(19) DANMARK

(10) **DK/EP 3604525 T3**

Patent- og
Varemærkestyrelsen

(12) Oversættelse af
europæisk patentskrift

(51) Int.Cl.: ***C 12 N 15/10 (2006.01)*** ***C 40 B 40/08 (2006.01)***

(45) Oversættelsen bekendtgjort den: **2021-05-31**

(80) Dato for Den Europæiske Patentmyndigheds
bekendtgørelse om meddelelse af patentet: **2021-03-10**

(86) Europæisk ansøgning nr.: **18186948.8**

(86) Europæisk indleveringsdag: **2018-08-02**

(87) Den europæiske ansøgnings publiceringsdag: **2020-02-05**

(84) Designerede stater: **AL AT BE BG CH CY CZ DE DK EE ES FI FR GB GR HR HU IE IS IT LI LT LU LV MC MK MT NL NO PL PT RO RS SE SI SK SM TR**

(73) Patenthaver: **TU Dresden, Helmholtzstrasse 10, 01069 Dresden, Tyskland**

(72) Opfinder: **Cui, Meiying, Pfotenhauerstr. 88 / 107, 01307 Dresden, Tyskland**
ANDRADE, Helena, Thielaustraße 3, 01309 Dresden, Tyskland
Heiden, Stephan, Zwickauer Straße 127, 01187 Dresden, Tyskland
REDDAVIDE, Francesco, Martin-Luther-Straße 18, 01099 Dresden, Tyskland
ZHANG, Yixin, Wormser Straße 63, 01309 Dresden, Tyskland

(74) Fuldmægtig i Danmark: **Dennemeyer & Associates S.A, P.O. Box 700425, DE-81304 Munich, Tyskland**

(54) Benævnelse: **Fremgangsmåde til tilvejebringelse af et dna-indkodet bibliotek, dna-inkodet bibliotek og fremgangsmåde til at afkode et dna-indkodet bibliotek**

(56) Fremdragne publikationer:
WO-A1-2014/065756
WO-A1-2017/075292
WO-A1-2018/005720
WO-A2-2016/008822
CN-A- 108 220 392
YIZHOU LI ET AL: "Quantitative PCR is a Valuable Tool to Monitor the Performance of DNA-Encoded Chemical Library Selections", CHEMBIOCHEM - A EUROPEAN JOURNAL OF CHEMICAL BIOLOGY., vol. 18, no. 9, 16 March 2017 (2017-03-16) , pages 848-852, XP055535160, DE ISSN: 1439-4227, DOI: 10.1002/cbic.201600626
RAPHAEL M. FRANZINI ET AL: "DNA-Encoded Chemical Libraries: Advancing beyond Conventional Small-Molecule Libraries", ACCOUNTS OF CHEMICAL RESEARCH, vol. 47, no. 4, 28 March 2014 (2014-03-28) , pages 1247-1255, XP055211954, ISSN: 0001-4842, DOI: 10.1021/ar400284t
J?RG SCHEUERMANN ET AL: "Dual-pharmacophore DNA-encoded chemical libraries", CURRENT OPINION IN CHEMICAL BIOLOGY, vol. 26, 1 June 2015 (2015-06-01), pages 99-103, XP055211951, ISSN: 1367-5931, DOI: 10.1016/j.cbpa.2015.02.021
MORENO WICHERT ET AL: "Dual-display of small molecules enables the discovery of ligand pairs and facilitates affinity maturation", NATURE CHEMISTRY, vol. 7, no. 3, 26 January 2015 (2015-01-26), pages 241-249, XP055184296, ISSN: 1755-4330, DOI: 10.1038/nchem.2158
FABIAN BULLER ET AL: "Drug discovery with DNA-encoded chemical libraries", BIOCONJUGATE CHEMISTRY, AMERICAN CHEMICAL SOCIETY, US, vol. 21, no. 9, 15 September 2010 (2010-09-15), pages 1571-1580, XP002695068, ISSN: 1043-1802, DOI: 10.1021/BC1001483 [retrieved on 2010-08-03]

Fortsættes ...

ALIX I CHAN ET AL: "Novel selection methods for DNA-encoded chemical libraries", CURRENT OPINION IN CHEMICAL BIOLOGY, vol. 26, 1 June 2015 (2015-06-01), pages 55-61, XP055535155, GB ISSN: 1367-5931, DOI: 10.1016/j.cbpa.2015.02.010
Helena Andrade ET AL: "Modelling Large, Dynamic, and Heterogeneous Populations Using DNA Libraries", bioRxiv, 5 March 2018 (2018-03-05), XP055534080, DOI: 10.1101/276873 Retrieved from the Internet: URL:https://www.biorxiv.org/content/biorxi v/early/2018/03/05/276873.full.pdf
LUCA MANNOCCI ET AL: "20 years of DNA-encoded chemical libraries", CHEMICAL COMMUNICATIONS, ROYAL SOCIETY OF CHEMISTRY, vol. 47, no. 48, 28 December 2011 (2011-12-28), pages 12747-12753, XP002695069, ISSN: 1359-7345, DOI: 10.1039/C1CC15634A [retrieved on 2011-11-14]
SCHEUERMANN J ET AL: "DNA-encoded chemical libraries for the discovery of MMP-3 inhibitors", BIOCONJUGATE CHEMISTRY, AMERICAN CHEMICAL SOCIETY, US, vol. 19, no. 3, 1 March 2008 (2008-03-01), pages 778-785, XP002737640, ISSN: 1043-1802, DOI: 10.1021/BC7004347 [retrieved on 2008-02-07]
JEAN PIERRE DAGUER ET AL: "DNA-templated combinatorial assembly of small molecule fragments amenable to selection/amplification cycles", CHEMICAL SCIENCE, vol. 2, no. 4, 24 January 2011 (2011-01-24), page 625, XP055212512, ISSN: 2041-6520, DOI: 10.1039/c0sc00574f
MIKE THOMPSON: "Adapting DNA-encoded library technology for fragment-based drug discovery", MEDICINAL CHEMISTRY, vol. 07, 1 June 2017 (2017-06-01), pages 83-83, XP055535198, DOI: 10.4172/2161-0444-C1-031

DESCRIPTION

[0001] A method for providing a DNA-encoding library, the DNA-encoding library and a method of decoding a DNA-encoded library are presented. Many different DNA molecules are synthesized which differ from each other by comprising different DNA barcode sequences, wherein each DNA barcode sequence comprises at least a first coding region DNA sequence comprising at least a first part, a second part and a third part, wherein the second part is located between the first and third part and the second part differs between all the DNA molecules by at least two nucleotides. Each of the many different DNA molecules is bonded to at least a specific substance forming different DNA-substance conjugates, wherein the DNA-substance conjugates differ from each other by the specific substance and by their DNA molecules, wherein the first part and the third part encode information regarding the second part of the first coding region and wherein a certain first part and/or a certain third part uniquely codes for a certain group of DNA-substance conjugates which is smaller than the group of all DNA-substance conjugates in the DNA-encoded library. The DNA-encoded library has the advantage that, for example after an enrichment experiment performed with the library, the library may be decoded in a faster and less expensive manner than known DNA-encoded libraries.

[0002] In drug discovery which aims at identifying high affinity binders from a pool of molecules, it is known in the prior art to use a DNA-encoded library ("DEL"). Ideally, said DEL can mimic the function-information relationship of cells, such as T cells and B cells in adaptive immunity and peptide/protein-display technologies (e.g. phage display, ribosome display, yeast display). In T cells, B cells and/or phages, the functions (mediated e.g. by proteins expressed on cell surface) and associated information (coded e.g. by genetic information) are both confined in individual cells. The function-information relationship can be studied even if there is only a single copy of an individual cell presented in a given cell mixture.

[0003] A DEL is composed of a pool of different molecules, each being a conjugate between a small organic molecule and a specific DNA sequence (a so-called "DNA barcode"), thus realizing a direct physical connection between function (function of the small organic molecule by its chemical structure) and information (information about the type of small organic molecule coded by the DNA sequence). The DNA sequences are designed to identify the associated chemical structures using various technologies, e.g. Sanger sequencing, DNA array and/or high throughput sequencing.

[0004] Although PCR (polymerase chain reaction) is mainly used to amplify the selected compounds, PCR and real-time PCR (rtPCR) can also be used as a validation technique to check whether and at which abundance one particular DNA barcode is present, e.g. before and/or after the DEL has been subjected to a selection experiment. A selection experiment seeks to enrich certain conjugates between small organic molecules and DNA barcodes based on isolating said conjugates after they have bound to one or more desired target(s). Since the conjugates are enriched, a DEL selection experiment may be regarded as an experiment

enriching certain DNA barcodes, namely those coding for small organic molecules having a high binding affinity to the target(s).

[0005] Similar like in the phage display technology, a usual DEL selection experiment provides tens to hundreds of DNA barcodes (DNA sequences) in one round of selection (one run). However, different from phage display technology, which regularly reveals organic molecules which are highly specific and potent binders (i.e. the k_D to the target lies in the pM to nM range), a DEL selection experiment frequently also reveals DNA barcodes coding for small organic molecules which are only moderate binders (e.g. the k_D to the target lies in the low to medium µM range).

[0006] In principle, Sanger sequencing provides a tool to decode DNA barcodes which have been found in a DEL selection experiment.

[0007] However, Sanger sequencing has the disadvantage that the throughput is low, i.e. the "reading" of the DNA barcode consumes a lot of time thus represents an uneconomical readout.

[0008] A further disadvantage of Sanger sequencing is its low sensitivity when analyzing complex mixtures of different DNA sequences. Assuming a DEL selection experiment using a DEL comprising 1 million different compounds, one compound is usually enriched 1000 times over the average and 100 sequences will be obtained from Sanger sequencing. In this case, there will be an approx. 90% chance that that one particular compound is not identified by the selection experiment, i.e. escapes identification, because its presence is not revealed by Sanger sequencing.

[0009] Moreover, even if a certain DNA barcode (e.g. coding for a certain small organic molecule) appears once in the enrichment process, Sanger sequence may identify said DNA barcode as coding for a small organic molecule which binds to the target. However, Sanger sequencing cannot reveal whether the identification of this specific small organic molecule has been a random event (i.e. an accidental hit) or is actually statistically significant (i.e. a true hit). In short, Sanger sequencing also suffers the disadvantage that false positives may not be distinguished from true positives without oversampling. While oversampling in the context of Sanger sequencing is apparently very important to obtain statistically meaningful results for hit identification in the decoding process (readout), it has become clear that Sanger sequencing is far from being efficient.

[0010] A DNA array provides an alternative solution to decode a DNA barcode sequence of binders identified in a DEL selection experiment. Since each DNA barcode sequence is associated with a certain physical location and evaluated according to its fluorescence intensity, the measurement avoids the requirement of oversampling using Sanger sequencing.

[0011] However, although fully complimentary sequences lead to highest signal intensity, strong background noise associated with mismatching DNA sequence interaction prevents the

use of this method to decode a large library of DNA barcode sequences. For example, with a library of only few hundreds compounds each having a DNA barcode sequence, great effort needs to be made to distinguish a specific pair from mismatching and background noises. In short, the DNA array identification method also suffers the disadvantage that false positives may not be distinguished well from true positives. In other words, the systemic error of this identification method is high.

[0012] High throughput sequencing ("HTS") has become the standard technology for decoding a DEL after a selection experiment. HTS applies a similar principle like Sanger sequencing and uses the count of a particular sequence as an indicator of its enrichment. Millions of sequence reads resulting from HTS make oversampling possible, even when a DEL of a relatively large size is used.

[0013] However, like the DNA array approach, HTS can only provide a semiquantitative analysis of a selection experiment, because it was found that the counts of DNA barcode sequences and the measured affinity of its bound small organic molecule to the desired target(s) only show a poor correlation. The identified poor correlation has not yet been fully understood. Principally, it could be caused by a low synthetic quality of the DNA barcodes, while biases during the PCR and sequencing process may play a role. In summary, HTS is prone to reveal many false positive hits during the identification process, i.e. the systemic error of this identification method is high. Moreover, as the size of a DEL has been increasing gradually in recent years, HTS will no longer fulfill the requirement of oversampling when a DEL has started to comprise billions of compounds.

[0014] Furthermore, although HTS has become cheaper in the last years, it is still very expensive for many academic researchers. The outsourced sequencing tasks normally take a few weeks while researchers have no control over the sequencing experiments.

[0015] PCR and rtPCR have been used in the prior art to overcome the problems of the Sanger sequencing, DNA array and HTS identification methods. The advantages of both PCR and rtPCR are that primer pairs can be designed for a certain code. In other words, different primers may be used which themselves can carry a "code" in the sense that some of the primers bind (at least partially) to certain codes and some other do not. Additionally, rtPCR has the advantage over PCR that it will reveal a difference between a positive control and a negative control (in real time) and thus allows a better discrimination between false and true positives.

[0016] However, although rtPCR provides a quantitative analysis of a DEL selection process, it can only be designed for a limited number of codes and compounds. Therefore rtPCR suffers the disadvantage that it cannot be used for decoding the results of de novo selection experiments.

[0017] Starting herefrom, it was the object of the present invention to provide a method for encoding and decoding DNA barcodes having been enriched in a selection experiment with a

DNA encoded library, wherein the method shall overcome the deficiencies of the prior art identification methods. Specifically, the method should be a facile, cost-efficient, quantitative, highly sensitive (i.e. be capable for revealing also weak binders), highly specific (i.e. be capable to reveal more true positives than false positives) and suitable to decode de novo selection experiments.

[0018] The object is solved by the method for providing a DNA-encoded library having the features of claim 1, the DNA-encoded library having the features of claim 5 and the method of decoding said DNA-encoded library having the features of claim 9. The subject-matter of the dependent claims illustrates advantageous embodiments of the invention.

[0019] According to the invention, a method for providing a DNA-encoded library (DEL) is provided, the method comprising

1. a) synthesizing many different DNA molecules which differ from each other by comprising different DNA barcode sequences, wherein each DNA barcode sequence comprises at least a first coding region DNA sequence comprising at least a first part, a second part and a third part, wherein the second part is located between the first and third part and the second part differs between all the DNA molecules by at least two nucleotides; and
2. b) bonding each of the many different DNA molecules to at least a specific substance forming different DNA-substance conjugates, wherein the DNA-substance conjugates differ from each other by the specific substance and by their DNA molecules;

characterized in that the first part and the third part encode information regarding the second part of the first coding region, wherein a certain first part and/or a certain third part uniquely codes for a certain group of DNA-substance conjugates which is smaller than the group of all DNA-substance conjugates in the DNA-encoded library.

[0020] The advantage of the DNA-encoded library ("DEL") provided by the inventive method is that both the first and third part of the DNA barcode sequence each encode for a certain subgroup of DNA-substance conjugates within the DEL. In qPCR, a primer binding to the first part of the DNA barcode sequence will give a strong signal (strong amplification) if the subgroup of DNA-substance conjugates for which the first part encodes (e.g. transcription factors) has been enriched in a previous selection experiment performed with the DEL. The same is true for the third part of the DNA barcode sequence, i.e. a primer binding to the third part of the DNA barcode sequence will give a strong signal (strong amplification) if the subgroup of DNA-substance conjugates for which the third part encodes (e.g. zinc finger proteins) has been enriched in a previous selection experiment performed with the DEL. If strong signal is obtained for both a primer binding to the first part and a primer binding to the third part after qPCR, the skilled person knows that DNA-substance conjugates belonging to both subgroups (e.g. zinc finger transcription factors) have been strongly enriched. The skilled person obtains this information only via qPCR with the inventive DEL and suitable primers, i.e. the skilled person does not have to perform a DNA sequencing. This allows a much faster and less expensive decoding of a DNA-encoded library after a selection experiment performed with

said library.

[0021] The DNA-encoded library can be used to construct many two-dimensional matrices in which different first primers which bind to different first parts of the barcode form the rows of the matrix, different second primers which bind to different second parts of the barcode are the columns of the matrix and the signal intensity after qPCR with each primer pair is given in each each field of the matrix (crossing point between rows and colums). The signal intensity obtained for each primer pairing allows a deconvolution of the mixture of DNA barcodes, i.e. of the DEL after the selection experiment. The possibility to deconvolute the mixture of DNA barcodes strongly improves the specificity of the identification method, i.e. its capability of distinguishing true positive hits from false positive hits and allows a quick determination of "hits" even without performing DNA sequencing.

[0022] Since performing qPCR without DNA sequencing is not expensive, it is estimated that a full decoding experiment will cost only approx. 50 €. Thus, the DEL produced with the inventive method allows a very cost-efficient "hit" detection after an enrichment experiment with said DEL and needs very little invenstment in instrumentation. Additionally, the DEL allows to obtain a more quantitative information on the abundance of a certain DNA barcodes after a selection experiment as compared with previously known DELs.

[0023] The inventive method can be characterized in that

1. i) the first coding region DNA sequence comprises at least a fourth part, wherein the second part is located between the fourth and third part and wherein both the combination of the first part and the fourth part and the combination of the first part and the third part of the first coding region encode information about the second part of the first coding region; and
2. ii) each barcode sequence comprises at least a second coding region DNA sequence comprising at least a first part, a second part, a third part, and a fourth part, wherein the second part is located between the fourth and third part and the second part differs between all the DNA molecules by at least two nucleotides, wherein both the combination of the first part and the fourth part and the combination of the first part and the third part of the second coding region encode information about the second part of the second coding region;

wherein a certain combination of a first part and fourth part in a certain coding region uniquely codes for a certain group of DNA-substance conjugates which is smaller than the group of all DNA-substance conjugates which is encoded by the first part alone.

[0024] In this embodiment of the invention, the DNA-encoded library can be used to construct more two-dimensional matrices because an additional primer can be used which anneals to the fourth part of the DNA barcode and because a further coding region with different four parts is present. Only with one single run of qPCR, very detailed information is obtained about the specific groups of DNA-substance conjugates that have been enriched in the selection experiment with the DEL.

[0025] Furthermore, the inventive method can be characterized in that

1. i) each barcode sequence comprises at least a second coding region DNA sequence comprising at least a first part, a second part, a third part, and a fourth part, wherein the second part is located between the fourth and third part and the second part differs between all the DNA molecules by at least two nucleotides, wherein both the combination of the first part and the fourth part and the combination of the first part and the third part of the second coding region encode information about the second part of the second coding region; and
2. ii) each barcode sequence comprises at least a third coding region DNA sequence comprising at least a first part, a second part, a third part, and a fourth part, wherein the second part is located between the fourth and third part and the second part differs between all the DNA molecules by at least two nucleotides, wherein both the combination of the first part and the fourth part and the combination of the first part and third part and the of the third coding region encode information about the second part of the third coding region;

wherein a certain combination of a first part and fourth part in a certain coding region uniquely codes for a certain group of DNA-substance conjugates which is smaller than the group of DNA-substance conjugates which is encoded by the first part.

[0026] In view of the further coding region and the separation of at least one coding region into five parts, more different primers can be used in one single qPCR and within one single run of qPCR, very detailed information can obtained which specific groups of DNA-substance conjugates have been enriched in the selection experiment with the DEL.

[0027] In a preferred embodiment of the invention, at least one coding region DNA sequence, optionally all coding region DNA sequences, comprise at least a first part, a second part, a third part, a fourth part and a fifth part, wherein the second part is located between the fourth and fifth part and the second part differs between all the DNA molecules by at least two nucleotides, wherein the combination of the first part and the fourth part and the combination of the fifth part and the third part of the coding region encode information about the second part of the coding region, preferably of all coding regions, wherein a certain combination of a first part and fourth part uniquely codes for a certain group of DNA-substance conjugates which is smaller than the group of DNA-substance conjugates which is encoded by the first part alone, and wherein a certain combination of a fifth part and third part uniquely codes for a certain group of DNA-substance conjugates which is smaller than the group of DNA-substance conjugates which is encoded by the third part alone.

[0028] Since in this embodiment, at least one coding region has not three or four, but actually five parts, a total of four primers can be used in each qPCR for amplifying the at least one coding region. In short, one primer annealing to the first part, one primer annealing to the third part, one primer annealing to the fourth part and one primer annealing to the fifth part can be used. This gives a total amount of 6 two-dimensional matrices. Thus, in one single qPCR, more

detailed information is obtained which specific groups of DNA-substance conjugates have been enriched in the selection experiment with the DEL.

[0029] Furthermore, according to the invention, a DNA-encoded library is provided. The DNA-encoded library comprises many different DNA-substance conjugates, wherein the DNA-substance conjugates differ from each other by their substance and by their DNA molecules, wherein the DNA molecules of the DNA-substance conjugates differ from each other by comprising different DNA barcode sequences, wherein each DNA barcode sequence comprises at least a first coding region DNA sequence comprising at least a first part, a second part and a third part, wherein the second part is located between the first and third part and the second part differs between all the DNA molecules by at least two nucleotides, characterized in that the first part and the third part encode information regarding the second part of the first coding region, wherein a certain first part and/or a certain third part uniquely codes for a certain group of DNA-substance conjugates which is smaller than the group of all DNA-substance conjugates in the DNA-encoded library.

[0030] The inventive DNA-encoded library can be characterized in that

1. i) the first coding region DNA sequence comprises at least a fourth part, wherein the second part is located between the fourth and third part and wherein both the combination of the first part and the fourth part and the combination of the first part and the third part of the first coding region encode information about the second part of the first coding region; and
2. ii) each barcode sequence comprises at least a second coding region DNA sequence comprising at least a first part, a second part, a third part, and a fourth part, wherein the second part is located between the fourth and third part and the second part differs between all the DNA molecules by at least two nucleotides, wherein both the combination of the first part and the fourth part and the combination of the first part and the third part of the second coding region encode information about the second part of the second coding region;

wherein a certain combination of a first part and fourth part in a certain coding region uniquely codes for a certain group of DNA-substance conjugates which is smaller than the group of all DNA-substance conjugates which is encoded by the first part alone.

[0031] Furthermore, the inventive DNA-encoded library can be characterized in that ach barcode sequence comprises at least a third coding region DNA sequence, which is on the same DNA strand as the second coding region, comprising at least a first part, a second part, a third part, and a fourth part, wherein the second part is located between the fourth and third part and the second part differs between all the DNA molecules by at least two nucleotides, wherein both the combination of the first part and the fourth part and the combination of the first part and the third part and the of the third coding region encode information about the second part of the third coding region, wherein a certain combination of a first part and fourth part in the second coding region and in the third coding region uniquely codes for a certain group of DNA-substance conjugates which is smaller than the group of DNA-substance

conjugates which is encoded by the first part alone.

[0032] In a preferred embodiment of the invention, the DNA-encoded library is characterized in that at least one coding region DNA sequence, optionally all coding region DNA sequences, comprise at least a first part, a second part, a third part, a fourth part and a fifth part, wherein the second part is located between the fourth and fifth part and the second part differs between all the DNA molecules by at least two nucleotides, wherein the combination of the first part and the fourth part and the combination of the fifth part and the third part of the coding region encode information about the second part of the coding region, preferably of all coding regions, and wherein a certain combination of a first part and fourth part uniquely codes for a certain group of DNA-substance conjugates which is smaller than the group of DNA-substance conjugates which is encoded by the first part alone and wherein a certain combination of a fifth part and third part uniquely codes for a certain group of DNA-substance conjugates which is smaller than the group of DNA-substance conjugates which is encoded by the third part alone.

[0033] In a further preferred embodiment, the DNA-encoded library is producible or produced by the inventive method for providing a DNA-encoded library,

[0034] Moreover, according to the invention, a method of decoding the inventive DNA-encoded library is provided. The method comprises

1. a) performing a qPCR with the DNA-encoded library according to one of claims 5 to 8 as template, wherein the following primers are used:

   a primer A and a primer B for amplifying the first coding region of every DNA-substance conjugate; and

   many different primers A-xN which anneal to the different first parts of the first coding region and many different primers B-yN which anneal to the different third parts of the first coding region, wherein primer A-xN has an identical length like the coding region primer A by shortening x nucleotides at its 5'-end, primer B-yN has an identical length like the coding region primer B by shortening y nucleotides at its 5'-end, N represents a A, T, G or C and x and y represent the total number of any one of A, T, G or C at the 3'-end of the primers, wherein x is an integer from 2 to 6, preferably 4;

2. b) calculating a mathematical product of the signal value of each primer A-xN and each primer B-xN by following equation:

   $\text{Value } (A\text{-}xN)_i = \text{signal value } [(A\text{-}xN)_i + B] \cdot \text{signal value } [(A\text{-}xN)_i + (B\text{-}xN)_i];$

   and

   $\text{Value } (B\text{-}yN)_i = \text{signal value } [(B\text{-}yN)_i + A] \cdot \text{signal value } [(B\text{-}yN)_i + (A\text{-}xn)_i],$

   wherein i is an integer and defines a specific primer, and the "+"-sign indicates a combination of two primers; wherein signal value is the percentage of abundance related to the whole set of qPCR quantification using different primers annealed to the same region; and
3. c) comparing the obtained mathematical products for each of the primers $(A\text{-}xN)_i$ and (B-

$yN)_i$, wherein those primers with high values code for DNA-substance conjugates which are present at a high concentration in the DNA-encoded library.

[0035] The method of decoding the inventive DNA-encoded library can be characterized in that the method comprises

1. i) calculating a mathematical product of the value obtained for each primer A-xN and each primer B-yN by following equation
   $\text{Value (A-B)}_i = \text{Value (A-xN)}_i \cdot \text{Value (B-yN)}_i$;
2. ii) comparing the obtained mathematical products for each of the combination of primers $(A\text{-}xN)_i$ and $(B\text{-}yN)_i$, wherein those primer combinations with high values code for DNA-substance conjugates which are present at a high concentration in the DNA-encoded library.

[0036] Furthermore, the method of decoding the inventive DNA-encoded library can be characterized in that the qPCR is performed with the inventive DNA-encoded library according and the method comprises

1. i) performing a qPCR with the following primers:

a first coding region primer A and a first coding region primer primer B for amplifying the first coding region of every DNA-substance conjugate; and many different primers A-xN which anneal to the different first parts, or first and fourth parts of the first coding region and many different primers B-yN which anneal to the different third parts of the first coding region, wherein A-xN has an identical length like the coding region primer A by shortening x nucleotides at its 5'-end, B-yN has an identical length like the coding region primer B by shortening y nucleotides at its 5'-end, N represents a A, T, G or C and x and y represent the total number of any one of A, T, G or C at the 3'-end of the primers, wherein x is an integer from 6 to 10, preferably 8, and y is an integer from 2 to 6, preferably 4; and

a second coding region primer C and a second coding region primer D for amplifying the second coding region of every DNA-substance conjugate; and

many different primers D-yN which anneal to the different first parts, or first and fourth parts of the second coding region and many different primers C-xN which anneal to the different third parts of the second coding region, wherein primer C-xN has an identical length like the coding region primer C by shortening x nucleotides at its 5'-end, primer D-yN has an identical length like the coding region primer D by shortening y nucleotides at its 5'-end, N represents a A, T, G or C and x and y represent the total number of any one of A, T, G or C at the 3'-end of the primers, wherein x is an integer from 6 to 10, preferably 8, and y is an integer from 2 to 6, preferably 4;

2. ii) calculating a mathematical product of the signal value of each primer A-xN, each primer B-yN, each primer C-xN and each primer D-yN by following equation:
   Value $(A\text{-}xN)_i$ = signal value $[(A\text{-}xN)_i+B] \cdot$ signal value $[(A\text{-}xN)_i+(B\text{-}xN)_i]$;

   Value $(B\text{-}yN)_i$ = signal value $[(B\text{-}yN)_i+A] \cdot$ signal value $[(B\text{-}yN)_i+(A\text{-}xN)_i]$,

   Value $(C\text{-}xN)_i$ = signal value $[(C\text{-}xN)_i+D] \cdot$ signal value $[(C\text{-}xN)_i+(D\text{-}xn)_i]$,

   Value $(D\text{-}yN)_i$ = signal value $[(D\text{-}yN)_i+C] \cdot$ signal value $[(D\text{-}yN)_i+(C\text{-}xn)_i]$,

   wherein i is an integer and defines a specific primer, and the "+"-sign indicates a combination of two primers; wherein signal value is the percentage of abundance related to the whole set of qPCR quantification using different primers annealed to the same region; and
3. iii) comparing the obtained mathematical products for each of the primers $(A\text{-}xN)_i$, $(B\text{-}yN)_i$, $(C\text{-}xN)_i$ and $(D\text{-}yN)_i$, wherein those primers with high values code for DNA-substance conjugates which are present at a high concentration in the DNA-encoded library.

[0037] In a preferred embodiment of the invention, the method of decoding the inventive DNA-encoded library comprises

1. i) calculating a mathematical product of the value obtained for each primer A-xN and each primer B-yN, for each primer A-xN and each primer D-xN and for each primer C-yN and D-xN by following equation
   Value $(A\text{-}B)_i$ = Value $(A\text{-}xN)_i \cdot$ Value $(B\text{-}yN)_i$;

   Value $(A\text{-}D)_i$ = Value $(A\text{-}xN)_i \cdot$ Value $(D\text{-}yN)_i$;

   Value $(C\text{-}D)_i$ = Value $(C\text{-}xN)_i \cdot$ Value $(D\text{-}yN)_i$;
2. ii) calculating the mathematical product of the Value $(A\text{-}B)_i$, $(A\text{-}D)_i$ and $(C\text{-}D)_i$ for each primer i by the following equation
   Value^i = value $(A\text{-}B)_i \cdot$ value $(A\text{-}D)_i \cdot$ value $(C\text{-}D)_i$
3. iii) comparing the obtained mathematical products Value^i, wherein those primer combinations i with high values code for DNA-substance conjugates which are present at a high concentration in the DNA-encoded library.

[0038] In a further preferred embodiment, the method of decoding the inventive DNA-encoded library is characterized in that the qPCR is performed with the inventive DNA-encoded library as template and the method comprises

1. i) performing a qPCR with the following primers:

a first coding region primer A and a first coding region primer B for amplifying the first coding region of every DNA-substance conjugate; and many different primers A-xN which anneal to the different first parts, or first and fourth parts of the first coding region, and many different primers B-yN which anneal to the different third parts of the first coding region, wherein A-xN has an identical length like the coding region primer A by shortening x nucleotides at its 5'-end, B has an identical length like the coding region primer B-yN by shortening y nucleotides at its 5'-end, N represents a A, T, G or C and x and y represent the total number of any one of A, T, G or C at the 3'-end of the primers, wherein x is an integer from 6 to 10, preferably 8, and y is an integer from 2 to 6, preferably 4; and

a second coding region primer C and a second coding region primer D for amplifying the second coding region of every DNA-substance conjugate; and

many different primers D-yN which anneal to the different first parts, or first and fourth parts of the second coding region and many different primers C-xN which anneal to the different third parts of the second coding region, wherein primer C-xN has an identical length like the coding region primer C by shortening x nucleotides at its 5'-end, primer D-yN has an identical length like the coding region primer D by shortening y nucleotides at its 5'-end, N represents a A, T, G or C and x and y represent the total number of any one of A, T, G or C at the 3'-end of the primers, wherein x is an integer from 6 to 10, preferably 8, and y is an integer from 2 to 6, preferably 4;

a third coding region primer E and a third coding region primer F for amplifying the third coding region of every DNA-substance conjugate; and many different primers E-xN which anneal to the different first parts of the third coding region and many different primers F-yN which anneal to the different third parts of the third coding region, wherein primer E-xN has an identical length like the coding region primer E by shortening x nucleotides at its 5'-end, primer F-yN has an identical length like the coding region primer F by shortening y nucleotides at its 5'-end, N represents a A, T, G or C and x and y represent the total number of any one of A, T, G or C at the 3'-end of the primers, wherein x is an integer from 6 to 10, preferably 8, and y is an integer from 2 to 6, preferably 4;

2. ii) calculating a mathematical product of the signal value of each primer A-xN, each primer B-yN, each primer C-xN, each primer D-yN, each primer E-xN and each primer F-yN by following equation:

$$\text{Value } (A\text{-}xN)_i = \text{signal value } [(A\text{-}xN)_i + B] \cdot \text{signal value } [(A\text{-}xN)_i + (B\text{-}xN)_i];$$

$$\text{Value } (B\text{-}yN)_i = \text{signal value } [(B\text{-}yN)_i + A] \cdot \text{signal value } [(B\text{-}yN)_i + (A\text{-}xN)_i],$$

$$\text{Value } (C\text{-}xN)_i = \text{signal value } [(C\text{-}xN)_i + D] \cdot \text{signal value } [(C\text{-}xN)_i + (D\text{-}xN)_i],$$

$$\text{Value } (D\text{-}yN)_i = \text{signal value } [(D\text{-}yN)_i + C] \cdot \text{signal value } [(D\text{-}yN)_i + (C\text{-}xN)_i],$$

Value $(E\text{-}xN)_i$ = signal value $[(E\text{-}xN)_i+F]$ · signal value $[(E\text{-}xN)_i+(F\text{-}xn)_i]$,

Value $(F\text{-}yN)_i$ = signal value $[(F\text{-}yN)_i+E]$ · signal value $[(F\text{-}yN)_i+(E\text{-}xN)_i]$, wherein i is an integer and defines a specific primer, and the "+"-sign indicates a combination of two primers; wherein signal value is the percentage of abundance related to the whole set of qPCR quantification using different primers annealed to the same region; and

3. iii) comparing the obtained mathematical products for each of the primers $(A\text{-}xN)_i$, $(B\text{-}yN)_i$, $(C\text{-}xN)_i$, $(D\text{-}yN)_i$, $(E\text{-}xN)_i$ and $(N\text{-}yN)_i$, wherein those primers with high values code for DNA-substance conjugates which are present at a high concentration in the DNA-encoded library.

[0039] The method of decoding the inventive DNA-encoded library may comprise

1. i) calculating a mathematical product of the value obtained for each primer A-xN and each primer B-yN, for each primer A-xN and each primer D-xN, for each primer C-yN and D-xN, for each primer A-xN and N-yN, for each primer M-xN and D-yN and for each primer M-xN and N-yN by following equation

$$\text{Value } (A\text{-}B)_i = \text{Value } (A\text{-}xN)_i \cdot \text{Value } (B\text{-}yN)_i;$$

$$\text{Value } (A\text{-}D)_i = \text{Value } (A\text{-}xN)_i \cdot \text{Value } (D\text{-}yN)_i;$$

$$\text{Value } (C\text{-}D)_i = \text{Value } (C\text{-}xN)_i \cdot \text{Value } (D\text{-}yN)_i;$$

$$\text{Value } (A\text{-}F)_i = \text{Value } (A\text{-}xN)_i \cdot \text{Value } (F\text{-}yN)_i;$$

$$\text{Value } (E\text{-}D)_i = \text{Value } (E\text{-}xN)_i \cdot \text{Value } (D\text{-}yN)_i;$$

$$\text{Value } (E\text{-}F)_i = \text{Value } (E\text{-}xN)_i \cdot \text{Value } (F\text{-}yN)_i;$$

2. ii) calculating the mathematical product of the values $(A\text{-}B)_i$, $(A\text{-}D)_i$, $(C\text{-}D)_i$, $(A\text{-}F)_i$, $(E\text{-}D)_i$ and $(E\text{-}F)_i$ for each primer combinations i by the following equation

$$\text{Value}^i = \text{value } (A\text{-}B)_i \cdot \text{value } (A\text{-}D)_i \cdot \text{value } (C\text{-}D)_i \cdot \text{value } (A\text{-}F)_i \cdot \text{value } (E\text{-}D)_i \cdot \text{value } (E\text{-}F)_i;$$

3. iii) comparing the obtained mathematical products Value^i, wherein those primer combinations i with high values code for DNA-substance conjugates which are present at a high concentration in the DNA-encoded library.

[0040] In a preferred embodiment, the method is characterized in that it comprises the calculation of a $\text{Value}^{i'}$ by the following calculation:

$$\text{Value}^{i'} = \log_{10}[\text{value } (A\text{-}B)_i \cdot \text{value } (A\text{-}D)_i \cdot \text{value } (C\text{-}D)_i \cdot \text{value } (A\text{-}F)_i \cdot \text{value } (E\text{-}D)_i \cdot$$

value $(E\text{-}F)_i$].

[0041] With reference to the following Figures and Examples, the subject according to the invention is intended to be explained in more detail without wishing to restrict said subject to the special embodiments shown here.

[0042] Figure 1A shows how the coding algorithm works for generating a qPCR-matrix for DNA codes 1 having one single coding region 2 (coding region III). E and F are primary primers and E_{xe} and F_{xf} are secondary primers. Primary primer E binds upstream (i.e. towards the 5'-end) of the first region #1 and primary primer F binds upstream (i.e. towards the 5'-end) of the third region #3. A qPCR with only the two primary primers E, F amplifies the DNA barcodes of all DNA-substance conjugates of the DNA-encoded library having coding region III. A qPCR with at least one primary primer E, F and at least one secondary primer E_{xe}, F_{xf} is called a "primary qPCR". Figure 1A illustrates a qPCR template containing one single coding region III having the three code parts (sub-codes) #1, #2, #3. The sequence of the second part #2 of the coding region III is a unique sub-code. Each combination of the first part #1 and the third part #3 can also represent a unique code. Therefore, a sequence of the second part #2 is corresponding to a combination of the first part #1 and the third part #3. For each code part (sub-code) #1, #2, #3, there is a minimal difference number n between any pair of sequences (e.g. between two different xe sequences), while n should be ≥ 2. This means that the code parts #1, #2, #3 differs from each other by at least two nucleotides.

[0043] Figure 1B shows how the coding algorithm works for generating a qPCR-matrix for DNA codes having two coding regions, namely coding region I and coding region II. A, B, C and D are primary primers, A_{xa}, B_{xb}, C_{xc} and D_{xd} are secondary primers and A_{xaya} and D_{xdyd} are tertiary primers. A qPCR comprising the use of at least two tertiary primers is called a "tertiary PCR". Figure 1B illustrates a qPCR template containing two different coding regions I, II, wherein the first coding region I has four code parts (sub-codes) #1, #2, #3, #4 and the second coding region II also has four code parts (sub-codes) #1, #2, #3, #4. The sequence of the second parts #2 of each coding region I, II represents a unique sub-code. Each combination of the first part #1 and the third part #3 of each coding region I, II can also represent a unique sub-code of each coding region I, II. In this case, a sequence of the second part #2 of each coding region I, II is corresponding to a combination of the sequence of the first part #1 and the sequence of the third part #3 of each coding region I, II. Each combination of the first part #1 the fourth part #4 can also represent one unique building block. In this case, a sequence of the second part #2 of each coding region I, II is also corresponding to a combination of the sequence of the first part #1 and the sequence of the fourth part #4 of each coding region I, II. For each code part (sub-code) #1, #2, #3, #4, there is a minimal difference number n between any pair of sequences (e.g. between two different #2 sequences), while n should be ≥ 2. This means that each code part #1, #2, #3, #4 differs from another code part #1, #2, #3, #4 by at least two nucleotides.

[0044] Figure 2 shows how the coding algorithm works for generating a qPCR-matrix for DNA codes having three coding regions I, II, III. A, B, C and D are each a primary primer. A_{xa}, B_{xb}, C_{xc}, D_{xd}, M_{xm} and N_{xn} are each a secondary primer. A_{xaya}, D_{xdyd}, M_{xmym} and N_{xnyn} are each a tertiary primer. A qPCR using at least two tertiary primers is called a "tertiary PCR". Figure 2 illustrates a qPCR template containing three different coding regions I, II, III, wherein the first coding region I has four code parts (sub-codes) #1, #2, #3, #4, the second coding region II also has four code parts (sub-codes) #1, #2, #3, #4 and the third coding region III has five code parts (sub-codes) #1, #2, #3, #4, #5. The sequence of each second code part #2 of each coding region I, II, III is a unique sub-code. Each combination of code parts #1 and code part #3, code part #1 with code part #4 and code part #1 with code part #5 can also represent a unique sub-code. For example, a sequence of code part #2 is corresponding to a combination of code part #1 and code part #3. For each code part (sub-code), there is a minimal difference number n between any pair of sequences (e.g. between two different ab sequences), while n should be ≥ 2. This means that each code part #1, #2, #3, #4, #5 differs from another code part #1, #2, #3, #4, #5 by at least two nucleotides.

[0045] Figure 3 shows three different qPCR matrices which were obtained after a qPCR using a DNA barcode having three coding regions I, II, III as template and 20 different primary primers A and 20 different primary primers B for binding to coding region I (see matrix "Q-PCR with A + B", columns = different primers A, lines = different primers B), 20 different primary primers C and 20 different primary primers D for coding region II (see matrix "Q-PCR with C + D", columns = different primers C, lines = different primers D) and 20 different primary primers E and 20 different primary primers F for coding region III (see matrox "Q-PCR with E + F", columns = different primers E, lines = different primers F). An exemplary result of the matrix is illustrated in the table "E+F" in which the primer pairs with the strongest amplification signal are listed together with their obtained (normalized) amplification signal. It can be derived from said table that the strong amplification signal have been obtained with the primer pairs E3 and F3 (25%), E18 and F11 (20%), E3 and F11 (15%), E11 and F3 (15%) and E11 and F17 (15%) and a medium amplification signal has been obtained with the primer pair E3 and F17 (10%). Below the table "E+F", the obtained result is also shown in a column diagram. It can be derived from the obtained result that DNA-substance conjugates to which e.g. the primer pair E3 and F3 binds had a high concentration in the DNA-encoded library (after the enrichment experiment) and DNA-substance conjugates to which e.g. the primer pair E3 and F17 binds had a lower concentration in the DNA-encoded library (after the enrichment experiment). It may also be concluded that DNA-substance conjugates to which primer pairs with no signal (e.g. E1 and F1) bind were not present in the DNA-encoded library (after the enrichment experiment). Since the substance connected to each specific DNA coding region is known, this approach allows a fast and sensitive identification of substances being present at a high concentration after a (DEL) selection experiment.

[0046] Figure 4 shows the results of the qPCR matrices "Q-PCR with A + B" (amplification of coding region I) and "Q-PCR with C + D" (amplification of coding region II) from Figure 3 and also the results of a secondary PCR with primer pairs A and D (coding region between I and II; see Figure 2). qPCR with the primer pair A + D gave strong amplification signals for the specific

primer pairs A11 and D15 (20%), A11 and D2 (15%), A2 and D18 (15%) and medium amplification signals for the primer pairs A2 and D8 (10%), A11 and D18 (10%), A17 and D18 (10%), A17 and D15 (10%) and A17 and D8 (10%). Each of said identified A primers binds to a specific coding region I and each of said identified D primers binds to a specific coding region II. This means that the primers A and D which gave strong signals code for coding regions I and II which must have been enriched in the DNA-encoded library after the enrichment experiment. It can also be concluded that the two coding regions I and II must be located on one single DNA strand because otherwise, no amplification signal would have been obtained. In order to combine the result obtained with the primer pair A and D with the result of the other primer pairs A and B and C and D, the mathematical product of the value obtained for each specific primer pair is calculated by the equation $Value_{coding\ region\ I\text{-}II} = Value_{matrix\text{-}A+D} \cdot Value_{matrix\text{-}A+B} \cdot Value_{matrix\text{-}C+D.}$ Specific primers A, B, C and D which resulted in a high amplification signal consequently have a high $Value_{coding\ region\ I\text{-}II}$. Thus, the obtained $Value_{coding\ region\ I\text{-}II}$ allows the identification of primers which must have bound to abundant DNA-barcodes and thus allows the identification of substances (bound to the DNA-barcode) which were abundantly present after the (DEL) selection experiment.

[0047] Figure 5 shows a plot of 36 different combinations of coding regions I (A + B) and II (C + D) which gave the highest mathematical product according to equation $Value_{coding\ region\ I\text{-}II} = Value_{matrix\text{-}A+D} \cdot Value_{matrix\text{-}A+B} \cdot Value_{matrix\text{-}C+D}$ (see absolute value in arbitrary unit on y-axis). In said plot, it can be visually identified that the combinations of coding region I and II with the numbers 2, 4, 23, 29 and 32 on the x-axis achieved the highest score. These numbers refer to the following five different combinations of coding regions I and II: A11B4-C19D2 (no. 2), A11B17-C19D2 (no. 4), A2B4-C10-D18 (no. 23), A11B4-C19D15 (no. 29) and A11B17-C19D15 (no. 32). Naturally, it is known for which substances (or plurality of substances) these five different combinations encode. Thus, it is possible to identify five different (groups of) substances which have been strongly enriched in a (DEL) selection experiment.

[0048] Figure 6 shows a part of qPCR matrices obtained after a qPCR using a DNA barcode having three coding regions I, II, III as template, 20 different primary primers A and B for coding region I, 20 different primary primers C and D for coding region II, the (same) 20 different primary primers A and D for coding region I to II and 20 different primary primers E and F for coding region III. After having caluculated the $Value_{coding\ region\ I\text{-}II} = Value_{matrix\text{-}A+D} \cdot Value_{matrix\text{-}A+B} \cdot Value_{matrix\text{-}C+D}$, it has become clear that significant values are obtained for coding regions I-II coded by the nine primer pairings A17B17-C1D15, A2B4-C10D18, A11B4-C1D15, A11B4-C19D15, A11B4-C19D2, A11B17-C1D15, A11B17-C19D15, A11B17-C10D18 and A11B17-C19D2. The highest value for coding region III has been determined as well by equation $Value_{coding\ region\ III} = Value_{matrix\text{-}F+E}$ and it has been found that high values for coding region III are obtained by the five primer pairings F3E3, F11E18, F3E11, F11E3 and F17E11. If the nine coding regions I to II identified above encode a first group of nine different substances and the five coding regions III identified above encode a second group of five different substances, it follows that the combination of the substances of the first group and second group must have been present at a high concentration before the qPCR experiment, i.e. must

have been strongly enriched by the (DEL) selection experiment.

[0049] Figure 7 shows a decoding process for a medium DEL having 306 compounds, each tagged with a DNA barcode. Before and after a DEL selection experiment, a primary qPCR was conducted with the primer pairs E and F, with the primer pairs E_{xe} and F, with the primer pairs E and F_{xf}, with the primer pairs E_{xe1} and F_{xf11} and with the primer pairs E_{xe3} and F_{xf17}. The obtained C_q values before selection are shown in the left matrix, the obtained C_q valued after selection are shown in the middle matrix and the ΔC_q values are shown in the right matrix in Figure 7. A ΔCq value of a primer pair which is below the ΔCq value of the primer pair E and F indicates an enrichment of the DNA-substance conjugate. As can be seen in the "ΔC_q"-matrix, subcoding region E-F_{xf11} has ΔC_q-value of 10.0 which is below the ΔC_q-value of 13.1 for subcoding region E-F (i.e. below the control). This means that the subcoding region E-F_{xf11} has been enriched. The same is true for the subcoding region E_{xe1}-F with its ΔC_q-value of 10.0 being below the ΔC_q-value of 13.1 for subcoding region E-F (i.e. being below the control). Thus, the results of the primary qPCR indicate that after the DEL selection experiment, substance(s) encoded by the subcoding regions to which primers E_{xe1} and F_{xf11} bind were enriched more strongly than substances encoded by the subcoding regions to which primers E_{xe3} and F_{xf17} bind. Additionally, for confirmation of said data, a secondary qPCR was conducted with the primer pair E_{xe} and F_{xf}. Said secondary qPCR confirmed that the subcoding region E_{xe1}-F_{xf11} is enriched more strongly than the subcoding region E_{xe3}-F_{xf17} (see matrix "ΔC_q" in Figure 7: value in field of column E_{xe1} and row F_{xf11} is much lower than value in field of column E_{xe3} and row F_{xf17} and much lower than the value in the field of column E and row F, which is the control). In summary, both the primary and secondary qPCR demonstrate that the substance(s) connected to the E_{xe1}-F_{xf11} subcoding region must have been strongly enriched after the DEL selection experiment.

[0050] Figure 8 shows an example of a large DEL having 4^{10} compounds, each tagged with a DNA barcode. The libraries were generated by partially degenerated synthesis of DNA. Figure 8 illustrates the setup for conducting a primary PCR with one (constant) primer E and various different primers F_n, wherein each primer Fn codes for a certain subgroup of the library, specifically $1/4^n$ compounds of the library (having 4^{10} compounds in total), wherein n is an integer from 0 to 5. This means that if six primers F are used, the first primer F_0 codes for $1/4^0$ of all compounds of the library, i.e. all compounds of the library (= 4^{10} = 1048576 compounds), the primer F_1 only codes for 1/4 of all compounds of the library (= 262144 compounds), the primer F_2 only codes for 1/16 of all compounds of the library (= 65536 compounds), the primer F_3 only codes for 1/64 of all compounds of the library (= 16384 compounds), the primer F_4 only codes for 1/256 of all compounds of the library (= 4096 compounds) and the primer F_5 only codes for 1/1024 of all compounds of the library (= 1024 compounds). This means that after the qPCR has been performed, the group of encoded substances which have been selected in the DEL experiment can be significantly narrowed because encoded substances which are not

amplifiable with a primer F_n, wherein n is 1 to 5, give no signal in qPCR. For example, if the combination of primer E and primer F_1 fails to provide a signal in primary qPCR, it is clear that 3/4 of all compounds, i.e. 786432 compounds of 1048576 compounds, are not amplifiable by said qPCR and thus were not enriched by the DEL selection experiment preceding the primary qPCR. Thus, there are only 1/4 of all compounds (= 262144 compounds) of the DEL library remaining coming into question for having been enriched in the (DEL) selection experiment.

[0051] Figure 9 shows another example of a very large DEL having 4^{20} compounds, each tagged with a DNA barcode. The principal procedure is the same like the one disclosed in Figure 8 for a DEL having 4^{10} compounds. However, due to the larger size of the DEL, it is beneficial if the primary PCR is carried out with more than five different primers F. Specifically, it is beneficial if n is an integer from 0 to 10 in this case. This means that if eleven primers F are used, the first primer F_0 codes for all compounds of the library and the last primer F_{10} only codes for $1/4^{10}$ of all compounds of the library (= 1048576 compounds). This means that after the qPCR has been performed, the group of encoded substances which have been selected in the DEL experiment has been significantly narrowed- For example, a DNA barcode of a DNA-substance conjugate which give no signal with the primer F_1 in qPCR means that said DNA barcode belongs to a group of 3/4 of 4^{20} compounds ($\approx 8.2 \cdot 10^{11}$ compounds) which have not been enriched in the DEL enrichment experiment. Thus, the group of relevant enriched DNA-substance conjugates has been narrowed to 1/4 of 4^{20} compounds ($\approx 2.7.10^{11}$ compounds). With each primer F_n increasing from n=1 to n=10, the group of relevant compounds is further narrowed. An amplification signal turning up with primer F_{10} means that the DNA-substance conjugate is within a subgroup of 4^{10} ($\approx 1.10^6$ compounds) of 4^{20} compounds ($\approx 1.1 \cdot 10^{12}$ compounds) in total.

[0052] Figure 10 shows the identification of the substance 4-carboxybenzenesulfonamide (in the following: "CBS") after a DNA-CBS-conjugate within a DNA-encoded library (DEL) has been enriched by selection with the enzyme carbonic anhydrase II. A small qPCR matrix has been built using three primers 1a, 2a, 3a pairing with the three primers 1b, 2b, 3b. The primer pair 1a, 1b anneals to all DNA-barcodes of the DNA-substance conjugates of the library and thus has the potential to amplify DNA barcodes of the entire library. Primer 2a covers a sub-library containing DNA-CBS-conjugate and primer 2b covers another sub-library containing DNA-CBS-conjugate. The combination of primers 2a and 2b can be assigned exclusively to CBS. Primer 3a covers a sub-library containing theobromine (in the following: "Theo") conjugated to a DNA barcode (= DNA-Theo-conjugate), primer 3b covers another sub-library containing the DNA-Theo-conjugate. The combination of primers 3a and 3b can be assigned exclusively to Theo. ΔCq is the difference in qPCR cycle before and after selection. A small number reflects large enrichment. $\Delta Cq(1a\text{-}1b) > \Delta Cq(2a\text{-}1b) \approx \Delta Cq(1a\text{-}2b) > \Delta Cq(2a\text{-}2b)$ indicated that CBS is remarkably enhanced. $\Delta Cq(1a\text{-}1b) < \Delta Cq(3a\text{-}1b) \approx \Delta Cq(1a\text{-}3b) \approx \Delta Cq(3a\text{-}3b)$ indicated that Theo is not enriched.

[0053] Figure 11 shows a DNA-substance conjugate in which a first substance S is conjugated

chemically covalently to a first coding region DNA sequence I and second coding region DNA sequence and in which a second substance S is conjugated chemically covalently to a third coding region DNA sequence III. Each coding region DNA sequence I, II, III has a first part #1 and a third part #3 to which certain primers can bind (i.e. anneal during qPCR). Primer P2' (5'-gctgttccca cattgcgt-3', SEQ-ID Nr. 1) binds to the first part #1 of first coding region DNA sequence I, primer P2Y (5'-ccttctggat tcggtcggag caccatc-3', SEQ-ID Nr. 2) binds to the third part #3 of first coding region DNA sequence I, primer P2Y' (5'-gatggtgctc cgaccgaatc cagaagg-3', SEQ-ID Nr. 3) binds to the first part #1 of second coding region DNA sequence II, primer P1Y (5'-ggaggtgtag acgacagagt atttgactgt cagg-3', SEQ-ID Nr. 4) binds to the third part #3 of second coding region DNA sequence II, primer P4' (5'-cagatcgagc aactccac-3', SEQ-ID Nr. 5) binds to the first part #1 of third coding region DNA sequence III and primer P5 (5'-tggtctcagc cgcctat-3', SEQ-ID Nr. 6) binds to the third part #3 of third coding region DNA sequence III. If substance S has been enriched after a selection experiment with the DNA-encoded library, amplification with primer pair P2' and P2Y, primer pair P2Y' and P1Y and primer pair P4' and P5 each gives a strong amplification signal in qPCR which allows identification of substance S.

Example 1 - DEL comprising DNA barcodes with one single coding region

[0054] For DNA codes containing only one single coding region, each code has 3 parts, #1 (first part), #2 (second part) and #3 (third part). Each #2 sequence is a unique code, while each combination of #1 and #3 can also represent a unique code (see e.g. Figure 1A). Therefore, a sequence of #2 is corresponding to a combination of #1 and #3.

[0055] For each part, there is a minimal difference number n between any pair of sequences (e.g. between two different #1 sequences), while n should be ≥ 2.

Example 2 - DEL comprising DNA barcodes with two coding regions

[0056] For DNA codes containing two coding regions, each sub-code has 4 parts, for example the first coding region #1 (first part), #2 (second part), #3 (third part) and #4 (fourth part) and the second coding region #1 (first part), #2 (second part), #3 (third part) and #4 (fourth part) (see e.g. Figure 1B). Each #2 (second part) sequence is a unique sub-code, while each combination of #1, #3 and #4 is corresponding to #2 and each combination of #1, #3 and #4 is corresponding to #2.

[0057] For each part, there is a minimal difference number n between any pair of sequences (e.g. between two different #1 sequences), while n should be ≥ 2.

Example 3 - DEL comprising DNA barcodes with more than two coding regions

[0058] For DNA codes containing more than two coding regions (see e.g. Figure 1B or Figure 2), two sub-codes at both ends are designed according Example 2, and the sub-code(s) in between is/are designed according to either Example 1, Example 2 or Example 4. It is very unlikely that a high quality DEL can be synthesized using the split-and-pool method. Therefore, a DEL containing less than 4 sub-codes is favorable.

Example 4 - DEL comprising DNA barcodes with 5 sub-codes

[0059] The DNA barcodes of this DEL have 5 parts, #1 (first part), #2 (second part), #3 (third part), #4 (fourth part) and #5 (fifth part).

[0060] Each #2 sequence (second part) is a unique sub-code, while each combination of #1 and #3 can also represent a unique sub-code. Therefore, a sequence of #2 is corresponding to a combination of #1 and #3. Each combination of #1 and #4 can also represent a unique sub-code. Therefore, a sequence of #2 is corresponding to a combination of #1 and #4. Each combination of #1 and #5 can also represent a unique sub-code. Therefore, a sequence of #2 is corresponding to a combination of #1 and #5.

[0061] For each part, there is a minimal difference number n between any pair of sequences (e.g. between two different #1 sequences), while n should be ≥ 2 in all designs.

Example 5 - Description of decoding process: Decoding one-(sub)-code

[0062] A primary qPCR matrix is built for the first coding region I using primer A with u different primers B-xb, and primer B with v different primers A-xa. Therefore, the size of resulting matrix is u·v (see e.g. Figure 3). Same matrices can also be built for the second coding region II and the third coding region III.

[0063] A secondary qPCR matrix is built for the first coding region I using pairs of B-xb and A-xa, while B-xb and A-xa are chosen according to the signal intensity in the primary matrix. Same secondary matrices can be built for the second coding region II and the third coding region III. The ranking for each building block can thus be concluded.

[0064] For sequence containing two sub-codes, an additional secondary qPCR matrix can be built using A-xa and D-xd, while A-xa and D-xd are chosen according to the signal intensity in the primary matrices.

[0065] In combination with the two sub-code matrices (A-xa+B-xb and C-xc+D-xd), the ranking of the combinations can be concluded based on certain algorithm, for example:

$$Value^{i} = Value^{i}_{matrix\text{-}A+D} \cdot Value^{i}_{matrix\text{-}A+B} \cdot Value^{i}_{matrix\text{-}C+D},$$

wherein the Value' is a value relating and being proportional to the amount of a certain DNA

barcode in the DEL. In other words, said Value' relates to an individual DNA sequence (barcode structure) which resulted from the combinatorial synthesis through joining two building blocks and two sub-codes.

[0066] To further validate the Value' ranking, an additional tertiary qPCR matrix can be built using A-xa-ya and D-xd-yd, while A-xa-ya and D-xd-yd are chosen according to the signal intensity in the primary and secondary matrices and the resulting $Value^i$ ranking.

[0067] A full matrix can also be built using A, D and all A-xa-ya and D-xd-yd, though it will be significantly more expensive than the method described before.

[0068] The method cannot provide a fully quantitative decoding solution for DEL containing more than two sub-codes. However, combining various primary, secondary, and tertiary rtPCR matrices can provide a $Value^i$ for certain compounds i, which is corresponding to a DNA code containing several sub-codes. All forward and backward primers can be combined to build a matrix.

[0069] For example, any primer A, A-xa, A-xa-ya can be combined with any primer B, B-xb, N, N-yn, N-xn-yn, D, D-xd, D-xd-yd to build QPCR matrices. A value for a particular compound can be calculated according to certain algorithm, for example:

$$Value^i = \log_{10}(Value^i_{matrix\text{-}A+D} \cdot Value^i_{matrix\text{-}A+N} \cdot Value^i_{matrix\text{-}M+D} \cdot Value^i_{matrix\text{-}A+B} \cdot Value^i_{matrix\text{-}C+D} \cdot Value^i_{matrix\text{-}M+N})$$

in which the $Value^i_{matrix\text{-}A+D}$, $Value^i_{matrix\text{-}A+N}$ and $Value^i_{matrix\text{-}M+D}$ can be either from the secondary, or tertiary matrices, or as a combination of them, and in which the $Value^i_{matrix\text{-}A+B} \cdot Value^i_{matrix\text{-}C+D} \cdot Value^i_{matrix\text{-}M+N}$ are from the secondary matrices.

List of reference signs

[0070]

DBC:
DNA barcode sequence;
S:
substance;
I:
first coding region DNA sequence;
II:
second coding region DNA sequence;
III:
third coding region DNA sequence;
#1:
first part of a coding region DNA sequence;

#2:
second part of a coding region DNA sequence;
#3:
third part of a coding region DNA sequence;
#4:
fourth part of a coding region DNA sequence;
#5:
fifth part of a coding region DNA sequence;
A, B, C, D, E, F, M, N:
primary primer;
A_{xa}, B_{xb}, C_{xc}, D_{xd}, E_{xe}, F_{xf}, M_{xm}, N_{yn}:
secondary primer;
A_{xaya}, D_{xdyd}, M_{xmym}, N_{xnyn}:
tertiary primer;
1a, 1b:
primary primer binding to all DBS;
2a, 2b:
secondary primer binding to DBS of CBS only;
3a, 3b:
secondary primer binding to Theo only;
P2':
primer annealing to first part #1 of coding region I;
P2Y:
primer annealing to third part #3 of coding region I;
P2Y':
primer annealing to first part #1 of coding region
P1Y:
primer annealing to third part #3 of coding region
P4':
primer annealing to first part #1 of coding region III;
P5:
primer annealing to third part #3 of coding region III.

SEQUENCE LISTING

[0071]

<110> Technische Universitat Dresden

<120> Method for providing a DNA-encoded library, DNA-encoded library and method of decoding a DNA-encoded library

<130> 179P 3353

<160> 6

<170> BiSSAP 1.3.6

<210> 1
<211> 18
<212> DNA
<213> Artificial Sequence

<220>
<223> primer annealing to first part #1 of coding region I

<400> 1
gctgttccca cattgcgt 18

<210> 2
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> primer annealing to third part #3 of coding region I

<400> 2
ccttctggat tcggtcggag caccatc 27

<210> 3
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> primer annealing to first part #1 of coding region II

<400> 3
gatggtgctc cgaccgaatc cagaagg 27

<210> 4
<211> 34
<212> DNA
<213> Artificial Sequence

<220>
<223> primer annealing to third part #3 of coding region II

<400> 4
ggaggtgtag acgacagagt atttgactgt cagg 34

<210> 5
<211> 18
<212> DNA
<213> Artificial Sequence

<220>
<223> primer annealing to first part #1 of coding region III

<400> 5
cagatcgagc aactccac 18

<210> 6
<211> 18
<212> DNA
<213> Artificial Sequence

<220>
<223> primer annealing to third part #3 of coding region III

<400> 6
tggtctcagc cgccctat 18

Fremgangsmåde til tilvejebringelse af et dna-indkodet bibliotek, dna-inkodet bibliotek og fremgangsmåde til at afkode et dna-indkodet bibliotek

Patentkrav

1. Fremgangsmåde til tilvejebringelse af et DNA-indkodet bibliotek, der omfatter

   a) syntetisering af mange forskellige DNA-molekyler, der adskiller sig fra hinanden ved at omfatte forskellige DNA-stregkodesekvenser, hvor hver DNA-stregkodesekvens omfatter mindst en første kodende DNA-regionsekvens, der omfatter mindst en første del, en anden del og en tredje del, hvor den anden del er placeret mellem den første og tredje del, og den anden del adskiller sig fra alle DNA-molekylerne med mindst to nukleotider; og

   b) binding af hvert af de mange forskellige DNA-molekyler til mindst et specifikt stof, der danner forskellige DNA-stofkonjugater, hvor DNA-stofkonjugaterne adskiller sig fra hinanden ved det specifikke stof og ved deres DNA-molekyler;

kendetegnet ved, at den første del og den tredje del koder for information vedrørende den anden del af den første kodende region, hvor en bestemt første del og/eller en bestemt tredje del udelukkende koder for en bestemt gruppe af DNA-stofkonjugater, der er mindre end gruppen af alle DNA-stofkonjugater i det DNA-indkodede bibliotek.

2. Fremgangsmåde ifølge krav 1, kendetegnet ved, at

   i) den første kodende DNA-regionsekvens omfatter mindst en fjerde del, hvor den anden del er placeret mellem den fjerde og tredje del, og hvor både kombinationen af den første del og den fjerde del og kombinationen af den første del og den tredje del af den første kodende region koder for information om den anden del af den første kodende region; og

   ii) hver stregkodesekvens omfatter mindst en anden kodende DNA-regionsekvens, der omfatter mindst en første del, en anden del, en tredje del og en fjerde del, hvor den anden del er placeret mellem den fjerde og tredje del, og den anden del adskiller sig fra alle DNA-molekylerne med mindst to nukleotider, hvor både kombinationen af den første del og den fjerde del og kombinationen af den første del og den tredje del af den anden kodende region koder for information om den anden del af den anden kodende region;

hvor en bestemt kombination af en første del og en fjerde del i et bestemt kodningsområde udelukkende koder for en bestemt gruppe af DNA-stofkonjugater, der er mindre end gruppen af alle DNA-stofkonjugater, der er indkodet af den første del alene.

3. Fremgangsmåde ifølge et af kravene 1 eller 2, kendetegnet ved, at

i) hver stregkodesekvens omfatter mindst en anden kodende DNA-regionsekvens, der omfatter mindst en første del, en anden del, en tredje del og en fjerde del, hvor den anden del er placeret mellem den fjerde og tredje del, og den anden del adskiller sig fra alle DNA-molekylerne med mindst to nukleotider, hvor både kombinationen af den første del og den fjerde del og kombinationen af den første del og den tredje del af den anden kodende region koder for information om den anden del af den anden kodende region; og

ii) hver stregkodesekvens omfatter mindst en tredje kodende DNA-regionsekvens, der omfatter mindst en første del, en anden del, en tredje del og en fjerde del, hvor den anden del er placeret mellem den fjerde og tredje del, og den anden del adskiller sig fra alle DNA-molekylerne med mindst to nukleotider, hvor både kombinationen af den første del og den fjerde del og kombinationen af den første del og den tredje del af den tredje kodende region koder for information om den anden del af den tredje kodende region;

hvor en bestemt kombination af en første del og en fjerde del i en bestemt kodende region udelukkende koder for en bestemt gruppe af DNA-stofkonjugater, der er mindre end gruppen af DNA-stofkonjugater, der er indkodet af den første del.

4. Fremgangsmåde ifølge et af de foregående krav, kendetegnet ved, at

mindst en kodende DNA-regionsekvens, eventuelt alle kodende DNA-regionsekvenser, omfatter mindst en første del, en anden del, en tredje del, en fjerde del og en femte del, hvor den anden del er placeret mellem den fjerde og femte del, og den anden del adskiller sig fra alle DNA-molekylerne med mindst to nukleotider,

hvor kombinationen af den første del og den fjerde del og kombinationen af den femte del og den tredje del af den kodende region koder for information om den anden del af den kodende region, fortrinsvis om alle kodende regioner,

hvor en bestemt kombination af en første del og en fjerde del udelukkende koder for en bestemt gruppe af DNA-stofkonjugater, der er mindre end gruppen af

DNA-stofkonjugater, der er indkodet af den første del alene, og
hvor en bestemt kombination af en femte del og en tredje del entydige koder for en bestemt gruppe af DNA-stofkonjugater, der er mindre end gruppen af DNA-stofkonjugater, der er indkodet af den tredje del.

5. DNA-indkodet bibliotek, der omfatter mange forskellige DNA-ligandkonjugater, hvor DNA-ligandkonjugaterne adskiller sig fra hinanden ved deres ligand og ved deres DNA-molekyler,
hvor DNA-molekylerne i DNA-ligandkonjugaterne adskiller sig fra hinanden ved at omfatte forskellige DNA-stregkodesekvenser, hvor hver DNA-stregkodesekvens omfatter mindst en første kodende DNA-regionsekvens, der omfatter mindst en første del, en anden del og en tredje del, hvor den anden del er placeret mellem den første og tredje del, og den anden del adskiller sig fra alle DNA-molekylerne med mindst to nukleotider;
kendetegnet ved, at den første del og den tredje del koder for information vedrørende den anden del af den første kodende region, hvor en bestemt første del og/eller en bestemt tredje del udelukkende koder for en bestemt gruppe af DNA-ligandkonjugater, der er mindre end gruppen af alle DNA-ligandkonjugater i det DNA-indkodede bibliotek.

6. DNA-indkodet bibliotek ifølge krav 5, kendetegnet ved, at
   i) den første kodende DNA-regionsekvens omfatter mindst en fjerde del, hvor den anden del er placeret mellem den fjerde og tredje del, og hvor både kombinationen af den første del og den fjerde del og kombinationen af den første del og den tredje del af den første kodende region koder for information om den anden del af den første kodende region; og
   ii) hver stregkodesekvens omfatter mindst en anden kodende DNA-regionsekvens, der omfatter mindst en første del, en anden del, en tredje del og en fjerde del, hvor den anden del er placeret mellem den fjerde og tredje del, og den anden del adskiller sig fra alle DNA-molekylerne med mindst to nukleotider, hvor både kombinationen af den første del og den fjerde del og kombinationen af den første del og den tredje del af den anden kodende region koder for information om den anden del af den anden kodende region;
hvor en bestemt kombination af en første del og en fjerde del i en bestemt kodende region udelukkende koder for en bestemt gruppe af DNA-ligandkonjugater, der er

mindre end gruppen af alle DNA-ligandkonjugater, der er indkodet af den første del alene.

7. DNA-indkodet bibliotek ifølge krav 6, kendetegnet ved, at hver stregkodesekvens omfatter mindst en tredje kodende DNA-regionsekvens, der er på den samme DNA-streng som den anden kodende region, der mindst omfatter en første del, en anden del, en tredje del og en fjerde del, hvor den anden del er placeret mellem den fjerde og tredje del, og den anden del adskiller sig fra alle DNA-molekylerne med mindst to nukleotider, hvor både kombinationen af den første del og den fjerde del og kombination af den første del og den tredje del af den tredje kodende region koder for information om den anden del af den tredje kodende region, hvor en bestemt kombination af en første del og fjerde del i den anden kodende region og i den tredje kodende region udelukkende koder for en bestemt gruppe af DNA-ligandkonjugater, der er mindre end gruppen af DNA-ligandkonjugater, der er indkodet af den første del alene.

8. DNA-indkodet bibliotek ifølge et af kravene 5 til 7, kendetegnet ved, at mindst en kodende DNA-regionsekvens, eventuelt alle kodende DNA-regionsekvenser, omfatter mindst en første del, en anden del, en tredje del, en fjerde del og en femte del, hvor den anden del er placeret mellem den fjerde og femte del, og den anden del adskiller sig fra alle DNA-molekylerne med mindst to nukleotider,
hvor kombinationen af den første del og den fjerde del og kombinationen af den femte del og den tredje del af den kodende region koder for information om den anden del af den kodende region, fortrinsvis om alle kodende regioner, og
hvor en bestemt kombination af en første del og en fjerde del udelukkende koder for en bestemt gruppe af DNA-ligandkonjugater, der er mindre end gruppen af DNA-ligandkonjugater, der er indkodet af den første del alene, og
hvor en bestemt kombination af en femte del og en tredje del entydige koder for en bestemt gruppe af DNA-ligandkonjugater, der er mindre end gruppen af DNA-ligandkonjugater, der er indkodet af den tredje del.

9. Fremgangsmåde til afkodning af et DNA-indkodet bibliotek ifølge et af kravene 5 til 8, der omfatter
   a) udførelse af en qPCR med det DNA-indkodede bibliotek ifølge et af kravene 5 til 8 som skabelon, hvor de følgende primere anvendes:

en primer A og en primer B til amplifikation af den første kodende region af hvert DNA-ligandkonjugat; og

mange forskellige primere A-xN, der annealerer til de forskellige første dele af den første kodende region, og mange forskellige primere B-yN, der annealerer til de forskellige tredje dele af den første kodende region,

hvor primer A-xN har samme længde som den kodende region-primer A ved at forkorte x nukleotider ved dens 5'-ende, primer B-yN har samme længde som den kodende region-primer B ved at forkorte y nukleotider i dens 5'-ende, N repræsenterer A, T, G eller C og x og y repræsenterer det samlede antal af et hvilken som helst af A, T, G eller C ved primernes 3'-ende, hvor x er et heltal fra 2 til 6, fortrinsvis 4;

b) beregning af et matematisk produkt af signalværdien af hver primer A-xN og hver primer B-xN med følgende ligning:

Værdi $(A\text{-}xN)_i$ = signalværdi $[(A\text{-}xN)_i+B]$ · signalværdi $[(A\text{-}xN)_i+(B\text{-}xN)_i]$; og

Værdi $(B\text{-}yN)_i$ = signalværdi $[(B\text{-}yN)_i+A]$ · signalværdi $[(B\text{-}yN)_i+(A\text{-}xn)_i]$,

hvor i er et heltal og definerer en specifik primer, og "+"-tegnet angiver en kombination af to primere, hvor signalværdien er procentdelen af hyppighed, der er relateret til hele sættet af qPCR-kvantificering, ved anvendelse af forskellige primere annealeret til den samme region; og

c) sammenligning af de opnåede matematiske produkter for hver af primerne $(A\text{-}xN)_i$ og $(B\text{-}yN)_i$, hvor primerne med høje værdier koder for DNA-ligandkonjugater, der er til stede i en høj koncentration i det DNA-indkodede bibliotek.

10. Fremgangsmåde ifølge krav 9, kendetegnet ved, at fremgangsmåden omfatter

i) beregning af et matematisk produkt af den opnåede værdi for hver primer A-xN og hver primer B-yN med følgende ligning

Værdi $(A\text{-}B)_i$ = Værdi $(A\text{-}xN)_i$ · Værdi $(B\text{-}yN)_i$;

ii) sammenligning af de opnåede matematiske produkter for hver af kombinationerne af primerne $(A\text{-}xN)_i$ og $(B\text{-}yN)_i$, hvor primerkombinationerne med høje værdier koder for DNA-ligandkonjugater, der er til stede i en høj koncentration i det DNA-indkodede bibliotek.

11. Fremgangsmåde ifølge et af kravene 9 eller 10, kendetegnet ved, at qPCR udføres med det DNA-indkodede bibliotek ifølge et af kravene 6 til 8 som skabelon, og at

fremgangsmåden omfatter

i) udførelsen af en qPCR med følgende primere:

en første kodende region-primer A og en første kodende region-primer B til amplifikation af den første kodende region af hvert DNA-ligandkonjugat; og

mange forskellige primere A-xN, der annealerer til de forskellige første dele eller første og fjerde dele af den første kodende region, og mange forskellige primere B-yN, der annealerer til de forskellige tredje dele af den første kodende region, hvor A-xN samme længde som den kodende region-primer A ved at forkorte x nukleotider ved dens 5'-ende, B-yN har samme længde som den kodende region-primer B ved at forkorte y nukleotider i dens 5'-ende, N repræsenterer A, T, G eller C og x og y repræsenterer det samlede antal af et hvilken som helst af A, T, G eller C ved primernes 3'-ende, hvor x er et heltal fra 6 til 10, fortrinsvis 8, og y er et heltal fra 2 til 6, fortrinsvis 4; og

en anden kodende region-primer C og en anden kodende-region primer D til amplifikation af den anden kodende region af hvert DNA-ligandkonjugat; og

mange forskellige primere D-yN, der annealerer til de forskellige første dele eller første og fjerde dele af den anden kodende region, og mange forskellige primere C-xN, der annealerer til de forskellige tredje dele af den anden kodende region, hvor primer C-xN samme længde som den kodende region-primer C ved at forkorte x nukleotider ved dens 5'-ende, primer D-yN har samme længde som den kodende region-primer D ved at forkorte y nukleotider i dens 5'-ende, N repræsenterer A, T, G eller C og x og y repræsenterer det samlede antal af et hvilken som helst af A, T, G eller C ved primernes 3'-ende, hvor x er et heltal fra 6 til 10, fortrinsvis 8, og y er et heltal fra 2 til 6, fortrinsvis 4;

ii) beregning af et matematisk produkt af signalværdien af hver primer A-xN, hver primer B-yN, hver primer C-xN og hver primer D-yN med følgende ligning:

Værdi $(A\text{-}xN)_i$ = signalværdi $[(A\text{-}xN)_i+B]$ · signalværdi $[(A\text{-}xN)_i+(B\text{-}xN)_i]$;

Værdi $(B\text{-}yN)_i$ = signalværdi $[(B\text{-}yN)_i+A]$ · signalværdi $[(B\text{-}yN)_i+(A\text{-}xN)_i]$,

Værdi $(C\text{-}xN)_i$ = signalværdi $[(C\text{-}xN)_i+D]$ · signalværdi $[(C\text{-}xN)_i+(D\text{-}xn)_i]$,

Værdi $(D\text{-}yN)_i$ = signalværdi $[(D\text{-}yN)_i+C]$ · signalværdi $[(D\text{-}yN)_i+(C\text{-}xn)_i]$,

hvor i er et heltal og definerer en specifik primer, og "+"-tegnet angiver en kombination af to primere, hvor signalværdien er procentdelen af hyppighed, der er relateret til hele sættet af qPCR-kvantificering, ved anvendelse af forskellige primere annealeret til den samme region; og

iii) sammenligning af de opnåede matematiske produkter for hver af primerne

$(A\text{-}xN)_i$, $(B\text{-}yN)_i$, $(C\text{-}xN)_i$ og $(D\text{-}yN)_i$, hvor primerne med høje værdier koder for DNA-ligandkonjugater, der er til stede i en høj koncentration i det DNA-indkodede bibliotek.

12. Fremgangsmåde ifølge det foregående krav, kendetegnet ved, at fremgangsmåden omfatter

i) beregning af et matematisk produkt af den opnåede værdi for hver primer A-xN og hver primer B-yN, for hver primer A-xN og hver primer D-xN og for hver primer C-yN og D-xN med følgende ligning

Værdi $(A\text{-}B)_i$ = Værdi $(A\text{-}xN)_i \cdot$ Værdi $(B\text{-}yN)_i$;

Værdi $(A\text{-}D)_i$ = Værdi $(A\text{-}xN)_i \cdot$ Værdi $(D\text{-}yN)_i$;

Værdi $(C\text{-}D)_i$ = Værdi $(C\text{-}xN)_i \cdot$ Værdi $(D\text{-}yN)_i$;

ii) beregning af det matematiske produkt af værdi $(A\text{-}B)_i$, $(A\text{-}D)_i$ og $(C\text{-}D)_i$ for hver primer i med følgende ligning

Værdi^i = værdi $(A\text{-}B)_i \cdot$ værdi $(A\text{-}D)_i \cdot$ værdi $(C\text{-}D)_i$

iii) sammenligning af de opnåede matematiske produkter Værdi^i, hvor primerkombinationerne i med høje værdier koder for DNA-ligandkonjugater, der er til stede i en høj koncentration i det DNA-indkodede bibliotek.

13. Fremgangsmåde ifølge et af kravene 9 eller 12, kendetegnet ved, at qPCR udføres med det DNA-indkodede bibliotek ifølge et af kravene 7 eller 8 som skabelon, og at fremgangsmåden omfatter

i) udførelse af en qPCR med følgende primere:

en første kodende region-primer A og en første kodende-region primer B til amplifikation af den første kodende region af hvert DNA-ligandkonjugat; og

mange forskellige primere A-xN, der annealerer til de forskellige første dele eller første og fjerde dele af den første kodende region, og mange forskellige primere B-yN, der annealerer til de forskellige tredje dele af den første kodende region, hvor A-xN samme længde som den kodende region-primer A ved at forkorte x nukleotider ved dens 5'-ende, B har samme længde som den kodende region-primer B-yN ved at forkorte y nukleotider i dens 5'-ende, N repræsenterer A, T, G eller C og x og y repræsenterer det samlede antal af et hvilken som helst af A, T, G eller C ved primernes 3'-ende, hvor x er et heltal fra 6 til 10, fortrinsvis 8, og y er et heltal fra 2 til 6, fortrinsvis 4; og

en anden kodende region-primer C og en anden kodende-region primer D til

amplifikation af den anden kodende region af hvert DNA-ligandkonjugat; og mange forskellige primere D-yN, der annealerer til de forskellige første dele eller første og fjerde dele af den anden kodende region, og mange forskellige primere C-xN, der annealerer til de forskellige tredje dele af den anden kodende region, hvor primer C-xN samme længde som den kodende region-primer C ved at forkorte x nukleotider ved dens 5'-ende, primer D-yN har samme længde som den kodende region-primer D ved at forkorte y nukleotider i dens 5'-ende, N repræsenterer A, T, G eller C og x og y repræsenterer det samlede antal af et hvilken som helst af A, T, G eller C ved primernes 3'-ende, hvor x er et heltal fra 6 til 10, fortrinsvis 8, og y er et heltal fra 2 til 6, fortrinsvis 4; en tredje kodende region-primer E og en tredje kodende-region primer F til amplifikation af den tredje kodende region af hvert DNA-ligandkonjuga; og mange forskellige primere E-xN, der annealerer til de forskellige første dele af den tredje kodende region, og mange forskellige primere F-yN, der annealerer til de forskellige tredje dele af den tredje kodende region, hvor primer E-xN samme længde som den kodende region-primer E ved at forkorte x nukleotider ved dens 5'-ende, primer F-yN har samme længde som den kodende region-primer F ved at forkorte y nukleotider i dens 5'-ende, N repræsenterer A, T, G eller C og x og y repræsenterer det samlede antal af et hvilken som helst af A, T, G eller C ved primernes 3'-ende, hvor x er et heltal fra 6 til 10, fortrinsvis 8, og y er et heltal fra 2 til 6, fortrinsvis 4;

ii) beregning af et matematisk produkt af signalværdien af hver primer A-xN, hver primer B-yN, hver primer C-xN, hver primer D-yN, hver primer E-xN og hver primer F-yN med følgende ligning:

$$\text{Værdi } (A\text{-}xN)_i = \text{signalværdi } [(A\text{-}xN)_i + B] \cdot \text{signalværdi } [(A\text{-}xN)_i + (B\text{-}xN)_i];$$

$$\text{Værdi } (B\text{-}yN)_i = \text{signalværdi } [(B\text{-}yN)_i + A] \cdot \text{signalværdi } [(B\text{-}yN)_i + (A\text{-}xN)_i],$$

$$\text{Værdi } (C\text{-}xN)_i = \text{signalværdi } [(C\text{-}xN)_i + D] \cdot \text{signalværdi } [(C\text{-}xN)_i + (D\text{-}xN)_i],$$

$$\text{Værdi } (D\text{-}yN)_i = \text{signalværdi } [(D\text{-}yN)_i + C] \cdot \text{signalværdi } [(D\text{-}yN)_i + (C\text{-}xN)_i],$$

$$\text{Værdi } (E\text{-}xN)_i = \text{signalværdi } [(E\text{-}xN)_i + F] \cdot \text{signalværdi } [(E\text{-}xN)_i + (F\text{-}xn)_i],$$

$$\text{Værdi } (F\text{-}yN)_i = \text{signalværdi } [(F\text{-}yN)_i + E] \cdot \text{signalværdi } [(F\text{-}yN)_i + (E\text{-}xN)_i],$$

hvor i er et heltal og definerer en specifik primer, og "+"-tegnet angiver en kombination af to primere, hvor signalværdien er procentdelen af hyppighed, der er relateret til hele sættet af qPCR-kvantificering, ved anvendelse af forskellige primere annealeret til den samme region; og

iii) at sammenligne de opnåede matematiske produkter for hver af primerne

$(A\text{-}xN)_i$, $(B\text{-}yN)_i$, $(C\text{-}xN)_i$, $(D\text{-}yN)_i$, $(E\text{-}xN)_i$ og $(N\text{-}yN)_i$, hvor primerne med høje værdier koder for DNA-ligandkonjugater, der er til stede i en høj koncentration i det DNA-indkodede bibliotek.

14. Fremgangsmåde ifølge krav 13, kendetegnet ved, at fremgangsmåden omfatter

i) beregning af et matematisk produkt af den opnåede værdi for hver primer A-xN og hver primer B-yN, for hver primer A-xN og hver primer D-xN, for hver primer C-yN og D-xN, for hver primer A-xN og N-yN, for hver primer M-xN og D-yN og for hver primer M-xN og N-yN med følgende ligning

$$\text{Værdi } (A\text{-}B)_i = \text{Værdi } (A\text{-}xN)_i \cdot \text{Værdi } (B\text{-}yN)_i;$$

$$\text{Værdi } (A\text{-}D)_i = \text{Værdi } (A\text{-}xN)_i \cdot \text{Værdi } (D\text{-}yN)_i;$$

$$\text{Værdi } (C\text{-}D)_i = \text{Værdi } (C\text{-}xN)_i \cdot \text{Værdi } (D\text{-}yN)_i;$$

$$\text{Værdi } (A\text{-}F)_i = \text{Værdi } (A\text{-}xN)_i \cdot \text{Værdi } (F\text{-}yN)_i;$$

$$\text{Værdi } (E\text{-}D)_i = \text{Værdi } (E\text{-}xN)_i \cdot \text{Værdi } (D\text{-}yN)_i;$$

$$\text{Værdi } (E\text{-}F)_i = \text{Værdi } (E\text{-}xN)_i \cdot \text{Værdi } (F\text{-}yN)_i;$$

ii) at beregne det matematiske produkt af værdierne $(A\text{-}B)_i$, $(A\text{-}D)_i$, $(C\text{-}D)_i$, $(A\text{-}F)_i$, $(E\text{-}D)_i$ og $(E\text{-}F)_i$ for hver primerkombination i med følgende ligning

$$\text{Værdi}^i = \text{værdi } (A\text{-}B)_i \cdot \text{værdi } (A\text{-}D)_i \cdot \text{værdi } (C\text{-}D)_i \cdot \text{værdi } (A\text{-}F)_i \cdot \text{værdi } (E\text{-}D)_i - \text{værdi } (E\text{-}F)_i;$$

iii) sammenligning af de opnåede matematiske produkter Værdi^i, hvor primerkombinationerne i med høje værdier koder for DNA-ligandkonjugater, der er til stede i en høj koncentration i det DNA-indkodede bibliotek.

15. Fremgangsmåde ifølge krav 14, kendetegnet ved, at fremgangsmåden yderligere omfatter beregningen af en $\text{Værdi}^{i'}$ med følgende beregning:

$$\text{Værdi}^{i'} = \log_{10}[\text{værdi } (A\text{-}B)_i \cdot \text{værdi } (A\text{-}D)_i \cdot \text{værdi } (C\text{-}D)_i \cdot \text{værdi } (A\text{-}F)_i \cdot \text{værdi } (E\text{-}D)_i \cdot \text{værdi } (E\text{-}F)_i].$$

DRAWINGS

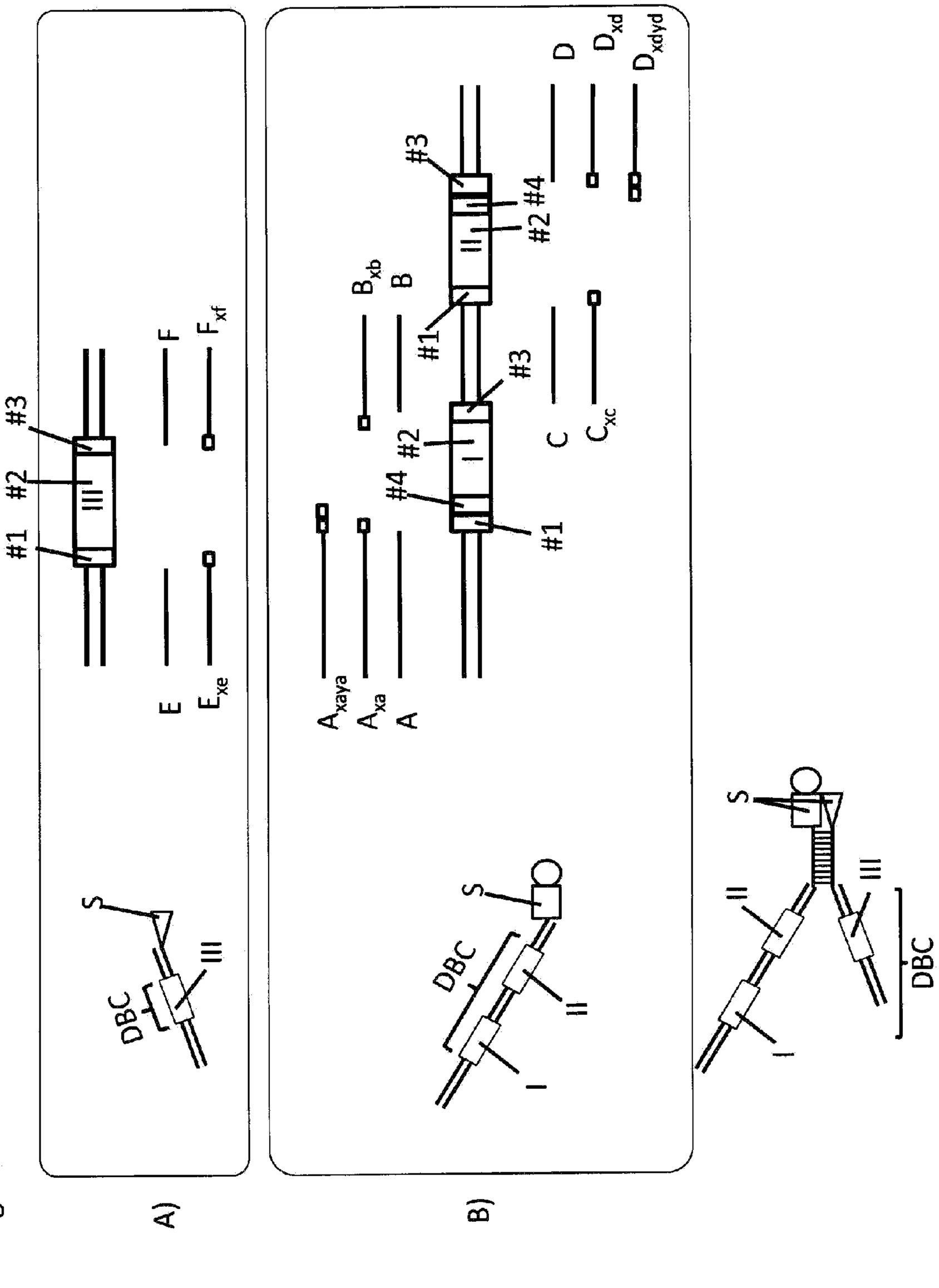
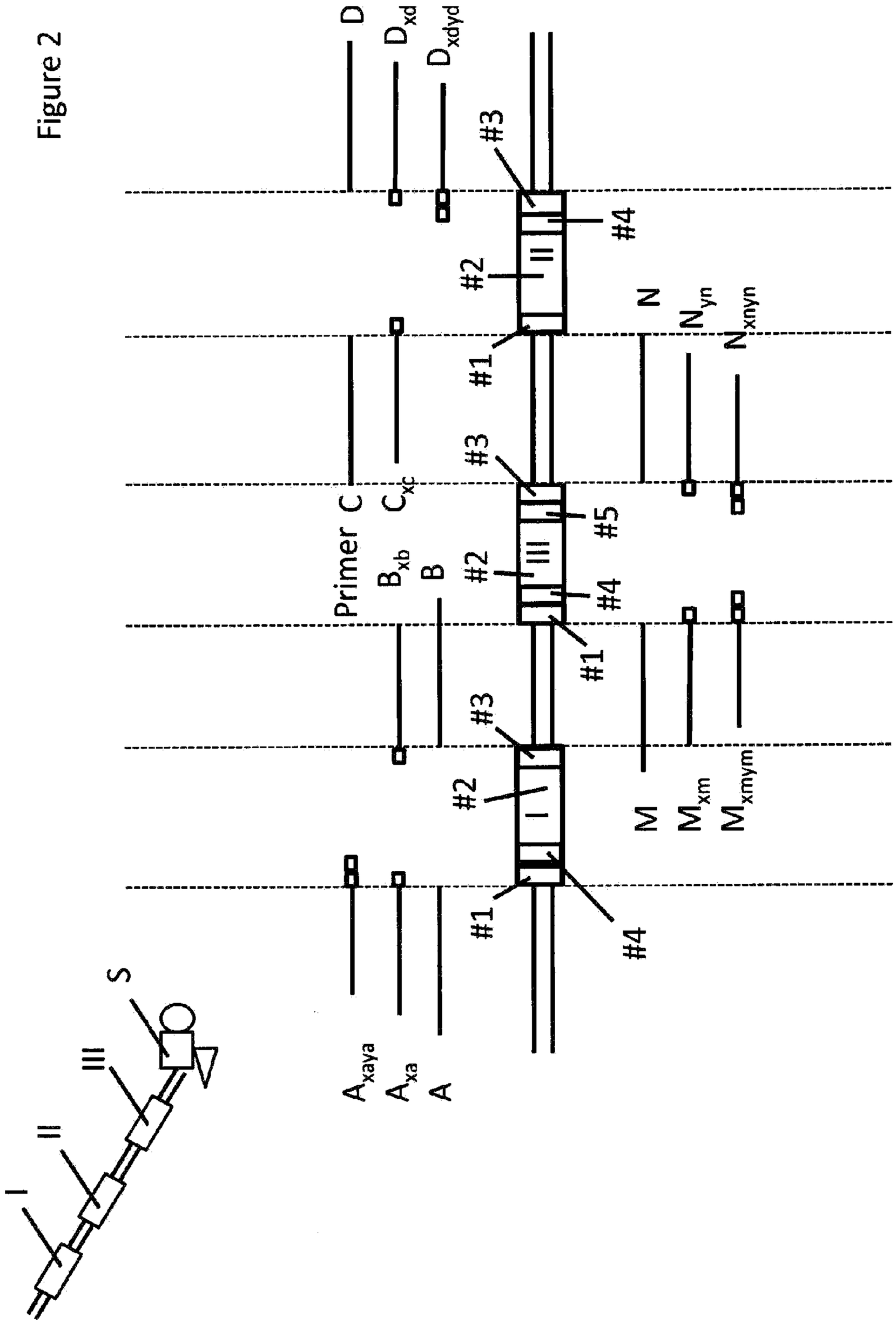


Figure 1

Figure 2


D
D_{xd}
D_{xdyd}
#3
#4
#2
N
N_{yn}
N_{xnyn}
#1
#3
C
C_{xc}
Primer
#5
B_{xb}
B
#2
III
#4
#1
#3
#2
I
M
M_{xm}
M_{xmym}
#1
#4
A_{xaya}
A_{xa}
A
S
III
II
I

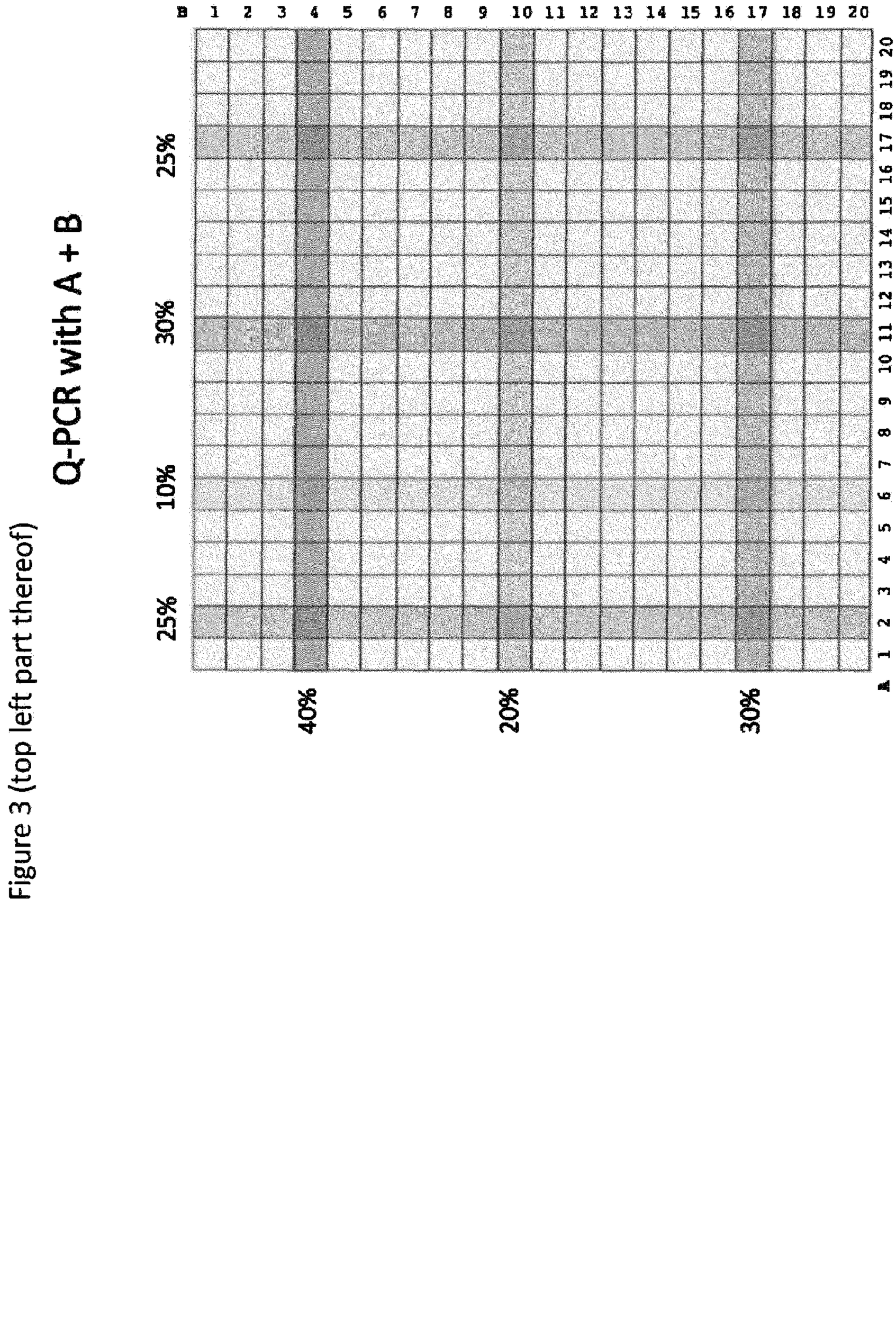


Figure 3 (top left part thereof)

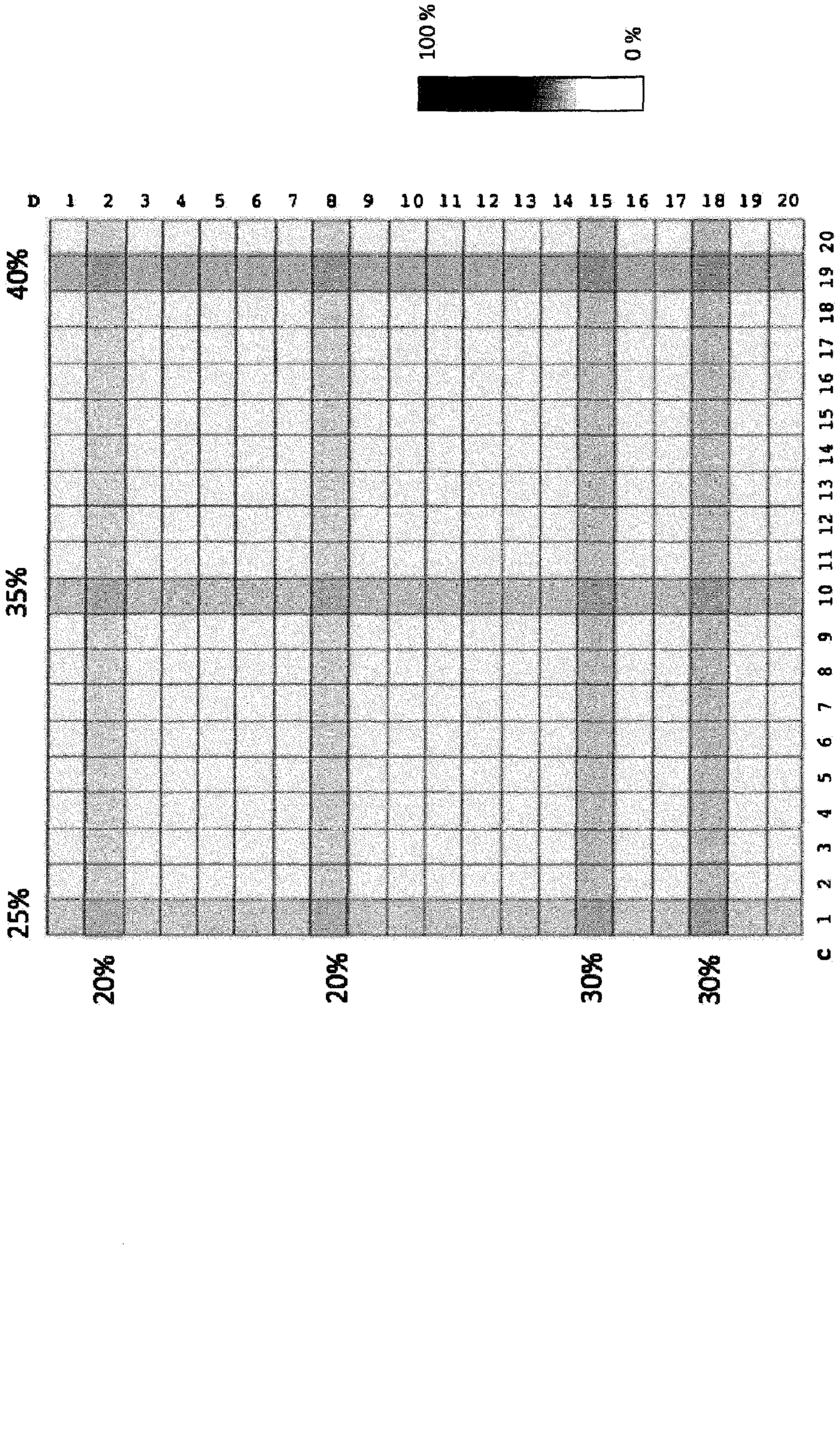


Figure 3 (top right part thereof)

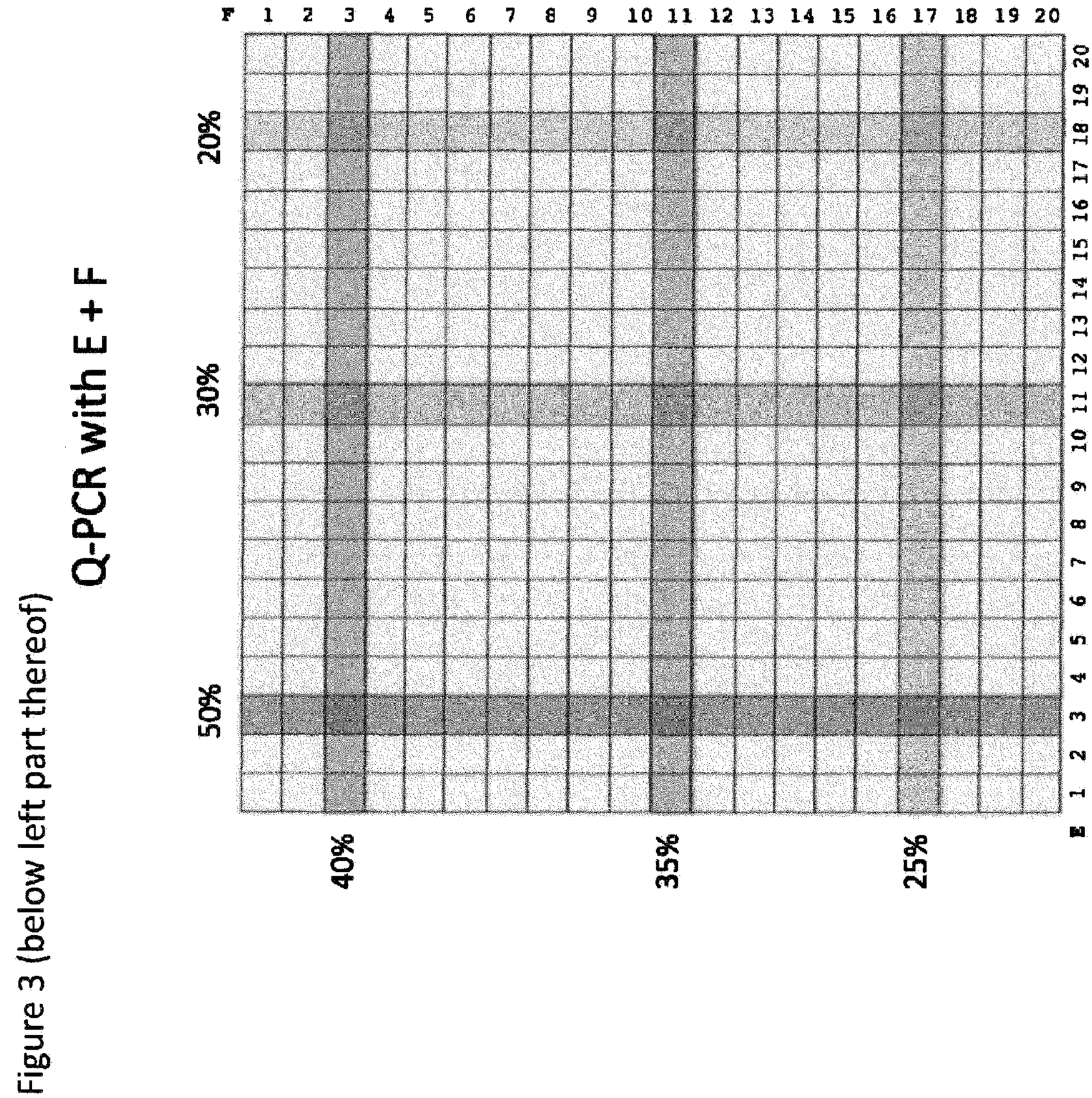


Figure 3 (below left part thereof)

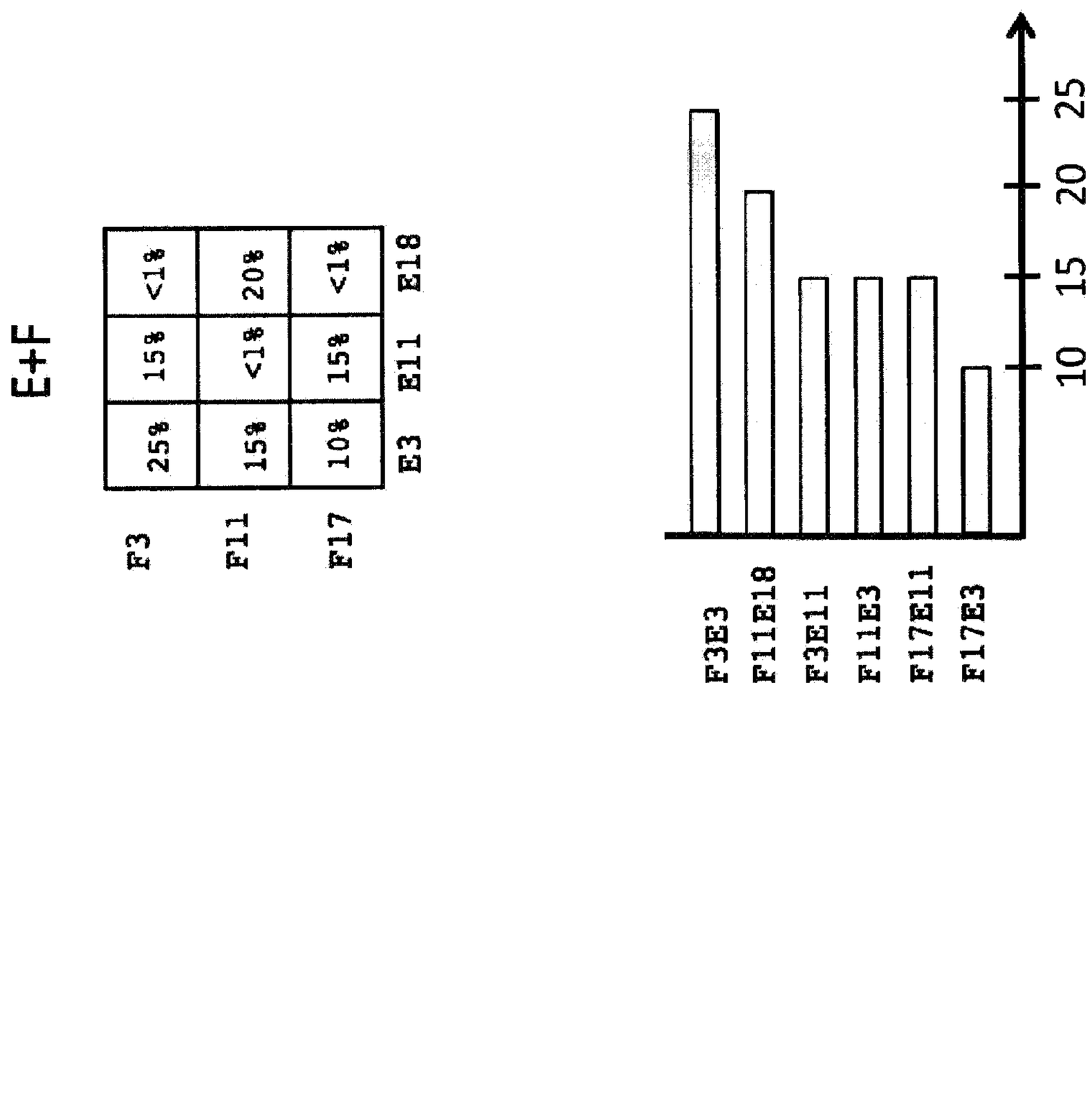


Figure 3 (below right part)

Figure 4

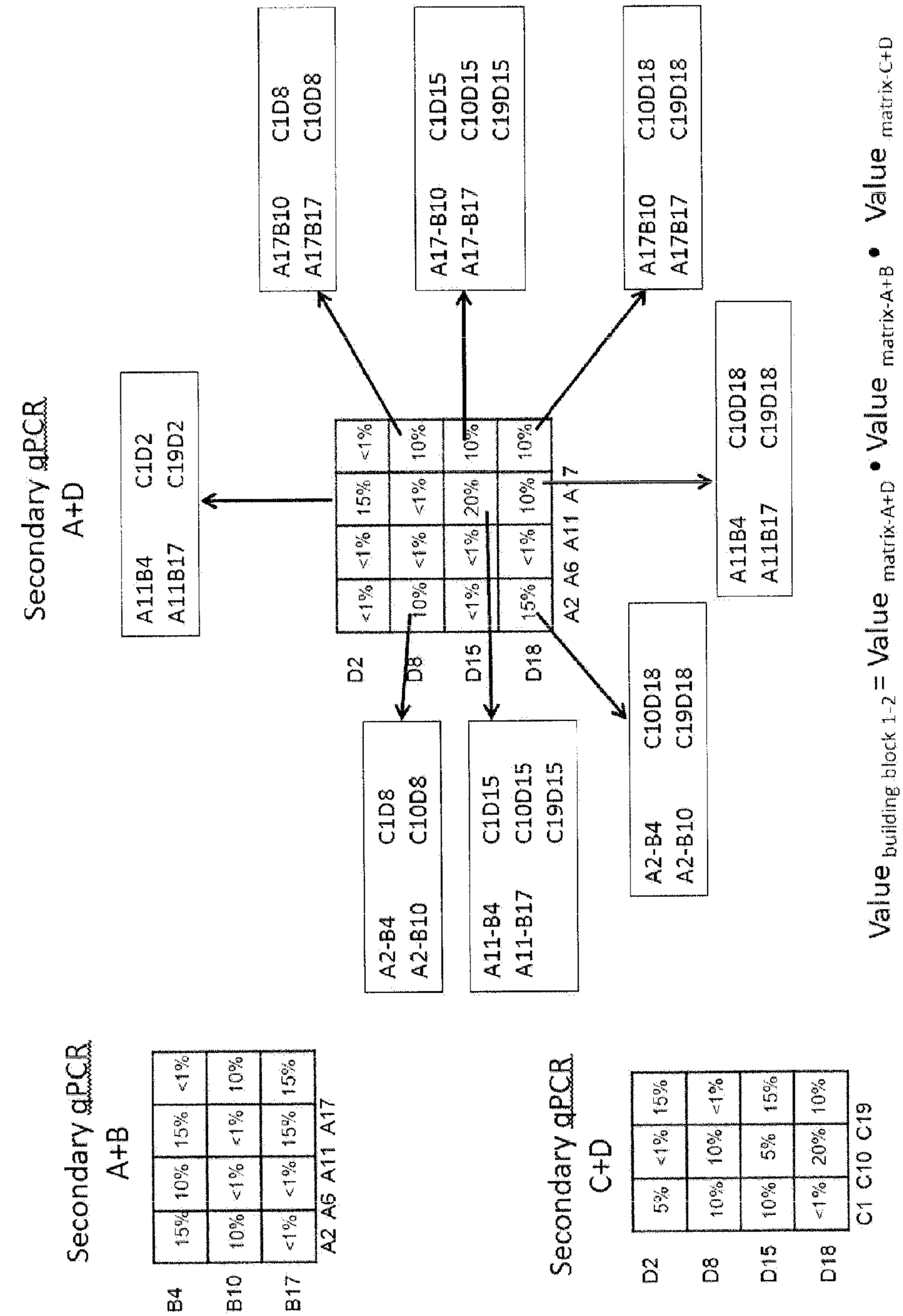

Secondary qPCR
A+D
A11B4 C1D2
A11B17 C19D2
A17B10 C1D8
A17B17 C10D8
A17-B10 C1D15
A17-B17 C10D15
C19D15
A17B10 C10D18
A17B17 C19D18
A11B4 C10D18
A11B17 C19D18
C10D18
C19D18
D2
D8
D15
D18
A2 A6 A11 A17
<1% <1% 15% <1%
10% <1% <1% 10%
<1% <1% 20% 10%
15% <1% 10% 10%
Secondary qPCR
A+B
B4
B10
B17
A2 A6 A11 A17
15% 10% 15% <1%
10% <1% <1% 10%
<1% <1% 15% 15%
Secondary qPCR
C+D
D2
D8
D15
D18
C1 C10 C19
5% <1% 15%
10% 10% <1%
10% 5% 15%
<1% 20% 10%
A2-B4 C1D8
A2-B10 C10D8
A11-B4 C1D15
A11-B17 C10D15
C19D15
A2-B4 C10D18
A2-B10 C19D18
Value building block 1-2 = Value matrix-A+D • Value matrix-A+B • Value matrix-C+D

Figure 5

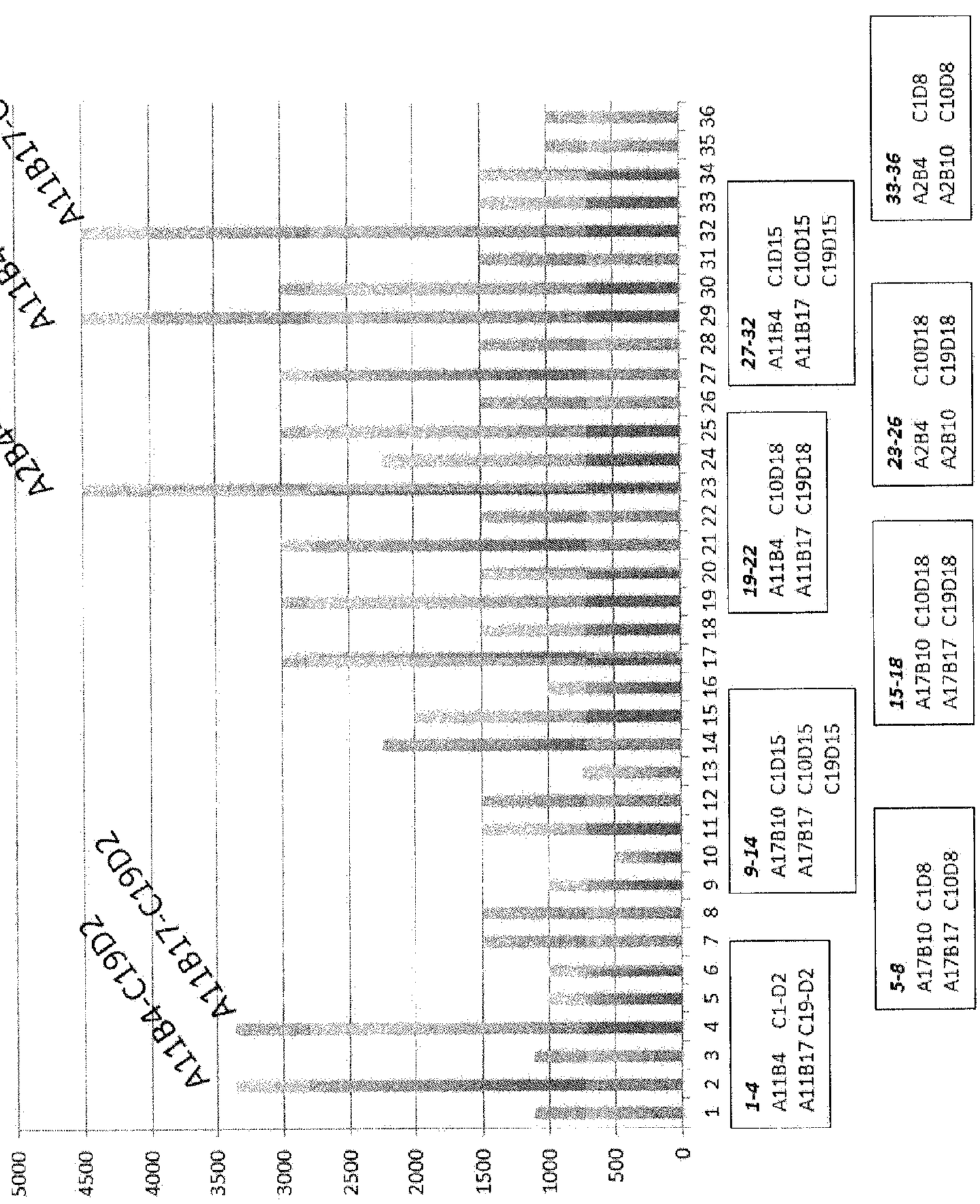
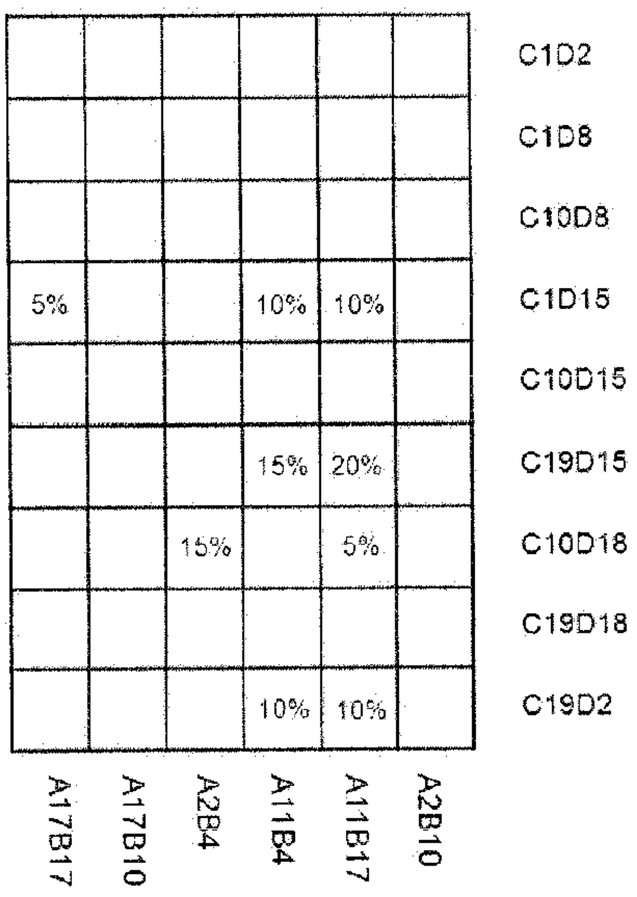
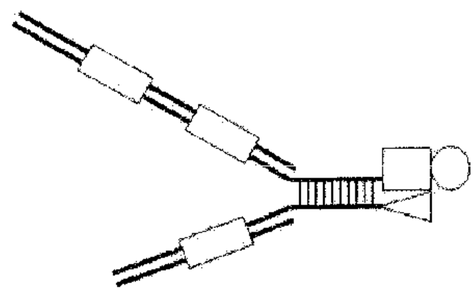

A2B4-C10D18
A11B4-C19D15
A11B17-C19D15
A11B4-C19D2
A11B17-C19D2
5000
4500
4000
3500
3000
2500
2000
1500
1000
500
0
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36
1-4 A11B4 C1-D2 A11B17 C19-D2
5-8 A17B10 C1D8 A17B17 C10D8
9-14 A17B10 C1D15 A17B17 C10D15 C19D15
15-18 A17B10 C10D18 A17B17 C19D18
19-22 A11B4 C10D18 A11B17 C19D18
23-25 A2B4 C10D18 A2B10 C19D18
27-32 A11B4 C1D15 A11B17 C10D15 C19D15
33-36 A2B4 C1D8 A2B10 C10D8

Figure 6

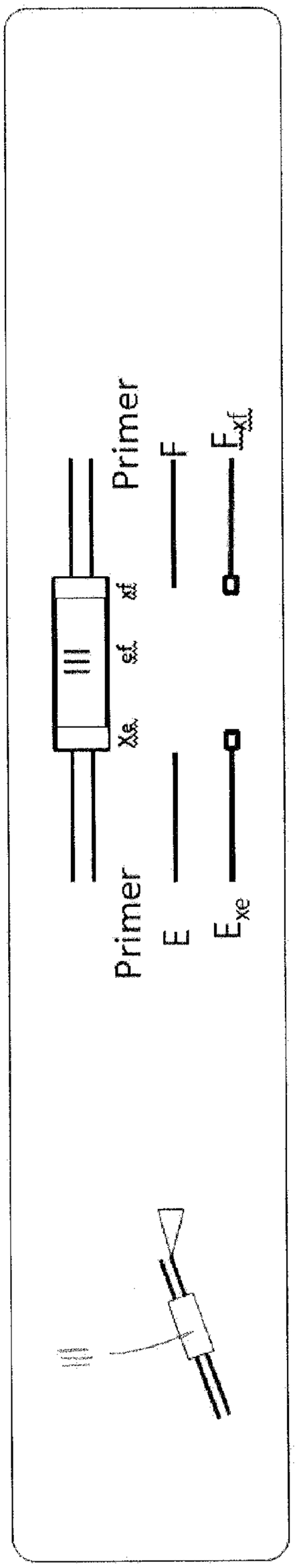


A11B17-C19D15
A11B4-C19D15
A2B4-C10D18
A11B4-C19D2
A11B17-C19D2
A11B4-C1D15
A11B17-C1D15



F3E3
F11E18
F3E11
F11E3
F17E11

Figure 7



	E_{xe1}	E_{xe3}	E
F_{xf11}	18.8		18.6
F_{xf17}		27.0	21.9
F	18.2	23.5	17.5

C_q After selection

	E_{xe1}	E_{xe3}	E
F_{xf11}	11.4		8.6
F_{xf17}		12.1	6.8
F	8.2	9.0	4.4

C_q Before selection

	E_{xe1}	E_{xe3}	E
F_{xf11}	7.4		10.0
F_{xf17}		14.9	15.1
F	10.0	14.5	13.1

ΔC_q

Figure 8

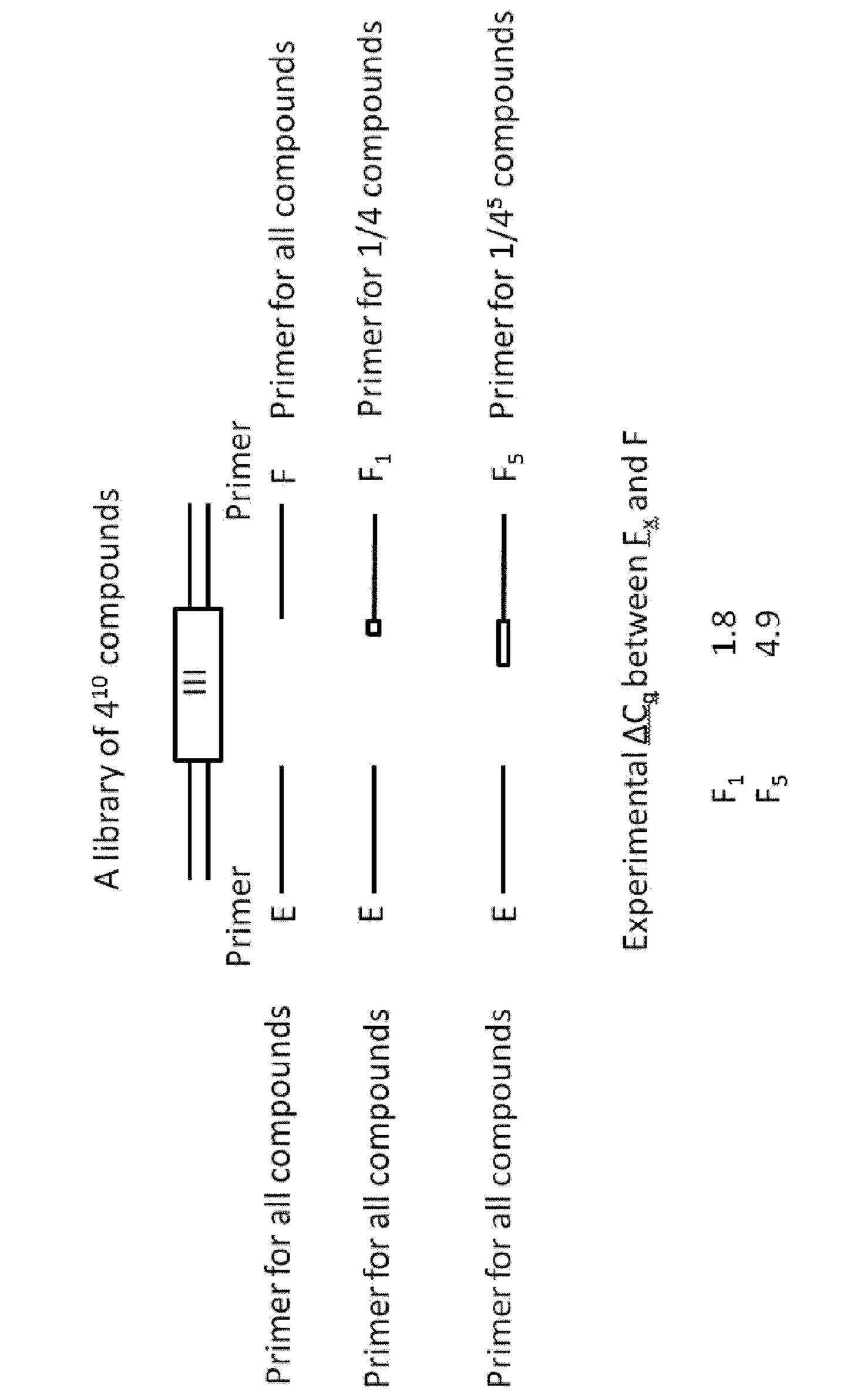
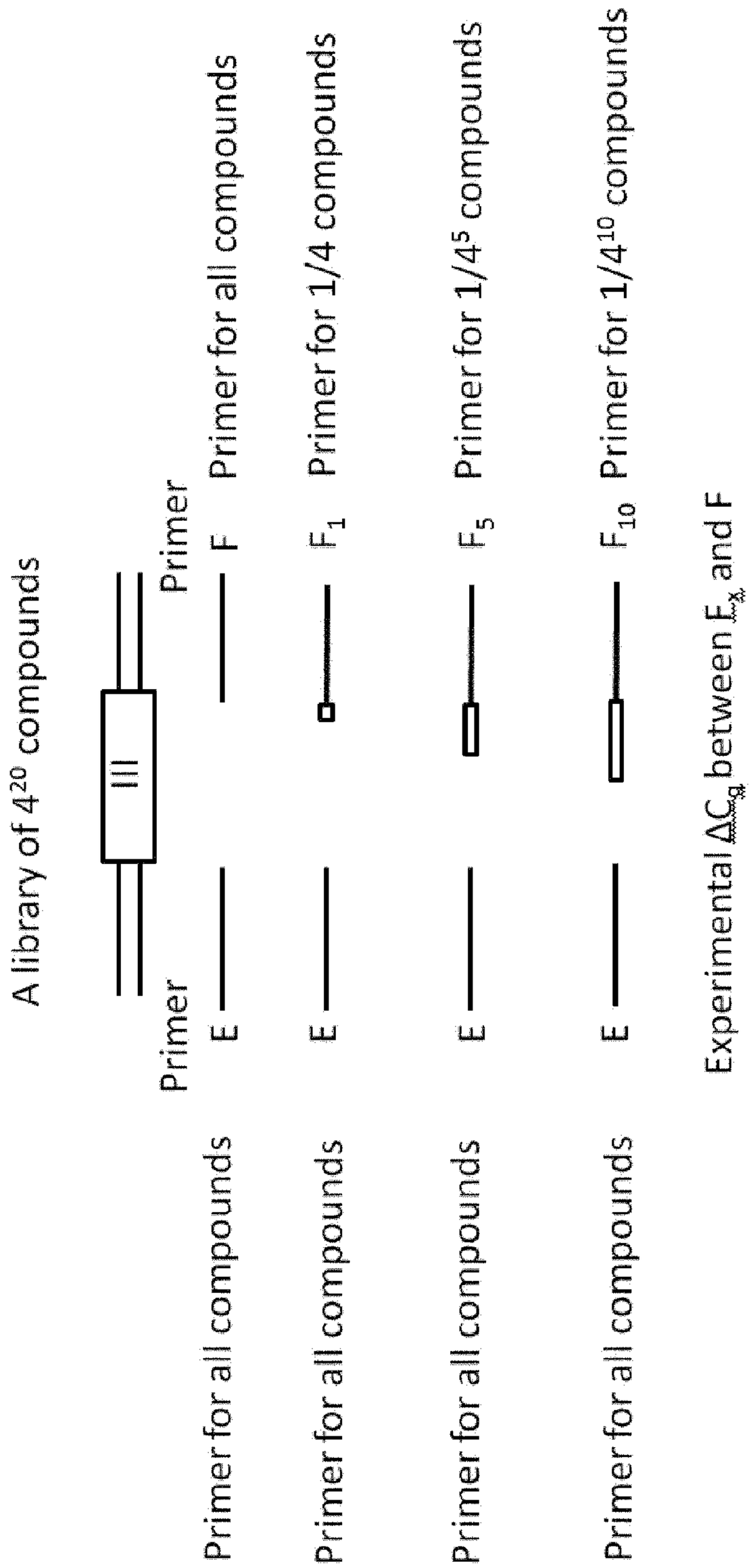

A library of 4^{10} compounds
Primer
III
Primer
Primer for all compounds
F
Primer for 1/4 compounds
F_1
Primer for $1/4^5$ compounds
F_5
Experimental ΔC_q between F_x and F
F_1 1.8
F_5 4.9

Figure 9


A library of 4^{20} compounds
Primer
Primer for all compounds
F
Primer for 1/4 compounds
F_1
Primer for $1/4^5$ compounds
F_5
Primer for $1/4^{10}$ compounds
F_{10}
Experimental ΔC_g between F_x and F
F_1 1.8
F_5 4.9
F_{10} 7.5
Primer for all compounds
Primer for all compounds
Primer for all compounds
Primer for all compounds

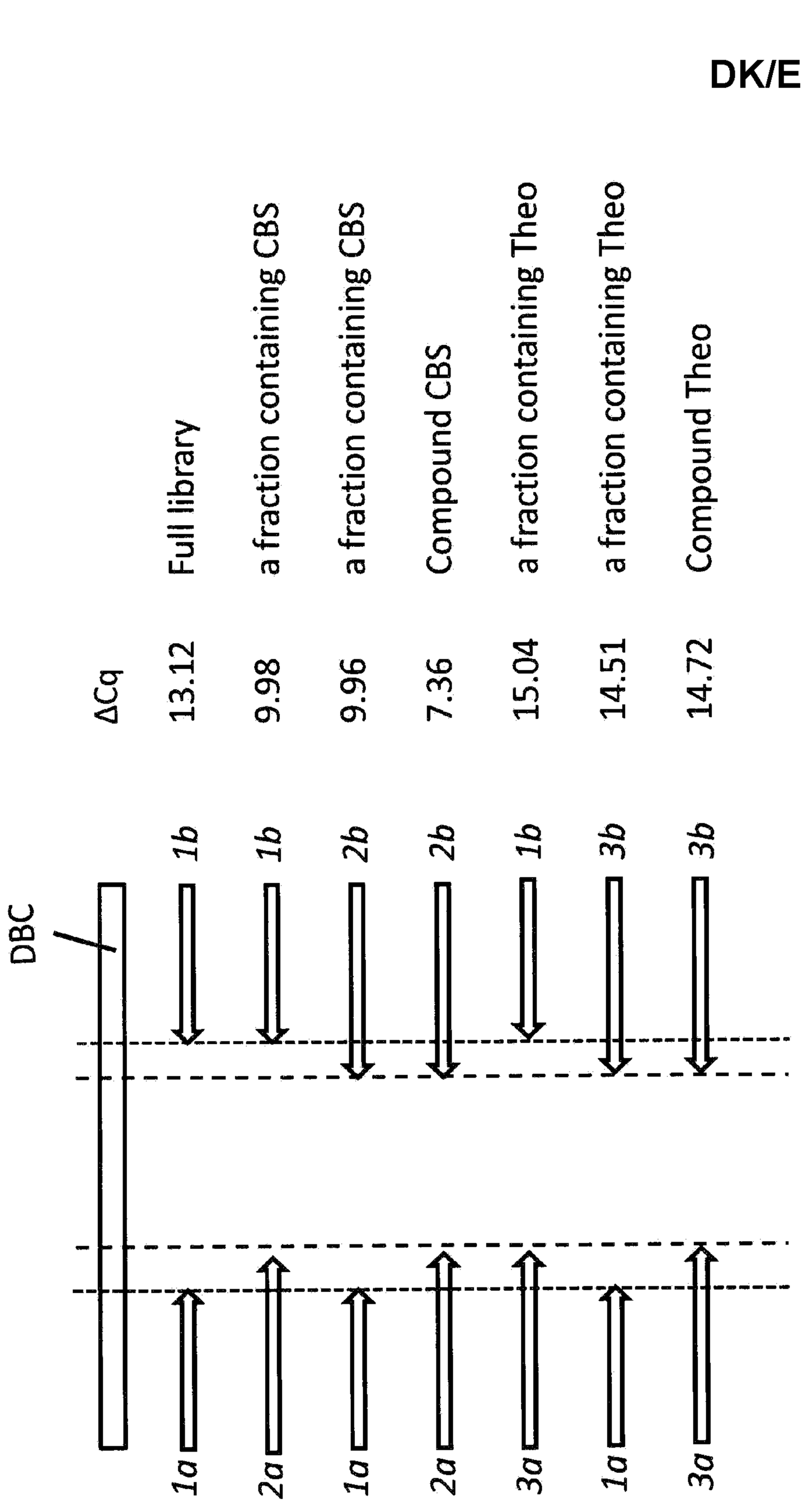

Full library
a fraction containing CBS
a fraction containing CBS
Compound CBS
a fraction containing Theo
a fraction containing Theo
Compound Theo
ΔCq
13.12
9.98
9.96
7.36
15.04
14.51
14.72
DBC
1b
1b
2b
2b
1b
3b
3b
1a
2a
1a
2a
3a
1a
3a

Figure 10

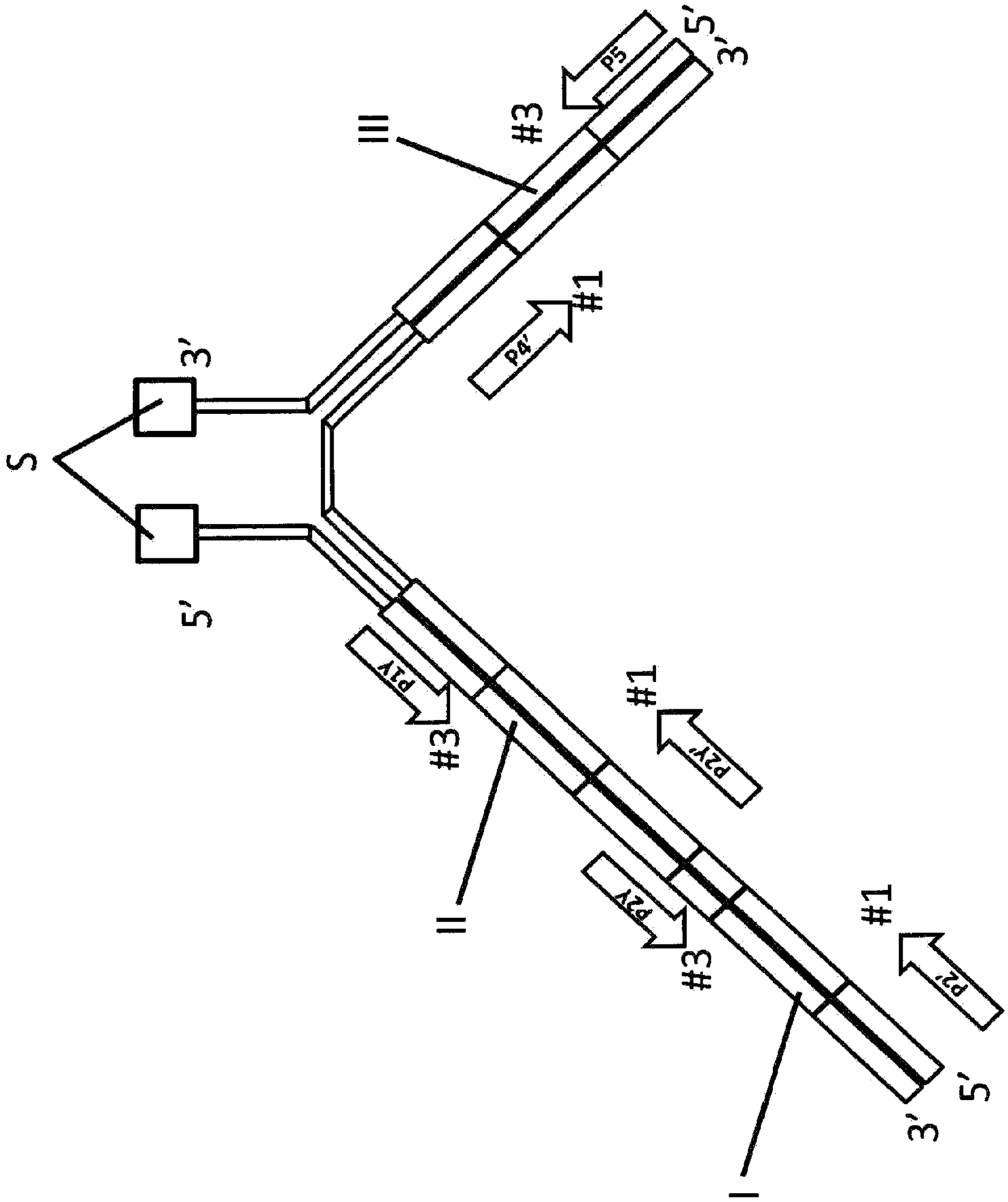


Figure 11