

Accordingly, the drawings and description are to be regarded as illustrative in nature, and not as restrictive.

INCORPORATION BY REFERENCE

[0047] All publications, patents, and patent applications mentioned in this specification are herein incorporated by reference to the same extent as if each individual publication, patent, or patent application was specifically and individually indicated to be incorporated by reference. To the extent publications and patents or patent applications incorporated by reference contradict the disclosure contained in the specification, the specification is intended to supersede and/or take precedence over any such contradictory material.

BRIEF DESCRIPTION OF THE DRAWINGS

[0048] The novel features of the invention are set forth with particularity in the appended claims. A better understanding of the features and advantages of the present invention will be obtained by reference to the following detailed description that sets forth illustrative embodiments, in which the principles of the invention are utilized, and the accompanying drawings (also “Figure” and “FIG.” herein), of which:

[0049] **FIG. 1** schematically illustrates an example workflow for peptide sequencing disclosed herein.

[0050] **FIG. 2** schematically illustrates another example workflow for peptide sequencing disclosed herein.

[0051] **FIG. 3** schematically illustrates another example workflow for peptide sequencing disclosed herein.

[0052] **FIG. 4** schematically illustrates an example composition comprising a linker and a polymerizable molecule.

[0053] **FIG. 5** shows a computer system that is programmed or otherwise configured to implement methods provided herein.

[0054] **FIG. 6A** shows a schematic of an example modified amino acid or derivative thereof comprising a polymerizable molecule and amino acid-linker complexes for three different amino acid types (Asp, Trp, and Tyr). Figure discloses SEQ ID NO: 2. **FIG. 6B** shows example current traces obtained from a nanopore sequencing system of two control molecules and three amino acid-linker-polymerizable molecule complexes. Figure discloses SEQ ID NOS 3, 3, 3, 3, and 3, respectively, in order of appearance. **FIG. 6C** shows example current traces as a function of time. Figure discloses SEQ ID NO: 2. **FIG. 6D** shows example current traces as a function of

base pair position. Figure discloses SEQ ID NO: 4. **FIG. 6E** shows example analyzed data of classification of the different molecule types.

[0055] FIG. 7A shows a schematic of another example modified amino acid or derivative thereof comprising a polymerizable molecule. Figure discloses SEQ ID NOS 5 and 6, respectively, in order of appearance. FIG. 7B shows example current traces as a function of time. FIG. 7C shows example current traces as a function of base pair position.

[0056] FIG. 8A shows a schematic of another example modified amino acid or derivative thereof comprising a polymerizable molecule. Figure discloses SEQ ID NOS 5 and 6, respectively, in order of appearance. FIG. 8B shows example current traces as a function of time and base pair position. FIG. 8C shows example current traces as a function of time comparing two polymerizable molecules comprising different linkers. FIG. 8D shows a t-SNE plot of a dynamic-time warped correlation matrix.

[0057] FIG. 9 shows example data on the dwell time of two control molecules and three amino acid-linker-polymerizable molecule complexes.

[0058] FIG. 10A shows a schematic of a stacked amino acid-linker-polymerizable molecule complex described herein. FIG. 10B shows a schematic of an example model stacked amino acid-linker-polymerizable molecule complex. Figure discloses SEQ ID NOS 7 and 7, respectively, in order of appearance. FIG. 10C shows a table of example amino acid types comprised by the model stacked amino acid-linker-polymerizable molecule complexes. FIG. 10D shows example current traces as a function of time and base pair position. FIG. 10E shows example classification data of the different molecule types.

[0059] FIG. 11A shows example current traces obtained from a polymerizable molecule translocating through a nanopore sequencing system at different temperatures. Figure discloses SEQ ID NOS 4, 3 and 3, respectively, in order of appearance. FIG. 11B shows histograms of the translocation duration of the molecules through the nanopore at the different temperatures. FIG. 11C shows histograms of the translocation durations of different sequences of the polymerizable molecule at the different temperatures. FIG. 11D shows example current traces of a portion of the polymerizable molecule. FIG. 11E shows example dwell times of a portion of the polymerizable molecule under the two different temperature conditions.

DETAILED DESCRIPTION

[0060] While various embodiments of the invention have been shown and described herein, it will be obvious to those skilled in the art that such embodiments are provided by way of example only. Numerous variations, changes, and substitutions may occur to those skilled in the art without departing from the invention. It should be understood that various alternatives to the embodiments of the invention described herein may be employed.

Definitions

[0061] Whenever the term “at least,” “greater than,” or “greater than or equal to” precedes the first numerical value in a series of two or more numerical values, the term “at least,” “greater

conjugate) is then purified from the reaction using reverse-phase High Pressure Liquid Chromatography (HPLC).

Example 2- Nanopore detection of amino acids conjugated to polymerizable molecules

[00247] To determine whether nanopore detection of a modified amino acid or derivative thereof, e.g., an AALC complex or stacked AALC complex, or portion thereof, is possible, an amino acid-linker-polymerizable molecule complex is generated and translocated through a nanopore sequencing system. The amino acid-linker-polymerizable molecule complex is generated by coupling a polymerizable molecule to an amino acid-linker complex. The polymerizable molecule comprises an alkyne-linked DNA molecule. More specifically, the polymerizable molecule comprises an alkyne linker to a T236 position within a 16 nt poly-dT (SEQ ID NO: 1) DNA oligo scaffold located in the middle of one strand of a 450 nt dsDNA backbone, as schematically illustrated in **FIG. 6A**. The amino acid-linker complex is generated using a bifunctional linker, 1-(2-azidoethyl)-4-isothiocyanatobenzene, which comprises (1) a PITC moiety (amino acid reactive group) and (2) an azide moiety that is capable of reacting with the alkyne of the polymerizable molecule. The bifunctional linker is reacted with three different amino acids: Asp, Trp, and Tyr via the PITC moiety, thereby generating an amino acid-linker complex, also shown in the bottom panel of **FIG. 6A**. The amino acid-linker complexes are then reacted with the alkyne-linked DNA molecule, thereby generating an amino acid-linker-polymerizable molecule conjugate.

[00248] Next, the amino acid-linker-polymerizable molecule conjugates are run through a nanopore sequencing system (Oxford Nanopore Technologies MinION) to interrogate whether the generated current signal can be assigned back to the amino acid type. The DNA backbone without modifications (“Unmodified”) is used as a baseline control, and the alkyne-linked DNA molecule (“Alkyne”) is used as a control for the alkyne linker-specific current blockage. **FIG. 6B** shows the current traces and assigned nucleotide sequence of the DNA backbone. Of note, all of the different molecules do pass through the nanopore, and a discernible qualitative difference or perturbation in the current signal can be visualized from the amino acid-linker-polymerizable molecule conjugates (labeled as “Linker-Asp”, “Linker-Trp”, or “Linker-Tyr” that is not present in the current signal of the control molecules. **FIG. 6C** shows the current traces (measured current signal) as a function of time for the five conditions: Unmodified, Alkyne, Linker-Asp, Linker-Trp, and Linker-Tyr. All the combined overlaid traces for all of the reads from the nanopores in the device are shown on the left and the right shows three representative individual traces. **FIG. 6D** shows the current traces (measured current signal) as a

processing of amino acids to generate modified amino acids comprising polymerizable molecules may be a viable approach for conducting highly accurate peptide sequencing using a nanopore or nanochannel approach.

Example 7- Temperature modulation for improving current signal-to-noise ratio

[00258] As described herein, a translocation velocity of a modified amino acid or derivative thereof may differ from that of an unmodified amino acid, and similarly, a translocation velocity of a modified amino acid or derivative thereof comprising a polymerizable molecule may differ from that of the polymerizable molecule without an amino acid coupled thereto. In one hypothesis, the translocation velocity of a molecule through a nanopore may be lowered by decreasing the temperature of the environment. To test this hypothesis, a polymerizable molecule without modifications (does not comprise an amino acid-linker complex) can be run through a nanopore sequencing system at ambient (room) temperature or on ice.

[00259] The polymerizable molecule includes a 450 nt dsDNA backbone that comprises a 16nt poly-dT sequence (SEQ ID NO: 1), as schematically illustrated in **FIG. 11A** (selected sequence from about ~225 bp to ~245 bp shown). The polymerizable molecule is run through a nanopore sequencing system (Oxford Nanopore Technologies MinION) to interrogate the effect of the temperature on the sequencing output. The measured temperature of the flow cell run at room temperature is 34 degrees Celsius, and the measured temperature of the flow cell run on ice is ~19-20 degrees Celsius. **FIG. 11A** shows the current traces (measured current signal) as a function of the base pair position and identified nucleotide of a selected sequence from the polymerizable molecule. Interestingly, the signal-to-noise ratio of the 16 nt poly-dT sequence (SEQ ID NO: 1) qualitatively appears higher.

[00260] **FIG. 11B** shows histograms of the translocation duration, which is a measure of the duration of time that the region or whole polymerizable molecule resides in the nanopore and corresponds to the translocation velocity of the region or whole polymerizable molecule. The translocation duration is measured in milliseconds (ms) of the polymerizable molecule for the two different temperature conditions (room temperature or ice). The left-hand plot shows the translocation durations of the entire polymerizable molecule and the right-hand plot shows the translocation durations of the selected site surrounding the 16nt poly-dT sequence (SEQ ID NO: 1) (200-275 nt position). As can be visualized from the histograms, the ice condition results in reduced translocation velocity (increased translocation duration) with greater variability or higher range of the measured duration values.

[00261] FIG. 11C shows histograms of the translocation duration in ms of the polymerizable molecule for the two different temperature conditions (room temperature or ice). The left-hand plot shows the translocation durations of the polymerizable molecule at the 200-225 nt position upstream of the 16nt poly-dT sequence (SEQ ID NO: 1). The middle plot shows the translocation durations for the selected site immediately adjacent to and encompassing the 16nt poly-dT sequence (SEQ ID NO: 1) (225-250 nt in position). The right-hand plot shows the translocation durations of the polymerizable molecule at the 250-275 nt position downstream of the 16nt poly-dT sequence (SEQ ID NO: 1). As can be visualized in all three plots, the ice condition results in reduced translocation velocity (increased translocation duration) with greater variability or higher range of the measured duration values for both the poly-dT region and adjacent regions.

[00262] FIG. 11D shows the current traces (measured current signal) as a function of base position from the 200-280 nt position. The left-hand plot shows the current trace for the room temperature condition and the right-hand plots shows the current trace for the ice condition. The ice condition improves the current signal-to-noise ratio, as evidenced by the decreased magnitude of the error bars in the poly-dT region (~225-245 nt region).

[00263] FIG. 11E shows the translocation duration in ms as a function of base position from the 200-280 nt position. The left-hand plot shows the current trace for the room temperature condition and the right-hand plots shows the current trace for the ice condition. Interestingly, while the ice condition results in an increased dwell time per base position, it also results in increased noise of the translocation duration except for the poly-dT region (~225-245 bp region).

[00264] Altogether, the results from FIGs. 11A-11E suggest that reducing the temperature may result in improved signal-to-noise ratio of current traces read out from polymerizable molecules from a nanopore sequencing system, which may be applicable for improved accuracy readout of modified amino acids or derivatives thereof.

[00265] While preferred embodiments of the present invention have been shown and described herein, it will be obvious to those skilled in the art that such embodiments are provided by way of example only. It is not intended that the invention be limited by the specific examples provided within the specification. While the invention has been described with reference to the aforementioned specification, the descriptions and illustrations of the embodiments herein are not meant to be construed in a limiting sense. Numerous variations, changes, and substitutions will now occur to those skilled in the art without departing from the invention. Furthermore, it shall be understood that all aspects of the invention are not limited to the specific depictions, configurations or relative proportions set forth herein which depend upon