention:

Thus, the new invention is directed to a real time PCR system with an extension of applicable detection formats and multicolor features. The system comprises specific reagent mixtures for multiplex/multicolor detection as well as an improved real time PCR instrument with an optical detection system comprising multiple
10 detection units. The optical system is designed to provide excitation sources and detection channels for multiple detection formats, for example: SybrGreenI, multiplex FRET-Hybridization Probes, and multiplex TaqMan probes.

More precisely, the invention is directed to a system for performing multi-color real time PCR, comprising a real time PCR instrument and a specific composition or
15 reaction mixture for performing multiplex PCR:

The composition or reaction mixture of the present invention comprises at least 3, preferably 4-5 and most preferably exactly 4 pairs of FRET hybridization probes. Each pair of said hybridization probes consists of a FRET donor probe carrying a FRET donor moiety and a FRET acceptor probe carrying a FRET acceptor moiety
20 having an emission maximum between 550 and 710 nm.

In one embodiment at least 3, preferably at least 4 and most preferably exactly 4 FRET donor moieties are identical. Most preferably, all FRET donor moieties are identical. In other words, if all FRET donor moieties of the different FRET pairs are identical, they can be excited with the same excitation source.

25 In another embodiment of the invention at least 3, preferably at least 4 and most preferably exactly 4 FRET donor moieties are Fluorescein. Most preferably, all FRET donor moieties are Fluorescein.

In a specific embodiment, at least one additional FRET donor moiety is selected from a group consisting of Atto425 and WI343.

- 9 -

In another specific embodiment, which is not at all mutual exclusive to those embodiments disclosed above, one FRET acceptor moiety is selected from a group consisting of LC-Red 705 , Cy5.5, JA286 and a sulfonated cyanine dye as disclosed in US 6,027,709, formula 1.

5 As a second FRET acceptor moiety Cy5 may be selected.

As a third FRET acceptor moiety, LC-Red 640 may be used.

As a fourth FRET acceptor moiety, the LCRed derivative LC-Red 610 is chosen.

In a particular embodiment, the fifth FRET acceptor moiety is selected from a group consisting of Rh6G and TAMRA.

10 The instrument part of the system is characterized in that said real time PCR instrument comprises

- at least 1 light source, preferably an LED
- at least 4 and preferably 5-6 fluorescent detector entities, each of said entities having central detection wavelengths which are distinct from each other by at least 25 and preferably at least 30 nm

15

characterized in that said detector entities are capable of

- simultaneously detecting maximum fluorescence emission of at least 3, preferably 4 and most preferably 5 differently labeled FRET Hybridization Probe pairs,

20

- simultaneously detecting maximum fluorescence emission of at least 2 differently labeled TaqMan hybridization probes, and
 - detecting maximum fluorescence emission of SybrGreenI
 - means for heating and cooling

- 10 -

- multiple reaction vessels for containing a reaction mixture.

In a preferred embodiment of the invention the instrument part of the system is characterized in that said real time PCR instrument comprises exactly one light source.

5 In this context it is understood that although the instrument is being capable of detecting fluorescence emission of the FRET format, the TaqMan format and the SybrGreen format, the term „simultaneously“ means that within the same thermocycling protocol (i. e. within one run of the instrument), either at least 3 differently labeled FRET Hybridization Probes or at least 2 differently TaqMan
10 probes or SybrGreenI are detected.

Preferably, the instrument comprises at least 24, more preferably 32 and most preferably 48 of said reaction vessels.

Also preferably, said centrally detected wavelengths are selected from a group of range of wavelengths, said group consisting of 520-540 nm, 545-565 nm, 570-590
15 nm, 600-620 nm, 630-650 nm, 660-680 nm, and 700-720 nm.

Also preferably, each reaction vessel is characterized in that fluorescence excitation by said fluorescent light sources and fluorescence monitoring by said fluorescent detector entities are obtained along the same axis of said reaction vessel.

It has been proven to be particular advantageous, if said light sources and said
20 detector entities are located in separate housings.

It has also been proven to be particular advantageous, if said means for heating and cooling are means of forced liquid or forced gas, preferentially means of forced air.

It has further been proven to be particular advantageous, if the reaction vessels are fixed in a rotating carousel.

25 In a last aspect, the invention is also directed to a corresponding method for amplifying and detecting multiple target DNA sequences comprising

- 11 -

- 5
- a) providing a reaction mixture as disclosed above,
 - b) subjecting said reaction mixture to a thermocycling protocol such that amplification of said multiple target sequences can take place
 - c) monitoring hybridization of each of said pairs of FRET hybridization probes at least once after a plurality of amplification cycles.

Hybridization may also be monitored at least once in a temperature dependent manner in order to perform a melting curve analysis.

Detailed description of the invention:

A) Instrument

10 In a first aspect, a system according to the invention comprises an instrument suitable for real time PCR. Optionally, the instrument is also suited for melting curve analysis, i. e. monitoring temperature dependent binding of a DNA binding entity such as a hybridization probe or a ds DNA binding dye.

15 The instrument basically consists of an optical part and a means for thermocycling which can subject multiple amplification solutions in their respective reaction vessels to a process of repeated thermocycling such that a polymerase chain reaction (PCR) can take place.

Preferably, the instrument part of the system is characterized in that said real time PCR instrument comprises

- 20
- at least 1 light source, preferably a light emitting diode
 - at least 4 and preferably 5-6 fluorescent detector entities, each of said entities having central detection wavelengths which are distinct from each other by at least 25 and preferably at least 30 nm

characterized in that said detector entities are capable of

- 12 -

- simultaneously detecting maximum fluorescence emission of at least 3, preferably 4 and most preferably 5 differently labeled FRET Hybridization Probe pairs,
- 5 ◦ simultaneously detecting maximum fluorescence emission of at least 2 differently labeled TaqMan hybridization probes, and
- detecting maximum fluorescence emission of SybrGreenI
- means for heating and cooling, and
- multiple reaction vessels for containing a reaction mixture.

10 Most preferably, the instrument part of the system is characterized in that said real time PCR instrument comprises exactly one light source.

15 The means for thermocycling can be chosen arbitrarily. For example, metal block cyclers which are well known in the art may be used. Yet it is advantageous if the means for thermocycling provide means for active heating as well as independent means for active cooling. Moreover it has also been proven to be particular advantageous, if said means for heating and cooling are means of forced liquid or forced gas. In a specific embodiments said means use forced air.

20 It is also preferred, if said means for thermocycling are capable of performing rapid thermocycling like it is the case in the LightCycler technology (WO 97/46712). As it can be deduced from WO 97/46712, this technology is based on a forced air heating characterized in that ambient air is heated by a heating coil and forced air cooling characterized in that ambient air is provided.

25 A further improvement with respect to rapid thermocycling is the usage of capillaries as reaction vessels due to their comparatively high surface/volume ratio. Capillaries with an outer diameter of less than 5 mm, preferably of less than 3,2 mm and most preferably of less than 1,6 mm have been proven by the inventors to be particularly advantageous, since they allow for reaction mixtures having total volumes between 10 and 100 µl which still can be subjected to a rapid thermocycling protocol with a reaction time of less than 2 minutes per cycle.

- 13 -

Moreover, each reaction vessel needs to be fabricated at least partially or completely of an optically clear material such as glass or plastic.

5 Positioning of each reaction vessel (capillary) in a monitoring position is preferably characterized in that fluorescence excitation and monitoring are performed along the same axis of said reaction vessel. Such a principle enhances signalling due to the phenomenon of total internal reflectance.

It has further been proven to be particularly advantageous, if the reaction vessels are fixed in a rotating carousel in order to generate uniform thermocycling profiles throughout the multiple reaction vessels.

10 The optical part of the instrument basically consists of an excitation unit with one or more appropriate light sources in order to excite the fluorescent compound(s) which are present in the reaction solutions and a detection unit for detecting fluorescence emission from said reaction vessels.

15 According to WO 97/46712, it is possible to construct an optical system wherein the excitation unit and the detection unit are placed in a common housing, provided that an appropriate number of dichroic mirrors is used. Yet, in the context of the present invention, it turned out to be advantageous, if the excitation unit comprising the light sources and the detection unit comprising the detector entities are located in separate housings with the result that the improved real time PCR
20 instrument comprises two separate optical units.

For the excitation unit, either monochromatic excitation or polychromatic excitation may be used according to the invention. For monochromatic excitation, either a monochromatic LED (light emitting diode) or a monochromatic laser may be used. It is also possible to use instruments with two or multiple LEDs or
25 monochromatic lasers but only using one at a time in an individual measurement. For polychromatic excitation, a white LED, a Halogen lamp or a Xenon lamp may be used. In this case either a filter wheel or other means for moving appropriate filters into and out of the excitation beam need to be installed in the instrument such that light with different selective excitation wavelengths can be provided.

- 14 -

The optical features of the excitation unit according to the invention are selected according to different parameters that have to be considered:

5 First, the shortest excitation wavelength should be longer than 360 nm and preferably longer than 400 nm, since short wavelength radiation has comparatively low transmission rates through conventionally used optical material which is used for real time PCR reaction vessels such as glass or plastic. In addition, excitation with higher wavelengths for example in the range of 400-500 nm avoids interference of intrinsic fluorescence from the material of the reaction vessels which might disturb monitoring of the reaction.

10 In a preferred embodiment of the invention, the excitation unit consists of only one light source, because according to the present invention even with a single light source it is possible to perform a multiplex detection in real time PCR. Moreover, a multiplex detection in real time PCR is even possible with a single, monochromatic light source.

15 In one embodiment, the excitation unit comprises a blue LED emitting at 470 nm. Fig. 1 shows the excitation spectra for different fluorescent dyes which may be used according to the present invention. Among these dyes, Fluorescein as a frequently used FRET donor compound or TaqMan reporter dye known by a person skilled in the art can be excited with a single light emitting diode having an excitation wavelength of 470 nm, whereas frequently used FRET acceptor dyes are only excited marginally with 470 nm LED.

In addition, a second UV-LED emitting in between 400 and 430 nm, preferably at 415 nm may be used optionally for exciting other FRET donor compounds having a shorter excitation maximum as compared to Fluorescein.

25 More precisely, the blue 470 nm LED is capable of exciting Fluorescein, which is used in the HybProbe format as donor dye in combination with 4 different acceptor (reporter) dyes. On the other hand, typical FRET acceptor dyes such as LC-Red-640 or Cy5 are only excited marginally by the 470 nm LED. The blue LED may also excite the reporter dyes FAM and HEX or VIC in dual color TaqMan assays. In addition, Fluorescein is used as reporter dye in the Single Label Probe (SLP) format.

Moreover, the blue LED is used to excite SybrGreenI for non probe based detection of amplification products.

5 The optional UV LED may be used in the HybProbe format to excite a short wavelength donor dye in combination with appropriate reporter dyes that emit fluorescence at comparatively short wavelengths. Moreover, a third LED emitting at 580 nm may be included.

10 For the detection unit, different modes of detection which are known in the art can be used according to the invention. For example, it is well known in the art to use several photodiodes characterized in that each photodiode is capable of detecting fluorescent light of a specific wavelength. In one aspect of the invention, a detection system having exactly five or six channels using photodiodes as detectors is provided. As it will be shown, such embodiments already enable for a simultaneous 4 color FRET Hybridization Probe detection.

15 The optical features of the detection unit according to the invention are selected according to the following different parameters:

20 First of all, the longest emission wavelength that should be detected is around less than 800 nm, and preferably less than 730 nm, since detection of higher wavelength signals is affected by a possible background of putative infrared radiation by the heating part of the instrument. Moreover, long wavelength infrared dyes which are known in the art have a suboptimal quantum yield, are less stable and thus more difficult to couple to a desired hybridization probe.

25 As a result, the different detection channels should have central detection maxima between 500 and 800 nm and preferably between 520 and 730 nm. In this context, fig. 2 shows the emission spectra for different fluorescent dyes which may be used according to the present invention.

Within the range of 500 and 800 nm, the detection channels can be defined exactly by means of selecting appropriate filters which the light beam has to pass prior to entering the respective photodiode. For appropriate selection, the following parameters have been identified by the inventors to be considered:

- 16 -

First, the central wavelength detection maximum of each channel should be chosen dependent on the maxima of the emission spectra of those fluorescent dyes that shall be detected predominantly. As a rule of thumb, for appropriate detection sensitivity, the maximum of the fluorescent emission from the dye to be detected
5 should not differ more than ± 10 nm from the centrally detected wavelength of each channel.

Moreover, all detection channels should be able to detect discrete emissions, i. e. the spectral overlap should be minimized as far as possible. In this regard, it is advantageous, if the half band widths of each channel is 40 nm or less. More
10 advantageously, the half band widths of each channel is 20 nm or less.

Preferably, crosstalk of the fluorophores to be detected in neighboring channels is less than 70%, more preferably less than 50% and most preferred less than 30%. Another important criterium in this regard is the blocking value, which is indicative for the maximum intensity that is detected outside the specified transmission range
15 of the channel (filter). Preferably, this value is less than 10^{-3} .

Based on these criteria, it has been proven to be advantageous, if the centrally detected wavelengths of each of the 5-6 channels according to the invention are separated from each other by more than 25 nm and preferably even more than 30 nm.

20 In one embodiment, said centrally detected wavelengths are selected from a group of range of wavelengths, said group consisting of 520-540 nm, 545-565 nm, 570-590 nm, 600-620 nm, 630-650 nm, 660-680 nm, and 700-720 nm.

Moreover, in one specific embodiment consisting of 6 detector channels said channels having centrally detected wavelengths at around 530, 580, 610, 640, 670,
25 and 710 nm ± 5 nm or 530, 555, 610, 640, 670, and 710 nm ± 5 nm or preferably ± 2 nm.

In another specific embodiment consisting of 5 detector channels said channels have centrally detected wavelengths at around 530, 610, 640, 670, and 710 nm ± 5 nm or preferably ± 2 nm.

- 17 -

In a third specific embodiment consisting of 4 detector channels said channels have centrally detected wavelengths at around 610, 640, 670, and 710 nm $\pm$ 5 nm or preferably $\pm$ 2 nm.

5 As an example, fig. 3 shows the 6 detection channels of a preferred embodiment and the emission spectra for different fluorescent dyes which may be used according to the present invention. In this embodiment, the six channels are separated by at least 25 nm spectral distance to minimize the crosstalk from each reporter dye into neighboring detection channels. The detection unit is capable of detecting SybrGreenI and the TaqMan reporter dye Fluorescein/FAM in the channel having
10 its central wavelength detection maximum around 530 nm. In addition, the system is preferentially capable of detecting HEX/VIC in a second channel around 560 nm. Further long wavelength detection channels are capable of detecting HybProbe acceptor dyes and optionally long wavelength SLP reporter dyes.

15 The excitation and detection unit are connected by a multiple-leg fiber bundle. Emitted light from the reaction vessels (e.g. the glass capillaries) is homogeneously distributed using a lightpipe and is transmitted into six glass fiber bundles. These bundles of 50 μ m single glass fibers transmit the light into each of the six detection channels.

20 This set up of excitation unit and detection unit located in separate housings provides two advantages compared to the optical unit as disclosed in WO 97/46712: Homogeneous distribution of emitted light into all six detection channels and mechanical decoupling of the excitation and detection unit. This enables highly precise positioning of the excitation unit towards the reaction vessels to become monitored (e. g. capillary tips) without moving the detection unit. Moreover, the
25 number of necessary dichroic mirrors is minimized. An example of such a set up is shown in fig. 4, which discloses a possible embodiment of the invention. As it can be seen, excitation and detection unit are located in different housings.

30 In addition, the system according to the invention comprises a color compensation tool as disclosed in detail in US 6,197,520 which is incorporated into this application in its entirety by reference. The corresponding software enables for measuring fluorescence throughout a range of temperatures and correcting for temperature dependent spectral overlap of the 6 detector channels disclosed above.

- 18 -

In principle this is achieved by

- a) providing calibrator solutions for each fluorescent entity which is incorporated in the multiplex experiment to become performed
- 5 b) monitoring temperature dependent fluorescence of each of said solutions in each of the detector channels in order to determine the temperature dependent amount of crosstalk between different channels, and
- c) generating a color compensation data set containing correction values.

B) Test formats and selection of dyes

10 In the context of the present invention, the term „multiplex PCR“ is understood as a PCR reaction characterized in that 2 or more different amplification products are generated by means of using 2 or more pairs of amplification primers in the same PCR reaction.

15 Also in the context of the present invention, the term „multicolor real time PCR“ is defined as an real time PCR assay characterized in that 1 or more different amplification products generated either in a multiplex PCR or in a monoplex PCR (using only 1 pair of amplification primers) are detected by differently labeled hybridization probes.

20 A major aspect of the present invention is based on usage of differently labeled hybridization reagents, each reagent comprising a pair of FRET hybridization probes comprising a pair of two fluorescent dyes which interact with each other on the principle of fluorescence resonance energy transfer (FRET). According to the invention, appropriate dye pairs can be selected for the labeling of a hybridization reagent which may be used for detection in at least three or four and preferably as much detection channels as possible.

25 More precisely, such a hybridization reagent is composed of two adjacently hybridizing oligonucleotides, appropriately labelled such that together they can act according to the FRET-HybProbe detection format as disclosed in WO 97/46707, WO 97/46712, and WO 97/46714). In many cases, it is sufficient if the

- 19 -

hybridization reagent consists of a single oligonucleotide or in case of the FRET hybprobe format, of a pair of oligonucleotides acting together as a donor probe and an acceptor probe. Yet, in other cases there may exist many other sequence variants in the target sequences which need to be detected. Thus it may be impossible to
5 detect the sequences of all members by just using just one pair of FRET oligonucleotide hybridization probes.

For those cases, a hybridization reagent may consist of 1, 2 or more hybridization probes, which are similar and bind to homologous sequences, but differ from each other by 1, 2, 3 or more mononucleotide-, dinucleotide- or trinucleotide exchanges, deletions or additions. In case of the hybridization reagent being a pair of FRET
10 hybridization probes, said hybridization reagent may consist of 1, 2, 3, or more FRET donor oligonucleotide probes and/or 1, 2, 3, or more acceptor oligonucleotide probes. In this case all donor probes may be similar, but differ from each other by mononucleotide-, dinucleotide- or trinucleotide exchanges, deletions or additions. As well, all acceptor probes may be similar, differ from each other by
15 mononucleotide-, dinucleotide- or trinucleotide exchanges, deletions or additions.

In addition it is also within the scope of the present invention, if multi-color FRET detection is performed wherein one member of said pair of labeled hybridization probes is replaced by an appropriately labeled primer as disclosed in Bernard, P. S.,
20 et al., Analytical Biochemistry 255 (1998) 101-107. Moreover, it is also within the scope of the present invention, if instead of an increase in fluorescent emission from a respective FRET acceptor compound, the decrease in fluorescent emission of one, several, or all FRET donor moieties is monitored as being indicative for the presence of the target nucleic acid to be detected.

In addition to the various possibilites for performing multi-color detection based on the FRET principle, the system of the invention also allows for the performance of other detection formates such as detection of the target nucleic acid by means of SybrGreenI, multiplex detection using differentially labeled Molecular Beacons, Single Labeled Probes or dual color TaqMan detection. As it known in the art, dual
25 color TaqMan assays may be based on the use of Cy 5 (Amersham) as Quencher compounds in combination with FAM as a first standard reporter dye and either HEX or alternatively VIC as a second reporter compound.
30

- 20 -

5 In principle, there exist multiple possibilities of fluorescent dye combinations, which together may act as a FRET pair (Review: Resonance Energy Transfer Editors: Meer, B Wieb van der et al., VCH Publishers INC., 1994). However, based on the instrumental preconditions it is not trivial for person skilled in the art to develop a multi-color test which combines multiple FRET hybridization reagents that do not interfere with each other in such a way that sensitivity and/or specificity of a respective real time PCR experiment are affected.

10 Regarding the selection of appropriate pairs of FRET hybridization probes for example, the dyes used for FRET donors and FRET acceptors can not be chosen arbitrarily but need to fulfill at least the following requirements:

- sufficient solubility and capability of being coupled to an oligonucleotide
- stability of the dyes under PCR thermocycling conditions
- sufficient and reproducible emission intensity and quantum yield
- low temperature dependence of emission intensity
- 15 • essentially no spectral shifts under different chemical conditions
- essentially no spectral shifts when coupled to different oligonucleotide sequences
- discrete spectral emission maxima
- availability of a suitable partner dye for performing a FRET process

20 As a consequence, one important aspect of the system according to the invention is focussed on identification of multiple Fluorescence Resonance Energy Transfer (FRET) dye pairs usable in the HybProbe format for as many detection channels as possible. Excitation and emission spectra of dyes FRET dye pairs that enable a 4- or 5-plex multi-color detection set up are already shown in Fig. 1-3. Alternatives will be
25 be described below.

- 21 -

In this regard, the following table summarizes various FRET acceptor dyes and TaqMan reporter dyes which have successfully been used so far in the indicated detection channels performing either a 4-color FRET Hybridization Probe assay or a dual color TaqMan assay:

5

Table 1

Channel	FRET Acceptor dye	TaqMan Reporter dye
530 (also detection of SybrGreenI)	-	Fluorescein (Fam) Coumarine
555	Rh6G Tamra	Hex Vic Tamra
580		Tamra ROX
610	LCRed610	TexasRed
640	LCRed640	
670	Cy5 Bodipi 655 Alexa 647	
710	Ja286 Cy5.5 LCRed705	

Yet, for several of the different reporter entities indicated above, different FRET donor moieties (Donor) and TaqMan quencher compounds (Q) respectively have to be used depending on the excitation wavelength which is available due to the presence of either one, two or three different LEDs. Therefore, different possibilities of appropriate FRET pairs or TaqMan combinations are summarized in the following more comprehensive table:

10

- 22 -

Table 2

	LED	530	555	580	610	640	670	710
FRET:								
Donor: Atto 425 WI 343	410	-	Rh6G TAMRA	-	LCRed 610	LCRed 640	Cy5 Bodipy655 Alexa647	JA286 LCRed705 Cy5.5
Donor: Fluores- cein	470	-	-	-	LCRed 610	LCRed 640	Cy5 Bodipy655 Alexa647	JA286 LCRed705 Cy5.5
Donor: LCRed 640	590	-	-	-	-	-	Cy5 Bodipy655 Alexa647	JA286 LC705 Cy5.5
TaqMan:								
Q: Arbi- trary	410	Cou ma- rine	-	-	-	-	-	-
Q: Tamra	470	Fam	Hex Vic	-	-	-	-	-
Q: Cy5	470	Fam	Hex Vic TAMRA	TAMRA Rox	-	-	-	-
Q: BHQ	470	Fam	Hex Vic TAMRA	TAMRA Rox	-	-	-	-
Q: Cy5	590	-	-	-	Texas- Red	-	-	-
Q: BHQ	590	-	-	-	Texas- Red	-	-	-

The dyes mentioned in the tables above are available as follows:

5 LC-Red-705 and LC Red-640 are available from Roche Applied Science (Cat. No. 2 015 161 and 2 157 594)

LC-Red 610 is synthesized according to standard protocols using a fluorescent dye as disclosed in US 5,750,409, compound II

- 23 -

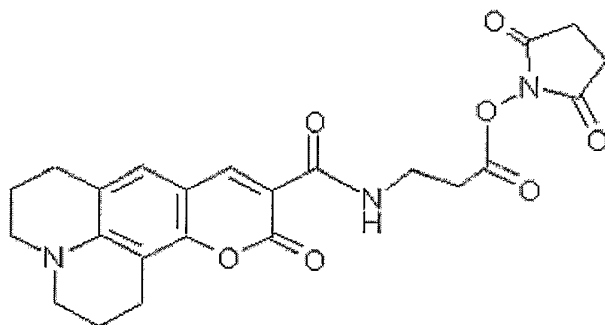
JA 286 is disclosed in EP 0 747 447, example 1. For this dye, it is advantageous, if this dye is connected to the hybridizing oligonucleotide by a long non hybridizing spacer moiety of five nucleotide residues.

5 Cy-5 NHS-ester is available from Amersham, (Cat. No. PA 15100) and Cyc 5.5 Phosphoramidite is also available from Amersham (Cat. No 271799 01).

Bodipy 650/665 NHS-Ester and Alexa 647 NHS Ester are available from Molecular Probes (Cat. Nos D1 000 1 and A 2 000 6).

Atto 425-NHS ester is available from Attotec (Cat. No. AD 425-3).

10 WI 343 is a derivative of Coumarin 343 (Aldrich, Cat. No. 55 804 654), which is coupled with a linker (beta alanin) and transformed to the succinimidyl ester according to standard methods:



Wi 343

15 In addition, all TaqMan dyes are available from different suppliers known by a person skilled in the art, for example Molecular probes.

The BlackHole Quencher (BHQ) is available from Biosearch Technologies (Cat. No. BHQ 1-3).

20 Appropriate labeling of oligonucleotide probes may be performed by conventional methods known by a person skilled in the art. In particular, labeling may be practiced using either fluorescent compounds that comprise an activated NHS-ester or fluorescent compounds linked to a Phosphoramidite, such that labeling can be

- 24 -

performed during a de-novo oligonucleotide synthesis itself. Furthermore, a Fluorescein-labeled probe may be prepared by using a Fluorescein Controlled Pore Glass particle as a solid support (Roche Applied Science Cat. No. 3 138 178) for oligonucleotide synthesis.

5 C) **Methods and Kits**

In a further aspect, the invention is directed to methods of performing multiplex PCR by means of using all different embodiments of the systems, instruments and compositions disclosed above.

10 In particular, the present invention is directed to a method for amplifying and detecting multiple target DNA sequences comprising

- providing a composition or reaction mixture according to the invention,
- subjecting said reaction mixture to a thermocycling protocol such that amplification of said multiple target sequences can take place, and
- 15 • monitoring hybridization of each of said pairs of FRET hybridization probes at least once after a plurality of amplification cycles.

In a specific embodiment, hybridization is monitored at least once in a temperature dependent manner.

20 The composition or reaction mixture predominantly comprises at least 3, preferably 4-5, and most preferably exactly 4 pairs of FRET hybridization probes, each pair of hybridization probes consisting of a FRET donor probe carrying a FRET donor moiety and a FRET acceptor probe carrying a FRET acceptor moiety having an emission maximum between 550 and 710 nm.

25 In addition, such a composition or reaction mixture according to the invention may comprise one or several or preferably all compounds and reagents selected from the following list:

- Buffer, applicable for a polymerase chain reaction

- 25 -

- Deoxynucleoside triphosphates
- Template dependent DNA polymerase, preferably thermostable,
- at least one pair or several pairs of amplification primers

5 In yet another aspect, the invention is directed to a kit comprising at least a first set of amplification primers and at least three, four, or five pairs of FRET hybridization probes.

In addition, such a kit according to the invention may comprise one or several other compounds and reagents selected from the following list:

- Buffer, applicable for a polymerase chain reaction
- 10 • Deoxynucleoside triphosphates
- Template dependent DNA polymerase, preferably thermostable,
- At least one or multiple pairs of amplification primers

15 Such a kit may also comprise an internal control DNA, which can be amplified and detected using the same primers and probes as used for the detection of any target nucleic acid. Each of the components disclosed above may be stored in a single storage vessel. Yet, any combination of components for storage within the same vessel is possible as well.

D) Best mode of the invention:

20 The best mode of the invention known by the inventors at the date of filing the application is as follows:

In this preferred embodiment, a system according to the invention provides a photometer with one or two excitation sources and six detection channels. A blue (470 nm) and optionally an UV (410 nm) light-emitting diode are combined with

- 26 -

detection channels having central detection wavelengths at 530, 555, 610, 640, 670 and 710 nm.

The photometer of the improved real time PCR instrument comprises two separate optical units. As disclosed in fig. 4, excitation and detection unit are located in different housings.

(I) An **excitation unit** with one or two LED's: A blue LED emitting at 470 nm and optionally a violet LED emitting at 415 nm. The blue LED is capable of exciting SybrGreenI and Fluorescein. Fluorescein is used in the HybProbe format as donor dye in combination with 4 different acceptor (reporter) dyes: Red 610, Red 640, Cy 5, Red 705. Furthermore the blue LED excites the reporter dyes FAM and HEX or VIC in dual color TaqMan assays. In addition, fluorescein is used as reporter dye in the Single Label Probe (SLP) format. The optional violet LED may be used in the HybProbe format to excite a short wavelength donor dye (Atto 425)) in combination with a reporter dye at 555 nm (Rh6G).

(II) A **detection unit** with six channels using photodiodes as detectors. The six channels are separated by at least 25 nm spectral distance to minimize the crosstalk from each reporter dye into neighboring detection channels. The detection unit is capable of detecting SybrGreenI and fluorescein/FAM in the 530 nm channel. In addition, the system is capable of detecting HEX/VIC and RH6G in the 555 nm channel. The long wavelength detection channels (610, 640, 670, 710 nm) are capable of detecting HybProbe acceptor dyes and optionally for long wavelength SLP reporter dyes.

The excitation and detection unit are connected by a six-leg fiber bundle. Emitted light from the glass capillaries is homogeneously distributed using a lightpipe and is transmitted into six glass fiber bundles. These bundles of 50 μ m single glass fibers transmit the light into each of the six detection channels. This setup provides two advantages compared to the LC 1.2 optical unit: Homogeneous distribution of emitted light into all six detection channels and mechanical decoupling of the excitation and detection unit. This enables highly precise positioning of the excitation unit towards the capillary tips without moving the detection unit. An effective color compensation function allows simultaneous detection of multiple targets in different channels.

Combined with these hardware and software features, multicolor capabilities of the system in order to perform Multiplex FRET Hybridization Probedetection have been established. In addition, the system of according to the invention allows for other non FRET-Hybprobe detection formats.

5 Based on the Hybridization Probe detection format, the system enables at least 4-color multiplexing of real-time quantification and melting curve analysis. Taking a melting curve resolution of six melting peaks per channel into consideration, up to 24 different PCR products are discernable using the photometer according to the invention. To enable this, new previously unknown FRET-acceptor dyes have been
10 developed for detection channels with central emission wavelengths at 610 and 670 nm. In addition, the optional 410 nm LED provides an option for using a short wavelength donor dye in combination with an acceptor dye emitting at 555 nm thus providing a fifth HybProbe detection channel. Excitation and emission spectra of specifically useful dyes are shown in Fig. 1 and 2.

15 Using fluorescein as FRET donor dye two new FRET acceptor dyes could be identified being detectable in the 610 nm channel (Red 610) and in the 670 nm channel (Cy 5) respectively. Both dyes show high signal dynamics when used in the HybProbe FRET format. In addition, the spectral properties of the dyes fit perfectly into the photometer disclosed above thus minimizing the crosstalk into
20 heterologous detection channels. Using a flexible multiple-channel color compensation algorithm it is shown that specific detection of all 4 FRET acceptor dyes (including the already existing dyes LC Red 640 and LC Red 705) from a single reaction is feasible over a temperature range from 40°C to 95°C.

25 When trying to identify a fifth FRET acceptor dye for the 555 nm channel it turned out that good FRET signal dynamics is inhibited by two effects:

(I) The fluorescein donor dye shows a high crosstalk into the 555 nm channel thus masking increasing signals of a potential FRET acceptor dye during the PCR process.

30 (II) A FRET acceptor dye emitting at 555 nm is itself directly excited by the 470 nm LED thus increasing the background fluorescence significantly and consequently lowering the detection sensitivity.

- 28 -

A solution to these drawbacks was found by using a 410 nm UV LED for excitation of a short wavelength donor dye (Atto 425) in combination with a suitable 555 nm acceptor dye (Rh6G). In this combination, crosstalk from the donor into the 555 nm detection channel and direct excitation of the FRET acceptor were minimized and sufficient signal dynamics was achieved after color compensation.

Furthermore, it turned out that the short wavelength donor Atto 425 can even be used in combination with all long wavelength acceptor dyes. Unexpectedly, FRET signal dynamics of long wavelength acceptors in combination with Atto 425 are comparable to signal dynamics when used in combination with fluorescein. This observation provides an option to extend multi color features to a 5-channel format using just one donor dye.

Using the TaqMan detection format, dual color applications based on standard TaqMan reporter dyes FAM (detected at 530 nm) and HEX or VIC (detected at 560 nm) can be performed without major modifications of the TaqMan chemistry known in the art. The obtained detection sensitivity is comparable to the COBAS TaqMan instrument (Roche Molecular Systems).

Furthermore, the Single Labeled Probes (SLP's) format for melting curve analysis (WO 02/14555) is applicable within a system according to the invention. This format is based on quenching or, respectively, de-quenching of the probe-label after hybridization to its target sequence. The format requires just one probe terminally labeled with a single dye. Compared to HybProbe or TaqMan formats, SLP's thus enable a significantly less expensive setup of assays for integrated analysis of Single Nucleotide Polymorphisms (SNP's). By reporting melting temperatures, different alleles of an SLP can unambiguously be distinguished in a single reaction. Multiplexing options are provided by using different reporter dyes.

Possibilities for Real time multicolor PCR detection in the above disclosed instrument are summarized in the following table:

- 29 -

Table 3

Assay Format	Detection Channels (Reporter Dyes)
SYBR Green I	530 nm (SYBR Green I)
HybProbes	610 nm (LC RED 610) 640 nm (LC RED 640) 670 nm (LC RED 670) 710 nm (LC RED 705)
Hydrolysis Probes (TaqMan Probes)	530 nm (FAM) 555 nm (VIC, HEX)
SimpleProbes	530 nm (Fluorescein)

5

- 30 -

The following examples, references, sequence listing and figures are provided to aid the understanding of the present invention, the true scope of which is set forth in the appended claims. It is understood that modifications can be made in the procedures set forth without departing from the spirit of the invention.

5 **Description of the Figures**

- Figure 1 Excitation spectra of FRET donor and acceptor dyes for the present invention
- Figure 2 Emission spectra of FRET donor and acceptor dyes for the present invention
- 10 Figure 3 Emission spectra of FRET donor and acceptor dyes and detection channels for the present invention
- Figure 4 Photometer setup
- Figure 5 4 color real time PCR quantification experiment of example 1
- Figure 6 Primer/probe design of example 2
- 15 Figure 7 4-color Real time PCR melting curve experiment of example 2
- Figure 8 Dual color TaqMan Real time PCR quantification experiment of example 3

Example 1

- 20 **Quantitative real time PCR of Factor V DNA using different pairs of FRET hybridization probes labeled with different fluorescent compounds**

For amplification of a Factor V DNA fragment, 4 different 100µl real time PCR reaction mixtures were set up as follows:

10⁶ copies of a plasmid containing Factor V gene

- 31 -

3 mM MgCl₂

500 nM each primers

200 nM FRET 3'Hybridization Probe according to Seq. A

5 200 nM FRET 5'Hybridization Probe according to Seq. B1, B2, B3 or B4, respectively.

In addition, PCR components from the LightCycler DNA Hyb Probes Kit (Roche Applied Science, Cat. No. 2158825) were used.

Primers and probes were used as follows:

Primer FactorV, forward:

10 5'-GAG AGA CAT CGC CTC TGG GCT A-3' (22-mer) (Seq. ID. No:1)

Primer FactorV, reverse:

5'-TGT TAT CAC ACT GGT GCT AA-3' (20-mer) (Seq. ID. No: 2)

A: 3'Fluorescein labeled Hybridization Probe

5'-AAT ACC TGT ATT CCT CGC CTG TC-3' (23-mer) (Seq. ID. No: 3)

15 B1: 5' Red610 Hybridization Probe

5'-AGG GAT CTG CTC TTA CAG ATT AGA AGT AGT CCT ATT-3' (36-mer)

B2: 5' Red640 Hybridization Probe

5'-AGG GAT CTG CTC TTA CAG ATT AGA AGT AGT CCT ATT-3' (36-mer)

B3: 5' Cy5 Hybridization Probe

- 32 -

5'-AGG GAT CTG CTC TTA CAG ATT AGA AGT AGT CCT ATT-3' (36-mer)

B4: 5' Red705 Hybridization Probe

5'-AGG GAT CTG CTC TTA CAG ATT AGA AGT AGT CCT ATT-3' (36-mer)

(B1-B4: Seq.Id. NO: 4)

- 5 Amplification was performed in a LightCycler instrument disclosed above as best mode of the invention according to the following thermocycling protocol:

	T[°C]	t[sec]	Ramp-rate[°C/sec]	Acquisition	Cycles
Denaturation	95	30	20.0	none	1
Amplification	95	10	20.0	none	
	55	20	20.0	single	45
	72	20	20.0	none	

15 Real time monitoring was performed with and without a color compensation algorithm and using the 2nd derivative threshold method over 45 cycles by measuring the fluorescence signals in a detection channel and using arithmetic background correction for normalization of initial fluorescence background intensities.

20 Results are shown in fig. 5a-5d. As can be seen in the figures, specific and quantitative amplification signals of all used FRET acceptor dyes could be obtained after color compensation in the corresponding 610, 640, 670 and 710 nm detection channels. Moreover, it is shown that the spectral overlap of the different dyes in the neighboring channels did not affect specific and quantitative signalling when appropriate color compensation is performed.

- 33 -

Example 24-color melting curve analysis

5 A 479 bp fragment of a specific plasmid containing a fragment of the human NAT-2 gene encoding N-Acetyl-Transsferase isozyme was amplified with specific primers and detected by fluorescence, using 4 different specific pairs of FRET Hybridization Probes. Whereas three probes were labeled at the 3'-end with Fluorescein, four detection probes were labeled 5' either with Light- Cyclor-Red 610, 640, 705 or Cy5 and modified at the 3'-end by phosphorylation. A schematic drawing of this experimental set up is shown in fig. 6. As it can be deduced from the figure, there were four differently labeled FRET acceptor probes, but only three detection sites. This resulted in a competition of annealing of two detection probes.

Assay conditions were basically the same as disclosed in example 1 with the alteration that 3x10⁶ copies of target DNA were used.

Primer forward:

15 5'-TGC CTT GCA TTT TCT GCT T-3' (19mer) (Seq. Id. No: 5)

Primer reverse:

5'-GAG TTG GGT GAT ACA TAC A-3' (19mer) (Seq. Id. No: 6)

3'-Fluorescein HybProbe 1:

20 5'-GAA ATT CTT TGT TTG TAA TAT ACT GCT CTC TC-Fluos-3' (32mer)
(Seq. Id. No:7)

5'-Red610 HybProbe 1a:

25 5'-Red610-TGA TTT GGT CCA CGT ACC-3' (18mer)
(Seq. Id. No:8)

5'-Red640 HybProbe 1b:

5'-Red640-TGA TTT GGT CCA AGT ACC C-3' (19mer)
(Seq. Id. No:9)

- 34 -

3'-Fluorescein HybProbe 2:

5'-GTT GGA GAC GTC TGC AGG TAT GTA TTC ATA GAC TCA A-Fluos-3'
(37mer)
(Seq. Id. No:10)

5

5'-Red705 HybProbe 2:

5'-Red705-ATC TTC AAT TGT TCG AGG TT-3' (20mer)
(Seq. Id. No:11)

10

3'-Fluorescein HybProbe 3:

5'-ATT TCC TTG GGG AGA AAT CTC GTG CCC A-Fluos-3' (28mer)
(Seq. Id. No:12)

5'-RedCy5 HybProbe 3:

15 5'-Cy5-ACC TGG TGA TGA ATC CCT TAC TAT TTA GAA TAA GGA AC-3'
(37mer)
(Seq. Id. No:13)

20 Amplification and subsequent melting curve analysis was performed in a
LightCycler instrument disclosed above as best mode of the invention according to
the following thermocycling protocol:

	T[°C]	t[sec]	Ramp-rate[°C/sec]	Acquisition	Cycles
Denaturation	95	600	20.0	none	1
Amplification	95	10	20.0	none	
	55	10	20.0	single	45
	72	20	3.0	none	
Melting Curve	95	60	20.0	none	
	40	60	20.0	none	1

25

- 35 -

	80	0	0.1	cont	
Cooling	40	30	20.0	none	1

Results are shown in fig. 7. As it can be deduced from the figure, 4 color melting curve analysis using Hybridization Probes with the 4 different acceptor dyes could be performed using an appropriate color compensation method. Hybridization Probe specific melting peaks could then be detected without any substantial crosstalk effects.

Example 3

Dual Color TaqMan Detection

A LightCycler instrument according to the best mode of the invention provided all means for sensitive detection of dual color TaqMan assays. The established TaqMan dye FAM and HEXs were well excited with the 470 nm LED and after using the instrument's color compensation function could differentially be identified in the 530 nm and 560 nm detection channels with high sensitivity. Transfer of dual color TaqMan assay protocols to the instrument according to the invention was simply performed by supplementing established assay conditions with BSA.

Different examples of successful dual color amplification and detection are shown in fig. 8. Fig. 8a shows that in mono-color experiments using either FAM or HEX, a detection sensitivity of about 1copy/ μ l for dual color TaqMan assays was obtained. (Only negative controls did not result in a significant increase in fluorescence after the PCR reaction). Fig. 8b shows a dual color experiment, wherein a Hex labeled TaqMan probe was always used to detect 500 copies of a first target DNA and in the same reaction a FAM labeled TaqMan probe was used to detect different amounts (10^0 - 10^6 copies) of a second target DNA.

The dynamic range (being indicative for the relative difference in concentrations of the two targets to become detected) was about 10^3 .

List of References

- Bernard, P.S., et al., Analytical Biochemistry 255 (1998) 101-107
EP 0 640 828
EP 0 747 447
- 5 Higuchi R, C Fockler G Dollinger and R Watson, Kinetic PCR analysis: real time monitoring of DNA amplification reactions, Bio/Technology 11 (1993) 1026-1030
- Higuchi R, G Dollinger, PS Walsh and R. Griffith, Simultaneous amplification and detection of specific DNA sequences, Bio/Technology 10 (1992) 413-417
- 10 Matthews, J.A., and Kricka, L.J., Analytical Biochemistry 169 (1988) 1-25
Pedersen, S., Bioradiations 107 (2001) 10-11
Review: Resonance Energy Transfer Editors: Meer, B Wieb van der et al., VCH Publishers INC., 1994
- 15 US 5,118,801
US 5,538,848
US 5,750,409
US 6,015,674
US 6,174,670
- 20 US 6,197,520
US 6,369,893
WO 02/14555
WO 97/46707
WO 97/46712
- 25 WO 98/46714

- 37 -

Patent Claims

1. A composition or reaction mixture for performing multi-color real time PCR comprising at least 3, preferably 4-5 most preferably exactly 4 pairs of FRET hybridization probes, each pair of hybridization probes consisting of a FRET donor probe carrying a FRET donor moiety and a FRET acceptor probe carrying a FRET acceptor moiety having an emission maximum between 550 and 710 nm.
5
2. A composition or reaction mixture according to claim 1, wherein at least 3, preferably at least 4 and most preferably exactly 4 FRET donor moieties are identical.
10
3. A composition or reaction mixture according to claim 1, wherein all FRET donor moieties are identical.
4. A composition or reaction mixture according to claim 1, wherein at least 3, preferably at least 4 and most preferably exactly 4 FRET donor moieties are Fluorescein.
15
5. A composition or reaction mixture according to claim 1, wherein all FRET donor moieties are Fluorescein.
6. A composition or reaction mixture according to claims 1-5, wherein at least one additional FRET donor moiety is selected from a group consisting of Atto425 and WI343.
20
7. A composition or reaction mixture according to claims 1-6, wherein one FRET acceptor moiety is selected from a group consisting of LC-Red 705, Cy5.5, and JA286.
8. A composition or reaction mixture according to claims 1-7, wherein at least one, two or three FRET acceptor moieties are selected from a group consisting of Cy5, LC-Red 640, and LC-Red 610
25
9. A composition or reaction mixture according to claims 1-8,

- 38 -

wherein one FRET acceptor moiety is selected from a group consisting of Rh6G and TAMRA.

10. A system for performing multi-color real time PCR, comprising
- 5 • a real time PCR instrument, and
 - a composition or reaction mixture according to claims 1-9.
11. A system according to claim 10, characterized in that said real time PCR instrument comprises
- 10 • at least 1 light source, preferably an LED
 - at least 4 and preferably 5-6 fluorescent detector entities, each of said entities having central detection wavelengths which are distinct from each other by at least 25 and preferably at least 30 nm
- characterized in that said detector entities are capable of
- 15 • simultaneously detecting maximum fluorescence emission of at least 3, preferably 4 and most preferably 5 differently labeled FRET Hybridization Probe pairs,
 - simultaneously detecting maximum fluorescence emission of at least 2 differently labeled TaqMan hybridization probes, and
 - detecting maximum fluorescence emission of SybrGreenI
- 20 • means for heating and cooling
 - multiple reaction vessels for containing a reaction mixture.
12. A system according to claim 11, characterized in that said real time PCR instrument comprises exactly one light source.
13. A method for amplifying and detecting multiple target DNA sequences comprising
- 25 a) providing a composition or reaction mixture according to claims 1-9,

- 39 -

- b) subjecting said reaction mixture to a thermocycling protocol such that amplification of said multiple target sequences can take place,
- c) monitoring hybridization of each of said pairs of FRET hybridization probes at least once after a plurality of amplification cycles.

5

14. A method according to claim 13, wherein hybridization is monitored at least once in a temperature dependent manner.

15. A real time PCR instrument comprising

10

- at least 1 light source, preferably an LED
- 5-6 fluorescent detector entities, each of said entities having central detection wavelengths which are distinct from each other by at least 25 and preferably at least 30 nm,

characterized in that said detector entities are capable of

15

- simultaneously detecting maximum fluorescence emission of at least 3, preferably 4 and most preferably 5 differently labeled FRET Hybridization Probe pairs,
- simultaneously detecting maximum fluorescence emission of at least 2 differently labeled TaqMan hybridization probes, and
- detecting maximum fluorescence emission of SybrGreenI

20

- means for heating and cooling
- multiple reaction vessels for containing a reaction mixture.

25

16. A real time PCR instrument according to claim 15 comprising exactly one light source.

17. An instrument according to claim 15-16, characterized in that said central detection wavelengths are selected from a group of range of wavelengths, said group consisting of 520-540 nm, 545-565 nm, 570-590 nm, 600-620 nm, 630-650 nm, 660-680 nm, and 700-720 nm.

Fig. 1

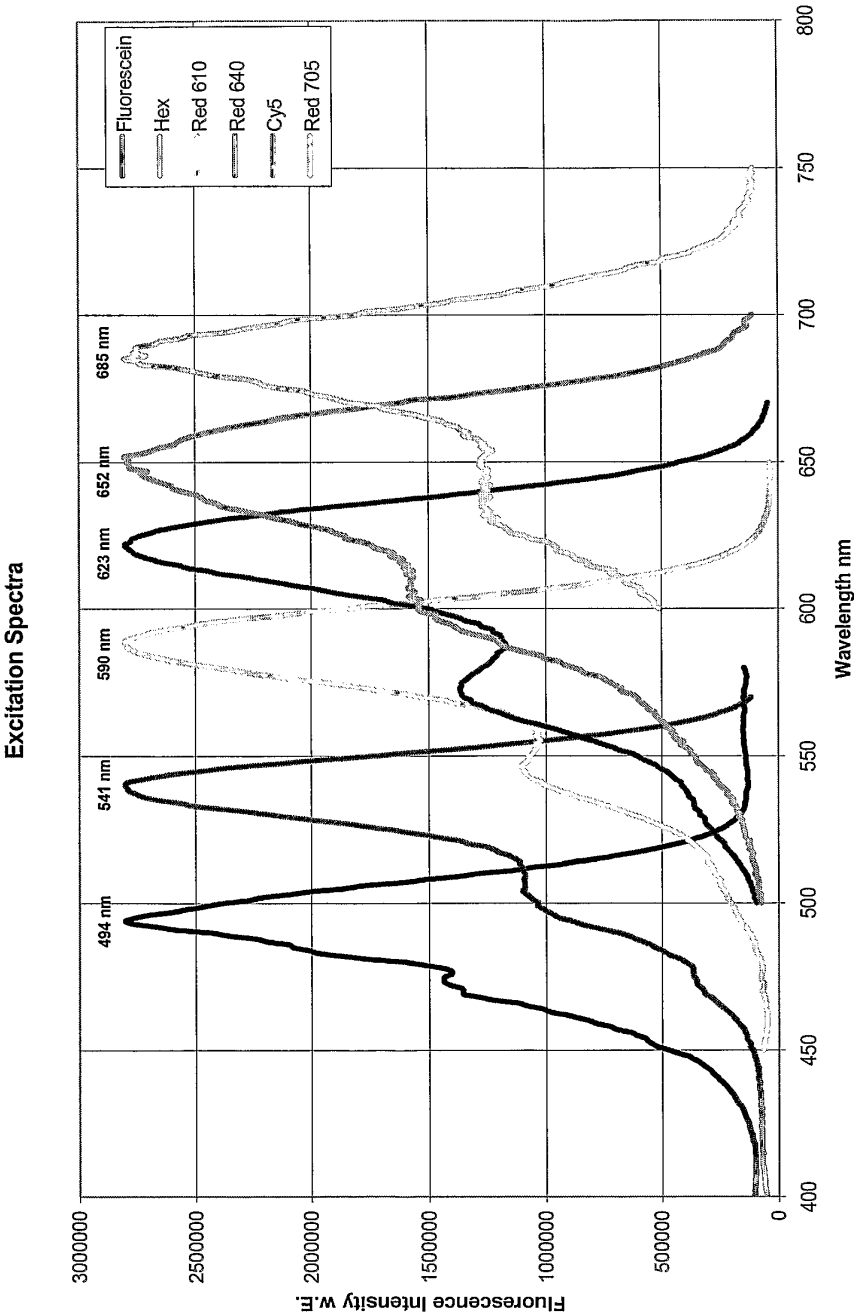


Fig. 2

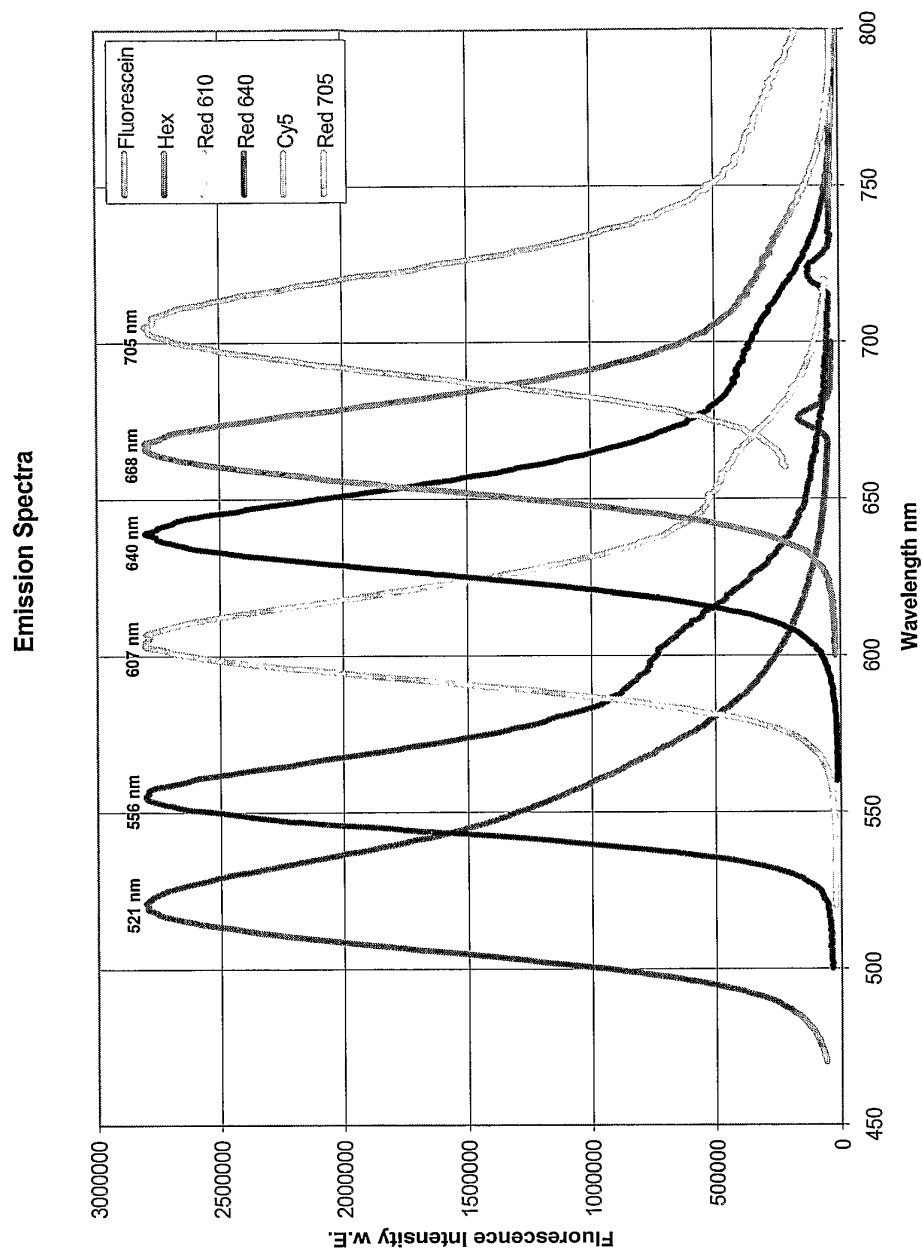
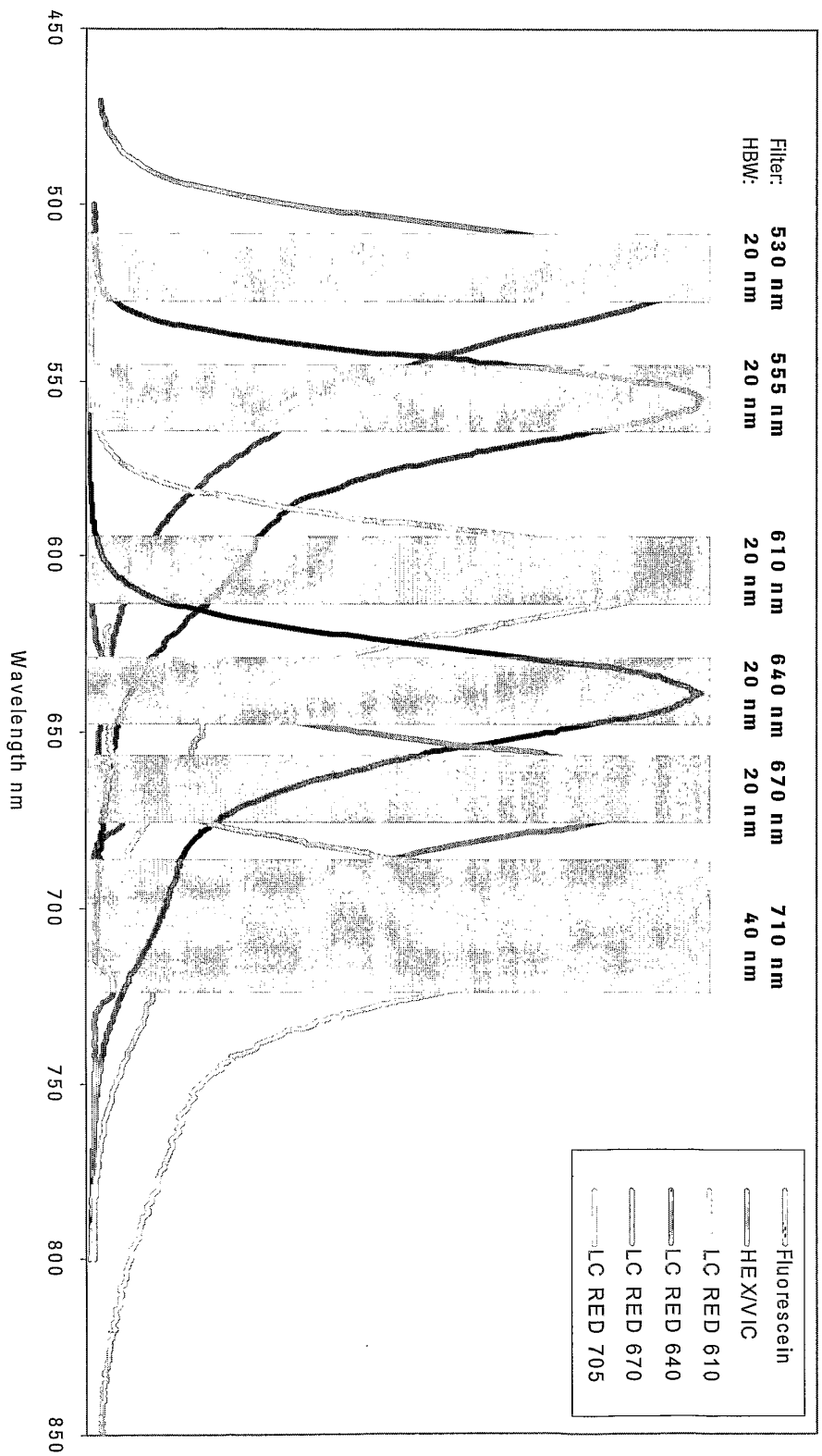


Fig. 3



WO 2004/087950

3/15

PCT/EP2004/003457

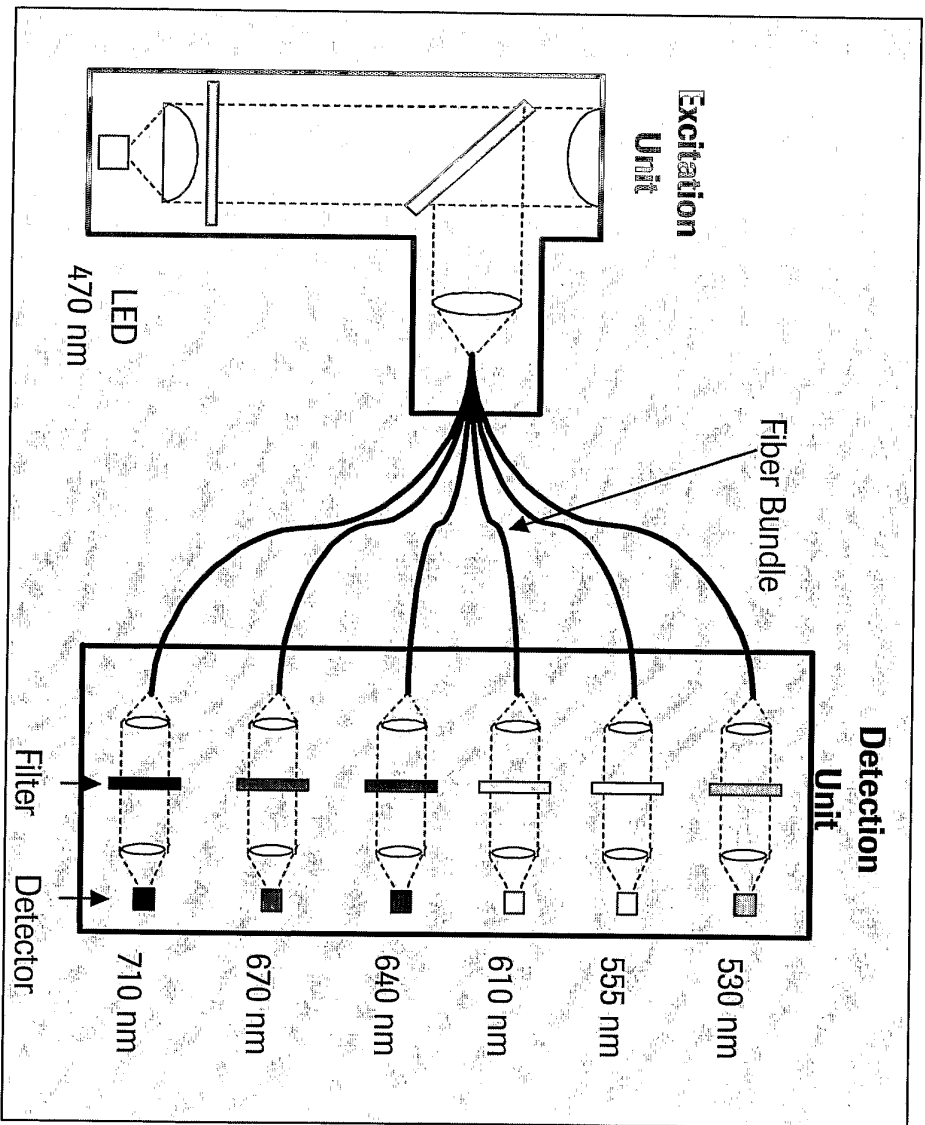


Fig. 4

WO 2004/087950

4/15

PCT/EP2004/003457

Fig. 5a**F3 = 610 nm**

Without Color Compensation

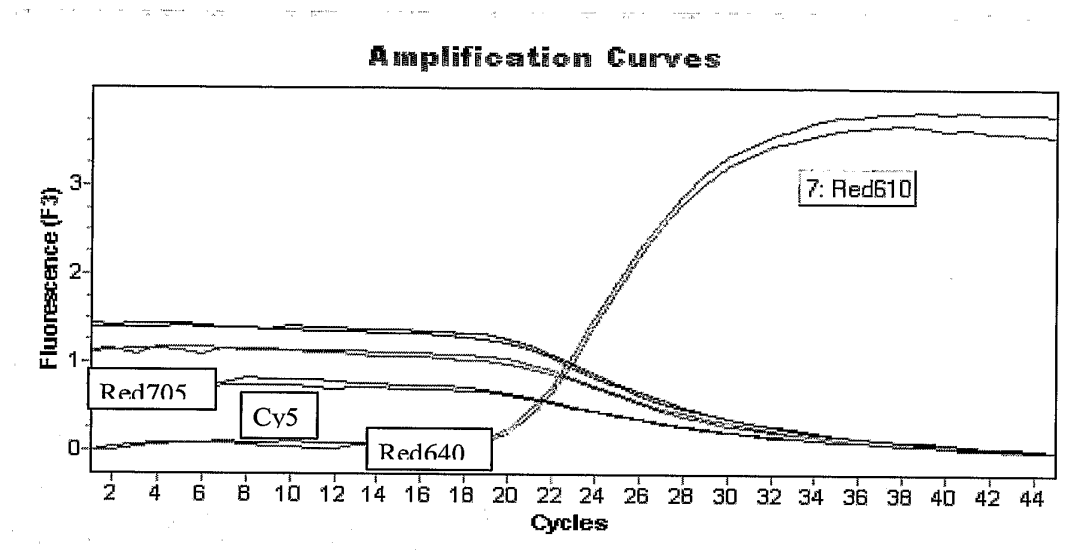
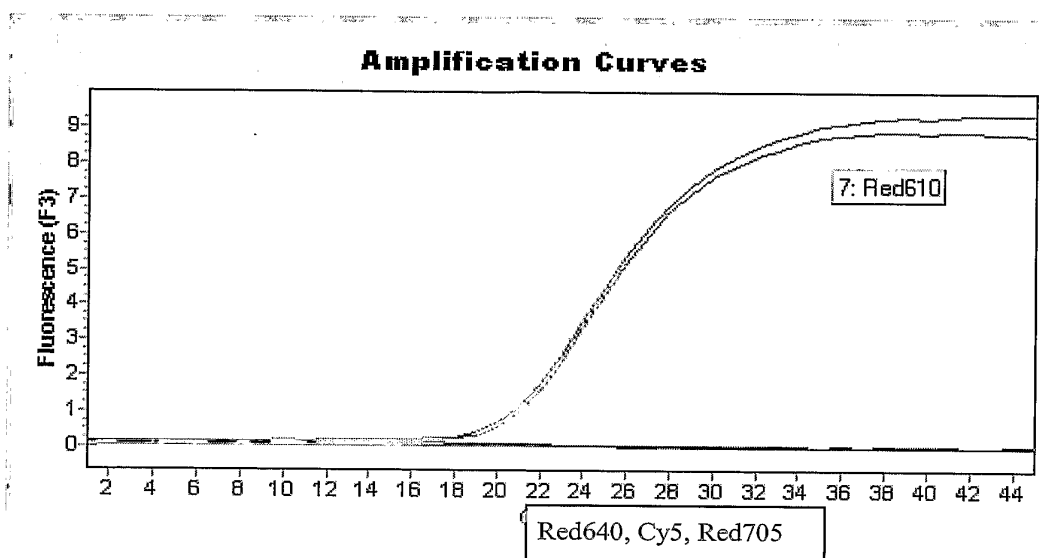
**With Color Compensation**

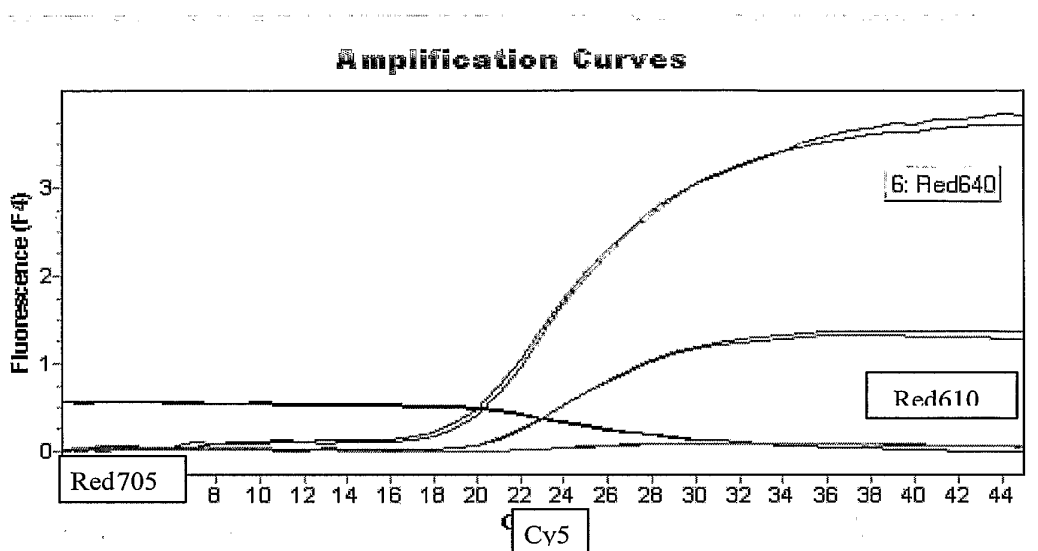
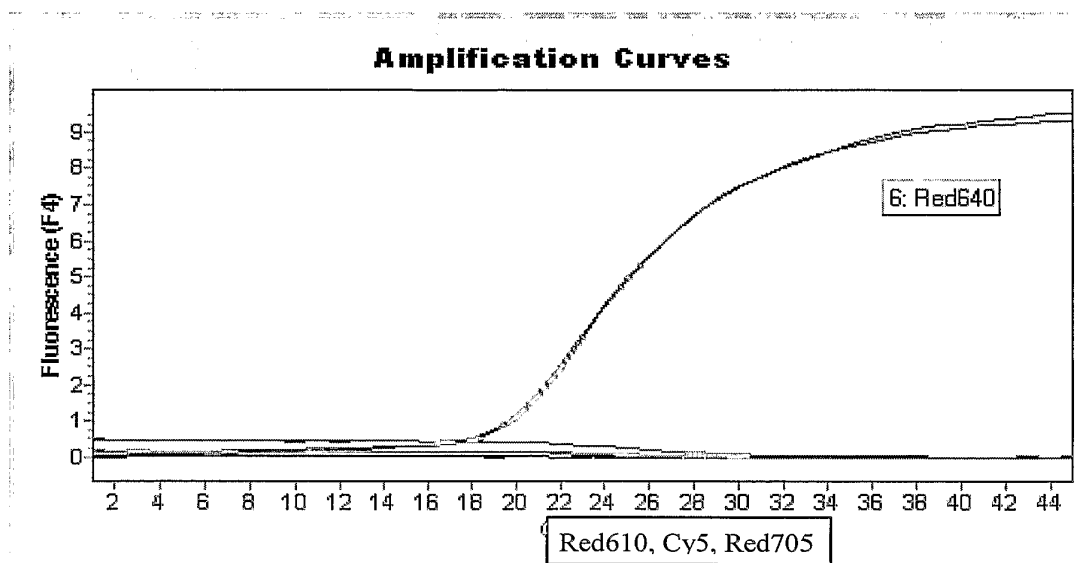
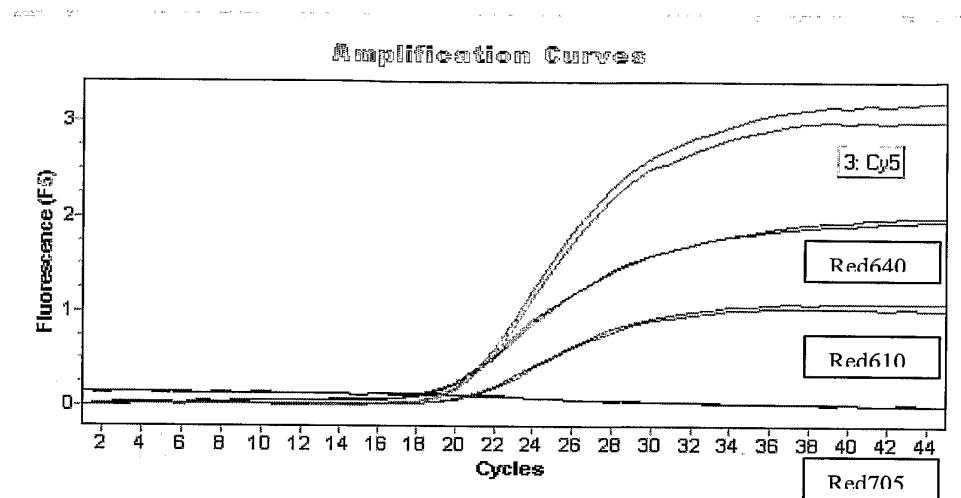
Fig. 5b**F4 = 640 nm****Without Color Compensation****With Color Compensation**

Fig. 5c

F5 = 670 nm
Without Color Compensation



With Color Compensation

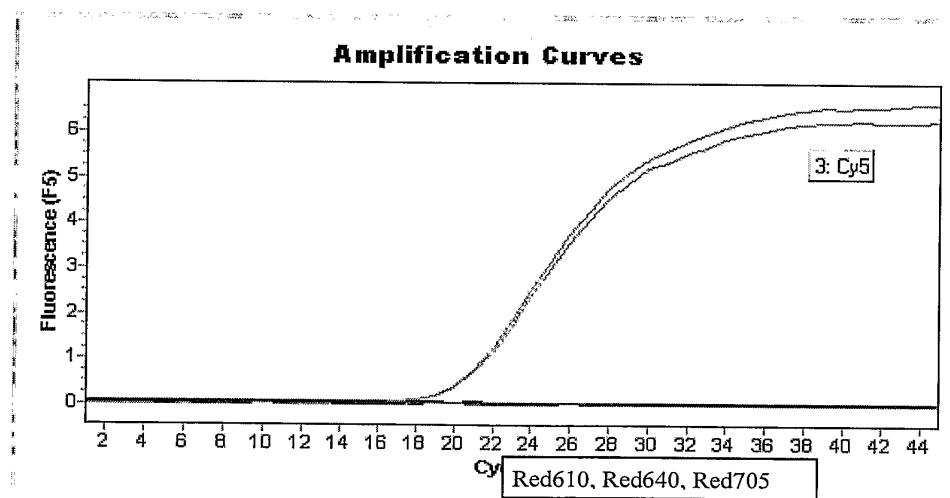
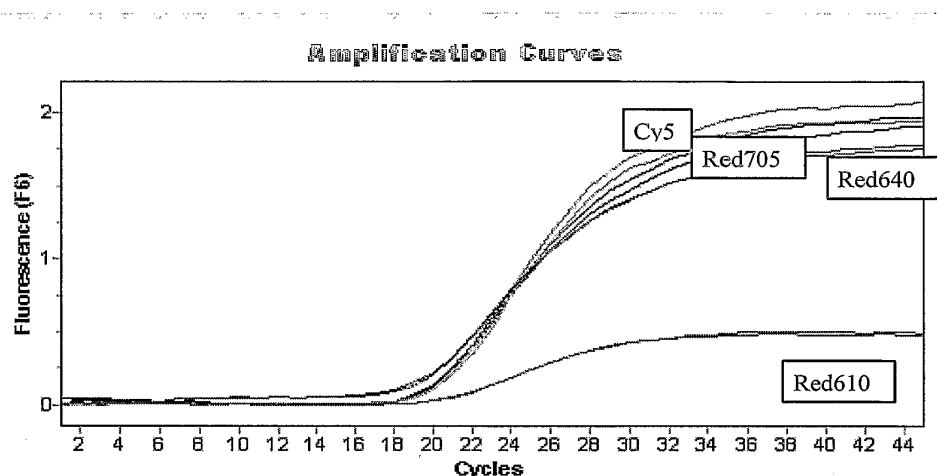


Fig. 5d

F6 = 710 nm
Without Color Compensation



With Color Compensation

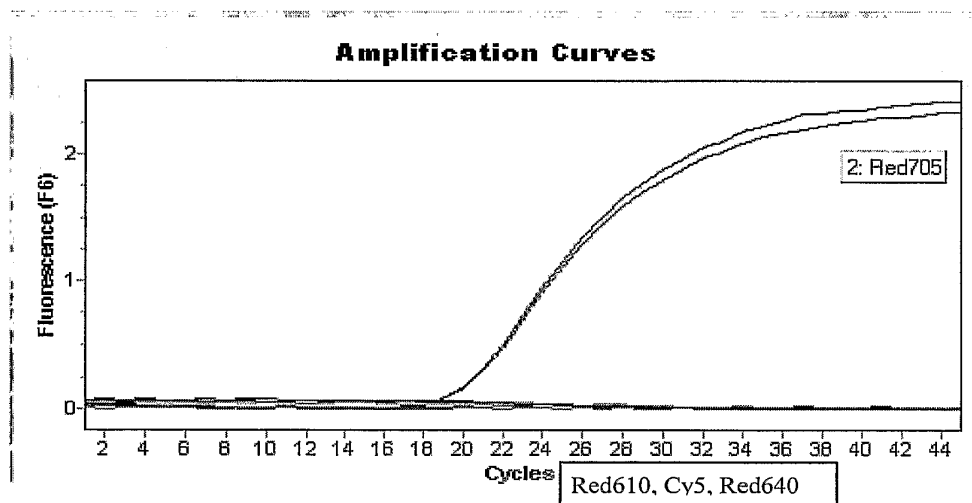


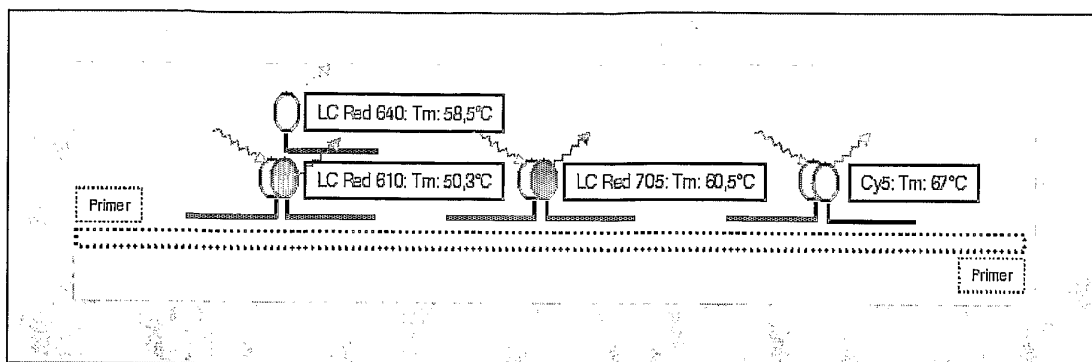
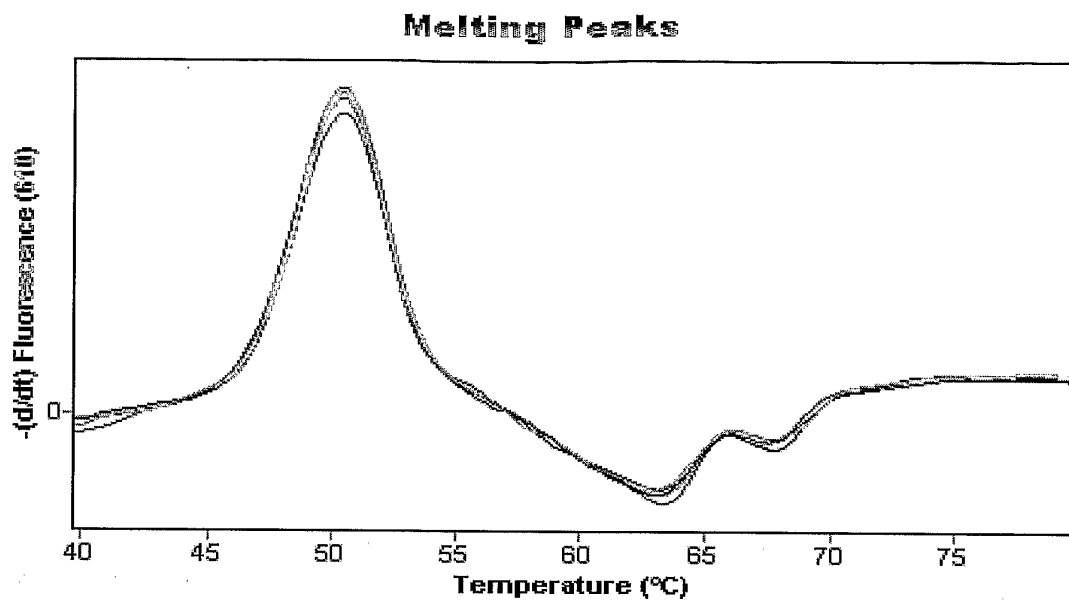
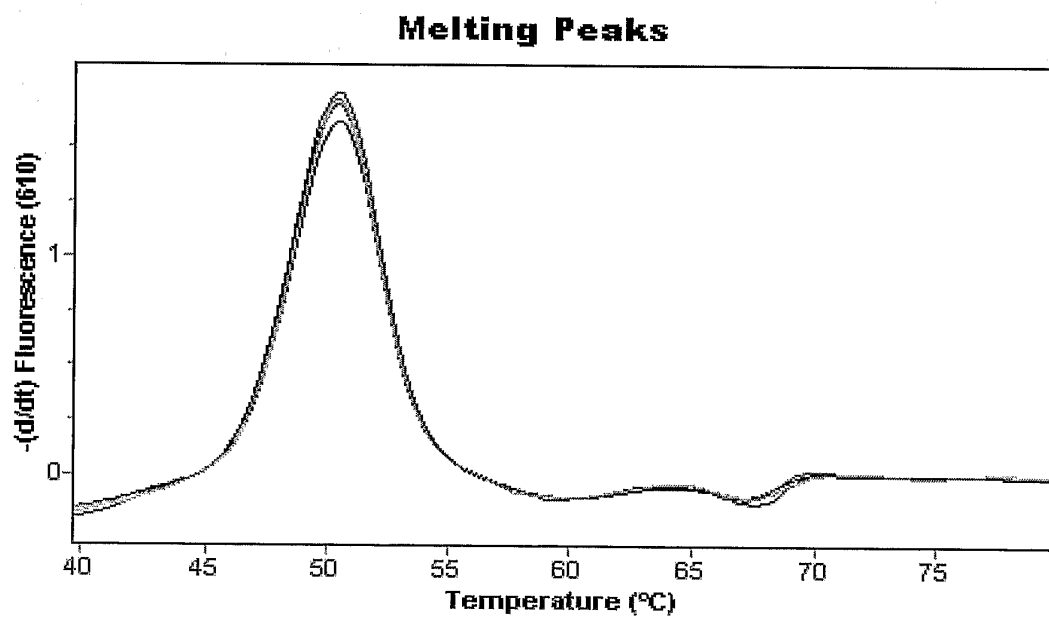
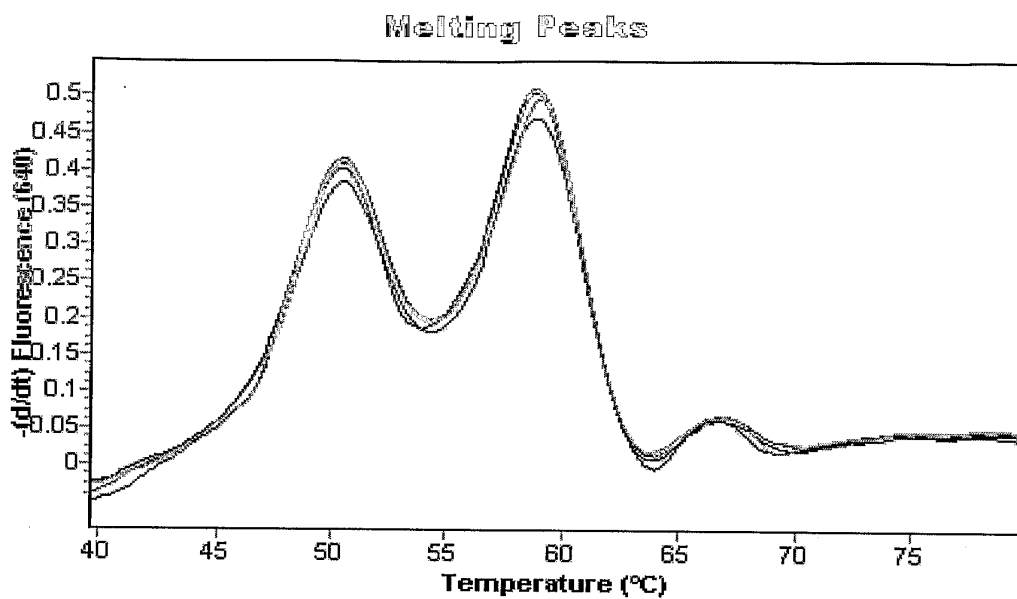
Fig. 6

Fig. 7a

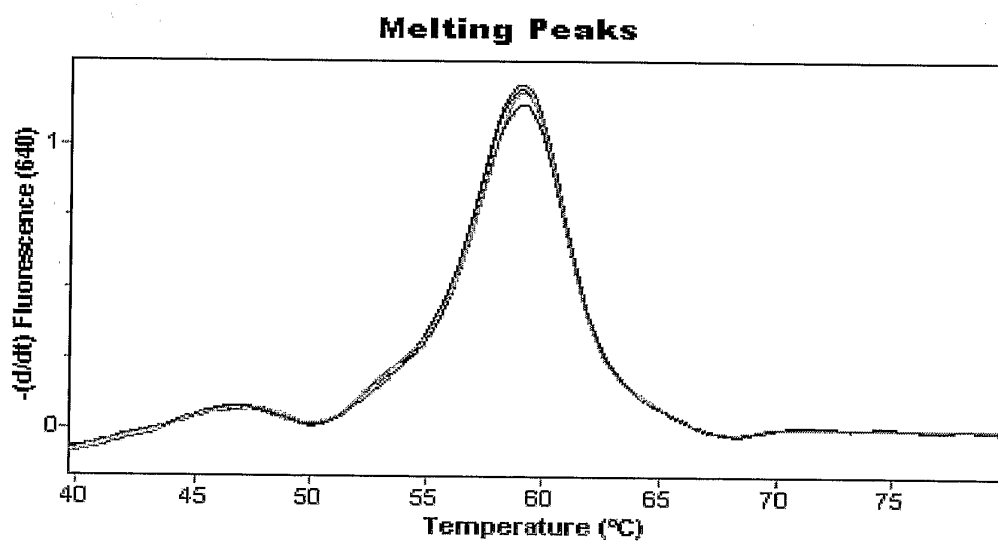
610 nm channel without color compensation



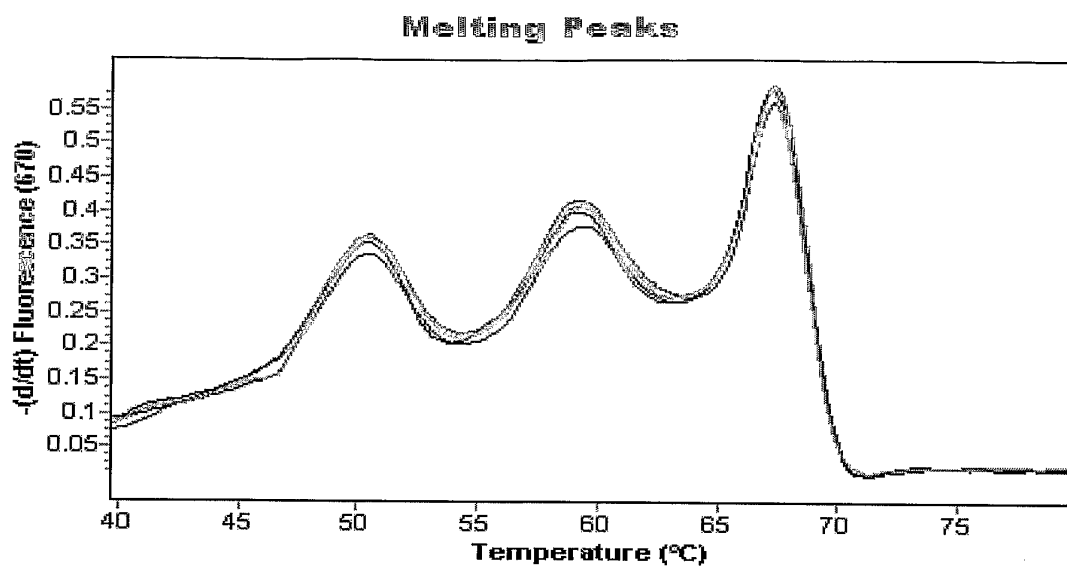
610 nm channel with color compensation

Fig. 7b

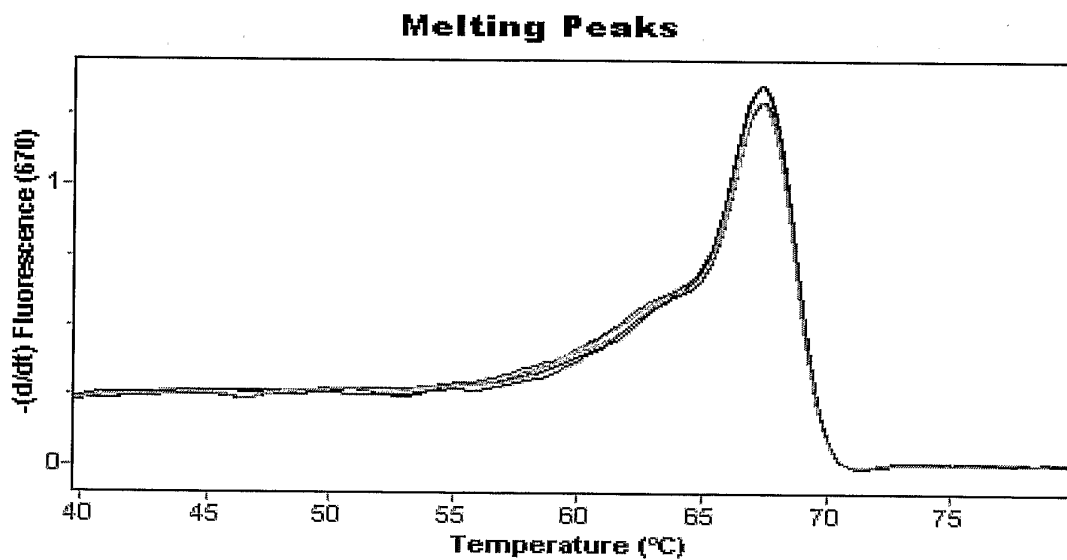
640 nm channel without color compensation



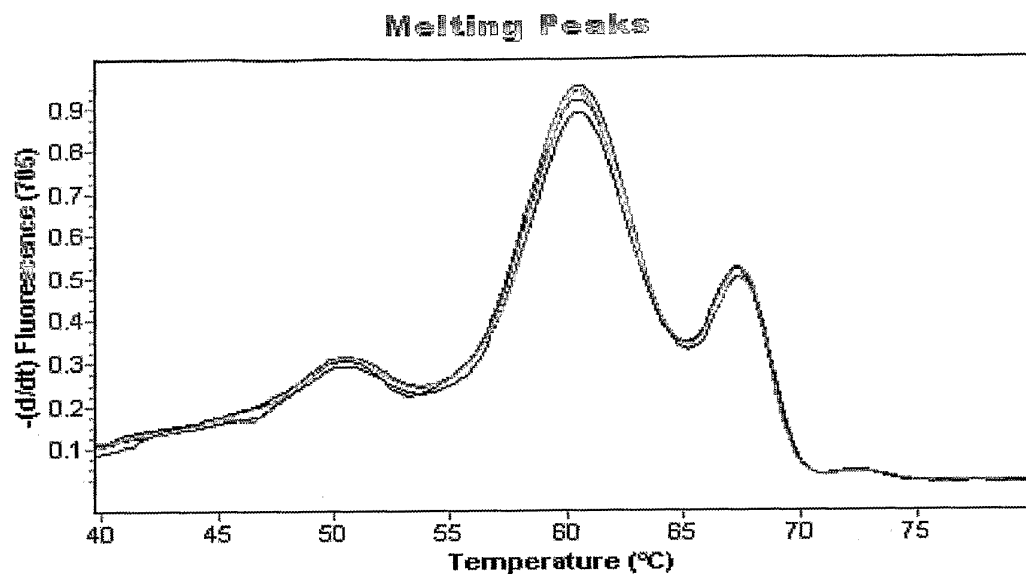
640 nm channel with color compensation

Fig. 7c

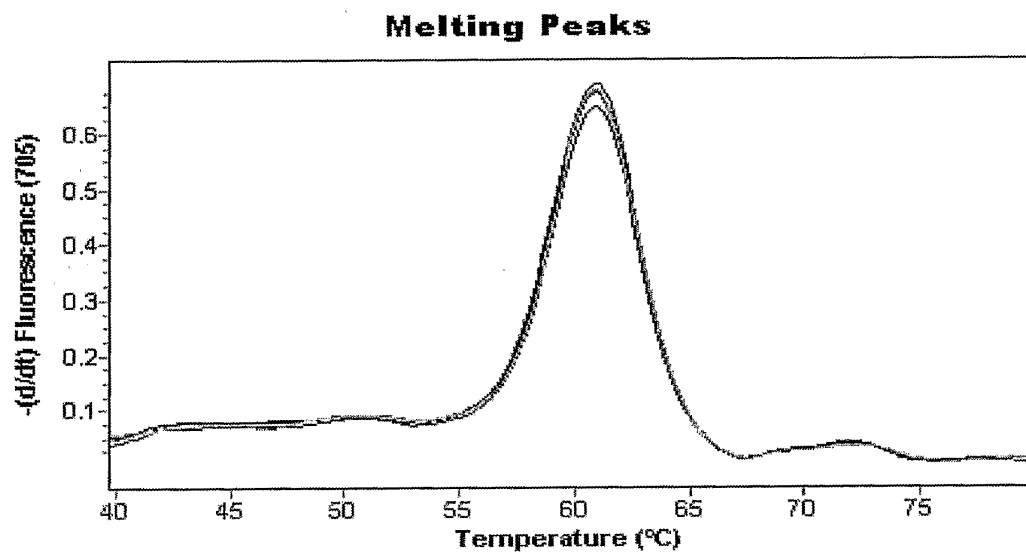
670 nm channel without color compensation



670 nm channel with color compensation

Fig. 7d

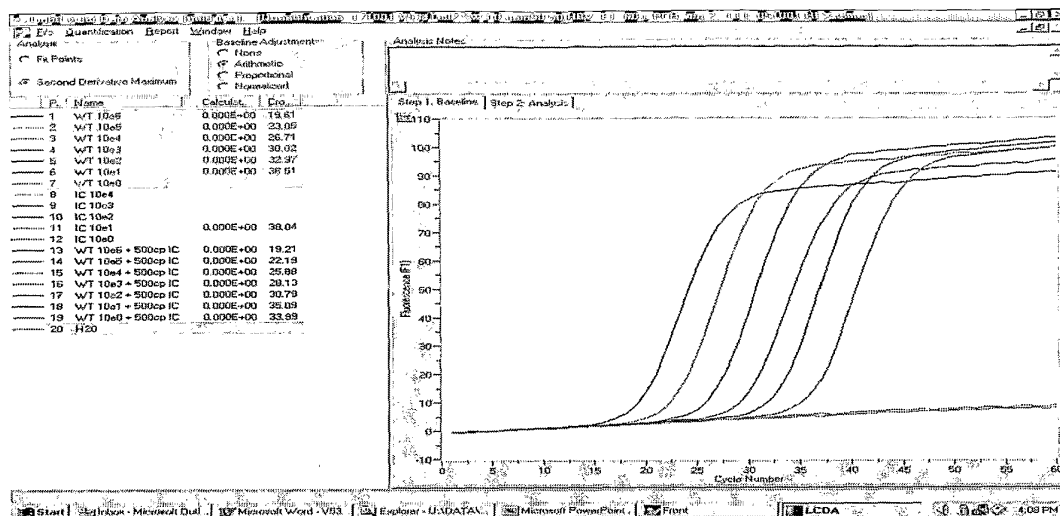
705 nm channel without color compensation



705 nm channel with color compensation

Fig. 8a

F1: 530nm (FAM)



F2: 560 nm (Hex)

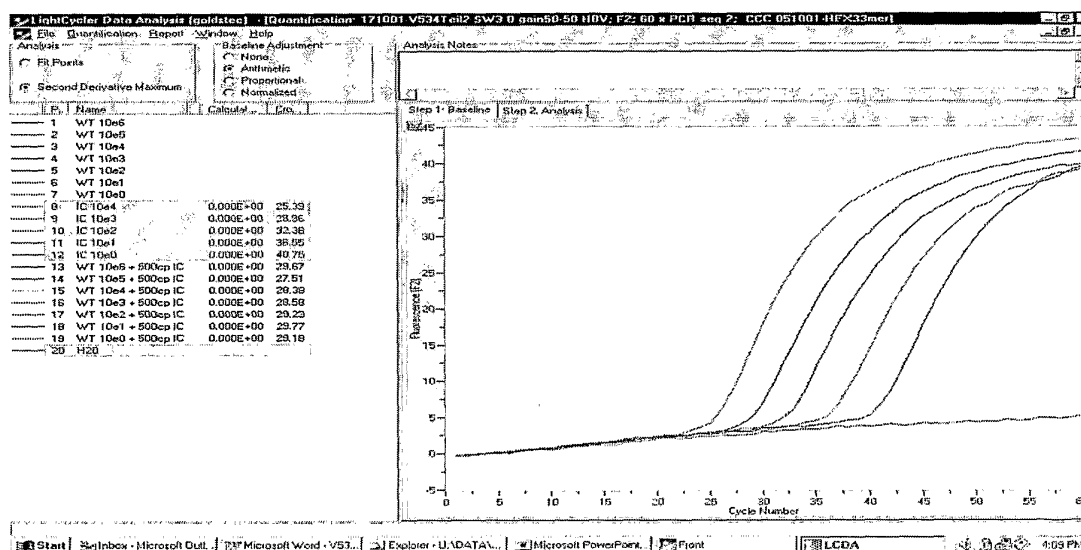
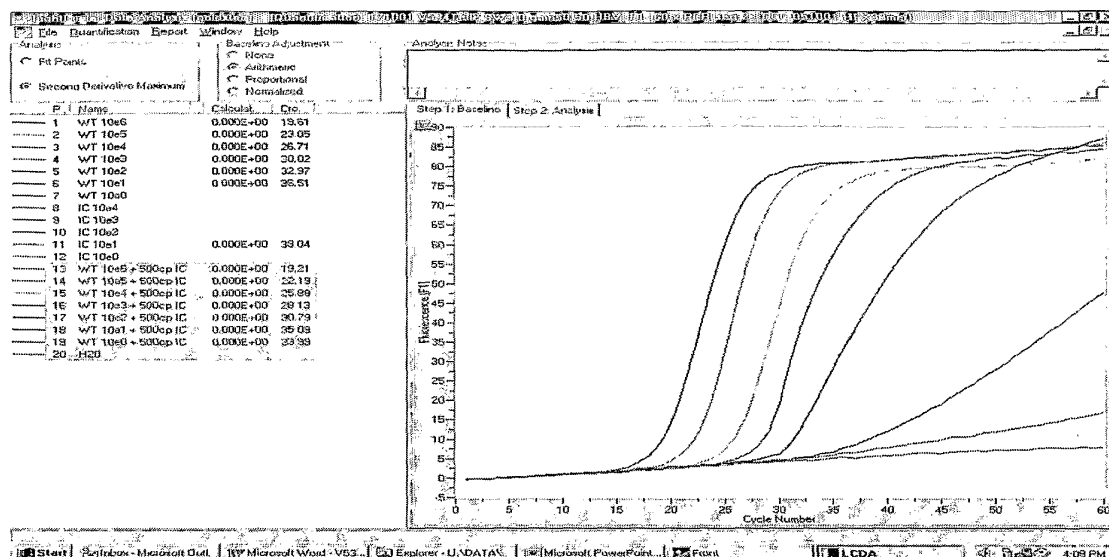
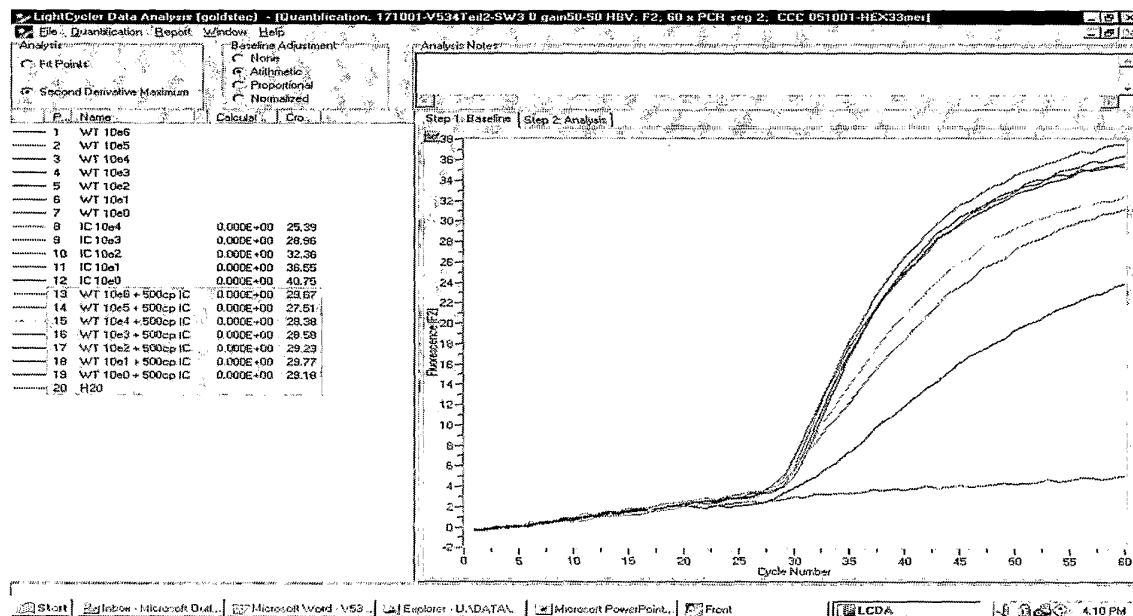


Fig. 8b

F1: 530nm (FAM)



F2: 560 nm (HEX)



- 1 -

SEQUENCE LISTING

<110> Roche Diagnostics GmbH
F. Hoffmann-La Roche AG

<120> Improved system for multi color real time PCR

<130> 21810 EP

<150> EP 03007458.7
<151> 2003-04-04

<150> EP 03014929.8
<151> 2003-08-07

<160> 13

<170> PatentIn version 3.1

<210> 1

<211> 22

<212> DNA

<213> Artificial

<220>

<221> misc_feature

<223> Primer FactorV, forward

<400> 1
gagagacatc gcctctgggc ta

22

<210> 2

<211> 20

- 2 -

<212> DNA

<213> Artificial

<220>

<221> misc_feature

<223> Primer FactorV, reverse

<400> 2

tggtatcaca ctggtgctaa

20

<210> 3

<211> 23

<212> DNA

<213> Artificial

<220>

<221> misc_feature

<223> 3'Fluorescein labeled Hybridization Probe

<400> 3

aatacctgta ttctctgcct gtc

23

<210> 4

<211> 36

<212> DNA

<213> Artificial

<220>

<221> misc_feature

<223> 5' Red610 Hybridization Probe

- 3 -

<220>

<221> misc_feature

<223> 5' Cy5 Hybridization Probe

<220>

<221> misc_feature

<223> 5' Red640 Hybridization Probe

<220>

<221> misc_feature

<223> 5' Red705 Hybridization Probe

<400> 4

agggatctgc tcttacagat tagaagtagt cctatt

36

<210> 5

<211> 19

<212> DNA

<213> Artificial

<220>

<221> misc_feature

<223> Primer forward

<400> 5

tgcccttgcat tttctgctt

19

<210> 6

<211> 19

<212> DNA

<213> Artificial

- 4 -

<220>

<221> misc_feature

<223> Primer reverse

<400> 6

gagttgggtg atacataca

19

<210> 7

<211> 32

<212> DNA

<213> Artificial

<220>

<221> misc_feature

<223> 3'-Fluorescein HybProbe 1

<400> 7

gaaattcttt gtttgtaata tactgctctc tc

32

<210> 8

<211> 18

<212> DNA

<213> Artificial

<220>

<221> misc_feature

<223> 5'-Red610 HybProbe 1a

<400> 8

tgatttggtc cacgtacc

18

- 5 -

<210> 9

<211> 19

<212> DNA

<213> Artificial

<220>

<221> misc_feature

<223> 5'-Red640 HybProbe 1b

<400> 9

tgatttggtc caagtaccc

19

<210> 10

<211> 37

<212> DNA

<213> Artificial

<220>

<221> misc_feature

<223> 3'-Fluorescein HybProbe 2

<400> 10

gttggagacg tctgcaggta tgtattcata gactcaa

37

<210> 11

<211> 20

<212> DNA

<213> Artificial

<220>

<221> misc_feature

- 6 -

<223> 5'-Red705 HybProbe 2

<400> 11
atcttcaatt gttcgagggt 20

<210> 12

<211> 28

<212> DNA

<213> Artificial

<220>

<221> misc_feature

<223> 3'-Fluorescein HybProbe 3

<400> 12
atttccttgg ggagaaatct cgtgccca 28

<210> 13

<211> 38

<212> DNA

<213> Artificial

<220>

<221> misc_feature

<223> 5'-RedCy5 HybProbe 3

<400> 13
acctggtgat gaatccctta ctatttagaa taaggaac 38

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
27 July 2006 (27.07.2006)

PCT

(10) International Publication Number
WO 2006/079049 A2

(51) International Patent Classification:
C12Q 1/68 (2006.01)

(21) International Application Number:
PCT/US2006/002393

(22) International Filing Date: 23 January 2006 (23.01.2006)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/645,696 21 January 2005 (21.01.2005) US

(71) Applicant (for all designated States except US): **THIRD WAVE TECHNOLOGIES, INC.** [US/US]; 502 South Rosa Road, Madison, WI 53719 (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **PETERSON, Patrick** [US/US]; 6716 Schroeder Road #3, Madison, WI 53711 (US). **ALLAWI, Hatim, Taysir** [JO/US]; 2325 Quartz Lane, Madison, WI 53719 (US). **LYAMICHEV, Victor** [RU/US]; 2523 Carriedale Court, Madison, WI 53711 (US). **LAW, Scott, M.** [US/US]; 438 Mineau Parkway, Madison, WI 53711 (US). **ELAGIN, Vecheslav, A.** [RU/US]; 1208 Wenzel Way, Waunakee, WI 53597 (US).

(74) Agents: **BROW, Mary, Ann, D.** et al.; Medlen & Carroll, LLP, 101 Howard Street, Suite 350, San Francisco, CA 94105 (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, KM, KN, KP, KR, KZ, 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, RU, SC, SD, SE, SG, SK, SL, SM, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, 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, 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:

— 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.

(54) Title: METHODS AND COMPOSITIONS FOR INCREASED DYNAMIC RANGE DETECTION OF NUCLEIC ACID MOLECULES

(57) Abstract: The present invention provides systems, methods and kits for increasing the dynamic range of detection of a target nucleic acid in a sample. In particular, the present invention provides methods and kits for increasing the dynamic range of detection of a target nucleic acid in a sample through the use of one or more probe oligonucleotides (e.g., analyte-specific probe oligonucleotides).



WO 2006/079049 A2

**COMPOSITIONS AND METHODS FOR INCREASED DYNAMIC RANGE
DETECTION OF NUCLEIC ACIDS**

5 The present invention claims priority to U.S. Patent Application Serial Number 60/645,696, filed January 21, 2005, the disclosure of which is herein incorporated by reference in its entirety.

FIELD OF THE INVENTION

10 The present invention provides systems, methods and kits for increasing the dynamic range of detection of a target nucleic acid in a sample. In particular, the present invention provides methods and kits for increasing the dynamic range of detection of a target nucleic acid in a sample through the use of one or more probe oligonucleotides.

BACKGROUND

15 All nucleic acid detection systems that rely on amplification of either the target being detected or the signal being generated inherently possess a dynamic range that limits their usefulness. At low concentrations of the target being detected, the signal generated is too low to detect or too low to be scored above background levels, and therefore is below the limit of detection, i.e., outside the dynamic range of the detection system. By contrast, at very high levels of the target being generated, the components of the detection system are exhausted such that the signal is said to be saturated, i.e. addition of still more target results in no increase in signal. In these cases, the quantity of target is said to be above the limit of detection, i.e. outside the dynamic range of the detection system.

25 In the real-world case of detection systems being used to detect targets from biological specimens, the range of target present in the sample being detected can be quite large, and is often either below or above the limit of detection of the system in use. Therefore, previous attempts to cover larger ranges of target concentration have required the generation of more than one detection system, to be used separately, that are
30 optimized for a given dynamic range. Because the quantity of target nucleic acid in the specimen is by definition an unknown quantity, this very frequently requires the use of

multiple detection systems sequentially to finally use the appropriate detection system that possesses the appropriate dynamic range for the specimen under examination.

As such, a single detection system with a broader dynamic range, if it was available, would significantly reduce costs, decrease labor time, and decrease expenditure of the specimen being examined. Even more, a method of increasing the dynamic range of an existing detection system would greatly aid the field of detection of targets within biological specimens generally.

SUMMARY OF THE INVENTION

The present invention provides systems, methods and kits for increasing the dynamic range of detection of a target nucleic acid in a sample. In particular, the present invention provides systems, methods and kits for increasing the dynamic range of detection of a target nucleic acid in a sample through the use of one or more probe oligonucleotides (e.g., analyte-specific probe oligonucleotides).

For example, in some embodiments, the present invention provides compositions, kits, and methods of quantitating nucleic acid targets (e.g., viral pathogens) using multiple probes that bind to a target nucleic acid at different strengths. In some embodiments, groups of probes are used in which each probe exhibits different binding affinities to the target sequence (e.g., by altering complementarity, length, concentration, additives, etc.). The use of multiple probes with different properties allows for an increase in the dynamic range of detection assays. In some embodiments, the multiple probes are used in invasive cleavage assays.

Accordingly, in some embodiments, the present invention provides a method for detecting the presence of, absence of, or amount of a target nucleic acid in a sample, comprising: incubating a sample suspected of containing a target nucleic acid with a plurality of first probe oligonucleotides and a plurality of second probe oligonucleotides, wherein each of the first and second probe oligonucleotides comprises an analyte specific region, wherein the plurality of second probe oligonucleotides are configured to occupy a probe hybridization site on the target nucleic acid at a different frequency than the plurality of first probe oligonucleotides; and measuring hybridization of the first and said second probe oligonucleotides over time, thereby measuring the amount of the target

nucleic acid. In some embodiments a plurality of third, fourth, fifth, etc. probe oligonucleotides are used. These additional oligonucleotides may be configured to bind to the same analyte-specific region of a target nucleic acid or may bind to different analyte-specific regions of the same or different target nucleic acids (e.g., the third and
5 fourth probes are configured to hybridize to a second analyte-specific region of the same target nucleic acid such that the third probe occupies the hybridization site at a different frequency than the fourth probe).

In some embodiments, the analyte specific regions of the first probe oligonucleotides are completely complementary to the target nucleic acid. In some
10 embodiments, the analyte specific regions of the second probe oligonucleotides are partially complementary to the target nucleic acid (e.g., contain a single mismatch). In some embodiments, the second probe oligonucleotide is shorter in length than the first probe oligonucleotide (e.g., by one, two, three, or four or more nucleotides). In some
15 embodiments, the second probe oligonucleotides are present at at least a 5 fold, and preferably at least a 10 fold lower concentration than the first probe oligonucleotides. In some embodiments, the second probe oligonucleotides are present at at least a 20 fold (e.g., 100 fold, 500 fold, 1000 fold, 10,000 fold, etc. lower concentration than the first
20 probe oligonucleotide). Where three or more probes of different concentrations are used, each probe may be separated by at at least 5 fold (10 fold, 20 fold, 100 fold, etc.) concentration from one another (e.g., a third probe 10000 fold more than a first probe and a second probe 100 fold more than a first probe). In some embodiments, one of the
25 mixtures comprises an agent known to increase or decrease hybridization efficiency (e.g., a charge tag, minor groove binding agent, or an intercalating agent). In other embodiments, one of the probes comprises one or more modified bases (e.g., amino T, indole, or nitropyrrole). In some embodiments, the analyte specific region of second
30 probe oligonucleotide is shorter than the analyte specific region of the first probe oligonucleotide (e.g., by one or more nucleotides). In other embodiments, the analyte specific region of the second probe oligonucleotide comprises increased secondary structure relative to the analyte specific region of the first probe oligonucleotide. In certain embodiments, the first probe oligonucleotides further comprise a non-analyte
specific region, wherein the non-analyte specific region comprises one or more

nucleotides that are not complementary to the target nucleic acid. In some embodiments, each of the second probe oligonucleotides further comprises a non-analyte specific region, wherein the non-analyte specific region comprises one or more nucleotides that are not complementary to the target nucleic acid. In some embodiments, incubating the sample with the second probe oligonucleotides comprises incubating the sample with competitor oligonucleotides, wherein the competitor oligonucleotides each comprise a region that is complementary to the non-analyte specific regions of the second probe oligonucleotides. The present invention is not limited by the nature of the competitor. The competitor may be a second target nucleic acid or a different region of the first oligonucleotide where, for example, hybridization of the non-analyte specific region of the second probe to the competitor does not generate a detectable event or generates a detectable event that is distinguishable from the detectable event generated by the first and/or second probes hybridizing to the analyte-specific region.

In some embodiments, incubating the sample with the second probe oligonucleotides comprises incubating the sample with competitor oligonucleotides, wherein the competitor oligonucleotides each comprise a region that is complementary to the non-analyte specific regions of the second probe oligonucleotides. In certain embodiments, one of the mixtures comprises altered reaction conditions that alter hybridization efficiency of a probe (e.g., altered pH, buffer, ionic strength or additional compositions (e.g., crowding agents)).

In some embodiments, the sample is a sample from an animal (e.g., a human) comprising blood, serum, stool, urine, or lymph known to or suspected of comprising a target nucleic acid (e.g., a virus or a bacterium). In some embodiments, the sample comprises a purified sample of nucleic acid (e.g., total DNA or RNA from a tissue, fluid or cell; genomic DNA; etc.). In some embodiments, the target nucleic acid is from a virus (e.g., human immunodeficiency virus (HIV) and other retroviruses, hepatitis C virus (HCV), hepatitis B virus (HBV), hepatitis A virus (HAV), human cytomegalovirus, (CMV), herpes simplex virus (HSV), Epstein bar virus (EBV), varicella zoster virus (VZV), human papilloma virus (HPV), bacteriophages (e.g., phage lambda), influenzaviruses, adenoviruses, or lentiviruses) or a bacterium (e.g., *Chlamydia sp.*, *N. gonorrhea*, or group B *streptococcus*). In other embodiments, the sample is from a plant.

For example, in some embodiments, the plant is infected with or suspected of being infected with a pathogenic microorganism (e.g., a fungus, a virus, or a bacteria).

In some embodiments, the methods of incubating the sample with the first and second probe oligonucleotides occur in the same reaction vessel (e.g., the first and second probe oligonucleotides are mixed in solution in the same reaction vessel). In some embodiments, the first and second probe oligonucleotides comprise labels. In some embodiments, the first and second labels are different from each other. In some embodiments, the first and second labels are the same label. In some embodiments, measuring the hybridization of the first and second probe oligonucleotides comprises performing an invasive cleavage structure type assay (e.g., an INVADER assay). In some such embodiments, the probes are unlabeled, but comprise a flap sequence that is removed from the probe upon cleavage during the invasive cleavage assay. In some embodiments, the removed flaps are configured to hybridize to a FRET cassette to trigger a detection reaction. In some embodiments, the first and second probes report to the same FRET cassette (e.g., the first and second probe generate identical flaps upon cleavage in the primary invasive cleavage reaction). In other embodiments, determining the amount of the target nucleic acid comprises performing a detection assay including, but not limited to, a hybridization assay, any real-time amplification assay that involves hybridization, a TAQMAN assay, SNP-IT assay, a Southern blot, a ligase assay, a microarray assay, a FULLVELOCITY assay, a cycling probe assay, NASBA, branched DNA assay, TMA, methods employing molecular beacons, capillary electrophoresis detection methods, microfluidic detection methods, and the like.

In other embodiments, the present invention provides a method for detecting the presence of, absence of, or amount of a target nucleic acid in a sample, comprising: providing a sample containing or suspected of containing a target nucleic acid; a first probe oligonucleotide comprising an analyte specific region and a first label, wherein the analyte specific region of the first probe oligonucleotide is completely complementary to the target nucleic acid; and a second probe oligonucleotide comprising an analyte specific region and a second label, wherein the analyte specific region of the second probe oligonucleotide is partially complementary to the target nucleic acid; and exposing the

sample to the first and second probe oligonucleotides; and, in some embodiments, determining the amount of the target nucleic acid in the sample.

The present invention further provides a kit comprising reagents and, in some embodiments, instructions, for performing the detection assays of the present invention. For example, in some embodiments, the present invention provide a kit for detecting the presence of, absence of, or quantitation of target nucleic acids in a sample, comprising: a plurality of first probe oligonucleotides comprising a first analyte specific region and, optionally, a first label, and a plurality of second probe oligonucleotides comprising a second analyte specific region and, optionally, a second label, wherein the second probe oligonucleotides are configured to occupy a probe hybridization site on the target nucleic acid at a different frequency than the first mixture of probe oligonucleotides; and reagents for performing an INVADER assay using the first and second probe oligonucleotides. In some embodiments, the analyte specific regions of the first probe oligonucleotides are completely complementary to the target nucleic acid. In other embodiments, the analyte specific regions of the second probe oligonucleotides are partially complementary to the target nucleic acid (e.g., contain one or more mismatches with the target nucleic acid). In still further embodiments, the second probe oligonucleotides are present at a lower concentration than the first probe oligonucleotides. In some embodiments, the kit further comprises instructions for using the kit for performing a nucleic acid detection assay. In some embodiments, the kit comprises reagents and/or instructions for use of the methods of the present invention with a one or more different detection assay technologies (e.g., an invasive cleavage assay (e.g., INVADER assay), a TAQMAN assay, SNP-IT assay, etc.)).

In some embodiments, the present invention provides methods for detecting a target nucleic acid, comprising: a) amplifying a target nucleic acid at two different levels of amplification to generate amplification products; b) hybridizing the amplification products to a first probe and second probe, wherein the first probe hybridizes to the amplification products at a different frequency than the second probe. In certain embodiments, the second probe is present at a 10-fold lower concentration than the first probe. In other embodiments, the at least two probes bind to the same sequence.

In additional embodiments, the present invention provides methods for detecting a target nucleic acid in a plurality of samples over a broad dynamic range, comprising: exposing a first sample having less than 10^3 copies of target nucleic acid and a second sample having greater than 10^5 copies of target nucleic acid to a set of reagents under conditions such that the target nucleic acid in the first and second samples is detected, wherein method comprises exposing each of the first and second samples to a first probe and a second probe, wherein the second probe hybridizes to the target nucleic acids at a different frequency than the first probe. In particular embodiments, the target nucleic acid in the first and second samples is quantitated. In further embodiments, the second probe is present at a 10-fold lower concentration than the first probe. In some embodiments, the target nucleic acids are treated under two or more different amplification conditions prior to detection. In other embodiments, the method is conducted without any amplification of the target nucleic acid.

In some embodiments, the present invention provides methods for detecting a target nucleic acid, comprising: a) amplifying a target nucleic acid to generate amplification products; b) contacting the amplification products with first and second probes, wherein the second probe hybridizes to the amplification products at a different frequency than the first probe; c) cleaving the first and second probes; and d) detecting the cleavage of the first and second probes.

In other embodiments, the present invention provides kits comprising: a polymerase, a 5' nuclease, and two probes configured to hybridize to an analyte-specific region of a target nucleic acid, wherein the second probe hybridizes to the analyte-specific region at a different frequency than the first probe oligonucleotide, and wherein the first and second probes are configured to both directly or indirectly generate a detectable signal in the presence of the target nucleic acid. In some embodiments, the first and second probes generate the same type of detectable signal. In certain embodiments, the first and second probes each comprise a flap sequence that is complementary to a FRET cassette. In other embodiments, the flap of the first probe is identical to the flap of the second probe.

In some embodiments, the present invention provides methods for detecting a target nucleic acid in a sample comprising: a) contacting a sample suspected of

containing a target nucleic acid with amplification reagents such that, if the target nucleic acid is present: i) a first region of the target nucleic acid is either not amplified, or is amplified at a first level to generate plurality of first product sequences; and ii) a second region of the target nucleic acid is amplified at a second level to generate a plurality of second product sequences, wherein the second level of amplification is greater than the first level of amplification (e.g. such that the second product sequences are present at a level of at least 10-fold ... 100-fold ... 1000-fold ... 10,000-fold ... or 100,000-fold higher concentration after amplification than the target nucleic acid, or first product sequences if produced); and b) incubating the sample with a plurality of first and second probe oligonucleotides, wherein: i) the first and second probe oligonucleotides hybridize to the first region of the target nucleic acid, and the first product sequences if produced, at different frequencies, or ii) the first and second probe oligonucleotides hybridize to the second product sequences at a different frequency; and c) measuring hybridization of the first and second probe oligonucleotides thereby detecting the target nucleic acid in the sample. In particular embodiments, the second product sequences are present at a level between 100-fold and 100,000 fold higher concentration after amplification than the target nucleic acid, or first product sequences if produced. In certain embodiments, the target nucleic acid is micro-RNA.

In certain embodiments, the present invention provides methods for detecting a target nucleic acid in a plurality of samples over a broad dynamic range, comprising: exposing a first sample having less than 10^3 copies of target nucleic acid and a second sample having greater than 10^5 copies of target nucleic acid to a set of reagents under conditions such that the target nucleic acid in the first and second samples is detected, wherein the method comprises exposing each of the first and second samples to a first probe and a second probe, wherein the second probe hybridize to the target at different frequencies.

In particular embodiments, the present invention provides methods for detecting a target nucleic acid, comprising: a) linearly amplifying a first region of the target nucleic acid to generate linearly amplified amplification products; b) exponentially amplifying a second region of the target nucleic acid to generate exponentially amplified amplification products; c) hybridizing the linearly amplified amplification products with a first set of

probes and the exponentially amplified amplification products with a second set of probes, wherein either the first or the second set of probes comprises a first plurality of probes that hybridize to amplified target nucleic acid and a second plurality of probes that hybridize to amplified target nucleic acid at a different frequency than the first plurality of probes. In certain embodiments, both the first set and the second set of probes comprises a first plurality of probes that hybridize to amplified target nucleic acid and a second plurality of probes that hybridize to amplified target nucleic acid at a different frequency than the first plurality of probes.

In some embodiments, the present invention provides methods for detecting a target nucleic acid, comprising: a) amplifying a target nucleic acid both linearly and exponentially to generate amplification products; b) hybridizing the amplification products to at least two probes, wherein the first probe hybridizes to amplified target nucleic acid at a different frequency than the second probe. In certain embodiments, the first and second probes both hybridize to the same probe binding site on the target nucleic acid.

In certain embodiments, the present invention provides methods for detecting a target nucleic acid in a sample comprising: a) contacting a sample suspected of containing a target nucleic acid with amplification reagents such that, if the target nucleic acid is present: i) a first region of the target nucleic acid comprising a first probe hybridization site is either not amplified, or is amplified at a first level to generate a plurality of first product sequences that comprise the first probe hybridization site; and ii) a second region of the target nucleic acid is amplified at a second level to generate a plurality of second product sequences that comprise a second probe hybridization site, wherein the second level of amplification is greater than the first level of amplification (e.g. such that the second product sequences are present at a level of at least 10-fold ... 100-fold ... 1000-fold ... 10,000-fold ... or 100,000-fold higher concentration after amplification than the target nucleic acid, or first product sequences if produced); and b) incubating the sample with a plurality of first and second probe oligonucleotides, wherein: i) the first and second probe oligonucleotides occupy the first probe hybridization site on the first region of the target nucleic acid, and the first product sequences if produced, at different frequencies, or ii) the first and second probe

oligonucleotides occupy the second probe hybridization site on the second product sequences at a different frequency; and c) measuring hybridization of the first and second probe oligonucleotides thereby detecting the target nucleic acid in the sample. In particular embodiments, the second product sequences are present at a level between 100-
5 fold and 100,000 fold higher concentration after amplification than the target nucleic acid, or first product sequences if produced.

In other embodiments, the methods further comprise incubating the sample with a third probe oligonucleotide that occupies the first probe hybridization site on the first region of the target nucleic acid, and the first product sequences if produced, at a first
10 frequency, and measuring the hybridization of the third probe oligonucleotide. In some embodiments, the methods further comprise incubating the sample with a fourth probe oligonucleotide that occupies the first probe hybridization site on the first region of the target nucleic acid, and the first product sequences if produced, at a second frequency, wherein the second frequency is different from the first frequency, and measuring the
15 hybridization of the fourth probe oligonucleotide.

In certain embodiments, the methods further comprise incubating the sample with a third probe oligonucleotide that occupies the second probe hybridization site on the second product sequences at a first frequency, and measuring the hybridization of the third probe oligonucleotide. In particular embodiments, the methods further comprise
20 incubating the sample with a fourth probe oligonucleotide that occupies the second probe hybridization site on the second product sequences at a second frequency, wherein the second frequency is different from the first frequency, and measuring the hybridization of the fourth probe oligonucleotide.

In some embodiments, the first level of amplification is achieved by linear
25 amplification, and the second level is achieved is achieved with logarithmic amplification (e.g., polymerase chain reaction). In further embodiments, the first level of amplification is achieved with compromised amplification (e.g. using inefficient primers and/or inefficient polymerases). In other embodiments, the second level of amplification is at least 10-fold greater than no amplification or the first level of amplification. In certain
30 embodiments, the target nucleic acid is micro-RNA.

In some embodiments, the present invention provides methods for detecting a target nucleic acid in a sample, comprising; a) contacting a sample suspected of containing a target nucleic acid with amplification reagents such that, if the target nucleic acid is present: i) a first region of the target nucleic acid is amplified non-logarithmically to generate a plurality of non-logarithmically amplified sequences that comprise a first probe hybridization site, and ii) a second region of the target nucleic acid is amplified logarithmically to generate a plurality of logarithmically amplified sequences that comprise a second probe hybridization site; b) incubating the sample with a plurality of first probe oligonucleotides, a plurality of second probe oligonucleotides, and a plurality of third probe oligonucleotides, wherein each of the first, second, and third probe oligonucleotides comprises an analyte specific region, wherein the plurality of second probe oligonucleotides are configured to occupy the second probe hybridization site on the logarithmically amplified sequences at a different frequency than the plurality of first probe oligonucleotides, and wherein the third probe oligonucleotides are configured to occupy the first probe hybridization site on the non-logarithmically amplified sequences at a first frequency; and c) measuring hybridization of the first, second, and third probe oligonucleotides, thereby detecting the target nucleic acid in the sample. In other embodiments, the target nucleic acid is initially present in the sample in an amount between about 10^1 and about 10^8 molecules (e.g. the dynamic range of the methods extend over at least about seven orders of magnitude). In certain embodiments, the target nucleic acid is micro-RNA.

In certain embodiments, the measuring detects the amount of the target nucleic acid in the sample. In other embodiments, the measuring is conducted over time. In further embodiments, the plurality of logarithmically amplified sequences do not contain the first probe hybridization site.

In particular embodiments, the analyte specific regions of the first probe oligonucleotides are completely complementary to the second probe hybridization site of the second product sequence (e.g. logarithmically amplified sequences). In other embodiments, the analyte specific regions of the second probe oligonucleotides are partially complementary to the second probe hybridization site of the second product sequences (e.g., logarithmically amplified sequences).

In certain embodiments, the methods further comprise incubating the sample with a plurality of fourth probe oligonucleotides comprising an analyte specific region, wherein the fourth probe oligonucleotides are configured to occupy the first probe hybridization site on the first product sequences (e.g., non-logarithmically amplified sequences) at a second frequency which is different from the first frequency of the third probe oligonucleotides. In further embodiments, the target nucleic acid is initially present in the sample in an amount between about 10^1 and about 10^{10} molecules (e.g. the dynamic range of the methods extend over at least about nine orders of magnitude).

In further embodiments, the analyte specific regions of the third probe oligonucleotides are completely complementary to the first probe hybridization site of the first product sequences (e.g., non-logarithmically amplified sequences). In other embodiments, the analyte specific regions of the third probe oligonucleotides are partially complementary to the first probe hybridization site of the first product sequences (e.g., non-logarithmically amplified sequences). In additional embodiments, the analyte specific regions of the third oligonucleotides are identical to the analyte specific regions of the fourth oligonucleotides.

In some embodiments, the second probe oligonucleotides are present in at least a 5-fold lower concentration than the first probe oligonucleotides (e.g. 5-fold, 6-fold, 7-fold, 8-fold, or 9-fold lower concentration). In certain embodiments, the second probe oligonucleotides are present in at least a 10-fold lower concentration than the first probe oligonucleotides (e.g. 10-fold ... 15-fold ... 25-fold ... 50-fold ... 75-fold ... or 95-fold lower concentration, or any range between 10-fold and 100-fold). In particular embodiments, the second probe oligonucleotides are present in at least a 100-fold lower concentration than the first probe oligonucleotides (e.g. 100-fold ... 125-fold ... 150-fold ... 250-fold ... 500-fold ... 750-fold ... or 900-fold lower concentration, or any range between 100-fold and 1000-fold). In further embodiments, the second probe oligonucleotides are present in at least a 1000-fold lower concentration than the first probe oligonucleotides (e.g., 1000-fold ... 1100-fold ... 1300-fold ... 1500-fold ... 10,000-fold ... 15,000-fold ... 25,000-fold ... 100,000-fold ... 500,000-fold ... or 1,000,000-fold, or any range between 1000-fold and 1,000,000-fold).

In some embodiments, the third probe oligonucleotides are present in at least a 5-fold lower concentration than the fourth probe oligonucleotides (e.g. 5-fold, 6-fold, 7-fold, 8-fold, or 9-fold lower concentration). In certain embodiments, the third probe oligonucleotides are present in at least a 10-fold lower concentration than the fourth probe oligonucleotides (e.g. 10-fold ... 15-fold ... 25-fold ... 50-fold ... 75-fold ... or 95-fold lower concentration, or any range between 10-fold and 100-fold). In particular embodiments, the third probe oligonucleotides are present in at least a 100-fold lower concentration than the fourth probe oligonucleotides (e.g. 100-fold ... 125-fold ... 150-fold ... 250-fold ... 500-fold ... 750-fold ... or 900-fold lower concentration, or any range between 100-fold and 1000-fold). In further embodiments, the third probe oligonucleotides are present in at least a 1000-fold lower concentration than the fourth probe oligonucleotides (e.g., 1000-fold ... 1100-fold ... 1300-fold ... 1500-fold ... 10,000-fold ... 15,000-fold ... 25,000-fold ... 100,000-fold ... 500,000-fold ... or 1,000,000-fold, or any range between 1000-fold and 1,000,000-fold).

In certain embodiments, the target nucleic acid is initially present in the sample in an amount between about 10^1 and about 10^3 molecules, and the amount of the target nucleic acid is determined by the measuring hybridization of the first probe oligonucleotides. In other embodiments, the target nucleic acid is initially present in the sample in an amount between about 10^3 and about 10^6 molecules, and the amount of the target nucleic acid is determined by the measuring hybridization of the second probe oligonucleotides. In some embodiments, the target nucleic acid is initially present in the sample in an amount between about 10^6 and about 10^8 molecules, and the amount of the target nucleic acid is determined by the measuring hybridization of the third probe oligonucleotides.

In certain embodiments, the method is conducted on two samples, wherein the target nucleic acid is initially present in one sample in an amount less than 10^3 and initially present in a second sample in an amount greater than 10^5 . In other embodiments, the method is conducted on two samples, wherein the target nucleic acid is initially present in one sample in an amount less than 10^2 and initially present in a second sample in an amount greater than 10^6 . In further embodiments, the method is conducted on two samples, wherein the target nucleic acid is initially present in one sample in an amount

less than 10^4 and initially present in a second sample in an amount greater than 10^7 , or greater than 10^8 , or greater than 10^9 .

In particular embodiments, the plurality of first product sequences (e.g., non-logarithmically amplified sequences) further comprise the second probe hybridization site. In other embodiments, the plurality of second product sequences (e.g., non-logarithmically amplified sequences) do not contain the second probe hybridization site. In certain embodiments, the non-logarithmic amplification of the first region comprises single-stranded PCR or compromised PCR.

In some embodiments, the first probe oligonucleotides further comprise a non-analyte specific region, wherein the non-analyte specific region comprises one or more nucleotides that are not complementary to the second product sequences (e.g., logarithmically amplified sequences). In other embodiments, the second probe oligonucleotides further comprise a non-analyte specific region, wherein the non-analyte specific region comprises one or more nucleotides that are not complementary to the second product sequences (e.g., logarithmically amplified sequences). In other embodiments, the third probe oligonucleotides further comprise a non-analyte specific region, wherein the non-analyte specific region comprises one or more nucleotides that are not complementary to the first product sequences (e.g., non-logarithmically amplified sequences). In further embodiments, the fourth probe oligonucleotides further comprise a non-analyte specific region, wherein the non-analyte specific region comprises one or more nucleotides that are not complementary to the first product sequences (e.g., non-logarithmically amplified sequences).

In certain embodiments, the analyte specific region of second probe oligonucleotide is shorter than the analyte specific region of the first probe oligonucleotide. In other embodiments, the analyte specific region of the fourth probe oligonucleotide is shorter than the analyte specific region of the third probe oligonucleotide.

In some embodiments, the first probe oligonucleotides comprise first labels and wherein the second probe oligonucleotides comprise second labels. In other embodiments, the third probe oligonucleotides comprise third labels and the fourth probe oligonucleotides comprise fourth labels. In particular embodiments, at least one of the

first, second, or third oligonucleotides is unlabeled. In additional embodiments, the first, second, and third probe oligonucleotides are unlabeled. In some embodiments, the fourth probe oligonucleotides are un-labeled. In certain embodiments, the fourth probe oligonucleotides comprises a label. In other embodiments, the first, the second, and the third labels are different from each other or are the same as each other. In certain
5 embodiments, the amplification reagents comprise first and second primers, and a polymerase.

 In some embodiments, the first and second probe oligonucleotides further comprise a non-analyte specific region configured to not hybridize to the second probe
10 hybridization site of the second product sequences (e.g., logarithmically amplified sequences), wherein the non-analyte specific region is 5' of the analyte specific region. In certain embodiments, the first and second probe oligonucleotides form an invasive cleavage structure with an upstream oligonucleotide, wherein the upstream oligonucleotide comprise a 5' portion and a 3' portion, wherein the 5' portion is
15 configured to hybridize to a region contiguous with the second probe hybridization site on the second product sequences (e.g., logarithmically amplified sequences), and wherein the 3' portion is configured to not hybridize to the second product sequences (e.g., logarithmically amplified sequences). In other embodiments, the methods further comprise incubating the sample with a plurality of additional probe oligonucleotides
20 comprising an analyte specific region, wherein the additional probe oligonucleotide is configured to occupy the second probe hybridization site on the second product sequences (e.g., logarithmically amplified sequences) at a frequency different that the first and second probe oligonucleotides.

 In particular embodiments, the present invention provides methods for detecting
25 an amount of a target nucleic acid in a sample, comprising; a) incubating a sample suspected of containing a target nucleic acid with a plurality of first probe oligonucleotides and a plurality of second probe oligonucleotides, wherein each of the first and the second probe oligonucleotides comprises an analyte specific region, wherein the plurality of second probe oligonucleotides are configured to occupy a probe
30 hybridization site on the target nucleic acid with the same affinity as the plurality of first probe oligonucleotides, and wherein the plurality of second probe oligonucleotides are

present in at least 5-fold lower concentration than the first probe oligonucleotides; and b) measuring hybridization of the first and the second probe oligonucleotides over time, thereby detecting the amount of the target nucleic acid. In some embodiments, the first probe oligonucleotides further comprise a first non-analyte specific region, and the
5 second probe oligonucleotides further comprise a second non-analyte specific region which is not identical to the first non-analyte specific region. In other embodiments, the analyte specific regions of the first and second oligonucleotides have an identical sequence.

In additional embodiments, the present invention provides methods for detecting
10 an amount of a target nucleic acid in a sample, comprising; a) incubating a sample suspected of containing a target nucleic acid with a plurality of un-labeled first probe oligonucleotides and a plurality of un-labeled second probe oligonucleotides, wherein each of the first and the second probe oligonucleotides comprises an analyte specific region, wherein the plurality of second probe oligonucleotides are configured to occupy a
15 probe hybridization site on the target nucleic acid at a different frequency than the plurality of first probe oligonucleotides; and b) measuring hybridization of the first and the second probe oligonucleotides over time, thereby detecting the amount of the target nucleic acid.

In further embodiments, the present invention provides methods for detecting an
20 initial amount of a target nucleic acid in a sample without amplifying initial amount of the target nucleic acid, comprising; a) incubating a sample initially containing 300 copies or less of a target nucleic acid with a plurality of first probe oligonucleotides and a plurality of second probe oligonucleotides, wherein each of the first and the second probe oligonucleotides comprises an analyte specific region, wherein the plurality of second
25 probe oligonucleotides are configured to occupy a probe hybridization site on the target nucleic acid at a different frequency than the plurality of first probe oligonucleotides; b) measuring hybridization of the first and the second probe oligonucleotides over time, thereby measuring the amount of the target nucleic acid, wherein the 300 copies or less of the target nucleic acid are not amplified prior to the measuring step. In particular
30 embodiments, the 300 copies or less is between 100 and 300 copies or between 100 and 200 copies.

In some embodiments, the present invention provides methods for detecting an amount of a target nucleic acid in a sample, comprising; a) contacting a sample suspected of containing target nucleic acid with amplification reagents such that, if the target nucleic acid is present, a region of the target nucleic acid containing a probe
5 hybridization site is amplified to generate a plurality of amplified sequences, b) incubating the sample with a plurality of first probe oligonucleotides and a plurality of second probe oligonucleotides, wherein each of the first and the second probe oligonucleotides comprises an analyte specific region, wherein the plurality of second probe oligonucleotides are configured to occupy the a probe hybridization site on the
10 amplified sequence at a different frequency than the plurality of first probe oligonucleotides; c) measuring hybridization of the first and the second probe oligonucleotides over time, thereby measuring the amount of the target nucleic acid, wherein the measuring is possible when the target nucleic acid is initially present in the sample in an amount between about 1 molecule and about 10^7 molecules.

15 In certain embodiments, the incubating and measuring steps are conducted in a single vessel. In other embodiments, the contacting, incubating, and measuring steps are conducted in a single vessel. In further embodiments, the analyte specific regions of the first oligonucleotides are identical to the analyte specific regions of the second oligonucleotides. In some embodiments, the analyte specific regions of the second probe
20 oligonucleotides contain a single mismatch with the logarithmically amplified sequences.

In some embodiments, the second or fourth probe oligonucleotides contain a charge tag. In other embodiments, the second or fourth probe oligonucleotide contains at least one modified nucleotide. In further embodiments, the second probe oligonucleotide has a lower or higher affinity for the second probe hybridization site than the first probe
25 oligonucleotide. In particular embodiments, the second probe oligonucleotide has a lower or higher T_m with the second probe hybridization site than the first probe oligonucleotide. In additional embodiments, the fourth probe oligonucleotide has a lower or higher affinity for the first probe hybridization site than the third probe oligonucleotide. In other embodiments, the fourth probe oligonucleotide has a lower or
30 higher T_m with the first probe hybridization site than the third probe oligonucleotide.

In some embodiments, the measuring hybridization of the first, second, and/or third, and/or fourth probe oligonucleotides comprises performing a hybridization assay. In particular embodiments, the hybridization assay is selected from the group consisting of a TAQMAN assay, SNP-IT assay, an invasive cleavage assay, a Southern blot, and a
5 microarray assay. In further embodiments, the invasive cleavage assay is an INVADER assay.

In certain embodiments, the present invention provides methods for genotyping a polymorphic locus in a target nucleic acid in a sample, comprising; a) contacting a sample suspected of containing the target nucleic acid with amplification reagents such
10 that, if the target nucleic acid is present, a region of the target nucleic acid containing the polymorphic locus is amplified to generate a plurality of amplified sequences, wherein the amplification is conducted until saturation; b) incubating the sample with a plurality of first probe oligonucleotides and a plurality of second probe oligonucleotides, wherein each of the first probe oligonucleotides comprises: i) a first analyte specific region
15 configured for detecting a first allele at the polymorphic locus, and ii) a label capable of generating a detectable signal or a cleavable portion configured to cause a detectable signal to be generated, and wherein the second probe oligonucleotides comprise: i) a second analyte specific region configured for detecting a second allele at the polymorphic locus, ii) a label capable of generating a detectable signal or a cleavable portion
20 configured to cause a detectable signal to be generated, wherein the plurality of second probe oligonucleotides are configured to occupy a probe hybridization site on the amplified sequences at a different frequency than the plurality of first probe oligonucleotides, and wherein the type of detectable signal from the first and second probe oligonucleotides is the same; c) measuring the strength of the detectable signal
25 generated, thereby determining the presence of the first allele, the second allele, or both the first and second alleles in the target nucleic acid. In certain embodiments, the polymorphic locus is a single nucleotide polymorphism. In other embodiments, the polymorphic locus is a repeat sequence.

In some embodiments, the present invention provides methods for detecting a
30 target nucleic acid in a sample, comprising; a) incubating a sample suspected of containing a target nucleic acid with a plurality of first and second probe

oligonucleotides, a plurality of upstream oligonucleotides, and a cleavage agent, wherein each of the first probe oligonucleotides comprise: i) a first analyte specific region configured to hybridize to a probe hybridization site on the target nucleic acid, and ii) a first non-analyte specific region configured to not hybridize to the target nucleic acid, wherein the first non-analyte specific region is 5' of the first analyte specific region, and wherein each of the second probe oligonucleotides comprises i) a second analyte specific region configured to hybridize to the probe hybridization site on the target nucleic acid, and ii) a second non-analyte specific region configured to not hybridize to the target nucleic acid, wherein the second non-analyte specific region is not identical to the first non-analyte specific region, and wherein the plurality of second probe oligonucleotides are configured to occupy the probe hybridization site on the target nucleic acid at a different frequency than the plurality of first probe oligonucleotides; wherein the incubating is under conditions such that invasive cleavage structures are formed resulting in the cleavage of both the first and second probe oligonucleotides by the cleavage agent to generate: i) first non-target cleavage products comprising the first non-analyte specific region, and ii) second non-target cleavage products comprising the second non-analyte specific region; and b) measuring hybridization of the first and the second probe oligonucleotides by detecting a signal generated by the first and second non-target cleavage products, thereby detecting the target nucleic acid. In some embodiments, the amount of the target is detected.

In certain embodiments, the present invention provides methods for detecting a target nucleic acid in a sample, comprising: a) incubating a sample suspected of containing a target nucleic acid with a plurality of first and second probe oligonucleotides, a plurality of first upstream oligonucleotides, a plurality of second upstream oligonucleotides, and a cleavage agent, wherein each of the first probe oligonucleotides comprise: i) a first analyte specific region configured to hybridize to a first probe hybridization site on the target nucleic acid, and ii) a first non-analyte specific region configured to not hybridize to the target nucleic acid, wherein the first non-analyte specific region is 5' of the first analyte specific region, and wherein each of the second probe oligonucleotides comprises i) a second analyte specific region configured to hybridize to a second probe hybridization site on the target nucleic acid, wherein the

second probe hybridization site is not the same as the first probe hybridization site, and ii) a second non-analyte specific region configured to not hybridize to the target nucleic acid, wherein the second non-analyte specific region is not identical to the first non-analyte specific region, and wherein the plurality of second probe oligonucleotides are
5 present in at least 5-fold lower concentration than the first probe oligonucleotides; wherein the incubating is under conditions such that invasive cleavage structures are formed resulting in the cleavage of both the first and second probe oligonucleotides by the cleavage agent to generate: i) first non-target cleavage products comprising the first non-analyte specific region, and ii) second non-target cleavage products comprising the
10 second non-analyte specific region; and b) measuring hybridization of the first and the second probe oligonucleotides over time by detecting a signal generated by the first and second non-target cleavage products, thereby measuring the amount of the target nucleic acid. In certain embodiments, the target nucleic acid is micro-RNA. In some embodiments, the amount of the target is detected.

15 In certain embodiments, the second probe oligonucleotides are present in at least a 10-fold, 100-fold, or 1000-fold, lower concentration than the first probe oligonucleotides. In further embodiments, the signal generated by the first and second non-target cleavage products is the same. In other embodiments, the signal generated by the first and second non-target cleavage products is different. In some embodiments, the upstream
20 oligonucleotides comprise a 5' portion and a 3' portion, wherein the 5' portion is configured to hybridize to a region contiguous with the probe hybridization site on the target nucleic acid, and wherein the 3' portion is configured to not hybridize to the target nucleic acid. In further embodiments, the methods further comprise incubating the sample with first and second labeled sequences, wherein the first labeled sequence is
25 configured to generate a first detectable signal when hybridized to the first non-target cleavage product, and wherein the second labeled sequence is configured to generate a second detectable signal when hybridized to the second non-target cleavage product. In particular embodiments, the first and second detectable signals are the same. In additional embodiments, the first and second labeled sequences comprise FRET cassettes.
30 In other embodiments, the plurality of upstream oligonucleotides are generated in the

sample (e.g., by a polymerase). In some embodiments, the upstream oligonucleotides are supplied pre-synthesized.

In certain embodiments, the present invention provides kits for quantitation of target nucleic acids in a sample, comprising: a) a plurality of first probe oligonucleotides, wherein each of the first probe oligonucleotides comprises a first analyte specific region, wherein the first probe oligonucleotides are un-labeled, or comprise a label, b) a plurality of second probe oligonucleotides, wherein each of the second probe oligonucleotides comprises a second analyte specific region, wherein the second probe oligonucleotides are un-labeled, or comprise a label, wherein the plurality of second probe oligonucleotides are configured to occupy a probe hybridization site on the target nucleic acids at a different frequency than the plurality of first probe oligonucleotides; and c) reagents for performing an INVADER assay using the pluralities of the first and second probe oligonucleotides.

In certain embodiments, the present invention provides kits or compositions comprising: i) a plurality of first oligonucleotides, and ii) a plurality of second probe oligonucleotides, wherein the first probe oligonucleotides comprise a first 5' region and a first 3' region, and the second probe oligonucleotides comprises a second 5' region and a second 3' region, wherein both of the first and second probe oligonucleotides will form an invasive cleavage structure in the presence of the same upstream oligonucleotide and target sequence, and will both be cleaved by the same cleavage agent to form a first 5' region product and a second 5' region product, wherein the second 5' region product is not identical to the first 5' region product. In some embodiments, the kit or composition further comprises iii) first and second labeled sequences, wherein the first labeled sequence is configured to generate a first detectable signal when hybridized to the first 5' region product, and wherein the second labeled sequence is configured to generate a second detectable signal when hybridized to the second 5' region product. In some embodiments, the kits further comprise the target sequence as a control.

In particular embodiments, the first and second probe oligonucleotides are provided in a first vessel. In further embodiments, the , and kit further comprises a second vessel containing a polymerase and FEN enzyme. In additional embodiments, the kit further comprises a third vessel containing a buffer.

In some embodiments, the first 3' region and the second 3' region have the identical sequence. In other embodiments, the first 3' region and the second 3' region do not have identical sequences. In particular embodiments, the second probe oligonucleotides are present in at least a 5 fold lower concentration than the first probe oligonucleotides. In other embodiments, the second probe oligonucleotides are present in at least a 10 fold ... 100-fold ... 1000-fold ... 10,000-fold ... or 500,000 lower concentration than the first probe oligonucleotides. In some embodiments, the first and second probe oligonucleotides are un-labeled.

In further embodiments, the kits or compositions further comprise a third probe oligonucleotide comprising a third 5' region and a third 3' region, wherein the third probe oligonucleotide will not form an invasive cleavage structure with the target and the upstream oligonucleotide that is cleavable by the cleavage agent. In some embodiments, the first and second detectable signals are the same or they are different.

In some embodiments, the present invention provides kits comprising i) a plurality of un-labeled first probe oligonucleotides and ii) a plurality of un-labeled second probe oligonucleotides, wherein the first and second probe oligonucleotides comprises an analyte specific region, wherein the plurality of second probe oligonucleotides are configured to occupy a probe hybridization site on a target nucleic acid at a different frequency than the plurality of first probe oligonucleotides. In further embodiments, the kits further comprise a polymerase and/or a FEN enzyme. In other embodiments, the kits further comprise a buffer.

In some embodiments, the present invention provides kits comprising; a) a first vessel comprising a plurality of first probe oligonucleotides (e.g., unlabeled) and a plurality of second probe oligonucleotides (e.g., unlabeled), wherein the first and second probe oligonucleotides comprises an analyte specific region, wherein the plurality of second probe oligonucleotides are configured to occupy a probe hybridization site on a target nucleic acid at a different frequency than the plurality of first probe oligonucleotides; b) a second vessel comprising a polymerase and/or a FEN enzyme, and c) a third vessel comprising a buffer. In certain embodiments, the kits further comprise d) a control target sequence comprising the probe hybridization site.

In particular embodiments, the present invention provides kits and compositions comprising; a) a plurality of first and second probe oligonucleotides, wherein the first probe oligonucleotides comprise a first 5' region and a first 3' region, and the second probe oligonucleotides comprises a second 5' region and a second 3' region, wherein
5 both of the first and second probe oligonucleotides will form an invasive cleavage structure in the presence of the same upstream oligonucleotide and target sequence, and will both be cleaved by the same cleavage agent to form a first 5' region product and a second 5' region product, wherein the second 5' region product is identical to the first 5' region product, and wherein the first 3' region is not identical to the second 3' region, and
10 b) first labeled sequences, wherein the first labeled sequence is configured to generate a detectable signal when hybridized to the first or second 5' region product.

Other embodiments of the invention are described in the Detailed Description of the Invention and the Examples.

15 DEFINITIONS

To facilitate an understanding of the present invention, a number of terms and phrases are defined below:

As used herein, the term "dynamic range" refers to the quantitative range of usefulness in a detection assay (e.g., a nucleic acid detection assay). For example, the
20 dynamic range of a viral detection assay is the range between the smallest number of viral particles (e.g., copy number) and the largest number of viral particles that the assay can distinguish between.

As used herein, the terms "subject" and "patient" refer to any organisms including plants, microorganisms and animals (e.g., mammals such as dogs, cats, livestock, and
25 humans).

The term "primer" refers to an oligonucleotide that is capable of acting as a point of initiation of synthesis when placed under conditions in which primer extension is initiated. An oligonucleotide "primer" may occur naturally, as in a purified restriction digest or may be produced synthetically.

30 The term "cleavage structure" as used herein, refers to a structure that is formed by the interaction of at least one probe oligonucleotide and a target nucleic acid, forming

a structure comprising a duplex, the resulting structure being cleavable by a cleavage means, including but not limited to an enzyme. The cleavage structure is a substrate for specific cleavage by the cleavage means in contrast to a nucleic acid molecule that is a substrate for non-specific cleavage by agents such as phosphodiesterases, which cleave nucleic acid molecules without regard to secondary structure (*i.e.*, no formation of a duplexed structure is required).

The term "invasive cleavage structure" as used herein refers to a cleavage structure comprising i) a target nucleic acid, ii) an upstream nucleic acid (*e.g.*, an INVADER oligonucleotide), and iii) a downstream nucleic acid (*e.g.*, a probe), where the upstream and downstream nucleic acids anneal to contiguous regions of the target nucleic acid, and where an overlap forms between the upstream nucleic acid and duplex formed between the downstream nucleic acid and the target nucleic acid. An overlap occurs where one or more bases from the upstream and downstream nucleic acids occupy the same position with respect to a target nucleic acid base, whether or not the overlapping base(s) of the upstream nucleic acid are complementary with the target nucleic acid, and whether or not those bases are natural bases or non-natural bases. In some embodiments, the 3' portion of the upstream nucleic acid that overlaps with the downstream duplex is a non-base chemical moiety such as an aromatic ring structure, *e.g.*, as disclosed, for example, in U.S. Patent No. 6,090,543, incorporated herein by reference in its entirety. In some embodiments, one or more of the nucleic acids may be attached to each other, *e.g.*, through a covalent linkage such as nucleic acid stem-loop, or through a non-nucleic acid chemical linkage (*e.g.*, a multi-carbon chain).

The term "cleavage means" or "cleavage agent" as used herein refers to any means that is capable of cleaving a cleavage structure, including but not limited to enzymes. "Structure-specific nucleases" or "structure-specific enzymes" are enzymes that recognize specific secondary structures in a nucleic molecule and cleave these structures. The cleavage means of the invention cleave a nucleic acid molecule in response to the formation of cleavage structures; it is not necessary that the cleavage means cleave the cleavage structure at any particular location within the cleavage structure.

The cleavage means may include nuclease activity provided from a variety of sources including the CLEAVASE enzymes, the FEN-1 endonucleases (including RAD2 and XPG proteins), *Taq* DNA polymerase and *E. coli* DNA polymerase I. The cleavage means may include enzymes having 5' nuclease activity (e.g., *Taq* DNA polymerase (DNAP), *E. coli* DNA polymerase I). The cleavage means may also include modified DNA polymerases having 5' nuclease activity but lacking synthetic activity. Examples of cleavage means suitable for use in the method and kits of the present invention are provided in U.S. Patent Nos. 5,614,402; 5,795,763; 5,843,669; PCT Appln. Nos WO 98/23774; WO 02/070755A2; and WO0190337A2, each of which is herein incorporated by reference in its entirety.

The term "thermostable" when used in reference to an enzyme, such as a 5' nuclease, indicates that the enzyme is functional or active (i.e., can perform catalysis) at an elevated temperature, i.e., at about 55°C or higher. In some embodiments the enzyme is functional or active at an elevated temperature of 65°C or higher (e.g., 75°C, 85°C, 95°C, etc.).

The term "cleavage products" as used herein, refers to products generated by the reaction of a cleavage means with a cleavage structure (i.e., the treatment of a cleavage structure with a cleavage means).

"Amplification" is a special case of nucleic acid replication involving template specificity. It is to be contrasted with non-specific template replication (i.e., replication that is template-dependent but not dependent on a specific template). Template specificity is here distinguished from fidelity of replication (i.e., synthesis of the proper polynucleotide sequence) and nucleotide (ribo- or deoxyribo-) specificity. Template specificity is frequently described in terms of "target" specificity. Target sequences are "targets" in the sense that they are sought to be sorted out from other nucleic acid. Amplification techniques have been designed primarily for this sorting out.

Template specificity is achieved in most amplification techniques by the choice of enzyme. Amplification enzymes are enzymes that, under conditions they are used, will process only specific sequences of nucleic acid in a heterogeneous mixture of nucleic acid. For example, in the case of Q β replicase, MDV-1 RNA is the specific template for the replicase (D.L. Kacian *et al.*, Proc. Natl. Acad. Sci. USA 69:3038 [1972]). Other

nucleic acid will not be replicated by this amplification enzyme. Similarly, in the case of T7 RNA polymerase, this amplification enzyme has a stringent specificity for its own promoters (Chamberlin *et al.*, Nature 228:227 [1970]). In the case of T4 DNA ligase, the enzyme will not ligate the two oligonucleotides or polynucleotides, where there is a mismatch between the oligonucleotide or polynucleotide substrate and the template at the ligation junction (D.Y. Wu and R. B. Wallace, Genomics 4:560 [1989]). Finally, *Taq* and *Pfu* polymerases, by virtue of their ability to function at high temperature, are found to display high specificity for the sequences bounded and thus defined by the primers; the high temperature results in thermodynamic conditions that favor primer hybridization with the target sequences and not hybridization with non-target sequences (H.A. Erlich (ed.), *PCR Technology*, Stockton Press [1989]).

As used herein, the term "amplifiable nucleic acid" is used in reference to nucleic acids that may be amplified by any amplification method. It is contemplated that "amplifiable nucleic acid" will usually comprise "sample template."

As used herein, the term "sample template" refers to nucleic acid originating from a sample that is analyzed for the presence of "target." In contrast, "background template" is used in reference to nucleic acid other than sample template that may or may not be present in a sample. Background template is most often inadvertent. It may be the result of carryover, or it may be due to the presence of nucleic acid contaminants sought to be purified away from the sample. For example, nucleic acids from organisms other than those to be detected may be present as background in a test sample.

The term "analyte specific region" as used in reference to an oligonucleotide, such as a probe oligonucleotide or an INVADER oligonucleotide, refers to a region of an oligonucleotide selected to hybridize to a specific sequence in a target nucleic acid or set of target nucleic acids. In some embodiments, an analyte specific region may be completely complementary to the segment of a target nucleic acid to which it hybridizes, while in other embodiments, an analyte specific region may comprise one or more mismatches to the segment of a target nucleic acid to which it hybridizes. In yet other embodiments, an analyte specific region may comprise one or more base analogs, *e.g.*, compounds that have altered hydrogen bonding, or that do not hydrogen bond, to the bases in the target strand. In some embodiments, the entire sequence of an

oligonucleotide is an analyte specific region, while in other embodiments an oligonucleotide comprises an analyte specific region and one or more regions not complementary the target sequence (*e.g.*, non-complementary flap regions).

The term "frequency" as used herein in reference to hybridization of nucleic acids refers to the probability that one particular nucleic acid (*e.g.*, a probe oligonucleotide) will be base-paired to a complementary nucleic acid (*e.g.*, a target nucleic acid) under particular hybridization conditions. The frequency of hybridization is influenced by many factors, including but not limited to the probability with which the complementary sequences will form a duplex under particular conditions (*e.g.*, likelihood of encounter and of successful duplex formation) and the stability of the duplex, once formed. Reaction conditions that increase the likelihood of initial duplex formation between a probe and a target (*e.g.*, increased concentration of one or both nucleic acids, absence of competitors such as other nucleic acids with sequences that can compete with a probe for binding to the target, or that can bind to the probe) can be said to increase the frequency of hybridization of between the probe and target (*i.e.*, increase the frequency with which the probe oligonucleotide will occupy, or hybridize to, the complementary target strand). Similarly, reaction conditions and probe features that increase the stability of a hybrid between an oligonucleotide and another nucleic acid strand (or that slow disassociation of the strands, *e.g.*, reduced reaction temperature, increased salt or divalent cation conditions, increased length of complementary regions, fewer mismatches, use of charged moieties favoring hybridization) can also be said to increase the frequency of hybridization of between the probe and target. Conversely, reaction conditions and probe features that decrease the likelihood of hybridization (*e.g.*, reduction in concentration of one or both nucleic acids, the presence of a competitor or other additive that reduces the effective concentration of a probe or target strand) or that reduce the stability and/or life time of hybrids that are formed (*e.g.*, increased reaction temperature, decreased salt or divalent cation conditions, decreased length of complementary regions, more mismatches, use of charged moieties disfavoring hybridization) are said to decrease the frequency of hybridization or occupation.

As used herein, the term "target," refers to a nucleic acid sequence or structure to be detected or characterized. Thus, the "target" is sought to be sorted out from other

nucleic acid sequences. A "segment" is defined as a region of nucleic acid within the target sequence.

The term "substantially single-stranded" when used in reference to a nucleic acid substrate means that the substrate molecule exists primarily as a single strand of nucleic acid in contrast to a double-stranded substrate which exists as two strands of nucleic acid which are held together by inter-strand base pairing interactions.

As used herein, the phrase "non-amplified oligonucleotide detection assay" refers to a detection assay configured to detect the presence or absence of a particular target sequence (e.g. genomic DNA or viral DNA or RNA) that has not been amplified (e.g. by PCR), without creating copies of the target sequence. A "non-amplified oligonucleotide detection assay" may, for example, amplify a signal used to indicate the presence or absence of a particular polymorphism in a target sequence, so long as the target sequence is not copied.

The term "liberating" as used herein refers to the release of a nucleic acid fragment from a larger nucleic acid fragment, such as an oligonucleotide, by the action of, for example, a 5' nuclease such that the released fragment is no longer covalently attached to the remainder of the oligonucleotide.

The term "microorganism" as used herein means an organism too small to be observed with the unaided eye and includes, but is not limited to bacteria, virus, protozoans, fungi, and ciliates.

The term "microbial gene sequences" refers to gene sequences derived from a microorganism.

The term "bacteria" refers to any bacterial species including eubacterial and archaeobacterial species.

The term "virus" refers to obligate, ultramicroscopic, intracellular parasites incapable of autonomous replication (i.e., replication requires the use of the host cell's machinery).

The term "multi-drug resistant" or "multiple-drug resistant" refers to a microorganism that is resistant to more than one of the antibiotics or antimicrobial agents used in the treatment of said microorganism.

The term "source of target nucleic acid" refers to any sample that contains nucleic acids (RNA or DNA). Particularly preferred sources of target nucleic acids are biological samples including, but not limited to blood, saliva, cerebral spinal fluid, pleural fluid, milk, lymph, sputum and semen.

5 A sample "suspected of containing" a first and a second target nucleic acid may contain either, both or neither target nucleic acid molecule.

The term "reactant" is used herein in its broadest sense. The reactant can comprise, for example, an enzymatic reactant, a chemical reactant or light (e.g., ultraviolet light, particularly short wavelength ultraviolet light is known to break
10 oligonucleotide chains). Any agent capable of reacting with an oligonucleotide to either shorten (i.e., cleave) or elongate the oligonucleotide is encompassed within the term "reactant."

As used herein the term "portion" when in reference to a protein (as in "a portion of a given protein") refers to fragments of that protein. The fragments may range in size
15 from four amino acid residues to the entire amino acid sequence minus one amino acid (e.g., 4, 5, 6, . . . , n-1).

The term "continuous strand of nucleic acid" as used herein is means a strand of nucleic acid that has a continuous, covalently linked, backbone structure, without nicks or other disruptions. The disposition of the base portion of each nucleotide, whether
20 base-paired, single-stranded or mismatched, is not an element in the definition of a continuous strand. The backbone of the continuous strand is not limited to the ribose-phosphate or deoxyribose-phosphate compositions that are found in naturally occurring, unmodified nucleic acids. A nucleic acid of the present invention may comprise modifications in the structure of the backbone, including but not limited to
25 phosphorothioate residues, phosphonate residues, 2' substituted ribose residues (e.g., 2'-O-methyl ribose) and alternative sugar (e.g., arabinose) containing residues.

The term "continuous duplex" as used herein refers to a region of double stranded nucleic acid in which there is no disruption in the progression of basepairs within the duplex (i.e., the base pairs along the duplex are not distorted to accommodate a gap,
30 bulge or mismatch with the confines of the region of continuous duplex). As used herein the term refers only to the arrangement of the basepairs within the duplex, without

implication of continuity in the backbone portion of the nucleic acid strand. Duplex nucleic acids with uninterrupted basepairing, but with nicks in one or both strands are within the definition of a continuous duplex.

The term "duplex" refers to the state of nucleic acids in which the base portions of the nucleotides on one strand are bound through hydrogen bonding the their complementary bases arrayed on a second strand. The condition of being in a duplex form reflects on the state of the bases of a nucleic acid. By virtue of base pairing, the strands of nucleic acid also generally assume the tertiary structure of a double helix, having a major and a minor groove. The assumption of the helical form is implicit in the act of becoming duplexed.

The term "template" refers to a strand of nucleic acid on which a complementary copy is built from nucleoside triphosphates through the activity of a template-dependent nucleic acid polymerase. Within a duplex the template strand is, by convention, depicted and described as the "bottom" strand. Similarly, the non-template strand is often depicted and described as the "top" strand.

As used herein, the term "sample" is used in its broadest sense. For example, in some embodiments, it is meant to include a specimen or culture (e.g., microbiological culture), whereas in other embodiments, it is meant to include both biological and environmental samples (e.g., suspected of comprising a target sequence, gene or template). In some embodiments, a sample may include a specimen of synthetic origin.

The present invention is not limited by the type of biological sample used or analyzed. The present invention is useful with a variety of biological samples including, but are not limited to, tissue (e.g., organ (e.g., heart, liver, brain, lung, stomach, intestine, spleen, kidney, pancreas, and reproductive (e.g., ovaries) organs), glandular, skin, and muscle tissue), cell (e.g., blood cell (e.g., lymphocyte or erythrocyte), muscle cell, tumor cell, and skin cell), gas, bodily fluid (e.g., blood or portion thereof, serum, plasma, urine, semen, saliva, etc), or solid (e.g., stool) samples obtained from a human (e.g., adult, infant, or embryo) or animal (e.g., cattle, poultry, mouse, rat, dog, pig, cat, horse, and the like). In some embodiments, biological samples may be solid food and/or feed products and/or ingredients such as dairy items, vegetables, meat and meat by-products, and waste. Biological samples may be obtained from all of the various families of domestic animals,

as well as feral or wild animals, including, but not limited to, such animals as ungulates, bear, fish, lagamorphs, rodents, etc.

Biological samples also include biopsies and tissue sections (e.g., biopsy or section of tumor, growth, rash, infection, or paraffin-embedded sections), medical or hospital samples (e.g., including, but not limited to, blood samples, saliva, buccal swab, cerebrospinal fluid, pleural fluid, milk, colostrum, lymph, sputum, vomitus, bile, semen, oocytes, cervical cells, amniotic fluid, urine, stool, hair and sweat), laboratory samples (e.g., subcellular fractions), and forensic samples (e.g., blood or tissue (e.g., spatter or residue), hair and skin cells containing nucleic acids), and archeological samples (e.g., fossilized organisms, tissue, or cells).

Environmental samples include, but are not limited to, environmental material such as surface matter, soil, water (e.g., freshwater or seawater), algae, lichens, geological samples, air containing materials containing nucleic acids, crystals, and industrial samples, as well as samples obtained from food and dairy processing instruments, apparatus, equipment, utensils, disposable and non-disposable items.

Other types of biological samples include bacteria (e.g., *Actinobacteria* (e.g., *Actinomyces*, *Arthrobacter*, *Corynebacterium* (e.g., *C. diphtheriae*)), *Mycobacterium* (e.g., *M. tuberculosis* and *M. leprae*), *Propionibacterium* (e.g., *P. acnes*), *Streptomyces*, *Chlamydiae* (e.g., *C. trachomatis* and *C. pneumoniae*), *Cyanobacteria*, *Deinococcus* (e.g., *Thermus* (e.g., *T. aquaticus*)), *Firmicutes* (e.g., *Bacilli* (e.g., *B. anthracis*, *B. cereus*, *B. thuringiensis*, and *B. subtilis*)), *Listeria* (e.g., *L. monocytogenes*), *Staphylococcus* (e.g., *S. aureus*, *S. epidermidis*, and *S. haemolyticus*), *Fusobacteria*, *Proteobacteria* (e.g., *Rickettsiales*, *Sphingomonadales*, *Bordetella* (e.g., *B. pertussis*), *Neisserisales* (e.g., *N. gonorrhoeae* and *N. meningitidis*), *Enterobacteriales* (e.g., *Escherichia* (e.g., *E. coli*), *Klebsiella*, *Plesiomonas*, *Proteus*, *Salmonella*, *Shigella*, and *Yersinia*), *Legionellales*, *Pasteurellales* (e.g., *Haemophilus influenzae*), *Pseudomonas*, *Vibrio* (e.g., *V. cholerae* and *V. vulnificus*), *Campylobacteriales* (e.g., *Campylobacter* (e.g., *C. jejuni*), and *Helicobacter* (e.g., *H. pylori*)), and *Spirochaetes* (e.g., *Leptospira*, *B. bergdorferi*, and *T. pallidum*)); Archaea (e.g., *Halobacteria* and *Methanobacteria*); Eucarya (e.g., *Animalia* (e.g., *Annelidia*, *Arthropoda* (e.g., *Chelicerata*, *Myriapoda*, *Insecta*, and *Crustacea*), *Mollusca*, *Nematoda*, (e.g., *C. elegans*, and *T. spiralis*) and *Chordata* (e.g.,

Actinopterygii, Amphibia, Aves, Chondrichthyes, Reptilia, and Mammalia (e.g., *Primates, Rodentia, Lagomorpha, and Carnivora*)); Fungi (e.g., *Dermatophytes, Fusarium, Penicillium, and Saccharomyces*); *Plantae* (e.g., *Magnoliophyta* (e.g., *Magnoliopsida* and *Liliopsida*)), and *Protista* (e.g., *Apicomplexa* (e.g., *Cryptosporidium, Plasmodium* (e.g., *P. falciparum, and Toxoplasma*), and *Metamonada* (e.g., *G. lambia*))); and Viruses (e.g., dsDNA viruses (e.g., *Bacteriophage, Adenoviridae, Herpesviridae, Papillomaviridae, Polyomaviridae, and Poxviridae*), ssDNA viruses (e.g., *Parvoviridae*), dsRNA viruses (including *Reoviridae*), (+)ssRNA viruses (e.g., *Coronaviridae, Astroviridae, Bromoviridae, Comoviridae, Flaviviridae, Picornaviridae, and Togaviridae*), (-) ssRNA viruses (e.g., *Bornaviridae, Filoviridae, Paramyxoviridae, Rhabdoviridae, Bunyaviridae, and Orthomyxoviridae*), ssRNA-reverse transcribing viruses (e.g., *Retroviridae*), and dsDNA-reverse transcribing viruses (e.g., *Hepadnaviridae and Caulomoviridae*)).

Sample may be prepared by any desired or suitable method. In some embodiments, nucleic acids are analyzed directly from bodily fluids or other samples using the methods described in U.S. Pat. Pub. Serial No. 20050186588, herein incorporated by reference in its entirety.

The above described examples are not, however, to be construed as limiting the sample (e.g., suspected of comprising a target sequence, gene or template (e.g., the presence or absence of which can be determined using the compositions and methods of the present invention)) types applicable to the present invention.

The terms "nucleic acid sequence" and "nucleic acid molecule" as used herein refer to an oligonucleotide, nucleotide or polynucleotide, and fragments or portions thereof. The terms encompasses sequences that include any of the known base analogs of DNA and RNA including, but not limited to, 4-acetylcytosine, 8-hydroxy-N6-methyladenosine, aziridinylcytosine, pseudoisocytosine, 5-(carboxyhydroxymethyl)uracil, 5-fluorouracil, 5-bromouracil, 5-carboxymethylaminomethyl-2-thiouracil, 5-carboxymethylaminomethyluracil, dihydrouracil, inosine, N6-isopentenyladenine, 1-methyladenine, 1-methylpseudouracil, 1-methylguanine, 1-methylinosine, 2,2-dimethylguanine, 2-methyladenine, 2-methylguanine, 3-methylcytosine, 5-methylcytosine, N6-methyladenine, 7-methylguanine, 5-methylaminomethyluracil, 5-methoxyaminomethyl-2-thiouracil, beta-D-mannosylqueosine,

5'-methoxycarbonylmethyluracil, 5-methoxyuracil, 2-methylthio-N6-isopentenyladenine, uracil-5-oxyacetic acid methylester, uracil-5-oxyacetic acid, oxybutoxosine, pseudouracil, queosine, 2-thiocytosine, 5-methyl-2-thiouracil, 2-thiouracil, 4-thiouracil, 5-methyluracil, N-uracil-5-oxyacetic acid methylester, uracil-5-oxyacetic acid, pseudouracil, queosine, 2-thiocytosine, and 2,6-diaminopurine.

A nucleic acid sequence or molecule may be DNA or RNA, of either genomic or synthetic origin, that may be single or double stranded, and represent the sense or antisense strand. Thus, nucleic acid sequence may be dsDNA, ssDNA, mixed ssDNA, mixed dsDNA, dsDNA made into ssDNA (e.g., through melting, denaturing, helicases, etc.), A-, B-, or Z- DNA, triple-stranded DNA, RNA, ssRNA, dsRNA, mixed ss and dsRNA, dsRNA made into ssRNA (e.g., via melting, denaturing, helicases, etc.), messenger RNA (mRNA), ribosomal RNA (rRNA), transfer RNA (tRNA), catalytic RNA, snRNA, or protein nucleic acid (PNA).

The present invention is not limited by the type or source of nucleic acid (e.g., sequence or molecule (e.g. target sequence and/or oligonucleotide)) utilized. For example, the nucleic acid sequence may be amplified or created sequence (e.g., amplification or creation of nucleic acid sequence via synthesis (e.g., polymerization (e.g., primer extension (e.g., RNA-DNA hybrid primer technology)) and reverse transcription (e.g., of RNA into DNA)) and/or amplification (e.g., polymerase chain reaction (PCR), rolling circle amplification (RCA), nucleic acid sequence based amplification (NASBA), transcription mediated amplification (TMA), ligase chain reaction (LCR), cycling probe technology, Q-beta replicase, strand displacement amplification (SDA), branched-DNA signal amplification (bDNA), hybrid capture, and helicase dependent amplification).

The terms "nucleotide" and "base" are used interchangeably when used in reference to a nucleic acid sequence, unless indicated otherwise herein.

The term "nucleotide analog" as used herein refers to modified or non-naturally occurring nucleotides including, but not limited to, analogs that have altered stacking interactions such as 7-deaza purines (i.e., 7-deaza-dATP and 7-deaza-dGTP); base analogs with alternative hydrogen bonding configurations (e.g., Iso-C and Iso-G and other non-standard base pairs described in U.S. Patent No. 6,001,983, herein incorporated

by reference in its entirety); non-hydrogen bonding analogs (e.g., non-polar, aromatic nucleoside analogs such as 2,4-difluorotoluene, described by B.A. Schweitzer and E.T. Kool, J. Org. Chem., 1994, 59, 7238-7242; B.A. Schweitzer and E.T. Kool, J. Am. Chem. Soc., 1995, 117, 1863-1872, each of which is herein incorporate by reference in its
5 entirety); "universal" bases such as 5-nitroindole and 3-nitropyrrole; and universal purines and pyrimidines (e.g., "K" and "P" nucleotides, respectively; See, e.g., P. Kong, et al., Nucleic Acids Res., 1989, 17, 10373-10383; P. Kong et al., Nucleic Acids Res., 1992, 20, 5149-5152). Still other nucleotide analogs include modified forms of deoxyribonucleotides as well as ribonucleotides. Various oligonucleotides of the present
10 invention (e.g., a primary probe or INVADER oligo) may contain nucleotide analogs.

The term "oligonucleotide" as used herein is defined as a molecule comprising two or more nucleotides (e.g., deoxyribonucleotides or ribonucleotides), preferably at least 5 nucleotides, more preferably at least about 10-15 nucleotides and more preferably at least about 15 to 30 nucleotides, or longer (e.g., oligonucleotides are typically less than
15 200 residues long (e.g., between 15 and 100 nucleotides), however, as used herein, the term is also intended to encompass longer polynucleotide chains). The exact size will depend on many factors, which in turn depend on the ultimate function or use of the oligonucleotide. Oligonucleotides are often referred to by their length. For example a 24 residue oligonucleotide is referred to as a "24-mer". Oligonucleotides can form
20 secondary and tertiary structures by self-hybridizing or by hybridizing to other polynucleotides. Such structures can include, but are not limited to, duplexes, hairpins, cruciforms, bends, and triplexes. Oligonucleotides may be generated in any manner, including chemical synthesis, DNA replication, reverse transcription, PCR, or a combination thereof. In some embodiments, oligonucleotides that form invasive
25 cleavage structures are generated in a reaction (e.g., by extension of a primer in an enzymatic extension reaction).

Because mononucleotides are reacted to make oligonucleotides in a manner such that the 5' phosphate of one mononucleotide pentose ring is attached to the 3' oxygen of its neighbor in one direction via a phosphodiester linkage, an end of an oligonucleotide is
30 referred to as the "5' end" if its 5' phosphate is not linked to the 3' oxygen of a mononucleotide pentose ring and as the "3' end" if its 3' oxygen is not linked to a 5'

phosphate or a subsequent mononucleotide pentose ring. As used herein, a nucleic acid sequence, even if internal to a larger oligonucleotide, also may be said to have 5' and 3' ends. A first region along a nucleic acid strand is said to be upstream of another region if the 3' end of the first region is before the 5' end of the second region when moving along a strand of nucleic acid in a 5' to 3' direction.

When two different, non-overlapping oligonucleotides anneal to different regions of the same linear complementary nucleic acid sequence, and the 3' end of one oligonucleotide points towards the 5' end of the other, the former may be called the "upstream" oligonucleotide and the latter the "downstream" oligonucleotide. Similarly, when two overlapping oligonucleotides are hybridized to the same linear complementary nucleic acid sequence, with the first oligonucleotide positioned such that its 5' end is upstream of the 5' end of the second oligonucleotide, and the 3' end of the first oligonucleotide is upstream of the 3' end of the second oligonucleotide, the first oligonucleotide may be called the "upstream" oligonucleotide and the second oligonucleotide may be called the "downstream" oligonucleotide.

As used herein, the terms "complementary" or "complementarity" are used in reference to polynucleotides (e.g., a sequence of two or more nucleotides (e.g., an oligonucleotide or a target nucleic acid)) related by the base-pairing rules. For example, the sequence "5'-A-G-T-3'," is complementary to the sequence "3'-T-C-A-5'."

Complementarity may be "partial," in which only some of the nucleic acid bases are matched according to the base pairing rules. Or, there may be "complete" or "total" complementarity between the nucleic acid bases. The degree of complementarity between nucleic acid strands has significant effects on the efficiency and strength of hybridization between nucleic acid strands. This is of particular importance in amplification reactions, as well as detection methods that depend upon the association of two or more nucleic acid strands. Either term may also be used in reference to individual nucleotides, especially within the context of polynucleotides. For example, a particular nucleotide within an oligonucleotide may be noted for its complementarity, or lack thereof, to a nucleotide within another nucleic acid sequence (e.g., a target sequence), in contrast or comparison to the complementarity between the rest of the oligonucleotide and the nucleic acid sequence.

The complement of a nucleic acid sequence as used herein refers to an oligonucleotide which, when aligned with the nucleic acid sequence such that the 5' end of one sequence is paired with the 3' end of the other, is in "antiparallel association." Certain bases not commonly found in natural nucleic acids may be included in the nucleic acids of the present invention and include, for example, inosine and 7-deazaguanine. Complementarity need not be perfect; stable duplexes may contain mismatched base pairs or unmatched bases. Those skilled in the art of nucleic acid technology can determine duplex stability empirically considering a number of variables including, for example, the length of the oligonucleotide, base composition and sequence of the oligonucleotide, ionic strength and incidence of mismatched base pairs.

The term "homology" refers to a degree of complementarity. There may be partial homology or complete homology (i.e., identity). A partially homologous sequence is one that is less than 100% identical to another sequence. A partially complementary sequence that is "substantially homologous" is a nucleic acid molecule that at least partially inhibits a completely complementary nucleic acid molecule from hybridizing to a target nucleic acid. The inhibition of hybridization of the completely complementary sequence to the target sequence may be examined using a hybridization assay (e.g., Southern or Northern blot, solution hybridization and the like) under conditions of low stringency. A substantially homologous sequence or probe will compete for and inhibit the binding (e.g., the hybridization) of a completely homologous nucleic acid molecule to a target under conditions of low stringency. This is not to say that conditions of low stringency are such that non-specific binding is permitted (e.g., the low stringency conditions may be such that the binding of two sequences to one another be a specific (e.g., selective) interaction). The absence of non-specific binding may be tested by the use of a second target that is substantially non-complementary (e.g., less than about 30% identity); in the absence of non-specific binding the probe will not hybridize to the second non-complementary target.

When used in reference to a double-stranded nucleic acid sequence such as a cDNA or genomic clone, the term "substantially homologous" refers to any probe that can hybridize to either or both strands of the double-stranded nucleic acid sequence under conditions of low stringency as described above.

When used in reference to a single-stranded nucleic acid sequence, the term "substantially homologous" refers to any probe that can hybridize (e.g., is complementary to) the single-stranded nucleic acid sequence under conditions of low stringency as described above.

5 The terms "target nucleic acid" and "target sequence," when used in reference to an invasive cleavage reaction, refer to a nucleic acid molecule containing a sequence that has at least partial complementarity with at least a first nucleic acid molecule (e.g. probe oligonucleotide) and may also have at least partial complementarity with a second nucleic acid molecule (e.g. INVADER oligonucleotide). Generally, the target nucleic acid (e.g.,
10 present within, isolated from, enriched from, or amplified from or within a sample (e.g., a biological or environmental sample)) is located within a target region and is identifiable via the successful formation of an invasive cleavage structure in combination with a first and second nucleic acid molecule (e.g., probe oligonucleotide and INVADER oligonucleotide) that is cleavable by a cleavage agent. Target nucleic acids from an
15 organism are not limited to genomic DNA and RNA. Target nucleic acids from an organism may comprise any nucleic acid species, including but not limited to genomic DNAs and RNAs, messenger RNAs, structural RNAs, ribosomal and tRNAs, and small RNAs such as snRNAs, siRNAs and microRNAs miRNAs). See, e.g., co-pending U.S. Patent Application Ser. No. 10/740,256, filed 12/18/03, which is incorporated herein by
20 reference in its entirety.

As used herein, the term "probe oligonucleotide," when used in reference to an invasive cleavage reaction, refers to an oligonucleotide that interacts with a target nucleic acid to form a cleavage structure in the presence or absence of an INVADER oligonucleotide. When annealed to the target nucleic acid, the probe oligonucleotide and
25 target form a cleavage structure and cleavage occurs within the probe oligonucleotide.

The term "INVADER oligonucleotide" refers to an oligonucleotide that hybridizes to a target nucleic acid at a location near the region of hybridization between a probe and the target nucleic acid, wherein the INVADER oligonucleotide comprises a portion (e.g., a chemical moiety, or nucleotide—whether complementary to that target or
30 not) that overlaps with the region of hybridization between the probe and target. In some

embodiments, the INVADER oligonucleotide contains sequences at its 3' end that are substantially the same as sequences located at the 5' end of a probe oligonucleotide.

The term "cassette," when used in reference to an invasive cleavage reaction, as used herein refers to an oligonucleotide or combination of oligonucleotides configured to
5 generate a detectable signal in response to cleavage of a probe oligonucleotide in an INVADER assay. In preferred embodiments, the cassette hybridizes to an cleavage product from cleavage of the probe oligonucleotide to form a second invasive cleavage structure, such that the cassette can then be cleaved.

In some embodiments, the cassette is a single oligonucleotide comprising a
10 hairpin portion (i.e., a region wherein one portion of the cassette oligonucleotide hybridizes to a second portion of the same oligonucleotide under reaction conditions, to form a duplex). In other embodiments, a cassette comprises at least two oligonucleotides comprising complementary portions that can form a duplex under reaction conditions. In preferred embodiments, the cassette comprises a label. In particularly preferred
15 embodiments, the cassette comprises labeled moieties that produce a fluorescence resonance energy transfer (FRET) effect.

An oligonucleotide is said to be present in "excess" relative to another oligonucleotide (or target nucleic acid sequence) if that oligonucleotide is present at a higher molar concentration than the other oligonucleotide (or target nucleic acid
20 sequence). When an oligonucleotide such as a probe oligonucleotide is present in a cleavage reaction in excess relative to the concentration of the complementary target nucleic acid sequence, the reaction may be used to indicate the amount of the target nucleic acid present. Typically, when present in excess, the probe oligonucleotide will be present in at least a 100-fold molar excess; typically at least 1 pmole of each probe
25 oligonucleotide would be used when the target nucleic acid sequence was present at about 10 fmoles or less.

As used herein, the term "gene" refers to a nucleic acid (e.g., DNA) sequence that comprises coding sequences necessary for the production of a polypeptide, precursor, or RNA (e.g., rRNA, tRNA). The polypeptide can be encoded by a full length coding
30 sequence or by any portion of the coding sequence so long as the desired activity or functional properties (e.g., enzymatic activity, ligand binding, signal transduction,

immunogenicity, etc.) of the full-length or fragment are retained. The term also encompasses the coding region of a structural gene and the sequences located adjacent to the coding region on both the 5' and 3' ends for a distance of about 1 kb or more on either end such that the gene corresponds to the length of the full-length mRNA. Sequences
5 located 5' of the coding region and present on the mRNA are referred to as 5' non-translated sequences. Sequences located 3' or downstream of the coding region and present on the mRNA are referred to as 3' non-translated sequences. The term "gene" encompasses both cDNA and genomic forms of a gene. A genomic form or clone of a gene contains the coding region interrupted with non-coding sequences termed "introns"
10 or "intervening regions" or "intervening sequences." Introns are segments of a gene that are transcribed into nuclear RNA (e.g., hnRNA); introns may contain regulatory elements (e.g., enhancers). Introns are removed or "spliced out" from the nuclear or primary transcript; introns therefore are absent in the messenger RNA (mRNA) transcript. The mRNA functions during translation to specify the sequence or order of amino acids in a
15 nascent polypeptide.

As used herein, the term "heterologous gene" refers to a gene that is not in its natural environment. For example, a heterologous gene includes a gene from one species introduced into another species (e.g., a viral or bacterial gene present within a human host (e.g., extrachromosomally or integrated into the host's DNA)). A heterologous gene also
20 includes a gene native to an organism that has been altered in some way (e.g., mutated, added in multiple copies, linked to non-native regulatory sequences, etc). In some embodiments, a heterologous gene can be distinguished from endogenous genes in that the heterologous gene sequences are typically joined to DNA sequences that are not found naturally associated with the gene sequences in the chromosome or are associated
25 with portions of the chromosome not found in nature (e.g., genes expressed in loci where the gene is not normally expressed).

As used herein, the term "gene expression" refers to the process of converting genetic information encoded in a gene into RNA (e.g., mRNA, rRNA, tRNA, or snRNA) through "transcription" of the gene (e.g., via the enzymatic action of an RNA
30 polymerase), and for protein encoding genes, into protein through "translation" of mRNA. Gene expression can be regulated at many stages in the process. "Up-

regulation" or "activation" refers to regulation that increases the production of gene expression products (e.g., RNA or protein), while "down-regulation" or "repression" refers to regulation that decrease production. Molecules (e.g., transcription factors) that are involved in up-regulation or down-regulation are often called "activators" and
5 "repressors," respectively.

In addition to containing introns, genomic forms of a gene may also include sequences located on both the 5' and 3' end of the sequences that are present on the RNA transcript. These sequences are referred to as "flanking" sequences or regions (e.g., these flanking sequences can be located 5' or 3' to the non-translated sequences present on the
10 mRNA transcript). The 5' flanking region may contain regulatory sequences such as promoters and enhancers that control or influence the transcription of the gene. The 3' flanking region may contain sequences that direct the termination of transcription, post-transcriptional cleavage and polyadenylation.

The term "wild-type" refers to a gene or gene product isolated from a naturally
15 occurring source. A wild-type gene is that which is most frequently observed in a population and is thus arbitrarily designed the "normal" or "wild-type" form of the gene. In contrast, the term "modified" or "mutant" refers to a gene or gene product that displays modifications in sequence and or functional properties (e.g., altered characteristics) when compared to the wild-type gene or gene product. It is noted that naturally occurring
20 mutants can be isolated (e.g., identified by the fact that they have altered characteristics (e.g., altered nucleic acid sequences) when compared to the wild-type gene or gene product).

The term "isolated" when used in relation to a nucleic acid (e.g., "an isolated oligonucleotide" or "isolated polynucleotide" or "an isolated nucleic acid sequence")
25 refers to a nucleic acid sequence that is separated from at least one component or contaminant with which it is ordinarily associated in its natural source. Thus, an isolated nucleic acid is present in a form or setting that is different from that in which it is found in nature. In contrast, non-isolated nucleic acids are nucleic acids such as DNA and RNA found in the state they exist in nature. For example, a given DNA sequence (e.g., a gene)
30 is found on the host cell chromosome in proximity to neighboring genes; RNA sequences, such as a specific mRNA sequence encoding a specific protein, are found in

the cell as a mixture with numerous other mRNAs that encode a multitude of proteins. However, isolated nucleic acid encoding a given protein includes, by way of example, such nucleic acid in cells ordinarily expressing the given protein where the nucleic acid is in a chromosomal location different from that of natural cells, or is otherwise flanked by a different nucleic acid sequence than that found in nature. The isolated nucleic acid, oligonucleotide, or polynucleotide may be present in single-stranded or double-stranded form. When an isolated nucleic acid, oligonucleotide or polynucleotide is to be utilized to express a protein, the oligonucleotide or polynucleotide will contain at a minimum the sense or coding strand (e.g., the oligonucleotide or polynucleotide may be single-stranded), but may contain both the sense and anti-sense strands (e.g., the oligonucleotide or polynucleotide may be double-stranded).

As used herein, the terms "purified" or "to purify" when used in reference to a sample (e.g., a molecule (e.g., a nucleic acid or amino acid sequence)) refers to removal (e.g., isolation and/or separation) of the sample from its natural environment. The term "substantially purified" refers to a sample (e.g., molecule (e.g. a nucleic acid or amino acid sequence) that has been removed (e.g., isolated and/or purified) from its natural environment and is at least 60% free, preferably 75% free, or most preferably 90% or more free from other components with which it is naturally associated. An "isolated polynucleotide" or "isolated oligonucleotide" may therefore be substantially purified if it is rendered free (e.g., 60%, 75% or more preferably 90% or more) from other components with which it is naturally associated.

The present invention is not limited to any particular means of purification (e.g., to generate purified or substantially purified molecules (e.g., nucleic acid sequences)). Indeed, a variety of purification techniques may be utilized including, but not limited to, centrifugation (e.g., isopycnic, rate-zonal, gradient, and differential centrifugation), electrophoresis (e.g., gel and capillary electrophoresis), gel filtration, matrix capture, charge capture, mass capture, antibody capture, magnetic separation, flow cytometry, and sequence-specific hybridization array capture.

As used herein, the term "hybridization" is used in reference to the pairing of complementary nucleic acids. Hybridization and the strength of hybridization (e.g., the strength of the association between the nucleic acids) is impacted by such factors as the

degree of complementary between the nucleic acids, stringency of the conditions involved, the T_m of the formed hybrid, and the G:C ratio within the nucleic acids. A single molecule that contains pairing of complementary nucleic acids within its structure is said to be "self-hybridized."

5 As used herein, the term " T_m " is used in reference to the "melting temperature." The melting temperature is the temperature at which a population of double-stranded nucleic acid molecules becomes half dissociated into single strands. Several equations for calculating the T_m of nucleic acids are well known in the art. As indicated by standard references, a simple estimate of the T_m value may be calculated by the equation:
10 $T_m = 81.5 + 0.41(\% G + C)$, when a nucleic acid is in aqueous solution at 1 M NaCl (See, e.g., Young and Anderson, (1985) in *Nucleic Acid Hybridisation: A Practical Approach* (Hames & Higgins, Eds.) pp 47-71, IRL Press, Oxford). Other computations for calculating T_m are known in the art and take structural and environmental, as well as sequence characteristics into account (See, e.g., Allawi, H.T. and SantaLucia, J., Jr.
15 *Biochemistry* 36, 10581-94 (1997)).

 As used herein, the term "INVADER assay reagents" refers to one or more reagents for detecting target sequences, said reagents comprising nucleic acid molecules capable of forming an invasive cleavage structure in the presence of the target sequence. In some embodiments, the INVADER assay reagents further comprise an agent for
20 detecting the presence of an invasive cleavage structure (e.g., a cleavage agent). In some embodiments, the nucleic acid molecules comprise first and second oligonucleotides, said first oligonucleotide comprising a 5' portion complementary to a first region of the target nucleic acid and said second oligonucleotide comprising a 3' portion and a 5' portion, said 5' portion complementary to a second region of the target nucleic acid downstream of and
25 contiguous to the first portion. In some embodiments, the 3' portion of the second oligonucleotide comprises a 3' terminal nucleotide not complementary to the target nucleic acid. In preferred embodiments, the 3' portion of the second oligonucleotide consists of a single nucleotide not complementary to the target nucleic acid. INVADER assay reagents may be found, for example, in U.S. Patent Nos. 5,846,717; 5,985,557;
30 5,994,069; 6,001,567; 6,913,881; and 6,090,543, WO 97/27214, WO 98/42873, U.S. Pat.

Publ. Nos. 20050014163, 20050074788, 2005016596, 20050186588, 20040203035, 20040018489, and 20050164177; U.S. Pat. Appln. Ser. No. 11/266,723; and Lyamichev et al., Nat. Biotech., 17:292 (1999), Hall et al., PNAS, USA, 97:8272 (2000), each of which is herein incorporated by reference in its entirety for all purposes.

5 In some embodiments, INVADER assay reagents are configured to detect a target nucleic acid sequence comprising first and second non-contiguous single-stranded regions separated by an intervening region comprising a double-stranded region. In certain embodiments, the INVADER assay reagents comprise a bridging oligonucleotide capable of binding to said first and second non-contiguous single-stranded regions of a
10 target nucleic acid sequence. In particularly preferred embodiments, either or both of said first and/or said second oligonucleotides of said INVADER assay reagents are bridging oligonucleotides.

In some embodiments, the INVADER assay reagents further comprise a solid support. For example, in some embodiments, the one or more oligonucleotides of the
15 assay reagents (e.g., first and/or second oligonucleotide, whether bridging or non-bridging) is attached to said solid support. The one or more oligonucleotides of the assay reagents may be linked to the solid support directly or indirectly (e.g., via a spacer molecule (e.g., an oligonucleotide)). Exemplary solid phase invasive cleavage reactions are described in U.S. Pat. Pub. Nos. 20050164177 and 20030143585, herein incorporated
20 by reference in their entireties.

As used herein, a "solid support" is any material that maintains its shape under assay conditions, and that can be separated from a liquid phase. The present invention is not limited by the type of solid support utilized. Indeed, a variety of solid supports are contemplated to be useful in the present invention including, but not limited to, a bead,
25 planar surface, controlled pore glass (CPG), a wafer, glass, silicon, plastic, paramagnetic bead, magnetic bead, latex bead, superparamagnetic bead, plurality of beads, microfluidic chip, a silicon chip, a microscope slide, a microplate well, a silica gel, a polymeric membrane, a particle, a derivatized plastic film, a glass bead, cotton, a plastic bead, an alumina gel, a polysaccharide, polyvinylchloride, polypropylene, polyethylene, nylon,
30 Sepharose, poly(acrylate), polystyrene, poly(acrylamide), polyol, agarose, agar, cellulose, dextran, starch, FICOLL, heparin, glycogen, amylopectin, mannan, inulin, nitrocellulose,

diazocellulose or starch, polymeric microparticle, polymeric membrane, polymeric gel, glass slide, styrene, multi-well plate, column, microarray, latex, hydrogel, porous 3D hydrophilic polymer matrix (e.g., HYDROGEL, Packard Instrument Company, Meriden, Conn.), fiber optic bundles and beads (e.g., BEADARRAY (Illumina, San Diego, Calif.), described in U.S. Pat. App. 20050164177), small particles, membranes, frits, slides, micromachined chips, alkanethiol-gold layers, non-porous surfaces, addressable arrays, and polynucleotide-immobilizing media (e.g., described in U.S. Pat. App. 20050191660). In some embodiments, the solid support is coated with a binding layer or material (e.g., gold or streptavidin).

In some embodiments, the INVADER assay reagents further comprise a buffer solution. In some preferred embodiments, the buffer solution comprises a source of divalent cations (e.g., Mn^{2+} and/or Mg^{2+} ions). Individual ingredients (e.g., oligonucleotides, enzymes, buffers, target nucleic acids) that collectively make up INVADER assay reagents are termed "INVADER assay reagent components."

In some embodiments, the INVADER assay reagents further comprise a third oligonucleotide complementary to a third portion of the target nucleic acid upstream of the first portion of the first target nucleic acid (e.g., a stacker oligonucleotides). In yet other embodiments, the INVADER assay reagents further comprise a target nucleic acid. In some embodiments, the INVADER assay reagents further comprise a second target nucleic acid. In yet other embodiments, the INVADER assay reagents further comprise a third oligonucleotide comprising a 5' portion complementary to a first region of the second target nucleic acid. In some specific embodiments, the 3' portion of the third oligonucleotide is covalently linked to the second target nucleic acid. In other specific embodiments, the second target nucleic acid further comprises a 5' portion, wherein the 5' portion of the second target nucleic acid is the third oligonucleotide. In still other embodiments, the INVADER assay reagents further comprise an ARRESTOR molecule (e.g., ARRESTOR oligonucleotide).

In some embodiments one or more of the INVADER assay reagents may be provided in a predispensed format (e.g., premeasured for use in a step of the procedure without re-measurement or re-dispensing). In some embodiments, selected INVADER assay reagent components are mixed and predispensed together. In preferred

embodiments, predisposed assay reagent components are predisposed and are provided in a reaction vessel (e.g., including, but not limited to, a reaction tube or a well (e.g., a microtiter plate)). In certain preferred embodiments, the INVADER assay reagents are provided in microfluidic devices such as those described in U.S. Pats., 6,627,159; 5 6,720,187; 6,734,401; and 6,814,935, as well as U.S. Pat. Pub. 2002/0064885, each of which is herein incorporated by reference in its entirety. In particularly preferred embodiments, predisposed INVADER assay reagent components are dried down (e.g., desiccated or lyophilized) in a reaction vessel.

In some embodiments, the INVADER assay reagents are provided as a kit. As 10 used herein, the term "kit" refers to any delivery system for delivering materials. In the context of reaction assays, such delivery systems include systems that allow for the storage, transport, or delivery of reaction reagents (e.g., oligonucleotides, enzymes, etc. in the appropriate containers) and/or supporting materials (e.g., buffers, written instructions for performing the assay etc.) from one location to another. For example, 15 kits include one or more enclosures (e.g., boxes) containing the relevant reaction reagents and/or supporting materials. As used herein, the term "fragmented kit" refers to delivery systems comprising two or more separate containers that each contains a subportion of the total kit components. The containers may be delivered to the intended recipient together or separately. For example, a first container may contain an enzyme for use in 20 an assay, while a second container contains oligonucleotides. The term "fragmented kit" is intended to encompass kits containing Analyte specific reagents (ASR's) regulated under section 520(e) of the Federal Food, Drug, and Cosmetic Act, but are not limited thereto. Indeed, any delivery system comprising two or more separate containers that each contains a subportion of the total kit components are included in the term 25 "fragmented kit." In contrast, a "combined kit" refers to a delivery system containing all of the components of a reaction assay in a single container (e.g., in a single box housing each of the desired components). The term "kit" includes both fragmented and combined kits.

In some embodiments, the present invention provides INVADER assay reagent 30 kits comprising one or more of the components necessary for practicing the present invention. For example, the present invention provides kits for storing or delivering the

enzymes and/or the reaction components necessary to practice an INVADER assay. The kit may include any and all components necessary or desired for assays including, but not limited to, the reagents themselves, buffers, control reagents (e.g., tissue samples, positive and negative control target oligonucleotides, etc.), solid supports, labels, written
5 and/or pictorial instructions and product information, inhibitors, labeling and/or detection reagents, package environmental controls (e.g., ice, desiccants, etc.), and the like. In some embodiments, the kits provide a sub-set of the required components, wherein it is expected that the user will supply the remaining components. In some embodiments, the kits comprise two or more separate containers wherein each container houses a subset of
10 the components to be delivered. For example, a first container (e.g., box) may contain an enzyme (e.g., structure specific cleavage enzyme in a suitable storage buffer and container), while a second box may contain oligonucleotides (e.g., INVADER oligonucleotides, probe oligonucleotides, control target oligonucleotides, etc.).

In some preferred embodiments, the INVADER assay reagents further comprise
15 reagents for detecting a nucleic acid cleavage product. In some embodiments, one or more oligonucleotides in the INVADER assay reagents comprise a label. In some preferred embodiments, said first oligonucleotide comprises a label. In other preferred embodiments, said third oligonucleotide comprises a label. In particularly preferred embodiments, the reagents comprise a first and/or a third oligonucleotide labeled with
20 moieties that produce a fluorescence resonance energy transfer (FRET) effect.

As used herein, the term "label" refers to any moiety (e.g., chemical species) that can be detected or can lead to a detectable response. In some preferred embodiments, detection of a label provides quantifiable information. Labels can be any known detectable moiety, such as, for example, a radioactive label (e.g., radionuclides), a ligand
25 (e.g., biotin or avidin), a chromophore (e.g., a dye or particle that imparts a detectable color), a hapten (e.g., digoxigenin), a mass label, latex beads, metal particles, a paramagnetic label, a luminescent compound (e.g., bioluminescent, phosphorescent or chemiluminescent labels) or a fluorescent compound.

A label may be joined, directly or indirectly, to an oligonucleotide or other
30 biological molecule. Direct labeling can occur through bonds or interactions that link the label to the oligonucleotide, including covalent bonds or non-covalent interactions such

as hydrogen bonding, hydrophobic and ionic interactions, or through formation of chelates or coordination complexes. Indirect labeling can occur through use of a bridging moiety or "linker", such as an antibody or additional oligonucleotide(s), which is/are either directly or indirectly labeled.

5 Labels can be used alone or in combination with moieties that can suppress (e.g., quench), excite, or transfer (e.g., shift) emission spectra (e.g., fluorescence resonance energy transfer (FRET)) of a label (e.g., a luminescent label).

As used herein, the term "FRET" refers to fluorescence resonance energy transfer, a process in which moieties (e.g., fluorophores) transfer energy (e.g., among themselves, or, from a fluorophore to a non-fluorophore (e.g., a quencher molecule)). In some
10 circumstances, FRET involves an excited donor fluorophore transferring energy to a lower-energy acceptor fluorophore via a short-range (e.g., about 10 nm or less) dipole-dipole interaction. In other circumstances, FRET involves a loss of fluorescence energy from a donor and an increase in fluorescence in an acceptor fluorophore. In still other
15 forms of FRET, energy can be exchanged from an excited donor fluorophore to a non-fluorescing molecule (e.g., a quenching molecule). FRET is known to those of skill in the art and has been described (See, e.g., Stryer et al., 1978, *Ann. Rev. Biochem.*, 47:819; Selvin, 1995, *Methods Enzymol.*, 246:300; Orpana, 2004 *Biomol Eng* 21, 45-50; Olivier, 2005 *Mutant Res* 573, 103-110, each of which is incorporated herein by reference in its
20 entirety).

As used herein, the term "donor" refers to a moiety (e.g., a fluorophore) that absorbs at a first wavelength and emits at a second, longer wavelength. The term "acceptor" refers to a moiety such as a fluorophore, chromophore, or quencher and that is able to absorb some or most of the emitted energy from the donor when it is near the
25 donor group (typically between 1-100 nm). An acceptor may have an absorption spectrum that overlaps the donor's emission spectrum. Generally, if the acceptor is a fluorophore, it then re-emits at a third, still longer wavelength; if it is a chromophore or quencher, it releases the energy absorbed from the donor without emitting a photon. In some preferred embodiments, alteration in energy levels of donor and/or acceptor
30 moieties are detected (e.g., via measuring energy transfer (e.g., by detecting light emission) between or from donors and/or acceptor moieties). In some preferred

embodiments, the emission spectrum of an acceptor moiety is distinct from the emission spectrum of a donor moiety such that emissions (e.g., of light and/or energy) from the moieties can be distinguished (e.g., spectrally resolved) from each other.

In some embodiments, a donor moiety is used in combination with multiple acceptor moieties. In a preferred embodiment, a donor moiety is used in combination with a non-fluorescing quencher moiety and with an acceptor moiety, such that when the donor moiety is close (e.g., between 1-100 nm, or more preferably, between 1-25 nm, or even more preferably around 10 nm or less) to the quencher, its excitation is transferred to the quencher moiety rather than the acceptor moiety, and when the quencher moiety is removed (e.g., by cleavage of a probe), donor moiety excitation is transferred to an acceptor moiety. In some preferred embodiments, emission from the acceptor moiety is detected (e.g., using wavelength shifting molecular beacons) (See, e.g., Tyagi, et al., *Nature Biotechnology* 18:1191 (2000); Mhlanga and Malmberg, 2001 *Methods* 25, 463-471; Olivier, 2005 *Mutant Res* 573, 103-110, and U.S. Pat. App. 20030228703, each of which is incorporated herein by reference in its entirety).

Detection of labels or a detectable response (e.g., provided by the labels) can be measured using a multitude of techniques, systems and methods known in the art. For example, a label may be detected because the label provides detectable fluorescence (e.g., simple fluorescence, FRET, time-resolved fluorescence, fluorescence quenching, fluorescence polarization, etc.), radioactivity, chemiluminescence, electrochemiluminescence, RAMAN, colorimetry, gravimetry, hybridization (e.g., to a sequence in a hybridization protection assay), X-ray diffraction or absorption, magnetism, enzymatic activity, characteristics of mass or behavior affected by mass (e.g., MALDI time-of-flight mass spectrometry), and the like.

A label may be a charged moiety (positive or negative charge) or alternatively, may be charge neutral. Labels can include or consist of nucleic acid or protein sequence, so long as the sequence comprising the label is detectable. In some embodiments, the label is not nucleic acid or protein.

In some embodiments, a label comprises a particle for detection. For example, in some embodiments, the particle is a phosphor particle. An example of a phosphor particle includes, but is not limited to, an up-converting phosphor particle (See, e.g.,

Ostermayer, Preparation and properties of infrared-to-visible conversion phosphors. Metall. Trans. 752, 747-755 (1971)). In some embodiments, rare earth-doped ceramic particles are used as phosphor particles. Phosphor particles may be detected by any suitable method, including but not limited to up-converting phosphor technology (UPT),
5 in which up-converting phosphors transfer low energy infrared (IR) radiation to high-energy visible light. Although an understanding of the mechanism is not necessary to practice the present invention and the present invention is not limited to any particular mechanism of action, in some embodiments the UPT up-converts infrared light to visible light by multi-photon absorption and subsequent emission of dopant-dependant
10 phosphorescence (See, e.g., U.S. Patent No. 6,399,397; van De Rijke, et al., Nature Biotechnol. 19(3):273-6 (2001); Corstjens, et al., IEE Proc. Nanobiotechnol. 152(2):64 (2005), each incorporated by reference herein in its entirety.

As used herein, the term "distinct" in reference to signals (e.g., of one or more labels) refers to signals that can be differentiated one from another, e.g., by spectral
15 properties such as fluorescence emission wavelength, color, absorbance, mass, size, fluorescence polarization properties, charge, etc., or by capability of interaction with another moiety, such as with a chemical reagent, an enzyme, an antibody, etc.

It will be apparent to one of skill in the art that there are a large number of methods (e.g., analytical procedures) that may be used to detect the presence or absence
20 of a nucleic acid sequence (e.g., a gene (e.g., wild-type, mutant (e.g., comprising one or more variant nucleotides at one or more positions), heterologous, etc.)). Such methods include, but are not limited to, nucleic acid discrimination techniques, amplification reactions and/or a signal generating systems. Such methods include, but are not limited to, DNA sequencing, hybridization sequencing, protein truncation test, single-strand
25 conformation polymorphism analysis (SSCP), denaturing gradient gel electrophoresis, temperature gradient gel electrophoresis, heteroduplex analysis, chemical mismatch cleavage, restriction enzyme digestion, and enzymatic mismatch cleavage, solid phase hybridization, dot blots, multiple allele specific diagnostic assays, reverse dot blots, oligonucleotide arrays (e.g., DNA chips), solution phase hybridization (e.g., TAQMAN
30 (See, e.g., U.S. Pat. Nos. 5,210,015 and 5,487,972, each of which is herein incorporated by reference in its entirety) and molecular beacons (See, e.g., Tyagi et al. 1996 Nature

Biotech, 14, 303 and Int. App. WO 95/13399, herein incorporated by reference), extension based amplification (e.g., amplification refractory mutation systems, amplification refractory mutation system linear extensions (See, e.g., EP 332435, herein incorporated by reference in its entirety), competitive oligonucleotide priming system (See, e.g., Gibbs et al., 1989 Nucleic Acids Research 17, 2347, herein incorporated by reference in its entirety), mini sequencing, restriction fragment length polymorphism, restriction site generating PCR, oligonucleotide ligation assay and many others described herein and elsewhere.

10 DESCRIPTION OF THE DRAWINGS

Figure 1 shows the dynamic range of detection of a first target using methods of the present invention.

Figure 2 shows the dynamic range of detection of a second target using methods of the present invention.

15 Figure 3 shows an overview of one exemplary embodiment of the INVADER assay.

Figure 4 shows an Excel graph showing results of an assay using two probe concentrations covering six orders of magnitude of target concentration, from 10 copies to 10,000,000. Data points represent: The resulting signal from 2 sets of probes and FRET cassettes in a combined reaction as described in Example 5 (Diamonds); the resulting signal from only the 1X probe (Squares); and the resulting signal from only the 1/150X probe (Triangles).

Figure 5 shows an Excel graph showing results of an assay using three probe concentrations covering six orders of magnitude of target concentration, from 10 copies to 10,000,000, as described in Example 6. Individual contributions of each of the three probe concentrations is shown with additional lines. Data points represent: The resulting signal from three probe concentrations (Squares); the 1X probe in isolation (Diamonds); the 1/125X probe in isolation (Squares); and the 1/1000X probe in isolation (Triangles).

Figure 6 shows an Excel graph showing the results of varying the length of incubation time of the invasive cleavage reaction, as described in Example 7. Data points represent: The resulting signal from 30 minutes incubation (Diamonds); 1 hour

incubation (Squares); 2 hours incubation (Triangles); 4 hours incubation (Circles); and 8 hours incubation (Asterisks).

Figure 7 shows an Excel graph showing the detection of two different targets in simultaneous multiplex using the methods of the present invention as described in

5 Example 8. CMV and EBV were simultaneously detected across over six orders of magnitude of dynamic range, in the same reaction vessel. In this experiment, CMV (Diamonds) and EBV (Squares) were detected in multiplex over a range from approximately 20 to 1,000,000 copies per reaction.

Figure 8 shows an Excel graph showing detection over nine orders of magnitude
10 of target concentration, as described in Example 9. Combining single strand amplification with standard PCR, and detecting both products simultaneously with two sets of two probes broadened the dynamic range to at least nine orders of magnitude (10-10,000,000,000 copies of target detected by a single reaction setup). As shown, basic target detection with no single strand product (detecting only double stranded PCR
15 product; Diamonds); basic reaction with the addition of a 1X probe to detect single strand product (Squares); the basic reaction with the addition of a 1/125X probe to detect single strand product (Triangles); and the basic reaction with the addition of both 1X and 1/125X probes to detect both PCR and single strand products.

Figure 9 shows an Excel graph showing the detection of RNA using multiple
20 probe concentrations, as described in Example 10. As shown, RNA target detection with 1X probe alone (Diamonds); 1/125X probe alone (Squares), and of both 1X and 1/125X probes to detect RT-PCR products (Triangles).

DESCRIPTION OF THE INVENTION

25 The present invention provides systems, methods and kits for increasing the dynamic range of detection of a target nucleic acid in a sample. In particular, the present invention provides methods and kits for increasing the dynamic range of detection of a target nucleic acid in a sample through the use of one or more probe oligonucleotides (e.g., analyte-specific probe oligonucleotides).

30 In some embodiments, the present invention achieves greater dynamic range of detection through the use of differential levels of amplification of regions of a target

nucleic acid (e.g., no amplification, linear amplification at one or more efficiencies, and/or exponential amplification at one or more efficiencies). In some embodiments, the present invention achieves greater dynamic range of detection through the use of probes with different hybridization properties to one or more analyte-specific regions of a target
5 nucleic acid or target nucleic acids. In some embodiments, the present invention achieves greater dynamic range of detection through the use of different signal generation methods. In some embodiments, the present invention achieves greater dynamic range of detection through the use of different signal detection methods. In preferred
10 embodiments, combinations of two or more of the methods are employed. For example, in some preferred embodiments, two or more probes (e.g., three, four, etc.) are contacted with first and second amplicons obtained via different levels of amplification. In some such embodiments, each probe generates the same type of signal so that one simply detects total signal generated by the reactions. The collective signal permits detection of target nucleic acid over a broad dynamic range. For example, experiments conducted
15 during the development of the present invention have demonstrated the ability to detect target nucleic acid from samples differing in over eight logs of copy number of target nucleic acid originally present in the sample.

In certain embodiments, the present invention provides methodologies for expansion of the dynamic range of hybridization assays, such as serial invasive cleavage
20 assays. In some embodiments, the upper limit of dynamic range may be expanded by the use of an additional probe that is present in the reaction at a lower concentration than another probe. In some embodiments, this additional probe will hybridize to the same region of the target. For invasive cleavage reactions, this probe may contain a different arm, or flap, sequence that is released after cleavage. In certain embodiments related to
25 invasive cleavage assays, a second FRET cassette will also be added to the reaction with the appropriate sequence to detect those cleaved flaps from the additional probe. Generally the concentration of the second FRET cassette is about the same as the first FRET cassette. For example, in certain embodiments, if probe B is present in the reaction at 100-fold lower concentration than probe A, this will enable the detection of
30 target nucleic acid when it is present at concentrations above the upper limit of detection of probe A. In this manner, each additional probe, present at 100-fold lower

concentrations will enable the detection of two additional orders of magnitude of probe concentration. This methodology is not limited to two primary probes, but may be expanded to three or more. Preferably, the methods are combined with amplification methods where one part of the target is amplified to a different level than a second part of the target.

As mentioned above, in certain embodiments, two probes are employed that are present at different concentrations that detect the same target nucleic acid molecule across a broad range of concentrations. In some embodiments involving invasive cleavage reactions, each of the two primary probes contain the same analyte specific region (ASR) but have different flap regions. Each of these two flap regions, when cleaved, reports to a different FRET cassette or other reporter sequence or system. In some embodiments, the two FRET cassettes both contain the same fluorophore molecule. In this system, an increase in dynamic range is achieved without the use of multiple different fluorophores. This system, therefore, offers a cost advantage over multiple fluorophore systems. Furthermore, expansion of dynamic range with a single fluorophore allows for multiplexing with multiple fluorophores for detection of different targets in the same vessel across a broader dynamic range than was previously feasible.

In certain embodiments, the concentration of each primary probe is present at 100-fold difference relative to each other, and the concentration of the two FRET cassettes are present at equivalent concentrations. In certain embodiments, as an example, the dynamic range with each of the primary probes present individually may be 10^4 - 10^6 and 10^6 - 10^8 , respectively, while the dynamic range of the assay when both are present at the requisite different concentrations may be 10^4 - 10^8 . The dynamic range of the serial invasive cleavage assay may be further expanded by the use of further additional primary probes, each present at different concentrations. In this manner, three, four, five or more primary probes, each having the same ASR and different flaps may report to the same number of different FRET cassettes, each reporting the same color or detection format. Such a combination of primary probes enables the expansion of the dynamic range to cover 5, 6, 7, 8, 9, 10, 11, 12, 13, 14 or 15 orders of magnitude.

The methods of the present invention are not limited by the type of target nucleic acid. For example, the target nucleic acid may include, for example, nucleic acid

material prepared from viruses having an RNA genome. Typically, the RNA target sequence will be converted to a cDNA molecule through the action of a reverse transcriptase, and then detected by the nucleic acid detection assay. Incorporation of the methods of the present invention will increase the dynamic range of detection of RNA target sequences to a breadth not previously feasible.

The methods of the present invention may be combined with amplification methods (e.g., PCR) to extend the lower limit of detection down to the theoretical limit of amplification, on the order of 1 copy per reaction vessel. Using this approach, the dynamic range of nucleic acid detection assay may be, for example, from 1 to 10^7 copies using a single set of reaction conditions and probe combinations in each reaction vessel being compared.

Additionally, methods of the present invention involve differential pre-amplification of target species prior to the detection assay. In certain embodiments, the use of differential semi-nested PCR using primers of different melting temperatures will result in a mixed population of different species, each containing the target region detected by the detection assay. The species present in higher numbers in the sample after this step can be detected by the probes present at lower concentration within its dynamic range, and the species present in lower numbers in the sample can be detected by the primary probe present at higher concentration within its dynamic range, as explained above. In addition, a population of target molecules present at different concentrations can also be generated by simultaneously combining linear and exponential amplification (or other types of amplification that lead to different levels of amplification). For example, two target-derived amplicons, both containing the target region detected by the nucleic acid detection assay, would be generated by producing one with a single PCR primer (for linear amplification) and the other with two PCR primers (for exponential amplification). As above, the different concentrations of targets can be detected with multiple primary probes tailored to detect those concentrations within their dynamic range. Further, non-amplified and amplified DNA can also be simultaneously detected using the above-described combination of probes.

Differential pre-amplification may also comprise multiple similar amplifications (e.g., exponential amplifications) that are performed at different efficiencies so as to

allow expanded dynamic range. By way of example and not to be limited to any particular embodiment or mechanism, target sequences may be selected such that one region to be amplified comprises few of a selected nucleotide, while another region to be amplified comprises an abundance of the same nucleotide. Pre-amplification under
5 conditions wherein the dNTP required to replicate the selected nucleotide is limited or omitted will favor amplification of the sequence that is largely free of the limiting nucleotide. Conditions can be selected to allow the other sequence to amplify inefficiently, *e.g.*, by mis-incorporating bases. This is but one way in which differential pre-amplification can be configured to allow.

10 In some embodiments, additional probes are used to further expand the dynamic range (*e.g.*, three probes of different concentrations that each bind to the same analyte-specific region). In some embodiments, the method detects one or more probes under each of three distinct amplification conditions: *e.g.*, one probe or probe set that detects exponentially amplified target nucleic acid; one probe or probe set that detects linearly
15 amplified target nucleic acid; and one probe or probe set that detects unamplified target nucleic acid. Additional amplification conditions may also be used (*e.g.*, exponential amplification using primers or other reaction conditions that provide different amplification efficiency per cycle—*e.g.*, a first set that is 90% efficient per cycle and a second set that is 70% efficient per cycle).

20 Where PCR or other amplification techniques are used, it may be desirable to use buffers and other agents and reaction conditions that minimize limitations of the respective amplification techniques. For example, where PCR is used, in some embodiments, a short amplicon is used. In some embodiments, the amplicon is less than one kilobase in length, although the present invention is not limited to such amplicons.
25 In some embodiments, where the target nucleic is RNA, the amplicon is less than 100 bases, although the present invention is not so limited.

Accordingly, in some embodiments, the present invention provides methods and compositions for performing probe hybridization assays. In some embodiments, the method utilizes a primary or first probe and preferably at least one additional probe
30 having different hybridization characteristics with respect to a target sequence than the primary probe. In some embodiments, a single probe that provides enhanced dynamic

range is utilized. In preferred embodiments, the compositions and methods of the present invention utilize a combination of two or more probe oligonucleotides to increase the dynamic range of detection of the amount of a target nucleic acid present in a sample. In preferred embodiments, combinations of two or more probe oligonucleotides include a mixture of probe oligonucleotides with varying degrees of hybridization to a target nucleic acid (e.g., frequency of occupation of a hybridization site). Exemplary probe oligonucleotides of the present invention are described in greater detail below.

In some embodiments, three or more probes are used (e.g., four, five, six, etc.). Two or more of the probes may be configured to hybridize to the same region of the target nucleic acid. However, one or more of the probes may be configured to hybridize to a second region of the target nucleic acid or to a different target nucleic acid. In some embodiments, the pluralities of different probes are configured to generate a detectable signal directly or indirectly. In some embodiments, the different probes use the same type of label so that the detected signal is an additive accumulation of the signal from the first and second probes. In some such embodiments, the user of the method observes the signal throughout the broader dynamic range without knowing or needing to know the contribution provided by each type or probe.

Using such systems and methods, detection of a target nucleic acid can be achieved through a very extensive dynamic range. In some embodiments, this permits detection of target nucleic acids without the need to amplify the target nucleic acid or without the need to extensively amplify the target nucleic acid. However, the systems and methods may further be employed with amplification methods, where desired. As described herein, the systems and methods of the present invention have been exemplified with a combination of polymerase chain extension amplification and invasive cleavage-based detection. Such methods experimentally demonstrated successful detection of target nucleic acids having over an eight-log difference in starting concentration. Thus, the systems and methods of the present invention are exceptionally well suited to the detection of target nucleic acids whose concentration differs dramatically from sample to sample. For example, patients infected with viruses such as HCV and HIV differ greatly the copy number of virus target nucleic acid present in sample (e.g., blood) from very low copy (as few as one copy) to very high copy (millions

to billions of copies or more). The ability of a single detection system to simultaneously detect viral target nucleic acid throughout this range is greatly desired. The present invention provides systems and methods that find use for such detection.

The compositions and methods are useful for the detection and quantitation of a wide variety of nucleic acid targets. The compositions and methods of the present invention are particularly useful for the quantitation of viral target nucleic acids (e.g., viral pathogens). Exemplary viral nucleic acids for which a clinical or research need for the detection of a large range of viral concentrations (e.g., viral load) include, but are not limited to, human immunodeficiency virus (HIV) and other retroviruses, hepatitis C virus (HCV), hepatitis B virus (HBV), hepatitis A virus (HAV), human cytomegalovirus, (CMV), Epstein bar virus (EBV), human papilloma virus (HPV), herpes simplex virus (HSV), Varicella Zoster Virus (VZV), bacteriophages (e.g., phage lambda), adenoviruses, and lentiviruses. In other embodiments, the compositions and methods of the present invention find use in the detection of bacteria (e.g., pathogens or bacteria important in commercial and research applications). Examples include, but are not limited to, *Chlamydia sp.*, *N. gonorrhea*, and group B *streptococcus*.

In some embodiments, the target sequence is a synthetic sequence. For example, a fragment generated in an enzymatic reaction (e.g., a restriction fragment, a cleaved flap from an invasive cleavage reaction, etc.) can be considered a target sequence. In some such embodiments, the detection of such a molecule indirectly detects a separate target nucleic acid from which the synthetic sequence was generated. For example, in an invasive cleavage reaction, a cleaved flap from a primary reaction may be detected with first and second probes that are FRET cassettes. The FRET cassettes differ in some characteristic (e.g., length, etc.) such that the cleaved flap differentially hybridizes to the first and second probes. By using both FRET cassettes (or a third, fourth, etc.), the dynamic range of the reaction is improved.

The quantitation of target nucleic acids using the methods and compositions of the present invention are utilized in a variety of clinical and research applications. For example, in some embodiments, the detection assays with increased dynamic range of the present invention are utilized in the detection and quantitation of viral pathogens in human samples. The detection assays of the present invention are suitable for use with a

variety of purified and unpurified samples including, but not limited to, urine, stool, lymph, whole blood, and serum. In preferred embodiments, the detection assays of the present invention are suitable for use in the presence of host cells.

5 In other embodiments, the detection assays of the present invention find use in research applications including, but not limited to, drug screening (e.g., for drugs against viral pathogens), animal models of disease, and in vitro quantitation of target nucleic acid (e.g., bacterial, viral, or genomic nucleic acids).

10 The probe oligonucleotides of the present invention find use in a variety of nucleic acid detection assays including, but not limited to, those described below. It should be understood that any nucleic acid detection method that employs hybridization can benefit from the systems and methods of the present invention.

I. Probe Oligonucleotides

15 In some embodiments, the present invention provides methods for altering (e.g., increasing) the dynamic range of a nucleic acid detection assay by altering probe oligonucleotides. In some embodiments, the present invention provides combinations of two or more probe oligonucleotides for use in the same detection assay. The present invention is not limited by the manner in which probes are modified to alter hybridization characteristics. Certain exemplary embodiments are provided below.

20

A. Mismatch Probes

In some embodiments, the present invention provides probes with one or more (preferably one) mismatch with the target sequence. The present invention is not limited to a particular mechanism. Indeed, an understanding of the mechanism is not necessary to practice the present invention. Nonetheless, it is contemplated that the presence of one or more mismatches allows the probe to bind to the target, but with a reduced affinity as compared to a corresponding probe lacking mismatches. This decreases the percent of the time that the mismatch probe occupies the target site, thus decreasing the signal generated (or increasing the signal, depending on the detection system used). The decrease in signal allows the detection assay to remain linear or accurate for quantitation at a higher target concentration.

25
30

In some embodiments, mismatch probes are utilized in combination with completely complementary probes. The completely complementary probes occupy the target-binding site a higher percentage of the time than the mismatch probes and thus generate more signal. The higher signal allows for the detection of lower concentrations of target nucleic acid. The use of both probes increases the dynamic range of the detection assay. In particular, as described above, it increases the linearity through a broader concentration of target molecules.

Example 1 and Figures 1 and 2 demonstrate how the use of mismatch probes can increase the dynamic range of an assay. A combination of match and mismatch probes was used in an INVADER assay to detect target nucleic acids. The mismatch probe increased the dynamic range by up to 16-fold over the use of a single completely complementary probe.

B. Lower probe concentrations

In other embodiments, the present invention provides a combination of probe concentrations to increase the dynamic range of a detection assay. In some embodiments a combination of two or more probe oligonucleotides, each of which is at a different concentration, is utilized. The probes present at a lower concentration generate a lower signal and are thus suitable for detecting higher target concentrations. The probes present at a higher concentration generate a higher signal and are thus suitable for detecting a lower concentration of target nucleic acids. By utilizing two or more probes at a range of concentrations, a broader dynamic range of target concentrations can be detected.

When probes are attached to a solid surface, lower probe concentration can be achieved, in some embodiments, through the use of different densities of probes attached to particular detection zones on the solid surface. For example, a first probe detection zone has a first density of the probe and a second probe detection zone has a lower density of the probe. Detection at the two detection zones provides enhanced dynamic range. In some embodiments, both detection zones generate the same type of signal and the total signal from the solid surface is detected (e.g., in real-time) to detect the target nucleic acid through an expanded dynamic range.

Example 1 and Figures 1 and 2 demonstrate how the use of multiple probes present at different concentrations probes can increase the dynamic range of an assay. A combination of concentrations of probes was used in an INVADER assay to detect target nucleic acids. The use of multiple concentrations of probes increased the dynamic range of the assay over the use of a single probe.

C. Charge Modified Probes

In other embodiments, the present invention utilizes charge modified probes to alter binding efficiency of probes (See e.g., US Patent 6,780,982, herein incorporated by reference in its entirety for all reasons). In some embodiments, the charge modified probes comprise "charge tags." Positively charged moieties need not always carry a positive charge. As used herein, the term "positively charged moiety" refers to a chemical structure that possesses a net positive charge under the reaction conditions of its intended use (e.g., when attached to a molecule of interest under the pH of the desired reaction conditions). Indeed, in some preferred embodiments of the present invention, the positively charged moiety does not carry a positive charge until it is introduced to the appropriate reaction conditions. This can also be thought of as "pH-dependent" and "pH-independent" positive charges. pH-dependent charges are those that possess the charge only under certain pH conditions, while pH-independent charges are those that possess a charge regardless of the pH conditions.

The positively charged moieties, or "charge tags," when attached to another entity, can be represented by the formula:

X-Y

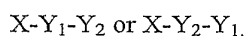
where X is the entity (e.g., a solid support, a nucleic acid molecule, etc.) and Y is the charge tag. The charge tags can be attached to other entities through any suitable means (e.g., covalent bonds, ionic interactions, etc.) either directly or through an intermediate (e.g., through a linking group). In preferred embodiments, where X is a nucleic acid molecule, the charge tag is attached to either the 3' or 5' end of the nucleic acid molecule.

The charge tags may contain a variety of components. For example, the charge tag Y can be represented by the formula:



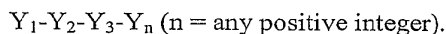
- 5 where Y_1 comprises a chemical component that provides the positive charge to the charge tag and where Y_2 is another desired component. Y_2 may be, for example, a dye, another chemical component that provides a positive charge to the charge tag, a functional group for attachment of other molecules to the charge tag, a nucleotide, etc. Where such a structure is attached to another entity, X, either Y_1 or Y_2 may be attached to X.

10



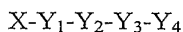
- The charge tags are not limited to two components. Charge tags may comprise any number of desired components. For example, the charge tag can be represented by the formula:

15



- where any of the Y_x groups comprises a chemical component that provides the positive charge to the charge tag and where the other Y groups are any other desired components. For example, in some embodiments, the present invention provides compositions of the structure:

20



25

- where X is an entity attached to the charge tag (e.g., a solid support, a nucleic acid molecule, etc.) and where Y_1 is a dye, Y_2 is a chemical component that provides the positive charge to the charge, Y_3 is a component containing a functional group that allows the attachment of other molecules, and Y_4 is a second chemical component that provides a positive charge. The identity of each of Y_1 - Y_4 can be interchanged (i.e., the present invention is not limited by the order of the components).

30

The present invention is not limited by the nature of the chemical components that provides the positive charge to the charge tag. Such chemical components include, but are not limited to, amines (primary, secondary, and tertiary amines), ammoniums, and phosphoniums. The chemical components may also comprise chemical complexes that
5 entrap or are otherwise associated with one or more positively charged metal ions.

In preferred embodiments of the present invention, charge tags are attached to nucleic acid molecules (e.g., DNA molecules). The charge tags may be synthesized directly onto a nucleic molecule or may be synthesized, for example, on a solid support or in liquid phase and then attached to a nucleic acid molecule or any other desired
10 molecule. In some preferred embodiments of the present invention, charge tags that are attached to nucleic acid molecules comprise one or more components synthesized by H-phosphonate chemistry, by incorporation of novel phosphoramidites, or a combination of both. For example, compositions of the present invention include structures such as:



where [X] is a nucleic acid molecule and [Y . . .] is a charge tag. In some embodiments, Y₁ is a dye, Y₂ is synthesized using H-phosphonate chemistry and comprises a chemical component that provides a positive charge to the charge tag, Y₃ is a positively charged
20 phosphoramidite, and Y₄ is a nucleotide or polynucleotide. Any of the Y components are interchangeable with one another.

As discussed above, one or more components of a charge tag can be synthesized using H-phosphonate chemistry. Production of charge tag using the methods described herein provides a convenient and flexible modular approach for the design of a wide
25 variety of charge tags. Since its introduction, solid phase H-phosphonate chemistry (B.C. Froehler, Methods in Molecular Biology, 20:33, S. Agrawal, Ed. Humana Press; Totowa, New Jersey[1993]) has been recognized as an efficient tool in the chemical synthesis of natural, modified and labeled oligonucleotides and DNA probes. Those skilled in the art know that this approach allows for the synthesis of the oligonucleotide fragments with a
30 fully modified phosphodiester backbone (e.g., oligonucleotide phosphorothioates; Froehler [1993], *supra*) or the synthesis of oligonucleotide fragments in which only

specific positions of the phosphodiester backbone are modified (Agrawal, *et al.*, Proc.Natl.Acad.Sci USA , 85:7079 [1988], Froehler, Tetrahedron Lett. 27:5575[1986], Froehler, *et al.*, Nucl.Acids Res. 16:4831 [1988]). The use of H-phosphonate chemistry allows for the introduction of different types of modifications into the oligonucleotide molecule (Agrawal, *et al.*, Froehler[1986], *supra*, Letsinger, *et al.*, J.Am. Chem.Soc., 110:4470 [1988], Agrawal and Zamecnik, Nucl. Acid Res. 18:5419 [1990], Handong, *et al.*, Bioconjugate Chem. 8:49 [1997], Vinogradov, *et al.*, Bioconjugate Chem. 7:3 [1995], Schultz, *et al.*, Tetrahedron Lett. 36:8407 [1995]), however the replacement of the phosphodiester linkage by the phosphoramidate linkage is one of the most frequent changes due to its effectiveness and synthetic flexibility. Froehler and Letsinger were among first to use this approach in the synthesis of modified oligonucleotides in which phosphodiester linkages were fully or partially replaced by the phosphoramidate linkages bearing positively charged groups (e.g., tertiary amino groups; Froehler [1986], Froehler, *et al.*, [1988], and Letsinger, *et al.*, *supra*).

In some embodiments of the present invention, charge tags are generated using H-phosphonate chemistry. The charge tags may be assembled on the end of a nucleic acid molecule or may be synthesized separately and attached to a nucleic acid molecule. Any suitable phosphorylating agent may be used in the synthesis of the charge tag. For example, the component to be added may contain the structure:



where A is a protecting group, B is any desired functional group (e.g., a functional group that provides a positive charge to the charge tag), and P is a chemical group containing phosphorous. In preferred embodiments, B comprises a chemical group that is capable of providing a positive charge to the charge tag. However, in some embodiments B is a functional group that allows post-synthetic attachment of a positively charged group to the charge tag.

In other embodiments, positively charged phosphoramidites (PCP) and neutral phosphoramidites (NP) are utilized to introduce both positive charge and structure modulation into the synthesized charge-balanced CRE probe (See e.g., US Patent 6,780,982, herein incorporated by reference in its entirety).

5 Standard coupling protocol with the use of the phosphoramidite reagents (which are compatible with the chemical synthesis of oligonucleotides) is associated with the introduction, into the growing molecule, of one negative charge per each performed coupling step, due to the formation of the phosphodiester linkage.

10 **D. Nucleic Acid Modification Agents**

The present invention is not limited to the use of charge tags as modifiers of probe hybridization efficiency. Any internal (e.g., to the probe) or external agent that alters the hybridization strength of probe binding is suitable for use with the methods and compositions of the present invention.

15 In some embodiments, the present invention provides probes comprising intercalating agents. Intercalating agents are agents that are capable of inserting themselves between the successive bases in DNA. In some embodiments, intercalating agents alter the binding properties of nucleic acid probes.

20 Examples of intercalating agents are known in the art and include, but are not limited to, ethidium bromide, psoralen and derivatives, acridines, proflavine, acridine orange, acriflavine, fluorcoumanin, ellipticine, daunomycin, chloroquine, distamycin D, chromomycin, homidium, mithramycin, ruthenium polypyridyls, and anthramycin.

25 In other embodiments, minor groove DNA binding agents are utilized to modify (e.g., increase or decrease) the hybridization efficiency of probes. Examples of minor groove binding agents include, but are not limited to, duocarmycins (See e.g., Boger, Pure & Appl. Chem., Vol. 66, No. 4, pp. 837-844, herein incorporated by reference in its entirety), netropsin, bisbenzimidazole, aromatic diamidines, lexitropsins, distamycin, and organic dications, based on unfused-aromatic systems (See e.g., US Patent 6,613,787, herein incorporated by reference in its entirety).

30 In still further embodiments, modified bases are utilized to alter the hybridization efficiency of probes. For example, in some embodiments, modified bases that include

charged groups are utilized. Examples include, but are not limited to, the substitution of a "t" nucleotide with "amino-T" in a probe and other modified nucleotides.

In yet other embodiments, one or more probe nucleotides are modified by the covalent attachment of groups that alter the hybridization properties of the probe.

5 Examples include, but are not limited to, the attachment of amino acids to nucleotides.

In yet other embodiments, probe oligonucleotides with base analogues are utilized to alter the hybridization characteristics of probes. For example, in some embodiments; nucleotides that do not form hydrogen bonds but that still participate in base stacking are utilized. Examples include but are not limited to non-polar, aromatic nucleoside analogs
10 such as 2,4-difluorotoluene and "universal" bases such as 5-nitroindole and 3-nitropyrrole. In other embodiments, base analogs that retain hydrogen bonding ability are utilized (See e.g., US Patent application US20040106108A1 and WO 04/065550A3, each of which is herein incorporated by reference in its entirety for all purposes).

15 E. Probe Length

In yet other embodiments, probe length is altered in order to alter the hybridization characteristics of a probe. For example, in some embodiments, two or more probes that hybridize to the same target sequence and share the same sequence are utilized. In some embodiments, one of the probes is shorter by one, two, three, or four or
20 more bases. It is preferred that the probes be truncated from one or both ends. Thus, the probes share sequence in all regions except the truncated 3' or 5' ends. It is contemplated that the shorter probes will anneal with decreased hybridization efficiency and will thus be useful in the detection of higher copy numbers of target sequences than the full length probe. In preferred embodiments, a combination of full length and truncated probes is
25 utilized to give the maximum range of target concentration detection. In some embodiments, the same length is employed, but the probe is split into two or more portions connected by linkers. Such probes hybridize with different affinity depending on a variety of factors, including secondary structure of the target nucleic acid in regions in which the probes or probe fragments hybridize.

30

F. Secondary Structure

In some embodiments, probes that comprise secondary structure are utilized to alter the hybridization efficiency of the probe. For example, in some embodiments, two or more probes are designed to hybridize the same target sequence. One of the probes is designed to have minimal secondary structure. Additional probes are designed that retain target sequence recognition, but that have secondary structure. It is contemplated that the probes with secondary structure will exhibit decreased hybridization properties and will thus be suitable for the detection of large copy numbers of target sequence. The combination of probes lacking and containing secondary structure serves to detect a larger dynamic range of target nucleic acids than a single probe. Likewise, probes that hybridize to regions of the target nucleic acid that differ in secondary structure may be used. For example, a probe that has 18 of 18 bases that bind to linear target nucleic acid will hybridize differently than a similar probe shifted two bases over on target nucleic acid such that the two bases on the end of the probe correspond to a region of the target nucleic acid occupied in an internal hairpin structure or other secondary structure.

G. Competitor Oligonucleotides

In yet other embodiments, additional oligonucleotides are utilized to modify hybridization efficiency of probes. For example, in some embodiments, two probes that recognize the same target sequence are designed. One of the probes further comprises additional nucleic acid sequence (e.g., at the 3' or 5' end) that does not hybridize to the target sequence. Competitor oligonucleotides are designed to hybridize to the extra region. The binding of the competitor oligonucleotide decreases the hybridization efficiency of the probe to the target. The combination of probes with and without competitor binding sequences serves to detect a larger dynamic range of target nucleic acids than a single probe.

H. Reaction Conditions

In still further embodiments, reaction conditions are modified to alter probe hybridization characteristics. For example, in some embodiments, identical probes are utilized in separate reaction vessels, chamber, or wells. One reaction vessel utilizes "standard" reaction conditions for the detection assay (e.g., those supplied by the

manufacturer or known in the art). The other reaction vessel comprises altered reaction conditions that increase or decrease the hybridization efficiency of the probe. Examples of parameters that affect nucleic acid hybridization conditions include, but are not limited to, ionic strength, buffer composition, pH, and additives (e.g., glycerol, polyethylene glycol, proteins).

I. Stacking Oligonucleotides

In still further embodiments, adjacently hybridizing oligonucleotides are used to alter probe hybridization characteristics. When short strands of nucleic acid align contiguously along a longer strand, the hybridization of each is stabilized by the hybridization of the neighboring fragments because the base pairs can stack along the helix as though the backbone was, in fact, uninterrupted. This cooperativity of binding can give each segment a stability of interaction in excess of what would be expected for the segment hybridizing to the longer nucleic acid alone. In the event of a perturbation in the cooperative binding, e.g., by a mismatch at or near the junction between the contiguous duplexes, this cooperativity can be reduced or eliminated. In some embodiments of the present invention, probes are configured to cooperate in distinct ways with one or more adjacently hybridizing oligonucleotides, so as to provide probes having different hybridization characteristics. In some embodiments, a probe comprises one or more mismatched bases at near the junction with the adjacent oligonucleotide, so as to alter or disrupt cooperativity of binding, as compared to a probe lacking the mismatches. In other embodiments, a probe comprises one or more base analogs selected to reduce stacking interactions with adjacent bases. In yet other embodiments, it is envisioned that gaps of one or more nucleotides (e.g., by the use of truncated probes) are used to alter cooperativity and thus alter hybridization characteristics. The use of a combination of probes that have a range of cooperativities of binding with an adjacently hybridized oligonucleotide, and thus having a range of different hybridization stabilities on the target, serves to detect a larger dynamic range of target nucleic acids than a single probe.

J. Multiplex Assays

In some embodiments utilizing multiple nucleic acid probes, the probes are utilized in a biplex or multiplex assay in which a plurality of probes is included in the same reaction vessel. In some embodiments, each probe in a biplex assay comprises a differently detectable label. For example, some embodiments, each probe in a set
5 comprises a different fluorescent label that fluoresces at a different wavelength. Many known probe binding assays are suitable for use in a multiplex format. Methods for performing multiplex assays that are unique to the particular assay format are described below.

10 **K. Others**

Any other method for altering the hybridization of characteristics of a probe may be used with the present invention. Other examples include, but are not limited to: use of sequences in probes or targets that render the sequence susceptible to differential hybridization behavior in response to buffer conditions (e.g., the use of guanosine-
15 quartets) or protein/nucleic acid interactions (e.g., by creating binding sites for nucleic acid binding proteins or enzyme that bind or alter nucleic acid sequences); use of dangling ends (e.g., for dangling-end stabilization and stacking); attachment of iron or other magnetic agents to allow concentration of the nucleic acid in a magnetic field; use of agents that titrate out a specific probe; and the like.

20 One may also use different labeling techniques to achieve a differential detection of signal, independent of the hybridization properties of the probe. For example, the location of labels and quenchers in a FRET detection system may be altered between first and second probes to alter the amount of signal detected from the probes. FRET signaling can also be affected by many other parameters, including, but not limited to, the
25 use of additional chemical moieties that influence the amount of quenching and the use of secondary structure in the probes. Additional methods for altering signal detection include the use of a helper oligonucleotide that is provided at low concentration, that when bound to a target occupied by a probe of the invention, changes the wavelength or otherwise alters the detectable aspects of the probe. The concentration of the helper can
30 be configured to only allow detection the alteration when a particular threshold level of

probe is hybridized to target. Any method or system that permits differential detection of hybridization events may be used in the systems and methods of the present invention.

II. Detection Assays

5 The present invention is not limited to a particular detection assay. Any number of suitable detection assays may be utilized. In some embodiments, the present invention provides methods and compositions for the detection of DNA or RNA (e.g., viral RNA). In some embodiments, the detection assays described below are suitable for direct
10 detection of RNA. In other embodiments, RNA is reverse transcribed (e.g., using a reverse transcriptase enzyme such as AMV or MMLV) into DNA and the detection assay is performed on the corresponding DNA. Methods for reverse transcription are known in the art. In some embodiments, a single enzyme having both reverse transcriptase and polymerase activities is used.

Exemplary assays that find use with the methods of the present invention are
15 described below.

A. Invader Assay

In some embodiments, the methods and compositions of the present invention are used to increase the dynamic range of the INVADER assay. The INVADER assay
20 provides means for forming a nucleic acid cleavage structure that is dependent upon the presence of a target nucleic acid and cleaving the nucleic acid cleavage structure so as to release distinctive cleavage products. 5' nuclease activity, for example, is used to cleave the target-dependent cleavage structure and the resulting cleavage products are indicative of the presence of specific target nucleic acid sequences in the sample. When two strands
25 of nucleic acid, or oligonucleotides, both hybridize to a target nucleic acid strand such that they form an overlapping invasive cleavage structure, as described below, invasive cleavage can occur. Through the interaction of a cleavage agent (e.g., a 5' nuclease) and the upstream oligonucleotide, the cleavage agent can be made to cleave the downstream oligonucleotide at an internal site in such a way that a distinctive fragment is produced.
30 Such embodiments have been termed the INVADER assay (Third Wave Technologies, Madison, WI) and are described in U.S. Patent Appl. Nos. 5,846,717, 5,985,557,

5,994,069, 6,001,567, and 6,090,543, WO 97/27214, WO 98/42873, Lyamichev et al., Nat. Biotech., 17:292 (1999), Hall et al., PNAS, USA, 97:8272 (2000), each of which is herein incorporated by reference in their entirety for all purposes.

The INVADER assay detects hybridization of probes to a target by enzymatic cleavage of specific structures by structure specific enzymes (*See*, INVADER assays, Third Wave Technologies; *See e.g.*, U.S. Patent Nos. 5,846,717; 6,090,543; 6,001,567; 5,985,557; 6,090,543; 5,994,069; Lyamichev et al., Nat. Biotech., 17:292 (1999), Hall et al., PNAS, USA, 97:8272 (2000), WO97/27214 and WO98/42873, each of which is herein incorporated by reference in their entirety for all purposes).

The INVADER assay detects specific DNA and RNA sequences by using structure-specific enzymes (*e.g.* FEN endonucleases) to cleave a complex formed by the hybridization of overlapping oligonucleotide probes. Elevated temperature and an excess of one of the probes enable multiple probes to be cleaved for each target sequence present without temperature cycling. In some embodiments, these cleaved probes then direct cleavage of a second labeled probe. The secondary probe oligonucleotide can be 5'-end labeled with fluorescent that is quenched by an internal dye. Upon cleavage, the de-quenched fluorescent labeled product may be detected using a standard fluorescence plate reader.

The INVADER assay detects specific target sequences in unamplified, as well as amplified, RNA and DNA including genomic DNA. In the embodiments shown schematically in Figure 3, the INVADER assay uses two cascading steps (a primary and a secondary reaction) both to generate and then to amplify the target-specific signal. For convenience, the alleles in the following discussion are described as wild-type (WT) and mutant (MT), even though this terminology does not apply to all genetic variations or target sequences. In the primary reaction (Figure 3, panel A), the WT primary probe and the INVADER oligonucleotide hybridize in tandem to the target nucleic acid to form an overlapping structure. An unpaired "flap" is included on the 5' end of the WT primary probe. A structure-specific enzyme (*e.g.* the CLEAVASE enzyme, Third Wave Technologies) recognizes the overlap and cleaves off the unpaired flap, releasing it as a target-specific product. In the secondary reaction, this cleaved product serves as an INVADER oligonucleotide on the WT fluorescence resonance energy transfer (WT-

FRET) probe to again create the structure recognized by the structure specific enzyme (panel A). When the two dyes on a single FRET probe are separated by cleavage (indicated by the arrow in Figure 3), a detectable fluorescent signal above background fluorescence is produced. Consequently, cleavage of this second structure results in an increase in fluorescence, indicating the presence of the WT allele (or mutant allele if the assay is configured for the mutant allele to generate the detectable signal). In some embodiments, FRET probes having different labels (*e.g.* resolvable by difference in emission or excitation wavelengths, or resolvable by time-resolved fluorescence detection) are provided for each allele or locus to be detected, such that the different alleles or loci can be detected in a single reaction. In such embodiments, the primary probe sets and the different FRET probes may be combined in a single assay, allowing comparison of the signals from each allele or locus in the same sample.

If the primary probe oligonucleotide and the target nucleotide sequence do not match perfectly at the cleavage site (*e.g.*, as with the MT primary probe and the WT target, Figure 3, panel B), the overlapped structure does not form and cleavage is suppressed. The structure specific enzyme (*e.g.*, CLEAVASE VIII enzyme, Third Wave Technologies) used cleaves the overlapped structure more efficiently (*e.g.* at least 340-fold) than the non-overlapping structure, allowing excellent discrimination of the alleles.

The probes turn over without temperature cycling to produce many signals per target (*i.e.*, linear signal amplification). Similarly, each target-specific product can enable the cleavage of many FRET probes.

The primary INVADER assay reaction is directed against the target DNA or RNA being detected. The target DNA is the limiting component in the first invasive cleavage, since the INVADER and primary probe are supplied in molar excess. In the second invasive cleavage, it is the released flap that is limiting. When these two cleavage reactions are performed sequentially, the fluorescence signal from the composite reaction accumulates linearly with respect to the target DNA amount.

In certain embodiments, the INVADER assay, or other nucleotide detection assays, are performed with accessible site designed oligonucleotides and/or bridging oligonucleotides. Such methods, procedures and compositions are described in U.S. Pat.

6,194,149, WO9850403, and WO0198537, all of which are specifically incorporated by reference in their entireties.

In certain embodiments, the target nucleic acid sequence is amplified prior to detection (e.g. such that synthetic nucleic acid is generated). In some embodiments, the target nucleic acid comprises genomic DNA. In other embodiments, the target nucleic acid comprises synthetic DNA or RNA. In some preferred embodiments, synthetic DNA within a sample is created using a purified polymerase. In some preferred embodiments, creation of synthetic DNA using a purified polymerase comprises the use of PCR. In other preferred embodiments, creation of synthetic DNA using a purified DNA polymerase, suitable for use with the methods of the present invention, comprises use of rolling circle amplification, (e.g., as in U.S. Pat. Nos. 6,210,884, 6,183,960 and 6,235,502, herein incorporated by reference in their entireties). In other preferred embodiments, creation of synthetic DNA comprises copying genomic DNA by priming from a plurality of sites on a genomic DNA sample. In some embodiments, priming from a plurality of sites on a genomic DNA sample comprises using short (e.g., fewer than about 8 nucleotides) oligonucleotide primers. In other embodiments, priming from a plurality of sites on a genomic DNA comprises extension of 3' ends in nicked, double-stranded genomic DNA (i.e., where a 3' hydroxyl group has been made available for extension by breakage or cleavage of one strand of a double stranded region of DNA). Some examples of making synthetic DNA using a purified polymerase on nicked genomic DNAs, suitable for use with the methods and compositions of the present invention, are provided in U.S. Patent Nos. 6,117,634, issued September 12, 2000, and 6,197,557, issued March 6, 2001, and in PCT application WO 98/39485, each incorporated by reference herein in their entireties for all purposes.

In some embodiments, synthetic DNA suitable for use with the methods and compositions of the present invention is made using a purified polymerase on multiply-primed genomic or other DNA, as provided, e.g., in U.S. Patent Nos. 6,291,187, and 6,323,009, and in PCT applications WO 01/88190 and WO 02/00934, each herein incorporated by reference in their entireties for all purposes. In these embodiments, amplification of DNA such as genomic DNA is accomplished using a DNA polymerase, such as the highly processive Φ 29 polymerase (as described, e.g., in US Patent Nos.

5,198,543 and 5,001,050, each herein incorporated by reference in their entireties for all purposes) in combination with exonuclease-resistant random primers, such as hexamers.

The present invention further provides assays in which the target nucleic acid is reused or recycled during multiple rounds of hybridization with oligonucleotide probes and cleavage of the probes without the need to use temperature cycling (i.e., for periodic
5 denaturation of target nucleic acid strands) or nucleic acid synthesis (i.e., for the polymerization-based displacement of target or probe nucleic acid strands). When a cleavage reaction is run under conditions in which the probes are continuously replaced on the target strand (e.g. through probe-probe displacement or through an equilibrium
10 between probe/target association and disassociation, or through a combination comprising these mechanisms, (Reynaldo et al., J. Mol. Biol. 97: 511-520 (2000)), multiple probes can hybridize to the same target, allowing multiple cleavages, and the generation of multiple cleavage products.

As described above, in some embodiments, the present invention provides
15 methods of utilizing the INVADER assay to quantitate the amount of target nucleic present in a sample. In some embodiments, the dynamic range of INVADER assays is increased using mismatch probes, alone or in combination with completely homologous probes. It is preferred that the mismatch is not present at the site of cleavage by the cleavage enzyme. In other embodiments, dynamic range of the INVADER assay is
20 increased by using probes of multiple concentrations. In preferred embodiments, each probe in a multiple probe INVADER assay comprises a different label, allowing the reactions to be run in the same well or tube of the reaction vessel and detected simultaneously. However, the probes may also share the same label, permitting the combined signal to be interpreted as one detection event. In some preferred
25 embodiments, a real time assay, in which signal is measured continuously or at time intervals, is utilized. In other embodiments, a single end-point detection is taken at a desired time point. In yet other embodiment, two or more time point readings are taken.

In other embodiments, composite or split probe oligonucleotides are utilized to increase the dynamic range is utilized in the INVADER-directed cleavage assay. For
30 example, the probe oligonucleotide may be split into two oligonucleotides that anneal in a contiguous and adjacent manner along a target oligonucleotide. The probe