J|A|C|S
ARTICLES
Published on Web 00/00/0000

Toward Single Molecule DNA Sequencing: Direct Identification of Ribonucleoside and Deoxyribonucleoside 5′-Monophosphates by Using an Engineered Protein Nanopore Equipped with a Molecular Adapter

Yann Astier, Orit Braha, and Hagan Bayley*

Contribution from the Department of Chemistry, University of Oxford, Chemistry Research Laboratory, Mansfield Road, OX1 3TA, Oxford, UK

Received October 19, 2005; E-mail: hagan.bayley@chem.ox.ac.uk.

Abstract: Individual nucleic acid molecules might be sequenced by the identification of nucleoside 5′-monophosphates as they are released by processive exonucleases. Here, we show that single molecule detection with a modified protein nanopore can be used to identify ribonucleoside and 2′-deoxyribonucleoside 5′-monophosphates, thereby taking a step along this path. Distinct levels of current block are observed for each of the four members of a set of nucleoside 5′-monophosphates when the molecules bind within a mutant α-hemolysin pore, $(M113R)_7$, equipped with the molecular adapter heptakis-(6-deoxy-6-amino)-β-cyclodextrin. While our results compare favorably with alternative approaches, further work will be required to improve the accuracy of identification of the nucleic acid bases, to feed each released nucleotide into the pore, and to ensure that every nucleotide is captured by the adapter.

Introduction

Cheaper and faster ways to sequence DNA would revolutionize the use of genetic information and the nature of biological and biomedical research. Toward this end, the U.S. National Institutes of Health have set goals for the low-cost sequencing and resequencing of entire human genomes by 2014: the "$1000 genome".[1] The current dideoxy chain termination method requires costly reagents such as fluorescent ddNTPs, dNTPs, oligonucleotide primers, and a DNA polymerase. The chain termination method also requires sophisticated equipment, nowadays a multichannel capillary electrophoresis system, which contributes significantly to the cost per base.[2] Various methods are being considered in attempts to reduce costs and increase the speed of sequencing, including capillary electrophoresis on microfabricated chips, sequencing by hybridization on microarrays, cyclic-array sequencing of amplified DNAs, and mass spectrometry.[2−5]

Concurrently, several interesting proposals for sequencing individual DNA molecules have been made, which would obviously save time and resources if they could be put into practice. For example, there would no longer be a need to amplify the target sequences, and information such as haplotype combinations of SNPs would be readily obtained. Cyclic array approaches using fluorescent base incorporation with polymerases can be adapted for single-molecule sequencing of molecules immobilized on surfaces[6] or in extremely small observation volumes.[7] In another approach, a processive exonuclease is used to digest a single DNA molecule, and the released nucleotides are identified one at a time.[8] In one manifestation, fluorescent nucleotides are first incorporated into the DNA and detected by single-molecule fluorescence upon release.[8−10] In practice, it has proved difficult to obtain full incorporation of the bulky nucleotides and to observe every one that is liberated by the nuclease.

In a different single-molecule approach, a DNA strand is threaded through a nanopore and its sequence identified from the fluctuations in ionic current.[11,12] While single-base recognition in a static DNA strand by the α-hemolysin (αHL) protein pore was achieved recently,[13] the sequencing concept has not worked so far because under an electrical potential DNA moves through the pore too rapidly for the identification of individual bases.[14] In the present work, we present the initial exploration of an approach that will combine detection with a nanopore and exonuclease sequencing. We show how individual unmodi-

(1) http://grants.nih.gov/grants/guide/rfa-files/RFA-HG-04-003.html.
(2) Shendure, J.; Mitra, R. D.; Varma, C.; Church, G. M. *Nat. Rev. Genet.* **2004**, *5*, 335−344.
(3) Edwards, J. R.; Ruparel, H.; Ju, J. *Mutat. Res.* **2005**, *573*, 3−12.
(4) Margulies, M.; et al. *Nature* **2005**, *437*, 376−380.
(5) Shendure, J.; Porreca, G. J.; Reppas, N. B.; Lin, X.; McCutcheon, J. P.; Rosenbaum, A. M.; Wang, M. D.; Zhang, K.; Mitra, R. D.; Church, G. M. *Science* **2005**, *309*, 1728−1732.
(6) Braslavsky, I.; Herbert, B.; Kartalov, E.; Quake, S. R. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 3960−3964.
(7) Levene, M. J.; Korlach, J.; Turner, S. W.; Foquet, M.; Craighead, H. G.; Webb, W. W. *Science* **2003**, *299*, 682−686.
(8) Jett, J. H.; Keller, R. A.; Martin, J. C.; Marrone, B. L.; Moyzis, R. K.; Ratliff, R. L.; Seitzinger, N. K.; Shera, E. B.; Stewart, C. C. *J. Biomol. Struct. Dyn.* **1989**, *7*, 301−309.
(9) Sauer, M. A., B.; Ankenbauer, W.; Foldes-Papp, Z.; Gobel, F.; Han, K.-T.; Rigler, R.; Schulz, A.; Wolfrum, J.; Zander, C. *J. Biotechnol.* **2001**, *86*, 181−201.
(10) Werner, J.; Cai, H.; Jett, J. H.; Reha-Krantz, L.; Keller, R. A.; Goodwin, P. M. *J. Biotechnol.* **2002**, *102*, 1−14.
(11) Kasianowicz, J. J.; Brandin, E.; Branton, D.; Deamer, D. W. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 13770−13773.
(12) Deamer, D. W.; Akeson, M. *Trends Biotechnol.* **2000**, *18*, 147−151.
(13) Ashkenasy, N.; Sanchez-Quesada, J.; Bayley, H.; Ghadiri, M. R. *Angew. Chem., Int. Ed.* **2005**, *44*, 1401−1404.
(14) Meller, A.; Nivon, L.; Branton, D. *Phys. Rev. Lett.* **2001**, *86*, 3435−3438.

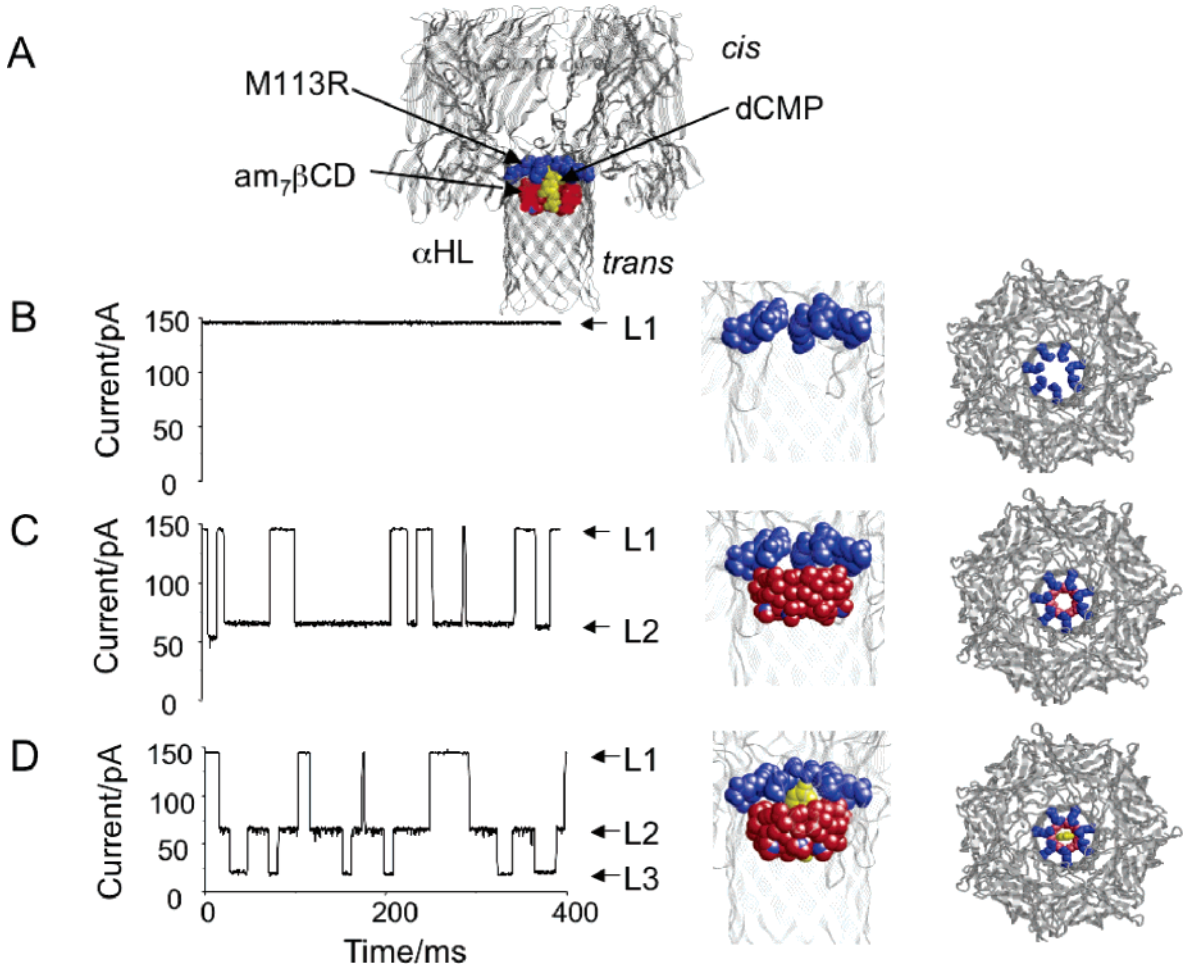

Figure 1. Interaction of the αHL pore with heptakis-(6-deoxy-6-amino)-β-cyclodextrin (am$_7\beta$CD) and dCMP. (A) Model of the heptameric αHL pore (7AHL), in which Met-113 has been substituted with Arg (blue). A model of am$_7\beta$CD in cross-section (red, nitrogen atoms in blue) generated in ChemDraw Ultra has been positioned manually at van der Waals distances from the Arg side chains, which block the passage of the cyclodextrin when it enters the pore from the trans side. The dCMP molecule (yellow) is positioned so that the phosphate group interacts with the protonated amines of am$_7\beta$CD and the cytosine ring interacts with the guanidinium groups of the Arg side chains. (B) Current trace from a single (M113R)$_7$ pore at +130 mV. L1 identifies the current flowing through the unoccupied protein nanopore, which is shown as a model on the right. (C) Current trace after the addition of 40 μM am$_7\beta$CD to the trans chamber. L2 indicates the current level observed when am$_7\beta$CD is bound inside the nanopore. (D) Current trace after the addition of 5 μM dCMP to the cis chamber. L3 shows the current level that is observed when dCMP binds to the (M113R)$_7$·am$_7\beta$CD complex.

fied nucleoside monophosphates can be identified by using an engineered α-hemolysin pore containing a molecular adapter as the detector, a variation of stochastic sensing.[15]

Stochastic sensing is achieved by placing a pore of nanometer dimensions in an insulating lipid bilayer membrane that separates two chambers of aqueous electrolyte and measuring the ionic current that passes through the pore under an applied potential. When an analyte interacts with a binding site within the pore, a change in the ionic current is detected.[15,16] The extent of current block and the mean duration of the binding events help to reveal the identity of an analyte, while the frequency of occurrence of the events reveals the analyte concentration. Various binding sites for analytes can be created within the pore by protein engineering.[15] Of significant value for small molecule detection is the use of host molecules as molecular adapters and carriers.[17,18] Adapters become lodged within the pore, and, while they are resident, analytes can associate and dissociate several times. In the case of carriers, the analytes exchange more slowly than the residence time of the host molecule within the pore. In the present work, we combine a mutant αHL pore with a positively charged cyclodextrin adapter (Figure 1) and show that this tailored nanopore can distinguish between all four 2′-deoxyribonucleoside 5′-monophosphates (dNMP) and all four ribonucleoside 5′-monophosphates (rNMP) on the basis of the extent of current block.

Materials and Methods

αHL Pore. The α-hemolysin mutant pore (M113R)$_7$ (RL2 background) was expressed, assembled, and purified as previously described.[19]

Chemicals. Reagents were obtained as follows: 1,2-diphytanoyl-*sn*-glycero-3-phosphocholine (Avanti Polar Lipids); pentane (JT Baker); hexadecane (99+%, Sigma-Aldrich); heptakis(6-deoxy-6-amino)-β-cyclodextrin·7HCl (am$_7\beta$CD, >99%, CYCLOLAB, Budapest, Hungary); 2′-deoxyguanosine 5′-monophosphate sodium salt (99%, Acros); 2′-deoxycytosine 5′-monophosphate disodium salt (>95%, Fluka); 2′-

(15) Bayley, H.; Cremer, P. S. *Nature* **2001**, *413*, 226−230.
(16) Braha, O.; Walker, B.; Cheley, S.; Kasianowicz, J. J.; Song, L.; Gouaux, J. E.; Bayley, H. *Chem. Biol.* **1997**, *4*, 497−505.
(17) Braha, O.; Webb, J.; Gu, L.-Q.; Kimoon, K.; Bayley, H. *Chem. Phys. Chem.* **2005**, *6*, 889−892.
(18) Gu, L.-Q.; Braha, O.; Conlan, S.; Cheley, S.; Bayley, H. *Nature* **1999**, *398*, 686−690.
(19) Cheley, S.; Gu, L.-Q.; Bayley, H. *Chem. Biol.* **2002**, *9*, 829−838.

deoxythymidine 5′-monophosphate disodium salt (>97%, Fluka); 2′-deoxyadenosine 5′-monophosphate disodium salt (>95%, Fluka); uridine 5′-monophosphate disodium salt (99%, Fluka); cytosine 5′-monophosphate (free acid >98%, Fluka); adenosine 5′-monophosphate (free acid 99%, Acros); guanosine 5′-monophosphate disodium salt (97%, Acros); Trizma base (99.9%, Sigma-Aldrich); concentrated HCl (analytical reagent grade, Fisher Scientific); potassium chloride (99%, Sigma-Aldrich); sodium chloride (99.9%, Sigma-Aldrich); potassium bromide (99.5%, Fluka); and cesium chloride (99%, Fluka).

Single Channel Recording. A bilayer of 1,2-diphytanoyl-*sn*-glycero-3-phosphocholine (Avanti Polar Lipids) was formed on an aperture 100−150 μm in diameter in a polycarbonate film (20-μm thickness from Goodfellow, Malvern, PA) that divided a planar bilayer chamber into two compartments, cis and trans. Both compartments contained 1.5 mL of buffer. Unless otherwise stated, both $(M113R)_7$ and rNMP or dNMP were added to the cis compartment, which was connected to ground. The am$_7\beta$CD was added to the trans compartment, which was connected to the head-stage of the amplifier. Experiments with dNMP were carried out at +130 mV, and experiments with rNMP experiments at +110 mV, in 25 mM Tris·HCl, 1 M KCl, pH 8.0, at 22.5 ± 2 °C, unless otherwise stated. Fresh aliquots of nucleotide stock solutions were used each day. Currents were recorded with a patch clamp amplifier (Axopatch 200B; Axon instruments, Foster City, CA), low pass filtered with a built-in 4-pole Bessel filter at 10 kHz, and sampled at 20 kHz by a computer equipped with a Digidata 1200 A/D converter (Axon instruments).

Data Analysis. Current traces were analyzed with pClamp 9.0 software (Axon Instuments). Events were detected using the Event Detection feature in pClamp 9.0, from which amplitude and dwell time histograms were constructed. Origin (Microcal, Northampton, MA) was used for curve fitting and graph presentation. Integrations were run on a personal calculator.

Molecular Models. The pdb file of the heptameric α-hemolysin pore (PDB file 7AHL) was modified in Swiss-PdbViewer (version 3.51) to create the $(M113R)_7$ mutations. The model of $(M113R)_7$ is displayed in RasWin Molecular Graphics version 2.7.1 as gray strands, with a spacefill representation of the arginine residues at position 113 in blue. The am$_7\beta$CD structure was constructed in ChemDraw Ultra (CambridgeSoft version 9.0.1) and imported into RasWin. The am$_7\beta$CD molecule was positioned immediately underneath the ring of arginines in $(M113R)_7$ with the primary amino groups (blue) facing downward, toward the trans entrance of the β barrel. A spacefill representation of dCMP (drawn in Chemdraw Ultra and imported into RasWin) was positioned inside the cyclodextrin in accord with the description of the complex dCMP·heptakis(6-amino-6-deoxy)-β-cyclodextrin.[20,21]

Results

The $(M113R)_7$·am$_7\beta$CD Pore. The homomeric αHL pore $(M113R)_7$ has a ring of seven arginines located near the constriction in the barrel of the pore (Figure 1A). $(M113R)_7$ interacts with various phosphate esters.[19] However, when this mutant was tested for the direct detection of rNMP and dNMP, it did not display any clear current blocking events. To bring the diameter of the conducting pathway closer to the size of rNMP and dNMP, we fitted a cyclodextrin adapter near the constriction.[22] We used heptakis-(6-deoxy-6-amino)-β-cyclodextrin (am$_7\beta$CD) (Figure 1A), which is β-cyclodextrin with the seven primary hydroxyls replaced with amino groups. When am$_7\beta$CD is present inside the pore, two rings of positive charge, one ring of seven primary amino groups contributed by the cyclodextrin, and a second ring of seven arginine side-chains, are separated by ~10 Å (Figure 1A). Aminocyclodextrins have previously been shown to bind nucleoside monophosphates with the phosphate group in an ionic interaction with the protonated amino groups.[20,21] We felt it was possible that the overall stability of such complexes would be enhanced by π-cation interactions between the nucleotide bases and the Arg side chains (Figure 1A).

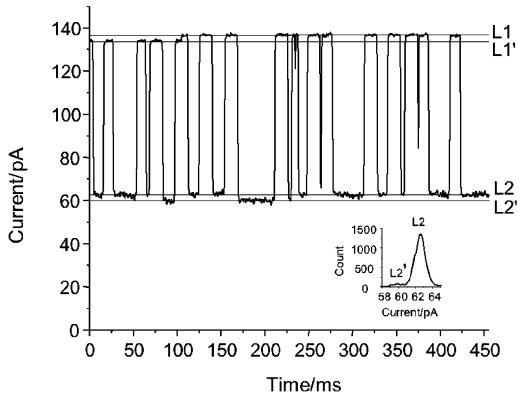

Figure 2. The αHL pore and am$_7\beta$CD produce rare L1′ and L2′ states. The current trace of a single $(M113R)_7$ pore with 40 μM am$_7\beta$CD in the trans chamber at +130 mV is shown. L1 and L1′ indicate two current levels of the unoccupied nanopore, and L2 and L2′ show two current levels resulting from the binding of am$_7\beta$CD to $(M113R)_7$. The insert shows an all-points amplitude histogram of a current trace containing more than 1000 events with the peaks corresponding to levels L2 and L2′.

2′-Deoxynucleoside 5′-Monophosphates Partially Block the $(M113R)_7$·am$_7\beta$CD Pore. We carried out single-channel electrical recordings on $(M113R)_7$ pores with am$_7\beta$CD applied from the trans side of the lipid bilayer (Figure 1B−D). In the absence of am$_7\beta$CD, the pore remained open (Figure 1B) with a unitary current (L1) of 141 ± 4 pA (n = 8) at +130 mV in 25 mM Tris·HCl, 1 M KCl, pH 8.0. The addition of 40 μM am$_7\beta$CD to the trans chamber led to reversible blocking events with a residual current level (L2) of 65 ± 2 pA (n = 8) (Figure 1C). Upon the addition of 5 μM dCMP to the cis chamber, a third current level (L3) was observed with a normalized value (see below) of 20 ± 1 pA (n = 5) (Figure 1D) originating from current level L2. L3 represents the binding of dCMP to the complex between $(M113R)_7$ and am$_7\beta$CD, $(M113R)_7$·am$_7\beta$CD. The other three dNMP and all four rNMP also produced events with a third current level (see below).

At pH 8.0, $(M113R)_7$ also exhibits a minor current level L1′ when the pore is unoccupied (Figure 2). am$_7\beta$CD binds to $(M113R)_7$ regardless of which current level the protein is in and generates two current levels L2, as already noted, and the minor level L2′ (Figure 2; three levels are observed at pH 7.5, not shown). L2′ accounts for <15% of the current passed when the cyclodextrin adapter is bound (Figure 2, inset; 8% in the recording shown), and in the present exploratory study we have ignored dNMP binding events originating from this level (but see Figure S1, for further analysis).

The addition of any dNMP at up to 300 μM to the trans, instead of the cis chamber, did not produce binding events with am$_7\beta$CD in the trans chamber (not shown). Current blockades were observed when unmodified β-cyclodextrin was added to the trans chamber (40 μM). However, no additional events were seen when dNMPs (up to 300 μM) were added either to the trans or to the cis chamber (not shown). In the absence of

(20) Eliseev, A. V.; Schneider, H.-J. *Angew. Chem., Int. Ed. Engl.* **1993**, *32*, 1331−1333.
(21) Eliseev, A. V.; Schneider, H.-J. *J. Am. Chem. Soc.* **1994**, *116*, 6081−6088.
(22) Song, L.; Hobaugh, M. R.; Shustak, C.; Cheley, S.; Bayley, H.; Gouaux, J. E. *Science* **1996**, *274*, 1859−1866.

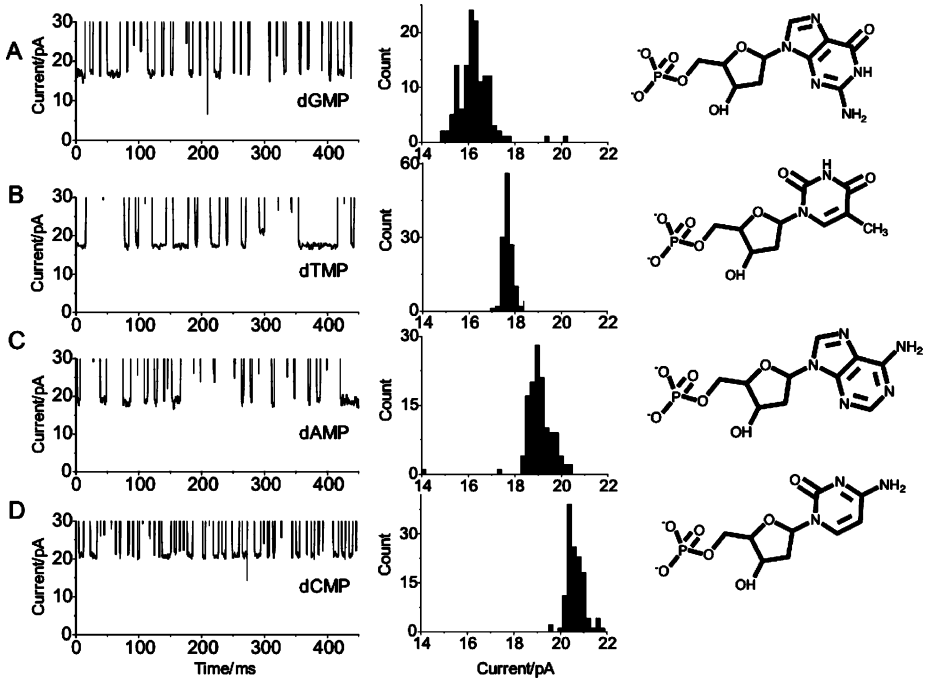

Figure 3. Current levels for each dNMP. The current traces from single $(M113R)_7$ pores at +130 mV with 40 μM $am_7\beta CD$ in the trans chamber are shown. dNMP (5 μM) was added to the cis chamber. (A) dGMP; (B) dTMP; (C) dAMP; (D) dCMP. Event histograms based on 133 events for each dNMP are shown at center, with the structure of the dNMP to the right. Events arising from the L2′ level of $(M113R)_7$·$am_7\beta CD$ were not included. The current levels for all four dNMP were normalized between zero and the $(M113R)_7$·$am_7\beta CD$ current level, by adjusting the $(M113R)_7$·$am_7\beta CD$ level to 65 pA.

$am_7\beta CD$ in the trans chamber, blocking events (<1 ms) were occasionally observed when 300 μM dGMP or dTMP was added to the cis chamber (not shown). However, these events were not observed over the course of a 2-h long experiment with 5 μM rNMP or dNMP in the cis chamber. Cyclodextrin binding events were observed with the WT αHL pore with $am_7\beta CD$ in the trans chamber, but no further alterations of the current were observed when dNMP was added to either the cis or the trans side.

2′-Deoxynucleoside 5′-Monophosphates Can Be Identified from the Amplitude of the Partial Block of $(M113R)_7$·$am_7\beta CD$ Pores. The partial block of the $(M113R)_7$·$am_7\beta CD$ complex differed in amplitude depending on which dNMP was added to the cis chamber (Figure 3, Table 1). The current amplitudes from separate experiments displayed variations that originated from the individualities of the protein nanopores. For example, the average current, at +130 mV, for a single $(M113R)_7$ pore is 141 pA, but some pores displayed currents as high as 146 pA and others currents as low as 135 pA. Therefore, to compare traces, we normalized the currents; the difference between zero current and the $(M113R)_7$·$am_7\beta CD$ current level (L2) was set to 65 pA. Dwell-time histograms were constructed over 500 events, and each was fit to a single exponential (not shown) to obtain the mean τ_{off} values for each dNMP (Table 1). The four different L3 states were observed when dGMP, dTMP, dAMP, and dCMP were added to the cis chamber simultaneously (Figure 4). Separate peaks were resolved in an events amplitude histogram (Figure 4B), even when blocks originating in L2′ were included (Figure S1).

Statistical Analysis. Current traces of $(M113R)_7$ in the presence of $am_7\beta CD$ (trans) and a single dNMP (cis) were digitally filtered at 1 kHz (low pass Gaussian filter), and event amplitude histograms were constructed. These histograms display a peak corresponding to the current level observed when $am_7\beta CD$ binds to $(M113R)_7$ (L2 in Figure 1). This current level (L2) varies slightly from one experiment to another (range ±5%). For this reason, the histograms were normalized between zero current and the L2 level, which was set at 65 pA. In the

Table 1. Mean Dwell Times and Mean Current Levels during Occupancies by DNMPs

	τ_{off} ±SD/ms[a]	current level ±SD/pA[b]	peak separation ±SD/pA[c]	Gaussian fit intercept ±SD/pA[d]
G	9.8 ± 0.2	16.2 ± 0.5	1.36 ± 0.05 (GT)	16.8 ± 0.1 (GT)
T	19.8 ± 0.8	17.6 ± 0.6	1.00 ± 0.09 (TA)	17.9 ± 0.1 (TA)
A	7.1 ± 0.2	18.6 ± 0.6	1.61 ± 0.12 (AC)	19.3 ± 0.1 (AC)
C	10.5 ± 0.4	20.2 ± 0.5		

[a] The currents passing through single $(M113R)_7$ pores were recorded in the presence of $am_7\beta CD$ and one of the four dNMPs as described in Materials and Methods. Each current trace was filtered digitally at 1 kHz. The dwell times for the nucleotide binding events originating from the L2 level of the $(M113R)_7$·$am_7\beta CD$ complex were recorded. The mean τ_{off} value for each dNMP was then obtained by fitting a dwell-time histogram (>500 events) to a single-exponential function. [b] Average peak positions were determined from current traces from five experiments in which all four nucleotides were mixed (5 μM dGMP, dCMP, and dAMP, and 10 μM dTMP). The traces were filtered digitally at 300 Hz, and all-points histograms were then generated. For each recording, the peak position of the $(M113R)_7$·$am_7\beta CD$ current level, L2, was determined and normalized to 65 pA. Four peaks were visible in each histogram and identified on the basis of the peak positions in experiments with individual nucleotides. The average peak position value in pA for each dNMP was obtained from the normalized values from the five different experiments. [c] Individual peak separation values were obtained from the five sets of peak position values determined as described above, and mean values were then determined. [d] The peak intercept values were obtained from three independent current traces with individual nucleotides. Each current trace was filtered digitally at 1 kHz. The nucleotide binding events arising from the L2′ of the $(M113R)_7$·$am_7\beta CD$ complex were discarded. The value of level L2 of the $(M113R)_7$·$am_7\beta CD$ complex was normalized to 65 pA. The events were plotted as amplitude histograms, and three Gaussian fits were obtained for each nucleotide from which the average σ values were determined. These σ values were used to fit a Gaussian distribution for each nucleotide that was centered on the mean peak value obtained from the experiments with all four nucleotides (this table). The intercepts between the four Gaussians were then determined (see Figure S2).

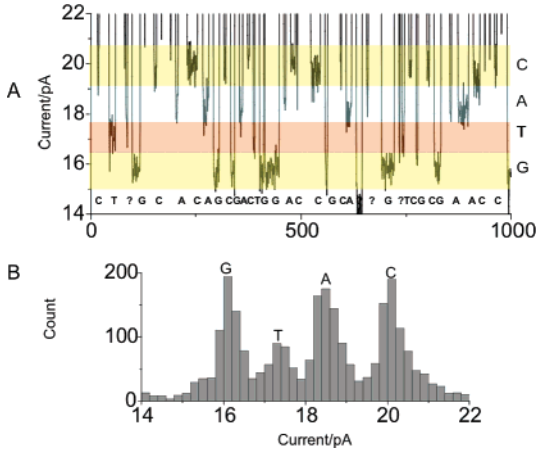

Figure 4. Simultaneous detection of dNMPs. (A) Recording from a single $(M113R)_7$ pore with 40 μM am$_7\beta$CD in the trans chamber. dGMP, dCMP, dAMP (all 5 μM), and dTMP (10 μM) were added to the cis chamber. The colored bands represent the width of the Gaussian distribution found to fit the current amplitudes of individual dNMPs cut off at the intersections with neighboring Gaussians (see the text). (B) Event histogram of the L3 levels from a current trace of 2000 binding events. Events arising from the L2′ level of $(M113R)_7$·am$_7\beta$CD were not included.

normalized histograms, the peaks representing the L3 level ($(M113R)_7$·am$_7\beta$CD·dNMP) were fitted to Gaussian distributions. The sigma (σ) values for each dNMP were averaged from at least three independent experiments, each containing at least 130 events (Table 1). We then examined the data from five independent experiments in which all four dNMPs were present (e.g., Figure 4) with at least 2000 binding events in each current trace. The traces were filtered (300 Hz low-pass Gaussian digital filter) and normalized between 0 and 65 pA, as described above for the experiments with individual dNMPs. All-points histograms were constructed to determine the peak current level in the presence of each nucleotide. The mean residual current values for each dNMP were averaged from the five experiments (Table 1). The standard deviations of the mean peak separations show that while the peak positions shifted from one experiment to another, the peak separations remained the same within ~10%. The mean peak positions for each dNMP taken from the events histograms with all four dNMPs and the mean σ values taken from the experiments with individual dNMPs (Table 1) were then used to produce new normalized Gaussians (i.e., curves enclosing the same area) for each nucleotide. The overlaps between the normalized Gaussians were then determined from their intercepts (Figure S2), to provide an estimate of the extent to which bases would be misidentified in the case where all four nucleotides are present in a sample at the same level (25%) (Table 2). A comparison of Table 2 and Figure 2 shows that this approach is likely to overestimate the extent of misidentification.

Detection of Ribonucleoside 5′-Monophosphates. rNMPs were also successfully discriminated by the same method at +110 mV, in 25 mM Tris·HCl, 1 M KCl, pH 8.0. The amplitude histogram peaks appear in the same order as those of the dNMPs, with rGMP displaying the lowest current (largest blocking), followed by rUMP, rAMP, and rCMP with the highest current level (smallest blocking). The discrimination between each nucleotide was inferior to that of the dNMPs, with peak separations smaller than 1 pA for U, A, and C.

Dependence of the Current Values on Experimental Conditions. The current level differences between the dNMPs are about 1 pA (Table 1). We found that this separation depends on a number of variables.

Table 2. The Overall Probability that a DNMP Is Detected as Itself or Another DNMP As Determined from the Overlap of the Gaussian Distributions for Each DNMP[a]

	G_{read}	T_{read}	A_{read}	C_{read}
G_{added}	0.98	0.02	0	0
T_{added}	0.03	0.93	0.04	0
A_{added}	0	0.03	0.97	0
C_{added}	0	0	0.04	0.96

[a] Gaussian distributions of the events corresponding to each nucleotide were constructed as described in Table 1. The area of each Gaussian distribution extending beyond the intercept(s) with the adjacent Gaussian(s) was integrated and represents the probability that a particular nucleotide would be assigned as its neighbor in the table (Figure S2).

First, the binding events are voltage dependent. At +50 mV, very few binding events are observed, suggesting that a minimum potential is required to drive the rNMP or dNMP to the binding site. At +150 and +200 mV, the current levels are too close to each other to allow the dNMP to be distinguished. A voltage of +130 mV proved to be optimal for dNMP, and +110 mV yielded the best resolution for rNMP.

Second, the current levels are pH dependent. Tris·HCl buffers containing 1 M KCl at pH values of 7.5, 8.0, 8.2, 8.5, 9.0, and 9.5 were tested. At pH 8.0 and above, two current levels are observed upon the binding of am$_7\beta$CD, but one (L2′) makes a minor contribution and has been ignored in this initial analysis (Figures 2, S1). At pH 7.5, am$_7\beta$CD generates a third current level (not shown). Under these conditions, dTMP binding generates two types of events with different current levels, one of which is similar to the dGMP current level. At pH 9.5, the dNMP binding events are no longer observed. The best peak separation is obtained at pH 8.0.

The effect of the electrolyte on the current levels was also examined. The separation of the current levels for the dNMPs and the rNMPs is better with KCl than with NaCl or CsCl (all 1 M). In the case of 1 M KBr, a complete block of $(M113R)_7$·am$_7\beta$CD was observed when dNMP bound, and therefore the nucleotides could not be distinguished. KCl solutions of 0.5, 1.0, and 2.0 M were tested. The best resolution in the events histogram (e.g., Figure 4) was obtained at 1.0 M KCl for both dNMP and rNMP. Lowering the temperature to 14 °C or increasing it to 50 °C did not improve the resolution.

Finally, αHL mutants other than $(M113R)_7$ were tested. $(M113N)_7$ bound am$_7\beta$CD, but dNMP and rNMP could not be detected. $(M113F)_7$ and $(M113F/K147N)_7$, whether complexed with am$_7\beta$CD or in its absence, did not bind dNMP or rNMP. $(M113K)_7$ gave results similar to $(M113R)_7$, but the peak separations between the dNMP were smaller.

Discussion

The results presented in this paper revive the possibility of using the identification of nucleotides generated by the exonuclease digestion of individual nucleic acid molecules as a cheap and rapid method for sequencing. Because it is a single-molecule approach, exonuclease sequencing circumvents the need for DNA amplification before sequence analysis, an expensive and time-consuming step. In previous attempts, fluorescent nucleotides have been first incorporated into the DNA, so that they can be detected subsequently by single-molecule fluorescence, but it has been difficult to obtain full incorporation and to

observe every molecule that is released.[8−10] Nevertheless, the approach remains attractive. For example, lambda exonuclease degrades one strand of duplex DNA in a highly processive manner moving in the 5′ to 3′ direction releasing bases one at a time as dNMP. The kinetics of the enzyme have been studied at the single molecule level;[23,24] one exonuclease molecule can digest all 48 502 nucleotides of lambda phage DNA at an average rate of 32 residues per second. A method for the detection of individual dNMPs might then be the key to the eventual success of exonuclease sequencing. We have shown that individual dNMP can be detected with an engineered protein nanopore sensor with an accuracy of at least 93−98% (Table 2) based on the extent of current block by each nucleotide. The dwell time values of the dNMP are too similar under our conditions to make a significant contribution to the identification of each dNMP.

Many problems must be solved before this approach can yield useful DNA sequences. First, the exonuclease must be coupled to the protein nanopore so that every dNMP that is released enters the pore. Second, exonuclease turnover must be slower than the dwell time of the nucleotide in the pore. This might be dealt with by mutagenesis of the enzyme or by manipulation of the identity and concentration of the required divalent metal ion. Third, the cyclodextrin adapter must be located in the nanopore while each dNMP enters. In the present experimental configuration, the am$_7\beta$CD adapter moves in and out of the pore (Figure 1). A dNMP molecule will not be detected if it passes through the pore in the absence of the adapter. This problem might be solved by covalently attaching the adapter inside the pore. The present recordings do show several dNMP binding events during a single cyclodextrin binding event, which proves that the adapter does not need to come off the pore for the dNMP to dissociate. Fourth, a dNMP that has just been detected must not interfere with the detection of the next incoming nucleotide. It remains to be demonstrated whether an exiting dNMP is released on the cis or the trans side of the bilayer, although it should be possible to control this by varying the applied potential.[25,26] Finally, while present accuracy of the identification of bases is surprisingly good for an initial assessment of the approach, it must be improved. One possibility might be to have several “looks” at the same nucleotide by pulling it in and out of the pore with an applied potential under feedback control.[27] Despite these difficulties, the ability to carry out long sequencing runs with some errors might have utility in the interim. For example, the data could be used to align short runs obtained through other approaches or to map DNA methylation sites. Once a nanopore sequencing device is operational, it should be relatively simple to produce multiplexed systems for ultrahigh throughput sequencing.

Acknowledgment. We thank Dr. S. Cheley, E. Mikhailova and O. Mates for valuable assistance. H.B. is the holder of a Royal Society-Wolfson Research Merit Award. This work was supported by a UK Research Councils Bionanotechnology IRC, the MRC, and the ONR.

Supporting Information Available: Figures S1 and S2 showing, respectively, the effect of the current level L2′ on the simultaneous detection of dNMPS and overlap between Gaussian distributions. Complete citation for ref 4. This material is available free of charge via the Internet at http://pubs.acs.org.

JA057123+

(23) Perkins, T. T.; Dalal, R. V.; Mitsis, P. G.; Block, S. M. *Science* **2003**, *301*, 1914−1918.
(24) Oijen, A. M. v.; Blainey, P. C.; Crampton, D. J.; Richardson, C. C.; Ellenberger, T.; Xie, X. S. *Science* **2003**, *301*, 1235−1238.
(25) Sanchez-Quesada, J.; Ghadiri, M. R.; Bayley, H.; Braha, O. *J. Am. Chem. Soc.* **2000**, *122*, 11757−11766.
(26) Gu, L.-Q.; Cheley, S.; Bayley, H. *Science* **2001**, *291*, 636−640.
(27) Bates, M.; Burns, M.; Meller, A. *Biophys. J.* **2003**, *84*, 2366−2372.