POLYMERIC MULTIVALENT CONJUGATES AND RELATED USES

CROSS-REFERENCE TO RELATED APPLICATIONS

This application claims the benefit of and priority to U.S. Provisional Patent Application No. 63/568,966, filed March 22, 2024, the entire contents of which are incorporated herein by reference.

BACKGROUND

Polynucleotide sequencing technology has applications in biomedical research and healthcare settings. Improved methods of polynucleotide sequencing require enhanced surface chemistry, on-support polynucleotide amplification, labeling and detection of nucleobase identities, and base calling. Currently, these elements produce barriers in existing sequencing technology that result in limits in throughput and poor signal-to-noise ratio, and ultimately to increased costs associated with polynucleotide sequencing.

There exists a need for new polynucleotide sequencing methods with improved surface chemistry, on-support amplification, labeling and detection of nucleobase identities, and base calling. The present disclosure provides methods and compositions to improve sequencing of polynucleotides.

SUMMARY

In an aspect, the present disclosure provides a method comprising:

a. contacting:

(i) a plurality of template nucleic acid molecules comprising two or more copies of a target sequence and two or more copies of a binding sequence for a forward sequencing primer,

(ii) a plurality of forward sequencing primers comprising a sequence complementary to the binding sequence for the forward sequencing primer,

(iii) a plurality of first polymerases, and

(iv) a plurality of polymeric molecules of Formula (I):

C-((P)-E)s

(I)

an ionized form thereof, or a salt thereof, wherein:

C is a central moiety;

each P independently is an optionally substituted polymeric side chain;

each E independently is an end moiety; and

s is an integer ranging from 1 to 10;

wherein each polymeric molecule comprises at least two nucleotide moieties and at one least detectable reporter moiety,

wherein the contacting occurs under conditions sufficient to form a plurality of multivalent binding complexes comprising a nucleic acid duplex between a template nucleic acid molecule and forward sequencing primer, a first polymerase, and a nucleotide moiety of a polymeric molecule that is complementary to the nucleotide in the template nucleic acid molecule immediately adjacent to the 3' end of the forward sequencing primer, and

wherein polymerase catalyzed incorporation of a complementary nucleotide moiety into the nucleic acid duplex is inhibited;

b. detecting the detectable reporter moieties; and

c. determining the identities of nucleotides in the nucleic acid template molecules based on the detectable reporter moieties of the polymeric molecules in the plurality of multivalent binding complexes formed in step (a).

In some embodiments, the compound of Formula (I) is of Formula (II):

$$C\text{-}((P)\text{-}E)_4$$

(II)

an ionized form thereof, or a salt thereof.

In some embodiments, at least one P is substituted with (i) one or more reporter moiety and (ii) one or more nucleotide moiety.

In some embodiments, each P is substituted with (i) one or more reporter moiety and (ii) one or more nucleotide moiety.

In some embodiments, at least one P is substituted with one or more blocking moiety, negative charge moiety, or PEG-Cap moiety.

In some embodiments, each P is substituted with one or more blocking moiety, negative charge moiety, or PEG-Cap moiety.

In some embodiments, at least one P is further substituted with (iii) one or more blocking moiety, (iv) one or more negative charge moiety, and (v) one or more PEG-Cap moiety.

In some embodiments, each P is further substituted with (iii) one or more blocking moiety, (iv) one or more negative charge moiety , and (v) one or more PEG-Cap moiety.

In some embodiments, the two or more copies of a target sequence in an individual of template nucleic acid molecule are the same target sequence.

In some embodiments, two or more multivalent binding complexes form on individual template nucleic acid molecules.

In some embodiments, the plurality of forward sequencing primers are soluble.

In some embodiments, the method comprises:

d. dissociating the multivalent binding complexes under conditions sufficient to retain the nucleic acid duplexes, thereby generating a plurality of nucleic acid duplexes;

e. contacting the plurality of nucleic acid duplexes with a plurality of second polymerases and a plurality of nucleotides or analogs thereof under conditions sufficient to incorporate nucleotides or analogs thereof complementary to the nucleotides of the nucleic acid template molecules immediately adjacent to the 3' ends of the forward sequencing primers in a primer extension reaction, thereby generating a plurality of extended nucleic acid duplexes comprising extended forward sequencing primer sequences.

In some embodiments, the method comprises:

f. dissociating the second polymerases from the extended nucleic acid duplexes under conditions sufficient to retain the plurality of extended nucleic acid duplexes.

In some embodiments, the template nucleic acid molecules comprise concatemers of two or more copies of a sequence comprising (i) the binding sequence for the forward sequencing primer and (ii) the target sequence.

In some embodiments, the two or more copies of (i) the binding sequence for the forward sequencing primer hybridize to the forward sequencing primers to form nucleic acid duplexes between the template nucleic acid molecules and the forward sequencing primers.

In some embodiments, in an individual polymeric molecule, the at least two nucleotide moieties are attached to different polymeric side chains.

In some embodiments, in an individual polymeric molecule, all nucleotide moieties are the same.

In some embodiments, all nucleotide moieties in an individual polymeric molecule are dATP.

In some embodiments, all nucleotide moieties in an individual polymeric molecule are dTTP.

In some embodiments, all nucleotide moieties in an individual polymeric molecule are dGTP.

In some embodiments, all nucleotide moieties in an individual polymeric molecule are dUTP.

In some embodiments, all nucleotide moieties in an individual polymeric molecule are dCTP.

In some embodiments, in an individual polymeric molecule, all detectable reporter moieties comprise the same fluorescent label.

In some embodiments, all detectable reporter moieties in an individual polymeric molecule are the same.

In some embodiments, in an individual polymeric molecule, all detectable reporter moieties are the same and all nucleotide moieties are the same.

In some embodiments, a polymeric molecule comprises two, three, or four nucleotide moieties.

In some embodiments, an individual polymeric molecule comprises two, three, or four detectable reporter moieties.

In some embodiments, two or more nucleotide moieties in an individual polymeric molecule are associated with two or more different multivalent binding complexes on the same template nucleic acid molecule.

In some embodiments, the method comprises:

i. contacting the plurality of extended nucleic acid duplexes with a plurality of first polymerases and a plurality of polymeric molecules of Formula (I), or an ionized form thereof, an isomer thereof, or a salt thereof,

wherein the contacting occurs under conditions sufficient to form a plurality of multivalent binding complexes comprising an extended nucleic acid duplex, a first polymerase, and a nucleotide moiety of a polymeric molecule that is complementary to a nucleotide in the template nucleic acid molecule immediately adjacent to the 3' end of the extended forward sequencing primer, and

wherein polymerase catalyzed incorporation of a complementary nucleotide moiety into the extended nucleic acid duplex is inhibited;

ii. detecting the detectable reporter moieties; and

iii. determining nucleobase identities of nucleotides in the nucleic acid template sequences complementary to the nucleotide moieties of the polymeric molecules based on the detectable reporter moieties of the polymeric molecules in the plurality of multivalent binding complexes formed in step (a);

iv. dissociating the multivalent binding complexes under conditions sufficient to retain the plurality extended nucleic acid duplexes;

v. contacting the plurality of extended nucleic acid duplexes with a plurality of second polymerases and a plurality of nucleotides or analogs thereof under conditions sufficient to incorporate nucleotides or analogs thereof complementary to the nucleotides of the nucleic acid template molecules immediately adjacent to the 3' ends of the extended forward sequencing primers in a primer extension reaction, thereby generating a plurality of extended nucleic acid duplexes comprising extended forward sequencing primers.

In some embodiments, in an individual polymeric molecule, the at least two nucleotide moieties are attached to different polymeric side chains.

In some embodiments, the method comprises repeating steps (i)-(v) at least 1, 10, 20, 30, 40, 50, 70, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 800, 900, 1000, or 1500 times.

In some embodiments, the method comprises repeating steps (i)-(v) until the identities of the nucleotides in the target sequences have been determined.

In some embodiments, the method comprises, before step (i), dissociating the second polymerases from the extended nucleic acid duplexes under conditions sufficient to retain the plurality of extended nucleic acid duplexes.

In some embodiments, the template nucleic acid molecules are single-stranded DNA molecules.

In some embodiments, the nucleotides or analogs thereof comprise a removable chain terminating moiety at the 3' sugar group.

In some embodiments, the removable chain terminating moiety comprises an alkyl group, alkenyl group, alkynyl group, allyl group, aryl group, benzyl group, azide group, azido group, O-azidomethyl group, amine group, amide group, keto group, isocyanate group, phosphate group, thio group, disulfide group, carbonate group, urea group, or silyl group, and wherein the removable chain terminating moiety is cleavable with a chemical compound to generate an extendible 3'OH moiety on the sugar group.

In some embodiments, the the nucleotides or analogs thereof comprise a mixture of any combination of two or more types of nucleotides selected from the group consisting of dATP, dGTP, dCTP, dTTP and dUTP.

In some embodiments, the nucleotides or analogs thereof comprise at least one fluorophore-labeled nucleotide analog.

In some embodiments, the plurality of template nucleic acid molecules are immobilized on a support.

In some embodiments, the template nucleic acid molecules are immobilized on the support through hybridization to first surface primers immobilized on the support.

In some embodiments, the template nucleic acid molecules are covalently joined to first surface primers immobilized on the support.

In some embodiments, the template nucleic acid molecules are clonally amplified template nucleic acid molecules.

In some embodiments, the template nucleic acid molecules are generated through rolling circle amplification.

In some embodiments, sequencing the plurality of template nucleic acid molecules generates a plurality of extended forward sequencing primer strands, and wherein the method comprises:

a. retaining the plurality of template nucleic acid molecules and replacing the plurality of extended forward sequencing primer strands with a plurality of forward extension strands that are hybridized to the plurality of nucleic acid template molecules by conducting a primer extension reaction;

b. removing the plurality of nucleic acid template molecules while retaining the plurality of forward extension strands and retaining the plurality of surface primers; and

c. sequencing the plurality of retained forward extension strands.

In some embodiments, the template nucleic acid molecules comprise:

i. two or more copies of the target sequence,

ii. two or more copies of the binding sequence for a forward sequencing primer, and

iii. two or more copies of a binding sequence for a reverse sequencing primer.

In some embodiments, the template nucleic acid molecules comprise binding sequences for an amplification primer, and wherein conducting the primer extension reaction comprises contacting the plurality of template nucleic acid molecules with a plurality of soluble amplification primers, a plurality of nucleotides and a plurality of polymerases, thereby generating a plurality of forward extension strands that are hybridized to template nucleic acid molecules.

In some embodiments, the plurality of amplification primers hybridize to the binding sequences for the amplification primers.

In some embodiments, the amplification primers are soluble.

In some embodiments, the polymerases comprise phi29 DNA polymerases, large fragment of Bst DNA polymerases, large fragment of Bsu DNA polymerases (exo-), Bca DNA polymerases (exo-), Klenow fragment of E. coli DNA polymerases, T5 polymerases, M-MuLV reverse transcriptases, HIV viral reverse transcriptases, Deep Vent DNA polymerases or KOD DNA polymerases.

In some embodiments, the nucleic acid template molecules comprise at least one nucleotide having a scissile moiety that can be cleaved to generate an abasic site.

In some embodiments, the surface primers lack a nucleotide having a scissile moiety.

In some embodiments, the nucleotide having a scissile moiety comprises uridine, 8-oxo-7,8-dihydrogunine, or deoxyinosine.

In some embodiments, removing the nucleic acid template molecules comprises generating abasic sites in the nucleic acid template molecules, followed by generating gaps at the abasic sites.

In some embodiments, the at least one nucleotide having a scissile moiety comprises uracil, and generating abasic sites comprises contacting the nucleic acid template molecules with uracil DNA glycosylase (UDG).

In some embodiments, generating gaps at the abasic sites comprises contacting the abasic sites with an endonuclease IV, AP lyase, FPG glycosylase/AP lyase and/or endo VIII glycosylase/AP lyase.

In some embodiments, individual template nucleic acid molecules comprise nucleic acid template molecules having up to 30% of thymidines replaced with uridine.

In some embodiments, sequencing the plurality of retained forward extension strands generates a plurality of extended reverse sequencing primer strands, wherein individual retained forward extension strands have two or more extended reverse sequencing primer strands hybridized thereon.

In some embodiments, sequencing the plurality of retained forward extension strands generates a plurality of extended reverse sequencing primer strands, wherein individual retained forward extension strands have two or more extended reverse sequencing primer strands hybridized thereon.

In some embodiments, sequencing the plurality of retained forward extension strands comprises a plurality of soluble reverse sequencing primers and (i) a plurality of a first polymerases and a plurality of polymeric molecules and (ii) a plurality of a second polymerases and a plurality of nucleotides or analogs thereof, thereby generating a plurality

of extended reverse sequencing primer strands, wherein individual retained forward extension strands have two or more extended reverse sequencing primer strands hybridized thereon.

In some embodiments, the nucleic acid template molecules comprise one or more copies of a binding sequence for a second surface primer.

In some embodiments, the method comprises a plurality of second surface primers immobilized on the support, whereby binding of the second surface primers to the binding sequence for the second surface primers immobilizes free ends of the plurality nucleic acid template molecules on the support.

In some embodiments, sequencing the plurality of retained forward extension strands comprises:

a. contacting:

(i) the plurality of retained forward extension strands,

(ii) a plurality of reverse sequencing primers comprising a sequence complementary to the binding sequence for the reverse sequencing primer,

(iii) a plurality of first polymerases, and

(iv) a plurality of polymeric molecules of Formula (I):

C-((P)-E)s

(I)

an ionized form thereof, or a salt thereof, wherein:

C is a central moiety;

each P independently is an optionally substituted polymeric side chain;

each E independently is an end moiety; and

s is an integer ranging from 1 to 10;

wherein each polymeric molecule comprises at least two nucleotide moieties and at least one detectable reporter moiety,

wherein the contacting occurs under conditions sufficient to form a plurality of multivalent binding complexes comprising a nucleic acid duplex between a retained forward extension strand and a reverse sequencing primer, a first polymerase, and a nucleotide moiety of a polymeric molecule that is complementary to a nucleotide in the retained forward extension strand immediately adjacent to the 3’ end of the reverse sequencing primer, and

wherein polymerase catalyzed incorporation of a complementary nucleotide moiety into the nucleic acid duplex is inhibited;

b. detecting the detectable reporter moieties; and

c. determining nucleobase identities of nucleotides in the retained forward extension strands complementary to the nucleotide moieties of the polymeric molecules based on the detectable reporter moieties of the polymeric molecules in the plurality of multivalent binding complexes formed in step (a).

In some embodiments, individual retained forward extension strands comprise two or more multivalent binding complexes.

In some embodiments, the plurality of reverse sequencing primers are soluble.

In some embodiments, the method comprises:

d. dissociating the multivalent binding complexes under conditions sufficient to retain the nucleic acid duplexes, thereby generating a plurality of nucleic acid duplexes;

e. contacting the plurality of nucleic acid duplexes with a plurality of second polymerases and a plurality of nucleotides or analogs thereof under conditions sufficient to incorporate nucleotides or analogs thereof complementary to the nucleotides of the retained forward extension strands immediately adjacent to the 3’ ends of the reverse sequencing primers in a primer extension reaction, thereby generating a plurality of extended nucleic acid duplexes comprising extended reverse sequencing primer sequences.

In some embodiments, the method comprises:

g. dissociating the second polymerases from the extended nucleic acid duplexes under conditions sufficient to retain the plurality of extended nucleic acid duplexes.

In some embodiments, the template nucleic acid molecules comprise concatemers of two or more copies of a sequence comprising (i) a sequence for the reverse sequencing primer, (ii) the target nucleic acid sequence, and (iii) a binding sequence for the forward sequencing primer.

In some embodiments, the two or more copies of a sequence complementary to (i) the sequence for the reverse sequencing primer hybridize to the reverse sequencing primers to form nucleic acid duplexes between the retained forward extension strands and the reverse sequencing primers.

In some embodiments, the two or more copies of a sequence complementary to (i) the sequence for the reverse sequencing primer hybridize to the reverse sequencing primers to form nucleic acid duplexes between the retained forward extension strands and the reverse sequencing primers.In some embodiments, the method comprises:

a. contacting the plurality of extended nucleic acid duplexes with a plurality of first polymerases and a plurality of polymeric molecules of Formula (I), or an ionized form thereof, an isomer thereof, or a salt thereof,

wherein the contacting occurs under conditions sufficient to form a plurality of multivalent binding complexes comprising an extended nucleic acid duplex, a first polymerase, and a nucleotide moiety of a polymeric molecule that is complementary to a nucleotide in the retained forward extension strand immediately adjacent to the 3' end of the extended reverse sequencing primer, and

wherein polymerase catalyzed incorporation of a complementary nucleotide moiety into the extended nucleic acid duplex is inhibited;

b. detecting the detectable reporter moieties;

c. determining nucleobase identities of nucleotides in the retained forward extension strands complementary to the nucleotide moieties of the polymeric molecules based on the detectable reporter moieties of the polymeric molecules in the plurality of multivalent binding complexes formed in step (a);

d. dissociating the multivalent binding complexes under conditions sufficient to retain the plurality extended nucleic acid duplexes; and

e. contacting the plurality of extended nucleic acid duplexes with a plurality of second polymerases and a plurality of nucleotides or analogs thereof under conditions sufficient to incorporate nucleotides or analogs thereof complementary to the nucleotides of the nucleic acid template sequences immediately adjacent to the 3' ends of the extended reverse sequencing primers in a primer extension reaction, thereby generating a plurality of extended nucleic acid duplexes comprising extended reverse sequencing primers.

In some embodiments, in an individual polymeric molecule, the at least two nucleotide moieties are attached to different X moieties.

In some embodiments, in an individual polymeric molecule, all detectable reporter moieties are the same and all nucleotide moieties are the same.

In some embodiments, an individual polymeric molecule comprises two, three, or four nucleotide moieties.

In some embodiments, in an individual polymeric molecule, all detectable reporter moieties comprise the same fluorescent.

In some embodiments, all detectable reporter moieties in an individual polymeric molecule label are the same.

In some embodiments, two or more nucleotide moieties in an individual polymeric molecule contact two or more different multivalent binding complexes on the same retained forward extension strand.

In some embodiments, the method comprises repeating steps (a)-(e) at least 1, 10, 20, 30, 40, 50, 70, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 800, 900, 1000, or 1500 times.

In some embodiments, the method comprises repeating steps (a)-(e) until the identities of the nucleotides in sequences of the retained forward extension strands complementary to the target sequences have been determined.

In some embodiments, the method comprises, before step (a), dissociating the second polymerases from the extended nucleic acid duplexes under conditions sufficient to retain the plurality of extended nucleic acid duplexes.

In some embodiments, template nucleic acid molecules comprise concatemers of two or more copies of a sequence comprising:

(i) a binding sequence for a forward sequencing primer,

(ii) sequence complementary to a binding sequence for a reverse sequencing primer,

(iii) a binding sequence for an first surface primer,

(iv) a binding sequence for a second surface primer,

(v) a binding sequence for a first amplification primer,

(vi) a binding sequence for a second amplification primer,

(vii) a binding sequence for a soluble compaction oligonucleotide,

(viii) a sample barcode sequence, and/or

(ix) a unique molecular index sequence.

In some embodiments, in an individual polymeric molecule, each polymeric side chain comprises one or more nucleotide moiety.

In some embodiments, in an individual polymeric molecule, each polymeric side chain comprises one or more detectable reporter moiety.

In some embodiments, in an individual polymeric molecule, each X moiety comprises one or more nucleotide moiety and one or more detectable reporter moiety.

In some embodiments, each individual polymeric molecule comprises two, three, or four detectable reporter moieties and two, three, or four nucleotide moieties.

In some embodiments, in an individual polymeric molecule, each polymeric side chain comprises one or more blocking moiety.

In some embodiments, in an individual polymeric molecule, each polymeric side chain comprises one or more negative charge moieties.

In some embodiments, in an individual polymeric molecule, each polymeric side chain comprises one or more PEG-Cap moieties.

In some embodiments, greater than 90%, greater than 95%, greater than 97%, greater than 98% or greater than 99% of bases have a quality score of Q30.

In some embodiments, greater than 80%, greater than 85%, greater than 87%, greater than 89%, greater than 90%, greater than 91%, greater than 92%, greater than 93%, greater than 94% or greater than 95% of bases have a quality score of Q40.

In an aspect, the present disclosure provides a polymeric molecule of Formula (I):

$$C\text{-}((P)\text{-}E)_s \quad (I)$$

an ionized form thereof, or a salt thereof, wherein:

C is a central moiety;

each P independently is an optionally substituted polymeric side chain;

each E independently is an end moiety; and

s is an integer ranging from 1 to 10.

In some embodiments, the polymeric molecule is of Formula (II):

$$C\text{-}((P)\text{-}E)_4 \quad (II)$$

an ionized form thereof, or a salt thereof.

In some embodiments, the polymeric molecule comprises at least two nucleotide moieties and at least one detectable reporter moiety.

In some embodiments, the at least two nucleotide moieties are attached to different polymeric side chains.

In some embodiments, all the nucleotide moieties are the same.

In some embodiments, all the nucleotide moieties are dATP.

In some embodiments, all the nucleotide moieties are dTTP.

In some embodiments, all the nucleotide moieties are dGTP.

In some embodiments, all the nucleotide moieties are dUTP.

In some embodiments, all the nucleotide moieties are dCTP.

In some embodiments, all the detectable reporter moieties are the same.

In some embodiments, all the detectable reporter moieties are the same and all the nucleotide moieties are the same.

In some embodiments, the molecule comprises two, three, or four detectable reporter moieties.

In some embodiments, the molecule comprises two, three, or four nucleotide moieties.

In some embodiments, the polymeric molecule further comprises one or more blocking moiety.

In some embodiments, the polymeric molecule further comprises one or more negative charge moiety.

In some embodiments, the polymeric molecule further comprises one or more PEG-Cap moiety.

In some embodiments, the present disclosure provides a complex comprising:

a. a polymeric molecule of the disclosure;

b. a polymerase;

c. a template nucleic acid molecule comprising at least one of a target sequence and a binding sequence for a sequencing primer; and

d. a sequence complementary to a portion of the template nucleic acid molecule comprising the sequencing primer sequence;

wherein the template nucleic acid molecule and the sequence complementary to a portion of the template nucleic acid molecule form a duplex, and wherein a nucleotide moiety of the polymeric molecule binds to a complementary to a nucleotide of the template nucleic acid molecule immediately adjacent to the 3' end of the sequence complementary to a portion of the template nucleic acid molecule.

In some embodiments, the polymeric molecule comprises at least two nucleotide moieties, and wherein the at least two nucleotide moieties are the same.

In some embodiments, the template nucleic acid molecule comprises a concatemer comprising at least two copies of a sequence comprising the target sequence and the binding sequence for a sequencing primer.

In some embodiments, at least two complexes form on the same template nucleic acid molecule.

In some embodiments, the at least two nucleotide moieties bind to complementary nucleotides of the template nucleic acid molecule in the at least two complexes.

In some embodiments, at least two complexes form on at least two different template nucleic acid molecules.

In some embodiments, the at least two different template nucleic acid molecules comprise the same target sequence.

In some embodiments, the at least two nucleotide moieties bind to complementary nucleotides of the template nucleic acid molecules in the at least two complexes.

BRIEF DESCRIPTION OF THE DRAWINGS

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

The novel features of the invention are set forth with particularity in the appended claims. A better understanding of the features and advantages of the present invention will be obtained by reference to the following detailed description that sets forth illustrative embodiments, in which the principles of the invention are utilized, and the accompanying drawings of which:

FIG. 1 is a schematic of an exemplary low binding support comprising a glass substrate and alternating layers of hydrophilic coatings which are covalently or non-covalently adhered to the glass, and which further comprises chemically-reactive functional groups that serve as attachment sites for oligonucleotide primers (e.g., capture oligonucleotides). In an alternative embodiment, the support can be made of any material such as glass, plastic or a polymer material.

FIG. 2 is a schematic of an immobilized single template stranded nucleic acid molecule with an exemplary first polymerase and a first polymerase bound to a duplex between a forward primer and the template nucleic acid molecule.

FIG. 3 is a schematic of an immobilized single template stranded nucleic acid molecule with an exemplary multivalent binding complex, which comprises a first polymerase bound to both a duplex between a forward primer and the template nucleic acid molecule, and a nucleotide unit of a polymeric molecule.

FIG. 4 is a schematic of an immobilized single template stranded nucleic acid molecule with an exemplary multivalent binding complex of FIG. 3, wherein two distinct first polymerases are bound to two different portions of the same template nucleic acid molecule, and the two first polymerases are bound to two different nucleotide units of the same polymeric molecule.

FIG. 5 is a schematic of an immobilized single template stranded nucleic acid molecule with an exemplary second polymerase and a second polymerase bound to a duplex between a forward primer and the template nucleic acid molecule.

FIG. 6 is a schematic showing contacting of a plurality of second polymerases bound to a duplex between a forward primer and the template nucleic acid molecule, with a plurality of nucleotides or analogs, under conditions that allow incorporation of the nucleotide.

FIG 7A is a schematic showing an exemplary polymeric molecule comprising a central moiety, four polymeric side chains, and four end moieties.

FIG. 7B is a schematic showing an exemplary polymeric molecule comprising four a central moiety, four polymeric side chains, and four end moieties, wherein each one polymeric side chain is an alternating copolymer, one polymeric side chain is a block copolymer, and one polymeric side chain is a random copolymer.

FIG. 8 is a schematic showing an exemplary polymeric molecule wherein the polymeric side chains are functionalized with blocking moieties, nucleotide moieties, detectable reporter moieties, and negative charge moieties.

FIG. 9 is a schematic showing an alternative exemplary polymeric molecule wherein the polymeric side chains are functionalized with blocking moieties, nucleotide moieties, detectable reporter moieties, and negative charge moieties.

FIG. 10 is a schematic showing an exemplary polymeric molecule wherein the polymeric side chains are functionalized by two blocking groups and two types of functional groups.

FIG. 11 is a schematic showing an exemplary polymeric molecule wherein the polymeric side chains are functionalized with blocking groups, detectable reporter moieties, PEG-Cap moieties, and negative charge moieties.

FIG. 12 is a schematic showing an exemplary single template stranded nucleic acid molecule, which is immobilized to a support using a first surface primer. The immobilized concatemer template molecule comprises at least one nucleotide having a scissile moiety that can be cleaved to generate an abasic site in the immobilized concatemer template molecule. The immobilized concatemer template molecule can be generated by conducting an on-support rolling circle amplification reaction. The arrangement of the various primer binding sequences is for illustration purposes. The skilled artisan will appreciate that many other arrangements are possible. FIGS. 13-16 show an exemplary workflow for pairwise sequencing the immobilized concatemer template molecule depicted in FIG. 12.

FIG. 13 is a schematic showing an exemplary forward sequencing reaction conducted on the immobilized concatemer template molecule shown in FIG. 12. The forward sequencing reaction can be conducted with a plurality of soluble forward sequencing primers and generates

a plurality of extended forward sequencing primer strands hybridized to the template nucleic acid molecule.

FIG. 14 is a schematic showing an exemplary method for replacing the extended forward sequencing primer strands by conducting a primer extension reaction with a strand displacing polymerase in the absence of a soluble primer thereby generating a forward extension strand.

FIG. 15 is a schematic showing an exemplary method for replacing the extended forward sequencing primer strands by conducting a primer extension reaction with a soluble forward sequencing primer thereby generating a forward extension strand.

FIG. 16 is a schematic showing an exemplary method for replacing the extended forward sequencing primer strands by conducting a primer extension reaction with a soluble amplification primer thereby generating a forward extension strand.

FIG. 17 is a schematic showing an exemplary method for generating abasic sites in the template nucleic acid molecules at the nucleotides having the scissile moiety, and generating gaps at the abasic sites to generate a plurality of gap-containing template molecules. The plurality of forward extension strands and first surface primers are retained. The forward extension strand can be generated by the method depicted in FIGS. 14 or 15.

FIG. 18 is a schematic showing an exemplary retained forward extension strand after removal of the gap-containing template molecule as shown in FIG. 17.

FIG. 19 is an exemplary schematic showing an exemplary method for generating abasic sites in the immobilized single stranded concatemer template molecules at the nucleotides having the scissile moiety and generating gaps at the abasic sites to generate a plurality of gap-containing concatemer template molecules while retaining the plurality of forward extension strands and retaining the plurality of immobilized first surface primers. The forward extension strand can be generated by the method depicted in FIG. 16.

FIG. 20 is a schematic showing an exemplary retained forward extension strand after removal of the template molecule as shown in FIG. 19.

FIG. 21 is a schematic showing an exemplary reverse sequencing reaction conducted on the retained forward extension strand shown in FIG. 18. The reverse sequencing reaction can be conducted with a plurality of soluble reverse sequencing primers. The retained forward extension strand can have two or more extended reverse sequencing primer strands hybridized thereon. The extended reverse sequencing primer strands are not hybridized to the first surface primer, or covalently joined to the first surface primer. Therefore, the extended reverse sequencing primer strands are not immobilized to the support. For the sake of simplicity, FIGS.

18-27 show an exemplary immobilized concatemer template nucleic acid molecule with one copy of the target and various primer binding sites. The skilled artisan will appreciate that the immobilized concatemer molecule can include two or more tandem copies containing the target and various primer binding sites.

FIG. 22 is a schematic showing an exemplary reverse sequencing reaction conducted on the retained forward extension strand shown in FIG. 20. The retained forward extension strand can have two or more extended reverse sequencing primer strands hybridized thereon. The extended reverse sequencing primer strands are not hybridized to the first surface primer, or covalently joined to the first surface primer. Therefore, the extended reverse sequencing primer strands are not immobilized to the support.

FIG. 23 is a schematic showing an exemplary support having a first and second surface primer immobilized thereon. A portion of the immobilized concatemer template nucleic acid molecule shown in FIG. 12 is hybridized to the immobilized second surface primer. The immobilized concatemer template molecule has two or more copies of a binding sequence for an immobilized second surface primer. The portion of the immobilized concatemer template molecule that includes the binding sequence for an immobilized second surface primer can hybridize to the immobilized second surface primer.

DETAILED DESCRIPTION

Next generation sequencing (NGS) frequently involves the simultaneous sequencing of a large library of target sequences that have been immobilized on a surface, such as a flow cell surface, and clonally amplified to produce PCR colonies (polonies). The polonies are sequenced in parallel by hybridizing sequencing primers to single stranded template strands in the polonies, followed by successive rounds of hybridizing labeled nucleotides to the templates ("trapping"), determining the identity of the labeled nucleotides, followed incorporating a nucleotide at the position of the labeled nucleotide to extend a strand that is the reverse complement of the template ("stepping"). The process is then repeated until the identities of the nucleotides of the target sequence have been determined.

In some cases, the library of target sequences is amplified in such a manner as to produce PCR colonies (polonies) of concatemerized template molecules which contain multiple copies of the binding sequence for the sequencing primer, the target sequence whose nucleotide identity is to be determined, and optionally, other sequences such as barcodes, which can be used to uniquely identify the source of the target sequence. The disclosure is based, at least in part, on the finding that using polymeric molecules that can simultaneously hybridize

to and label multiple copies of a target sequence in a concatemerized template molecule at the trapping step can increase the accuracy of NGS methods.

Definitions

The headings provided herein are not limitations of the various aspects of the disclosure, which aspects can be understood by reference to the specification as a whole.

Unless defined otherwise, technical and scientific terms used herein have meanings that are commonly understood by those of ordinary skill in the art unless defined otherwise. Generally, terminologies pertaining to techniques of molecular biology, nucleic acid chemistry, protein chemistry, genetics, microbiology, transgenic cell production, and hybridization described herein are those well-known and commonly used in the art. Techniques and procedures described herein are generally performed according to conventional methods well known in the art and as described in various general and more specific references that are cited and discussed throughout the instant specification. For example, see Sambrook et al., Molecular Cloning: A Laboratory Manual (Third ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. 2000). See also Ausubel et al., Current Protocols in Molecular Biology, Greene Publishing Associates (1992). The nomenclatures utilized in connection with, and the laboratory procedures and techniques described herein are those well-known and commonly used in the art.

Unless otherwise required by context herein, singular terms shall include pluralities and plural terms shall include the singular. Singular forms “a”, “an” and “the”, and singular use of any word, include plural referents unless expressly and unequivocally limited on one referent.

It is understood the use of the alternative term (e.g., “or”) is taken to mean either one or both or any combination thereof of the alternatives.

The term “and/or” used herein is to be taken mean specific disclosure of each of the specified features or components with or without the other. For example, the term “and/or” as used in a phrase such as “A and/or B” herein is intended to include: “A and B”; “A or B”; “A” (A alone); and “B” (B alone). In a similar manner, the term “and/or” as used in a phrase such as “A, B, and/or C” is intended to encompass each of the following aspects: “A, B, and C”; “A, B, or C”; “A or C”; “A or B”; “B or C”; “A and B”; “B and C”; “A and C”; “A” (A alone); “B” (B alone); and “C” (C alone).

As used herein and in the appended claims, terms “comprising”, “including”, “having” and “containing”, and their grammatical variants, as used herein are intended to be non-limiting so that one item or multiple items in a list do not exclude other items that can be substituted or

added to the listed items. It is understood that wherever aspects are described herein with the language "comprising," otherwise analogous aspects described in terms of "consisting of" and/or "consisting essentially of" are also provided.

As used herein, the terms "about" and "approximately" refer to a value or composition that is within an acceptable error range for the particular value or composition as determined by one of ordinary skill in the art, which will depend in part on how the value or composition is measured or determined, i.e., the limitations of the measurement system. For example, "about" or "approximately" can mean within one or more than one standard deviation per the practice in the art. Alternatively, "about" or "approximately" can mean a range of up to 10% (i.e., ±10%) or more depending on the limitations of the measurement system. For example, about 5 mg can include any number between 4.5 mg and 5.5 mg. Furthermore, particularly with respect to biological systems or processes, the terms can mean up to an order of magnitude or up to 5-fold of a value. When particular values or compositions are provided in the instant disclosure, unless otherwise stated, the meaning of "about" or "approximately" should be assumed to be within an acceptable error range for that particular value or composition. Also, where ranges and/or subranges of values are provided, the ranges and/or subranges can include the endpoints of the ranges and/or subranges.

As used herein, the term "sequencing" and its variants refers to methods for obtaining sequence information from a nucleic acid strand, typically by determining the ordered identity of at least some nucleotides (including their nucleobase components) within the nucleic acid template molecule. "Sequencing" a given region of a nucleic acid molecule includes identifying each and every nucleotide within the region that is sequenced, as well as methods whereby the identity of only some of the nucleotides in a region are determined, while the identity of some nucleotides remains undetermined or incorrectly determined. Any suitable method of sequencing may be used with the polymeric molecules described herein. Sequencing can include label-free or ion based sequencing methods, as well as labeled or dye-containing nucleotide or fluorescent based nucleotide sequencing methods. Sequencing can include polony-based sequencing or bridge sequencing methods. Sequencing includes massively parallel sequencing platforms that employ sequence-by-synthesis, sequence-by-hybridization or sequence-by-binding procedures. Examples of massively parallel sequence-by-synthesis procedures include polony sequencing, pyrosequencing (e.g., from 454 Life Sciences; U.S. Patent Nos. 7,211,390, 7,244,559 and 7,264,929), chain-terminator sequencing (e.g., from Illumina; U.S. Patent No. 7,566,537; Bentley 2006 Current Opinion Genetics and Development 16:545-552; and Bentley, et al., 2008 Nature 456:53-59, ion-sensitive sequencing (e.g., from

Ion Torrent), probe-anchor ligation sequencing (e.g., Complete Genomics), DNA nanoball sequencing, nanopore DNA sequencing. Examples of single molecule sequencing include Heliscope single molecule sequencing, and single molecule real time (SMRT) sequencing from Pacific Biosciences (Levene, et al., 2003 Science 299(5607):682-686; Eid, et al., 2009 Science 323(5910): 133-138; U.S. patent Nos. 7,170,050; 7,302,146; and 7,405,281). An example of sequence-by-hybridization includes SOLiD sequencing (e.g., from Life Technologies; WO 2006/084132). An example of sequence-by-binding includes Omniome sequencing (e.g., U.S patent No. 10,246,744).

During sequencing by synthesis, each DNA strand in a cluster extends by one base per cycle. A small proportion of strands may become out of phase with the current cycle, either falling a base behind (phasing) or jumping a base ahead (prephasing). The phasing and prephasing rates define the fraction of molecules that become phased or prephased per cycle.

A Quality score (Q score, or Q) is defined by the equation Q = -10log10(e), where e is the estimated probability of the base call being wrong. Higher Q scores indicate a smaller probability of error. Lower Q scores can result in a significant portion of the reads being unusable. They may also lead to increased false-positive variant calls, resulting in inaccurate conclusions. A quality score of Q20 represents an error rate of 1 in 100 (i.e., every 100 bp of sequencing read may contain an error), while a score of Q30 represents an error rate of 1 in 1,000 and a score of Q40 represents an error rate of 1 in 10,000.

A z-score, also referred to as standard score, z-value, Zscore, or normal score, is a dimensionless quantity that is used to indicate the signed, fractional, number of standard deviations by which an event is above the mean value being measured.

A polony (or PCR colony) refers to a population of molecules fixed to a substrate, such as a microscope slide or acrylamide gel, that have been derived through amplification from a single parental molecule. Amplification of a dilute mixture of single template molecules leads to the formation of distinct polonies. Thus, all molecules within a given polony are amplicons of the same molecule, but molecules in two distinct polonies are amplicons of different single molecules.

A “concatemer” refers to a contiguous nucleic acid molecule that contains multiple copies of the same polynucleotide sequence linked in a series. Suitable concatemers can be generated by any suitable methods known in the art, including, but not limited to, rolling circle amplification of circular library molecules comprising adaptor sequences and target sequences. Suitable methods of generating library of concatemer template nucleic acid molecules are described, for example, in WO2022/266470 and WO2023/168444.

The term "polymerase" and its variants, as used herein, comprises any enzyme that can catalyze polymerization of nucleotides (including analogs thereof) into a nucleic acid strand. Typically, but not necessarily, such nucleotide polymerization can occur in a template-dependent fashion. Typically, a polymerase comprises one or more active sites at which nucleotide binding and/or catalysis of nucleotide polymerization can occur. In some embodiments, a polymerase includes other enzymatic activities, such as for example, 3′ to 5′ exonuclease activity or 5′ to 3′ exonuclease activity. In some embodiments, a polymerase has strand displacing activity. A polymerase can include, without limitation, naturally occurring polymerases and any subunits and truncations thereof, mutant polymerases, variant polymerases, recombinant, fusion or otherwise engineered polymerases, chemically modified polymerases, synthetic molecules or assemblies, and any analogs, derivatives or fragments thereof that retain the ability to catalyze nucleotide polymerization (e.g., catalytically active fragment). Polymerases can be isolated from cells, or generated using recombinant DNA technology or chemical synthesis methods. Polymerases can be expressed in prokaryote, eukaryote, viral, or phage organisms. Polymerases can be post-translationally modified proteins or fragments thereof. A polymerase can be derived from a prokaryote, eukaryote, virus or phage. The term "polymerase" encompasses DNA-directed DNA polymerases and RNA-directed DNA polymerases.

As used herein, the term "binding complex" refers to a complex formed by binding together a nucleic acid duplex, a polymerase, and a free nucleotide or a nucleotide moiety (sometimes referred to as nucleotide unit) of a polymeric molecule, where the nucleic acid duplex comprises a template nucleic acid molecule hybridized to a nucleic acid primer. "Multivalent binding complex" refers to a binding complex comprising a nucleic acid duplex, a polymerase, and a nucleotide moiety of a polymeric molecule. Nucleic acid primer sequences, and nucleic acid primer sequences that have undergone one or more rounds of primer extension (or "stepping") reactions, are both encompassed by these terms. In the multivalent complex, the free nucleotide or the nucleotide moiety of the polymeric molecule may or may not be bound at 3' end of the nucleic acid primer at a position that is opposite a complementary nucleotide in the template nucleic acid molecule. A "ternary complex" is an example of a binding complex which is formed by binding together a nucleic acid duplex, a polymerase, and a free nucleotide or nucleotide moiety of a polymeric molecule, where the free nucleotide or nucleotide moiety is bound at the 3' end of the nucleic acid primer or extended primer (as part of the nucleic acid duplex) at a position that is opposite a complementary nucleotide in the template nucleic acid molecule.

As used herein, “avidity complex” refers to a complex in which two or more nucleotide moieties of a polymeric molecule of the disclosure are associated with two or more multivalent binding complexes.

The term “persistence time” and related terms refers to the length of time that a binding complex remains stable without dissociation of any of the components. In some cases, the binding complex is a multivalent binding complex as described herein, and the components of the binding complex include a template nucleic acid molecule or its reverse complement and a nucleic acid primer such as a sequence primer or alternatively an extension product, a polymerase, and a nucleotide moiety of a polymeric molecule or a free (e.g., unconjugated) nucleotide. The nucleotide moiety or the free nucleotide can be complementary or non-complementary to a nucleotide residue in the template nucleic acid molecule. The nucleotide moiety or the free nucleotide can bind to the 3’ end of the nucleic acid primer (or extended primer) at a position that is opposite a complementary nucleotide in the template nucleic acid molecule. The persistence time is indicative of the stability of the binding complex and strength of the binding interactions. Persistence time can be measured by observing the onset and/or duration of a binding complex, such as by observing a signal from a labeled component of the binding complex. For example, a labeled nucleotide or a labeled reagent comprising one or more nucleotides may be present in a binding complex, thus allowing the signal from the label to be detected during the persistence time of the binding complex. One exemplary label is a fluorescent label. The binding complex (e.g., ternary complex) remains stable until subjected to a condition that causes dissociation of interactions between any of the polymerase, template molecule, primer and/or the nucleotide moiety or the free nucleotide. For example, a dissociating condition comprises contacting the binding complex with any one or any combination of a detergent, EDTA and/or water.

The terms “nucleic acid”, "polynucleotide" and "oligonucleotide" and other related terms used herein are used interchangeably and refer to polymers of nucleotides and are not limited to any particular length. Nucleic acids include recombinant and chemically-synthesized forms. Nucleic acids include DNA molecules (e.g., cDNA or genomic DNA), RNA molecules (e.g., mRNA), analogs of the DNA or RNA generated using nucleotide analogs (e.g., peptide nucleic acids and non-naturally occurring nucleotide analogs), and chimeric forms containing DNA and RNA. Nucleic acids can be single-stranded or double-stranded. Nucleic acids comprise polymers of nucleotides, where the nucleotides include natural or non-natural bases and/or sugars. Nucleic acids comprise naturally-occurring internucleosidic linkages, for example phosphodiester linkages. Nucleic acids comprise non-natural internucleoside

linkages, including phosphorothioate, phosphorothiolate, or peptide nucleic acid (PNA) linkages. In some embodiments, nucleic acids comprise a one type of polynucleotide or a mixture of two or more different types of polynucleotides.

The term "primer" and related terms used herein refers to an oligonucleotide, either natural or synthetic, that is capable of hybridizing with a DNA and/or RNA polynucleotide template to form a duplex (double stranded) molecule. Primers may have any length, but typically range from 4-50 nucleotides. A typical primer comprises a 5' end and 3' end. The 3' end of the primer can include a 3' OH moiety which serves as a nucleotide polymerization initiation site in a polymerase-mediated primer extension reaction. Alternatively, the 3' end of the primer can lack a 3' OH moiety, or can include a terminal 3' blocking group that inhibits nucleotide polymerization in a polymerase-mediated reaction. Any one nucleotide, or more than one nucleotide, along the length of the primer can be labeled with a detectable reporter moiety. A primer can be in solution (e.g., a soluble primer) or can be immobilized on a support (e.g., a surface or capture primer).

The term "target sequence" or "target polynucleotide", sometimes also referred to herein as "sequence of interest" refers to a sequence whose nucleotide identity is to be determined by the sequencing methods described herein. The term "template nucleic acid", "template polynucleotide", "template strand" and other variations refer to a nucleic acid strand that serves as the basis nucleic acid molecule for generating a complementary nucleic acid strand. The template nucleic acid can be single-stranded or double-stranded, or the template nucleic acid can have single-stranded or double-stranded portions. The sequence of the template nucleic acid can be partially or wholly complementary to the sequence of the complementary strand. The template nucleic acid can be obtained from a naturally-occurring source, recombinant form, or chemically synthesized to include any type of nucleic acid analog. The template nucleic acid can be linear, circular, or other forms. The template sequence can be single-stranded, or double-stranded, for example a single-stranded DNA molecule. Template nucleic acid molecules of the disclosure can include an insert region having an insert sequence comprising the target sequence. In addition, the template nucleic acids can also include at least one adaptor sequence, such as an adaptor sequencing comprising a primer binding sequence. The template nucleic acid can be a concatemer having two or tandem copies of a target sequence and at least one adaptor sequence. The target sequence can be isolated in any form, including chromosomal, genomic, organellar (e.g., mitochondrial, chloroplast or ribosomal), recombinant molecules, cloned, amplified, cDNA, RNA such as precursor mRNA or mRNA, oligonucleotides, whole genomic DNA, obtained from fresh frozen paraffin embedded tissue,

needle biopsies, cell free circulating DNA, or any type of nucleic acid library. The target sequence can be isolated from any source including from organisms such as prokaryotes, eukaryotes (e.g., humans, plants and animals), fungus, viruses cells, tissues, normal or diseased cells or tissues, body fluids including blood, urine, serum, lymph, tumor, saliva, anal and vaginal secretions, amniotic samples, perspiration, semen, environmental samples, culture samples, or synthesized nucleic acid molecules prepared using recombinant molecular biology or chemical synthesis methods. The target sequence can be isolated from any organ, including head, neck, brain, breast, ovary, cervix, colon, rectum, endometrium, gallbladder, intestines, bladder, prostate, testicles, liver, lung, kidney, esophagus, pancreas, thyroid, pituitary, thymus, skin, heart, larynx, or other organs. The template nucleic acid can be subjected to nucleic acid analysis, including sequencing and composition analysis.

When used in reference to nucleic acid molecules, the terms “hybridize” or “hybridizing” or “hybridization” or other related terms refers to hydrogen bonding between two different nucleic acids to form a duplex (double-stranded) nucleic acid. Hybridization also includes hydrogen bonding between two different regions of a single nucleic acid molecule to form a self-hybridizing molecule having a duplex region. Hybridization can comprise Watson-Crick or Hoogstein binding to form a duplex double-stranded nucleic acid, or a double-stranded region within a nucleic acid molecule. The double-stranded nucleic acid, or the two different regions of a single nucleic acid, may be wholly complementary, or partially complementary. Complementary nucleic acid strands need not hybridize with each other across their entire length. The complementary base pairing can be the standard A-T or C-G base pairing, or can be other forms of base-pairing interactions. Duplex nucleic acids can include mismatched base-paired nucleotides.

The term “nucleotides” and related terms refers to a molecule comprising an aromatic base, a five carbon sugar (e.g., ribose or deoxyribose), and at least one phosphate group. Canonical and non-canonical nucleotides are consistent with use of the term. The phosphate, in some cases, comprises a monophosphate, diphosphate, or triphosphate, or corresponding phosphate analog. In some cases, the nucleotide comprises 1, 2, 3, 4, 5, 6, 7, 8, 9 or 10 phosphate groups. The term “nucleoside” refers to a molecule comprising an aromatic base and a sugar.

Nucleotides (and nucleosides) typically comprise a heterocyclic base including a substituted or unsubstituted nitrogen-containing parent heteroaromatic ring which are commonly found in nucleic acids, including naturally-occurring, substituted, modified, or engineered variants, or analogs of the same. The base of a nucleotide (or nucleoside) is capable

of forming Watson-Crick and/or Hoogstein hydrogen bonds with an appropriate complementary base. Exemplary bases include, but are not limited to, purines and pyrimidines such as: 2-aminopurine, 2,6-diaminopurine, adenine (A), ethenoadenine, N^6-Δ^2-isopentenyladenine (6iA), N^6-Δ^2-isopentenyl-2-methylthioadenine (2ms6iA), N^6-methyladenine, guanine (G), isoguanine, N^2-dimethylguanine (dmG), 7-methylguanine (7mG), 2-thiopyrimidine, 6-thioguanine (6sG), hypoxanthine and O^6-methylguanine; 7-deaza-purines such as 7-deazaadenine (7-deaza-A) and 7-deazaguanine (7-deaza-G); pyrimidines such as cytosine (C), 5-propynylcytosine, isocytosine, thymine (T), 4-thiothymine (4sT), 5,6-dihydrothymine, O^4-methylthymine, uracil (U), 4-thiouracil (4sU) and 5,6-dihydrouracil (dihydrouracil; D); indoles such as nitroindole and 4-methylindole; pyrroles such as nitropyrrole; nebularine; inosines; hydroxymethylcytosines; 5-methycytosines; base (Y); as well as methylated, glycosylated, and acylated base moieties; and the like. Additional exemplary bases can be found in Fasman, 1989, in "Practical Handbook of Biochemistry and Molecular Biology", pp. 385-394, CRC Press, Boca Raton, Fla.

Nucleotides (and nucleosides) typically comprise a sugar moiety, such as carbocyclic moiety (Ferraro and Gotor 2000 Chem. Rev. 100: 4319-48), acyclic moieties (Martinez, et al., 1999 Nucleic Acids Research 27: 1271-1274; Martinez, et al., 1997 Bioorganic & Medicinal Chemistry Letters vol. 7: 3013-3016), and other sugar moieties (Joeng, et al., 1993 J. Med. Chem. 36: 2627-2638; Kim, et al., 1993 J. Med. Chem. 36: 30-7; Eschenmosser 1999 Science 284:2118-2124; and U.S. Pat. No. 5,558,991). Exemplary sugar moieties comprise ribosyl; 2′-deoxyribosyl; 3′-deoxyribosyl; 2′,3′-dideoxyribosyl; 2′,3′-didehydrodideoxyribosyl; 2′-alkoxyribosyl; 2′-azidoribosyl; 2′-aminoribosyl; 2′-fluororibosyl; 2′-mercaptoriboxyl; 2′-alkylthioribosyl; 3′-alkoxyribosyl; 3′-azidoribosyl; 3′-aminoribosyl; 3′-fluororibosyl; 3′-mercaptoriboxyl; 3′-alkylthioribosyl carbocyclic; acyclic or other modified sugars.

In some cases, nucleotides comprise a chain of one, two or three phosphorus atoms. The chain is typically attached to the 5' carbon of the sugar moiety via an ester or phosphoramide linkage. In some cases, the nucleotide is an analog having a phosphorus chain in which the phosphorus atoms are linked together with intervening O, S, NH, methylene or ethylene. The phosphorus atoms in the chain can include substituted side groups, including O, S or BH_3. In some cases, the chain includes phosphate groups substituted with analogs including phosphoramidate, phosphorothioate, phosphordithioate, and O-methylphosphoroamidite groups.

When used in reference to nucleic acids, the terms "extend", "extending", "extension" and other variants, refers to incorporation of one or more nucleotides into a nucleic acid

molecule. Nucleotide incorporation comprises polymerization of one or more nucleotides into the terminal 3' OH end of a nucleic acid strand, resulting in extension of the nucleic acid strand. Nucleotide incorporation can be conducted with natural nucleotides and/or nucleotide analogs. Typically, but not necessarily, nucleotide incorporation occurs in a template-dependent fashion. Any suitable method of extending a nucleic acid molecule may be used, including primer extension catalyzed by a DNA polymerase or RNA polymerase.

The terms "reporter moiety", "reporter moieties", "detectable reporter moieties," and related terms refer to a compound that generates, or causes to generate, a detectable signal. A reporter moiety is sometimes called a "label." Any suitable reporter moiety may be used, including luminescent, photoluminescent, electroluminescent, bioluminescent, chemiluminescent, fluorescent, phosphorescent, chromophore, radioisotope, electrochemical, mass spectrometry, Raman, hapten, affinity tag, atom, or enzymatic. A reporter moiety generates a detectable signal resulting from a chemical or physical change (e.g., heat, light, electrical, pH, salt concentration, enzymatic activity, or proximity events). A proximity event includes two reporter moieties approaching each other, or associating with each other, or binding each other. It is well known to one skilled in the art to select multiple reporter moieties so that each absorbs excitation radiation and/or emits fluorescence at a wavelength distinguishable from the other reporter moieties to permit monitoring of the presence of different reporter moieties in the same reaction, or in different reactions. Two or more different reporter moieties can be selected having spectrally distinct emission profiles, or having minimal overlapping spectral emission profiles. Reporter moieties can be linked (e.g., operably linked) to the polymeric molecules described herein, as well as nucleotides, nucleosides, nucleic acids, enzymes (e.g., polymerases or reverse transcriptases), or supports (e.g., surfaces) by any suitable methods.

Detectable reporter moieties, or labels, may include fluorescent labels and/or fluorophores. Exemplary fluorescent moieties which may serve as fluorescent labels or fluorophores include, but are not limited to fluorescein and fluorescein derivatives such as carboxyfluorescein, tetrachlorofluorescein, hexachlorofluorescein, carboxynapthofluorescein, fluorescein isothiocyanate, NHS-fluorescein, iodoacetamidofluorescein, fluorescein maleimide, SAMSA-fluorescein, fluorescein thiosemicarbazide, carbohydrazinomethylthioacetyl-amino fluorescein, rhodamine and rhodamine derivatives such as TRITC, TMR, lissamine rhodamine, Texas Red, rhodamine B, rhodamine 6G, rhodamine 10, NHS-rhodamine, TMR-iodoacetamide, lissamine rhodamine B sulfonyl chloride, lissamine rhodamine B sulfonyl hydrazine, Texas Red sulfonyl chloride, Texas Red

hydrazide, coumarin and coumarin derivatives such as AMCA, AMCA-NHS, AMCA-sulfo-NHS, AMCA-HPDP, DCIA, AMCE-hydrazide, BODIPY™ and derivatives such as BODIPY™ FL C3-SE, BODIPY™ 530/550 C3, BODIPY™ 530/550 C3-SE, BODIPY™ 530/550 C3 hydrazide, BODIPY™ 493/503 C3 hydrazide, BODIPY™ FL C3 hydrazide, BODIPY™ FL IA, BODIPY™ 530/551 IA, Br-BODIPY™ 493/503, Cascade Blue™ and derivatives such as Cascade Blue™ acetyl azide, Cascade Blue™ cadaverine, Cascade Blu™e ethylenediamine, Cascade Blue™ hydrazide, Lucifer Yellow and derivatives such as Lucifer Yellow iodoacetamide, Lucifer Yellow CH, cyanine and derivatives such as indolium based cyanine dyes, benzo-indolium based cyanine dyes, pyridium based cyanine dyes, thiozolium based cyanine dyes, quinolinium based cyanine dyes, imidazolium based cyanine dyes, Cy 3, Cy5, lanthanide chelates and derivatives such as BCPDA, TBP, TMT, BHHCT, BCOT, Europium chelates, Terbium chelates, Alexa Fluor® dyes, DyLight™ dyes, Atto dyes, LightCycler® Red dyes, CAL Flour dyes, JOE and derivatives thereof, Oregon Green dyes, WellRED dyes, IRD dyes, phycoerythrin and phycobilin dyes, Malachite green, stilbene, DEG dyes, NR dyes, near-infrared dyes and others known in the art such as those described in Haugland, Molecular Probes Handbook, (Eugene, Oreg.) 6th Edition; Lakowicz, Principles of Fluorescence Spectroscopy, 2nd Ed., Plenum Press New York (1999), or Hermanson, Bioconjugate Techniques, 2nd Edition, or derivatives thereof, or any combination thereof. Cyanine dyes may exist in either sulfonated or non-sulfonated forms, and consist of two indolenin, benzo-indolium, pyridium, thiozolium, and/or quinolinium groups separated by a polymethine bridge between two nitrogen atoms. Commercially available cyanine fluorophores include, for example, Cy3, (which may comprise 1-[6-(2,5-dioxopyrrolidin-1-yloxy)-6-oxohexyl]-2-(3-{1-[6-(2,5-dioxopyrrolidin-1-yloxy)-6-oxohexyl]-3,3-dimethyl-1,3-dihydro-2H-indol-2-ylidene}prop-1-en-1-yl)-3,3-dimethyl-3H-indolium or 1-[6-(2,5-dioxopyrrolidin-1-yloxy)-6-oxohexyl]-2-(3-{1-[6-(2,5-dioxopyrrolidin-1-yloxy)-6-oxohexyl]-3,3-dimethyl-5-sulfo-1,3-dihydro-2H-indol-2-ylidene}prop-1-en-1-yl)-3,3-dimethyl-3H-indolium-5-sulfonate), Cy5 (which may comprise 1-(6-((2,5-dioxopyrrolidin-1-yl)oxy)-6-oxohexyl)-2-((1E,3E)-5-((E)-1-(6-((2,5-dioxopyrrolidin-1-yl)oxy)-6-oxohexyl)-3,3-dimethyl-5-indolin-2-ylidene)penta-1,3-dien-1-yl)-3,3-dimethyl-3H-indol-1-ium or 1-(6-((2,5-dioxopyrrolidin-1-yl)oxy)-6-oxohexyl)-2-((1E,3E)-5-((E)-1-(6-((2,5-dioxopyrrolidin-1-yl)oxy)-6-oxohexyl)-3,3-dimethyl-5-sulfoindolin-2-ylidene)penta-1,3-dien-1-yl)-3,3-dimethyl-3H-indol-1-ium-5-sulfonate), and Cy7 (which may comprise 1-(5-carboxypentyl)-2-[(1E,3E,5E,7Z)-7-(1-ethyl-1,3-dihydro-2H-indol-2-ylidene)hepta-1,3,5-trien-1-yl]-3H-indolium or 1-(5-carboxypentyl)-2-[(1E,3E,5E,7Z)-7-(1-ethyl-5-sulfo-1,3-dihydro-2H-indol-2-ylidene)hepta-1,3,5-trien-1-yl]-

3H-indolium-5-sulfonate), where “Cy” stands for 'cyanine', and the first digit identifies the number of carbon atoms between two indolenine groups. Cy2 which is an oxazole derivative rather than indolenin, and the benzo-derivatized Cy3.5, Cy5.5 and Cy7.5 are exceptions to this rule.

Detectable reporter moieties may include fluorescence resonance energy transfer (FRET) pairs, such that multiple classifications can be performed under a single excitation and imaging step. As used herein, FRET may comprise excitation exchange (Forster) transfers, or electron-exchange (Dexter) transfers.

The terms “linked”, “joined”, “attached”, and variants thereof comprise any type of fusion, bond, adherence or association between any combination of compounds or molecules that is of sufficient stability to withstand use in the particular procedure. The procedure can include but are not limited to: nucleotide transient-binding; nucleotide incorporation; de-blocking; washing; removing; flowing; detecting; imaging and/or identifying. Such linkage can comprise, for example, covalent, ionic, hydrogen, dipole-dipole, hydrophilic, hydrophobic, or affinity bonding, bonds or associations involving van der Waals forces, mechanical bonding, and the like. Linkage can occur intramolecularly, for example linking together the ends of a single-stranded or double-stranded linear nucleic acid molecule to form a circular molecule. Linkage can also occur between a combination of different molecules, or between a molecule and a non-molecule, including but not limited to: linkage between a nucleic acid molecule and a solid surface; linkage between a protein and a detectable reporter moiety; linkage between a nucleotide and detectable reporter moiety; and the like. Some examples of linkages can be found, for example, in Hermanson, G., “Bioconjugate Techniques”, Second Edition (2008); Aslam, M., Dent, A., “Bioconjugation: Protein Coupling Techniques for the Biomedical Sciences”, London: Macmillan (1998); Aslam, M., Dent, A., “Bioconjugation: Protein Coupling Techniques for the Biomedical Sciences”, London: Macmillan (1998).

The term “adaptor” and related terms refers to oligonucleotides that can be operably linked (appended) to a target polynucleotide, where the adaptor confers a function to the co-joined adaptor-target molecule. Adaptors comprise DNA, RNA, chimeric DNA/RNA, or analogs thereof. Adaptors can include at least one ribonucleoside residue. Adaptors can be single-stranded, double-stranded, or have single-stranded and/or double-stranded portions. Adaptors can be configured to be linear, stem-looped, hairpin, or Y-shaped forms. Adaptors can be any length, including 4-100 nucleotides or longer. Adaptors can have blunt ends, overhang ends, or a combination of both. Overhang ends include 5’ overhang and 3’ overhang ends. The 5’ end of a single-stranded adaptor, or one strand of a double-stranded adaptor, can

have a 5' phosphate group or lack a 5' phosphate group. Adaptors can include a 5' tail that does not hybridize to a target polynucleotide (e.g., tailed adaptor), or adaptors can be non-tailed. An adaptor can include a sequence that is complementary to at least a portion of a primer, such as an amplification primer, a sequencing primer, or a capture primer (e.g., soluble or immobilized capture primers). Adaptors can include a random sequence or degenerate sequence. Adaptors can include at least one inosine residue. Adaptors can include at least one phosphorothioate, phosphorothiolate and/or phosphoramidate linkage. Adaptors can include a barcode sequence which can be used to distinguish polynucleotides (e.g., target sequences) from different sample sources in a multiplex assay such as high throughput sequencing. Adaptors can include a unique identification sequence (e.g., unique molecular index, UMI; or a unique molecular tag) that can be used to uniquely identify a nucleic acid molecule to which the adaptor is appended. In some cases, a unique identification sequence can be used to increase error correction and accuracy, reduce the rate of false-positive variant calls and/or increase sensitivity of variant detection. Adaptors can include at least one restriction enzyme recognition sequence, including any one or any combination of two or more selected from a group consisting of type I, type II, type III, type IV, type Hs or type IIB.

The term "universal sequence", "universal adaptor sequences" and related terms refers to a sequence in a nucleic acid molecule that is common among two or more polynucleotide molecules, for example sequences shared amongst nucleic acid molecules in a library. For example, adaptors having the same universal sequence can be joined to a plurality of polynucleotides so that the population of co-joined molecules carry the same universal adaptor sequence. Examples of universal adaptor sequences include binding sequences amplification primers, sequencing primersor capture primers (e.g., soluble or support-immobilized capture primers).

As used herein, "support" refers to a substrate that is solid, semi-solid, or a combination of both, to which a plurality of nucleic acid molecules (e.g. template nucleic acid molecules and/or capture primers) can be affixed. The support can be porous, semi-porous, non-porous, or any combination of porosity. The support can be substantially planar, concave, convex, or any combination thereof. The support can be cylindrical, for example comprising a capillary or interior surface of a capillary. The support can be a surface of a flow cell. The surface of the support can be substantially smooth. Alternatively, the support can be regularly or irregularly textured, including bumps, etched, pores, three-dimensional scaffolds, or any combination thereof. The term "support" also encompasses beads having any shape, including spherical,

hemi-spherical, cylindrical, barrel-shaped, toroidal, disc-shaped, rod-like, conical, triangular, cubical, polygonal, tubular or wire-like.

The support can be fabricated from any material, including but not limited to glass, fused-silica, silicon, a polymer (e.g., polystyrene (PS), macroporous polystyrene (MPPS), polymethylmethacrylate (PMMA), polycarbonate (PC), polypropylene (PP), polyethylene (PE), high density polyethylene (HDPE), cyclic olefin polymers (COP), cyclic olefin copolymers (COC), polyethylene terephthalate (PET)), or any combination thereof. Various compositions of both glass and plastic substrates are contemplated.

The surface of the support can be coated with one or more compounds to produce a passivated layer on the support. Supports comprising a low non-specific binding surface that enable improved nucleic acid hybridization and amplification performance on the support are envisaged as within the scope of the instant disclosure. In general, the support may comprise one or more layers of a covalently or non-covalently attached low-binding, chemical modification layers, e.g., silane layers, polymer films, and one or more covalently or non-covalently attached oligonucleotides that may be used for immobilizing a plurality of nucleic acid template molecules to the support.

The degree of hydrophilicity (or "wettability" with aqueous solutions) of the surface coatings of the support may be assessed, for example, through the measurement of water contact angles in which a small droplet of water is placed on the surface and its angle of contact with the surface is measured using, e.g., an optical tensiometer. In some cases, a static contact angle may be determined. In some cases, an advancing or receding contact angle may be determined. The water contact angle for the hydrophilic, low-binding support surfaced disclosed herein may range from about 0 degrees to about 30 degrees, for example the water contact angle for the hydrophilic, low-binding support surface disclosed herein may no more than 50 degrees, 40 degrees, 30 degrees, 25 degrees, 20 degrees, 18 degrees, 16 degrees, 14 degrees, 12 degrees, 10 degrees, 8 degrees, 6 degrees, 4 degrees, 2 degrees, or 1 degree. In many cases the contact angle is no more than 40 degrees. Those of skill in the art will realize that a given hydrophilic, low-binding support surface of the present disclosure may exhibit a water contact angle having a value of anywhere within this range.

The present disclosure provides a plurality (e.g., two or more) of template nucleic acid molecules immobilized to a support. The immobilized plurality of template nucleic acid molecules have the same sequence or have different sequences (e.g., template nucleic acid molecules in the plurality have different target sequences). Individual nucleic acid template molecules in the plurality of nucleic acid templates can be immobilized to a different site on

the support, for example in an array. The term "array" refers to a support comprising a plurality of sites located at pre-determined locations on the support to form an array of sites. The sites can be discrete and separated by interstitial regions. The pre-determined sites on the support can be arranged in one dimension in a row or a column, or arranged in two dimensions in rows and columns. The plurality of pre-determined sites can be arranged on the support in an organized fashion, for example in any organized pattern, including rectilinear, hexagonal patterns, grid patterns, patterns having reflective symmetry, patterns having rotational symmetry, or the like. The pitch between different pairs of sites can be that same or can vary. The support can have template nucleic acid molecules immobilized at a plurality of sites at a surface density of about 10^2 – 10^{15} sites per mm^2, or more, to form a template nucleic acid array. For example, the support comprises at least 10^2 sites, at least 10^3 sites, at least 10^4 sites, at least 10^5 sites, at least 10^6 sites, at least 10^7 sites, at least 10^8 sites, at least 10^9 sites, at least 10^{10} sites, at least 10^{11} sites, at least 10^{12} sites, at least 10^{13} sites, at least 10^{14} sites, at least 10^{15} sites, or more, where the sites are located at pre-determined locations on the support. In some cases, a plurality of pre-determined sites on the support (e.g., 10^2 – 10^{15} sites or more) are immobilized with nucleic acid templates to form a nucleic acid template array. The nucleic acid templates can be immobilized at a plurality of pre-determined sites by any methods known in the art, including hybridization to immobilized surface capture primers, or covalent attachment to immobilized surface capture primers. In some cases, the template nucleic acid molecules are immobilized at a plurality of pre-determined sites, for example 10^2 – 10^{15} sites or more. The template nucleic acid molecules that are immobilized at a plurality of sites on the support can comprise linear or circular molecules, or a mixture of both linear and circular molecules. The immobilized template nucleic acid molecules can be clonally-amplified to generate immobilized polonies at the plurality of pre-determined sites. In some embodiments, individual immobilized template nucleic acid molecules comprise one copy of a target sequence of interest, or comprise concatemers having two or more tandem copies of a target sequence of interest. Optionally, the concatemerized sequence can include additional sequences, such as binding sequences for sequencing primers, amplification primers, and/or barcodes.

Alternatively, a support can comprise a plurality of sites located at random locations (referred to herein as a support having randomly located sites thereon). In a support having randomly located sites, the locations of the sites on the support are not pre-determined. The plurality of randomly-located sites is arranged on the support in a disordered and/or unpredictable fashion. A support with randomly located sites can comprise at least 10^2 sites, at

least 10^3 sites, at least 10^4 sites, at least 10^5 sites, at least 10^6 sites, at least 10^7 sites, at least 10^8 sites, at least 10^9 sites, at least 10^{10} sites, at least 10^{11} sites, at least 10^{12} sites, at least 10^{13} sites, at least 10^{14} sites, at least 10^{15} sites, or more. In some cases, a plurality of template nucleic acid molecules are randomly located on the support (e.g., at 10^2 – 10^{15} sites or more) and are immobilized to form a support comprising immobilized template nucleic acid molecules. Template nucleic acid molecules can be immobilized randomly on a support by hybridization to immobilized, randomly-located surface capture primers, or by covalently attached to the surface capture primers. The template nucleic acid molecules can be immobilized at a plurality of randomly located sites on the support, for example immobilized at 10^2 – 10^{15} sites or more. The nucleic acid templates that are immobilized at a plurality of sites on the support can comprise linear or circular molecules, or a mixture of both linear and circular molecules. The immobilized template nucleic acid molecules can be clonally-amplified to generate immobilized nucleic acid polonies at the plurality of randomly located sites. In some embodiments, individual immobilized template nucleic acid molecules comprise one copy of a target sequence of interest, or comprise concatemers having two or more tandem copies of a target sequence of interest. Optionally, the concatemerized sequence can include additional sequences, such as binding sequences for sequencing primers, amplification primers, and/or barcodes.

In general, during NGS methods described herein, nucleic acid molecules immobilized to the support are in fluid communication with each other to permit flowing a solution of reagents (e.g., enzymes including polymerases, polymeric molecules, nucleotides, divalent cations and/or buffers and the like) onto the support so that the plurality nucleic acid molecules on the support can be reacted with the reagents in a massively parallel manner. For example, a library of template nucleic acid molecules, or the reverse complements thereof, depending upon the sequencing protocol, can be immobilized to a surface of a flow cell, and are in fluid communication with each other as reagents are flowed into and out of the flow cell. The fluid communication of the plurality of immobilized nucleic acid molecules can be used to conduct nucleotide binding assays and/or conduct nucleotide polymerization reactions (e.g., primer extension or sequencing) on the plurality of immobilized nucleic acid template molecules, and to conduct detection and imaging for massively parallel sequencing.

The term “immobilized” and related terms, refers to nucleic acid molecules or enzymes (e.g., polymerases) that are attached to the support. The attachment can be directly to the support through covalent bond or non-covalent interaction, or indirectly to a coating on the support.

As used herein, the term "clonally amplified" and it variants refers to a nucleic acid molecule that has been subjected to one or more amplification reactions either in-solution or on-support. In the case of in-solution amplified template nucleic acid molecules, the resulting amplicons are distributed onto the support. Prior to amplification, the template nucleic acid molecule comprises a target sequence and at least one universal adaptor sequence (e.g., an adaptor sequence comprising binding sequence for a forward sequencing primer). Clonal amplification comprises the use of a polymerase chain reaction (PCR), multiple displacement amplification (MDA), transcription-mediated amplification (TMA), nucleic acid sequence-based amplification (NASBA), strand displacement amplification (SDA), real-time SDA, bridge amplification, isothermal bridge amplification, rolling circle amplification (RCA), circle-to-circle amplification, helicase-dependent amplification, recombinase-dependent amplification, single-stranded binding (SSB) protein-dependent amplification, or any combination thereof.

As used herein, "alkyl", "C_1, C_2, C_3, C_4, C_5 C_6 C_7 C_8 C_9 C_{10} C_{11} C_{12} C_{13} C_{14} C_{15} C_{16} C_{17} C_{18} C_{19} C_{20} alkyl" or "C_1-C_{20} alkyl" is intended to include C_1, C_2, C_3, C_4, C_5, C_6, C_7, C_8, C_9, C_{10}, C_{11}, C_{12}, C_{13}, C_{14}, C_{15}, C_{16}, C_{17}, C_{18}, C_{19}, or C_{20} straight chain (linear) saturated aliphatic hydrocarbon groups and C_3, C_4, C_5, C_6, C_7, C_8, C_9, C_{10}, C_{11}, C_{12}, C_{13}, C_{14}, C_{15}, C_{16}, C_{17}, C_{18}, C_{19}, or C_{20} branched saturated aliphatic hydrocarbon groups. For example, C_1-C_6 alkyl is intends to include C_1, C_2, C_3, C_4, C_5 and C_6 alkyl groups. Examples of alkyl include, moieties having from one to six carbon atoms, such as, but not limited to, methyl, ethyl, n-propyl, i-propyl, n-butyl, s-butyl, t-butyl, n-pentyl, i-pentyl, or n-hexyl. In some embodiments, a straight chain or branched alkyl has twenty or fewer carbon atoms (*e.g.*, C_1-C_{20} for straight chain, C_3-C_{20} for branched chain). In some embodiments, a straight chain or branched alkyl has six or fewer carbon atoms (*e.g.*, C_1-C_6 for straight chain, C_3-C_6 for branched chain), and in another embodiment, a straight chain or branched alkyl has four or fewer carbon atoms. The term "alkylene" refers to a multivalent alkyl group, e.g., a bivalent, trivalent, or tetravalent alkyl group.

As used herein, the term "optionally substituted alkyl" refers to unsubstituted alkyl or alkyl having designated substituents replacing one or more hydrogen atoms on one or more carbons of the hydrocarbon backbone. Such substituents can include, for example, alkyl, alkenyl, alkynyl, halogen, hydroxyl, alkylcarbonyloxy, arylcarbonyloxy, alkoxycarbonyloxy, aryloxycarbonyloxy, carboxylate, alkylcarbonyl, arylcarbonyl, alkoxycarbonyl, aminocarbonyl, alkylaminocarbonyl, dialkylaminocarbonyl, alkylthiocarbonyl, alkoxyl,

phosphate, phosphonato, phosphinato, amino (including alkylamino, dialkylamino, arylamino, diarylamino and alkylarylamino), acylamino (including alkylcarbonylamino, arylcarbonylamino, carbamoyl and ureido), amidino, imino, sulfhydryl, alkylthio, arylthio, thiocarboxylate, sulfates, alkylsulfinyl, sulfonato, sulfamoyl, sulfonamido, nitro, trifluoromethyl, cyano, azido, heterocyclyl, alkylaryl, or an aromatic or heteroaromatic moiety.

As used herein, the term "alkenyl" includes unsaturated aliphatic groups analogous in length and possible substitution to the alkyls described above, but that contain at least one double bond. For example, the term "alkenyl" includes straight chain alkenyl groups (*e.g.*, ethenyl, propenyl, butenyl, pentenyl, hexenyl, heptenyl, octenyl, nonenyl, decenyl), and branched alkenyl groups. In certain embodiments, a straight chain or branched alkenyl group has twenty or fewer carbon atoms in its backbone (*e.g.*, C_2-C_{20} for straight chain, C_3-C_{20} for branched chain). In certain embodiments, a straight chain or branched alkenyl group has six or fewer carbon atoms in its backbone (*e.g.*, C_2-C_6 for straight chain, C_3-C_6 for branched chain). The term "C_2-C_6" includes alkenyl groups containing two to six carbon atoms. The term "C_3-C_6" includes alkenyl groups containing three to six carbon atoms. The term "alkenylene" refers to a multivalent (e.g., bivalent, trivalent, tetravalent) alkyl group.

As used herein, the term "optionally substituted alkenyl" refers to unsubstituted alkenyl or alkenyl having designated substituents replacing one or more hydrogen atoms on one or more hydrocarbon backbone carbon atoms. Such substituents can include, for example, alkyl, alkenyl, alkynyl, halogen, hydroxyl, alkylcarbonyloxy, arylcarbonyloxy, alkoxycarbonyloxy, aryloxycarbonyloxy, carboxylate, alkylcarbonyl, arylcarbonyl, alkoxycarbonyl, aminocarbonyl, alkylaminocarbonyl, dialkylaminocarbonyl, alkylthiocarbonyl, alkoxyl, phosphate, phosphonato, phosphinato, amino (including alkylamino, dialkylamino, arylamino, diarylamino and alkylarylamino), acylamino (including alkylcarbonylamino, arylcarbonylamino, carbamoyl and ureido), amidino, imino, sulfhydryl, alkylthio, arylthio, thiocarboxylate, sulfates, alkylsulfinyl, sulfonato, sulfamoyl, sulfonamido, nitro, trifluoromethyl, cyano, heterocyclyl, alkylaryl, or an aromatic or heteroaromatic moiety.

As used herein, the term "alkynyl" includes unsaturated aliphatic groups analogous in length and possible substitution to the alkyls described above, but which contain at least one triple bond. For example, "alkynyl" includes straight chain alkynyl groups (*e.g.*, ethynyl, propynyl, butynyl, pentynyl, hexynyl, heptynyl, octynyl, nonynyl, decynyl), and branched alkynyl groups. In certain embodiments, a straight chain or branched alkynyl group has twenty or fewer carbon atoms in its backbone (*e.g.*, C_2-C_{20} for straight chain, C_3-C_{20} for branched chain). In certain embodiments, a straight chain or branched alkynyl group has six or fewer

carbon atoms in its backbone (*e.g.*, C_2-C_6 for straight chain, C_3-C_6 for branched chain). The term "C_2-C_6" includes alkynyl groups containing two to six carbon atoms. The term "C_3-C_6" includes alkynyl groups containing three to six carbon atoms. The term "alkynylene" refers to a multivalent (e.g., bivalent, trivalent, tetravalent) alkyl group.

As used herein, the term "optionally substituted alkynyl" refers to unsubstituted alkynyl or alkynyl having designated substituents replacing one or more hydrogen atoms on one or more hydrocarbon backbone carbon atoms. Such substituents can include, for example, alkyl, alkenyl, alkynyl, halogen, hydroxyl, alkylcarbonyloxy, arylcarbonyloxy, alkoxycarbonyloxy, aryloxycarbonyloxy, carboxylate, alkylcarbonyl, arylcarbonyl, alkoxycarbonyl, aminocarbonyl, alkylaminocarbonyl, dialkylaminocarbonyl, alkylthiocarbonyl, alkoxyl, phosphate, phosphonato, phosphinato, amino (including alkylamino, dialkylamino, arylamino, diarylamino and alkylarylamino), acylamino (including alkylcarbonylamino, arylcarbonylamino, carbamoyl and ureido), amidino, imino, sulfhydryl, alkylthio, arylthio, thiocarboxylate, sulfates, alkylsulfinyl, sulfonato, sulfamoyl, sulfonamido, nitro, trifluoromethyl, cyano, azido, heterocyclyl, alkylaryl, or an aromatic or heteroaromatic moiety.

Other optionally substituted moieties (such as optionally substituted cycloalkyl, heterocycloalkyl, aryl, or heteroaryl) include both the unsubstituted moieties and the moieties having one or more of the designated substituents. For example, substituted heterocycloalkyl includes those substituted with one or more alkyl groups, such as 2,2,6,6-tetramethyl-piperidinyl and 2,2,6,6-tetramethyl-1,2,3,6-tetrahydropyridinyl.

As used herein, the term "cycloalkyl" refers to a saturated or partially unsaturated hydrocarbon monocyclic or polycyclic (*e.g.*, fused, bridged, or spiro rings) system having 3 to 30 carbon atoms (*e.g.*, C_3-C_{12}, C_3-C_{10}, or C_3-C_8). Examples of cycloalkyl include, but are not limited to, cyclopropyl, cyclobutyl, cyclopentyl, cyclohexyl, cycloheptyl, cyclooctyl, cyclopentenyl, cyclohexenyl, cycloheptenyl, 1,2,3,4-tetrahydronaphthalenyl, and adamantyl. In the case of polycyclic cycloalkyl, only one of the rings in the cycloalkyl needs to be non-aromatic. The term "cycloalkylene" refers to a multivalent (e.g., bivalent, trivalent, tetravalent) cycloalkylene group.

As used herein, the term "heterocycloalkyl" refers to a saturated or partially unsaturated 3-8 membered monocyclic, 7-12 membered bicyclic (fused, bridged, or spiro rings), or 11-14 membered tricyclic ring system (fused, bridged, or spiro rings) having one or more heteroatoms (such as O, N, S, P, or Se), *e.g.*, 1 or 1-2 or 1-3 or 1-4 or 1-5 or 1-6 heteroatoms, or *e.g.*, 1, 2, 3, 4, 5, or 6 heteroatoms, independently selected from the group consisting of nitrogen, oxygen and sulfur, unless specified otherwise. Examples of heterocycloalkyl groups include, but are

not limited to, piperidinyl, piperazinyl, pyrrolidinyl, dioxanyl, tetrahydrofuranyl, isoindolinyl, indolinyl, imidazolidinyl, pyrazolidinyl, oxazolidinyl, isoxazolidinyl, triazolidinyl, oxiranyl, azetidinyl, oxetanyl, thietanyl, 1,2,3,6-tetrahydropyridinyl, tetrahydropyranyl, dihydropyranyl, pyranyl, morpholinyl, tetrahydrothiopyranyl, 1,4-diazepanyl, 1,4-oxazepanyl, 2-oxa-5-azabicyclo[2.2.1]heptanyl, 2,5-diazabicyclo[2.2.1]heptanyl, 2-oxa-6-azaspiro[3.3]heptanyl, 2,6-diazaspiro[3.3]heptanyl, 1,4-dioxa-8-azaspiro[4.5]decanyl, 1,4-dioxaspiro[4.5]decanyl, 1-oxaspiro[4.5]decanyl, 1-azaspiro[4.5]decanyl, 3'H-spiro[cyclohexane-1,1'-isobenzofuran]-yl, 7'H-spiro[cyclohexane-1,5'-furo[3,4-b]pyridin]-yl, 3'H-spiro[cyclohexane-1,1'-furo[3,4-c]pyridin]-yl, 3-azabicyclo[3.1.0]hexanyl, 3-azabicyclo[3.1.0]hexan-3-yl, 1,4,5,6-tetrahydropyrrolo[3,4-c]pyrazolyl, 3,4,5,6,7,8-hexahydropyrido[4,3-d]pyrimidinyl, 4,5,6,7-tetrahydro-1H-pyrazolo[3,4-c]pyridinyl, 5,6,7,8-tetrahydropyrido[4,3-d]pyrimidinyl, 2-azaspiro[3.3]heptanyl, 2-methyl-2-azaspiro[3.3]heptanyl, 2-azaspiro[3.5]nonanyl, 2-methyl-2-azaspiro[3.5]nonanyl, 2-azaspiro[4.5]decanyl, 2-methyl-2-azaspiro[4.5]decanyl, 2-oxa-azaspiro[3.4]octanyl, 2-oxa-azaspiro[3.4]octan-6-yl, 5,6-dihydro-4H-cyclopenta[b]thiophenyl, and the like. In the case of multicyclic heterocycloalkyl, only one of the rings in the heterocycloalkyl needs to be non-aromatic (*e.g.*, 4,5,6,7-tetrahydrobenzo[c]isoxazolyl). The term "heterocycloalkylene" refers to a multivalent (e.g., bivalent, trivalent, tetravalent) heterocycloalkyl group.

It is understood that when a variable has two attachments to the rest of the formula of the compound, the two attachments could be at the same atom or different atoms of the variable. For example, when a variable (*e.g.*, variable X) is cycloalkyl or heterocycloalkyl, and has two attachments to the rest of the formula of the compound, the two attachments could be at the same atom or different atoms of the cycloalkyl or heterocycloalkyl.

As used herein, the term "aryl" includes groups with aromaticity, including "conjugated," or multicyclic systems with one or more aromatic rings and do not contain any heteroatom in the ring structure. The term aryl includes both monovalent species and divalent species. Examples of aryl groups include, but are not limited to, phenyl, biphenyl, naphthyl and the like. For example, an aryl is phenyl. The term "aryl" refers to a multivalent (e.g., bivalent, trivalent, or tetravalent) aryl group.

As used herein, the term "heteroaryl" is intended to include a stable 5-, 6-, or 7-membered monocyclic or 7-, 8-, 9-, 10-, 11- or 12-membered bicyclic aromatic heterocyclic ring which consists of carbon atoms and one or more heteroatoms, *e.g.*, 1 or 1-2 or 1-3 or 1-4 or 1-5 or 1-6 heteroatoms, or *e.g.*, 1, 2, 3, 4, 5, or 6 heteroatoms, independently selected from the group consisting of nitrogen, oxygen and sulfur. The nitrogen atom may be substituted or

unsubstituted (*i.e.*, N or NR wherein R is H or other substituents, as defined). The nitrogen and sulfur heteroatoms may optionally be oxidized (*i.e.*, N→O and $S(O)_p$, where p = 1 or 2). It is to be noted that total number of S and O atoms in the aromatic heterocycle is not more than 1. Examples of heteroaryl groups include pyrrole, furan, thiophene, thiazole, isothiazole, imidazole, triazole, tetrazole, pyrazole, oxazole, isoxazole, isothiazole, pyridine, pyrazine, pyridazine, pyrimidine, and the like. Heteroaryl groups can also be fused or bridged with alicyclic or heterocyclic rings, which are not aromatic so as to form a multicyclic system (*e.g.*, 4,5,6,7-tetrahydrobenzo[c]isoxazolyl). In some embodiments, the heteroaryl is thiophenyl or benzothiophenyl. In some embodiments, the heteroaryl is thiophenyl. In some embodiments, the heteroaryl benzothiophenyl. The term "heteroarylene" refers to a multivalent (e.g., bivalent, trivalent, or tetravalent) heteroaryl group.

Furthermore, the terms "aryl" and "heteroaryl" include multicyclic aryl and heteroaryl groups, *e.g.*, tricyclic, bicyclic, *e.g.*, naphthalene, benzoxazole, benzodioxazole, benzothiazole, benzoimidazole, benzothiophene, quinoline, isoquinoline, naphthrydine, indole, benzofuran, purine, benzofuran, deazapurine, indolizine.

The cycloalkyl, heterocycloalkyl, aryl, or heteroaryl ring can be substituted at one or more ring positions (*e.g.*, the ring-forming carbon or heteroatom such as N) with such substituents as described above, for example, alkyl, alkenyl, alkynyl, halogen, hydroxyl, alkoxy, alkylcarbonyloxy, arylcarbonyloxy, alkoxycarbonyloxy, aryloxycarbonyloxy, carboxylate, alkylcarbonyl, alkylaminocarbonyl, aralkylaminocarbonyl, alkenylaminocarbonyl, alkylcarbonyl, arylcarbonyl, aralkylcarbonyl, alkenylcarbonyl, alkoxycarbonyl, aminocarbonyl, alkylthiocarbonyl, phosphate, phosphonato, phosphinato, amino (including alkylamino, dialkylamino, arylamino, diarylamino and alkylarylamino), acylamino (including alkylcarbonylamino, arylcarbonylamino, carbamoyl and ureido), amidino, imino, sulfhydryl, alkylthio, arylthio, thiocarboxylate, sulfates, alkylsulfinyl, sulfonato, sulfamoyl, sulfonamido, nitro, trifluoromethyl, cyano, azido, heterocyclyl, alkylaryl, or an aromatic or heteroaromatic moiety. Aryl and heteroaryl groups can also be fused or bridged with alicyclic or heterocyclic rings, which are not aromatic so as to form a multicyclic system (*e.g.*, tetralin, methylenedioxyphenyl such as benzo[d][1,3]dioxole-5-yl).

As used herein, the term "substituted," means that any one or more hydrogen atoms on the designated atom is replaced with a selection from the indicated groups, provided that the designated atom's normal valency is not exceeded, and that the substitution results in a stable compound. When a substituent is oxo or keto (*i.e.*, =O), then 2 hydrogen atoms on the atom

are replaced. Keto substituents are not present on aromatic moieties. Ring double bonds, as used herein, are double bonds that are formed between two adjacent ring atoms (*e.g.*, C=C, C=N or N=N). "Stable compound" and "stable structure" are meant to indicate a compound that is sufficiently robust to survive isolation to a useful degree of purity from a reaction mixture, and formulation into an efficacious therapeutic agent.

When a bond to a substituent is shown to cross a bond connecting two atoms in a ring, then such substituent may be bonded to any atom in the ring. When a substituent is listed without indicating the atom via which such substituent is bonded to the rest of the compound of a given formula, then such substituent may be bonded via any atom in such formula. Combinations of substituents and/or variables are permissible, but only if such combinations result in stable compounds.

When any variable (*e.g.*, R) occurs more than one time in any constituent or formula for a compound, its definition at each occurrence is independent of its definition at every other occurrence. Thus, for example, if a group is shown to be substituted with 0-2 R moieties, then the group may optionally be substituted with up to two R moieties and R at each occurrence is selected independently from the definition of R. Also, combinations of substituents and/or variables are permissible, but only if such combinations result in stable compounds.

As used herein, the term "hydroxy" or "hydroxyl" includes groups with an -OH or $-O^-$.

As used herein, the term "halo" or "halogen" refers to fluoro, chloro, bromo and iodo.

The term "haloalkyl" or "haloalkoxyl" refers to an alkyl or alkoxyl substituted with one or more halogen atoms.

As used herein, the term "optionally substituted haloalkyl" refers to unsubstituted haloalkyl having designated substituents replacing one or more hydrogen atoms on one or more hydrocarbon backbone carbon atoms. Such substituents can include, for example, alkyl, alkenyl, alkynyl, halogen, hydroxyl, alkylcarbonyloxy, arylcarbonyloxy, alkoxycarbonyloxy, aryloxycarbonyloxy, carboxylate, alkylcarbonyl, arylcarbonyl, alkoxycarbonyl, aminocarbonyl, alkylaminocarbonyl, dialkylaminocarbonyl, alkylthiocarbonyl, alkoxyl, phosphate, phosphonato, phosphinato, amino (including alkylamino, dialkylamino, arylamino, diarylamino and alkylarylamino), acylamino (including alkylcarbonylamino, arylcarbonylamino, carbamoyl and ureido), amidino, imino, sulfhydryl, alkylthio, arylthio, thiocarboxylate, sulfates, alkylsulfinyl, sulfonato, sulfamoyl, sulfonamido, nitro, trifluoromethyl, cyano, azido, heterocyclyl, alkylaryl, or an aromatic or heteroaromatic moiety.

As used herein, the term "alkoxy" or "alkoxyl" includes substituted and unsubstituted alkyl, alkenyl and alkynyl groups covalently linked to an oxygen atom. Examples of alkoxy groups or alkoxyl radicals include, but are not limited to, methoxy, ethoxy, isopropyloxy, propoxy, butoxy and pentoxy groups. Examples of substituted alkoxy groups include halogenated alkoxy groups. The alkoxy groups can be substituted with groups such as alkenyl, alkynyl, halogen, hydroxyl, alkylcarbonyloxy, arylcarbonyloxy, alkoxycarbonyloxy, aryloxycarbonyloxy, carboxylate, alkylcarbonyl, arylcarbonyl, alkoxycarbonyl, aminocarbonyl, alkylaminocarbonyl, dialkylaminocarbonyl, alkylthiocarbonyl, alkoxyl, phosphate, phosphonato, phosphinato, amino (including alkylamino, dialkylamino, arylamino, diarylamino, and alkylarylamino), acylamino (including alkylcarbonylamino, arylcarbonylamino, carbamoyl and ureido), amidino, imino, sulfhydryl, alkylthio, arylthio, thiocarboxylate, sulfates, alkylsulfinyl, sulfonato, sulfamoyl, sulfonamido, nitro, trifluoromethyl, cyano, azido, heterocyclyl, alkylaryl, or an aromatic or heteroaromatic moieties. Examples of halogen substituted alkoxy groups include, but are not limited to, fluoromethoxy, difluoromethoxy, trifluoromethoxy, chloromethoxy, dichloromethoxy and trichloromethoxy.

As used herein, the term "alkyl aryl" refers to an alkyl group, as defined herein, substituted with an aryl group, as defined herein. A "C_{1-6} alkyl C_{6-10} aryl" group refers to a C_{1-6} alkyl group, as defined herein, substituted with a C_{6-10} aryl group, as defined herein. Unless otherwise specified, alkyl group and aryl group may be optionally substituted as described herein.

As used herein, the term "alkyl cycloalkyl" refers to an alkyl group, as defined herein, substituted with a cycloalkyl group, as defined herein. A "C_{1-6} alkyl C_{3-10} aryl" group refers to a C_{1-6} alkyl group, as defined herein, substituted with a C_{3-10} cycloalkyl group, as defined herein. Unless otherwise specified, alkyl group and cycloalkyl group may be optionally substituted as described herein.

As used herein, the term "alkyl heteroaryl" refers to an alkyl group, as defined herein, substituted with a heteroaryl group, as defined herein. A "C_{1-6} alkyl C_{2-9} heteroaryl" group refers to a C_{1-6} alkyl group, as defined herein, substituted with a C_{2-9} heteroaryl group, as defined herein. Unless otherwise specified, alkyl group and heteroaryl group may be optionally substituted as described herein.

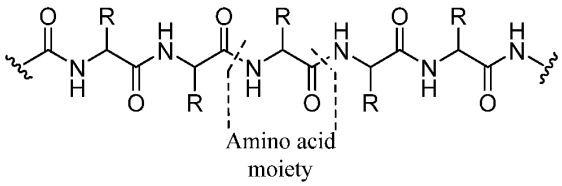
As used herein, the term "alkyl heterocycloalkyl" refers to an alkyl group, as defined herein, substituted with a heteroaryl group, as defined herein. A "C_{1-6} alkyl C_{2-9} heterocycloalkyl" group refers to a C_{1-6} alkyl group, as defined herein, substituted with a C_{2-9}

heterocycloalkyl group, as defined herein. Unless otherwise specified, alkyl group and heterocycloalkyl group may be optionally substituted as described herein.

As used herein, the term "heteroalkyl" refers to an alkyl group, as defined herein, that further comprises at least one heteroatom (e.g., at least one nitrogen, oxygen, or sulfur atom). In some embodiments, a C_{1-20} heteroalkyl group has twenty or fewer carbon atoms and twenty or fewer heteroatoms. In some embodiments, a straight branched alkyl has six or fewer carbon atoms and six or fewer heteroatoms, and in another embodiment, a straight chain or branched chain heteroalkyl has four or fewer carbon atoms and four or fewer heteroatoms. A C_1 heteroalkyl comprises one carbon atom and at least one heteroatom. The term "heteroalkylene" refers to a multivalent heteroalkyl group, e.g., a bivalent, trivalent, or tetravalent heteroalkyl group.

As used herein, the term "amino acid" refers to an organic molecule that comprises an amino group and a carboxylic acid that separated by one, two ,or three methylene units, wherein the methylene units are optionally substituted with a C_{1-6} alkyl group, C_{1-6} heteroalkyl group, C_{1-6} alkyl C_{6-10} aryl group, C_{1-6} alkyl C_{2-9} heteroaryl group, C_{1-6} alkyl C_{3-10} cycloalkyl group, or C_{1-6} alkyl C_{2-9} heterocycloalkyl group, wherein the C_{1-6} alkyl group, C_{1-6} heteroalkyl group, C_{1-6} alkyl C_{6-10} aryl group, C_{1-6} alkyl C_{2-9} heteroaryl group, C_{1-6} alkyl C_{3-10} cycloalkyl group, or C_{1-6} alkyl C_{2-9} heterocycloalkyl group is optionally substituted with a hydroxy, halo, cyano, or nitro group. Amino acids in which the carboxylic acid and amino group are separated by one optionally substituted methylene unit are referred to herein as "alpha amino acids." Amino acids in which the carboxylic acid and amino group are separated by two optionally substituted methylene units are referred to herein as "beta amino acids." Amino acids in which the carboxylic acid and amino group are separated by optionally substituted methylene units are referred to herein as "gamma amino acids." This disclosure contemplates the use of naturally occurring and non-naturally occurring amino acids. This disclosure contemplates the use of (L) and (D) amino acids. Exemplary and nonlimiting amino acids include alanine, arginine, asparagine, aspartic acid, cysteine, glutamine, glutamic acid, glycine, histidine, isoleucine, leucine, lysine, methionine, phenylalnine, proline, serine, threonine, tryptophan, tyrosine, valine, pyrrolocysteine, selenocysteine, pyrrolysine, 2-naphthyl-alanine, statine, homoalanine, 3-pyridyl-alanine, 4-fluorophenyl-alanine, cyclohexyl-alanine, homo-cysteine, penicillamine, 3-nitro-tyrosine, homo-phenyl-alanine, *t*-leucine, and hydroxy-proline. As used herein, the term "amino acid moiety" refers to the portion of a chain of linked amino acids that would correspond to the atoms in one individual amino acid, were the amino acids unlinked. Scheme AA-1 illustrates the relationship between a linked chain of amino acids an amino acid moiety:

Scheme AA-1



This disclosure contemplates embodiments in which amino acids are linked to one or more moieties that are not amino acids. In such instance, the term "amino acid moiety" refers to the atoms that would correspond to atoms within the individual amino acid, were the amino acid unlinked from the moiety or moieties that are not amino acids.

As used herein, the term "homopolymer" refers to a polymer consisting of one type of monomer. The constituent monomers of the homopolymer may be optionally substituted, e.g., with a nucleotide moiety, reporter moiety, blocking moiety, PEG-Cap moiety, or negative charge moiety.

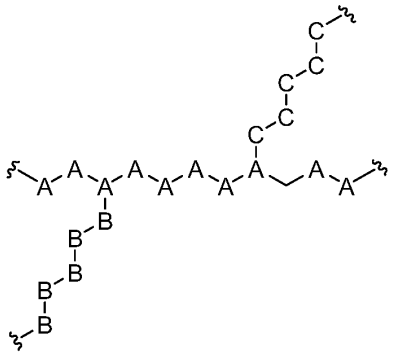
As used herein, the term "copolymer" refers to a polymer consisting of a plurality of more than one type of monomer. The constituent monomers of the copolymer may be optionally substituted, e.g., with a nucleotide moiety, reporter moiety, blocking moiety, PEG-Cap moiety, or negative charge moiety.

As used herein the term "alternating copolymer" refers to a copolymer, as described herein, wherein the more than one types of monomer, e.g., A and B or A, B, and C, are present in an alternating and repeating order of monomer types, e.g., ...A-B-A-B-A-B.... or ...A-B-C-A-B-C... The constituent monomers of the alternating copolymer be optionally substituted, e.g., with a nucleotide moiety, reporter moiety, blocking moiety, PEG-Cap moiety, or negative charge moiety.

As used herein, the term "random copolymer" refers to a copolymer, as described herein, wherein the order of constituent monomers is random. The constituent monomers in the random copolymer may be optionally substituted, e.g., with a nucleotide moiety, reporter moiety, blocking moiety, PEG-Cap moiety, or negative charge moiety.

As used herein, the term "block copolymer" refers to a copolymer, as described herein, wherein the more than one type of monomer, e.g., A and B, are present in an uninterrupted sequence of sequence of monomer type, e.g., ...A-A-A-A-A-A-A-A-A-B-B-B-B-B-B-B-B... The constituent monomers in the block copolymer may be optionally substituted, e.g., with a nucleotide moiety, reporter moiety, blocking moiety, PEG-Cap moiety, or negative charge moiety.

As used herein, the term "graft copolymer" refers to a first homopolymer, as described herein, wherein the first homopolymer is substituted with one or more second homopolymer. An exemplary and non-limiting structure graft copolymer consisting of monomer types A, B, and C is shown in Scheme GP-1:



GP-1

The constituent monomers in the graft copolymer may be optionally substituted, e.g., with a nucleotide moiety, reporter moiety, blocking moiety, PEG-Cap moiety, or negative charge moiety.

Unless defined otherwise, technical and scientific terms used herein have meanings that are commonly understood by those of ordinary skill in the art unless defined otherwise. Generally, terminologies pertaining to techniques of molecular biology, nucleic acid chemistry, protein chemistry, genetics, microbiology, transgenic cell production, and hybridization described herein are those well-known and commonly used in the art. Techniques and procedures described herein are generally performed according to conventional methods well known in the art and as described in various general and more specific references that are cited and discussed throughout the instant specification. For example, see Sambrook et al., Molecular Cloning: A Laboratory Manual (Third ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. 2000). See also Ausubel et al., Current Protocols in Molecular Biology, Greene Publishing Associates (1992). The nomenclatures utilized in connection with, and the laboratory procedures and techniques described herein are those well-known and commonly used in the art.

Throughout this application various publications, patents, and/or patent applications are referenced. The disclosures of the publications, patents and/or patent applications are hereby incorporated by reference in their entireties into this application in order to more fully describe the state of the art to which this disclosure pertains.

Polymeric Molecules

The disclosure provides polymeric molecules (e.g., compositions comprising compounds of Formula (I), below), and methods of using same in high throughput sequencing. In some embodiments, the methods comprising contacting a plurality of template nucleic acid molecules comprising two or more copies of a target sequence, or its complement, and two or more copies of a binding sequence for a sequencing primer, wherein the binding sequences of the template nucleic acid molecules comprise duplexes with sequencing primers or extended strands thereof, a plurality of polymerases, and a plurality of polymeric molecules of Formula (I) below, or an ionized form thereof, an isomer thereof, or a salt thereof, under conditions sufficient to form a plurality of multivalent binding complexes comprising a nucleic acid duplex between a template nucleic acid molecule and sequencing primer or extended sequencing primer strand, a first polymerase, and a nucleotide moiety of a polymeric molecule that is complementary to a nucleotide in the template nucleic acid molecule immediately adjacent to a 3' end of the sequencing primer or extended sequencing primer strand. In some embodiments, the methods comprise detecting the detectable reporter moieties of the polymeric molecules, and determining nucleobase identities of nucleotides in the nucleic acid template molecules complementary to the nucleotide moieties of the polymeric molecules based on the detectable reporter moieties of the polymeric molecules in the plurality of multivalent binding complexes. In some embodiments, the polymeric molecule comprises at least two nucleotide moieties and at least one detectable reporter moiety, and two or more nucleotide moieties in an individual polymeric molecule contact two or more different multivalent binding complexes. In some embodiments, the two or more different multivalent binding complexes are on the same template nucleic acid molecule. In some embodiments, the two or more multivalent binding complexes are on different template nucleic acid molecules.

In some aspects, the present disclosure provides a compound of Formula (I):

$$C\text{-}((P)\text{-}E)_s$$

(I)

an ionized form thereof, or a salt thereof, wherein:

C is a central moiety;

each P independently is an optionally substituted polymeric side chain;

each E independently is an end moiety; and

s is an integer ranging from 1 to 10.

In some embodiments, the present disclosure provides a compound of Formula (II):

$$C\text{-}((P)\text{-}E)_4$$

(II)

an ionized form thereof, or a salt thereof.

In some embodiments, at least one P is substituted with one or more reporter moiety.

In some embodiments, each P is substituted with one or more reporter moiety .

In some embodiments, at least one P is substituted with one or more nucleotide moiety.

In some embodiments, each P is substituted with one or more nucleotide moiety .

In some embodiments, at least one P is substituted with (i) one or more reporter moiety and (ii) one or more nucleotide moiety.

In some embodiments, each P is substituted with (i) one or more reporter moiety and (ii) one or more nucleotide moiety.

In some embodiments, at least one P (e.g., each P) is further substituted with one or more blocking moiety, negative charge moiety, or PEG-Cap moiety.

In some embodiments, at least one P (e.g., each P) is further substituted with one or more blocking moiety.

In some embodiments, at least one P (e.g., each P) is further substituted with one or more negative charge moiety .

In some embodiments, at least one P (e.g., each P) is further substituted with one or more a PEG-Cap moiety.

In some embodiments, at least one P (e.g., each P) is further substituted with (iii) one or more blocking moiety, (iv) one or more negative charge moiety , and (v) one or more PEG-Cap moiety.

In some embodiments, at least one P (e.g., each P) is substituted with (i) one or more reporter moiety, (ii) one or more nucleotide moiety, (iii) one or more blocking moiety, (iv) one or more negative charge moiety, and (v) one or more PEG-Cap moiety.

Polymeric Side Chain

In some embodiments, at least one polymeric side chain (e.g., each polymeric side chain) comprises a polymer.

In some embodiments, at least one polymeric side chain (e.g., each polymeric side chain) comprises a homopolymer.

In some embodiments, at least one polymeric side chain (e.g., each polymeric side chain) comprises a homopolymer and one or more reporter moiety or nucleotide moiety.

In some embodiments, at least one polymeric side chain (e.g., each polymeric side chain) comprises a homopolymer, and (i) one or more reporter moiety and (ii) one or more nucleotide moiety.

In some embodiments, at least one polymeric side chain (e.g., each polymeric side chain) further comprises one or more blocking moiety, negative charge moiety, or PEG-Cap moiety.

In some embodiments, at least one polymeric side chain (e.g., each polymeric side chain) comprises a homopolymer, and (i) one or more reporter moiety, (ii) one or more nucleotide moiety, (iii) one or more blocking moiety, (iv) one or more negative charge moiety, and (v) one or more PEG-Cap moiety.

In some embodiments, the polymeric side chain comprises a copolymer.

In some embodiments, at least one polymeric side chain (e.g., each polymeric side chain) comprises a copolymer and one or more (i) reporter moiety or (ii) nucleotide moiety.

In some embodiments, at least one polymeric side chain (e.g., each polymeric side chain) comprises a copolymer and (i) one or more reporter moiety and (ii) one or more nucleotide moiety.

In some embodiments, at least one polymeric side chain (e.g., each polymeric side chain) further comprises one or more blocking moiety, negative charge moiety, or PEG-Cap moiety.

In some embodiments, at least one polymeric side chain (e.g., each polymeric side chain) comprises a copolymer, and (i) one or more reporter moiety, (ii) one or more nucleotide moiety, (iii) one or more blocking moiety, (iv) one or more negative charge moiety, and (v) one or more PEG-Cap moiety.

In some embodiments, the polymeric side chain comprises an alternating copolymer.

In some embodiments, at least one polymeric side chain (e.g., each polymeric side chain) comprises an alternating copolymer and one or more reporter moiety or nucleotide moiety.

In some embodiments, at least one polymeric side chain (e.g., each polymeric side chain) comprises an alternating copolymer and (i) one or more reporter moiety and (ii) one or more nucleotide moiety.

In some embodiments, at least one polymeric side chain (e.g., each polymeric side chain) further comprises one or more blocking moiety, negative charge moiety, PEG-Cap moiety.

In some embodiments, at least one polymeric side chain (e.g., each polymeric side chain) comprises an alternating copolymer, and (i) one or more reporter moiety, (ii) one or more nucleotide moiety, (iii) one or more blocking moiety, (iv) one or more negative charge moiety, and (v) one or more PEG-Cap moiety.

In some embodiments, the polymeric side chain comprises a random copolymer.

In some embodiments, at least one polymeric side chain (e.g., each polymeric side chain) comprises a random copolymer, and one or more reporter moiety or nucleotide moiety.

In some embodiments, at least one polymeric side chain (e.g., each polymeric side chain) comprises a random copolymer and (i) one or more reporter moiety and (ii) one or more nucleotide moiety.

In some embodiments, at least one polymeric side chain (e.g., each polymeric side chain) further comprises one or more blocking moiety, negative charge moiety, or PEG-Cap moiety.

In some embodiments, at least one polymeric side chain (e.g., each polymeric side chain) comprises a random copolymer, and (i) one or more reporter moiety, (ii) one or more nucleotide moiety, (iii) one or more blocking moiety, (iv) one or more negative charge moiety, and (v) one or more PEG-Cap moiety.

In some embodiments, the polymeric side chain comprises a block copolymer.

In some embodiments, at least one polymeric side chain (e.g., each polymeric side chain) comprises a block copolymer and one or more reporter moiety or nucleotide moiety.

In some embodiments, at least one polymeric side chain (e.g., each polymeric side chain) comprises a block copolymer, and (i) one or more reporter moiety and (ii) one or more nucleotide moiety.

In some embodiments, at least one polymeric side chain (e.g., each polymeric side chain) further comprises one or more blocking moiety, negative charge moiety, or PEG-Cap moiety.

In some embodiments, at least one polymeric side chain (e.g., each polymeric side chain) comprises a block copolymer, and (i) one or more reporter moiety, (ii) one or more nucleotide moiety, (iii) one or more blocking moiety, (iv) one or more negative charge moiety, and (v) one or more PEG-Cap moiety.

In some embodiments, the polymeric side chain comprises a graft copolymer.

In some embodiments, at least one polymeric side chain (e.g., each polymeric side chain) comprises a graft copolymer and one or more reporter moiety or nucleotide moiety.

In some embodiments, at least one polymeric side chain (e.g., each polymeric side chain) comprises a graft copolymer, and (i) one or more reporter moiety and (ii) one or more nucleotide moiety.

In some embodiments, at least one polymeric side chain (e.g., each polymeric side chain) further comprises one or more blocking moiety, negative charge moiety, or PEG-Cap moiety.

In some embodiments, at least one polymeric side chain (e.g., each polymeric side chain) comprises a graft copolymer, and (i) one or more reporter moiety, (ii) one or more nucleotide moiety, (iii) one or more blocking moiety, (iv) one or more negative charge moiety, and (v) one or more PEG-Cap moiety.

Reporter Moiety

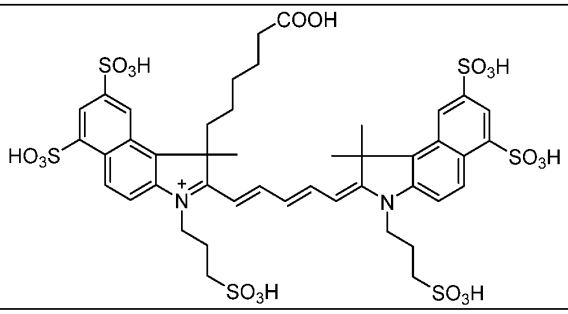
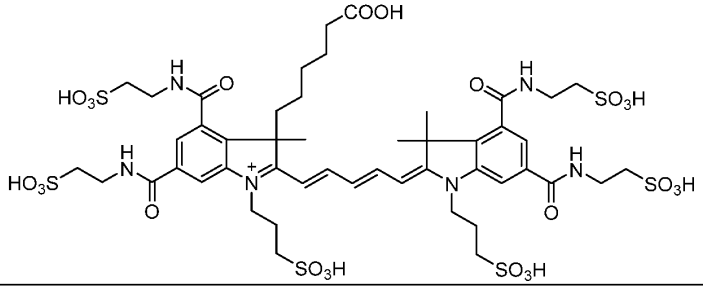
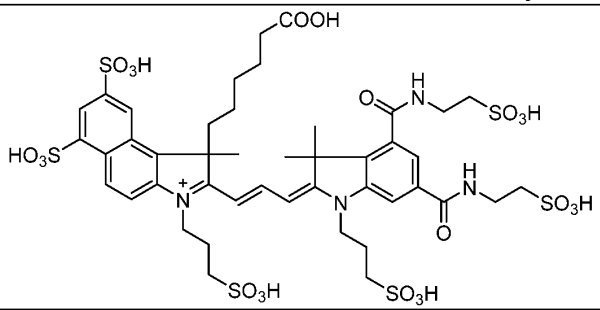
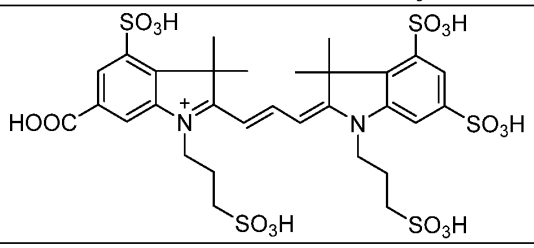
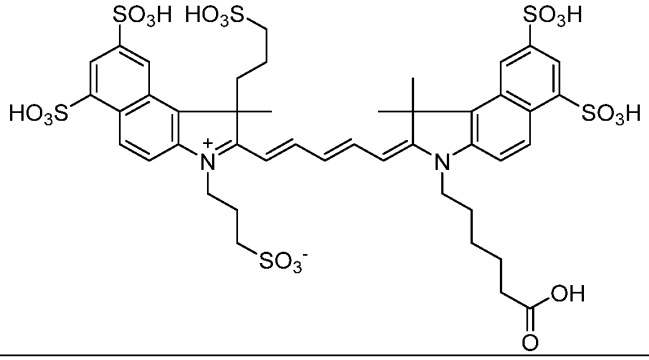
In some embodiments, the reporter moiety is a moiety that is detectable (e.g., being capable of emitting a signal). In some embodiments, when bound to a surface (e.g., by way of a base-pairing interaction between nucleotides), the reporter moiety is capable of allowing the compound to be detected.

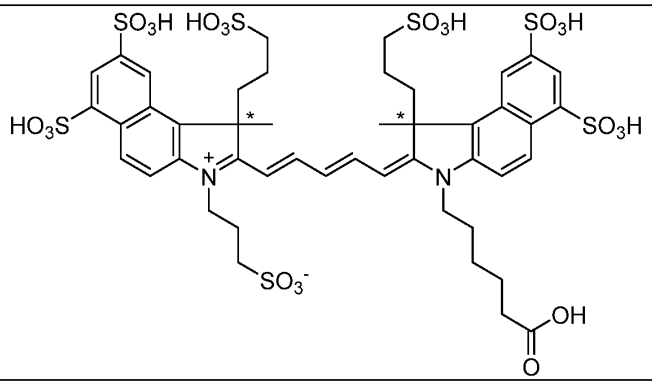
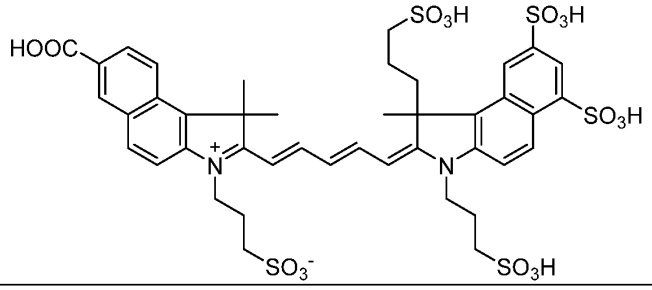
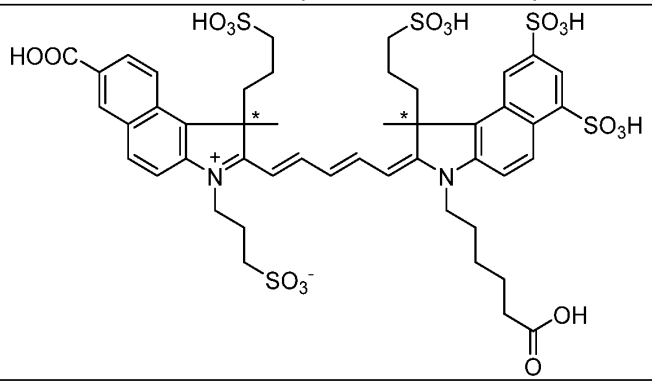
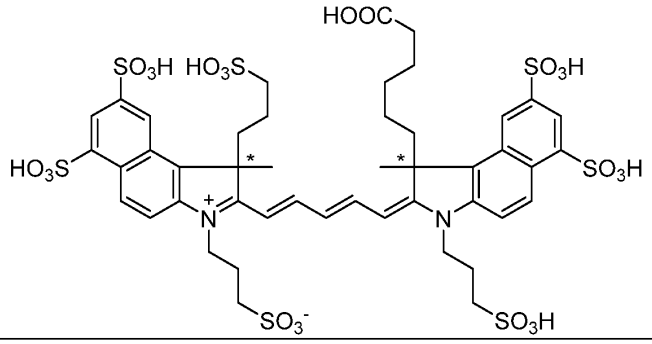
In some embodiments, the reporter moiety comprises a dye (e.g., a fluorescent dye). As used herein, the term "fluorophore" is used synonymously with "fluorescent dye."

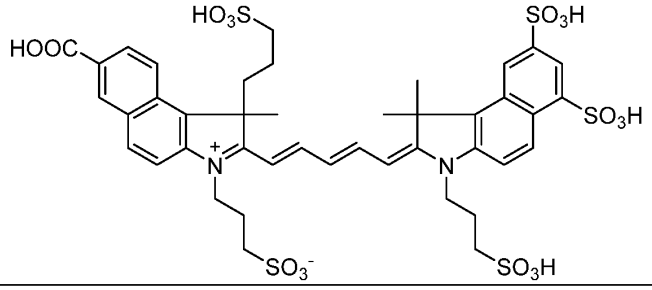
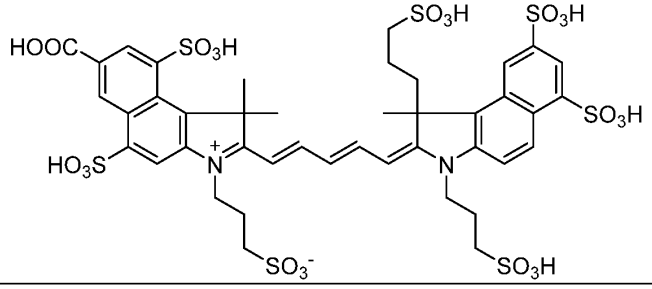
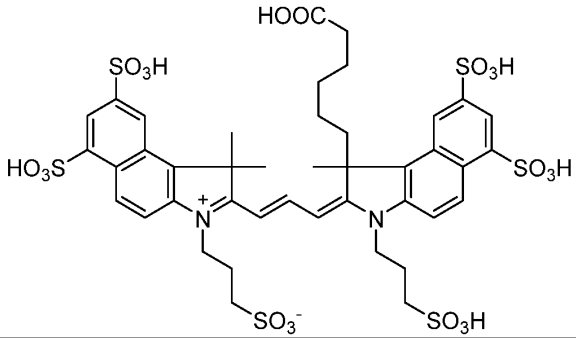
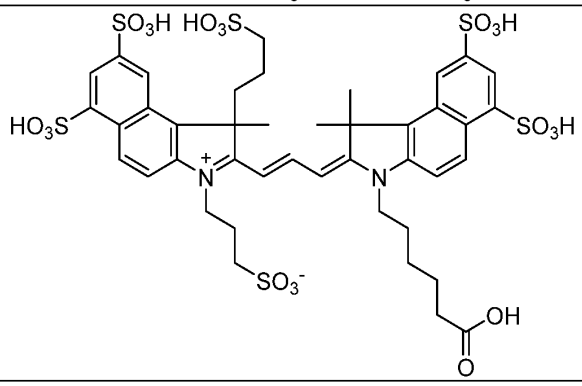
In some embodiments, the fluorophore is a cyanine dye. In some embodiments, the dye is directly bound to the polymeric side chain, e.g., by way of a chemical bond. In some embodiments, the dye is bound to the polymeric side chain by way of an amide bond. In some embodiments, the dye is bound to the polymeric side chain by way of an ester bond. In some embodiments, the dye is bound to the polymeric side chain by way of a thioester bond. In some embodiments, the dye is bound to the polymeric side chain by way of a bivalent connection moiety, such as the bivalent connection moiety disclosed below.

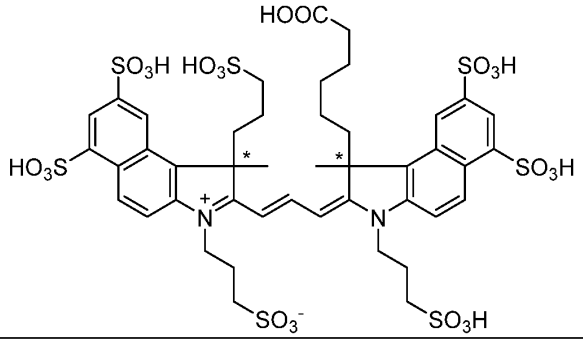
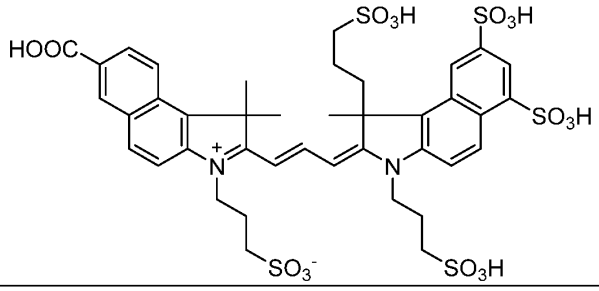
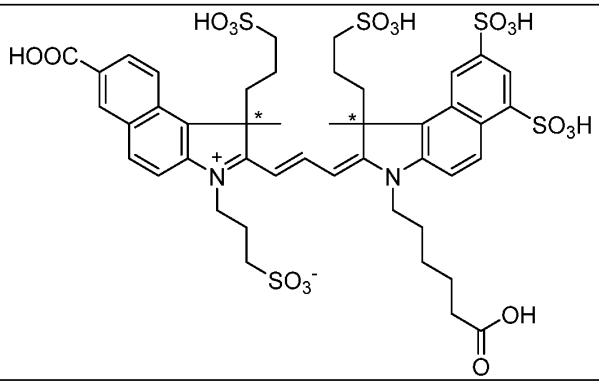
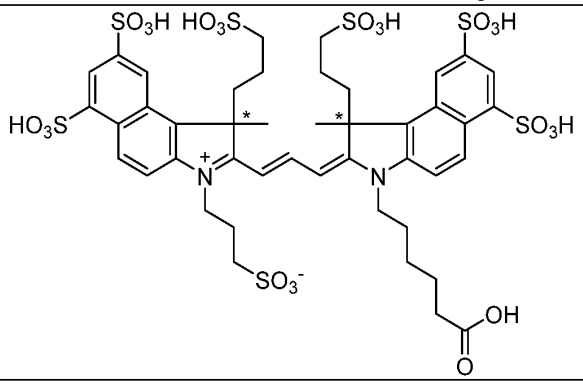
In some embodiments, the fluorophore is CF570. In some embodiments, the fluorophore comprises the structure of any one of the compounds in Table 1:

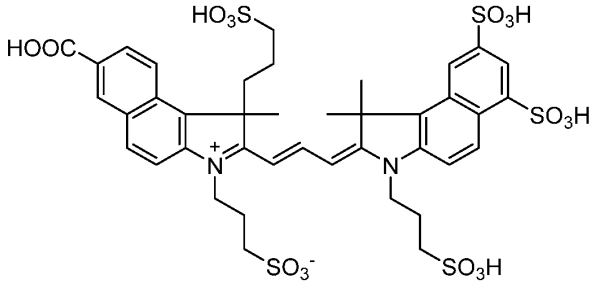
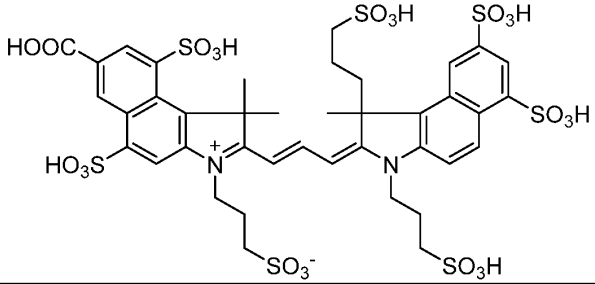
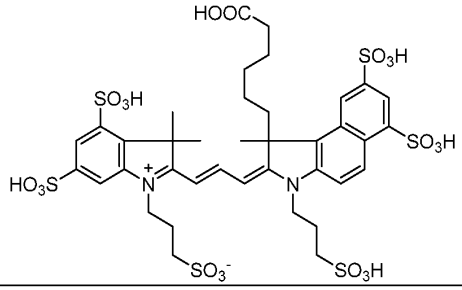
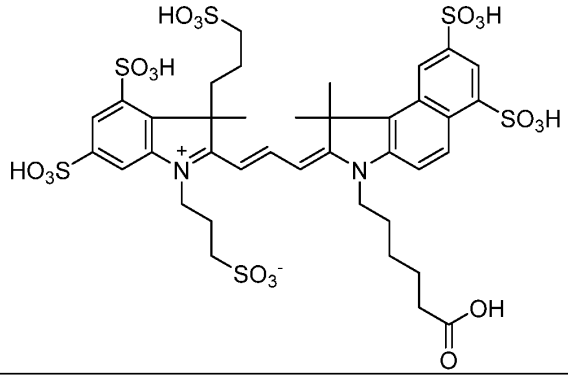
Table 1: Fluorophores

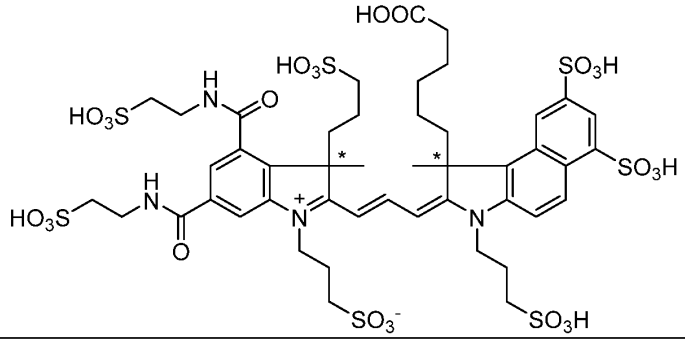
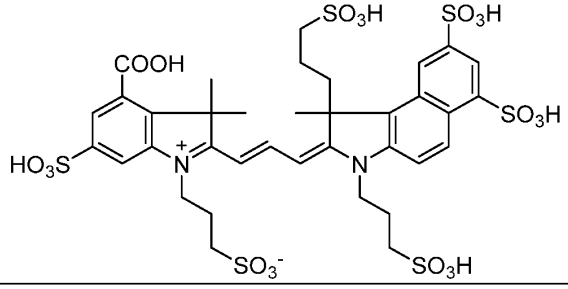
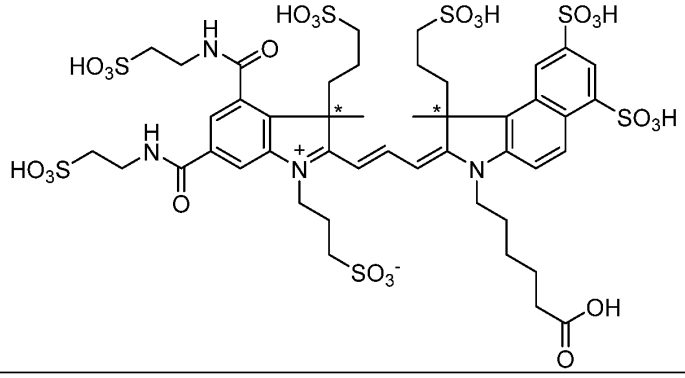
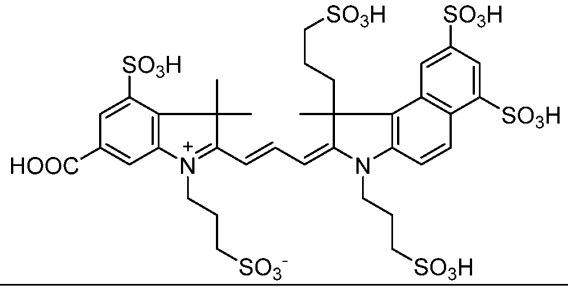
Compound No.	Structure
1	
2	
3	
4	
9	

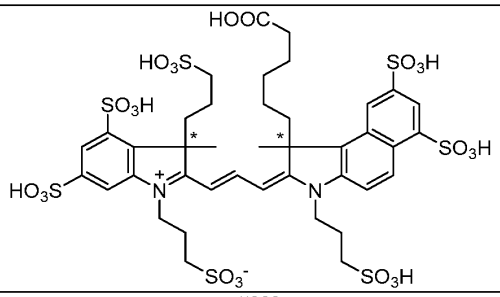
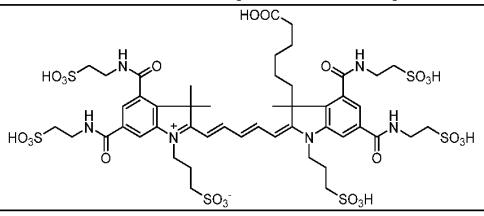
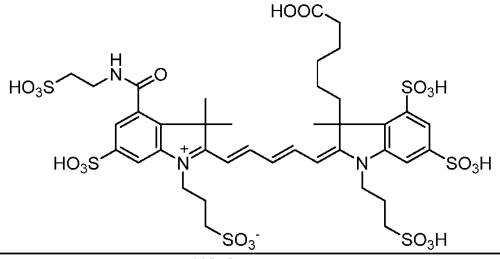
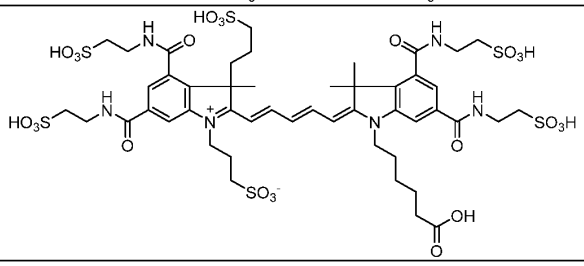
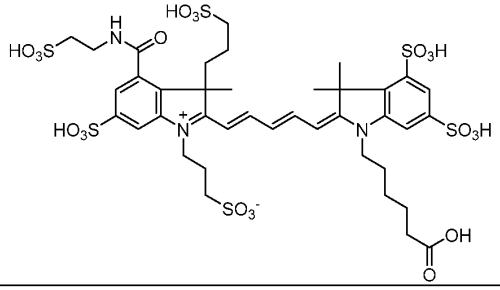
Compound No.	Structure
10	
11	
12	
13	

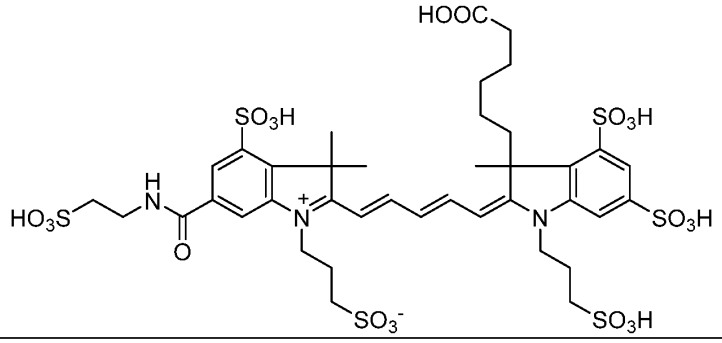
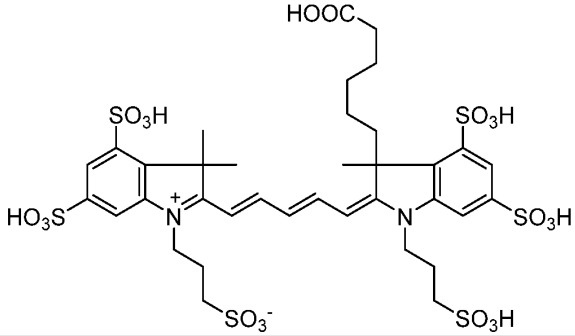
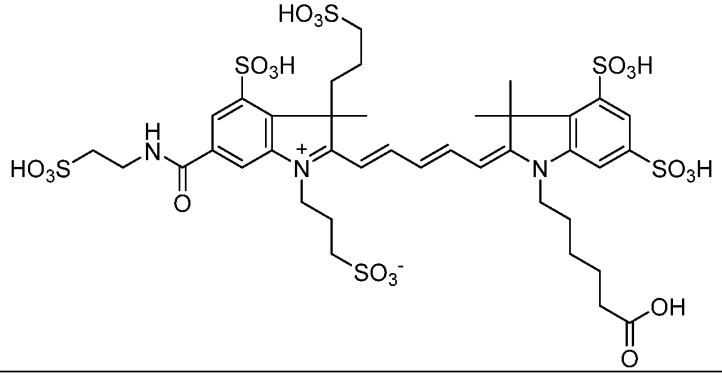
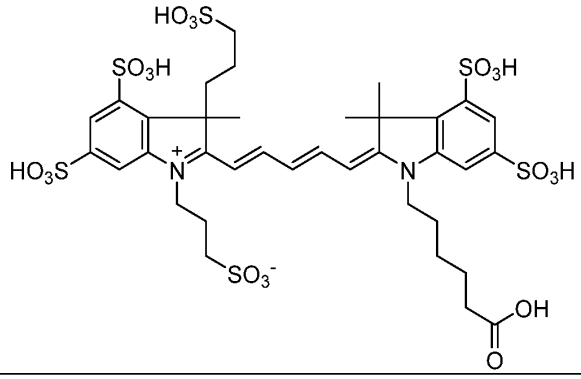
Compound No.	Structure
14	
15	
16	
17	

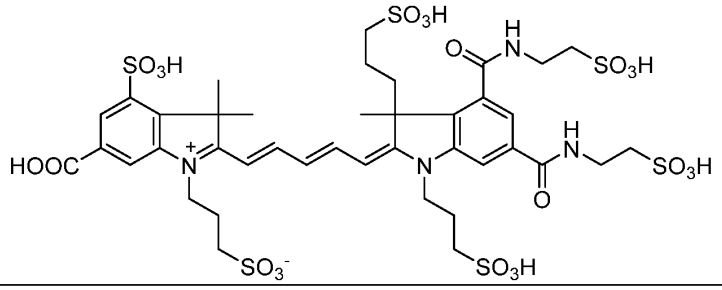
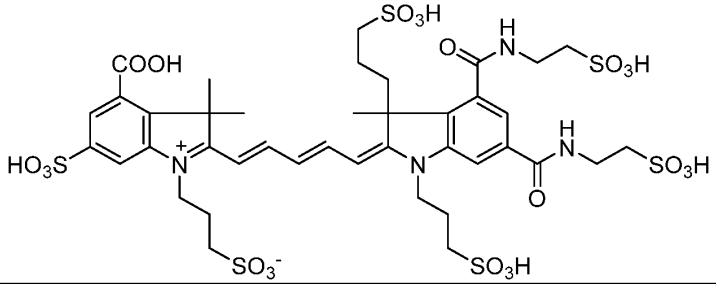
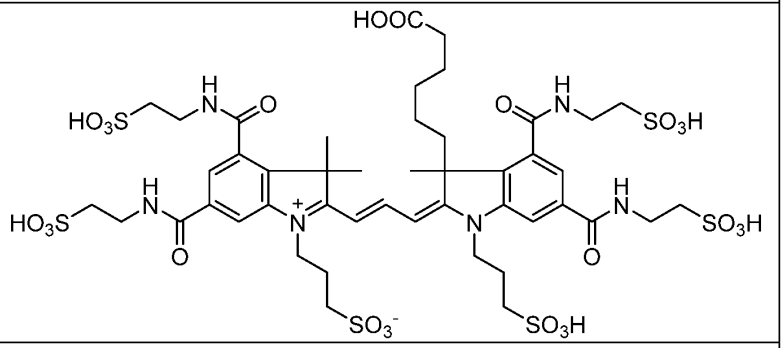
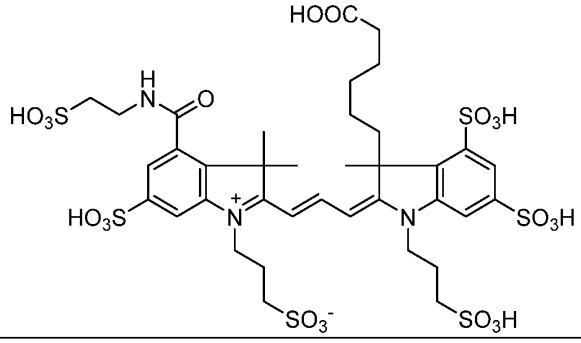
Compound No.	Structure
18	
19	
20	
21	

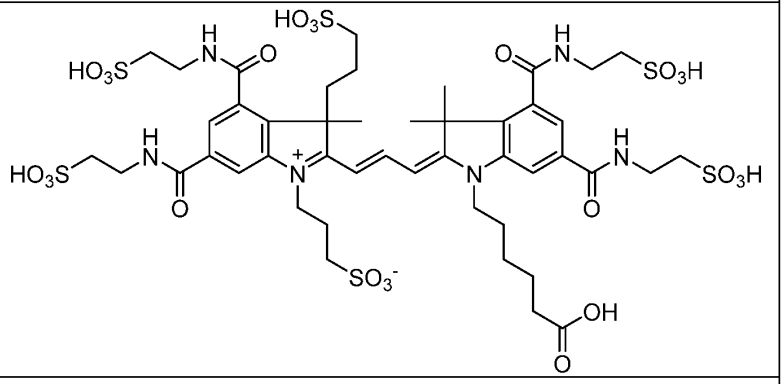
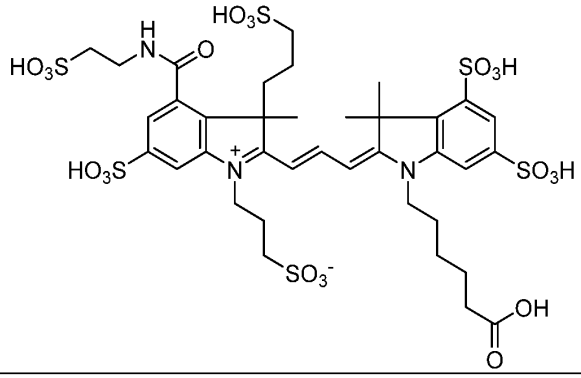
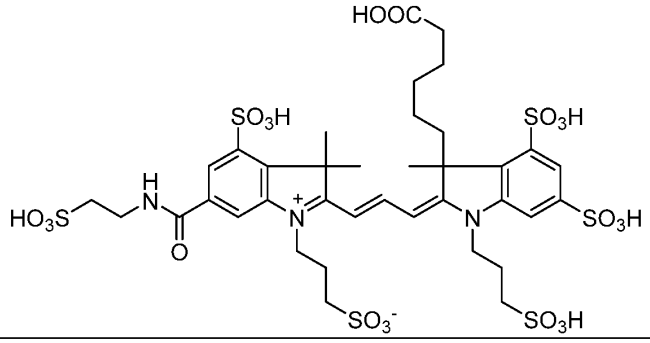
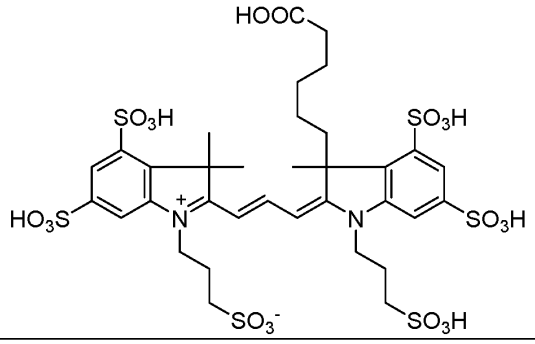
Compound No.	Structure
22	
23	
24	
25	

Compound No.	Structure
26	
27	
28	
29	

Compound No.	Structure
30	
31	
32	
33	
34	

Compound No.	Structure
35	
36	
37	
38	

Compound No.	Structure
39	
40	
41	
42	

Compound No.	Structure
43	
44	
45	
46	

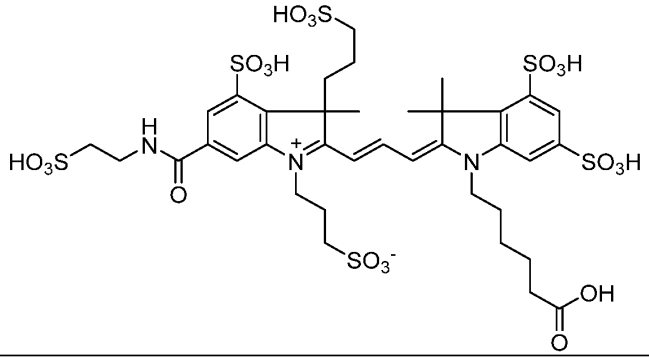
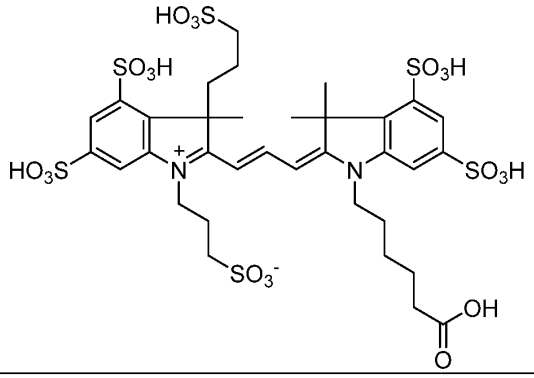
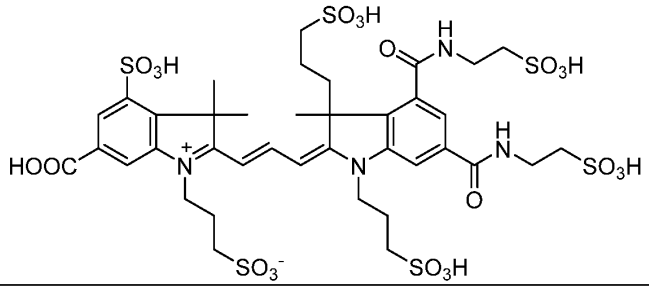
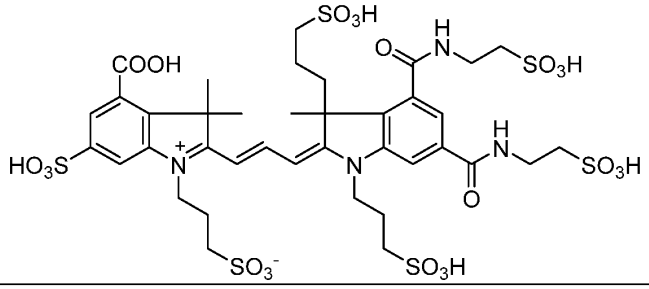
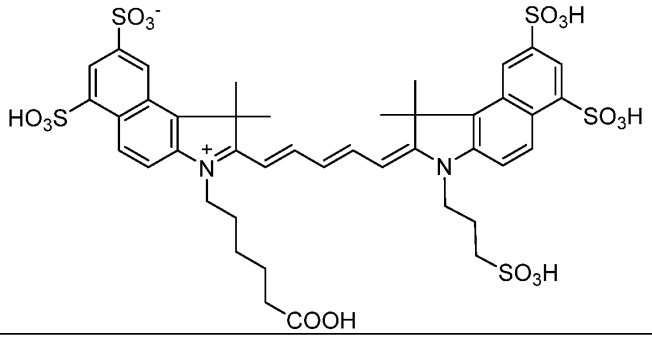
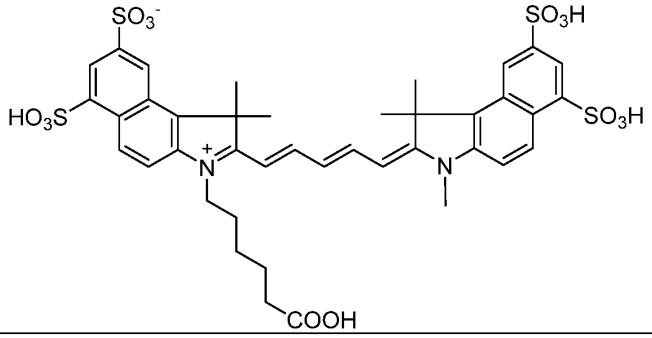
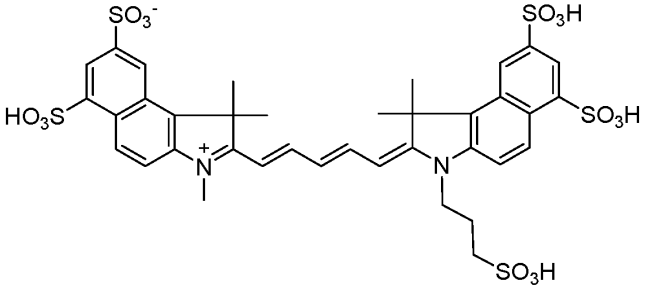
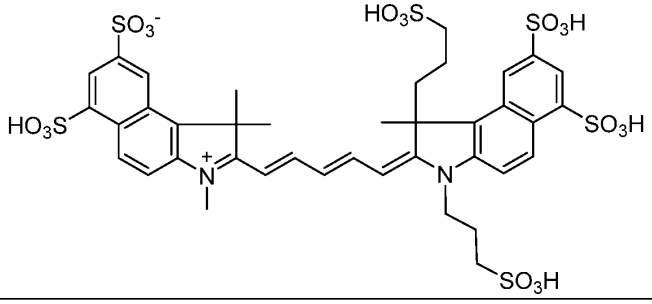
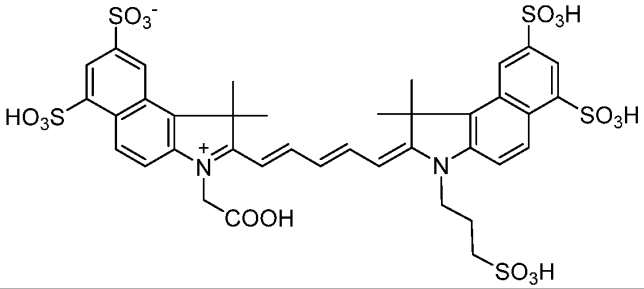
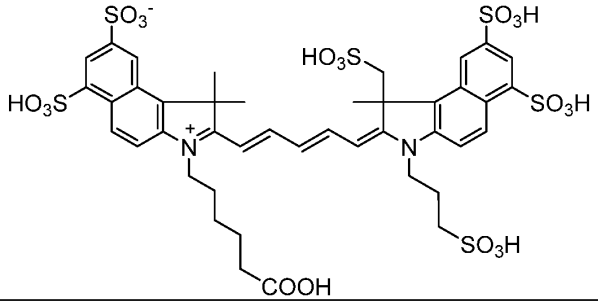
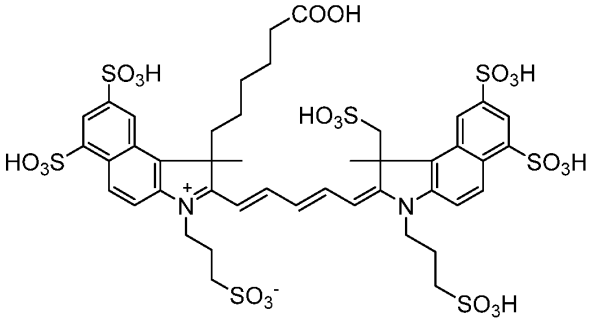
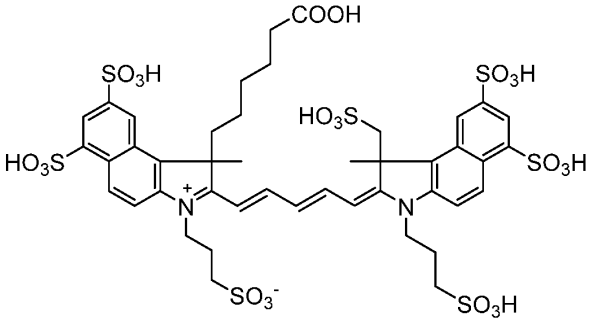
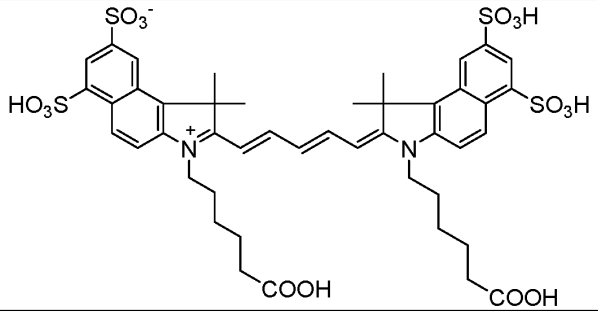
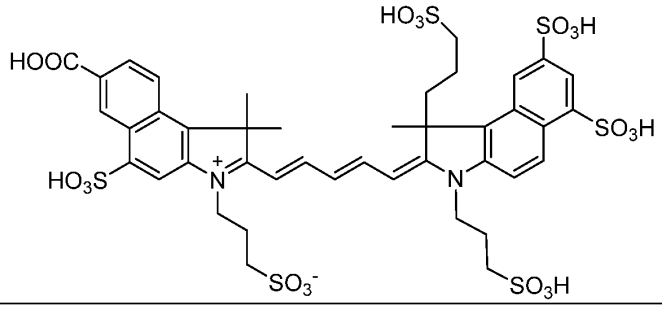
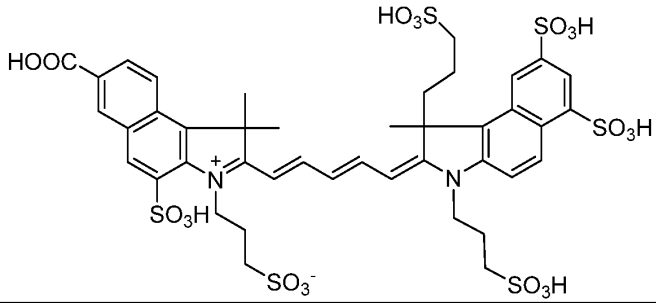
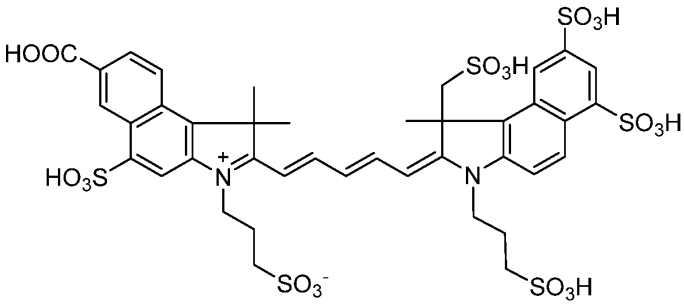
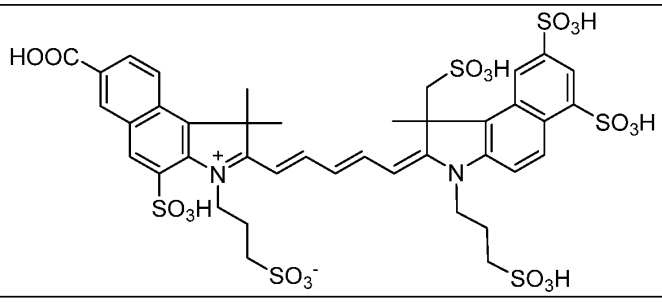
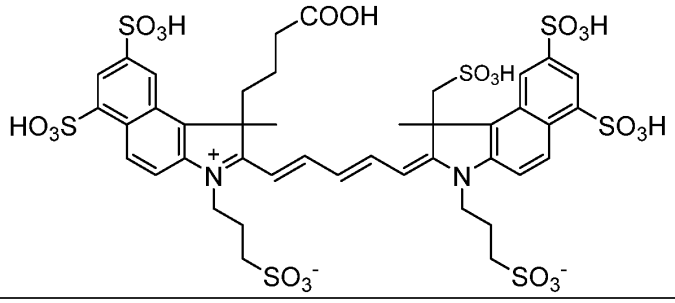
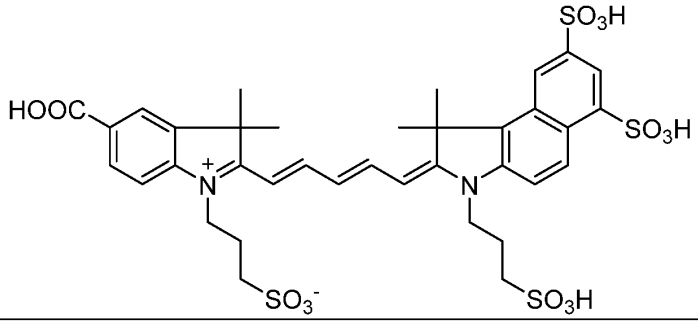
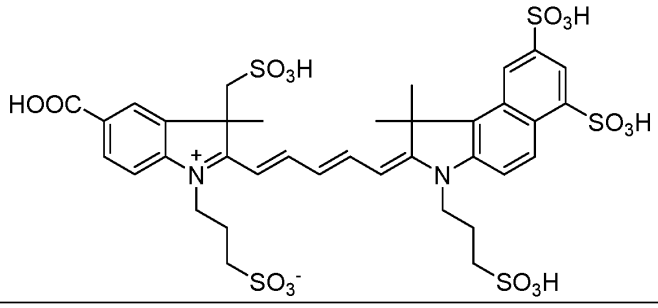
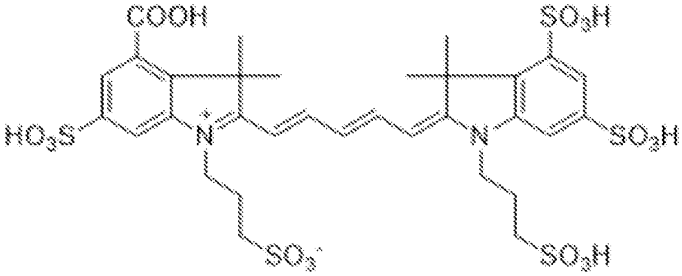
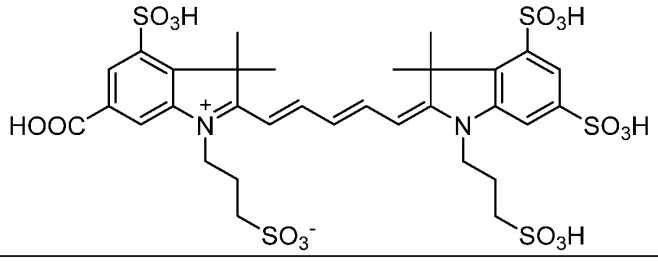
Compound No.	Structure
47	
48	
49	
50	

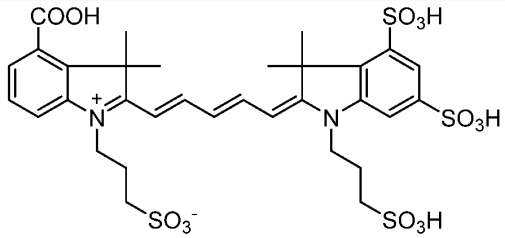
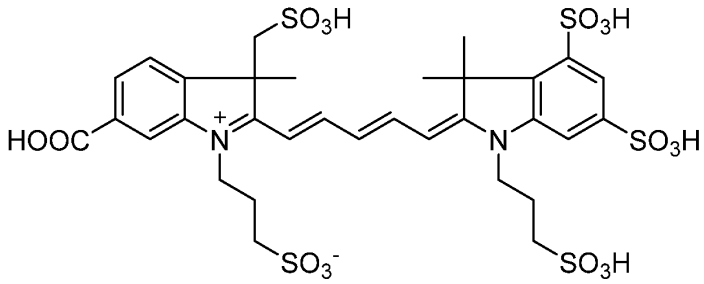
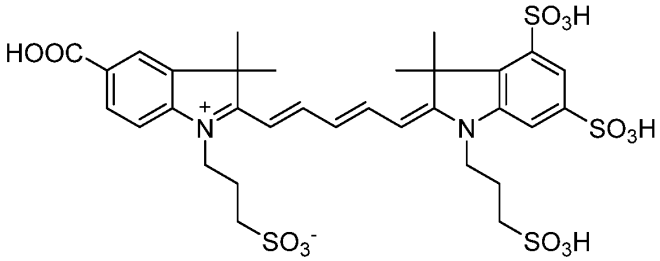
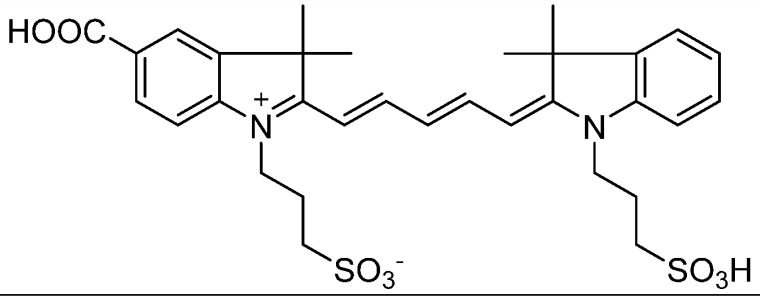
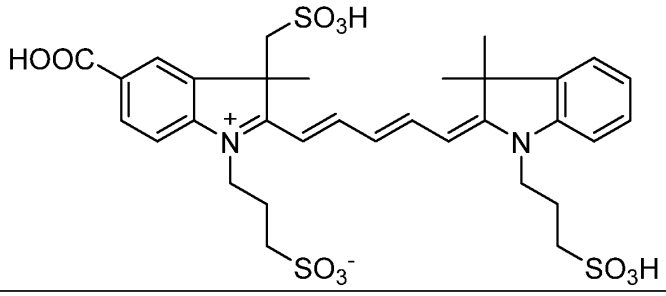
Table 2

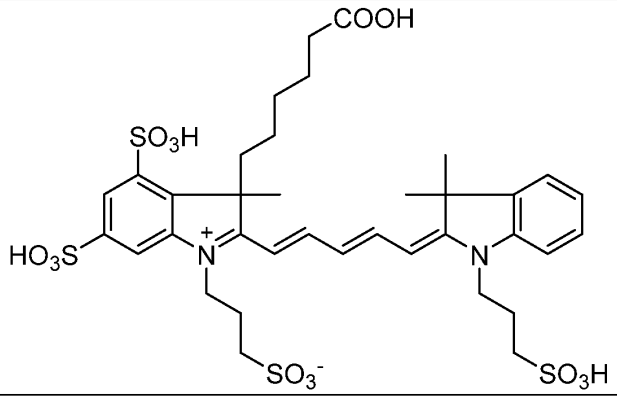
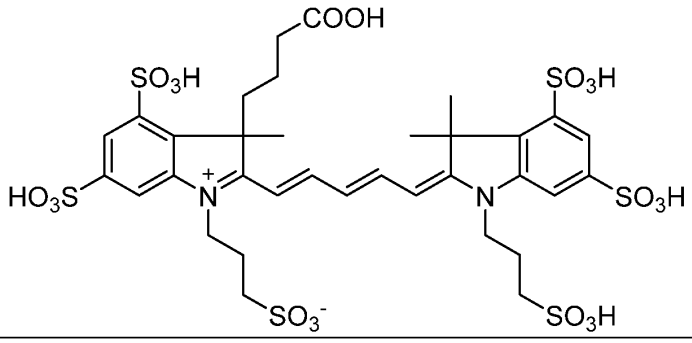
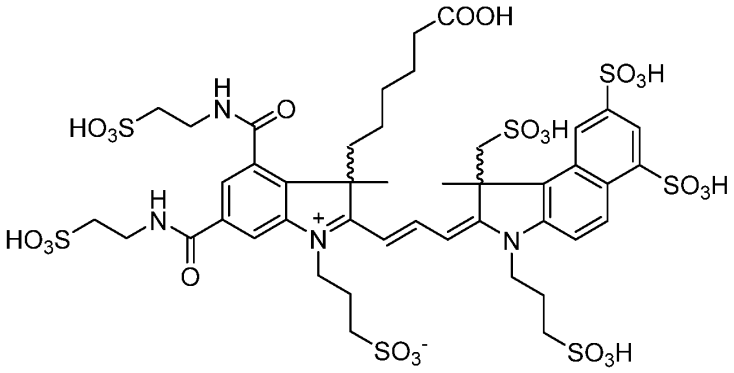
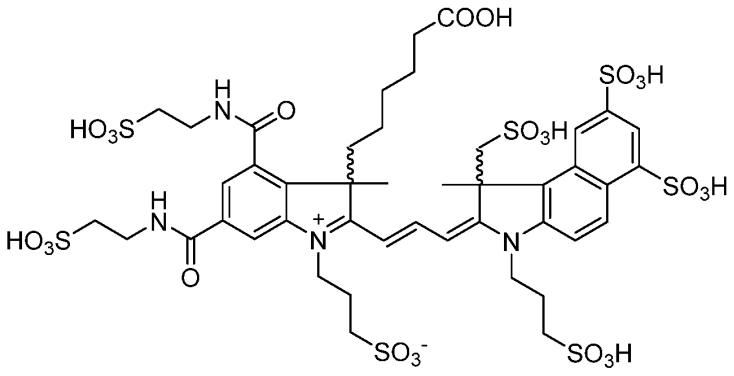
Compound No.	Structure
51	
52	
53	
54	

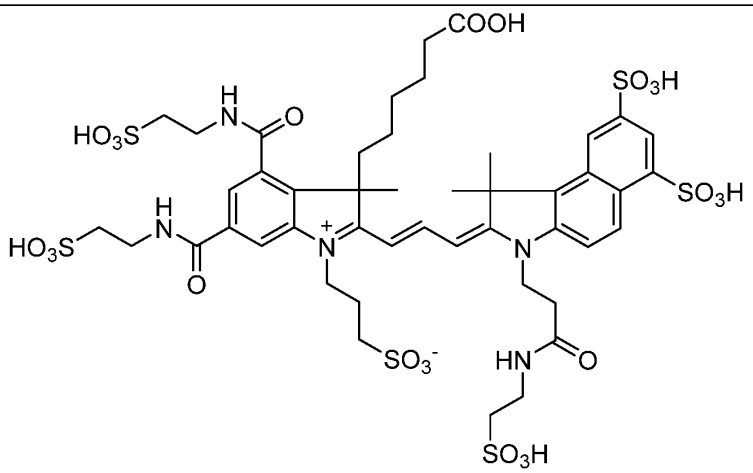
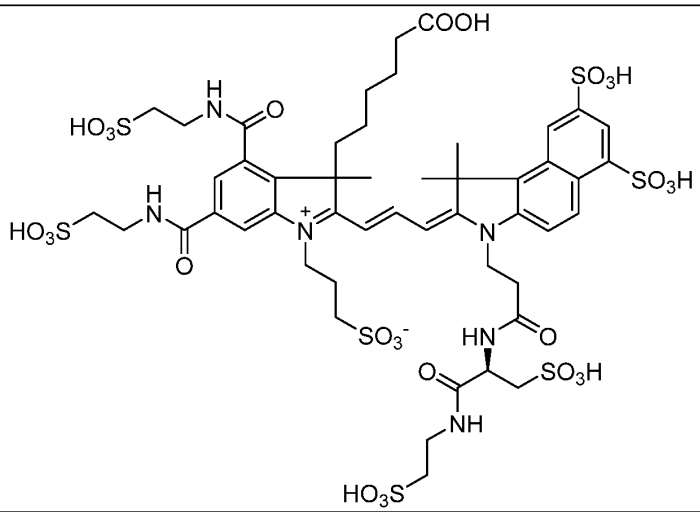
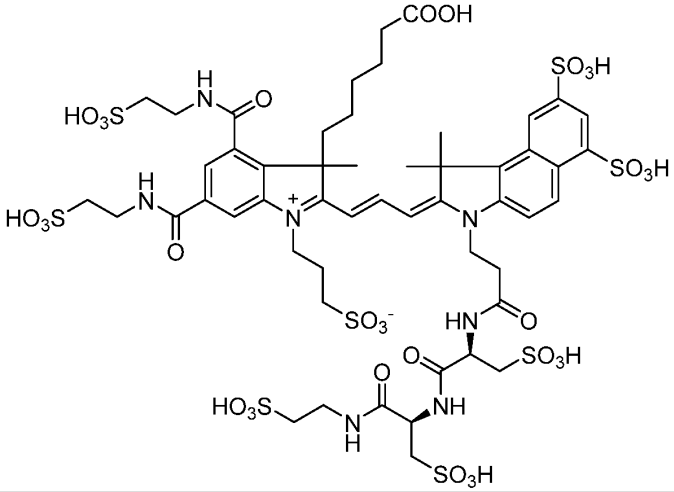
55	
56	
57	 Diastereomer 1
57a	 Diastereomer 2

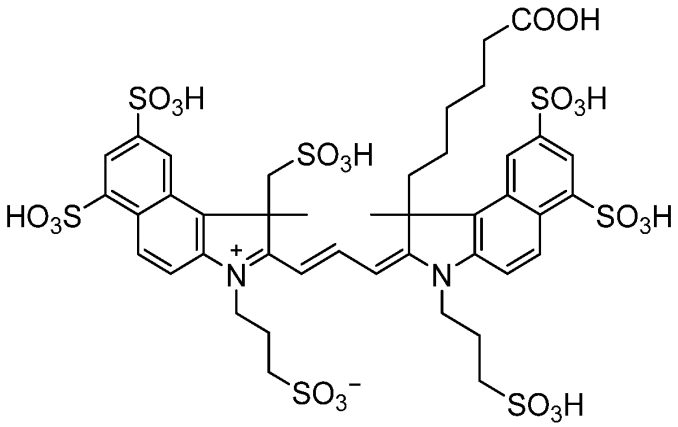
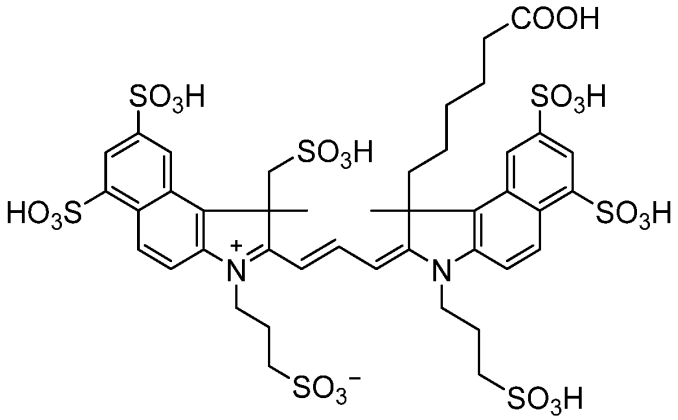
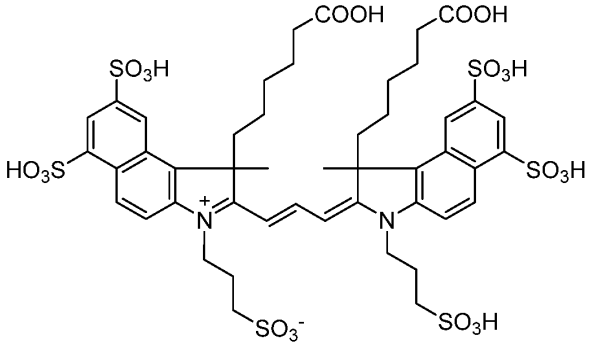
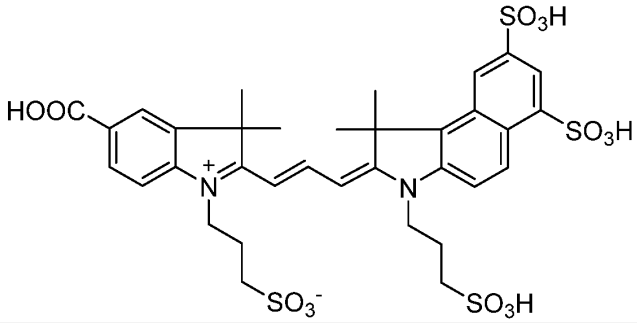
58	
59	
60	
61	
62	

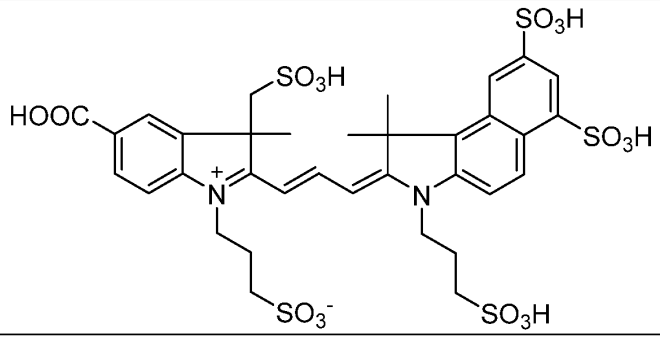
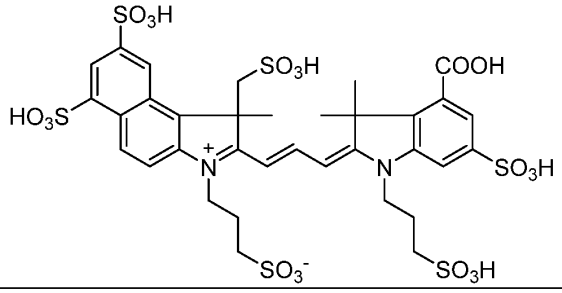
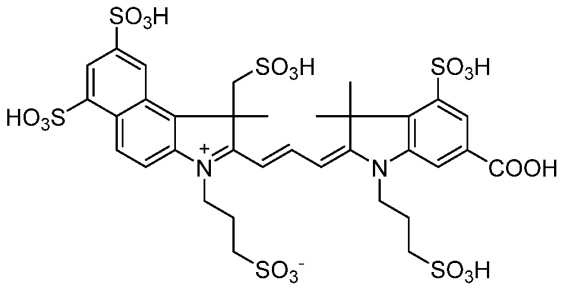
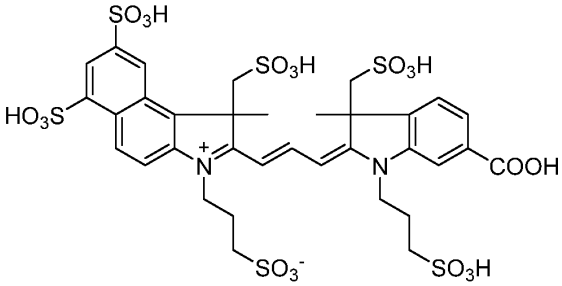
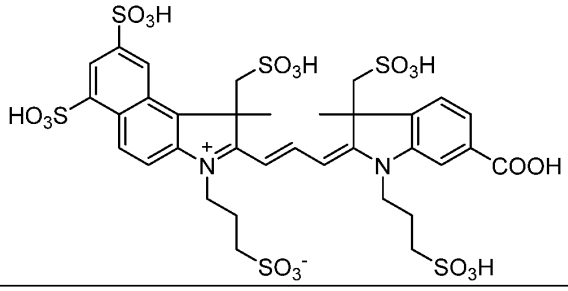
63	
64	
65	
66	
67	

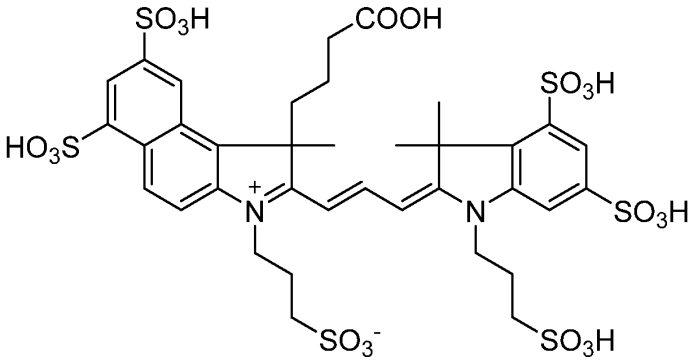
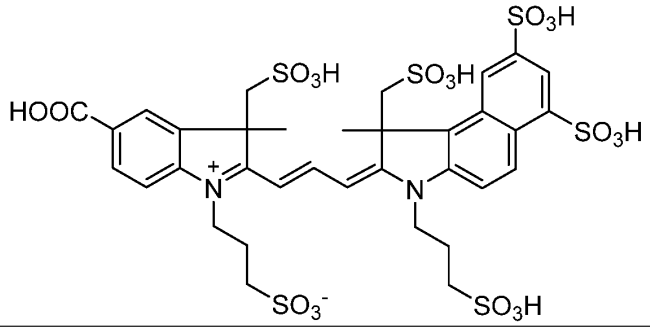
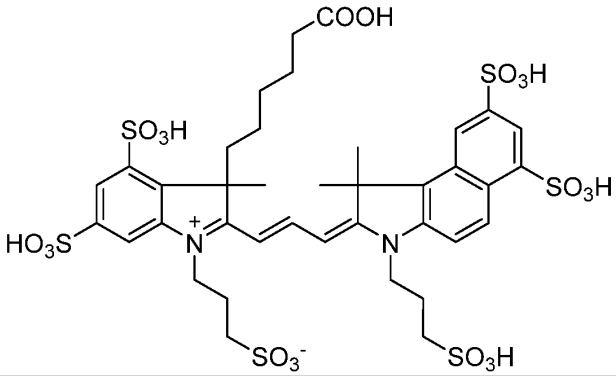
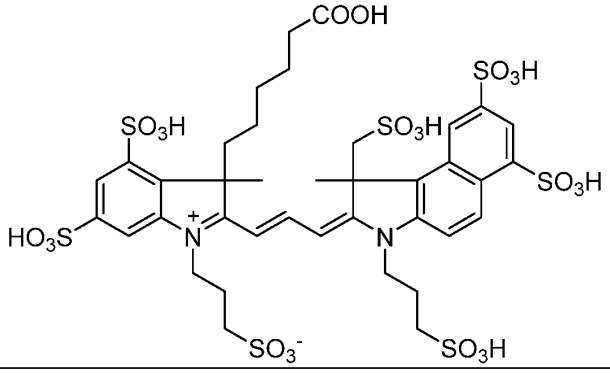
68	
69	
70	
71	
72	

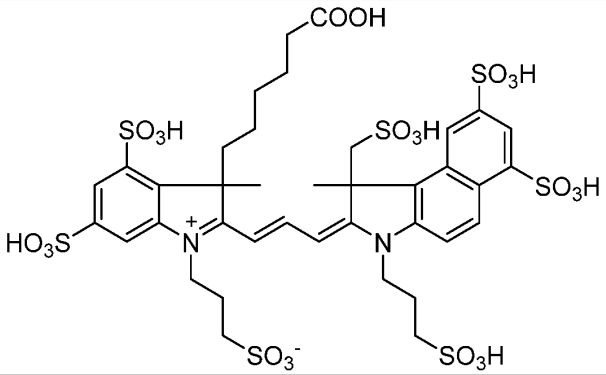
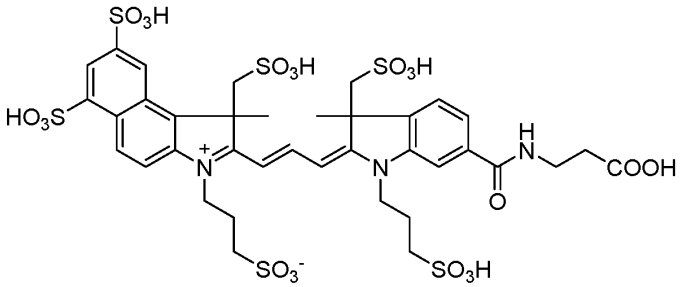
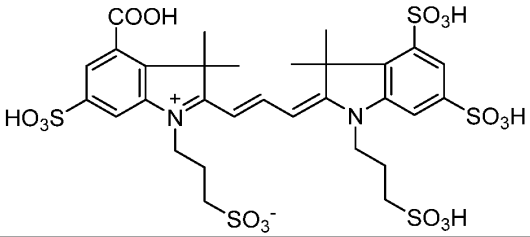
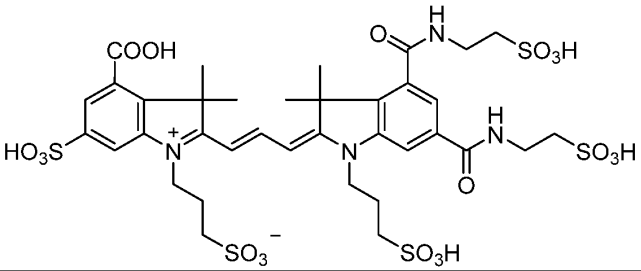
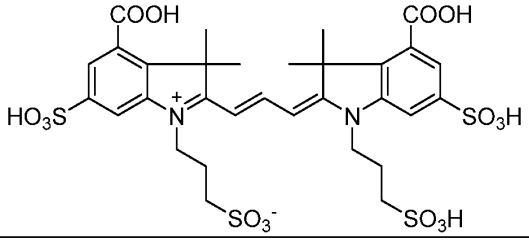
73	
74	
75	
76	

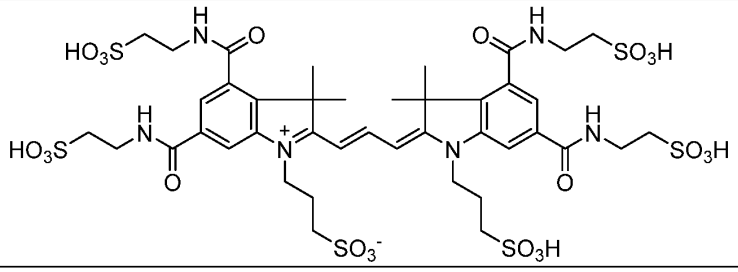
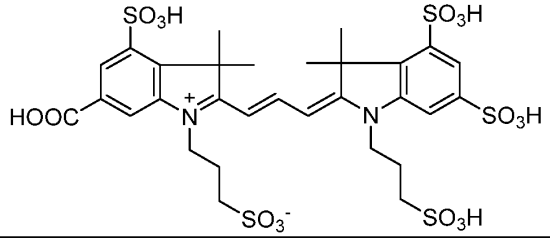
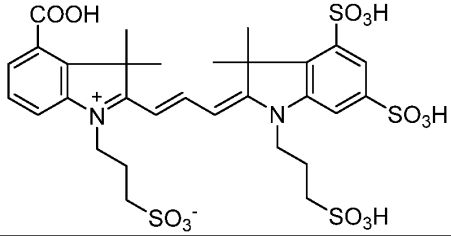
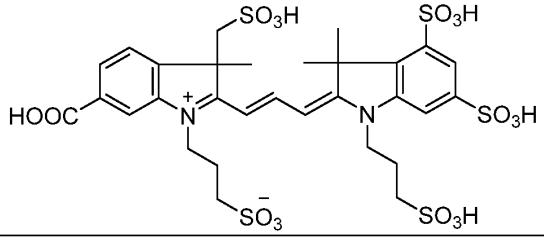
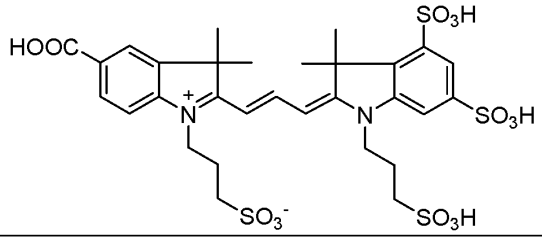
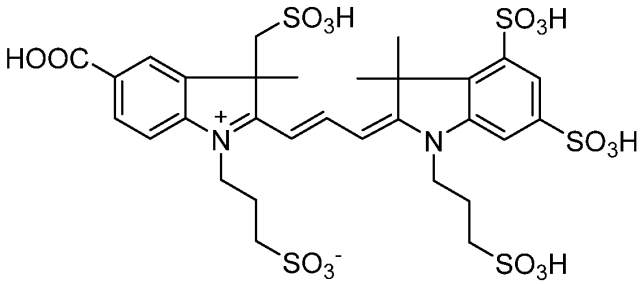
77	
78	
79	

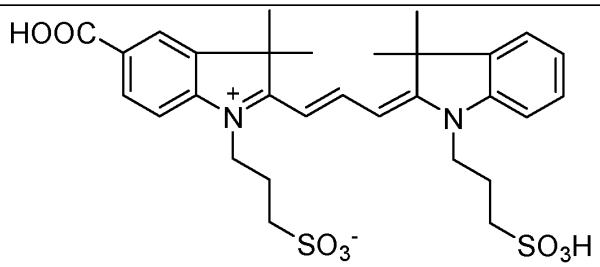
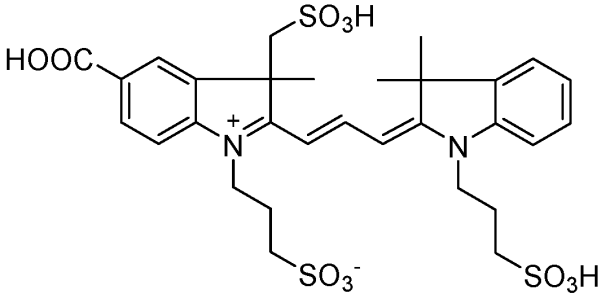
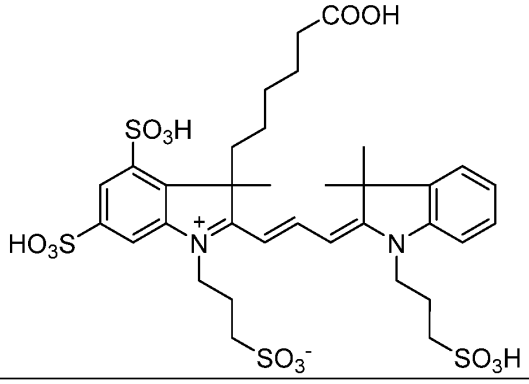
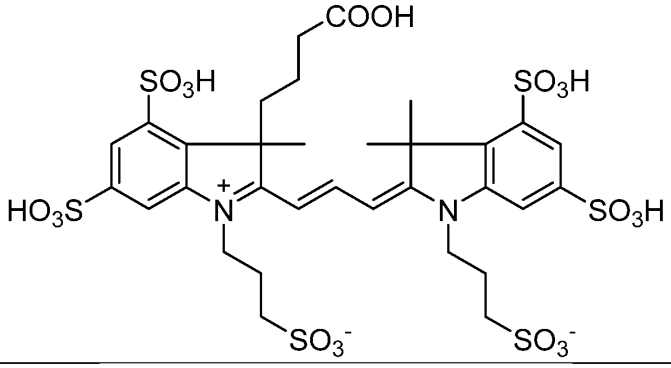
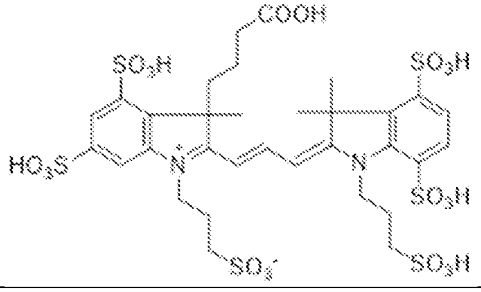
80	
81	
82	
83	

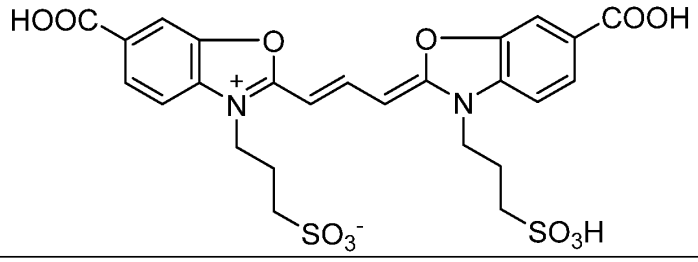
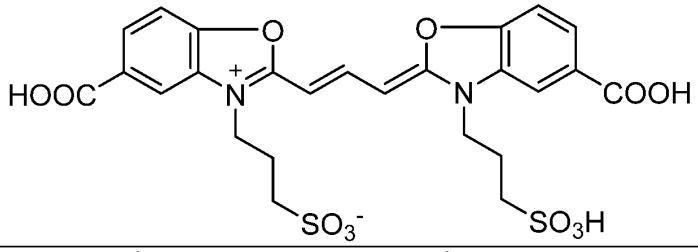
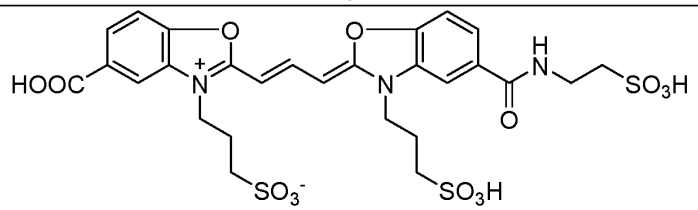
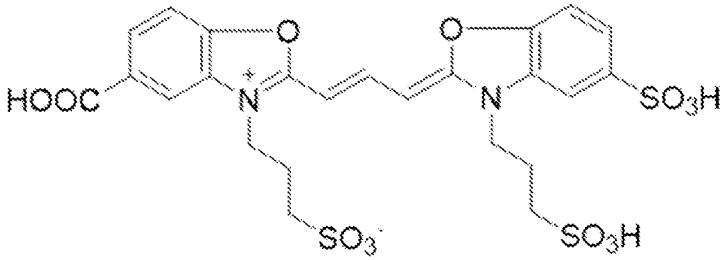
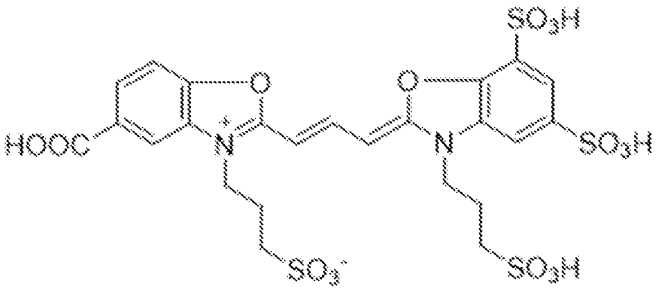
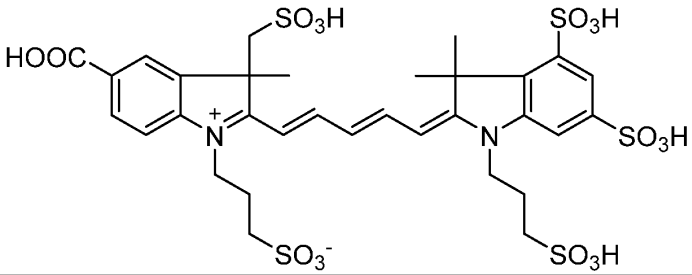
84	
85	
86	
87	
88	

89	
90	
91	
92	

93	
94	
95	
96	
97	

98	
99	
100	
101	
102	
103	

104	
105	
106	
107	
108	

109	
110	
111	
112	
113	
114	

Nucleotide Moiety

In some embodiments, the nucleotide moiety is a moiety that comprises a nucleotide and is capable of binding a second nucleotide, e.g., by way of a base-pairing interaction. For example, when used in a sequencing reaction, the nucleotide moiety enables the compositions of Formula (I) to engage in base-pairing interactions with nucleotides from the polynucleotide that is being sequenced.

In some embodiments, nucleotide moiety comprises a heteroarylbase, a five carbon sugar (e.g., ribose or deoxyribose), and one or more phosphate groups (e.g., 1-10 phosphate groups).

In some embodiments, the nucleotide moiety comprises a chain comprising 1-10 phosphorus atoms wherein the chain is attached to the 5' carbon of the sugar moiety by way of an ester or phosphoramide linkage. In some embodiments, at least one nucleotide moiety is a nucleotide analog comprises a phosphorus chain in which the phosphorus atoms are linked together by way of intervening -O-, -S-, -NH-, methylene or ethylene moieties. In some embodiments, the phosphorus atoms in the chain are optionally substituted by moieties comprising O, S or BH_3. In some embodiments, the phosphorous chain comprises phosphate group analogs, e.g., phosphoramidate, phosphorothioate, phosphordithioate, and O-methylphosphoroamidite groups.

In some embodiments, the nucleotide moiety is a nucleotide analog comprising a chain terminating moiety (e.g., a blocking moiety) at the sugar 2' position, at the sugar 3' position, or at the sugar 2' and 3' positions. In some embodiments, the nucleotide moiety comprises a chain terminating moiety (e.g., blocking moiety) at the sugar 2' position, at the sugar 3' position, or at the sugar 2' and 3' positions. In some embodiments, the chain terminating moiety inhibits polymerase-catalyzed incorporation of a subsequent nucleotide moiety or free nucleotide in a nascent strand during a primer extension reaction. In some embodiments, the sugar comprises a ribose or deoxyribose sugar moiety and the chain terminating moiety is attached to the 3' position of the ribose or deoxyribose moiety. In some embodiments, the chain terminating moiety is removable/cleavable from the 3' sugar position. In some embodiments, removal/cleavage of the chain terminating moiety generates a nucleotide having a 3'OH sugar group which is extendible with a subsequent nucleotide in a polymerase-catalyzed nucleotide incorporation reaction. In some embodiments, the chain terminating moiety comprises an alkyl group, alkenyl group, alkynyl group, allyl group, aryl group, benzyl group, azide group, amine group, amide group, keto group, isocyanate group, phosphate group, thio group, disulfide group, carbonate group, urea group, or silyl group.

In some embodiments, the chain terminating moiety is cleavable/removable from the nucleotide moiety, for example by reacting the chain terminating moiety with a chemical agent, pH change, light or heat. In some embodiments, this disclosure contemplates the removal of chain terminating moieties such as alkyl, alkenyl, alkynyl and allyl groups by treatment with a suitable set of reaction conditions, for example, tetrakis(triphenylphosphine)palladium(0) ($Pd(PPh_3)_4$) with piperidine, or with 2,3-Dichloro-5,6-dicyano-1,4-benzo-quinone (DDQ). In some embodiments, this disclosure contemplates the removal of the chain terminating moieties such as aryl and benzyl by way of treatment with a suitable set of reaction conditions, for example, hydrogenolysis (e.g., treatment with H_2) in the presence of a suitable catalyst, e.g., palladium supported on carbon (Pd/C). In some embodiments, this disclosure contemplates the removal of chain terminating moieties such as amine, amide, keto, isocyanate, phosphate, thio, or disulfide groups by treatment with a suitable set of reaction conditions, for example, treatment with phosphine or with a thiol group including beta-mercaptoethanol or dithiothritol (DTT). In some embodiments, the chain terminating moiety carbonate is cleavable with potassium carbonate (K_2CO_3) in MeOH, with triethylamine in pyridine, or with Zn in acetic acid (AcOH). In some embodiments, this disclosure contemplates the removal of chain terminating moieties such as urea and silyl by treatment with a suitable set of reaction conditions, for example, treatment with tetrabutylammonium fluoride, pyridine-HF, with ammonium fluoride, or with triethylamine trihydrofluoride.

In some embodiments, the nucleotide moiety comprises a chain terminating moiety (e.g., blocking moiety) at the sugar 2' position, at the sugar 3' position, or at the sugar 2' and 3' position. In some embodiments, the chain terminating moiety comprises an azide, azido or azidomethyl group. In some embodiments, the chain terminating moiety comprises a 3'-O-azido or 3'-O-azidomethyl group. In some embodiments, this disclosure contemplates the removal of chain terminating moieties such as azide, azido and azidomethyl group by treatment with a suitable set of reaction conditions, for example, with a phosphine compound. In some embodiments, the phosphine compound comprises a derivatized tri-alkyl phosphine moiety or a derivatized tri-aryl phosphine moiety. In some embodiments, the phosphine compound comprises Tris(2-carboxyethyl)phosphine (TCEP) or bis-sulfo triphenyl phosphine (BS-TPP) or Tri(hydroxyproyl)phosphine (THPP). In some embodiments, the cleaving agent comprises 4-dimethylaminopyridine (4-DMAP).

In some embodiments, the nucleotide moiety comprises a chain terminating moiety selected from a group consisting of 3'-deoxy nucleotides, 2',3'-dideoxynucleotides, 3'-methyl, 3'-azido, 3'-azidomethyl, 3'-O-azidoalkyl, 3'-O-ethynyl, 3'-O-aminoalkyl, 3'-O-fluoroalkyl,

3'-fluoromethyl, 3'-difluoromethyl, 3'-trifluoromethyl, 3'-sulfonyl, 3'-malonyl, 3'-amino, 3'-O-amino, 3'-sulfhydral, 3'-aminomethyl, 3'-ethyl, 3'butyl, 3'-*tert* butyl, 3'-Fluorenylmethyloxycarbonyl, 3' *tert*-Butyloxycarbonyl, 3'-O-alkyl hydroxylamino group, 3'-phosphorothioate, and 3-O-benzyl, or derivatives thereof.

In some embodiments, the nucleotide moiety comprises a nucleotide. In some embodiments, the nucleotide moiety comprises a nucleotide triphosphate. In some embodiments, the nucleotide moiety comprises adenosine (e.g., adenosine triphosphate). In some embodiments, the nucleotide moiety comprises guanosine (e.g., guanosine triphosphate). In some embodiments, the nucleotide moiety comprises thymidine (e.g., thymidine triphosphate). In some embodiments, the nucleotide moiety comprises cytosine (e.g., cytosine triphosphate).

In some embodiments, the nucleotide moiety is directly bonded to the polymeric side chain. In some embodiments, the nucleotide moiety further comprises a bivalent connection moiety which connects the nucleotide to the polymeric side chain.

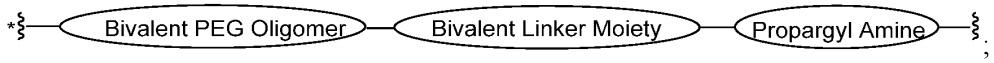
In some embodiments, the bivalent connection moiety comprises one or more PEG oligomer moieties. In some embodiments, the bivalent connection moiety comprises an oligomer of a PEG substitute, such as PEG (e.g., poly(vinyl alcohol) (PVA), poly(vinyl pyridine), poly(vinyl pyrrolidone) (PVP), poly(acrylic acid) (PAA), polyacrylamide, poly(N-isopropylacrylamide) (PNIPAM), PEG-acrylate, poly(methyl methacrylate) (PMA), poly(2-hydroxylethyl methacrylate) (PHEMA), poly(oligo(ethylene glycol) methyl ether methacrylate) (POEGMA), polyglutamic acid (PGA), or PEG-diacrylate.

In some embodiments, the bivalent connection moiety comprises a linker moiety. In some embodiments, the linker moiety comprises one or more PEG-moieties (e.g., a Bivalent PEG Oligomer). In some embodiments, the linker moiety comprises a Bivalent Linker Moiety. some embodiments, the linker moiety comprises (i) one or more PEG-moieties (e.g., a Bivalent PEG Oligomer) and (ii) a Bivalent Linker Moiety.

In some embodiments, the bivalent connection moiety comprises a propargyl amine moiety.

In some embodiments, the bivalent connection moiety comprises (i) one or more PEG-moieties (e.g., a Bivalent PEG Oligomer), (ii) a Bivalent Linker Moiety moiety, and (iii) a propargyl amine moiety.

In some embodiments, the bivalent connection moiety has the structure:

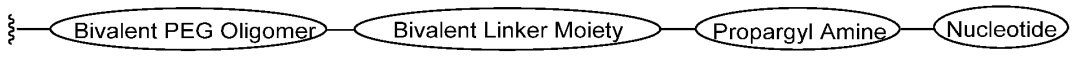


wherein *§ indicates connection to the polymeric side chain.

In some embodiments, the use of a sufficiently long Bivalent PEG Oligomer to link the nucleotide moiety with the remainder of the multivalent molecule has the benefits of preventing potential dye photochemistry from damaging enzyme and main DNA during the binding event between the multivalent molecule and the template nucleic acid molecule.

In some embodiments, attaching the nucleotide moiety to the remainder of the polymeric multivalent conjugate disclosed herein by way of a oligomeric chain comprising several hundred atoms or several thousand atoms, e.g., a PEG chain, may provide the nucleotide moiety sufficient flexibility to engage in binding, e.g., Watson-Crick base-pairing, with nucleotides in a polynucleotide chain, e.g., a polynucleotide chain to be sequenced.

In some embodiments, the Nucleotide Moiety has the structure:



.

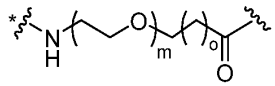
In some embodiments, the Bivalent PEG Oligomer moiety and polymeric side chain are joined by way of an amide bond. In some embodiments, the PEG oligomer moiety and polymeric side chain are joined by way of an ester bond. In some embodiments, the PEG oligomer and polymeric side chain are joined by way of an ether bond.

In some embodiments, the PEG oligomer moiety and linker moiety are joined by way of an amide bond. In some embodiments, the PEG oligomer moiety and linker moiety are joined by way of an ester bond. In some embodiments, the PEG oligomer and linker moiety are joined by way of an ether bond.

In some embodiments, the bivalent linker moiety and propargyl amine moiety are joined by way of an amide bond. In some embodiments, the bivalent linker moiety and propargyl amine moiety are joined by way of an ester bond. In some embodiments, the bivalent linker moiety and propargyl amine moiety are joined by way of an ether bond.

In some embodiments, the PEG oligomer moiety is linear. In some embodiments, the PEG oligomer is branched. In some embodiments, the PEG oligomer comprises a PEG chain that ranges between about 1,000 molecular weight (MW) (PEG 1000) and about 10,000 MW (PEG 10,000). In some embodiments, the PEG oligomer comprises a PEG chain that ranges between about PEG 2000 and about PEG 8000. In some embodiments, the PEG oligomer comprises a PEG chain that ranges between about PEG 3000 and about PEG 7000. In some embodiments, the PEG oligomer comprises a PEG chain that ranges between about PEG 4000 and about PEG 6000. In some embodiments, the PEG oligomer comprises a PEG chain that

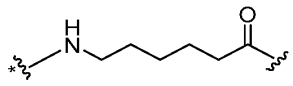
ranges between about PEG 4500 and about PEG 5500. In some embodiments, the PEG oligomer moiety comprises a PEG 1000 chain. In some embodiments, the PEG oligomer moiety comprises a PEG 2000 chain. In some embodiments, the PEG oligomer moiety comprises a PEG 3000 chain. In some embodiments, the PEG oligomer moiety comprises a PEG 4000 chain. In some embodiments, the PEG oligomer moiety comprises a PEG 5000 chain. In some embodiments, the PEG oligomer moiety comprises a PEG 6000 chain. In some embodiments, the PEG oligomer moiety comprises a PEG 7000 chain. In some embodiments, the PEG oligomer moiety comprises a PEG 8000 chain. In some embodiments, the PEG oligomer moiety comprises a PEG 9000 chain. In some embodiments, the PEG oligomer moiety comprises a PEG 10,000 chain. In some embodiments, the PEG oligomer moiety comprises a PEG 11,000 chain. In some embodiments, the PEG oligomer moiety comprises a PEG 12,000 chain. In some embodiments, the PEG oligomer moiety comprises a PEG 13,000 chain. In some embodiments, the PEG oligomer moiety comprises a PEG 14,000 chain. In some embodiments, the PEG oligomer moiety comprises a PEG 15,000 chain. In some embodiments, the PEG oligomer moiety comprises a PEG 16,000 chain. In some embodiments, the PEG oligomer moiety comprises a PEG 17,000 chain. In some embodiments, the PEG oligomer moiety comprises a PEG 18,000 chain. In some embodiments, the PEG oligomer moiety comprises a PEG 19,000 chain. In some embodiments, the PEG oligomer moiety comprises a PEG 20,000 chain. In some embodiments, the PEG oligomer moiety comprises a PEG 21,000 chain. In some embodiments, the PEG oligomer moiety comprises a PEG 22,000 chain. In some embodiments, the PEG oligomer moiety comprises a PEG 23,000 chain. In some embodiments, the PEG oligomer moiety comprises a PEG 24,000 chain. In some embodiments, the PEG oligomer moiety comprises a PEG 25,000 chain. In some embodiments, the PEG oligomer further comprises a -NH- moiety. In some embodiments, the -NH- moiety is bound to a carbonyl group that is a portion of the polymeric side chain. In some embodiments, the PEG oligomer further comprises a C_{1-10} alkyl moiety that terminates in a carbonyl group. In some embodiments, the C_{1-10} alkyl moiety that terminates in a carbonyl group joins the PEG oligomer moiety to the linker moiety.

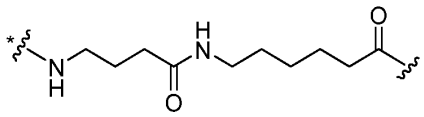
In some embodiments, the PEG oligomer moiety has the structure , wherein m is 2-2000 and o is 1-10 and * indicates attachment to the polymeric side chain. In some embodiments, the PEG oligomer moiety has the structure

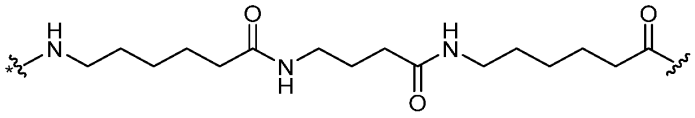


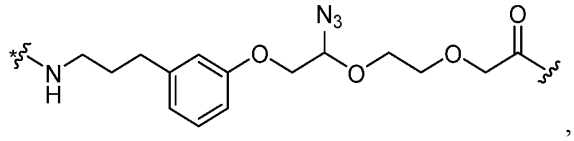
, wherein * indicates attachment to the polymeric side chain.

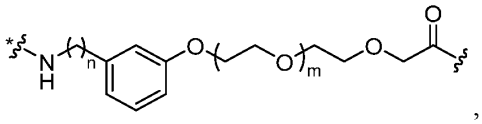
In some embodiments, the bivalent linker moiety comprises a linear or branched C_{1-20} alkyl group. In some embodiments, the bivalent linker moiety comprises a linear or branched C_{1-20} heteroalkyl group. In some embodiments, the bivalent linker moiety comprises from 1 to 20 cyclic moieties, wherein the cyclic moieties are linked to one another by way of chemical bonds. In some embodiments, the bivalent linker moiety comprises the structure

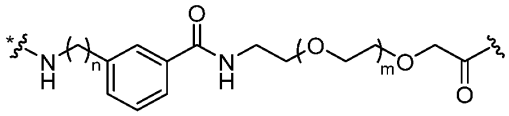

,

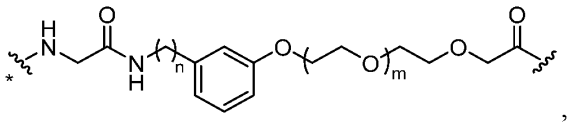

,

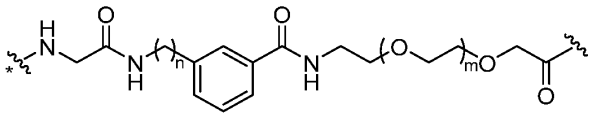

,

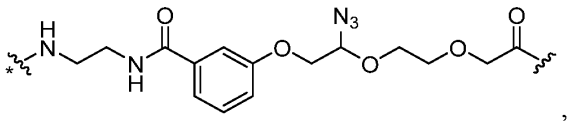

,

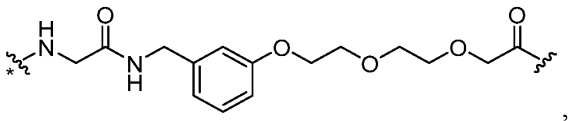

,

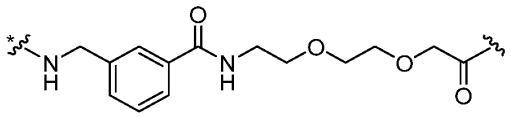

,


,


,

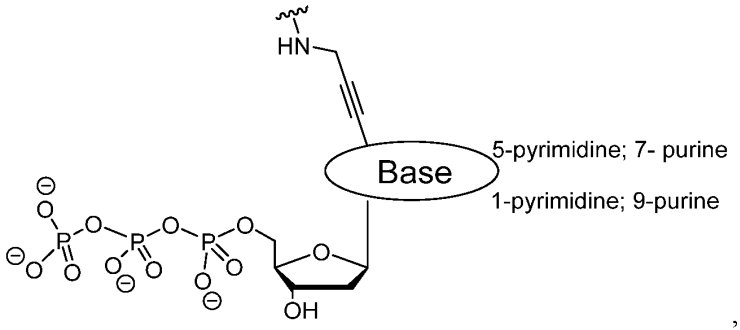

,


, or

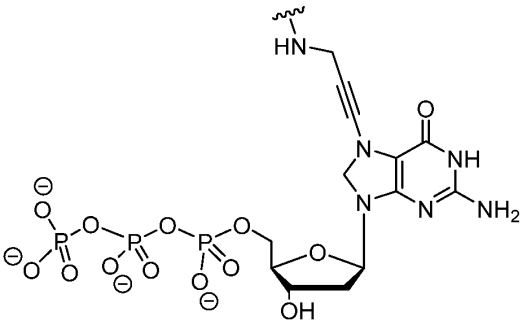

,

wherein m is 2-2000 and n is 1-10, and * indicates attachment to the polymeric side chain.

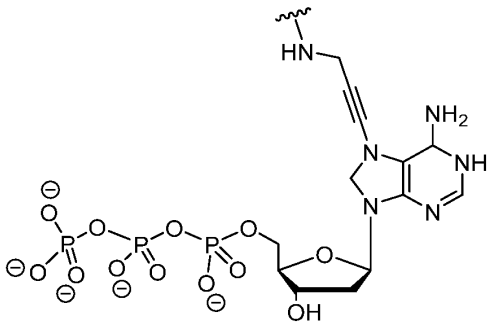
In some embodiments, the Nucleotide Moiety comprises the structure:



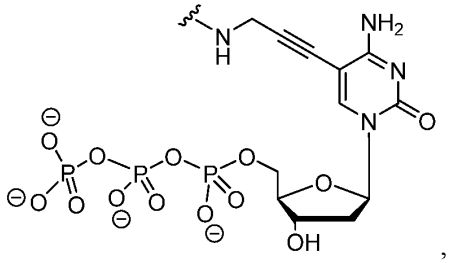
,



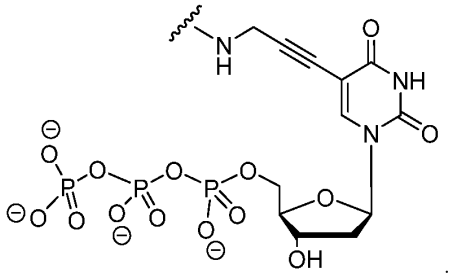
,



,

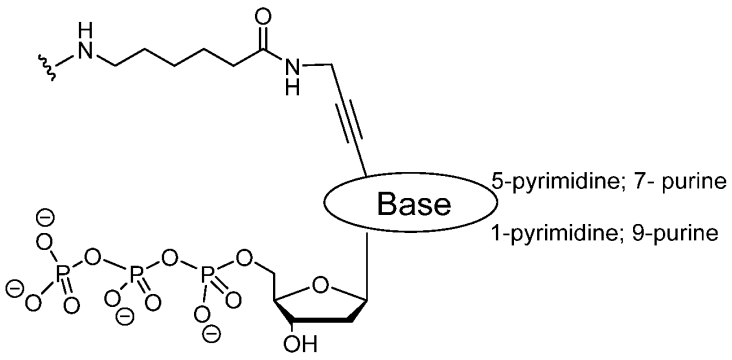


, or

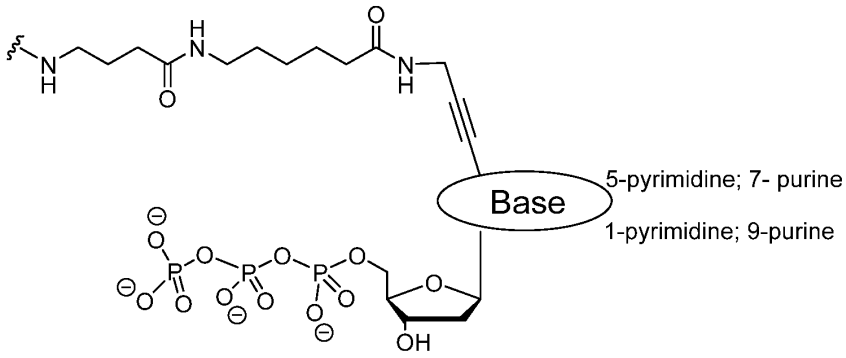


.

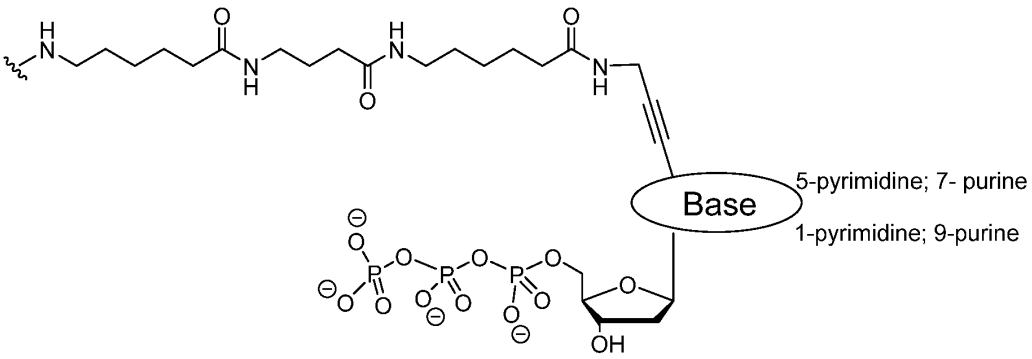
In some embodiments, the Nucleotide Moiety comprises the structure:



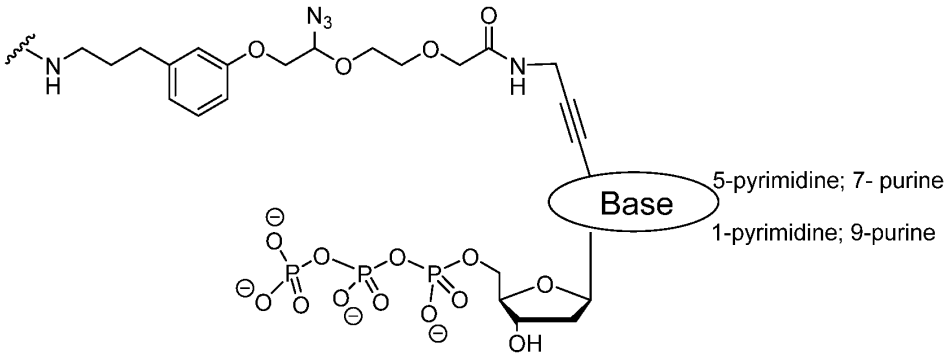
,



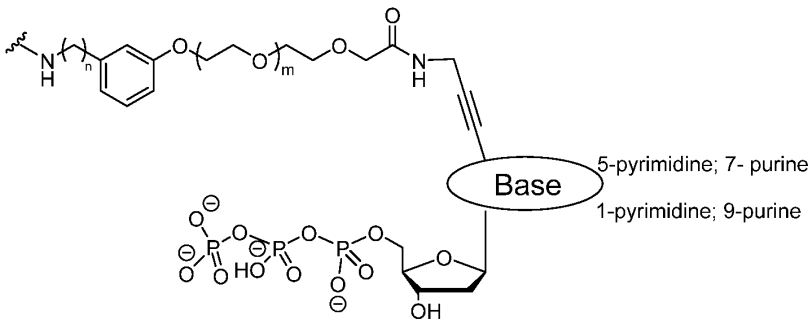
5 ,


Base
5-pyrimidine; 7- purine
1-pyrimidine; 9-purine
OH

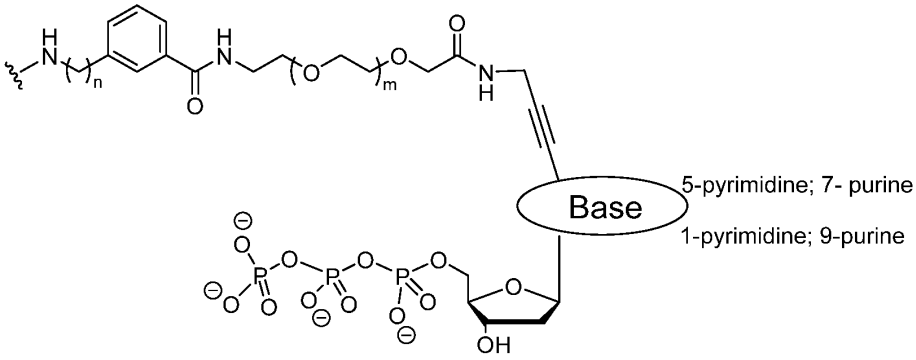
,


N3
Base
5-pyrimidine; 7- purine
1-pyrimidine; 9-purine
OH

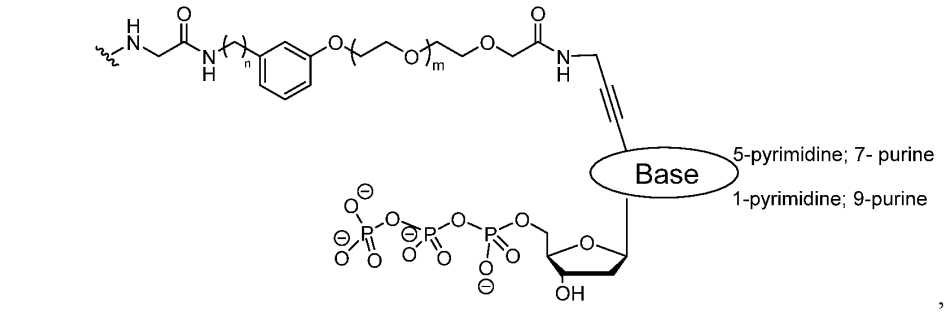
,


Base
5-pyrimidine; 7- purine
1-pyrimidine; 9-purine
HO
OH

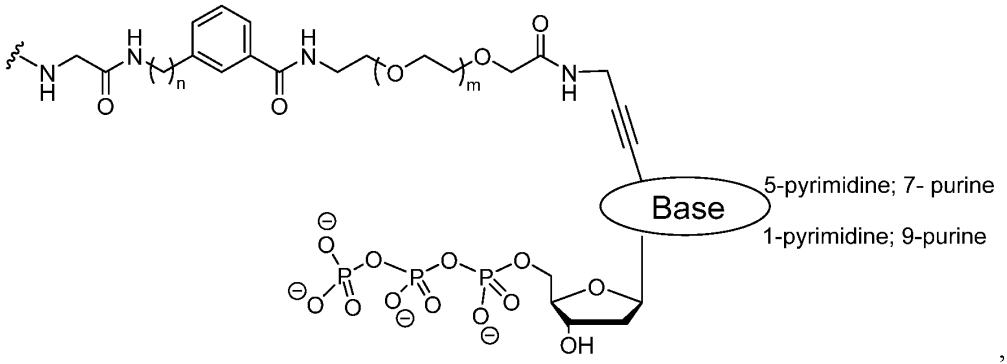
,


Base
5-pyrimidine; 7- purine
1-pyrimidine; 9-purine
OH

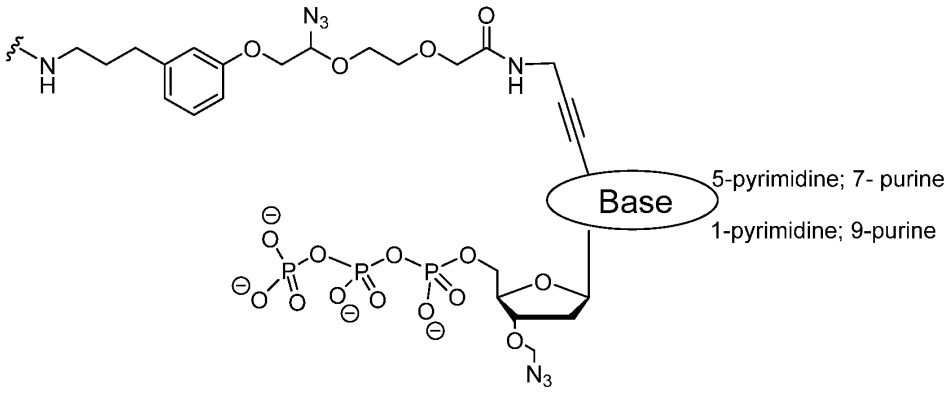
,



,



, or

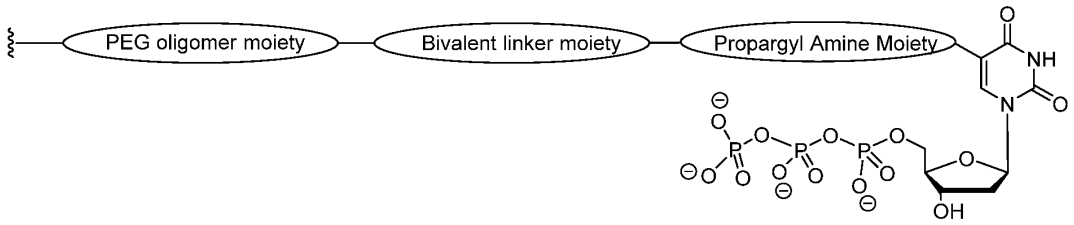


,

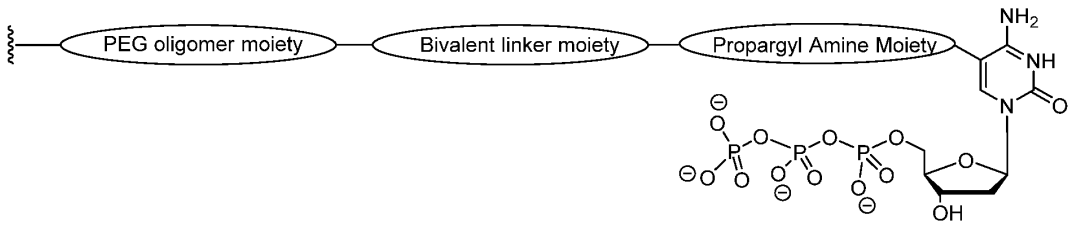
wherein m is 2-2000 and n is 1-10.

In some embodiments, the nucleotide is joined to the bivalent linker moiety by way of a propargyl amine linkage that attaches to the 5' position of a pyrimidine base or the 7' position of a purine base. In some embodiments, the propargyl amine linkages are stable, e.g., they are not cleavable under the conditions in which the polymeric multivalent conjugates are used in sequencing reactions. For example, polymeric multivalent conjugates wherein the nucleotide is attached by way of a propargyl amine linkage covalently bound to the nucleobase are more stable than polymeric multivalent conjugates wherein the nucleotide is attached way of a covalent bond to the triphosphate portion of the nucleotide. Linkages to the triphosphate portion of the nucleotide are more labile than the propargyl amine-type linkages to the nucleobase that are disclosed herein.

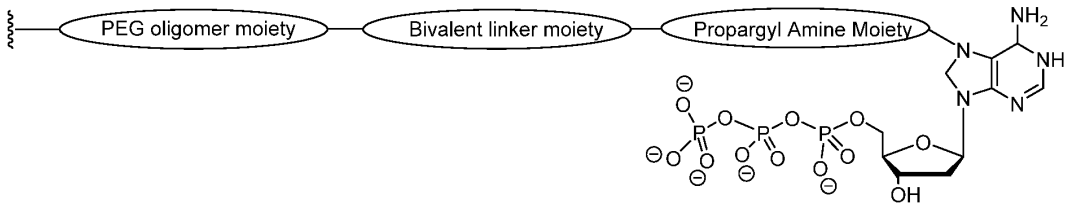
In some embodiments, the Nucleotide Moiety has the structure:



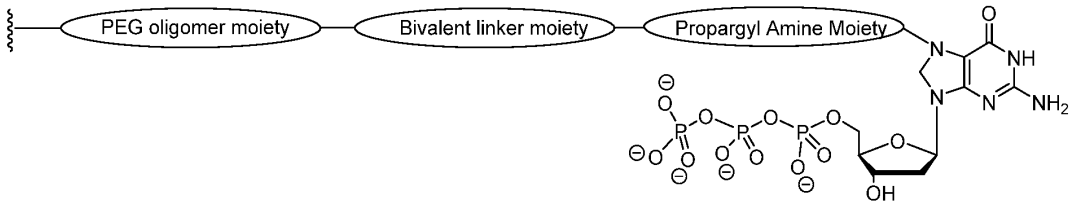
,



,



, or

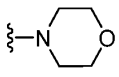


.

Blocking Moiety

In some embodiments, the blocking moiety is a moiety that is inert to further functionalization, e.g., inert to functionalization with a nucleotide moiety, reporter moiety, negative charge moiety, or PEG-Cap moiety. In some embodiments, the blocking moiety prevents the functionalization of the polymeric backbone with a nucleotide moiety, reporter moiety, negative charge moiety, or PEG-Cap moiety. The present disclosure contemplates that the degree of incorporation of blocking moieties may be tuned in order to modify the extent of functionalization of the polymeric side chain with alternative moieties, e.g., with nucleotide moieties, reporter moieties, negative charge moieties, or PEG-Cap moieties .

In some embodiments, the blocking moiety comprises an unreactive moiety (e.g., a non-nucleophilic, non-electrophilic, non-basic moiety). In some embodiments, the blocking moiety comprises a heterocyclyl, e.g., a non-reactive heterocyclyl. In some embodiments, the blocking moiety comprises a morpholinyl moiety.

In some embodiments, the blocking moiety comprises morpholine. In some embodiments, the blocking moiety comprises . In some embodiments, the blocking moiety is joined to the polymeric side chain by way of attachment to a carbonyl moiety of the polymeric side chain.

Negative Charge Moiety

In some embodiments, the negative charge moiety is a moiety that increases the concentration of negative charge on the polymeric side chain(s). In some embodiments, incorporation of negative charge moieties increases the degree of negative charge-charge repulsion on the polymeric side chain and/or between polymeric side chains.

Without wishing to be bound by theory, it is believed that increasing the concentration of negative charge on the polymeric side chain may have the effects of:

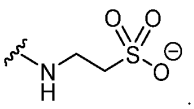
increasing the stiffness of the polymeric side chain(s);

decreasing the quenching of reporter moieties affixed to the polymeric side chain(s);

increasing water solubility of the composition; and

reducing hydrophobicity of the composition.

In some embodiments, the negative charge moiety comprises a functional moiety that is negatively charged, e.g., a functional moiety that is negatively charged at the pH at which the sequencing reaction takes place. In some embodiments, the negative charge moiety comprises a carboxylic acid moiety, a sulfonic acid moiety, or a phosphoric acid moiety. In some embodiments, the negative charge moiety comprises taurine. In some embodiments, the negative charge moiety comprises cystic acid. In some embodiments, the negative charge moiety comprises an amino acid. In some embodiments, the negative charge moiety comprises an amino carboxylic acid. In some embodiments, the negative charge moiety comprises an amino phosphate. In some embodiments, the negative charge moiety comprises an amino sugar.

In some embodiments, the negative charge moiety has the structure .

PEG-Cap

In some embodiments, the PEG-Cap moiety is a moiety that increases the degree of PEGylation on the polymeric side chain.

Without wishing to be bound by theory, it is believed that increasing the degree of PEGylation on the polymeric side chain may have the effects of:

increasing the stiffness of the polymeric side chain(s); and

reducing contact between the polymeric side chain(s) surfaces, e.g., the surface of a reaction vessel in which the sequencing reaction occurs.

In some embodiments, the PEG-Cap moiety comprises a PEG (polyethylene glycol) oligomer that terminates in a non-reactive "cap" moiety. The PEG oligomer may be of any suitable length. For example, in some embodiments, the PEG oligomer comprises In some embodiments, the "cap" moiety is an alkyl moiety, e.g., a C_{1-6} alkyl moiety. In some embodiments, the "cap" moiety is an alkoxy moiety, e.g., a C_{1-6} alkoxy moiety.

In some embodiments, the PEG oligomer is linear. In some embodiments, the PEG oligomer is branched. In some embodiments, the PEG oligomer comprises a PEG chain between about 1,000 molecular weight (MW) (PEG 1000) and about 10,000 MW (PEG 10,000). In some embodiments, the PEG oligomer comprises a PEG chain between about PEG 2000 and PEG 8000. In some embodiments, the PEG oligomer comprises a PEG chain between about PEG 3000 and PEG 7000. In some embodiments, the PEG oligomer comprises a PEG chain between about PEG 4000 and PEG 6000. In some embodiments, the PEG oligomer comprises a PEG chain between about PEG 4500 and PEG 5500. In some embodiments, the PEG oligomer moiety comprises a PEG 1000 chain. In some embodiments, the PEG oligomer moiety comprises a PEG 2000 chain. In some embodiments, the PEG oligomer moiety comprises a PEG 3000 chain. In some embodiments, the PEG oligomer moiety comprises a PEG 4000 chain. In some embodiments, the PEG oligomer moiety comprises a PEG 5000 chain. In some embodiments, the PEG oligomer moiety comprises a PEG 6000 chain. In some embodiments, the PEG oligomer moiety comprises a PEG 7000 chain. In some embodiments, the PEG oligomer moiety comprises a PEG 8000 chain. In some embodiments, the PEG oligomer moiety comprises a PEG 9000 chain. In some embodiments, the PEG oligomer moiety comprises a PEG 10,000 chain.

Polymeric Side Chain

The polymeric side chain may be synthesized by any suitable polymerization technique. In some embodiments, the polymeric side chain is synthesized by way of radical

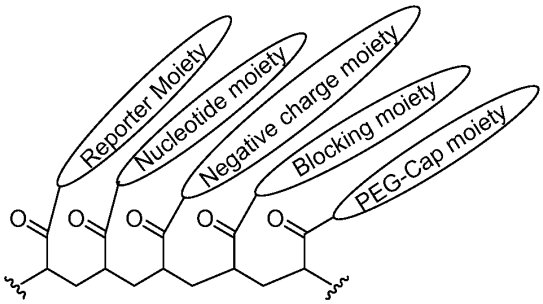
polymerization. In some embodiments, the polymeric side chain is synthesized by way of condensation polymerization. In some embodiments, the polymeric side chain is synthesized by way of Reversible Addition-Fragmentation Chain-Transfer (RAFT) polymerization.

In some embodiments, the polymeric side chain comprises polyethylene, wherein the ethylene monomers independently are optionally substituted with (i) a reporter moiety, (ii) nucleotide moiety, (iii), negative charge moiety, (iv) blocking moiety, or (v) PEG-Cap moiety. In some embodiments, the polymeric side chain comprises a homopolymer of ethylene monomers, wherein the acrylamide monomers are optionally substituted with (i) a reporter moiety, (ii) nucleotide moiety, (iii), negative charge moiety, (iv) blocking moiety, or (v) PEG-Cap moiety.

In some embodiments, the polymeric side chain comprises polyacrylamide, wherein the acrylamide monomers independently are optionally substituted with (i) a reporter moiety, (ii) nucleotide moiety, (iii), negative charge moiety, (iv) blocking moiety, or (v) PEG-Cap moiety. In some embodiments, the polymeric side chain comprises a homopolymer of acrylamide monomers, wherein the acrylamide monomers are optionally substituted with (i) a reporter moiety, (ii) nucleotide moiety, (iii), negative charge moiety, (iv) blocking moiety, or (v) PEG-Cap moiety.

In some embodiments, the polymeric side chain comprises poly-4-acryloylmorpholine. In some embodiments, the polymeric side chain comprises poly N-acryloxysuccinamide. The present disclosure contemplates the functionalization of a polymeric side chain comprising of poly N-acryloxysuccinamide by way of contacting the polymeric side chain comprising poly N-acryloxysuccinamide with various moieties that comprise an amino group, e.g., a reporter moiety comprising an amino group, a nucleotide moiety comprising an amino group, a negative charge moiety comprising an amino group, and/or a PEG-Cap moiety comprising an amino group.

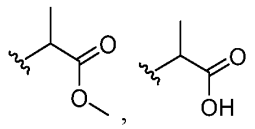
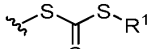
In some embodiments, the structure of an exemplary and non-limiting portion of a polymeric side chain is shown in Scheme PSC-1:



Scheme PSC-1.

End Moiety

In some embodiments, the polymeric side chain terminates in an end moiety. In some embodiments, the end moiety tunes the hydrophobicity or hydrophilicity of the polymeric side chain. Without wishing to be bound by theory, it is believed that the end moiety may modify the degree of interaction, e.g., binding, between the polymeric side chain and a surface, e.g., the surface of a flow cell in which a sequencing reaction takes place.

In some embodiments, the end moiety comprises a C_{1-6} alkyl group optionally substituted with a carboxylic acid moiety or ester moiety. In some embodiments, the end moiety comprises a trithiocarbonate optionally substituted with a C_{1-6} alkyl group, wherein the C_{1-6} alkyl group is branched or linear and is optionally substituted with a carboxylic acid moiety or ester moiety. In some embodiments, the end moiety has the structure , , , or , wherein R^1 is a C_{1-6} branched or linear alkyl group optionally substituted with a carboxylic acid moiety or an ester moiety.

In some embodiments, the end moiety is a nucleotide moiety.

Central Moiety

In some embodiments, the polymeric side chains are attached to a single central moiety.

In some embodiments, the central moiety is bivalent, trivalent, tetravalent, pentavalent, or hexavalent.

In some embodiments, the central moiety is a bond.

In some embodiments, the central moiety is inorganic. In some embodiments, the central moiety is a quantum dot. In some embodiments, the central moiety is a nanoparticle.

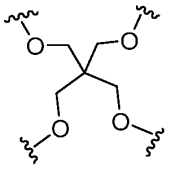
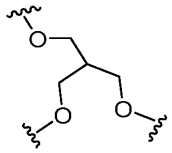
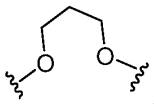
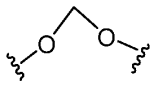
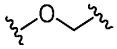
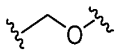
In some embodiments, the central moiety is a cyclic moiety. In some embodiments, the central moiety is a 5- to 10-membered heteroarylene. In some embodiments, the central moiety is an C_{6-10} arylene. In some embodiments, the central moiety is a 3- to 10-membered heterocyclene. In some embodiment, the central moiety is a C_{3-10} cycloalkylene.

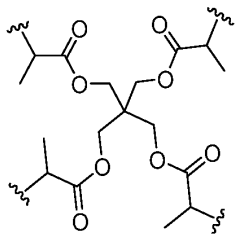
In some embodiments, the central moiety is an acyclic moiety. In some embodiments, the central moiety is a C_{1-20} alkyl moiety. In some embodiments, the central moiety is a C_{1-20} heteroalkyl moiety.

In some embodiments, the central moiety is C_{1-6} alkylene. In some embodiments, the central moiety is methylene, ethylene, propylene, butylene, pentylene, or hexylene.

In some embodiments, the central moiety is C_{2-6} alkynylene. In some embodiments, the central moiety is ethynylene, propynylene, butynylene, pentynylene, or hexynylene.

In some embodiments, the central moiety is C_{2-6} alkenylene. In some embodiments, the central moiety is ethenylene, propenylene, butenylene, pentenylene, or hexenylene.

In some embodiments, C is C_{1-6} heteroalkylene. In some embodiments, the central moiety is , , , , , or .

In some embodiments, the central moiety is .

In some embodiments, the central moiety is C_{6-10} arylene. In some embodiments, the central moiety is phenylene, napthylene, anthracenylene, phenanthrenylene, chrysenylene, pyrenylene, corannulenylene, coronenylene, or hexahelicenylene.

In some embodiments, the central moiety is 5- to 10-membered heteroarylene. In some embodiments, the central moiety is pyrrolylene, furanylene, thiophenylene, thiazolylene, isothiazolylene, imidazolylene, triazolylene, tetrazolylene, pyrazolylene, oxazolylene, isoxazolylene, isothiazolylene, pyridinylene, pyrazinylene, pyridazinylene, pyrimidinylene, benzoxazolylene, benzodioxazolylene, benzothiazolylene, benzoimidazolylene, benzothiophenylene, quinolinylene, isoquinolinylene, naphthrydinylene, indolylene, benzofuranylene, purinylene, benzofuranylene, deazapurinylene, or indolizinylene.

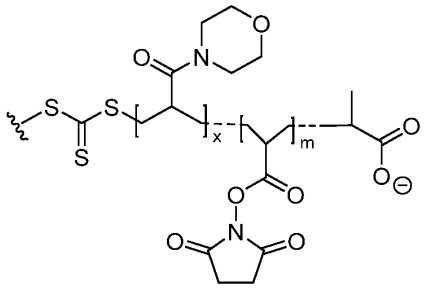
In some embodiments, the central moiety is C_{3-10} cycloalkylene. In some embodiments, the central moiety is cyclopropylene, cyclobutylene, cyclopentylene, cyclohexylene, cycloheptylene, cyclooctylene, or adamantylene.

In some embodiments, the central moiety is C_{5-10} cycloalkenylene. In some embodiments, the central moiety is cyclopentenylene, cyclohexenylene, cycloheptenylene, or 1,2,3,4-tetrahydronaphthalenylene.

In some embodiments, the central moiety is 3- to 10-membered heterocycloalkylene. In some embodiments, the central moiety is piperidinylene, piperazinylene, pyrrolidinylene, dioxanylene, tetrahydrofuranylene, isoindolinylene, indolinylene, imidazolidinylene, pyrazolidinylene, oxazolidinylene, isoxazolidinylene, triazolidinylene, oxiranylene, azetidinylene, oxetanylene, thietanylene, 1,2,3,6-tetrahydropyridinylene, tetrahydropyranylene, dihydropyranylene, pyranylene, morpholinylene, tetrahydrothiopyranylene, 1,4-diazepanylene, 1,4-oxazepanylene, 2-oxa-5-azabicyclo{2.2.1}heptanylene, 2,5-diazabicyclo{2.2.1}heptanylene, 2-oxa-6-azaspiro{3.3}heptanylene, 2,6-diazaspiro{3.3}heptanylene, 1,4-dioxa-8-azaspiro{4.5}decanylene, 1,4-dioxaspiro{4.5}decanylene, 1-oxaspiro{4.5}decanylene, 1-azaspiro{4.5}decanylene, 3'H-spiro{cyclohexane-1,1'-isobenzofuran}-ylene, 7'H-spiro{cyclohexane-1,5'-furo{3,4-b}pyridin}-ylene, 3'H-spiro{cyclohexane-1,1'-furo{3,4-c}pyridin}-ylene, 3-azabicyclo{3.1.0}hexanylene, 3-azabicyclo{3.1.0}hexan-3-ylene, 1,4,5,6-tetrahydropyrrolo{3,4-c}pyrazolylene, 3,4,5,6,7,8-hexahydropyrido{4,3-d}pyrimidinylene, 4,5,6,7-tetrahydro-1H-pyrazolo{3,4-c}pyridinylene, 5,6,7,8-tetrahydropyrido{4,3-d}pyrimidinylene, 2-azaspiro{3.3}heptanylene, 2-methylene-2-azaspiro{3.3}heptanylene, 2-azaspiro{3.5}nonanylene, 2-methylene-2-azaspiro{3.5}nonanylene, 2-azaspiro{4.5}decanylene, 2-methylene-2-azaspiro{4.5}decanylene, 2-oxa-azaspiro{3.4}octanylene, 2-oxa-azaspiro{3.4}octan-6-ylene, or 5,6-dihydro-4H-cyclopenta{b}thiophenylene.

Exemplary Embodiments of Polymeric Side Chain

In some embodiments, at least one P (e.g., each P) is:



or an ionic derivative thereof,

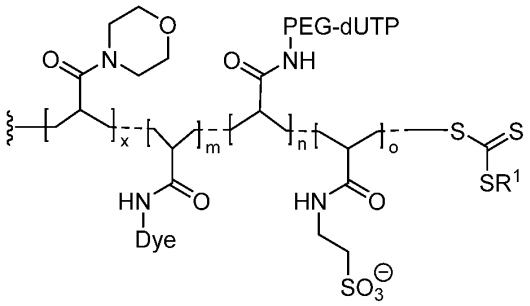
wherein:

x is 2-500; and

m is 2-500;

---- is a non-functionalized portion of the polymeric side chain.

In some embodiments, at least one P (e.g., each P) is:



or an ionic derivative thereof,

wherein:

x is 2-500;

m is 2-500;

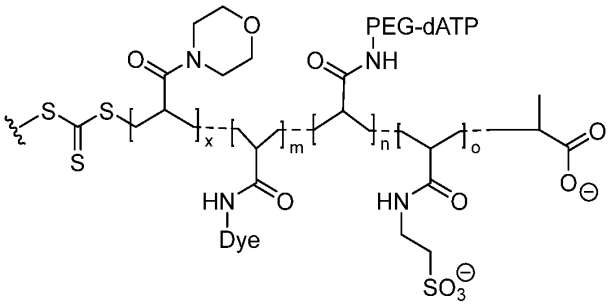
n is 2-500;

o is 2-500;

---- is a non-functionalized portion of the polymeric side chain; and

R^1 is a C_{1-6} branched or linear alkyl group optionally substituted with a carboxylic acid moiety or an ester moiety.

In some embodiments, at least one P (e.g., each P) is:



or an ionic derivative thereof,

wherein:

x is 2-500;

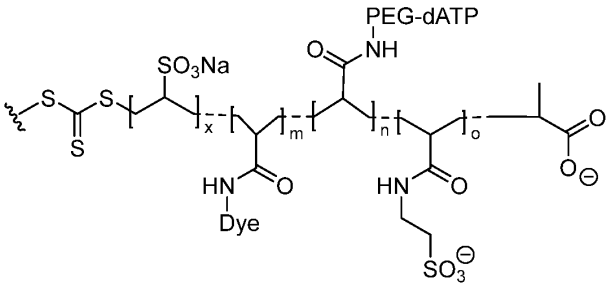
m is 2-500;

n is 2-500;

o is 2-500; and

---- is a non-functionalized portion of the polymeric side chain.

In some embodiments, at least one P (e.g., each P) is:



or an ionic derivative thereof,

wherein:

x is 2-500;

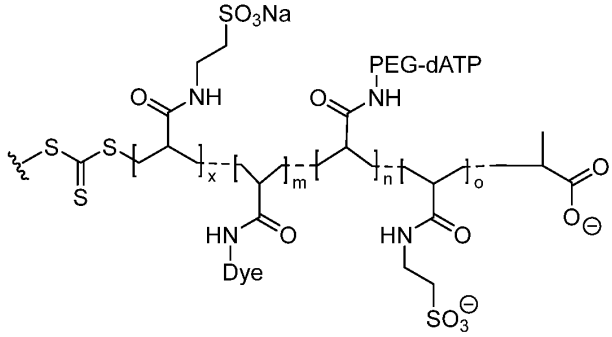
m is 2-500;

n is 2-500;

o is 2-500; and

---- is a non-functionalized portion of the polymeric side chain.

In some embodiments, at least one P (e.g., each P) is:



or an ionic derivative thereof,

wherein:

x is 2-500;

m is 2-500;

n is 2-500;

is 2-500; and

---- is a non-functionalized portion of the polymeric side chain.

In some embodiments of the polymeric side chains of the present disclosure, m is 2-2000, e.g., 2-1500, 2-1000, 2-950, 2-900, 2-850, 2-800, 2-750, 2-700, 2-650, 2-600, 2-550, 2-500, 2-450, 2-400, 2-350, 2-300, 2-250, 2-200, 2-150, 2-100, 2-50, or 2-10.

In some embodiments of the polymeric side chains of the present disclosure, x is 2-2000, e.g., 2-1500, 2-1000, 2-950, 2-900, 2-850, 2-800, 2-750, 2-700, 2-650, 2-600, 2-550, 2-500, 2-450, 2-400, 2-350, 2-300, 2-250, 2-200, 2-150, 2-100, 2-50, or 2-10.

In some embodiments of the polymeric side chains of the present disclosure, n is 2-2000, e.g., 2-1500, 2-1000, 2-950, 2-900, 2-850, 2-800, 2-750, 2-700, 2-650, 2-600, 2-550, 2-500, 2-450, 2-400, 2-350, 2-300, 2-250, 2-200, 2-150, 2-100, 2-50, or 2-10.

In some embodiments of the polymeric side chains of the present disclosure, o is 2-2000, e.g., 2-1500, 2-1000, 2-950, 2-900, 2-850, 2-800, 2-750, 2-700, 2-650, 2-600, 2-550, 2-500, 2-450, 2-400, 2-350, 2-300, 2-250, 2-200, 2-150, 2-100, 2-50, or 2-10.

Pluralities of Polymeric Molecules

The disclosure provides pluralities of polymeric molecules, e.g., compositions of Formula (I), for use in the sequencing methods described herein. In some embodiments, individual polymeric molecules in the plurality comprise nucleotide moieties and detectable reporter moieties. In some embodiments, all of the nucleotide moieties in an individual polymeric molecule are the same. In some embodiments, all of the detectable reporter moieties in an individual polymeric molecules are the same. Through selection of nucleotide moieties and detectable reporter moieties, the detectable reporter moiety can be associated with a particular nucleotide moiety, and this relationship can be used to identify the corresponding complementary nucleobase identity in the target sequencing during the sequencing reactions described herein.

In some embodiments, all polymeric molecules with the same nucleotide moiety in the plurality comprise the same detectable reporter moiety. As a non-limiting example, a plurality of polymeric molecules of the disclosure can comprise four types of polymeric molecules: a first type comprising dATP nucleotide moieties (or analogs thereof), and a first detectable reporter moiety, a second type comprising dCTP nucleotide moieties (or analogs thereof), and a second detectable reporter moiety, a third type comprising dGTP nucleotide moieties (or analogs or derivatives thereof), and a third detectable reporter moiety, and a fourth type comprising dTTP nucleotide moieties (or analogs thereof), and a fourth detectable reporter moiety. In this example, the first, second, third and fourth detectable reporter moieties are different, and their identities can be determined simultaneously during a massively parallelized sequencing reaction., For example, the first, second, third and fourth detectable reporter moieties can be fluorescent labels with clearly separable emission spectra, such as far red, red, green and blue dyes. Selection of appropriate, and separable labels, for use in the methods described herein will be apparent to persons of ordinary sill in the art.

Sequencing Methods using Polymeric Molecules

The present disclosure provides compositions comprising polymeric molecules, e.g., compositions of Formula (I) described *supra*, and methods that employ the compositions of Formula (I) for sequencing a target sequence. Any suitable sequencing methods are envisaged as within the scope of the instant disclosure, including, but not limited to, two-stage sequencing methods, sequencing-by-binding methods and zero mode waveguide based sequencing methods. In some embodiments, the sequencing methods comprise paired end sequencing. Alternatively, the polymeric molecules can be used in single read sequencing methods.

Two-Stage Sequencing Methods

The methods described below can be used to sequence single-stranded template nucleic acid molecules. In some cases, the single-stranded template nucleic acid molecules comprise DNA. In some cases, the single-stranded template nucleic acid molecules are concatemers comprising two or more copies of a target sequence, and a binding site for a sequencing primer.

In some embodiments, the methods for sequencing comprise a two-stage sequencing reaction. The first stage comprises binding polymeric molecules and polymerases to a DNA duplex, one of whose strands is the template nucleic acid molecule, or depending upon which strand is being sequenced, its complement, to form a multivalent binding complex. The contacting takes place under conditions sufficient to inhibit polymerase-catalyzed nucleotide incorporation, and is followed by detecting the multivalent-complexed polymerases using the detectable reporter moieties. The polymeric molecules in the multivalent binding complex that forms comprise a nucleotide moiety complementary to a nucleotide in a template nucleic acid molecule adjacent to a duplex region. Detection of the detectable reporter moiety in the polymeric molecule allows for the identification complementary nucleotide in the template nucleic acid molecule. The second stage generally comprises conducting polymerase-catalyzed nucleotide incorporation, following dissociation of the complex formed in the first step. The first and second stages are repeated at least once, and usually until a desired read length is obtained.

First Stage Reaction (Complexes with Polymeric molecules)

In some embodiments, the methods for sequencing comprise a two-stage sequencing reaction. In some embodiments, the first stage comprises **step (a)** contacting (i) a plurality of template nucleic acid molecules comprising two or more copies of a target sequence, (ii) a plurality of forward sequencing primers, (iii) a plurality of first polymerases, and (iv) a plurality of polymeric molecules of Formula (I) under conditions sufficient to form a plurality of

multivalent binding complexes comprising a nucleic acid duplex between a template nucleic acid molecule and forward sequencing primer, a first polymerase, and a nucleotide moiety of a polymeric molecule that is complementary to a nucleotide in the template nucleic acid molecule immediately 3' of an end of the sequencing primer. In some embodiments, the plurality of template nucleic acid molecules comprising two or more copies of a target sequence and two or more copies of a binding sequence for a forward sequencing primer, e.g., as a concatemer, and the plurality of sequence primers comprise a sequence complementary to the binding sequence for the forward sequencing primer. In some embodiments, individual polymeric molecules in the plurality comprises at least two nucleotide moieties and at one least detectable reporter moiety. In some embodiments, polymerase catalyzed incorporation of a complementary nucleotide moiety into the nucleic acid duplex is inhibited. In some embodiments, the first stage comprises **step (b)** detecting the detectable reporter moieties. In some embodiments, the first stage comprises **step (c)** determining nucleobase identities of nucleotides in the nucleic acid template molecules complementary to the nucleotide moieties of the polymeric molecules based on the detectable reporter moieties of the polymeric molecules in the plurality of multivalent binding complexes formed in step (a). For example, individual polymeric molecules in the plurality can be designed such that the detectable reporter moieties of an individual polymeric molecule correspond to the identities of the nucleotide moieties in the same molecule, allowing for the identification of the complementary nucleotide in the template nucleic acid molecule through detection of the detectable reporter moiety.

In some embodiments, the first stage comprises **step (a):** contacting a plurality of a first polymerase with (i) a plurality of template nucleic acid molecules and (ii) a plurality of forward sequencing primers, wherein the contacting is conducted under conditions suitable to bind the plurality of first polymerases to the plurality of template nucleic acid molecules and the plurality of forward sequencing primers, thereby forming a plurality of first polymerase complexes, wherein individual complexes comprise a first polymerase bound to a nucleic acid duplex, and wherein the nucleic acid duplex comprises a template nucleic acid molecule hybridized to a forward sequencing primer. In some embodiments, the template nucleic molecules in the plurality of template nucleic acid molecules of step (a) comprise the same target sequence or different target sequences. In some embodiments, the plurality of template nucleic acid molecules and/or the plurality of forward sequencing primers of step (a) are in solution or are immobilized to a support. In some embodiments, for example those embodiments wherein the plurality of template nucleic acid molecules and/or the plurality of

sequencing primers are immobilized to a support, the binding with the first polymerase generates a plurality of immobilized first polymerase complexes. In some embodiments, the plurality of template nucleic acid molecules and/or sequencing primers are immobilized to 10^2 – 10^{15} different sites on a support. In some embodiments, the binding of the plurality of template nucleic acid molecules and sequencing primers with the plurality of first polymerases generates a plurality of first polymerase complexes immobilized to 10^2 – 10^{15} different sites on the support. In some embodiments, the plurality of immobilized first polymerase complexes on the support are immobilized to pre-determined or to random sites on the support. In some embodiments, the plurality of immobilized first polymerase complexes are in fluid communication with each other to permit flowing a solution of reagents (e.g., enzymes including sequencing polymerases, polymeric molecules, nucleotides, and/or divalent cations) onto the support so that the plurality of immobilized polymerase complexes on the support are reacted with the solution of reagents in a massively parallel manner.

In some embodiments, the methods for sequencing comprise **step (b)** contacting the plurality of first polymerase complexes with a plurality of polymeric molecules, e.g., compositions of formula (I), to form a plurality of multivalent binding complexes. In some embodiments, individual polymeric molecules in the plurality of polymeric molecules comprise a central moiety operably linked by way of a polymeric side chain to a plurality of nucleotide moieties. In some embodiments, the individual polymeric molecules comprise at least one detectable reporter moiety. In some embodiments, the contacting of step (b) is conducted under conditions suitable for binding complementary nucleotide moieties of the polymeric molecules to at least two of the plurality of first polymerase complexes thereby forming a plurality of multivalent binding complexes. In some embodiments, the complementary nucleotide moieties of the polymeric molecules bind to a complementary nucleotide in the template nucleic acid molecule that is immediately 3' of the sequencing primer. In some embodiments, the conditions are suitable for inhibiting polymerase-catalyzed incorporation of the complementary nucleotide moieties into the duplex, e.g. through a polymerase-catalyzed extension reaction.

In some embodiments, the methods for sequencing further comprise **step (c)** detecting a detectable reporter moiety of the polymeric molecules in the plurality of multivalent binding complexes. In some embodiments, the detecting includes detecting signals emitted by the detectable reporter moieties of the polymeric molecules that are bound to the first polymerases, where the nucleotide moieties of the polymeric molecules are bound to complementary nucleotides of the template nucleic acid molecules, but incorporation of the nucleotide moieties

is inhibited. In some embodiments, the polymeric molecules are labeled with a detectable reporter moiety to permit detection.

In some embodiments, the methods for sequencing further comprise **step (d)** identifying the nucleobase of the complementary nucleotide in the template nucleic acid molecules that are bound to the plurality of first polymerases, thereby determining the sequence of the template nucleic acid molecules. In some embodiments, the polymeric molecules are labeled with a detectable reporter moiety that corresponds to the particular nucleotide moieties of the individual polymeric molecule to permit identification of the complementary nucleotide moieties (e.g., nucleotide base adenine, guanine, cytosine, thymine or uracil) that are bound to the plurality of first pol In some embodiments, in the methods for sequencing, the plurality of template nucleic acid molecules comprise amplified template nucleic acid molecules (e.g., clonally amplified template molecules). In some embodiments, individual template nucleic acid molecules comprise two or more tandem copies of a target sequence and at least one universal primer binding site (e.g., concatemers). In some embodiments, individual template nucleic acid molecules comprise one copy of a target sequence and at least one universal primer binding site, such as a forward sequencing primer binding site. In some embodiments, individual nucleic acid template molecules comprise circularized nucleic acid molecules having at least one copy of a target sequence and at least one universal primer binding site. In some embodiments, the sequencing primer comprises an oligonucleotide having a 3' extendible end or a 3' non-extendible end.

In some embodiments, for example those embodiments where the template nucleic acid molecules comprise concatemers comprising two or more copies of a target sequence, the two or more copies of a target sequence in an individual of template nucleic acid molecule are the same target sequence, or substantially the same sequence. The person of ordinary skill in the art will understand that copies of target sequences can differ by, e.g., 1, 2, 3, 4, 5 or more nucleotides, depending on the length of the target sequence, and still be considered the same sequence. Such differences can be introduced during library preparation (e.g., by fragmentation or truncation), or during amplification reactions (e.g., by polymerase error).

In some embodiments, two or more multivalent binding complexes form on individual template nucleic acid molecules. For example, when the template nucleic acid molecules are concatemers of two or more copies of a sequence comprising (i) the binding sequence for the forward sequencing primer and (ii) the target nucleic acid sequence, duplexes can form at the multiple forward primer binding sites in the concatemer (see, e.g., FIG. 29). Accordingly, when the template nucleic acid molecules are contacted with the plurality of forward sequencing

primers and plurality of first polymerases, a plurality of multivalent binding complexes corresponding to copies of the concatemerized primer and target sequences can form. In some embodiments, for example those embodiments in which an individual polymeric molecule comprises two or more nucleotide moieties, the two or more nucleotide moieties in an individual polymeric molecule contact two or more different multivalent binding complexes on the same template nucleic acid molecule. In some embodiments, the two or more nucleotide moieties are the same, and contact the same nucleotide in the two or more copies of the target sequence in the concatemer template nucleic acid molecule.

In some embodiments, the plurality of template nucleic acid molecules and/or the plurality of sequencing primers are in solution or are immobilized on a support. In some embodiments, for example those embodiments wherein the plurality of template nucleic acid molecules and/or the plurality of sequencing primers are immobilized on a support, the binding with the first polymerases and the polymeric molecules generates a plurality of immobilized multivalent binding complexes. In some embodiments, the plurality of template nucleic acid molecules and/or sequencing primers are immobilized to $10^2 – 10^{15}$ different sites on a support. In some embodiments, the binding of the plurality of template nucleic acid molecules and sequencing primers with the plurality of first polymerases and plurality of polymeric molecules generates a plurality of multivalent binding complexes immobilized to $10^2 – 10^{15}$ different sites on the support. In some embodiments, the plurality multivalent binding complexes on the support are immobilized to pre-determined or to random sites on the support. In some embodiments, the plurality of multivalent binding complexes are in fluid communication with each other to permit flowing a solution of reagents (e.g., enzymes including sequencing polymerases, polymeric molecules, nucleotides, and/or divalent cations) onto the support so that the plurality of multivalent binding complexes on the support are reacted with the solution of reagents in a massively parallel manner.

In some embodiments, contacting the polymeric molecules with the plurality of first polymerases, and template nucleic acid molecules and sequencing primers occurs under conditions that inhibit polymerase-catalyzed incorporation of the nucleotide moieties of the polymeric molecules into the duplex. In some embodiments, the plurality of polymeric molecules comprise at least one polymeric molecule wherein the nucleotide moiety comprises a nucleotide analog. In some embodiments, the nucleotide analog comprises a chain terminating moiety at the sugar 2’ and/or 3’ position. In some embodiments, the plurality of polymeric molecules comprises at least one polymeric molecule comprising a nucleotide moiety that lacks a chain terminating moiety. In some embodiments, at least one of the

polymeric molecules in the plurality of polymeric molecules is labeled with a detectable reporter moiety that emits a signal. In some embodiments, the detectable reporter moiety comprises a fluorophore. In some embodiments, at least one of the polymeric molecules in the plurality of polymeric molecules is unlabeled (e.g., "dark"). In some embodiments, contacting the polymeric molecules with the plurality of first polymerases, and template nucleic acid molecules and sequencing primers, or duplexes, is the conducted in the presence of at least one non-catalytic cation which inhibits polymerase-catalyzed nucleotide incorporation, where the at least one non-catalytic cation comprises strontium, barium, calcium, scandium, titanium, vanadium, chromium, iron, cobalt, nickel, copper, zinc, gallium, germanium, arsenic, selenium, rhodium, europium and/or terbium.

In some embodiments, the plurality of forward sequencing primers are soluble.

In some embodiments, the first polymerases comprises wild type polymerases or recombinant mutant polymerases. In some embodiments, the first polymerases comprise phi29 DNA polymerases, large fragment of Bst DNA polymerases, large fragment of Bsu DNA polymerases (exo-), Bca DNA polymerases (exo-), Klenow fragment of E. coli DNA polymerases, T5 polymerases, M-MuLV reverse transcriptases, HIV viral reverse transcriptases, Deep Vent DNA polymerases or KOD DNA polymerases. In some embodiments, the first polymerases comprise at least one amino acid substitution that confers exonuclease-minus activity.

In some embodiments, the methods comprise dissociating the multivalent binding complexes under conditions sufficient to retain the nucleic acid duplexes, thereby generating a plurality of nucleic acid duplexes. In some embodiments, the methods comprise removing the plurality of first polymerases and the plurality of polymeric molecules, and retaining the plurality of nucleic acid duplexes. In some embodiments, a dissociating condition comprises contacting the multivalent binding complex with any one or any combination of a detergent, EDTA and/or water.

Second Stage Reaction (Extension)

In some embodiments, the methods for sequencing comprise a two-stage sequencing reaction. Following dissociation of the first polymerases and polymeric molecules used in the first stage from the duplex between the template nucleic acid molecule and the sequencing primer, the sequencing primer (or extended strand comprising the sequencing primer, if the template has already undergone one or more rounds of sequencing reactions) is extended in the second stage by one nucleotide by nucleotide incorporation into the duplex region.

In some embodiments, the methods comprise, at the second stage, contacting the plurality of nucleic acid duplexes with a plurality of second sequencing polymerases and a plurality of nucleotides or analogs thereof under conditions sufficient to incorporate nucleotides or analogs thereof complementary to the nucleotides of the template nucleic acid molecules immediately adjacent to the 3' ends of the forward sequencing primers in a primer extension reaction, thereby generating a plurality of extended nucleic acid duplexes comprising extended forward sequencing primer sequences. The person of skill in the art will understand that the nucleotides incorporated into the duplex, and extending the duplex region, are added to the 3' end of the sequencing primer (or extended sequencing primer) strand, and are complementary to the corresponding nucleotides in the template nucleic acid molecules.

In some embodiments, the second stage of the two-stage sequencing reaction comprises nucleotide incorporation. In some embodiments, the methods comprise **step (a)** contacting the plurality of the nucleic acid duplexes from the first stage that have been retained following dissociation of the first polymerases and polymeric molecules with a plurality of second polymerases. In some embodiments, the contacting is conducted under a condition suitable for binding the plurality of second polymerases to the plurality of the nucleic acid duplexes, thereby forming a plurality of second polymerase complexes comprising a second polymerase bound to a nucleic acid duplex.

In some embodiments, the methods comprise **step (b)** contacting the plurality of second polymerase complexes with a plurality of nucleotides or analogs thereof, wherein the contacting is conducted under conditions suitable for binding complementary nucleotides or analogs thereof from the plurality of nucleotides or analogs thereof to at least two of the second polymerase complexes. In some embodiments, the complementary nucleotides or analogs thereof bound by the second polymerase complexes comprise nucleotides complementary to a nucleotide of the template nucleic acid sequence immediately adjacent to the 3' end of the forward sequencing primer. In some embodiments, the contacting of step (b) is conducted under conditions suitable for promoting polymerase-catalyzed incorporation of the bound complementary nucleotides or analogs thereof into the duplex, thereby extending the sequencing primer by one nucleotide. In some embodiments, incorporating the nucleotides or analogs thereof into the 3' end of the sequencing primer in step (b) comprises a primer extension reaction.

In some embodiments, for example when the plurality of nucleotides or analogs thereof in step (b) are detectably labeled, the methods for sequencing further comprise **step (c)** detecting the complementary nucleotides or analogs thereof which are incorporated into the

primers (or extended primers). In some embodiments, the plurality of nucleotides or analogs thereof are labeled with a detectable reporter moiety to permit detection. In some embodiments, when the plurality of nucleotides or analogs thereof in step (b) are non-labeled, the detecting of step (c) is omitted.

In some embodiments, when the plurality of nucleotides or analogs thereof in step (b) are detectably labeled, the methods for sequencing further comprise **step (d)** identifying the nucleobases of the complementary nucleotides which are incorporated into the duplexes. In some embodiments, the identification of the incorporated nucleotides or analogs thereof in step (d) can be used to confirm the identity of the nucleotides of the polymeric molecules from the plurality of multivalent binding complexes in the first stage of the sequencing reaction. In some embodiments, the identifying of step (d) can be used to determine the sequence of the template nucleic acid molecules. In some embodiments, when the plurality of nucleotides or analogs thereof in step (b) are non-labeled, the identifying of step (d) is omitted.

In some embodiments, the methods comprise **step (e)** removing the chain terminating moiety from the incorporated nucleotide analogs when step (b) is conducted by contacting the plurality of second polymerase complexes with a plurality of nucleotide analogs that comprise at least one nucleotide having a 2' and/or 3' chain terminating moiety.

In some embodiments, the second polymerases comprises wild type polymerases or recombinant mutant polymerases. In some embodiments, the second polymerases comprise phi29 DNA polymerases, large fragment of Bst DNA polymerases, large fragment of Bsu DNA polymerases (exo-), Bca DNA polymerases (exo-), Klenow fragment of E. coli DNA polymerases, T5 polymerases, M-MuLV reverse transcriptases, HIV viral reverse transcriptases, Deep Vent DNA polymerases or KOD DNA polymerases. In some embodiments, the second polymerases comprise at least one amino acid substitution that confers exonuclease-minus activity. In some embodiments, the plurality of first polymerases comprise polymerases which have an amino acid sequence that is 100% identical to the amino acid sequence as the plurality of the second polymerases. In some embodiments, the plurality of first polymerases have an amino acid sequence that differs from the amino acid sequence of the plurality of the second polymerases.

In some embodiments, the contacting of the plurality of nucleic acid duplexes with the plurality of second polymerases and the plurality of nucleotides or analogs thereof occurs under conditions sufficient to incorporate nucleotides or analogs thereof into the duplex in a primer extension reaction. In some embodiments, the contacting is conducted in the presence of at

least one catalytic cation which promotes polymerase-catalyzed nucleotide incorporation, where the at least one catalytic cation comprises magnesium and/or manganese.

In some embodiments, the plurality of nucleotides or analogs thereof comprise one or more native nucleotides (e.g., non-analog nucleotides) or nucleotide analogs. In some embodiments, the plurality of nucleotides comprise a 2' and/or 3' chain terminating moiety which is removable or is not removable. In some embodiments, at least one of the nucleotides in the plurality is not labeled with a detectable reporter moiety (e.g., "dark"). In some embodiments, the plurality of nucleotides are non-labeled. In some embodiments, the plurality of nucleotides comprises a plurality of nucleotides or analogs thereof labeled with detectable reporter moiety. The detectable reporter moiety comprises a fluorophore. In some embodiments, the fluorophore is attached to the nucleotide base. In some embodiments, the fluorophore is attached to the nucleotide base with a linker which is cleavable/removable from the base or is not removable from the base. In some embodiments, the fluorophore is attached to the terminal phosphate group of the phosphate chain. In some embodiments, a particular detectable reporter moiety (e.g., fluorophore) that is attached to the nucleotide can correspond to the nucleotide base (e.g., dATP, dGTP, dCTP, dTTP or dUTP) to permit detection and identification of the nucleotide base.

In some embodiments the methods comprise dissociating the second polymerases from the extended nucleic acid duplexes under conditions sufficient to retain the plurality of extended nucleic acid duplexes. In some embodiments, a dissociating condition comprises contacting the second polymerases and extended nucleic acid duplexes with any one or any combination of a detergent, EDTA and/or water.

In some embodiments, the nucleotides or analogs thereof comprise a mixture of any combination of two or more types of nucleotides selected from the group consisting of dATP, dGTP, dCTP, dTTP and dUTP.

In some embodiments, one or more nucleotides in the plurality of nucleotides is a nucleotide analog comprising a chain terminating moiety (e.g., blocking moiety) at the sugar 2' position, at the sugar 3' position, or at the sugar 2' and 3' position. In some embodiments, the chain terminating moiety inhibits polymerase-catalyzed incorporation of a subsequent nucleotide moiety or free nucleotide in a nascent strand during a primer extension reaction. In some embodiments, the chain terminating moiety is attached to the 3' sugar position where the sugar comprises a ribose or deoxyribose sugar moiety. In some embodiments, the chain terminating moiety is removable/cleavable from the 3' sugar position to generate a nucleotide having a 3'OH sugar group which is extendible with a subsequent nucleotide in a polymerase-

catalyzed nucleotide incorporation reaction. In some embodiments, the chain terminating moiety comprises an alkyl group, alkenyl group, alkynyl group, allyl group, aryl group, benzyl group, azide group, amine group, amide group, keto group, isocyanate group, phosphate group, thio group, disulfide group, carbonate group, urea group, or silyl group. In some embodiments, the chain terminating moiety is cleavable/removable from the nucleotide moiety, for example by reacting the chain terminating moiety with a chemical agent, pH change, light or heat. In some embodiments, the chain terminating moieties alkyl, alkenyl, alkynyl and allyl are cleavable with tetrakis(triphenylphosphine)palladium(0) ($Pd(PPh_3)_4$) with piperidine, or with 2,3-Dichloro-5,6-dicyano-1,4-benzo-quinone (DDQ). In some embodiments, the chain terminating moieties aryl and benzyl are cleavable with H2 Pd/C. The chain terminating moieties amine, amide, keto, isocyanate, phosphate, thio, disulfide are cleavable with phosphine or with a thiol group including beta-mercaptoethanol or dithiothritol (DTT). The chain terminating moiety carbonate is cleavable with potassium carbonate (K_2CO_3) in MeOH, with triethylamine in pyridine, or with Zn in acetic acid (AcOH). The chain terminating moieties urea and silyl are cleavable with tetrabutylammonium fluoride, pyridine-HF, with ammonium fluoride, or with triethylamine trihydrofluoride.

In some embodiments, one or more nucleotides in the plurality of nucleotides comprises a chain terminating moiety (e.g., blocking moiety) at the sugar 2' position, at the sugar 3' position, or at the sugar 2' and 3' position. In some embodiments, the chain terminating moiety comprises an azide, azido or azidomethyl group. In some embodiments, the chain terminating moiety comprises a 3'-O-azido or 3'-O-azidomethyl group. The chain terminating moieties azide, azido and azidomethyl group are cleavable/removable with a phosphine compound. In some embodiments, the phosphine compound comprises a derivatized tri-alkyl phosphine moiety or a derivatized tri-aryl phosphine moiety. In some embodiments, the phosphine compound comprises Tris(2-carboxyethyl)phosphine (TCEP) or bis-sulfo triphenyl phosphine (BS-TPP) or Tri(hydroxyproyl)phosphine (THPP). In some embodiments, the cleaving agent comprises 4-dimethylaminopyridine (4-DMAP).

In some embodiments, the chain terminating moiety is selected from a group consisting of 3'-deoxy nucleotides, 2',3'-dideoxynucleotides, 3'-methyl, 3'-azido, 3'-azidomethyl, 3'-O-azidoalkyl, 3'-O-ethynyl, 3'-O-aminoalkyl, 3'-O-fluoroalkyl, 3'-fluoromethyl, 3'-difluoromethyl, 3'-trifluoromethyl, 3'-sulfonyl, 3'-malonyl, 3'-amino, 3'-O-amino, 3'-sulfhydral, 3'-aminomethyl, 3'-ethyl, 3'butyl, 3'-*tert* butyl, 3'- Fluorenylmethyloxycarbonyl, 3' *tert*-Butyloxycarbonyl, 3'-O-alkyl hydroxylamino group, 3'-phosphorothioate, and 3-O-benzyl, or derivatives thereof.

In some embodiments, the methods for sequencing further comprise repeating the steps of the first stage and the second stage, described *supra*, at least once. In some embodiments, the sequence of the template nucleic acid molecules can be determined by detecting and identifying the polymeric molecules that bind the sequencing polymerases but do not incorporate into the 3' end of the sequencing primer, or extended sequencing primer, . In some embodiments, the sequence of the template nucleic acid template molecule can be determined (or confirmed) by detecting and identifying the nucleotide that incorporates into the 3' end of the sequencing primer, or extended sequencing primer.

Repeating Complex Formation and Extension Steps

Following sequencing primer strand extension and dissociation of the second polymerases, the first and second stage sequencing reactions described *supra* can be repeated to determine the identity of the nucleotides in the nucleic acid template molecules located immediately 5' of the template nucleic acid nucleotides whose identity was determined in the previous sequencing reaction (i.e., the identity of the nucleotides and located immediately adjacent to the 3' end of the extended sequencing primer, on the complementary strand).

In some embodiments, the methods comprise repeating the two stages of the sequencing reaction at least once. In some embodiments, the methods comprise **step (a)** contacting the plurality of extended nucleic acid duplexes with a plurality of first polymerases and a plurality of polymeric molecules of Formula (I), or an ionized form, isomer, or salt thereof, under conditions sufficient to form a plurality of multivalent binding complexes comprising an extended nucleic acid duplex, a first polymerase, and a nucleotide moiety of a polymeric molecule that is complementary to a nucleotide in the template nucleic acid molecule immediately adjacent to the 3' end of the extended forward sequencing primer. The extended duplex is produced by incorporation of a nucleotide complementary to the template nucleic acid molecules at the 3' end of the sequencing primer, or extended sequencing primer (the "sequencing primer strand") during the previous second stage sequencing reaction. In some embodiments, the method comprises contacting the plurality of extended nucleic acid duplexes with a plurality of first polymerases and a plurality of polymeric molecules, wherein the two strands of the duplex have not been dissociated. In alternative embodiments, the duplexes may be partially or fully dissociated, and reform during the contacting step. In some embodiments, polymerase catalyzed incorporation of a complementary nucleotide moiety into the extended nucleic acid duplex is inhibited. In some embodiments, the methods comprise **step (b)** detecting the detectable reporter moieties. In some embodiments, the methods comprise **step**

(c) determining nucleobase identities of nucleotides in the nucleic acid template sequences complementary to the nucleotide moieties of the polymeric molecules based on the detectable reporter moieties of the polymeric molecules in the plurality of multivalent binding complexes formed in step (a). In some embodiments, the methods comprise **step (d)** dissociating the multivalent binding complexes under conditions sufficient to retain the plurality extended nucleic acid duplexes. In some embodiments the methods comprise **step (e)** contacting the plurality of extended nucleic acid duplexes with a plurality of second polymerases and a plurality of nucleotides or analogs thereof under conditions sufficient to incorporate nucleotides or analogs thereof complementary to the nucleotides of the nucleic acid template molecules immediately adjacent to the 3' of the ends of the extended forward sequencing primers in a primer extension reaction, thereby generating a plurality of extended nucleic acid duplexes comprising extended forward sequencing primers.

In some embodiments, the methods comprise repeating the two stages of the sequencing reaction at least once. In some embodiments, the first stage comprises **step (a)** contacting a plurality of a first polymerases with (i) a plurality of duplexes comprising a template nucleic acid molecule and a sequencing primer strand (e.g., a sequencing primer extended 3' by one or more nucleotides during a previous sequencing reaction), wherein the contacting is conducted under conditions suitable to bind the plurality of first polymerases to the plurality of duplexes, thereby forming a plurality of first polymerase complexes, wherein individual complexes comprise a first polymerase bound to a nucleic acid duplex, and wherein the nucleic acid duplex comprises a template nucleic acid molecule hybridized to a forward sequencing primer strand. In some embodiments, the methods comprise **step (b)** contacting the plurality of first polymerase complexes with a plurality of polymeric molecules, e.g., compositions of formula (I), to form a plurality of multivalent binding complexes. In some embodiments, the contacting of step (b) is conducted under conditions suitable for binding complementary nucleotide moieties of the polymeric molecules to at least two of the plurality of first polymerase complexes thereby forming a plurality of multivalent binding complexes. In some embodiments, the complementary nucleotide moieties of the polymeric molecules bind to a complementary nucleotide in the template nucleic acid molecules that are adjacent to a 3' ends of the sequencing primer strands. In some embodiments, the conditions are suitable for inhibiting polymerase-catalyzed incorporation of the complementary nucleotide moieties into the duplex, e.g. through a polymerase-catalyzed extension reaction. In some embodiments, the methods comprise **step (c)** detecting a detectable reporter moiety of the plurality of multivalent binding complexes. In some embodiments, the detecting includes detecting signals emitted by

the detectable reporter moieties of the polymeric molecules that are bound to the first polymerases, where the nucleotide moieties of the polymeric molecules are bound to complementary nucleotides of the template nucleic acid molecules, but incorporation of the nucleotide moieties is inhibited. In some embodiments, the methods for sequencing comprise **step (d)** identifying the nucleobase of the complementary nucleotide in the template nucleic acid molecules that are bound to the plurality of first polymerases, thereby determining the sequence of the template nucleic acid molecules, using the methods described herein. In some embodiments, the methods comprise **step (e)** contacting the plurality of the nucleic acid duplexes from the first stage and polymeric molecules with a plurality of second polymerases. The plurality of first polymerases and polymeric molecules can be dissociated from the duplexes prior to step (e) using any of the methods described herein. In some embodiments, the contacting is conducted under conditions suitable for binding the plurality of second polymerases to the plurality of the nucleic acid duplexes, thereby forming a plurality of second polymerase complexes comprising a second polymerase bound to a nucleic acid duplex. In some embodiments, the methods comprise **step (f)** contacting the plurality of second polymerase complexes with a plurality of nucleotides or analogs thereof, wherein the contacting is conducted under conditions suitable for binding complementary nucleotides or analogs thereof from the plurality of nucleotides or analogs thereof to at least two of the second polymerase complexes. In some embodiments, the complementary nucleotides or analogs thereof bound by the second polymerase complexes comprise nucleotides or analogs thereof complementary to a nucleotide of the template nucleic acid sequence immediately adjacent to the 3' end of the forward sequencing primer strand. In some embodiments, the contacting of step (f) is conducted under conditions suitable for promoting polymerase-catalyzed incorporation of the bound complementary nucleotides or analogs thereof into the duplex, thereby extending the sequencing primer by one nucleotide. In some embodiments, incorporating the nucleotide into the 3' end of the sequencing primer in step (f) comprises a primer extension reaction. In some embodiments, for example when the plurality of nucleotides in step (b) are detectably labeled, the methods for sequencing comprise **step (g)** detecting the complementary nucleotides which are incorporated into the sequencing primer strand. In some embodiments, the plurality of nucleotides are labeled with a detectable reporter moiety to permit detection. In some embodiments, when the plurality of nucleotides in step (f) are non-labeled, the detecting of step (g) is omitted. In some embodiments, when the plurality of nucleotides in step (f) are detectably labeled, the methods for sequencing comprise **step (h)** identifying the bases of the complementary nucleotides which are incorporated into the

duplexes. In some embodiments, the identification of the incorporated complementary nucleotides in step (h) can be used to confirm the identity of the complementary nucleotides of the polymeric molecules that are bound in the plurality of multivalent binding complexes in the first stage of the sequencing reaction. In some embodiments, the identifying of step (h) can be used to determine the sequence of the template nucleic acid molecules. In some embodiments, when the plurality of nucleotides in step (f) are non-labeled, the identifying of step (h) is omitted. In some embodiments, the methods comprise **step (i)** removing the chain terminating moiety from the incorporated nucleotide analogs when step (f) is conducted by contacting the plurality of second polymerase complexes with a plurality of nucleotides or analogs thereof that comprise at least one nucleotide analog having a 2' and/or 3' chain terminating moiety.

In some embodiments, the methods comprise, before starting at step (a) and repeating the first and second stage sequencing reactions, dissociating the second polymerases from the nucleic acid duplexes under conditions sufficient to retain the plurality of nucleic acid duplexes.

In some embodiments, the methods for sequencing comprise repeating the steps of the first stage and the second stage, described *supra*, at least once. In some embodiments, the sequence of the template nucleic acid molecules can be determined by detecting and identifying the polymeric molecules that bind the sequencing polymerases but do not incorporate into the 3' end of the sequencing primer, or extended sequencing primer. In some embodiments, the sequence of the template nucleic acid template molecule can be determined (or confirmed) by detecting and identifying the nucleotide that incorporates into the 3' end of the sequencing primer, or extended sequencing primer.

In some embodiments, the methods for sequencing comprise repeating the steps of the first and second stage sequencing reactions at least 1, 10, 20, 30, 40, 50, 70, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 800, 900, 1000, or 1500 times. In some embodiments, the methods comprise repeating the steps at least 100 times. In some embodiments, the methods comprise repeating the steps at least 150 times. In some embodiments, the methods comprise repeating the steps at least 200 times. In some embodiments, the methods comprise repeating the steps at least 250 times. In some embodiments, the methods comprise repeating the steps at least 300 times. In some embodiments, the methods comprise repeating the steps at least 400 times. In some embodiments, the methods comprise repeating the steps at least 500 times. In some embodiments, the methods for sequencing comprise repeating the steps of the first and second

stage sequencing reaction until the identities of the nucleotides in the target sequences have been determined.

Paired End Sequencing

In some embodiments, the methods comprise sequencing the sequences complementary to the target sequences. Exemplary methods of sequencing complementary sequences are illustrated in FIGS. 28-39. The plurality of template nucleic acid molecules, e.g. single-stranded template nucleic acid molecules, can be amplified in an extension reaction using a universal binding site for an amplification primer. The template nucleic acid molecules are then removed, and the complementary strand is retained, and subjected to the sequencing methods described herein, using the polymeric molecules of the disclosure. The template nucleic acid molecule can be amplified by any suitable means, including using soluble primers, such as an amplification primer or sequencing primer, to prime a polymerase-catalyzed amplification reaction, or by extending the forward sequencing primer strands produced by the previous sequencing reactions. In those cases where sequencing takes place on a support, the resulting amplification products can also be immobilized and sequenced using the methods described herein.

In some embodiments, sequencing the plurality of template nucleic acid molecules generates a plurality of extended forward sequencing primer strands. In some embodiments, the methods comprise (a) retaining the plurality of template nucleic acid molecules and replacing the plurality of extended forward sequencing primer strands with a plurality of forward extension strands that are hybridized to the plurality of nucleic acid template molecules by conducting a primer extension reaction; (b) removing the plurality of nucleic acid template molecules while retaining the plurality of forward extension strands and (c) sequencing the plurality of retained forward extension strands. In some embodiments, for example those embodiments where the plurality of template nucleic acid molecules are immobilized on a support using a plurality of surface primers, the plurality of surface primers are retained.

In some embodiments, for example those embodiments wherein the template nucleic acid molecules are concatemers, the template nucleic acid molecules comprise (i) two or more copies of the target sequence, (ii) two or more copies of the binding sequence for a forward sequencing primer, and (iii) two or more copies of a binding sequence for a reverse sequencing primer. The reverse sequencing primer binding sites can be used in conjunction with a plurality of reverse sequencing primers that bind thereto to sequence the plurality of retained forward extension strands. The forward sequencing primer binding sites can be used, together with a

plurality forward sequencing primers, to generate the plurality of retained forward extension strands in a polymerase-catalyzed extension reaction.

Alternatively, the template nucleic acid molecules comprise one or more of a binding sequence for an amplification primer, which, with a plurality of amplification primers, can be used to generate the plurality of retained forward extension strands in a polymerase-catalyzed extension reaction. In some embodiments, the template nucleic acid molecules comprise binding sequences for an amplification primer, and conducting the primer extension reaction comprises contacting the plurality of template nucleic acid molecules with a plurality of amplification primers, a plurality of nucleotides and a plurality of polymerases, thereby generating a plurality of forward extension strands that are hybridized to the template nucleic acid molecules. In some embodiments, wherein the plurality of amplification primers hybridize to the binding sequences for the amplification primers.

In some embodiments, the amplification primers are soluble.

In some embodiments, the polymerases comprises wild type polymerases or recombinant mutant polymerases. In some embodiments, the polymerases comprise phi29 DNA polymerases, large fragment of Bst DNA polymerases, large fragment of Bsu DNA polymerases (exo-), Bca DNA polymerases (exo-), Klenow fragment of E. coli DNA polymerases, T5 polymerases, M-MuLV reverse transcriptases, HIV viral reverse transcriptases, Deep Vent DNA polymerases or KOD DNA polymerases. In some embodiments, the polymerases comprise at least one amino acid substitution that confers exonuclease-minus activity.

In some embodiments, the plurality of template nucleic acid molecules are removed by generating abasic sites in the template nucleic molecules, followed by generating gaps at the abasic sites. When the template nucleic acid molecules include scissile moieties that can be cleaved to generate abasic sites, but the surface primers used to immobilize the template nucleic acid molecules on the support lack the scissile moiety, the template nucleic acid molecules can be removed while retaining the surface primers. In some embodiments, the nucleic acid template molecules comprise at least one nucleotide having a scissile moiety that can be cleaved to generate an abasic site. In some embodiments, the surface primers lack a nucleotide having a scissile moiety. In some embodiments, the nucleotide having a scissile moiety comprises uridine, 8-oxo-7,8-dihydrogunine, or deoxyinosine. In some embodiments, removing the nucleic acid template molecules comprises generating abasic sites in the nucleic acid template molecules, followed by generating gaps at the abasic sites.

In some embodiments, the at least one nucleotide having a scissile moiety comprises uracil, and generating abasic sites comprises contacting the nucleic acid template molecules with uracil DNA glycosylase (UDG). In some embodiments, generating gaps at the abasic sites comprises contacting the abasic sites with an endonuclease IV, AP lyase, FPG glycosylase/AP lyase and/or endo VIII glycosylase/AP lyase. In some embodiments, , wherein individual template nucleic acid molecules comprise nucleic acid template molecules having up to 30% of thymidines replaced with uridine. In some embodiments, 0.01 – 30% of the thymidine nucleotides in the template nucleic acid molecules are replaced with uridine. In some embodiments, 0.01 – 30% of the guanosine nucleotides in the individual template nucleic acid molecules are replaced with 8-oxo-7,8-dihydrogunine or deoxyinosine.

The plurality of gap-containing template nucleic acid molecules can be removed using any suitable method known in the art, for example by using an enzyme, chemical compound and/or heat. After the gap-removal procedure, the plurality of retained forward extension strands can hybridized to the retained surface primers.

In some embodiments, the plurality of gap-containing molecules can be enzymatically degraded using a 5' to 3' double-stranded DNA exonuclease, including T7 exonuclease (e.g., from New England Biolabs, catalog # M0263S). When a 5' to 3' double-stranded DNA exonuclease is used for removing gap-containing molecules, then the plurality of amplification primers can comprise at least one phosphorothioate diester bond at their 5' ends which can render the soluble amplification primers resistant to exonuclease degradation. In some embodiments, the plurality of amplification primers comprise 2-5 or more consecutive phosphorothioate diester bonds at their 5' ends. In some embodiments, the plurality amplification primers comprise at least one ribonucleotide and/or at least one 2'-O-methyl or 2'-O-methoxyethyl (MOE) nucleotide which can render the primers resistant to exonuclease degradation.

In some embodiments, the plurality of gap-containing template nucleic acid molecules can be removed using a chemical reagent that favors nucleic acid denaturation. The denaturation reagent can include any one or any combination of compounds such as formamide, acetonitrile, guanidinium chloride and/or a buffering agent (e.g., Tris-HCl, MES, HEPES, or the like).

In some embodiments, the plurality of gap-containing template nucleic acid molecules can be removed using an elevated temperature (e.g., heat) with or without a nucleic acid denaturation reagent. The gap-containing template molecules can be subjected to a temperature

of about 45-50 °C, or about 50-60 °C, or about 60-70 °C, or about 70-80 °C, or about 80-90 °C, or about 90-95 °C, or higher temperature.

In some embodiments, the plurality of gap-containing template nucleic acid molecules can be removed using 100% formamide at a temperature of about 65 °C for about 3 minutes, and washing with a reagent comprising about 50 mM NaCl or equivalent ionic strength and having a pH of about 6.5 – 8.5.

In some embodiments, sequencing the plurality of retained forward extension strands generates a plurality of extended reverse sequencing primer strands, wherein individual retained forward extension strands have two or more extended reverse sequencing primer strands hybridized thereon. Optionally the two or more extended reverse sequencing primer strands hybridized thereon can also be sequenced using the methods described herein.

In some embodiments, sequencing the plurality of retained forward extension strands comprises a plurality of soluble reverse sequencing primers and (i) a plurality of a first polymerases and a plurality of polymeric molecules and (ii) a plurality of a second polymerases and a plurality of nucleotides or analogs thereof, thereby generating a plurality of extended reverse sequencing primer strands, wherein individual retained forward extension strands have two or more extended reverse sequencing primer strands hybridized thereon.

In some embodiments, sequencing the complement of the target sequence comprises a two-stage sequencing reaction. In some embodiments, sequencing the plurality of retained forward extension strands comprises first stage comprising (a) contacting (i) the plurality of retained forward extension strands, (ii) a plurality of reverse sequencing primers comprising a sequence complementary to the binding sequence for the reverse sequencing primer, (iii) a plurality of first polymerases, and (iv) a plurality of polymeric molecules of Formula (I) or an ionized form, salt, or isomer thereof, wherein each polymeric molecule comprises at least two nucleotide moieties and at least one detectable reporter moiety, wherein the contacting occurs under conditions sufficient to form a plurality of multivalent binding complexes comprising a nucleic acid duplex between a retained forward extension strand and a reverse sequencing primer, a first polymerase, and a nucleotide moiety of a polymeric molecule that is complementary to a nucleotide in the retained forward extension strand immediately adjacent to the 3’ of an end of the reverse sequencing primer, and wherein polymerase catalyzed incorporation of a complementary nucleotide moiety into the nucleic acid duplex is inhibited; (b) detecting the detectable reporter moieties; and (c) determining nucleobase identities of nucleotides in the retained forward extension strands complementary to the nucleotide moieties

of the polymeric molecules based on the detectable reporter moieties of the polymeric molecules in the plurality of multivalent binding complexes formed in step (a).

In some embodiments, individual retained forward extension strands comprise two or more multivalent binding complexes. In some embodiments, the plurality of reverse sequencing primers are soluble.

In some embodiments, the methods comprise (d) dissociating the multivalent binding complexes under conditions sufficient to retain the nucleic acid duplexes, thereby generating a plurality of nucleic acid duplexes; and (e) contacting the plurality of nucleic acid duplexes with a plurality of second polymerases and a plurality of nucleotides or analogs thereof under conditions sufficient to incorporate nucleotides or analogs thereof complementary to the nucleotides of the retained forward extension strands immediately adjacent to the 3' of the ends of the reverse sequencing primers in a primer extension reaction, thereby generating a plurality of extended nucleic acid duplexes comprising extended reverse sequencing primer sequences

In some embodiments, the methods comprise dissociating the second polymerases from the extended nucleic acid duplexes under conditions sufficient to retain the plurality of extended nucleic acid duplexes.

In some embodiments, the template nucleic acid molecules comprise concatemers of two or more copies of a sequence comprising (i) a sequence for the reverse sequencing primer (i.e., the complementary sequence, when produced, will contain the binding sequence for the primer), (ii) the target nucleic acid sequence, and (iii) a binding sequence for the forward sequencing primer. In some embodiments, retained forward extension strands comprise the two or more copies of a sequence complementary to (i) the sequence for the reverse sequencing primer hybridize to the reverse sequencing primers to form nucleic acid duplexes between the retained forward extension strands and the reverse sequencing primer.

In some embodiments, two or more nucleotide moieties in an individual polymeric molecule contact two or more different multivalent binding complexes on the same retained forward extension strand.

In some embodiments, the first stage comprises **step (a):** contacting a plurality of a first polymerases with (i) a plurality of retained forward extension strands and (ii) a plurality of reverse sequencing primers, wherein the contacting is conducted under conditions suitable to bind the plurality of first polymerases to the plurality of retained forward extension strands and the plurality of reverse sequencing primers, thereby forming a plurality of first polymerase complexes, wherein individual complexes comprise a first polymerase bound to a nucleic acid duplex, and wherein the nucleic acid duplex comprises a retained forward extension strands

hybridized to a reverse sequencing primer. In some embodiments, the retained forward extension strands in the plurality of retained forward extension strands of step (a) comprise a sequence complementary to the same target sequence or sequences complementary to different target sequences. In some embodiments, the retained forward extension strands and/or the plurality of reverse sequencing primers of step (a) are in solution or are immobilized to a support, as described herein for the plurality of template nucleic acid molecules. In some embodiments, the plurality of retained forward extension strands and/or sequencing primers are immobilized to $10^2 – 10^{15}$ different sites on a support. In some embodiments, the plurality of immobilized first polymerase complexes are in fluid communication with each other to permit flowing a solution of reagents (e.g., enzymes including sequencing polymerases, polymeric molecules, nucleotides, and/or divalent cations) onto the support so that the plurality of immobilized polymerase complexes on the support are reacted with the solution of reagents in a massively parallel manner.

In some embodiments, the methods comprise **step (b)** contacting the plurality of first polymerase complexes with a plurality of polymeric molecules, e.g., compositions of formula (I) to form a plurality of multivalent binding complexes. In some embodiments, individual polymeric molecules in the plurality of polymeric molecules comprise a central moiety operably linked by a polymeric side chain to a plurality of nucleotide moieties. In some embodiments, the individual polymeric molecules comprise at least one detectable reporter moiety. In some embodiments, the contacting of step (b) is conducted under conditions suitable for binding complementary nucleotide moieties of the polymeric molecules to at least two of the plurality of first polymerase complexes thereby forming a plurality of multivalent binding complexes. In some embodiments, the complementary nucleotide moieties of the polymeric molecules bind to a complementary nucleotide in the template nucleic acid molecule that is immediately 3’ of the sequencing primer. In some embodiments, the conditions are suitable for inhibiting polymerase-catalyzed incorporation of the complementary nucleotide moieties into the duplex, e.g. through a polymerase-catalyzed extension reaction.

In some embodiments, the methods for sequencing further comprise **step (c)** detecting a detectable reporter moiety of the plurality of multivalent binding complexes. In some embodiments, the detecting includes detecting signals emitted by the detectable reporter moieties of the polymeric molecules that are bound to the first polymerases, where the nucleotide moieties of the polymeric molecules are bound to complementary nucleotides of the template nucleic acid molecules, but incorporation of the nucleotide moieties is inhibited. In

some embodiments, the polymeric molecules are labeled with a detectable reporter moiety to permit detection.

In some embodiments, the methods for sequencing comprise **step (d)** identifying the nucleobase of the complementary nucleotide in the retained forward extension strands that are bound to the plurality of first polymerases, thereby determining the sequence of the template nucleic acid molecules. In some embodiments, the polymeric molecules are labeled with a detectable reporter moiety that corresponds to the particular nucleotide moieties of the individual polymeric molecule to permit identification of the complementary nucleotide moieties (e.g., nucleotide base adenine, guanine, cytosine, thymine or uracil) that are bound to the plurality of first polymerases.

In some embodiments, the second stage of the two-stage sequencing reaction comprises nucleotide incorporation. In some embodiments, the methods comprise **step (e)** contacting the plurality of the nucleic acid duplexes from the first stage that have been retained following dissociation of the first polymerases and polymeric molecules with a plurality of second polymerases. In some embodiments, the contacting is conducted under conditions suitable for binding the plurality of second polymerases to the plurality of the nucleic acid duplexes, thereby forming a plurality of second polymerase complexes comprising a second polymerase bound to a nucleic acid duplex.

In some embodiments, the methods comprise **step (f)** contacting the plurality of second polymerase complexes with a plurality of nucleotides or analogs thereof, wherein the contacting is conducted under conditions suitable for binding complementary nucleotides or analogs thereof from the plurality of nucleotides or analogs thereof to at least two of the second polymerase complexes. In some embodiments, the complementary nucleotides or analogs thereof bound by the second polymerase complexes comprise nucleotides complementary to a nucleotide of the retained forward extension strand immediately adjacent to the 3’ end of the reverse sequencing primer. In some embodiments, the contacting of step (f) is conducted under conditions suitable for promoting polymerase-catalyzed incorporation of the bound complementary nucleotides or analogs thereof into the duplex, thereby extending the sequencing primer strand by one nucleotide. In some embodiments, incorporating the nucleotides or analogs thereof into the 3’ end of the sequencing primer strand in step (f) comprises a primer extension reaction.

In some embodiments, for example when the plurality of nucleotides or analogs thereof in step (f) are detectably labeled, the methods for sequencing further comprise **step (g)** detecting the complementary nucleotides or analogs thereof which are incorporated into the

primers (or extended primers). In some embodiments, the plurality of nucleotides or analogs thereof are labeled with a detectable reporter moiety to permit detection. In some embodiments, when the plurality of nucleotides or analogs thereof in step (f) are non-labeled, the detecting of step (g) is omitted.

In some embodiments, when the plurality of nucleotides or analogs thereof in step (f) are detectably labeled, the methods for sequencing further comprise **step (h)** identifying the nucleobases of the complementary nucleotides which are incorporated into the duplexes. In some embodiments, the identification of the incorporated nucleotides or analogs thereof in step (g) can be used to confirm the identity of the nucleotides of the polymeric molecules from the plurality of multivalent binding complexes in the first stage of the sequencing reaction. In some embodiments, the identifying of step (g) can be used to determine the sequence of the retained forward extension strands. In some embodiments, when the plurality of nucleotides or analogs thereof in step (f) are non-labeled, the identifying of step (h) is omitted.

In some embodiments, the methods comprise **step (i)** removing the chain terminating moiety from the incorporated nucleotide analogs when step (f) is conducted by contacting the plurality of second polymerase complexes with a plurality of nucleotide analogs that comprise at least one nucleotide having a 2' and/or 3' chain terminating moiety.

In some embodiments, the methods comprise repeating the first and second stage reactions one or more times. In some embodiments, the methods comprise (a) contacting the plurality of extended nucleic acid duplexes with a plurality of first polymerases and a plurality of polymeric molecules of Formula (I) or (I'), or an ionized form thereof, an isomer thereof, or a salt thereof, wherein the contacting occurs under conditions sufficient to form a plurality of multivalent binding complexes comprising an extended nucleic acid duplex, a first polymerase, and a nucleotide moiety of a polymeric molecule that is complementary to a nucleotide in the retained forward extension strand immediately adjacent to the 3' of the extended reverse sequencing primer, and wherein polymerase catalyzed incorporation of a complementary nucleotide moiety into the extended nucleic acid duplex is inhibited; (b) detecting the detectable reporter moieties; (c) determining nucleobase identities of nucleotides in the retained forward extension strands complementary to the nucleotide moieties of the polymeric molecules based on the detectable reporter moieties of the polymeric molecules in the plurality of multivalent binding complexes formed in step (a); (d) dissociating the multivalent binding complexes under conditions sufficient to retain the plurality extended nucleic acid duplexes; and (e) contacting the plurality of extended nucleic acid duplexes with a plurality of second polymerases and a plurality of nucleotides or analogs thereof under conditions sufficient to

incorporate nucleotides or analogs thereof complementary to the nucleotides of the nucleic acid template sequences immediately adjacent to the 3' of the ends of the extended reverse sequencing primers in a primer extension reaction, thereby generating a plurality of extended nucleic acid duplexes comprising extended reverse sequencing primers.

In some embodiments, the methods comprise repeating the two stages of the sequencing reaction at least once. In some embodiments, the first stage comprises **step (a)** contacting a plurality of a first polymerase with (i) a plurality of duplexes comprising a retained forward extension strand and a sequencing primer strand (e.g., a sequencing primer, or a sequencing primer extended 3' by one or more nucleotides during previous rounds of sequencing reactions), wherein the contacting is conducted under conditions suitable to bind the plurality of first polymerases to the plurality of duplexes, thereby forming a plurality of first polymerase complexes, wherein individual complexes comprise a first polymerase bound to a nucleic acid duplex, and wherein the nucleic acid duplex comprises a retained forward extension strand hybridized to a reverse sequencing primer strand. In some embodiments, the methods comprise **step (b)** contacting the plurality of first polymerase complexes with a plurality of polymeric molecules, e.g., compositions of formula (I) and/or formula (I'), to form a plurality of multivalent binding complexes. In some embodiments, the contacting of step (b) is conducted under conditions suitable for binding complementary nucleotide moieties of the polymeric molecules to at least two of the plurality of first polymerase complexes thereby forming a plurality of multivalent binding complexes. In some embodiments, the complementary nucleotide moieties of the polymeric molecules bind to a complementary nucleotide in the retained forward extension strand that is immediately adjacent to the 3' end of the sequencing primer strand. In some embodiments, the conditions are suitable for inhibiting polymerase-catalyzed incorporation of the complementary nucleotide moieties into the duplex, e.g. through a polymerase-catalyzed extension reaction. In some embodiments, the methods for sequencing comprise **step (c)** detecting a detectable reporter moiety of the plurality of multivalent binding complexes. In some embodiments, the methods for sequencing further comprise **step (d)** identifying the nucleobase of the complementary nucleotide in retained forward extension strands that are bound to the plurality of first polymerases, thereby determining the sequence of the template nucleic acid molecules, using the methods described herein. In some embodiments, the methods comprise **(e)** contacting the plurality of the nucleic acid duplexes from the first stage and polymeric molecules with a plurality of second polymerases. The plurality of first polymerases and polymeric molecules can be dissociated from the duplexes prior to step (e) using any of the methods described herein. In some embodiments, the

contacting is conducted under conditions suitable for binding the plurality of second polymerases to the plurality of the nucleic acid duplexes, thereby forming a plurality of second polymerase complexes comprising a second polymerase bound to a nucleic acid duplex. In some embodiments, the methods comprise **step (f)** contacting the plurality of second polymerase complexes with a plurality of nucleotides or analogs thereof, wherein the contacting is conducted under conditions suitable for binding complementary nucleotides or analogs thereof from the plurality of nucleotides or analogs thereof to at least two of the second polymerase complexes. In some embodiments, the complementary nucleotides bound by the second polymerase complexes comprise nucleotides or analogs thereof complementary to a nucleotide of the retained forward extension strand immediately adjacent to the 3' end of the reverse sequencing primer strand. In some embodiments, the contacting of step (f) is conducted under conditions suitable for promoting polymerase-catalyzed incorporation of the bound complementary nucleotides or analogs thereof into the duplex, thereby extending the sequencing primer strand by one nucleotide. In some embodiments, incorporating the nucleotides into the 3' ends of the sequencing primer strands in step (f) comprises a primer extension reaction. In some embodiments, for example when the plurality of nucleotides in step (b) are detectably labeled, the methods for sequencing further comprise **step (g)** detecting the complementary nucleotides which are incorporated into the sequencing primer strands. In some embodiments, the plurality of nucleotides are labeled with a detectable reporter moiety to permit detection. In some embodiments, when the plurality of nucleotides in step (f) are non-labeled, the detecting of step (g) is omitted. In some embodiments, when the plurality of nucleotides in step (f) are detectably labeled, the methods for sequencing comprise **step (h)** identifying the bases of the complementary nucleotides which are incorporated into the duplexes. In some embodiments, the identification of the incorporated complementary nucleotides in step (h) can be used to confirm the identity of the complementary nucleotides of the polymeric molecules that are bound in the plurality of multivalent binding complexes in the first stage of the sequencing reaction. In some embodiments, the identifying of step (h) can be used to determine the sequence of the retained forward extension strands. In some embodiments, when the plurality of nucleotides in step (f) are non-labeled, the identifying of step (h) is omitted. In some embodiments, the methods comprise **step (i)** removing the chain terminating moieties from the incorporated nucleotide analogs when step (f) is conducted by contacting the plurality of second polymerase complexes with a plurality of nucleotides that comprise at least one nucleotide analog having a 2' and/or 3' chain terminating moiety.

In some embodiments, the methods comprise, before starting at step (a) and repeating the first and second stage sequencing reactions, dissociating the second polymerases from the nucleic acid duplexes under conditions sufficient to retain the plurality of nucleic acid duplexes.

In some embodiments, the methods for sequencing comprise repeating the steps of the first stage and the second stage, described *supra*, at least once. In some embodiments, the sequence of the template nucleic acid molecules can be determined by detecting and identifying the polymeric molecules that bind the sequencing polymerases but do not incorporate into the 3' end of the sequencing primer, or extended sequencing primer. In some embodiments, the sequence of the template nucleic acid template molecule can be determined (or confirmed) by detecting and identifying the nucleotide that incorporates into the 3' end of the sequencing primer, or extended sequencing primer.

In some embodiments, the methods for sequencing comprise repeating the steps of the first and second stage sequencing reactions at least 1, 10, 20, 30, 40, 50, 70, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 800, 900, 1000, or 1500 times. In some embodiments, the methods comprise repeating the steps at least 100 times. In some embodiments, the methods comprise repeating the steps at least 150 times. In some embodiments, the methods comprise repeating the steps at least 200 times. In some embodiments, the methods comprise repeating the steps at least 250 times. In some embodiments, the methods comprise repeating the steps at least 300 times. In some embodiments, the methods comprise repeating the steps at least 400 times. In some embodiments, the methods comprise repeating the steps at least 500 times. In some embodiments, the methods for sequencing comprise repeating the steps of the first and second stage sequencing reaction until the identities of the nucleotides in the sequences complementary to the target sequences have been determined.

In some embodiments, the methods comprise before step (a), dissociating the second polymerases from the extended nucleic acid duplexes under conditions sufficient to retain the plurality of extended nucleic acid duplexes.

Second Surface Primers

In any of the sequencing methods described *supra*, a second surface primer can be used to immobilize an end of the template nucleic acid molecules.

In some embodiments, for example those embodiments wherein the template nucleic acid molecules are immobilized on a support, the support further comprises a plurality of a second surface primer immobilized thereon. The second surface primers have a sequence that

differs from the first surface primer. In some embodiments, the second surface primers comprise single stranded oligonucleotides comprising DNA, RNA or a combination of DNA and RNA. The second surface primers comprise a sequence that is wholly complementary or partially complementary along their lengths to at least a portion of template nucleic acid molecule. The second surface primers can be immobilized on the support or immobilized via a coating on the support. The second surface primers can be embedded and attached (coupled) to the coating on the support. In some embodiments, the 5' end of the second surface primers are immobilized to a support or immobilized to a coating on the support. Alternatively, an interior portion or the 3' end of the second surface primers can be immobilized to a support or immobilized to a coating on the support. The support comprises a plurality of second surface primers having the same sequence. The immobilized second surface primers can be any length, for example 4-50 nucleotides, or 50-100 nucleotides, or 100-150 nucleotides, or longer lengths. In some embodiments, the 3' terminal ends of the second surface primers comprise an extendible 3' OH moiety. In some embodiments, the 3' terminal ends of the second surface primers comprise a 3' non-extendible moiety. The 3' terminal ends of the second surface primers comprise a moiety that blocks primer extension, such as for example a phosphate group, a dideoxycytidine group, an inverted dT, or an amino group. The immobilized second surface primers are not extendible in a primer extension reaction. The immobilized second surface primers lack a nucleotide having a scissile moiety.

In some embodiments, the plurality of second surface primers comprise at least one phosphorothioate diester bond at their 5' ends which can render the second surface primers resistant to exonuclease degradation. In some embodiments, the plurality of second surface primers comprise 2-5 or more consecutive phosphorothioate diester bonds at their 5' ends. In some embodiments, the plurality of second surface primers comprise at least one ribonucleotide and/or at least one 2'-O-methyl or 2'-O-methoxyethyl (MOE) nucleotide which can render the second surface primers resistant to exonuclease degradation.

In some embodiments, the plurality of template nucleic acid molecules are single stranded. In some embodiments, individual template nucleic acid molecules are covalently joined to a first surface primer, and at least one portion of the individual template nucleic acid molecule is hybridized to second surface primer. The second surface primers serve to pin down a portion of the immobilized template molecules to the support. The template molecule can have two or more copies of a universal binding sequence for the second surface primer, e.g. as part of the concatemerized sequence when the template nucleic acid molecules are concatemers. The portion of the template nucleic molecule that includes the universal binding

sequence for a second surface primer can hybridize to the second surface primer. In some embodiments, the second surface primers include a terminal 3' blocking group that renders them non-extendible. In some embodiments, the second surface primers have terminal 3' extendible ends.

In some embodiments, the support comprises about $10^2 – 10^{15}$ first surface primers per mm^2. In some embodiments, the support comprises about $10^2 – 10^{15}$ second surface primers per mm^2. In some embodiments, the support comprises about $10^2 – 10^{15}$ first surface primers and second surface primers per mm^2.

In some embodiments, the template nucleic acid molecules comprise two or more copies of a universal binding sequence (or complementary sequence thereof) for a second surface primer having a sequence that differs from a universal binding sequence for the first surface primer.

In some embodiments, the 3' terminal end of the second surface primers comprise an extendible 3' OH moiety. In some embodiments, the 3' terminal end of the second surface primers comprise a 3' non-extendible moiety. In some embodiments, the 3' terminal end of the second surface primers comprise a moiety that blocks primer extension (e.g., non-extendible terminal 3' end), such as for example a phosphate group, a dideoxycytidine group, an inverted dT, or an amino group such that thee second surface primers are not extendible in a primer extension reaction. In some embodiments, the second surface primers lack a nucleotide having a scissile moiety.

Soluble Compaction Oligonucleotides

In some embodiments, the template nucleic acid molecules comprise a binding site for a soluble compaction oligonucleotide. In some embodiments, the method comprises generating the plurality of template nucleic acid molecules through rolling circle amplification to generate concatemers, wherein the rolling circle amplification (RCA) comprises contacting soluble compaction oligonucleotides. Alternatively, or in addition, the methods described herein comprise contacting the plurality of template nucleic acid molecules with a plurality of soluble compaction oligonucleotides.

Without wishing to be bound by theory, it is thought that template nucleic acid molecules that are concatemers can self-collapse into a compact nucleic acid nanoball. Inclusion of one or more compaction oligonucleotides during the RCA reaction to produce the concatemer can further compact the size and/or shape of the nanoball. An increase in the number of tandem copies in a given concatemer increases the number of sites along the

concatemer for hybridizing to multiple sequencing primers which serve as multiple initiation sites for polymerase-catalyzed sequencing reactions. When the sequencing reaction employs detectably labeled nucleotides or analogs thereof, and/or detectably labeled polymeric molecules, the signals emitted by the labeled nucleotides or analogs thereof or polymeric molecules that participate in the parallel sequencing reactions along the concatemer yields an increased signal intensity for each concatemer. Multiple portions of a given concatemer can be simultaneously sequenced. Furthermore, a plurality of binding complexes can form along a particular concatemer molecule, each binding complex comprising a sequencing polymerase bound to a polymeric molecule wherein the plurality of binding complexes remain stable without dissociation resulting in increased persistence time which increases signal intensity and reduces imaging time.

The primer extension reaction to generate the retained forward extension strands can also optionally include a plurality of compaction oligonucleotides and/or hexamine (e.g., cobalt hexamine III). Individual forward extension strands can collapse into a nanoball having a more compact size and/or shape compared to a nanoball generated from a primer extension reaction conducted without compaction oligonucleotides and/or hexamine (e.g., cobalt hexamine III). Inclusion of compaction oligonucleotides and/or hexamine (e.g., cobalt hexamine III) in the primer extension reaction can improve FWHM (full width half maximum) of a spot image of the nanoball. The spot image can be represented as a Gaussian spot and the size can be measured as a FWHM. A smaller spot size as indicated by a smaller FWHM typically correlates with an improved image of the spot. In some embodiments, the FWHM of a nanoball spot can be about 10 μm or smaller.

In some embodiments, the rolling circle amplification step that generates the template nucleic acid molecules comprises a plurality of compaction oligonucleotides and/or hexamine to generate immobilized template nucleic acid molecules having a more compact size and/or shape compared to a rolling circle amplification reaction in the absence of compaction oligonucleotides and/or hexamine.

In some embodiments, the primer extension reaction comprises a plurality of compaction oligonucleotides and/or hexamine to generate a plurality of retained forward extension strands having a more compact size and/or shape compared to a primer extension reaction in the absence of compaction oligonucleotides and/or hexamine.

Sequencing-by-Binding (SBB)

The present disclosure provides methods for sequencing, wherein the sequencing methods comprise a sequencing-by-binding (SBB) procedure which employs labeled and/or

non-labeled polymeric molecules, wherein the Terminal Moieties comprising chain-terminating nucleotide moieties. In some embodiments, the sequencing-by-binding (SBB) method comprises **step (a)** sequentially contacting a template nucleic acid molecule associated with a primer (a primed template nucleic acid molecule) with at least two separate mixtures under ternary complex stabilizing conditions, wherein the at least two separate mixtures comprise a polymerase and at least one type of polymeric molecule, whereby the sequentially contacting results in the primed template nucleic acid being contacted, under the ternary complex stabilizing conditions, with nucleotide cognates for first, second and third base type base types in the template; **step (b)** examining the at least two separate mixtures to determine whether a ternary complex formed; **step (c)** identifying the next correct nucleotide for the primed template nucleic acid molecule, wherein the next correct nucleotide is identified as a cognate of the first, second or third base type if ternary complex is detected in step (b), and wherein the next correct nucleotide is imputed to be a nucleotide cognate of a fourth base type based on the absence of a ternary complex in step (b); **step (d)** adding a next correct nucleotide to the primer of the primed template nucleic acid after step (b) (e.g., by conducting a nucleotide incorporation reaction), thereby producing an extended primer; and **step (e)** repeating steps (a) through (d) at least once on the primed template nucleic acid that comprises the extended primer. Exemplary sequencing-by-binding methods are described in U.S. patent Nos. 10,246,744 and 10,731,141 (where the contents of both patents are hereby incorporated by reference in their entireties).

Complexes

The disclosure provides binding complexes comprising the polymeric molecules of the disclosure, and methods of using same in sequencing methods.

In some embodiments, any of the methods for sequencing nucleic acid molecules described herein can include forming a binding complex, where the binding complex comprises a polymerase, a nucleic acid template molecule associated with a primer, such as a sequencing primer to form a duplex (nucleic acid template molecule duplexed with a primer), and a nucleotide or analog thereof. Alternatively, or in addition, the binding complex comprises a polymerase, a nucleic acid template molecule duplexed with a primer, and a nucleotide moiety of a polymeric molecule as described herein (referred to herein as a “multivalent binding complex”). In some embodiments, the binding complex has a persistence time of greater than about 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9 or 1 second. The binding complex has a persistence time of greater than about 0.1-0.25 seconds, or about 0.25-0.5 seconds, or about 0.5-0.75

seconds, or about 0.75-1 second, or about 1-2 seconds, or about 2-3 seconds, or about 3-4 second, or about 4-5 seconds, and/or wherein the method is or may be carried out at a temperature of at or above 15 °C, at or above 20 °C, at or above 25 °C, at or above 35 °C, at or above 37 °C, at or above 42 °C at or above 55 °C at or above 60 °C, or at or above 72 °C, or at or above 80 °C, or within a range defined by any of the foregoing. The binding complex (e.g., ternary complex) remains stable until subjected to a condition that causes dissociation of interactions between any of the polymerase, template molecule, primer and/or the nucleotide moiety of the polymeric molecule or the free nucleotide or analog thereof. For example, a dissociating condition comprises contacting the binding complex with any one or any combination of a detergent, EDTA and/or water. In some embodiments, the present disclosure provides said method wherein the binding complex is deposited on, attached to, or hybridized to, a surface showing a contrast to noise ratio in the detecting step of greater than 20. In some embodiments, the present disclosure provides said method wherein the contacting is performed under a condition that stabilizes the binding complex when the free nucleotide or nucleotide moiety is complementary to a next base of the template nucleic acid, and destabilizes the binding complex when the free nucleotide or nucleotide moiety is not complementary to the next base of the template nucleic acid.

Forming Avidity Complexes

In some embodiments, the binding of the plurality of first polymerase complexes (e.g., polymerases associated with a duplexed template nucleic acid molecule or its complement and a sequencing primer) with the plurality of polymeric molecules forms at least one multivalent binding complex. In some embodiments, the method comprises **step (a)** binding a first nucleic acid primer, a first polymerase, and a first polymeric molecule to a first portion of a template nucleic acid molecule thereby forming a first multivalent binding complex, wherein a first nucleotide moiety of the first polymeric molecule binds to the first polymerase; and **step (b)** binding a second nucleic acid primer, a second polymerase, and the first polymeric molecule to a second portion of the same template molecule thereby forming a second multivalent binding complex, wherein a second nucleotide moiety of the first polymeric molecule binds to the second polymerase, wherein the first and second multivalent binding complexes include the same polymeric molecule(sometimes referred to herein as an “avidity complex”). In some embodiments, the template nucleic acid molecule is a concatemer, and the first portion comprises a first copy of the concatemerized sequence, and the second portion comprises a second coy of the concatemerized sequence. In some embodiments, the two copies are identical, or substantially identical. In some embodiments, the first polymerase comprises a

wild type or mutant polymerase. In some embodiments, the second polymerase comprises a wild type or mutant polymerase. In some embodiments, the template nucleic acid molecule is a concatemer, and comprises two or more tandem repeat sequences of a target sequence and at least one universal sequencing primer binding site. In some embodiments, the first and second nucleic acid primers can bind to a sequencing primer binding site along the concatemer template molecule.

The skilled artisan will understand that the foregoing steps can also be used to when sequencing the complementary strand to the template nucleic acid molecule in a paired end sequencing reaction.

Detecting and Identifying Avidity Complexes on a Concatemer

In some embodiments, the methods include binding the plurality of first polymerase complexes with a plurality of polymeric molecules to form at least one avidity complex, and the methods comprise **step (a)** contacting the plurality of polymerases and the plurality of nucleic acid primers with different portions of a template nucleic acid molecule that is a concatemer to form at least a first and second polymerase complexes on the same template nucleic acid molecule. In some embodiments, the methods comprise **step (b)** contacting a plurality of polymeric molecules with the at least first and second polymerase complexes on the same concatemer template molecule, under conditions suitable to bind at least one single polymeric molecule from the plurality to the first and second polymerase complexes, wherein at least a first nucleotide moiety of the polymeric molecule is bound to the first polymerase complex which includes a first primer hybridized to a first portion of the template nucleic acid molecule thereby forming a first multivalent binding complex (e.g., first ternary complex), and wherein at least a second nucleotide moiety of the single polymeric molecule is bound to the second polymerase complex which includes a second primer hybridized to a second portion of the template nucleic acid molecule thereby forming a second multivalent binding complex (e.g., second ternary complex). In some embodiments, the contacting is conducted under conditions suitable to inhibit polymerase-catalyzed incorporation of the nucleotide moieties in the first and second multivalent binding complexes. In some embodiments, the first and second multivalent binding complexes which comprise nucleotide moieties of the same polymeric molecule form an avidity complex. In some embodiments, the methods comprise **step (c)** detecting the first and second multivalent binding complexes on the same template nucleic acid molecule. In some embodiments, the methods comprise **step (d)** identifying the first nucleotide moiety in the first multivalent binding complex, thereby determining the identity of

corresponding complementary nucleotide in the first portion of the template molecule, and identifying the second nucleotide moiety in the second multivalent binding complex thereby determining the identity of the corresponding complementary nucleotide in the second portion of the template molecule. In some embodiments, the identities of the first and second nucleotides are the same. In some embodiments, the plurality of polymerases comprise a wild type or mutant sequencing polymerase. In some embodiments, the template nucleic acid molecule comprises a concatemer comprising two or more tandem repeat sequences of a target sequence and at least one universal sequencing primer binding site. The plurality of nucleic acid primers can bind to a sequencing primer binding site along the concatemer template molecule.

The skilled artisan will understand that the foregoing steps can also be used to when sequencing the complementary strand to the template nucleic acid molecule in a paired end sequencing reaction.

Forming Complexes on Two Template Nucleic Acid Molecules

In some embodiments, the binding of the plurality of first polymerase complexes with the plurality of polymeric molecules forms at least one avidity complex (i.e., a complex comprising nucleotide moieties of the same polymeric molecule associated with two or more multivalent binding complexes). In some embodiments, the method comprises **step (a)** binding a first nucleic acid primer, a first polymerase, and a first polymeric molecule to a first template nucleic acid molecule thereby forming a first multivalent binding complex, wherein a first nucleotide moiety of the first polymeric molecule binds the first multivalent binding complex; and **step (b)** binding a second nucleic acid primer, a second polymerase, and the first polymeric molecule to a second template nucleic molecule thereby forming a second multivalent binding complex, wherein a second nucleotide moiety of the first polymeric molecule binds to the multivalent binding complex, wherein the first and second multivalent binding complexes which include the same polymeric molecule form an avidity complex. In some embodiments, the first polymerase and second polymerases comprises a wild type or mutant polymerases as described herein. In some embodiments, the first and second template nucleic molecules each comprise a target sequence (e.g., one or more copies of a target sequence) and at least one universal sequencing primer binding site. In some embodiments, the first nucleic acid primer can bind to a sequencing primer binding site on the first template molecule. In some embodiments, the second nucleic acid primer can bind to a sequencing primer binding site on the second template molecule. In some embodiments, the first and second template nucleic acid molecules are not the same molecule. In some embodiments, the first and second template

nucleic acid molecules are localized in close proximity to each other. For example, the clonally-amplified first and second template nucleic molecules comprise linear template molecules that are generated via bridge amplification and are immobilized to the same location or feature on a support.

The skilled artisan will understand that the foregoing steps can also be used to when sequencing the complementary strand to the template nucleic acid molecule in a paired end sequencing reaction.

Detecting and identifying Avidity Complexes on Two Template Nucleic Acid Molecules

In some embodiments, methods comprise binding the plurality of first polymerase complexes with the plurality of polymeric molecules to form at least one avidity complex. In some embodiments, the methods comprise **step (a)** (i) contacting a first polymerase and a first nucleic acid primer with a first template nucleic acid molecule to form a first polymerase complex on the first template nucleic acid molecule, and (ii) contacting a second polymerase and a second nucleic acid primer with a second template nucleic acid molecule to form a second polymerase complex on the second template nucleic acid molecule. In some embodiments, the methods comprise **step (b)** contacting a plurality of polymeric molecules with the first and second polymerase complexes, under conditions sufficient to bind at least one polymeric molecule to the first and second polymerase complexes, wherein at least a first nucleotide moiety of the single polymeric molecule is bound to the first polymerase complex thereby forming a first multivalent binding complex (e.g., first ternary complex), and wherein at least a second nucleotide moiety of the single polymeric molecule is bound to the second polymerase complex thereby forming a second multivalent binding complex (e.g., second ternary complex), thereby forming the avidity complex. In some embodiments, the contacting is conducted under conditions sufficient to inhibit polymerase-catalyzed incorporation of the first and second nucleotide moieties into the first and second binding complexes. In some embodiments, the methods comprise **step (c)** detecting the first and second multivalent binding complexes on the first and second template nucleic molecules respectively. In some embodiments the methods comprise **step (d)** identifying the first nucleotide moiety in the first multivalent binding complex, thereby determining the identity of the corresponding complementary nucleotide of the first template nucleic acid molecule, and identifying the second nucleotide moiety in the second multivalent binding complex thereby determining the identity of the corresponding complementary nucleotide of the second template nucleic acid molecule. In some embodiments, the plurality of polymerases comprise wild type or mutant polymerases. In some embodiments, the first template nucleic acid molecule comprises one

copy of a first target sequence and at least one universal primer binding site (e.g., universal sequencing primer binding site). In some embodiments, the first nucleic acid primer can bind to a sequencing primer binding site on the first template nucleic acid molecule. In some embodiments, the second template nucleic acid molecule comprises one copy of a second target sequence and at least one universal primer binding site (e.g., universal sequencing primer binding site). In some embodiments, the second nucleic acid primer can bind to a sequencing primer binding site on the second template molecule. In some embodiments, the first and second template molecules comprise the same, or substantially the same, target sequence.

The skilled artisan will understand that the foregoing steps can also be used to when sequencing the complementary strand to the template nucleic acid molecule in a paired end sequencing reaction.

Template Nucleic Acid Molecules

The disclosure provides pluralities of template nucleic acid molecules for use in the methods of sequencing described herein.

In some embodiments, template nucleic acid molecules in the plurality comprise a target sequence. In some embodiments, different template nucleic acid molecules in the plurality comprise different target sequences.

In some embodiments, template nucleic acid molecules in the plurality have been clonally amplified. In some embodiments, template nucleic acid molecules in the plurality comprise the same target sequence. In some embodiments, some template nucleic acid molecules in the plurality comprise the same target sequence, while other template nucleic acid molecules in the plurality comprise different target sequence.

In some embodiments, the template nucleic acid molecules comprise concatemers. In some embodiments the concatemers comprise at least 2 copies, at least 3 copies, at least 4 copies, at least 5 copies, at least 10 copies, at least 50 copies, at least 100 copies, at least 500 copies, at least 700 copies, at least 1000 copies, at least 1500 copies, at least 2000 copies, at least 5000 copies or at least 1000 copies of the target sequence.

In some embodiments, the template nucleic acid molecules comprise concatemers comprising template nucleic acid molecules comprise concatemers of two or more copies of a sequence comprising: (i) a binding sequence for a forward sequencing primer, (ii) sequence complementary to a binding sequence for a reverse sequencing primer, (iii) a binding sequence for an first surface primer, (iv) a binding sequence for a second surface primer, (iv) a binding sequence for a first amplification primer, (v) a binding sequence for a second amplification

primer, (vii) a binding sequence for a soluble compaction oligonucleotide, (viii) a sample barcode sequence, and/or (ix) a unique molecular index sequence. Selection of appropriate sequences, depending on the method of sequencing, will be apparent to persons of ordinary skill in the art. The selected sequences can be introduced into the concatemer during amplification to produce the concatemer and these methods are described, for example in WO 2022/266470 and PCT/US2023/063736.

In some embodiments, the target sequence is between about 50 and 2000 basepairs, between about 100 and 1500 basepairs, between about 150 and 1000 basepairs, between about 200 and 800 basepairs, between about 200 and 500 base pairs, between about 100 and 700 basepairs, or between about 100 and 500 basepairs in length.

The target sequence can be isolated or derived from any suitable source, including genomic DNA, cDNA, mitochondrial DNA and chloroplast DNA. The plurality of target sequences can be from a eukaryote, prokaryote, virus or transposable element. The plurality of target sequences can be human, simian, ape, canine, feline, bovine, equine, murine, porcine, caprine, lupine, canine, piscine, plant, insect, bacterial or viral. The plurality of target sequences can comprise sequences from a plurality of sources, such as are found in samples isolated from hosts with parasites, commensal organisms, or communities such as biofilms.

Supports

Solid Supports Passivated with a Coating

The disclosure provides methods of sequencing pluralities of template nucleic acid molecules, wherein the plurality of template nucleic acid molecules are immobilized on a support. Exemplary supports include, but are not limited to, a surface of a flow cell.

In some embodiments, the support comprises a planar or non-planar support. The support can be solid or semi-solid. In some embodiments, the support can be porous, semi-porous or non-porous. In some embodiments, the surface of the support can be coated with one or more compounds to produce a passivated layer on the support. In some embodiments, the passivated layer forms a porous or semi-porous layer. In some embodiments, the nucleic acid primer or template, or the polymerase, can be attached to the passivated layer to immobilize the primer, template and/or polymerase to the support. In some embodiments, the support comprises a low non-specific binding surface that enable improved nucleic acid hybridization and amplification performance on the support. In general, the support may comprise one or more layers of a covalently or non-covalently attached low-binding, chemical modification layers, e.g., silane layers, polymer films, and one or more covalently or non-covalently attached

oligonucleotides that can be used for immobilizing a plurality of nucleic acid template molecules to the support. In some embodiments, the support can comprise a functionalized polymer coating layer covalently bound at least to a portion of the support via a chemical group on the support, a primer grafted to the functionalized polymer coating, and a water-soluble protective coating on the primer and the functionalized polymer coating. In some embodiments, the functionalized polymer coating comprises a poly(N-(5-azidoacetamidylpentyl)acrylamide-co-acrylamide (PAZAM). In some embodiments, the support comprises a surface coating having at least one hydrophilic polymer coating layer and at least one layer of a plurality of oligonucleotides. The hydrophilic polymer coating layer can comprise polyethylene glycol (PEG). The hydrophilic polymer coating layer can comprise branched PEG having at least 4 branches. In some embodiments, the low non-specific binding coating has a degree of hydrophilicity which can be measured as a water contact angle, where the water contact angle is no more than 45 degrees.

Zero Mode Waveguide Supports

In some embodiments, in the methods for sequencing, the support comprises a plurality of separate compartments and a sequencing polymerase is immobilized to the bottom of a compartment. Such supports are used in zero mode waveguide sequencing methods, which are contemplated as within the scope of the instant disclosure. In some embodiments, the separate compartments comprise a silica bottom through which light can penetrate. In some embodiments, the separate compartments comprise a silica bottom configured with a nanophotonic confinement structure comprising a hole in a metal cladding film (e.g., aluminum cladding film). In some embodiments, the hole in the metal cladding has a small aperture, for example, approximately 70 nm. In some embodiments, the height of the nanophotonic confinement structure is approximately 100 nm. In some embodiments, the nanophotonic confinement structure comprises a zero mode waveguide (ZMW). In some embodiments, the nanophotonic confinement structure contains a liquid.

In some embodiments, the detecting step comprises detecting the fluorescent signal emitted by the labeled multivalent bound by the polymerase complex.

Supports and Low Non-Specific Coatings

The present disclosure provides sequencing compositions and methods which employ a support comprising a plurality of surface primers immobilized thereon.

In some embodiments, the support is passivated with a low non-specific binding coating. The surface coatings described herein exhibit very low non-specific binding to reagents typically used for nucleic acid capture, amplification and sequencing workflows, such

as dyes, nucleotides, enzymes, and nucleic acid primers. The surface coatings exhibit low background fluorescence signals or high contrast-to-noise (CNR) ratios compared to conventional surface coatings.

The low non-specific binding coating comprises one layer or multiple layers. In some embodiments, the plurality of surface primers are immobilized to the low non-specific binding coating. In some embodiments, at least one surface primer is embedded within the low non-specific binding coating. The low non-specific binding coating enables improved nucleic acid hybridization and amplification performance. In general, the supports comprise a substrate (or support structure), one or more layers of a covalently or non-covalently attached low-binding, chemical modification layers, e.g., silane layers, polymer films, and one or more covalently or non-covalently attached surface primers that can be used for tethering single-stranded nucleic acid library molecules to the support. In some embodiments, the formulation of the coating, e.g., the chemical composition of one or more layers, the coupling chemistry used to cross-link the one or more layers to the support and/or to each other, and the total number of layers, may be varied such that non-specific binding of proteins, nucleic acid molecules, and other hybridization and amplification reaction components to the coating is minimized or reduced relative to a comparable monolayer. The formulation of the coating described herein may be varied such that non-specific hybridization on the coating is minimized or reduced relative to a comparable monolayer. The formulation of the coating may be varied such that non-specific amplification on the coating is minimized or reduced relative to a comparable monolayer. The formulation of the coating may be varied such that specific amplification rates and/or yields on the coating are maximized. Amplification levels suitable for detection are achieved in no more than 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, or more than 30 amplification cycles in some cases disclosed herein.

The support structures that comprise the one or more chemically-modified layers, e.g., layers of a low non-specific binding polymer, may be independent or integrated into another structure or assembly. For example, in some embodiments, the support structure may comprise one or more surfaces within an integrated or assembled microfluidic flow cell. The support structure may comprise one or more surfaces within a microplate format, e.g., the bottom surface of the wells in a microplate. In some embodiments, the support structure comprises the interior surface (such as the lumen surface) of a capillary. In some embodiments, the support structure comprises the interior surface (such as the lumen surface) of a capillary etched into a planar chip.

The attachment chemistry used to graft a first chemically-modified layer to the surface of the support will generally be dependent on both the material from which the surface is fabricated and the chemical nature of the layer. In some embodiments, the first layer may be covalently attached to the surface. In some embodiments, the first layer may be non-covalently attached, e.g., adsorbed to the support through non-covalent interactions such as electrostatic interactions, hydrogen bonding, or van der Waals interactions between the support and the molecular components of the first layer. In either case, the support may be treated prior to attachment or deposition of the first layer. Any of a variety of surface preparation techniques known to those of skill in the art may be used to clean or treat the surface. For example, glass or silicon surfaces may be acid-washed using a Piranha solution (a mixture of sulfuric acid (H_2SO_4) and hydrogen peroxide (H_2O_2)), base treatment in KOH and NaOH, and/or cleaned using an oxygen plasma treatment method.

Silane chemistries constitute non-limiting approaches for covalently modifying the silanol groups on glass or silicon surfaces to attach more reactive functional groups (e.g., amines or carboxyl groups), which may then be used in coupling linker molecules (e.g., linear hydrocarbon molecules of various lengths, such as C6, C12, C18 hydrocarbons, or linear polyethylene glycol (PEG) molecules) or layer molecules (e.g., branched PEG molecules or other polymers) to the surface. Examples of suitable silanes that may be used in creating any of the disclosed low binding coatings include, but are not limited to, (3-Aminopropyl) trimethoxysilane (APTMS), (3-Aminopropyl) triethoxysilane (APTES), any of a variety of PEG-silanes (e.g., comprising molecular weights of 1K, 2K, 5K, 10K, 20K, etc.), amino-PEG silane (i.e., comprising a free amino functional group), maleimide-PEG silane, biotin-PEG silane, and the like.

Any of a variety of molecules known to those of skill in the art including, but not limited to, amino acids, peptides, nucleotides, oligonucleotides, other monomers or polymers, or combinations thereof may be used in creating the one or more chemically-modified layers on the support, where the choice of components used may be varied to alter one or more properties of the layers, e.g., the surface density of functional groups and/or tethered oligonucleotide primers, the hydrophilicity/hydrophobicity of the layers, or the three three-dimensional nature (i.e., “thickness”) of the layer. Examples of polymers that may be used to create one or more layers of low non-specific binding material in any of the disclosed coatings include, but are not limited to, polyethylene glycol (PEG) of various molecular weights and branching structures, streptavidin, polyacrylamide, polyester, dextran, poly-lysine, and poly-lysine copolymers, or any combination thereof. Examples of conjugation chemistries that may be used to graft one

or more layers of material (e.g. polymer layers) to the surface and/or to cross-link the layers to each other include, but are not limited to, biotin-streptavidin interactions (or variations thereof), his tag – Ni/NTA conjugation chemistries, methoxy ether conjugation chemistries, carboxylate conjugation chemistries, amine conjugation chemistries, NHS esters, maleimides, thiol, epoxy, azide, hydrazide, alkyne, isocyanate, and silane.

The low non-specific binding surface coating may be applied uniformly across the support. Alternatively, the surface coating may be patterned, such that the chemical modification layers are confined to one or more discrete regions of the support. For example, the coating may be patterned using photolithographic techniques to create an ordered array or random pattern of chemically-modified regions on the support. Alternately or in combination, the coating may be patterned using, e.g., contact printing and/or ink-jet printing techniques. In some embodiments, an ordered array or random pattern of chemically-modified regions may comprise at least 1, 5, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 2000, 3000, 4000, 5000, 6000, 7000, 8000, 9000, or 10,000 or more discrete regions.

In some embodiments, the low nonspecific binding coatings comprise hydrophilic polymers that are non-specifically adsorbed or covalently grafted to the support. Typically, passivation is performed utilizing poly(ethylene glycol) (PEG, also known as polyethylene oxide (PEO) or polyoxyethylene) or other hydrophilic polymers with different molecular weights and end groups that are linked to a support using, for example, silane chemistry. The end groups distal from the surface can include, but are not limited to, biotin, methoxy ether, carboxylate, amine, NHS ester, maleimide, and bis-silane. In some embodiments, two or more layers of a hydrophilic polymer, e.g., a linear polymer, branched polymer, or multi-branched polymer, may be deposited on the surface. In some embodiments, two or more layers may be covalently coupled to each other or internally cross-linked to improve the stability of the resulting coating. In some embodiments, surface primers with different nucleotide sequences and/or base modifications (or other biomolecules, e.g., enzymes or antibodies) may be tethered to the resulting layer at various surface densities. In some embodiments, for example, both surface functional group density and surface primer concentration may be varied to attain a desired surface primer density range. Additionally, surface primer density can be controlled by diluting the surface primers with other molecules that carry the same functional group. For example, amine-labeled surface primers can be diluted with amine-labeled polyethylene glycol in a reaction with an NHS-ester coated surface to reduce the final primer density. Surface primers with different lengths of linker between the hybridization region and the surface attachment functional group can also be applied to control surface density. Example of suitable

linkers include poly-T and poly-A strands at the 5' end of the primer (e.g., 0 to 20 bases), PEG linkers (e.g., 3 to 20 monomer units), and carbon-chain (e.g., C6, C12, C18, etc.). To measure the primer density, fluorescently-labeled primers may be tethered to the surface and a fluorescence reading then compared with that for a dye solution of known concentration.

In some embodiments, the low nonspecific binding coatings comprise a functionalized polymer coating layer covalently bound at least to a portion of the support via a chemical group on the support, a primer grafted to the functionalized polymer coating, and a water-soluble protective coating on the primer and the functionalized polymer coating. In some embodiments, the functionalized polymer coating comprises a poly(N-(5-azidoacetamidylpentyl)acrylamide-co-acrylamide (PAZAM).

In order to scale primer surface density and add additional dimensionality to hydrophilic or amphoteric coatings, supports comprising multi-layer coatings of PEG and other hydrophilic polymers have been developed. By using hydrophilic and amphoteric surface layering approaches that include, but are not limited to, the polymer/co-polymer materials described below, it is possible to increase primer loading density on the support significantly. Traditional PEG coating approaches use monolayer primer deposition, which have been generally reported for single molecule applications, but do not yield high copy numbers for nucleic acid amplification applications. As described herein "layering" can be accomplished using traditional crosslinking approaches with any compatible polymer or monomer subunits such that a surface comprising two or more highly crosslinked layers can be built sequentially. Examples of suitable polymers include, but are not limited to, streptavidin, poly acrylamide, polyester, dextran, poly-lysine, and copolymers of poly-lysine and PEG. In some embodiments, the different layers may be attached to each other through any of a variety of conjugation reactions including, but not limited to, biotin-streptavidin binding, azide-alkyne click reaction, amine-NHS ester reaction, thiol-maleimide reaction, and ionic interactions between positively charged polymer and negatively charged polymer. In some embodiments, high primer density materials may be constructed in solution and subsequently layered onto the surface in multiple steps.

Examples of materials from which the support structure may be fabricated include, but are not limited to, glass, fused-silica, silicon, a polymer (e.g., polystyrene (PS), macroporous polystyrene (MPPS), polymethylmethacrylate (PMMA), polycarbonate (PC), polypropylene (PP), polyethylene (PE), high density polyethylene (HDPE), cyclic olefin polymers (COP), cyclic olefin copolymers (COC), polyethylene terephthalate (PET)), or any combination thereof. Various compositions of both glass and plastic support structures are contemplated.

The support structure may be rendered in any of a variety of geometries and dimensions known to those of skill in the art, and may comprise any of a variety of materials known to those of skill in the art. For example, the support structure may be locally planar (e.g., comprising a microscope slide or the surface of a microscope slide). Globally, the support structure may be cylindrical (e.g., comprising a capillary or the interior surface of a capillary), spherical (e.g., comprising the outer surface of a non-porous bead), or irregular (e.g., comprising the outer surface of an irregularly-shaped, non-porous bead or particle). In some embodiments, the surface of the support structure used for nucleic acid hybridization and amplification may be a solid, non-porous surface. In some embodiments, the surface of the support structure used for nucleic acid hybridization and amplification may be porous, such that the coatings described herein penetrate the porous surface, and nucleic acid hybridization and amplification reactions performed thereon may occur within the pores.

The support structure that comprises the one or more chemically-modified layers, e.g., layers of a low non-specific binding polymer, may be independent or integrated into another structure or assembly. For example, the support structure may comprise one or more surfaces within an integrated or assembled microfluidic flow cell. The support structure may comprise one or more surfaces within a microplate format, e.g., the bottom surface of the wells in a microplate. In some embodiments, the support structure comprises the interior surface (such as the lumen surface) of a capillary. In some embodiments the support structure comprises the interior surface (such as the lumen surface) of a capillary etched into a planar chip.

As noted, the low non-specific binding supports of the present disclosure exhibit reduced non-specific binding of proteins, nucleic acids, and other components of the hybridization and/or amplification formulation used for solid-phase nucleic acid amplification. The degree of non-specific binding exhibited by a given support surface may be assessed either qualitatively or quantitatively. For example, exposure of the surface to fluorescent dyes (e.g., cyanins such as Cy3, or Cy5, etc., fluoresceins, coumarins, rhodamines, etc. or other dyes disclosed herein), fluorescently-labeled nucleotides, fluorescently-labeled oligonucleotides, and/or fluorescently-labeled proteins (e.g. polymerases) under a standardized set of conditions, followed by a specified rinse protocol and fluorescence imaging may be used as a qualitative tool for comparison of non-specific binding on supports comprising different surface formulations. In some embodiments, exposure of the surface to fluorescent dyes, fluorescently-labeled nucleotides, fluorescently-labeled oligonucleotides, and/or fluorescently-labeled proteins (e.g. polymerases) under a standardized set of conditions, followed by a specified rinse protocol and fluorescence imaging may be used as a quantitative tool for comparison of non-

specific binding on supports comprising different surface formulations—provided that care has been taken to ensure that the fluorescence imaging is performed under conditions where fluorescence signal is linearly related (or related in a predictable manner) to the number of fluorophores on the support surface (e.g., under conditions where signal saturation and/or self-quenching of the fluorophore is not an issue) and suitable calibration standards are used. In some embodiments, other techniques known to those of skill in the art, for example, radioisotope labeling and counting methods may be used for quantitative assessment of the degree to which non-specific binding is exhibited by the different support surface formulations of the present disclosure.

Some surfaces disclosed herein exhibit a ratio of specific to nonspecific binding of a fluorophore such as Cy3 of at least 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 30, 35, 40, 50, 75, 100, or greater than 100, or any intermediate value spanned by the range herein. Some surfaces disclosed herein exhibit a ratio of specific to nonspecific fluorescence of a fluorophore such as Cy3 of at least 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 30, 35, 40, 50, 75, 100, or greater than 100, or any intermediate value spanned by the range herein.

The degree of non-specific binding exhibited by the disclosed low-binding supports may be assessed using a standardized protocol for contacting the surface with a labeled protein (e.g., bovine serum albumin (BSA), streptavidin, a DNA polymerase, a reverse transcriptase, a helicase, a single-stranded binding protein (SSB), etc., or any combination thereof), a labeled nucleotide, a labeled oligonucleotide, etc., under a standardized set of incubation and rinse conditions, followed be detection of the amount of label remaining on the surface and comparison of the signal resulting therefrom to an appropriate calibration standard. In some embodiments, the label may comprise a fluorescent label. In some embodiments, the label may comprise a radioisotope. In some embodiments, the label may comprise any other detectable label known to one of skill in the art. In some embodiments, the degree of non-specific binding exhibited by a given support surface formulation may thus be assessed in terms of the number of non-specifically bound protein molecules (or nucleic acid molecules or other molecules) per unit area. In some embodiments, the low-binding supports of the present disclosure may exhibit non-specific protein binding (or non-specific binding of other specified molecules, (e.g., cyanins such as Cy3, or Cy5, etc., fluoresceins, coumarins, rhodamines, etc. or other dyes disclosed herein)) of less than 0.001 molecule per μm^2, less than 0.01 molecule per μm^2, less than 0.1 molecule per μm^2, less than 0.25 molecule per μm^2, less than 0.5 molecule per μm^2, less than 1 molecule per μm^2, less than 10 molecules per μm^2, less than 100 molecules per μm^2,

or less than 1,000 molecules per μm^2. Those of skill in the art will realize that a given support surface of the present disclosure may exhibit non-specific binding falling anywhere within this range, for example, of less than 86 molecules per μm^2. For example, some modified surfaces disclosed herein exhibit nonspecific protein binding of less than 0.5 molecule/μm^2 following contact with a 1 μM solution of Cy3 labeled streptavidin (GE Amersham) in phosphate buffered saline (PBS) buffer for 15 minutes, followed by 3 rinses with deionized water. Some modified surfaces disclosed herein exhibit nonspecific binding of Cy3 dye molecules of less than 0.25 molecules per μm^2. In independent nonspecific binding assays, 1 μM labeled Cy3 SA (ThermoFisher), 1 μM Cy5 SA dye (ThermoFisher), 10 μM Aminoallyl-dUTP-ATTO-647N (Jena Biosciences), 10 μM Aminoallyl-dUTP-ATTO-Rhol 1 (Jena Biosciences), 10 μM Aminoallyl-dUTP-ATTO-Rhol 1 (Jena Biosciences), 10 μM 7-Propargylamino-7-deaza-dGTP-Cy5 (Jena Biosciences, and 10 μM 7-Propargylamino-7-deaza-dGTP-Cy3 (Jena Biosciences) were incubated on the low binding coated supports at 37° C. for 15 minutes in a 384 well plate format. Each well was rinsed 2-3× with 50 ul deionized RNase/DNase Free water and 2-3× with 25 mM ACES buffer pH 7.4. The 384 well plates were imaged on a GE Typhoon instrument using the Cy3, AF555, or Cy5 filter sets (according to dye test performed) as specified by the manufacturer at a PMT gain setting of 800 and resolution of 50-100 μm. For higher resolution imaging, images were collected on an Olympus IX83 microscope (e.g., inverted fluorescence microscope) (Olympus Corp., Center Valley, Pa.) with a total internal reflectance fluorescence (TIRF) objective (100×, 1.5 NA, Olympus), a CCD camera (e.g., an Olympus EM-CCD monochrome camera, Olympus XM-10 monochrome camera, or an Olympus DP80 color and monochrome camera), an illumination source (e.g., an Olympus 100W Hg lamp, an Olympus 75W Xe lamp, or an Olympus U-HGLGPS fluorescence light source), and excitation wavelengths of 532 nm or 635 nm. Dichroic mirrors were purchased from Semrock (IDEX Health & Science, LLC, Rochester, N.Y.), e.g., 405, 488, 532, or 633 nm dichroic reflectors/beamsplitters, and band pass filters were chosen as 532 LP or 645 LP concordant with the appropriate excitation wavelength. Some modified surfaces disclosed herein exhibit nonspecific binding of dye molecules of less than 0.25 molecules per μm^2. In some embodiments, the coated support was immersed in a buffer (e.g., 25 mM ACES, pH 7.4) while the image was acquired.

In some embodiments, the surfaces disclosed herein exhibit a ratio of specific to nonspecific binding of a fluorophore such as Cy3 of at least 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 30, 35, 40, 50, 75, 100, or greater than 100, or any intermediate value spanned by the range herein. In some embodiments, the surfaces disclosed herein exhibit

a ratio of specific to nonspecific fluorescence signals for a fluorophore such as Cy3 of at least 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 30, 35, 40, 50, 75, 100, or greater than 100, or any intermediate value spanned by the range herein.

The low-background surfaces consistent with the disclosure herein may exhibit specific dye attachment (e.g., Cy3 attachment) to non-specific dye adsorption (e.g., Cy3 dye adsorption) ratios of at least 4:1, 5:1, 6:1, 7:1, 8:1, 9:1, 10:1, 15:1, 20:1, 30:1, 40:1, 50:1, or more than 50 specific dye molecules attached per molecule nonspecifically adsorbed. Similarly, when subjected to an excitation energy, low-background surfaces consistent with the disclosure herein to which fluorophores, e.g., Cy3, have been attached may exhibit ratios of specific fluorescence signal (e.g., arising from Cy3-labeled oligonucleotides attached to the surface) to non-specific adsorbed dye fluorescence signals of at least 4:1, 5:1, 6:1, 7:1, 8:1, 9:1, 10:1, 15:1, 20:1, 30:1, 40:1, 50:1, or more than 50:1.

In some embodiments, the degree of hydrophilicity (or "wettability" with aqueous solutions) of the disclosed support surfaces may be assessed, for example, through the measurement of water contact angles in which a small droplet of water is placed on the surface and its angle of contact with the surface is measured using, e.g., an optical tensiometer. In some embodiments, a static contact angle may be determined. In some embodiments, an advancing or receding contact angle may be determined. In some embodiments, the water contact angle for the hydrophilic, low-binding support surfaced disclosed herein may range from about 0 degrees to about 30 degrees. In some embodiments, the water contact angle for the hydrophilic, low-binding support surfaced disclosed herein may no more than 50 degrees, 40 degrees, 30 degrees, 25 degrees, 20 degrees, 18 degrees, 16 degrees, 14 degrees, 12 degrees, 10 degrees, 8 degrees, 6 degrees, 4 degrees, 2 degrees, or 1 degree. In many cases the contact angle is no more than 40 degrees. Those of skill in the art will realize that a given hydrophilic, low-binding support surface of the present disclosure may exhibit a water contact angle having a value of anywhere within this range.

In some embodiments, the hydrophilic surfaces disclosed herein facilitate reduced wash times for bioassays, often due to reduced nonspecific binding of biomolecules to the low-binding surfaces. In some embodiments, adequate wash steps may be performed in less than 60, 50, 40, 30, 20, 15, 10, or less than 10 seconds. For example, adequate wash steps may be performed in less than 30 seconds.

Some low-binding surfaces of the present disclosure exhibit significant improvement in stability or durability to prolonged exposure to solvents and elevated temperatures, or to repeated cycles of solvent exposure or changes in temperature. For example, the stability of the

disclosed surfaces may be tested by fluorescently labeling a functional group on the surface, or a tethered biomolecule (e.g., an oligonucleotide primer) on the surface, and monitoring fluorescence signal before, during, and after prolonged exposure to solvents and elevated temperatures, or to repeated cycles of solvent exposure or changes in temperature. In some embodiments, the degree of change in the fluorescence used to assess the quality of the surface may be less than 1%, 2%, 3%, 4%, 5%, 10%, 15%, 20%, or 25% over a time period of 1 minute, 2 minutes, 3 minutes, 4 minutes, 5 minutes, 10 minutes, 20 minutes, 30 minutes, 40 minutes, 50 minutes, 60 minutes, 2 hours, 3 hours, 4 hours, 5 hours, 6 hours, 7 hours, 8 hours, 9 hours, 10 hours, 15 hours, 20 hours, 25 hours, 30 hours, 35 hours, 40 hours, 45 hours, 50 hours, or 100 hours of exposure to solvents and/or elevated temperatures (or any combination of these percentages as measured over these time periods). In some embodiments, the degree of change in the fluorescence used to assess the quality of the surface may be less than 1%, 2%, 3%, 4%, 5%, 10%, 15%, 20%, or 25% over 5 cycles, 10 cycles, 20 cycles, 30 cycles, 40 cycles, 50 cycles, 60 cycles, 70 cycles, 80 cycles, 90 cycles, 100 cycles, 200 cycles, 300 cycles, 400 cycles, 500 cycles, 600 cycles, 700 cycles, 800 cycles, 900 cycles, or 1,000 cycles of repeated exposure to solvent changes and/or changes in temperature (or any combination of these percentages as measured over this range of cycles).

In some embodiments, the surfaces disclosed herein may exhibit a high ratio of specific signal to nonspecific signal or other background. For example, when used for nucleic acid amplification, some surfaces may exhibit an amplification signal that is at least 4, 5, 6, 7, 8, 9, 10, 15, 20, 30, 40, 50, 75, 100, or greater than 100 fold greater than a signal of an adjacent unpopulated region of the surface. Similarly, some surfaces exhibit an amplification signal that is at least 4, 5, 6, 7, 8, 9, 10, 15, 20, 30, 40, 50, 75, 100, or greater than 100 fold greater than a signal of an adjacent amplified nucleic acid population region of the surface.

In some embodiments, fluorescence images of the disclosed low background surfaces when used in nucleic acid hybridization or amplification applications to create polonies of hybridized or clonally-amplified nucleic acid molecules (e.g., that have been directly or indirectly labeled with a fluorophore) exhibit contrast-to-noise ratios (CNRs) of at least 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 20, 210, 220, 230, 240, 250, or greater than 250.

One or more types of primer may be attached or tethered to the support surface. In some embodiments, the one or more types of adapters or primers may comprise spacer sequences, adapter sequences for hybridization to adapter-ligated target library nucleic acid sequences, forward amplification primers, reverse amplification primers, sequencing primers, and/or

molecular barcoding sequences, or any combination thereof. In some embodiments, 1 primer or adapter sequence may be tethered to at least one layer of the surface. In some embodiments, at least 2, 3, 4, 5, 6, 7, 8, 9, 10, or more than 10 different primer or adapter sequences may be tethered to at least one layer of the surface.

In some embodiments, the tethered adapter and/or primer sequences may range in length from about 10 nucleotides to about 100 nucleotides. In some embodiments, the tethered adapter and/or primer sequences may be at least 10, at least 20, at least 30, at least 40, at least 50, at least 60, at least 70, at least 80, at least 90, or at least 100 nucleotides in length. In some embodiments, the tethered adapter and/or primer sequences may be at most 100, at most 90, at most 80, at most 70, at most 60, at most 50, at most 40, at most 30, at most 20, or at most 10 nucleotides in length. Any of the lower and upper values described in this paragraph may be combined to form a range included within the present disclosure, for example, in some embodiments the length of the tethered adapter and/or primer sequences may range from about 20 nucleotides to about 80 nucleotides. Those of skill in the art will recognize that the length of the tethered adapter and/or primer sequences may have any value within this range, e.g., about 24 nucleotides.

In some embodiments, the resultant surface density of primers (e.g., capture primers) on the low binding support surfaces of the present disclosure may range from about 100 primer molecules per μm^2 to about 100,000 primer molecules per μm^2. In some embodiments, the resultant surface density of primers on the low binding support surfaces of the present disclosure may range from about 1,000 primer molecules per μm^2 to about 1,000,000 primer molecules per μm^2. In some embodiments, the surface density of primers may be at least 1,000, at least 10,000, at least 100,000, or at least 1,000,000 molecules per μm^2. In some embodiments, the surface density of primers may be at most 1,000,000, at most 100,000, at most 10,000, or at most 1,000 molecules per μm^2. Any of the lower and upper values described in this paragraph may be combined to form a range included within the present disclosure, for example, in some embodiments the surface density of primers may range from about 10,000 molecules per μm^2 to about 100,000 molecules per μm^2. Those of skill in the art will recognize that the surface density of primer molecules may have any value within this range, e.g., about 455,000 molecules per μm^2. In some embodiments, the surface density of target library nucleic acid sequences initially hybridized to adapter or primer sequences on the support surface may be less than or equal to that indicated for the surface density of tethered primers. In some embodiments, the surface density of clonally-amplified target library nucleic acid sequences

hybridized to adapter or primer sequences on the support surface may span the same range as that indicated for the surface density of tethered primers.

Local densities as listed above do not preclude variation in density across a surface, such that a surface may comprise a region having an oligo density of, for example, 500,000/μm^2, while also comprising at least a second region having a substantially different local density.

In some embodiments, the performance of nucleic acid hybridization and/or amplification reactions using the disclosed reaction formulations and low-binding supports may be assessed using fluorescence imaging techniques, where the contrast-to-noise ratio (CNR) of the images provides a key metric in assessing amplification specificity and non-specific binding on the support. CNR is commonly defined as: CNR=(Signal-Background)/Noise, and is described in US 202/0149095. The background term is commonly taken to be the signal measured for the interstitial regions surrounding a particular feature (diffraction limited spot, DLS) in a specified region of interest (ROI). While signal-to-noise ratio (SNR) is often considered to be a benchmark of overall signal quality, it can be shown that improved CNR can provide a significant advantage over SNR as a benchmark for signal quality in applications that require rapid image capture (e.g., sequencing applications for which cycle times must be minimized), as shown in the example below. At high CNR the imaging time required to reach accurate discrimination (and thus accurate base-calling in the case of sequencing applications) can be drastically reduced even with moderate improvements in CNR. Improved CNR in imaging data on the imaging integration time provides a method for more accurately detecting features such as clonally-amplified nucleic acid colonies on the support surface.

In most ensemble-based sequencing approaches, the background term is typically measured as the signal associated with “interstitial” regions. In addition to "interstitial" background (B_{inter}), "intrastitial" background (B_{intra}) exists within the region occupied by an amplified DNA colony. The combination of these two background signals dictates the achievable CNR, and subsequently directly impacts the optical instrument requirements, architecture costs, reagent costs, run-times, cost/genome, and ultimately the accuracy and data quality for cyclic array-based sequencing applications. The B_{inter} background signal arises from a variety of sources; a few examples include auto-fluorescence from consumable flow cells, non-specific adsorption of detection molecules that yield spurious fluorescence signals that may obscure the signal from the ROI, the presence of non-specific DNA amplification products (e.g., those arising from primer dimers). In typical next generation sequencing (NGS)

applications, this background signal in the current field-of-view (FOV) is averaged over time and subtracted. The signal arising from individual DNA colonies (i.e., (Signal)-B(interstial) in the FOV) yields a discernable feature that can be classified. In some embodiments, the intrastitial background (B(intrastitial)) can contribute a confounding fluorescence signal that is not specific to the target of interest, but is present in the same ROI thus making it far more difficult to average and subtract.

Nucleic acid amplification on the low-binding coated supports described herein may decrease the B(interstitial) background signal by reducing non-specific binding, may lead to improvements in specific nucleic acid amplification, and may lead to a decrease in non-specific amplification that can impact the background signal arising from both the interstitial and intrastitial regions. In some embodiments, the disclosed low-binding coated supports, optionally used in combination with the disclosed hybridization and/or amplification reaction formulations, may lead to improvements in CNR by a factor of 2, 5, 10, 100, 250, 500 or 1000-fold over those achieved using conventional supports and hybridization, amplification, and/or sequencing protocols. Although described here in the context of using fluorescence imaging as the read-out or detection mode, the same principles apply to the use of the disclosed low-binding coated supports and nucleic acid hybridization and amplification formulations for other detection modes as well, including both optical and non-optical detection modes.

Polymerases

The present disclosure provides methods for sequencing nucleic acid molecules, where any of the sequencing methods described herein employ at least one type of polymerase and a plurality of nucleotides, or employ at least one type of polymerase and a plurality of nucleotides and a plurality of polymeric molecules. In some embodiments, the polymerase(s) is/are capable of incorporating a complementary nucleotide opposite a nucleotide in a template molecule. In some embodiments, the polymerase(s) is/are capable of binding a complementary nucleotide moiety of a polymeric molecule opposite a nucleotide in a template nucleic acid molecule (or the complement thereof, in paired end sequencing). In some embodiments, the plurality of polymerases comprise recombinant mutant polymerases.

Examples of suitable polymerases for use in sequencing with nucleotides and/or polymeric molecules include but are not limited to: Klenow DNA polymerase; Thermus aquaticus DNA polymerase I (Taq polymerase); KlenTaq polymerase; Candidatus altiarchaeales archaeon; Candidatus Hadarchaeum Yellowstonense; Hadesarchaea archaeon; Euryarchaeota archaeon; Thermoplasmata archaeon; Thermococcus polymerases such as

Thermococcus litoralis, bacteriophage T7 DNA polymerase; human alpha, delta and epsilon DNA polymerases; bacteriophage polymerases such as T4, RB69 and phi29 bacteriophage DNA polymerases; Pyrococcus furiosus DNA polymerase (Pfu polymerase); Bacillus subtilis DNA polymerase III; *E. coli* DNA polymerase III alpha and epsilon; 9 degree N polymerase; reverse transcriptases such as HIV type M or O reverse transcriptases; avian myeloblastosis virus reverse transcriptase; Moloney Murine Leukemia Virus (MMLV) reverse transcriptase; or telomerase. Further non-limiting examples of DNA polymerases include those from various Archaea genera, such as, Aeropyrum, Archaeglobus, Desulfurococcus, Pyrobaculum, Pyrococcus, Pyrolobus, Pyrodictium, Staphylothermus, Stetteria, Sulfolobus, Thermococcus, and Vulcanisaeta and the like or variants thereof, including such polymerases as are known in the art such as 9 degrees N, VENT, DEEP VENT, THERMINATOR, Pfu, KOD, Pfx, Tgo and RB69 polymerases.

Kits and Articles of Manufacture

The disclosure provides kits comprising the polymeric molecules, primers and/or reagents for carrying out the sequencing methods described herein. In some embodiments, the kits comprise target nucleic acid sequences for use a positive control. In some embodiments, the kits comprise vials, tubes, boxes and the like.

In some embodiments, the kits comprise instructions for use.

It is to be appreciated that the Detailed Description section, and not any other section, is intended to be used to interpret the claims. Other sections may set forth one or more but not all exemplary embodiments as contemplated by the inventor(s), and thus, are not intended to limit this disclosure or the appended claims in any way.

While this disclosure describes exemplary embodiments for exemplary fields and applications, it should be understood that the disclosure is not limited thereto. Other embodiments and modifications thereto are possible, and are within the scope and spirit of this disclosure. For example, and without limiting the generality of this paragraph, embodiments are not limited to the software, hardware, firmware, and/or entities illustrated in the figures and/or described herein. Further, embodiments (whether or not explicitly described herein) have significant utility to fields and applications beyond the examples described herein.

Embodiments have been described herein with the aid of functional building blocks illustrating the implementation of specified functions and relationships thereof. The boundaries

of these functional building blocks have been arbitrarily defined herein for the convenience of the description. Alternate boundaries may be defined as long as the specified functions and relationships (or equivalents thereof) are appropriately performed. Also, alternative embodiments may perform functional blocks, steps, operations, methods, etc. using orderings different from those described herein.

References herein to "one embodiment," "an embodiment," "an example embodiment," "some embodiments," or similar phrases, indicate that the embodiment described may include a particular feature, structure, or characteristic, but every embodiment may not necessarily include the particular feature, structure, or characteristic. Moreover, such phrases are not necessarily referring to the same embodiment. Further, when a particular feature, structure, or characteristic is described in connection with an embodiment, it would be within the knowledge of persons skilled in the relevant art(s) to incorporate such feature, structure, or characteristic into other embodiments whether or not explicitly mentioned or described herein.

Additionally, some embodiments may be described using the expression "coupled" and "connected" along with their derivatives. These terms are not necessarily intended as synonyms for each other. For example, some embodiments may be described using the terms "connected" and/or "coupled" to indicate that two or more elements are in direct physical or electrical contact with each other. The term "coupled," however, may also mean that two or more elements are not in direct contact with each other, but yet still co-operate or interact with each other.

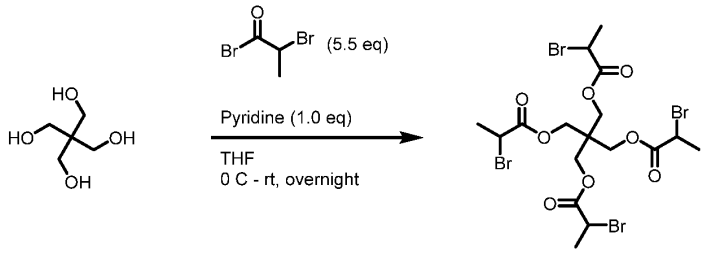
While preferred embodiments of the present invention have been shown and described herein, it will be obvious to those skilled in the art that such embodiments are provided by way of example only. The breadth and scope of this disclosure should not be limited by any of the above-described exemplary embodiments. Numerous variations, changes, and substitutions will now occur to those skilled in the art without departing from the invention. It should be understood that various alternatives to the embodiments of the invention described herein may be employed in practicing the invention. It is intended that the following claims define the scope of the invention and that methods and structures within the scope of these claims and their equivalents be covered thereby.

EXAMPLES

Example 1: Synthesis of Chain Transfer Agent

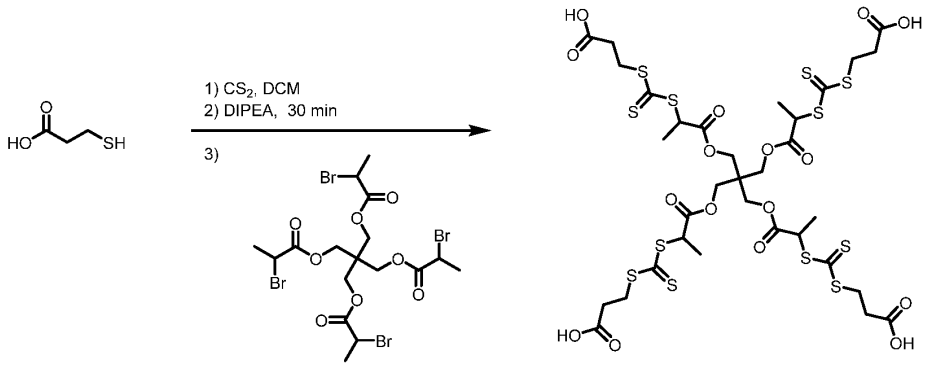
The chain transfer agent was prepared as described in Schemes 1 and 2 and the accompanying descriptions.

Scheme 1: Synthesis of brominated precursor



To a 250 mL round bottom flask containing a stir bar was added pentaerythritol (2.5 g, 18.38 mmol, 1.0 eq.) and THF (50 mL). The solution was placed under an Argon balloon, cooled to 0 C with an ice bath, and stirred vigorously. Then 2-bromopropionyl bromide (21.84 g, 101.1 mmol, 5.5 eq.) was added dropwise and the reaction allowed to stir for 30 minutes. Pyridine (1.45 g, 18.38 mmol, 1.0 eq) was then added dropwise and the reaction allowed to stir overnight. The reaction was then diluted with diethyl ether (130 mL) and transferred to a separatory funnel (250 mL). The organic layer was then sequentially washed with water (40 mL), saturated sodium bicarbonate (3 x 40 mL), and finally water (40 mL). The organic layer was then dried over sodium sulfate, filtered, and concentrated under reduced pressure. The crude oil was taken up in a minimal amount of hot ethanol and product is crystallized out upon cooling with an ice bath. The white precipitate obtained was filtered off and dried under vacuum overnight yielding the desired product as a white powder (6.3 g, 51% yield).

Scheme 2: Synthesis of Chain Transfer Agent



To a 25 mL round bottom flask equipped with a stirbar was added 3-mercaptopropionic acid (700 mg, 6.6 mmol, 4.5 eq), DCM (15 mL) and diisopropyl ethylamine (DIPEA, 1.3 mL, 7.4 mmol, 5 eq). The solution was stirred vigorously and carbon disulfide (CS_2, 676 mg, 8.9

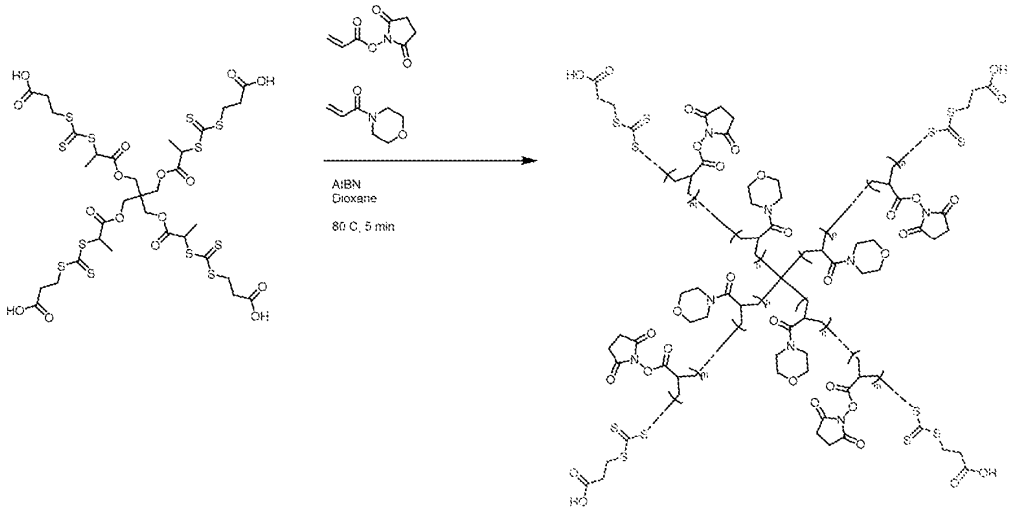
mmol, 6.0 eq) was added dropwise. Upon addition of CS_2 the reaction became yellow-orange indicating the formation of the tri-thiocarbonate intermediate. After 30 minutes 2,2-bis(((2-bromopropanoyl)oxy)methyl)propane-1,3-diyl bis(2-bromopropanoate (1.0 g, 1.48 mmol, 1.0 eq) was added to the solution in one portion as a solid and the reaction allowed to stir vigorously. After 3 h RP-HPLC indicated the full conversion of intermediates to the desired product. The reaction was then diluted with DCM (100 mL) and transferred to a separatory funnel (125 mL). The organic layer was then washed sequentially with water (25 mL), brine (25 mL), and 0.1 M TEAB (3 x 25 mL). The organic layer was then concentrated under reduced pressure and the material was purified on silica gel (0-20% Methanol/DCM gradient). The desired product was obtained as a sticky yellow semi-solid/oil (402 mg, 25.2% yield).

If the reaction stalls, steps 1 and 2 can be repeated in a second flask (20% of above amounts) and then added to the reaction to push to completion.

Example 2: Synthesis of Polymeric Side Chain

The polymeric side chain was synthesized as described in Scheme 3 and accompanying description.

Scheme 3: Synthesis of polymeric side chain



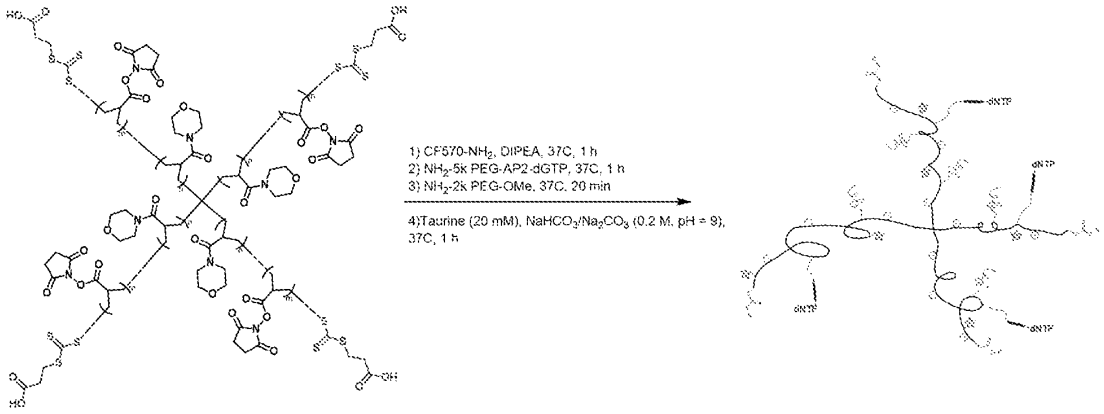
To a 1-dram vial containing a stir bar was added N-Acryloxysuccinimide (78.1 mg, 0.462 mmol), 4-acryloylmorpholine (97.9 mg, 0.694 mmol), CTA (10 mg, 9.25 umol), AIBN (0.15 mg, 0.925 umol), and dry dioxane (600 uL). The vial was fitted with a 14/20 septum and sealed with electrical tape. The reaction was then sparged with Argon for 15 minutes, then kept under inert atmosphere with a balloon. The vial was then added to a hotplate (preheated to 80 C) and the reaction allowed to stir vigorously for 5 minutes. The reaction was then opened to

air and the solution was added dropwise to a falcon tube containing diethyl ether (12 mL), precipitating the polymer as a white solid. The tube was centrifuged, and the solvent decanted. The polymer was placed under vacuum for 5 minutes before being taken back up in a minimal amount of dichloromethane (~1.5 mL). The dichloromethane solution was then added dropwise to a falcon tube containing diethyl ether (12 mL). The process repeated a total of three times. The polymer was then allowed to dry under vacuum overnight and stored in a desiccator.

Example 3: Functionalization of Polymeric Side Chain

The functionalization of the polymeric side chain is described in Scheme 4 and accompanying description.

Scheme 4: functionalization of polymeric side chain



To a 1-dram vial containing a stir bar was added 4-arm star polymer (1 mg/25 uL dry DMSO, 160 uL, 284 nmol) followed by CF570-NH_2 (20.6 mM in DMSO, 87.4 uL, 1800 nmol) and diisopropylethylamine (12 uL, 1.9 umol). The vial was sealed and the solution was stirred at 37 °C for 1 h. Then NH_2-5kPEG-AP2-dGTP (27.5 mM in H_2O, 43.6 uL, 1200 nmol) was added to above reaction solution. The resulting reaction mixture was stirred at 37 °C for 1 h. NH_2-2kPEG-OMe (20 mM in DMSO, 160 uL, 3200 nmol) was added and the resulting reaction mixture was stirred at 37C for 20 min. Taurine (0.2 M) in $NaHCO_3/Na_2CO_3$ (0.2M, pH=8.9, 160 uL) buffer was added to above reaction. The reaction was further stirred at 37 °C over 1 h. The reaction was transferred to a 30 kDa molecular weight cut off spin filter and was then diluted with HW buffer (3 mL), and was then centrifuged. Spin filtration process repeated until no free dye was detected by SEC.

CLAIMS

What is claimed is:

1. A method comprising:

a. contacting:

(i) a plurality of template nucleic acid molecules comprising two or more copies of a target sequence and two or more copies of a binding sequence for a forward sequencing primer,

(ii) a plurality of forward sequencing primers comprising a sequence complementary to the binding sequence for the forward sequencing primer,

(iii) a plurality of first polymerases, and

(iv) a plurality of polymeric molecules of Formula (I):

$$C\text{-}((P)\text{-}E)_s$$

(I)

an ionized form thereof, or a salt thereof, wherein:

C is a central moiety;

each P independently is an optionally substituted polymeric side chain;

each E independently is an end moiety; and

s is an integer ranging from 1 to 10;

wherein each polymeric molecule comprises at least two nucleotide moieties and at one least detectable reporter moiety,

wherein the contacting occurs under conditions sufficient to form a plurality of multivalent binding complexes comprising a nucleic acid duplex between a template nucleic acid molecule and forward sequencing primer, a first polymerase, and a nucleotide moiety of a polymeric molecule that is complementary to the nucleotide in the template nucleic acid molecule immediately adjacent to the 3' end of the forward sequencing primer, and

wherein polymerase catalyzed incorporation of a complementary nucleotide moiety into the nucleic acid duplex is inhibited;

b. detecting the detectable reporter moieties; and

c. determining the identities of nucleotides in the nucleic acid template molecules based on the detectable reporter moieties of the polymeric molecules in the plurality of multivalent binding complexes formed in step (a).

2. The method of claim 1, wherein the compound of Formula (I) is of Formula (II):

$C\text{-}((P)\text{-}E)_4$

(II)

an ionized form thereof, or a salt thereof.

3. The method of claim 1 or 2, wherein at least one P is substituted with (i) one or more reporter moiety and (ii) one or more nucleotide moiety.
4. The method of any one of claims 1-3, wherein each P is substituted with (i) one or more reporter moiety and (ii) one or more nucleotide moiety.
5. The method of any one of claims 1-4, wherein at least one P is substituted with one or more blocking moiety, negative charge moiety, or PEG-Cap moiety.
6. The method of any one of claims 1-5, wherein each P is substituted with one or more blocking moiety, negative charge moiety, or PEG-Cap moiety.
7. The method of any one of claims 1-6, wherein at least one P is further substituted with (iii) one or more blocking moiety, (iv) one or more negative charge moiety, and (v) one or more PEG-Cap moiety.
8. The method of any one of claims 1-7, wherein each P is further substituted with (iii) one or more blocking moiety, (iv) one or more negative charge moiety , and (v) one or more PEG-Cap moiety.
9. The method of any one of claims 1-8, wherein the two or more copies of a target sequence in an individual of template nucleic acid molecule are the same target sequence.
10. The method of any one of claims 1-9, wherein two or more multivalent binding complexes form on individual template nucleic acid molecules.
11. The method of any one of any one of claims 1-10, wherein the plurality of forward sequencing primers are soluble.
12. The method of any one of claims 1-11, comprising:
    d. dissociating the multivalent binding complexes under conditions sufficient to retain the nucleic acid duplexes, thereby generating a plurality of nucleic acid duplexes;
    e. contacting the plurality of nucleic acid duplexes with a plurality of second polymerases and a plurality of nucleotides or analogs thereof under conditions sufficient to incorporate nucleotides or analogs thereof complementary to the nucleotides of the nucleic acid template molecules immediately adjacent to the 3’ ends of the forward sequencing primers in a primer extension reaction, thereby generating a plurality of extended nucleic acid duplexes comprising extended forward sequencing primer sequences.

13. The method of claim 12, comprising:

    f. dissociating the second polymerases from the extended nucleic acid duplexes under conditions sufficient to retain the plurality of extended nucleic acid duplexes.

14. The method of any one of claims 1-13, wherein the template nucleic acid molecules comprise concatemers of two or more copies of a sequence comprising (i) the binding sequence for the forward sequencing primer and (ii) the target sequence.

15. The method of claim 14, wherein the two or more copies of (i) the binding sequence for the forward sequencing primer hybridize to the forward sequencing primers to form nucleic acid duplexes between the template nucleic acid molecules and the forward sequencing primers.

16. The method of any one of claims 1-15, wherein, in an individual polymeric molecule, the at least two nucleotide moieties are attached to different polymeric side chains.

17. The method of any one of claims 1-16 wherein, in an individual polymeric molecule, all nucleotide moieties are the same.

18. The method of claim 17, wherein all nucleotide moieties in an individual polymeric molecule are dATP.

19. The method of claim 17, wherein all nucleotide moieties in an individual polymeric molecule are dTTP.

20. The method of claim 17, wherein all nucleotide moieties in an individual polymeric molecule are dGTP.

21. The method of claim 17, wherein all nucleotide moieties in an individual polymeric molecule are dUTP

22. The method of claim 17, wherein all nucleotide moieties in an individual polymeric molecule are dCTP.

23. The method of any one of claims 1-22, wherein, in an individual polymeric molecule, all detectable reporter moieties comprise the same fluorescent label.

24. The method of claim 23, wherein all detectable reporter moieties in an individual polymeric molecule are the same.

25. The method of any one of claims 1-24, wherein, in an individual polymeric molecule, all detectable reporter moieties are the same and all nucleotide moieties are the same.

26. The method of any one of claims 1-25, wherein a polymeric molecule comprises two, three, or four nucleotide moieties.

27. The method of any one of claims 1-25, wherein an individual polymeric molecule comprises two, three, or four detectable reporter moieties.

28. The method of any one of claims 1-27, wherein two or more nucleotide moieties in an individual polymeric molecule are associated with two or more different multivalent binding complexes on the same template nucleic acid molecule.

29. The method of any one of claims 13-28, comprising

i. contacting the plurality of extended nucleic acid duplexes with a plurality of first polymerases and a plurality of polymeric molecules of Formula (I), or an ionized form thereof, an isomer thereof, or a salt thereof,

wherein the contacting occurs under conditions sufficient to form a plurality of multivalent binding complexes comprising an extended nucleic acid duplex, a first polymerase, and a nucleotide moiety of a polymeric molecule that is complementary to a nucleotide in the template nucleic acid molecule immediately adjacent to the 3' end of the extended forward sequencing primer, and

wherein polymerase catalyzed incorporation of a complementary nucleotide moiety into the extended nucleic acid duplex is inhibited;

ii. detecting the detectable reporter moieties; and

iii. determining nucleobase identities of nucleotides in the nucleic acid template sequences complementary to the nucleotide moieties of the polymeric molecules based on the detectable reporter moieties of the polymeric molecules in the plurality of multivalent binding complexes formed in step (a);

iv. dissociating the multivalent binding complexes under conditions sufficient to retain the plurality extended nucleic acid duplexes;

v. contacting the plurality of extended nucleic acid duplexes with a plurality of second polymerases and a plurality of nucleotides or analogs thereof under conditions sufficient to incorporate nucleotides or analogs thereof complementary to the nucleotides of the nucleic acid template molecules immediately adjacent to the 3' ends of the extended forward sequencing primers in a primer extension reaction, thereby generating a plurality of extended

nucleic acid duplexes comprising extended forward sequencing primers.

30. The method of claim 29, wherein, in an individual polymeric molecule, the at least two nucleotide moieties are attached to different polymeric side chains.

31. The method of claim 29 or 30, wherein, in an individual polymeric molecule, all nucleotide moieties are the same.

32. The method of claim 31, wherein all nucleotide moieties in an individual polymeric molecule are dATP.

33. The method of claim 31, wherein all nucleotide moieties in an individual polymeric molecule are dTTP.

34. The method of claim 31, wherein all nucleotide moieties in an individual polymeric molecule are dGTP.

35. The method of claim 31, wherein all nucleotide moieties in an individual polymeric molecule are dUTP.

36. The method of claim 31, wherein all nucleotide moieties in an individual polymeric molecule are dCTP.

37. The method of any one of claims 29-36, wherein, in an individual polymeric molecule, all detectable reporter moieties comprise the same fluorescent label.

38. The method of claim 37, wherein all detectable reporter moieties in an individual polymeric molecule are the same.

39. The method of any one of claims 29-38, wherein, in an individual polymeric molecule, all detectable reporter moieties are the same and all nucleotide moieties are the same.

40. The method of any one of claims 29-39, wherein an individual polymeric molecule comprises two, three, or four nucleotide moieties.

41. The method of any one of claims 29-40, wherein an individual polymeric molecule comprises two, three, or four detectable reporter moieties.

42. The method of any one of claims 29-41, wherein two or more nucleotide moieties in an individual polymeric molecule are associated with two or more different multivalent binding complexes on the same template nucleic acid molecule.

43. The method of any one of claims 29-42, comprising repeating steps (i)-(v) at least 1, 10, 20, 30, 40, 50, 70, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 800, 900, 1000, or 1500 times.

44. The method of any one of claims 29-43, comprising repeating steps (i)-(v) until the identities of the nucleotides in the target sequences have been determined.

45. The method of any one of claims 29-44, comprising, before step (i), dissociating the second polymerases from the extended nucleic acid duplexes under conditions sufficient to retain the plurality of extended nucleic acid duplexes.
46. The method of any one of claims 29-45, wherein the template nucleic acid molecules are single-stranded DNA molecules.
47. The method of any one of claims 29-46, wherein the nucleotides or analogs thereof comprise a removable chain terminating moiety at the 3' sugar group.
48. The method of claim 47, wherein the removable chain terminating moiety comprises an alkyl group, alkenyl group, alkynyl group, allyl group, aryl group, benzyl group, azide group, azido group, O-azidomethyl group, amine group, amide group, keto group, isocyanate group, phosphate group, thio group, disulfide group, carbonate group, urea group, or silyl group, and wherein the removable chain terminating moiety is cleavable with a chemical compound to generate an extendible 3'OH moiety on the sugar group.
49. The method of any one of claims 1-48, wherein the nucleotides or analogs thereof comprise a mixture of any combination of two or more types of nucleotides selected from the group consisting of dATP, dGTP, dCTP, dTTP and dUTP.
50. The method of any one of claims 1-49, wherein the nucleotides or analogs thereof comprise at least one fluorophore-labeled nucleotide analog.
51. The method of any one of claims 1-50, wherein the plurality of template nucleic acid molecules are immobilized on a support.
52. The method of claim 51, wherein the template nucleic acid molecules are immobilized on the support through hybridization to first surface primers immobilized on the support.
53. The method of claim 51, wherein the template nucleic acid molecules are covalently joined to first surface primers immobilized on the support.
54. The method of any one of claims 1-53, wherein the template nucleic acid molecules are clonally amplified template nucleic acid molecules.
55. The method of any one of claims 1-54, wherein the template nucleic acid molecules are generated through rolling circle amplification.
56. The method of any one of claims 1-55, wherein sequencing the plurality of template nucleic acid molecules generates a plurality of extended forward sequencing primer strands, and wherein the method comprises:
    a. retaining the plurality of template nucleic acid molecules and replacing the plurality of extended forward sequencing primer strands with a plurality of

forward extension strands that are hybridized to the plurality of nucleic acid template molecules by conducting a primer extension reaction;

b. removing the plurality of nucleic acid template molecules while retaining the plurality of forward extension strands and retaining the plurality of surface primers; and

c. sequencing the plurality of retained forward extension strands.

57. The method of claim 56, wherein the template nucleic acid molecules comprise:

i. two or more copies of the target sequence,

ii. two or more copies of the binding sequence for a forward sequencing primer, and

iii. two or more copies of a binding sequence for a reverse sequencing primer.

58. The method of claim 57, wherein the template nucleic acid molecules comprise binding sequences for an amplification primer, and wherein conducting the primer extension reaction comprises contacting the plurality of template nucleic acid molecules with a plurality of soluble amplification primers, a plurality of nucleotides and a plurality of polymerases, thereby generating a plurality of forward extension strands that are hybridized to template nucleic acid molecules.

59. The method of claim 58, wherein the plurality of amplification primers hybridize to the binding sequences for the amplification primers.

60. The method of claim 58 or 59, wherein the amplification primers are soluble.

61. The method of any one of claims 58-60, wherein the polymerases comprise phi29 DNA polymerases, large fragment of Bst DNA polymerases, large fragment of Bsu DNA polymerases (exo-), Bca DNA polymerases (exo-), Klenow fragment of E. coli DNA polymerases, T5 polymerases, M-MuLV reverse transcriptases, HIV viral reverse transcriptases, Deep Vent DNA polymerases or KOD DNA polymerases.

62. The method of any one of claims 1-61, wherein the nucleic acid template molecules comprise at least one nucleotide having a scissile moiety that can be cleaved to generate an abasic site.

63. The method of any one of claims 52-62, wherein the surface primers lack a nucleotide having a scissile moiety.

64. The method of claim 62 or claim 63, wherein the nucleotide having a scissile moiety comprises uridine, 8-oxo-7,8-dihydrogunine, or deoxyinosine.

65. The method of any one of claims 62-64, wherein removing the nucleic acid template molecules comprises generating abasic sites in the nucleic acid template molecules, followed by generating gaps at the abasic sites.

66. The method of claim 65, wherein the at least one nucleotide having a scissile moiety comprises uracil, and generating abasic sites comprises contacting the nucleic acid template molecules with uracil DNA glycosylase (UDG).

67. The method of claim 65 or 66, wherein generating gaps at the abasic sites comprises contacting the abasic sites with an endonuclease IV, AP lyase, FPG glycosylase/AP lyase and/or endo VIII glycosylase/AP lyase.

68. The method of any one of claims 1-67, wherein individual template nucleic acid molecules comprise nucleic acid template molecules having up to 30% of thymidines replaced with uridine.

69. The method of any one of claims 56-68, wherein sequencing the plurality of retained forward extension strands generates a plurality of extended reverse sequencing primer strands, wherein individual retained forward extension strands have two or more extended reverse sequencing primer strands hybridized thereon.

70. The method of any one of claims 56-69, wherein sequencing the plurality of retained forward extension strands comprises a plurality of soluble reverse sequencing primers and (i) a plurality of a first polymerases and a plurality of polymeric molecules and (ii) a plurality of a second polymerases and a plurality of nucleotides or analogs thereof, thereby generating a plurality of extended reverse sequencing primer strands, wherein individual retained forward extension strands have two or more extended reverse sequencing primer strands hybridized thereon.

71. The method of any one of claims 56-70, wherein the nucleic acid template molecules comprise one or more copies of a binding sequence for a second surface primer.

72. The method of claim 71, comprising a plurality of second surface primers immobilized on the support, whereby binding of the second surface primers to the binding sequence for the second surface primers immobilizes free ends of the plurality nucleic acid template molecules on the support.

73. The method of any one of claims 56-72, wherein sequencing the plurality of retained forward extension strands comprises

a. contacting:

(i) the plurality of retained forward extension strands,

(ii) a plurality of reverse sequencing primers comprising a sequence complementary to the binding sequence for the reverse sequencing primer,

(iii) a plurality of first polymerases, and

(iv) a plurality of polymeric molecules of Formula (I):

$$C\text{-}((P)\text{-}E)_s \quad (I)$$

an ionized form thereof, or a salt thereof, wherein:

C is a central moiety;

each P independently is an optionally substituted polymeric side chain;

each E independently is an end moiety; and

s is an integer ranging from 1 to 10;

wherein each polymeric molecule comprises at least two nucleotide moieties and at least one detectable reporter moiety,

wherein the contacting occurs under conditions sufficient to form a plurality of multivalent binding complexes comprising a nucleic acid duplex between a retained forward extension strand and a reverse sequencing primer, a first polymerase, and a nucleotide moiety of a polymeric molecule that is complementary to a nucleotide in the retained forward extension strand immediately adjacent to the 3' end of the reverse sequencing primer, and

wherein polymerase catalyzed incorporation of a complementary nucleotide moiety into the nucleic acid duplex is inhibited;

b. detecting the detectable reporter moieties; and

c. determining nucleobase identities of nucleotides in the retained forward extension strands complementary to the nucleotide moieties of the polymeric molecules based on the detectable reporter moieties of the polymeric molecules in the plurality of multivalent binding complexes formed in step (a).

74. The method of claim 73, wherein individual retained forward extension strands comprise two or more multivalent binding complexes.

75. The method of claim 73 or 74, wherein the plurality of reverse sequencing primers are soluble.

76. The method of any one of claims 73-75, comprising:

d. dissociating the multivalent binding complexes under conditions sufficient to retain the nucleic acid duplexes, thereby generating a plurality of nucleic acid duplexes;

e. contacting the plurality of nucleic acid duplexes with a plurality of second polymerases and a plurality of nucleotides or analogs thereof under conditions sufficient to incorporate nucleotides or analogs thereof complementary to the nucleotides of the retained forward extension strands immediately adjacent to the 3' ends of the reverse sequencing primers in a primer extension reaction, thereby generating a plurality of extended nucleic acid duplexes comprising extended reverse sequencing primer sequences.

77. The method of claim 76, comprising:

g. dissociating the second polymerases from the extended nucleic acid duplexes under conditions sufficient to retain the plurality of extended nucleic acid duplexes.

78. The method of any one of claims 73-77, wherein the template nucleic acid molecules comprise concatemers of two or more copies of a sequence comprising (i) a sequence for the reverse sequencing primer, (ii) the target nucleic acid sequence, and (iii) a binding sequence for the forward sequencing primer.

79. The method of claim 78, wherein the two or more copies of a sequence complementary to (i) the sequence for the reverse sequencing primer hybridize to the reverse sequencing primers to form nucleic acid duplexes between the retained forward extension strands and the reverse sequencing primers.

80. The method of any one of claims 76-79, wherein, in an individual polymeric molecule, the at least two nucleotide moieties are attached to different polymeric side chains.

81. The method of any one of claims 76-80, wherein, in an individual polymeric molecule, all nucleotide moieties are the same.

82. The method of claim 81, wherein all nucleotide moieties in an individual polymeric molecule are dATP.

83. The method of claim 81, wherein all nucleotide moieties in an individual polymeric molecule are dTTP.

84. The method of claim 81, wherein all nucleotide moieties in an individual polymeric molecule are dGTP.

85. The method of claim 81, wherein all nucleotide moieties in an individual polymeric molecule are dUTP.

86. The method of claim 81, wherein all nucleotide moieties in an individual polymeric molecule are dCTP.

87. The method of any one of claims 76-86, wherein, in an individual polymeric molecule, all detectable reporter moieties in the polymeric molecule comprise the same fluorescent label.
88. The method of claim 87, wherein all detectable reporter moieties in an individual polymeric molecule are the same.
89. The method of any one of claims 76-88, wherein, in an individual polymeric molecule, all detectable reporter moieties are the same and all nucleotide moieties are the same.
90. The method of any one of claims 76-89, wherein an individual polymeric molecule comprises two, three, or four nucleotide moieties.
91. The method of any one of claims 76-90, wherein an individual polymeric molecule comprises two, three, or four detectable reporter moieties.
92. The method of any one of claims 76-91, wherein two or more nucleotide moieties in an individual polymeric molecule are associated with two or more different multivalent binding complexes on the same template nucleic acid molecule.
93. The method of any one of claims 76-92, wherein two or more nucleotide moieties in an individual polymeric molecule are associated with two or more different multivalent binding complexes on the same retained forward extension strand.
94. The method of any one of claims 73-93, comprising

a. contacting the plurality of extended nucleic acid duplexes with a plurality of first polymerases and a plurality of polymeric molecules of Formula (I), or an ionized form thereof, an isomer thereof, or a salt thereof,

wherein the contacting occurs under conditions sufficient to form a plurality of multivalent binding complexes comprising an extended nucleic acid duplex, a first polymerase, and a nucleotide moiety of a polymeric molecule that is complementary to a nucleotide in the retained forward extension strand immediately adjacent to the 3' end of the extended reverse sequencing primer, and

wherein polymerase catalyzed incorporation of a complementary nucleotide moiety into the extended nucleic acid duplex is inhibited;

b. detecting the detectable reporter moieties;
c. determining nucleobase identities of nucleotides in the retained forward extension strands complementary to the nucleotide moieties of the polymeric molecules based on the detectable reporter moieties of the polymeric molecules in the plurality of multivalent binding complexes formed in step (a);

d. dissociating the multivalent binding complexes under conditions sufficient to retain the plurality extended nucleic acid duplexes; and

e. contacting the plurality of extended nucleic acid duplexes with a plurality of second polymerases and a plurality of nucleotides or analogs thereof under conditions sufficient to incorporate nucleotides or analogs thereof complementary to the nucleotides of the nucleic acid template sequences immediately adjacent to the 3' ends of the extended reverse sequencing primers in a primer extension reaction, thereby generating a plurality of extended nucleic acid duplexes comprising extended reverse sequencing primers.

95. The method of claim 94, wherein, in an individual polymeric molecule, the at least two nucleotide moieties are attached to different X moieties.

96. The method of claim 94 or 95, wherein, in an individual polymeric molecule, all nucleotide moieties polymeric molecule are the same.

97. The method of claim 96, wherein all nucleotide moieties in an individual polymeric molecule are dATP.

98. The method of claim 96, wherein all nucleotide moieties in an individual polymeric molecule are dTTP.

99. The method of claim 96, wherein all nucleotide moieties in an individual polymeric molecule are dGTP.

100. The method of claim 96, wherein all nucleotide moieties in an individual polymeric molecule are dUTP.

101. The method of claim 96, wherein all nucleotide moieties in an individual polymeric molecule are dCTP.

102. The method of any one of claims 94-101, wherein, in an individual polymeric molecule, all detectable reporter moieties are the same and all nucleotide moieties are the same.

103. The method of any one of claims 94-102, wherein an individual polymeric molecule comprises two, three, or four nucleotide moieties.

104. The method of any one of claims 94-103, wherein an individual polymeric molecule comprises two, three, or four detectable reporter moieties.

105. The method of any one of claims 94-104, wherein, in an individual polymeric molecule, all detectable reporter moieties comprise the same fluorescent label.

106. The method of claim 105, wherein all detectable reporter moieties in an individual polymeric molecule label are the same.

107. The method of any one of claims 94-106, wherein two or more nucleotide moieties in an individual polymeric molecule contact two or more different multivalent binding complexes on the same retained forward extension strand.

108. The method of any one of claims 94-107, comprising repeating steps (a)-(e) at least 1, 10, 20, 30, 40, 50, 70, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 800, 900, 1000, or 1500 times.

109. The method of any one of claims 94-108, comprising repeating steps (a)-(e) until the identities of the nucleotides in sequences of the retained forward extension strands complementary to the target sequences have been determined.

110. The method of any one of claims 94-109, comprising, before step (a), dissociating the second polymerases from the extended nucleic acid duplexes under conditions sufficient to retain the plurality of extended nucleic acid duplexes.

111. The method of any one of claims 1-110, wherein template nucleic acid molecules comprise concatemers of two or more copies of a sequence comprising:

(i) a binding sequence for a forward sequencing primer,

(ii) sequence complementary to a binding sequence for a reverse sequencing primer,

(iii) a binding sequence for an first surface primer,

(iv) a binding sequence for a second surface primer,

(v) a binding sequence for a first amplification primer,

(vi) a binding sequence for a second amplification primer,

(vii) a binding sequence for a soluble compaction oligonucleotide,

(viii) a sample barcode sequence, and/or

(ix) a unique molecular index sequence.

112. The method of any of claims 1-111, wherein, in an individual polymeric molecule, each polymeric side chain comprises one or more nucleotide moiety.

113. The method of any one of claims 1-112, wherein, in an individual polymeric molecule, each polymeric side chain comprises one or more detectable reporter moiety.

114. The method of any one of claims 1-113, wherein, in an individual polymeric molecule, each X moiety comprises one or more nucleotide moiety and one or more detectable reporter moiety.

115. The method of any one of claims 1-114, wherein each individual polymeric molecule comprises two, three, or four detectable reporter moieties and two, three, or four nucleotide moieties.

116. The method of any one of claims 1-115, wherein, in an individual polymeric molecule, each polymeric side chain comprises one or more blocking moiety.

117. The method of any one of claims 1-116, wherein, in an individual polymeric molecule, each polymeric side chain comprises one or more negative charge moieties.

118. The method of any one of claims 1-117, wherein, in an individual polymeric molecule, each polymeric side chain comprises one or more PEG-Cap moieties.

119. The method of any one of claims 1-118, wherein greater than 90%, greater than 95%, greater than 97%, greater than 98% or greater than 99% of bases have a quality score of Q30.

120. The method of any one of claims 1-119, wherein greater than 80%, greater than 85%, greater than 87%, greater than 89%, greater than 90%, greater than 91%, greater than 92%, greater than 93%, greater than 94% or greater than 95% of bases have a quality score of Q40.

121. A polymeric molecule of Formula (I):

$$C\text{-}((P)\text{-}E)_s \quad (I)$$

an ionized form thereof, or a salt thereof, wherein:

C is a central moiety;

each P independently is an optionally substituted polymeric side chain;

each E independently is an end moiety; and

s is an integer ranging from 1 to 10.

122. The polymeric molecule of claim 121, wherein the polymeric molecule is of Formula (II):

$$C\text{-}((P)\text{-}E)_4 \quad (II)$$

an ionized form thereof, or a salt thereof.

123. The polymeric molecule of claim 121 or 122, wherein the polymeric molecule comprises at least two nucleotide moieties and at least one detectable reporter moiety.

124. The polymeric molecule of claim 123, wherein the at least two nucleotide moieties are attached to different polymeric side chains.

125. The polymeric molecule of any one of claims 121-124, wherein all the nucleotide moieties are the same.

126. The polymeric molecule of any one of claims 121-125, wherein all the nucleotide moieties are dATP.

127. The polymeric molecule of any one of claims 121-125, wherein all the nucleotide moieties are dTTP.

128. The polymeric molecule of any one of claims 121-125, wherein all the nucleotide moieties are dGTP.

129. The polymeric molecule of any one of claims 121-125, wherein all the nucleotide moieties are dUTP.

130. The polymeric molecule of any one of claims 121-125, wherein all the nucleotide moieties are dCTP.

131. The polymeric molecule of any one of claims 121-130, wherein all detectable reporter moieties comprise the same fluorescent label.

132. The polymeric molecule of any one of claims 121-131, wherein all the detectable reporter moieties are the same.

133. The polymeric molecule of any one of claims 121-132, wherein all the detectable reporter moieties are the same and all the nucleotide moieties are the same.

134. The polymeric molecule of any one of claims 121-133, the molecule comprises two, three, or four detectable reporter moieties.

135. The polymeric molecule of any one of claims 121-134, the molecule comprises two, three, or four nucleotide moieties.

136. The polymeric molecule of any one of claims 121-135, wherein the polymeric molecule further comprises one or more blocking moiety.

137. The polymeric molecule of any one of claims 121-136, wherein the polymeric molecule further comprises one or more negative charge moiety.

138. The polymeric molecule of any one of claims 121-137, wherein the polymeric molecule further comprises one or more PEG-Cap moiety.

139. A complex comprising:

a. the polymeric molecule of any one of claims 122-138;

b. a polymerase;

c. a template nucleic acid molecule comprising at least one of a target sequence and a binding sequence for a sequencing primer; and

d. a sequence complementary to a portion of the template nucleic acid molecule comprising the sequencing primer sequence;

wherein the template nucleic acid molecule and the sequence complementary to a portion of the template nucleic acid molecule form a duplex, and wherein a nucleotide moiety of the polymeric molecule binds

to a complementary to a nucleotide of the template nucleic acid molecule immediately adjacent to the 3' end of the sequence complementary to a portion of the template nucleic acid molecule.

140. The complex of claim 139, wherein the polymeric molecule comprises at least two nucleotide moieties, and wherein the at least two nucleotide moieties are the same.

141. The complex of claim 139 or 140, wherein the template nucleic acid molecule comprises a concatemer comprising at least two copies of a sequence comprising the target sequence and the binding sequence for a sequencing primer.

142. The complex of claim 141, wherein at least two complexes form on the same template nucleic acid molecule.

143. The complex of claim 141 or 142, wherein the at least two nucleotide moieties bind to complementary nucleotides of the template nucleic acid molecule in the at least two complexes.

144. The complex of claim 142, wherein at least two complexes form on at least two different template nucleic acid molecules.

145. The complex of claim 144, wherein the at least two different template nucleic acid molecules comprise the same target sequence.

146. The complex of claim 145, wherein the at least two nucleotide moieties bind to complementary nucleotides of the template nucleic acid molecules in the at least two complexes.