

CHEMICAL COMPOSITIONS AND METHODS OF USING SAME

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application claims priority to, and the benefit of, U.S. Provisional Application No. 62/424,887, filed November 21, 2016; U.S. Provisional Application No. 62/457,237, filed February 10, 2017; and U.S. Provisional Application No. 62/536,147, filed July 24, 2017. The contents of each of the aforementioned patent applications are incorporated herein by reference in their entireties.

SEQUENCE LISTING

[0002] The instant application contains a Sequence Listing which has been submitted in ASCII format via EFS-Web and is hereby incorporated by reference in its entirety. Said ASCII copy, created on November 20, 2017, is named NATE033-001WO_SeqList_ST25.txt and is 19,351 bytes in size.

BACKGROUND OF THE INVENTION

[0003] There are currently a variety of methods for nucleic acid sequencing, *i.e.*, the process of determining the precise order of nucleotides within a nucleic acid molecule. Current methods require amplifying a nucleic acid enzymatically, *e.g.*, PCR, and/or by cloning. Further enzymatic polymerizations are required to produce a detectable signal by a light detection means. Such amplification and polymerization steps are costly and/or time-consuming. Thus, there is a need in the art for a method of nucleic acid sequencing that is rapid and amplification- and enzyme-free. The present disclosure addresses these needs.

SUMMARY

[0004] The present disclosure provides sequencing probes, methods, kits, and apparatuses that provide rapid enzyme-free, amplification-free, and library-free nucleic acid sequencing that has long-read-lengths and with low error rate. The sequencing probes described herein include barcode domains in which each position in the barcode domain corresponds to at least two nucleotides in the target binding domain. Moreover, the methods, kits, and apparatuses have rapid sample-to-answer capability. These features are particularly useful for sequencing in a

clinical setting. The present disclosure is an improvement of the disclosure disclosed in Patent Publication No. U.S. 2016/0194701, the contents of which are herein incorporated by reference in their entirety.

[0004a] In one aspect, the present disclosure provides a sequencing probe comprising a target binding domain and a barcode domain; wherein the target binding domain comprises at least eight nucleotides and is capable of binding a target nucleic acid, wherein at least six nucleotides in the target binding domain are capable of identifying a corresponding nucleotide in the target nucleic acid molecule and wherein at least two nucleotides in the target binding domain do not identify a corresponding nucleotide in the target nucleic acid molecule; wherein at least two nucleotides of the at least six nucleotides in the target binding domain are modified nucleotides or nucleotide analogues; wherein the barcode domain comprises a synthetic backbone, the barcode domain comprising at least three attachment positions, each attachment position comprising at least one attachment region comprising at least one nucleic acid sequence capable of being bound by a complementary nucleic acid molecule, wherein the nucleic acid sequence of the at least three attachment positions determines the position and identity of the at least six nucleotides in the target nucleic acid that is bound by the target binding domain, and wherein each of the at least three attachment positions have a different nucleic acid sequence.

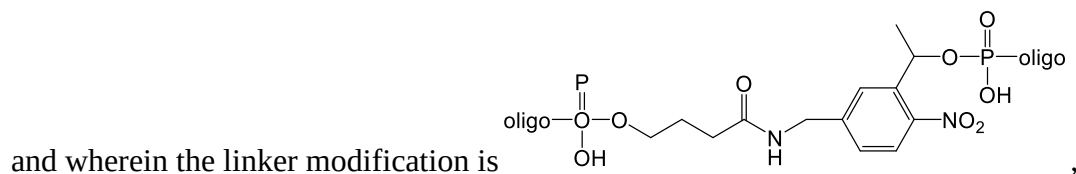
[0004b] In another aspect, the present disclosure provides a sequencing probe comprising a target binding domain and a barcode domain; wherein the target binding domain comprises at least eight nucleotides and is capable of binding a target nucleic acid, wherein at least six nucleotides in the target binding domain are capable of identifying a corresponding nucleotide in the target nucleic acid molecule and wherein at least two nucleotides in the target binding domain do not identify a corresponding nucleotide in the target nucleic acid molecule; wherein at least two nucleotides of the at least six nucleotides in the target binding domain are modified nucleotides or nucleotide analogues; wherein the barcode domain comprises a synthetic backbone, the barcode domain comprising at least three attachment positions, each attachment position comprising at least one attachment region comprising at least one nucleic acid sequence capable of being bound by a complementary nucleic acid molecule, wherein each attachment position of the at least three attachment positions corresponds to two nucleotides of the at least six nucleotides in the target binding domain and each of the at least three attachment positions have a different nucleic acid sequence, and wherein the nucleic acid sequence of each position of the

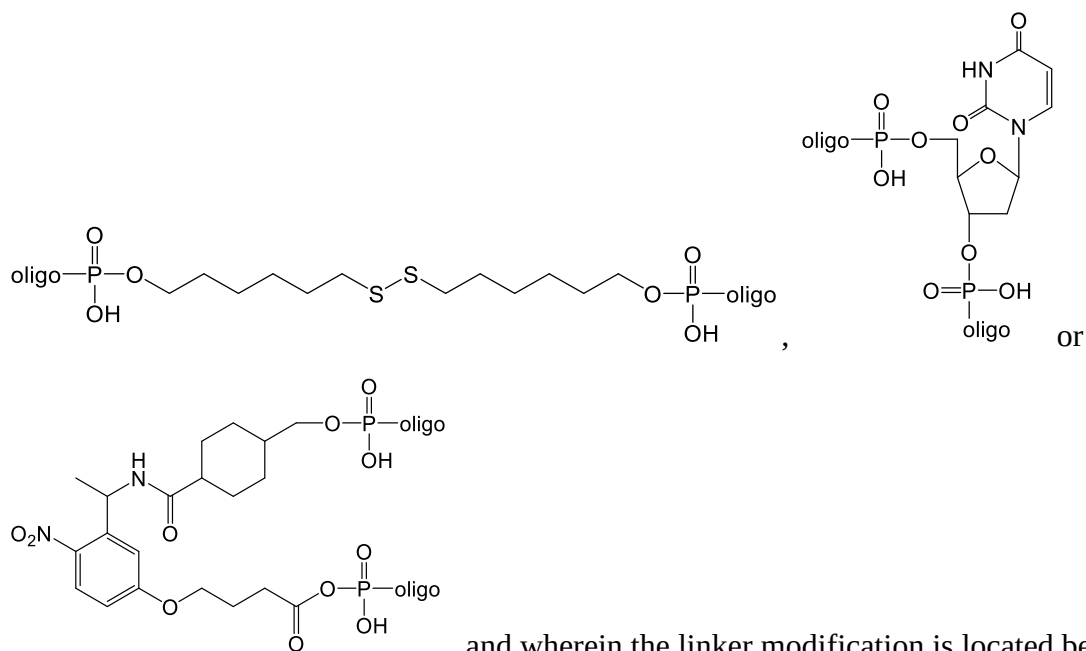
at least three attachment positions determines the position and identity of the corresponding two nucleotides of the at least six nucleotides in the target nucleic acid that is bound by the target binding domain.

[0004c] In another aspect, the present disclosure provides a sequencing probe comprising a target binding domain and a barcode domain; wherein the target binding domain comprises at least ten nucleotides and is capable of binding a target nucleic acid, wherein at least six nucleotides in the target binding domain are capable of identifying a corresponding nucleotide in the target nucleic acid molecule and wherein at least four nucleotides in the target binding domain do not identify a corresponding nucleotide in the target nucleic acid molecule; wherein the barcode domain comprises a synthetic backbone, the barcode domain comprising at least three attachment positions, each attachment position comprising at least one attachment region comprising at least one nucleic acid sequence capable of being bound by a complementary nucleic acid molecule, wherein the nucleic acid sequence of the at least three attachment positions determines the position and identity of the at least six nucleotides in the target nucleic acid that is bound by the target binding domain, and wherein each of the at least three attachment positions have a different nucleic acid sequence.

[0004d] In another aspect, the present disclosure also provides a sequencing probe comprising a target binding domain and a barcode domain; wherein the target binding domain comprises at least ten nucleotides and is capable of binding a target nucleic acid, wherein at least six nucleotides in the target binding domain are capable of identifying a corresponding nucleotide in the target nucleic acid molecule and wherein at least four nucleotides in the target binding domain do not identify a corresponding nucleotide in the target nucleic acid molecule; wherein the barcode domain comprises a synthetic backbone, the barcode domain comprising at least three attachment positions, each attachment position comprising at least one attachment region comprising at least one nucleic acid sequence capable of being bound by a complementary nucleic acid molecule, wherein each attachment position of the at least three attachment positions corresponds to two nucleotides of the at least six nucleotides in the target binding domain and each of the at least three attachment positions have a different nucleic acid sequence, and wherein the nucleic acid sequence of each position of the at least three attachment positions determines the position and identity of the corresponding two nucleotides of the at least six nucleotides in the target nucleic acid that is bound by the target binding domain.

[0005] In another aspect, the present disclosure provides a complex comprising a) a composition comprising a target binding domain and a barcode domain, wherein the target binding domain comprises at least eight nucleotides and is capable of binding a target nucleic acid, wherein at least six nucleotides in the target binding domain are capable of identifying a corresponding nucleotide in the target nucleic acid molecule and wherein at least two nucleotides in the target binding domain do not identify a corresponding nucleotide in the target nucleic acid molecule; wherein at least two nucleotides of the at least six nucleotides in the target binding domain are modified nucleotides or nucleotide analogues; wherein the barcode domain comprises a synthetic backbone, the barcode domain comprising at least three attachment positions, each attachment position comprising at least one attachment region comprising at least one nucleic acid sequence capable of being bound by a complementary nucleic acid molecule, wherein the nucleic acid sequence of the at least three attachment positions determines the position and identity of the at least six nucleotides in the target nucleic acid that is bound by the target binding domain, and wherein each of the at least three attachment positions have a different nucleic acid sequence; and a first complementary primary nucleic acid molecule hybridized to a first attachment position of the at least three attachment positions, wherein the first primary complementary nucleic acid molecule comprises at least two domains and a linker modification, wherein the first domain is hybridized to the first attachment position of the barcode domain and the second domain capable of hybridizing to at least one complementary secondary nucleic acid molecule,



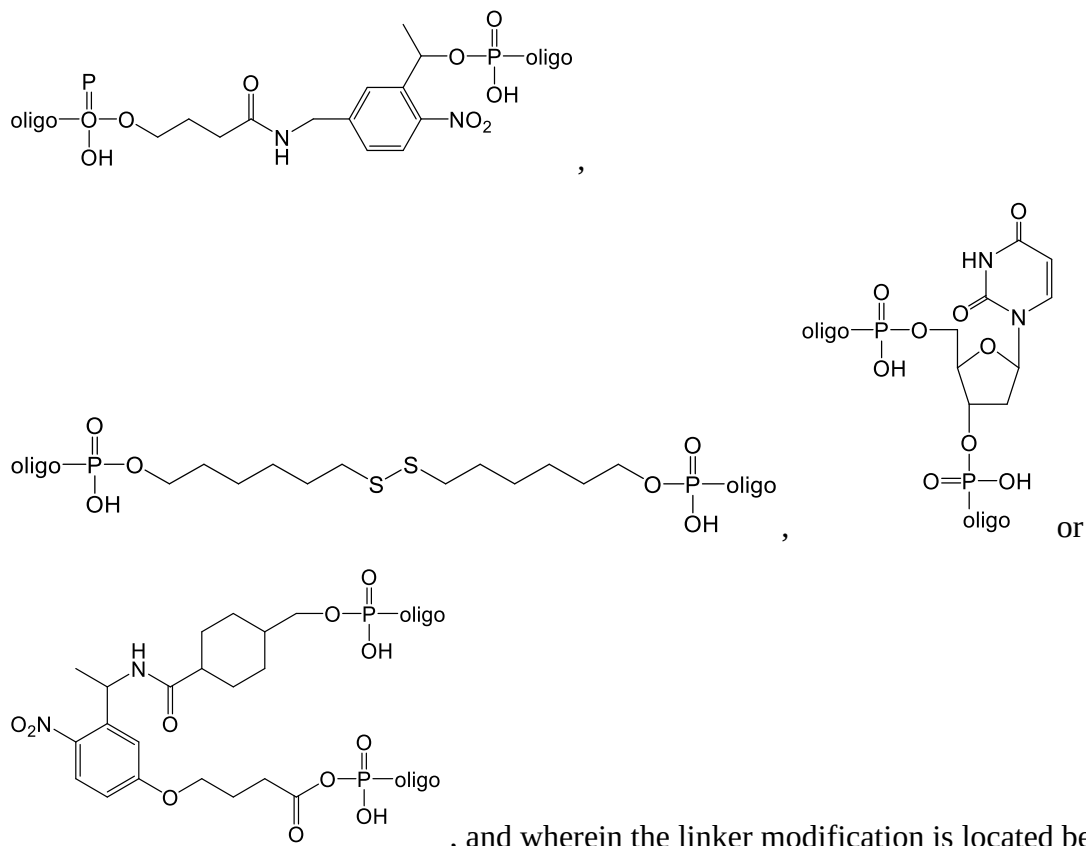


first and second domains.

[0006] In another aspect, the present disclosure provides a complex comprising a) a composition comprising a target binding domain and a barcode domain, wherein the target binding domain comprises at least eight nucleotides and is capable of binding a target nucleic acid, wherein at least six nucleotides in the target binding domain are capable of identifying a corresponding nucleotide in the target nucleic acid molecule and wherein at least two nucleotides in the target binding domain do not identify a corresponding nucleotide in the target nucleic acid molecule; wherein at least two nucleotides of the at least six nucleotides in the target binding domain are modified nucleotides or nucleotide analogues; wherein the barcode domain comprises a synthetic backbone, the barcode domain comprising at least three attachment positions, each attachment position comprising at least one attachment region comprising at least one nucleic acid sequence capable of being bound by a complementary nucleic acid molecule, wherein the nucleic acid sequence of the at least three attachment positions determines the position and identity of the at least six nucleotides in the target nucleic acid that is bound by the target binding domain, wherein each attachment position of the at least three attachment positions corresponds to two nucleotides of the at least six nucleotides in the target binding domain and each of the at least three attachment positions have a different nucleic acid sequence, and wherein the nucleic acid sequence of each position of the at least three attachment positions

determines the position and identity of the corresponding two nucleotides of the at least six nucleotides in the target nucleic

acid that is bound by the target binding domain; and a first complementary primary nucleic acid molecule hybridized to a first attachment position of the at least three attachment positions, wherein the first primary complementary nucleic acid molecule comprises at least two domains and a linker modification, wherein the first domain is hybridized to the first attachment position of the barcode domain and the second domain capable of hybridizing to at least one complementary secondary nucleic acid molecule, and wherein the linker modification is



, and wherein the linker modification is located between the first and second domains.

[0007] The present disclosure provides a sequencing probe comprising a target binding domain and a barcode domain; wherein the target binding domain comprises at least eight nucleotides and is capable of binding a target nucleic acid, wherein at least six nucleotides in the target binding domain are capable of identifying a corresponding nucleotide in the target nucleic acid molecule and wherein at least two nucleotides in the target binding domain do not identify a corresponding nucleotide in the target nucleic acid molecule; wherein at least two nucleotides of the at least six nucleotides in the target binding domain are modified nucleotides or nucleotide

analogues; wherein the barcode domain comprises a synthetic backbone, the barcode domain comprising at least three attachment positions, each attachment position comprising at least one

molecule and to a first detectable label and an at least second detectable label. The first and at least second detectable labels can have the same emission spectrum or can have different emission spectra. When the secondary nucleic molecule is hybridized to at least one primary nucleic acid molecule, at least one tertiary nucleic acid molecule comprising a first detectable label and an at least second detectable label, and to a first detectable label and an at least second detectable label, the at least first and second detectable labels located on the secondary nucleic acid molecule can have the same emission spectra and the at least first and second detectable labels located on the tertiary nucleic acid molecule can have the same emission spectra, and wherein the emission spectra of the detectable labels on the secondary nucleic acid molecule can be different than the emission spectra of the detectable labels on the tertiary nucleic acid molecule.

[0028] The secondary nucleic acid molecule can comprise a cleavable linker. Preferably, the cleavable linker is located between the first domain and the second domain. The cleavable linker can be a photo-cleavable linker (e.g., UV-light cleavable linker), a reducing agent cleavable linker, or an enzymatically cleavable linker. Preferably, the linker is a photo-cleavable linker.

[0029] A tertiary nucleic acid molecule can comprise at least two domains, a first domain capable of binding to complementary sequence in at least one secondary nucleic acid molecule; and a second domain capable of binding to a first detectable label and an at least second detectable label.

[0030] The tertiary nucleic molecule can be hybridized to at least one secondary nucleic acid molecule and can comprise a first detectable label and an at least second detectable label. The first and at least second detectable labels can have the same emission spectrum or can have different emission spectra.

[0031] The tertiary nucleic acid molecule can comprise a cleavable linker. Preferably, the cleavable linker is located between the first domain and the second domain. The cleavable linker can be a photo-cleavable linker (e.g., UV-light cleavable linker), a reducing agent cleavable linker, or an enzymatically cleavable linker. Preferably, the linker is a photo-cleavable linker.

[0031a] In another aspect, the present disclosure also provides a population of the sequencing probes comprising a plurality of the sequencing probes disclosed herein.

[0032] The present disclosure also provides a population of the sequencing probes comprising a plurality of the sequencing probes disclosed herein. Preferably, each sequencing probe within

the plurality of sequencing probes comprises a different target binding domain and binds a different region within a target nucleic acid.

[0042] The method can further comprise assembling each identified linear order of nucleotides in the at least first region and at least second region of the immobilized target nucleic acid, thereby identifying a sequence for the immobilized target nucleic acid.

[0043] Steps (5) and (6) can occur sequentially or concurrently. The first and at least second detectable labels can have the same emission spectrum or can have different emission spectra. The third and at least fourth detectable labels can have the same emission spectrum or can have different emission spectra.

[0044] The first complementary nucleic acid molecule can comprise a cleavable linker. The second complementary nucleic acid molecule can comprise a cleavable linker. The first complementary nucleic acid molecule and the second complementary nucleic acid molecule can each comprise a cleavable linker. Preferably, the cleavable linker is photo-cleavable. The release force can be light. Preferably, UV light. The light can be provided by a light source selected from the group consisting of an arc-lamp, a laser, a focused UV light source, and light emitting diode.

[0045] The first complementary nucleic acid molecule and the first hybridizing nucleic acid molecule lacking a detectable label can comprise the same nucleic acid sequence. The first hybridizing nucleic acid molecule lacking a detectable label can comprise a nucleic acid sequence complementary to a flanking single-stranded polynucleotide adjacent to the first attachment position in the barcode domain.

[0046] The second complementary nucleic acid molecule and the second hybridizing nucleic acid molecule lacking a detectable label can comprise the same nucleic acid sequence. The second hybridizing nucleic acid molecule lacking a detectable label can comprise a nucleic acid sequence complementary to a flanking single-stranded polynucleotide adjacent to the second attachment position in the barcode domain.

[0047] In another aspect, the present disclosure also provides a method for determining a nucleotide sequence of a nucleic acid comprising (1) hybridizing a first sequencing probe as disclosed herein to a target nucleic acid that is optionally immobilized to a substrate at one or more positions; (2) hybridizing a first complementary nucleic acid molecule comprising a first detectable label and a second detectable label to a first attachment position of the at least three attachment positions of the barcode domain; (3) identifying the first and the second detectable

label of the first complementary nucleic acid molecule hybridized to the first attachment position; (4) removing

complementary nucleic acid molecule hybridized to the third attachment position; and (20) identifying the linear order of at least six nucleotides in the second region of the optionally immobilized target nucleic acid hybridized to the target binding domain of the at least second sequencing probe based on the identity of the first detectable label, second detectable label, third detectable label, fourth detectable label, fifth detectable label and sixth detectable label.

[0049] The method can further comprise assembling each identified linear order of nucleotides in the at least first region and at least second region of the immobilized target nucleic acid, thereby identifying a sequence for the immobilized target nucleic acid.

[0050] Steps (4) and (5) can occur sequentially or concurrently. The first and at least second detectable labels can have the same emission spectrum or can have different emission spectra. The third and at least fourth detectable labels can have the same emission spectrum or can have different emission spectra. The fifth and at least sixth detectable labels can have the same emission spectrum or can have different emission spectra.

[0051] The first complementary nucleic acid molecule can comprise a cleavable linker. The second complementary nucleic acid molecule can comprise a cleavable linker. The third complementary nucleic acid molecule can comprise a cleavable linker. The first complementary nucleic acid molecule, the second complementary nucleic acid molecule and the at least third complementary nucleic molecule can each comprise a cleavable linker. Preferably, the cleavable linker is photo-cleavable. A method of removing any one of the first complementary nucleic acid molecule, the second complementary nucleic acid molecule and the at least third complementary nucleic molecule can comprise contact with light. Preferably, UV light. The light can be provided by a light source selected from the group consisting of an arc-lamp, a laser, a focused UV light source, and light emitting diode.

[0052] In another aspect, the present disclosure also provides a method for determining the nucleotide sequence of a nucleic acid comprising (1) hybridizing a first sequencing probe as disclosed herein to a first region of a target nucleic acid obtained from a predetermine gene, wherein the target nucleic acid is optionally immobilized to a substrate at one or more positions; (2) hybridizing a first complementary nucleic acid molecule comprising a first detectable label and a second detectable label to a first attachment position of the at least three attachment positions of the barcode domain; (3) detecting the first and the second detectable label of the first

complementary nucleic acid molecule hybridized to the first attachment position; (4) removing the first and the second

complementary nucleic acid molecule hybridized to the third attachment position; and (20) identifying the linear order of at least six nucleotides in the second region of the optionally immobilized target nucleic acid hybridized to the target binding domain of the at least second sequencing probe based on the identity of the first detectable label, second detectable label, third detectable label, fourth detectable label, fifth detectable label and sixth detectable label.

[0054] The method can further comprise assembling each identified linear order of nucleotides in the at least first region and at least second region of the immobilized target nucleic acid, thereby identifying a sequence for the immobilized target nucleic acid.

[0055] Steps (4) and (5) can occur sequentially or concurrently. The first and at least second detectable labels can have the same emission spectrum or can have different emission spectra. The third and at least fourth detectable labels can have the same emission spectrum or can have different emission spectra. The fifth and at least sixth detectable labels can have the same emission spectrum or can have different emission spectra.

[0056] The first complementary nucleic acid molecule can comprise a cleavable linker. The second complementary nucleic acid molecule can comprise a cleavable linker. The third complementary nucleic acid molecule can comprise a cleavable linker. The first complementary nucleic acid molecule, the second complementary nucleic acid molecule and the at least third complementary nucleic acid molecule can each comprise a cleavable linker. Preferably, the cleavable linker is photo-cleavable. A method of removing any one of the first complementary nucleic acid molecule, the second complementary nucleic acid molecule and the at least third complementary nucleic acid molecule can comprise contact with light. Preferably, UV light. The light can be provided by a light source selected from the group consisting of an arc-lamp, a laser, a focused UV light source, and light emitting diode.

[0057] In another aspect, the present disclosure also provides an apparatus for performing any of the methods disclosed herein.

[0057a] In another aspect, the present disclosure provides a kit comprising a substrate; a population of sequencing probes disclosed herein, at least one capture probe, at least three complementary nucleic acid molecules comprising a first detectable label and at least two second detectable labels, and instructions for use.

[0058] The present disclosure also provides one or more kits comprising a substrate, a population of sequencing probes disclosed herein, at least three complementary nucleic acid

molecules comprising a first detectable label and at least two second detectable labels, and instructions for use. The one or more kits can further comprise at least one capture probe. The one or more kits can further comprise at least two capture probes.

[0059] Any of the above examples can be combined with any other example.

[0060] Unless otherwise defined, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this disclosure belongs. In the Specification, the singular forms also include the plural unless the context clearly dictates otherwise; as examples, the terms “a,” “an,” and “the” are understood to be singular or plural and the term “or” is understood to be inclusive. By way of example, “an element” means one or more element. Throughout the specification the word “comprising,” or variations such as “comprises” or “comprising,” will be understood to imply the inclusion of a stated element, integer or step, or group of elements, integers or steps, but not the exclusion of any other element, integer or step, or group of elements, integers or steps. About can be understood as within 10%, 9%, 8%, 7%, 6%, 5%, 4%, 3%, 2%, 1%, 0.5%, 0.1%, 0.05%, or 0.01% of the stated value. Unless otherwise clear from the context, all numerical values provided herein are modified by the term “about.”

[0061] Although methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present disclosure, suitable methods and materials are described below. All publications, patent applications, patents, and other references mentioned herein are incorporated by reference in their entirety. The references cited herein are not admitted to be prior art to the disclosure. In the case of conflict, the present Specification, including definitions, will control. In addition, the materials, methods, and examples are illustrative only and are not intended to be limiting. Other features and advantages of the disclosure will be apparent from the following detailed description and claim.

BRIEF DESCRIPTION OF THE DRAWINGS

[0062] The patent or application file contains at least one drawing executed in color. Copies of this patent or patent application publication with color drawings will be provided by the Office upon request and payment of the necessary fee.

[0063] The above and further features will be more clearly appreciated from the following detailed description when taken in conjunction with the accompanying drawings.

[0064] FIG. 1 is an illustration of one exemplary sequencing probe of the present disclosure.

[0065] FIG. 2 shows the design of standard and three-part sequencing probes of the present disclosure.