nature nanotechnology

ARTICLES

PUBLISHED ONLINE: 22 FEBRUARY 2009 | DOI: 10.1038/NNANO.2009.12

Continuous base identification for single-molecule nanopore DNA sequencing

James Clarke[1], Hai-Chen Wu[2], Lakmal Jayasinghe[1,2], Alpesh Patel[1], Stuart Reid[1] and Hagan Bayley[2]*

A single-molecule method for sequencing DNA that does not require fluorescent labelling could reduce costs and increase sequencing speeds. An exonuclease enzyme might be used to cleave individual nucleotide molecules from the DNA, and when coupled to an appropriate detection system, these nucleotides could be identified in the correct order. Here, we show that a protein nanopore with a covalently attached adapter molecule can continuously identify unlabelled nucleoside 5′-monophosphate molecules with accuracies averaging 99.8%. Methylated cytosine can also be distinguished from the four standard DNA bases: guanine, adenine, thymine and cytosine. The operating conditions are compatible with the exonuclease, and the kinetic data show that the nucleotides have a high probability of translocation through the nanopore and, therefore, of not being registered twice. This highly accurate tool is suitable for integration into a system for sequencing nucleic acids and for analysing epigenetic modifications.

Dramatic improvements in the cost and speed of DNA sequencing are having a profound effect on genome research. For example, the growing understanding of the genetic components of disease is leading to the development of new medical diagnostics and treatments. Genomic information also has many applications outside human medicine, for example, in agriculture, security and defence, and evolutionary biology. However, the costs of sequencing are still too high for routine application. For example, we estimate that the full cost, including instrumentation, sample preparation and labour, of resequencing a haploid human genome to high quality with various second-generation technologies is currently in the region of $100,000 to $1,000,000. In contrast, in 2004, the US National Institutes of Health set a 10-year goal of a $1,000 human genome, a cost believed to be sufficiently low to revolutionize genomic medicine (http://www.genome.gov/12513210).

Nanopore DNA sequencing offers the possibility of a label-free, single-molecule approach that can be performed without the need for sample amplification. Like second-generation systems, nanopore technology is amenable to parallelization, and several cost estimates place nanopore sequencing in the $1,000 range for a complete human genome[1,2]. Furthermore, because the sequence quality should be constant throughout a read, long reads from single molecules of DNA will be possible by using nanopores, offering many advantages including the possibility of *de novo* sequencing, the high-resolution analysis of chromosomal structure variation, and long-range haplotype mapping.

The initial concept for nanopore sequencing involved threading an individual single stranded DNA (ssDNA) molecule through the staphylococcal α-haemolysin (αHL) protein pore under an applied potential while recording modulations of the ionic current passing through the pore. Each base would be registered, in sequence, by a characteristic decrease in current amplitude[1–4]. It has been demonstrated that an individual nucleobase can be identified on a static strand in a nanopore[5]; however, the rate of DNA translocation under a potential is too high for the necessary current resolution in a moving strand unless the bases are subjected to chemical modification with bulky groups[6]. Recent work on strand sequencing has focused on using enzymes to reduce the speed of DNA translocation[7–9]. Mutations have also been added to the αHL pore to improve the rate of DNA capture from solution[10].

An alternative approach to single-molecule sequencing uses an exonuclease to liberate individual nucleoside monophosphates from a strand of DNA. The bases are identified in order of release. The original concept, which proved difficult to implement in practice, used a DNA strand substituted with fluorescently labelled nucleotides to be detected downstream by single-molecule fluorescence[11–15]. A means to detect unlabelled nucleotides at the single-molecule level would be a valuable alternative. Engineered nanopores have been used extensively to detect a variety of small molecules[16] through non-covalent interactions with the pore[17], covalent reaction with a cysteine residue[18] or adapter-mediated binding[19]. Although the principle of label-free nucleotide detection had been demonstrated by using a nanopore in combination with a non-covalent adapter[20], a number of difficult problems remained to be solved before nanopore detection could be established as a reliable readout method. First, the adapter was not covalently attached, leading to periods when no bases were observed. Second, the resolution of the bases was not sufficient for high-quality sequencing. Third, the operating conditions were not optimal for exonuclease activity. The activity of *Escherichia coli* exonuclease I is suitable for this approach, with the enzyme liberating 275 nucleotides per second at 37 °C (refs 15,21,22). The enzyme rate can be controlled with temperature for compatibility with the nanopore detection rate.

In the present work, we solve the key technical problems required for real-time, high-resolution nucleoside monophosphate detection. By covalent attachment of the adapter, continuous base detection has been achieved. The position of the adapter within the barrel of the pore has been optimized to improve the base discrimination at high data acquisition rates. Base discrimination is maintained under operating conditions compatible with exonuclease activity. Finally, we identify 5-methylcytosine, the 'fifth base', in the presence of G, A, T and C, which will be useful in epigenetics for the investigation of methylation patterns. These advances represent the realization of a complete nanopore base detector, which, when combined with a compatible exonuclease

[1]Oxford Nanopore Technologies Ltd, Begbroke Science Park, Sandy Lane, Oxford OX5 1PF, UK, [2]Department of Chemistry, University of Oxford, Chemistry Research Laboratory, Mansfield Road, Oxford OX1 3TA, UK; *e-mail: hagan.bayley@chem.ox.ac.uk

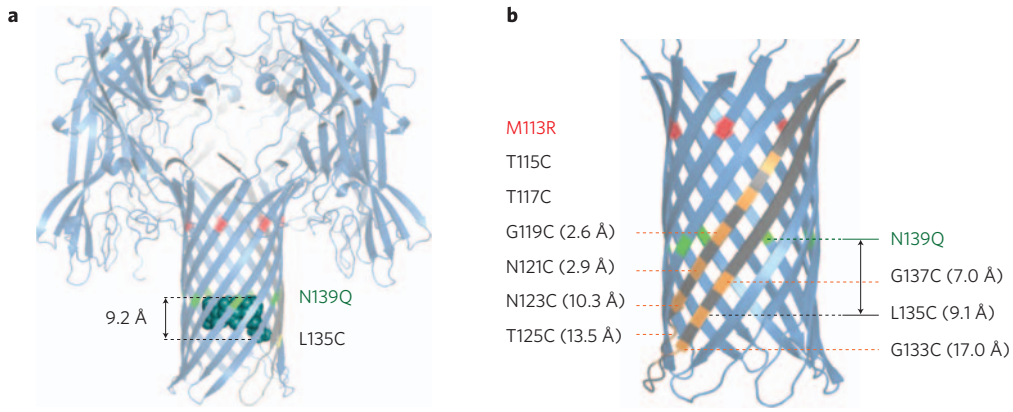

Figure 1 | Structures of haemolysin mutants. a, Structure of the WT-$(M113R/N139Q)_6(M113R/N139Q/L135C)_1$ mutant (cartoon view) showing the cyclodextrin covalently attached at position 135 (space filling model), the glutamines at residue 139, and the vertical distance between the secondary hydroxyls and the disulphide bond of the cyclodextrin. **b**, Close-up of the β barrel showing the arginines at position 113 and the location of the cysteines in the mutants tested in this study. The C_α–C_α distance between N139Q and each of the mutated residues is shown in brackets. Mutants are described using standard single-letter amino-acid codes (M113R denotes Met-113 replaced by Arg, N139Q denotes Asn-139 replaced by Gln and L135C denotes Lys-135 replaced by Cys).

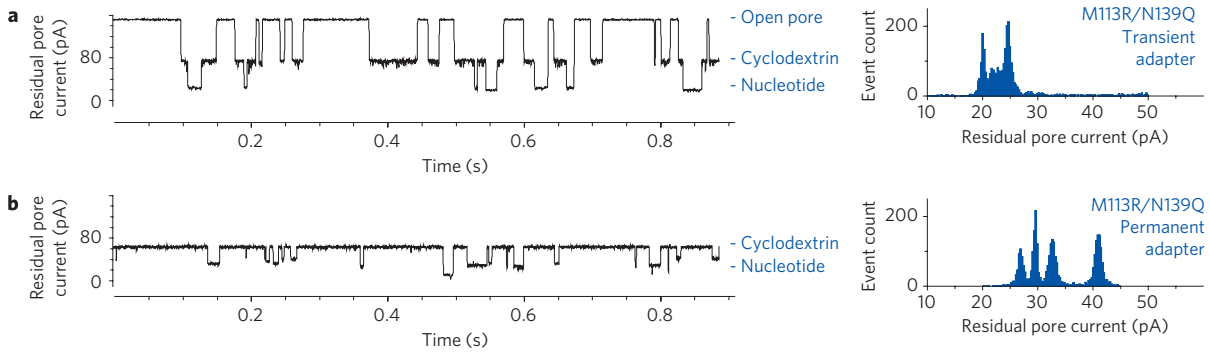

Figure 2 | Single-channel recordings comparing permanent and transient adapters. a, The WT-$(M113R/N139Q)_7$ αHL pore showing transient adapter binding (40 μM am_7βCD) and nucleotide detection. Histogram of the residual current of nucleotide binding events. **b**, Corresponding data for the WT-$(M113R/N139Q)_6(M113R/N139Q/L135C)_1$ mutant with a covalently attached am_6amPDP_1βCD, allowing continuous nucleotide detection and enhanced nucleotide discrimination (see histogram). The traces were recorded in 800 mM KCl, 25 mM Tris HCl, pH 7.5, at +160 mV in the presence of 10 μM dGMP, 10 μM dTMP, 10 μM dAMP and 10 μM dCMP.

DNA processing system, will provide the basis of a complete nanopore sequencer.

Results and discussion

Mutagenesis and adapter attachment. Our previous work demonstrated that discrimination of 2′-deoxynucleoside 5′-monophosphates (dNMPs) is possible in principle with an engineered αHL nanopore[20], (M113R), equipped with a non-covalent adapter. However, the discrimination of nucleotides could only be achieved by examining the longer events (>4 ms), using low data acquisition rates of 1 kHz. At this level of acquisition, only 67% of nucleotides could be observed, compared to 98% at 20,000 Hz (assuming an exponential distribution of dwell times with a mean dwell time of 10 ms and a requirement of four data points to assign a nucleotide binding event). In order for the αHL nanopore to be suitable for DNA sequencing, the system needed to be improved to permit the identification of nearly every nucleotide with high accuracy.

To improve nucleotide discrimination at high acquisition rates, we explored variations in the αHL polypeptide sequence. We examined the role of the protein background in $(M113R)_7$ by comparing the RL2 background, used in previous studies[20], to the wild type (WT). When the RL2 background was used, the adapter-mediated base discrimination was similar to the published work. However, when a WT sequence was used, the four nucleotides could not be resolved (see Supplementary Data). As reported elsewhere, the RL2 background contains five mutations with respect to the WT sequence[23]. We proposed that the N139Q mutation, present in the RL2 background, was crucial for nucleotide discrimination. To test this hypothesis, WT-$(M113R/N139Q)_7$ was produced and showed almost identical behaviour to RL2-$(M113R)_7$, confirming that the N139Q mutation is involved in base discrimination (see Supplementary Results).

The N139Q mutation is located near the middle of the β barrel, ~1.7 nm from the M113R mutation (Fig. 1)[24], making it unlikely that am_7βCD interacts with both mutations simultaneously. In previous work, mutants of αHL were created to bind different cyclodextrins at various positions within the β barrel, including the N139Q position[25]. Given that am_7βCD carries a net positive charge and the arginines at residue 113 are also positively charged, it seems unlikely that am_7βCD resides near the top of the barrel. We reasoned that am_7βCD is instead located near the N139Q mutation, and is perhaps stabilized though hydrogen bonding to the glutamine residues. This was confirmed by examining constructs lacking the N139Q mutation (see Supplementary Results).

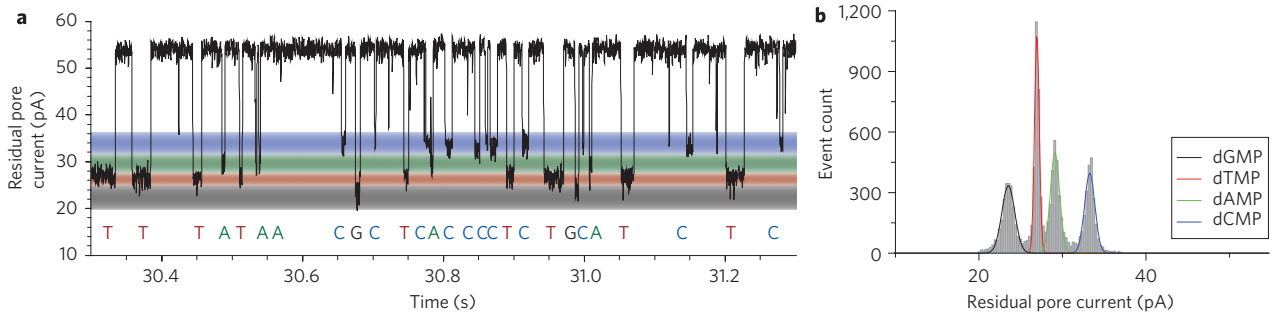

Figure 3 | Nucleotide event distributions with the permanent adapter. a, Single-channel recording from the WT-(M113R/N139Q)$_6$(M113R/N139Q/L135C)$_1$-am$_6$amDP$_1$βCD pore showing dGMP, dTMP, dAMP and dCMP discrimination, with coloured bands (three standard deviations from the centre of the individual Gaussian fits) added to represent the residual current distribution for each nucleotide. **b**, Corresponding residual current histogram of nucleotide binding events, including Gaussian fits. Data acquired in 400 mM KCl, 25 mM Tris HCl, pH 7.5, at +180 mV in the presence of 10 μM dGMP, 10 μM dTMP, 10 μM dAMP and 10 μM dCMP.

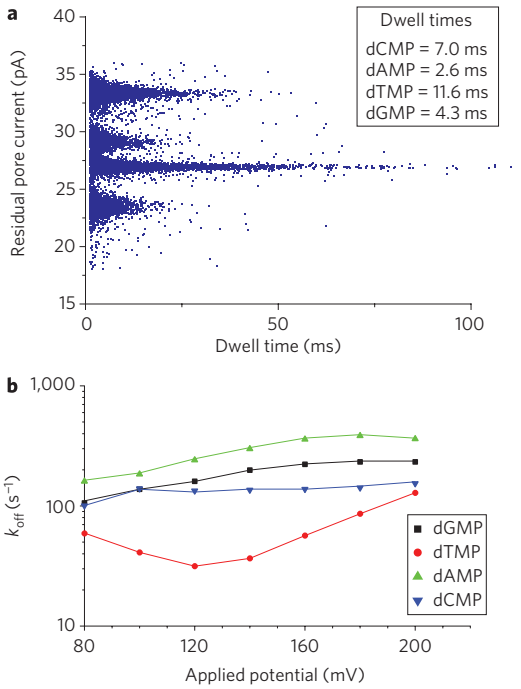

Figure 4 | Nucleotide dwell times and kinetics with a permanent adapter. a, Scatter plot showing residual currents and dwell times for dNMP binding events to the WT-(M113R/N139Q)$_6$(M113R/N139Q/L135C)$_1$-am$_6$amDP$_1$βCD pore. Dwell times were measured for each of the four standard nucleotides at 400 mM KCl, +180 mV. **b**, Variation of k_{off} with applied potential for each nucleotide. Lines have been added to aid the eye.

For sequencing applications, a permanent adapter molecule is required for continuous base detection. Recent work has demonstrated that a cyclodextrin can be attached to the αHL nanopore to permit the continuous stochastic detection of various analytes[26]. However, these constructs were not able to discriminate between nucleoside monophosphates, for which a non-covalent aminocyclodextrin had been used[20]. We therefore made heptakis(6-deoxy-6-amino)-6-*N*-mono(2-pyridyl)dithiopropanoyl-β-cyclodextrin (am$_6$amPDP$_1$βCD), which contains the primary amino groups required for base detection and a reactive linker for covalent attachment to a cysteine residue on the nanopore (see Supplementary Fig. S1). To test the hypothesis that the cyclodextrin must be attached near the N139Q position, a series of mutants (XnnnC) were made with cysteines at various positions in the β barrel (G119C, N121C, N123C, T125C, G133C, L135C, G137C). The mutants were assembled as 6 : 1 heteroheptamers of the form WT-(M113R/N139Q)$_6$(M113R/N139Q/XnnnC)$_1$ to ensure that only one cysteine was present in each nanopore and that all subunits had a WT background.

The cysteine-containing nanopores were screened for discrimination of the four standard dNMPs after attachment of the am$_6$amPDP$_1$βCD adapter. Cyclodextrin attachment could not be achieved with the cysteines at positions 125 or 133. Nucleotides were detected after attachment of the adapter to a cysteine at positions 119, 121, 123, 135 or 137, but the discrimination varied considerably between these positions. The best separation between nucleotides was observed with the adapter attached to the L135C residue, which gave a significant improvement in nucleotide discrimination over the non-covalent adapter (Fig. 2). With a permanent adapter, continuous detection of nucleotides was achieved with a single nanopore.

Nucleotide discrimination and translocation. The WT-(M113R/N139Q)$_6$(M113R/N139Q/L135C)$_1$-am$_6$amDP$_1$βCD construct was evaluated under a range of physical conditions. The discrimination of dGMP, dTMP, dAMP and dCMP was tested with concentrations of KCl from 350 to 1,000 mM, over a pH range of 6.0 to 8.5, and at temperatures from 15 to 40 °C. The applied potential was also varied to provide fine tuning of the nucleotide separation. A combination of conditions was sought that would be compatible with exonuclease activity while maintaining base discrimination. If the applied potential was increased to +180 mV, the salt concentration could be as low as 400 mM KCl, without adversely affecting the separation of the dNMP current levels (Fig. 3). Experiments using individual nucleotides and mixtures of nucleotides confirmed that each base has a characteristic peak position in the current histogram and identified the order of decreasing pore blockage by the dNMP to be dGMP, dTMP, dAMP and dCMP (data not shown).

The dNMP distributions were fitted to Gaussians (see Supplementary Methods), and the areas of peak overlap were determined to give confidence values for base identity. The percentages of binding events that could be assigned to each base with a confidence approaching 100% at a high salt concentration (800 mM) were 99.9, 99.7, 99.8 and 99.99% for dGMP, dTMP, dAMP and dCMP, respectively (see Supplementary Fig. S2). For a low salt concentration (400 mM) the percentages were 99.9, 97.9, 98.0 and 99.99% for dGMP, dTMP, dAMP and dCMP, respectively. It should be noted that when there is ambiguity in a base call, the alternative identities of the base are known. For example, if the conditions in Fig. 3 are used, an event at ~28 pA could either be dTMP or dAMP, but is extremely unlikely to be dGMP or dCMP.

In addition to the residual pore current, the different dNMPs also show variations in mean dwell time within the adapter (τ_{off}). The clearest example of this is dTMP, which displays binding events

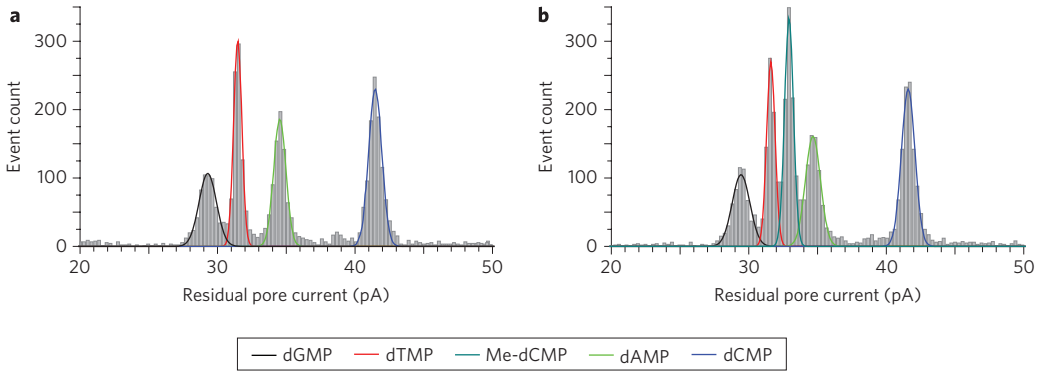

Figure 5 | Detection of methyl-dCMP. a, Residual current histograms for the WT-$(M113R/N139Q)_6(M113R/N139Q/L135C)_1$-$am_6amDP_1\beta CD$ pore in the presence of a mixture of dGMP, dTMP, dAMP and dCMP. **b**, Histogram from the same nanopore following the addition of Me-dCMP. Data were acquired in 400 mM KCl, 25 mM Tris HCl, pH 7.5, at +200 mV after reaction with 5 μM $am_6amPDP_1\beta CD$, and in the presence of 5 μM dGMP, 5 μM dTMP, 5 μM dAMP, 5 μM dCMP and 5 μM Me-dCMP.

with mean dwell times two to three times longer than the other nucleotides (Fig. 4). Long binding events are desirable for accurate base calling as they allow a better estimate of the residual pore current. However, longer τ_{off} values decrease the overall rate of sequencing. The dwell time of an individual nucleotide also provides additional information for base calling. The long mean dwell time of dTMP may be due to the interaction of the methyl group at the 5′ position of the thymine base with the adapter. Accordingly, a long dwell-time is also seen with 5-methyl-2′-deoxycytidine 5′-monophosphate (Me-dCMP) which exhibited a τ_{off} of 15.5 ms, compared to 8.9 ms for dCMP under similar conditions (see below).

For DNA sequencing applications, it is preferable that a nucleotide exit the nanopore to the *trans* side of the bilayer to remove any possibility of the nucleotide being reread. The variation in the off rate constant, k_{off} $(1/\tau_{off})$, of a dNMP with the applied potential can be used to determine whether the molecule is exiting the nanopore at the *cis* or *trans* side of the membrane[25,27]. There are two voltage regimes. In the first regime, at low potentials, the nucleotide returns to the *cis* chamber. A greater applied potential promotes binding of the charged dNMP to the adapter for a longer period of time, resulting in a decrease in k_{off} with increasing potential. In the second regime, at high potentials, the dNMP is 'pushed' through the cyclodextrin and translocated into the *trans* chamber. In this case, an increase in the applied potential reduces the time that a dNMP resides in the adapter, causing k_{off} to increase at higher potentials. In some cases, both mechanisms may occur, leading to a minimum in k_{off} versus the applied potential[25,27]. During translocation, the dNMP and cyclodextrin most probably undergo conformational changes to facilitate the movement of the nucleotide through the pore.

The voltage dependence of k_{off} was examined for dGMP, dTMP, dAMP and dCMP (Fig. 4b). An increase in k_{off} at higher potentials was seen for dGMP and dAMP, indicating that these nucleotides cross the membrane to the *trans* chamber. For dCMP, k_{off} appeared relatively invariant, suggesting that k_{off} is not limited by the electrical potential. A clearer picture was seen for dTMP which exhibited a minimum in k_{off} versus applied potential, suggesting that at low potentials dTMP returns to the *cis* chamber (<120 mV), and at higher potentials (>120 mV) dTMP exits to the *trans* chamber. The optimal base discrimination was recorded at +180 mV, where it is expected that a very high proportion of dNMPs translocate through the nanopore.

DNA methylation, manifested as 5-methylcytosine (m^5C), is a critical factor in epigenetics. Up to 5% of cytosines in mammalian genomes are methylated, which results in long-term transcriptional silencing[28–30]. The current method for mapping methylation sites involves treatment of the DNA with bisulphite and then base, which converts C to U while sparing m^5C, followed by sequence analysis[31]. We set out to determine the presence of 5-methyl-2′-deoxycytidine 5′-monophosphate (Me-dCMP) directly, by using our preferred protein construct WT-$(M113R/N139Q)_6(M113R/N139Q/L135C)_1$-$am_6amDP_1\beta CD$. Me-dCMP alone gave a distinct peak of binding events in an amplitude histogram, producing a current block larger than that of dCMP and similar to dTMP (data not shown). At a high transmembrane potential, in the presence of dGMP, dAMP, dTMP and dCMP, all five nucleotides could be distinguished (Fig. 5).

Nucleotide detection from ssDNA. To use αHL for exonuclease sequencing, enzymatic cleavage of nucleotides from a DNA strand in close proximity to the mouth of the nanopore is required. Therefore, the physical conditions for base discrimination must be compatible with exonuclease activity. To achieve this, we used asymmetric salt conditions. The salt concentration on the *cis* side of the bilayer was reduced to 200 mM KCl to promote enzyme activity, while the salt concentration in the *trans* chamber was increased to 500 mM KCl to maintain a high ionic conductance through the nanopore and thereby maintain base discrimination. Good dNMP separation was achieved; the percentages of single-molecule events that could be unambiguously assigned to a particular base were 99.4, 90.3, 90.9 and 99.99% for dGMP, dTMP, dAMP and dCMP, respectively (200/500 mM KCl, 25 mM Tris HCl, pH 7.5, +180 mV, room temperature; see Supplementary Fig. 2).

The protein nanopore construct was used to identify dNMPs produced in solution from ssDNA and ExoI from *E. coli* (Fig. 6). Two different DNAs were used: one contained only G, A and C bases (85mer), and a second only G, T and C bases (76mer). A solution containing ExoI and a ssDNA template was added to the *cis* chamber. The dNMPs produced by the exonuclease were observed as binding events and plotted in a residual current histogram. As expected, the ssDNA lacking the T base showed three peaks corresponding to dGMP, dAMP and dCMP (Fig. 6b). The second DNA, lacking the A base, produced a different distribution of three peaks corresponding to dGMP, dTMP and dCMP (Fig. 6c). This experiment demonstrates the ability of the nanopore to detect nucleotides released from a DNA strand under physical conditions compatible with exonuclease activity.

Conclusions

We have engineered a nanopore with a covalently attached adapter that is capable of continuous nucleoside monophosphate detection without the need for labelling. The nanopore shows accurate discrimination of the four standard dNMPs, reading raw bases with over 99% confidence under optimal operating conditions.

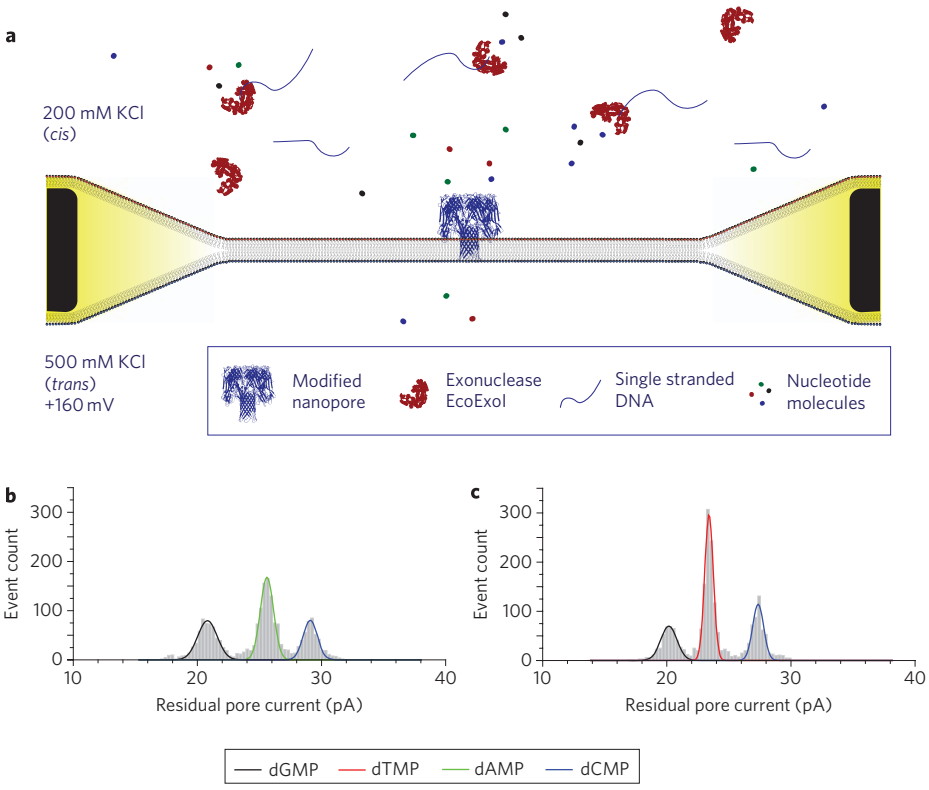

Figure 6 | Detection of nucleotides cleaved from ssDNA by exonuclease I. a, Experimental setup for the exonuclease experiments. Nucleotides liberated by the enzyme are detected by the WT-$(M113R/N139Q)_6(M113R/N139Q/L135C)_1$-$am_6amDP_1\beta CD$ pore. **b**, Residual pore current histogram of nucleotide binding events for a ssDNA containing G, A and C. **c**, Residual pore current histogram of nucleotide binding events for a ssDNA containing G, T and C. Experimental conditions: 200/500 mM KCl, 25 mM Tris HCl, pH 7.5, 1 mM $MgCl_2$, 5 μM ssDNA and 80 U exonuclease I, at +180 mV and room temperature.

The pore can also discriminate between other nucleotides, including methylated dCMP. It has also been demonstrated that the detected dNMPs have a high probability of exiting the nanopore into the *trans* chamber, which will prevent duplicate detection events. In addition, our detector can operate under the low salt conditions (200 mM KCl) compatible with exonuclease activity, while maintaining base detection with high confidence. It should be noted that while most enzymes prefer low salt conditions, exonuclease I has been shown to work in high salt conditions[8], which may allow data acquisition in high salt for optimal base discrimination. Enzyme engineering could also be performed to improve salt tolerance and control other properties such as processivity, digestion rate and stability.

To integrate the base-detecting nanopore into an exonuclease sequencing system, a number of requirements must be met. The exonuclease must be in close proximity to the nanopore sensor; this could be achieved by chemical attachment or a genetic fusion of the nanopore and the exonuclease. Additionally, the location of the exonuclease active site must be manipulated to align the trajectory of the released dNMP with the nanopore entrance. Furthermore, it may be necessary to vary the rate of the enzyme, either by mutation or changes to physical conditions such as temperature, to ensure that a nucleotide is recorded before the next base enters the nanopore.

If these technical challenges can be overcome, a nanopore DNA sequencing system would offer many advantages. The sequencing of single molecules removes the need for amplification of the target DNA, lowering the time and cost of sample preparation and avoiding amplification errors. Single-molecule detection also allows the direct observation of modified nucleotides such as methylated dCMP, which would significantly benefit the field of epigenetics. The nanopore approach will not require expensive fluorescent labels or imaging technologies, further reducing the overall cost and workflow time. In addition, certain exonucleases have high processivity, which will allow long read lengths that will simplify sequence alignment and reassembly and offer additional benefits such as direct information on haplotypes. Furthermore, the simple signal from the nanopore requires little processing, allowing base assignment in real time. Therefore, in contrast to optical sequencing techniques, the requirements for data processing and storage in nanopore sequencing are low. Additionally, the electrical nature of the signal means that the approach is scaleable using established technology and cheap fabrication techniques. These challenges are currently being addressed, with a view to establishing a new generation of sequencing technology.

Methods

Chemicals and enzymes. Reagents were obtained from the following sources and used without further purification: 1,2-diphytanoyl-*sn*-glycero-3-phosphocholine (Avanti Polar Lipids); pentane (Sigma-Aldrich); hexadecane (99 + %, Sigma-Aldrich); heptakis(6-deoxy-6-amino)-β-cyclodextrin.7HCl ($am_7\beta CD$, >99%, Cyclolab); 2′-deoxyguanosine 5′-monophosphate sodium salt (99%, Acros); 2′-deoxycytidine 5′-monophosphate disodium salt (>95%, Fluka); 2′-deoxythymidine 5′-monophosphate disodium salt (>97%, Fluka); 2′-deoxyadenosine 5′-monophosphate disodium salt (>95%, Fluka); 5-methyl-2′-deoxycytidine 5′-monophosphate (USB Europe); Trizma base (99.9%, Sigma-Aldrich); concentrated HCl (analytical reagent grade, Fisher Scientific); potassium chloride (99%, Sigma-Aldrich); 3-(2-pyridyldithio)propionic acid N-hydroxysuccinimide (SPDP, Sigma-Aldrich). Exonuclease I (*E. coli*) was from New England Biolabs. The heptakis(6-deoxy-6-amino)-6-*N*-mono(2-pyridyl)dithiopropanoyl-β-cyclodextrin was synthesized in-house and then out-sourced to Chromatide. This compound was obtained with a final purity of 98% (see Supplementary Methods). Synthetic oligonucleotides with the following sequences were used (Sigma-Aldrich): 5′ GGCCGAAGAAGCCAACACGAAAGACGAAACCGCAACAGCCGAAAAAC GCCACAAAGCCAACAGGAACAGCAAGCCCGCCCAAGCG 3′ and 5′ GGCGG GCTTGCTGTTCCTGTTGGCTTTGTGGCGTTTTTCGGCTGTTGCGGTTTCG TCTTTCGTGTTGGCTTCTTCG 3′.

Cysteine mutants of α-haemolysin. The mutant αHL pores used in this study were expressed, assembled and purified as described previously[32]. To prepare the

heteroheptamers, a sequence encoding a C-terminal tail of eight aspartate residues was added to the genes of the cysteine mutants. The aspartate tail allowed the resolution of heptamers with different combinations of subunits by SDS-polyacrylamide gel electrophoresis[32]. Further details are contained in the Supplementary Methods. Mutants are described using standard single-letter amino-acid codes (for example, M113R denotes Met-113 replaced by Arg).

Unless otherwise stated, all proteins were used with a wild-type background. In one case, $(M113R)_7$, the pore was made in an RL2 background. RL2 is the product of a semisynthetic gene that was devised to permit cassette mutagenesis of the sequence encoding the transmembrane β barrel[23]. It contains five altered amino acids in the encoded polypeptide sequence. The K8A mutation is in the cap region of αHL; the others are in the β barrel. The G130S, V124L and I142L side chains face the hydrocarbon core of the bilayer, whereas the side chain of the N139Q mutation projects into the hydrophilic interior of the β barrel.

Single-channel recordings. Single-channel recordings were obtained using standard methods[20] with an apparatus designed and produced in-house. In short, a bilayer of 1,2-diphytanoyl-*sn*-glycero-3-phosphocholine (Avanti Polar Lipids) was formed on an aperture 60–150 μm in diameter in a Teflon film (25 μm thickness; Goodfellow) that divided a planar bilayer chamber into two compartments, *cis* and *trans*. Unless otherwise stated, αHL pores (0.2 μl, ~1 μg ml^{-1}) and dNMPs (20–40 μM) were added to the *cis* compartment, which was connected to ground. am_7βCD or am_6amPDP_1βCD was added to the *trans* compartment (20–40 μM), which was connected to the head-stage of the amplifier. Both compartments contained 1 ml of buffer, which unless stated otherwise, was 25 mM Tris HCl, 400 mM KCl, pH 7.5, at 22 ± 2 °C. Runs were typically acquired for 5 to 15 min. Bilayers that lasted hours were routinely obtained, even at high applied potentials.

During the reaction of am_6amPDP_1βCD with a single cysteine-containing pore, non-covalent binding events were first observed as previously described[20], before reaction of the cyclodextrin with the cysteine of the nanopore resulted in a permanent reduction of the open pore current of approximately 65% (ref. 26). The open pore current could be restored by the addition of 1,4-dithiothreitol (DTT, 1 mM, Sigma-Aldrich) to cleave the disulphide bond[26].

Data acquisition and analysis. Single-channel recordings were made with a patch clamp amplifier (Axopatch 200B; Molecular Devices). The signal was low-pass filtered with a built-in four-pole Bessel filter at 10 kHz, and sampled at 20 kHz by a PC equipped with a Digidata 1440A A/D converter (Molecular Devices) running Clampex 10 software (Molecular Devices). Single-channel traces shown in the figures were filtered at 2 kHz for display purposes. All histograms were obtained from data filtered at 10 kHz. Event detection was achieved by using in-house software (see Supplementary Methods).

Received 19 September 2008; accepted 14 January 2009; published online 22 February 2009

References

1. Bayley, H. Sequencing single molecules of DNA. *Curr. Opin. Chem. Biol.* **10,** 628–637 (2006).
2. Branton, D. *et al.* The potential and challenges of nanopore sequencing. *Nature Biotechnol.* **26,** 1146–1153 (2008).
3. Kasianowicz, J. J., Brandin, E., Branton, D. & Deamer, D. W. Characterization of individual polynucleotide molecules using a membrane channel. *Proc. Natl Acad. Sci. USA* **93,** 13770–13773 (1996).
4. Deamer, D. W. & Branton, D. Characterization of nucleic acids by nanopore analysis. *Acc. Chem. Res.* **35,** 817–825 (2002).
5. Ashkenasy, N., Sanchez-Quesada, J., Bayley, H. & Ghadiri, M. R. Recognizing a single base in an individual DNA strand: a step toward DNA sequencing in nanopores. *Angew Chem. Int. Ed. Engl.* **44,** 1401–1404 (2005).
6. Mitchell, N. & Howorka, S. Chemical tags facilitate the sensing of individual DNA strands with nanopores. *Angew Chem. Int. Ed. Engl.* **47,** 5565–5568 (2008).
7. Astier, Y., Kainov, D. E., Bayley, H., Tuma, R. & Howorka, S. Stochastic detection of motor protein–RNA complexes by single-channel current recording. *Chem. Phys. Chem.* **8,** 2189–2194 (2007).
8. Hornblower, B. *et al.* Single-molecule analysis of DNA–protein complexes using nanopores. *Nature Methods* **4,** 315–317 (2007).
9. Cockroft, S. L., Chu, J., Amorin, M. & Ghadiri, M. R. A single-molecule nanopore device detects DNA polymerase activity with single-nucleotide resolution. *J. Am. Chem. Soc.* **130,** 818–820 (2008).
10. Maglia, G., Restrepo, M. R., Mikhailova, E. & Bayley, H. Enhanced translocation of single DNA molecules through alpha-hemolysin nanopores by manipulation of internal charge. *Proc. Natl Acad. Sci. USA* **105,** 19720–19725 (2008).
11. Jett, J. H. *et al.* High-speed DNA sequencing: an approach based upon fluorescence detection of single molecules. *J. Biomol. Struct. Dyn.* **7,** 301–309 (1989).
12. Sauer, M. *et al.* Single molecule DNA sequencing in submicrometer channels: state of the art and future prospects. *J. Biotechnol.* **86,** 181–201 (2001).
13. Stephan, J. *et al.* Towards a general procedure for sequencing single DNA molecules. *J. Biotechnol.* **86,** 255–267 (2001).
14. Werner, J. H. *et al.* Progress towards single-molecule DNA sequencing: a one color demonstration. *J. Biotechnol.* **102,** 1–14 (2003).
15. Werner, J. H., Cai, H., Keller, R. A. & Goodwin, P. M. Exonuclease I hydrolyzes DNA with a distribution of rates. *Biophys. J.* **88,** 1403–1412 (2005).
16. Bayley, H. & Cremer, P. S. Stochastic sensors inspired by biology. *Nature* **413,** 226–230 (2001).
17. Cheley, S., Gu, L. Q. & Bayley, H. Stochastic sensing of nanomolar inositol 1,4,5-trisphosphate with an engineered pore. *Chem. Biol.* **9,** 829–838 (2002).
18. Shin, S. H., Luchian, T., Cheley, S., Braha, O. & Bayley, H. Kinetics of a reversible covalent-bond-forming reaction observed at the single-molecule level. *Angew Chem. Int. Ed. Engl.* **41,** 3707–3709, 3523 (2002).
19. Gu, L. Q., Braha, O., Conlan, S., Cheley, S. & Bayley, H. Stochastic sensing of organic analytes by a pore-forming protein containing a molecular adapter. *Nature* **398,** 686–690 (1999).
20. Astier, Y., Braha, O. & Bayley, H. Toward single molecule DNA sequencing: direct identification of ribonucleoside and deoxyribonucleoside 5′-monophosphates by using an engineered protein nanopore equipped with a molecular adapter. *J. Am. Chem. Soc.* **128,** 1705–1710 (2006).
21. Brody, R. S., Doherty, K. G. & Zimmerman, P. D. Processivity and kinetics of the reaction of exonuclease I from *Escherichia coli* with polydeoxyribonucleotides. *J. Biol. Chem.* **261,** 7136–7143 (1986).
22. Enderlein, J. Nucleotide specificity versus complex heterogeneity in exonuclease activity measurements. *Biophys. J.* **92,** 1556–1558 (2007).
23. Cheley, S., Braha, O., Lu, X., Conlan, S. & Bayley, H. A functional protein pore with a 'retro' transmembrane domain. *Protein Sci.* **8,** 1257–1267 (1999).
24. Song, L. *et al.* Structure of staphylococcal alpha-hemolysin, a heptameric transmembrane pore. *Science* **274,** 1859–1866 (1996).
25. Gu, L. Q., Cheley, S. & Bayley, H. Capture of a single molecule in a nanocavity. *Science* **291,** 636–640 (2001).
26. Wu, H. C., Astier, Y., Maglia, G., Mikhailova, E. & Bayley, H. Protein nanopores with covalently attached molecular adapters. *J. Am. Chem. Soc.* **129,** 16142–16148 (2007).
27. Sanchez-Quesada, J., Ghadiri, M. R., Bayley, H. & Braha, O. Cyclic peptides as molecular adapters for a pore-forming protein. *J. Am. Chem. Soc.* **122,** 11757–11766 (2000).
28. Brena, R. M., Huang, T. H. & Plass, C. Toward a human epigenome. *Nature Genet.* **38,** 1359–1360 (2006).
29. Schaefer, C. B., Ooi, S. K., Bestor, T. H. & Bourc'his, D. Epigenetic decisions in mammalian germ cells. *Science* **316,** 398–399 (2007).
30. Ooi, S. K. & Bestor, T. H. The colorful history of active DNA demethylation. *Cell* **133,** 1145–1148 (2008).
31. Hayatsu, H. The bisulfite genomic sequencing used in the analysis of epigenetic states, a technique in the emerging environmental genotoxicology research. *Mutat. Res.* **659,** 77–82 (2008).
32. Howorka, S., Cheley, S. & Bayley, H. Sequence-specific detection of individual DNA strands using engineered nanopores. *Nature Biotechnol.* **19,** 636–639 (2001).

Acknowledgements

The authors would like to thank O. Braha, S. Cheley, T. Reid and G. Sanghera for valuable discussion, S. Lewis for cyclodextrin preparation and analysis, M. Crawford, G. Hall and L. Woodward for nanopore testing, J. Kilgour and J. White for protein production, Z. McDougall for assistance with the manuscript, and all the staff at ONT. This work was supported by grants from the NIH, the European Commission's seventh Framework Programme (FP7) READNA Consortium and Oxford Nanopore Technologies. H.B. is the holder of a Royal Society Wolfson Research Merit Award.

Author contributions

J.C. and H.B. conceived the experiments and wrote the manuscript. J.C. designed the mutant constructs and analysed the data. H.W. designed and synthesized the cyclodextrin. L.J. engineered the proteins. A.P. performed the single-channel recordings. S.R. wrote data analysis algorithms and software.

Additional information

Supplementary Information accompanies this paper at www.nature.com/naturenanotechnology. Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/. Correspondence and requests for materials should be addressed to H.B.