Osmylated DNA, a novel concept for sequencing DNA using nanopores

To cite this article: Anastassia Kanavarioti 2015 Nanotechnology 26 134003

View the article online for updates and enhancements.

Related content
- Threading DNA through nanopores for biosensing applications
  Maria Fyta
- DNA stretching and optimization of nucleobase recognition in enzymatic nanopore sequencing
  David Stoddart, Lorenzo Franceschini, Andrew Heron et al.
- DNA sequencing by nanopores: advances and challenges
  Shaghayegh Agah, Ming Zheng, Matteo Pasquali et al.

Recent citations
- Osmium-Based Pyrimidine Contrast Tags for Enhanced Nanopore-Based DNA Base Discrimination
  Robert Y. Henley et al.
Osmylated DNA, a novel concept for sequencing DNA using nanopores

Anastassia Kanavarioti

Yenos Analytical LLC 1350 Terracina Dr El Dorado Hills, CA, USA
E-mail: tessi.kanavarioti@gmail.com

Received 29 September 2014, revised 6 December 2014
Accepted for publication 13 February 2015
Published 11 March 2015

Abstract
Saenger sequencing has led the advances in molecular biology, while faster and cheaper next generation technologies are urgently needed. A newer approach exploits nanopores, natural or solid-state, set in an electrical field, and obtains base sequence information from current variations due to the passage of a ssDNA molecule through the pore. A hurdle in this approach is the fact that the four bases are chemically comparable to each other which leads to small differences in current obstruction. ‘Base calling’ becomes even more challenging because most nanopores sense a short sequence and not individual bases. Perhaps sequencing DNA via nanopores would be more manageable, if only the bases were two, and chemically very different from each other; a sequence of 1s and 0s comes to mind. Osmylated DNA comes close to such a sequence of 1s and 0s. Osmylation is the addition of osmium tetroxide bipyridine across the C5–C6 double bond of the pyrimidines. Osmylation adds almost 400% mass to the reactive base, creates a sterically and electronically notably different molecule, labeled 1, compared to the unreactive purines, labeled 0. If osmylated DNA were successfully sequenced, the result would be a sequence of osmylated pyrimidines (1), and purines (0), and not of the actual nucleobases. To solve this problem we studied the osmylation reaction with short oligos and with M13mp18, a long ssDNA, developed a UV–vis assay to measure extent of osmylation, and designed two protocols. Protocol A uses mild conditions and yields osmylated thymidines (1), while leaving the other three bases (0) practically intact. Protocol B uses harsher conditions and effectively osmylates both pyrimidines, but not the purines. Applying these two protocols also to the complementary of the target polynucleotide yields a total of four osmylated strands that collectively could define the actual base sequence of the target DNA.

Keywords: nanopore DNA sequencing, osmium tetroxide bipyridine, chemically modified nucleobases, next generation sequencing, pyrimidine osmylation

(Some figures may appear in colour only in the online journal)

1. Introduction

A concerted effort by government, industry, and academia over the last 10 years has sought to speed up and reduce the cost of DNA sequencing [1], in the hope that dependable genomic datasets will improve our understanding of biology and drive personalized medicine. The goal is to make whole genome sequencing as easy, fast, accurate, and cheap as a routine blood test. This prospect is made plausible by advances in computation to handle the enormity of the obtained data and advances in nanofabrication that yield massively parallel structures functional for single molecule applications.

A few nucleic acid sequencing methods are currently in use, most of them a variation of the original method by Saenger [2], based on the enzymatic synthesis of the complementary strand. These methods have resulted in faster and cheaper sequencing, readings of good accuracy, but are limited by their inability to read certain repetitive sequences, and
generally result in short reads from which assembly of unknown genomes is challenging [3].

A major breakthrough came with the development and implementation of both natural and solid-state nanopores to allow for the translocation and detection of a single strand of DNA, in its entirety, via the pore. It has been demonstrated that alpha-hemolysin protein self-assembles in a lipid bilayer membrane to form a pore with a 2.6 nm diameter vestibule and 1.5 nm limiting aperture (the narrowest point of the pore) [4]. The limiting aperture of the nanopore allows linear single-stranded but not double-stranded, nucleic acid molecules (diameter ∼ 2.0 nm) to pass through. In an aqueous ionic salt solution, when an appropriate voltage is applied across the membrane, the pore formed by an alpha-hemolysin channel conducts a sufficiently strong and steady ionic current [4]. The polyanionic nucleic acid strand is driven through the pore by the applied electric field, thus reducing or even blocking the ionic current that would be otherwise unimpeded. This translocation process generates an electronic signature that varies based on steric and electronic effects and hence is sequence dependent [4]. In this system the sequence dependent information does not read one base at a time due to the length of the pore. Devices that sequence via natural nanopores are now commercially available and have resulted in single base calling interpretation, but with substantial and, perhaps, unacceptable error rates [5].

A separate technology involves fabricating pores made of insulated single atom graphene layers that can sense chemical changes by transverse current electronics or tunneling effects, but these efforts are still on the drawing table due to the challenges in manufacturing sensors for single layer atoms [8]. Although nanopore approaches show promise as full length DNA sensing methods, the more demanding goal of accurate base-to-base sensing has not yet been achieved, primarily due to (i) the relative similarity of the bases with respect to steric and electronic effects, (ii) the questionable sensitivity of the applied probe, i.e. water and salt ions, to those effects and (iii) the fact that the translocation via the pores is very fast and in the same order of magnitude as detection speed [4, 9].

In an attempt to circumvent those issues we are proposing the use of labeled DNA instead of native. In the 1960s a far-reaching approach to whole genome sequencing was postulated and explored. The concept was to react a nucleic acid with an organo-metallic molecule that exhibited selectivity for one of the four bases [10]. Labeled DNA would then be stretched on suitable surfaces, imaged by electron microscopy to detect the position of the metal and infer the position of the labeled base [11, 12]. This powerful approach encountered a number of obstacles, such as the need for expensive instrumentation and special lab environmental provisions; implausible miniaturization, issues with Brownian motion of the label, etc. Nevertheless the effort recently produced at least one suitable DNA label, the osmium tetroxide 2,2′-bipyridine (Osbipy) [13]. Osbipy is a thymidine (T)-specific label that is known to react preferentially with the C5–C6 double bond of exposed or unpaired T in DNA.

In addition to the above protein nanopores research has moved towards man-made solid-state pores that are functional in the absence of a membrane, robust towards environmental parameters, and can be customized with regard to material composition, as well as pore width and length [7]. A separate system involves using Mycobacterium smegmatis porin (MspA) including modifications thereof, that appears better suited due to a shorter pore that fits fewer bases at a time [6]. Nevertheless even in this system the signal apparently corresponds to a sequence of four bases at a time and therefore base calling depends on discriminating among 4^4 = 256 different signals [6]. Considering the chemical comparability of the four nucleobases such discrimination may not lead to unambiguous sequence determination for long reads.

In an attempt to circumvent those issues we are proposing the use of labeled DNA instead of native. In the 1960s a far-reaching approach to whole genome sequencing was postulated and explored. The concept was to react a nucleic acid with an organo-metallic molecule that exhibited selectivity for one of the four bases [10]. Labeled DNA would then be stretched on suitable surfaces, imaged by electron microscopy to detect the position of the metal and infer the position of the labeled base [11, 12]. This powerful approach encountered a number of obstacles, such as the need for expensive instrumentation and special lab environmental provisions; implausible miniaturization, issues with Brownian motion of the label, etc. Nevertheless the effort recently produced at least one suitable DNA label, the osmium tetroxide 2,2′-bipyridine (Osbipy) [13]. Osbipy is a thymidine (T)-specific label that is known to react preferentially with the C5–C6 double bond of exposed or unpaired T in DNA.

Scheme 1. Reaction of osmium tetroxide with 2,2′-bipyridine to form the reactive complex (bipy-OsO₄), which in a second step reacts with a thymidine derivative (dTMP here) at the C5–C6 double bond to form the osmylated-dTMP (see [13] and references therein). Similarly, cytidine (C) yields osmylated-cytidine also at the C5–C6 double bond (not shown), but at a much slower rate. This work reports on deoxynucleotides and deoxyoligos, so that C in the text refers to dC. A simplistic way to illustrate the difference between osmylated and intact bases is to compare (molecular weight) of each: C (111), T (126), A (135), G (151); C-Osbipy (521), T-Osbipy (536), i.e. osmylation adds about 400% mass to the reactive base compared to the unreactive one.

2
Using these materials we could further extend our tests to randomly selected oligonucleotides of different composition, length, and sequence. The reaction mixtures can be easily purified and their validity for C-osmylation. We then develop methods to assess extent of osmylation, i.e. fraction of osmylated DNA over time, and (vi) a UV–vis assay to include quantitative assessment of total osmylated DNA.

In this report we substantiate the prediction of labeling long DNA by extending our tests to randomly selected oligos as well as to M13mp18, a circular ssDNA 7459 bases long. The reactions were monitored automatically by capillary electrophoresis (CE). Product distributions as well as rates of reactions were obtained. Here is a summary of the conclusions: [13] (i) the osmylation has excellent selectivity for T over C, with no competing reaction or coordination with the purine bases. (ii) Labeling conditions for T are mild and yield undetectable degradation of the phosphodiester bond. (iii) The reaction mixtures can be easily purified from the unreacted Osbipy using TrimGen spin columns. (iv) Osmylated DNA is stable and soluble in water, (v) no detectable amount of Osbipy or 2,2′-bipyridine is released from osmylated DNA over time, and (vi) a UV–vis assay was successfully developed to assess extent of osmylation, i.e. fraction of modified nucleobases over the total number of nucleotides in any partially or fully T-osmylated strand. Most importantly, it was established (vii) that the rate of T-osmylation using a large excess of Osbipy and under the same set of conditions, such as concentration of reactants and temperature, is insensitive to the relative abundance of T, the length, as well as the sequence of the reacting oligo. The observation of equal rates across all tested oligos including deoxythymidine triphosphate (dTTP) implies that the same osmylation protocol would become 24 = 16, a much simpler task.

Proposing to use labeled/modified DNA for sequencing in place of the actual target DNA places stringent requirements on the label to be used, so that the information encoded in the DNA is transferred to the new product with greatest accuracy. Some of these requirements are: (i) stability of the DNA backbone under the labeling conditions, and independence of the labeling rate with (ii) length and (iii) sequence. Our earlier work as well as the work reported here strongly support the idea that these three requirements are met with Osbipy, and hence allowed the design of protocols that are expected to work with any DNA.

What would be the anticipated error rate of sequencing osmylated DNA using scheme 2? Assuming a strong signal (1) to noise (0) ratio, and considering the independence of osmylation with respect to length and sequence, then the expected error rate should be a function of the intrinsic signal to noise ratio, the apparent signal to noise ratio based on the coverage. For example, Protocol A, as detailed below, results in approximately 90% osmylated T (1) and about 6.5% of osmylated C (1) and practically 100% intact purines (0) (see under 4.3.1). This apparent selectivity implies that a specific T has 90% probability to register as 1 and only 10% probability to register as 0. For the same strand a specific C

5′----0-0-0-0-0-0-0-0-0-0-0-0-0-0-3′... Protocol A
5′----0-0-0-0-0-0-0-0-0-0-0-0-0-0-3′... Protocol B

3′----0-0-0-0-0-0-0-0-0-0-0-0-0-0-5′... Protocol A
3′----0-0-0-0-0-0-0-0-0-0-0-0-0-0-5′... Protocol B

2. A novel concept for sequencing DNA

In the context of modifying DNA one might have favored a scenario with four different labels each selective for one of the four bases, in analogy to Sanger sequencing. However four labels have proven difficult to find, and may even complicate the detection method. A few alternatives have been evaluated using surrogates instead of the actual bases [17]. Scheme 2 illustrates our vision of how sequence information can be obtained with a single label. The concept is based on the apparent dramatic chemical, both steric and electronic, difference between an osmylated and an intact base (see caption in scheme 1).

A Kanavarioti
has only 6.5% probability to register as 1 and 93.5% probability to register as 0. We speculate that the coverage necessary to establish unequivocally T and C in a sequence is in the tens, while strands passing via nanopores per minute are in the hundreds. Depending upon the signal to noise ratio as defined above, the error rate in sequencing osmylated DNA may not be a major issue.

Last, but not least, scheme 2 proposes the sequencing of four strands, even though sequencing of three should suffice; the fourth serves as control. Specifically, Protocol A on the target strand defines the position of T and Protocol B on the target strand defines the position of both T and C. In principle, only one more sequencing is necessary, for example Protocol A of the complementary strand, to define A. Positions not identified as T, C, or A, are presumably G. Sequencing of the fourth strand serves as control to assess accuracy in base calling.

### 3. Materials and methods

#### 3.1. Materials and purification

HPCE grade solutions of 100 mM sodium phosphate pH 7.0, 50 mM sodium tetraborate pH 9.3 and 1 N sodium hydroxide were purchased from Agilent Technologies. A 4% aqueous osmium tetroxide solution was purchased from Electron Microscopy Sciences. 2,2'-bipyridine 99+% (bipyridine) was purchased from Acros Organics. Solutions of the triphosphates, dATP, dCTP, dGTP and dTTP, as well as an equimolar mixture of the four dNTPs were purchased from Zymo Research Corporation. A solution of M13mp18 was purchased from Bayou Labs as well as from New England Biolabs. Oligos were purchased from Integrated DNA Technologies, diluted with DNase/RNase-free water (from MP Biomedical) to 1 μg μL⁻¹. Solutions of mononucleotides, oligos, and M13mp18, a circular 7249 nucleotide long ssDNA, were all stored at −20 °C. One group of oligos, identified as Oligo1 through 10 was of relatively low purity, and the other group, identified as primers, was of purity 88% or higher. Oligos used in this study, their sequences, and percent purity of main peak is listed in table 1.

Aqueous reaction mixtures were prepared with DNase/RNase-free water. A stock solution of Osbipy at 15.75 mM (OsO₄: bipyridine = 1:1) was prepared by mixing bipyridine in OsO₄ solution. This stock solution was dispensed in small glass vials, sealed with parafilm, and kept at −20 °C until use; no detectable change in reactivity was observed with solutions that were stored sealed at 4 °C for two weeks.

<table>
<thead>
<tr>
<th>ODN</th>
<th>% purity</th>
<th>Sequence: deoxyoligonucleotide (5′—3′)</th>
</tr>
</thead>
<tbody>
<tr>
<td>dTTP</td>
<td>100</td>
<td>T</td>
</tr>
<tr>
<td>dCPT</td>
<td>100</td>
<td>C</td>
</tr>
<tr>
<td>Oligo1</td>
<td>77</td>
<td>CGC GGG GTG GAG CAG CCT GG</td>
</tr>
<tr>
<td>Oligo2</td>
<td>84</td>
<td>TAG CTC GTC GGG CTC ATA ACC CGA AGA TCG TCG GTT CAA-ATC CGG CCC CCG CAA CCA</td>
</tr>
<tr>
<td>Oligo3</td>
<td>88</td>
<td>CCA GGC TGG TCC ACC CCG CG</td>
</tr>
<tr>
<td>Oligo4</td>
<td>47</td>
<td>TGG TTG CGG GGG CCG GAT TTG AAC CGA GTA CTT TCG GGT-TAT GAG CCC GAC GAG CTA</td>
</tr>
<tr>
<td>Oligo5</td>
<td>77</td>
<td>ATA CTG GTG GAG CAG TTG GG</td>
</tr>
<tr>
<td>Oligo6</td>
<td>90</td>
<td>CGC GAA GTG GAG CAG (5-Me-dC)CT GG</td>
</tr>
<tr>
<td>Oligo7</td>
<td>91</td>
<td>CGC GAA GTG GAG CAG CCT GG</td>
</tr>
<tr>
<td>Oligo8</td>
<td>90</td>
<td>AAA ACC CAA CCC AAA A</td>
</tr>
<tr>
<td>Oligo9</td>
<td>86</td>
<td>AAA AAC CCC CCA AAA A</td>
</tr>
<tr>
<td>Oligo10</td>
<td>88</td>
<td>AAA CAC ACA CAC ACA A</td>
</tr>
<tr>
<td>OligodT(15)</td>
<td>91</td>
<td>TTT TTT TTT TTT TTT</td>
</tr>
<tr>
<td>3RACE PCR</td>
<td>91</td>
<td>GCC CAC CGG TCG ACT AGT AC</td>
</tr>
<tr>
<td>pGEX 3′</td>
<td>90</td>
<td>CCG GGA GCT CTA TGT GTC AGA GG</td>
</tr>
<tr>
<td>pGEX 5′</td>
<td>91</td>
<td>GGG CTC GCA ACG CAC GTT TGG TG</td>
</tr>
<tr>
<td>Bluescript SK</td>
<td>88</td>
<td>CGC TCT AGA ACT AGT GGA TC</td>
</tr>
<tr>
<td>Neomycin for</td>
<td>93</td>
<td>CTT GGG TGG AGA GGC TAT TC</td>
</tr>
<tr>
<td>Neomycin rev</td>
<td>89</td>
<td>ACG TGA GAT GAC AGC TAA AG</td>
</tr>
<tr>
<td>16S RNA for</td>
<td>93</td>
<td>AGA GTT TGA TCC TGG CTC AG</td>
</tr>
<tr>
<td>M13 rev −48</td>
<td>97</td>
<td>AGC GGA TAA CAA TTT CAC ACA GG</td>
</tr>
<tr>
<td>M13mp18b, c</td>
<td>100</td>
<td>c</td>
</tr>
<tr>
<td>M13mp18b, d</td>
<td>100</td>
<td>d</td>
</tr>
</tbody>
</table>

See section 3.

a M13mp18, a circular ssDNA 7459 bases long, is longer than bacteriophage ssM13 that is only 6407 bases long. M13 is not commercially available.

b M13mp18 from Bayou Labs.

c From New England Biolabs, both share the same sequence.

Table 1. List of Oligos/DNA used in this study, their purity, and sequence. The data obtained can be found in table 2.
Experiments were initiated by mixing the Osbipy stock solution and the oligo stock solution directly in a CE glass vial filled with the appropriate amount of distilled water. The final concentration of Osbipy and oligo in the reaction mixture is reported with each experiment. No buffer was added and reaction mixtures were incubated at 27 ± 2 °C. Reproducible rate determinations in unbuffered solutions confirmed that buffer is unnecessary and showed that some buffers react with Osbipy [13]. Analyses were carried out automatically. The aliquoting time together with the product distribution, in the form of an electropherogram, were recorded by the CE. Spin columns TC-100 from TrimGen were used to remove excess Osbipy [13] according to the manufacturer’s instructions. Practically 100% recovery of labeled oligo and removal of up to 15 mM Osbipy down to detectability levels after just one round was achieved.

3.2. CE methods, analyses, and peak resolution

Analyses of the reaction mixtures were carried out using an Agilent G1600A CE instrument equipped with Diode Array Detector (DAD) and Chemstation software, Rev.B.04.02SP1 [212], for data acquisition and processing. Only glass type CE vials were used, as Osbipy was found to react with polyurethane vials, lowering its effective concentration. Reaction mixtures were also prepared in sealed glass microvials and aliquoted as needed; aliquots were typically diluted with water before analysis. Untreated fused silica capillaries (50 μm × 40 cm) with extended light path were purchased from Agilent Technologies.

Analysis of reaction mixtures was conducted using capillary zone electrophoresis methods: either in 50 mM sodium phosphate buffer pH 7.0 with 20 kV, or in 50 mM sodium tetraborate pH 9.3 with 20 kV. All the data reported here were obtained using the borate buffer due to its superior resolution and the fact that capillaries showed substantially higher resilience compared to when phosphate buffer was used. The capillary’s temperature was matched to the autosampler’s temperature and ranged from 25 to 29 °C. A typical method was 18 min long and included a 4 min automatic buffer conditioning ahead of sample analysis. Osbipy migrates early together with any other neutral compound present, the triphosphates of the nucleotides migrate last, and oligos/DNA in between. Osbipy-labeled products migrate between the Osbipy peak and the corresponding DNA peak. Resolution between starting material and osmylated product, is among other parameters, a function of the length and the number of reacting bases. With the longer DNA, a shift of the peak to earlier migration is detected as a function of reaction progress. Peaks on CE were detected and identified using the DAD in the UV–vis region 200–450 nm and the electropherograms were recorded at several wavelengths including 272 and 312 nm. In contrast to DNA that exhibits no

<table>
<thead>
<tr>
<th>ODN</th>
<th># T</th>
<th># C</th>
<th># N_total</th>
<th>T/N_total</th>
<th>(T+C)/N_total</th>
<th>R1 (312/272) (T)</th>
<th>R2 (312/272) (T+C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>dTTP</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2.77</td>
<td>—</td>
</tr>
<tr>
<td>dCTP</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>—</td>
<td>2.2</td>
<td>—</td>
</tr>
<tr>
<td>Oligo1</td>
<td>2</td>
<td>5</td>
<td>20</td>
<td>0.10</td>
<td>0.35</td>
<td>0.25</td>
<td>0.63</td>
</tr>
<tr>
<td>Oligo2</td>
<td>10</td>
<td>21</td>
<td>57</td>
<td>0.18</td>
<td>0.54</td>
<td>0.42</td>
<td>1.12</td>
</tr>
<tr>
<td>Oligo3</td>
<td>2</td>
<td>11</td>
<td>20</td>
<td>0.10</td>
<td>0.65</td>
<td>0.35</td>
<td>1.26</td>
</tr>
<tr>
<td>Oligo4</td>
<td>13</td>
<td>13</td>
<td>57</td>
<td>0.23</td>
<td>0.46</td>
<td>0.50</td>
<td>0.88</td>
</tr>
<tr>
<td>Oligo5</td>
<td>6</td>
<td>2</td>
<td>20</td>
<td>0.30</td>
<td>0.40</td>
<td>0.59</td>
<td>0.82</td>
</tr>
<tr>
<td>Oligo6</td>
<td>2</td>
<td>5</td>
<td>20</td>
<td>0.10</td>
<td>0.35</td>
<td>0.26</td>
<td>0.73</td>
</tr>
<tr>
<td>Oligo7</td>
<td>2</td>
<td>5</td>
<td>20</td>
<td>0.10</td>
<td>0.35</td>
<td>0.24</td>
<td>0.65</td>
</tr>
<tr>
<td>Oligo8</td>
<td>0</td>
<td>6</td>
<td>16</td>
<td>0</td>
<td>0.375</td>
<td>0.76</td>
<td>—</td>
</tr>
<tr>
<td>Oligo9</td>
<td>0</td>
<td>6</td>
<td>16</td>
<td>0</td>
<td>0.375</td>
<td>—</td>
<td>0.76</td>
</tr>
<tr>
<td>Oligo10</td>
<td>0</td>
<td>6</td>
<td>16</td>
<td>0</td>
<td>0.375</td>
<td>—</td>
<td>0.72</td>
</tr>
<tr>
<td>OligoT(15)</td>
<td>15</td>
<td>0</td>
<td>15</td>
<td>1</td>
<td>3</td>
<td>2.77</td>
<td>—</td>
</tr>
<tr>
<td>3’RACE PCR</td>
<td>3</td>
<td>7</td>
<td>20</td>
<td>0.15</td>
<td>0.50</td>
<td>0.37</td>
<td>0.98</td>
</tr>
<tr>
<td>pGEX 3’</td>
<td>4</td>
<td>5</td>
<td>23</td>
<td>0.17</td>
<td>0.39</td>
<td>0.40</td>
<td>0.79</td>
</tr>
<tr>
<td>pGEX 5’</td>
<td>5</td>
<td>5</td>
<td>23</td>
<td>0.22</td>
<td>0.43</td>
<td>0.42</td>
<td>0.85</td>
</tr>
<tr>
<td>Bluescript SK</td>
<td>5</td>
<td>5</td>
<td>20</td>
<td>0.25</td>
<td>0.50</td>
<td>0.57</td>
<td>1.05</td>
</tr>
<tr>
<td>Neomycin for</td>
<td>6</td>
<td>3</td>
<td>20</td>
<td>0.30</td>
<td>0.45</td>
<td>0.65</td>
<td>0.94</td>
</tr>
<tr>
<td>Neomycin rev</td>
<td>3</td>
<td>2</td>
<td>20</td>
<td>0.15</td>
<td>0.25</td>
<td>0.32</td>
<td>0.50</td>
</tr>
<tr>
<td>16S RNA for</td>
<td>6</td>
<td>4</td>
<td>20</td>
<td>0.30</td>
<td>0.50</td>
<td>0.64</td>
<td>1.05</td>
</tr>
<tr>
<td>M13 rev −48</td>
<td>4</td>
<td>5</td>
<td>23</td>
<td>0.17</td>
<td>0.39</td>
<td>0.41</td>
<td>0.78</td>
</tr>
<tr>
<td>M13mp18b</td>
<td>2410</td>
<td>1527</td>
<td>7249</td>
<td>0.33</td>
<td>0.54</td>
<td>0.75</td>
<td>1.18</td>
</tr>
<tr>
<td>M13mp18c</td>
<td>2410</td>
<td>1527</td>
<td>7249</td>
<td>0.33</td>
<td>0.54</td>
<td>0.77</td>
<td>1.17</td>
</tr>
</tbody>
</table>

R1 values for Oligo3, Oligo5, and pGEX 5’ deviate more than average from the line that statistically fits all the data (see figure 3).

The difference between Oligo6 and Oligo7 is that in Oligo6 one of the Cs is a 5-Me-dC.

A circular ssDNA 7459 bases long; b from Bayou Labs; c from New England BioLabs.
detectable absorbance at 312 nm, osmylated DNA absorbs at 312 nm. Interestingly Osbipy exhibits strong absorption at 272 nm and weak absorption at 312 nm; it is the osmylated product that absorbs strongly at 312 nm [13]. We have exploited this feature and developed an UV/vis assay to monitor extent of modification.

4. Results and discussion

Our earlier work [13] made two important discoveries: (i) the feasibility to osmylate T in oligos as long as 90-mer effectively, predictably, and selectively and (ii) the UV–vis assay to measure extent of T-osmylation. Here we confirmed the reported attributes of T osmylation, evaluated osmylation of C, developed the conditions to do it reproducibly, and extended the UV–vis assay to measure both, osmylated T and C. While it was known that harsher conditions will label C in addition to T [16], the objective was to find conditions so that C is practically 100% osmylated, while the backbone and the purines remain intact. For this purpose we exploited a series of oligos (Oligo1 through 10) of variable length and composition. Because the purity of these oligos became a concern, we purchased and evaluated a second set of oligos, of the highest commercially available purity, and commonly used as primers. M13mp18, a circular ssDNA 7459 bases long, was also included (see table 1).

4.1. Quantitative assay to assess extent of osmylation

The difficulty in labeling DNA or RNA is that these polymers are composed of four repeating units and there is no routine analytical tool that discriminates among positional isomers and determines extent of modification. One approach is to use enzymatic degradation to monomers or dimers and assess percent of intact and labeled fragments. However this is not a trivial pursuit as the enzyme may be partially inhibited by the presence of the label. Hence the observation that the incremental addition of Osbipy to an oligo results in a proportional increase in absorbance, or the equivalent peak area, at wavelengths in the range 300–330 nm, where DNA does not absorb, was a game changer. The proportionality between number of osmylated bases and increase in peak area in the range 300–330 nm was also documented by CE analysis that resolves (n+1)-from n-osmylated and from intact starting material in short oligos [13].

A rather extreme example of resolution is shown in figure 1 where the osmylation reaction of a 16-mer oligo

![Figure 1. CE profiles of sequential analyses monitoring the osmylation of Oligo (here Oligo10 with 6 Cs) at 260 nm every 17 min, top to bottom. Mixture contains 300 ng μL−1 Oligo10 with 7.9 mM Osbipy at 27 °C in water. T1 is obtained after only 6 min from mixing the reactants. Still one can clearly detect two groups of peaks: Group 1 is believed to be the singly osmylated oligo. Group 1 has multiple peaks due to the six plausibe positional isomers. Please note that CE is known to resolve positional isomers. Group of peaks, labeled 2, represents products with two osmylated Cs, group 3 with three osmylate Cs, group 4 with four, group 5 with five, and group 6 with six osmylated Cs. Even after many hours there was no additional peak migrating ahead of group 6. However these conditions, i.e., relatively high concentration of oligo and only 7.9 mM Osbipy do not lead to complete osmylation and the reaction levels off. The reason we chose to show these conditions is because the reaction is relatively slow compared to the CE analysis and one can clearly follow the appearance of higher osmylated products and the accumulation of the material initially from the Oligo to group 1, and then to group 2, followed by accumulation to group 3 (see CE profile T3) and to C4 (see CE profile T4, bottom). Please note that in T4 the oligo peak is barely detectable. Identification of products, as specified here, is supported by the observation that R(312/272) of a certain group peak is proportional to the proposed number of labeled Cs. Separate and well resolved group of peaks, such as the ones observed with Oligo10, were not observed with either Oligo8 or Oligo9, even though their composition is identical to Oligo10.

![Figure 1](image-url)
composed of 6 Cs and 10 As is monitored by CE. The addition of Osbipy creates a group peak (composed of all different positional isomers), distinct from the group peak of the products with one less or one more Osbipy moiety. It is noticeable that increasing number of conjugated Osbipy moieties leads to decreasing migration time of the corresponding peak (see figure caption for discussion). The earlier work exploited 320 nm as the wavelength to measure the appearance of products and also proposed and implemented a more accurate measure, the ratio of peak area at 320 nm versus peak area at 260 nm [13]. This measure represents a normalized absorbance, a thermodynamic property of the material at hand, independent of concentration, and independent of instrument variability. This measure ($R$ for ratio) represents the cornerstone of the quantitative assay. In order to improve the sensitivity of $R$, wavelengths were selected, by inspection of the UV–vis spectra of osmylated dTTP and dCTP, where $R$ is largest. Those wavelengths are 312 and 272 nm, and correspondingly replace 320 and 260 nm discussed above. The new measure is ratio of peak area at 312 versus peak area at 272 nm, $R$ (312/272). Evidently, any spectrophotometer can be used to determine $R$ (312/272), as long as the osmylated DNA has been purified from unreacted Osbipy (see 3.1).

4.2. Experiments

We conducted two types of experiments: (i) reaction mixtures were prepared directly in CE vials and monitored automatically for a few hours; this was done to confirm rates of reactions, or (ii) reaction mixtures were prepared in sealed glass minivials; Protocol A or B, or another test protocol was followed, and samples were analyzed in order to assess $R$ (312/272) values. Most of the reaction mixtures were analyzed before TrimGen purification, as the Osbipy peak migrates very early and does not interfere with the reactants or products. The table in the supplementary material (SM) includes all the experiments together with the number of analyses performed for each sample and the resulting information.

Earlier work established that 3 mM Osbipy is necessary to bring T-osmylation to completion [13]. Here we used routinely 3.15 mM Osbipy. We then evaluated osmylation of C at 7.88, 9.45, 12.6, and 14.2 mM Osbipy. It was observed that 7.88 mM Osbipy leaves 3% intact dCTP, even after 17 h of incubation, whereas the higher concentrations lead to complete disappearance of the dCTP peak. Hence the choice was made to use 14.2 mM Osbipy and 11 h incubation at room temperature to osmylate both T and C, even though somewhat lower concentrations worked too. Earlier work also established that DNA concentration should be kept below 200 ng μL$^{-1}$, and so here all the experiments, but one (see figure 1), were conducted with 200 ng μL$^{-1}$ DNA or less.

The series of Oligos1 through 10 were used without additional purification, and their purity, as tested by us, was found to vary from 77 to 91%, with the exception of Oligo4 that was only 47% pure. As the method applied does not test length, the actual purity of the material may be even lower. The low purity of these materials was a concern, and for this reason we purchased and evaluated a series of oligos that are used as primers and exhibited purity better than 88%. Not surprisingly, if the data are plotted as two groups instead of one, the standard deviation to the line that fits the data is much tighter for the purer oligos. Nevertheless every single oligo tested has been included in the correlations (see figures 3 and 4 and later discussion).

4.3. Oligo/DNA denaturation

Initially it was presumed that oligos with high CG content, such as Oligos 1, 3, 6 and 7, as well as M13mp18 will require some form of denaturation for effective osmylation. Hence osmylation of oligos and M13mp18 was done in the presence/absence of urea. Reactions with the mononucleotides in the presence of urea were used as controls (see SM). In the presence of urea (2.7, 3.0, or 4.2 M) reactivity was reduced by about 20%, and the data exhibited more scatter compared to the data obtained in the absence of urea. The lower reactivity and the scatter in the data may be attributed to the higher viscosity of the medium. Still the presence/absence of urea did not change the outcome, as measured by $R$ (312/272) of the fully osmylated DNA. The observation that the presence of urea did not increase the degree of osmylation in M13mp18 was intriguing. It is plausible that relatively high Osbipy concentration, such as 10 mM and higher, is playing a denaturing role. For simplicity then the proposed protocols do not include urea.

4.3.1. Protocol A. The recommendation is to use 50–200 ng μL$^{-1}$ DNA with 3.15 mM Osbipy in a glass vial, incubate 60 min at room temperature and purify immediately using TrimGen (Protocol A). Following this protocol the purified osmylated DNA will exhibit a ratio of peak area at 312 nm versus 272 nm identified as $R_1$ (312/272). This material may be kept in a regular test tube refrigerated for several months. As control dTTP or any other oligo/DNA, for which $R_1$ (312/272) is known, could be used. After 60 min T-osmylation has completed approximately 3.5 half lives, about 90% of T is osmylated, and about 6.5% of C is osmylated (see table 3 in reference [13]). The exact product distribution depends on the actual preparation of the Osbipy stock solution and the final concentration of Osbipy in the reaction mixture, and can only be precisely assessed by running controls. The apparent selectivity of T over C using Protocol A is about 14:1 in favor of T. This selectivity may yield excellent discrimination of T over C for a sequencing experiment with coverage in the tens. Please note that at this level T-osmylation has dramatically slowed down, and C-osmylation is very slow due to the low concentration of Osbipy. So small variations, let us say ±2 min, to the 60 min interval, from sample to sample will still result in practically comparable product distributions.

4.3.2. Protocol B. The recommendation is to use 50–200 ng μL$^{-1}$ DNA with 14.2 mM Osbipy in a glass vial and incubate for 11 h at room temperature before purification.
Osmylated products in the presence of excess Osbipy were shown by CE analysis to be stable for days at room temperature. However it is recommended that products be purified with TrimGen soon after incubation and stored at 4 °C. Evaluation of new purifications methods may be conducted using CE. The conditions of Protocol B are not as strict as the ones for Protocol A. Other comparable conditions can be selected by inspection of the table (see SM). The purified osmylated DNA, following Protocol B, will exhibit a ratio of peak area at 312 nm versus 272 nm, identified as R2 (312/272). Figure 2 illustrates the overlapping CE profiles of three materials: intact M13mp18 (peak identified as M13), M13mp18 osmylated based on Protocol A (peak identified as M13 (R1)) and also based on Protocol B (peak identified as M13 (R2)). For each material CE profiles monitored at two wavelengths, 312 and 272 nm, are included (see figure caption for discussion).

4.4. Determination of T- and C-chromophore, and hypochromicity

It is relatively easy to determine R (312/272) for T using dTTP, because osmylated-dTTP is well resolved from dTTP and hence there is no need to wait until all the starting material is osmylated. As seen in the table (SM) eight analyses with dTTP provided R1 (312/272) = 2.77 ± 0.05. This value is identical to the plateau value of R1 (312/272) = 2.77 ± 0.02 with oligodT(15) (see SM). The comparable measure at the other two wavelengths provided R1 (320/260) = 1.53 ± 0.05 and it was valid for dTTP, oligos with T and A, as well as obtained from the extrapolation of the sloping infinity line of a plot of R (320/260) for longer oligos that included all four bases [13]. The change from R (320/260) to R (312/272) increased the sensitivity to measure extent of T- osmylation by about 180%.

Multiple measurements from the osmylated-dCTP product peak (see SM) provided an average R2 (312/272) = 2.2 ± 0.1. Oligo8 through 10 are 16 bases long each and contain 10 A and 6 C each, but they have different sequence. In Oligo9 all Cs are in line, whereas Oligo8 has two groups of 3 C separated by two A. In Oligo10 C and A bases alternate. Incubation of these three oligos with 14.2 mM Osbipy for 10 or 22 h provided average values of R2 (312/272) equal to 0.76, 0.76 and 0.72, respectively for Oligo8, Oligo9 and Oligo10. Accounting for the presence of only 6 Cs out of a total number of 16 nucleobases, provided R (312/272) = 0.76 x 16/6 = 2.0 per C from the first two oligos and 1.9 from Oligo10. The agreements between the values obtained from dCTP and the Oligo8–10 is not as perfect as the agreement obtained for T, but it is acceptable and taken as 2.1 ± 0.2 per C. Comparable measurements provided R (320/260) ≈ 1 per C [13], so here the change of wavelengths not only doubled the sensitivity, but actually made C-osmylation detectable and quantifiable.

The critical question is whether or not one can practically osmylate all the pyrimidines in M13mp18, a prototype for osmylating long ssDNA. Towards this end we conducted nine different experiments with the material from Bayou Labs and two experiments with the material from New England Biolabs (see SM). Most of the experiments were aimed at obtaining R2 (312/272) and about a third of those experiments were done in the presence of urea as denaturant and using prolonged incubation for up to 5 days. These experiments provided R2 (312/272) = 1.18 ± 0.01 from 10 different conditions. It was only in the presence of 7.16 mM Osbipy (with 2.7 M urea) that R2 = 1.14, presumably because of the low Osbipy concentration. M13mp18 from New England Biolabs in a single experiment following Protocol B exhibited R2 (312/272) = 1.17. The fact that different experimental...
conditions, with increasing concentrations of the label and the denaturant, consistently provided identical $R^2$ values was considered evidence for having achieved complete osmylation. Urea as a denaturant was also included in experiments with Oligo2 and Oligo5 and no detectable difference was observed there either (SM). Hence Protocol B can be used with any ssDNA and achieve practically complete osmylation.

Figure 3 illustrates the relationship between extent of T-osmylation and the fraction of T bases in an oligo/DNA by plotting $R_1$ (312/272), as obtained using Protocol A, versus T/Ntotal. The line is the best fit to all the data and forced to go via the intercept. The correlation appears very good considering that synthetically made oligos are intrinsically less pure than biologically produced polymers. The good fit of 17 oligos/DNA in a single correlation strongly supports the validity and robustness of Protocol A. The fact that the slope of the line is 2.2 and not 2.77, as found for complete T osmylation, can be attributed to the fact that Protocol A does not lead to complete T-osmylation. As seen in table 3 ([13]) 102 min, instead of 60, would yield practically complete T-osmylation, in addition to 10% of osmylated C. Whether or not Protocol A or a protocol with a longer incubation is more useful for material to be exploited for sequencing is left to the experts. Nevertheless it is apparent from figure 3 that Oigos and ssDNA osmylate comparably and reproducibly under these conditions.

Figure 4 illustrates the relationship between extent of (T+C)-osmylation and the fraction of pyrimidines in an oligo by plotting $R_2$ (312/272), as obtained from Protocol B, versus (T+C)/Ntotal. The line is the best fit to all the data from 20 oligos/DNA and was forced through the intercept. The correlation is excellent and strongly supports the validity of Protocol B for ssDNA. Our initial expectation was that $R_2$ (312/272) for a given DNA can be calculated from the pyrimidine content, the values for T-and C-chromophore determined above, and given from the theoretical equation $R_2 = 2.77x(T/Ntotal) + 2.1x(C/Ntotal)$. Hence one calculates a theoretical $R_2 = 1.36$ for M13mp18, markedly higher compared to the value of observed $R_2 = 1.17 \pm 0.01$. The observed lower value was attributed to hypochromicity [18] due to proximity of the bulky/hydrophobic Osbipy moiety to a purine base. Actually all observed $R_2$ values are seen to be about 15% lower compared to the theoretical value (not included here; see equation above) and this strongly supported the proposition of a hypochromicity effect.

4.5. Stability of materials and DNA backbone integrity

Repeated CE analyses strongly suggest that partially or fully osmylated material is stable in the presence of excess Osbipy. Not only the product peak did not change its electrophoretic properties, but also the UV–vis spectrum of the material did not change. Moreover new small peaks increasing with time have not been detected even when the unpurified reaction mixtures were left at room temperature for several days. New peaks with areas increasing as a function of time would have been detected at levels as low as 1% of DNA and would be consistent with backbone degradation. The absence of such peaks strongly suggests that Osbipy has no detectable effect on the backbone and the backbone remains intact with concentrations of Osbipy as high as 14.3 mM and incubation of several days at room temperature. This last feature of backbone integrity during and after osmylation is critically important in endorsing the tentative use of osmylated DNA for sequencing purposes.

5. Conclusions

This report presented two protocols, one to partially and another to fully osmylate ssDNA, as well as an UV–vis assay to determine extent of labeling. This set of tools is complete and can efficiently, predictably, and reproducibly osmylate
any nucleic acid and create two products. In the first product practically all Ts are osmylated, and in the second product both pyrimidines are exhaustively osmylated. Similar labeling of the complementary of the target DNA will yield two additional products. This set of four labeled DNAs, as shown in scheme 2, may serve as testing articles, in place of the native target strand. It is plausible that the bulky/hydrophobic Osbipy moiety will fit better in nanopores of width larger than the width required for translocation of ssDNA, and snugly fit nanopores suited for translocation of dsDNA. The steric and electronic features of the Osbipy moiety may result in solvent reorganization and perhaps slow down translocation via the pore. The presence or absence of Osbipy at a base along the strand may be dramatically more discriminating than one base from the others; high discrimination is expected to improve accuracy in ‘base calling’. Scheme 2 uses inexpensive reagents, in contrast to the expensive reagents currently required by the available technologies. Last but not least, the known electroactivity of osmium may result in a markedly stronger signal for detection by transverse current nanopore systems. Among the questions to be answered are: (i) what nanopore set-up might be suitable, (ii) how strong is the current change, i.e. discrimination, between osmylated base and intact one, (iii) whether a molecular motor is needed to process the strand one step at a time, and (iv) what type of molecular motor could be used with such unnatural/labeled ssDNA. Osmylated DNA may tentatively change the task of discriminating among four comparable nucleobases to the substantially simpler task of discriminating two markedly different molecules. Regarding third-generation sequencing approaches, osmylated DNA may turn out to be better suited than native DNA.

Acknowledgments

The work was supported with private funding. The author is grateful to her colleagues and friends, from the former Halcyon Molecular, who introduced her to next generation sequencing. Specifically the author acknowledges collaboration and insightful discussions with Chris R Melville (Greentecaspoon), Kent Kemmish (Stealth Biotech) and Mark Hamalainen (Syntheygo) that led to this work. The author has declared no conflict of interest.

References

Anderson M W and Schrijver I 2010 Next generation DNA sequencing and the future of genomic medicine Rev. Genes 1 38–69
Moorthie S, Christopher J, Mattocks C J and Wright C F 2011 Review of massively parallel DNA sequencing technologies HUGO J. 5 1–12
Akeson M, Branton D, Kasianowicz J J, Brandin E and Deamer D W 1999 Microsecond time-scale discrimination among polycytidylic acid, polyadenylic acid, and polyuridylic acid as homopolymers or as segments within single RNA molecules Biophys. J. 77 3227–33
Nivala J, Marks D B and Akeson M 2013 Unfoldase-mediated protein translocation through an a-hemolysin nanopore Nat. Biotechnol. 31 247–50
Quick J, Quinlan A R and Loman N J 2014 A reference bacterial genome dataset generated on the MinION™ portable single-molecule nanopore sequencer GigaScience 3 22
Quick J, Quinlan A R and Loman N J 2015 Erratum: A reference bacterial genome dataset generated on the MinION™ portable single-molecule nanopore sequencer GigaScience 4 6
Laszlo A H et al 2014 Decoding long nanopore sequencing reads of natural DNA Nat. Biotechnol. 32 829–33
Wanunu M 2012 Nanopores: a journey towards DNA sequencing Phys. Life Rev. 9 125–58
Liang L, Wang Q, Agren H and Tu Y 2014 Computational studies of DNA sequencing with solid-state nanopores: key issues and future prospects Frontiers Chemistry 2 5
nanopores with insulated transverse nanoelectrodes for DNA sensing in salt solution *Electrophoresis* **33** 3488–96


Banerjee S *et al* 2013 Electrochemistry at edge of single graphene layer in a nanopore *ACS Nano* **7** 834–43


Luan B and Stolovitzky G 2013 An Electro-hydrodynamics-based model for the ionic conductivity of solid-state nanopores during DNA translocation *Nanotechnology* **24** 195702

Akahori R, Haga T, Hatano T, Yanagi I, Ohura T, Hamamura H, Iwasaki T, Yokoi T and Anazawa T 2014 Slowing single-stranded DNA translocation through a solid-state nanopore by decreasing the nanopore diameter *Nanotechnology* **25** 275501


Moudrianakis E N and Beer M 1964 A selective reagent for the study of base sequence in nucleic acids *Nature* **204** 685–6


Payne A C *et al* 2013 Molecular threading: mechanical extraction, stretching and placement of DNA molecules from a liquid-air interface *PLoS One* **8** e69058


Borsenberger V, Mitchell N and Howorka S 2009 Chemically labeled nucleotides and oligonucleotides encode DNA for sensing with nanopores *J. Am. Chem. Soc.* **131** 7530–1


Juan E C M *et al* 2010 Insights into the DNA stabilizing contributions of a bicyclic cytosine analogue: crystal structures of DNA duplexes containing 7,8-dihydropyrido [2,3-d]pyrimidin-2-one *Nucleic Acids Res.* **38** 6737–45