Title: HEAVY ATOM LABELED NUCLEOSIDES, NUCLEOTIDES, AND NUCLEIC ACID POLYMERS, AND USES THEREOF

Abstract: The present disclosure provides compositions and methods to sequence nucleic acid polymers for improving sequencing read length using electron microscopy, e.g., high-resolution scanning transmission electron microscopy (STEM). The present disclosure further provides heavy-atom labeled compounds of Formula (I): nucleic acid polymers comprising one or more heavy-atom labeled units of Formula (II): such as heavy-atom labeled nucleic acid polymers of Formula (II); and salts thereof, wherein each of G^1, G^2, G^3, M^1, M^2, Base, and n are as defined herein, optionally for use in the methods described.

Published: — with international search report (Art. 21(3))
HEAVY ATOM LABELED NUCLEOSIDES, NUCLEOTIDES, AND NUCLEIC ACID POLYMERS, AND USES THEREOF

RELATED APPLICATIONS


BACKGROUND OF THE INVENTION

[0002] Advances over the last decade have greatly improved the speed and reduced the costs of DNA sequencing. Currently, they are limited to molecules less than 1,000 base pairs long, principally due to the inefficiency or incomplete nature of the fluorescent labeling reactions of "next generation" approaches (Schuster, 2008). Richard Feynman (1959) famously suggested that the incredible magnification power of electron microscopes might be harnessed to read DNA sequence; until now this challenge had not been met. The limiting issue has not been the small size of DNA, but the fact that the four different base types differ by only a few atoms, and all of the differing atoms are light elements, differences particularly indistinguishable for electron microscopy. Standard techniques used to increase sample contrast for electron microscopy have not been able to do so in a reliably sequence-specific manner, even after 40 years of effort (Gal-Or et al. 1967; ASTA, 2010).

[0003] The ADF-STEM was the method of choice for Crewe and co-workers to originally image single heavy atoms in anticipating that the method might be used for sequencing DNA. (Crewe, 1970; Crewe et al., 1970). Recent STEM improvements now allow studies of atomic-level and single atom imaging (Batson et al., 2002; Voyles et al., 2002; Jia et al., 2003). In an ADF-STEM, a very small electron beam is raster-scanned across the sample. Most of the electrons pass through the sample with only subtle changes of energy, direction, and/or phase. However, some electrons scatter at a high angle. The high angle scattering process (Rutherford scattering) scales with the atomic number (Z) of the atom (MuUer et al., 2008) raised to the power of approximately 1.5. The $Z^{1.5}$ dependence allows heavy nuclei to be definitively discriminated from light nuclei. The direct identification of unlabeled DNA base pairs, with average $Z \approx 5.5$, has proven to be difficult, and to-date unsuccessful. There is simply not enough difference between the base types to be detected without suitable contrast enhancement. Various groups have worked to overcome this problem, chiefly by chemically
modifying single-stranded DNA with clusters of heavy atoms (Beer & Moudrianakis, 1962; Moudrianakis & Beer, 1965; Ottensmeyer, 1979).

SUMMARY OF THE INVENTION

[0004] The present disclosure provides compositions and methods to sequence nucleic acid molecules including improving sequencing read length by directly visualizing DNA as long, intact molecules using electron microscopy, such as high-resolution scanning transmission electron microscopy (STEM). In some aspects, template-directed polymerase enzymes are used to incorporate heavy-atom labeled bases directly into a long DNA molecule. As shown herein, the incorporation of heavy-atom labeled bases provides annular dark-field imaging (ADF-STEM) contrast substantially greater than in unlabeled DNA. The methods disclosed also simplify the challenge of making the labeling reactions sequence-specific because polymerase reactions are intrinsically sequence specific.

[0005] The present disclosure further provides inventive heavy-atom labeled compounds optionally for use in the inventive methods as described herein. Exemplary heavy-atom labeled compounds include compounds of Formula (I):

![Formula I](image)

and nucleic acid polymers comprising one or more heavy-atom labeled units of Formula (II'):

![Formula II'](image)

such as heavy-atom labeled nucleic acid polymers of Formula (II):

![Formula II](image)

and salts thereof;

wherein:
each instance of $G_1$ is independently -0-, -S-, -Se-, -CH$_2$-, or -NH-;
each instance of $G_2$ is independently hydrogen, halogen, -OR$_A$, -SR$_A$, -N(R$_A$)$_2$, -SHg, -S0$_2$SHg, -SHgR$_D$, -SeR$_D$ or -TeR$_D$;
each instance of $R_A$ is independently hydrogen, substituted or unsubstituted C$_{1-20}$alkyl, substituted or unsubstituted C$_{2-20}$alkenyl, substituted or unsubstituted C$_{2-20}$alkynyl, substituted or unsubstituted carbocycyl, substituted or unsubstituted heterocyclyl, substituted or unsubstituted aryl, or substituted or unsubstituted heteroaryl, an oxygen protecting group when attached to an oxygen atom, a sulfur protecting group when attached to a sulfur atom, a nitrogen protecting group when attached to a nitrogen atom; or two $R_A$ groups are joined to form a substituted or unsubstituted heterocyclic or substituted or unsubstituted heteroaryl ring;
each instance of $M^1$ is independently -0-, -S-, -NH-, -Se-, or -C(R$_M$)$_2$-, wherein each instance of $R^M$ is independently hydrogen or halogen;
each instance of $G_3$ is independently hydrogen, substituted or unsubstituted C$_{\leq 20}$alkyl, substituted or unsubstituted C$_{2-20}$alkenyl, substituted or unsubstituted C$_{2-20}$alkynyl, substituted or unsubstituted carbocycyl, substituted or unsubstituted heterocyclyl, substituted or unsubstituted aryl, or substituted or unsubstituted heteroaryl, or a monophosphate, diphosphate, or triphosphate of formula:

$$\text{HO-P} \xrightarrow{\text{OH}} \text{HO-P-O-P} \xrightarrow{\text{OH}} \text{HO-P-O-P-O-P} \xrightarrow{\text{OH}} \text{M}^2-\text{H}$$

wherein each instance of $M^2$ is independently -0-, -S-, or -Se-; and
each instance of Base is independently:

- Adenine
- Guanine
- Cytosine
- Uracil
- Thymine

or an analog thereof selected from the group consisting of:

(i) \hspace{1cm} (ii)
wherein:

each instance of $R^1$, $R^2$, $R^4$, and $R^5$ is independently hydrogen, substituted or unsubstituted $C^\text{a}alkyl$, substituted or unsubstituted $C_{2-20}alkenyl$, substituted or unsubstituted $C_{2-20}alkynyl$, substituted or unsubstituted carbocycyl, substituted or unsubstituted heterocyclyl, substituted or unsubstituted aryl, or substituted or unsubstituted heteroaryl, a nitrogen protecting group, -OR $^B$, or -SR $^B$, wherein each instance of $R^B$ is independently hydrogen, substituted or unsubstituted $C^\text{a}alkyl$, substituted or unsubstituted $C_{2-20}alkenyl$, substituted or unsubstituted $C_{2-20}alkynyl$, substituted or unsubstituted carbocycyl, substituted or unsubstituted heterocyclyl, substituted or unsubstituted aryl, or substituted or unsubstituted heteroaryl, an oxygen protecting group when attached to an oxygen group, or a sulfur protecting group when attached to a sulfur group; or $R^1$ and $R^2$
and/or R^4 and R^5 are joined to form a substituted or unsubstituted heterocyclic or substituted or unsubstituted heteroaryl ring;

each instance of R^3 is independently substituted or unsubstituted C_{1-20}alkyl, substituted or unsubstituted C_{2-20}alkenyl, substituted or unsubstituted C_{2-20}alkynyl, substituted or unsubstituted carbocycyl, substituted or unsubstituted heterocyclyl, substituted or unsubstituted aryl, substituted or unsubstituted heteroaryl, halogen, -OR^ε, -SR^ε, -N(R^ε)^2, -SHg, -SO_{2}SHg, -SHgR^D, -SeR^D, or -TeR^D wherein each instance of R^C is hydrogen, substituted or unsubstituted C^oalkyl, substituted or unsubstituted C_{2-20}alkenyl, substituted or unsubstituted C_{2-20}alkynyl, substituted or unsubstituted heteroaryl, substituted or unsubstituted aryl, or substituted or unsubstituted heteroaryl, an oxygen protecting group when attached to an oxygen atom, a sulfur protecting group when attached to a sulfur atom, a nitrogen protecting group when attached to a nitrogen atom; or two R^C groups are joined to form a substituted or unsubstituted heterocyclic or substituted or unsubstituted heteroaryl ring;

each instance of L^1 is independently absent or a linking moiety selected from the group consisting of substituted or unsubstituted C^oalkylene, substituted or unsubstituted C_{2-20}alkenylene, substituted or unsubstituted C_{2-20}alkynylene, substituted or unsubstituted heteroC_{1-20}alkylene, substituted or unsubstituted heteroC_{2-20}alkenylene, substituted or unsubstituted heteroC_{2-20}alkynylene, substituted or unsubstituted carbocycylene, substituted or unsubstituted heterocyclylylene, substituted or unsubstituted arylene, or substituted or unsubstituted heteroarylene, or a combination thereof;

each instance of R^D is independently hydrogen, substituted or unsubstituted C^oalkyl, substituted or unsubstituted C_{2-20}alkenyl, substituted or unsubstituted C_{2-20}alkynyl, substituted or unsubstituted carbocycyl, substituted or unsubstituted heterocyclyl, substituted or unsubstituted aryl, or substituted or unsubstituted heteroaryl; and

each instance of M^1 and M^4 are independently O, Se, Te, CH_2, CF_2, CCl_2, CBr_2, or Cl_2; and

n is 1 to 200,000, inclusive;

provided that the compound comprises at least one instance of a heavy atom selected from the group consisting of bromine, iodine, selenium, tellurium, or mercury.

[0006] In some aspects, the invention provides compositions of heavy atom labeled nucleic acids, as well as systems and methods of identifying, sequencing and/or detecting nucleic acid polymers, as well as related components (e.g., substrates, software and the like).
According to one aspect of the invention, methods of determining the sequence of a nucleic acid polymer labeled with heavy atoms are provided. The methods include forming a complementary strand of the nucleic acid polymer comprising one or more heavy-atom labeled compounds as described herein and identifying a sequence of nucleotides in the nucleic acid polymer and/or in the complementary strand using a particle beam.

In certain embodiments, the nucleic acid polymer and/or the complementary strand is DNA or RNA. In other embodiments, the nucleic acid polymer and/or its complementary strand is formed by a nucleic acid polymerase enzyme, such as using polymerase chain reaction (PCR).

In preferred embodiments, the nucleotides of the nucleic acid polymer and/or the complementary strand are modified to include labels comprising one or more heavy-atom labeled compounds as described herein. Preferably the labels are specific for each type of nucleotide. However, in some embodiments, at least two types of nucleotides are labeled with the same type of heavy-atom label. In other embodiments, one type of nucleotide is labeled, two types of nucleotides are labeled, three types of nucleotides are labeled, or four types of nucleotides are labeled. In some embodiments, substantially all of the nucleotides of the nucleic acid polymer and/or complementary strand are modified such that all nucleotides are labeled.

Preferably nucleotide specific labels are incorporated in the nucleic acid polymer and/or the complementary strand during formation of the nucleic acid polymer and/or the complementary strand.

In further embodiments, the nucleic acid polymer and/or the complementary strand are affixed to a substrate, and prior to the step of identification the nucleotides of the nucleic acid polymer and/or its complementary strand are substantially removed from the substrate, leaving the labels of the labeled nucleotides affixed to the substrate.

In still other embodiments, the step of identifying a sequence of nucleotides includes generating a particle beam, exposing the nucleic acid polymer and/or the complementary strand to the particle beam, and identifying the nucleotides due to characteristic changes to the particle beam. Preferably the nucleotides of the nucleic acid polymer and/or the complementary strand are modified to include labels as described herein, and more preferably the step of identifying the nucleotides includes detecting characteristic changes to the particle beam due to the heavy atom label(s) on the nucleotides. In certain embodiments, the particle beam is a lepton beam; more preferably the lepton beam is an
electron beam. In some embodiments the particle beam is an ion beam; more preferably a helium ion beam or a gallium ion beam.

[0013] In other embodiments the nucleic acid polymer and/or the complementary strand are affixed to a substrate. The nucleic acid polymer and/or the complementary strand can be affixed to a substrate at one end of the nucleic acid polymer and/or the complementary strand, at both ends of the nucleic acid polymer and/or the complementary strand, and/or at a plurality of locations along the length of the nucleic acid polymer and/or the complementary strand.

[0014] In certain embodiments, the nucleic acid polymer and/or the complementary strand are substantially straightened prior to identifying the sequence. Preferably the nucleic acid polymer and/or the complementary strand are straightened by fluid flow, and more preferably the fluid flow includes molecular combing. The fluid can include one or more liquids, gases, phases or a combination thereof. In some embodiments, the nucleic acid polymer and/or the complementary strand are attached to a substrate and straightened by hybridization in the fluid flow to oligonucleotides that are attached to the substrate.

[0015] In additional embodiments, the step of identifying the nucleotides in the nucleic acid polymer and/or its complementary strand includes interpreting changes in the particle beam resulting from interactions with the nucleotides to detect the nucleotides in the nucleic acid polymer and/or its complementary strand, whereby the sequence of the nucleic acid polymer is determined. Preferably the nucleotides are labeled as described herein. The changes in the particle beam include changes in absorbance, reflection, deflection, energy or direction. The changes in the particle beam also can be changes in a spatial pattern, for example, a one dimensional pattern, a two dimensional pattern or a three dimensional pattern.

[0016] In further embodiments, the method also includes attaching the complementary strand and/or the nucleic acid polymer to a substrate. Preferably the attachment is by nucleic acid sequence-specific molecules, which preferably are oligonucleotides. In other preferred the substrate is derivatized to provide attachment points that are sequence non-specific. The complementary strand and optionally the nucleic acid polymer can be attached to the substrate in a grid pattern. Preferably the substrate includes a carbon thin film.

[0017] In other embodiments, the step of identifying the sequence of nucleotides includes performing a plurality of scans of the nucleic acid polymer and/or the complementary strand using the particle beam. Preferably at least 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1200, 1400, 1600, 1800, 2000, 3000, 4000, 5000, 6000, 7000, 8000, 9000, 10000, or more nucleotides are identified in each scan.
According to another aspect of the invention, methods of determining the sequence of a nucleic acid polymer are provided. The methods include synthesizing the nucleic acid polymer and/or its complementary strand using labeled ribonucleotide and/or deoxyribonucleotide triphosphates as described herein, and identifying labeled ribonucleotides and/or deoxyribonucleotides in the nucleic acid polymer and/or its complementary strand using a particle beam, wherein the labeled ribonucleotides and/or deoxyribonucleotides, when incorporated in the nucleic acid polymer and/or its complementary strand, are identifiable using the particle beam.

In certain embodiments, the nucleic acid polymer and/or the complementary strand is DNA or RNA. In other embodiments, the nucleic acid polymer and/or its complementary strand is synthesized by a nucleic acid polymerase enzyme, such as using polymerase chain reaction (PCR).

In preferred embodiments, the labels are specific for each type of nucleotide. However, in some embodiments, at least two types of nucleotides are labeled with the same type of heavy-atom label. In other embodiments, one type of nucleotide is labeled, two types of nucleotides are labeled, three types of nucleotides are labeled, or four types of nucleotides are labeled. In some embodiments, substantially all of the nucleotides of the nucleic acid polymer and/or complementary strand are modified such that all nucleotides are labeled.

Preferably, the labels are incorporated in the ribonucleotide and/or deoxyribonucleotide triphosphates used in synthesis of the nucleic acid polymer and/or the complementary strand.

In further embodiments, the step of identifying the labeled ribonucleotides and/or deoxyribonucleotides includes generating a particle beam, exposing the nucleic acid polymer and the complementary strand to the particle beam, and identifying the ribonucleotides and/or deoxyribonucleotides due to characteristic changes to the particle beam due to the heavy atom label(s) on the nucleotides. Preferably the step of detecting the ribonucleotides and/or deoxyribonucleotides includes detecting characteristic changes to the particle beam. In certain embodiments, the particle beam is a lepton beam; more preferably the lepton beam is an electron beam. In some embodiments the particle beam is an ion beam; more preferably a helium ion beam or a gallium ion beam.

In other embodiments the nucleic acid polymer and/or the complementary strand are affixed to a substrate. In certain embodiments, prior to the step of identification the ribonucleotides and/or deoxyribonucleotides of the nucleic acid polymer and/or its complementary strand are substantially removed from the substrate, leaving the labels of the
labeled ribonucleotides and/or deoxyribonucleotides affixed to the substrate. The nucleic acid polymer and/or the complementary strand can be affixed to a substrate at one end of the nucleic acid polymer and/or the complementary strand, at both ends of the nucleic acid polymer and/or the complementary strand, and/or at a plurality of locations along the length of the nucleic acid polymer and/or the complementary strand.

[0024] In certain embodiments, the nucleic acid polymer and/or the complementary strand are substantially straightened prior to identifying the labeled ribonucleotides and/or deoxyribonucleotides. Preferably the nucleic acid polymer and/or the complementary strand are straightened by fluid flow, and more preferably the fluid flow includes molecular combing. The fluid can include one or more liquids, gases, phases or a combination thereof. In some embodiments, the nucleic acid polymer and/or the complementary strand are attached to a substrate and straightened by hybridization in the fluid flow to oligonucleotides that are attached to the substrate.

[0025] In additional embodiments, the step of identifying the nucleotides in the nucleic acid polymer and/or its complementary strand includes interpreting changes in the particle beam resulting from interactions with the nucleotides to detect the ribonucleotides and/or deoxyribonucleotides in the nucleic acid polymer and/or its complementary strand, whereby the sequence of the nucleic acid polymer is determined. Preferably the nucleotides are labeled as described herein. The changes in the particle beam include changes in absorbance, reflection, deflection, energy or direction. The changes in the particle beam also can be changes in a spatial pattern, for example, a one dimensional pattern, a two dimensional pattern or a three dimensional pattern.

[0026] In further embodiments, the method also includes attaching the complementary strand and/or the nucleic acid polymer to a substrate. Preferably the attachment is by nucleic acid sequence-specific molecules, which preferably are oligonucleotides. In other preferred the substrate is derivatized to provide attachment points that are sequence non-specific. The complementary strand and optionally the nucleic acid polymer can be attached to the substrate in a grid pattern. Preferably the substrate includes a carbon thin film.

[0027] In other embodiments, the step of identifying the sequence of nucleotides includes performing a plurality of scans of the nucleic acid polymer and/or the complementary strand using the particle beam. Preferably at least 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1200, 1400, 1600, 1800, 2000, 3000, 4000, 5000, 6000, 7000, 8000, 9000, 10000, or more nucleotides are identified in each scan.
According to another aspect of the invention, methods of determining the sequence of a nucleic acid polymer are provided. The methods include synthesizing a complementary strand of the nucleic acid polymer using labeled ribonucleotide triphosphates or deoxyribonucleotide triphosphates as described herein, attaching the nucleic acid polymer and/or the complementary strand to a substrate, substantially straightening the nucleic acid polymer and/or the complementary strand using molecular combing, generating a particle beam, exposing the nucleic acid polymer and the complementary strand to the particle beam through the complementary strand on the substrate, and interpreting changes in the particle beam resulting from interactions with the nucleotides to detect the labeled nucleotides in the complementary strand, whereby the sequence of a nucleic acid polymer is determined.

According to another aspect of the invention, methods of detecting the presence and/or identifying a nucleic acid polymer are provided. The methods include forming a complementary strand of the nucleic acid polymer, attaching the complementary strand and, optionally, the nucleic acid polymer to a substrate, and detecting the presence and/or identifying the complementary strand and/or the nucleic acid polymer using a particle beam.

In some embodiments, the step of identifying includes measuring the length or determining at least a partial sequence of the complementary strand and/or the nucleic acid polymer.

In certain embodiments, the nucleic acid polymer and/or its complementary strand is DNA or RNA. In other embodiments, the nucleic acid polymer and/or its complementary strand is formed by a nucleic acid polymerase enzyme, e.g., using polymerase chain reaction (PCR); preferably the nucleic acid polymerase enzyme is a DNA-dependent DNA polymerase, a RNA-dependent DNA polymerase or a RNA-dependent RNA polymerase.

In other embodiments, the nucleotides of the nucleic acid polymer and/or the complementary strand are modified to include labels as described herein. In preferred embodiments, the labels are specific for each type of nucleotide. However, in some embodiments, at least two types of nucleotides are labeled with the same type of heavy-atom label. In other embodiments, one type of nucleotide is labeled, two types of nucleotides are labeled, three types of nucleotides are labeled, or four types of nucleotides are labeled. In some embodiments, substantially all of the nucleotides of the nucleic acid polymer and/or complementary strand are modified such that all nucleotides are labeled.
[0033] Preferably nucleotide specific labels are incorporated in the nucleic acid polymer and/or the complementary strand during formation of the nucleic acid polymer and/or the complementary strand.

[0034] In further embodiments, the step of detecting the presence and/or identifying of the complementary strand and/or the nucleic acid polymer using a particle beam includes generating a particle beam, exposing the nucleic acid polymer and/or the complementary strand to the particle beam, and detecting the nucleotides of the complementary strand and/or the nucleic acid polymer due to characteristic changes to the particle beam.

[0035] In some embodiments, the nucleotides of the nucleic acid polymer and/or the complementary strand are modified to include labels as described herein. Preferably the step of detecting the ribonucleotides and/or deoxyribonucleotides includes detecting characteristic changes to the particle beam due to the heavy atom label(s) on the nucleotides. In certain embodiments, the particle beam is a lepton beam; more preferably the lepton beam is an electron beam. In some embodiments the particle beam is an ion beam; more preferably a helium ion beam or a gallium ion beam.

[0036] In certain embodiments, the nucleic acid polymer and/or the complementary strand are substantially straightened prior to identifying the sequence. Preferably the nucleic acid polymer and/or the complementary strand are straightened by fluid flow, and more preferably the fluid flow includes molecular combing. The fluid can include one or more liquids, gases, phases or a combination thereof. In some embodiments, the nucleic acid polymer and/or the complementary strand are attached to a substrate and straightened by hybridization in the fluid flow to oligonucleotides that are attached to the substrate.

[0037] In additional embodiments, the step of identifying the nucleotides in the nucleic acid polymer and/or its complementary strand includes interpreting changes in the particle beam resulting from interactions with the nucleotides to detect the nucleotides in the nucleic acid polymer and/or its complementary strand, whereby the presence of the nucleic acid polymer is determined and/or the nucleic acid polymer is identified. Preferably the nucleotides are labeled as described herein and the characteristic changes to the particle beam due to the heavy atom label(s) on the nucleotides. The changes in the particle beam include changes in absorbance, reflection, deflection, energy or direction. The changes in the particle beam also can be changes in a spatial pattern, for example, a one dimensional pattern, a two dimensional pattern or a three dimensional pattern.

[0038] In further embodiments, the method also includes attaching the complementary strand and/or the nucleic acid polymer to a substrate. Preferably the attachment is by nucleic
acid sequence-specific molecules, which preferably are oligonucleotides. In other preferred
the substrate is derivatized to provide attachment points that are sequence non-specific. The
complementary strand and optionally the nucleic acid polymer can be attached to the
substrate in a grid pattern. Preferably the substrate includes a carbon thin film.

In other embodiments, the method also includes quantifying the amount of the
complementary strand and/or the nucleic acid polymer.

According to another aspect of the invention, a device is provided that includes a
substrate that is substantially transparent to a particle beam, and nucleic acid polymer binding
sites on a surface of the substrate.

In some embodiments the substrate is substantially transparent to an electron
beam. Preferably the substrate includes a carbon thin film.

In other embodiments, the device also includes a support that is substantially
transparent to a particle beam.

Preferably the substrate is less than 5 nm thick, more preferably less than 2 nm
thick, still more preferably less than 1.5 nm thick, and yet more preferably less than 1.1 nm
thick.

In other embodiments, the nucleic acid polymer binding sites are formed at
predetermined positions on the surface of the substrate, preferably in a grid pattern. In
certain embodiments, the nucleic acid polymer binding sites are sequence specific, preferably
oligonucleotides. In other embodiments, the nucleic acid polymer binding sites are not
sequence specific.

In further embodiments, the device also includes one or more nucleic acid
polymers affixed to the nucleic acid polymer binding sites. Preferably the one or more
nucleic acid polymers are modified to include labels.

According to another aspect of the invention, methods for making a device are
provided. The methods include obtaining a substrate that is substantially transparent to a
particle beam, and forming nucleic acid polymer binding sites on a surface of the substrate.

In some embodiments the substrate is substantially transparent to an electron
beam. Preferably the substrate includes a carbon thin film. In some embodiments, the
nucleic acid polymer binding sites are formed at predetermined positions on the surface of the
substrate, preferably in a grid pattern.

In other embodiments, the method also includes attaching to the substrate a
support that is substantially transparent to a particle beam.
Preferably the substrate is less than 5 nm thick, more preferably less than 2 nm thick, still more preferably less than 1.5 nm thick, and yet more preferably less than 1.1 nm thick.

In certain embodiments, the nucleic acid polymer binding sites are sequence specific, preferably oligonucleotides. In other embodiments, the nucleic acid polymer binding sites are not sequence specific.

In still other embodiments, the methods also include affixing one or more nucleic acid polymers to the nucleic acid polymer binding sites. Preferably, the one or more nucleic acid polymers are modified to include labels.

According to another aspect of the invention, systems designed to detect the presence of, determine the sequence of and/or identify a nucleic acid polymer are provided. The systems include: a sample chamber; a particle beam generator associated with the chamber; a sample comprising a labeled complementary strand of a nucleic acid polymer, wherein the sample, when positioned in the chamber, is exposed to a particle beam generated by the particle beam generator resulting in an interaction between the particle beam and the complementary strand; and a detector constructed and arranged to collect particle beam species after the interaction.

In some embodiments, the system also includes a data analysis module operative to receive and analyze signals from the detector. Preferably the data analysis module is operative to analyze signals related to absorbance, reflection, deflection, energy or direction. In other embodiments, the data analysis module is operative to analyze pattern recognition techniques to analyze the signals.

In further embodiments, the system also includes a user interface operative to control a display of information received and/or generated by the data analysis module.

In preferred embodiments, the particle beam generator is an electron beam generator.

The system in other embodiments also includes a feedback module designed to calibrate the system based on nucleic acid polymer data.

According to another aspect of the invention, systems designed to detect the presence of, determine the sequence of and/or identify a nucleic acid polymer are provided. The systems include: a sample chamber; a particle beam generator associated with the chamber; a detector constructed and arranged to collect particle beam species after interaction between the particle beam and a sample comprising the nucleic acid polymer and/or a complementary strand of the nucleic acid polymer; a data analysis module designed to
analyze signals related to the particle beam species to determine information related to the nucleic acid polymer; and a feedback module designed to calibrate the system based on the information.

[0058] In some embodiments, the sample includes a labeled complementary strand of a nucleic acid polymer.

[0059] In certain embodiments, the feedback module is designed to calibrate the system based on a base-base distance of the nucleic acid polymer. In other embodiments, the feedback module is designed to calibrate the system based on known geometries of the nucleic acid polymer.

[0060] Also provided in accordance with another aspect of the invention are methods for calibrating a particle beam instrument. The methods include acquiring data related to a nucleic acid polymer; and calibrating the instrument based on the data. Preferably the data is related to a base-base distance of the nucleic acid polymer. In some embodiments, the calibrating includes calibrating the instrument based on known geometries of the nucleic acid polymer.

[0061] According to another aspect of the invention, systems are provided for detecting, sequencing and/or identifying a nucleic acid polymer based on particle beam species detected by a detector, the particle beam species resulting from exposure of a sample comprising a nucleic acid polymer and/or its complementary strand to a particle beam. The systems include a data analysis module operative to receive one or more signals from the detector, the one or more signals representing the particle beam species, and to detect, sequence and/or identify the nucleic acid polymer and/or its complementary strand comprised in the sample based at least in part on the received one or more signals. Preferably the nucleic acid polymer and/or its complementary strand is labeled.

[0062] In some embodiments, the particle beam species has one or more of the following properties: absorbance, reflection, deflection, energy and direction, and the data analysis module is operative to analyze the one or more signals to determine values of the one or more properties.

[0063] In other embodiments, the data analysis module is operative to access a data resource comprising nucleic acid polymer information, the data resource including a data structure having a plurality of entries, each entry specifying information about a respective nucleic acid polymer sequence. Preferably the data analysis module is operative to partially sequence the nucleic acid polymer based on the one or more signals, the data analysis module further comprising: a combining module to combine the partial sequence with sequencing
information of the nucleic acid polymer accessed from the data resource. In preferred
embodiments the data analysis module includes a comparison module operative to compare
information determined from the one or more signals to the information specified by one or
more of the data structure entries. Preferably the comparison module is operative to use
pattern recognition techniques to compare the information determined from the one or more
signals to the information specified by the one or more the data structure entries.

[0064] In other embodiments the data analysis module includes a user interface module
to display information received and/or generated by the data analysis module to a user.

[0065] In further embodiments the particle beam to which the sample is exposed is
generated by a particle beam generator, and the data analysis module includes a feedback
module operative to provide one or more feedback signals to the particle beam generator
and/or the detector, the one or more feedback signals specifying information determined at
least in part from the one or more signals received from the detector. Preferably the one or
more feedback signals include information for calibrating the particle beam generator. In
preferred embodiments the feedback module is operative to generate the one or more
feedback signals based at least in part on known geometries of the nucleic acid polymer. The
data analysis module preferably includes a storage module operative to store information
received and/or generated by the data analysis module on a computer-readable medium.

[0066] In some embodiments the sample includes a plurality of molecules of a same
nucleic acid polymer and/or its complementary strand, and a plurality of particle beam
species results from exposure of the plurality of molecules of the sample to the particle beam,
the one or more signals representing the plurality of particle beam species, wherein the data
analysis module is operative to partially sequence the nucleic acid polymer based on a first of
the plurality of molecules to produce a first partial sequence, and to partially sequence the
nucleic acid polymer based on a second of the plurality of molecules to produce a second
partial sequence, and wherein the data processing module further includes a combining
module to combine the first and second partial sequences.

[0067] According to another aspect of the invention, a computer-readable medium is
provided having computer-readable signals stored thereon that define instructions that, as a
result of being executed by a computer, control the computer to perform a process of
detecting, sequencing and/or identifying a nucleic acid polymer based on particle beam
species detected by a detector, the particle beam species resulting from exposure of a sample
comprising a nucleic acid polymer and/or its complementary strand to a particle beam. The
process includes: receiving one or more signals from the detector, the one or more signals
representing the particle beam species; and detecting, sequencing and/or identifying the nucleic acid polymer and/or its complementary strand comprised in the sample based at least in part on the received one or more signals. Preferably the nucleic acid polymer and/or its complementary strand is labeled.

[0068] In some embodiments, the particle beam species has one or more of the following properties: absorbance, reflection, deflection, energy and direction, and the act of detecting, sequencing and/or identifying includes analyzing the one or more signals to determine values of the one or more properties.

[0069] In other embodiments, the act of detecting, sequencing and/or identifying includes accessing a data resource comprising nucleic acid polymer information, the data resource including a data structure having a plurality of entries, each entry specifying information about a respective nucleic acid polymer sequence. Preferably the act of detecting, sequencing and/or identifying includes partially sequencing the nucleic acid polymer based on the one or more signals to produce a partial sequence; accessing partial sequence information of the nucleic acid polymer from the data resource; and combining the partial sequence with the partial sequence information. In preferred embodiments the act of detecting, sequencing and/or identifying includes comparing information determined from the one or more signals to the information specified by one or more of the entries. In some of these embodiments, the act of detecting, sequencing and/or identifying preferably includes using pattern recognition techniques to compare the information determined from the one or more signals to the information specified by the one or more entries.

[0070] In further embodiments, the process further includes displaying information determined from the one or more received signals to a user.

[0071] In other embodiments the particle beam to which the sample is exposed is generated by a particle beam generator, and the process further includes providing one or more feedback signals to the particle beam generator and/or the detector, the one or more feedback signals specifying information determined at least in part from the one or more signals received from the detector. Preferably the act of providing includes providing one or more feedback signals that include information for calibrating the particle beam generator. In some embodiments the process further includes generating the one or more feedback signals based at least in part on known geometries of the nucleic acid polymer.

[0072] In other embodiments the process further includes storing information determine from the one or more signals on a computer-readable medium.
In further embodiments the sample includes a plurality of molecules of a same nucleic acid polymer and/or its complementary strand, and a plurality of particle beam species result from exposure of the plurality of molecules of the sample to the particle beam, the one or more signals representing the plurality of particle beam species, and the act of detecting, sequencing and/or identifying includes partially sequencing the nucleic acid polymer based on a first of the plurality of molecules to produce a first partial sequence; partially sequencing the nucleic acid polymer based on a second of the plurality of molecules to produce a second partial sequence; combining the first and second partial sequences.

The details of one or more embodiments of the invention are set forth in the accompanying Figures and the Detailed Description. Other features, objects, and advantages of the invention will be apparent from the description and from the claims.

**BRIEF DESCRIPTION OF THE DRAWINGS**

*Figure 1* depicts a non-limiting example of heavy atoms labels detected within DNA molecules. (A): Schematic showing heavy atoms deflecting portion of the raster scanned electron beam. Highly deflected electrons are detected on the ADF detector. (B): Unlabeled DNA bases scatter fewer electrons than the heavy-atom-labeled bases, distinguished by detector current.

*Figure 2* depicts a non-limiting example of heavy-atom-labeling strategy. A single stranded template is primed with a complementary oligonucleotide primer. For simplicity, the lengths of the primer and the template have been shortened. In the presence of polymerase, the template directs the synthesis of a complementary strand. Thymine deoxyribose nucleotide triphosphates in the primer extension reaction have been completely replaced with a heavy-atom-modified analog. Consequently, the resulting double-stranded DNA molecule is modified with heavy atoms on the thymine bases of the synthetic strand. These heavy atoms provide signal to the dark-field detector of a STEM system.

*Figure 3* depicts DNA alignment of a prepared and linearized DNA molecule on a thin amorphous carbon substrate. (A) Bright-field TEM image of multiple DNA molecules linearized on amorphous carbon surface. (B) Darkfield STEM image of linearized DNA molecule on thin amorphous carbon substrate.

*Figure 4* depicts heavy-atom locations and contrast distribution. (A): ADF detector readings from labeled M13 viral DNA (NEB product #N4040S) showing thymine discrimination with red-colored band, 1.66 to 5 standard deviations (95% to 99.99+% confidence) from unlabeled DNA regions. (B): Corresponding residual current histogram of
labeled DNA molecule scan. Data acquired in Carl Zeiss Libra 200-80kV aberration-corrected EF-STEM with Cs = - 1.2 μm, 80 kV with elastic scattering using in-column energy-filter retaining only zero energy-loss electrons. DNA labeled with 5-Me-Hg-S-dUTP replacing dTTP in primer extension labeling reaction.

**Figure 5** depicts a schematic of repeating "test pattern" molecule. Heavy atoms are attached to thymine/uridine bases of one strand of double-stranded DNA molecules. The labels nearest one another are separated by one unlabeled base pair; the theoretical pitch between the heavy atoms is 0.7 to 1.2 nm. These doublets repeat every 12 base pairs, for a theoretical pitch of 4.1 to 7.3 nm. Actual spacing of both patterns depends on local stretching, predicted to be 0% to 80%.

**Figure 6** depicts sequence data from repeating "test pattern" molecule. (A): Partial sequence of DNA molecule. Yellow lines (starred, *) show heavy atoms in predicted large-scale test pattern positions, where distances to neighbors in both directions match the large-scale test pattern. White circles show pairs of atoms matching small-scale pattern. Red lines (indicated by arrows) show atoms of the large-scale pattern in positions predicted by spacing with one rather than two neighbors. (B): Schematic of repeating test pattern shows both small-scale and large-scale patterns. Not to scale.

**DEFINITIONS**


**Compounds** described herein can comprise one or more asymmetric centers, and thus can exist in various stereoisomeric forms, e.g., enantiomers and/or diastereomers. For example, the compounds described herein can be in the form of an individual enantiomer, diastereomer or geometric isomer, or can be in the form of a mixture of stereoisomers, including racemic mixtures and mixtures enriched in one or more stereoisomer. Isomers can
be isolated from mixtures by methods known to those skilled in the art, including chiral high pressure liquid chromatography (HPLC) and the formation and crystallization of chiral salts; or preferred isomers can be prepared by asymmetric syntheses. See, for example, Jacques et al., Enantiomers, Racemates and Resolutions (Wiley Interscience, New York, 1981); Wilen et al., Tetrahedron 33:2725 (1977); Eliel, E.L. Stereochemistry of Carbon Compounds (McGraw-Hill, NY, 1962); and Wilen, S.H. Tables of Resolving Agents and Optical Resolutions p. 268 (E.L. Eliel, Ed., Univ. of Notre Dame Press, Notre Dame, IN 1972). The invention additionally encompasses compounds as individual isomers substantially free of other isomers, and alternatively, as mixtures of various isomers.

[0083] When a range of values is listed, it is intended to encompass each value and sub-range within the range. For example "Ci-6 alkyl" is intended to encompass, C1, C2, C3, C4, C5, C6, Ci-6, Ci-5, Ci^, C1-3, Ci-2, C2-6, C2-5, C2-4, C2-3, C3-6, C3-5, C3-4, C4-6, C4-5, and C5-6 alkyl.

[0084] As used herein, "alkyl" refers to a radical of a straight-chain or branched saturated hydrocarbon group having from 1 to 20 carbon atoms ("C^0 alkyl"). In some embodiments, an alkyl group has 1 to 10 carbon atoms ("Ci-w alkyl"). In some embodiments, an alkyl group has 1 to 9 carbon atoms ("C1-9 alkyl"). In some embodiments, an alkyl group has 1 to 8 carbon atoms ("C1-8 alkyl"). In some embodiments, an alkyl group has 1 to 7 carbon atoms ("Ci-7 alkyl"). In some embodiments, an alkyl group has 1 to 6 carbon atoms ("Ci^ alkyl"). In some embodiments, an alkyl group has 1 to 5 carbon atoms ("Ci-5 alkyl"). In some embodiments, an alkyl group has 1 to 4 carbon atoms ("C^ alkyl"). In some embodiments, an alkyl group has 1 to 3 carbon atoms ("Ci-3 alkyl"). In some embodiments, an alkyl group has 1 to 2 carbon atoms ("Ci 2 alkyl"). In some embodiments, an alkyl group has 1 carbon atom ("Ci alkyl"). In some embodiments, an alkyl group has 2 to 6 carbon atoms ("Ci 1-6 alkyl"). Examples of Ci-6 alkyl groups include methyl (CO, ethyl (C2), n-propyl (C3), isopropyl (C3), n-butyl (C4), tert-butyl (C4), sec-butyl (C4), iso-butyl (C4), n-pentyl (C5), 3-pentanyl (C5), amy1 (C5), neopentyl (C5), 3-methyl-2-butanyl (C5), tertiary amyl (C5), and n-hexyl (C6). Additional examples of alkyl groups include n-heptyl (C7), n-octyl (C8) and the like. Unless otherwise specified, each instance of an alkyl group is independently unsubstituted (an "unsubstituted alkyl") or substituted (a "substituted alkyl") with one or more substituents. In certain embodiments, the alkyl group is an unsubstituted Ci-2o alkyl (e.g., -CH3). In certain embodiments, the alkyl group is a substituted Ci-2o alkyl.

[0085] As used herein, "haloalkyl" is an alkyl group as defined herein wherein one or more of the hydrogen atoms are independently replaced by a halogen, e.g., fluoro, bremo,
chloro, or iodo. "Perhaloalkyl" is a subset of haloalkyl, and refers to an alkyl group wherein all of the hydrogen atoms are independently replaced by a halogen, e.g., fluoro, bromo, chloro, or iodo. In some embodiments, the haloalkyl moiety has 1 to 20 carbon atoms ("Ca_o haloalkyl"). In some embodiments, the haloalkyl moiety has 1 to 10 carbon atoms ("Cl_w haloalkyl"). In some embodiments, the haloalkyl moiety has 1 to 8 carbon atoms ("Ci_1-8 haloalkyl"). In some embodiments, the haloalkyl moiety has 1 to 6 carbon atoms ("Ci_6 haloalkyl"). In some embodiments, the haloalkyl moiety has 1 to 4 carbon atoms ("Ca haloalkyl"). In some embodiments, the haloalkyl moiety has 1 to 3 carbon atoms ("Ca haloalkyl"). In some embodiments, the haloalkyl moiety has 1 to 2 carbon atoms ("Ci_2 haloalkyl"). In some embodiments, all of the haloalkyl hydrogen atoms are replaced with fluoro to provide a perfluoroalkyl group. In some embodiments, all of the haloalkyl hydrogen atoms are replaced with chloro to provide a "perchloroalkyl" group. Examples of haloalkyl groups include -CF_3, -CF_2CF_3, -CF_2CF_2CF_3, -CCl_3, -CFC_1_2, - CF_2Cl, and the like.

Haloalkenyl, haloalkynyl, halocarbocyclyl, haloheterocyclyl, haloaryl, and haloheteroaeryl follow the definition of haloalkyl, and refer to an alkenyl, alkynyl, carbocyclyl, heterocyclyl, aryl, and heteroaryl group, as defined herein, wherein one or more of the hydrogen atoms are independently replaced by a halogen, e.g., fluoro, bromo, chloro, or iodo. Likewise, perhaloalkenyl, perhaloalkynyl, perhalocarbocyclyl, perhaloheterocyclyl, perhaloaryl, and perhaloheteroaeryl follow the definition of perhaloalkyl, and refer to an alkenyl, alkynyl, carbocyclyl, heterocyclyl, aryl, and heteroaryl group, as defined herein, wherein all of the hydrogen atoms are independently replaced by a halogen, e.g., fluoro, bromo, chloro, or iodo.

As used herein, "heteroalkyl" refers to an alkyl group as defined herein which further includes at least one heteroatom (e.g., 1, 2, 3, or 4 heteroatoms) selected from oxygen, nitrogen, or sulfur within (i.e., inserted between adjacent carbon atoms of) and/or placed at one or more terminal position(s) of the parent chain. In certain embodiments, a heteroalkyl group refers to a saturated group having from 1 to 10 carbon atoms and 1 or more heteroatoms within the parent chain ("heteroC_1-o alkyl"). In some embodiments, a heteroalkyl group is a saturated group having 1 to 9 carbon atoms and 1 or more heteroatoms within the parent chain ("heteroC_1-9 alkyl"). In some embodiments, a heteroalkyl group is a saturated group having 1 to 8 carbon atoms and 1 or more heteroatoms within the parent chain ("heteroC_1-8 alkyl"). In some embodiments, a heteroalkyl group is a saturated group having 1 to 7 carbon atoms and 1 or more heteroatoms within the parent chain ("heteroC_1-7 alkyl"). In some embodiments, a heteroalkyl group is a saturated group having 1 to 6 carbon...
atoms and 1 or more heteroatoms within the parent chain ("heteroC<sub>1-6</sub> alkyl"). In some embodiments, a heteroalkyl group is a saturated group having 1 to 5 carbon atoms and 1 or 2 heteroatoms within the parent chain ("heteroC<sub>i.5</sub> alkyl"). In some embodiments, a heteroalkyl group is a saturated group having 1 to 4 carbon atoms and 1 or 2 heteroatoms within the parent chain ("heteroC<sub>3</sub> alkyl"). In some embodiments, a heteroalkyl group is a saturated group having 1 to 3 carbon atoms and 1 heteroatom within the parent chain ("heteroC<sub>2</sub> alkyl"). In some embodiments, a heteroalkyl group is a saturated group having 1 to 2 carbon atoms and 1 heteroatom within the parent chain ("heteroC<sub>1</sub> alkyl"). In some embodiments, a heteroalkyl group is a saturated group having 1 carbon atom and 1 heteroatom ("heteroC<sub>0</sub> alkyl"). In some embodiments, a heteroalkyl group is a saturated group having 2 to 6 carbon atoms and 1 or 2 heteroatoms within the parent chain ("heteroC<sub>2-6</sub> alkyl"). Unless otherwise specified, each instance of a heteroalkyl group is independently unsubstituted (an "unsubstituted heteroalkyl") or substituted (a "substituted heteroalkyl") with one or more substituents. In certain embodiments, the heteroalkyl group is an unsubstituted heteroC<sup>0</sup> alkyl. In certain embodiments, the heteroalkyl group is a substituted heteroC<sup>0</sup> alkyl.

[0088] As used herein, "alkenyl" refers to a radical of a straight-chain or branched hydrocarbon group having from 2 to 20 carbon atoms and one or more carbon-carbon double bonds (e.g., 1, 2, 3, or 4 double bonds). In some embodiments, an alkenyl group has 2 to 10 carbon atoms ("C<sub>2-10</sub> alkenyl"). In some embodiments, an alkenyl group has 2 to 9 carbon atoms ("C<sub>2-9</sub> alkenyl"). In some embodiments, an alkenyl group has 2 to 8 carbon atoms ("C<sub>2-8</sub> alkenyl"). In some embodiments, an alkenyl group has 2 to 7 carbon atoms ("C<sub>2-7</sub> alkenyl"). In some embodiments, an alkenyl group has 2 to 6 carbon atoms ("C<sub>2-6</sub> alkenyl"). In some embodiments, an alkenyl group has 2 to 5 carbon atoms ("C<sub>2-5</sub> alkenyl"). In some embodiments, an alkenyl group has 2 to 4 carbon atoms ("C<sub>2-4</sub> alkenyl"). In some embodiments, an alkenyl group has 2 to 3 carbon atoms ("C<sub>2-3</sub> alkenyl"). In some embodiments, an alkenyl group has 2 carbon atoms ("C<sub>2</sub> alkenyl"). The one or more carbon-carbon double bonds can be internal (such as in 2-butene) or terminal (such as in 1-butene). Examples of C<sub>2-4</sub> alkenyl groups include ethenyl (C<sub>2</sub>), 1-propenyl (C<sub>3</sub>), 2-propenyl (C<sub>3</sub>), 1-butenyl (C<sub>4</sub>), 2-butenyl (C<sub>4</sub>), butadienyl (C<sub>4</sub>), and the like. Examples of C<sub>2-6</sub> alkenyl groups include the aforementioned C<sub>2-4</sub> alkenyl groups as well as pentenyl (C<sub>5</sub>), pentadienyl (C<sub>5</sub>), hexenyl (C<sub>6</sub>), and the like. Additional examples of alkenyl include heptenyl (C<sub>7</sub>), octenyl (C<sub>8</sub>), octatrienyl (C<sub>9</sub>), and the like. Unless otherwise specified, each instance of an alkenyl group is independently unsubstituted (an "unsubstituted alkenyl") or substituted (a
"substituted alkenyl") with one or more substituents. In certain embodiments, the alkenyl group is an unsubstituted C₂₀ alkenyl. In certain embodiments, the alkenyl group is a substituted C₂₀ alkenyl.

[0089] As used herein, "heteroalkenyl" refers to an alkenyl group as defined herein which further includes at least one heteroatom (e.g., 1, 2, 3, or 4 heteroatoms) selected from oxygen, nitrogen, or sulfur within (i.e., inserted between adjacent carbon atoms of) and/or placed at one or more terminal position(s) of the parent chain. In certain embodiments, a heteroalkenyl group refers to a group having from 2 to 20 carbon atoms, at least one double bond, and 1 or more heteroatoms within the parent chain ("heteroC₂₀ alkenyl"). In certain embodiments, a heteroalkenyl group refers to a group having from 2 to 10 carbon atoms, at least one double bond, and 1 or more heteroatoms within the parent chain ("heteroC₂₀ alkenyl"). In some embodiments, a heteroalkenyl group has 2 to 7 carbon atoms, at least one double bond, and 1 or more heteroatoms within the parent chain ("heteroC₂₋₇ alkenyl"). In some embodiments, a heteroalkenyl group has 2 to 6 carbon atoms, at least one double bond, and 1 or more heteroatoms within the parent chain ("heteroC₂₋₆ alkenyl"). In some embodiments, a heteroalkenyl group has 2 to 5 carbon atoms, at least one double bond, and 1 or 2 heteroatoms within the parent chain ("heteroC₂₋₅ alkenyl"). In some embodiments, a heteroalkenyl group has 2 to 4 carbon atoms, at least one double bond, and 1 or 2 heteroatoms within the parent chain ("heteroC₂₋₄ alkenyl"). In some embodiments, a heteroalkenyl group has 2 to 3 carbon atoms, at least one double bond, and 1 heteroatom within the parent chain ("heteroC₂₋₃ alkenyl"). In some embodiments, a heteroalkenyl group has 2 to 6 carbon atoms, at least one double bond, and 1 or 2 heteroatoms within the parent chain ("heteroC₂₋₆ alkenyl"). Unless otherwise specified, each instance of a heteroalkenyl group is independently unsubstituted (an "unsubstituted heteroalkenyl") or substituted (a "substituted heteroalkenyl") with one or more substituents. In certain embodiments, the heteroalkenyl group is an unsubstituted heteroC₂₋₀ alkenyl. In certain embodiments, the heteroalkenyl group is a substituted heteroC₂₋₀ alkenyl.

[0090] As used herein, "alkynyl" refers to a radical of a straight-chain or branched hydrocarbon group having from 2 to 20 carbon atoms and one or more carbon-carbon triple bonds (e.g., 1, 2, 3, or 4 triple bonds) ("C₂₋₂₀ alkynyl"). In some embodiments, an alkynyl
group has 2 to 10 carbon atoms ("C\textsubscript{2-10} alkynyl"). In some embodiments, an alkynyl group has 2 to 9 carbon atoms ("C\textsubscript{2-9} alkynyl"). In some embodiments, an alkynyl group has 2 to 8 carbon atoms ("C\textsubscript{2-8} alkynyl"). In some embodiments, an alkynyl group has 2 to 7 carbon atoms ("C\textsubscript{2-7} alkynyl"). In some embodiments, an alkynyl group has 2 to 6 carbon atoms ("C\textsubscript{2-6} alkynyl"). In some embodiments, an alkynyl group has 2 to 5 carbon atoms ("C\textsubscript{2-5} alkynyl"). In some embodiments, an alkynyl group has 2 to 4 carbon atoms ("C\textsubscript{2-4} alkynyl"). In some embodiments, an alkynyl group has 2 to 3 carbon atoms ("C\textsubscript{2-3} alkynyl"). In some embodiments, an alkynyl group has 2 carbon atoms ("C\textsubscript{2} alkynyl"). The one or more carbon-carbon triple bonds can be internal (such as in 2-butynyl) or terminal (such as in 1-butynyl).

Examples of C\textsubscript{2-4} alkynyl groups include, without limitation, ethynyl (C\textsubscript{2}), 1-propynyl (C\textsubscript{3}), 2-propynyl (C\textsubscript{3}), 1-butynyl (C\textsubscript{4}), 2-butynyl (C\textsubscript{4}), and the like. Examples of C\textsubscript{2-6} alkenyl groups include the aforementioned C\textsubscript{2-4} alkynyl groups as well as pentynyl (C\textsubscript{5}), hexynyl (C\textsubscript{6}), and the like. Additional examples of alkynyl include heptynyl (C\textsubscript{7}), octynyl (C\textsubscript{8}), and the like. Unless otherwise specified, each instance of an alkynyl group is independently unsubstituted (an "unsubstituted alkynyl") or substituted (a "substituted alkynyl") with one or more substituents. In certain embodiments, the alkynyl group is an unsubstituted C\textsubscript{2-20} alkynyl. In certain embodiments, the alkynyl group is a substituted C\textsubscript{2-20} alkynyl.

[0091] As used herein, "heteroalkynyl" refers to an alkynyl group as defined herein which further includes at least one heteroatom (e.g., 1, 2, 3, or 4 heteroatoms) selected from oxygen, nitrogen, or sulfur within (i.e., inserted between adjacent carbon atoms of) and/or placed at one or more terminal position(s) of the parent chain. In certain embodiments, a heteroalkynyl group refers to a group having from 2 to 20 carbon atoms, at least one triple bond, and 1 or more heteroatoms within the parent chain ("heteroC\textsubscript{2-20} alkynyl"). In certain embodiments, a heteroalkynyl group refers to a group having from 2 to 10 carbon atoms, at least one triple bond, and 1 or more heteroatoms within the parent chain ("heteroC\textsubscript{2-10} alkynyl"). In some embodiments, a heteroalkynyl group has 2 to 9 carbon atoms, at least one triple bond, and 1 or more heteroatoms within the parent chain ("heteroC\textsubscript{2-9} alkynyl"). In some embodiments, a heteroalkynyl group has 2 to 8 carbon atoms, at least one triple bond, and 1 or more heteroatoms within the parent chain ("heteroC\textsubscript{2-8} alkynyl"). In some embodiments, a heteroalkynyl group has 2 to 7 carbon atoms, at least one triple bond, and 1 or more heteroatoms within the parent chain ("heteroC\textsubscript{2-7} alkynyl"). In some embodiments, a heteroalkynyl group has 2 to 6 carbon atoms, at least one triple bond, and 1 or more heteroatoms within the parent chain ("heteroC\textsubscript{2-6} alkynyl"). In some embodiments, a heteroalkynyl group has 2 to 5 carbon atoms, at least one triple bond, and 1 or 2 heteroatoms
within the parent chain ("heteroC₆₋₅ alkynyl"). In some embodiments, a heteroalkynyl group has 2 to 4 carbon atoms, at least one triple bond, and 1 or 2 heteroatoms within the parent chain ("heteroC₂₋₅ alkynyl"). In some embodiments, a heteroalkynyl group has 2 to 3 carbon atoms, at least one triple bond, and 1 heteroatom within the parent chain ("heteroC₂₋₃ alkynyl"). In some embodiments, a heteroalkynyl group has 2 to 6 carbon atoms, at least one triple bond, and 1 or 2 heteroatoms within the parent chain ("heteroC₂₋₆ alkynyl"). Unless otherwise specified, each instance of a heteroalkynyl group is independently unsubstituted (an "unsubstituted heteroalkynyl") or substituted (a "substituted heteroalkynyl") with one or more substituents. In certain embodiments, the heteroalkynyl group is an unsubstituted heteroC₂₋₀ alkynyl. In certain embodiments, the heteroalkynyl group is a substituted heteroC₂₋ₙ alkynyl.

[0092] As used herein, "carbocyclyl" or "carbocyclic" refers to a radical of a non-aromatic cyclic hydrocarbon group having from 3 to 10 ring carbon atoms (\(\text{C}_{3₋₁₀}\) carbocyclyl") and zero heteroatoms in the non-aromatic ring system. In some embodiments, a carbocyclyl group has 3 to 8 ring carbon atoms ("C₃₋₈ carbocyclyl"). In some embodiments, a carbocyclyl group has 3 to 7 ring carbon atoms ("C₃₋₇ carbocyclyl"). In some embodiments, a carbocyclyl group has 3 to 6 ring carbon atoms ("C₃₋₆ carbocyclyl"). In some embodiments, a carbocyclyl group has 4 to 6 ring carbon atoms ("C₄₋₆ carbocyclyl"). In some embodiments, a carbocyclyl group has 5 to 6 ring carbon atoms ("C₅₋₆ carbocyclyl"). In some embodiments, a carbocyclyl group has 5 to 10 ring carbon atoms ("C₅₋₁₀ carbocyclyl"). Exemplary C₃₋₆ carbocyclyl groups include, without limitation, cyclopentyl (C₅), cyclopropenyl (C₃), cyclobutyl (C₄), cyclobutenyl (C₄), cyclopentyl (C₅), cyclopentenyl (C₅), cyclohexyl (C₆), cyclohexenyl (C₆), cyclohexadienyl (C₆), and the like. Exemplary C₃₋₈ carbocyclyl groups include, without limitation, the aforementioned C₃₋₆ carbocyclyl groups as well as cycloheptyl (C₇), cycloheptenyl (C₇), cycloheptadienyl (C₇), cycloheptatrienyl (C₇), cyclooctyl (C₈), cyclooctenyl (C₈), bicyclo[2.2.1]heptyl (C₇), bicyclo[2.2.2]octyl (C₈), and the like. Exemplary C₃₋₁₀ carbocyclyl groups include, without limitation, the aforementioned C₃₋₈ carbocyclyl groups as well as cyclononyl (C₉), cyclononanyl (C₉), cyclodecyl (C₁₀), cyclodecenyl (C₁₀), octahydro-1 H-indenyl (C₉), decahydro-1 napthalenyl (C₁₀), spiro[4,5]decanyl (C₁₀), and the like. As the foregoing examples illustrate, in certain embodiments, the carbocyclyl group is either monocyclic ("monocyclic carbocyclyl") or polycyclic (e.g., containing a fused, bridged or spiro ring system such as a bicyclic system ("bicyclic carbocyclyl") or tricyclic system ("tricyclic carbocyclyl")). and can be saturated or can contain one or more carbon-carbon double or triple bonds. "Carbocyclyl" also includes
ring systems wherein the carbocyclyl ring, as defined above, is fused with one or more aryl or heteroaryl groups wherein the point of attachment is on the carbocyclyl ring, and in such instances, the number of carbons continue to designate the number of carbons in the carbocyclic ring system. Unless otherwise specified, each instance of a carbocyclyl group is independently unsubstituted (an "unsubstituted carbocyclyl") or substituted (a "substituted carbocyclyl") with one or more substituents. In certain embodiments, the carbocyclyl group is an unsubstituted C_{3-10} carbocyclyl. In certain embodiments, the carbocyclyl group is a substituted C_{3-10} carbocyclyl.

[0093] In some embodiments, "carbocyclyl" is a monocyclic, saturated carbocyclyl group having from 3 to 10 ring carbon atoms ("C_{3-10} cycloalkyl"). In some embodiments, a cycloalkyl group has 3 to 8 ring carbon atoms ("C_{3-8} cycloalkyl"). In some embodiments, a cycloalkyl group has 3 to 6 ring carbon atoms ("C_{3-6} cycloalkyl"). In some embodiments, a cycloalkyl group has 4 to 6 ring carbon atoms ("C_{4-6} cycloalkyl"). In some embodiments, a cycloalkyl group has 5 to 6 ring carbon atoms ("C_{5-6} cycloalkyl"). In some embodiments, a cycloalkyl group has 5 to 10 ring carbon atoms ("C_{5-10} cycloalkyl"). Examples of C_{5-6} cycloalkyl groups include cyclopentyl (C_{5}) and cyclohexyl (C_{6}). Examples of C_{3-6} cycloalkyl groups include the aforementioned C_{3-6} cycloalkyl groups as well as cyclopropyl (C_{3}) and cyclobutyl (C_{4}). Examples of C_{3-8} cycloalkyl groups include the aforementioned C_{3-6} cycloalkyl groups as well as cycloheptyl (C_{7}) and cyclooctyl (C_{8}). Unless otherwise specified, each instance of a cycloalkyl group is independently unsubstituted (an "unsubstituted cycloalkyl") or substituted (a "substituted cycloalkyl") with one or more substituents. In certain embodiments, the cycloalkyl group is an unsubstituted C_{3-10} cycloalkyl. In certain embodiments, the cycloalkyl group is a substituted C_{3-10} cycloalkyl.

[0094] As used herein, "heterocyclyl" or "heterocyclic" refers to a radical of a 3- to 14-membered non-aromatic ring system having ring carbon atoms and 1 to 4 ring heteroatoms, wherein each heteroatom is independently selected from nitrogen, oxygen, and sulfur ("3-14 membered heterocyclyl"). In heterocyclyl groups that contain one or more nitrogen atoms, the point of attachment can be a carbon or nitrogen atom, as valency permits. A heterocyclyl group can either be monocyclic ("monocyclic heterocyclyl") or polycyclic (e.g., a fused, bridged or spiro ring system such as a bicyclic system ("bicyclic heterocyclyl") or tricyclic system ("tricyclic heterocyclyl")), and can be saturated or can contain one or more carbon-carbon double or triple bonds. Heterocyclyl polycyclic ring systems can include one or more heteroatoms in one or both rings. "Heterocyclyl" also includes ring systems wherein the heterocyclyl ring, as defined above, is fused with one or more carbocyclyl groups wherein the
point of attachment is either on the carbocyclyl or heterocyclyl ring, or ring systems wherein the heterocyclyl ring, as defined above, is fused with one or more aryl or heteroaryl groups, wherein the point of attachment is on the heterocyclyl ring, and in such instances, the number of ring members continue to designate the number of ring members in the heterocyclyl ring system. Unless otherwise specified, each instance of heterocyclyl is independently unsubstituted (an "unsubstituted heterocyclyl") or substituted (a "substituted heterocyclyl") with one or more substituents. In certain embodiments, the heterocyclyl group is an unsubstituted 3-14 membered heterocyclyl. In certain embodiments, the heterocyclyl group is a substituted 3-14 membered heterocyclyl.

In some embodiments, a heterocyclyl group is a 5-10 membered non-aromatic ring system having ring carbon atoms and 1-4 ring heteroatoms, wherein each heteroatom is independently selected from nitrogen, oxygen, and sulfur ("5-10 membered heterocyclyl"). In some embodiments, a heterocyclyl group is a 5-8 membered non-aromatic ring system having ring carbon atoms and 1-4 ring heteroatoms, wherein each heteroatom is independently selected from nitrogen, oxygen, and sulfur ("5-8 membered heterocyclyl"). In some embodiments, a heterocyclyl group is a 5-6 membered non-aromatic ring system having ring carbon atoms and 1-4 ring heteroatoms, wherein each heteroatom is independently selected from nitrogen, oxygen, and sulfur ("5-6 membered heterocyclyl"). In some embodiments, the 5-6 membered heterocyclyl has 1-3 ring heteroatoms selected from nitrogen, oxygen, and sulfur. In some embodiments, the 5-6 membered heterocyclyl has 1-2 ring heteroatoms selected from nitrogen, oxygen, and sulfur. In some embodiments, the 5-6 membered heterocyclyl has 1 ring heteroatom selected from nitrogen, oxygen, and sulfur.

Exemplary 3-membered heterocyclyl groups containing 1 heteroatom include, without limitation, aziridinyl, oxiranyl, thiorenyl. Exemplary 4-membered heterocyclyl groups containing 1 heteroatom include, without limitation, azetidinyl, oxetanyl and thietanyl. Exemplary 5-membered heterocyclyl groups containing 1 heteroatom include, without limitation, tetrahydrofuranyl, dihydrofuran, tetrahydrothiophenyl, dihydrothiophenyl, pyrrolidinyl, dihydropyrrolyl and pyrrolyl-2,5-dione. Exemplary 5-membered heterocyclyl groups containing 2 heteroatoms include, without limitation, dioxolanyl, oxathiolanyl and dithiolanyl. Exemplary 5-membered heterocyclyl groups containing 3 heteroatoms include, without limitation, triazolinyl, oxadiazolinyl, and thiazolinyl. Exemplary 6-membered heterocyclyl groups containing 1 heteroatom include, without limitation, piperidinyl, tetrahydropyran, dihydropyridinyl, and thianyl. Exemplary 6-membered heterocyclyl groups containing 2 heteroatoms include, without limitation, piperidinyl, tetrahydropyran, dihydropyridinyl, and thianyl.
limitation, piperazinyl, morpholinyl, dithianyl, dioxanyl. Exemplary 6-membered heterocyclyl groups containing 2 heteroatoms include, without limitation, triazinanyl. Exemplary 7-membered heterocyclyl groups containing 1 heteroatom include, without limitation, azepanyl, oxepanyl and thiepanyl. Exemplary 8-membered heterocyclyl groups containing 1 heteroatom include, without limitation, azocanyl, oxecanyl and thioecanyl. Exemplary bicyclic heterocyclyl groups include, without limitation, indolinyl, isoindolinyl, dihydrobenzofuranyl, dihydrobenzothienyl, tetrahydrobenzothiényl, tetrahydrobenzofuranyl, tetrahydroindolyl, tetrahydroquinolinyl, tetrahydroisoquinolinyl, decahydroquinolinyl, decahydrosiquinolinyl, octahydrochromenyl, octahydroisochromenyl, decahydrophenanthridinyl, decahydro-1,8-naphthyridinyl, octahydropyrrolo[3,2-b]pyrrole, indoliny1, phthalimidyl, naphthalimidyl, chromanyl, chromenyl, 1H-benzo[e][1,4]diazepinyl, 1,4,5,7-tetrahydropyranofurano[3,4-b]pyrrolyl, 5,6-dihydro-4H-furo[3,2-b]pyrrolyl, 6,7-dihydro-5H-furo[3,2-b]pyrany1, 5,7-dihydro-4H-thieno[2,3-c]pyranyl, 2,3-dihydro-1H-pyrrolo[2,3-b]pyridinyl, 2,3-dihydrofuro[2,3-b]pyrrolyl, 4,5,6,7-tetrahydro-1H-pyrrolo[2,3-b]pyridinyl, 4,5,6,7-tetrahydrofuro[3,2-c]pyrindinyl, 4,5,6,7-tetrahydrothieno[3,2-b]pyrindinyl, 1,2,3,4-tetrahydro-1,6-naphthyridinyl, and the like.

As used herein, "aryl" refers to a radical of a monocyclic or polycyclic (e.g., bicyclic or tricyclic) 4n+2 aromatic ring system (e.g., having 6, 10, or 14 π electrons shared in a cyclic array) having 6-14 ring carbon atoms and zero heteroatoms provided in the aromatic ring system ("C₆₋₁₄ aryl"). In some embodiments, an aryl group has 6 ring carbon atoms ("C₆ aryl"; e.g., phenyl). In some embodiments, an aryl group has 10 ring carbon atoms ("C₁₀ aryl"; e.g., naphthyl such as 1-naphthyl and 2-naphthyl). In some embodiments, an aryl group has 14 ring carbon atoms ("C₁₄ aryl"; e.g., anthracyl). "Aryl" also includes ring systems wherein the aryl ring, as defined above, is fused with one or more carbocyclic or heterocyclic groups wherein the radical or point of attachment is on the aryl ring, and in such instances, the number of carbon atoms continue to designate the number of carbon atoms in the aryl ring system. Unless otherwise specified, each instance of an aryl group is independently unsubstituted (an "unsubstituted aryl") or substituted (a "substituted aryl") with one or more substituents. In certain embodiments, the aryl group is an unsubstituted C₆₋₁₄ aryl. In certain embodiments, the aryl group is a substituted C₆₋₁₄ aryl.

As used herein, "heteroaryl" refers to a radical of a 5-14 membered monocyclic or polycyclic (e.g., bicyclic, tricyclic) 4n+2 aromatic ring system (e.g., having 6, 10, or 14 π electrons shared in a cyclic array) having ring carbon atoms and 1-4 ring heteroatoms.
provided in the aromatic ring system, wherein each heteroatom is independently selected from nitrogen, oxygen and sulfur ("5-14 membered heteroaryl"). In heteroaryl groups that contain one or more nitrogen atoms, the point of attachment can be a carbon or nitrogen atom, as valency permits. Heteroaryl polycyclic ring systems can include one or more heteroatoms in one or both rings. "Heteroaryl" includes ring systems wherein the heteroaryl ring, as defined above, is fused with one or more carbocyclyl or heterocyclyl groups wherein the point of attachment is on the heteroaryl ring, and in such instances, the number of ring members continue to designate the number of ring members in the heteroaryl ring system. "Heteroaryl" also includes ring systems wherein the heteroaryl ring, as defined above, is fused with one or more aryl groups wherein the point of attachment is either on the aryl or heteroaryl ring, and in such instances, the number of ring members designates the number of ring members in the fused polycyclic (aryl/heteroaryl) ring system. Polycyclic heteroaryl groups wherein one ring does not contain a heteroatom (e.g., indolyl, quinolinyl, carbazolyl, and the like) the point of attachment can be on either ring, i.e., either the ring bearing a heteroatom (e.g., 2-indolyl) or the ring that does not contain a heteroatom (e.g., 5-indolyl).

[0099] In some embodiments, a heteroaryl group is a 5-10 membered aromatic ring system having ring carbon atoms and 1-4 ring heteroatoms provided in the aromatic ring system, wherein each heteroatom is independently selected from nitrogen, oxygen, and sulfur ("5-10 membered heteroaryl"). In some embodiments, a heteroaryl group is a 5-8 membered aromatic ring system having ring carbon atoms and 1-4 ring heteroatoms provided in the aromatic ring system, wherein each heteroatom is independently selected from nitrogen, oxygen, and sulfur ("5-8 membered heteroaryl"). In some embodiments, a heteroaryl group is a 5-6 membered aromatic ring system having ring carbon atoms and 1-4 ring heteroatoms provided in the aromatic ring system, wherein each heteroatom is independently selected from nitrogen, oxygen, and sulfur ("5-6 membered heteroaryl"). In some embodiments, the 5-6 membered heteroaryl has 1-3 ring heteroatoms selected from nitrogen, oxygen, and sulfur. In some embodiments, the 5-6 membered heteroaryl has 1-2 ring heteroatoms selected from nitrogen, oxygen, and sulfur. In some embodiments, the 5-6 membered heteroaryl has 1 ring heteroatom selected from nitrogen, oxygen, and sulfur. Unless otherwise specified, each instance of a heteroaryl group is independently unsubstituted (an "unsubstituted heteroaryl") or substituted (a "substituted heteroaryl") with one or more substituents. In certain embodiments, the heteroaryl group is an unsubstituted 5-14 membered heteroaryl. In certain embodiments, the heteroaryl group is a substituted 5-14 membered heteroaryl.
Exemplary 5-membered heteroaryl groups containing 1 heteroatom include, without limitation, pyrrolyl, furanyl and thiophenyl. Exemplary 5-membered heteroaryl groups containing 2 heteroatoms include, without limitation, imidazolyl, pyrazolyl, oxazolyl, isoxazolyl, thiazolyl, and isothiazolyl. Exemplary 5-membered heteroaryl groups containing 3 heteroatoms include, without limitation, triazolyl, oxadiazolyl, and thiadiazolyl. Exemplary 5-membered heteroaryl groups containing 4 heteroatoms include, without limitation, tetrazolyl. Exemplary 6-membered heteroaryl groups containing 1 heteroatom include, without limitation, pyridinyl. Exemplary 6-membered heteroaryl groups containing 2 heteroatoms include, without limitation, pyridazinyl, pyrimidinyl, and pyrazinyl. Exemplary 6-membered heteroaryl groups containing 3 or 4 heteroatoms include, without limitation, triazinyl and tetrazinyl, respectively. Exemplary 7-membered heteroaryl groups containing 1 heteroatom include, without limitation, azepinyl, oxepinyl, and thiepinyl. Exemplary 5,6-bicyclic heteroaryl groups include, without limitation, indolyl, isoindolyl, indazolyl, benzotriazolyl, benzothiophenyl, isobenzothiophenyl, benzofuranyl, benzisofuranyl, benzimidazolyl, benzoxazolyl, benzisoxazolyl, benzoisothiazolyl, benzothiazolyl, benzisothiazolyl, benzthiadiazolyl, indolizinyl, and purinyl. Exemplary 6,6-bicyclic heteroaryl groups include, without limitation, naphthyridinyl, pteridinyl, quinolinyl, isoquinolinyl, cinnolinyl, quinoxalinyl, phthalazinyl, and quinazolinyl. Exemplary tricyclic heteroaryl groups include, without limitation, phenanthridinyl, dibenzofuranyl, carbazolyl, acridinyl, phenothiazinyl, phenoxazinyl and phenazinyl.

As used herein, the term "partially unsaturated" refers to a ring moiety that includes at least one double or triple bond. The term "partially unsaturated" is intended to encompass rings having multiple sites of unsaturation, but is not intended to include aromatic groups (e.g., aryl or heteroaryl moieties) as herein defined.

As used herein, the term "saturated" refers to a ring moiety that does not contain a double or triple bond, i.e., the ring contains all single bonds.

Affixing the suffix "-ene" to a group indicates the group is a divalent moiety, e.g., alkylene is the divalent moiety of alkyl, alkenylene is the divalent moiety of alkenyl, alkynylene is the divalent moiety of alkynyl, heteroalkylene is the divalent moiety of heteroalkyl, heteroalkenyne is the divalent moiety of heteroalkenyl, heteroalkynylene is the divalent moiety of heteroalkynyl, carbocyclylene is the divalent moiety of carbocyclyl, heterocyclylene is the divalent moiety of heterocyclyl, arylene is the divalent moiety of aryl, and heteroarylene is the divalent moiety of heteroaryl.
As understood from the above, alkyl, alkenyl, alkynyl, heteroalkyl, heteroalkenyl, heteroalkynyl, carbocyclyl, heterocyclyl, aryl, and heteroaryl groups, as defined herein, are, in certain embodiments, optionally substituted. Optionally substituted refers to a group which may be substituted or unsubstituted (e.g., "substituted" or "unsubstituted" alkyl, "substituted" or "unsubstituted" alkenyl, "substituted" or "unsubstituted" alkynyl, "substituted" or "unsubstituted" heteroalkenyl, "substituted" or "unsubstituted" heteroalkynyl, "substituted" or "unsubstituted" heteroalkyl, "substituted" or "unsubstituted" carbocyclyl, "substituted" or "unsubstituted" heterocyclyl, "substituted" or "unsubstituted" aryl or "substituted" or "unsubstituted" heteroaryl group). In general, the term "substituted" means that at least one hydrogen present on a group is replaced with a permissible substituent, e.g., a substituent which upon substitution results in a stable compound, e.g., a compound which does not spontaneously undergo transformation such as by rearrangement, cyclization, elimination, or other reaction. Unless otherwise indicated, a "substituted" group has a substituent at one or more substitutable positions of the group, and when more than one position in any given structure is substituted, the substituent is either the same or different at each position. The term "substituted" is contemplated to include substitution with all permissible substituents of organic compounds, any of the substituents described herein that results in the formation of a stable compound. The present invention contemplates any and all such combinations in order to arrive at a stable compound. For purposes of this invention, heteroatoms such as nitrogen may have hydrogen substituents and/or any suitable substituent as described herein which satisfy the valencies of the heteroatoms and results in the formation of a stable moiety.
14 membered heterocyclyl, C_{6-14} aryl, and 5-14 membered heteroaryl, wherein each alkyl, alkenyl, alkynyl, heteroalkyl, heteroalkenyl, heteroalkynyl, carbocyclyl, heterocyclyl, aryl, and heteroaryl is independently substituted with 0, 1, 2, 3, 4, or 5 R^{dd} groups;

or two geminal hydrogens on a carbon atom are replaced with the group =S, =NN(R^{bb})_2, =NNR^{bb}C(=0)R^{aa}, =NNR^{bb}C(=0)OR^{aa}, =NNR^{bb}S(=0)_2R^{aa}, =NR^{bb}, or =NOR^{cc};
each instance of R^{aa} is, independently, selected from C^\alpha alkyl, C^\alpha perhaloalkyl, C_{2-10} alkenyl, C_{2-10} alkynyl, C_i.w heteroalkyl, C_{2-10} heteroalkenyl, C_i.o.heteroalkynyl, C_{3-10} carbocyclyl, 3-14 membered heterocyclyl, C_{6-14} aryl, and 5-14 membered heteroaryl, or two R^{aa} groups are joined to form a 3-14 membered heterocyclyl or 5-14 membered heteroaryl ring, wherein each alkyl, alkenyl, alkynyl, heteroalkyl, heteroalkenyl, heteroalkynyl, carbocyclyl, heterocyclyl, aryl, and heteroaryl is independently substituted with 0, 1, 2, 3, 4, or 5 R^{dd} groups;
each instance of R^{bb} is, independently, selected from hydrogen, -OH, -OR^{aa}, -N(R^{cc})_2, -CN, -(C(=0)R^{aa})_2, -(C(=0)N(R^{cc}))_2, -C0_2R^{aa}, -SO_2R^{aa}, -C(NR^{cc})OR^{aa}, -C(NR^{cc})N(R^{cc})_2, -(C(=0)R^{cc})_2, -C(=0)NR^{cc}OR^{aa}, -C(=0)NR^{cc}N(R^{cc})_2, -(C(=0)SR^{cc})_2, -C(SR^{cc})_2, -P(=0)(R^{aa})_2, -P(=0)(R^{cc})_2, -N_2, perhaloalkyl, C_{2-10} alkenyl, C_{2-10} alkynyl, C^\alpha heteroalkyl, C_{2-10} heteroalkenyl, C_{2-10} heteroalkynyl, C_{3-10} carbocyclyl, 3-14 membered heterocyclyl, C_{6-14} aryl, and 5-14 membered heteroaryl, or two R^{bb} groups are joined to form a 3-14 membered heterocyclyl or 5-14 membered heteroaryl ring, wherein each alkyl, alkenyl, alkynyl, heteroalkyl, heteroalkenyl, heteroalkynyl, carbocyclyl, heterocyclyl, aryl, and heteroaryl is independently substituted with 0, 1, 2, 3, 4, or 5 R^{dd} groups;
each instance of R^{cc} is, independently, selected from hydrogen, C^\alpha alkyl, C_{1-10} perhaloalkyl, C_{2-10} alkenyl, C_{2-10} alkynyl, C_i.w heteroalkyl, C_{2-10} heteroalkenyl, C_{2-10} heteroalkynyl, C_{3-10} carbocyclyl, 3-14 membered heterocyclyl, C_{6-14} aryl, and 5-14 membered heteroaryl, or two R^{cc} groups are joined to form a 3-14 membered heterocyclyl or 5-14 membered heteroaryl ring, wherein each alkyl, alkenyl, alkynyl, heteroalkyl, heteroalkenyl, heteroalkynyl, carbocyclyl, heterocyclyl, aryl, and heteroaryl is independently substituted with 0, 1, 2, 3, 4, or 5 R^{dd} groups;
each instance of R^{dd} is, independently, selected from halogen, -CN, -N0_2, -N_3, -SO_2H, -SO_3H, -OH, -OR^{ee}, -ON(R^{ff})_2, -N(R^{ff})_2, -N(R^{ff})_2X^-, -N(OR^{ee})R^{ff}, -SH, -SR^{ee}, -SSR^{ee}, -(C(=0)R^{ee})_2, -(C(=0)C_0_2R^{ee})_2, -(C(=0)C_0_2R^{ee})_2, -(C(=0)C_0_2R^{ee})_2, -(C(=0)N(R^{ff})_2, -OC(=0)(R^{ff})_2, -NR^{ff}C(=0)R^{ee}, -NR^{ff}C(=0)R^{ee}, -NR^{ff}C(=0)N(R^{ff})_2, -(C(=NR^{ff})OR^{ee}, -OC(=NR^{ff})R^{ee}, -(C(=NR^{ff})OR^{ee}, -C(=NR^{ff})N(R^{ff})_2, -(C(=NR^{ff})N(R^{ff})_2, -OC(=NR^{ff})N(R^{ff})_2, -OC(=NR^{ff})N(R^{ff})_2, -OC(=NR^{ff})N(R^{ff})_2,
NR"C(=NR")N(R") 2 , -NR"S0 2 , -S0 2 R , -S0 2 OR , -OS0 2 R , -S(=0)R ,
-Si(R ) 3 , -OSi(R ) 3 , ... alkyl) 3 , -OSi(Ci_6 alkyl) 3 -
C(=S)N(Ci _6 alkyl) 2 , C(=S)NH(d_6 alkyl), C(=S)NH 2 , -C(=0)S(C^ alkyl), -C(=S)SC^ heteroalkenyl,
[72x366] heterocyclyl,
[72x511] heterocyclyl,
[72x738] -P(=0)(R

each instance of R^ee is, independently, selected from C_{3-10} alkyl, C_2-6 alkyl, C_2-6 alkynyl, C_{1-6} heteroalkyl, C_{2-6} heteroalkenyl, C_3-10 carbocycl, C_6-io aryl, 3-10 membered heterocyclyl, and 3-10 membered heteroaryl, wherein each alkyl, alkenyl, alkynyl, heteroalkyl, heteroalkenyl, heteroalkynyl, carbocycl, heterocycl, aryl, and heteroaryl is independently substituted with 0, 1, 2, 3, 4, or 5 R^{gg} groups; and

each instance of R^ff is, independently, selected from hydrogen, C_{1-6} alkyl, C_{1-6} perhaloalkyl, C_{2-6} alkenyl, C_{2-6} alkynyl, C_{1-6} heteroalkyl, C_{2-6} heteroalkenyl, C_{2-6} heteroalkynyl, C_3-10 carbocycl, C_6-io aryl and 5-10 membered heteroaryl, or two R^ff groups are joined to form a 3-14 membered heterocycl or 5-14 membered heterocycl ring, wherein each alkyl, alkenyl, alkynyl, heteroalkyl, heteroalkenyl, heteroalkynyl, carbocycl, heterocycl, aryl, and heteroaryl is independently substituted with 0, 1, 2, 3, 4, or 5 R^{gg} groups; and

each instance of R^{gg} is, independently, halogen, -CN, -N0 2 , -N3, -S0 2 H, -S0 3 H, -OH, -OCI_6 alkyl, -ON(Ci_6 alkyl) 2 , -N(Ci_6 alkyl) 2 , -N(Ci_6 alkyl) 3 +X^- , -NH(Ci_{1-6}
[72x303]C_{1-6} alkyl) 2 +X^- , -NH_{2-6} (Ci_6 alkyl) 2 +X^- , -NH_{3-10} X^- , -N(OCi_{1-6} alkyl)(Ci_{1-6} alkyl), -N(OH)(Ci_{1-6} alkyl),
[72x531] -NH(OH), -SH, -SCI_6 alkyl, -SS(Ci_e alkyl), -C(=0)(Ci_{1-6} alkyl), -C0 2 H, -C0 2 (Ci_{1-6}
[72x325]OC(=0)(Ci_{1-6} alkyl), -OC0 2 (C_{1-6} alkyl), -C(=0)NH 2 , -C(=0)N(d_{1-6} alkyl), -
[72x717]OC(=0)(Ci_{1-6} alkyl), -NH(=0)( C_{1-6} alkyl), -N(Ci_{1-6} alkyl)C(=0)( C_{1-6} alkyl), -
[72x96]NHC0 2 (Ci_{1-6} alkyl), -NHC(=0)(N(Ci_{1-6} alkyl) 2 , -NHC(=0)NH(Ci_{1-6} alkyl), -NHC(=0)NH 2 ,
[72x116]C(=0)(NH)(Ci_{1-6} alkyl), -OC(=0)NH(Ci_{1-6} alkyl), -OC(=0)NHOC^ alkyl, -alkyl, -C(=NH)N(Ci_{1-6}
[72x179]C(=NH)(NH)NCi_{1-6} alkyl), -C(=NH)NH 2 , -OC(=0)NHNCi_{1-6} alkyl), -
[72x325]OC(=0)NH(Ci^ alkyl), -OC(=0)NH 2 , -NHC(=0)(N(Ci_{1-6} alkyl) 2 , -NHC(=0)NH 2 ,
[72x524]NHS0 2 (Ci_{1-6} alkyl), -S0 2 N(Ci_{1-6} alkyl) 2 , -S0 2 NH(Ci_{1-6} alkyl), -S0 2 NH 2 , -S0 2 Ci_{1-6} alkyl, -
[72x593]S0 2 OC^ alkyl, -OS0 2 Ci_{1-6} alkyl, -SOCl_{1-6} alkyl, -Si(Ci_e alkyl) 3 , -OSi(Ci_{1-6} alkyl) 3 -
[72x76]C(=S)N(Ci_{1-6} alkyl) 2 , C(=S)NH(d_{1-6} alkyl), C(=S)NH 2 , -C(=0)SC^ alkyl, -C(=S)SC^
alkyl, -SC(=S)Salkyl, -P(=0)(d_6 alkyl), -OP(=0)(d_6 alkyl), -OP(=0)(0Ci_6alkyl), C_{1-6}alkyl, d_{6}perhaloalkyl, C_{2-6}alkenyl, C_{2-6}alkynyl, C_{6}heteroalkynyl, C_{6}heteroalkenyl, C_{2-6}heteroalkynyl, C_{3-10}carbocyclyl, C_{6}i'oaryl, 3-10membered heterocyclyl, 5-10membered heteroaryl; or two geminal R^{R'}substituents can be joined to form =0 or =S; wherein X^- is a counterion.

[00106] As used herein, the term "halo" or "halogen" refers to fluorine (fluoro, -F), chlorine (chloro, -Cl), bromine (bromo, -Br), or iodine (iodo, -I).

[00107] As used herein, a "counterion" is a negatively charged group associated with a positively charged quaternary amine in order to maintain electronic neutrality. Exemplary counterions include halide ions (e.g., F-, Cl-, Br-, I-), N0_3-, C10_4-, OCT, H2PO_4-, HS0_4-, sulfonate ions (e.g., methansulfonate, trifluoromethanesulfonate, p-toluenesulfonate, benzenesulfonate, 10-camphor sulfonate, naphthalene-2-sulfonate, naphthalene-1-sulfonic acid-5-sulfonate, ethan-1-sulfonic acid-2-sulfonate, and the like), and carboxylate ions (e.g., acetate, ethanoate, propanoate, benzoate, glycerate, lactate, tartrate, glycolate, and the like).

[00108] In certain embodiments, the substituent present on the nitrogen atom is an nitrogen protecting group (also referred to as an "amino protecting group"). Nitrogen protecting groups include, but are not limited to, -OH, -OR, -N(R^c)C2-, -C(=0)R^aa, -C(=0)N(R^cc)2-, -CO_2R^aa, -SO_2R^aa, -C(=NR^cc)R^aa, -C(=NR^cc)OR^aa, -C(=NR^cc)N(R^cc)2-, -SO_2N(R^cc)2, -SO_2RC^cc, -SO_2OR^cc, -SOR^aa, -C(=S)N(R^cc)2, -C(=S)SR^cc, -C(=S)CR^cc, Ci_i'0alkyl (e.g., aralkyl, heteroaralkyl), C_{2,io}alkenyl, C_{2,io}alkynyl, C^o heteroalkyl, C_{2,io}heteroalkenyl, C_{2,io}heteroalkynyl, C_{3,io}carbocyclyl, 3-14membered heterocyclyl, C_{6}i'4aryl, and 5-14membered heteroaryl groups, wherein each alkyl, alkenyl, alkynyl, heteroalkyl, heteroalkenyl, heteroalkynyl, carbocyclyl, heterocyclyl, aralkyl, aryl, and heteroaryl is independently substituted with 0, 1, 2, 3, 4, or 5 R^{dd}groups, and wherein R^{aa}, R^{bb}, R^{cc}and R^{dd}are as defined herein. Nitrogen protecting groups are well known in the art and include those described in detail in Protecting Groups in Organic Synthesis, T.W. Greene and P.G.M. Wuts, 3rd edition, John Wiley & Sons, 1999, incorporated herein by reference.

[00109] For example, nitrogen protecting groups such as amide groups (e.g., -C(=0)R^{aa}) include, but are not limited to, formamide, acetamide, chloroacetamide, trichloroacetamide, trifluoroacetamide, phenylacetamide, 3-phenylpropanamide, picolinamide, 3-pyridylcarboxamide, N-benzylophenylalanyl derivative, benzamide, p-phenylbenzamide, o-nitophenylacetamide, o-nitrophenoxyacetamide, acetoacetamide, (N'-'dithiobenzyloxyacrylamino)acetamide, 3-(p-hydroxyphenyl)propanamide, 3-(o-
nitrophenyl)propanamide, 2-methyl-2-(o-nitrophenoxy)propanamide, 2-methyl-2-(o-
phenylazophenoxy)propanamide, 4-chlorobutanamide, 3-methyl-3-nitrobutanamide, o-
nitrocinnamide, N-acetylmethionine derivative, o-nitrobenzamide and o-
(benzoyloxymethyl)benzamide.

[00110] Nitrogen protecting groups such as carbamate groups (e.g., -C(=0)OR 
include, but are not limited to, methyl carbamate, ethyl carbamate, 9-fluorenylmethyl 
carbamate (Fmoc), 9-(2-sulfo)fluorenymethyl carbamate, 9-(2,7-dibromo)fluorenymethyl 
carbamate, 2,7-di-i-butyl-[9-(10,10-dioxo-1 0,10,10-tetrahydrothioxanthyl)]methyl 
carbamate (DBD-Tmoc), 4-methoxyphenacyl carbamate (Phenoc), 2,2,2-trichloroethyl 
carbamate (Troce), 2-trimethylsilylethyl carbamate (Teoce), 2-phenylethyl carbamate (hZ), 
1-(l-adamantyl)-l-methylethyl carbamate (Adpoc), 1,l-dimethyl-2-haloethyl carbamate, 
1,l-dimethyl-2,2-dibromoethyl carbamate (DB-i-BOC), 1,l-dimethyl-2,2,2-trichloroethyl 
carbamate (TCBOC), 1-methyl-1-(4-biphenylyl)ethyl carbamate (Bpoc), 1-(5,3—di—l—
butylphenyl)-l-methylethyl carbamate (i-Bumeoc), 2-(2'— and 4'-pyridyl)ethyl carbamate 
(Pyoc), 2-(N,N'-dicyclohexylcarboxamido)ethyl carbamate, i-butyl carbamate (BOC), 1-
adamantyl carbamate (Adoc), vinyl carbamate (Voc), allyl carbamate (Alloc), 1-
isopropylallyl carbamate (Ipaoc), cinnamyl carbamate (Coc), 4-nitrocinnamyl carbamate 
(Noc), 8-quinolyl carbamate, N-hydroxypiperidinyl carbamate, alkylidithio carbamate, 
benzyl carbamate (Cbz), p-methoxybenzyl carbamate (Moz), /?-nitrobenzyl carbamate, p-
bromobenzyl carbamate, p-chlorobenzyl carbamate, 2,4-dichlorobenzyl carbamate, 4-
methylsulfinylbenzyl carbamate (Msz), 9-anthrylmethyl carbamate, diphenylmethyl 
carbamate, 2-methylthioethyl carbamate, 2-methylsulfonyethyl carbamate, 2-(p-
toluene-sulfonyl)ethyl carbamate, [2-(1,3-dithianyl)]methyl carbamate (Dmoc), 4-
methylthiophenyl carbamate (Mtpc), 2,4-dimethylthiophenyl carbamate (Bmpc), 
2-phosphonioethyl carbamate (Peoc), 2-triphenylphosphonoisopropyl carbamate (Ppoc), 1,1-
dimethyl-2-cyanoethyl carbamate, m-chloro-p-acyloxybenzyl carbamate, p-
(dihydroxyboryl)benzyl carbamate, 5-benzisoxazolylmethyl carbamate, 2-(trifluoromethyl)6-
chromonylmethyl carbamate (Tcroc), m-nitrophenyl carbamate, 3,5-dimethoxybenzyl 
carbamate, o-nitrobenzyl carbamate, 3,4-dimethoxy-6-nitrobenzyl carbamate, phenyl(o-
nitrophenyl)methyl carbamate, i-amyl carbamate, 5-benzyl thiocarbamate, p-cyanobenzyl 
carbamate, cyclobutyl carbamate, cyclohexyl carbamate, cyclopentyl carbamate, 
cyclopropylmethyl carbamate, p-decylxybenzyl carbamate, 2,2-dimethoxyacetylvinyl 
carbamate, o-(N,N’-dimethylcarboxamido)benzyl carbamate, 1,l-dimethyl-3 - (N,N-
dimethylcarboxamido)propyl carbamate, 1,1-dimethylpropynyl carbamate, di(2-
pyridyl)methyl carbamate, 2-furanylmethyl carbamate, 2-iodoethyl carbamate, isoborynl carbamate, isobutyl carbamate, isonicotinyl carbamate, p-(p'-methoxyphenylazo)benzyl carbamate, 1-methylcyclobutyl carbamate, 1-methylcyclohexyl carbamate, 1-methyl-1-cyclopropylmethyl carbamate, 1-methyl-1-(3,5-dimethoxyphenyl)ethyl carbamate, 1-methyl-1-(p-phenylazophenyl)ethyl carbamate, 1-methyl-1-(4-pyridyl)ethyl carbamate, phenyl carbamate, /?-(phenylazo)benzyl carbamate, 2,4,6-tri-i-butylphenyl carbamate, 4-(trimethylammonium)benzyl carbamate, and 2,4,6-trimethylbenzyl carbamate.

[00111] Nitrogen protecting groups such as sulfonamide groups (e.g., -S(=O) 2R^m) include, but are not limited to, p-toluenesulfonamide (Ts), benzenesulfonamide, 2,3,6,3-trimethyl-4-methoxybenzenesulfonamide (Mtr), 2,4,6-trimethoxybenzenesulfonamide (MtB), 2,6-dimethyl-4-methoxybenzenesulfonamide (Pme), 2,3,5,6-tetramethyl-4-methoxybenzenesulfonamide (Mte), 4-methoxybenzenesulfonamide (Mbs), 2,4,6-trimethylbenzenesulfonamide (Mts), 2,6-dimethoxy-4-methylbenzenesulfonamide (iMds), 2,5,7,8-pentamethylchroman-6-sulfonamide (Pmc), methanesulfonamide (Ms), β-trimethylsilylethanesulfonamide (SES), 9-anthracenesulfonamide, 4-(4',8'-dimethoxynaphthylmethyl)benzenesulfonamide (DNMBS), benzylsulfonamide, trifluoromethylsulfonamide, and phenacylsulfonamide.

[00112] Other nitrogen protecting groups include, but are not limited to, phenothiazinyl-(10)-acyl derivative, N'-p-toluene sulfonylaminoacyl derivative, N'-phenylaminothioacyl derivative, N-benzoylphenylalanyl derivative, N-acetylmethionine derivative, 4,5-diphenyl-3-oxazolin-2-one, N-phthalimide, N-dithiasuccinimide (Dts), N-2,3-diphenylmaleimide, N-2,5-dimethylpyrrole, N-1,1,4,4-tetramethylidisilylazacyclopentane adduct (STABASE), 5-substituted 1,3-dimethyl-1,3,5-triazacyclohexan-2-one, 5-substituted 1,3-dibenzyl-1,3,5-triazacyclohexan-2-one, 1-substituted 3,5-dinitro-4-pyridone, N-methylamine, N-allylamine, N-[2-(trimethylsilyl)ethoxy]methylamine (SEM), N-3-acetoxypropylamine, N-(1-isopropyl-4-nitro-2-oxo-3-pyroolin-3-yl)amine, quaternary ammonium salts, N-benzylamine, N-di(4-methoxyphenyl)methylamine, N-5-dibenzosuberylamine, N-triphenylmethylamine (Tr), N-[(4-methoxyphenyl)diphenylmethyl] amine (MMTr), N-9-phenylfluorenylamine (PhF), N-2,7-dichloro-9-fluorenylmethyleneamine, N-ferrocenylmethylamino (Fcm), N-2-picolylamino N'-oxide, N-1,1,1-dimethylthioureyleneamine, N-benzylideneamine, N-p-methoxybenzylideneamine, N-diphenylmethyleneamine, N-[(2-pyridyl)mesityl)methyleneamine, N-(N',N'-dimethylaminomethylene)amine, N,N'-isopropylidenediamine, N-p-nitrobenzylideneamine,
N-salicylideneamine, N-5-chlorosalicylideneamine, N-(5-chloro-2-hydroxyphenyl)phenylmethyleneamine, N-cyclohexylideneamine, N-(5,5-dimethyl-3-oxo-1-cyclohexenyl)amine, N-borane derivative, N-diphenylborinic acid derivative, N-[phenyl(pentaacetylchromium- or tungsten)acetyl]amine, N-copper chelate, N-zinc chelate, N-nitroamine, N-nitrosoamine, amine N-oxide, diphenylphosphinamide (Dpp), dimethylthiophosphinamide (Mpt), diphenylthiophosphinamide (Ppt), dialkyl phosphoramidates, dibenzyl phosphoramidate, diphenyl phosphoramidate, benzenesulfenamide, o-nitrobenzenesulfenamide (Nps), 2,4-dinitrobenzenesulfenamide, pentachlorobenzenesulfenamide, 2-nitro-4-methoxybenzenesulfenamide, triphenylmethylsulfenamide, and 3-nitropyridinesulfenamide (Nyps).

[00113] In certain embodiments, the substituent present on an oxygen atom is an oxygen protecting group (also referred to as a "hydroxyl protecting group"). Oxygen protecting groups include, but are not limited to, -R<sup>a</sup>, -N(R<sup>bb</sup>)<sub>2</sub>, -C(=0)SR<sup>aa</sup>, -C(=0)OR<sup>aa</sup>, -C(=0)N(R<sup>cc</sup>)<sub>2</sub>, -Si(R<sup>aa</sup>)<sub>2</sub>, -P(=0)(OR<sup>cc</sup>)<sub>2</sub>, -P(=0)(SR<sup>aa</sup>)<sub>2</sub>, -P(=0)(NR<sup>cc</sup>)<sub>2</sub>, and -P(=0)(NR<sup>bb</sup>)<sub>2</sub>, wherein R<sup>aa</sup>, R<sup>bb</sup>, and R<sup>cc</sup> are as defined herein. Oxygen protecting groups are well known in the art and include those described in detail in Protecting Groups in Organic Synthesis, T. W. Greene and P. G. M. Wuts, 3rd edition, John Wiley & Sons, 1999, incorporated herein by reference.

[00114] Exemplary oxygen protecting groups include, but are not limited to, methyl, methoxymethyl (MOM), methyliothioethyl (MTM), t-butyliothioethyl, (phenyldimethylsilyl)methoxymethyl (SMOM), benzyloxymethyl (BOM), p-methoxybenzyloxymethyl (PMBM), (4-methoxyphenoxymethyl) (p-AOM), guaiacolmethyl (GUM), i-butoxymethyl, 4-pentenyloxymethyl (POM), siloxymethyl, 2-methoxymethoxymethyl (MEM), 2,2,2-trichloroethoxymethyl, bis(2-chloroethoxy)methyl, 2-(trimethylsilyl)ethoxymethyl (SEPOR), tetrahydropyranyl (THP), 3-bromotetrahydropropyranyl, tetrahydrothiopyranyl, 1-methoxycyclohexyl, 4-methoxytetrahydropropyranyl (MTDP), 4-methoxytetrahydrothiopyranyl, 4-methoxytetrahydrothiopyranyl S,S-dioxide, 1-[2-chloro-4-methylphenyl]-4-methoxypiperidin-4-yl (CTMP), 1,4-dioxan-2-yl, tetrahydrofuran, tetrahydrothiofuran, 2,3,3a,4,5,6,7,7a-octahydro-7,8,8-trimethyl-4,7-methanobenzofuran-2-yl, 1-ethoxyethyl, 1-(2-chloroethoxy)ethyl, 1-methyl-1-methoxycarbonyl, 1-methyl-1-benzylxoyethyl, 1-methyl-1-benzylxoy-2-fluoroethyl, 2,2,2-trichloroethyl, 2-trimethylsilylthyl, 2-(phenylseleny)ethyl, i-butyl, allyl, p-chlorophenyl, p-methoxyphenyl, 2,4-dinitrophenyl,
benzyl (Bn), \(\beta\)-methoxybenzyl, 3,4-dimethoxybenzyl, o-nitrobenzyl, p-nitrobenzyl, \(p\)-halobenzyl, 2,6-dichlorobenzyl, p-cyanobenzyl, p-phenylbenzyl, 2-picoly, 4-picoly, 3-methyl-2-picoly \(N\)-oxido, diphenylmethyl, \(p,p'\)-dinitrobenzhydryl, 5-dibenzosuberyl, triphenylmethyl, \(\alpha\)-napthyl diphenylmethyl, \(p\)-methoxyphenyl diphenylmethyl, \(di(p\)-methoxyphenyl)phenylmethyl, tri(p-methoxyphenyl)methyl, 4-(4'-bromophenacyloxyphenyl)diphenylmethyl, 4,4',4'-tris(4,5-dichlorophthalimido)phenyl)methyl, 4,4',4'-tris(levulinoxyloxyphenyl)methyl, 4,4',4'-tris(benzoyloxyphenyl)methyl, 3-(imidazol-1-yl)bis(4',4''-dimethoxyphenyl)methyl, 1,1-bis(4-methoxyphenyl)-r-pyrenylmethyl, 9-anthryl, 9-(9-phenyl)xanthenyl, 9-(9-phenyl-10-oxo)anthryl, 1,3-benzodithiolan-2-yl, benzisothiazolyl S,S-dioxido, trimethylsilyl (TMS), triethylsilyl (TES), triisopropylsilyl (TIPS), dimethylisopropylsilyl (IPDMS), diethylisopropylsilyl (DEIPS), dimethylethylsilyl, i-butyldimethylsilyl (TBDMS), t-butylidiphenylsilyl (TBDPS), tribenzylsilyl, tri-p-xyllysilyl, triphenylsilyl, diphenylmethysilyl (DPMS), i-butylmethoxyphenylsilyl (TBMPS), formate, benzoylformate, acetate, chloroacetate, dichloroacetate, trichloroacetate, trifluoroacetate, methoxyacetate, triphenylmethoxyacetate, phenoxyacetate, p-chlorophenoxyacetate, 3-phenylpropionate, 4-oxopentanoate (levulinate), 4,4’-(ethylenedithio)pentanoate (levulinoylthioacetol), pivaloate, adamantanoate, crotonate, 4-methoxycrotonate, benzoate, \(p\)-phenylbenzoate, 2,4,6-trimethylbenzoate (mesitoate), alkyl methyl carbonate, 9-fluorenymethyl carbonate (Fmoc), alkyl ethyl carbonate, alkyl 2,2,2-trichloroethyl carbonate (Troc), 2-(trimethylsilyl)ethyl carbonate (TMSEC), 2-(phenylsulfonyl) ethyl carbonate (Psec), 2-(triphenylphosphonio) ethyl carbonate (Peoc), alkyl isobutyl carbonate, alkyl vinyl carbonate alkyl allyl carbonate, alkyl p-nitrophenyl carbonate, alkyl benzyl carbonate, alkyl \(p\)-methoxybenzyl carbonate, alkyl 3,4-dimethoxybenzyl carbonate, alkyl o-nitrobenzyl carbonate, alkyl p-nitrobenzyl carbonate, alkyl 5-benzyl thiocarbonate, 4-ethoxy-1-napththyl carbonate, methyl dithiocarbonate, 2-iodobenzoate, 4-azidobutyrate, 4-nitro-4-methylpentanoate, o-(dibromomethyl)benzoate, 2-formylbenzenesulfonate, 2-(methylthiomethoxy)ethyl, 4-(methylthiomethoxy)butyrate, 2-(methylthiomethoxymethyl)benzoate, 2,6-dichloro-4-methylphenoxyacetate, 2,6-dichloro-4-(1,1,3,3-tetramethylbutyl)phenoxyacetate, 2,4-bis(1,1-dimethylpropyl)phenoxyacetate, chlorodiphenylacetate, isobutyrate, monosuccinoate, \((E\)-)2-methyl-2-butoenoate, \(\alpha\)-(methoxyacetyl)benzoate, \(a\)-napththoate, nitrate, alkyl \(N,N,N':N'\)-tetramethylphosphorodiamidate, alkyl \(N\)-phenylcarbamate, borate, dimethylphosphinothioyl,
alkyl 2,4-dinitrophenylsulfenate, sulfate, methanesulfonate (mesylate), benzylsulfonate, and tosylate (Ts).

**[00115]** In certain embodiments, the substituent present on an sulfur atom is a sulfur protecting group (also referred to as a "thiol protecting group"). Sulfur protecting groups include, but are not limited to, -R\(^{\text{aa}}\), -N(R\(^{\text{bb}}\))\(^{-}\), -C(=0)SR \(^{\text{aa}}\), -C(=0)R \(^{\text{aa}}\), -C0 \(^{\text{aa}}\), -C(=0)N(R\(^{\text{bb}}\))\(^{-}\), -C(=0)O(R\(^{\text{bb}}\))\(^{-}\), -S(=0)R \(^{\text{aa}}\), -SO\(^{\text{aa}}\), -Si(R\(^{\text{aa}}\))\(^{3} \cdot \text{P}(R^{\text{cc}})\(^{2}\), -P(R\(^{\text{cc}}\))\(^{2}\), -P(=0)\(^{2}\)R\(^{\text{aa}}\), -P(=0)(R\(^{\text{aa}}\))\(^{2}\), -P(=0)(OR\(^{\text{cc}}\))\(^{2}\), -P(=0)\(^{2}\)N(R\(^{\text{bb}}\))\(^{-}\), and -P(=0)(NR\(^{\text{bb}}\))\(^{-}\), wherein R\(^{\text{aa}}\), R\(^{\text{bb}}\), and R\(^{\text{cc}}\) are as defined herein. Sulfur protecting groups are well known in the art and include those described in detail in *Protecting Groups in Organic Synthesis*, T. W. Greene and P. G. M. Wuts, 3rd edition, John Wiley & Sons, 1999, incorporated herein by reference.

**[00116]** These and other exemplary substituents are described in more detail in the Detailed Description, Examples, and claims. The invention is not intended to be limited in any manner by the above exemplary listing of substituents.

**[00117]** As used herein, the term "salt" refers to any and all salts.

**[00118]** Exemplary acid-addition salts include, but are not limited to, acid-addition salt between an amino substituent and an inorganic acid such as hydrochloric acid, hydrobromic acid, phosphoric acid, sulfuric acid and perchloric acid, or with organic acids such as acetic acid, oxalic acid, maleic acid, tartaric acid, citric acid, succinic acid or malonic acid or by using other methods used in the art such as ion exchange. Other acid addition salts include salts formed from adipate, alginite, ascorbate, aspartate, benzenesulfonate, benzoate, bisulfate, borate, butyrate, camphorate, camphorsulfonate, citrate, cyclopentanepropionate, digluconate, dodecylsulfate, ethanesulfonate, formate, fumarate, glucoheptonate, glycerophosphate, gluconate, hemisulfate, heptanoate, hexanoate, hydroiodide, 2-hydroxyethanesulfonate, lactobionate, lactate, laurate, lauryl sulfate, malate, maleate, malonate, methanesulfonate, 2-naphthalenesulfonate, nicotinate, nitrate, oleate, oxalate, palmitate, pamoate, pectinate, persulfate, 3-phenylpropionate, phosphate, picrate, pivalate, propionate, stearate, succinate, sulfate, tartrate, thiocyanate, p-toluenesulfonate, undecanoate, valerate salts, and the like.

**[00119]** Exemplary salts derived from appropriate bases include amino acids having a net positive charge, metals, and quaternary amine salts (e.g., \(+\text{NH}_{4}\) and \(+\text{N} \text{(Ci^alkyl)}_{4}\) salts). Representative metals include, but are not limited to, alkali metals (e.g., Li, Na, K, Cs), alkaline earth metals (e.g., Mg, Ca, Ba), and transition metals (e.g., Hg). Exemplary amino acids include, but are not limited to, arginine, histidine, lysine, aspartic acid, glutamic acid,
serine, threonine, asparagine, glutamine, cysteine, selenocysteine, glycine, proline, alanine, valine, isoleucine, leucine, methionine, phenylalanine, tyrosine, and tryptophan.

**DETAILED DESCRIPTION OF CERTAIN EMBODIMENTS OF THE INVENTION**

[00120] Compositions of heavy-atom labeled nucleic acids for use in systems and methods of sequencing, identifying and/or detecting nucleic acid polymers, such as DNA, are provided. The methods can involve using a particle beam, such as an electron beam, or ion beam, to obtain information regarding the heavy-atom labeled nucleic acid polymer. Examples of such methods using particle beams to obtain information can be found in U.S. Patent Publication Nos. 2006/0024716, 2006/0024717, 2006/0024718, 2006/0029957, which correspond to PCT Application Publication No. WO06019903; and U.S. Patent Publication 2007/0190557, which corresponds to PCT Application Publication No.WO07089542, all entitled, "Systems and Methods of Analyzing Nucleic Acid Polymers and Related Components," each of which is incorporated by reference in its entirety. For example, a sample of heavy-atom labeled DNA can be exposed to a particle beam and changes in the beam resulting from interaction with the sample may form a pattern which can be interpreted to provide the information. In some embodiments, a particle beam instrument (e.g., an electron microscope) can be used to directly view samples of DNA. As described further below, the methods can enable nucleic acid sequencing, identifying and/or detection at high speeds, low costs, and high accuracy, amongst other advantages.

[00121] In some embodiments, a complementary strand of a nucleic acid polymer may be analyzed to determine the sequence and/or presence of a nucleic acid polymer. In certain embodiments, it is preferred that the sample be formed of one or more complementary strands of the nucleic acid polymer. In other embodiments, the sample may be formed of one or more strands of the nucleic acid polymer along with or separate from the complementary strand.

[00122] Conventional techniques may be used to form a complementary strand of a nucleic acid polymer and/or the polymer itself. Typically, the first step in forming the complementary strand is to obtain a single strand of a nucleic acid polymer. Any suitable technique may be used to obtain a single strand. Standard denaturing processes (e.g., thermal, enzymatic) which break the hydrogen bonding between the strands may be used. In other embodiments, a single strand can be created by synthesizing it from a template. For example, polymerase chain reaction (PCR) or reverse transcriptase processes that are well known in the art may be used. In other embodiments, a single strand may be chemically
synthesized one nucleotide at a time, for example, in an oligonucleotide synthesis process. Such synthetic processes are well known in the art and can be automated. It is also possible to obtain a single strand by purifying it from a natural source, such as single stranded RNA from cells. Combinations of the foregoing (and other methods known to those of skill in the art) also can be used.

00123 A complementary strand of a nucleic acid polymer can be created from the single strand using any suitable conventional technique. For example, standard polymerization techniques may be used including polymerase chain reaction (PCR) (e.g., standard PCR, long PCR protocols). The techniques generally involve exposing the single strand to an excess of nucleotides under the proper reaction conditions. The nucleotides may be labeled, as described in further below, and shown schematically in FIG. 2. In some embodiments, single or multiple polymerase enzymes are used to facilitate reactions. Polymerase enzymes include DNA-dependent DNA polymerases (including thermostable enzymes such as Taq polymerase), RNA-dependent DNA polymerases (e.g., reverse transcriptases) and RNA-dependent RNA polymerases. In other embodiments, enzymes need not be used (e.g., in vitro chemical synthesis). Other suitable components (e.g., nucleotide primers, other enzymes such as primases, and the like) may also be present.

00124 It should be understood that complementary strands may be modified to include other components that would not otherwise be present in a DNA strand. For example, the complementary strand may be modified to include labels (e.g., during formation) that facilitate detection and identification of nucleotides in methods of the invention. Labels (e.g., atoms or molecules) when exposed to a particle beam create characteristic particle beam species that may be detected and identified using the systems and methods of the invention. Similarly, the nucleic acid polymer also can be modified to include labels as described herein. This advantageously is done during synthesis of the nucleic acid, for example using PCR, which typically results in the synthesis of both strands (i.e., the nucleic acid polymer and its complementary strand).

00125 When labels are present, it may be preferable to attach the labels to nucleotides of the complementary strand only (e.g., as shown in FIG. 2) or to both strands of the nucleic acid. Labels can be incorporated in the complementary strand only (e.g., using a single round of PCR) or in both strands of the nucleic acid (e.g., using two or more rounds of PCR). In certain embodiments, specific types of label are respectively attached to each type of nucleotide (e.g., cytosine triphosphate (CTP), adenosine triphosphate (ATP), thymine triphosphate (TTP), uracil triphosphate (UTP), guanosine triphosphate (GTP); conventionally
these nucleotides as incorporated into nucleic acid molecules are referred to by a single letter, e.g., A, C, G, T or U). For example, for labeling DNA, a first type of label is attached to a first nucleotide type (e.g., CTP); a second type of label is attached to a second nucleotide type (e.g., ATP); a third type of label is attached to a third nucleotide type (e.g., TTP); and a fourth type of label is attached to a fourth nucleotide type (e.g., GTP). Thus, as described further below, nucleotide types may be identified by identifying a particular label or labels on the labeled nucleotide. Modified (non-natural) or atypical natural nucleotides also can be used, in which the bases, sugars or phosphate moieties can be different than those present in typical naturally occurring nucleotides (e.g., in A, C, G, T and U). One example of this is "locked" nucleic acids, which for example can be a bicyclic nucleic acid where a ribonucleoside is linked between the 2'-oxygen and the 4'-carbon atoms with a methylene unit. Mixtures of the foregoing can be employed in the invention.

[00126] It should be understood that, as used herein, a "nucleotide" comprises a nitrogenous base, a sugar molecule (e.g., deoxyribose in DNA, ribose in RNA) and one or more (typically 1-3) linking groups (e.g., phosphate, peptide). A typical nucleotide is a nucleotide triphosphate, such as cytosine triphosphate as referred to above. As used herein, a "nucleoside" comprises a nitrogenous base and a sugar molecule, as described above, but no linking group. As used herein, a "base" comprises a nitrogenous base, but not the sugar molecule or linking group. Because of these composition differences, a nucleotide can be polymerized into a nucleic acid polymer, but a nucleoside or base cannot. As described further below, one advantage of certain embodiments of the present invention is that labels may be attached to nucleotides, which may be polymerized into nucleic acid polymer, as opposed to nucleic acid bases. Note, however, that a "base pair" is conventionally used to denote pairs of nucleotides that are bound in a sequence specific manner, e.g., Watson-Crick pairing such as A-T and C-G, in a double stranded nucleic acid polymer. However, this term also can refer to pairings of nucleosides or bases, which by definition are not part of nucleic acid polymers.

[00127] One of the advantages of having each nucleotide type bearing a unique label is that only a single "data read" is needed to obtain the sequence directly. Some interpretation as to which strand a given nucleotide is on may be required. Labeling each type of nucleotide uniquely also allows for some flexibility in data interpretation, as each base pair is identified twice: each nucleotide is identified directly and there are two nucleotides per base pair, which provides an internal control for the correctness of the data read and sequence.
In other embodiments, each nucleotide type (e.g., C, A, T, U, G) in a given strand bears a unique label, but the labels on the other strand are different. This can be accomplished by using different sets of labeled nucleotides in sequential PCR cycles, or other synthetic methods, and allows for greater ease in tracking the strand to which a nucleotide belongs.

In certain embodiments, not all nucleotide types need to be labeled. For example, if three nucleotide types (e.g., C, A, T) are labeled and the fourth (e.g., G) is unlabeled, then each "unlabeled" type may readily be identified as the fourth nucleotide type (e.g., G). The position of the unlabeled nucleotides can be inferred from observation of the distances between labeled nucleotides, given the highly regular spacing of nucleotides in nucleic acid polymers. In other embodiments, only two of the nucleotide types may be labeled. For example, a first set of sequencing data may be generated with two nucleotide types labeled (e.g., C, A) and a second set of sequencing data may be generated with the other two nucleotide types labeled (e.g., T, G). Both data sets may be processed to provide information regarding the entire sequence.

Alternatively, by labeling only two nucleotides (e.g., A, C) on both strands of a nucleic acid polymer, the sequence of either strand can be inferred from the sequence of the other strand. For example, all labeled adenines in one strand of a double stranded nucleic acid polymer will be bound to thymines on the opposite strand in accordance with Watson-Crick nucleotide binding rules. Thus, observation of an adenine on one strand allows one to infer the existence of a thymine in the corresponding position of the other strand of a double stranded nucleic acid. The positions of other nucleotides can likewise be directly read or inferred from observing a double stranded nucleic acid that incorporates only two nucleotide-specific labels.

The labels may be attached to nucleotides in a variety of different locations. In some embodiments, labels are attached to the nucleotides on, or within, the nitrogenous base (e.g., adenine, guanine, thymine, cytosine, uracil). For example, in these embodiments, labels may be attached to carbon/nitrogen rings in the base or may replace carbon or nitrogen atoms in the base. In other embodiments, labels are attached to the nucleotides on, or within, the sugar molecule (e.g., ribose in RNA, or deoxyribose in DNA). In other embodiments, labels are attached on, or within, linking groups of the nucleotides. For example, the labels may be attached on, or within, a phosphate linking group. The labels may be attached to oxygen substitutes, such as sulfur (e.g., alpha substituted phosphates, aS) or may replace the phosphorous atom at certain sites.
In certain embodiments, the labels are attached to the nucleotides by covalent bonding. As described further below, covalent bonding provides strong attachment between labels and nucleotides which can enable labeled samples to withstand exposure to relatively high particle beam energies (e.g., greater than about 50 kV for electron beams, for example about 80-120 kV) that may be important to detection and/or identification of nucleic acids.

In certain embodiments, it is preferable that the labels are attached to nucleotides prior to the nucleotides forming the complementary strand (and/or copies of the first strand of the nucleic acid polymer). In these embodiments, the labels may be selected from types, as described further below, that do not prevent polymerase reactions that form the complementary strand (and/or copies of the first strand of the nucleic acid polymer). Thus, in these cases, the complementary strand is labeled during its formation.

However, in other embodiments, it may be desired to attach additional labels to nucleotides after formation of the complementary strand (and/or copies of the first strand of the nucleic acid polymer). In these cases, the nucleotides may have been modified (prior to formation of the complementary strand and/or copies of the first strand of the nucleic acid polymer) to include a suitable attachment site which can be bound, preferably covalently, to a desired label type. After formation, the nucleic acid strand(s) may be exposed to the labels which attach to the sites.

In certain methods of the invention, the complementary strand is separated from first strand to form a single complementary strand as shown which is used as the sample. The complementary strand may be separated from the first strand using conventional denaturing techniques (e.g., thermal, enzymatic). After separation, the first strand may be discarded, or may be retained and otherwise used.

In some cases, separation and use of the complementary strand can simplify detection and/or identification and/or quantitation in subsequent method steps. Although, in some embodiments, the complementary strand and the first strand are not separated, and the double-stranded structure is used as a sample in the detection and/or identification steps.

In certain embodiments, when the complementary strand is separated from the first strand, the complementary strand is used as a template to create another strand which may be labeled. This can create a double-stranded structure which includes two labeled strands (i.e., the complementary strand and the new strand created from the complementary strand). In certain methods, this double-stranded structure is used as the sample in the detection and/or identification steps.
[00138] Methods of the invention may involve attaching a sample (e.g., complementary strand, complementary strand and first strand, complementary strand and new strand), or more than one sample, to a substrate. When more than one sample is attached, the sample may be the same (i.e., based on the same sequence) or different. In general, the substrate should be suitable for exposure to a particle beam. In embodiments in which particle beam species transmitted through the sample are detected, the substrate should permit sufficient transmission of the particle beam.

[00139] The substrate is generally thin to enable sufficient particle beam transmission therethrough. For example, the substrate may be less than 5 nanometers (nm); in some cases, less than 2 nm; or, even less than 1.5 or 1.1 nm. The substrate may be formed of a single layer or multiple layers. In certain cases, the layer(s) may be cross-linked. Conventional techniques can be used to form the substrates including vapor deposition and FIB milling, amongst others.

[00140] Suitable substrate materials are known to those of skill in the art and can include carbon (e.g., pure carbon, graphene, diamond), boron nitride (e.g., having a cubic structure), aluminum and certain polymeric resins (e.g., FORMVAR® (polyvinyl formal)). In other embodiments, the substrate is formed from organic materials such as a lipid, natural protein or synthetic protein. The substrate material may be doped with chemicals, for example, to cross-link layers or to facilitate attachment of the sample as described further below.

[00141] Samples may be attached to the substrate by chemically bonding at least a portion of the sample to the substrate. Suitable techniques are known to those of skill in the art. For example, molecules present on the surface of the substrate (e.g., pre-existing as part of the substrate or following derivatization of the substrate) may be used to bind to the sample. The molecules may be nucleic acid sequence specific molecules (e.g., oligonucleotides). In other cases, the substrate surface may be derivatized to provide attachment points that are sequence non-specific. In other cases, electrical charge may be used to bind the sample to the substrate surface. The attachment points for the samples can be spaced apart in a predetermined pattern, such as a grid or microarray.

[00142] A portion, or portions, of a sample may be attached to the substrate. In some cases, both ends of the sample (e.g., complementary strand, complementary strand and first strand, complementary strand and new strand) may be attached; in other cases, only one end of the sample may be attached; in some cases, one or more non-end portions along the length of the sample may be attached. The attachment at the end(s) or along the length of the nucleic acid
molecule(s) can be facilitated, if desired, by including in the nucleic acid during synthesis nucleotides capable of forming bonds with the substrate.

[00143] Certain methods of the invention involve substantially straightening a sample (e.g., labeled double strand) prior to, during, or even after, attachment to the substrate. This can facilitate detection and/or identification. The labeled double strand may be attached to the substrate, for example, via a linking bond to a bonding site as described further below. Conventional techniques may be used to straighten the sample. For example, a sample may be straightened using fluid flow (e.g., molecular combing). The fluid may comprise one or more liquids, gases, or combinations thereof. In certain embodiments, the sample is attached and straightened by hybridization in a fluid flow to oligonucleotides present on the substrate surface. In some cases, electrical fields may be used (either in the presence of fluid flow, or alone) to promote sample straightening. In embodiments in which more than one sample is attached to the substrate, it may be preferred for each sample to be aligned substantially parallel to one another to facilitate exposure to the beam. Methods exist to perform molecular alignment of nucleic acid molecules in a thin or monolayer on a substrate. Some focus on isolating one or a few strands of materials and stretching them out for observation and genetic analysis. Examples of such methods are molecular combing using an air-water meniscus developed by the Pasteur Institute (e.g., US Patents to Bensimon et al. 5,840,862, 6,265,153 and 6,548,255) and a molecular alignment technique for optical mapping used by OpGen, Inc. Methods also exist to attach nucleic acid molecules in high density patterns on a substrate with a thickness of tens to millions of atoms. An example would be oligo synthesis or spotting on a microarray.

[00144] In certain embodiments, methods and compositions of the present disclosure may be combined with methods to perform high-density molecular alignment of nucleic acid molecules on substrates or surfaces as embodied in PCT Publication Nos. WO 2009/002506 A2 and WO 2010/144128 A2, entitled "High Density Molecular Alignment of Nucleic Acid Molecules," and "Molecular Alignment and Attachment of Nucleic Acid Molecules," respectively, both of which are incorporated herein by reference in their entirety.

[00145] In some embodiments, the disclosure provides compositions and methods aside from nucleic acid sequencing and/or identification, such as gene expression analysis. Procedures used for gene expression are generally based on immobilizing mRNA or cDNA (prepared via reverse transcriptase PCR from mRNA) to microarrays, and estimating quantity from fluorescent images. Some of these procedures are described in U.S. Patent Nos.
5,405,783; 5,424,186; 5,445,934; 5,744,305; 6,261,776; 6,406,844; 6,416,952; 6,506,558; and 5,143,854.

[00146] One aspect of the disclosure provides a substrate having a combination of materials and dimensions that allows the substrate to have distinct physical properties. Specifically, in one embodiment, the materials and dimensions of the substrate allow it to be used for imaging samples with a particle beam instrument such as a transmission electron microscope. The substrate can include one or more ligands (e.g., nucleic acids, polypeptides, oligosaccharides, and synthetic polymers) which may form an array. Corresponding changes in labeling chemistry can allow for ligands, binding partners and other relevant materials to be identifiable, quantitatable, and even sequenceable via modified forms of electron microscopy. In certain embodiments, the array dimensions are on the order of nanometers per functional region rather than micrometers as in certain conventional arrays. With these dimensions, smaller amounts of sample material can be used and more accurate genetic analyses performed. These smaller substrate dimensions may also give rise to dramatically reduced production costs, amongst other advantages. The transparency of the substrate, due to thinness, material type and other factors, may provide a suitable contrast ratio between the labeled molecules and the substrate that result in higher quality readings and lower cost analysis than some conventional techniques.

[00147] Certain embodiments of the invention may be used for identification, quantification, sequencing, fingerprinting, and mapping of polymers, particularly biological polymers. Various embodiments of the invention may be applied, for example, in the sequencing, fingerprinting, identification, quantification, or mapping of nucleic acids, polypeptides, oligosaccharides, and synthetic polymers.

[00148] Aspects of the present disclosure may be combined with the description of certain embodiments in U.S. Patent Publication Nos. 2006/0024716, 2006/0024717, 2006/0024718, 2006/0029957, which correspond to PCT Application Publication No. WO06019903, all entitled, "Systems and Methods of Analyzing Nucleic Acid Polymers and Related Components," as well as 2007/0134699, which corresponds to WO07120202, entitled "Nano-Scale Ligand Arrays on Substrates for Particle Beam Instruments and Related Methods," each of which is incorporated herein by reference in its entirety. These references may provide, for example, methods and devices for incorporating contrast heavy atom labels in a biologic sample that are designed to interfere with a beam from a particle beam instrument. In certain embodiments, the labeled sample materials are binding partners, which can be bound to ligands in an array on a suitable substrate. A particle beam may be
directed through the array and the labels can create interference patterns that are then read by a detector instrument and processed by a data analysis module.

[00149] Methods of the invention involve exposing the sample to a particle beam. In certain embodiments, it is preferred that the particle beam is a lepton beam such as an electron beam. In other cases, the particle beam may be an x-ray beam. Yet in other embodiments, the particle beam may be an ion beam such as a helium or gallium ion beam. When an electron beam is used, a beam generator produces a beam having a desired voltage which, for example, can be greater than 50 kV, *e.g.*, 80-300 kV, preferably 80-120 kV. Beam energies are a function of both voltage and current. The beam current typically ranges between 5 to 25 μA, preferably between 8 and 15 μA. The specific beam energy depends, in part, on the specific analysis being performed.

[00150] Methods can include properly focusing the beam on the sample using a lens arrangement as known to those of skill in the art. Methods may also include a calibration step. In certain cases, the system may be automatically calibrated based on known information from nucleic acid molecules in the sample (such as known molecular geometries and structures) using a feedback loop. For example, data obtained from a nucleic acid sample using an electron beam may include internucleotide (*e.g.*, interlabel) distances. As used herein, an internucleotide distance is the distance from one nucleotide base in one strand to the adjacent nucleotide base in the same strand. While the internucleotide distances of, for example, a DNA molecule are generally known, the internucleotide distance in any given sample may not correspond to the generally known distance, but will typically by substantially uniform within a sample as affixed to a substrate, particularly a sample that has been straightened, *e.g.*, by treatment using molecular combing or like methods. Thus, after obtaining a data read on a given sample, various aspects of the system can be calibrated or adjusted using a feedback control system. For example, knowing the internucleotide distances permits feedback relevant to focusing the particle beam and movement of the sample relative to the particle beam.

[00151] Though systems of the invention may include several components similar to that of a conventional transmission electron microscope (*e.g.*, beam generator, lens, etc.), certain systems of the invention may be more simple than typical conventional TEMs. For example, in some embodiments, the systems are simplified by limiting the magnification range, accelerating voltages, probe diameter, beam current, and sample flexibility, amongst other features. Also, problems related to spherical aberration in conventional TEMs may be
limited, or eliminated, by using a lens arrangement that is pre-set for typical operating
conditions for the system.

[00152] Characteristics of the particle beam are changed when the beam interacts with the
sample. For example, one or more of the following characteristics of the particle beam may
change: energy, direction, absorbance, reflection and deflection. Such changes may result
from interactions between the particle beam and labels attached to nucleotides as described
above. Specific types of labels may produce specific or characteristic changes. Thus, a label
(and, the specific nucleotide to which it is attached) may be identified by recognizing the
specific or characteristic beam changes.

[00153] A detector collects particle beam species after the interaction between the particle
beam and the sample. The detector typically collects beam species that have been transmitted
through the sample, though also can collect beam species that are reflected and/or scattered.
The detector may include a charge coupled device (CCD). The CCD may directly convert
the beam species into digital information. Technologies other than CCD technology may be
used to convert the beam species into digital information, and are intended to fall within the
scope of the invention.

[00154] In some embodiments of the invention, a nucleic acid polymer may be detected,
and/or sequenced and/or identified based on particle beam species detected by a detector
(e.g., the detector described above). Particle beam species may result from exposure of a
sample comprising a nucleic acid polymer and/or its complimentary strand to a particle beam
(e.g., a lepton beam such as an electron beam). Methods, systems, computers, computer
systems, computer storage media, software, and components for analyzing digital information
generated by a detector and e.g. a CCD are known in the art, and are exemplified in U.S.
2006/0024718, 2006/0029957, which correspond to PCT Application Publication No.
WO06019903, each of which are incorporated by reference in their entirety.

Compounds

[00155] Heavy-atom labeled compounds contemplated herein include, but are not limited
to, heavy-atom labeled nucleosides, heavy-atom labeled nucleotides, and heavy-atom labeled
nucleic acid polymers. Such compounds may be useful in the inventive methods as described
herein.

[00156] In one aspect, provided are heavy-atom labeled compounds of Formula (I):
and salts thereof;

wherein:

- each instance of \( G_i \) is independently \(-O-, -S-, -Se-, -CH_2-, \) or \(-NH-\);
- each instance of \( G_2 \) is independently hydrogen, halogen, \(-OR^A, -SR^A, -N(R^A)_2, -SHg, -S0_2SHg, -SHgR^D, -SeR^D \) or \(-TeR^D;\)
- each instance of \( R^A \) is independently hydrogen, substituted or unsubstituted \( C_{1-20} \) alkyl, substituted or unsubstituted \( C_{2-20} \) alkenyl, substituted or unsubstituted \( C_{2-20} \) alkynyl, substituted or unsubstituted carbocyclic, substituted or unsubstituted heterocyclic, substituted or unsubstituted aryl, or substituted or unsubstituted heteroaryl, an oxygen protecting group when attached to an oxygen atom, a sulfur protecting group when attached to a sulfur atom, a nitrogen protecting group when attached to a nitrogen atom; or two \( R^A \) groups are joined to form a substituted or unsubstituted heterocyclic or substituted or unsubstituted heteroaryl ring;
- each instance of \( M_1 \) is independently \(-O-, -S-, -NH-, -Se-, or \(-C(R^M)_2;\) where each instance of \( R^M \) is independently hydrogen or halogen;
- each instance of \( G_3 \) is independently hydrogen, substituted or unsubstituted \( C_{1-20} \) alkyl, substituted or unsubstituted \( C_{2-20} \) alkenyl, substituted or unsubstituted \( C_{2-20} \) alkynyl, substituted or unsubstituted carbocyclic, substituted or unsubstituted heterocyclic, substituted or unsubstituted aryl, or substituted or unsubstituted heteroaryl, or a monophosphate, diphosphate, or triphosphate of formula:

\[
\text{HO-P—} \quad \text{HO-P—} \quad \text{HO-P—}
\]

\( M^2 \) with \( M^2 \) being \( H \), \( OH \), or \( O^2-\) and \( M^2 \) being \( H \) or \( OH \);

- each instance of \( M^2 \) is independently \(-O-, -S-, or -Se-; and
- each instance of \( Base \) is independently:

- Adenine
- Guanine
- Cytosine
- Uracil
- Thymine

or an analog thereof selected from the group consisting of:
wherein:

each instance of $R_1$, $R_2$, $R_4$, and $R_5$ is independently hydrogen, substituted or unsubstituted $C^\alpha$alkyl, substituted or unsubstituted $C_2$-$C_alkenyl$, substituted or unsubstituted $C_2$-$C$alkynyl, substituted or unsubstituted carbocycyl, substituted or unsubstituted heterocyclyl, substituted or unsubstituted aryl, or substituted or unsubstituted heteroaryl, a nitrogen protecting group, $-OR^B$, or $-SR^B$, wherein each instance of $R^B$ is independently hydrogen, substituted or unsubstituted $C_1$-$C_alkyl$, substituted or unsubstituted...
C$_2$-oalkenyl, substituted or unsubstituted C$_{2-20}$ alkynyl, substituted or unsubstituted carbocycyl, substituted or unsubstituted heterocyclyl, substituted or unsubstituted aryl, or substituted or unsubstituted heteroaryl, an oxygen protecting group when attached to an oxygen group, or a sulfur protecting group when attached to a sulfur group; or R$^1$ and R$^2$ and/or R$^4$ and R$^5$ are joined to form a substituted or unsubstituted heterocyclic or substituted or unsubstituted heteroaryl ring;

  each instance of R$^3$ is independently substituted or unsubstituted C$_{1-20}$alkyl, substituted or unsubstituted C$_{2-20}$oalkenyl, substituted or unsubstituted C$_{2-20}$alkynyl, substituted or unsubstituted carbocycyl, substituted or unsubstituted heterocyclyl, substituted or unsubstituted aryl, substituted or unsubstituted heteroaryl, halogen, -OR$^c$, -SR$^c$, -N(R$^c$)$_2$, -SHg, -S0$_2$SHg, -SHgR$^D$, -SeR$^D$, or -TeR$^D$ wherein each instance of R$^C$ is hydrogen, substituted or unsubstituted C$^a$oalkyl, substituted or unsubstituted C$_{2-20}$oalkenyl, substituted or unsubstituted C$_{2-20}$alkynyl, substituted or unsubstituted carbocycyl, substituted or unsubstituted heterocyclyl, substituted or unsubstituted aryl, or substituted or unsubstituted heteroaryl, an oxygen protecting group when attached to an oxygen atom, a sulfur protecting group when attached to a sulfur atom, a nitrogen protecting group when attached to a nitrogen atom; or two R$^C$ groups are joined to form a substituted or unsubstituted heterocyclic or substituted or unsubstituted heteroaryl ring;

  each instance of L$^1$ is independently absent or a linking moiety selected from the group consisting of substituted or unsubstituted C$^a$oalkylene, substituted or unsubstituted C$_{2-20}$oalkenyne, substituted or unsubstituted C$_{2-20}$alkynylene, substituted or unsubstituted heteroC$_{1-20}$oalkylene, substituted or unsubstituted heteroC$_{2-20}$oalkenyne, substituted or unsubstituted heteroC$_{2-20}$alkynylene, substituted or unsubstituted heterocyclylene, substituted or unsubstituted arylene, or substituted or unsubstituted heteroarylene, or a combination thereof;

  each instance of R$^D$ is independently hydrogen, substituted or unsubstituted C$^a$oalkyl, substituted or unsubstituted C$_{2-20}$oalkenyl, substituted or unsubstituted C$_{2-20}$alkynyl, substituted or unsubstituted carbocycyl, substituted or unsubstituted heterocyclyl, substituted or unsubstituted aryl, or substituted or unsubstituted heteroaryl; and

  each instance of M$^3$ and M$^4$ are independently O, Se, Te, CH$_2$, CF$_2$, CCl$_2$, CBr$_2$, or CI$_2$;

  provided that the compound comprises at least one instance of a heavy atom selected from the group consisting of bromine, iodine, selenium, tellurium, or mercury.
[00157] In another aspect, provided are nucleic acid polymers comprising one or more heavy-atom labeled units of Formula (IF), wherein one or more units may be the same or different:

\[
\text{(II')}
\]

wherein Base, \(G_1, G_2, G_3, M^1,\) and \(M^2\) are as defined herein.

[00158] An exemplary nucleic acid polymers is a heavy-atom labeled nucleic acid polymers of Formula (II):

\[
\text{(II)}
\]

and salts thereof; wherein Base, \(G_1, G_2, G_3, M^1,\) and \(M^2\) are as defined herein; and

\(n\) is 1 to 200,000, inclusive;

provided that the polymer comprises at least one instance of a heavy atom selected from the group consisting of bromine, iodine, selenium, tellurium, or mercury.

[00159] In certain embodiments of Formula (II), \(n\) is 1 to 180,000, inclusive; \(n\) is 1 to 160,000, inclusive; \(n\) is 1 to 140,000, inclusive; \(n\) is 1 to 120,000, inclusive; \(n\) is 1 to 100,000, inclusive; \(n\) is 1 to 80,000, inclusive; \(n\) is 1 to 60,000, inclusive; \(n\) is 1 to 40,000, inclusive; \(n\) is 1 to 20,000, inclusive; \(n\) is 1 to 18,000, inclusive; \(n\) is 1 to 16,000, inclusive; \(n\) is 1 to 14,000, inclusive; \(n\) is 1 to 12,000, inclusive; \(n\) is 1 to 10,000, inclusive; \(n\) is 1 to 9,000, inclusive; \(n\) is 1 to 8,000, inclusive; \(n\) is 1 to 7,000, inclusive; \(n\) is 1 to 6,000, inclusive; \(n\) is 1 to 5,000, inclusive; \(n\) is 1 to 4,000, inclusive; \(n\) is 1 to 3,000, inclusive; \(n\) is 1 to 2,000, inclusive; \(n\) is 1 to 1,000, inclusive; \(n\) is 1 to 900, inclusive; \(n\) is 1 to 800, inclusive; \(n\) is 1 to 700, inclusive; \(n\) is 1 to 600, inclusive; \(n\) is 1 to 500, inclusive; \(n\) is 1 to 400, inclusive; \(n\) is 1 to 300, inclusive; \(n\) is 1 to 200, inclusive; \(n\) is 1 to 100, inclusive; \(n\) is 1 to 90, inclusive; \(n\) is 1 to 80, inclusive; \(n\) is 1 to 70, inclusive; \(n\) is 1 to 60, inclusive; \(n\) is 1 to 50, inclusive; \(n\) is 1 to 40, inclusive; \(n\) is 1 to 30, inclusive; \(n\) is 1 to 20, inclusive; \(n\) is 1 to 10, inclusive; \(n\) is 1 to 5, inclusive;
inclusive; \( n \) is 5,000 to 15,000, inclusive; \( n \) is 5,000 to 50,000, inclusive; \( n \) is 5,000 to 200,000, inclusive.

[00160] As depicted herein, it is understood that the compound of Formula (I), polymer of Formula (II), or unit of Formula (II') encompasses any number of stereoisomers. However, in certain embodiments, Formula (I), Formula (II), and Formula (II'), respectively, encompasses the stereoisomers (I-a), (II-a), and (II'-a):

![Chemical structures](image)

the enantiomer thereof, and/or salt thereof.

[00161] Such compounds may be employed using the inventive methods as described herein. However, in certain embodiments, any one of the following compounds of Formula (I) are specifically excluded:

![Chemical structures](image)
In certain embodiments, nucleic acid polymers, such as polymers of Formula (II), comprising one or more units of the below formula are also specifically excluded:

![Chemical structures]

and salts thereof.

Further compounds excluded include, but are not limited to, 2'MeSe-ATP, 2'-TePh, 2'-SeCR, and C5-TePh, and selenium compounds as disclosed in JP2008195648 and JP200700003211.

Heavy atom labels

As generally described herein, the heavy-atom labeled compound of Formula (I), polymer of Formula (II), or unit of Formula (II') comprises at least one instance of a heavy atom selected from the group consisting of bromine, iodine, selenium, tellurium, or mercury.

In certain embodiments, there is only one instance of a heavy atom provided in a compound of Formula (I) or unit of Formula (II') as described herein. In certain embodiments, there are more than one instances of a heavy atom provided in a compound of Formula (I) or unit of Formula (II') as described herein (e.g. 2, 3, 4 or more instances). In certain preferred embodiments, the Base region comprises at least one instance of the heavy atom. In certain embodiments, the sugar region comprises at least one instance of the heavy atom. In certain embodiments the phosphate region comprises at least one instance of the heavy atom. In embodiments, at least one instance of the heavy atom is provided in the Base
region. Labeling in the Base region as described herein is contemplated to provide clearer and unambiguous imaging results compared to labeling elsewhere in the molecule.

[00166] In certain embodiments, the compound comprises at least one instance (e.g., 1, 2, 3, 4 or more instances) of bromine. In certain embodiments, bromine is attached to a carbon atom which optionally comprises one or two additional instances of a halogen, e.g., for example, -CBr\(_3\), CBr\(_2\)H, -CBrH\(_2\), -CBr\(_2\)X, or -CBr\(_2\)X\(_2\), wherein each instance of X is independently -Cl, -F, or -I.

[00167] In certain embodiments, the compound comprises at least one instance (e.g., 1, 2, 3, 4 or more instances) of iodine. In certain embodiments, iodine is attached to a carbon atom which optionally comprises one or two additional instances of a halogen, e.g., for example, -Cl\(_3\), Cl\(_2\)H, -ClH\(_2\), -Cl\(_2\)X, or -Cl\(_2\)X\(_2\), wherein each instance of X is independently -Cl, -F, or -Br.

[00168] In certain embodiments, the compound comprises at least one instance (e.g., 1, 2, 3, 4 or more instances) of selenium, e.g., for example, a divalent =Se or -Se- group, or a monovalent -SeR\(^D\) group. In certain embodiments, the compound comprises a divalent =Se group. In certain embodiments, the compound comprises a divalent -Se- group. In certain embodiments, the compound comprises a monovalent -SeR\(^D\) group.

[00169] In certain embodiments, the compound comprises at least one instance (e.g., 1, 2, 3, 4 or more instances) of tellurium, e.g., for example, a divalent =Te or -Te- group, or -TeR\(^D\). In certain embodiments, the compound comprises a divalent =Te group. In certain embodiments, the compound comprises a divalent -Te- group. In certain embodiments, the compound comprises a monovalent -TeR\(^D\) group.

[00170] As understood from the above, in certain embodiments, the compound comprises at least one instance of -SHgR\(^D\), -SeR\(^D\), or -TeR\(^D\).

[00171] In certain embodiments, the compound comprises at least one instance of -SeR\(^D\) or -TeR\(^D\), wherein R\(^D\) is hydrogen, i.e., to provide -SeH or -TeH.

[00172] In certain embodiments, the compound comprises at least one instance of -SHgR\(^D\), -SeR\(^D\), or -TeR\(^D\), wherein R\(^D\) is substituted or unsubstituted C\(^\#\)alkyl, e.g., R\(^D\) is substituted or unsubstituted C\(_1\)i\(_5\)alkyl, substituted or unsubstituted C\(_1\)i\(_6\)alkyl, substituted or unsubstituted C\(^\#\)\(_1\)i\(_2\)alkyl, substituted or unsubstituted C\(_1\)i\(_3\)alkyl, substituted or unsubstituted C\(_1\)i\(_4\)alkyl, substituted or unsubstituted C\(_1\)i\(_5\)alkyl, substituted or unsubstituted C\(_1\)i\(_6\)alkyl, substituted or unsubstituted C\(^\#\)\(_1\)i\(_2\)alkyl, substituted or unsubstituted C\(_1\)i\(_3\)alkyl, substituted or unsubstituted C\(_1\)i\(_4\)alkyl, substituted or unsubstituted C\(_1\)i\(_5\)alkyl, substituted or unsubstituted C\(_1\)i\(_6\)alkyl, substituted or unsubstituted C\(_1\)i\(_3\)alkyl, substituted or unsubstituted C\(_1\)i\(_4\)alkyl, substituted or unsubstituted C\(_1\)i\(_5\)alkyl, substituted or unsubstituted C\(_1\)i\(_6\)alkyl, substituted or unsubstituted C\(_1\)i\(_3\)alkyl, substituted or unsubstituted C\(_1\)i\(_4\)alkyl, substituted or unsubstituted C\(_1\)i\(_5\)alkyl, subscribed or unsubstituted C\(_1\)i\(_6\)alkyl. In certain embodiments, R\(^D\) is substituted or unsubstituted C\(_1\), C\(_2\), C\(_3\), C\(_4\), C\(_5\), or C\(_6\)-alkyl. In certain embodiments, R\(^D\) is alkyl substituted with at least one or
more halogen atoms (i.e., one or more -Br, -I, -F, or -Cl atoms), e.g., R\(^D\) is substituted or unsubstituted C\(^n\)haloalkyl, substituted or unsubstituted C\(^1\)haloalkyl, substituted or unsubstituted C\(^1\)haloalkyl, substituted or unsubstituted C\(^1\)haloalkyl, substituted or unsubstituted C\(^1\)haloalkyl, substituted or unsubstituted C\(^1\)haloalkyl, substituted or unsubstituted C\(^1\)haloalkyl, substituted or unsubstituted C\(^1\)haloalkyl, or substituted or unsubstituted C\(^n\)haloalkyl. In certain embodiments, R\(^D\) is substituted or unsubstituted C\(_1\), C\(_2\), C\(_3\), C\(_4\), C\(_5\), or C\(_6\)-haloalkyl. In certain embodiments, the haloalkyl is a perhaloalkyl group. In certain embodiments, R\(^D\) is -CX\(_3\), wherein X is halogen. In certain embodiments, R\(^D\) is -CBr\(_3\), CBr\(_2\)H, -CBr\(_2\)X, or -CBrX\(_2\), wherein each instance of X is independently -Cl, -F, or -I. In certain embodiments, R\(^D\) is -Cl\(_3\), Cl\(_2\)H, -Cl\(_2\)X, or -ClX\(_2\), wherein each instance of X is independently -Cl, -F, or -Br. In certain embodiments, R\(^D\) is -CBr\(_3\), -Cl\(_3\), -CFClBr, or -CCIBr.

[00173] In certain embodiments, the compound comprises at least one instance of -SHgR\(^D\), -SeR\(^D\), or -TeR\(^D\), wherein R\(^D\) is substituted or unsubstituted C\(_2\)-alkenyl, e.g., R\(^D\) is substituted or unsubstituted C\(^n\)-alkenyl, substituted or unsubstituted C\(^n\)-alkenyl, substituted or unsubstituted C\(_2\)-alkenyl, substituted or unsubstituted C\(_2\)-alkenyl, substituted or unsubstituted C\(_2\)-alkenyl, substituted or unsubstituted C\(_2\)-alkenyl, substituted or unsubstituted C\(_2\)-alkenyl, substituted or unsubstituted C\(_2\)-alkenyl, or substituted or unsubstituted C\(_2\)-alkenyl. In certain embodiments, R\(^D\) is substituted or unsubstituted C\(_2\)-alkenyl, C\(_4\), C\(_5\), or C\(_6\)-alkenyl. In certain embodiments, R\(^D\) is alkenyl substituted with at least one or more halogen atoms (e.g., 1, 2, 3, 4, 5, 6, or more -Br, -I, -F, or -Cl atoms), e.g., R\(^D\) is substituted or unsubstituted C\(_2\)-haloalkenyl, substituted or unsubstituted C\(_2\)-haloalkenyl, substituted or unsubstituted C\(_2\)-haloalkenyl, substituted or unsubstituted C\(_2\)-haloalkenyl, substituted or unsubstituted C\(_2\)-haloalkenyl, substituted or unsubstituted C\(_2\)-haloalkenyl, substituted or unsubstituted C\(_2\)-haloalkenyl, substituted or unsubstituted C\(_2\)-haloalkenyl, substituted or unsubstituted C\(_2\)-haloalkenyl, substituted or unsubstituted C\(_2\)-haloalkenyl, substituted or unsubstituted C\(_2\)-haloalkenyl, or substituted or unsubstituted C\(_2\)-haloalkenyl. In certain embodiments, the halogenalkenyl is a perhaloalkenyl group. In certain embodiments, R\(^D\) is substituted or unsubstituted C\(_2\), C\(_3\), C\(_4\), C\(_5\), or C\(_6\)-haloalkenyl. In certain embodiments, 

R\(^D\) is -CH\(_2\)CX=CH\(_2\), -CH\(_2\)CH=CH\(_2\), -CH\(_2\)CX=CHX, -CH\(_2\)CH=CHX, -CH\(_2\)CX=CH\(_2\), wherein each instance of X is independently -Cl, -F, -Br, or -I.

[00174] In certain embodiments, the compound comprises at least one instance of -SHgR\(^D\), -SeR\(^D\), or -TeR\(^D\), wherein R\(^D\) is substituted or unsubstituted C\(_2\)-alkynyl, e.g., R\(^D\) is substituted or unsubstituted C\(_2\)-alkynyl, C\(_4\), C\(_5\), or C\(_6\)-alkynyl.
substituted or unsubstituted C_{1-4} alkynyl, substituted or unsubstituted C_{2-12} alkynyl,
substituted or unsubstituted C_{2-10} alkynyl, substituted or unsubstituted C_{2-2} alkynyl, substituted
or unsubstituted C_{2-6} alkynyl, substituted or unsubstituted C_{2-3} alkynyl. In certain embodiments,
R^D is alkynyl substituted or unsubstituted C_{2}, C_{3}, C_{4}, C_{5}, or C_{6}-alkynyl. In certain
embodiments, R^D is alkynyl substituted with at least one or more halogen atoms (e.g., 1, 2, 3, 4, 5, 6,
or more -Br, -I, -F, or -CI atoms), e.g., R^D is substituted or unsubstituted C_{2-2} haloalkynyl,
substituted or unsubstituted C_{2-12} haloalkynyl, substituted or unsubstituted C_{2-6} haloalkynyl,
substituted or unsubstituted C_{2-3} haloalkynyl. In certain embodiments, the haloalkynyl is a
perhaloalkynyl group. In certain embodiments, R^D is substituted or unsubstituted C_{2}, C_{3}, C_{4}, C_{5},
or C_{6}-haloalkynyl. In certain embodiments, R^D is -CH_2X_2, -CHX_2, or -CX_3, wherein each X is
independently -Cl, -F, -Br, or -I.

[00175] In certain embodiments, the compound comprises at least one instance of -SHgR^D,
-SeR^D, or -TeR^D, wherein R^D is substituted or unsubstituted carbocyclyl, e.g., substituted or
unsubstituted C_{3} carbocyclyl, substituted or unsubstituted C_{4} carbocyclyl, substituted or
unsubstituted C_{6} carbocyclyl. In certain embodiments, R^D is carbocyclyl substituted with at least
one or more halogen atoms (e.g., 1, 2, 3, 4, 5, 6, or more -Br, -I, -F, or -Cl atoms), e.g., R^D is
substituted or unsubstituted C_{3} halocarbocyclyl, substituted or unsubstituted C_{4} halocarbocyclyl,
substituted or unsubstituted C_{6} halocarbocyclyl.

[00176] In certain embodiments, the compound comprises at least one instance of -SeR^D or
-TeR^D, wherein R^D is substituted or unsubstituted heterocyclyl, e.g., substituted or
unsubstituted 3-membered heterocyclyl, substituted or unsubstituted 4-membered heterocyclyl,
substituted or unsubstituted 5-membered heterocyclyl, or substituted or
unsubstituted 6-membered heterocyclyl. In certain embodiments, R^D is heterocyclyl
substituted with at least one or more halogen atoms (e.g., 1, 2, 3, 4, 5, 6, or more -Br, -I, -F,
or -Cl atoms), e.g., R^D is substituted or unsubstituted 3-membered haloheterocyclyl,
substituted or unsubstituted 4-membered haloheterocyclyl, substituted or unsubstituted 5-
membered haloheterocyclyl, or substituted or unsubstituted 6-membered haloheterocyclyl.

[00177] In certain embodiments, the compound comprises at least one instance of -SHgR^D,
-SeR^D, or -TeR^D, wherein R^D is substituted or unsubstituted aryl, e.g., substituted or
unsubstituted phenyl or substituted or unsubstituted naphthyl. In certain embodiments, \( R^D \) is ary1 substituted with at least one or more halogen atoms (e.g., 1, 2, 3, 4, 5, 6, or more -Br, -I, -F, or -Cl atoms), e.g., \( R^D \) is substituted or unsubstituted haloaryl. In certain embodiments, \( R^D \) is substituted or unsubstituted halophenyl, such as mono substituted halophenyl (e.g., ortho, meta, or para substituted with halogen atoms, substitution relative to the point of attachment of the halophenyl ring), disubstituted halophenyl (e.g., 1,2-, 1,3-, 1,4-, 1,5-, 2,3-, 2,4-, or 2,5- substituted with halogen atoms, substitution relative to the point of attachment of the halophenyl ring), or trisubstituted halophenyl (e.g., 1,3,5-, 1,2,3-, 1,2,4-, 1,2,5-, or 2,3,4- substituted with halogen atoms, substitution relative to the point of attachment of the halophenyl ring).

In certain embodiments, the compound comprises at least one instance of -SHgR \( ^D \), -SeR \( ^D \), or -TeR \( ^D \), wherein \( R^D \) is substituted or unsubstituted heteroaryl, e.g., substituted or unsubstituted 5-membered heteroaryl or substituted or unsubstituted 6-membered heteroaryl. In certain embodiments, \( R^D \) is heteroaryl substituted with at least one or more halogen atoms (e.g., 1, 2, 3, 4, 5, 6, or more -Br, -I, -F, or -Cl atoms), e.g., \( R^D \) is substituted or unsubstituted 5-membered haloheteroaryl or substituted or unsubstituted 6-membered haloheteroaryl.

In certain embodiments, the compound comprises at least one instance (e.g., 1, 2, 3, 4 or more instances) of mercury, e.g., for example, in certain embodiments, the compound comprises at least one instance of -SHg, -SO_2SHg, or -SHgR \( ^D \) (e.g., -SHgMe).

In certain embodiments, one or more heavy atoms (e.g., 1, 2, 3, 4, or more heavy atoms) are present in the sugar region of the compound. In certain embodiments, one or more heavy atoms (e.g., 1, 2, 3, or 4 heavy atoms) are present in the phosphate region of the compound. In certain embodiments, one or more heavy atoms (e.g., 1, 2, 3, or 4 heavy atoms) are present in the base region of the compound.

In the instance wherein the compound is an nucleotide of Formula (I), in certain embodiments, the compound may comprise heavy atoms in either the sugar, the phosphate, or the base region. In certain embodiments, the compound may comprise only heavy atoms in the sugar or phosphate region. In certain embodiments, the compound may comprise only heavy atoms in the base. In certain embodiments, the compound may comprises no heavy atoms in the base, and in that instance, heavy atoms are necessarily present in the sugar or phosphate region.

In the instance wherein the compound is a nucleic acid polymer comprising one or more units of Formula (IF), such as a compound of Formula (II), the 5' and/or 3' terminating group and/or one or more repeating units, e.g., 1 to 25,000 units, of the nucleic
acid polymer may comprise heavy atoms. In certain embodiments, the nucleic acid polymer comprises one or more instances of a heavy-atom labeled nucleotide in combination with one or more instances of an unlabeled nucleotide. In certain embodiments, there are multiple instances, e.g., 2 or more instances, of the same heavy-atom labeled nucleotide. In certain embodiments, each instance of a particular nucleotide, for example, A, G, T, C, or U, is replaced with a different heavy-atom labeled nucleotide as described herein. For example, in certain embodiments, each instance of A is replaced with a heavy-atom labeled nucleotide as described herein. In certain embodiments, each instance of G is replaced with a heavy-atom labeled nucleotide as described herein. In certain embodiments, each instance of T is replaced with a heavy-atom labeled nucleotide as described herein. In certain embodiments, each instance of C is replaced with a heavy-atom labeled nucleotide as described herein. In certain embodiments, each instance of U is replaced with a heavy-atom labeled nucleotide as described herein. In these instances, in certain embodiments, one of the heavy-atom labeled compounds is labeled in the sugar or phosphate region, and one of the heavy-atom labeled compounds is labeled in the base region, in order to better distinguish between A, G, T, C, or U. In certain embodiments, one of the heavy-atom labeled compounds is labeled in the sugar or phosphate region with one type of label, and one of the heavy-atom labeled compounds is labeled in the base region with a different type of label, in order to better distinguish between A, G, T, C, or U.

The Sugar Region and Groups G, G' and M

[00183] As generally described herein, the "sugar region" of the heavy-atom labeled compound of Formula (I), polymer of Formula (II), or unit of Formula (II'), may comprise a heavy atom, or may not comprise a heavy atom. If the sugar region does not comprise a heavy atom, the phosphate and/or base region of the heavy-atom labeled compound of Formula (I), polymer of Formula (II), or unit of Formula (II'), comprises a heavy atom.

"sugar region" of Formula (I)  "sugar region" of Formula (II)

[00184] As generally described herein, each instance of Gi is independently -O-, -S-, -Se-, -CH2-, or -NH-. In certain embodiments, at least one instance (e.g., 1, 2, 3, 4 or more instances, or each instance) of Gi is -O-. In certain embodiments, at least one instance (e.g.,
1. 2, 3, 4 or more instances, or each instance) of G₁ is -S-. In certain embodiments, at least one instance (e.g., 1, 2, 3, 4 or more instances, or each instance) of G₁ is -Se-. In certain embodiments, at least one instance (e.g., 1, 2, 3, 4 or more instances, or each instance) of Gᵢ is -CH₂-. In certain embodiments, at least one instance (e.g., 1, 2, 3, 4 or more instances, or each instance) of Gᵢ is -NH-.

[00185] As generally described herein, each instance of G₂ is independently hydrogen, halogen, -OR, -SR, -N(Rₐ)₂, -SHg, -S0₂SHg, -SHgR, -SeR or -TeR.

[00186] In certain embodiments, at least one instance (e.g., 1, 2, 3, 4 or more instances, or each instance) of G₂ is hydrogen.

[00187] In certain embodiments, at least one instance (e.g., 1, 2, 3, 4 or more instances, or each instance) of G₂ is halogen, i.e., G₂ is -Br, -I, -F, or -Cl.

[00188] In certain embodiments, at least one instance (e.g., 1, 2, 3, 4 or more instances, or each instance) of G₂ is -OR, wherein Rₐ is hydrogen, substituted or unsubstituted Ci₂₀alkyl, substituted or unsubstituted C₂₋₀alkenyl, substituted or unsubstituted C₂₋₀alkynyl, substituted or unsubstituted carbocycyl, substituted or unsubstituted heterocyclyl, substituted or unsubstituted aryl, or substituted or unsubstituted heteroaryl, or an oxygen protecting group.

[00189] In certain embodiments, G₂ is -OR and Rₐ is hydrogen, i.e., G₂ is -OH.

[00190] In certain embodiments, G₂ is -OR and Rₐ is an oxygen protecting group, as defined herein.

[00191] In certain embodiments, G₂ is -OR and Rₐ is substituted or unsubstituted Ci₂₀alkyl, e.g., G₂ is -OR and Rₐ is substituted or unsubstituted C₁₋₀alkyl, substituted or unsubstituted Ci₁₆alkyl, substituted or unsubstituted Ci₁₄alkyl, substituted or unsubstituted Ci₁₂alkyl, substituted or unsubstituted C₁₀alkyl, substituted or unsubstituted C₈alkyl, substituted or unsubstituted C₆alkyl, substituted or unsubstituted C₅alkyl, substituted or unsubstituted C₃alkyl, or substituted or unsubstituted C₂alkyl. In certain embodiments, G₂ is -OR and Rₐ is substituted or unsubstituted Ci₀alkyl. In certain embodiments, Rₐ is alkyl substituted with at least one or more halogen atoms (i.e., one or more -Br, -I, -F, or -Cl atoms), e.g., Rₐ is substituted or unsubstituted Ci₀haloalkyl, substituted or unsubstituted Ci₁₀haloalkyl, substituted or unsubstituted Ci₂₀haloalkyl, substituted or unsubstituted Ci₃₀haloalkyl, substituted or unsubstituted Ci₄₀haloalkyl, substituted or unsubstituted Ci₅₀haloalkyl, substituted or unsubstituted Ci₆₀haloalkyl, substituted or unsubstituted Ci₇₀haloalkyl, or substituted or unsubstituted Ci₈₀haloalkyl. In certain embodiments, G₂ is -OR and Rₐ is substituted or unsubstituted Ci₀alkyl.
certain embodiments, $G_2$ is -OR and $R^A$ is substituted or unsubstituted C$_1$, C$_2$, C$_3$, C$_4$, C$_5$, or C$_6$-haloalkyl. In certain embodiments, the haloalkyl is a perhaloalkyl group. In certain embodiments, $R^A$ is -C$X_m$, wherein X is halogen. In certain embodiments, $R^A$ is -CBr$_3$. CBr$_2$H, -CBrH$_2$, -CBr$_2$X, or -CBrX$_2$, wherein each instance of X is independently -Cl, -F, or -I. In certain embodiments, $R^A$ is -Cl$_3$, ClH$_2$, -CH$_2$X, or -ClX$_2$, wherein each instance of X is independently -Cl, -F, or -Br. In certain embodiments, $R^A$ is -CBr$_3$, -Cl$_3$, -CFClBr, or -CClBr.

[00192] In certain embodiments, $G_2$ is -OR and $R^A$ is substituted or unsubstituted C$_2$-alkenyl, e.g., $G_2$ is -OR and $R^A$ is substituted or unsubstituted C$_2$-alkenyl, substituted or unsubstituted C$_2$-alkenyl, substituted or unsubstituted C$_2$-alkenyl, substituted or unsubstituted C$_2$-alkenyl, substituted or unsubstituted C$_2$-alkenyl, substituted or unsubstituted C$_2$-alkenyl, substituted or unsubstituted C$_2$-alkenyl, substituted or unsubstituted C$_2$-alkenyl, substituted or unsubstituted C$_2$-alkenyl, substituted or unsubstituted C$_2$-alkenyl, or substituted or unsubstituted C$_2$-alkenyl. In certain embodiments, $G_2$ is -OR and $R^A$ is substituted or unsubstituted C$_2$, C$_3$, C$_4$, C$_5$, or C$_6$-alkenyl. In certain embodiments, $R^A$ is alkyl substituted with at least one or more halogen atoms (e.g., 1, 2, 3, 4, 5, 6, or more -Br, -I, -F, or -Cl atoms), e.g., $R^A$ is substituted or unsubstituted C$_2$-haloalkenyl, substituted or unsubstituted C$_2$-haloalkenyl, substituted or unsubstituted C$_2$-haloalkenyl, substituted or unsubstituted C$_2$-haloalkenyl, substituted or unsubstituted C$_2$-haloalkenyl, substituted or unsubstituted C$_2$-haloalkenyl, substituted or unsubstituted C$_2$-haloalkenyl, substituted or unsubstituted C$_2$-haloalkenyl, substituted or unsubstituted C$_2$-haloalkenyl, or substituted or unsubstituted C$_2$-haloalkenyl. In certain embodiments, the haloalkenyl is a perhaloalkenyl group. In certain embodiments, $G_2$ is -OR and $R^A$ is substituted or unsubstituted C$_2$, C$_3$, C$_4$, C$_5$, or C$_6$-haloalkenyl. In certain embodiments, $R^A$ is -CH$_2$CX=$CX_2$, -CH$_2$CH=$CX_2$, -CH$_2$CX=$CHX$, -CH$_2$CH=$CHX$, -CH$_2$CX=$CH_2$, wherein each instance of X is independently -Cl, -F, -Br, or -I.

[00193] In certain embodiments, $G_2$ is -OR and $R^A$ is substituted or unsubstituted C$_2$-alkynyl, e.g., $G_2$ is -OR and $R^A$ is substituted or unsubstituted C$_2$-alkynyl, substituted or unsubstituted C$_2$-alkynyl, substituted or unsubstituted C$_2$-alkynyl, substituted or unsubstituted C$_2$-alkynyl, substituted or unsubstituted C$_2$-alkynyl, substituted or unsubstituted C$_2$-alkynyl, substituted or unsubstituted C$_2$-alkynyl, substituted or unsubstituted C$_2$-alkynyl, or substituted or unsubstituted C$_2$-alkynyl. In certain embodiments, $G_2$ is -OR and $R^A$ is substituted or unsubstituted C$_2$, C$_3$, C$_4$, C$_5$, or C$_6$-alkynyl. In certain embodiments, $R^A$ is alkynyl substituted with at least one or more halogen atoms (e.g., 1, 2, 3, 4, 5, 6, or more -Br, -I, -F, or -Cl atoms), e.g., $R^A$ is substituted or
unsubstituted C$_2$ haloalkynyl, substituted or unsubstituted C$_2$ haloalkynyl, substituted or unsubstituted C$_2$ haloalkynyl, substituted or unsubstituted C$_2$ haloalkynyl, substituted or unsubstituted C$_2$ haloalkynyl, substituted or unsubstituted C$_2$ haloalkynyl, substituted or unsubstituted C$_2$ haloalkynyl, substituted or unsubstituted C$_2$ haloalkynyl, or substituted or unsubstituted C$_2$ haloalkynyl. In certain embodiments, the haloalkynyl is a perhaloalkynyl group. In certain embodiments, G$_2$ is -OR$^A$ and R$^A$ is substituted or unsubstituted carbcycl, e.g., substituted or unsubstituted C$_3$ carbcycl, substituted or unsubstituted C$_4$ carbcycl, substituted or unsubstituted C$_5$ carbcycl, or substituted or unsubstituted C$_6$ carbcycl. In certain embodiments, R$^A$ is carbcycl substituted with at least one or more halogen atoms (e.g., 1, 2, 3, 4, 5, 6, or more -Br, -I, -F, or -Cl atoms), e.g., R$^A$ is substituted or unsubstituted C$_3$ halocarbcycl, substituted or unsubstituted C$_4$ halocarbcycl, substituted or unsubstituted C$_5$ halocarbcycl, or substituted or unsubstituted C$_6$ halocarbcycl.

[00195] In certain embodiments, G$_2$ is -OR$^A$ and R$^A$ is substituted or unsubstituted heterocycl, e.g., substituted or unsubstituted 3-membered heterocycl, substituted or unsubstituted 4-membered heterocycl, substituted or unsubstituted 3-membered heterocycl, or substituted or unsubstituted 6-membered heterocycl. In certain embodiments, R$^A$ is heterocycl substituted with at least one or more halogen atoms (e.g., 1, 2, 3, 4, 5, 6, or more -Br, -I, -F, or -Cl atoms), e.g., R$^A$ is substituted or unsubstituted 3-membered haloheterocycl, substituted or unsubstituted 4-membered haloheterocycl, substituted or unsubstituted 5-membered haloheterocycl, or substituted or unsubstituted 6-membered haloheterocycl.

[00196] In certain embodiments, G$_2$ is -OR$^A$ and R$^A$ is substituted or unsubstituted aryl, e.g., substituted or unsubstituted phenyl or substituted or unsubstituted naphthyl. In certain embodiments, R$^A$ is aryl substituted with at least one or more halogen atoms (e.g., 1, 2, 3, 4, 5, 6, or more -Br, -I, -F, or -Cl atoms), e.g., R$^A$ is substituted or unsubstituted haloaryl. In certain embodiments, R$^A$ is substituted or unsubstituted halophenyl, such as mono substituted halophenyl (e.g., ortho, meta, or para-substituted with halogen atoms, substitution relative to the point of attachment of the halophenyl ring), disubstituted halophenyl (e.g., 1,2-, 1,3-, 1,4-, 1,5-, 2,3-, 2,4-, or 2,5- substituted with halogen atoms, substitution relative to the point of
attachment of the halophenyl ring), or trisubstituted halophenyl (e.g., 1,3,5-, 1,2,3-, 1,2,4-, 1,2,5-, or 2,3,4-substituted with halogen atoms, substitution relative to the point of attachment of the halophenyl ring).

[00197] In certain embodiments, G₂ is -ORₐ and Rₐ is substituted or unsubstituted heteroaryl, e.g., substituted or unsubstituted 5-membered heteroaryl or substituted or unsubstituted 6-membered heteroaryl. In certain embodiments, Rₐ is heteroaryl substituted with at least one or more halogen atoms (e.g., 1, 2, 3, 4, 5, 6, or more -Br, -I, -F, or -Cl atoms), e.g., Rₐ is substituted or unsubstituted 5-membered haloheteroaryl or substituted or unsubstituted 6-membered halo heteroaryl.

[00198] In certain embodiments, G₂ is -SRₐ and Rₐ is hydrogen, i.e., G₂ is -SH.

[00199] In certain embodiments, G₂ is -SRₐ and Rₐ is a sulfur protecting group, as defined herein.

[00200] In certain embodiments, G₂ is -SRₐ and Rₐ is substituted or unsubstituted C₁₋₂alkyl, e.g., G₂ is -SRₐ and Rₐ is substituted or unsubstituted C₁₋₈alkyl, substituted or unsubstituted C₅₋₁₄alkyl, substituted or unsubstituted C₁₋₁₂alkyl, substituted or unsubstituted C₁₋₁₀alkyl, substituted or unsubstituted C₁₋₈alkyl, substituted or unsubstituted C₅₋₁₄alkyl, substituted or unsubstituted C₁₋₁₀alkyl, or substituted or unsubstituted C₁₋₂alkyl. In certain embodiments, G₂ is -SRₐ and Rₐ is substituted or unsubstituted C₁₋₂haloalkyl, substituted or unsubstituted C₁₋₁₀haloalkyl, substituted or unsubstituted C₁₋₈haloalkyl, substituted or unsubstituted C₁₋₁₂haloalkyl, substituted or unsubstituted C₁₋₁₀haloalkyl, substituted or unsubstituted C₁₋₈haloalkyl, substituted or unsubstituted C₁₋₁₂haloalkyl, substituted or unsubstituted C₁₋₁₀haloalkyl, substituted or unsubstituted C₁₋₈haloalkyl, substituted or unsubstituted C₁₋₂haloalkyl, or substituted or unsubstituted C₁₋₂haloalkyl. In certain embodiments, G₂ is -SRₐ and Rₐ is substituted or unsubstituted C₁₋₂haloalkyl. In certain embodiments, G₂ is -SRₐ and Rₐ is substituted or unsubstituted C₁₋₈haloalkyl, substituted or unsubstituted C₁₋₁₂haloalkyl, substituted or unsubstituted C₁₋₁₀haloalkyl, substituted or unsubstituted C₁₋₈haloalkyl, substituted or unsubstituted C₁₋₁₂haloalkyl, substituted or unsubstituted C₁₋₁₀haloalkyl, substituted or unsubstituted C₁₋₈haloalkyl, substituted or unsubstituted C₁₋₂haloalkyl, or substituted or unsubstituted C₁₋₂haloalkyl. In certain embodiments, G₂ is -SRₐ and Rₐ is substituted or unsubstituted C₁₋₈haloalkyl. In certain embodiments, the haloalkyl is a perhaloalkyl group. In certain embodiments, Rₐ is -CX₃, wherein X is halogen. In certain embodiments, Rₐ is -CBr₃, CBr₂H, -CBrH, -CBr₂X, or -CBrX₂, wherein each instance of X is independently -Cl, -F, or -I. In certain embodiments, Rₐ is -Cl₃, Cl₂H, -ClH₂, -Cl₂X, or -ClX₂, wherein each instance of X is independently -Cl, -F, or -Br. In certain embodiments, Rₐ is -CBr₃, -Cl₃, -CFCIBr, or -CClBrl.
In certain embodiments, $G_2$ is -SR and $R^A$ is substituted or unsubstituted C$_{2-6}$alkenyl, e.g., $G_2$ is -SR and $R^A$ is substituted or unsubstituted C$_2$-alkenyl, substituted or unsubstituted C$_{2-10}$alkenyl, substituted or unsubstituted C$_{2-14}$alkenyl, substituted or unsubstituted C$_{2-12}$alkenyl, substituted or unsubstituted C$_{2-15}$alkenyl, substituted or unsubstituted C$_{2-9}$alkenyl, substituted or unsubstituted C$_{2-6}$alkenyl, substituted or unsubstituted C$_{2-3}$alkenyl, or substituted or unsubstituted C$_{2-3}$haloalkenyl. In certain embodiments, $G_2$ is -SR and $R^A$ is substituted or unsubstituted C$_2$, C$_3$, C$_4$, C$_5$, or C$_6$-alkenyl.

In certain embodiments, $R^A$ is alkyl substituted with at least one or more halogen atoms (e.g., 1, 2, 3, 4, 5, 6, or more -Br, -I, -F, or -Cl atoms), e.g., $R^A$ is substituted or unsubstituted C$_2$-haloalkenyl, substituted or unsubstituted C$_{2-9}$haloalkenyl, substituted or unsubstituted C$_{2-14}$haloalkenyl, substituted or unsubstituted C$_{2-12}$haloalkenyl, substituted or unsubstituted C$_{2-15}$haloalkenyl, substituted or unsubstituted C$_{2-9}$haloalkenyl, substituted or unsubstituted C$_{2-6}$haloalkenyl, substituted or unsubstituted C$_{2-3}$haloalkenyl, or substituted or unsubstituted C$_{2-3}$haloalkenyl. In certain embodiments, the haloalkenyl is a perhaloalkenyl group. In certain embodiments, $G_2$ is -SR and $R^A$ is substituted or unsubstituted C$_2$-alkynyl, e.g., $G_2$ is -SR and $R^A$ is substituted or unsubstituted C$_{2-14}$alkynyl, substituted or unsubstituted C$_{2-12}$alkynyl, substituted or unsubstituted C$_{2-15}$alkynyl, substituted or unsubstituted C$_{2-9}$alkynyl, substituted or unsubstituted C$_{2-6}$alkynyl, substituted or unsubstituted C$_{2-3}$alkynyl, or substituted or unsubstituted C$_{2-3}$alkynyl. In certain embodiments, $G_2$ is -SR and $R^A$ is substituted or unsubstituted C$_2$, C$_3$, C$_4$, C$_5$, or C$_6$-alkynyl.

In certain embodiments, $R^A$ is alkynyl substituted with at least one or more halogen atoms (e.g., 1, 2, 3, 4, 5, 6, or more -Br, -I, -F, or -Cl atoms), e.g., $R^A$ is substituted or unsubstituted C$_2$-haloalkynyl, substituted or unsubstituted C$_{2-9}$haloalkynyl, substituted or unsubstituted C$_{2-14}$haloalkynyl, substituted or unsubstituted C$_{2-12}$haloalkynyl, substituted or unsubstituted C$_{2-15}$haloalkynyl, substituted or unsubstituted C$_{2-9}$haloalkynyl, substituted or unsubstituted C$_{2-6}$haloalkynyl, substituted or unsubstituted C$_{2-3}$haloalkynyl, or substituted or unsubstituted C$_{2-3}$haloalkynyl. In certain embodiments, the haloalkynyl is a perhaloalkynyl group. In certain embodiments, $G_2$ is -SR and $R^A$ is substituted or unsubstituted C$_2$, C$_3$, C$_4$, C$_5$, or C$_6$-haloalkynyl. In certain embodiments, $R^A$ is
-CH₂X₁, -CHX₂, or -CX₃, wherein each X is independently -Cl, -F, -Br, or -I.

[00203] In certain embodiments, G₂ is -SR and R₈ is substituted or unsubstituted carbocyclyl, e.g., substituted or unsubstituted C₃ carbocyclyl, substituted or unsubstituted C₄ carbocyclyl, substituted or unsubstituted C₅ carbocyclyl, or substituted or unsubstituted C₆ carbocyclyl. In certain embodiments, R₈ is carbocyclyl substituted with at least one or more halogen atoms (e.g., 1, 2, 3, 4, 5, 6, or more -Br, -I, -F, or -Cl atoms), e.g., R₈ is substituted or unsubstituted C₃ halocarbocyclyl, substituted or unsubstituted C₄ halocarbocyclyl, substituted or unsubstituted C₅ halocarbocyclyl, or substituted or unsubstituted C₆ halocarbocyclyl.

[00204] In certain embodiments, G₂ is -SR and R₈ is substituted or unsubstituted heterocyclyl, e.g., substituted or unsubstituted 3-membered heterocyclyl, substituted or unsubstituted 4-membered heterocyclyl, substituted or unsubstituted 5-membered heterocyclyl, or substituted or unsubstituted 6-membered heterocyclyl. In certain embodiments, R₈ is heterocyclyl substituted with at least one or more halogen atoms (e.g., 1, 2, 3, 4, 5, 6, or more -Br, -I, -F, or -Cl atoms), e.g., R₈ is substituted or unsubstituted 3-membered haloheterocyclyl, substituted or unsubstituted 4-membered haloheterocyclyl, substituted or unsubstituted 5-membered haloheterocyclyl, or substituted or unsubstituted 6-membered haloheterocyclyl.

[00205] In certain embodiments, G₂ is -SR and R₈ is substituted or unsubstituted aryl, e.g., substituted or unsubstituted phenyl or substituted or unsubstituted naphthyl. In certain embodiments, R₈ is aryl substituted with at least one or more halogen atoms (e.g., 1, 2, 3, 4, 5, 6, or more -Br, -I, -F, or -Cl atoms), e.g., R₈ is substituted or unsubstituted haloaryl. In certain embodiments, R₈ is substituted or unsubstituted halophenyl, such as mono-substituted halophenyl (e.g., ortho, meta, or para-substituted) with halogen atoms, substitution relative to the point of attachment of the halophenyl ring), disubstituted halophenyl (e.g., 1,2-., 1,3-., 1,4-., 1,5-., 2,3-., 2,4-., or 2,5- substituted with halogen atoms, substitution relative to the point of attachment of the halophenyl ring), or trisubstituted halophenyl (e.g., 1,3,5-., 1,2,3-., 1,2,4-., 1,2,5-., or 2,3,4- substituted with halogen atoms, substitution relative to the point of attachment of the halophenyl ring).

[00206] In certain embodiments, G₂ is -SR and R₈ is substituted or unsubstituted heteroaryl, e.g., substituted or unsubstituted 5-membered heteroaryl or substituted or unsubstituted 6-membered heteroaryl. In certain embodiments, R₈ is heteroaryl substituted with at least one or more halogen atoms (e.g., 1, 2, 3, 4, 5, 6, or more -Br, -I, -F, or -Cl
atoms), e.g., $R^A$ is substituted or unsubstituted 5-membered haloheteroaryl or substituted or unsubstituted 6-membered haloheteroaryl.

[00207] In certain embodiments, $G_2$ is $-N(R^A)_2$ and at least one $R^A$ is hydrogen, i.e., $G_2$ is $-NHR^A$ or $-NH_2$.

[00208] In certain embodiments, $G_2$ is $-N(R^A)_2$ and at least one $R^A$ is a nitrogen protecting group, as defined herein.

[00209] In certain embodiments, $G_2$ is $-N(R^A)_2$ and at least one $R^A$ is substituted or unsubstituted $C_{1-20}$alkyl, e.g., $G_2$ is $-N(R^A)_2$ and at least one $R^A$ is substituted or unsubstituted $C_1$ i$_k$alkyl, substituted or unsubstituted C$_3$alkyl, substituted or unsubstituted C$_2$-alkenyl, substituted or unsubstituted C$_2$-alkenyl, substituted or unsubstituted C$_2$-alkenyl, substituted or unsubstituted C$_2$-alkenyl, or substituted or unsubstituted C$_2$-alkenyl. In certain embodiments, at least one is $R^A$ is alkyl substituted with at least one or more halogen atoms (i.e., one or more -Br, -I, -F, or -Cl atoms), e.g., at least one is $R^A$ is substituted or unsubstituted $C_{1-2}$haloalkyl, substituted or unsubstituted $C_1$ i$_k$haloalkyl, substituted or unsubstituted $C_1$ i$_k$haloalkyl, substituted or unsubstituted $C_1$ i$_k$haloalkyl, substituted or unsubstituted $C_1$ i$_k$haloalkyl, substituted or unsubstituted $C_1$ i$_k$haloalkyl, or substituted or unsubstituted $C_1$ i$_k$haloalkyl, substituted or unsubstituted $C_1$ i$_k$haloalkyl, or substituted or unsubstituted $C_1$ i$_k$haloalkyl, or substituted or unsubstituted $C_1$ i$_k$haloalkyl. In certain embodiments, the haloalkyl is a perhaloalkyl group. In certain embodiments, at least one $R^A$ is $-CX_3$, wherein $X$ is halogen. In certain embodiments, at least one $R^A$ is $-CF_{2}Br$, $-CF_{3}H$, $-CF_{2}H$, $-CF_{2}X$, or $-CF_{2}X_2$, wherein each instance of $X$ is independently -CI, -F, or -I. In certain embodiments, at least one $R^A$ is $-Cl_3$, $-Cl_2H$, $-Cl_2H$, $-CIH_2$, $-Cl_2X$, or $-CIH_2X$, wherein each instance of $X$ is independently -CI, -F, or -Br. In certain embodiments, at least one $R^A$ is $-CF_{3}Br$, $-Cl_3$, $-CF_{2}Br$, or $-ClF_{2}Br$.

[00210] In certain embodiments, $G_2$ is $-N(R^A)_2$ and at least one $R^A$ is substituted or unsubstituted $C_{2-2}$alkenyl, e.g., $G_2$ is $-N(R^A)_2$ and at least one $R^A$ is substituted or unsubstituted $C_{2-16}$alkenyl, substituted or unsubstituted $C_{2-8}$alkenyl, substituted or unsubstituted $C_{2-7}$alkenyl, substituted or unsubstituted $C_{2-12}$alkenyl, substituted or unsubstituted $C_{2-7}$alkenyl, substituted or unsubstituted $C_{2-6}$alkenyl, substituted or unsubstituted $C_{2-4}$alkenyl, or substituted or
unsubstituted C$_2$-alkenyl. In certain embodiments, G$_2$ is -N(R$_A$)$_2$ and at least one R$_A$ is substituted or unsubstituted C$_2$, C$_3$, C$_4$, C$_5$, or C$_6$-alkenyl. In certain embodiments, at least one R$_A$ is alkenyl substituted with at least one or more halogen atoms (e.g., 1, 2, 3, 4, 5, 6, or more -Br, -I, -F, or -Cl atoms), e.g., R$_A$ is substituted or unsubstituted C$_2$-haloalkenyl, substituted or unsubstituted C$_2$-alkenyl, substituted or unsubstituted C$_2$-alkenyl, substituted or unsubstituted C$_2$-alkenyl, substituted or unsubstituted C$_2$-alkenyl, substituted or unsubstituted C$_2$-alkenyl, substituted or unsubstituted C$_2$-alkenyl, substituted or unsubstituted C$_2$-alkenyl, substituted or unsubstituted C$_2$-alkenyl, substituted or unsubstituted C$_2$-alkenyl, or substituted or unsubstituted C$_2$-alkenyl. In certain embodiments, the alkenyl is a perhaloalkenyl group. In certain embodiments, G$_2$ is -N(R$_A$)$_2$ and at least one R$_A$ is substituted or unsubstituted C$_2$-alkynyl, e.g., G$_2$ is -N(R$_A$)$_2$ and at least one R$_A$ is substituted or unsubstituted C$_2$-alkynyl, substituted or unsubstituted C$_2$-alkynyl, substituted or unsubstituted C$_2$-alkynyl, substituted or unsubstituted C$_2$-alkynyl, substituted or unsubstituted C$_2$-alkynyl, substituted or unsubstituted C$_2$-alkynyl, substituted or unsubstituted C$_2$-alkynyl, substituted or unsubstituted C$_2$-alkynyl, or unsubstituted C$_2$-alkynyl. In certain embodiments, G$_2$ is -N(R$_A$)$_2$ and at least one R$_A$ is substituted or unsubstituted C$_2$, C$_3$, C$_4$, C$_5$, or C$_6$-alkynyl. In certain embodiments, at least one R$_A$ is alkynyl substituted with at least one or more halogen atoms (e.g., 1, 2, 3, 4, 5, 6, or more -Br, -I, -F, or -Cl atoms), e.g., at least one R$_A$ is substituted or unsubstituted C$_2$-haloalkynyl, substituted or unsubstituted C$_2$-haloalkynyl, substituted or unsubstituted C$_2$-haloalkynyl, substituted or unsubstituted C$_2$-haloalkynyl, substituted or unsubstituted C$_2$-haloalkynyl, substituted or unsubstituted C$_2$-haloalkynyl, substituted or unsubstituted C$_2$-haloalkynyl, substituted or unsubstituted C$_2$-haloalkynyl, substituted or unsubstituted C$_2$-haloalkynyl, or substituted or unsubstituted C$_2$-haloalkynyl. In certain embodiments, the haloalkynyl is a perhaloalkynyl group. In certain embodiments, G$_2$ is -N(R$_A$)$_2$ and at least one R$_A$ is substituted or unsubstituted C$_2$, C$_3$, C$_4$, C$_5$, or C$_6$-haloalkynyl. In certain embodiments, at least one R$_A$ is alkynyl substituted with at least one or more halogen atoms (e.g., 1, 2, 3, 4, 5, 6, or more -Br, -I, -F, or -Cl atoms), e.g., at least one R$_A$ is substituted or unsubstituted C$_2$-haloalkynyl, substituted or unsubstituted C$_2$-haloalkynyl, substituted or unsubstituted C$_2$-haloalkynyl, substituted or unsubstituted C$_2$-haloalkynyl, substituted or unsubstituted C$_2$-haloalkynyl, substituted or unsubstituted C$_2$-haloalkynyl, substituted or unsubstituted C$_2$-haloalkynyl, substituted or unsubstituted C$_2$-haloalkynyl, or unsubstituted C$_2$-haloalkynyl. In certain embodiments, the haloalkynyl is a perhaloalkynyl group. In certain embodiments, G$_2$ is -N(R$_A$)$_2$ and at least one R$_A$ is substituted or unsubstituted carbocyclyl, e.g., substituted or unsubstituted C$_2$carbocyclyl, substituted or
unsubstituted C₄ carbocycyl, substituted or unsubstituted C₅ carbocycyl, or substituted or unsubstituted C₆ carbocycyl. In certain embodiments, at least one Rᴬ is carbocycyl substituted with at least one or more halogen atoms (e.g., 1, 2, 3, 4, 5, or more -Br, -I, -F, or -Cl atoms), e.g., at least one Rᴬ is substituted or unsubstituted C-halocarbocycyl, substituted or unsubstituted C₅ halocarbocycyl, substituted or unsubstituted C₆ halocarbocycyl, or substituted or unsubstituted C-aryl.

[00213] In certain embodiments, G₂ is -N(Rᴬ)₂ and at least one Rᴬ is substituted or unsubstituted heterocyclyl, e.g., substituted or unsubstituted 3-membered heterocyclyl, substituted or unsubstituted 4-membered heterocyclyl, substituted or unsubstituted 5-membered heterocyclyl, or substituted or unsubstituted 6-membered heterocyclyl. In certain embodiments, at least one Rᴬ is heterocyclyl substituted with at least one or more halogen atoms (e.g., 1, 2, 3, 4, 5, 6, or more -Br, -I, -F, or -Cl atoms), e.g., at least one Rᴬ is substituted or unsubstituted 3-membered haloheterocyclyl, substituted or unsubstituted 4-membered haloheterocyclyl, substituted or unsubstituted 5-membered haloheterocyclyl, or substituted or unsubstituted 6-membered haloheterocyclyl.

[00214] In certain embodiments, G₂ is -N(Rᴬ)₂ and at least one Rᴬ is substituted or unsubstituted aryl, e.g., substituted or unsubstituted phenyl or substituted or unsubstituted naphthyl. In certain embodiments, at least one Rᴬ is aryl substituted with at least one or more halogen atoms (e.g., 1, 2, 3, 4, 5, 6, or more -Br, -I, -F, or -Cl atoms), e.g., at least one Rᴬ is substituted or unsubstituted haloaryl. In certain embodiments, at least one Rᴬ is substituted or unsubstituted halophenyl, such as monosubstituted halophenyl (e.g., ortho, meta, or para-substituted with halogen atoms, substitution relative to the point of attachment of the halophenyl ring), disubstituted halophenyl (e.g., 1,2-, 1,3-, 1,4-, 1,5-, 2,3-, 2,4-, or 2,5-substituted with halogen atoms, substitution relative to the point of attachment of the halophenyl ring), or trisubstituted halophenyl (e.g., 1,3,5-, 1,2,3-, 1,2,4-, 1,2,5-, or 2,3,4-substituted with halogen atoms, substitution relative to the point of attachment of the halophenyl ring).

[00215] In certain embodiments, G₂ is -N(Rᴬ)₂ and at least one Rᴬ is substituted or unsubstituted heteroaryl, e.g., substituted or unsubstituted 5-membered heteroaryl or substituted or unsubstituted 6-membered heteroaryl. In certain embodiments, at least one Rᴬ is heteroaryl substituted with at least one or more halogen atoms (e.g., 1, 2, 3, 4, 5, 6, or more -Br, -I, -F, or -Cl atoms), e.g., at least one Rᴬ is substituted or unsubstituted 5-membered haloheteroaryl or substituted or unsubstituted 6-membered haloheteroaryl.
[00216] In certain embodiments, G₂ is -N(R¹)₂, and two R groups are joined to form a substituted or unsubstituted heterocyclic or substituted or unsubstituted heteroaryl ring, e.g., a 5- to 6-membered substituted or unsubstituted heterocyclic or substituted or unsubstituted heteroaryl ring.

[00217] In certain embodiments, at least one instance (e.g., 1, 2, 3, 4 or more instances, or each instance) of G₂ is -SHg or -SO₂SHg, or -SHgR²D, wherein R²D is as defined herein.

[00218] In certain embodiments, at least one instance (e.g., 1, 2, 3, 4 or more instances, or each instance) of G₂ is -SeR²D, wherein R²D is as defined herein.

[00219] In certain embodiments, at least one instance (e.g., 1, 2, 3, 4 or more instances, or each instance) of G₂ is -TeR²D, wherein R²D is as defined herein.

[00220] As generally described herein, each instance of M¹ is independently -O-, -S-, -NH-, -Se-, or -C(R³M)₂⁻, wherein each instance of R³M is independently hydrogen or halogen. In certain embodiments, at least one instance (e.g., 1, 2, 3, 4 or more instances, or each instance) of M¹ is -O-. In certain embodiments, at least one instance (e.g., 1, 2, 3, 4 or more instances, or each instance) of M¹ is -S-. In certain embodiments, at least one instance (e.g., 1, 2, 3, 4 or more instances, or each instance) of M¹ is -NH-. In certain embodiments, at least one instance (e.g., 1, 2, 3, 4 or more instances, or each instance) of M¹ is -Se-. In certain embodiments, at least one instance (e.g., 1, 2, 3, 4 or more instances, or each instance) of M¹ is -C(R³M)₂⁻, wherein each instance of R³M is independently hydrogen or halogen. In certain embodiments, each instance of R³M is hydrogen. In certain embodiments, at least one instance of R³M is halogen, e.g., -Br, -I, -F, or -Cl.

[00221] Various combinations of the above described embodiments of the "sugar regions" are further contemplated herein.

[00222] For example, in certain embodiments, each instance of Gi is O to provide a compound of Formula (I-b), polymer of Formula (II-b), or unit of Formula (II'-b):

![Diagram of structures](image-url)
In certain embodiments, $G_2$ is hydrogen. In certain embodiments, $G_2$ is $\text{-SHgR}_D$ (e.g., $\text{-SHgMe}$), $\text{-SHg}$, or $\text{-SO}_2\text{SHg}$. In certain embodiments, $G_2$ is $\text{-SeR}_D$, e.g., $\text{-SeCX}_3$, wherein $X$ is halogen. In certain embodiments, $G_2$ is $\text{-TeR}_D$, e.g., $\text{-TeCX}_3$, wherein $X$ is halogen.

In certain embodiments, each instance of $G_1$ and $M^1$ is $O$ to provide a compound of Formula (I-c), polymer of Formula (II-c), or unit of Formula (II'-c):

or salt thereof. In certain embodiments, $G_2$ is hydrogen. In certain embodiments, $G_2$ is $\text{-SHgR}_D$ (e.g., $\text{-SHgMe}$), $\text{-SHg}$, or $\text{-SO}_2\text{SHg}$. In certain embodiments, $G_2$ is $\text{-SeR}_D$, e.g., $\text{-SeCX}_3$, wherein $X$ is halogen. In certain embodiments, $G_2$ is $\text{-TeR}_D$, e.g., $\text{-TeCX}_3$, wherein $X$ is halogen. In certain embodiments, $G_2$ is $\text{Se-CBr}_3$ or $\text{-TeBr}_3$.

In certain embodiments, each instance of $G_1$ and $M^1$ is $O$ to provide a compound of Formula (I-d), polymer of Formula (II-d), or unit of Formula (II'-d) with the specified stereochemistry:
or the enantiomer thereof and/or salt thereof. In certain embodiments, G₂ is hydrogen. In
certain embodiments, G₂ is -SHgRᵗ (e.g., -SHgMe), -SHg, or -S0₂ SHg. In certain
embodiments, G₂ is -SeRᵗ, e.g., -SeCX₃, wherein X is halogen. In certain embodiments, G₂
is -TeRᵗ, e.g., -TeCX₃, wherein X is halogen.

The Phosphate Region and G³

[00225] As generally described herein, each instance of G₃ independently is hydrogen,
substituted or unsubstituted Cᵣalkyl, substituted or unsubstituted C₂₋₂ oalkenyl, substituted
or unsubstituted C₂₋₂ oalkynyl, substituted or unsubstituted carbocycyl, substituted or
unsubstituted heterocyclyl, substituted or unsubstituted aryl, or substituted or unsubstituted
heteroaryl, or may be a monophosphate, diphosphate, or triphosphate group.

[00226] In certain embodiments, at least one instance of G₃ is hydrogen.

[00227] In certain embodiments, at least one instance of G₃ is substituted or unsubstituted
C₁₋₂ oalkyl, e.g., substituted or unsubstituted C₁₋₁₆ oalkyl, substituted or unsubstituted C₁₋₁₆
alkyl, substituted or unsubstituted C₁₋₁₄ oalkyl, substituted or unsubstituted C₁₋₁₆ oalkyl,
substituted or unsubstituted C₁₋₁₄ oalkyl, substituted or unsubstituted C₁₋₁₆ oalkyl, substituted or
unsubstituted C₁₋₁₆ oalkyl, substituted or unsubstituted C₁₋₁₆ oalkyl, substituted or unsubstituted
C^alkyl, or substituted or unsubstituted C^alkyl. In certain embodiments, at least one instance of G_3 is substituted or unsubstituted C_1, C_2, C_3, C_4, C_5, or C_6-alkyl. In certain embodiments, at least one instance of G_3 is alkyl substituted with at least one or more halogen atoms (i.e., one or more -Br, -I, -F, or -Cl atoms), e.g., at least one instance of G_3 is substituted or unsubstituted C^haloalkyl, substituted or unsubstituted C_i,haloalkyl, substituted or unsubstituted C_i_alkoalkyl, substituted or unsubstituted C_i_alkoalkyl, substituted or unsubstituted C_i_alkoalkyl, substituted or unsubstituted C_i_alkoalkyl, substituted or unsubstituted C_i_alkoalkyl, substituted or unsubstituted C_i_alkoalkyl, substituted or unsubstituted C_i_alkoalkyl, substituted or unsubstituted C_i_alkoalkyl, or substituted or unsubstituted C_i_alkoalkyl. In certain embodiments, at least one instance of G_3 is substituted or unsubstituted C_2, C_3, C_4, C_5, or C_6-alkenyl. In certain embodiments, at least one instance of G_3 is alkyl substituted with at least one or more halogen atoms (e.g., 1, 2, 3, 4, 5, 6, or more -Br, -I, -F, or -Cl atoms), e.g., at least one instance of G_3 is substituted or unsubstituted C_2, C_3, C_4, C_5, or C_6-alkenyl. In certain embodiments, at least one instance of G_3 is alkyl substituted with at least one or more halogen atoms (e.g., 1, 2, 3, 4, 5, 6, or more -Br, -I, -F, or -Cl atoms), e.g., at least one instance of G_3 is substituted or unsubstituted C_2, C_3, C_4, C_5, or C_6-alkenyl. In certain embodiments, at least one instance of G_3 is substituted or unsubstituted C_2, C_3, C_4, C_5, or C_6-alkenyl. In certain embodiments, at least one instance of G_3 is substituted or unsubstituted C_2, C_3, C_4, C_5, or C_6-alkenyl. In certain embodiments, at least one instance of G_3 is substituted or unsubstituted C_2, C_3, C_4, C_5, or C_6-alkenyl. In certain embodiments, at least one instance of G_3 is substituted or unsubstituted C_2, C_3, C_4, C_5, or C_6-alkenyl.
CH₂CX=CX₂, -CH₂CH=CX₂, -CH₂CX=CHX, -CH₂CH=CHX, -CH₂CX=CH₂, wherein each instance of X is independently -Cl, -F, -Br, or -I.

[00229] In certain embodiments, at least one instance of G₃ is substituted or unsubstituted C₂₋₅alkynyl, e.g., at least one instance of G₃ is substituted or unsubstituted C₂₋₆alkynyl, substituted or unsubstituted C₂₋₅alkynyl, substituted or unsubstituted C₂₋₆alkynyl, substituted or unsubstituted C₂₋₅alkynyl, substituted or unsubstituted C₂₋₆alkynyl, substituted or unsubstituted C₂₋₅alkynyl, substituted or unsubstituted C₂₋₅alkynyl, or substituted or unsubstituted C₂₋₅alkynyl. In certain embodiments, at least one instance of G₃ is substituted or unsubstituted C₂, C₃, C₄, C₅, or C₆-alkynyl. In certain embodiments, at least one instance of G₃ is alkynyl substituted with at least one or more halogen atoms (e.g., 1, 2, 3, 4, 5, 6, or more -Br, -I, -F, or -Cl atoms), e.g., at least one instance of G₃ is substituted or unsubstituted C₂₋₅haloalkynyl, substituted or unsubstituted C₂₋₅haloalkynyl, substituted or unsubstituted C₂₋₅haloalkynyl, substituted or unsubstituted C₂₋₅haloalkynyl, substituted or unsubstituted C₂₋₅haloalkynyl, substituted or unsubstituted C₂₋₅haloalkynyl, substituted or unsubstituted C₂₋₅haloalkynyl, substituted or unsubstituted C₂₋₅haloalkynyl, or substituted or unsubstituted C₂₋₅haloalkynyl. In certain embodiments, the haloalkynyl is a perhaloalkynyl group. In certain embodiments, at least one instance of G₃ is substituted or unsubstituted C₂, C₃, C₄, C₅, or C₆-haloalkynyl. In certain embodiments, at least one instance of G₃ is -CH₂X -CHX₂, or -CX₃, wherein each X is independently -Cl, -F, -Br, or -I.

[00230] In certain embodiments, at least one instance of G₃ is substituted or unsubstituted carbocycyl, e.g., substituted or unsubstituted C₃carbocycyl, substituted or unsubstituted C₄carbocycyl, substituted or unsubstituted C₅carbocycyl, substituted or unsubstituted C₆carbocycyl. In certain embodiments, at least one instance of G₃ is carbocycyl substituted with at least one or more halogen atoms (e.g., 1, 2, 3, 4, 5, 6, or more -Br, -I, -F, or -Cl atoms), e.g., at least one instance of G₃ is substituted or unsubstituted C₃halocarbocycyl, substituted or unsubstituted C₄halocarbocycyl, substituted or unsubstituted C₅halocarbocycyl, or substituted or unsubstituted C₆halocarbocycyl.

[00231] In certain embodiments, at least one instance of G₃ is substituted or unsubstituted heterocyclyl, e.g., substituted or unsubstituted 3-membered heterocyclyl, substituted or unsubstituted 4-membered heterocyclyl, substituted or unsubstituted 5-membered heterocyclyl, or substituted or unsubstituted 6-membered heterocyclyl. In certain embodiments, at least one instance of G₃ is heterocyclyl substituted with at least one or more halogen atoms (e.g., 1, 2, 3, 4, 5, 6, or more -Br, -I, -F, or -Cl atoms), e.g., at least one
instance of $G_3$ is substituted or unsubstituted 3-membered haloheterocyclyl, substituted or unsubstituted 4-membered haloheterocyclyl, substituted or unsubstituted 5-membered haloheterocyclyl, or substituted or unsubstituted 6-membered haloheterocyclyl.

[00232] In certain embodiments, at least one instance of $G_3$ is substituted or unsubstituted aryl, e.g., substituted or unsubstituted phenyl or substituted or unsubstituted naphthyl. In certain embodiments, at least one instance of $G_3$ is aryl substituted with at least one or more halogen atoms (e.g., 1, 2, 3, 4, 5, 6, or more -Br, -I, -F, or -Cl atoms), e.g., at least one instance of $G_3$ is substituted or unsubstituted haloaryl. In certain embodiments, at least one instance of $G_3$ is substituted or unsubstituted halophenyl, such as monosubstituted halophenyl (e.g., ortho, meta, or para -substituted with halogen atoms, substitution relative to the point of attachment of the halophenyl ring), disubstituted halophenyl (e.g., 1,2-, 1,3-, 1,4-, 1,5-, 2,3-, 2,4-, or 2,5- substituted with halogen atoms, substitution relative to the point of attachment of the halophenyl ring), or trisubstituted halophenyl (e.g., 1,3,5-, 1,2,3-, 1,2,4-, 1,2,5-, or 2,3,4- substituted with halogen atoms, substitution relative to the point of attachment of the halophenyl ring).

[00233] In certain embodiments, at least one instance of $G_3$ is substituted or unsubstituted heteroaryl, e.g., substituted or unsubstituted 5-membered heteroaryl or substituted or unsubstituted 6-membered heteroaryl. In certain embodiments, at least one instance of $G_3$ is heteroaryl substituted with at least one or more halogen atoms (e.g., 1, 2, 3, 4, 5, 6, or more -Br, -I, -F, or -Cl atoms), e.g., at least one instance of $G_3$ is substituted or unsubstituted 5-membered haloheteroaryl or substituted or unsubstituted 6-membered haloheteroaryl.

[00234] In certain embodiments, at least one instance of $G_3$ is a monophosphate, diphosphate, or triphosphate, referred to herein as the "phosphate region" of Formula (I) and (II). In certain embodiments, wherein group $G_3$ is a monophosphate, diphosphate, or triphosphate, $G_3$ may comprise a heavy atom, or may not comprise a heavy atom. If the "phosphate region" does not comprise a heavy atom, the sugar and/or base region of Formula (I) or (II) comprises a heavy atom.

[00235] As generally described herein, each instance of the "phosphate region" $G_3$ is independently a monophosphate, diphosphate, or triphosphate of formula:

\[
\begin{align*}
\text{HO-} & \quad \text{O} \\
\text{O} & \quad \text{O} \\
\text{M}^2- & \quad \text{H}, \quad \text{OH}, \quad \text{M}^2-\text{H}, \quad \text{or} \quad \text{OH}, \quad \text{OH}, \quad \text{M}^2-\text{H}
\end{align*}
\]

wherein each instance of $M^2$ is independently -0-, -S-, or -Se-.
In certain embodiments, at least one instance (e.g., 1, 2, 3, 4 or more instances, or each instance) of $M^2$ is -O-.

In certain embodiments, at least one instance (e.g., 1, 2, 3, 4 or more instances, or each instance) of $M^2$ is -S-.

In certain embodiments, at least one instance (e.g., 1, 2, 3, 4 or more instances, or each instance) of $M^2$ is -Se-.

In certain embodiments, at least one instance (e.g., 1, 2, 3, 4 or more instances, or each instance) of $G_3$ is a monophosphate group. In certain embodiments, at least one instance (e.g., 1, 2, 3, 4 or more instances, or each instance) of $G_3$ is a diphosphate group. In certain embodiments, at least one instance (e.g., 1, 2, 3, 4 or more instances, or each instance) of $G_3$ is a triphosphate group. In certain embodiments, each instance of $G_3$ is a triphosphate group.

It is understood that the compound of Formula (I) and (II) may also be provided as a salt, and in this instance, in certain embodiments, the monophosphate, diphosphate, or triphosphate groups may be provided as a salt form:

\[ O \quad O \quad O \quad O \]
\[ YO-P-\overset{\|}{\xi} \quad YO-P-O-P-\overset{\|}{\xi} \quad YO-P-O-P-O-P-\overset{\|}{\xi} \]
\[ M^2-Y, \quad \bar{O}Y \quad M^2-Y, \quad \bar{O}Y \quad \bar{O}Y \quad \bar{O}Y \]

wherein $M^2$ is as defined herein, and each $Y$ is independently hydrogen or an electropositive group (e.g., a quaternary amine, an amino acid, a metal) provided at least one instance of $Y$ (e.g., at least 1, 2, 3, or all instances of $Y$) is an electropositive group in order to provide the salt. In certain embodiments, at least one instance of $Y$ is a metal. Exemplary metals include alkali metals (e.g., Li, Na, K, Cs), alkaline earth metals (e.g., Mg, Ca, Ba), a transition metal (e.g., Hg). In certain embodiments, at least one instance of $Y$ is a quaternary amine (e.g., ammonium, NH$_4$$^+$). In certain embodiments, at least one instance of $Y$ is an amino acid having a net positive charge, e.g., for example, wherein the zwitterionic form which predominates in equilibrium is the amino acid with a quaternized alpha-amino group and the protonated alpha-carboxylic acid group. Exemplary amino acids include, but are not limited to, arginine, histidine, lysine, aspartic acid, glutamic acid, serine, threonine, asparagine, glutamine, cysteine, selenocysteine, glycine, proline, alanine, valine, isoleucine, leucine, methionine, phenylalanine, tyrosine, and tryptophan.

Various combinations of the above described embodiments of the "phosphate region" are further contemplated herein.

In certain embodiments, each instance of $G_3$ is independently a hydrogen or a triphosphate to provide a compound of Formula (I-e1), (I-e2), (II-e1), (II-e2), (II-e3), or (II-e4):
In certain embodiments, G₁ is O. In certain embodiments, M¹ is O. In certain embodiments, G₂ is hydrogen. In certain embodiments, G₂ is -SHg or -SO₂SHg. In certain
embodiments, \( G_2 \) is -SeR\(^D\), e.g., -SeCX\(_3\), wherein \( X \) is halogen. In certain embodiments, \( G_2 \) is -TeR\(^D\), e.g., -TeCX\(_3\), wherein \( X \) is halogen. In certain embodiments, \( G_2 \) is Se-CBr or -TeBr\(_3\). In certain embodiments, \( M^2 \) is O. In certain embodiments, the compound is a salt, e.g., a salt of a quaternary amine, an amino acid, or a metal.

[00241] In certain embodiments, each instance of \( G_1 \) is independently a hydrogen or a triphosphate, and \( G_i, M^1, \) and \( M^2 \) are O, to provide a compound of Formula (I-f1), (I-f2), (II-f1), (II-f2), (II-\( \beta \)), or (II-f4):

![Diagram](image-url)
or salt thereof. In certain embodiments, \( G_2 \) is hydrogen. In certain embodiments, \( G_2 \) is -SHg or -S0\(_2\)Hg. In certain embodiments, \( G_3 \) is -SeR\(_n\), e.g., -SeCX\(_3\), wherein X is halogen. In certain embodiments, \( G_2 \) is -TeR\(_n\), e.g., -TeCX\(_3\), wherein X is halogen. In certain embodiments, \( G_2 \) is Se-CBr\(_3\) or -TeBr\(_3\). In certain embodiments, the compound is a salt, e.g., a salt of a quaternary amine, an amino acid, or a metal.

[00242] In certain embodiments, each instance of \( G_3 \) is independently a hydrogen or a triphosphate, and \( G_i, M^1, \) and \( M^2 \) are O, to provide a compound of Formula (I-g1), (I-g2), (II-g1), (II-g2), (II-g3), or (II-g4) with the specified stereochemistry:
or the enantiomer thereof and/or salt thereof. In certain embodiments, \( G_2 \) is hydrogen. In certain embodiments, \( G_2 \) is -SHg or \( -S\text{O}_2\text{SHg} \). In certain embodiments, \( G_2 \) is \(-\text{SeR}_n\), e.g., \(-\text{SeCX}_3\), wherein X is halogen. In certain embodiments, \( G_2 \) is \(-\text{TeR}_n\), e.g., \(-\text{TeCX}_3\), wherein X is halogen. In certain embodiments, \( G_2 \) is \(-\text{Se-CBr}_3\) or \(-\text{TeBr}_3\). In certain embodiments, the compound is a salt, e.g., a salt of a quaternary amine, an amino acid, or a metal.

The Base Region

[00243] As generally described herein, the "base region" of a compound of Formula (I-b), polymer of Formula (II-b), or unit of Formula (II'-b) may comprise a heavy atom, or may not comprise a heavy atom. If the "base region" does not comprise a heavy atom, the phosphate and/or sugar region of a compound of Formula (I-b), polymer of Formula (II-b), or unit of Formula (II'-b) comprises a heavy atom.

[00244] In certain embodiments, the Base does not comprise a heavy atom, and is selected from the group consisting of:

A nucleic acid polymer, such as a polymer of Formula (II), may have one or more instances of any of the above formula.

[00245] In certain embodiments, the Base is an analog of adenine and guanine, and which optionally comprises a heavy atom, selected from the group consisting of:
wherein R\textsubscript{1}, R\textsubscript{2}, R\textsubscript{3}, L\textsubscript{1}, R\textsubscript{4}, R\textsubscript{5}, and M\textsubscript{3} are as defined herein. A nucleic acid polymer, such as a polymer of Formula (II), may have one or more instances of any of the above formula.

[00246] In certain embodiments, the Base is an analog of cytosine, uracil, and thymine, and which optionally comprises a heavy atom, selected from the group consisting of:

wherein R\textsubscript{3}, L\textsubscript{1}, R\textsubscript{4}, R\textsubscript{5}, M\textsubscript{3}, and M\textsubscript{4} are as defined herein. A nucleic acid polymer, such as a polymer of Formula (II), may have one or more instances of any of the above formula.

(i) **Groups R\textsuperscript{1} and R\textsuperscript{2}**

[00247] In certain embodiments of formula (iii), (iv), (v), and (vi), each instance of R\textsuperscript{1} and R\textsuperscript{2} is independently hydrogen, substituted or unsubstituted C\textsubscript{o}alkyl, substituted or unsubstituted C\textsubscript{2-2}oalkenyl, substituted or unsubstituted C\textsubscript{2-20} alkynyl, substituted or unsubstituted carboxycyclyl, substituted or unsubstituted heterocyclyl, substituted or unsubstituted aryl, or substituted or unsubstituted heteroaryl, a nitrogen protecting group, -OR\textsuperscript{B}, or -SR\textsuperscript{B}, or R\textsuperscript{1} and R\textsuperscript{2} are joined to form a substituted or unsubstituted heterocyclic or substituted or unsubstituted heteroaryl ring.

[00248] In certain embodiments, R\textsuperscript{1} is hydrogen. In certain embodiments, both R\textsuperscript{1} and R\textsuperscript{2} are hydrogen.

[00249] In certain embodiments, at least one of R\textsuperscript{1} and R\textsuperscript{2} is substituted or unsubstituted C\textsubscript{1-20} alkyl, e.g., at least one of R\textsuperscript{1} and R\textsuperscript{2} is substituted or unsubstituted C\textsubscript{1-18}alkyl, substituted or
unsubstituted C\textsubscript{1-16}alkyl, substituted or unsubstituted C\textsubscript{1-14}alkyl, substituted or unsubstituted C\textsubscript{1-12}alkyl, substituted or unsubstituted C\textsubscript{1-10}alkyl, substituted or unsubstituted C\textsubscript{1-8}alkyl, substituted or unsubstituted C\textsubscript{\alpha}alkyl, substituted or unsubstituted C\textsubscript{\alpha}alkyl, substituted or unsubstituted C\textsubscript{\alpha}alkyl, substituted or unsubstituted C\textsubscript{\alpha}alkyl, substituted or unsubstituted C\textsubscript{\alpha}alkyl. In certain embodiments, at least one of R\textsuperscript{1} and R\textsuperscript{2} is substituted or unsubstituted C\textsubscript{1}, C\textsubscript{2}, C\textsubscript{3}, C\textsubscript{4}, C\textsubscript{5}, or C\textsubscript{6}-alkyl. In certain embodiments, at least one of R\textsuperscript{1} and R\textsuperscript{2} is alkyl substituted with at least one or more halogen atoms (i.e., one or more -Br, -I, -F, or -Cl atoms), e.g., at least one of R\textsuperscript{1} and R\textsuperscript{2} is substituted or unsubstituted C\textsubscript{1}haloalkenyl, substituted or unsubstituted C\textsubscript{1,3}haloalkenyl, substituted or unsubstituted C\textsubscript{1,5}haloalkenyl, substituted or unsubstituted C\textsubscript{1,7}haloalkenyl, substituted or unsubstituted C\textsubscript{1,9}haloalkenyl, substituted or unsubstituted C\textsubscript{1,11}haloalkenyl, substituted or unsubstituted C\textsubscript{1,13}haloalkenyl, substituted or unsubstituted C\textsubscript{1,15}haloalkenyl, or substituted or unsubstituted C\textsubscript{1,17}haloalkenyl. In certain embodiments, at least one of R\textsuperscript{1} and R\textsuperscript{2} is substituted or unsubstituted C\textsubscript{1}, C\textsubscript{2}, C\textsubscript{3}, C\textsubscript{4}, C\textsubscript{5}, or C\textsubscript{6}-haloalkenyl. In certain embodiments, the haloalkenyl is a perhaloalkenyl group. In certain embodiments, at least one of R\textsuperscript{1} and R\textsuperscript{2} is -CX\textsubscript{3}, wherein X is halogen. In certain embodiments, at least one of R\textsuperscript{1} and R\textsuperscript{2} is -CBr\textsubscript{3}, CBr\textsubscript{3}H, -CBrH\textsubscript{2}, -CBr\textsubscript{2}X, or -CBrX\textsubscript{2}, wherein each instance of X is independently -Cl, -F, or -I. In certain embodiments, at least one of R\textsuperscript{1} and R\textsuperscript{2} is -Cl\textsubscript{3}, Cl\textsubscript{2}H, -CIH\textsubscript{2}, -Cl\textsubscript{2}X, or -CIX\textsubscript{2}, wherein each instance of X is independently -Cl, -F, or -Br. In certain embodiments, at least one of R\textsuperscript{1} and R\textsuperscript{2} is -CBr\textsubscript{3}, -Cl\textsubscript{3}, -CFCIBr, or -CCIBr. In any of the above instances, in certain embodiments, R\textsuperscript{1} is as defined above, and R\textsuperscript{2} is hydrogen.

[00250] In certain embodiments, at least one of R\textsuperscript{1} and R\textsuperscript{2} is substituted or unsubstituted C\textsubscript{2-20}alkenyl, e.g., substituted or unsubstituted C\textsubscript{2-16}alkenyl, substituted or unsubstituted C\textsubscript{2-14}alkenyl, substituted or unsubstituted C\textsubscript{2-12}alkenyl, substituted or unsubstituted C\textsubscript{2-10}alkenyl, substituted or unsubstituted C\textsubscript{2-8}alkenyl, substituted or unsubstituted C\textsubscript{2-6}alkenyl, substituted or unsubstituted C\textsubscript{2-4}alkenyl, substituted or unsubstituted C\textsubscript{2-2}alkenyl. In certain embodiments, at least one of R\textsuperscript{1} and R\textsuperscript{2} is substituted or unsubstituted C\textsubscript{2}, C\textsubscript{3}, C\textsubscript{4}, C\textsubscript{5}, or C\textsubscript{6}-alkenyl. In certain embodiments, at least one of R\textsuperscript{1} and R\textsuperscript{2} is alkyl substituted with at least one or more halogen atoms (e.g., 1, 2, 3, 4, 5, 6, or more -Br, -I, -F, or -Cl atoms), e.g., at least one of R\textsuperscript{1} and R\textsuperscript{2} is substituted or unsubstituted C\textsubscript{2-2}haloalkenyl, substituted or unsubstituted C\textsubscript{2-16}haloalkenyl, substituted or unsubstituted C\textsubscript{2-14}haloalkenyl, substituted or unsubstituted C\textsubscript{2-12}haloalkenyl, substituted or unsubstituted C\textsubscript{2-10}haloalkenyl, substituted or unsubstituted C\textsubscript{2-8}haloalkenyl, substituted or unsubstituted C\textsubscript{2-6}haloalkenyl, substituted or unsubstituted C\textsubscript{2-4}haloalkenyl, substituted or unsubstituted C\textsubscript{2-2}haloalkenyl.
4 haloalkenyl, or substituted or unsubstituted C₂₃ haloalkenyl. In certain embodiments, the haloalkenyl is a perhaloalkenyl group. In certain embodiments, at least one of R¹ and R² is substituted or unsubstituted C₂, C₃, C₄, C₅, or C₆ haloalkenyl. In certain embodiments, at least one of R¹ and R² is -CH₂CX=CH₂, -CH₂CH=CH₂, -CH₂CHX=CH₂, -CH₂CHX=CHX, -CH₂CHX=CHX, -CH₂CHX=CHX, -CH₂CHX=CHX, -CH₂CHX=CHX, -CH₂CHX=CHX, -CH₂CHX=CHX, -CH₂CHX=CHX, -CH₂CHX=CHX, -CH₂CHX=CHX, -CH₂CHX=CHX, -CH₂CHX=CHX, -CH₂CHX=CHX, -CH₂CHX=CHX, -CH₂CHX=CHX, -CH₂CHX=CHX, -CH₂CHX=CHX, -CH₂CHX=CHX, -CH₂CHX=CHX, -CH₂CHX=CHX, or -CH₂CHX=CHX, wherein each instance of X is independently -Cl, -F, -Br, or -I. In certain embodiments, the alkenyl group is trans or the E-isomer. In any of the above instances, in certain embodiments, R¹ is as defined above, and R² is hydrogen.

[00251] In certain embodiments, at least one of R¹ and R² is substituted or unsubstituted C₂-alkynyl, e.g., at least one of R¹ and R² is substituted or unsubstituted C₆^alkynyl, substituted or unsubstituted C₂₆-alkynyl, substituted or unsubstituted C₂₄-alkynyl, substituted or unsubstituted C₂₂-alkynyl, substituted or unsubstituted C₂₀-alkynyl, substituted or unsubstituted C₂₈-alkynyl, substituted or unsubstituted C₂₆-alkynyl, substituted or unsubstituted C₂₄-alkynyl, substituted or unsubstituted C₂₂-alkynyl, substituted or unsubstituted C₂₀-alkynyl, substituted or unsubstituted C₂₈-alkynyl, substituted or unsubstituted C₂₆-alkynyl, substituted or unsubstituted C₂₄-alkynyl, substituted or unsubstituted C₂₂-alkynyl. In certain embodiments, at least one of R¹ and R² is substituted or unsubstituted C₂, C₃, C₄, C₅, or C₆-alkynyl. In certain embodiments, at least one of R¹ and R² is alkynyl substituted with at least one or more halogen atoms (e.g., 1, 2, 3, 4, 5, 6, or more -Br, -I, -F, or -Cl atoms), e.g., at least one of R¹ and R² is substituted or unsubstituted C₂₂-haloalkynyl, substituted or unsubstituted C₂₁-haloalkynyl, substituted or unsubstituted C₂₀-haloalkynyl, substituted or unsubstituted C₂₉-haloalkynyl, substituted or unsubstituted C₂₈-haloalkynyl, substituted or unsubstituted C₂₇-haloalkynyl, substituted or unsubstituted C₂₆-haloalkynyl, substituted or unsubstituted C₂₅-haloalkynyl, or substituted or unsubstituted C₂₄-haloalkynyl. In certain embodiments, the haloalkynyl is a perhaloalkynyl group. In certain embodiments, at least one of R¹ and R² is substituted or unsubstituted C₂, C₃, C₄, C₅, or C₆-haloalkynyl. In certain embodiments, at least one of R¹ and R² is substituted or unsubstituted C₂, C₃, C₄, C₅, or C₆-haloalkynyl. In certain embodiments, at least one of R¹ and R² is substituted or unsubstituted C₃ carbocycyl, substituted or unsubstituted C₄ carbocycyl, or substituted or unsubstituted C₅ carbocycyl, substituted or unsubstituted C₆ carbocycyl. In certain embodiments, at least one of R¹ and R² is carbocycyl substituted...
with at least one or more halogen atoms (e.g., 1, 2, 3, 4, 5, 6, or more -Br, -I, -F, or -Cl atoms), e.g., at least one of \( R^1 \) and \( R^2 \) is substituted or unsubstituted C\(_6\)halocarbocycyl, substituted or unsubstituted C\(_4\)halocarbocycyl, substituted or unsubstituted Chalocarbocycyl, or substituted or unsubstituted C\(_6\)halocarbocycyl. In any of the above instances, in certain embodiments, \( R^1 \) is as defined above, and \( R^2 \) is hydrogen.

[00253] In certain embodiments, at least one of \( R^1 \) and \( R^2 \) is substituted or unsubstituted heterocyclyl, e.g., substituted or unsubstituted 3-membered heterocyclyl, substituted or unsubstituted 4-membered heterocyclyl, or substituted or unsubstituted 5-membered heterocyclyl. In certain embodiments, at least one of \( R^1 \) and \( R^2 \) is heterocyclyl substituted with at least one or more halogen atoms (e.g., 1, 2, 3, 4, 5, 6, or more -Br, -I, -F, or -Cl atoms), e.g., at least one of \( R^1 \) and \( R^2 \) is substituted or unsubstituted 3-membered haloheterocyclyl, substituted or unsubstituted 4-membered haloheterocyclyl, or substituted or unsubstituted 5-membered haloheterocyclyl. In any of the above instances, in certain embodiments, \( R^1 \) is as defined above, and \( R^2 \) is hydrogen.

[00254] In certain embodiments, at least one of \( R^1 \) and \( R^2 \) is substituted or unsubstituted aryl, e.g., substituted or unsubstituted phenyl or substituted or unsubstituted naphthyl. In certain embodiments, at least one of \( R^1 \) and \( R^2 \) is aryl substituted with at least one or more halogen atoms (e.g., 1, 2, 3, 4, 5, 6, or more -Br, -I, -F, or -Cl atoms), e.g., at least one of \( R^1 \) and \( R^2 \) is substituted or unsubstituted haloaryl. In certain embodiments, at least one of \( R^1 \) and \( R^2 \) is substituted or unsubstituted halophenyl, such as monosubstituted halophenyl (e.g., ortho, meta, or para -substituted with halogen atoms, substitution relative to the point of attachment of the halophenyl ring), disubstituted halophenyl (e.g., 1,2-, 1,3-, 1,4-, 1,5-, 2,3-, 2,4-, or 2,5- substituted with halogen atoms, substitution relative to the point of attachment of the halophenyl ring), or trisubstituted halophenyl (e.g., 1,3,5-, 1,2,3-, 1,2,4-, 1,2,5-, or 2,3,4-substituted with halogen atoms, substitution relative to the point of attachment of the halophenyl ring). In any of the above instances, in certain embodiments, \( R^1 \) is as defined above, and \( R^2 \) is hydrogen.

[00255] In certain embodiments, at least one of \( R^1 \) and \( R^2 \) is or substituted or unsubstituted heteroaryl, e.g., substituted or unsubstituted 5-membered heteroaryl or substituted or unsubstituted 6-membered heteroaryl. In certain embodiments, at least one of \( R^1 \) and \( R^2 \) is heteroaryl substituted with at least one or more halogen atoms (e.g., 1, 2, 3, 4, 5, 6, or more -Br, -I, -F, or -Cl atoms), e.g., at least one of \( R^1 \) and \( R^2 \) is substituted or unsubstituted 5-
membered haloheteroaryl or substituted or unsubstituted 6-membered haloheteroaryl. In any of the above instances, in certain embodiments, R\textsuperscript{1} is as defined above, and R\textsuperscript{2} is hydrogen.

**00256** In certain embodiments, at least one of R\textsuperscript{1} and R\textsuperscript{2} is a nitrogen protecting group, as defined herein.

**00257** In certain embodiments, at least one of R\textsuperscript{1} and R\textsuperscript{2} is -OR\textsuperscript{B}, wherein R\textsuperscript{B} is independently hydrogen, substituted or unsubstituted C\textsubscript{1-20}alkyl, substituted or unsubstituted C\textsubscript{2-20}alkenyl, substituted or unsubstituted C\textsubscript{2-20}alkynyl, substituted or unsubstituted carbocycyl, substituted or unsubstituted heterocyclyl, substituted or unsubstituted aryl, or substituted or unsubstituted heteroaryl, or an oxygen protecting group.

**00258** In certain embodiments, at least one of R\textsuperscript{1} and R\textsuperscript{2} is -SR\textsuperscript{B}, wherein R\textsuperscript{B} is independently hydrogen, substituted or unsubstituted C\textsubscript{1-20}alkyl, substituted or unsubstituted C\textsubscript{2-20}alkenyl, substituted or unsubstituted C\textsubscript{2-20}alkynyl, substituted or unsubstituted carbocycyl, substituted or unsubstituted heterocyclyl, substituted or unsubstituted aryl, or substituted or unsubstituted heteroaryl, or a sulfur protecting group.

**00259** In certain embodiments, R\textsuperscript{1} and R\textsuperscript{2} are joined to form a substituted or unsubstituted heterocyclic or substituted or unsubstituted heteroaryl ring, e.g., a 5- to 6- membered substituted or unsubstituted heterocyclic or substituted or unsubstituted heteroaryl ring.

**00260** In certain embodiments of formula (i), (ii), (v), (vi), (vii), and (viii) each instance of R\textsuperscript{4} and R\textsuperscript{5} is independently hydrogen, substituted or unsubstituted C\textsubscript{1-20}alkyl, substituted or unsubstituted C\textsubscript{2-20}alkenyl, substituted or unsubstituted C\textsubscript{2-20}alkynyl, substituted or unsubstituted carbocycyl, substituted or unsubstituted heterocyclyl, substituted or unsubstituted aryl, or substituted or unsubstituted heteroaryl, a nitrogen protecting group, -OR\textsuperscript{B}, or -SR\textsuperscript{B}, or R\textsuperscript{4} and R\textsuperscript{5} are joined to form a substituted or unsubstituted heterocyclic or substituted or unsubstituted heteroaryl ring.

(ii) **Groups R\textsuperscript{4} and R\textsuperscript{5}**

**00261** In certain embodiments, R\textsuperscript{4} is hydrogen. In certain embodiments, both R\textsuperscript{4} and R\textsuperscript{5} are hydrogen.

**00262** In certain embodiments, at least one of R\textsuperscript{4} and R\textsuperscript{5} is substituted or unsubstituted C\textsubscript{1-20}alkyl, e.g., at least one of R\textsuperscript{4} and R\textsuperscript{5} is substituted or unsubstituted C\textsubscript{1-16}alkyl, substituted or unsubstituted C\textsubscript{1-16}alkyl, substituted or unsubstituted C\textsubscript{1-14}alkyl, substituted or unsubstituted C\textsubscript{1-12}alkyl, substituted or unsubstituted C\textsubscript{1-10}alkyl, substituted or unsubstituted C\textsubscript{1-8}alkyl, substituted or unsubstituted C\textsubscript{1-6}alkyl, or substituted or unsubstituted C\textsubscript{1-3}alkyl, or substituted or unsubstituted C\textsubscript{1-3}alkyl. In certain embodiments, at
least one of R⁴ and R⁵ is substituted or unsubstituted C₁, C₂, C₃, C₄, C₅, or C₆-alkyl. In certain embodiments, at least one of R⁴ and R⁵ is alkyl substituted with at least one or more halogen atoms (i.e., one or more -Br, -I, -F, or -Cl atoms), e.g., at least one of R⁴ and R⁵ is substituted or unsubstituted C₁-haloalkyl, substituted or unsubstituted C₁₃-haloalkyl, substituted or unsubstituted C₁₈-haloalkyl, substituted or unsubstituted C₁₉-haloalkyl, substituted or unsubstituted C₁₂-haloalkyl, substituted or unsubstituted C₁-haloalkyl, substituted or unsubstituted C₁₈-haloalkyl, substituted or unsubstituted C₁₉-haloalkyl, substituted or unsubstituted C₉-haloalkyl, substituted or unsubstituted C₉-haloalkyl, or substituted or unsubstituted C₉-haloalkyl. In certain embodiments, at least one of R⁴ and R⁵ is substituted or unsubstituted C₁, C₂, C₃, C₄, C₅, or C₆-haloalkyl. In certain embodiments, the haloalkyl is a perhaloalkyl group. In certain embodiments, at least one of R⁴ and R⁵ is -CX₃, wherein X is halogen. In certain embodiments, at least one of R⁴ and R⁵ is -CBr₃, CBr₂H, -CBrH₂, -CBr₂X, or -CBrX₂, wherein each instance of X is independently -Cl, -F, or -I. In certain embodiments, at least one of R⁴ and R⁵ is -Cl₃, Cl₂H, -ClH₂, -Cl₂X, or -ClX₂, wherein each instance of X is independently -Cl, -F, or -Br. In certain embodiments, at least one of R⁴ and R⁵ is -CBr₃, -Cl₃, -CFClBr, or -CCIBr₃. In any of the above instances, in certain embodiments, R⁴ is as defined above, and R⁵ is hydrogen.

[00263] In certain embodiments, at least one of R⁴ and R⁵ is substituted or unsubstituted C₂-2-alkenyl, e.g., substituted or unsubstituted C₂-2, alkyl, substituted or unsubstituted C₂-2, i₈-alkenyl, substituted or unsubstituted C₂-2, i₄-alkenyl, substituted or unsubstituted C₂-2, i₂-alkenyl, substituted or unsubstituted C₂-2, g-alkenyl, substituted or unsubstituted C₂-2, g-alkenyl, substituted or unsubstituted C₂-2, g-alkenyl, or substituted or unsubstituted C₂-2, g-alkenyl. In certain embodiments, at least one of R⁴ and R⁵ is substituted or unsubstituted C₂, C₃, C₄, C₅, or C₆-alkenyl. In certain embodiments, at least one of R⁴ and R⁵ is alkyl substituted with at least one or more halogen atoms (e.g., 1, 2, 3, 4, 5, 6, or more -Br, -I, -F, or -Cl atoms), e.g., at least one of R⁴ and R⁵ is substituted or unsubstituted C₂-2-haloalkenyl, substituted or unsubstituted C₂-2, haloalkenyl, substituted or unsubstituted C₂-2, i₈-haloalkenyl, substituted or unsubstituted C₂-2, haloalkenyl, substituted or unsubstituted C₂-2, i₈-haloalkenyl, substituted or unsubstituted C₂-2, i₈-haloalkenyl, substituted or unsubstituted C₂-2, i₈-haloalkenyl, substituted or unsubstituted C₂-2, i₈-haloalkenyl, or substituted or unsubstituted C₂-2, i₈-haloalkenyl. In certain embodiments, the haloalkenyl is a perhaloalkenyl group. In certain embodiments, at least one of R⁴ and R⁵ is substituted or unsubstituted C₂, C₃, C₄, C₅, or C₆-haloalkenyl. In certain embodiments, at least one of R⁴ and R⁵ is -CH₂CX=CHX, -CH₂CX=CHX, -CH₂CX=CHX, -CH₂CH=CHX, -
CH₂CX=CH₂, -CX₂CH=CH₂, -CX₂CX=CH₂, -CX₂CH=CH₂, -CX₂CX=CHX, -CX₂CH=CHX, -CHXCH=CH₂, -CHXCH=CH₂, -CHXCH=CH₂, -CHXCH=CH₂, -CHXCH=CH₂, CHXCH=CH₂... C₃ halocarbocycyl, substituted or unsubstituted C₄ halocarbocycyl, substituted or unsubstituted C₅ halocarbocycyl, unsubstituted or substituted C₆ halocarbocycyl, wherein each instance of X is independently -Cl, -F, -Br, or -I. In certain embodiments, the alkenyl group is trans or the E-isomer. In any of the above instances, in certain embodiments, R⁴ is as defined above, and R⁵ is hydrogen.

[00264] In certain embodiments, at least one of R⁴ and R⁵ is substituted or unsubstituted C₂-alkynyl, e.g., at least one of R⁴ and R⁵ is substituted or unsubstituted C^alkynyl, substituted or unsubstituted C₂-alkynyl, substituted or unsubstituted C₂-alkynyl, substituted or unsubstituted C₂-alkynyl, substituted or unsubstituted C₂-alkynyl, substituted or unsubstituted C₂-alkynyl, substituted or unsubstituted C₂-alkynyl, substituted or unsubstituted C₂-alkynyl, substituted or unsubstituted C₂-alkynyl, or substituted or unsubstituted C₂-alkynyl. In certain embodiments, at least one of R⁴ and R⁵ is substituted or unsubstituted C₂, C₃, C₄, C₅, or C₆-alkynyl. In certain embodiments, at least one of R⁴ and R⁵ is alkynyl substituted with at least one or more halogen atoms (e.g., 1, 2, 3, 4, 5, 6, or more -Br, -I, -F, or -Cl atoms), e.g., at least one of R⁴ and R⁵ is substituted or unsubstituted C₂-alkynyl, substituted or unsubstituted C₂-alkynyl, substituted or unsubstituted C₂-alkynyl, substituted or unsubstituted C₂-alkynyl, substituted or unsubstituted C₂-alkynyl, substituted or unsubstituted C₂-alkynyl, substituted or unsubstituted C₂-alkynyl, substituted or unsubstituted C₂-alkynyl, substituted or unsubstituted C₂-alkynyl, substituted or unsubstituted C₂-alkynyl, substituted or unsubstituted C₂-alkynyl, or substituted or unsubstituted C₂-alkynyl. In certain embodiments, the alkynyl group is a perhaloalkynyl group. In certain embodiments, at least one of R⁴ and R⁵ is substituted or unsubstituted C₂, C₃, C₄, C₅, or C₆-alkynyl. In certain embodiments, at least one of R⁴ and R⁵ is substituted or unsubstituted C₂, C₃, C₄, C₅, or C₆-alkynyl. In any of the above instances, in certain embodiments, R⁴ is as defined above, and R⁵ is hydrogen.

[00265] In certain embodiments, at least one of R⁴ and R⁵ is substituted or unsubstituted carbocycyl, e.g., substituted or unsubstituted C₃ carbocycyl, substituted or unsubstituted C₄ carbocycyl, substituted or unsubstituted C₅ carbocycyl, substituted or unsubstituted C₆ carbocycyl. In certain embodiments, at least one of R⁴ and R⁵ is carbocycyl substituted with at least one or more halogen atoms (e.g., 1, 2, 3, 4, 5, 6, or more -Br, -I, -F, or -Cl atoms), e.g., at least one of R⁴ and R⁵ is substituted or unsubstituted C₃ halocarbocycyl, substituted or unsubstituted C₄ halocarbocycyl, substituted or unsubstituted C₅ halocarbocycyl, or substituted or unsubstituted C₆ halocarbocycyl.
or substituted or unsubstituted Cehalocarbocycyl. In any of the above instances, in certain embodiments, \( R^4 \) is as defined above, and \( R^5 \) is hydrogen.

[00266] In certain embodiments, at least one of \( R^4 \) and \( R^5 \) is substituted or unsubstituted heterocyclyl, \( e.g., \) substituted or unsubstituted 3-membered heterocyclyl, substituted or unsubstituted 4-membered heterocyclyl, or substituted or unsubstituted 5-membered heterocyclyl. In certain embodiments, at least one of \( R^4 \) and \( R^5 \) is heterocyclyl substituted with at least one or more halogen atoms (\( e.g., \) 1, 2, 3, 4, 5, 6, or more -Br, -I, -F, or -Cl atoms), \( e.g., \) at least one of \( R^4 \) and \( R^5 \) is substituted or unsubstituted 3-membered haloheterocyclyl, substituted or unsubstituted 4-membered haloheterocyclyl, substituted or unsubstituted 5-membered haloheterocyclyl, or substituted or unsubstituted 6-membered haloheterocyclyl. In any of the above instances, in certain embodiments, \( R^4 \) is as defined above, and \( R^5 \) is hydrogen.

[00267] In certain embodiments, at least one of \( R^4 \) and \( R^5 \) is substituted or unsubstituted aryl, \( e.g., \) substituted or unsubstituted phenyl or substituted or unsubstituted naphthyl. In certain embodiments, at least one of \( R^4 \) and \( R^5 \) is aryl substituted with at least one or more halogen atoms (\( e.g., \) 1, 2, 3, 4, 5, 6, or more -Br, -I, -F, or -Cl atoms), \( e.g., \) at least one of \( R^4 \) and \( R^5 \) is substituted or unsubstituted haloaryl. In certain embodiments, at least one of \( R^4 \) and \( R^5 \) is substituted or unsubstituted halophenyl, such as monosubstituted halophenyl (\( e.g., \) ortho, meta, or para -substituted with halogen atoms, substitution relative to the point of attachment of the halophenyl ring), disubstituted halophenyl (\( e.g., \) 1,2-, 1,3-, 1,4-, 1,5-, 2,3-, 2,4-, or 2,5- substituted with halogen atoms, substitution relative to the point of attachment of the halophenyl ring), or trisubstituted halophenyl (\( e.g., \) 1,3,5-, 1,2,3-, 1,2,4-, 1,2,5-, or 2,3,4- substituted with halogen atoms, substitution relative to the point of attachment of the halophenyl ring). In any of the above instances, in certain embodiments, \( R^4 \) is as defined above, and \( R^5 \) is hydrogen.

[00268] In certain embodiments, at least one of \( R^4 \) and \( R^5 \) is or substituted or unsubstituted heteroaryl, \( e.g., \) substituted or unsubstituted 5-membered heteroaryl or substituted or unsubstituted 6-membered heteroaryl. In certain embodiments, at least one of \( R^4 \) and \( R^5 \) is heteroaryl substituted with at least one or more halogen atoms (\( e.g., \) 1, 2, 3, 4, 5, 6, or more -Br, -I, -F, or -Cl atoms), \( e.g., \) at least one of \( R^4 \) and \( R^5 \) is substituted or unsubstituted 5-membered haloheteroaryl or substituted or unsubstituted 6-membered haloheteroaryl. In any of the above instances, in certain embodiments, \( R^4 \) is as defined above, and \( R^5 \) is hydrogen.

[00269] In certain embodiments, at least one of \( R^4 \) and \( R^5 \) is a nitrogen protecting group, as defined herein.
In certain embodiments, at least one of R⁴ and R⁵ is -OR, wherein R is independently hydrogen, substituted or unsubstituted C₁₋₂₀alkyl, substituted or unsubstituted C₂₋₂₀alkenyl, substituted or unsubstituted C₂₋₂₀alkynyl, substituted or unsubstituted carbocycyl, substituted or unsubstituted heterocyclyl, substituted or unsubstituted aryl, or substituted or unsubstituted heteroaryl, or an oxygen protecting group.

In certain embodiments, at least one of R⁴ and R⁵ is -SR, wherein R is independently hydrogen, substituted or unsubstituted C₁₋₂₀alkyl, substituted or unsubstituted C₂₋₂₀alkenyl, substituted or unsubstituted C₂₋₂₀alkynyl, substituted or unsubstituted carbocycyl, substituted or unsubstituted heterocyclyl, substituted or unsubstituted aryl, or substituted or unsubstituted heteroaryl, or a sulfur protecting group.

In certain embodiments, R⁴ and R⁵ are joined to form a substituted or unsubstituted heterocyclic or substituted or unsubstituted heteroaryl ring, e.g., a 5- to 6- membered substituted or unsubstituted heterocyclic or substituted or unsubstituted heteroaryl ring.

(iii) Groups L¹ and Group R²

In certain embodiments of formula (i), (iii), (v), (vii), and (ix), R¹ is independently substituted or unsubstituted Cⁿalkyl, substituted or unsubstituted C₂₋₂₀alkenyl, substituted or unsubstituted C₂₋₂₀alkynyl, substituted or unsubstituted carbocycyl, substituted or unsubstituted heterocyclyl, substituted or unsubstituted aryl, substituted or unsubstituted heteroaryl, halogen, -OR, -SR, -N(R)₂, -SHg, -SO₂SHg, -SHgR, -SeR, or -TeR.

In certain embodiments of formula (i), (iii), (v), (vii), and (ix), R¹ is substituted or unsubstituted C₁₋₁₈alkyl, e.g., substituted or unsubstituted C₁₋₁₈alkyl, substituted or unsubstituted C₁₋₁₆alkyl, substituted or unsubstituted C₁₋₄alkyl, substituted or unsubstituted C₁₋₁₄alkyl, substituted or unsubstituted C₁₋₁₂alkyl, substituted or unsubstituted C₁₋₁₀alkyl, substituted or unsubstituted C₁₋₈alkyl, substituted or unsubstituted C¹alkyl, substituted or unsubstituted C¹alkyl, or substituted or unsubstituted C¹alkyl. In certain embodiments, R¹ is alkyl substituted with at least one or more halogen atoms (i.e., one or more -Br, -I, -F, or -Cl atoms), e.g., R¹ is substituted or unsubstituted C₁₋₁₈haloalkyl, substituted or unsubstituted C₁₋₁₄haloalkyl, substituted or unsubstituted C₁₋₁₀haloalkyl, substituted or unsubstituted C₁₋₈haloalkyl, substituted or unsubstituted C₁₋₆haloalkyl, substituted or unsubstituted C¹haloalkyl, substituted or unsubstituted C¹haloalkyl, or substituted or unsubstituted C¹haloalkyl. In certain embodiments, R¹ is
substituted or unsubstituted C\textsubscript{1}, C\textsubscript{2}, C\textsubscript{3}, C\textsubscript{4}, C\textsubscript{5}, or C\textsubscript{6}-haloalkyl. In certain embodiments, the haloalkyl is a perhaloalkyl group. In certain embodiments, R\textsuperscript{3} is -CX\textsubscript{3}, wherein X is halogen.

In certain embodiments, R\textsuperscript{3} is -CBr\textsubscript{3}, CBr\textsubscript{2}H, -CBrH\textsubscript{2}, -CBr\textsubscript{2}X, or -CBrX\textsubscript{2}, wherein each instance of X is independently -Cl, -F, or -I. In certain embodiments, R\textsuperscript{3} is -Cl\textsubscript{3}, Cl\textsubscript{2}H, -ClH\textsubscript{2}, -Cl\textsubscript{2}X, or -ClX\textsubscript{2}, wherein each instance of X is independently -Cl, -F, or -Br. In certain embodiments, R\textsuperscript{3} is -CBr\textsubscript{3}, -Cl\textsubscript{3}, -CFClBr, or -CClBr.

In certain embodiments of formula (i), (iii), (v), (vii), and (ix), R\textsuperscript{3} is substituted or unsubstituted C\textsubscript{2}-alkenyl, e.g., substituted or unsubstituted C\textsuperscript{\&\textsubscript{\&}}-alkenyl, substituted or unsubstituted C\textsubscript{2}-alkenyl, substituted or unsubstituted C\textsubscript{2}-alkenyl, substituted or unsubstituted C\textsubscript{2}-alkenyl, substituted or unsubstituted C\textsubscript{2}-alkenyl, substituted or unsubstituted C\textsubscript{2}-alkenyl, substituted or unsubstituted C\textsubscript{2}-alkenyl, or substituted or unsubstituted C\textsubscript{2}-alkenyl. In certain embodiments, R\textsuperscript{3} is substituted or unsubstituted C\textsubscript{2}, C\textsubscript{3}, C\textsubscript{4}, C\textsubscript{5}, or C\textsubscript{6}-alkenyl. In certain embodiments, R\textsuperscript{3} is alkenyl substituted with at least one or more halogen atoms (e.g., 1, 2, 3, 4, 5, 6, or more -Br, -I, -F, or -Cl atoms), e.g., R\textsuperscript{3} is substituted or unsubstituted C\textsubscript{2}-haloalkenyl, substituted or unsubstituted C\textsubscript{2}haloalkenyl, substituted or unsubstituted C\textsubscript{2}haloalkenyl, substituted or unsubstituted C\textsubscript{2}haloalkenyl, substituted or unsubstituted C\textsubscript{2}haloalkenyl, substituted or unsubstituted C\textsubscript{2}haloalkenyl, substituted or unsubstituted C\textsubscript{2}haloalkenyl, substituted or unsubstituted C\textsubscript{2}haloalkenyl, substituted or unsubstituted C\textsubscript{2}haloalkenyl, or substituted or unsubstituted C\textsubscript{2}haloalkenyl. In certain embodiments, the haloalkenyl is a perhaloalkenyl group. In certain embodiments, R\textsuperscript{3} is substituted or unsubstituted C\textsubscript{2}, C\textsubscript{3}, C\textsubscript{4}, C\textsubscript{5}, or C\textsubscript{6}-haloalkenyl. In certain embodiments, R\textsuperscript{3} is -CH\textsubscript{2}CX=CH\textsubscript{2}, -CH\textsubscript{2}CH=CH\textsubscript{2}, -CH\textsubscript{2}CH=CH\textsubscript{2}, -CH\textsubscript{2}CH=CH\textsubscript{2}.

In certain embodiments of formula (i), (iii), (v), (vii), and (ix), R\textsuperscript{3} is substituted or unsubstituted C\textsubscript{2}alkynyl, e.g., R\textsuperscript{3} is substituted or unsubstituted C\textsubscript{2}alkynyl, substituted or unsubstituted C\textsubscript{2}alkynyl, substituted or unsubstituted C\textsubscript{2}alkynyl, substituted or unsubstituted C\textsubscript{2}alkynyl, substituted or unsubstituted C\textsubscript{2}alkynyl, substituted or unsubstituted C\textsubscript{2}alkynyl, substituted or unsubstituted C\textsubscript{2}alkynyl, substituted or unsubstituted C\textsubscript{2}alkynyl, or substituted or unsubstituted C\textsubscript{2}alkynyl. In certain embodiments, R\textsuperscript{3} is substituted or unsubstituted C\textsubscript{2}, C\textsubscript{3}, C\textsubscript{4}, C\textsubscript{5}, or C\textsubscript{6}-alkynyl. In certain
embodiments, R^3 is alkynyl substituted with at least one or more halogen atoms (e.g., 1, 2, 3, 4, 5, 6, or more -Br, -I, -F, or -Cl atoms), e.g., R^3 is substituted or unsubstituted C_2- 
_2 haloalkynyl, substituted or unsubstituted C^4 haloalkynyl, substituted or unsubstituted C_2- 
_2 haloalkynyl, substituted or unsubstituted C_2- 
_2 haloalkynyl, substituted or unsubstituted C_2- 
_2 haloalkynyl, substituted or unsubstituted C_2- 
_2 haloalkynyl, substituted or unsubstituted C_2- 
_2 haloalkynyl, substituted or unsubstituted C_2- 
_2 haloalkynyl, substituted or unsubstituted C_2- 
_2 haloalkynyl, or substituted or unsubstituted C_2- 
_2 haloalkynyl. In certain embodiments, the haloalkynyl is a perhaloalkynyl group. In certain embodiments, R^3 is substituted or unsubstituted C_2, C_3, C_4, C_5, or C_6-haloalkynyl. In certain embodiments, R^3 is substituted or unsubstituted carbocycyl, e.g., substituted or unsubstituted C_2-carbocycyl, substituted or unsubstituted C_4-carbocycyl, substituted or unsubstituted C_6-carbocycyl. In certain embodiments, R^3 is substituted or unsubstituted C_2-halocarbocycyl, substituted or unsubstituted C_3-halocarbocycyl, substituted or unsubstituted C_4-halocarbocycyl, substituted or unsubstituted C_6-halocarbocycyl.

[00276] In certain embodiments of formula (i), (iii), (v), (vii), and (ix), R^3 is substituted or unsubstituted heterocyclyl, e.g., substituted or unsubstituted 3-membered heterocyclyl, substituted or unsubstituted 4-membered heterocyclyl, substituted or unsubstituted 5-membered heterocyclyl, or substituted or unsubstituted 6-membered heterocyclyl. In certain embodiments, R^3 is heterocyclyl substituted with at least one or more halogen atoms (e.g., 1, 2, 3, 4, 5, 6, or more -Br, -I, -F, or -Cl atoms), e.g., R^3 is substituted or unsubstituted 3-membered haloheterocyclyl, substituted or unsubstituted 4-membered haloheterocyclyl, substituted or unsubstituted 5-membered haloheterocyclyl, or substituted or unsubstituted 6-membered haloheterocyclyl.

[00278] In certain embodiments of formula (i), (iii), (v), (vii), and (ix), R^3 is substituted or unsubstituted aryl, e.g., substituted or unsubstituted phenyl or substituted or unsubstituted naphthyl. In certain embodiments, R^3 is aryl substituted with at least one or more halogen atoms (e.g., 1, 2, 3, 4, 5, 6, or more -Br, -I, -F, or -Cl atoms), e.g., R^3 is substituted or unsubstituted haloaryl. In certain embodiments, R^3 is substituted or unsubstituted halophenyl, such as monosubstituted halophenyl (e.g., ortho, meta, or para -substituted with halogen atoms, substitution relative to the point of attachment of the halophenyl ring),
disubstituted halophenyl (e.g., 1,2-, 1,3-, 1,4-, 1,5-, 2,3-, 2,4-, or 2,5- substituted with halogen atoms, substitution relative to the point of attachment of the halophenyl ring), or trisubstituted halophenyl (e.g., 1,3,5-, 1,2,3-, 1,2,4-, 1,2,5-, or 2,3,4- substituted with halogen atoms, substitution relative to the point of attachment of the halophenyl ring).

[00279] In certain embodiments of formula (i), (iii), (v), (vii), and (ix), R3 is substituted or unsubstituted heteroaryl, e.g., substituted or unsubstituted 5-membered heteroaryl or substituted or unsubstituted 6-membered heteroaryl. In certain embodiments, R3 is heteroaryl substituted with at least one or more halogen atoms (e.g., 1, 2, 3, 4, 5, 6, or more -Br, -I, -F, or -Cl atoms), e.g., R3 is substituted or unsubstituted 5-membered halo heteroaryl or substituted or unsubstituted 6-membered halo heteroaryl.

[00280] In certain embodiments of formula (i), (iii), (v), (vii), and (ix), R3 is halogen, i.e., R3 is -Br, -I, -F, or -Cl.

[00281] In certain embodiments of formula (i), (iii), (v), (vii), and (ix), R3 is -OR, and R is hydrogen, i.e., R3 is -OH.

[00282] In certain embodiments of formula (i), (iii), (v), (vii), and (ix), R3 is -OR, and R is an oxygen protecting group, as defined herein.

[00283] In certain embodiments of formula (i), (iii), (v), (vii), and (ix), R3 is -OR, and R is substituted or unsubstituted Cαalkyl, e.g., R3 is -OR, and R is substituted or unsubstituted C1-16alkyl, substituted or unsubstituted C1-12alkyl, substituted or unsubstituted C1-alkyl, substituted or unsubstituted C1-8alkyl, substituted or unsubstituted C1-6alkyl, substituted or unsubstituted C1-4alkyl, substituted or unsubstituted Cαalkyl, or substituted or unsubstituted Cαalkyl. In certain embodiments, R3 is -OR, and R is substituted or unsubstituted C1, C2, C3, C4, C5, or C6-alkyl. In certain embodiments, R is alkyl substituted with at least one or more halogen atoms (i.e., one or more -Br, -I, -F, or -Cl atoms), e.g., R is substituted or unsubstituted Cαhaloalkyl, substituted or unsubstituted C1-alkyl, substituted or unsubstituted C1-haloalkyl, substituted or unsubstituted C1-haloalkyl, substituted or unsubstituted C1-haloalkyl, substituted or unsubstituted C1-12haloalkyl, substituted or unsubstituted C1-12haloalkyl, substituted or unsubstituted C1-12haloalkyl, substituted or unsubstituted C1-12haloalkyl, substituted or unsubstituted C1-12haloalkyl, substituted or unsubstituted C1-12haloalkyl, substituted or unsubstituted C1-12haloalkyl, substituted or unsubstituted C1-12haloalkyl, substituted or unsubstituted C1-12haloalkyl. In certain embodiments, R3 is -OR, and R is substituted or unsubstituted C1, C2, C3, C4, C5, or C6-haloalkyl. In certain embodiments, the haloalkyl is a perhaloalkyl group. In certain embodiments, R is -CX3, wherein X is halogen. In certain embodiments, R is -CBr3, CBr2H, -CBrH2, -CBr2X, or -CBrX2, wherein each instance of X
is independently -Cl, -F, or -I. In certain embodiments, R is -Cl, Cl₂H, -ClH₂, -Cl₂X, or -ClX₂, wherein each instance of X is independently -Cl, -F, or -Br. In certain embodiments, R is -Br, -Cl₂, -CF₂Br, or -CClBr.

[00284] In certain embodiments of formula (i), (iii), (v), (vii), and (ix), R is -OR, and R is substituted or unsubstituted C₂₅-alkenyl, e.g., R is -OR, and R is substituted or unsubstituted C₂-alkenyl, substituted or unsubstituted C₂₅-alkenyl, substituted or unsubstituted C₂-alkenyl, substituted or unsubstituted C₂-alkenyl, substituted or unsubstituted C₂-alkenyl, substituted or unsubstituted C₂-alkenyl, substituted or unsubstituted C₂-alkenyl, substituted or unsubstituted C₂-alkenyl, substituted or unsubstituted C₂-alkenyl, or substituted or unsubstituted C₂-alkenyl. In certain embodiments, R is -OR, and R is substituted or unsubstituted C₂₅-alkenyl, e.g., R is -OR, and R is substituted or unsubstituted C₂₅-alkenyl, substituted or unsubstituted C₂₅-alkenyl, substituted or unsubstituted C₂₅-alkenyl, substituted or unsubstituted C₂₅-alkenyl, substituted or unsubstituted C₂₅-alkenyl, substituted or unsubstituted C₂₅-alkenyl, substituted or unsubstituted C₂₅-alkenyl, substituted or unsubstituted C₂₅-alkenyl, substituted or unsubstituted C₂₅-alkenyl, or substituted or unsubstituted C₂₅-alkenyl. In certain embodiments, R is -OR, and R is substituted or unsubstituted C₂₅-alkenyl, substituted or unsubstituted C₂₅-alkenyl, substituted or unsubstituted C₂₅-alkenyl, substituted or unsubstituted C₂₅-alkenyl, substituted or unsubstituted C₂₅-alkenyl, substituted or unsubstituted C₂₅-alkenyl, or substituted or unsubstituted C₂₅-alkenyl. In certain embodiments, the alkynyl is a perhaloalkenyl group. In certain embodiments, R is -OR, and R is substituted or unsubstituted C₂₅-alkenyl, substituted or unsubstituted C₂₅-alkenyl, substituted or unsubstituted C₂₅-alkenyl, substituted or unsubstituted C₂₅-alkenyl, substituted or unsubstituted C₂₅-alkenyl, substituted or unsubstituted C₂₅-alkenyl, substituted or unsubstituted C₂₅-alkenyl, or substituted or unsubstituted C₂₅-alkenyl. In certain embodiments, R is -OR, and R is substituted or unsubstituted C₂₅-alkenyl, substituted or unsubstituted C₂₅-alkenyl, substituted or unsubstituted C₂₅-alkenyl, substituted or unsubstituted C₂₅-alkenyl, substituted or unsubstituted C₂₅-alkenyl, substituted or unsubstituted C₂₅-alkenyl, substituted or unsubstituted C₂₅-alkenyl, substituted or unsubstituted C₂₅-alkenyl, or substituted or unsubstituted C₂₅-alkenyl. In certain embodiments, the alkynyl is a perhaloalkenyl group.

[00285] In certain embodiments of formula (i), (iii), (v), (vii), and (ix), R is -OR, and R is substituted or unsubstituted C₂₅-alkenyl, e.g., R is -OR, and R is substituted or unsubstituted C₂₅-alkenyl, substituted or unsubstituted C₂₅-alkenyl, substituted or unsubstituted C₂₅-alkenyl, substituted or unsubstituted C₂₅-alkenyl, substituted or unsubstituted C₂₅-alkenyl, substituted or unsubstituted C₂₅-alkenyl, or substituted or unsubstituted C₂₅-alkenyl. In certain embodiments, R is -OR, and R is substituted or unsubstituted C₂₅-alkenyl, substituted or unsubstituted C₂₅-alkenyl, substituted or unsubstituted C₂₅-alkenyl, substituted or unsubstituted C₂₅-alkenyl, substituted or unsubstituted C₂₅-alkenyl, substituted or unsubstituted C₂₅-alkenyl, or substituted or unsubstituted C₂₅-alkenyl. In certain embodiments, the alkynyl is a perhaloalkenyl group.
atoms), e.g., $R^C$ is substituted or unsubstituted $C_{2-3}$ohaloalkynyl, substituted or unsubstituted $C_{2-4}$haloalkynyl, substituted or unsubstituted $C_{2-4}$haloalkynyl, substituted or unsubstituted $C_{2-5}$haloalkynyl, substituted or unsubstituted $C_{2-5}$haloalkynyl, substituted or unsubstituted $C_{2-5}$haloalkynyl, substituted or unsubstituted $C_{2-5}$haloalkynyl, substituted or unsubstituted $C_{2-5}$haloalkynyl. In certain embodiments, the haloalkynyl is a perhaloalkynyl group. In certain embodiments, $R^3$ is -OR $C$, and $R^C$ is substituted or unsubstituted carbocyclyl, e.g., substituted or unsubstituted C$_n$carbocyclyl, substituted or unsubstituted C$_n$carbocyclyl, substituted or unsubstituted C$_n$carbocyclyl, substituted or unsubstituted C$_n$carbocyclyl, substituted or unsubstituted C$_n$carbocyclyl, substituted or unsubstituted C$_n$carbocyclyl, substituted or unsubstituted C$_n$carbocyclyl, substituted or unsubstituted C$_n$carbocyclyl, or substituted or unsubstituted C$_n$carbocyclyl. In certain embodiments, $R^C$ is carbocyclyl substituted with at least one or more halogen atoms (e.g., 1, 2, 3, 4, 5, 6, or more -Br, -I, -F, or -CI atoms), e.g., $R^C$ is substituted or unsubstituted C$_n$halocarbocyclyl, substituted or unsubstituted C$_n$halocarbocyclyl, substituted or unsubstituted C$_n$halocarbocyclyl, substituted or unsubstituted C$_n$halocarbocyclyl, substituted or unsubstituted C$_n$halocarbocyclyl, or substituted or unsubstituted C$_n$halocarbocyclyl. In certain embodiments, $R^C$ is heterocyclyl substituted with at least one or more halogen atoms (e.g., 1, 2, 3, 4, 5, 6, or more -Br, -I, -F, or -CI atoms), e.g., $R^C$ is substituted or unsubstituted 3-membered haloheterocyclyl, substituted or unsubstituted 3-membered haloheterocyclyl, substituted or unsubstituted 3-membered haloheterocyclyl, substituted or unsubstituted 3-membered haloheterocyclyl, substituted or unsubstituted 3-membered haloheterocyclyl, or substituted or unsubstituted 3-membered haloheterocyclyl. In certain embodiments, $R^C$ is substituted or unsubstituted aryl, e.g., substituted or unsubstituted phenyl or substituted or unsubstituted naphthyl. In certain embodiments, $R^C$ is aryl substituted with at least one or more halogen atoms (e.g., 1, 2, 3, 4, 5, 6, or more -Br, -I, -F, or -CI atoms), e.g., $R^C$ is substituted or unsubstituted 3-membered haloheterocyclyl, substituted or unsubstituted 3-membered haloheterocyclyl, substituted or unsubstituted 3-membered haloheterocyclyl, substituted or unsubstituted 3-membered haloheterocyclyl, or substituted or unsubstituted 3-membered haloheterocyclyl.
substituted with halogen atoms, substitution relative to the point of attachment of the halophenyl ring), or trisubstituted halophenyl (e.g., 1,3,5-, 1,2,3-, 1,2,4-, 1,2,5-, or 2,3,4-
substituted with halogen atoms, substitution relative to the point of attachment of the halophenyl ring).

[00289] In certain embodiments of formula (i), (iii), (v), (vii), and (ix), R³ is -OR C, and R⁵ is substituted or unsubstituted heteroaryl, e.g., substituted or unsubstituted 5-membered heteroaryl or substituted or unsubstituted 6-membered heteroaryl. In certain embodiments, R⁵ is heteroaryl substituted with at least one or more halogen atoms (e.g., 1, 2, 3, 4, 5, 6, or more -Br, -I, -F, or -Cl atoms), e.g., R⁵ is substituted or unsubstituted 5-membered haloheteroaryl or substituted or unsubstituted 6-membered haloheteroaryl.

[00290] In certain embodiments of formula (i), (iii), (v), (vii), and (ix), R³ is -SR C, and R⁵ is hydrogen, i.e., R³ is -SH.

[00291] In certain embodiments of formula (i), (iii), (v), (vii), and (ix), R³ is -SR C, and R⁵ is a sulfur protecting group, as defined herein.

[00292] In certain embodiments of formula (i), (iii), (v), (vii), and (ix), R³ is -SR C, and R⁵ is substituted or unsubstituted C^alkyl, e.g., R³ is -SR C, and R⁵ is substituted or unsubstituted C1-16alkyl, substituted or unsubstituted C1-16alkyl, substituted or unsubstituted C1-16alkyl, substituted or unsubstituted C1-16alkyl, substituted or unsubstituted C1-16alkyl, substituted or unsubstituted C1-16alkyl, substituted or unsubstituted C1-16alkyl, substituted or unsubstituted C1-16alkyl, substituted or unsubstituted C1-16alkyl, substituted or unsubstituted C1-16alkyl, substituted or unsubstituted C1-16alkyl, substituted or unsubstituted C1-16alkyl, substituted or unsubstituted C1-16alkyl, substituted or unsubstituted C1-16alkyl, or substituted or unsubstituted C1-16alkyl. In certain embodiments, R³ is -SR C, and R⁵ is substituted or unsubstituted C1-C2, C3, C4, C5, or C6-alkyl. In certain embodiments, R⁵ is alkyl substituted with at least one or more halogen atoms (i.e., one or more -Br, -I, -F, or -Cl atoms), e.g., R⁵ is substituted or unsubstituted C^haloalkyl, substituted or unsubstituted C^haloalkyl, substituted or unsubstituted C^haloalkyl, substituted or unsubstituted C^haloalkyl, substituted or unsubstituted C^haloalkyl, substituted or unsubstituted C^haloalkyl, substituted or unsubstituted C^haloalkyl, substituted or unsubstituted C^haloalkyl, substituted or unsubstituted C^haloalkyl, substituted or unsubstituted C^haloalkyl, substituted or unsubstituted C^haloalkyl, substituted or unsubstituted C^haloalkyl, or substituted or unsubstituted C^haloalkyl. In certain embodiments, R³ is -SR C, and R⁵ is substituted or unsubstituted C1-C2, C3, C4, C5, or C6-haloalkyl. In certain embodiments, the haloalkyl is a perhaloalkyl group. In certain embodiments, R⁵ is -CX 3, wherein X is halogen. In certain embodiments, R⁵ is -CBr 3, CBr2H, -CBrH 2, -CBr2X, or -CBrX 2, wherein each instance of X is independently -Cl, -F, or -I. In certain embodiments, R⁵ is -Cl 3, Cl 2H, -CIH 2, -ClX, or
-CI\textsubscript{X}, wherein each instance of X is independently -CI, -F, or -Br. In certain embodiments, 
R\textsuperscript{C} is -CBr\textsubscript{3}, -Cl\textsubscript{3}, -CFClBr, or -CCl\textsubscript{3}.

[00293] In certain embodiments of formula (i), (iii), (v), (vii), and (ix), R\textsuperscript{3} is -OR\textsuperscript{C}, and R\textsuperscript{C} is substituted or unsubstituted C\textsubscript{2}-alkenyl, e.g., R\textsuperscript{3} is -SR\textsuperscript{C}, and R\textsuperscript{C} is substituted or unsubstituted C\textsubscript{2}-alkenyl, substituted or unsubstituted C\textsubscript{2}-alkenyl, substituted or unsubstituted C\textsubscript{2}-alkenyl, substituted or unsubstituted C\textsubscript{2}-alkenyl, substituted or unsubstituted C\textsubscript{2}-alkenyl, substituted or unsubstituted C\textsubscript{2}-alkenyl, substituted or unsubstituted C\textsubscript{2}-alkenyl, substituted or unsubstituted C\textsubscript{2}-alkenyl, substituted or unsubstituted C\textsubscript{2}-alkenyl, substituted or unsubstituted C\textsubscript{2}-alkenyl, substituted or unsubstituted C\textsubscript{2}-alkenyl, substituted or unsubstituted C\textsubscript{2}-alkenyl. In certain embodiments, R\textsuperscript{3} is -SR\textsuperscript{C}, and R\textsuperscript{C} is substituted or unsubstituted C\textsubscript{2}, C\textsubscript{3}, C\textsubscript{4}, C\textsubscript{5}, or C\textsubscript{6}-alkenyl. In certain embodiments, R\textsuperscript{C} is alkyl substituted with at least one or more halogen atoms (e.g., 1, 2, 3, 4, 5, 6, or more -Br, -I, -F, or -Cl atoms), e.g., R\textsuperscript{C} is substituted or unsubstituted C\textsubscript{2}-haloalkenyl, substituted or unsubstituted C\textsubscript{2}-haloalkenyl, substituted or unsubstituted C\textsubscript{2}-haloalkenyl, substituted or unsubstituted C\textsubscript{2}-haloalkenyl, substituted or unsubstituted C\textsubscript{2}-haloalkenyl, substituted or unsubstituted C\textsubscript{2}-haloalkenyl, substituted or unsubstituted C\textsubscript{2}-haloalkenyl, substituted or unsubstituted C\textsubscript{2}-haloalkenyl, substituted or unsubstituted C\textsubscript{2}-haloalkenyl, substituted or unsubstituted C\textsubscript{2}-haloalkenyl, substituted or unsubstituted C\textsubscript{2}-haloalkenyl, or substituted or unsubstituted C\textsubscript{2}-haloalkenyl. In certain embodiments, the haloalkenyl is a perhaloalkenyl group. In certain embodiments, R\textsuperscript{3} is -SR\textsuperscript{C}, and R\textsuperscript{C} is substituted or unsubstituted C\textsubscript{2}, C\textsubscript{3}, C\textsubscript{4}, C\textsubscript{5}, or C\textsubscript{6}-haloalkenyl. In certain embodiments, R\textsuperscript{C} is -CH\textsubscript{2}CX=CHX, -CH\textsubscript{2}CX=CHX, -CH\textsubscript{2}CH=CHX, -CH\textsubscript{2}CH=CHX, -CH\textsubscript{2}CH=CHX, -CH\textsubscript{2}CH=CHX, -CH\textsubscript{2}CH=CHX, -CH\textsubscript{2}CH=CHX, where each instance of X is independently -Cl, -F, -Br, or -I. In certain embodiments, the alkenyl group is trans or the E-isomer.

[00294] In certain embodiments of formula (i), (iii), (v), (vii), and (ix), R\textsuperscript{3} is -SR\textsuperscript{C}, and R\textsuperscript{C} is substituted or unsubstituted C\textsubscript{2}-alkynyl, e.g., R\textsuperscript{3} is -SR\textsuperscript{C}, and R\textsuperscript{C} is substituted or unsubstituted C\textsuperscript{2}-alkynyl, substituted or unsubstituted C\textsuperscript{2}-alkynyl, substituted or unsubstituted C\textsuperscript{2}-alkynyl, substituted or unsubstituted C\textsuperscript{2}-alkynyl, substituted or unsubstituted C\textsuperscript{2}-alkynyl, substituted or unsubstituted C\textsuperscript{2}-alkynyl, substituted or unsubstituted C\textsuperscript{2}-alkynyl, substituted or unsubstituted C\textsuperscript{2}-alkynyl, substituted or unsubstituted C\textsuperscript{2}-alkynyl, substituted or unsubstituted C\textsuperscript{2}-alkynyl, substituted or unsubstituted C\textsuperscript{2}-alkynyl, or substituted or unsubstituted C\textsuperscript{2}-alkynyl. In certain embodiments, R\textsuperscript{3} is -SR\textsuperscript{C}, and R\textsuperscript{C} is substituted or unsubstituted C\textsubscript{2}, C\textsubscript{3}, C\textsubscript{4}, C\textsubscript{5}, or C\textsubscript{6}-alkynyl. In certain embodiments, R\textsuperscript{C} is alkynyl substituted with at least one or more halogen atoms (e.g., 1, 2, 3, 4, 5, 6, or more -Br, -I, -F, or -Cl atoms), e.g., R\textsuperscript{C} is substituted or unsubstituted C\textsubscript{2}-haloalkynyl, substituted or unsubstituted C\textsubscript{2}-haloalkynyl, substituted or unsubstituted C\textsubscript{2}-haloalkynyl, substituted or unsubstituted C\textsubscript{2}-haloalkynyl, substituted or unsubstituted C\textsubscript{2}-haloalkynyl, substituted or unsubstituted C\textsubscript{2}-haloalkynyl, substituted or unsubstituted C\textsubscript{2}-haloalkynyl, substituted or unsubstituted C\textsubscript{2}-haloalkynyl, substituted or unsubstituted C\textsubscript{2}-haloalkynyl, or substituted or unsubstituted C\textsubscript{2}-haloalkynyl.
C$_2$$_i$$_8$ haloalkynyl, substituted or unsubstituted C$_2$$_i$$_6$ haloalkynyl, substituted or unsubstituted C$_2$$_i$$_4$ haloalkynyl, substituted or unsubstituted C$_2$$_i$$_2$ haloalkynyl, substituted or unsubstituted C$_2$$_i$$_0$ haloalkynyl, substituted or unsubstituted C$_2$. In certain embodiments, the haloalkynyl is a perhaloalkynyl group. In certain embodiments, $R^3$ is -SR$_C$, and $R^C$ is substituted or unsubstituted C$_3$ carbocycyl, e.g., substituted or unsubstituted C$_3$ carbocycyl, substituted or unsubstituted C$_3$ carbocycyl, substituted or unsubstituted C$_3$ carbocycyl, substituted or unsubstituted C$_3$ carbocycyl. In certain embodiments, $R^C$ is carbocycyl substituted with at least one or more halogen atoms (e.g., 1, 2, 3, 4, 5, 6, or more -Br, -I, -F, or -Cl atoms), e.g., $R^C$ is substituted or unsubstituted C$_3$ halocarbocycyl, substituted or unsubstituted C$_3$ halocarbocycyl, substituted or unsubstituted C$_3$ halocarbocycyl, or substituted or unsubstituted C$_3$ halocarbocycyl.

In certain embodiments of formula (i), (iii), (v), (vii), and (ix), $R^3$ is -SR$_C$, and $R^C$ is substituted or unsubstituted heterocycyl, e.g., substituted or unsubstituted 3-membered heterocycyl, substituted or unsubstituted 4-membered heterocycyl, or substituted or unsubstituted 5-membered heterocycyl, or substituted or unsubstituted 6-membered heterocycyl. In certain embodiments, $R^C$ is heterocycyl substituted with at least one or more halogen atoms (e.g., 1, 2, 3, 4, 5, 6, or more -Br, -I, -F, or -Cl atoms), e.g., $R^C$ is substituted or unsubstituted 3-membered haloheterocycyl, substituted or unsubstituted 4-membered haloheterocycyl, substituted or unsubstituted 5-membered haloheterocycyl, or substituted or unsubstituted 6-membered haloheterocycyl.

In certain embodiments of formula (i), (iii), (v), (vii), and (ix), $R^3$ is -SR$_C$, and $R^C$ is substituted or unsubstituted aryl, e.g., substituted or unsubstituted phenyl or substituted or unsubstituted naphthyl. In certain embodiments, $R^C$ is aryl substituted with at least one or more halogen atoms (e.g., 1, 2, 3, 4, 5, 6, or more -Br, -I, -F, or -Cl atoms), e.g., $R^C$ is substituted or unsubstituted halophenyl, such as mono substituted halophenyl (e.g., ortho, meta, or para-substituted with halogen atoms), disubstituted halophenyl (e.g., 1,2-, 1,3-, 1,4-, 1,5-, 2,3-, 2,4-, or 2,5- substituted with halogen atoms), substitution relative to the point of attachment of the halophenyl ring), or trisubstituted halophenyl (e.g., 1,3,5-, 1,2,3-, 1,2,4-, 1,2,5-, or 2,3,4-
substituted with halogen atoms, substitution relative to the point of attachment of the halophenyl ring).

In certain embodiments of formula (i), (iii), (v), (vii), and (ix), $R^3$ is -SR$^C$, and $R^C$ is substituted or unsubstituted heteroaryl, e.g., substituted or unsubstituted 5-membered heteroaryl or substituted or unsubstituted 6-membered heteroaryl. In certain embodiments, $R^C$ is heteroaryl substituted with at least one or more halogen atoms (e.g., 1, 2, 3, 4, 5, 6, or more -Br, -I, -F, or -Cl atoms), e.g., $R^C$ is substituted or unsubstituted 5-membered haloheteroaryl or substituted or unsubstituted 6-membered haloheteroaryl.

In certain embodiments of formula (i), (iii), (v), (vii), and (ix), $R^3$ is -$N(R^C)_2$, and at least one $R^C$ is hydrogen, i.e., $R^3$ is -$NHR^C$ or -$NH_2$.

In certain embodiments, $R^3$ is -$N(R^C)_2$, and at least one $R^C$ is a nitrogen protecting group, as defined herein.

In certain embodiments, $R^3$ is -$N(R^C)_2$, and at least one $R^C$ is substituted or unsubstituted $C_{1-2}$alkyl, e.g., $R^3$ is -$N(R^C)_2$, and at least one $R^C$ is substituted or unsubstituted $C_{1-18}$alkyl, substituted or unsubstituted $C_{1-16}$alkyl, substituted or unsubstituted $C_{1-14}$alkyl, substituted or unsubstituted $C_{1-12}$alkyl, substituted or unsubstituted $C_{1-10}$alkyl, substituted or unsubstituted $C_{1-8}$alkyl, substituted or unsubstituted $C_{1-6}$alkyl, substituted or unsubstituted $C_{1-4}$alkyl, substituted or unsubstituted $C^a$alkyl, or substituted or unsubstituted $C_{1-2}$alkyl. In certain embodiments, $R^3$ is -$N(R^C)_2$, and at least one $R^C$ is substituted or unsubstituted $C_1$, $C_2$, $C_3$, $C_4$, $C_5$, or $C_6$-alkyl. In certain embodiments, at least one is $R^A$ is alkyl substituted with at least one or more halogen atoms (i.e., one or more -Br, -I, -F, or -Cl atoms), e.g., at least one is $R^C$ is substituted or unsubstituted $C_{1-2}$haloalkyl, substituted or unsubstituted $C_{1-14}$haloalkyl, substituted or unsubstituted $C_{1-12}$haloalkyl, substituted or unsubstituted $C_{1-10}$haloalkyl, substituted or unsubstituted $C_{1-8}$haloalkyl, substituted or unsubstituted $C_{1-6}$haloalkyl, substituted or unsubstituted $C^a$haloalkyl, substituted or unsubstituted $C_{1-4}$haloalkyl, or substituted or unsubstituted $C^a$haloalkyl. In certain embodiments, $R^3$ is -$N(R^C)_2$, and at least one $R^C$ is substituted or unsubstituted $C_1$, $C_2$, $C_3$, $C_4$, $C_5$, or $C_6$-haloalkyl. In certain embodiments, the haloalkyl is a perhaloalkyl group. In certain embodiments, at least one $R^C$ is -$CX_3$, wherein X is halogen. In certain embodiments, at least one $R^C$ is -CBr$_3$, CBr$_2$H, -CBrH$_2$, -CBr$_2$X, or -CBrX$_2$, wherein each instance of X is independently -Cl, -F, or -I. In certain embodiments, at least one $R^C$ is -Cl$_3$, ClH$_2$, -ClH$_2$, -Cl$_2$X, or -ClX$_2$, wherein each instance of X is independently -Cl, -F, or -Br. In certain embodiments, at least one $R^C$ is -CBr$_3$, -Cl$_3$, -CFCIBr, or -CClBr$_3$. 

97
In certain embodiments, \( R^3 \) is -N(R'^C)_2, and at least one \( R^C \) is substituted or unsubstituted \( C_{2-20} \) alkenyl, e.g., \( R^3 \) is -N(R'^C)_2, and at least one \( R^C \) is substituted or unsubstituted \( C_{2-18} \) alkenyl, substituted or unsubstituted \( C_{2-16} \) alkenyl, substituted or unsubstituted \( C_{2-14} \) alkenyl, substituted or unsubstituted \( C_{2-12} \) alkenyl, substituted or unsubstituted \( C_{2-10} \) alkenyl, substituted or unsubstituted \( C_{2-8} \) alkenyl, substituted or unsubstituted \( C_{2-6} \) alkenyl, substituted or unsubstituted \( C_{2-4} \) alkenyl, or substituted or unsubstituted \( C_{2-2} \) alkenyl. In certain embodiments, \( R^3 \) is -N(R'^C)_2, and at least one \( R^C \) is substituted or unsubstituted \( C_{2-18} \) alkenyl, substituted or unsubstituted \( C_{2-16} \) alkenyl, substituted or unsubstituted \( C_{2-14} \) alkenyl, substituted or unsubstituted \( C_{2-12} \) alkenyl, substituted or unsubstituted \( C_{2-10} \) alkenyl, substituted or unsubstituted \( C_{2-8} \) alkenyl, substituted or unsubstituted \( C_{2-6} \) alkenyl, substituted or unsubstituted \( C_{2-4} \) alkenyl, or substituted or unsubstituted \( C_{2-2} \) alkenyl. In certain embodiments, at least one \( R^C \) is alkenyl substituted with at least one or more halogen atoms (e.g., 1, 2, 3, 4, 5, 6, or more -Br, -I, -F, or -Cl atoms), e.g., \( R^C \) is substituted or unsubstituted \( C_{2-20} \) haloalkenyln, substituted or unsubstituted \( C_{2-18} \) haloalkenyln, substituted or unsubstituted \( C_{2-16} \) haloalkenyln, substituted or unsubstituted \( C_{2-14} \) haloalkenyln, substituted or unsubstituted \( C_{2-12} \) haloalkenyln, substituted or unsubstituted \( C_{2-10} \) haloalkenyln, substituted or unsubstituted \( C_{2-8} \) haloalkenyln, substituted or unsubstituted \( C_{2-6} \) haloalkenyln, substituted or unsubstituted \( C_{2-4} \) haloalkenyln, or substituted or unsubstituted \( C_{2-2} \) haloalkenyln. In certain embodiments, the haloalkenyln is a perhaloalkenyln group. In certain embodiments, \( R^3 \) is -N(R'^C)_2, and at least one \( R^C \) is substituted or unsubstituted \( C_{2-18} \) alkenyl, substituted or unsubstituted \( C_{2-16} \) alkenyl, substituted or unsubstituted \( C_{2-14} \) alkenyl, substituted or unsubstituted \( C_{2-12} \) alkenyl, substituted or unsubstituted \( C_{2-10} \) alkenyl, substituted or unsubstituted \( C_{2-8} \) alkenyl, substituted or unsubstituted \( C_{2-6} \) alkenyl, substituted or unsubstituted \( C_{2-4} \) alkenyl, or substituted or unsubstituted \( C_{2-2} \) alkenyl. In certain embodiments, the alkenyl group is trans or the E-isomer.

In certain embodiments, \( R^3 \) is -N(R'^C)_2, and at least one \( R^C \) is substituted or unsubstituted \( C_{2-20} \) alkenyl, e.g., \( R^3 \) is -N(R'^C)_2, and at least one \( R^C \) is substituted or unsubstituted \( C_{2-18} \) alkenyl, substituted or unsubstituted \( C_{2-16} \) alkenyl, substituted or unsubstituted \( C_{2-14} \) alkenyl, substituted or unsubstituted \( C_{2-12} \) alkenyl, substituted or unsubstituted \( C_{2-10} \) alkenyl, substituted or unsubstituted \( C_{2-8} \) alkenyl, substituted or unsubstituted \( C_{2-6} \) alkenyl, substituted or unsubstituted \( C_{2-4} \) alkenyl, or substituted or unsubstituted \( C_{2-2} \) alkenyl. In certain embodiments, at least one \( R^A \) is alkenyl substituted with at least one or more halogen atoms (e.g., 1, 2, 3, 4, 5, 6, or more -Br, -I, -F, or -Cl atoms), e.g., at least one \( R^C \) is substituted or unsubstituted \( C_{2-20} \) haloalkenyln, substituted or unsubstituted \( C_{2-18} \) haloalkenyln, substituted or unsubstituted \( C_{2-16} \) haloalkenyln, or substituted or unsubstituted \( C_{2-14} \) haloalkenyln, substituted or unsubstituted \( C_{2-12} \) haloalkenyln, substituted or unsubstituted \( C_{2-10} \) haloalkenyln, substituted or unsubstituted \( C_{2-8} \) haloalkenyln, substituted or unsubstituted \( C_{2-6} \) haloalkenyln, or substituted or unsubstituted \( C_{2-4} \) haloalkenyln.
haloalkynyl, substituted or unsubstituted C\(^{ohaloalkynyl}\), substituted or unsubstituted C\(_{2-4}\)haloalkynyl, substituted or unsubstituted C\(_{2-6}\)haloalkynyl, substituted or unsubstituted C\(_{2-4}\)haloalkynyl, or substituted or unsubstituted C\(_{2-3}\)haloalkynyl. In certain embodiments, the haloalkynyl is a perhaloalkynyl group. In certain embodiments, R\(^{3}\) is \(-N(R^{C})_{2}\), and at least one R\(^{C}\) is substituted or unsubstituted C\(_{2}\)carbocyclyl, substituted or unsubstituted C\(_{3}\)carbocyclyl, substituted or unsubstituted C\(_{4}\)carbocyclyl, substituted or unsubstituted C\(_{5}\)carbocyclyl, or substituted or unsubstituted C\(_{6}\)carbocyclyl. In certain embodiments, at least one R\(^{C}\) is carbocyclyl substituted with at least one or more halogen atoms (\(e.g., 1, 2, 3, 4, 5, 6\), or more -Br, -I, -F, or -Cl atoms), \(e.g., at least one R^{C}\) is substituted or unsubstituted C\(_{3}\)halocarbocyclyl, substituted or unsubstituted C\(_{4}\)halocarbocyclyl, substituted or unsubstituted C\(_{5}\)halocarbocyclyl, or substituted or unsubstituted C\(_{6}\)halocarbocyclyl.

[00304] In certain embodiments, R\(^{3}\) is \(-N(R^{C})_{2}\), and at least one R\(^{C}\) is substituted or unsubstituted heterocyclyl, \(e.g., substituted or unsubstituted 3\)-membered heterocyclyl, substituted or unsubstituted 4-membered heterocyclyl, substituted or unsubstituted 5-membered heterocyclyl, or substituted or unsubstituted 6-membered heterocyclyl. In certain embodiments, at least one R\(^{C}\) is heterocyclyl substituted with at least one or more halogen atoms (\(e.g., 1, 2, 3, 4, 5, 6\), or more -Br, -I, -F, or -Cl atoms), \(e.g., at least one R^{C}\) is substituted or unsubstituted 3-membered haloheterocyclyl, substituted or unsubstituted 4-membered haloheterocyclyl, substituted or unsubstituted 5-membered haloheterocyclyl, or substituted or unsubstituted 6-membered haloheterocyclyl.

[00305] In certain embodiments, R\(^{3}\) is \(-N(R^{C})_{2}\), and at least one R\(^{C}\) is substituted or unsubstituted aryl, \(e.g., substituted or unsubstituted phenyl or substituted or unsubstituted naphthyl. In certain embodiments, at least one R\(^{C}\) is aryl substituted with at least one or more halogen atoms (\(e.g., 1, 2, 3, 4, 5, 6\), or more -Br, -I, -F, or -Cl atoms), \(e.g., at least one R^{C}\) is substituted or unsubstituted haloaryl. In certain embodiments, at least one R\(^{C}\) is substituted or unsubstituted halophenyl, such as monosubstituted halophenyl (\(e.g., ortho, meta, or para\) substituted with halogen atoms, substitution relative to the point of attachment of the halophenyl ring), disubstituted halophenyl (\(e.g., 1,2-, 1,3-, 1,4-, 1,5-, 2,3-, 2,4-, or 2,5-\)substituted with halogen atoms, substitution relative to the point of attachment of the halophenyl ring), or trisubstituted halophenyl (\(e.g., 1,3,5-, 1,2,3-, 1,2,4-, 1,2,5-, or 2,3,4-\).
substituted with halogen atoms, substitution relative to the point of attachment of the halophenyl ring).

[00307] In certain embodiments, R³ is -N(R²)₂, and at least one R² is substituted or unsubstituted heteroaryl, e.g., substituted or unsubstituted 5-membered heteroaryl or substituted or unsubstituted 6-membered heteroaryl. In certain embodiments, at least one R² is heteroaryl substituted with at least one or more halogen atoms (e.g., -Br, -I, -F, or -Cl atoms), e.g., at least one R² is substituted or unsubstituted 5-membered haloheteroaryl or substituted or unsubstituted 6-membered haloheteroaryl.

[00308] In certain embodiments, R³ is -N(R²)₂, and two R² groups are joined to form a substituted or unsubstituted heterocyclic or substituted or unsubstituted heteroaryl ring, e.g., a 5- to 6-membered substituted or unsubstituted heterocyclic or substituted or unsubstituted heteroaryl ring.

[00309] In certain embodiments, R³ is -SHgR, wherein RD is

[00310] In certain embodiments of formula (i), (iii), (v), (vii), and (ix), R³ is -SHgMe, -SHg or -SO₂SHg.

[00311] In certain embodiments of formula (i), (iii), (v), (vii), and (ix), R³ is -SeR, as defined herein.

[00312] In certain embodiments of formula (i), (iii), (v), (vii), and (ix), R³ is -TeR, as defined herein.

[00313] Furthermore, in certain embodiments of formula (i), (iii), (v), (vii), and (ix), each instance of L¹ is independently absent or a linking moiety selected from the group consisting of substituted or unsubstituted C₄₀alkylene, substituted or unsubstituted C₂₀alkenylene, substituted or unsubstituted C₂₀alkynylene, substituted or unsubstituted heteroC₁₂₀alkynylene, substituted or unsubstituted heteroC₂₀alkenylene, substituted or unsubstituted heteroC₂₀alkynylene, substituted or unsubstituted carbocycylene, substituted or unsubstituted heterocyclylene, substituted or unsubstituted areylene, or substituted or unsubstituted heteroarylene, or a combination thereof.

[00314] In certain embodiments, L¹ is absent, and R³ is attached directly to the ring system.

[00315] However, in certain embodiments, L¹ is a linking moiety selected from the group consisting of substituted and unsubstituted alkylene; substituted and unsubstituted alkenylene; substituted and unsubstituted alkynylene; substituted and unsubstituted heteroalkylene; substituted and unsubstituted heteroalkynylene; substituted and unsubstituted heteroalkynylene; substituted and unsubstituted heterocyclylene; substituted and unsubstituted heterocyclylene; substituted and
unsubstituted carbocyclylene; substituted and unsubstituted arylene; substituted and unsubstituted heteroarylene; and combinations thereof.

[00316] As used herein, reference to a linking moiety consisting of "a combination" refers to a linking moiety comprising 1, 2, 3, 4 or more of the recited moieties. For example, the linking moiety may consist of an alkylnylene attached to an alkenylene. As used herein "at least one instance" refers to 1, 2, 3, 4, or more instances of the recited moiety.

[00317] In certain embodiments, L¹ is a linking moiety selected from the group consisting of substituted and unsubstituted alkenylene; substituted and unsubstituted alkenylene; and combinations thereof.

[00318] In certain embodiments, L¹ comprises at least one instance of substituted or unsubstituted alkenylene, e.g., substituted or unsubstituted C^alkylene, substituted or unsubstituted C^alkylene, substituted or unsubstituted C⁵^alkylene, substituted or unsubstituted C⁶^alkylene, substituted or unsubstituted C⁷^alkylene, substituted or unsubstituted C⁸^alkylene, substituted or unsubstituted C⁹^alkylene, or substituted or unsubstituted C₁₀^alkylene. Exemplary alkenylene groups include unsubstituted alkenylene groups such as methylene -CH₂-, ethylene -(CH₂)₂-, n-propylene -(CH₂)₃-, n-butylene -(CH₂)₄-, n-pentylene -(CH₂)₅-, and n-hexylene -(CH₂)₆-.

[00319] In certain embodiments, L¹ is a linking moiety selected from the group consisting of substituted and unsubstituted alkenylene and substituted and unsubstituted alkenylene, and combinations thereof.

[00320] In certain embodiments, L¹ comprises at least one instance of substituted or unsubstituted alkenylene, e.g., substituted or unsubstituted C⁵^alkenylene, substituted or unsubstituted C⁶^alkenylene, substituted or unsubstituted C⁷^alkenylene, substituted or unsubstituted C⁸^alkenylene, substituted or unsubstituted C⁹^alkenylene, or substituted or unsubstituted C₁₀^alkenylene.

[00321] In certain embodiments, L¹ is a linking moiety selected from the group consisting of substituted and unsubstituted alkenylene and substituted and unsubstituted alkenylene, and combinations thereof.

[00322] In certain embodiments, L¹ comprises at least one instance of substituted or unsubstituted alkenylene, e.g., substituted or unsubstituted C⁵^alkynylene, substituted or unsubstituted C⁶^alkynylene, substituted or unsubstituted C⁷^alkynylene, substituted or unsubstituted C⁸^alkynylene, substituted or unsubstituted C⁹^alkynylene, or substituted or unsubstituted C₁₀^alkynylene.

[00323] In certain embodiments, L¹ comprises at least one instance of substituted or unsubstituted heteroalkylene, e.g., substituted or unsubstituted heteroC₁⁰^alkylene, substituted or unsubstituted heteroC¹₀^alkylene, substituted or unsubstituted heteroC²⁻alkylene, substituted or unsubstituted heteroC³⁻alkylene,
substituted or unsubstituted heteroC₃ алкylene, substituted or unsubstituted heteroC₄-alkylene, or substituted or unsubstituted heteroC₅₋₆-alkylene. Exemplary heteroalkylene groups include unsubstituted alkylene groups such as -(CH₂)₂⁻₀(CH₂)₂⁻₀-, -OCH₂⁻₀-, -CH₂⁻₀-, -O(CH₂)₂⁻₀-, -(CH₂)₃⁻₀-, -(CH₂)₄⁻₀-, -(CH₂)₅⁻₀-, -(CH₂)₆⁻₀-. 

[00324] In certain embodiments, L¹ comprises at least one instance of substituted or unsubstituted heteroalkenylen e, e.g., substituted or unsubstituted heteroC₂₋₆-alkenylen e, substituted or unsubstituted heteroC₂₋₃-alkenylen e, substituted or unsubstituted heteroC₃₋₄-alkenylen e, substituted or unsubstituted heteroC₄₋₅-alkenylen e, or substituted or unsubstituted heteroC₅₋₆-alkenylen e.

[00325] In certain embodiments, L¹ comprises at least one instance of substituted or unsubstituted heteroalkynylene, e.g., substituted or unsubstituted heteroC₂₋₆-alkynylene, substituted or unsubstituted heteroC₂₋₃-alkynylene, substituted or unsubstituted heteroC₃₋₄-alkynylene, substituted or unsubstituted heteroC₄₋₅-alkynylene, or substituted or unsubstituted heteroC₅₋₆-alkynylene.

[00326] In certain embodiments, L¹ comprises at least one instance of substituted or unsubstituted carbocyclylene, e.g., substituted or unsubstituted C₃₋₆-carbocyclylene, substituted or unsubstituted C₃₋₆-carbocyclylene, substituted or unsubstituted C₄₋₅-carbocyclylene, or substituted or unsubstituted C₅₋₆-carbocyclylene.

[00327] In certain embodiments, L¹ comprises at least one instance of substituted or unsubstituted heterocyclylene, e.g., substituted or unsubstituted C₃₋₆-heterocyclylene, substituted or unsubstituted C₃₋₄-heterocyclylene, substituted or unsubstituted C₄₋₅-heterocyclylene, or substituted or unsubstituted C₅₋₆-heterocyclylene.

[00328] In certain embodiments, L¹ comprises at least one instance of substituted or unsubstituted arylene, e.g., substituted or unsubstituted phenylene.

[00329] In certain embodiments, L¹ comprises at least one instance of substituted or unsubstituted heteroarylene, e.g., substituted or unsubstituted 5- to 6-membered heteroarylene.

[00330] In certain embodiments, L¹ represents a linking moiety consisting of a combination of one or more consecutive covalently bonded groups of the formula:
wherein:

each instance of $m$ is independently an integer between 1 to 10, inclusive;

each instance of $p$ is independently an integer between 1 to 4, inclusive;

each instance of $R^w_1$ is independently hydrogen; substituted or unsubstituted alkyl; substituted or unsubstituted alkenyl; substituted or unsubstituted alkynyl; substituted or unsubstituted carbocyclyl; substituted or unsubstituted heterocyclyl; substituted or unsubstituted aryl; substituted or unsubstituted heteroaryl; or a nitrogen protecting group;

each instance of $R^w_2$ is independently hydrogen; halogen; substituted or unsubstituted alkyl; substituted or unsubstituted alkenyl; substituted or unsubstituted alkynyl; substituted or unsubstituted carbocyclyl; substituted or unsubstituted heterocyclyl; substituted or unsubstituted aryl; substituted or unsubstituted heteroaryl; or two $R^w_2$ groups are joined to form a substituted or unsubstituted 5- to 6-membered ring.

[00331] In certain embodiments, $L^1$ represents a linking moiety consisting of a combination of 1 to 10 consecutive covalently bonded groups of the above described formulae, e.g., $L^1$ represents a linking moiety consisting of a combination of 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 consecutive covalently bonded groups of the above described formulae. It should be generally understood that multiple instances of a given variable or group present in a linking moiety may optionally differ.

[00332] As described herein, each instance of $R^w_1$ is independently hydrogen; substituted or unsubstituted alkyl; substituted or unsubstituted alkenyl; substituted or unsubstituted alkynyl; substituted or unsubstituted carbocyclyl; substituted or unsubstituted heterocyclyl; substituted or unsubstituted aryl; substituted or unsubstituted heteroaryl; a nitrogen protecting group if attached to a nitrogen atom, or an oxygen protecting group if attached to an oxygen atom. In any of the above formulae, as described herein, in certain embodiments, each instance of $R^w_1$ is independently hydrogen; substituted or unsubstituted alkyl (e.g., methyl); or a nitrogen protecting group.

[00333] As described herein, each instance of $R^w_2$ is independently hydrogen; halogen; substituted or unsubstituted alkyl; substituted or unsubstituted alkenyl; substituted or unsubstituted alkynyl; substituted or unsubstituted carbocyclyl; substituted or unsubstituted heterocyclyl; substituted or unsubstituted aryl; substituted or unsubstituted heteroaryl; or two $R^w_2$ groups are joined to form a 5-6 membered ring. In any of the above formulae, as
described herein, in certain embodiments, each instance of \( R^{W2} \) is independently hydrogen, halogen (e.g., -Br, -Cl, -F, or -I), or substituted or unsubstituted alkyl (e.g., methyl).

[00334] As described herein, each instance of \( m \) is independently an integer between 1 to 10, inclusive. In certain embodiments, \( m \) is 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10.

[00335] As described herein, each instance of \( p \) is independently an integer between 1 to 4, inclusive. In certain embodiments, \( p \) is 1, 2, 3, or 4.

[00336] In certain embodiments, the group \(-L^1-R^3\) represents a group of formula:

\[
\begin{align*}
\text{wherein } R^{W2}, m, \text{ and } R^3 \text{ are as defined herein. In certain embodiments, } m \text{ is 1, 2, 3, or 4. In certain embodiments, each instance of } R^{W2} \text{ is hydrogen. In certain embodiments, } R^3 \text{ is aryl substituted with at least one or more halogen atoms (e.g., -Br, -I). In certain embodiments, } R^3 \text{ is halogen (e.g., -Br, -I), } -OR^C, -SR^C, -N(R^C)_2, -SHg, -SO_2SHg, -SHgR^D, -SeR^D, \text{ or } -TeR^D. \\
\text{In certain embodiments, } R^3 \text{ is } -OR^C. \text{ In certain embodiments, } R^3 \text{ is } -SR^C. \text{ In certain embodiments, } R^3 \text{ is } -N(R^C)_2. \text{ In certain embodiments, } R^C \text{ is } -CX_3, \text{ wherein } X \text{ is halogen. In certain embodiments, } R^3 \text{ is } -SHgR^D \text{ (e.g., } -SHgMe), -SHg \text{ or } -SO_2SHg. \text{ In certain embodiments, } R^3 \text{ is } -SeR^D, \text{ e.g., } -SeCX_3, \text{ wherein } X \text{ is halogen. In certain embodiments, } R^3 \text{ is } -TeR^D, \text{ e.g., } -TeCX_3, \text{ wherein } X \text{ is halogen. In certain embodiments, } R^3 \text{ is } Se-CBr_3 \text{ or } -TeBr_3.
\end{align*}
\]

(iv) Group \( M^3 \) and \( M^4 \)

[00337] As generally defined herein, each instance of \( M^3 \) and \( M^4 \) is independently O, Se, Te, CH\(_2\), CF\(_2\), CC\(_1\)\(_2\), CBr\(_2\), or Cl\(_2\).

[00338] In certain embodiments of formula (vii), (viii), (ix), and (x), \( M^3 \) is O. In certain embodiments of formula (vii), (viii), (ix), and (x), \( M^3 \) is Se. In certain embodiments of formula (vii), (viii), (ix), and (x), \( M^3 \) is Te. In certain embodiments of formula (vii), (viii), (ix), and (x), \( M^3 \) is CH\(_2\). In certain embodiments of formula (vii), (viii), (ix), and (x), \( M^3 \) is CF\(_2\). In certain embodiments of formula (vii), (viii), (ix), and (x), \( M^3 \) is CC\(_1\)\(_2\). In certain embodiments of formula (vii), (viii), (ix), and (x), \( M^3 \) is CBr\(_2\). In certain embodiments of formula (vii), (viii), (ix), and (x), \( M^3 \) is Cl\(_2\).
In certain embodiments of formula (iii), (iv), (ix), and (x), M⁴ is O. In certain embodiments of formula (iii), (iv), (ix), and (x), M⁴ is Se. In certain embodiments of formula (iii), (iv), (ix), and (x), M⁴ is Te. In certain embodiments of formula (iii), (iv), (ix), and (x), M⁴ is CH₂. In certain embodiments of formula (vii), (viii), (ix), and (x), M⁴ is CF₂. In certain embodiments of formula (vii), (viii), (ix), and (x), M⁴ is CCl₂. In certain embodiments of formula (vii), (viii), (ix), and (x), M⁴ is CBr₂. In certain embodiments of formula (vii), (viii), (ix), and (x), M⁴ is Cl₂.

In certain embodiments of formula (ix) and (x), M³ is O and M⁴ is O. In certain embodiments of formula (ix) and (x), M³ is Se and M⁴ is O. In certain embodiments of formula (ix) and (x), M³ is Te and M⁴ is O. In certain embodiments of formula (ix) and (x), M³ is CH₂ and M⁴ is O. In certain embodiments of formula (ix) and (x), M³ is CF₂ and M⁴ is O. In certain embodiments of formula (ix) and (x), M³ is CCl₂ and M⁴ is O. In certain embodiments of formula (ix) and (x), M³ is CBr₂ and M⁴ is O. In certain embodiments of formula (ix) and (x), M³ is Cl₂ and M⁴ is O.

In certain embodiments of formula (ix) and (x), M³ is O and M⁴ is Se. In certain embodiments of formula (ix) and (x), M³ is Se and M⁴ is Se. In certain embodiments of formula (ix) and (x), M³ is Te and M⁴ is Se. In certain embodiments of formula (ix) and (x), M³ is CH₂ and M⁴ is Se. In certain embodiments of formula (ix) and (x), M³ is CF₂ and M⁴ is Se. In certain embodiments of formula (ix) and (x), M³ is CCl₂ and M⁴ is Se. In certain embodiments of formula (ix) and (x), M³ is CBr₂ and M⁴ is Se. In certain embodiments of formula (ix) and (x), M³ is Cl₂ and M⁴ is Se.

In certain embodiments of formula (ix) and (x), M³ is O and M⁴ is Te. In certain embodiments of formula (ix) and (x), M³ is Se and M⁴ is Te. In certain embodiments of formula (ix) and (x), M³ is Te and M⁴ is Te. In certain embodiments of formula (ix) and (x), M³ is CH₂ and M⁴ is Te. In certain embodiments of formula (ix) and (x), M³ is CF₂ and M⁴ is Te. In certain embodiments of formula (ix) and (x), M³ is CCl₂ and M⁴ is Te. In certain embodiments of formula (ix) and (x), M³ is CBr₂ and M⁴ is Te. In certain embodiments of formula (ix) and (x), M³ is Cl₂ and M⁴ is Te.

In certain embodiments of formula (ix) and (x), M³ is O and M⁴ is CH₂. In certain embodiments of formula (ix) and (x), M³ is Se and M⁴ is CH₂. In certain embodiments of formula (ix) and (x), M³ is Te and M⁴ is CH₂. In certain embodiments of formula (ix) and (x), M³ is CH₂ and M⁴ is CH₂. In certain embodiments of formula (ix) and (x), M³ is CF₂ and M⁴ is CH₂. In certain embodiments of formula (ix) and (x), M³ is CCl₂ and M⁴ is CH₂.
certain embodiments of formula (ix) and (x), $M^3$ is CBr$_2$ and $M^4$ is CH$_2$. In certain embodiments of formula (ix) and (x), $M^3$ is Cl$_2$ and $M^4$ is CH$_2$.

[00344] In certain embodiments of formula (ix) and (x), $M^3$ is O and $M^4$ is CF$_2$. In certain embodiments of formula (ix) and (x), $M^3$ is Se and $M^4$ is CF$_2$. In certain embodiments of formula (ix) and (x), $M^3$ is CH$_2$ and $M^4$ is CF$_2$. In certain embodiments of formula (ix) and (x), $M^3$ is CF$_2$ and $M^4$ is CF$_2$. In certain embodiments of formula (ix) and (x), $M^3$ is CC$_1$$_2$ and $M^4$ is CF$_2$. In certain embodiments of formula (ix) and (x), $M^3$ is CBr$_2$ and $M^4$ is CF$_2$. In certain embodiments of formula (ix) and (x), $M^3$ is Cl$_2$ and $M^4$ is CF$_2$.

[00345] In certain embodiments of formula (ix) and (x), $M^3$ is O and $M^4$ is CC$_1$$_2$. In certain embodiments of formula (ix) and (x), $M^3$ is Se and $M^4$ is CC$_1$$_2$. In certain embodiments of formula (ix) and (x), $M^3$ is Te and $M^4$ is CC$_1$$_2$. In certain embodiments of formula (ix) and (x), $M^3$ is CH$_2$ and $M^4$ is CC$_1$$_2$. In certain embodiments of formula (ix) and (x), $M^3$ is CF$_2$ and $M^4$ is CC$_1$$_2$. In certain embodiments of formula (ix) and (x), $M^3$ is CBr$_2$ and $M^4$ is CC$_1$$_2$. In certain embodiments of formula (ix) and (x), $M^3$ is Cl$_2$ and $M^4$ is CC$_1$$_2$.

[00346] In certain embodiments of formula (ix) and (x), $M^3$ is O and $M^4$ is CBr$_2$. In certain embodiments of formula (ix) and (x), $M^3$ is Se and $M^4$ is CBr$_2$. In certain embodiments of formula (ix) and (x), $M^3$ is Te and $M^4$ is CBr$_2$. In certain embodiments of formula (ix) and (x), $M^3$ is CH$_2$ and $M^4$ is CBr$_2$. In certain embodiments of formula (ix) and (x), $M^3$ is CF$_2$ and $M^4$ is CBr$_2$. In certain embodiments of formula (ix) and (x), $M^3$ is CBr$_2$ and $M^4$ is CBr$_2$. In certain embodiments of formula (ix) and (x), $M^3$ is Cl$_2$ and $M^4$ is CBr$_2$.

[00347] In certain embodiments of formula (ix) and (x), $M^3$ is O and $M^4$ is Cl$_2$. In certain embodiments of formula (ix) and (x), $M^3$ is Se and $M^4$ is Cl$_2$. In certain embodiments of formula (ix) and (x), $M^3$ is Te and $M^4$ is Cl$_2$. In certain embodiments of formula (ix) and (x), $M^3$ is CH$_2$ and $M^4$ is Cl$_2$. In certain embodiments of formula (ix) and (x), $M^3$ is CF$_2$ and $M^4$ is Cl$_2$. In certain embodiments of formula (ix) and (x), $M^3$ is CBr$_2$ and $M^4$ is Cl$_2$. In certain embodiments of formula (ix) and (x), $M^3$ is Cl$_2$ and $M^4$ is Cl$_2$.

(v) Exemplary combinations of the base region

[00348] Various combinations of the above described embodiments of the "base region" are further contemplated herein.
For example, in certain embodiments, at least one instance of a Base is of formula (i), (iii), (v), (vii), or (ix):

wherein the base optionally comprises a heavy atom. In certain embodiments, the base comprises a heavy atom. In certain embodiments, the base does not comprise a heavy atom, but the sugar region or phosphate region comprises a heavy atom. In certain embodiments, \( M_3 \) is O. In certain embodiments, \( M_3 \) is Se. In certain embodiments, \( M^4 \) is O. In certain embodiments, \( M^4 \) is Se. In certain embodiments, \( R^1 \) and \( R^2 \) are hydrogen. In certain embodiments, \( R^4 \) and \( R^5 \) are hydrogen. In certain embodiments, \( L^1 \) is absent. In certain embodiments, \( L^1 \) is substituted or unsubstituted alkynylene. In certain embodiments, \( L^1 \) is substituted or unsubstituted alkenylene. In certain embodiments, \( L^1 - R^3 \) is of formula:

wherein \( R^w_2, m, \) and \( R^3 \) are as defined herein. In certain embodiments, \( m \) is 1, 2, 3, or 4. In certain embodiments, each instance of \( R^w_2 \) is hydrogen. In certain embodiments, \( R^3 \) is aryl substituted with at least one or more halogen atoms (e.g., -Br, -I). In certain embodiments, \( R^3 \) is halogen (e.g., -Br, -I, -OR\(^C\), -SR\(^C\), -N(R\(^C\))_2, -SHg, -SO\(^2\)SHg, -SHgR\(^D\), -SeR\(^D\), or -TeR\(^D\). In certain embodiments, \( R^3 \) is -OR\(^C\). In certain embodiments, \( R^3 \) is -SR\(^C\). In certain embodiments, \( R^3 \) is -N(R\(^C\))_2. In certain embodiments, \( R^C \) is -CX\(^3\), wherein X is halogen. In certain embodiments, \( R^3 \) is -SHgR\(^D\) (e.g., -SHgMe), -SHg or -SO\(^2\)SHg. In certain
In certain embodiments, $R^3$ is $-SeR^D$, e.g., $-SeCX_3$, wherein $X$ is halogen. In certain embodiments, $R^3$ is $-TeR^D$, e.g., $-TeCX_3$, wherein $X$ is halogen. In certain embodiments, $R^3$ is $Se-CBr_3$ or $-TeBr_3$.

In certain embodiments of formula (ix), wherein $M^4$ is O, wherein $L^I$ is absent, or $L^I$-$R^3$ is a group of formula $R^3$-\(\begin{array}{c} R^W \text{ or } \text{a heavy atom} \end{array}\), or \(\begin{array}{c} R^W \text{ or } \text{a heavy atom} \end{array}\), $m$ is 1, and each $R^W$ is hydrogen, provided is a base of formula (ix-a), (ix-b), (ix-c), or (ix-d):

\[
\begin{array}{c}
\text{(ix-a)}, \\
\text{(ix-b)}, \\
\text{(ix-c)}, \text{ or } \text{(ix-d)},
\end{array}
\]

wherein the base optionally comprises a heavy atom. In certain embodiments, the base comprises a heavy atom. In certain embodiments, the base does not comprise a heavy atom, but the sugar region or phosphate region comprises a heavy atom. In certain embodiments, $R^3$ is aryl substituted with at least one or more halogen atoms (e.g., -Br, -I). In certain embodiments, $R^3$ is halogen (e.g., -Br, -I), -OR, -SR, -N(R$^C$)$_2$, -SHg, -SO$_2$SHg, -SHgR$^D$, -SeR$^D$, or -TeR$^D$. In certain embodiments, $R^3$ is -OR. In certain embodiments, $R^3$ is -SR. In certain embodiments, $R^3$ is -$N(R^C)_2$. In certain embodiments, $R^C$ is -CX$_3$, wherein $X$ is halogen. In certain embodiments, $R^3$ is -SHgR$^D$ (e.g., -SHgMe), -SHg or -SO$_2$SHg. In certain embodiments, $R^3$ is -SeR$^D$, e.g., -SeCX$_3$, wherein $X$ is halogen. In certain embodiments, $R^3$ is -TeR$^D$, e.g., -TeCX$_3$, wherein $X$ is halogen. In certain embodiments, $R^3$ is Se-CBr$_3$ or -TeBr$_3$.

In certain embodiments of Formula (I-d), wherein the base is of formula (ix-a), (ix-b), (ix-c), or (ix-d), provided is a compound of Formula (I-d-ix-a), (I-d-ix-b), (I-d-ix-c), or (I-d-ix-d):
wherein the compound comprises a heavy atom. In certain embodiments, R⁴ and R⁵ are hydrogen. In certain embodiments, G₂ of the sugar region is hydrogen. In certain embodiments, G₂ of the sugar region is -SHgMe, -SHg or -SO₂SHg. In certain embodiments, G₂ of the sugar region is -SeR³, e.g., -SeCX₃, wherein X is halogen. In certain embodiments, G₂ of the sugar region is -TeR³, e.g., -TeCX₃, wherein X is halogen. In certain embodiments, G₂ is Se-CBr₃ or -TeBr₃. In certain embodiments, G³ is a monophosphate, diphosphate, or triphosphate of formula:

\[
\begin{align*}
\text{HO-} & \text{P—HO—P—O—P—O—P—O—} \\
\text{M—H} & , \quad \text{OH—M—H} \text{, or OH—OH—M—H}.
\end{align*}
\]

In certain embodiments, M² is -O-. In certain embodiments, R³ is aryl substituted with at least one or more halogen atoms (e.g., -Br, -I). In certain embodiments, R³ is halogen (e.g., -Br, -I), -OR, -SR, -N(R²)₂, -SHg, -SO₂SHg, -SHgR³, -SeR³ or -TeR³. In certain embodiments, R³ is -OR. In certain embodiments, R³ is -SR. In certain embodiments, R³ is -N(R²)₂. In certain embodiments, R³ is -CX₃, wherein X is halogen. In certain embodiments, R³ is -SHgR³ (e.g., -SHgMe), -SHg or -SO₂SHg. In certain embodiments, R³ is -SeR³, e.g., -SeCX₃, wherein X is halogen. In certain embodiments, R³ is -TeR³, e.g., -TeCX₃, wherein X is halogen. In certain embodiments, R³ is Se-CBr₃ or -TeBr₃.

[00352] In certain embodiments, wherein the Base is of formula (ix-a), (ix-b), (ix-c), or (ix-d), provided is a compound of Formula (II'-d), comprising at least one instance of (II'-d-ix-a), (II'-d-ix-b), (II'-d-ix-c), and/or (II'-d-ix-d):
wherein the at least one instance of (II'-d-ix-a), (II'-d-ix-b), (II'-d-ix-c), and/or (II'-d-ix-d) comprises a heavy atom. In certain embodiments, M₂ is -0-. In certain embodiments, R⁴ and R⁵ are hydrogen. In certain embodiments, G₂ of the sugar region is hydrogen. In certain embodiments, G₂ of the sugar region is -SHgMe, -SHg or -S0₂SHg. In certain embodiments, G₂ of the sugar region is -SeR₃, e.g., -SeCX₃, wherein X is halogen. In certain embodiments, G₂ of the sugar region is -TeR₃, e.g., -TeCX₃, wherein X is halogen. In certain embodiments, G₂ is Se-CBr₃ or -TeBr₃. In certain embodiments, R³ is aryl substituted with at least one or more halogen atoms (e.g., -Br, -I). In certain embodiments, R³ is halogen (e.g., -Br, -I), -OR, -SR, -N(R⁵)₂, -SHg, -S0₂SHg, -SHgR₃, -SeR₃ or -TeR₃. In certain embodiments, R³ is -OR. In certain embodiments, R³ is -SR. In certain embodiments, R³ is -N(R⁵)₂. In certain embodiments, R⁵ is -CX₃, wherein X is halogen. In certain embodiments, R³ is -SHgR₃ (e.g., -SHgMe), -SHg or -S0₂SHg. In certain embodiments, R³ is -SeR₃, e.g., -SeCX₃, wherein X is halogen. In certain embodiments, R³ is -TeR₃, e.g., -TeCX₃, wherein X is halogen. In certain embodiments, R³ is Se-CBr₃ or -TeBr₃.

[00353] In certain embodiments of formula (i), wherein L¹ is absent, or -L¹-R₃ is of formula
wherein the base optionally comprises a heavy atom. In certain embodiments, the base comprises a heavy atom. In certain embodiments, the base does not comprise a heavy atom, but the sugar region or phosphate region comprises a heavy atom. In certain embodiments, R⁴ and R⁵ are hydrogen. In certain embodiments, R³ is aryl substituted with at least one or more halogen atoms (e.g., -Br, -I). In certain embodiments, R³ is halogen (e.g., -Br, -I), -OR C, -SR C, -N(R C), -SHg, -S0₂ SHg, -SHgR D, -SeR D, or -TeR D. In certain embodiments, R³ is -OR C. In certain embodiments, R³ is -SR C. In certain embodiments, R³ is -N(R C)₂. In certain embodiments, R C is -CX ₃, wherein X is halogen. In certain embodiments, R³ is -SHgR D (e.g., -SHgMe), -SHg or -S0₂ SHg. In certain embodiments, R³ is -SeR D, e.g., -SeCX ₃, wherein X is halogen. In certain embodiments, R³ is -TeR D, e.g., -TeCX ₃, wherein X is halogen. In certain embodiments, R³ is Se-CBr₃ or -TeBr₃.

[00354] In certain embodiments of Formula (I-d), wherein the Base is of formula (i-a), (i-b), (i-c) or (i-d), provided is a compound of Formula (I-d-i-a), (I-d-i-b), (I-d-i-c), or (I-d-i-d):
wherein the compound comprises a heavy atom. In certain embodiments, \( R^4 \) and \( R^5 \) are hydrogen. In certain embodiments, \( G_2 \) of the sugar region is hydrogen. In certain embodiments, \( G_2 \) of the sugar region is -SHgMe, -SHg or -SO\textsubscript{2}SHg. In certain embodiments, \( G_2 \) of the sugar region is -SeR\textsuperscript{D}, e.g., -SeCX\textsubscript{3}, wherein X is halogen. In certain embodiments, \( G_2 \) of the sugar region is -TeR\textsuperscript{D}, e.g., -TeCX\textsubscript{3}, wherein X is halogen. In certain embodiments, \( G_2 \) is Se-CBr\textsubscript{3} or -TeBr\textsubscript{3}. In certain embodiments, \( G^3 \) is a monophosphate, diphosphate, or triphosphate of formula:

\[
\begin{align*}
\text{HO-P—} & \quad \text{HO-P—} \quad \text{HO-P—} \\
\text{O} & \quad \text{O} & \quad \text{O} \\
M^2-H, & \quad \text{OH} & \quad M^2-H, \text{ or } & \quad \text{OH} & \quad \text{OH} & \quad M^2-H.
\end{align*}
\]

In certain embodiments, \( M^2 \) is -0-. In certain embodiments, \( R^3 \) is aryl substituted with at least one or more halogen atoms (e.g., -Br, -I). In certain embodiments, \( R^3 \) is halogen (e.g., -Br, -I), -0R\textsuperscript{C}, -SR\textsuperscript{C}, -N(R\textsuperscript{C})\textsubscript{2}, -SHg, -SO\textsubscript{2}SHg, -SHgR\textsuperscript{D}, -SeR\textsuperscript{D}, or -TeR\textsuperscript{D}. In certain embodiments, \( R^3 \) is -OR\textsuperscript{C}. In certain embodiments, \( R^3 \) is -SR\textsuperscript{C}. In certain embodiments, \( R^3 \) is -N(R\textsuperscript{C})\textsubscript{2}. In certain embodiments, \( R^3 \) is -CX\textsubscript{3}, wherein X is halogen. In certain embodiments, \( R^3 \) is -SeR\textsuperscript{D}, -TeR\textsuperscript{D}, e.g., -TeCX\textsubscript{3}, wherein X is halogen. In certain embodiments, \( R^3 \) is Se-CBr\textsubscript{3} or -TeBr\textsubscript{3}.

[00355] In certain embodiments, wherein the Base is of formula (i-a), (i-b), (i-c), or (i-d), provided is a compound of Formula (II'-d), comprising at least one instance of (II'-d-i-a), (ir-d-i-b), (ir-d-i-c), and/or (IΓ-d-i-d):
wherein the at least one instance of (I’-d-i-a), (II’-d-i-b), (I’-d-i-c), and/or (II’-d-i-d) comprises a heavy atom. In certain embodiments, M2 is -0-. In certain embodiments, R4 and R5 are hydrogen. In certain embodiments, G2 of the sugar region is hydrogen. In certain embodiments, G2 of the sugar region is -SHgMe, -SHg or -SO2SHg. In certain embodiments, G2 of the sugar region is -SeR, e.g., -SeCX3, wherein X is halogen. In certain embodiments, G2 of the sugar region is -TeR, e.g., -TeCX3, wherein X is halogen. In certain embodiments, G2 is Se-CBr3 or -TeBr3. In certain embodiments, R3 is aryl substituted with at least one or more halogen atoms (e.g., -Br, -I). In certain embodiments, R3 is halogen (e.g., -Br, -I), -OR, -SR, -N(R)2, -SHg, -SO2SHg, -SHgR, -SeR, or -TeR. In certain embodiments, R3 is -OR. In certain embodiments, R3 is -SR. In certain embodiments, R3 is -N(R)2. In certain embodiments, R5 is -CX3, wherein X is halogen. In certain embodiments, R5 is -SHgR (e.g., -SHgMe), -SHg or -SO2SHg. In certain embodiments, R5 is -SeR, e.g., -SeCX3, wherein X is halogen. In certain embodiments, R5 is -TeR, e.g., -TeCX3, wherein X is halogen. In certain embodiments, R5 is Se-CBr3 or -TeBr3.

[00356] In certain embodiments of formula (iii), wherein L1 is absent, or -L1-R3 is a group of formula ..., or ..., m is 1, and each R2W is hydrogen, provided is a Base of formula (i-a), (i-b), (i-c), or (i-d):
wherein the base optionally comprises a heavy atom. In certain embodiments, the base comprises a heavy atom. In certain embodiments, the base does not comprise a heavy atom, but the sugar region or phosphate region comprises a heavy atom. In certain embodiments, R^1 and R^2 are hydrogen. In certain embodiments, R^3 is aryl substituted with at least one or more halogen atoms (e.g., -Br, -I). In certain embodiments, R^3 is halogen (e.g., -Br, -I), -OR^C, -SR^C, -N(R^C)_2, -SHg, -SO_2SHg, -SHgR^D, -SeR^D, or -TeR^D. In certain embodiments, R^3 is -OR^C. In certain embodiments, R^3 is -SR^C. In certain embodiments, R^3 is -N(R^C)_2. In certain embodiments, R^3 is -CX_3, wherein X is halogen. In certain embodiments, R^3 is -SHgR^D (e.g., -SHgMe), -SHg or -SO_2SHg. In certain embodiments, R^3 is -SeR^D, e.g., -SeCX_3, wherein X is halogen. In certain embodiments, R^3 is -TeR^D, e.g., -TeCX_3, wherein X is halogen. In certain embodiments, R^3 is Se-CBr_3 or -TeBr_3. In certain embodiments, M^4 is O. In certain embodiments, M^4 is Se.

[00357] In certain embodiments of Formula (I-d), wherein the Base is of formula (i-a), (i-b), (i-c), or (i-d), provided is a compound of Formula (I-d-i-a), (I-d-i-b), (I-d-i-c), or (I-d-i-d):

![Chemical structures](image-url)
wherein the compound comprises a heavy atom. In certain embodiments, \( R^4 \) and \( R^5 \) are hydrogen. In certain embodiments, \( G_2 \) of the sugar region is hydrogen. In certain embodiments, \( G_2 \) of the sugar region is \(-\text{SHgMe}, \ -\text{SHg} \ or \ -\text{SO}_2\text{SHg}\). In certain embodiments, \( G_2 \) of the sugar region is \(-\text{SeR}^D\), \(-\text{SeCX}_3\), wherein \( X \) is halogen. In certain embodiments, \( G_2 \) of the sugar region is \(-\text{TeR}^D\), \(-\text{TeCX}_3\), wherein \( X \) is halogen. In certain embodiments, \( G_2 \) is \(-\text{SeCBr}_3\) or \(-\text{TeBr}_3\). In certain embodiments, \( G^3 \) is a monophosphate, diphosphate, or triphosphate of formula:

\[
\text{HO-P-O-P-O-P-O-H} \quad \text{M}^1-\text{H}, \quad \text{M}^2-\text{H}, \quad \text{or} \quad \text{OH} \quad \text{OH} \quad \text{OH} \quad \text{M}^3-\text{H}.
\]

In certain embodiments, \( M^2 \) is \(-\text{O}\.\) In certain embodiments, \( R^3 \) is aryl substituted with at least one or more halogen atoms (\( \text{e.g.,} \ -\text{Br, I} \)). In certain embodiments, \( R^3 \) is halogen (\( \text{e.g.,} \ -\text{Br, I} \), \(-\text{OR}^C\), \(-\text{SR}^C\), \(-\text{N}(\text{R}^C)_2\), \(-\text{SHg} \), \(-\text{SO}_2\text{SHg}\), \(-\text{SHgR}^D\), \(-\text{SeR}^D\), or \(-\text{TeR}^D\). In certain embodiments, \( R^3 \) is \(-\text{OR}^C\). In certain embodiments, \( R^3 \) is \(-\text{SR}^C\). In certain embodiments, \( R^3 \) is \(-\text{N}(\text{R}^C)_2\). In certain embodiments, \( R^3 \) is \(-\text{CX}_3\), wherein \( X \) is halogen. In certain embodiments, \( R^3 \) is \(-\text{SHgR}^D\) (\( \text{e.g.,} \ -\text{SHgMe}\), \(-\text{SHg} \) or \(-\text{SO}_2\text{SHg}\). In certain embodiments, \( R^3 \) is \(-\text{SeR}^D\), \(-\text{SeCX}_3\), wherein \( X \) is halogen. In certain embodiments, \( R^3 \) is \(-\text{TeR}^D\), \(-\text{TeCX}_3\), wherein \( X \) is halogen. In certain embodiments, \( R^3 \) is \(-\text{SeCBr}_3\) or \(-\text{TeBr}_3\). In certain embodiments, \( M^4 \) is \( \text{O}\.\) In certain embodiments, \( M^4 \) is \( \text{Se}\.\)

[00358] In certain embodiments, wherein the Base is of formula (I-a), (I-b), (I-c), or (I-d), provided is a compound of Formula (II-d-i-a), (II'-d-i-a), (II'-d-i-b), (II'-d-i-c), and/or (H'-d-i-d):
wherein the at least one instance of (I'-d-i-a), (II'-d-i-b), (I'-d-i-c), and/or (II'-d-i-d) comprises a heavy atom. In certain embodiments, M₂ is -0-. In certain embodiments, R⁴ and R⁵ are hydrogen. In certain embodiments, G₂ of the sugar region is hydrogen. In certain embodiments, G₂ of the sugar region is -SHgMe, -SHg or -SO₂SHg. In certain embodiments, G₂ of the sugar region is -SeR², e.g., -SeCX₃, wherein X is halogen. In certain embodiments, G₂ of the sugar region is -TeR², e.g., -TeCX₃, wherein X is halogen. In certain embodiments, G₂ is Se-CBr₃ or -TeBr₃. In certain embodiments, R³ is aryl substituted with at least one or more halogen atoms (e.g., -Br, -I). In certain embodiments, R³ is halogen (e.g., -Br, -I), -OR, -SR, -N(R²)₂, -SHg, -SO₂SHg, -SHgR², -SeR², or -TeR². In certain embodiments, R³ is -OR. In certain embodiments, R³ is -SR. In certain embodiments, R³ is -N(R²)₂. In certain embodiments, R⁵ is -CX₃, wherein X is halogen. In certain embodiments, R⁵ is -SHgR² (e.g., -SHgMe), -SHg or -SO₂SHg. In certain embodiments, R⁵ is -SeR², e.g., -SeCX₃, wherein X is halogen. In certain embodiments, R⁵ is -TeR², e.g., -TeCX₃, wherein X is halogen. In certain embodiments, R⁵ is Se-CBr₃ or -TeBr₃. In certain embodiments, M⁴ is O. In certain embodiments, M⁴ is Se.

[00359] In certain embodiments, at least one instance of a Base is of formula (ii), (iv), (vi), (viii), or (x):
wherein the base optionally comprises a heavy atom. In certain embodiments, the base comprises a heavy atom. In certain embodiments, the base does not comprise a heavy atom, but the sugar region or phosphate region comprises a heavy atom. In certain embodiments, \( M^3 \) is O. In certain embodiments, \( M^4 \) is O. In certain embodiments, \( M^3 \) and \( M^4 \) are both O. In certain embodiments, \( M^3 \) is Se or Te. In certain embodiments, \( M^4 \) is Se or Te. In certain embodiments, \( R^4 \) is \(-CX_3\) wherein X is halogen.

[00360] In other embodiments, at least one instance of a Base is:

wherein the sugar region comprises a heavy atom. In certain embodiments, \( M^3 \) and \( M^4 \) are O. In certain embodiments, \( R^1 \) and \( R^2 \) are hydrogen. In certain embodiments, \( R^4 \) and \( R^5 \) are hydrogen. In certain embodiments, \( G_2 \) of the sugar region is hydrogen. In certain embodiments, \( G_2 \) of the sugar region is \(-SHgMe, -SHg \) or \(-SO_2SHg\). In certain embodiments, \( G_2 \) of the sugar region is \(-SeR^D, e.g., -SeCX_3\), wherein X is halogen. In certain embodiments, \( G_2 \) of the sugar region is \(-TeR^D, e.g., -TeCX_3\), wherein X is halogen. In certain embodiments, \( G_2 \) is Se-CBr_3 or -TeBr_3.
Exemplary Compounds of Formula (I) and (II)

Exemplary compounds of Formula (I) include, but are not limited to:
and salts thereof.

[00362] Exemplary compounds of Formula (I) include, but are not limited to:
and salts thereof.

[00363] Exemplary compounds of Formula (II), and salts thereof, comprise at least one instance of any one of the formula:
and/or salts thereof.

EXEMPLIFICATION

[00364] In order that the invention described herein may be more fully understood, the following examples are set forth. It should be understood that these examples are for illustrative purposes only and are not to be construed as limiting this invention in any manner.

DNA BASE IDENTIFICATION BY ELECTRON MICROSCOPY

[00365] Advances in DNA sequencing, based on fluorescent microscopy, have transformed many areas of biological research. However, only relatively short molecules can be sequenced by these technologies. Dramatic improvements in genomic research will require accurate sequencing of long (>10,000 base-pairs), intact DNA molecules. Our approach
directly visualizes the sequence of DNA molecules using electron microscopy. This
disclosure represents the first identification of DNA base pairs within intact DNA molecules
by electron microscopy. By enzymatically incorporating modified bases, which contain atoms
of increased atomic number, direct visualization and identification of individually labeled
bases within a synthetic 3,272 base-pair DNA molecule and a 7,249 base-pair viral genome
have been accomplished. This proof of principle is made possible by the use of a dUTP
nucleotide, substituted with a single mercury atom attached to the nitrogenous base. One of
these contrast-enhanced, heavy-atom-labeled bases is paired with each adenosine base in the
template molecule and then built into a double-stranded DNA molecule by a template-
directed DNA polymerase enzyme. This modification is small enough to allow very long
molecules with labels at each A-U position. Image contrast is further enhanced by using
annular dark-field scanning transmission electron microscopy (ADF-STEM). Further
refinements to identify additional base types and more precisely determine the location of
identified bases would allow full sequencing of long, intact DNA molecules, significantly
improving the pace of complex genomic discoveries. The inventors have published this work
in Bell et al., Micros. Microanai (2012) 18:1049-1053, published online October 9, 2012,

Introduction
[00366] Advances over the last decade have greatly improved the speed and reduced the
costs of DNA sequencing. Currently, they are limited to molecules less than 1,000 base pairs
long, principally due to the inefficiency or incomplete nature of the fluorescent labeling
reactions of "next generation" approaches (Schuster, 2008).
[00367] The approach taken in this article aims to improve read length by directly
visualizing DNA as long, intact molecules using high-resolution scanning transmission
electron microscopy (STEM). Richard Feynman (1999) famously suggested that the
incredible magnification power of electron microscopes might be harnessed to read DNA
sequence; until now this challenge had not been met. The limiting issue has not been the
small size of DNA, but the fact that the four different base types differ by only a few atoms,
and all of the differing atoms are light elements, differences particularly indistinguishable for
electron microscopy. Standard techniques used to increase sample contrast for electron
microscopy have not been able to do so in a reliably sequence specific manner, even after 40
years of effort (Gal-Or et al., 1967; ASTA, 2010).
[00368] In the present work, annular dark-field (ADF) imaging is utilized in the
monochromated, spherical-aberration corrected scanning transmission electron microscope
(MCSTEM) for high-resolution atomic identification (see FIG. 1). The ADF-STEM was the method of choice for Crewe and co-workers to originally image single heavy atoms in 1970 (Crewe, 1970; Crewe et al., 1970), anticipating that the method might be used for sequencing DNA. Recent STEM improvements now allow studies of atomic-level and single atom imaging (Batson et al., 2002; Voyles et al., 2002; Jia et al., 2003). In an ADF-STEM, a very small electron beam is raster-scanned across the sample. Most of the electrons pass through the sample with only subtle changes of energy, direction, and/or phase. However, some electrons scatter at a high angle. The high angle scattering process (Rutherford scattering) scales with the atomic number (Z) of the atom (Muller et al., 2008) raised to the power of approximately 1.5. The $Z^{1.5}$ dependence allows heavy nuclei to be definitively discriminated from light nuclei. The direct identification of unlabeled DNA base pairs, with average $Z \approx 5.5$, has proven to be difficult, and to-date unsuccessful. There is simply not enough difference between the base types to be detected without suitable contrast enhancement. Various groups have worked to overcome this problem, chiefly by chemically modifying single-stranded DNA with clusters of heavy atoms (Beer & Moudrianakis, 1962; Moudrianakis & Beer, 1965; Ottensmeyer, 1979). The approach employed here uses a standard, template-directed polymerase enzyme to incorporate heavy-atom-modified bases directly into a long DNA molecule (FIG. 2). The modification, with a single mercury atom on each thymine/uridine, provides ADF-STEM contrast substantially greater than in natural DNA. This also simplifies the challenge of making the labeling reactions sequence-specific because polymerase reactions are intrinsically sequence specific.

**Methods**

[00369] The "test pattern" DNA was built from a synthetic gene (provided by DNA 2.0, Menlo Park, CA, USA) with a 3,072 base-pair segment with all the thymines of one strand in a repeating pattern,...TNTNNNNNNNNN. ..., where T represents thymine and N represents any of the other three nucleobases. This pseudo-repeating region was amplified using flanking priming sites and standard polymerase chain reaction methods, with one standard primer and one biotinylated primer. The product was purified by centrifugation filter and bound to Dynabeads, Single-stranded DNA was obtained by denaturation, and the template strand was then used as template in a one-primer, one-cycle polymerization reaction using Bst polymerase standard reaction conditions, replacing 1 µM of dTTP with 1.5 µM CH$_3$-Hg-S-dUTP (Livingston et al., 1976). The DNA product was gel purified. Final concentration and buffer exchange was done on centrifugation filter. The efficacy of label inclusion was
tested with restriction enzymes, which were seen not to react with modified recognition sites (Banfalvi & Sarkar, 1995), confirming the presence of modifications at those sites. The DNA was also assayed by inductively coupled plasma mass spectrometry, which confirmed the presence of mercury. Single stranded M13 and primers were processed in the same manner.

Mercury-labeled DNA was deposited and aligned on an amorphous carbon film on a 400 mesh Au transmission electron microscopy (TEM) grid using a method similar to (Bensimon et al. 1994). The sample was vacuum dried for 2 min, then placed immediately into the STEM apparatus. TEM imaging was conducted on FEI T-12 TEM (FEI Company, Hillsboro, OR) at 80 kV. ADF-STEM imaging was performed by an aberration-corrected STEM, Carl Zeiss Libra 200-80kV (Carl Zeiss, Oberkochen, Germany) with Cs = -1.2 μm, 80 kV with elastic scattering using the in-coil energy-filter retaining only zero energy-loss electrons.

Results and Discussion

Double-stranded DNA was prepared that had been completely substituted with mercury labeled nucleotides on one strand, using 5-MeHgS-dUTP. This "Z-dNTP" is labeled on the 5 carbon of the uridines and is known to readily incorporate into DNA (Bridgman & Petersen, 1996), taking the place of the thymines. In this work, we have labeled M13 DNA, a 7,249 base-pair viral genome molecule, and a 3,272 base-pair synthetic molecule, with a visually identifiable "test pattern." Success with both confirmed the efficient incorporation of labels into DNA molecules substantially longer than sequenceable via other technologies.

Labeled DNA molecules were mounted on a thin supporting substrate, using a method (Bensimon et al., 1995) that separates, linearizes, and may partially stretch individual DNA molecules. FIG 3B shows a bright-field TEM image of a prepared and linearized DNA molecule on a thin amorphous carbon substrate. A critical distinguishing factor in identifying these molecules is their general morphology. Specifically, at relatively low TEM magnifications (12,000 to 80,000X), the labeled DNA molecules are seen to be 2 nm in width. In separate experiments, the lengths of the molecules match the known lengths of the M13 and test pattern molecules, with an allowance for elongation. These observed widths and lengths correspond to the known dimensions of DNA molecules; no such features are found in control samples that do not include labeled DNA.

The resulting STEM images were despeckled and thresholded to identify the features with the greatest contrast. Features that match the known morphology of linearized DNA were selected. Features that did not match known DNA morphology were not included.
in subsequent analysis, A trace was drawn over the centerline of the resulting linear features. The individual features are assessed to be individual mercury atoms, or in the case of the M13 molecules, adjacent mercury atoms. This continuous trace, including dark-field current values for both high contrast features and low contrast gaps, is shown in FIG. 4 as the dark-field current.

[00374] In ADF-STEM imaging, individual heavy atoms were clearly visible in the labeled M13 molecules. The heavy atoms create a substantially and statistically higher current in the ADF detector as the electron probe passes through these atoms (FIG. 4). The detection events are mostly distinguishable from background fluctuations, but not perfectly, with a slight overlap between event and nonevent histograms. The test pattern molecule is a synthetic gene (Villalobos et al., 2006), with a sequence specified to include two identifiable patterns. In this molecule, a pair of uridine bases, each with one heavy atom, are separated by exactly one nonlabeled base. Depending on local stretching, the distance between atoms separated by only one unlabeled base pair should be between 0.68 and 1.2 nm (Bensimon et al., 1995) (FIG. 5). This small-scale pattern repeats every 12 base pairs, creating a large-scale pattern. The large-scale pattern should have a period of 4.1 to 7.3 nm. FIG. 6 shows a region of the test pattern molecule. The ADF-STEM intensity is noticeable above the substrate background, in large part, this is due to the in-column energy filter, which selectively eliminates the inelastic scattering contribution, substantially reducing the intensity of substrate background noise. It should be noted, however, that the exact thresholding intensity-required varies between individual ADF-STEM images due to variations of the direct current offset on the ADF detector.

[00375] With image thresholds and a median filter applied, individual mercury atom labels become distinct due to their scattering intensity and size. Under these imaging conditions, only individual atoms, or clusters of atoms, are capable of producing such high contrast in such a small cross-sectional area. In this case, these features are known to be individual atoms, rather than columns, because the morphology is independent of viewing angle. Moreover, the track of the atoms follows an approximately linear pattern, which corresponds to the known position of the linearized DNA. This allows us to conclude not only that the features represent individual, high-/low- energy atoms, but that those atoms are collocated with the DNA molecule.

[00376] Every heavy atom in this sequence is in a location predicted by the pattern built into the synthetic sequence. While there are missing mercury atoms, there are none where they should not be. More specifically, within this 180 base-pair segment of DNA, the test
pattern repeats, in part, 15 times. Of the 30 predicted labels, 17 are present. Fifteen occur in positions predicted by proximity to neighbors on both sides; the other two labels appear in locations predicted by neighbors on only one side. These two also are larger in cross section than the others, probably indicating a tangle that has added to the contrast of the mercury atom, and shortened the distance on one side of the label. All labels are found within the DNA molecule itself. The mercury atom contrast is consistent and statistically above the contrast found either outside the molecule or elsewhere within the molecule.

A fraction of the labels are missing due to thermal damage; the samples are heated prior to imaging to prevent adsorbed materials from interfering with STEM imaging. Greater heating has been observed to drive off a higher proportion of the labeling atoms. Nevertheless, the labels that remain, and the corresponding gaps between them, follow precisely the predicted pattern from the synthetic DNA. The smaller scale pattern is expected to have a characteristic distance of 0.7 to 1.2 nm. The larger scale pattern is predicted to have a pitch of between 4.1 and 7.3 nm. In fact, three characteristic modes are observed, around 1 nm, around 4.5 to 7.5 nm, and between 14 and 16 nm. All three of these distances match the test pattern. Where doublets are intact, the mercury atoms are seen to be between 0.7 and 1.1 nm apart (FIG. 6A). These measurements closely match the predicted spacing of the small-scale test pattern. The cluster of spacing between 4.5 and 7.5 nm matches the predicted large-scale pattern. This is more varied than the doublet pattern, likely because these spacing represent a thermal loss of one of the atoms in the doublet; the distances between them depend on which atom of the two was lost. The mode at 14 to 16 nm is very close to twice the large-scale pattern and corresponds to instances in which both atoms of a small-scale pair were lost.

The variability within the smaller scale pattern results in less ambiguity in determining the local sequence of the DNA molecule than variability in the larger scale pattern. This suggests a potential limit in the use of this technique for determining local sequence. DNA base pairs have a nominal linear pitch of 0.34 nm. If the distance between labeled bases is less than 1 nm or so, the number of unlabeled bases this distance indicates is likely to be fairly certain, whereas if the distance is greater, the number of unlabeled bases in between could become unpractically ambiguous. A higher labeling density might be able to overcome this problem, either using the same label for multiple base types or distinct labels to identify distinct base types. In certain embodiments, using the same label for multiple base types would require parallel experiments to deduce actual sequence. For example, labeling C's and T's in one experiment, then C's and A's in the next, then combining the information
to deduce the identity of the bases. However, using different labels for different bases would
avoid the issue of parallel experiments. Using different labels for distinct bases allows for
differentiating distinct signals, either by number of atoms in the label, or by atomic number
of individual labeling atoms.

[00379] Simultaneous use of multiple labeled bases allows for greater sequence information
determination from a DNA molecule. For example, complete labeling of a molecule, with
distinct labels for each of the four base types, allows for total sequence information to be
extracted. If one of the bases were unlabeled, such that the other three were differentially
labeled, the identity of the fourth could be determined by the absence of labels in a given
location. However, this could become limiting in certain conditions, such as an extended
segment (homopolymer read) section in which multiple unlabeled bases were all in a row. It
would be knowable that the region corresponds only to the unlabeled base, but the precise
number could be ambiguous. With fewer than three labeled bases, the challenge of
interpreting unlabeled regions would become more difficult and require Sanger-style
evaluation of multiple molecules representing the same region in order to extract complete
sequence information.

References
National Human Genome Research Institute. Available at genome.gov/27541189.
susceptibility to cleavage by restriction endonucleases. DNA Cell Biol 14, 5.
(1994). Alignment and sensitive detection of DNA by a moving interface. Science 265, 2096-
2098.
mercurated dUTP. J Sequencing and Mapping 6, 199-209.


SYNTHESIS OF EXEMPLARY HEAVY-ATOM LABELED NUCLEOTIDES

Example 1. Selenium-labeled base

Example 2. Selenium-labeled sugar
Example 3. Tellurium labeled base

3.1

3.2

3.4

3.3
Example 4. Tellurium labeled sugar

Example 5. Mercury labeled base
Example 6. Iodine labeled base

Example 7. Iodine labeled base [dUTP-Ethyne-PhI]
Reagents and Conditions: a) HC≡C-TMS, Et₃N, Pd(0), Cul, DMF, rt, 12h; b) K₂C₈O₃, THF/MeOH; c) 1,4-Diiodobenzene, Et₃N, Pd(0), Cul, DMF, rt, 12h; d) i. (MeO)₃P, Proton sponge, POCl₃, 0 °C - rt, 3h; ii. Pyrophosphate, iPr-n-BuNH, DMF, 0 °C - rt, 1h; iii. TEAB, rt, lh.

5-(4-Iodophenylethynyl)-2'-deoxyuridine: 5-Ethynyl-dU (0.17 g, 0.674 mmol) was dissolved in DMF (5 mL) and maintained under nitrogen atmosphere. To this solution NEt₃ (0.427 mL, 3.03 mmol), 1,4-diiodobenzene (1.12 g, 3.37 mmol), Pd(Ph₃)₄ (78 mg, 0.068 mmol) and Cul (25.7 mg, 0.135 mmol) were added sequentially with stirring under nitrogen. The reaction was continued at rt for 2 h and TLC (10% MeOH in DCM) and LCMS (ES+) indicated complete disappearance of starting material. After removing the solvent under reduced pressure, the residue was chromatographed on silica gel column using 0 - 20% MeOH gradient over DCM to get pure product (0.178 g, 58%). TLC: (10% MeOH in DCM): Rᵣ = 0.68. LCMS (ES+): (M+H) calculated mass: 455.22 and observed mass: 454.67. 1H-NMR (DMSO-d₆): δ 11.70 (bs, 1H, 3'-NH-), 8.37 (s, 1H, 6-H), 7.75 (d, 2H, Ar-H), 7.23 (d, 2H, Ar-H), 6.09 (t, 1H, 1'-H), 5.44 (d, 1H, 2'-OH), 5.28 (t, 1H, 5'-OH), 4.24 (m, 1H, 4H), 3.53 - 3.80 (m, 3H, 3'-H, 5'-H & 5''-H), 2.45 (m, 1H, 2'-H), 2.14 (m, 1H, 2' & 2''-H).

5-(4-Iodophenylethynyl)-2'-deoxyuridine-5'-triphosphate: 5-(4-Iodophenylethynyl)-2'-deoxyuridine (150 mg, 0.33 mmol) and proton sponge (108 mg, 0.495 mmol) were dissolved in trimethylphosphate (2 mL) and cooled to -10 °C and maintained under a nitrogen atmosphere. POCl₃ (63 µL, 0.66 mmol) was added, reaction mixture brought to rt and stirred for lh. A solution of bis-tri-w-butylammonium pyrophosphate (0.784 g, 1.65 mmol) and tri-w-butylamine (0.397 mL, 1.65 mmol) in anhydrous DMF (3 mL) was added at 0 °C. After stirring for 2h at rt, TEAB buffer (0.5 M, pH 8.5; 15 mL) was added. The reaction was stirred at room temperature for 30 min and lyophilized. The residue was dissolved in water (2 mL), filtered, and purified on RP-HPLC using 50 mM TEAB buffer pH 8.5 and...
acetonitrile to yield triphosphate product. LCMS (ES−): [M-H] calculated mass was 693.16, and observed mass was 692.51.

**[00383]** Compounds which may be synthesized following the Example 7 procedure include:

![dCTP-Ethyne-Ph](image)

**Example 8. Iodine and Bromine labeled base fdUTP-Ethyne-hBr^+**

![Chemical structures](image)

**[00384]** **Reagents and Conditions:**

a) i-BuONO, TMS-N₃, ACN, 0 °C - rt; b) CuSO₄, Na ascorbate, THF/H₂O/t-BuOH 3:1:1; c) (i) POCl₃, (MeO)₃P, Proton sponge, 0 °C - rt, 2h; (ii) Pyrophosphate, i-n-BuNH, DMF, 0 °C - rt, lh; (iii) TEAB, rt, lh.

**[00385]** **1-Azido-2,4,6-tribromo-3,5-diodobenzene:** 2,4,6-Tribromo-3,5-diiodoaniline (272 mg, 0.468 mmol) was suspended in CH₃CN (5 mL) and cooled to 0°C in an ice bath. To
this stirred mixture was added i-BuONO (247 µL, 1.87 mmol) followed by TMSN₃ (196 µL, 1.40 mmol) dropwise. The resulting suspension was stirred at room temperature for 2 h. The solvent and volatile reagents were redistilled under reduced pressure and the product was dried under high vacuum. TLC (10% EtOAc in hexanes) indicated product (brown solid, 296 mg, 94%) was pure and used without further purification. TLC: 10% EtOAc in hexanes, Rf: 0.92. LCMS (ES+): M+H calculated mass 609.62, and observed mass 609.29. ¹³C-NMR (DMSO-d₆): δ 135.86, 135.57, 128.92, 116.77, 116.78, 109.93, 93.07, 90.22, 76.06, 66.77 and 60.32.

[00386] 5-[l-(2,4,6-Tribromo-3,5-diiodobenzene)-l,2,3-triazol-4-yl]-2'-deoxyuracil or 1-(2-Deoxy-β-D-erythro-3-hydroxyfuranosyl)-5-[l-(2,4,6-Tribromo-3,5-diiodobenzene)-l,2,3-triazol-4-yl]uracil: 5-Ethynyl-dU (0.12 g, 0.476 mmol) and 1-azido-2,4,6-tribromo-3,5-diiodobenzene (0.29 g, 0.476 mmol) were dissolved in THF/H2O/i-BuOH (3:1:1, v/v, 5 mL). To this stirring solution, freshly prepared 1 M sodium ascorbate solution in water (243 µL, 0.238 mmol), followed by CuSO4·5H2O 7.5% in water (396 µL, 0.119 mmol) were added at rt. The reaction mixture was stirred for 20 h at room temperature. The solvent was evaporated, and the residue was purified by silica gel column chromatography using 2 - 20% MeOH over DCM to get pure product (0.226 g, 55%) as a brown solid. TLC: 10% MeOH in DCM: Rf: 0.58. LCMS (ES+): [M+H] calculated mass 859.84, and observed mass 859.23. 1H-NMR (DMSO-d₆): δ 11.72 (1H, -NH-), 8.72 (s, 1H, 6-H), 8.66 (s, 1H, Triazole-H), 6.24 (t, 1H, 1'-H), 5.29 (d, 1H, 2'-OH), 5.08 (t, 1H, 5'-OH), 4.29 (m, 1H, 4'-H), 3.86 (m, 1H, 5''-H), 3.58 - 3.64 (m, 2H, 3'-H, 5'''-H), 2.20 (m, 2H, 2' & 2''-H). ¹³C-NMR (DMSO-d₆): δ 166.42, 155.04, 146.66, 144.72, 141.95, 139.52, 137.13, 137.08, 128.92, 116.77, 116.78, 109.93, 93.07, 90.22, 76.06, 66.77 and 60.32.

[00387] 5-[l-(2,4,6-Tribromo-3,5-diiodobenzene)-l,2,3-triazol-4-yl]-2'-deoxyuracil-5'-triphosphate (10): 5-[l-(2,4,6-Tribromo-3,5-diiodobenzene)-l,2,3-triazol-4-yl]-2'-deoxyuracil or 1-(2-Deoxy-P-D-erythro-3-hydroxyfuranosyl)-5-[l-(2,4,6-Tribromo-3,5-diiodobenzene)-l,2,3-triazol-4-yl]uracil (75 mg, 0.087 mmol) and proton sponge (27 mg, 0.131 mmol) were dissolved in trimethylphosphate (2 mL) and cooled to -10 °C and maintained under a nitrogen atmosphere. POCl₃ (17 µL, 0.175 mmol) was added, reaction mixture brought to rt and stirred for lh. A solution of bis-tri-n-butylammonium pyrophosphate (207 mg, 0.436 mmol) and tri-w-butylamine (105 µL, 0.436 mmol) in anhydrous DMF (3.0 mL) was added at 0 °C. After stirring for 2h at rt, triethylammonium bicarbonate (TEAB) buffer (0.5 M, pH 8.5; 15 mL) was added. The reaction was stirred at room temperature for 30 min and lyophilized. The residue was dissolved in water (3 mL),
filtered, and purified on RP-HPLC using 50 mM TEAB buffer pH 8.5 and acetonitrile to yield the triphosphate product. LCMS (ES-): [M-H] calculated mass 1097.78, and observed mass 1097.50. $^3$P-NMR ($D_2$O): $\delta$ 8.35, 8.90 and -20.70.

**Example 9. Iodine labeled base fdCTP-Ph-I$_2$**

![Chemical structures](image)

**[00388]** **Reagents and Conditions:**

a) 1,3,5-Triiodobenzene, Et$_3$N, Pd(Ph$_3$P)$_4$, Cul, DMF, rt, 12h; b) i. (MeO)$_3$P, Proton sponge, POCl$_3$, 0 °C - rt, 3h; ii. Pyrophosphate, $i$rt-n-BuNH, DMF, 0 °C - rt, lh; iii. TEAB, rt, lh.

**[00389]** **5-(3,5-Diiodophenylethynyl)-2'-deoxycytidine:** Ethynyl-dC was synthesized following the procedure of Dodd *et al.*, *Org. Biomol. Chem.* (2010) 8:663-6665. Ethynyl-dC (0.2 g, 0.796 mmol) was then dissolved in DMF (10 mL) and maintained under nitrogen atmosphere. To this solution NEt$_3$ (0.1.12 mL, 7.96 mmol), 1,3,5-triiodobenzene (1.11 g, 2.39 mmol), Pd(Ph$_3$P)$_4$ (92 mg, 0.080 mmol) and Cul (31 mg, 0.16 mmol) were added sequentially with stirring under nitrogen. The reaction was continued at rt for 2 h and TLC (10% MeOH in DCM) and LCMS (ES+) indicated complete disappearance of starting material. After removing the solvent under reduced pressure, the residue was chromatographed on silica gel column using 0 - 20% MeOH gradient over DCM) to get pure product (0.324 g, 70%). TLC:
(10% MeOH in DCM): R_f = 0.52. LCMS (ES+): (M+H) calculated mass: 579.14 and observed mass: 579.31. 1H-NMR (DMSO-d_6): δ 8.82 (bs, 2H, 4-NH_2), 8.38 (s, 1H, 6-H), 8.06 (t, 1H, Ar-H), 8.00 (d, 2H, Ar-H), 6.10 (t, 1H, l’-H), 5.22 (d, 1H, 2'-OH), 5.13 (t, 1H, 5'-OH), 4.22 (m, 1H, 4H), 3.80 (m, 1H, 3'-H), 3.63 - 3.68 (m, 1H, 5'-H), 3.55 - 3.61 (m, 1H, 5''-H), 2.15 - 2.22 (m, 1H, 2'-H), 1.98 - 2.04 (m, 1H, 2''-H).

[00390] 5-(3,5-Diiodophenylethynyl)-2'-deoxycytidine-5'-triphosphate: 5-(3,5-Diiodophenylethynyl)-2'-deoxycytidine (226 mg, 0.39 mmol) and proton sponge (128 mg, 0.585 mmol) were dissolved in trimethylphosphate (6 mL) and cooled to -10 °C and maintained under a nitrogen atmosphere. POCl_3 (73 µL, 0.78 mmol) was added, reaction mixture brought to rt and stirred for 1h. A solution of bis-tri-w-butylammonium pyrophosphosphate (0.927 g, 1.95 mmol) and tri-w-butylamine (0.47 mL, 1.95 mmol) in anhydrous DMF (3 mL) was added at 0 °C. After stirring for 1.5h at rt, TEAB buffer (0.5 M, pH 8.5; 30 mL) was added. The reaction was stirred at room temperature for 30 min and lyophilized. The residue was dissolved in water (3 mL), filtered, and purified on RP-HPLC using 50 mM TEAB buffer pH 8.5 and acetonitrile to yield the triphosphate. LCMS (ES-): [M-H] calculated mass was 818.08, and observed mass was 817.98.

[00391] Compounds which may be synthesized following the Example 9 procedure include:
Example 10. Te labeled base fdATP-Ethyne-Te-Ph

Reagents and Conditions: a) TMSC≡CH, Et₃N, Pd(0), Cul, DMF, rt, 12h; b) K₂C₅, THF/MeOH 3:1; c) Ph-Te-Te-Ph, Cul, K₂C₅, DMSO, rt, 12h; d) (i) POCl₃, (MeO)₃P, Proton sponge, 0 °C - rt, 2h; (ii) Pyrophosphate, tri-n-BuNH, DMF, 0 °C - rt, lh; (iii) TEAB, rt, lh.

Synthesis of 7-Deaza-7-(phenyltelluro)ethynyl-2'-deoxyadenosine: Deaza-7-ethynyl-2'-dA synthesized following Seela and Zulauf, Synthesis (1996) 726 - 730. A mixture of 7-Deaza-7-ethynyl-2'-dA (0.19 g, 0.69 mmol), (PhTe)₂ (0.14 g, 0.34 mmol) Cul (13.1 mg, 0.069 mmol) and K₂C₅ (0.19 g, 1.37 mmol) in 1.0 mL of commercial grade, undried DMSO was stirred at room temperature for the 30 min. TLC (solvent: MeOH/DCM 1:9) and LCMS (ES+) indicated disappearance of most of the starting material and product formation. Continued stirring at rt for overnight, reaction was quenched with water (1 mL) and lyophilized to dryness. Residue was dissolved in 10%MeOH in DCM and purified on silica gel using 0 - 20% MeOH gradient over DCM to get pure compound 5 (0.29 g, 64%). TLC (1:9 MeOH/DCM): Rₜ = 0.74. LCMS (ES+): (M+H) calculated: 478.98 and observed:
478.71. 1H-NMR (DMSO-d$_6$): δ 8.10 (s, IH, 2-H), 7.77 (m, 3H, Ar-H), 7.27 (m, 2H, Ar-H), 6.44 (m, 2H, l'-H & 8-H), 5.43 (bs, IH, 2'-OH), 5.20 (bs, IH, 5'-OH), 4.32 (bs, 2H, 6-NH$_2$), 3.48 - 3.85 (m, 4H, 3'-H, 4'-H, 5'-H & 5''-H), 2.45 (m, IH, 2'-H), 2.19 (m, IH, 2''-H).

[00394] Synthesis of 7-Deaza-7-(phenyltelluro)ethynyl-2'-deoxyadenosine-5'-triphosphate: 7-Deaza-7-(phenyltelluro)ethynyl-2'-dA (60 mg, 0.126 mmol) and proton sponge (41 mg, 0.188 mmol) were dissolved in trimethylphosphate (1 mL) and cooled to -10 oC and maintained under a nitrogen atmosphere. POCl$_3$ (23 uL, 0.251 mmol) was added, reaction mixture brought to rt and stirred for 1h. A solution of bis-tri-n-butylammonium pyrophosphate (0.45g, 0.941 mmol) and tri-n-butylamine (0.23 mL, 0.941 mmol) in anhydrous DMF (1.5 mL) was added at 0 oC. After stirring for 1h at rt, triethylammonium bicarbonate buffer (0.5 M, pH 8.5; 15 mL) was added. The reaction was stirred at room temperature for 1h and lyophilized. The residue was dissolved in water (2.0 mL), filtered, and purified on RP-HPLC using 50 mM TEAB buffer pH 8.5 and acetonitrile to yield triphosphate. LCMS (ES-): [M-H] calculated mass was 716.92, and observed mass was 716.35.

Example 11. Te labeled base fdUTP-Ethyne-Te-Ph

[00395] Reagents and Conditions: a) TMSC≡CH, Et$_3$N, Pd(0), Cul, DMF, rt, 12h; b) K$_2$CO$_3$, THF/MeOH 3:1; c) Ph-Te-Te-Ph, Cul, K$_2$CO$_3$, DMSO, rt, 12h; d) (i) POCl$_3$,
(MeO)$_3$P, Proton sponge, 0 °C - rt, 2h; (ii) Pyrophosphate, tri-n-BuNH, DMF, 0 °C - rt, lh; (iii) TEAB, rt, lh.

**[00396]** **5-(Phenyltelluro)ethynyl-2'-deoxyuridine:** 5-Ethynyl-2'-dU was synthesized following the procedure of Yu, Synlett (2000) 86-88. A mixture of 5-Ethynyl-2'-dU (50 mg, 0.17 mmol), (PhTe)$_2$ (42 mg, 0.1 mmol), Cul (4 mg, 0.02 mmol) and K$_2$C0$_3$ (55 mg, 0.4 mmol) in 1.0 mL of commercial grade, undried DMSO was stirred at room temperature for the 30 min. TLC (solvent: MeOH/DCM 1:9) and LCMS (ES+) indicated disappearance of most of the starting material and product formation. Continued stirring at rt for overnight, reaction was quenched with water (1 mL) and lyophilitized to dryness. Residue was dissolved in 10%MeOH in DCM and purified on silica gel using 0 - 20% MeOH gradient over DCM to get pure compound (0.38 g, 42%). TLC (10% MeOH in DCM): R$_f$ = 0.63. LCMS (ES+): (M+H) calculated: 456.93 and observed: 456.82. 1H-NMR (DMSO-$d_6$): δ 8.26 (s, 1H, 6-H), 7.78 (m, 2H, Ar-H), 7.28 (m, 3H, Ar-H), 6.11 (t, 1H, 1'-H), 5.29 (bs, 1H, 2'-OH), 5.14 (bs, 1H, 5'-OH), 4.24 (bs, 1H, 3-NH-), 3.80 (m, 1H, 4'-H) 3.54 - 3.56 (m, 2H, 3'-H, 5'-H), 3.29 (m, 1H, 5''-H), 2.14 (m, 2H, 2' & 2''-H).

**[00397]** **5-(Phenyltelluro)ethynyl-2'-deoxyuridine-5'-triphosphate:** 5-(Phenyltelluro)ethynyl-2'-dU (26 mg, 0.057 mmol) and proton sponge (18.7 mg, 0.086 mmol) were dissolved in trimethylphosphate (1 mL) and cooled to -10 °C and maintained under a nitrogen atmosphere. POCl$_3$ (11 µL, 0.115 mmol) was added, reaction mixture brought to rt and stirred for lh. A solution of bis-tri-n-butylammonium pyrophosphate (135 mg, 0.285 mmol) and tri-n-butylamine (69 µL, 0.285 mmol) in anhydrous DMF (1.0 mL) was added at 0 °C. After stirring for 1h at rt, triethylammonium bicarbonate (TEAB) buffer (0.5 M, pH 8.5; 10 mL) was added. The reaction was stirred at room temperature for 30 min and lyophilized. The residue was dissolved in water (2 mL), filtered, and purified on RP-HPLC using 50 mM TEAB buffer pH 8.5 and acetonitrile to yield triphosphate product. LCMS (ES+): [M+H] calculated mass was 696.87, [M-sugar] calculated mass was 338.80, and observed mass was 338.21.
Example 12. SHsMe labeled base fdCTP-SHsMel

![Chemical Structures](https://example.com/structures.png)

[00398] **Reagents and Conditions:**

a) Propargyl S-benzoate, Et₃N, Pd(PPh₃)₄, Cul, DMF, rt, 12h; b) (i) POCl₃, (MeO)₃P, Proton sponge, 0 °C - rt, 2h; (ii) Pyrophosphate, iPr-n-BuNH, DMF, 0 °C - rt, 2h; (iii) TEAB, rt, 1h; c) NH₄OH, rt, 2h; d) (i) TCEP, H₂O, 30 min; (ii) MeHgOH, H₂O, rt, 1h.

[00399] **5-(Propargylthiobenzoate)-2'-deoxycytidine:** 5-Iodo-dC (0.25 g, 0.708 mmol) was suspended in DMF (5 mL) and maintained under nitrogen atmosphere. To this solution NEt₃ (0.45 mL, 3.2 mmol), propargyl S-thiobenzoate (0.374 g, 2.12 mmol), Pd(PPh₃)₄ (82 mg, 0.071 mmol) and Cul (27 mg, 0.142 mmol) were added sequentially with stirring under nitrogen. The reaction mixture was stirred at rt for 12 h. TLC (10% MeOH in DCM) and LCMS (ES+) indicated complete disappearance of starting material. After removing the solvent under reduced pressure, the residue was chromatographed on silica gel column using 0 - 20% MeOH gradient over DCM) to get pure product (0.14 g, 49%). TLC: (10% MeOH in DCM): Rₜ = 0.45. LCMS (ES+): (M+H) calculated mass: 402.45 and observed mass: 402.14. 1H-NMR (CD₃OD): δ 8.08 (bs, 2H, 4-NH₂), 8.31 (s, 1H, 6-H), 7.97 (dd, 2H, Ar-H), 7.66 (m, 1H, Ar-H), 7.54 (m, 2H, Ar-H), 6.19 (t, 1H, l'-H), 4.89 (bs, 1H, 2'-OH), 4.59 (bs, 1H, 5'-OH), 4.35 (m, 1H, 4'-H), 3.94 (m, 1H, 3'-H), 3.70 - 3.83 (m, 2H, 5'&5"-H), 2.38 (m, 1H, 2'-H), 2.13 (m, 1H, 2"-H).

[00400] **5-(Propargylthiobenzoate)-2'-deoxycytidine-5'-triphosphate:** 5-(Propargylthiobenzoate)-2'-deoxycytidine (135 mg, 0.336 mmol) and proton sponge (108 mg,
0.504 mmol) were dissolved in trimethylphosphate (2 mL) and cooled to -10 °C and maintained under a nitrogen atmosphere. POCl₃ (63 µL, 0.673 mmol) was added, reaction mixture brought to rt and stirred for lh. A solution of bis-tri-w-butylammonium pyrophosphate (0.799 g, 1.68 mmol) and tri-w-butylamine (0.405 mL, 1.68 mmol) in anhydrous DMF (3 mL) was added at 0 °C. After stirring for 1 h at rt, TEAB buffer (0.5 M, pH 8.5; 15 mL) was added. The reaction was stirred at room temperature for 30 min and lyophilized. The residue was dissolved in water (3 mL), filtered, and purified on RP-HPLC using 50 mM TEAB buffer pH 8.5 and acetonitrile to yield the triphosphate product (16.4 mg, 7.6%) as white solid. LCMS (ES+): [M+H] calculated mass was 642.39, and observed mass was 642.34.

[00401] 5-(Propargyl S-thiomethylmercury)-2'-deoxycytidine-5'-triphosphate: 5-(Propargylthiobenzoate)-2'-deoxycytidine-5'-triphosphate (8.5 mg, 0.013 mmol) was dissolved in ammonium hydroxide 28 - 30% (5 mL), capped the vial tightly and occasionally agitated for 1 h at room temperature. LCMS (ES+) indicated complete disappearance of starting material. Ammonia was removed under vacuum and lyophilized. The resulting solid was dissolved in water (1 mL) and TCEP (8.2, 0.028 mmol) was added and agitated for 30 min at room temperature. To this solution methylmercury (II) hydroxide (1M, 60 µL, 56 µmol) was added, and agitated occasionally for 1 h at room temperature. The solution was purified on RP-HPLC using 50 mM TEAB buffer pH 8.5 and acetonitrile to yield triphosphate product. LCMS (ES-): [M-H] calculated mass was 750.89, and observed mass was 750.83.

**Example 13. Iodine labeled base [dUTP-Ethyne-Phli]**
**00402** Reagents and Conditions: a) H≡C-TMS, Et$_3$N, Pd(0), Oil, DMF, rt, 12h; b) K$_2$CO$_3$, THF/MeOH; c) 1,2,4,5-Tetraiodobenzene, Et$_3$N, Pd(Ph$_3$P)$_4$, Cul, DMF, rt, 12h; d) i. (MeO)$_3$P, Proton sponge, POCl$_3$, 0 °C - rt, 3h; ii. Pyrophosphate, tri-n-BuNH, DMF, 0 °C - rt, lh; iii. TEAB, rt, lh.

**00403** 5-(2,4,5-Triiodophenylethynyl)-2'-deoxyuridine: 5-Ethynyl-dU was synthesized following the procedure of Yu, *Synlett* 2000, 86-88. 1,2,4,5-tetraiodobenzene may be synthesized following the procedure of Mattern, *J. Org. Chem.*, 1983, 48, 4773-4774.

**00404** 5-Ethynyl-dU (0.1 g, 0.396 mmol) was suspended in DMF (10 mL) and maintained under nitrogen atmosphere. To this solution NEt$_3$ (0.558 mL, 3.96 mmol), 1,2,4,5-tetraiodobenzene (0.941 g, 1.59 mmol), Pd(Ph$_3$P)$_4$ (46 mg, 0.04 mmol) and Cul (16 mg, 0.08 mmol) were added sequentially with stirring under nitrogen. The reaction mixture was heated to 100 °C and the solution was stirred for 2 h. TLC (10% MeOH in DCM) and LCMS (ES+) indicated complete disappearance of starting material. Reaction mixture was cooled to rt and after removing the solvent under reduced pressure, the residue was chromatographed on silica gel column using 0 - 20% MeOH gradient over DCM) to get pure product (0.106 g, 38%).

TLC: (10% MeOH in DCM): R$_f$ = 0.68. LCMS (ES+): (M+H) calculated mass: 707.02 and observed mass: 706.36. 1H-NMR (DMSO-d$_6$): δ 11.71 (bs, 1H, 3-NH-), 8.44 (s, 1H, 6-H), 8.37 (s, 1H, Ar-H), 7.90 (s, 1H, Ar-H), 6.11 (t, 1H, l'-H), 5.24 (d, 1H, 2'-OH), 5.07 (t, 1H, 5'-OH), 4.24 (m, 1H, 4-H), 3.81 (m, 1H, 3'-H), 3.55 - 3.67 (m, 2H, 5'&5''-H), 2.17 - 2.22 (m, 2H, 2'&2''-H).

**00405** 5-(2,4,5-Triiodophenylethynyl)-2'-deoxycytidine-5'-triphosphate: 5-(2,4,5-Triiodophenylethynyl)-2'-deoxyuridine (90 mg, 0.127 mmol) and proton sponge (42 mg, 0.191 mmol) were dissolved in trimethylphosphate (3 mL) and cooled to 0 °C and maintained under a nitrogen atmosphere. POCl$_3$ (24 µL, 0.255 mmol) was added, reaction mixture brought to rt and stirred for 2h. Reaction became clear and LCMS (ES-) indicated monophosphate formation. A solution of bis-tri-w-butylammonium pyrophosphate (0.303 g,
0.637 mmol) and tri-w-butylamine (0.153 mL, 0.637 mmol) in anhydrous DMF (3 mL) was added at 0 °C. After stirring at rt for 2 h, TEAB buffer (0.5 M, pH 8.5; 30 mL) was added. The reaction was stirred at room temperature for 30 min and lyophilized. The residue was dissolved in water (3 mL), filtered, and purified on RP-HPLC using 50 mM TEAB buffer pH 8.5 and acetonitrile to yield triphosphate product. LCMS (ES-): [M-H] calculated mass was 944.96, and observed mass was 944.29.

**Example 14. Se labeled base**

![Chemical structures](image)

**[00406] Reagents and Conditions:** a) 1,3-Dichloro-1,1,3,3-tetraisopropylsiloxane, DMAP; b) POCl$_3$, with heat; c) NaHSe, DMF, with heat; d) TABF; e) i) (MeO)$_3$P=0/POCl$_3$, Proton sponge, ii) tetrabutylammonium pyrophosphate
OTHER EMBODIMENTS

[00407] In the claims articles such as "a," "an," and "the" may mean one or more than one unless indicated to the contrary or otherwise evident from the context. Claims or descriptions that include "or" between one or more members of a group are considered satisfied if one, more than one, or all of the group members are present in, employed in, or otherwise relevant to a given product or process unless indicated to the contrary or otherwise evident from the context. The invention includes embodiments in which exactly one member of the group is present in, employed in, or otherwise relevant to a given product or process. The invention includes embodiments in which more than one, or all of the group members are present in, employed in, or otherwise relevant to a given product or process.

[00408] Furthermore, the invention encompasses all variations, combinations, and permutations in which one or more limitations, elements, clauses, and descriptive terms from one or more of the listed claims is introduced into another claim. For example, any claim that is dependent on another claim can be modified to include one or more limitations found in any other claim that is dependent on the same base claim. Where elements are presented as lists, e.g., in Markush group format, each subgroup of the elements is also disclosed, and any element(s) can be removed from the group. It should be understood that, in general, where the invention, or aspects of the invention, is/are referred to as comprising particular elements and/or features, certain embodiments of the invention or aspects of the invention consist, or consist essentially of, such elements and/or features. For purposes of simplicity, those embodiments have not been specifically set forth in haec verba herein. It is also noted that the terms "comprising" and "containing" are intended to be open and permits the inclusion of additional elements or steps. Where ranges are given, endpoints are included. Furthermore, unless otherwise indicated or otherwise evident from the context and understanding of one of ordinary skill in the art, values that are expressed as ranges can assume any specific value or sub-range within the stated ranges in different embodiments of the invention, to the tenth of the unit of the lower limit of the range, unless the context clearly dictates otherwise.

[00409] This application refers to various issued patents, published patent applications, journal articles, and other publications, all of which are incorporated herein by reference. If there is a conflict between any of the incorporated references and the instant specification, the specification shall control. In addition, any particular embodiment of the present invention that falls within the prior art may be explicitly excluded from any one or more of the claims. Because such embodiments are deemed to be known to one of ordinary skill in the art, they may be excluded even if the exclusion is not set forth explicitly herein. Any particular
embodiment of the invention can be excluded from any claim, for any reason, whether or not related to the existence of prior art.

[00410] Those skilled in the art will recognize or be able to ascertain using no more than routine experimentation many equivalents to the specific embodiments described herein. The scope of the present embodiments described herein is not intended to be limited to the above Description, but rather is as set forth in the appended claims. Those of ordinary skill in the art will appreciate that various changes and modifications to this description may be made without departing from the spirit or scope of the present invention, as defined in the following claims.
CLAIMS

What is claimed is:

1. A heavy-atom labeled nucleotide of Formula (I):

   \[ \text{G}_3 - \text{M}^1 - \text{G}_1 \text{Base} - \text{G}_2 \]  

   or a salt thereof;

wherein:

   each instance of \( \text{G}_i \) is independently \(-0-, -S-, -Se-, -\text{CH}_2^-, \text{or} -\text{NH}^-\);

   each instance of \( \text{G}_2 \) is independently hydrogen, halogen, \(-\text{OR}^A, -\text{SR}^A, -\text{N(R}^A)_2, -\text{SHg}, -\text{SO}_2^2\text{Hg}, -\text{SeR}^D \text{ or} -\text{TeR}^D; \)

   each instance of \( \text{R}^A \) is independently hydrogen, substituted or unsubstituted \( \text{C}^\text{oalkyl}, \)
   substituted or unsubstituted \( \text{C}_2\text{-oalkenyl}, \) substituted or unsubstituted \( \text{C}_2\text{-alkynyl}, \)
   substituted or unsubstituted carbocyclyl, substituted or unsubstituted heterocyclyl, substituted or unsubstituted aryl, or substituted or unsubstituted heteroaryl, an oxygen protecting group when attached to an oxygen atom, a sulfur protecting group when attached to a sulfur atom, a nitrogen protecting group when attached to a nitrogen atom; or two \( \text{R}^A \) groups are joined to form a substituted or unsubstituted heterocyclic or substituted or unsubstituted heteroaryl ring;

   each instance of \( \text{M}^1 \) is independently \(-0-, -S-, -\text{NH}^-, -\text{Se}-, \text{or} -\text{C(M}^\text{M}^2)\text{H}^-\), wherein each instance of \( \text{R}^M \) is independently hydrogen or halogen;

   each instance of \( \text{G}_3 \) is independently hydrogen, hydrogen, substituted or unsubstituted \( \text{C}_2\text{-oalkyl}, \) substituted or unsubstituted \( \text{C}_2\text{-oalkenyl}, \) substituted or unsubstituted \( \text{C}_2\text{-alkynyl}, \)
   substituted or unsubstituted carbocyclyl, substituted or unsubstituted heterocyclyl, substituted or unsubstituted aryl, or substituted or unsubstituted heteroaryl, or a monophosphate, diphosphate, or triphosphate of formula:

   \[
   \begin{align*}
   & \text{HO-P-O-P-O-P-} \quad \text{HO-P-O-P-} \quad \text{HO-P-O-P-} \\
   & \text{M}^\text{H}, \quad \text{OH} \quad \text{M}^\text{H}^2, \quad \text{or} \quad \text{OH} \quad \text{OH} \quad \text{M}^\text{H}
   \end{align*}
   \]

   wherein each instance of \( \text{M}^2 \) is independently \(-0-, -S-, \text{or} -\text{Se}^-\); and

   each instance of \( \text{Base} \) is independently:
or an analog thereof selected from the group consisting of:

(i), (ii), (iii), (iv), (v), (vi), (vii), (viii), (ix), and (x),
wherein:
each instance of \( R^1 \), \( R^2 \), \( R^4 \), and \( R^5 \) is independently hydrogen, substituted or unsubstituted \( C_{1-20} \)alkyl, substituted or unsubstituted \( C_{2-20} \)alkenyl, substituted or unsubstituted \( C_{2-20} \)alkynyl, substituted or unsubstituted heterocyclyl, substituted or unsubstituted carbocycyl, substituted or unsubstituted heteroarylene, a nitrogen protecting group, -OR \(^B\), or -SR \(^B\), wherein each instance of \( R^B \) is independently hydrogen, substituted or unsubstituted \( C^0 \)alkyl, substituted or unsubstituted \( C_{2-20} \)alkenyl, substituted or unsubstituted alkynyl, substituted or unsubstituted heteroaryl, an oxygen protecting group when attached to an oxygen group, or a sulfur protecting group when attached to a sulfur group; or \( R^1 \) and \( R^2 \) and/or \( R^4 \) and \( R^5 \) are joined to form a substituted or unsubstituted heterocyclic or substituted or unsubstituted heteroarylene ring;
each instance of \( R^3 \) is independently substituted or unsubstituted \( C^0 \)alkyl, substituted or unsubstituted \( C_{2-20} \)alkenyl, substituted or unsubstituted \( C_{2-20} \)alkynyl, substituted or unsubstituted carbocycyl, substituted or unsubstituted heterocyclylene, substituted or unsubstituted heteroarylene, halogen, -OR \(^C\), -SR \(^C\), -N(\( R^C \))\(^2\), -SHg, -S0 \(^2\)SHg , -SeR \(^D\), or -TeR \(^D\) wherein each instance of \( R^C \) is hydrogen, substituted or unsubstituted \( C_{1-20} \)alkyl, substituted or unsubstituted \( C_{2-20} \)alkenyl, substituted or unsubstituted \( C_{2-20} \)alkynyl, substituted or unsubstituted carbocycyl, substituted or unsubstituted heterocyclylene, substituted or unsubstituted heteroarylene, an oxygen protecting group when attached to an oxygen atom, a sulfur protecting group when attached to a sulfur atom, a nitrogen protecting group when attached to a nitrogen atom; or two \( R^C \) groups are joined to form a substituted or unsubstituted heterocyclic or substituted or unsubstituted heteroarylene ring;
each instance of \( L^1 \) is independently absent or a linking moiety selected from the group consisting of substituted or unsubstituted \( C^0 \)alkylene, substituted or unsubstituted \( C_{2-20} \)alkynylene, substituted or unsubstituted \( C_{2-20} \)alkynylene, substituted or unsubstituted hetero\( C_{1-20} \)alkylene, substituted or unsubstituted hetero\( C_{2-20} \)alkenylene, substituted or unsubstituted hetero\( C_{2-20} \)alkenylene, substituted or unsubstituted hetero\( C_{2-20} \)alkenyne, substituted or unsubstituted carbocycylene, substituted or unsubstituted heterocyclylene, substituted or unsubstituted arylene, or substituted or unsubstituted heteroarylene, or a combination thereof;
each instance of \( R^D \) is independently hydrogen, substituted or unsubstituted \( C^0 \)alkyl, substituted or unsubstituted \( C_{2-20} \)alkenyl, substituted or unsubstituted \( C_{2-20} \)alkynyl,
substituted or unsubstituted carbocycyl, substituted or unsubstituted heterocyclyl, substituted or unsubstituted aryl, or substituted or unsubstituted heteroaryl; and

each instance of $M^1$ and $M^4$ are independently O, Se, Te, CH$_2$, CF$_2$, CC$_2$, CBr$_2$, or Cl$_2$;

provided that the compound comprises at least one instance of a heavy atom selected from the group consisting of bromine, iodine, selenium, tellurium, or mercury; and

further provided that the following compounds are specifically excluded:

![Chemical structures](image)

and salts thereof.

2. The heavy-atom labeled nucleotide of claim 1 having the following stereochemistry:

![Stereochemistry](image)

or the enantiomer thereof and/or salt thereof.

3. The heavy-atom labeled nucleotide of claim 1, wherein the nucleotide is a quaternary amine, an amino acid, or a metal salt.

4. The heavy-atom labeled nucleotide of claim 1, comprising at least one instance of -SeR$_D$ or -TeR$_D$, wherein R$_D$ is substituted or unsubstituted C^haloalkyl.

5. The heavy-atom labeled nucleotide of claim 1, comprising at least one instance of -SHg or -S0$_2$SHg.
6. The heavy-atom labeled nucleotide of claim 1, comprising at least one instance of -Br or -I.

7. The heavy-atom labeled nucleotide of claim 1, wherein the base comprises a heavy atom.

8. The heavy-atom labeled nucleotide of claim 1, wherein the base does not comprise a heavy atom.

9. The heavy-atom labeled nucleotide of claim 1, wherein G₂ is hydrogen.

10. The heavy-atom labeled nucleotide of claim 1, wherein G₂ is -SHg or -SO₂SHg.

11. The heavy-atom labeled nucleotide of claim 1, wherein G₂ is -SeR² or -TeR², wherein R² is substituted or unsubstituted C₁₋₂haloalkyl.

12. The heavy-atom labeled nucleotide of claim 1, wherein G₃ is hydrogen or a triphosphate group.

13. The heavy-atom labeled nucleotide of claim 1, wherein the base is selected from the group consisting of:

wherein the base comprises a heavy atom.
14. The heavy-atom labeled nucleotide of claim 13, wherein $L^1$ is a linking moiety selected from the group consisting of substituted and unsubstituted alkyylene; substituted and unsubstituted alkenylene; substituted and unsubstituted alkynylene.

15. The heavy-atom labeled nucleotide of claim 13, wherein $L^1$ represents a linker consisting of a combination of one or more consecutive covalently bonded groups of the formula:

```
R^1_1 - R^3_1 - R^2_2
```

wherein:
- each instance of $m$ is independently an integer between 1 to 10, inclusive;
- each instance of $p$ is independently an integer between 1 to 4, inclusive;
- each instance of $R^1$ is independently hydrogen; substituted or unsubstituted alkyl; substituted or unsubstituted alkenyl; substituted or unsubstituted alkynyl; substituted or unsubstituted carbocyclyl; substituted or unsubstituted heterocyclyl; substituted or unsubstituted aryl; substituted or unsubstituted heteroaryl; or a nitrogen protecting group;
- each instance of $R^2$ is independently hydrogen; halogen; substituted or unsubstituted alkyl; substituted or unsubstituted alkenyl; substituted or unsubstituted alkynyl; substituted or unsubstituted carbocyclyl; substituted or unsubstituted heterocyclyl; substituted or unsubstituted aryl; substituted or unsubstituted heteroaryl; or two $R^2$ groups are joined to form a substituted or unsubstituted 5- to 6-membered ring.

16. The heavy-atom labeled nucleotide of claim 15, wherein $L^1$-$R^3$ is a group of formula:

```
R^1_1 - R^3_1 - R^2_2
```

17. The heavy-atom labeled nucleotide of claim 13, wherein $M^3$ is O.
18. The heavy-atom labeled nucleotide of claim 13, wherein $M_4$ is $O$.

19. The heavy-atom labeled nucleotide of claim 13, wherein $R_1$ and $R_2$ are hydrogen.

20. The heavy-atom labeled nucleotide of claim 13, wherein $R_4$ and $R_5$ are hydrogen.

21. The heavy-atom labeled nucleotide of claim 13, wherein $R_3$ is $-\text{Br}$, $-\text{I}$, $-\text{OR}^C$, $-\text{SR}^C$, $-\text{N}(\text{R}^C)_2$, $-\text{SHg}$, $-\text{S}0_2\text{SHg}$, $-\text{SeR}^D$, or $-\text{TeR}^D$.

22. The heavy-atom labeled nucleotide of claim 1 selected from the group consisting of:

![Chemical Structures]

and salts thereof.
23. A heavy-atom labeled nucleic acid polymer of Formula (II):

\[
\begin{array}{c}
\text{Base} \\
\text{G}_1 \quad \text{G}_2 \\
\text{O} \quad \text{M}_1^{-} \\
\text{M}_2^{-} \\
\text{G}_3 \\
\end{array}
\]

or a salt thereof;

wherein:

- each instance of $\text{G}_i$ is independently $-0^-, -S^-, -Se^-, -CH_2^-, \text{ or } -NH^-;
- each instance of $\text{G}_2$ is independently hydrogen, halogen, $-\text{OR}^A$, $-\text{SR}^A$, $-\text{N(R^A)}_2$, $-\text{SHg}$, $-\text{S}_2\text{SHg}$, $-\text{SeR}^D \text{ or } -\text{TeR}^D$;
- each instance of $\text{R}^A$ is independently hydrogen, substituted or unsubstituted $\text{C}_{2-20}\text{alkyl}$, substituted or unsubstituted $\text{C}_{2-20}\text{alkenyl}$, substituted or unsubstituted $\text{C}_{2-20}\text{alkynyl}$, substituted or unsubstituted carbocycyl, substituted or unsubstituted heterocyclyl, substituted or unsubstituted aryl, or substituted or unsubstituted heteroaryl, an oxygen protecting group when attached to an oxygen atom, a sulfur protecting group when attached to a sulfur atom, a nitrogen protecting group when attached to a nitrogen atom; or two $\text{R}^A$ groups are joined to form a substituted or unsubstituted heterocyclic or substituted or unsubstituted heteroaryl ring;
- each instance of $\text{M}_1$ is independently $-0^-, -S^-, -NH^-, -Se^-, \text{ or } -\text{C(R}^M)_2^-$, wherein each instance of $\text{R}^M$ is independently hydrogen or halogen;
- each instance of $\text{G}_3$ is independently hydrogen, substituted or unsubstituted $\text{C}^\text{oalkyl}$, substituted or unsubstituted $\text{C}_{2-20}\text{alkenyl}$, substituted or unsubstituted $\text{C}_{2-20}\text{alkynyl}$, substituted or unsubstituted carbocycyl, substituted or unsubstituted heterocyclyl, substituted or unsubstituted aryl, or substituted or unsubstituted heteroaryl, or a monophosphate, diphosphate, or triphosphate of formula:

\[
\begin{array}{c}
\text{O}^0 \quad \text{O}^0 \quad \text{O}^0 \\
\text{OH} \quad \text{M}^{-\text{H}}, \text{ or } \text{OH} \quad \text{OH} \quad \text{M}^{-\text{H}}
\end{array}
\]

wherein each instance of $\text{M}^2$ is independently $-0^-, -S^-, \text{ or } -\text{Se}^-$; and
- each instance of $\text{Base}$ is independently:
or an analog thereof selected from the group consisting of:

(i), (ii), (iii), (iv), (v), (vi), (vii), (viii), (ix), (x),
wherein:

each instance of $R_1$, $R_2$, $R_4$, and $R_5$ is independently hydrogen, substituted or unsubstituted $C_{1-20}$ alkyl, substituted or unsubstituted $C_{2-2}$ alkenyl, substituted or unsubstituted $C_{2-20}$ alkynyl, substituted or unsubstituted heterocyclyl, substituted or unsubstituted aryl, or substituted or unsubstituted heteroaryl, a nitrogen protecting group, -$OR^B$, or -$SR^B$, wherein each instance of $R^B$ is independently hydrogen, substituted or unsubstituted $C^\circ$ alkyl, substituted or unsubstituted carbocyclyl, substituted or unsubstituted heterocyclyl, substituted or unsubstituted alkynyl, substituted or unsubstituted alkenyl, substituted or unsubstituted carbocyclyl, substituted or unsubstituted heterocyclyl, substituted or unsubstituted aryl, or substituted or unsubstituted heteroaryl, an oxygen protecting group when attached to an oxygen group, or a sulfur protecting group when attached to a sulfur group; or $R_1$ and $R_2$ and/or $R_4$ and $R_5$ are joined to form a substituted or unsubstituted heterocyclic or substituted or unsubstituted heteroaryl ring;

each instance of $R^3$ is independently substituted or unsubstituted $C^\circ$ alkyl, substituted or unsubstituted $C_{2-2}$ alkenyl, substituted or unsubstituted $C_{2-20}$ alkynyl, substituted or unsubstituted carbocyclyl, substituted or unsubstituted heterocyclyl, substituted or unsubstituted aryl, substituted or unsubstituted heteroaryl, halogen, -$OR^C$, -$SR^C$, -$N(R^D)_2$, -$SHg$, -$SO\_2SHg$, -$SeR^D$, or -$TeR^D$ wherein each instance of $R^C$ is hydrogen, substituted or unsubstituted $C_{1-20}$ alkyl, substituted or unsubstituted $C_{2-2}$ alkenyl, substituted or unsubstituted $C_{2-20}$ alkynyl, substituted or unsubstituted carbocyclyl, substituted or unsubstituted heterocyclyl, substituted or unsubstituted aryl, or substituted or unsubstituted heteroaryl, an oxygen protecting group when attached to an oxygen atom, a sulfur protecting group when attached to a sulfur atom, a nitrogen protecting group when attached to a nitrogen atom; or two $R^C$ groups are joined to form a substituted or unsubstituted heterocyclic or substituted or unsubstituted heteroaryl ring;

each instance of $L^1$ is independently absent or a linking moiety selected from the group consisting of substituted or unsubstituted $C^\circ$ alkylene, substituted or unsubstituted $C_{2-2}$ alkenylene, substituted or unsubstituted $C_{2-20}$ alkynylene, substituted or unsubstituted hetero$C_{1-2}$ alkylene, substituted or unsubstituted hetero$C_{2-2}$ alkenylene, substituted or unsubstituted hetero$C_{2-20}$ alkynylene, substituted or unsubstituted carbocycylene, substituted or unsubstituted heterocyclylene, substituted or unsubstituted arylene, or substituted or unsubstituted heteroarylene, or a combination thereof;

each instance of $R^D$ is independently hydrogen, substituted or unsubstituted $C^\circ$ alkyl, substituted or unsubstituted $C_{2-2}$ alkenyl, substituted or unsubstituted $C_{2-20}$ alkynyl,
substituted or unsubstituted carbocycyl, substituted or unsubstituted heterocyclyl, substituted or unsubstituted aryl, or substituted or unsubstituted heteroaryl;

each instance of M³ and M⁴ are independently O, Se, Te, CH₂, CF₂, CCl₂, CBr₂, or Cl₂; and

and n is 1 to 25,000, inclusive;

provided that the compound comprises at least one instance of a heavy atom selected from the group consisting of bromine, iodine, selenium, tellurium, or mercury; and

further provided that the compounds of Formula (II) comprising one or more instances of the formula:

24. The heavy-atom labeled nucleic acid polymer of claim 23 comprising one or more instances of formula:
25. A method of determining the sequence of a nucleic acid polymer comprising forming a complementary strand of the nucleic acid polymer from one or more heavy-atom labeled compounds of Formula (I):

or a salt thereof;
wherein:

each instance of $G_1$ is independently $-\text{O}^-, -\text{S}^-, -\text{Se}^-, -\text{CH}_2^-$, or $-\text{NH}^-$;

each instance of $G_2$ is independently hydrogen, halogen, $-\text{OR}^A$, $-\text{SR}^A$, $-\text{N}(\text{R}^A)_2$, $-\text{SHg}$, $-\text{S0}_2\text{SHg}$, $-\text{SeR}^D$, or $-\text{TeR}^D$;

each instance of $\text{R}^A$ is independently hydrogen, substituted or unsubstituted $\text{C}^\text{oalkyl}$, substituted or unsubstituted $\text{C}_2\text{2oalkenyl}$, substituted or unsubstituted $\text{C}_2\text{2oalkynyl}$, substituted or unsubstituted carbocycyl, substituted or unsubstituted heterocyclyl, substituted or unsubstituted aryl, or substituted or unsubstituted heteroaryl, an oxygen protecting group when attached to an oxygen atom, a sulfur protecting group when attached to a sulfur atom, a nitrogen protecting group when attached to a nitrogen atom; or two $\text{R}^A$ groups are joined to form a substituted or unsubstituted heterocyclic or substituted or unsubstituted heteroaryl ring;

each instance of $M^1$ is independently $-\text{O}^-$, $-\text{S}^-$, $-\text{NH}^-$, $-\text{Se}^-$, or $-\text{C}(\text{R}^M)^2^-$, wherein each instance of $\text{R}^M$ is independently hydrogen or halogen;

each instance of $G_3$ is independently hydrogen, substituted or unsubstituted $\text{C}_{1-20}\text{alkyl}$, substituted or unsubstituted $\text{C}_{2-20}\text{oalkenyl}$, substituted or unsubstituted $\text{C}_{2-20}\text{alkynyl}$, substituted or unsubstituted carbocycyl, substituted or unsubstituted heterocyclyl, substituted or unsubstituted aryl, or substituted or unsubstituted heteroaryl, or a monophosphate, diphosphate, or triphosphate of formula:

\[
\begin{align*}
\text{HO-P—} & \quad \text{HO-P—} & \quad \text{O-P—} \\
\text{M—}^\text{OH} & \quad \text{M—}^\text{OH} & \quad \text{M—}^\text{OH}
\end{align*}
\]

wherein each instance of $M^2$ is independently $-\text{O}^-$, $-\text{S}^-$, or $-\text{Se}^-$; and

each instance of Base is independently:

- [Adenine](#)
- [Guanine](#)
- [Cytosine](#)
- [Uracil](#)
- [Thymine](#)

or an analog thereof selected from the group consisting of:
wherein:
each instance of $R_1$, $R_2$, $R_4$, and $R_5$ is independently hydrogen, substituted or unsubstituted $C^1$alkyl, substituted or unsubstituted $C_2$-alkenyl, substituted or unsubstituted $C_2$-alkynyl, substituted or unsubstituted carbocyclyl, substituted or unsubstituted heterocyclyl, substituted or unsubstituted $C_2$-alkyl, substituted or unsubstituted heteroaryl, a nitrogen protecting group, $-OR^B$, or $-SR^B$, wherein each instance of $R^B$ is independently hydrogen, substituted or unsubstituted $C_1$-alkyl, substituted or unsubstituted...
C₆₂₀alkenyl, substituted or unsubstituted C₂₋₂₀ alkynyl, substituted or unsubstituted carbocycyl, substituted or unsubstituted heterocyclyl, substituted or unsubstituted aryl, or substituted or unsubstituted heteroaryl, an oxygen protecting group when attached to an oxygen group, or a sulfur protecting group when attached to a sulfur group; or R¹ and R² and/or R⁴ and R⁵ are joined to form a substituted or unsubstituted heterocyclic or substituted or unsubstituted heteroaryl ring;

   each instance of R³ is independently substituted or unsubstituted C₁₋₋₂₀alkyl, substituted or unsubstituted C₂₋₂₀alkenyl, substituted or unsubstituted C₂₋₂₀ alkynyl, substituted or unsubstituted carbocycyl, substituted or unsubstituted heterocyclyl, substituted or unsubstituted aryl, substituted or unsubstituted heteroaryl, halogen, -OR⁻, -SR⁻, -N(R⁻)₂, -SHg, -SO₂SHg, -SHgR⁻, -SeR⁻, or -TeR⁻ wherein each instance of R⁻ is hydrogen, substituted or unsubstituted C₆ₒalkyl, substituted or unsubstituted C₂₋₂₀alkenyl, substituted or unsubstituted C₂₋₂₀ alkynyl, substituted or unsubstituted carbocycyl, substituted or unsubstituted heterocyclyl, substituted or unsubstituted aryl, or substituted or unsubstituted heteroaryl, an oxygen protecting group when attached to an oxygen atom, a sulfur protecting group when attached to a sulfur atom, a nitrogen protecting group when attached to a nitrogen atom; or two R⁻ groups are joined to form a substituted or unsubstituted heterocyclic or substituted or unsubstituted heteroaryl ring;

   each instance of L¹ is independently absent or a linking moiety selected from the group consisting of substituted or unsubstituted C₆ₒalkylene, substituted or unsubstituted C₂₋₂ₒalkenylene, substituted or unsubstituted C₂₋₂₀ alkynylene, substituted or unsubstituted heteroC₁₋₋₂₀alkylene, substituted or unsubstituted heteroC₂₋₂ₒalkenylene, substituted or unsubstituted heteroC₂₋₂₀ alkynylene, substituted or unsubstituted heterocyclylene, substituted or unsubstituted arylene, or substituted or unsubstituted heteroarylene, or a combination thereof;

   each instance of R⁶ is independently hydrogen, substituted or unsubstituted C₆ₒalkyl, substituted or unsubstituted C₂₋₂ₒalkenyl, substituted or unsubstituted C₂₋₂₀ alkynyl, substituted or unsubstituted carbocycyl, substituted or unsubstituted heterocyclyl, substituted or unsubstituted aryl, or substituted or unsubstituted heteroaryl; and

   each instance of M³ and M⁴ are independently O, Se, Te, CH₂, CF₂, CCl₂, CBr₂, or Cl₂;

   provided that the compound comprises at least one instance of a heavy atom selected from the group consisting of bromine, iodine, selenium, tellurium, or mercury; and
identifying a sequence of nucleotides in the nucleic acid polymer and/or in the complementary strand using a particle beam.

26. The method of claim 25, wherein the nucleic acid polymer is DNA or RNA.

27. The method of claim 25, wherein the complementary strand is DNA or RNA.

28. The method of claim 25, wherein the nucleic acid polymer and/or its complementary strand is formed by a nucleic acid polymerase enzyme.

29. The method of claim 25, wherein the complementary strand of the nucleic acid polymer is formed using polymerase chain reaction (PCR).

30. The method of claim 25, wherein the nucleotides of the nucleic acid polymer and/or the complementary strand are modified to include labels comprising one or more heavy-atom labeled compounds of Formula (I).

31. The method of claim 30, wherein at least two types of nucleotides are labeled with the same type of heavy-atom label.

32. The method of claim 30, wherein one type of nucleotide is labeled.

33. The method of claim 30, wherein two types of nucleotides are labeled.

34. The method of claim 30, wherein three types of nucleotides are labeled.

35. The method of claim 30, wherein all the nucleotides are labeled.

36. The method of claim 25, wherein nucleotide specific labels are incorporated in the nucleic acid polymer and/or the complementary strand during formation of the nucleic acid polymer and/or the complementary strand.

37. The method of claim 25, wherein the step of identifying a sequence of nucleotides comprises generating a particle beam, exposing the nucleic acid polymer and/or the
complementary strand to the particle beam, and identifying the nucleotides due to characteristic changes to the particle beam.

38. The method of claim 37, wherein the step of identifying the nucleotides comprises detecting characteristic changes to the particle beam.

39. The method of claim 25, 37 or 38, wherein the particle beam is a lepton beam.

40. The method of claim 39, wherein the lepton beam is an electron beam.
Fig. 5
**INTERNATIONAL SEARCH REPORT**

**International application No.**
PCT/US2013/070299

**A. CLASSIFICATION OF SUBJECT MATTER**

<table>
<thead>
<tr>
<th>Category</th>
<th>Citation of document, with indication, where appropriate, of the relevant passages</th>
<th>Relevant to claim No.</th>
</tr>
</thead>
</table>

**B. FIELDS SEARCHED**

- Minimum documentation searched (classification system followed by classification symbols)
  - C07H, C12Q

**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

- **X** Further documents are listed in the continuation of Box C.
- **X** See patent family annex.

* Special categories of cited documents:
  - **X** later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
  - **Y** document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
  - **X** document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
  - **A** document member of the same patent family

**Date of the actual completion of the international search**
18 December 2013

**Date of mailing of the international search report**
03/01/2014

**Name and mailing address of the ISA**
European Patent Office, P.B. 5818 Patentlaan 2
NL-2280 HN Rijswijk
Tel. (+31-70) 340-2040
Fax: (+31-70) 340-3016

**Authorized officer**
Ni kol ai, Joachim
INTERNATIONAL SEARCH REPORT

DOCUMENTS CONSIDERED TO BE RELEVANT

<table>
<thead>
<tr>
<th>Category</th>
<th>Citation of document, with indication, where appropriate, of the relevant passages</th>
<th>Relevant to claim No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>X</td>
<td>MOHANNAD ABD0 ET AL: &quot;Electrophilic Aromatic Selective Dehydro-3'-Deoxythymidine derivatives&quot;, ORGANIC LETTERS, vol. 12, no. 13, 2 July 2010 (2010-07-02), pages 2982-2985, XP55093658, ISSN: 1523-7060, DOI: 10.1021/ol100321</td>
<td>1, 2, 7, 12</td>
</tr>
</tbody>
</table>

Form PCT/ISA/210 (continuation of second sheet) (April 2008)
<table>
<thead>
<tr>
<th>Category</th>
<th>Citation of document, with indication, where appropriate, of the relevant passages</th>
<th>Relevant to claim No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>X</td>
<td>wo 2010/135564 A2 (SENA RES INC [US] ; GEORGIA STATE UNIVERSITY OFFICE OF LEGAL AFFAIRS [U]) 25 November 2010 (2010-11-25) scheme 2</td>
<td>1, 2, 7, 9, 12, 23</td>
</tr>
<tr>
<td>X</td>
<td>wo 02/077002 A2 (UNIV CITY [US] ; HUANG ZHEN [US]) 3 October 2002 (2002-10-03) claim 1 figures 1, 6, 7, 8</td>
<td>1, 2, 7-9, 12, 23</td>
</tr>
</tbody>
</table>

Form PCT/ISA/210 (continuation of second sheet) (April 2008)
<table>
<thead>
<tr>
<th>Category</th>
<th>Citation of document, with indication, where appropriate, of the relevant passages</th>
<th>Relevant to claim No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patent document cited in search report</td>
<td>Publication date</td>
<td>Patent family member(s)</td>
</tr>
<tr>
<td>---------------------------------------</td>
<td>----------------</td>
<td>-------------------------</td>
</tr>
<tr>
<td></td>
<td></td>
<td>WO 2010135564 A2</td>
</tr>
<tr>
<td>WO 2009117668 A1</td>
<td>24-09-2009</td>
<td>CN 101977611 A</td>
</tr>
<tr>
<td></td>
<td></td>
<td>US 20111105743 A1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>WO 2009117668 A1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>US 2013211067 A1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>WO 2012050264 A1</td>
</tr>
<tr>
<td>WO 02077002 A2</td>
<td>03-10-2002</td>
<td>AU 2002305073 A1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>US 2003055016 A1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>US 2010137573 A1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>US 2012065387 A1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>WO 02077002 A2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>AU 2874902 A</td>
</tr>
<tr>
<td></td>
<td></td>
<td>AU 2002228749 B2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>AU 2008203296 A1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>AU 2012203287 A1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>BR 0114837 A</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CA 2426187 A</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CN 1646141 A</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CN 101862345 A</td>
</tr>
<tr>
<td></td>
<td></td>
<td>DK 1411954 T3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>DK 2251015 T3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>EP 1411954 A2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>EP 2251015 A1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ES 2402597 T3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>JP 2004533406 A</td>
</tr>
<tr>
<td></td>
<td></td>
<td>JP 2009161541 A</td>
</tr>
<tr>
<td></td>
<td></td>
<td>JP 2011012062 A9</td>
</tr>
<tr>
<td></td>
<td></td>
<td>JP 2013079250 A</td>
</tr>
<tr>
<td></td>
<td></td>
<td>KR 20040028657 A</td>
</tr>
<tr>
<td></td>
<td></td>
<td>KR 20080041296 A</td>
</tr>
<tr>
<td></td>
<td></td>
<td>KR 2009007650 A</td>
</tr>
<tr>
<td></td>
<td></td>
<td>KR 20090089922 A</td>
</tr>
<tr>
<td></td>
<td></td>
<td>KR 20100080866 A</td>
</tr>
<tr>
<td></td>
<td></td>
<td>KR 20110039585 A</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PT 1411954 E</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PT 2251015 E</td>
</tr>
<tr>
<td></td>
<td></td>
<td>US 2003087873 A1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>US 2011269707 A1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>WO 0232920 A2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>WO 2007089542 A2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CA 2523919 A1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CN 1813070 A</td>
</tr>
<tr>
<td></td>
<td></td>
<td>EP 1620574 A2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>US 2004038261 A1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>WO 2004097032 A2</td>
</tr>
</tbody>
</table>