



- (51) **International Patent Classification:**
C07H 21/00 (2006.01) *C12Q 1/68* (2006.01)
- (21) **International Application Number:**
PCT/DK2014/050242
- (22) **International Filing Date:**
13 August 2014 (13.08.2014)
- (25) **Filing Language:** English
- (26) **Publication Language:** English
- (30) **Priority Data:**
PA 2013 70449 16 August 2013 (16.08.2013) DK
- (71) **Applicant:** UNIVERSITY OF COPENHAGEN
[DK/DK]; Nørregade 10, DK-1165 Copenhagen K (DK).
- (72) **Inventors:** VINTHER, Jeppe; Petersborgvej 4, 3.tv., DK-2100 Copenhagen Ø (DK). KIELPINSKI, Lukasz Jan; Rolfsvej 14, 3.th., DK-2000 Frederiksberg (DK). POULSEN, Line Dahl; Bagsværd Hovedgade 196, 2.th., DK-2880 Bagsværd (DK).
- (74) **Agent:** HØIBERG A/S; St. Kongensgade 59 A, DK-1264 Copenhagen K (DK).
- (81) **Designated States** (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT,

HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

- (84) **Designated States** (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Declarations under Rule 4.17:

— of inventorship (Rule 4.17(iv))

Published:

- with international search report (Art. 21(3))
- before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))
- with sequence listing part of description (Rule 5.2(a))

(54) **Title:** RNA PROBING METHOD AND REAGENTS

(57) **Abstract:** The present invention relates to a method for probing RNA structure and accessibility. Also provided are reagents that are useful for performing the method. The application provides a method of analyzing the structure of RNA polynucleotides by allowing for selection of full length cDNA.



RNA probing method and reagents

Field of invention

- 5 The present invention relates to a method for probing RNA structure and accessibility. Also provided are reagents that are useful for performing the method.

Background of invention

- 10 RNA structure, RNA accessibility and RNA-protein interactions are of fundamental importance to the function of RNA molecules. Gaining knowledge about RNA structure is essential to understand the interactions between RNA and RNA-binding therapeutic molecules. Thus over the last decades numerous attempts have been made to improve the methods for probing RNA structure and accessibility and for mapping RNA-protein
15 interactions.

Existing RNA structure probing methods (Wan et al., 2011) can broadly be divided in two different categories:

- 20 1) Methods based on ribonucleases having different specificity for single- or double-stranded RNA. Data readout from these methods relies on end-labelling of the probed RNA, followed by detection by electrophoresis, or on ligating adapters to the RNA, followed by PCR amplification and sequencing.
- 25 2) Methods based on reacting RNA in a structure-dependent manner with a reagent, which has the ability to terminate reverse transcription (Weeks et al., 2010). The intended reverse transcription termination sites (IRTTS) can then be detected by labelling the primer used for reverse transcription and doing electrophoresis or by ligating an adapter to the 3' end of the cDNA, doing PCR amplification and sequencing.
- 30 Methods based on chemical reagents have the advantage that they are not restricted to conditions (temperature, ion concentrations, pH) that support enzymatically active proteins and they do not require an extended stretch of single-stranded RNA to react. Moreover, some chemical reagents, such as Selective 2'-Hydroxyl Acylation analyzed by Primer Extension (SHAPE) reagents and hydroxyl radicals, have the advantage that
35 they react well with all 4 different RNA nucleotides and thus provide more information

on the RNA structure. SHAPE probing is based on the selective reaction of SHAPE reagents with the 2' OH group of the RNA Ribose group. When the nucleotide is not engaged in base-pairing, i.e. when the RNA is single-stranded at the position of the nucleotide, the 2' OH group is nucleophile and is therefore in general highly reactive with electrophilic SHAPE reagents. SHAPE reagents can be designed to be less reactive and more hydrophilic, so that they can enter bacterial, yeast or mammalian cells, thus allowing in vivo probing of RNA structure (Spitale et al., 2013).

Radical probing is another method for the analysis of RNA structure. Radicals lead to strand breakage, which is thought to be dependent on the accessibility of hydrogens to the solvent, but independent of the nature of the base. It is likely that the radical reactions leading to the breakage of RNA are similar and radical probing has been used for decades to monitor the accessibility of RNA.

However, for the methods described above, which are based on the detection of IRTTS, premature termination of the reverse transcriptase causes a background signal resulting in noise. This means that the analysis typically is restricted to one RNA and primer at a time, which severely limits the throughput. Moreover, the background noise increases when probing complex RNA mixtures and when using random priming. Thus new methods are urgently needed for structure probing of complex RNA populations.

Global mapping of RNA-protein interactions is of major importance for the understanding of cellular regulation. Current methods are largely based on cross-linking and immunoprecipitation (CLIP) methods, often followed by sequencing (CLIP-seq). The current CLIP-Seq protocols remain quite work-intensive and require a lot of optimisation. Thus there is a need for methods that enable global probing of RNA structure, RNA accessibility and RNA-protein interactions, in particular probing methods allowing analysis of complex RNA populations.

Summary of invention

The invention presented herein provides a method of analyzing the structure of RNA polynucleotides by allowing for selection of full length cDNA. New compounds and derivatives thereof are also disclosed.

In a first aspect the invention relates to a method of analyzing the structure of RNA polynucleotides by allowing for selection of full length cDNA, said method comprising:

a. covalently binding a probing reagent to said RNA polynucleotide, said probing reagent comprising

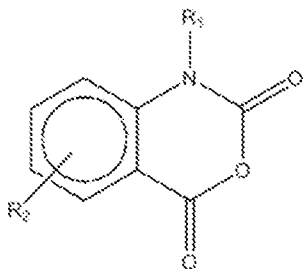
- 5 i. a group capable of being coupled to a solid support matrix; or
- ii. a solid support matrix,

and

b. adding primers to said RNA, and extending said primers by reverse transcriptase thereby generating a RNA/cDNA duplex,

10 wherein said probing reagent is capable of terminating the synthesis of cDNA from said RNA by reverse transcriptase.

In another aspect, the invention relates to a compound of formula I



15

(I)

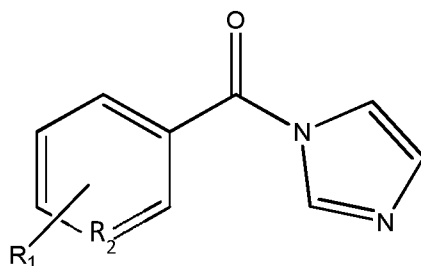
wherein R_1 is selected from the group consisting of: H, CH_3 , a ketone, an aldehyde, a primary amine, an azide, an alkene, an alkyne, a biotin, a PEG-biotin; and

20 wherein R_2 is selected from the group consisting of NO_2 , CF_3 , CCl_3 , H, CH_3 , a ketone, an aldehyde, a primary amine, an azide, an alkene, an alkyne, a biotin, a PEG-biotin;

with the proviso that if one of R_1 or R_2 is H, CH_3 , a ketone, an aldehyde, a primary amine, an azide, an alkene, an alkyne, a biotin, or a PEG-biotin, then the other of R_1 or R_2 is NO_2 , CF_3 or CCl_3 .

25

In yet another aspect, the invention relates to a compound of formula II



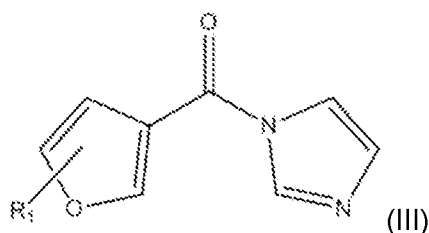
30

(II)

wherein R_1 is selected from the group consisting of: H, CH_3 , a ketone, an aldehyde, a primary amine, an azide, an alkene, an alkyne, a biotin, a PEG-biotin, and wherein R_2 is C or N.

5

In yet another aspect, the invention relates to a compound of formula III



10

wherein R_1 is selected from the group consisting of: H, CH_3 , a ketone, an aldehyde, a primary amine, an azide, an alkene, an alkyne, a biotin, a PEG-biotin.

15

In yet another aspect, the invention relates to the use of a compound of formula I, II, IIa or III and compounds of the invention for binding covalently to an RNA molecule or as a probing reagent in a method of the invention.

20

In yet another aspect, the invention relates to a compound comprising:

- a. a cDNA/RNA duplex
- b. a probing reagent of the invention comprising:
 - i. a group capable of being coupled to a solid support matrix; or
 - ii. a solid support matrix

said probing reagent being covalently bound to the RNA.

25

In yet another aspect, the invention relates to a kit for analysing the structure of RNA polynucleotides by allowing selection of full length cDNA, said kit comprising:

30

- a. a probing reagent capable of binding specifically with the base part of an RNA monomer or with the 2'OH group on the ribose ring of an RNA monomer, said probing reagent further comprising a solid support, or
- b. a probing reagent capable of binding specifically with the base part of an RNA monomer or with the 2'OH group on the ribose ring of an RNA monomer, a solid support capable of being bound covalently to said

- probing reagent, and a reagent to enable covalent binding of said solid support to said probing reagent,
- and
- a. a mixture of primers
- b. a reverse transcriptase
- c. buffers
- d. at least two control template RNAs.

Description of Drawings

- Fig. 1. General principle of probing RNA structure according to the invention.
- Fig. 2. Examples of electrophilic groups that can be comprised in probing reagents.
- Fig. 3. Examples of electrophilic groups coupled to propanone to allow subsequent binding to solid support.
- Fig. 4. Method of probing RNA with NPIA and binding to solid support.
- Fig. 5. Probing of RNA accessibility by coupling of aldehydes/ketones produced by radicals to solid support.
- Fig. 6. Detection of RNA-protein interactions.
- Fig. 7. Reaction of NPIA and Biotin hydrazide with ATP.
- Fig. 8: Hydrolysis of NPIA and NMIA.
- Fig. 9: SHAPES strategy for enrichment of structure signal.
- Fig. 10: RNaseP specificity domain probing.
- Fig. 11: ROC curve
- Fig. 12: Synthesis of N-Biotinyl-4-aminobenzoic acid imidazolid (BABAI)
- Fig. 13: BABAI can cross the cell membrane and reacts with RNA in vivo.
- Fig. 14: BABAI in vivo selection.

Detailed description of the drawings

- Figure 1. General principle of probing RNA structure according to the invention.
- RNA is probed with a reagent (black circle) that 1) reacts with the RNA in a structure and/or accessibility dependent manner, 2) can terminate reverse transcriptase and 3) allows for coupling to solid support and selection of full length cDNA. (A) After probing a primer (grey bar) is extended by reverse transcriptase to produce cDNA. Some cDNA reactions will terminate prematurely and thereby produce background signals. (B) The probing reagents used allow for coupling to solid support (greyed circle) as described in

the invention. (C) RNA-cDNA hybrids are treated with RNase specific for single stranded RNA and washed extensively. The arrows symbolise degradation of single-stranded RNA. (D) The hybrids remaining on the solid support comprise cDNAs that extend from the position of priming all the way to the position of probing. (E) The RNA is removed from the duplexes e.g. with RNaseH or by denaturation. The cDNA is analysed by capillary electrophoresis or prepared for analysis by massive parallel sequencing by ligating adaptors and PCR amplification.

Figure 2. Examples of electrophilic groups that can be comprised in probing reagents.

(A) General strategy for probing. The probing reagents (SHAPES) comprise (i) an electrophile (R) that can react selectively with the 2' OH on the RNA ribose ring in a structure-dependent manner and (ii) a group (SS), which may be a solid support matrix or a reactive group that can be coupled to a solid support matrix either directly or via a smaller molecule having high affinity for a solid support matrix. (B) and (C) Different classes of probing reagents that can react specifically with the RNA ribose 2' OH and in addition comprise one or more functional groups (X and Y). At least one of the functional groups should bind to a solid support matrix or be a reactive group that can be coupled to a solid support matrix either directly or via a smaller molecule having high affinity for a solid support matrix.

Figure 3. Examples of electrophilic groups coupled to propanone to allow binding to solid support.

NPIA, 1P7 and PAI are specific examples of probing reagents according to the invention. Here they are coupled to a propanone group which allows subsequent binding to a solid support.

Figure 4. Method of probing RNA with NPIA and binding to solid support.

N-propionylisatoic anhydride (NPIA) can react with RNA in a structure-dependent manner and facilitate coupling to solid support via reaction with Biotin Hydrazide.

Figure 5. Probing of RNA accessibility by coupling of aldehydes/ketones produced by radicals to solid support.

RNA is treated with radicals. In some cases the treatment will produce strand cleavage and a 5' aldehyde/ketone group (symbolised by =O). In other cases an aldehyde/ketone group can be produced by base elimination. The RNA is isolated

using standard methods. (A) A primer (grey bar) is extended by reverse transcriptase to produce cDNA. Some reactions will terminate prematurely and thereby produce background signals. (B) An aldehyde/ketone specific reagent is reacted with the RNA to couple RNA to solid support. (C) RNA-cDNA hybrids are treated with RNase specific for single-stranded RNA and washed extensively. The arrows symbolise degradation of single-stranded RNA. (D) The hybrids remaining on the solid support comprise cDNAs that extend from the position of priming all the way to the position of probing. (E) The RNA is removed from the duplexes e.g. with RNaseH or by denaturation. The cDNA is analysed by capillary electrophoresis or prepared for analysis by massive parallel sequencing by ligating adaptors and PCR amplification.

Fig. 6. Detection of RNA-protein interactions.

Proteins (black and grey) are UV crosslinked to RNA and RNA is isolated. (A) A primer (grey bar) is extended by reverse transcriptase to produce cDNA. Some reactions will terminate prematurely and thereby produce background signals. (B) An antibody (Y shape) specific for the protein in question (shown here is the black protein) is used to couple RNA to solid support. (C) RNA-cDNA hybrids are treated with RNase specific for single-stranded RNA and washed extensively. The arrows symbolise degradation of single-stranded RNA. (D) The hybrids remaining on the solid support comprise cDNAs that extend from the position of priming all the way to the position of the crosslinked protein. (E) The RNA is removed from the duplex with RNaseH. The cDNA is analysed by capillary electrophoresis or prepared for analysis by massive parallel sequencing by ligating adaptors and PCR amplification.

Fig. 7. Reaction of NPIA and Biotin hydrazide with ATP.

(A) Audiogram of a gel with reactions of radioactively labelled ATP with DMSO (control) or NPIA, which were subsequently reacted with the indicated reagents overnight. (B) The structure of ATP is shown with the reactive 2' and 3' hydroxyl groups indicated.

Fig. 8: Hydrolysis of NPIA and NMIA.

The plot shows fluorescence normalized to the fluorescence observed at the last time point as measured for NPIA (black dot) and NMIA (white dots).

Fig. 9: SHAPES strategy for enrichment of structure signal.

An RNA or a mixture of RNAs are probed with NPIA. A primer or a mix of primers is extended, but some of the reverse transcription reactions will terminate prematurely giving rise to background. After reverse transcription, biotin hydrazide is coupled to the propanone group of NPIA. Finally the sample is subjected to cleavage with RNaseI, which will cleave RNA not protected by cDNA, followed by selection on streptavidin bead to enrich for the RNA structure signal.

Fig. 10: RNaseP specificity domain probing.

In vitro transcribed RNaseP specificity domain RNA was folded and probed with NMIA, NPIA or DMSO (control). After probing, the samples were reverse-transcribed and part of the samples was subjected to selection as outlined in figure 13. For all samples, adaptors were ligated to the cDNA and after PCR amplification the libraries were sequenced on the Illumina platform. The resulting reads were mapped back to the RNaseP specificity domain sequence and the position causing Reverse transcriptase termination was identified and aggregated. The barplots show the resulting raw counts for the different positions across the RNaseP specificity domain sequence. From the top the following experiments are plotted: DMSO unselected, NMIA unselected, NPIA unselected, NMIA selected, NPIA selected; the term 'selected' refers to the selection of full-length compounds as described herein, by use of a solid support. The bottom plot shows the unpaired positions of the RNaseP specificity domain as annotated in the crystal structure (Krasilnikov et al., 2003).

Fig. 11: ROC curve

The data shown in figure 13 plotted as receiver operating characteristic (ROC) curves using the base pairing from the crystal structure as the binary classifier. The area under the ROC curve (AUC) is equal to the probability that a data point randomly chosen among the positions that are unpaired in the crystal structure will have a higher count in experimental data than a data point randomly chosen among the positions that are paired in the crystal structure.

Fig. 12: Synthesis of N-Biotinyl-4-aminobenzoic acid imidazolide (BABAI)

BABAI is the imidazolide of N-Biotinyl-4-aminobenzoic acid and can be synthesized through the reaction of equimolar amounts of the commercially available carbonyldiimidazole and N-Biotinyl-4-aminobenzoic acid in DMSO to give BABAI plus

imidazole and CO₂. The resulting solution can be used directly for probing experiments.

Fig. 13: BABAI can cross the cell membrane and reacts with RNA in vivo.

- 5 A) Total RNA extracted from zebrafish embryos and run on native agarose gel. B) Samples run on native agarose gel as in A) and transferred to a membrane where biotin was detected with HRP-conjugated streptavidin. CP: control for probing during extraction; L: 1 kb ladder; PC: positive control.

10 Fig. 14: BABAI in vivo selection.

- A) Percentage of RNaseP RNA (left) and E. Coli RNA spike-in controls (right) before (U, light grey boxes) and after selection (S, dark grey boxes) with BABAI, for replicate 1 (R1) and replicate 2 (R2). Y-axis shows percent RNA. B) and C) Termination counts (unique barcodes) mapping to the FTH1 mRNA IRE structure for the DMSO control sample, the unselected BABAI sample and two replicates of BABAI selected sample. TC: termination counts; U: unselected; S: selected with BABAI; R1: replicate 1; R2: replicate 2.

Definitions

20

Complementary DNA (cDNA)

As understood herein, complementary DNA (cDNA) is DNA synthesized from a messenger RNA (mRNA) template in a reaction catalysed by the enzymes reverse transcriptase. Full-length cDNAs are cDNA molecules in which the reverse transcription reaction catalysed by the reverse transcriptase did not terminate prematurely.

25

Electrophile

An electrophile is a reagent attracted to electrons that participates in a chemical reaction by accepting an electron pair in order to bond to a nucleophile. Most electrophiles are positively charged, have an atom that carries a partial positive charge, or have an atom that does not have an octet of electrons. Electrophiles are attacked by the most electron-populated part of a nucleophile.

30

Full-length cDNA

The term 'full-length cDNA' as understood herein refers to a cDNA molecule obtained by reverse transcription of an mRNA and extending from the priming site to the intended reverse transcription termination site (IRTTS). Thus the term does not encompass cDNAs for which the reverse transcription reaction terminates prematurely, i.e. before the IRTTS.

Nucleic acid

The term 'nucleic acid' as understood herein may refer to a molecule comprising ribonucleotides or deoxyribonucleotides, such as an RNA molecule, a DNA molecule or a cDNA molecule. The term 'duplex' refers to a double-stranded nucleic acid molecule, in which one strand is an RNA and the other is a DNA or cDNA strand.

Primers

A primer as understood herein is a nucleic acid that serves as a starting point for DNA synthesis reactions, such as PCR or reverse transcription. The primers may be degenerate primers, which are mixtures of similar, but not identical primers, or the primers may be targeted, i.e. they may be specific for a particular sequence.

Probing reagent

As understood herein, a probing reagent is a reagent that is capable of i) interacting with RNA, ii) terminating reverse transcription of RNA to cDNA by a reverse transcriptase, and iii) facilitating the binding of the RNA to a solid support.

Radical

As understood herein, a radical (or a free radical) is an atom, molecule, or ion that has unpaired valence electrons or an open electron shell, and thus is highly chemically reactive towards other substances, or towards itself.

Reverse transcription

Reverse transcription refers to the reaction involving reverse transcriptase and allowing synthesis of a cDNA molecule from an RNA template. Reverse transcription can be performed with targeted or random (degenerate) primers. The site at which the reaction terminates is termed 'intended reverse transcription termination site' (IRTTS).

RNA monomer

The term RNA monomer is to be understood as referring to adenine, uracil, cytosine or guanine.

SHAPE

5 The term 'SHAPE' stands for Selective 2'-Hydroxyl Acylation analysed by Primer Extension and refers to a reaction involving SHAPE reagents. SHAPE reagents are electrophiles which are capable of selectively reacting with the 2'OH group of the ribose group of RNA bases. The nucleophilicity of the 2'OH of the ribose is influenced by the 3' phosphodiester group, and thus nucleotides participating in base pairing are
10 unreactive.

SHAPES

The term 'SHAPES' refers to 'Selective 2'-Hydroxyl Acylation analysed by Primer Extension plus Selection' and refers to a reaction involving SHAPES reagents.
15 SHAPES reagents retain the ability of SHAPE reagents to react selectively with the 2'OH of the RNA ribose in a structure dependent fashion and terminate Reverse transcriptase, but in addition allow coupling to solid support.

Structure of RNA

20 As used herein, the term 'structure of RNA' may refer to the two-dimensional structure or the three-dimensional structure of RNA or to the accessibility of RNA to radicals or proteins. The term may also relate to the presence of binding sites, such as miRNA binding sites, siRNA binding sites, DNA binding sites, protein binding sites, RNA base reacting reagent binding sites. The term does not, however, refer to tagging of the 5'-
25 end of the RNA as known by the cage method (Takahashi et al., 2012).

Detailed description of the invention

The invention relates to a method of analyzing the structure of RNA polynucleotides by
30 allowing for selection of full length cDNA, said method comprising:

- a. Covalently binding a probing reagent to said RNA polynucleotide, said probing reagent comprising
 - i. a group capable of being coupled to a solid support matrix; or
 - 35 ii. a solid support matrix,

and

- b. adding primers to said RNA, and extending said primers by reverse transcriptase thereby generating a RNA/cDNA duplex,

wherein said probing reagent is capable of terminating the synthesis of cDNA from said RNA by reverse transcriptase.

5

The RNA to be probed may be any kind of coding or non-coding RNA, such as regulatory RNA, messenger RNA (mRNA) or pre-messenger RNA (pre-mRNA), micro RNA (miRNA), small interfering RNA (siRNA) and interfering RNA (RNAi), antisense RNA (asRNA), transfer RNA (tRNA), ribosomal RNA (rRNA), Piwi-interacting RNA (piRNA), Clustered Regularly Interspaced Short Palindromic Repeats RNA (CRISPR RNA), small nucleolar RNA (snRNA).

10

The RNA may originate from a prokaryote or a eukaryote. For example, the RNA may originate from *Bacillus subtilis*, *Escherichia coli*, and other pathogenic or non-pathogenic bacteria. The RNA may also originate from organisms such as yeasts and fungi, insects, birds, mammals, fish. The RNA may also originate from a human. The scope of the invention should not be restricted to a particular organism. Preferably, the RNA is comprised in a biological sample.

15

20

Thus the method of the invention relates to a method of generating an RNA/cDNA duplex. Probing reagents which can covalently bind RNA are contacted with the RNA to be probed (figure 1A). In some embodiments, probing reagents can only bind to unpaired nucleotides. In other embodiments, probing reagents can bind to paired nucleotides, so that they allow probing of e.g. double-stranded RNA.

25

In some embodiments, the method further comprises the step of fractionating the RNA molecules prior to covalent binding to the probing reagent. Such fractionation may be performed by methods known in the art, such as, but not limited to, using RNase or hydrolysis.

30

In all embodiments, primers are then added to the RNA and an RNA/cDNA duplex is generated by reverse transcription. The primers may comprise random primers, having random sequences and expected to hybridize at numerous positions. The primers may

also be targeted, i.e. designed for hybridizing in a specific location on the RNA. The primers may also be a mixture of random and targeted primers.

5 For each strand of RNA, the reverse transcription may terminate spontaneously, resulting in short duplexes, or it may terminate at the site where the probing reagent is bound, resulting in full-length duplexes (of which the DNA strand is a full-length cDNA as defined above). The present method allows selection of the full-length duplexes.

10 The probing reagents disclosed herein either comprise a solid support matrix or a group capable of being coupled to a solid support matrix. Thus the probing reagents bound to the duplexes can be coupled to a solid support.

15 The duplexes may then be treated with RNase specific for single-stranded RNA. If the reverse transcription terminated at the IRTTS created by the probing reagent, the RNA is single-stranded only 'after' the IRTTS, i.e. after the site where the probing reagent is bound (figure 1C). The term 'after' the IRTTS is here to be understood as referring to the part of the RNA which cannot be reverse transcribed due to the presence of the probing reagent (figure 1B). Upon treatment with RNase specific for single-stranded RNA, the full length duplex comprising RNA, cDNA and further comprising the probing agent is double-stranded along its whole length, from the priming site to the IRTTS (figure 1D).

25 If the reverse transcription terminates early, resulting in a short duplex, single-stranded RNA is present both after the IRTTS and before the IRTTS. The term 'before' the IRTTS refers herein to the part of the RNA which can potentially be reverse transcribed. Thus upon treatment with RNase specific for single-stranded RNA, the probing reagent, bound to single-stranded RNA, and not to the RNA-cDNA duplex, is removed from the duplex. The resulting double-stranded duplex comprises RNA and cDNA only, and does not further comprise the probing agent. The short double-stranded duplex extends from the priming site to a site upstream the IRTTS.

35 On the other hand, if the reverse transcription proceeds until the IRTTS, treatment of the full-length duplex with RNase specific for single-stranded RNA will result in a double-stranded duplex comprising RNA, cDNA and further comprising the probing agent.

The full length duplexes further comprising the probing reagent can now be coupled to a solid support. The short duplexes which do not comprise the probing agent cannot be coupled to a solid support. This allows specific selection of full-length duplexes
5 extending all the way from the reverse transcription priming site to the IRTTS. In some embodiments, the solid support to which the full length duplexes are coupled may be captured in order to allow removal of the short duplexes.

In order to analyse the cDNA, the RNA is removed in preferred embodiments. Methods
10 of removing the RNA include, but are not limited to, treatment with an RNase, denaturation and washing away of the RNA from the solid support, and any method known to the skilled person.

Thus in one aspect the invention relates to a compound comprising:

- 15 a. a cDNA/RNA duplex
b. a probing reagent, wherein the probing reagent is as defined below and comprises:
i. a group capable of being coupled to a solid support matrix; or
ii. a solid support matrix

20 said probing reagent being covalently bound to the RNA.

In some embodiments, the method allows enrichment of full-length cDNA with a factor of at least 10-fold, such as at least 20-fold, such as at least 30-fold, such as at least 40-
25 fold, such as at least 50-fold, such as at least 60-fold, such as at least 70-fold, such as at least 80-fold, such as at least 90-fold, such as at least 100-fold, compared to the method performed without covalently binding of a probing reagent.

Probing reagents

The probing reagents of the invention are reagents capable of covalently binding to an
30 RNA polynucleotide. Preferably, the probing reagent is capable of interacting with RNA polynucleotides even when the polynucleotides are at low concentration. Probing reagents include, but are not limited to, RNA binding proteins. Contacting the probing reagent with the RNA to be probed may be performed in vitro or in vivo. The probing reagents of the invention do not relate to reagents for detecting promoters and

transcriptional networks by the Cap Analysis of Gene Expression (CAGE) method (Takahashi et al., 2012), which involve tagging of the 5'-end of the RNA.

5 The probing reagents disclosed herein either comprise a solid support matrix or a group capable of being coupled to a solid support matrix. Thus the probing reagents bound to the duplexes can be coupled to a solid support, and the method of the invention comprises in some embodiments a step of covalently linking said group capable of being coupled to a solid support matrix to a solid support matrix. Such groups capable of being coupled to a solid support matrix are well known in the art. In 10 some embodiments, said group is selected from the group consisting of: amine, amide, aldehyde, ketone, ester, ether, hydrocarbon, propanone and biotin.

In some embodiments, the probing reagent comprises an electrophile that can react selectively with the 2'-OH group of the RNA ribose ring. In other embodiments, the 15 probing reagent can bind covalently to the base part of an RNA monomer. In preferred embodiments, the probing reagent can react with all four RNA monomers.

Electrophiles and SHAPE reagents

In some embodiments, the probing reagents are electrophiles such as SHAPE 20 reagents. SHAPE reagents are reagents that can selectively react with the 2' hydroxyl group of the RNA ribose, as defined above. The 2' hydroxyl group on the RNA ribose is a nucleophile when the nucleotide it is comprised in is not engaged in Watson-Crick base pairing, i.e. when the RNA at this position is single-stranded. SHAPE reagents preferably react with the 2' hydroxyl group with the same efficiency, independently of 25 the nature of the nucleotide and thus can provide information about the global structure of the probed RNA, with a single nucleotide resolution.

In some embodiments, the SHAPE reagent is selected from the group comprising isatoic anhydrides and derivatives thereof, acid imidazolides and derivatives thereof, 30 acid chlorides and derivatives thereof, isocyanates and derivatives thereof, acid cyanides and derivatives thereof, phthalic cyanides and derivatives thereof, benzoyl cyanide and derivatives thereof, benzoyl chloride and derivatives thereof, benzyl isocyanate and derivatives thereof. Thus in some embodiments, the electrophile that can react to the 2'OH group of the RNA ribose ring is selected from the group 35 consisting of: acid imidazolides, acid cyanides, acid chlorides, isocyanates, anhydrides,

isatoic anhydrides and phthalic anhydrides. For example, the SHAPE reagent may be N-methylisatoic anhydride (NMIA), or 1-methyl-7-nitroisatoic anhydride (1M7), or benzoyl-cyanide, or any other SHAPE reagent known in the art, such as, but not limited to, the SHAPE reagents disclosed in US2010/0035761.

5

The SHAPE reagents may be adapted or modified so that they are suitable for *in vivo* probing. In these embodiments, the properties of the SHAPE reagents may be modified so that they display properties such as, but not limited to, one or more of the following: increased solubility, e.g. in aqueous solution; decreased cross-reactivity (i.e. the probing reagent has low reactivity toward the nucleophiles in the cell which are not RNA nucleotides); increased ability to enter the cell; increased ability to enter the nucleus. Such reagents will thus be particularly well adapted for *in vivo* probing. Thus in specific embodiments, the SHAPE reagent is selected from the group comprising 2-methyl-3-furoic acid imidazolidine (FAI), 2-methylnicotinic acid imidazolidine (NAI), N-propanoneisatoic anhydride (NPIA), 1-propanone-7-nitroisatoic anhydride (1P7), N-Biotinyl-4-aminobenzoic acid imidazolidine (BABAI) and 2-propanone nicotinic acid imidazolidine (PAI). Reagents adapted for *in vivo* probing have reduced reactivity. Reagents adapted for *ex vivo* probing typically have high reactivity, and may be particularly relevant for embodiments in which snapshot-like information is needed as to RNA structure and/or accessibility.

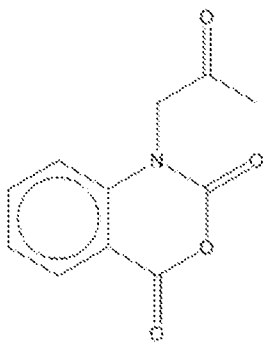
20

Embodiments of the invention further relate to probing reagents comprising an electrophile group R (figure 2A), such as a SHAPE reagent, and either a solid support matrix or at least one coupling group (X or Y, figure 2C) capable of being coupled to a solid support matrix.

25

Some embodiments relate to derivatives of NPIA, the structure of which is shown below:

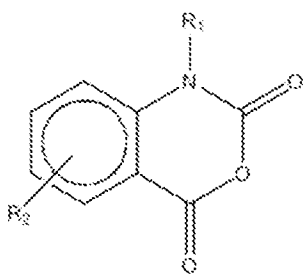
30



NPIA

Examples of NPIA derivatives are isatoic anhydride derivatives of formula I, comprising a group R_1 which may consist of a ketone, an aldehyde, a primary amine, an azide, an alkene, an alkyne, a biotin, a PEG-biotin or other relevant groups. For isatoic anhydride derivatives the coupling group can be added to the nitrogen in position 1 or to the aromatic ring. The derivative may also comprise an electron-withdrawing group R_2 , such as NO_2 , CF_3 , CCl_3 , which increases the reactivity of the reagent. Such compounds may not be desirable for embodiments where in vivo probing is performed.

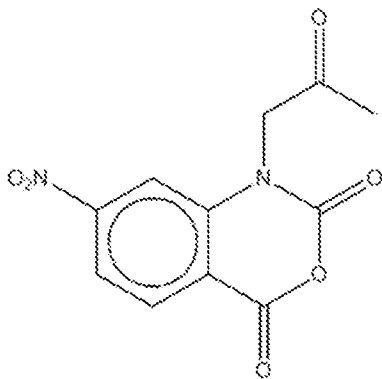
Thus in a further aspect, the invention provides compounds of formula I wherein R_1 is selected from the group consisting of: H, CH_3 , a ketone, an aldehyde, a primary amine, an azide, an alkene, an alkyne, a biotin, a PEG-biotin; and wherein R_2 is selected from the group consisting of NO_2 , CF_3 , CCl_3 , H, CH_3 , a ketone, an aldehyde, a primary amine, an azide, an alkene, an alkyne, a biotin, a PEG-biotin; with the proviso that if one of R_1 or R_2 is H, CH_3 , a ketone, an aldehyde, a primary amine, an azide, an alkene, an alkyne, a biotin, or a PEG-biotin, then the other of R_1 or R_2 is NO_2 , CF_3 or CCl_3 . Embodiments in which one of R_1 or R_2 is an azide, an alkene or an alkyne are particularly relevant for coupling the compound to a solid support by click chemistry.



(I)

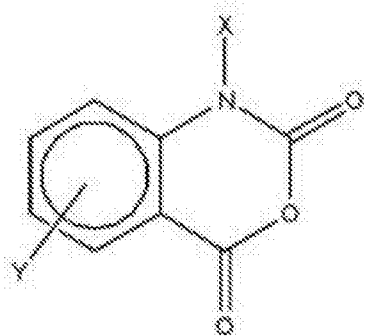
Isatoic anhydride derivatives of NPIA

Another suitable SHAPES reagent is 1P7 (1-propanone-7-nitroisatoic anhydride):



1P7 is a preferred SHAPES reagent.

Derivatives of 1P7 relevant for the invention are:

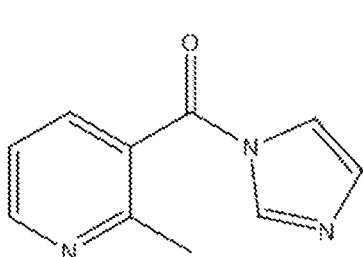


5

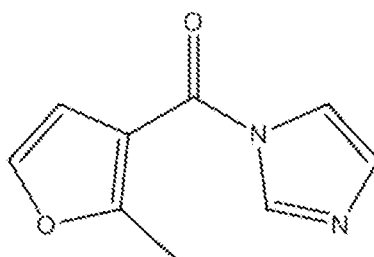
wherein at least one of the functional groups X or Y should bind to a solid support matrix or be a reactive group that can be coupled to a solid support matrix either directly or via a smaller molecule having high affinity for a solid support matrix.

10

Examples of known SHAPE reagents that have higher solubility and lower reactivity and are thus suited for probing RNA in vivo are the imidazolides 2-methylnicotinic acid imidazolide (NAI) and 2-methyl-3-furoic acid imidazolide (FAI):



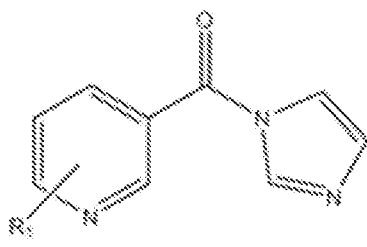
NAI



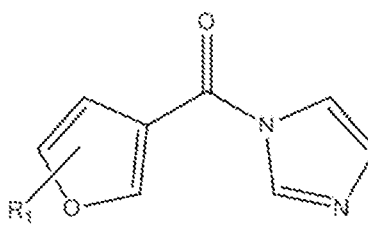
FAI

15

These can be modified to SHAPES reagents of formula IIa or III as illustrated here:



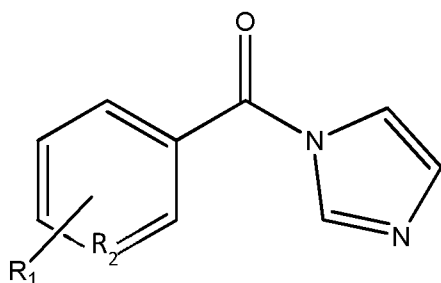
(IIa)



(III)

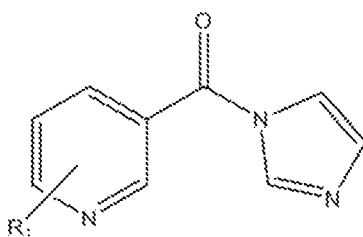
wherein R_1 is selected from the group consisting of: H, CH_3 , a ketone, an aldehyde, a primary amine, an azide, an alkene, an alkyne, a biotin, a PEG-biotin, or another relevant group that can be coupled to a solid support. Azides, alkenes and alkynes are particularly well suited for coupling the compound to a solid support by click chemistry.

Thus in another aspect the invention relates to a compound of formula II:



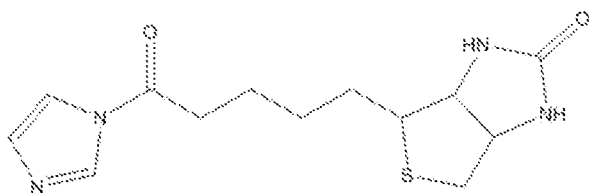
wherein R_1 is selected from the group consisting of: H, CH_3 , a ketone, an aldehyde, a primary amine, an azide, an alkene, an alkyne, a biotin, a PEG-biotin, and wherein R_2 is C or N. In preferred embodiments, R_1 is not in position 2.

In some embodiments, R_2 is N and the compound of the invention is of formula IIa:



In yet another aspect the invention relates to a compound of formula III, wherein R_1 is selected from the group consisting of: H, CH_3 , a ketone, an aldehyde, a primary amine, an azide, an alkene, an alkyne, a biotin, a PEG-biotin. In preferred embodiments, R_1 is not in position 2.

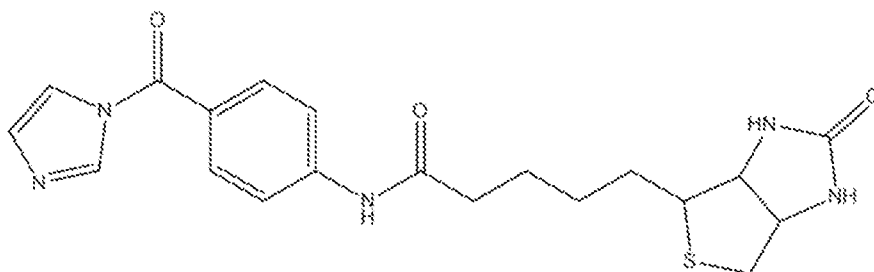
Other embodiments relate to derivatives of biotin. Biotin is an acid, which can be used to synthesise the corresponding imidazolides by reaction with diimidazole.



Biotin imidazolidine

5

N-(+)-biotinyl-4-aminobenzoic acid is an enzyme substrate used in the assay of biotinidase. This acid can also be converted to the corresponding imidazolidine, N-Biotinyl-4-aminobenzoic acid imidazolidine (BABAI; see example 15):



10

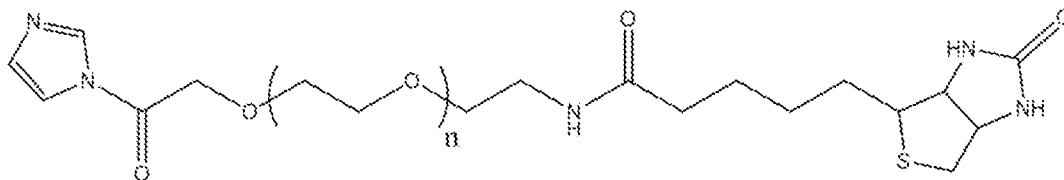
The reactivity of such compounds decreases because of the presence of the aromatic ring, which can be desirable when the probing is to be performed in vivo. In a preferred embodiment, the probing reagent is N-Biotinyl-4-aminobenzoic acid imidazolidine (BABAI).

15

There are different biotin molecules which have extended "arms" to facilitate binding to streptavidin and enhance solubility. Polyethylene glycol (PEG) is the most used group for extension and addition of hydrophilic groups. Thus some embodiments relate to biotin derivatives such as imidazolides, comprising one or more (n) PEG group.

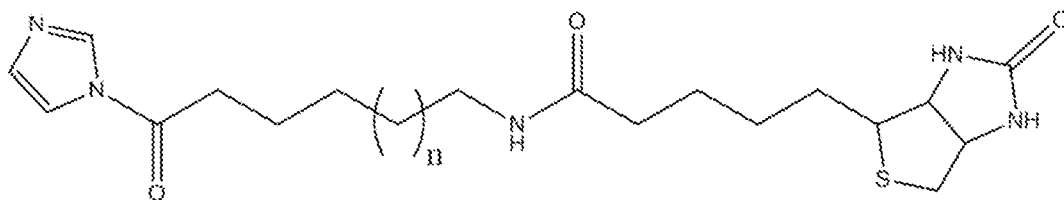
20

Increasing n results in increased hydrophilicity but also results in reduced capacity of the compound to enter the cells. Thus considerations as to whether the probing is to be performed in vivo or ex vivo will direct the choice of the value of n.



n is an odd integer superior or equal to 1.

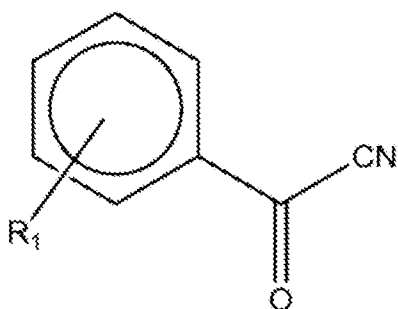
- 5 In some embodiments where compounds are desired that are membrane permeable, other groups than PEG may be used for extending the biotin arm. For example, the biotin imidazolidine may be extended by a C6 linker or longer carbon linkers:



- 10 n is any integer superior or equal to 1.

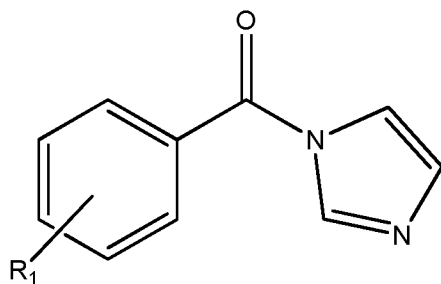
Other SHAPES reagents may be derived from:

- aromatic acid cyanides:

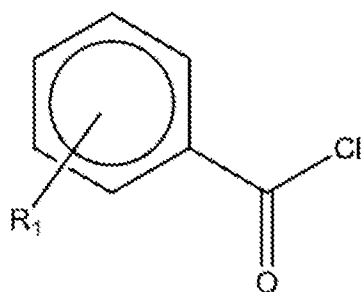


15

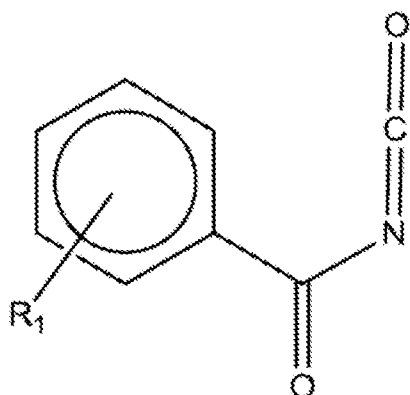
- from aromatic imidazolidines:



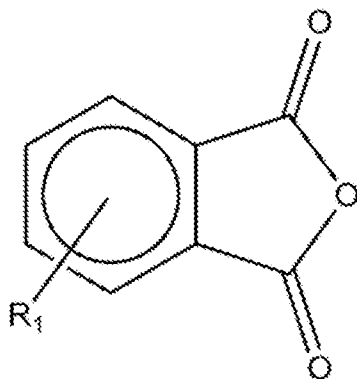
- from aromatic acid chlorides:



- from aromatic isocyanates:

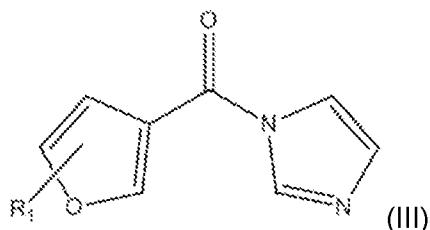


5 - from phthalic anhydrides:



10 R_1 is a group that can be coupled to solid support. R_1 can be a group containing a ketone, an aldehyde, a primary amine, an azide, an alkene, an alkyne, a biotin, a PEG-biotin or other relevant groups that can be coupled to solid support.

In yet another aspect, the invention relates to a compound of formula III



wherein R₁ is selected from the group consisting of: H, CH₃, a ketone, an aldehyde, a primary amine, an azide, an alkene, an alkyne, a biotin, a PEG-biotin. In some
5 embodiments, R₁ is not in position 2.

Solid support

The solid support is selected from the group comprising a bead, a dish, a well, a stick,
10 a membrane, a resin, and any solid support which will appear suitable to the skilled
person. The solid support matrix preferably has a hydrophilic surface. In some
embodiments, the solid support is a magnetic bead or a resin. The solid support matrix
may be selected from the group consisting of: diatomaceous earth, celite, squalane,
hexadecane, dialkyl phthalates, tetrachlorophthalates, polyethylene glycol, and
15 polysiloxanes. As explained above, the solid support may allow selection of full-length
duplexes comprising a probing reagent. In some embodiments, capturing of the solid
support allows removal of the shorter duplexes which do not comprise a probing
reagent. Such shorter may be removed by washing and/or centrifuging or any other
method known in the art.

The group capable of being coupled to a solid support matrix may be any group which
can be added to the electrophile R in order to facilitate or enable coupling to a solid
support. Such groups allowing binding to a solid support should preferably not modify
the reactivity of the probing reagent. Such groups capable of being coupled to a solid
25 support matrix are well known in the art. In some embodiments, said group is selected
from the group consisting of: amine, amide, aldehyde, ketone, ester, ether,
hydrocarbon, propanone and biotin. Thus preferred embodiments relate to coupling
groups which are substantially unreactive but which can be reacted in order to enable
coupling to a solid support. Other embodiments relate to coupling groups which are
30 small and do not provide steric hindrances. A preferred coupling group is propanone,

which is essentially unreactive but can be reacted with hydrazide to allow for coupling to solid support. Another preferred coupling group is biotin.

5 The addition of a coupling group to the electrophile group of the probing reagent should be performed so that the coupling group has a minimal chance of disturbing the electrophilic status of the electrophile group, or otherwise perturbing the reactivity of the electrophile group with the RNA polynucleotides. For example, if the electrophile comprises a heterocyclic or an aromatic ring, preferably the coupling group will be added to said ring, for example the coupling group may replace a methyl group on a
10 heterocyclic ring.

Reaction of the coupling groups, such as the ketone or aldehyde group of a propanone group, with biotin hydrazide to form a hydrazone, or an alkoxyamine group with a carbonyl group to form an oxime, may be accelerated by the use of catalysts known in
15 the art, such as aniline and *m*-phenylenediamine.

Radical probing

In some embodiments, the probing reagent comprises a radical (or free radical). Radicals are very reactive in solution. They can either be produced by radiation or by
20 chemical means. The reaction of radicals with RNA is complex and diverse and can lead to formation of many different reaction products with some of these leading to the breakage of the RNA backbone. Radicals react by abstracting hydrogens from other molecules. For DNA, strand breaks induced by hydroxyl radicals correlate well with the accessibility of the C5' and C4' hydrogens, meaning that the strand breakage is
25 dependent on the accessibility of the C4' and C5' hydrogens. For DNA it is known that radicals lead to the formation of products having 5' aldehydes and in some cases ketones, as well as to the formation of abasic sites, which also result in aldehyde groups being produced. It is generally assumed that the effect of radicals on RNA is similar. Thus radicals are capable of modifying the RNA molecule in such a way that
30 the reverse transcription will terminate.

In some embodiments of the invention, the probing reagent comprises a radical and the aldehyde and/or ketone group that it induces. The aldehyde and/or ketone group may be coupled to a solid support after selective reaction with reagents allowing coupling.
35 Preferred embodiments relate to coupling reagents such as hydrazide-biotin or N'-

aminoxymethylcarbonylhydrazino-D-biotin. The aldehyde and/or ketone group may be at the 5'-end. Thus in some embodiments, the method further comprises treating the RNA polynucleotide with a radical to create RNA fragments with ketone or aldehyde groups, wherein the probing reagent is capable of binding to said aldehyde or ketone group.

Radical formation may be induced by chemicals such as the Fenton's reagents or by radiation, such as UV radiation, ionizing radiation, see for example Shcherbakova and Mitra, Methods Enzymol. 2009;468:31-46, and Adilakshmi et al, Methods Enzymol. 2009;468:239-58.

Proteins

In some embodiments, the probing reagent is an RNA binding protein, which in some embodiments may be covalently attached to RNA by UV crosslinking the sample or adding a chemical that creates a covalent bond between the RNA and the RNA binding protein.

The protein can further be coupled to a coupling group such as an antibody, which allows coupling to a solid support. For example, the RNA binding protein is crosslinked to the RNA to be probed by the use of UV light. After the antibody binds specifically to the RNA binding protein, the antibody is coupled to solid support via a molecule that binds specifically to the Fc region of antibodies, such as Protein A or Protein G. The antibody may also be covalently coupled to solid support or the antibody may comprise a group allowing for coupling to a solid support, such as but not limited to biotin.

Methods of purifying RNA

The RNA may be purified prior to probing, if the probing reagents are not suited for in vivo probing. Thus in some embodiments the method is performed on purified RNA. Methods of purifying the RNA are methods known in the art. Such methods include, but are not limited to: acid phenol-guanidium thiocyanate-chloroform extraction; selection of poly(A)⁺ RNA by oligo(dT)-cellulose chromatography or by batch chromatography; separation of RNA according to size with agarose gels, optionally including enzymatic digestion of the RNA (Sambrook and Russell). In some embodiments, purification of the RNA prior to probing comprises a step of lysing the cell.

In some embodiments, the probing reagent is bound to RNA in a living cell. Preferably, the probing agent is capable of crossing physical barriers such as membranes. The probing agent may thus be capable of crossing e.g. the cellular membrane, the nuclear membrane, the chorion. In such embodiments where RNA probing is performed in vivo, the RNA may be purified after probing, prior to the reverse transcription reaction, using methods known in the art including, but not limited to: acid phenol-guanidium thiocyanate-chloroform extraction; selection of poly(A)⁺ RNA by oligo(dT)-cellulose chromatography or by batch chromatography; separation of RNA according to size with agarose gels, optionally including enzymatic digestion of the RNA (Sambrook and Russell). In some embodiments, the probing reagent is bound to RNA within a living cell and the remaining steps of the method are performed after cell lysis.

Methods of analyzing full length cDNA

The full-length cDNAs selected by the present method may be analysed by methods known in the art. For example, the cDNA may be digested by restriction enzymes and submitted to restriction fragment analysis, or it may be sequenced.

In one embodiment, the invention relates to a method as described above, further comprising the steps of:

- a. treating said bound RNA/cDNA duplex with RNase;
- b. washing said bound duplex;
- c. isolating said cDNA;
- d. optionally, digesting said cDNA with at least one restriction enzyme;
- e. analysing said cDNA by restriction fragment length analysis or sequencing.

Use of compounds of formula I, II, III as probing reagents

In another aspect the invention relates to the use of a compound of formula I as a probing reagent in the method disclosed herein.

In yet another aspect the invention relates to the use of a compound of formula II as a probing reagent in the method disclosed herein.

In yet another aspect the invention relates to the use of a compound of formula III as a probing reagent in the method disclosed herein.

In yet another aspect the invention relates to the use of a compound of formula I, II or III for binding covalently to an RNA molecule.

5 Preferred embodiments relate to the use of N-propanone isatoic anhydride (NPiA) as a probing reagent in the method disclosed herein. Other preferred embodiments relate to the use of N-Biotinyl-4-aminobenzoic acid imidazolide (BABAI) in the method disclosed herein. Other preferred embodiments relate to the use of 1-propanone-7-nitroisatoic anhydride (1P7) in the method disclosed herein.

10

In yet another aspect the invention relates to a compound comprising:

- a. a cDNA/RNA duplex
- b. a probing reagent, wherein the probing reagent is as defined above and comprises:

15

- i. a group capable of being coupled to a solid support matrix; or
- ii. a solid support matrix

said probing reagent being covalently bound to the RNA.

In yet another aspect the invention relates to a compound comprising:

20

- a. a cDNA/RNA duplex
- b. a probing reagent, wherein the probing reagent is an electrophile as defined above and comprises:
 - i. a group capable of being coupled to a solid support matrix; or
 - ii. a solid support matrix

25

said probing reagent being covalently bound to the RNA.

Kit of parts

It is another object of the invention to provide a kit of parts for analyzing the structure of RNA polynucleotides by allowing selection of full length cDNA, said kit comprising:

30

- a. a probing reagent capable of binding specifically with the base part of an RNA monomer or with the 2'OH group on the ribose ring of an RNA monomer, said probing reagent further comprising a solid support, or

35

- b. a probing reagent capable of binding specifically with the base part of an RNA monomer or with the 2'OH group on the ribose ring of an RNA

monomer, a solid support capable of being bound covalently to said probing reagent, and a reagent to enable covalent binding of said solid support to said probing reagent,

and

5

- a. a mixture of primers
- b. a reverse transcriptase
- c. buffers
- d. at least two control template RNAs.

10

The mixture of primers may be such that it allows random priming of the reverse transcriptase, as described earlier. The buffers comprised in the kit should be such that they allow the reverse transcriptase to perform reverse transcription. The kit further comprises at least two control template RNAs. One of the control RNAs may be used as a negative control and the other as a positive control. The negative control RNA should be added to the reaction after the probing has been performed. If the reaction is selective, the final mixture containing the selected cDNAs does not contain the negative control RNA or the corresponding cDNA. The positive control RNA should be added to the reaction before the probing is performed. The corresponding cDNA should always be detected in the final mixture if the reaction occurred properly.

20

Preferably, the control template RNAs are RNAs for which the structure is known. In preferred embodiments, the RNAs are ribosomal RNAs. For example, the ribosomal RNAs originate from a model organism such as *Escherichia coli* or *Saccharomyces cerevisiae*.

25

In some embodiments, the probing reagent comprised in the kit comprises an electrophile that can react selectively with the 2' OH group of the RNA ribose ring. The probing reagent is capable of binding selectively to unpaired DNA. The probing reagent can react with all four RNA monomers.

30

In some embodiments, the electrophile that can react to the 2'OH group of the RNA ribose ring is selected from the group consisting of: acid imidazolides, acid cyanides, acid chlorides, isocyanates, anhydrides, isatoic anhydrides and phthalic anhydrides.

35

The reactive group capable of being coupled to a solid support matrix may be selected from the group consisting of: amine, amide, aldehyde, ketone, ester, ether,

hydrocarbon, propanone and biotin. The solid support matrix preferably has a hydrophilic surface. The solid support may be selected from a bead, a dish, a well, a stick, a resin. In some embodiments, the solid support is a magnetic bead or a resin. The solid support matrix may be selected from the group consisting of: diatomaceous
5 earth, celite, squalane, hexadecane, dialkyl phthalates, tetrachlorophthalates, polyethylene glycol, and polysiloxanes.

In preferred embodiments, the probing reagent comprises: N-propanone-isatoic anhydride (NPIA), 1-propanone-7-nitroisatoic anhydride (1P7), or N-Biotinyl-4-
10 aminobenzoic acid imidazolidine (BABA).

Examples

Examples 1 to 13 describe experiments to validate N-propanone-isatoic anhydride
15 (NPIA) as a probing reagent (SHAPES)

Example 1

This example describes the materials used.

RNase-free water

[Gamma-32P]ATP

20 Reaction buffer: 100 mM HEPES pH 8.0, 6 mM MgCl₂ and 100 mM NaCl

10 M m-phenylenediamine in water

6 x loading buffer

10 x TBE buffer (Tris-borate-EDTA)

29:1 Acrylamide/Bis-isacrylamide

25 3.3 x Folding buffer: 333 mM HEPES pH 8.0, 333 mM NaCl

10 x Mg²⁺: 100 mM MgCl₂

DMSO anhydrous

10 x NPIA in DMSO: The concentration of the NPIA solution can vary, depending on the preferred extent of RNA modification. Here a 10xNPIA solution of 60 mM or 500
30 mM was used.

pEX-A-RNaseP-SD plasmid

DNA primer for reverse transcription

PrimeScript Reverse Transcriptase (TAKARA)

5 x Reverse Transcriptase buffer: 250 mM HEPES pH 8.3, 375 mM KCl, 15 mM MgCl₂

35 3.3 M/0.6 M Sorbitol/Trehalose mix

10 mM dNTPs

Agencourt RNAClean XP Kit (BECKMAN COULTER)

1 M Na-Citrate pH 6.0

Biotin (long arm) hydrazide (VECTOR lab)

1 M Tris-HCl pH 7.0

5 1 M Tris-HCl pH 8.5

0.5 M EDTA pH 8.0

RNase I ribonuclease

MPG Streptavidin (TAKARA)

20 µg/µl *E. Coli* tRNA mix: 30 mg *E. Coli* tRNA (ribonucleic acid, transfer from
10 *Escherichia Coli*, lyophilized powder (Sigma-Aldrich) was dissolved in 400 µl water, and
45 µl 10 × RQ1 DNase buffer and 30 µl RQ1 RNase-Free DNase were added. The
sample was incubated at 37°C for 2 h. After incubation, 10 µl of 0.5 M EDTA (pH 8.0),
10 µl of 10% SDS, and 10 µl of 10 ng/ml Proteinase K were added to the tRNA
15 solution, and the sample was incubated at 45°C for 30 min. 500 µl of phenol–
chloroform was added, and the sample was centrifuged at 20,000 g for 5 min at room
temperature. The aqueous phase was collected and 1 volume of chloroform was
added. The sample was then centrifuged at 20,000 g for 5 min at room temperature.
The supernatant was collected. 1/10 volume of 3 M NaCl, 1 µl 5 mg/ml glycogen, and
1 volume isopropanol were added, and the sample was centrifuged once again at
20 20,000 g for 5 min at room temperature. The supernatant was removed and 900 µl of
80% ethanol was added to the tRNA pellet. It was centrifuged at 20,000 g for 5 min at
room temperature, the supernatant was removed, and another 900 µl of 80% ethanol
was added to the tRNA pellet. After centrifugation at 20,000 g for 5 min at room
temperature and removing of the supernatant, the tRNA pellet was dissolved in 20 µl
25 water. The tRNA concentration was measured using a Nanodrop 1000 (Thermo
scientific). The tRNA was then diluted to 20 µg/µl in water and stored at –20°C.

Wash buffer 1: 4.5 M NaCl, 50 mM EDTA pH 8.0

Wash buffer 2: 300 mM NaCl, 1 mM EDTA pH 8.0

30 Wash buffer 3: 20 mM Tris-HCl pH 8.5, 1 mM EDTA pH 8.0, 500 mM NaOAc pH 6.1,
0.4% SDS

Wash buffer 4: 10 mM Tris-HCl pH 8.5, 1 mM EDTA pH 8.0, 500 mM NaOAc pH 6.1
50 mM NaOH

Ligation adapter: 5' PHO-AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT-3NHC₃
(SEQ ID NO: 1), where PHO is a 5'-phosphate modification and 3NHC₃ is a 3'-amino
35 modification.

- The cDNA ligation mix was scaled up to the required volume mixture made of 1 volume of 10x ligation buffer (0.5 M MOPS (pH 7.5), 0.1 M KCl, 50 mM MgCl₂, and 10 mM DTT), 0.5 volume of 1 mM ATP, half of the volume of 50 mM MnCl₂, two volumes of 50% PEG6000, two volumes of 5M betaine, 0.5 volume of the 100 μM ligation adapter, 0.5 volume of CircLigase enzyme (100U/μl, Epicentre). The ligation mix was prepared right before the reaction and stored on ice. In this example we used CircLigase enzyme to ligate the adapter to the 3'-end of cDNA molecule. Other methods of ligation, such as single-strand linker ligation (Takahashi H, Lassmann T, Murata M, Carninci P. 2012. Nature protocols 7: 542-61), can also be used.
- Salt solution used for nucleic acid precipitation, such as 7.5M NH₄Ac (ammonium acetate).
- 5mg/ml glycogen
- Ethanol
- Phusion DNA polymerase (Finnzymes)
- 5x HF buffer (Finnzymes)
- Primer 1 (SEQ ID NO: 2): (5' – AATGATACGGCGACCAACGAGATCTACACTCTTTCCCTACACGACGCT – 3')
- Primer 2: (5'- CAAGCAGAAGACGGCATACGAGATXXXXXXGTGACTGGAGTTCAGACGTGTGCTC TTCCGATCT-3', where XXXXXX denotes the reaction-specific barcode, it is one of the sequences GGACGG (SEQ ID NO: 3), TGACAT (SEQ ID NO: 4), GGAAC (SEQ ID NO: 5), TACAAG (SEQ ID NO: 6), GTAGCC (SEQ ID NO: 7), ATTGGC (SEQ ID NO: 8) or CACTGT (SEQ ID NO: 9)). Overhangs match the Solexa Illumina sequencing platform and are copyrighted by Illumina Inc.
- The PCR mix was scaled up to the required volume mixture made of 3 volumes of 10 μM primer 1, 10 volumes of 5x HF buffer (Finnzymes), 1 volume of 10 mM dNTPs, 27.5 volumes of H₂O, 1 volume of Phusion DNA polymerase (Finnzymes) and 2.5 volumes of 10 μM primer 2. The PCR mix was prepared right before the reaction and stored on ice.
- Ampure XP beads (Beckman Coulter)
- Bioanalyzer system with DNA 1000 chip (Agilent)

Example 2

2'-acylation of ATP with NP1A, followed by biotinylation with biotin (long arm) hydrazide.

10000 counts per min (cpm)/ μ l of radiolabeled ATP were incubated with 50 mM NP1A in 100 mM HEPES pH 8.0, 6 mM $MgCl_2$ and 100 mM NaCl (1 hour, 37°C, total volume 10 μ l). 1 μ l of the reaction was mixed with 2 μ l 1 M Na-Citrate pH 6.0, 6.75 μ l 30 mM biotin (long arm) hydrazide, and water to a final volume of 28.75 μ l. A control reaction with DMSO instead of biotin (long arm) hydrazide was included. Also, a reaction with 500 mM of the catalyst m-phenylenediamine (mPDA) (Rashidian et al., 2013) was carried out. After mixing by pipetting, the reactions were incubated 15 hours at 23°C in the dark.

1 μ l of each reaction was mixed with 9 μ l water and 2 μ l 6 x loading buffer, and loaded on a 30% native polyacrylamide gel (29:1 Acrylamide/Bis-acrylamide, 1 x TBE). After electrophoresis (14 W, 1 hour), the results were analysed with phosphorimaging (STORM, Molecular Dynamics).

Example 3

Determination of the rate of hydrolysis for SHAPE reagents

1.5 μ l 10 μ M NP1A or NM1A was hydrolysed in reaction buffer (100 mM potassium phosphate pH 8.0, 10% v/v DMSO, 250 mM NaCl) for 1 hour, and the excitation source and emission profile were determined using the Fluorescence Profiler feature on NanoDrop 3300 (Thermo Scientific). Reactions were initiated by addition of 1 μ l 10 mM NP1A or NM1A to 1 ml reaction buffer, and the formation of the hydrolysis product was measured as an increase in fluorescence (excitation 375 nm, emission 440 nm, 30 sec. time points).

Example 4

Synthesis of Bacillus Subtilis RNase P RNA

A DNA template of the specificity domain of *B. Subtilis* RNase P inserted in a 5' and 3' flanking cassette structure (Lucks et al., 2011) was synthesized *de novo* (Eurofins MWG Operon) and inserted into the standard vector pEX-A. The synthesized DNA fragment contained a T7 RNA promoter sequence in the 5'-end for transcription and a *Bsal* restriction site in the 3'-end for linearization. The plasmid was transformed into One Shot TOP10 chemically competent *Escherichia Coli* (Invitrogen), and grown on agar plates containing ampicillin. One clone was picked, and after culturing and plasmid isolation, the sequence was verified with DNA sequencing. 5 μ g plasmid were linearized with *Bsal*-HF (20 μ l, 37°C, 1 hour) in a reaction containing 1 x NEB buffer 4 (New England Biolabs), 100 μ g/ml BSA, 20 U *Bsal*-HF (New England Biolabs). The

DNA template was purified with phenol/chloroform extraction and ethanol precipitated. Transcription reactions (200 µl, 37°C, 4 hours) contained 40 mM Tris-HCl pH 8.0, 6 mM MgCl₂, 1 mM Spermidin, 5 mM DTT, 2 mM of each NTP, 1 µg linearized DNA template and 20 U of T7 RNA polymerase. After transcription, the RNA was purified by denaturing polyacrylamide gel electrophoresis (7 M urea, 5% 29:1 acrylamide:bisacrylamide, 14 W, 2 hours), excised from the gel, recovered by phenol/chloroform extraction and ethanol precipitation. The purified RNA (25 µg) was eluted in water and stored at -80°C

10 **Example 5**

Structure selective RNA modification

1.5 µg RNaseP RNA in 16.1 µl water were heated to 95°C for 2 min, and placed on ice for 1 min. 8.2 µl 3.3x folding buffer (333 mM HEPES pH 8.0, 333 mM NaCl) was added, and the RNA solution was incubated 10 min at 37°C. 2.7 µl prewarmed (37°C) 10 x Mg₂ (100 mM MgCl₂) was added, and the RNA solution was incubated 10 min at 37°C. The RNA solution was treated with NPfA (3 µl, 60 mM in anhydrous DMSO) and incubated 25 min at 37°C. For comparison, RNA was treated with NMfA (3 µl, 60 mM in anhydrous DMSO), and incubated 25 min at 37°C. A control reaction was prepared by treating RNA with DMSO (3 µl). The modified RNA was recovered by ethanol precipitation (69 µl RNase-free water, 1 µl 20 mg/ml glycogen, 10 µl 3 M NaOAc, 300 µl ethanol, 2 hours at -20°C). After centrifugation at 12,000 g for 30 min at 4°C, the pellet was washed with 1 ml 70% ethanol, centrifuged 3 min at 12,000 g, and dissolved in 5 µl water.

25 **Example 6**

Primer extension

2.5 µl of a 100 µM DNA primer (5'-AGACGTGTGCTCTTCCGATCTGAACCGGACCGAAGCCCG-3'; SEQ ID NO: 10) was added to 5 µl RNA from the previous step. For annealing of the primer to the RNA, the sample was heated to 65°C for 5 min, 37°C for 1 min and then placed on ice. 30 µl Enzyme mix (7.5 µl 5 x Reverse transcription buffer (250 mM HEPES pH 8.3, 375 mM KCl, 15 mM MgCl₂), 7.5 µl 2.5 mM dNTP mix, 7.5 µl 3.3 M/0.6 M Sorbitol/Trehalose mix, 2.5 µl PrimeScript Reverse Transcriptase (TAKARA), 5 µl water) were added. After mixing, the sample was incubated at 45°C for 5 min, 50°C for 25 min, 60°C for 10 min, and kept on ice before purification. The cDNA/RNA hybrids were purified using

Agencourt RNAClean XP kit (cDNA/RNA:beads ratio 1:1.8, washed twice with 70% ethanol, eluted in 40 µl water).

Example 7

5 *Biotinylation*

4 µl of 1 M Na-Citrate (pH 6.0), and 13.5 µl 15 mM biotin (Long arm) hydrazide were added to the 40 µl cDNA/RNA sample. The sample was mixed by pipetting, and incubated at 23°C for 15 hours in the dark.

10 **Example 8**

RNase I treatment

After biotinylation, 6 µl of Tris-HCl (pH 8.5), 1 µl of 0.5 mM EDTA (pH 5.0), and 5 µl of RNase I were added to the sample. It was then mixed by pipetting and incubated at 37°C for 30 min. The enzyme was heat-inactivated (65°C for 5 min). The cDNA/RNA
15 hybrids were then purified by ethanol precipitation (29.5 µl RNase-free water, 1 µl 20 mg/ml glycogen, 10 µl 3 M NaOAc, 300 µl ethanol, 1 hour at -80°C). After centrifugation (12,000 g, 30 min, 4°C), the pellet was washed in 1 ml 70% ethanol, briefly centrifuged (12,000 g, 3 min, 4°C), and resuspended in 40 µl water.

20 **Example 9**

Selection of full length cDNA

100 µl MPG Streptavidin beads were blocked with 1.5 µl 20 µg/µl *E. Coli* tRNA mix (60 minutes, room temperature, mixing every 10 minutes by pipetting). The beads were then separated from the supernatant on a magnetic stand. The beads were washed
25 twice with 50 µl wash buffer 1 (see Example 1) and resuspended in 80 µl wash buffer 1. 40 µl of cDNA treated with RNase I were added to the beads, and the sample was incubated 30 min at room temperature (vortexing every 5 min). The beads were separated on the magnetic stand for 3 min, and the supernatant was removed. The beads were then extensively washed (wash buffer 1 (one time), wash buffer 2 (one
30 time), wash buffer 3 (two times), wash buffer 4 (two times); see compositions in Example 1) using 150 µl of the buffers in each wash. To fragment the RNA and elute the full length cDNA, 60 µl 50 mM NaOH were added to the beads. The beads were then incubated 10 min at room temperature, with mixing every 2 min. The beads were separated on a magnetic stand, and the supernatant was transferred to new tubes and
35 kept on ice. 12 µl of Tris-HCl (pH 7) were added to neutralize the solution. The cDNA

was then ethanol-precipitated ((27 µl RNase-free water, 1 µl 20 mg/ml glycogen, 10 µl 3 M NaOAc, 300 µl ethanol, 2 hours at -80°C). After centrifugation (12,000 g, 30 min, 4°C), the pellet was washed in 1 ml 70% ethanol, briefly centrifuged (12,000 g, 3 min, 4°C), and resuspended in 8 µl water.

5

Example 10*cDNA ligation*

Selected and purified cDNA solution was diluted to the concentration 0.66 ng/µl. 3 µl of the cDNA solution were mixed with 7 µl of ligation mix. The mixture was incubated for 2 hours at 60°C, then 1 hour at 68°C and 10 minutes at 80°C. After the incubation 340 µl H₂O, 1 µl 5mg/ml glycogen, 25 µl 7.5M NH₄Cl and 1 ml of absolute ethanol were added followed by overnight incubation at -20°C followed by 30 min of centrifugation at 14000 g. The supernatant was carefully discarded and the pellet air-dried for 10 minutes and dissolved in 30 µl H₂O.

15

Example 11*PCR and sequencing*

5 µl of the dissolved cDNA ligation pellet was mixed with the 45 µl of the PCR mix with primer 2 with the specific barcode (different barcode for each reaction). Reactions underwent thermal cycling (Biorad C1000) with the following program: One time (98°C for 3 min), four times (98°C for 80 sec; 64°C for 15 sec; 72°C for 30 sec), fifteen times (98°C for 80 sec; 72°C for 45 sec), one time (72°C for 5 min). The generated PCR amplicons were purified with Ampure XP beads using the ratio 1:1.8, washed twice with 70% ethanol solution and eluted by 10 minutes incubation with 20 µl 10 mM Tris-HCl pH 8.3 and binding the beads on the magnet. The purified samples were analyzed on the DNA 1000 chip to estimate the concentrations, mixed in the desired ratio and sequenced on the Illumina HiSeq 2000 instrument according to the genomic DNA sequencing protocol with single end 50 bp long reads including index reading.

25

Example 12*Data analysis*

FASTQ files obtained from the Illumina sequencing (split by the barcodes) were processed first with the "cutadapt v1.0" (<http://code.google.com/p/cutadapt/>) program to remove the sequences matching overhang from the primer introduced in the reverse transcription step (options "-a AGATCGGAAGAGCACACGTCT -m 15 -O 10"; SEQ ID

35

NO: 11). Trimmed reads were mapped to the sequence of the assayed RNA fragment using “bowtie v0.12.7” (Langmead et al., 2009) with default options and choosing SAM format for the output. It is known that the reverse transcriptase terminal transferase activity may lead to addition of untemplated nucleotides to the 3'-end of cDNA molecules – those nucleotides would be detected as first sequenced positions. We wrote the AWK script that removed untemplated nucleotides at the beginning of the reads based on the information in the MD field of the SAM file (SAM format specification available under <http://samtools.sourceforge.net>). Briefly, the sequenced nucleotide (within first three sequenced positions) not matching the template lead to trimming of all the positions up to and including the mismatched nucleotide (trimming of 1, 2 or 3 positions or no trimming if all three nucleotides matched; between 87% to 93% (different barcodes) of the mapped reads didn't require trimming). Following trimming for each location in the sequence of RNA fragment the number of reads for 5' (left) ends (corresponding to cDNA 3'-end) was summed and reported. Plotting and further data analysis was carried out in R. For correlation the Pearson correlation was used and ROC curves and AUC values were calculated with homemade scripts. The basepair annotation of the specificity domain of RNaseP used in the analysis was obtained from the NDB database entry UR0027. The data is based on the B assembly of the published crystal structure (Krasilnikov et al., 2003).

Example 13

Validation of NPIA as a probing reagent (SHAPES)

To validate the SHAPE plus selection (SHAPES) approach, we obtained N-propanone-isatoic anhydride (NPIA), which is identical to the canonical SHAPE reagent N-methyl-isatoic anhydride (NMIA) except that the methyl group in the 1-N position of the isatoic anhydride is replaced by a propanone group. To test that the propanone group in NPIA would allow coupling of the reagent to solid support without changing the ability to react with the RNA 2' OH in a structure dependent fashion (figure 4), we reacted NPIA with gamma P-32 ATP and subsequently with biotin-hydrazide in the presence or absence of mPDA, which has been described to catalyse the ketone-hydrazide reaction. Just like NMIA (Merino et al., 2005), NPIA reacted with the 2' and 3' OH of ATP, causing a double band on the gel (Figure 11A). Subsequent reaction with biotin hydrazide (BH) caused an additional gel shift, which was not observed in the control, indicating that BH did indeed react with propane group of NPIA. Under the reaction circumstances used

here (overnight incubation with BH), the addition of m-phenylenediamine (mPDA) to the reaction did not affect the amount of biotin label added to the ATP.

SHAPE reagents based on isatoic anhydrides undergo spontaneous hydrolysis in aqueous solution to give fluorescent aminobenzoates. We determined the rate of hydrolysis for NPIA and compared it to the well characterised SHAPE molecule NMIA (figure 12). We found that the hydrolysis rate of NPIA is slightly faster than NMIA, requiring around 30 min to react to completion.

Next, we wanted to validate that NPIA would indeed work as a SHAPE plus Selection (SHAPES) reagent that decreases the background in RNA structure probing experiments. The crystal structure of the specificity domain of *Bacillus Subtilis* RNaseP has been determined and showed that this RNA folds into a very well defined and compact structure (Krasilnikov et al., 2003). Moreover, the specificity domain of *Bacillus Subtilis* has previously been probed with SHAPE (Mortimer and Weeks, 2007). We therefore chose this well characterised structural RNA for our experiments. The RNA was in vitro transcribed, refolded and probed with NPIA, NMIA or DMSO control. After probing, the samples were reverse-transcribed and part of the samples was subjected to selection as outlined in figure 13. For all samples, adaptors were ligated to the 3'-ends of cDNA molecules and after PCR amplification the libraries were sequenced on the Illumina platform. The resulting reads were mapped back to the RNaseP specificity domain sequence and aggregated for each position (figure 14). The data from the control DMSO experiment (top barplot), shows that the different positions in the RNaseP specificity domain vary greatly in their propensity to terminate reverse transcriptase. By comparison with the data from the NMIA and NPIA probed sample (second and third barplots), it is clear that these datasets both contain abundant background signal caused by premature termination of the Reverse transcriptase. The Pearson correlation between the DMSO and the NMIA and NPIA samples are 0.90 and 0.85, respectively. Also worth noting is the very strong correlation ($R = 0.98$) between the data from the NMIA and NPIA samples, suggesting that the reaction of these two reagents with RNA is virtually indistinguishable. The next two barplots (fourth and fifth) shows the data from the samples that have been selected as outlined in figure 13. For the NMIA reacted sample, the selection still correlates strongly with the background sample ($R = 0.72$). The reduction in correlation is most likely caused by a reduction in the starting material for building the sequencing library that is left after selection. In

contrast, the data from the NPIA selection sample shows little correlation with the background ($R = 0.05$) seems to enrich for positions known to be unpaired in the crystal structure (black bars in the bottom barplot).

5 To further validate that SHAPES reagents reduce background and enrich for structure signal, we plotted the data from the different experiments as a receiver operating characteristic (ROC) curves using the base pairing from the crystal structure as the binary classifier (figure 15). The area under the ROC curve (AUC) is equal to the probability that a data point randomly chosen among the positions that are unpaired in
10 the crystal structure will have a higher count in experimental data than a data point randomly chosen among the positions that are paired in the crystal structure. As expected there is virtually no structure signal present in the data from the DMSO experiment (AUC of 0.55), whereas both NMIA and NPIA treatment increase the structure signal in the data to AUCs of 0.67 and 0.64, respectively. For the selected
15 samples the AUC values increased 0.02 for the NMIA experiment and 0.09 for NPIA experiment, demonstrating that the NPIA selection reduces background and enriches for the structure signal in the data.

Conclusion

20 We demonstrate that NPIA works as a SHAPES reagent, which just like canonical SHAPE reagents allows structure-specific reaction with the 2'-OH of the RNA ribose, but at the same time facilitates coupling to solid support via the propanone group. Moreover, we show that combination of these two features allows full length cDNAs that reach from the priming position to the probing position to be specifically selected
25 among the background. We expect that SHAPES reagents will significantly improve the quality of RNA structure probing data and facilitate the use of massive parallel sequencing technology for RNA structure probing.

Examples 14 to 32 describe experiments showing that N-(+)-Biotinyl-4-aminobenzoic
30 imidazole (BABAI) can be used as a probing agent.

Examples 14 to 20 described experiments performed on zebrafish embryo RNA.

Example 14

35 *Materials and methods*

N-(+)-Biotinyl-4-aminobenzoic imidazole

1,1'-carbonyldiimidazole

Pronase

DMSO anhydrous

5 Beta-mercaptoethanol

TRI reagent (Sigma-Aldrich): this reagent allows for simultaneous isolation of DNA, RNA and protein

Chloroform

Liquid nitrogen

10 Isopropanol

70% ethanol

Agarose

1 x TBE (Tris-Borate-EDTA)

Chemiluminescent Nucleic Acid Detection Module (Thermo Scientific)

15 X-ray film

Example 15

Synthesis of N-(+)-biotinyl-4-aminobenzoic imidazole (BABAI)

20 Equimolar quantities of N-(+)-Biotinyl-4-aminobenzoic acid (Sigma-Aldrich) and 1,1'-carbonyldiimidazole (Sigma-Aldrich) were mixed in a 1.5 ml Eppendorf tube, and enough DMSO was added to obtain a 1 M N-(+)-biotinyl-4-aminobenzoic imidazole (BABAI) stock. The solution was incubated 2 hours at room temperature, to ensure completion of the reaction.

25 BABAI is the imidazolide of N-Biotinyl-4-aminobenzoic acid and can be synthesized through the reaction of equimolar amounts of the commercially available carbonyldiimidazole and N-Biotinyl-4-aminobenzoic acid in DMSO to give BABAI plus imidazole and CO₂ (Figure 12). The resulting solution can be used directly for probing experiments.

30

Example 16

Probing zebrafish embryo RNA with N-(+)-biotinyl-4-aminobenzoic imidazole (BABAI)

500 healthy zebrafish embryos (mainly shield stage) were collected, and half of these were treated with 1 mg/ml pronase to remove the chorion. The embryos were treated
35 with DMSO (control) or DMSO containing a 5 mM mixture of N-(+)-biotinyl-4-

aminobenzoic acid imidazolidine and imidazole for 15 min at 28°C. They were then washed twice with water (removing as much water as possible, but ensuring that the embryos were covered with water at all times) and washed once with water containing 0.7 M beta-mercaptoethanol. After washing, the embryos were kept in 50 µl water, and a reagent for simultaneous isolation of DNA, RNA and protein, such as 250 µl TRI reagent (Sigma-Aldrich), was added. The embryos were then grinded with a pestle before repeating the addition of the reagent for simultaneous isolation of DNA, RNA and protein (e.g. an additional 750 µl TRI reagent was added). To check that RNA probing did not occur during RNA extraction, a control in which N-(+)-Biotinyl-4-aminobenzoic imidazole was incubated in water for 15 min and then treated with 0.7 M beta-mercaptoethanol was included. This solution was added to embryos in the reagent for simultaneous isolation of DNA, RNA and protein (e.g. 250 µl TRI reagent). All samples were flash frozen in liquid nitrogen and stored at -80°C until RNA extraction.

Example 17

RNA extraction

The samples were thawed 5 min at room temperature, and then spun at 12000 g for 20 min at 4°C. The supernatant was transferred to a new tube and 0.2 volume chloroform was added. After mixing by inversion, the solution was spun at 12000 for 30 min at 4°C. The supernatant was transferred to a new tube and 0.5 ml isopropanol was added. After incubation 10 min at room temperature, the RNA was recovered by centrifugation at 12,000 g for 30 min at 4°C. The pellet was washed with 1 ml 70% ethanol, centrifuged 3 min at 12,000 g, and dissolved in 15 µl water.

Example 18

Gel electrophoresis

The concentration of RNA was measured with Nanodrop, and 400 ng of each sample were run on a native agarose gel (such as 1% agarose, 1 x TBE). The samples were run in duplicates on the same gel. A positive control containing N-(+)-biotinyl-4-aminobenzoic imidazole in vitro probed RNA from HeLa cells was included. The gel was run 50 min at 150 V and after electrophoresis the gel was cut in the middle to separate the two sets of samples. One set was visualized with ethidium bromide to check RNA quality (figure 13A). The RNA from the other half of the gel was blotted on a hybond-N-membrane (Amersham) in a semidry blotting apparatus at 3 mA/cm² for 1

hour (figure 13B). After transfer, the RNA was crosslinked to the membrane using the autocrosslink program (254 nm, 120 mJ/cm²) on a crosslinker such as a stratalinker. The blot was wrapped in vita wrap and stored in a cassette before detection of biotinylated RNA.

5

Example 19*Detection of biotinylated RNA*

Biotinylated RNA was detected with the Chemiluminescent Nucleic Acid Detection Module (Thermo Scientific), according to manufacturer's instructions (figure 14). The chemiluminescent signal was detected with an X-ray film (Agfa).

10

Example 20*BABAI can cross the chorion and cellular membrane to react with RNA*

To demonstrate that BABAI enters cells and reacts with RNA, we used zebrafish embryos (mainly shield stage) and embryos that were treated with pronase to remove the chorion (eggshell). DMSO or DMSO containing a BABAI/imidazole mixture was added to embryos for 15 min at 28°C. After washing with water and with 0.7 M beta-mercaptoethanol to quench remaining reagents, total RNA was purified. To check that RNA probing did not occur during RNA extraction, we performed a control where BABAI was incubated 15 min in water, treated with 0.7 M beta-mercaptoethanol and added to embryos at the very beginning of RNA purification.

15

20

The resulting RNA was separated on an agarose gel (figure 13A) and blotted to a membrane before chemiluminescent detection of biotin, using a streptavidin coupled horseradish peroxidase (HRP) (Figure13B).

25

For the embryos without chorion, we find that BABAI very efficiently labels the RNA with biotin demonstrating that BABAI is able to cross the embryo membrane and react with RNA. The labelling efficiency is reduced for embryos with an intact chorion, but still robust labelling is observed (figure 13B). Importantly, no labelling is observed in the no-reagent and reagent-during-purification controls.

30

Examples 21 to 32 describe experiments performed on HeLa cells.

Example 21

35

Materials and methods

Materials

PBS

Isopropanol

5 Beta-mercaptoethanol

3.3 x Folding buffer: 333 mM HEPES pH 8.0, 333 mM NaCl

10 x Mg²⁺: 100 mM MgCl₂

DMSO anhydrous

10 10 x N-(+)-biotinyl-4-aminobenzoic imidazole (BABAI) in DMSO: The concentration of
BABAI can vary, depending on the preferred extent of RNA modification. Here a
10xBABAI solution of 50 mM was used.

10 x N-methylisatoic anhydride (NMIA) in DMSO: The concentration of NMIA can vary,
depending on the preferred extent of RNA modification. Here a 10xNMIA solution of 50
mM was used.

15 DNA primer for reverse transcription: RT_15xN:

AGACGTGTGCTCTTCCGATCTNNNNNNNNNNNNNNNN (SEQ ID NO: 12)

PrimeScript Reverse Transcriptase (TAKARA)

5 x Reverse Transcriptase buffer: 250 mM HEPES pH 8.3, 375 mM KCl, 15 mM MgCl₂

3.3 M/0.6 M Sorbitol/Trehalose mix

20 10 mM dNTPs

Agencourt RNAClean XP Kit (BECKMAN COULTER)

1 M Na-Citrate pH 6.0

Biotin (long arm) hydrazide (VECTOR lab)

1 M Tris-HCl pH 7.0

25 1 M Tris-HCl pH 8.5

0.5 M EDTA pH 8.0

RNase I ribonuclease

MPG Streptavidin (TAKARA)

30 20 µg/µl E. Coli tRNA mix: 30 mg *E. Coli* tRNA (ribonucleic acid, transfer from
Escherichia Coli, lyophilized powder (Sigma-Aldrich) was dissolved in 400 µl water, and
45 µl 10 x RQ1 DNase buffer and 30 µl RQ1 RNase-Free DNase were added. The
sample was incubated at 37°C for 2 h. After incubation, 10 µl of 0.5 M EDTA (pH 8.0),
10 µl of 10% SDS, and 10 µl of 10 ng/ml Proteinase K were added to the tRNA
35 25 solution, and the sample was incubated at 45°C for 30 min. 500 µl of phenol–

chloroform was added, and the sample was centrifuged at 20,000 g for 5 min at room temperature. The aqueous phase was collected and 1 volume of chloroform was added. The sample was then centrifuged at 20,000 g for 5 min at room temperature. The supernatant was collected. 1/10 volume of 3 M NaCl, 1 µl 5 mg/ml glycogen, and 1
5 volume isopropanol were added, and the sample was centrifuged once again at 20,000 g for 5 min at room temperature. The supernatant was removed and 900 µl of 80% ethanol was added to the tRNA pellet. It was centrifuged at 20,000 g for 5 min at room temperature, the supernatant was removed, and another 900 µl of 80% ethanol was added to the tRNA pellet. After centrifugation at 20,000 g for 5 min at room temperature
10 and removing of the supernatant, the tRNA pellet was dissolved in 20 µl water. The tRNA concentration was measured using a Nanodrop 1000 (Thermo scientific). The tRNA was then diluted to 20 µg/µl in water and stored at -20°C.

Wash buffer 1: 4.5 M NaCl, 50 mM EDTA pH 8.0
15 Wash buffer 2: 300 mM NaCl, 1 mM EDTA pH 8.0
Wash buffer 3: 20 mM Tris-HCl pH 8.5, 1 mM EDTA pH 8.0, 500 mM NaOAc pH 6.1, 0.4% SDS
Wash buffer 4: 10 mM Tris-HCl pH 8.5, 1 mM EDTA pH 8.0, 500 mM NaOAc pH 6.1 50 mM NaOH
20 Ligation adapter: 5' PHO-
NNNNNNNAGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT-3NHC₃ (SEQ ID NO: 13), where PHO is a 5'-phosphate modification and 3NHC₃ is a 3'-amino modification.

The cDNA ligation mix was scaled up to the required volume mixture made of 1 volume
25 of 10x ligation buffer (0.5 M MOPS (pH 7.5), 0.1 M KCl, 50 mM MgCl₂, and 10 mM DTT), 0.5 volume of 1 mM ATP, half of the volume of 50 mM MnCl₂, two volumes of 50% PEG6000, two volumes of 5M betaine, 0.5 volume of the 100 µM ligation adapter, 15 0.5 volume of CircLigase enzyme (100U/µl, Epicentre). The ligation mix was prepared right before the reaction and stored on ice. In this example we used
30 CircLigase enzyme to ligate the adapter to the 3'-end of cDNA molecule. Other methods of ligation, such as single-strand linker ligation (Takahashi H, Lassmann T, Murata M, Carninci P. 2012. Nature protocols 7: 542-61), can also be used.

Salt solution used for nucleic acid precipitation, such as 7.5M NH₄Ac (ammonium
35 Acetate)

5mg/ml glycogen

Ethanol

Phusion DNA polymerase (Finnzymes)

5x HF buffer (Finnzymes)

5 Primer 1 (SEQ ID NO: 2): (5' –

AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCT – 3')

Primer 2: (5'-

AAGCAGAAGACGGCATACGAGATXXXXXXGTGACTGGAGTTCAGACGTGTGCTC

TTCCGATCT-3', where XXXXXX denotes the reaction-specific barcodes (SEQ ID NO:

10 3, SEQ ID NO: 4, SEQ ID NO: 5, SEQ ID NO: 6, SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 9)

Overhangs match the Solexa Illumina sequencing platform.

15 The PCR mix was scaled up to the required volume mixture made of 3 volumes of 10 uM primer 1, 10 volumes of 5X HF buffer (Finnzymes), 1 volume of 10 mM dNTPs, 27.5 volumes of H₂O, 1 volume of Phusion DNA polymerase (Finnzymes) and 2.5 volumes of 10 uM primer 2. The PCR mix was prepared right before the reaction and stored on ice.

20 Ampure XP beads (Beckman Coulter)
Bioanalyzer system with DNA 1000 chip (Agilent)

Example 22

Cell culturing

25 HeLa cells were grown at 37°C with 5% CO₂ in culture flasks with a surface area of 175cm², to a confluence of 90-100%. Cells from 1/3rd of a flask were used per in vitro probed sample, and cells from 1 flask were used per in vivo probing sample.
RNA structure probing in vivo.

30 Before in vivo RNA structure probing, HeLa cells were detached by scraping and washed three times in PBS (prewarmed to 37°C). Between each wash, the cells were spun at 700 rpm for 5 min before PBS removal. After washing, the cells were resuspended in 1350 µl PBS, and 150 µl DMSO (control) or 150 µl 50 mM N-(+)-biotinyl-4-aminobenzoic imidazole (BABAI) was added. The cells were then incubated
35 for 30 min at 37°C. For investigating RNA structure during heat-shock response, HeLa

cells were incubated 1 hour at 42°C before structure probing with N-(+)-biotinyl-4-aminobenzoic imidazole for 30 min at 42°C. After structure probing, cells were washed three times with PBS with 0.7 M beta-mercaptoethanol and resuspended in 400 µl PBS with 0.7 M beta-mercaptoethanol.

5

Example 23*RNA structure probing in vitro*

Total RNA was isolated from HeLa cells as described in 'RNA extraction from HeLa cells', but without adding the E. coli spike-in. Total RNA or poly(A) purified RNA in 16.1 µl water were heated to 95°C for 2 min, and placed on ice for 1 min. 8.2 µl 3.3x folding buffer (333 mM HEPES pH 8.0, 333 mM NaCl) was added, and the RNA solution was incubated 10 min at 37°C. 2.7 µl prewarmed (37°C) 10 x Mg²⁺ (100 mM MgCl₂) was added, and the RNA solution was incubated 10 min at 37°C. The RNA solution was treated with BABAI (3 µl, 50 mM in anhydrous DMSO) and incubated 25 min at 37°C. For comparison, RNA was treated with NMIA (3 µl, 50 mM in anhydrous DMSO), and incubated 25 min at 37°C. A control reaction was prepared by treating RNA with DMSO (3 µl). After adding beta-mercaptoethanol to 0.7 M and incubating for 2 min, E. coli total RNA was spiked in. The modified RNA was recovered by ethanol precipitation (69 µl RNase-free water, 1 µl 20 mg/ml glycogen, 10 µl 3 M NaOAc, 300 µl ethanol, 2 hours at -20°C). After centrifugation at 12,000 g for 30 min at 4°C, the pellet was washed with 1 ml 70% ethanol, centrifuged 3 min at 12,000 g, and dissolved in 5 µl water.

Example 24*RNA extraction from HeLa cells*

900 µl TRI reagent and 10 µl 150 ng/µl E. Coli total RNA (in vivo probed samples) (strain MRE600, a gift from Birte Vester) were added to the cells, and solution was incubated 5 min at room temperature. After vortexing, 0.2 volume chloroform was added. The solution was spun at 12000 for 30 min at 4°C, and the supernatant was transferred to a new tube. To precipitate the RNA, 1 µl 5 mg/ml glycogen and 600 µl isopropanol was added. After incubation 10 min at room temperature, the RNA was recovered by centrifugation at 12,000 g for 30 min at 4°C. The pellet was washed with 1 ml 70% ethanol, centrifuged 3 min at 12,000 g, air-dried for 5 min, and dissolved in 60 µl water.

Example 25

35

PolyA enrichment

PolyA RNA was enriched with a suitable kit, such as the Poly(A) Purist Mag Kit (Ambion), following manufacturer's instructions. Poly(A) purification was performed twice on each sample, and purified poly(A) was eluted in 11 µl water before primer extension.

Primer extension

To estimate the efficiency of selection, *B. subtilis* RNase P RNA was spiked in before primer extension. 2.5 µl of a 100 µM DNA primer (5'-

AGACGTGTGCTCTTCCGATCTNNNNNNNNNNNNNNNN-3'; SEQ ID NO: 12) was

added to 5 µl RNA from the previous step. For annealing of the primer to the RNA, the sample was heated to 65°C for 5 min, 37°C for 1 min and then placed on ice. 30 µl Enzyme mix (7.5 µl 5 x Reverse transcription buffer (250 mM HEPES pH 8.3, 375 mM KCl, 15 mM MgCl₂), 7.5 µl 2.5 mM dNTP mix, 7.5 µl 3.3 M/0.6 M Sorbitol/Trehalose mix, 2.5 µl PrimeScript Reverse Transcriptase (TAKARA), 5 µl water) were added.

After mixing, the sample was incubated at 45°C for 5 min, 50°C for 25 min, 60°C for 10 min, and kept on ice before purification. The cDNA/RNA hybrids were purified using a suitable kit, such as Agencourt RNAClean XP kit (cDNA/RNA:beads ratio 1:1.8, washed twice with 70% ethanol, eluted in 40 µl water).

Example 26*Biotinylation*

4 µl of 1 M Na-Citrate (pH 6.0), and 13.5 µl 15 mM biotin (Long arm) hydrazide were added to the 40 µl cDNA/RNA sample. The sample was mixed by pipetting, and incubated at 23°C for 15 hours in the dark.

Example 27*RNase I treatment*

After biotinylation, 6 µl of Tris-HCl (pH 8.5), 1 µl of 0.5 mM EDTA (pH 5.0), and 5 µl of RNase I were added to the sample. It was then mixed by pipetting and incubated at 37°C for 30 min. The enzyme was heat-inactivated (65°C for 5 min). The cDNA/RNA hybrids were then purified by ethanol precipitation (29.5 µl RNase-free water, 1 µl 20 mg/ml glycogen, 10 µl 3 M NaOAc, 300 µl ethanol, 1 hour at -80°C). After centrifugation (12,000 g, 30 min, 4°C), the pellet was washed in 1 ml 70% ethanol, briefly centrifuged (12,000 g, 3 min, 4°C), and resuspended in 40 µl water.

Example 28*Selection of full length cDNA*

100 µl MPG Streptavidin beads were blocked with 1.5 µl 20 µg/µl *E. Coli* tRNA mix (60 minutes, room temperature, mixing every 10 minutes by pipetting). The beads were
5 then separated from the supernatant on a magnetic stand. The beads were washed twice with 50 µl wash buffer 1 (see Example 1) and resuspended in 80 µl wash buffer 1. 40 µl of cDNA treated with RNase I were added to the beads, and the sample was incubated 30 min at room temperature (vortexing every 5 min). The beads were separated on the magnetic stand for 3 min, and the supernatant was removed. The
10 beads were then extensively washed (wash buffer 1 (one time), wash buffer 2 (one time), wash buffer 3 (two times), wash buffer 4 (two times); see compositions in Example 1) using 150 µl of the buffers in each wash. To fragment the RNA and elute the full length cDNA, 60 µl 50 mM NaOH were added to the beads. The beads were then incubated 10 min at room temperature, with mixing every 2 min. The beads were
15 separated on a magnetic stand, and the supernatant was transferred to new tubes and kept on ice. 12 µl of Tris-HCl (pH 7) were added to neutralize the solution. The cDNA was then ethanol-precipitated ((27 µl RNase-free water, 1 µl 20 mg/ml glycogen, 10 µl 3 M NaOAc, 300 µl ethanol, 2 hours at -80°C). After centrifugation (12,000 g, 30 min, 4°C), the pellet was washed in 1 ml 70% ethanol, briefly centrifuged (12,000 g, 3 min, 4°C), and resuspended in 8 µl water.
20

Example 29*cDNA ligation*

Selected and purified cDNA solution was diluted to the concentration 0.66 ng/µl. 3 µl of
25 the cDNA solution were mixed with 7 µl of ligation mix. The mixture was incubated for 2 hours at 60°C, then 1 hour at 68°C and 10 minutes at 80°C. After the incubation 340 µl H₂O, 1 µl 5mg/ml glycogen, 25 µl 7.5M NH₄Cl and 1 ml of absolute ethanol were added followed by overnight incubation at -20°C followed by 30 min of centrifugation at 14000 g. The supernatant was carefully discarded and the pellet air-dried for 10
30 minutes and dissolved in 30 µl H₂O.

Example 30*PCR and sequencing*

5 µl of the dissolved cDNA ligation pellet was mixed with the 45 µl of the PCR mix with
35 primer 2 with the specific barcode (different barcode for each reaction). Reactions

underwent thermal cycling (Biorad C1000) with the following program: One time (98°C for 3 min), four times (98°C for 80 sec; 64°C for 15 sec; 72°C for 30 sec), fifteen times (98°C for 80 sec; 72°C for 45 sec), one time (72°C for 5 min). The generated PCR amplicons were purified using magnetic beads such as Ampure XP beads using the ratio 1:1.8, washed twice with 70% ethanol solution and eluted by 10 minutes incubation with 20 µl 10 mM Tris-HCl pH 8.3 and binding the beads on the magnet. The purified samples were analyzed on the DNA 1000 chip to estimate the concentrations, mixed in the desired ratio and sequenced on the Illumina HiSeq 2000 instrument according to the genomic DNA sequencing protocol with paired-end 100 bp long reads including index reading.

Example 31

Data analysis

FASTQ files obtained from the Illumina sequencing (split by the barcodes) were processed first to remove the sequences matching overhang from primer introduced in the reverse transcription step. This was done with the “cutadapt v1.0” (<http://code.google.com/p/cutadapt/>) program (from read 1 in each pair, options “-a AGATCGGAAGAGCACACGTCT -q 17”; SEQ ID NO 11) and from the read 2 from the forward primer during PCR (from read 2 in each pair, options “-a AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT -q 17; SEQ ID NO: 1). The 7 nucleotide barcode was removed from each read with an awk script. The script also removed reads smaller than 20 nucleotides, after barcode removal. Trimmed reads were mapped to the sequence of the assayed RNA fragment. This was done using “bowtie v2.2.2” (Langmead et al., 2012) with default options and choosing SAM format for the output. The output SAM file was converted to a BAM file using samtools v0.1.17 (options “view -bS), and the percentage of mapped reads were determined with samtools flagstat command as the percentage of concordantly mapped reads.

For *E. coli* mapping of, reads were trimmed as described previously; however, reads shorter than 40 nucleotides were removed. Reads were mapped “bowtie v2.2.2” (Langmead et al., 2012) (options “-N 1 -L 15 --norc -X 700”), and samtools were used to calculate percentage of concordantly mapped reads as described above.

Example 32

BABAI allows selection of mRNA probed in vivo

We wanted to demonstrate that RNA treated with BABAI can be specifically selected and that SHAPES selection of “in vivo” probed RNA removes background signal from SHAPE probing experiments. We treated human HeLa cells with 5 mM of the BABAI/imidazole mixture or with a DMSO control for 30 min at 37°C. To control for probing during RNA purification, *E. coli* total RNA was added just prior to RNA purification. Following total RNA purification, the RNA was polyA enriched. Moreover, in vitro transcribed *B. subtilis* RNase P specificity domain RNA was added to the purified mRNA to control for selection efficiency. The SHAPES probing was performed and sequencing libraries were constructed. After Illumina sequencing the resulting reads were mapped to human mRNAs, *E. coli* genome and the *B. subtilis* RNase P specificity domain sequence.

Selection was performed in replicates and for both samples with selection we observed a strong depletion of the spike-in RNAs with replicate 2 only having about 1% of RNA remaining of the RNase P RNA and no detectable *E. coli* RNA (Figure 14A). This level of selection corresponds to at least a 100 fold enrichment of the probed RNA over the unprobed RNA.

The sequencing depth in this experiment was not high, however among the highly expressed mRNAs in HeLa is the ferritin heavy chain (FTH1), which has a conserved RNA structure in its 5' UTR. The structure is an Iron Response Element (IRE) and it is supported by comparative evidence and experimental validation. The counts from the DMSO and the unselected BABAI sample correlate, but do not support the annotated FTH1 IRE RNA structure. In contrast, for the BABAI selected samples the sequencing counts support the annotated structure, demonstrating that in vivo SHAPES with BABAI select for RNA structure signal (Figure 14B).

In conclusion, we demonstrate that BABAI works as a SHAPES reagent.

References

- Adilakshmi T, Soper SF, Woodson SA, 2009. *Methods Enzymol.*;468:239-58
Krasilnikov AS, Yang X, Pan T, Mondragon A. 2003. *Nature* 421: 760-4
Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. *Genome biology* 10

- Lucks JB, Mortimer SA, Trapnell C, Luo SJ, Aviran S, et al. 2011. *Proc Natl Acad Sci U S A*. 108: 11063-8
- Merino EJ, Wilkinson KA, Coughlan JL, Weeks KM. 2005. *J Am Chem Soc*. 127: 4223-31
- 5 Mortimer SA, Weeks KM. 2007. *J Am Chem Soc* 129: 4144-5
- Rashidian M, Mahmoodi MM, Shah R, Dozier JK, Wagner CR, Distefano MD. 2013. *Bioconjug Chem* 24: 333-42
- Sambrook and Russell. Molecular Cloning, a Laboratory Manual, third edition, *Cold Spring Harbour Laboratory Press*.
- 10 Shcherbakova and Mitra. 2009. *Methods Enzymol*. 468:31-46
- Spitale RC, Crisalli P, Flynn RA, Torre EA, Kool ET, Chang HY. 2013. *Nat Chem Biol*. 9(1):18-20
- Takahaschi ,H Kato S, Murata M, Carninci P. 2012. *Methods Mol Biol*. 786: 181-200
- Wan Y, Kertesz M, Spitale RC, Segal E, Chang HY. 2011. *Nat Rev Genet*. 12: 641-55
- 15 Weeks KM. 2010. *Curr Op Struc Biol* 20: 295-304

Claims

1. A method of analyzing the structure of RNA polynucleotides by allowing for selection of full length cDNA, said method comprising:
- 5 a. covalently binding a probing reagent to said RNA polynucleotide, said probing reagent comprising
- i. a group capable of being coupled to a solid support matrix; or
- ii. a solid support matrix,
- 10 and
- b. adding primers to said RNA, and extending said primers by reverse transcriptase thereby generating a RNA/cDNA duplex,
- wherein said probing reagent is capable of terminating the synthesis of cDNA from said RNA by reverse transcriptase.
- 15
2. The method of claim 1, wherein said method further comprises covalently linking of said group to a solid support.
3. The method of any of the preceding claims, wherein the primers comprise a
- 20 group of random primers.
4. The method of any of the preceding claims, wherein the primers comprise one or more targeted primers.
5. The method of any of the preceding claims, wherein the probing reagent
- 25 comprises an RNA binding protein covalently bound to the RNA.
6. The method of claim 5, wherein the RNA binding protein is bound to a solid support via an antibody specific for said RNA binding protein.
- 30
7. The method of any of the preceding claims, wherein the probing reagent comprises an electrophile that can react selectively with the 2' OH group of the RNA ribose ring.
8. The method of any of the preceding claims, wherein the probing reagent can
- 35 bind covalently to the base part of an RNA monomer.

9. The method of any of the preceding claims, further comprising treating the RNA polynucleotide with a radical to create RNA fragments with ketone or aldehyde groups and wherein the probing reagent is capable of binding to said aldehyde or ketone group.
- 5
10. The method of claim 9, wherein one or more ketone or aldehyde group is in the 5'-end of the RNA fragments.
11. The method of claim 9, wherein aldehyde or ketone groups are reacted with a hydrazide-biotin reagent or an N'-aminooxymethylcarbonylhydrazino D-biotin reagent.
- 10
12. The method of any of the preceding claims, wherein the probing reagent can bind selectively to unpaired RNA.
- 15
13. The method of any of the preceding claims, wherein the probing reagent can react with all four RNA monomers.
14. The method of any of the preceding claims, wherein the electrophile that can react to the 2'OH group of the RNA ribose ring is selected from the group consisting of: acid imidazolides, acid cyanides, acid chlorides, isocyanates, anhydrides, isatoic anhydrides and phthalic anhydrides.
- 20
15. The method of any of the preceding claims, wherein the group capable of being coupled to a solid support matrix is selected from the group consisting of: amine, amide, aldehyde, ketone, ester, ether, hydrocarbon, propanone and biotin.
- 25
16. The method of any of the preceding claims, wherein the solid support matrix has a hydrophilic surface.
- 30
17. The method of any of the preceding claims, wherein the solid support is selected from a bead, a dish, a well, a stick, a resin.
18. The method of claim 17, wherein the solid support is a magnetic bead or a resin.
- 35

19. The method of any of the preceding claims, wherein the solid support matrix is selected from the group consisting of diatomaceous earth, celite, squalane, hexadecane, dialkyl phthalates, tetrachlorophthalates, polyethylene glycol, and polysiloxanes.
- 5 20. The method of any of the preceding claims, wherein the probing reagent comprises: N-propanone-isatoic anhydride (NPIA), 1-propanone-7-nitroisatoic anhydride (1P7), or N-Biotinyl-4-aminobenzoic acid imidazolide (BABAI).
- 10 21. The method of any of the preceding claims, further comprising capturing said solid support to allow removal of duplexes without a probing reagent.
22. The method of any of the preceding claims, further comprising fractionating the RNA molecules using RNase or hydrolysis prior to covalently binding said
- 15 probing reagent.
23. The method of any of the preceding claims, wherein said method further comprises:
- 20 a. treating said bound RNA/cDNA duplex with RNase; and
- b. washing said bound duplex and
- c. isolating said cDNA, and
- d. optionally, digesting said cDNA with at least one restriction enzyme, and
- e. analysing said cDNA by restriction fragment length analysis or
- 25 sequencing.
24. The method of any of the preceding claims, wherein single stranded RNA is removed with RNase specific for single strands.
- 30 25. The method of any of the preceding claims, wherein the RNA strand is removed from the RNA/cDNA duplex by RNase treatment.
26. The method of any of the preceding claims, wherein the RNA strand is removed from the RNA/cDNA duplex by denaturation of said duplex.
- 35 27. The method of any of the preceding claims, wherein said RNA is of eukaryotic or prokaryotic origin.

28. The method of any of the preceding claims, wherein said RNA is of human origin.

5 29. The method of any of the preceding claims, wherein said method is performed on purified RNA.

30. The method of any of the preceding claims, wherein the probing reagent is bound to RNA within a living cell.

10

31. The method of any of the preceding claims, wherein the probing agent is capable of crossing membranes such as cellular membranes, nuclear membranes, and embryonic membranes.

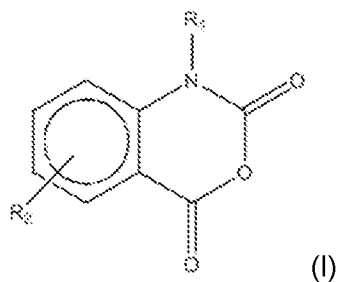
15

32. The method of claim 30, wherein the remaining steps are performed after lysis of said living cell.

20 33. The method of any of the preceding claims, wherein full length cDNA is enriched with a factor of at least 10-fold, such as at least 20-fold, such as at least 30-fold, such as at least 40-fold, such as at least 50-fold, such as at least 60-fold, such as at least 70-fold, such as at least 80-fold, such as at least 90-fold, such as at least 100-fold, compared to a method performed without covalently binding of a probing reagent.

25

34. A compound of formula I



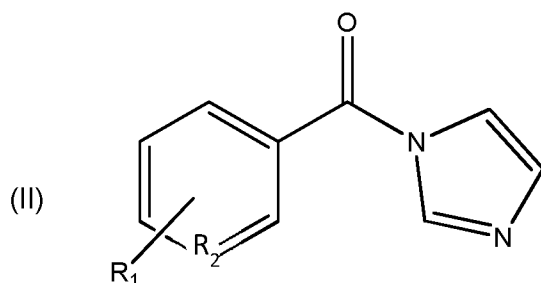
30 wherein R₁ is selected from the group consisting of: H, CH₃, a ketone, an aldehyde, a primary amine, an azide, an alkene, an alkyne, a biotin, a PEG-biotin; and

wherein R_2 is selected from the group consisting of NO_2 , CF_3 , CCl_3 , H, CH_3 , a ketone, an aldehyde, a primary amine, an azide, an alkene, an alkyne, a biotin, a PEG-biotin;

with the proviso that if one of R_1 or R_2 is H, CH_3 , a ketone, an aldehyde, a primary amine, an azide, an alkene, an alkyne, a biotin, or a PEG-biotin, then the other of R_1 or R_2 is NO_2 , CF_3 or CCl_3 .

35. The compound according to claim 34, wherein said compound is 1-propanone-7-nitroisatoic anhydride.

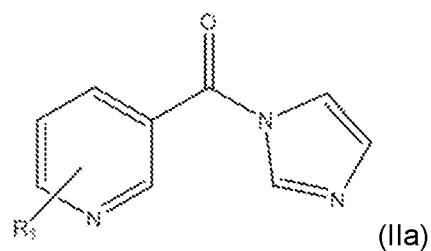
36. A compound of formula II



wherein R_1 is selected from the group consisting of: H, CH_3 , a ketone, an aldehyde, a primary amine, an azide, an alkene, an alkyne, a biotin, a PEG-biotin, and wherein R_2 is C or N.

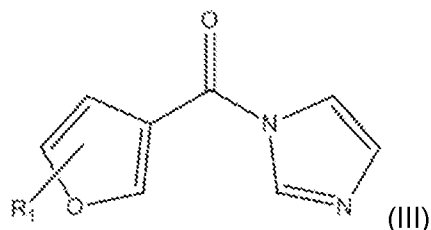
37. The compound of claim 36, wherein R_1 is not in position 2.

38. The compound of any of claims 36 to 37, of formula IIa:



39. The compound of any of claims 37 to 38, wherein said compound is N-Biotinyl-4-aminobenzoic acid imidazolid (BABA).

40. A compound of formula III



wherein R_1 is selected from the group consisting of: H, CH_3 , a ketone, an aldehyde, a primary amine, an azide, an alkene, an alkyne, a biotin, a PEG-biotin.

41. The compound according to claim 40, wherein R_1 is not in position 2.

42. Use of a compound according to any of the preceding claims 34-41, for binding covalently to an RNA molecule or as a probing reagent in the method according to any of claims 1 to 33.

43. Use of N-propanone isatoic anhydride (NPIA) for binding covalently to an RNA molecule or as a probing reagent in the method according to any of claims 1 to 33.

44. Use of N-Biotinyl-4-aminobenzoic acid imidazolid (BABAI) for binding covalently to an RNA molecule or as a probing reagent in the method according to any of claims 1 to 33.

45. Use of 1-propanone-7-nitroisatoic anhydride (1P7) for binding covalently to an RNA molecule or as a probing reagent in the method according to any of claims 1 to 33.

46. A compound comprising:

- a. a cDNA/RNA duplex
- b. a probing reagent, wherein the probing reagent is as defined in any one of claims 5 to 20 or 34 to 41 and comprises:
 - i. a group capable of being coupled to a solid support matrix; or
 - ii. a solid support matrix

said probing reagent being covalently bound to the RNA.

47. A compound comprising:

- a. a cDNA/RNA duplex
- b. a probing reagent, wherein the probing reagent is as defined in any one of claims 7 to 20 or 34 to 41 and comprises:

- 5 i. a group capable of being coupled to a solid support matrix; or
- ii. a solid support matrix

said probing reagent being covalently bound to the RNA.

48. A kit for analysing the structure of RNA polynucleotides by allowing selection of
10 full length cDNA, said kit comprising:

- a. a probing reagent capable of binding specifically with the base part of an RNA monomer or with the 2'OH group on the ribose ring of an RNA monomer, said probing reagent further comprising a solid support,

15 or

- b. a probing reagent capable of binding specifically with the base part of an RNA monomer or with the 2'OH group on the ribose ring of an RNA monomer, a solid support capable of being bound covalently to said probing reagent, and a reagent to enable covalent binding of said solid
20 support to said probing reagent,

and

- a. a mixture of primers
- b. a reverse transcriptase
- c. buffers

25 d. at least two control template RNAs.

49. The kit of claim 48, wherein the at least two control template RNAs are ribosomal RNAs.

30 50. The kit of claim 49, wherein the ribosomal RNAs originates from *Escherichia coli* or *Saccharomyces cerevisiae*.

51. The kit according to any of claims 48 to 50, wherein the probing reagent comprises an electrophile that can react selectively with the 2' OH group of the
35 RNA ribose ring.

52. The kit according to any of claims 48 to 51, wherein the probing reagent can bind selectively to unpaired RNA.
53. The kit according to any of claims 48 to 52, wherein the probing reagent can react with all four RNA monomers.
54. The kit according to any of claims 48 to 53, wherein the electrophile that can react to the 2'OH group of the RNA ribose ring is selected from the group consisting of: acid imidazolides, acid cyanides, acid chlorides, isocyanates, anhydrides, isatoic anhydrides and phthalic anhydrides.
55. The kit according to any of claims 48 to 54, wherein the reactive group capable of being coupled to a solid support matrix is selected from the group consisting of: amine, amide, aldehyde, ketone, ester, ether, hydrocarbon, propanone and biotin.
56. The kit according to any of claims 48 to 55, wherein the solid support matrix has a hydrophilic surface.
57. The kit according to any of claims 48 to 56, wherein the solid support is selected from a bead, a dish, a well, a stick, a resin.
58. The kit according to claim 57, wherein the solid support is a magnetic bead or a resin.
59. The kit according to any of claims 48 to 58, wherein the solid support matrix is selected from the group consisting of: diatomaceous earth, celite, squalane, hexadecane, dialkyl phthalates, tetrachlorophthalates, polyethylene glycol, and polysiloxanes.
60. The kit according to any of claims 48 to 59, wherein the probing reagent comprises: N-propanone-isatoic anhydride (NPIA), 1-propanone-7-nitroisatoic anhydride (1P7), or N-Biotinyl-4-aminobenzoic acid imidazolid (BABAI).

FIG. 1

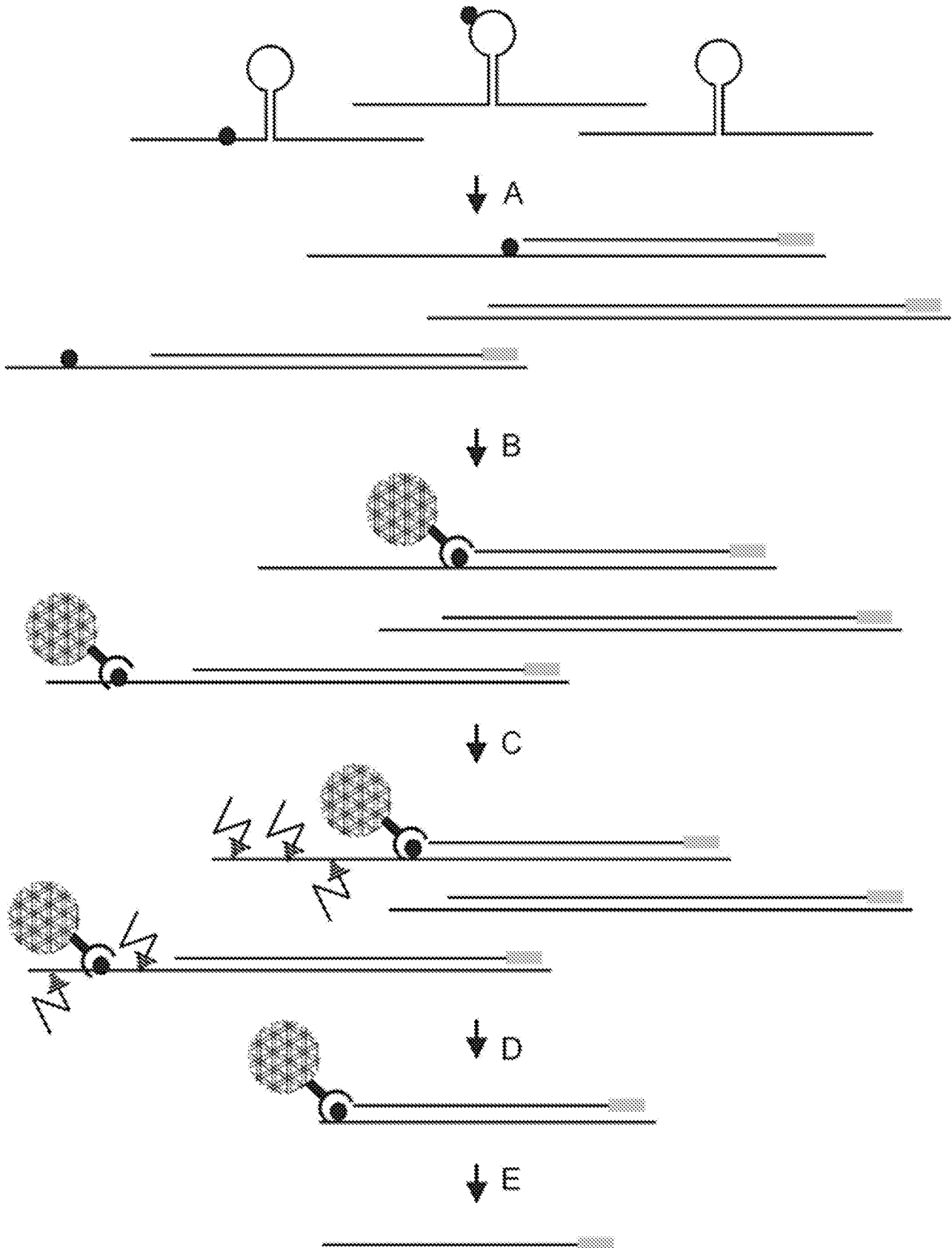


FIG. 2

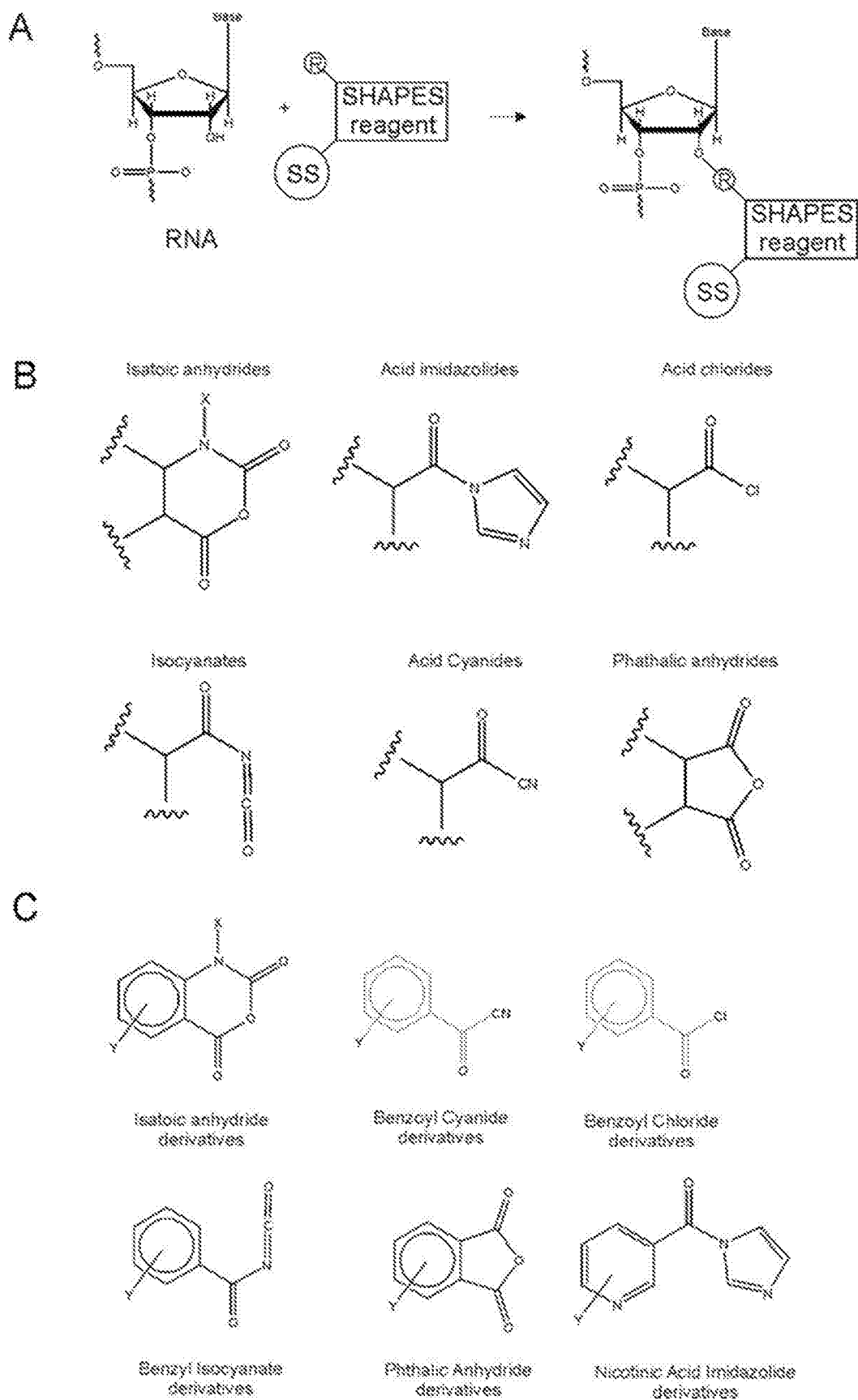
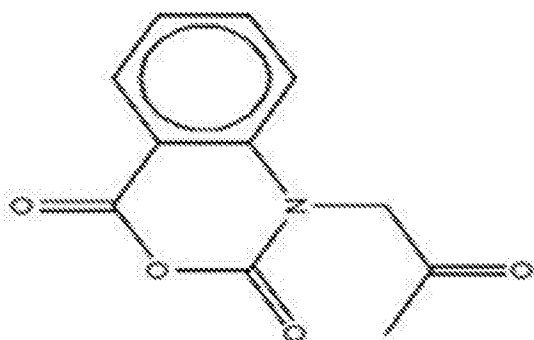
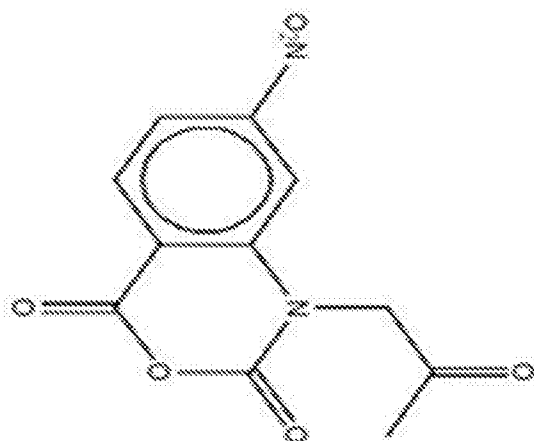


FIG. 3

N-propanoneisatoic anhydride
(NP1A)



1-propanone-7-nitroisatoic anhydride
(1P7)



2-Propanone Nicotinic Acid
Imidazolide (PAI)

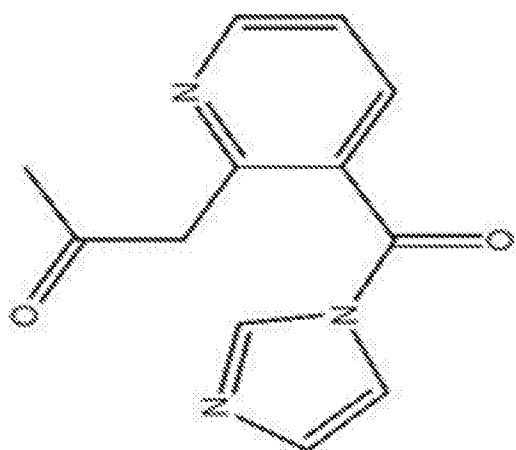


FIG. 4

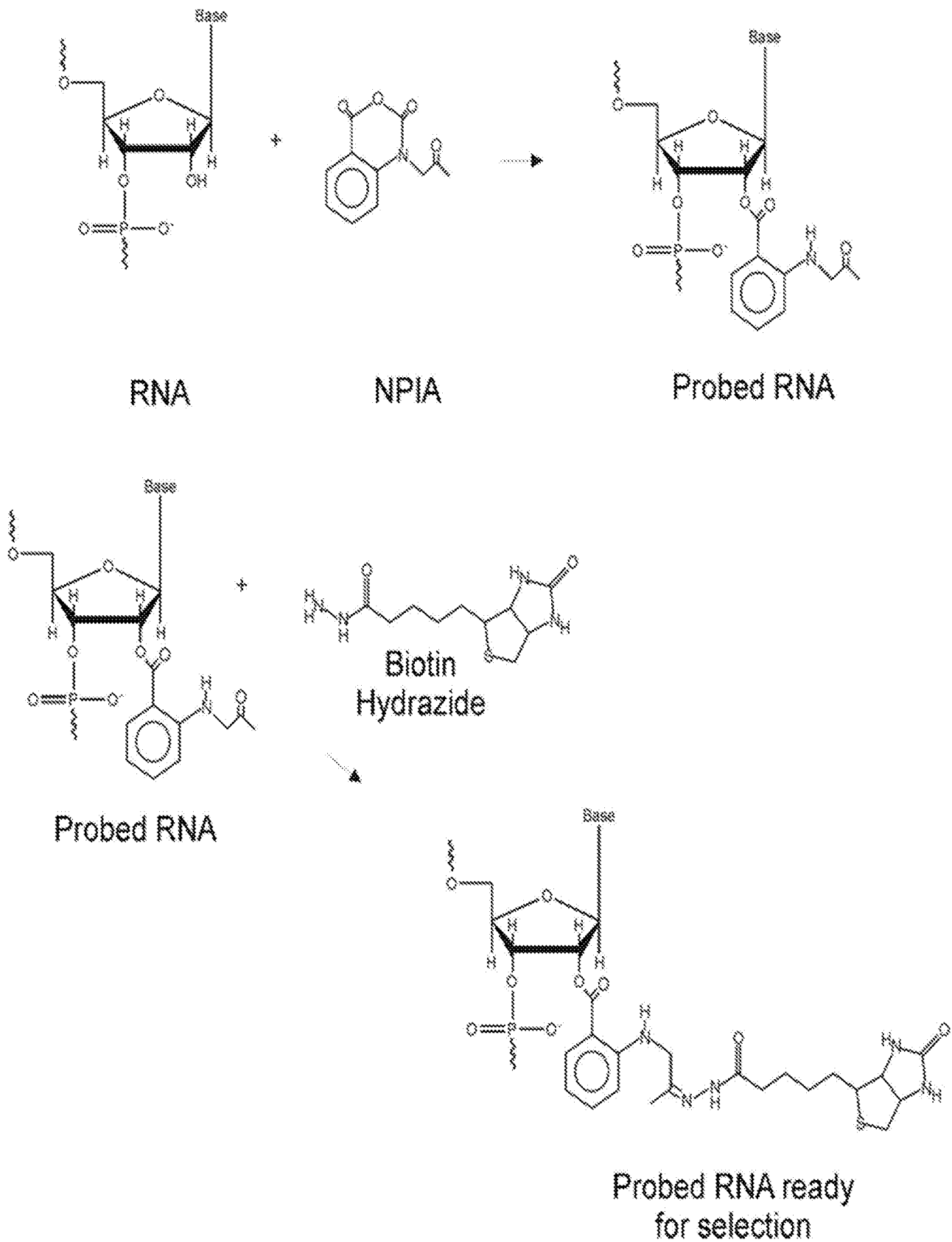


FIG. 5

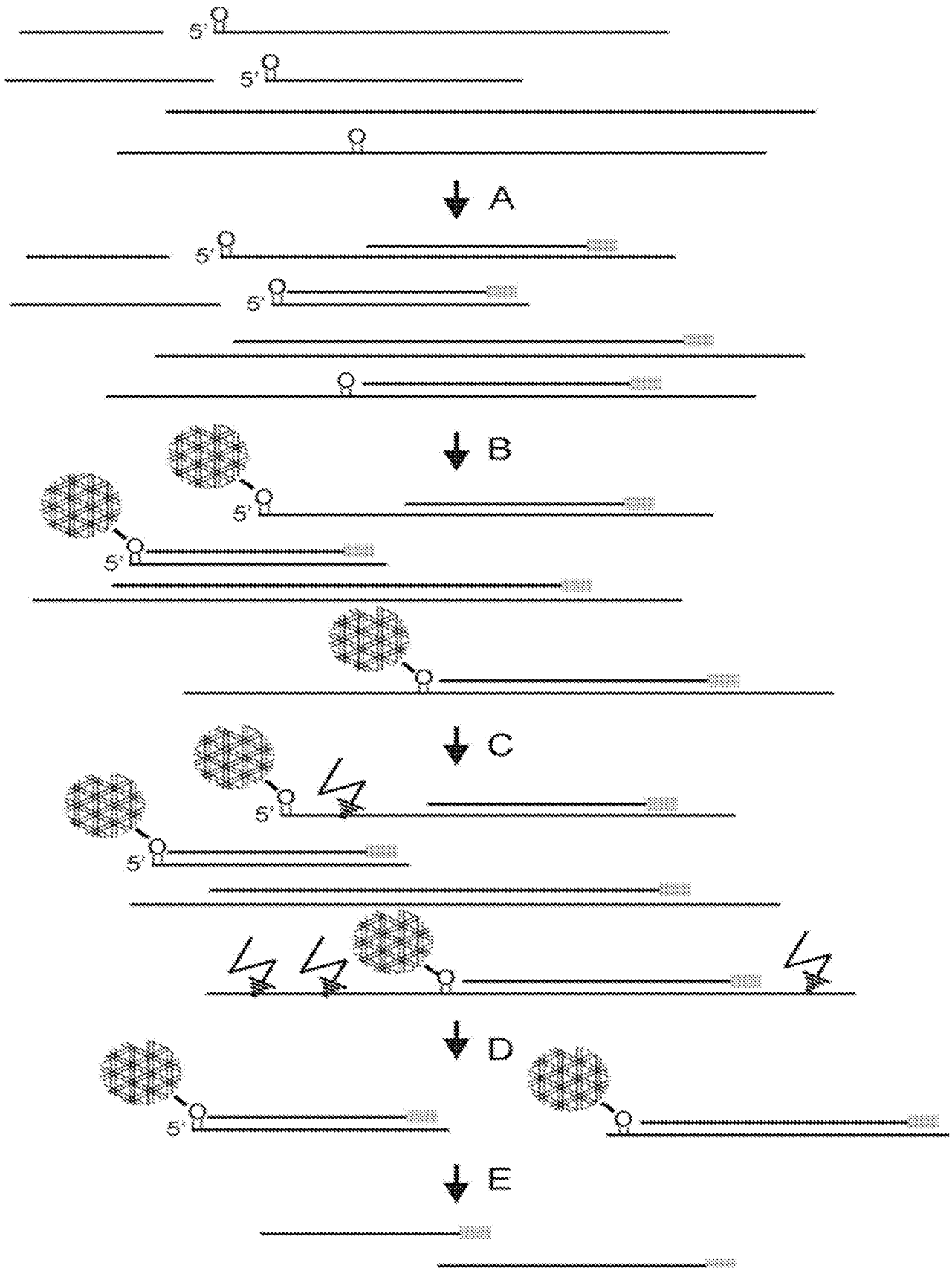


FIG. 6

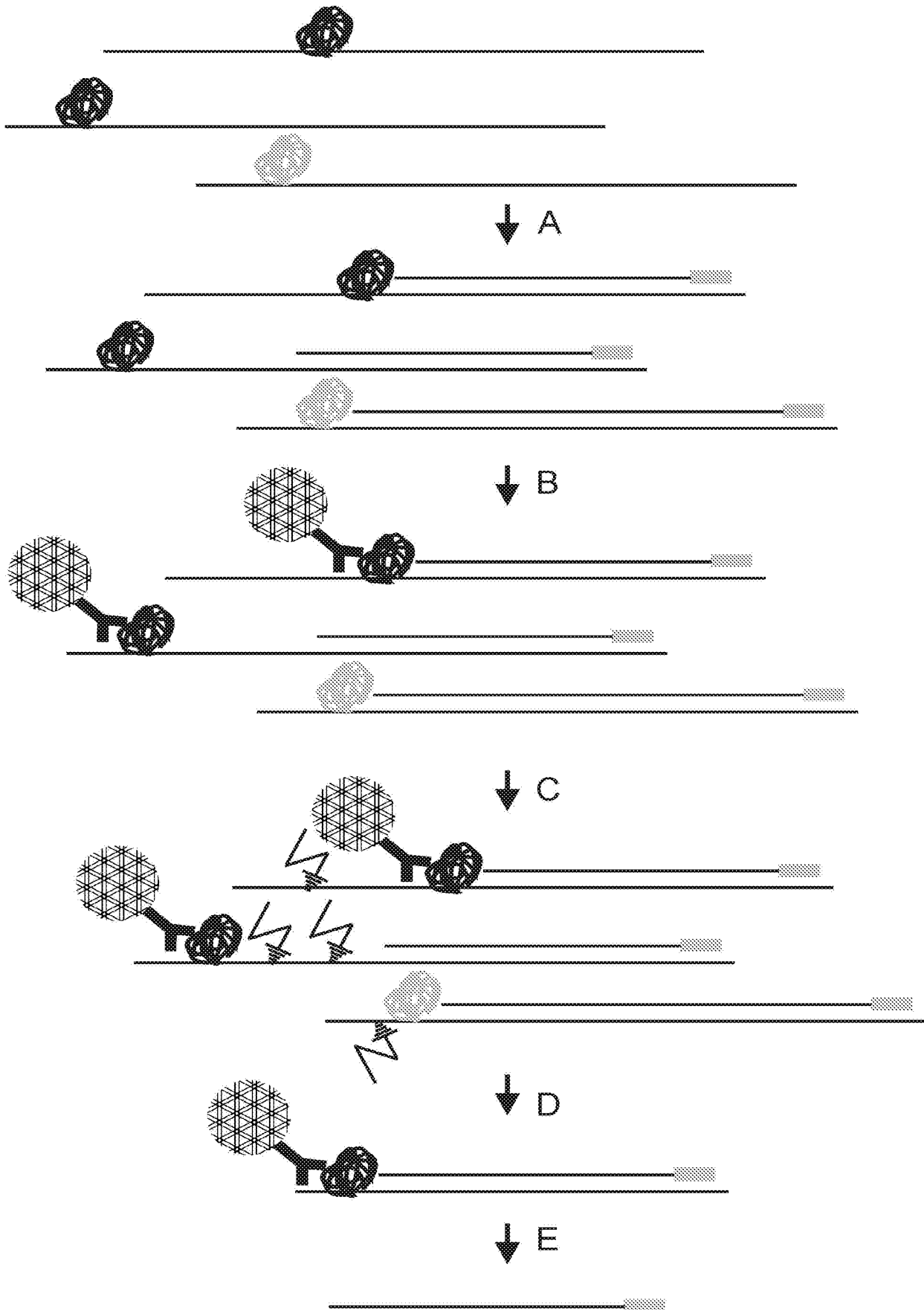


FIG. 7

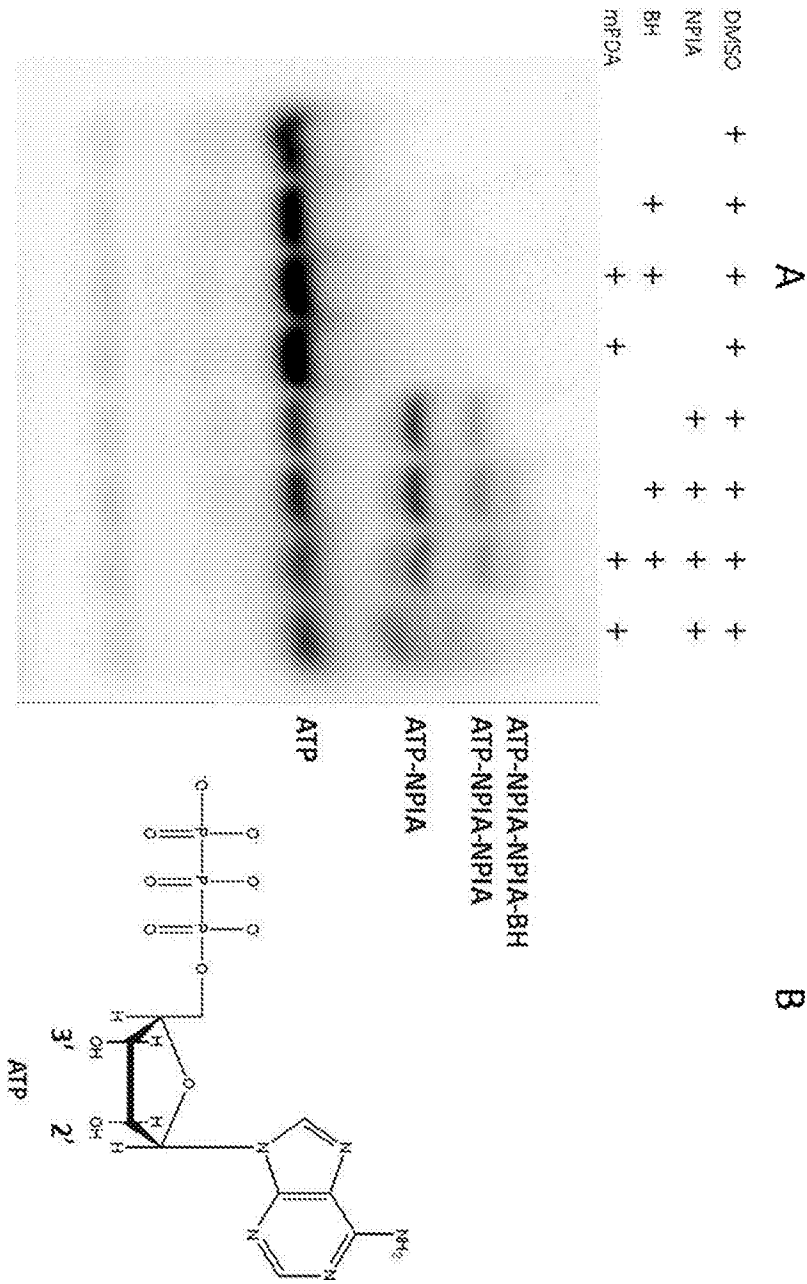


FIG. 8

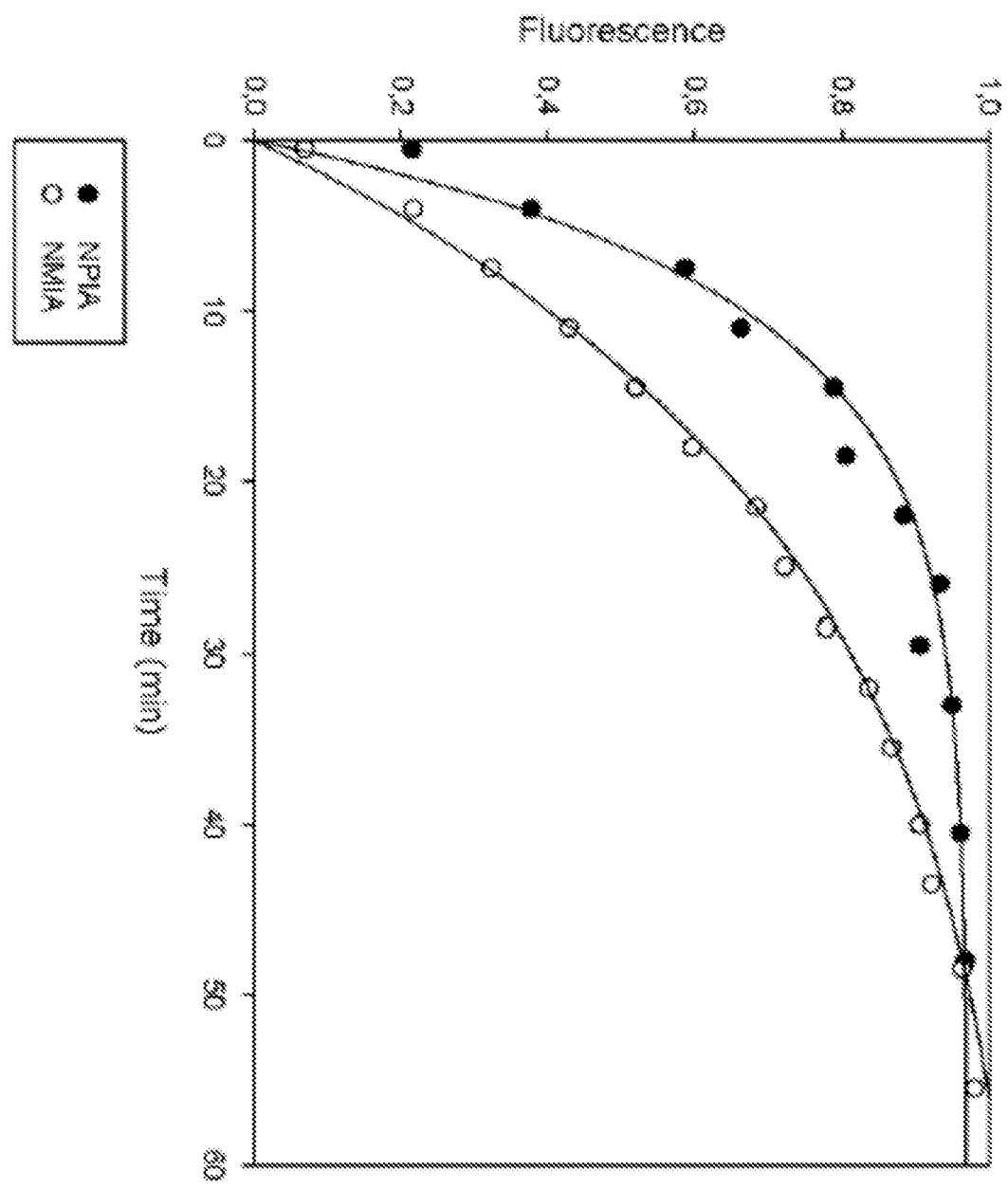


FIG. 9

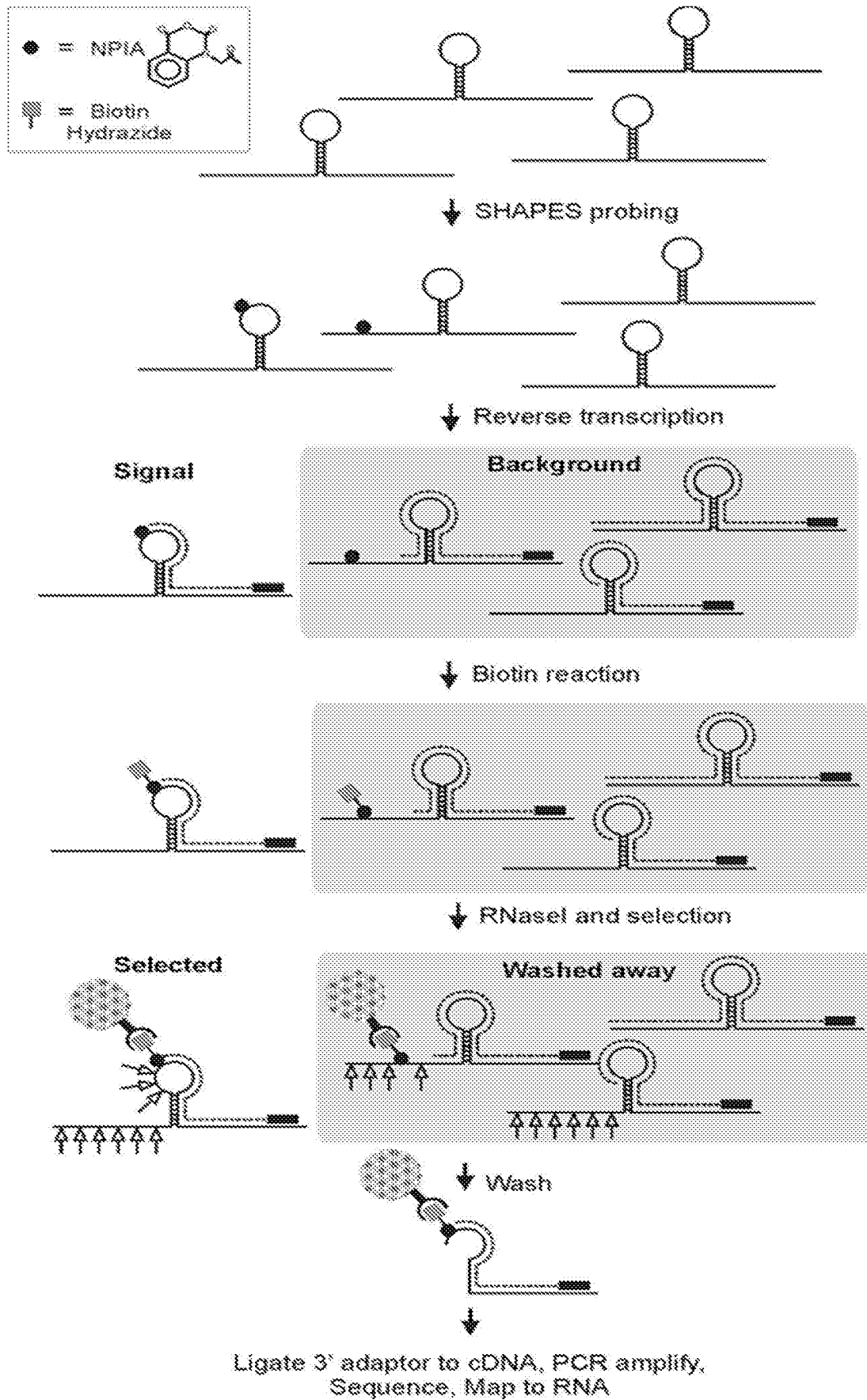


FIG. 10

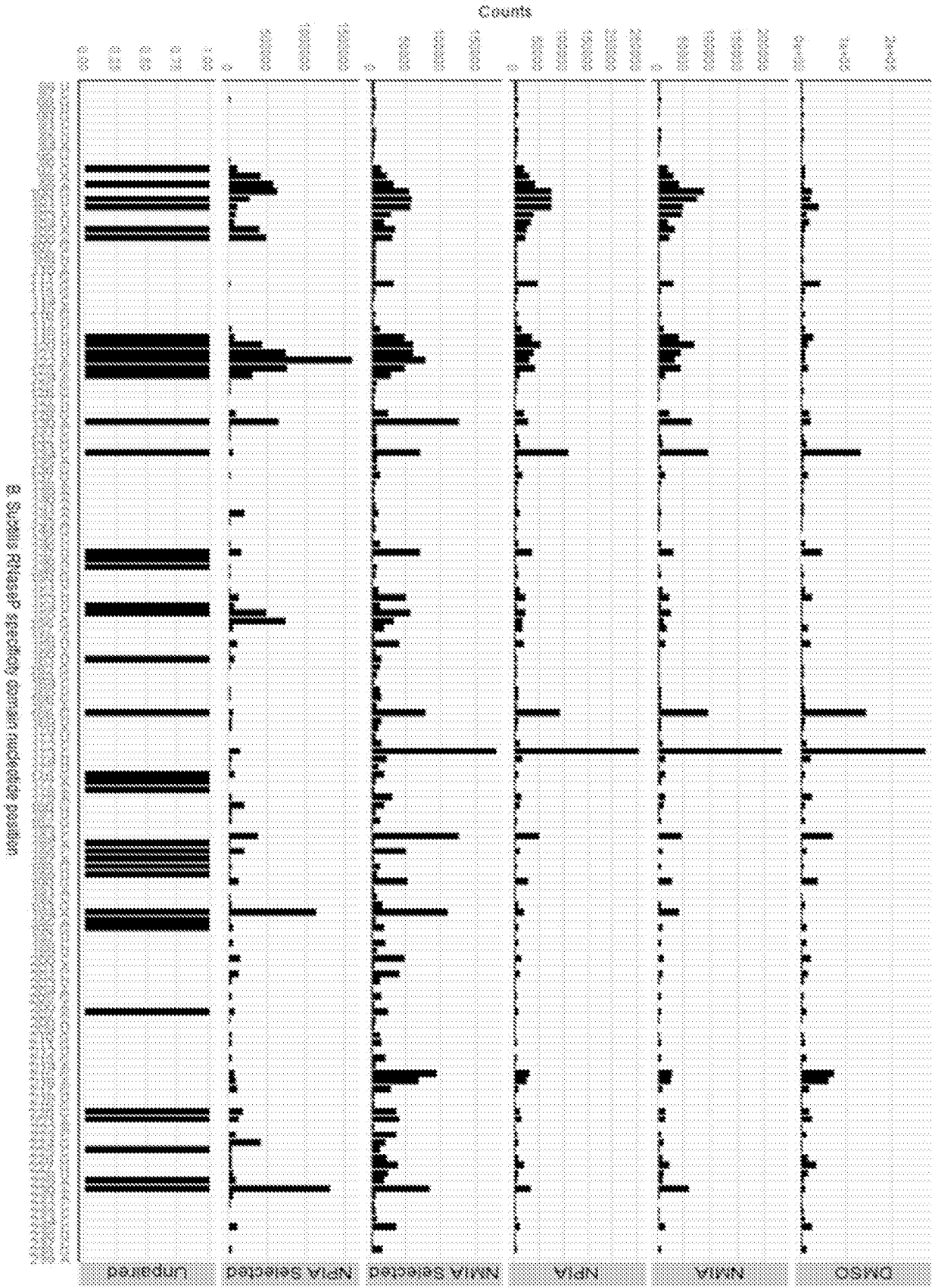


FIG. 11

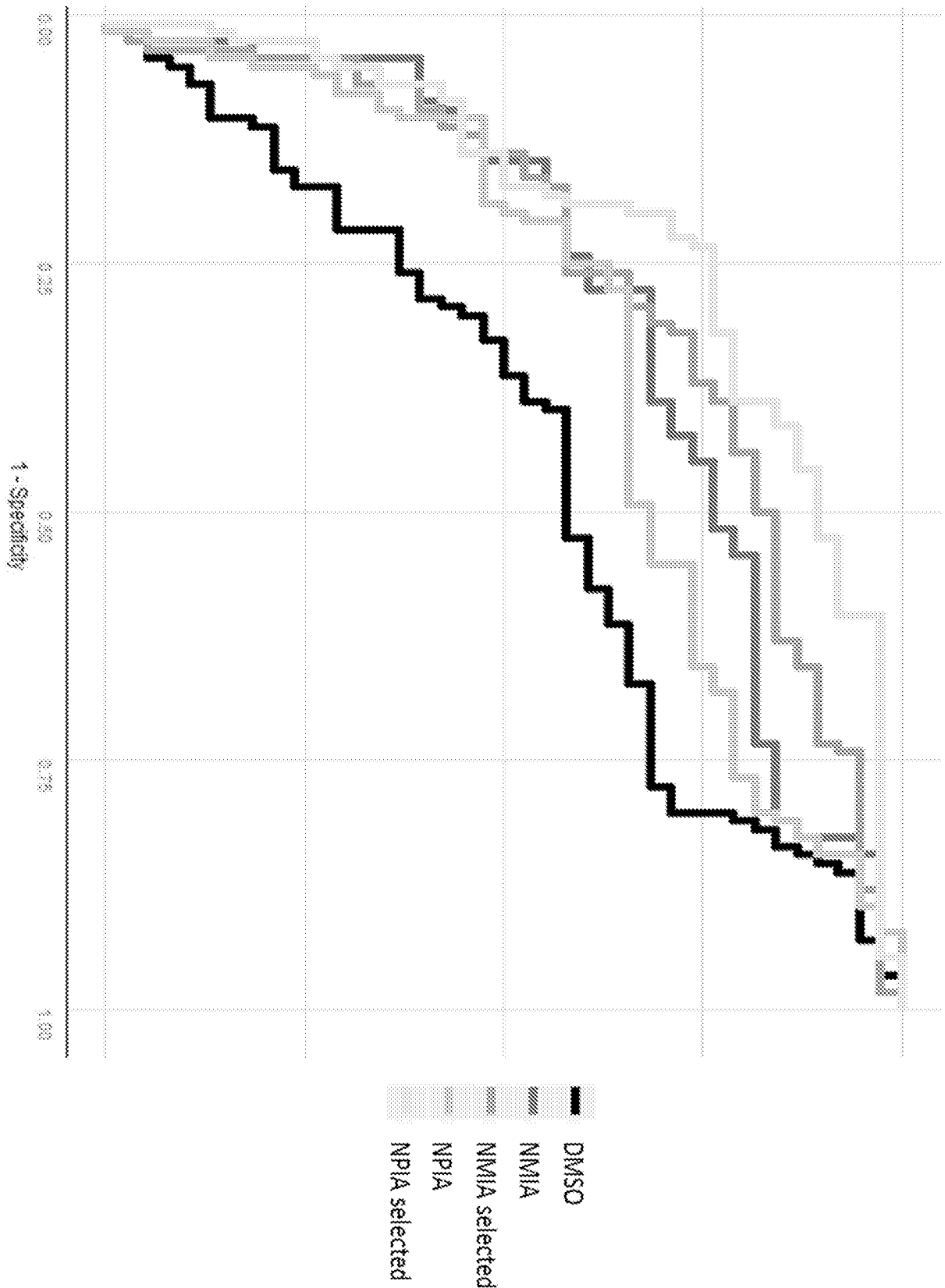


FIG. 12

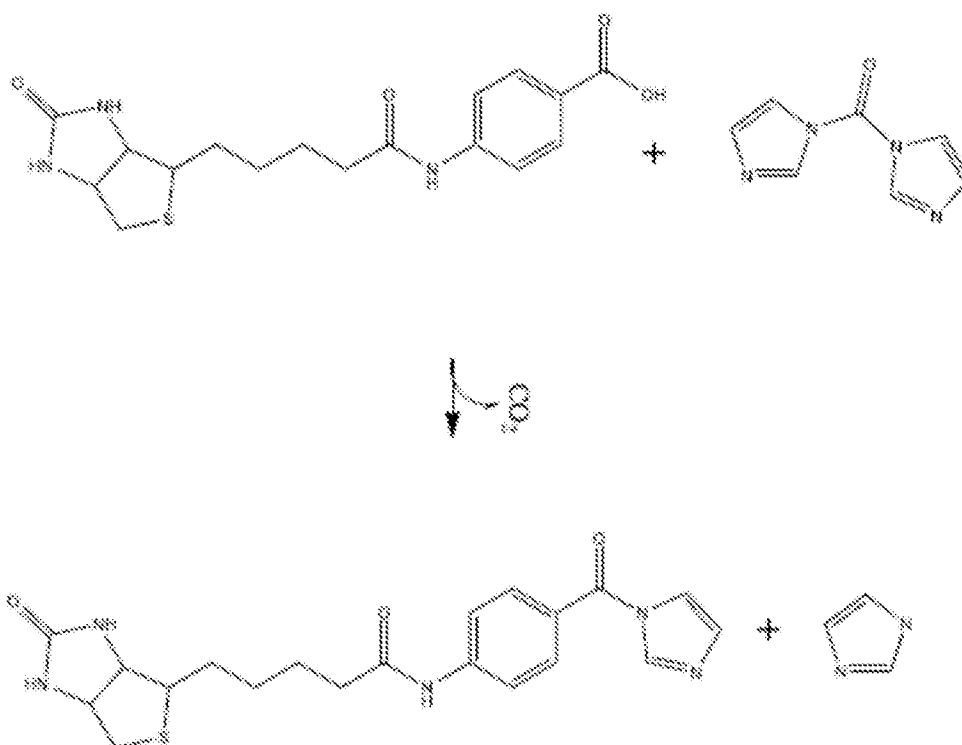
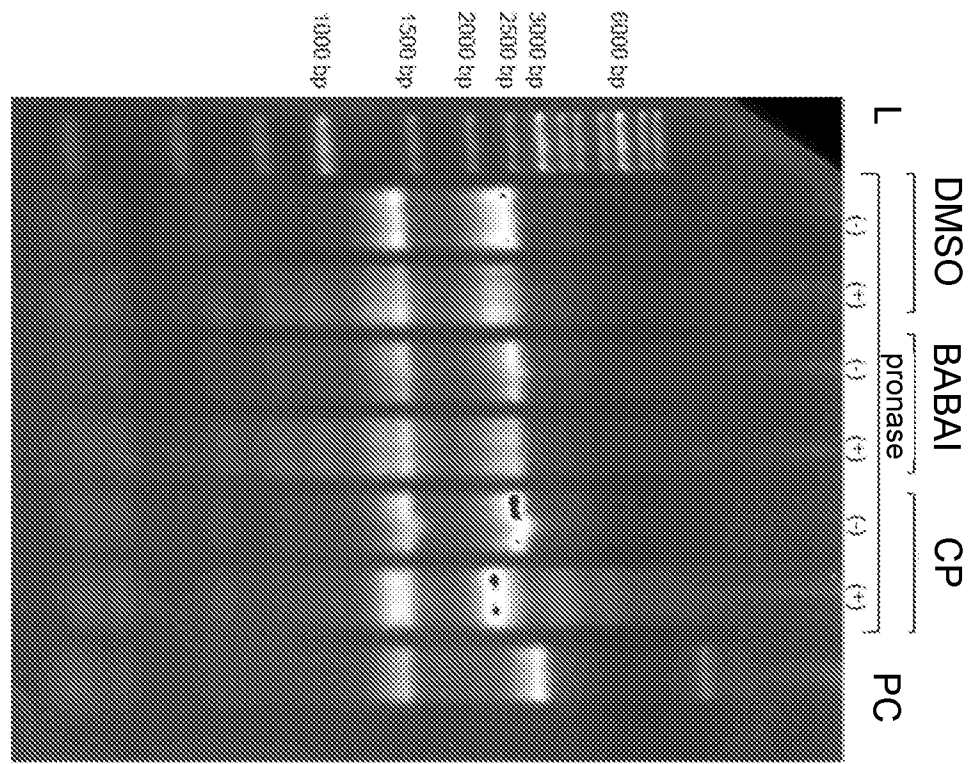


FIG. 13

A



B

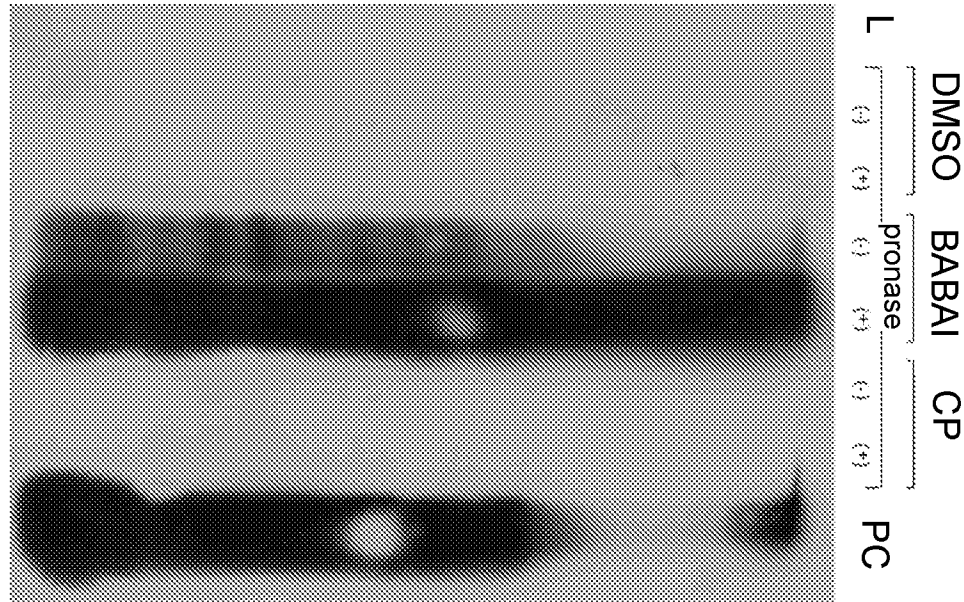
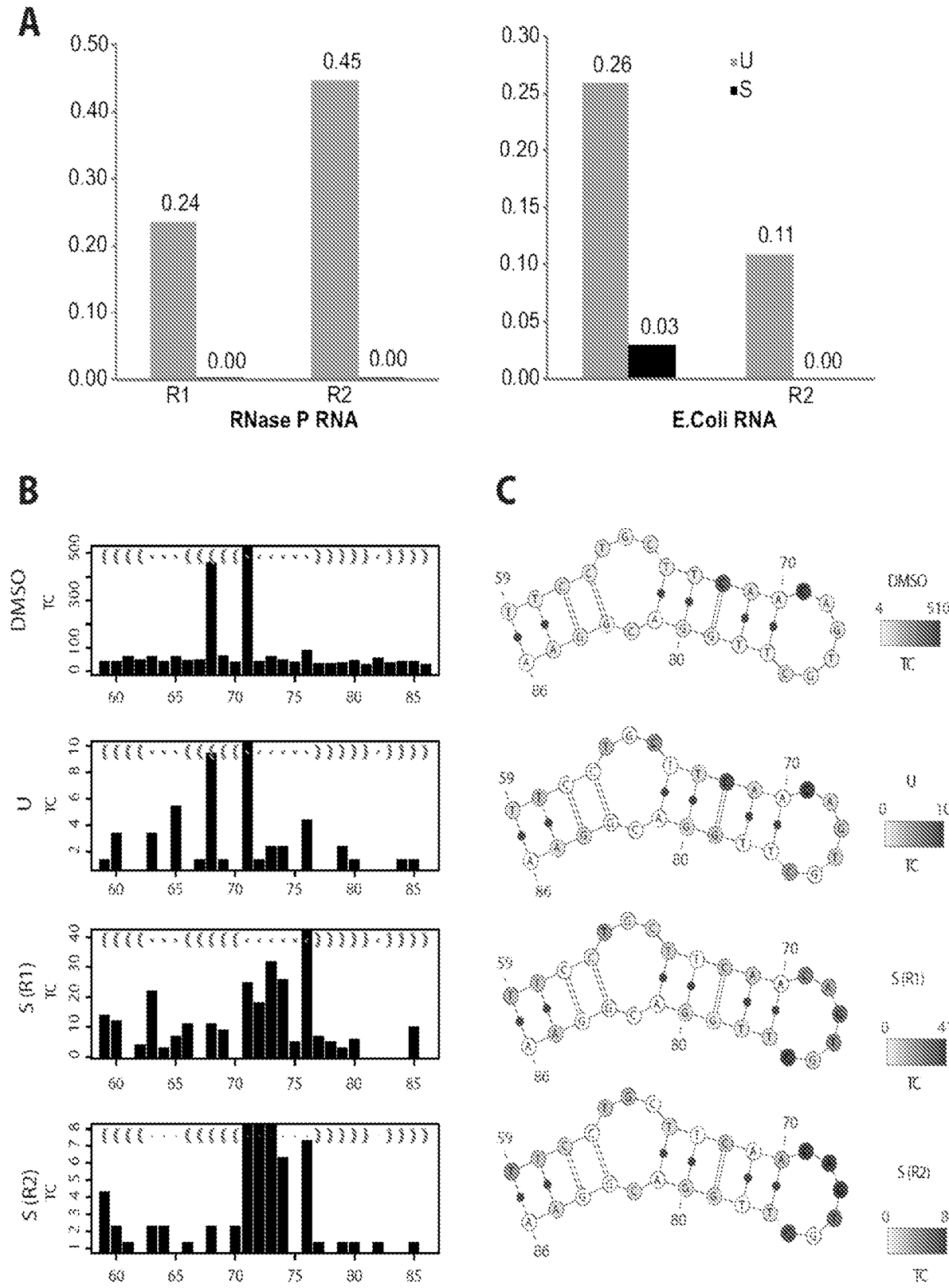


FIG. 14



INTERNATIONAL SEARCH REPORT

International application No

PCT/DK2014/050242

A. CLASSIFICATION OF SUBJECT MATTER

INV. C07H21/00 C12Q1/68
ADD.

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

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

C07H C12Q

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

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

EP0-Internal, WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	SPITALE ROBERT C ET AL: "RNA SHAPE analysis in living cells.", NATURE CHEMICAL BIOLOGY JAN 2013, vol. 9, no. 1, January 2013 (2013-01), pages 18-20, XP002732729, ISSN: 1552-4469	20-26, 34-60
Y	abstract; figure 1 ----- -/-	7-11,14, 15



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents :

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

"E" earlier application or patent but published on or after the international filing date

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

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

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

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

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

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

"&" document member of the same patent family

Date of the actual completion of the international search

26 November 2014

Date of mailing of the international search report

08/12/2014

Name and mailing address of the ISA/

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

Authorized officer

Lapopin, Laurence

INTERNATIONAL SEARCH REPORT

International application No

PCT/DK2014/050242

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	P. S. PANG ET AL: "Simplified RNA secondary structure mapping by automation of SHAPE data analysis", NUCLEIC ACIDS RESEARCH, vol. 39, no. 22, 30 September 2011 (2011-09-30), pages e151-e151, XP55153424, ISSN: 0305-1048, DOI: 10.1093/nar/gkr773	34
Y	a method for analysis RNA structure using as probing reagent NMIA, and RT-PCR analysis. Said molecule reacts selectively with the 2'-OH group of the RNA ribose ring, thus, once covalently linked to the RNA strand the probing blocks the cDNA synthesis reaction	7-11,14, 15
X	----- WO 2007/145940 A2 (UNIV NORTH CAROLINA [US]; WEEKS KEVIN M [US]; WILKINSON KEVIN [US]; VA) 21 December 2007 (2007-12-21)	34
Y	abstract, and claims 1-47	7-11,14, 15
X	----- KEVIN A WILKINSON ET AL: "High-Throughput SHAPE Analysis Reveals Structures in HIV-1 Genomic RNA Strongly Conserved across Distinct Biological States", PLOS BIOLOGY, PUBLIC LIBRARY OF SCIENCE, US, vol. 6, no. 4, 1 April 2008 (2008-04-01), pages 883-899, XP007913443, ISSN: 1544-9173	34
Y	abstract, and "material and methods"	7-11,14, 15
A	----- US 2004/067492 A1 (PENG ALLAN [US] ET AL) 8 April 2004 (2004-04-08) abstract	1-60
A	----- YUE WAN ET AL: "Understanding the transcriptome through RNA structure", NATURE REVIEWS GENETICS, vol. 12, no. 9, 18 August 2011 (2011-08-18), pages 641-655, XP55153916, ISSN: 1471-0056, DOI: 10.1038/nrg3049 cited in the application the whole document	1-60
	----- -/--	

INTERNATIONAL SEARCH REPORT

International application No

PCT/DK2014/050242

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	KÖNIG JULIAN ET AL: "Protein-RNA interactions: new genomic technologies and perspectives.", NATURE REVIEWS. GENETICS FEB 2011, vol. 13, no. 2, February 2011 (2011-02), pages 77-83, XP002732730, ISSN: 1471-0064	1-6,12, 13,16-33
Y	the whole document -----	7-11,14, 15

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No

PCT/DK2014/050242

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 2007145940 A2	21-12-2007	US 2010035761 A1 WO 2007145940 A2	11-02-2010 21-12-2007
US 2004067492 A1	08-04-2004	EP 1556511 A2 JP 4718181 B2 JP 2006501844 A US 2004067492 A1 WO 2004033629 A2	27-07-2005 06-07-2011 19-01-2006 08-04-2004 22-04-2004