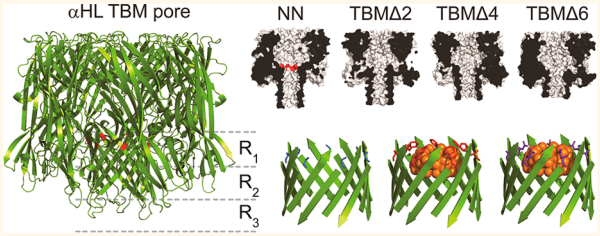
Nucleobase Recognition by Truncated α-Hemolysin Pores

Mariam Ayub,[†] David Stoddart,[†,‡] and Hagan Bayley*

Department of Chemistry, University of Oxford, Oxford, OX1 3TA, United Kingdom. [†]M.A. and D.S. contributed equally to this work. [‡]Present address: Oxford Nanopore Technologies, Oxford OX4 4GA, United Kingdom.

ABSTRACT The α-hemolysin (αHL) protein nanopore has been investigated previously as a base detector for the strand sequencing of DNA and RNA. Recent findings have suggested that shorter pores might provide improved base discrimination. New work has also shown that truncated-barrel mutants (TBM) of αHL form functional pores in lipid bilayers. Therefore, we tested TBM pores for the ability to recognize bases in DNA strands immobilized within them. In the case of TBMΔ6, in which the barrel is shortened by ~16 Å, one of the three recognition sites found in the wild-type pore, R1, was almost eliminated. With further mutagenesis (Met113 → Gly), R1 was completely removed, demonstrating that TBM pores can mediate sharpened recognition. Remarkably, a second mutant of TBMΔ6 (Met113 → Phe) was able to bind the positively charged β-cyclodextrin, am$_7$βCD, unusually tightly, permitting the continuous recognition of individual nucleoside monophosphates, which would be required for exonuclease sequencing mediated by nanopore base identification.



KEYWORDS: alpha-hemolysin · nanopore · base identification · truncated pore · toroidal lipid pore

Engineered protein pores are being developed for use in biotechnology, including applications in molecular sensing.[1−3] There has been particular interest in a new generation of nanopore DNA sequencers that would operate cheaply and quickly at the single-molecule level,[4,5] and recently this approach has proved successful.[6−10] The aim of the present study was to examine truncated-barrel mutants (TBM) of αHL[11] for their ability to identify RNA and DNA bases,[12−18] with the view to facilitate the electrical read-out of the sequences of nucleic acid molecules. The 5 nm-long transmembrane β barrel of the αHL pore comprises the base recognition region of the protein, and the wild-type (WT) pore contains three broad recognition sites, R_1, R_2 and R_3.[15]

Previous work has shown that individual nucleobases, presented in a fixed DNA homopolymer or heteropolymer background, can be identified at each of the three sites.[15,16,19,20] However, base identification is context dependent, and the signal ($I_{RES\%}$) from a given base is shifted when neighboring bases are changed. Context-dependent signals include additional information that is useful for sequence determination.[6,8,10,21−23] However, the signal will be uninterpretable unless the number of reading heads and their width is restricted. The *Mycobacterium smegmatis* MspA pore has favorable properties for reading DNA sequences on single strands, because changes in the ionic current are dominated by a single reading head that spans 3−4 bases.[9,10,23] In the present work, we attempted to reduce the number of reading heads in the αHL pore by using truncated pores[24] and thereby demonstrate an approach that might be generally useful for improving protein pores as sequence readers.

Our recent work has demonstrated that the αHL pore is able to withstand substantial truncations in the β barrel region and still form channels in lipid bilayers.[11] The β barrel contains 14 antiparallel β strands, with each protomer of the heptameric pore contributing two adjacent strands, which are connected by a turn (amino acids Gly-126 through Ile-132, Figure 1). The strands themselves largely consist of alternating hydrophilic and hydrophobic amino acids, with the side chains of the hydrophilic amino acids pointing into the lumen of the pore, and the side chains of the hydrophobic amino acids pointing into the lipid bilayer (Figure 1B). To truncate the β barrel,

* Address correspondence to hagan.bayley@chem.ox.ac.uk.

Received for review October 22, 2014 and accepted June 26, 2015.

Published online June 26, 2015
10.1021/nn5060317

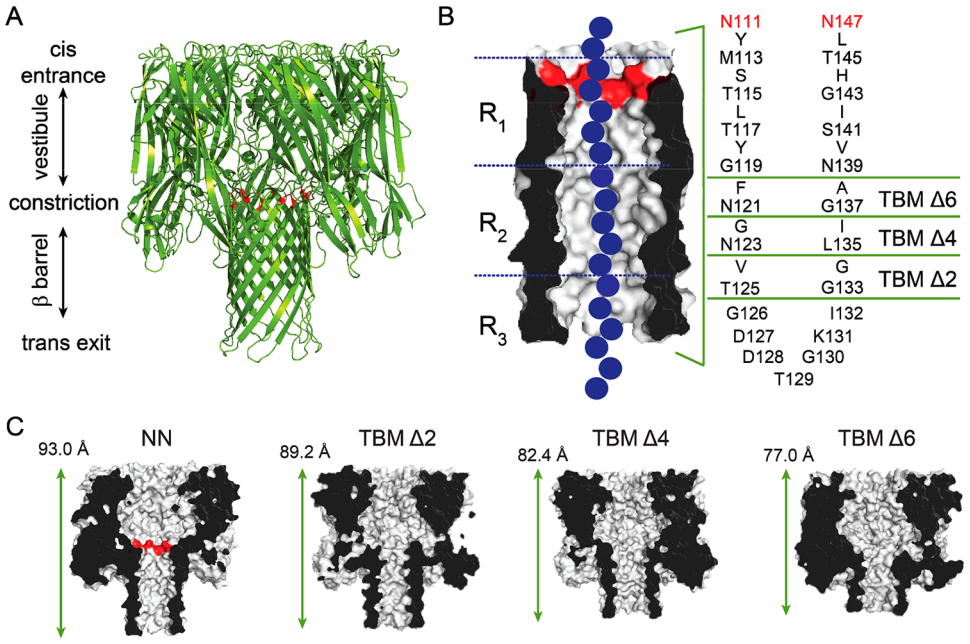


Figure 1. α-Hemolysin (αHL) protein nanopore. (A) Cartoon representation of the αHL pore (pdb: 7AHL). The αHL protein forms heptameric nanopores in lipid bilayers. The pore consists of an upper cap domain, which contains a roughly spherical, water-filled vestibule, and a transmembrane domain. The E111N and K147N mutations[15] at the top of the β barrel are highlighted in red, and replace the charged residues (Glu and Lys) at the central constriction. (B) The transmembrane domain is a 14-stranded, antiparallel β barrel. Each of the seven protomers contributes a pair of adjacent β strands (separated by a turn sequence; amino acids 126−132) to the barrel. The amino acid sequence of the transmembrane portion of the β strands for the most part alternates between hydrophilic residues (which face inward toward the water-filled lumen of the pore) and hydrophobic residues (which face outward toward the hydrocarbon core of the bilayer). To truncate the β barrel, rings of inward and outward facing residues from each of the two strands were deleted by PCR mutagenesis (leaving the turn sequence intact) to form truncated barrel mutant (TBM) proteins.[11] The three nucleobase recognition sites within the β barrel, R_1, R_2 and R_3, are also indicated.[15,16,19] (C) Cut-through representations of the truncated mutants, TBMΔ2, Δ4 and Δ6, used in the present study. The length indicated is the distance between the Cα atoms of amino acids N17 (located at the top of the cap domain), of the 3rd subunit, and T129 (located at the bottom of the transmembrane domain), of the 7th subunit.[11]

rings of inward and outward facing residues from each of the two strands were sequentially deleted by PCR mutagenesis (leaving the turn sequence intact) to form truncated barrel mutant (TBM) proteins. The rest of the TBM sequences were unaltered, except the charged residues at the central constriction (E111 and K147), which were mutated to neutral asparagines (NN). In TBMΔ2, amino acids V124 and T125 were deleted from the "down" strand and amino acids G133 and G134 were deleted from the "up" strand. TBMΔ4 and Δ6 were formed by deleting additional pairs of amino acids from each β strand (Figure 1B). As the mutant proteins have been demonstrated to adopt WT-like folds,[11] it is estimated that with each sequential truncation, the protein becomes ∼5 Å shorter in length (Figure 1C). To test the integrity of the barrel in TBM mutants, cyclodextrin[11] (CD) binding experiments were also carried out using β-cyclodextrin (βCD), heptakis-(6-deoxy-6-amino)-β-cyclodextrin (am_7βCD) and γ-cyclodextrin (γCD). CD binding within the β barrel of the αHL pore[24−26] is sensitive to small perturbations in the structure of the pore[26] or the cyclodextrin itself.[27] Interestingly, while the TBMΔ6 bound am_7βCD weakly, the mutation Met-113 → Phe, which strengthens βCD binding in the untruncated pore,[26] dramatically improved am_7βCD binding to TBMΔ6, allowing am_7βCD to remain bound to TBMΔ6/M113F for more than 1.5 h (at potentials of +60 to +140 mV).

RESULTS AND DISCUSSION

Defining Recognition Elements within the TBM Pores. The TBMΔ2, Δ4 and Δ6 pores were examined for the ability to discriminate single adenine bases, within immobilized poly(dC) oligonucleotides, in a similar manner to that previously established.[14,15,28] A set of 14 poly(dC) oligonucleotides was used, each containing a single adenine nucleobase. The adenine substitutions were in positions 7 to 20 relative to a 3′ biotin tag (Figure S1, S6 and Table S1), positions that span the entire length of the β barrel in full-length αHL pores. The residual current difference, $\Delta I_{RES\%}$ (with respect to poly(dC)), was plotted against the position of the adenine nucleobase for each of the truncated pores (Figure 2 and Table S2).

With each sequential truncation, the recognition region of the protein is reduced. The last nucleobase recognized by the full-length NN pore is at position 19 (relative to the 3′ biotin-tag) and after this position it is assumed that the immobilized DNA chain protrudes

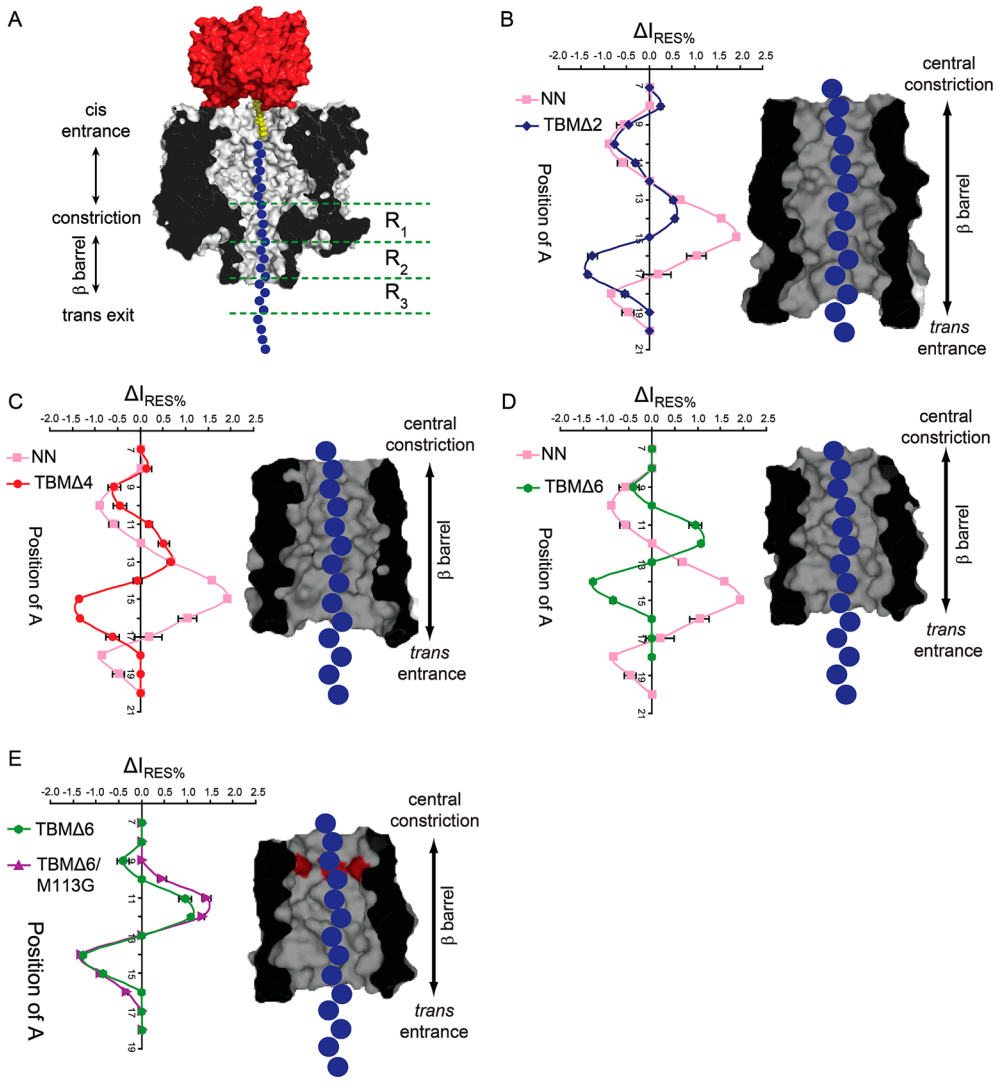


Figure 2. Effect of β barrel truncations on adenine recognition along the length of the β barrel. (A) Schematic representation of a homopolymeric DNA oligonucleotide (blue circles), immobilized inside the TBMΔ6 αHL pore (gray, cross-section) through the use of a 3′ biotin-TEG (yellow)·streptavidin (red) complex. The nucleobase recognition sites (R_1, R_2 and R_3) within the β barrel of the untruncated pore are shown alongside.[15,16,19] The differences in residual current ($\Delta I_{RES\%}$) between blockades caused by poly(dC) oligonucleotides containing a single adenine base (A_x) and poly(dC) (Table S2) for the αHL pores NN (pink) and (B) TBMΔ2 (blue), (C) TBMΔ4 (red) and (D) TBMΔ6 (green) are plotted. $I_{RES\%}$ values are mean values derived from Gaussian fits to event histograms. $I_{RES\%} = (I_{RES}/I_O) \times 100$. $\Delta I_{RES\%}$ is defined as the difference in residual current between an A_x oligonucleotide and poly(dC) ($I_{RES\%}^{Ax\ oligo} - I_{RES\%}^{pC}$) from an individual experiment. The means of the individual $\Delta I_{RES\%}$ values are plotted with s.d. values as error bars. A cross-section of the β barrel domain of the truncated αHL pores, filled with DNA indicating the positions of the immobilized bases, is shown in each case. (E) The differences in residual current ($\Delta I_{RES\%}$) between blockades caused by poly(dC) oligonucleotides containing a single adenine base and poly(dC) for the αHL pores TBMΔ6 (green) and TBMΔ6/3G (purple) (Table S3). A cross-section of the β barrel domain of the truncated αHL pore is shown (indicating the position of the mutation M113G) filled with DNA indicating the position of the immobilized bases. The data were collected by using a voltage protocol as described in Experimental Methods. Briefly, +160 mV was applied for 900 ms to drive the negatively charged, DNA into the pore, followed by −140 mV for 50 ms, to eject the immobilized DNA and a final step to 0 mV for 50 ms.

from the β barrel into the *trans* compartment.[15] However, in the truncated mutants, TBMΔ2, Δ4 and Δ6, the last nucleobase positions recognized are 18, 17, and 15 respectively. This suggests that the DNA protrudes from the β barrel sooner in the truncated mutants, and as expected the length of the recognition region has been reduced. However, adenine recognition by the truncated mutants is remarkable; the progressive changes in the patterns suggest that recognition site R_1 (near the central constriction) has been weakened, with R_2 and R_3 remaining, despite the removal of amino acids from the *trans* entrance of the pore. This suggests that recognition at sites R_2 and R_3 is not solely due to the interaction of nucleobases with specific amino acid side chains located at the bottom of the β barrel in the full-length pore. In the case of R_3, the DNA conformation upon exit from the pore or interaction with lipid head groups may affect the ionic current.

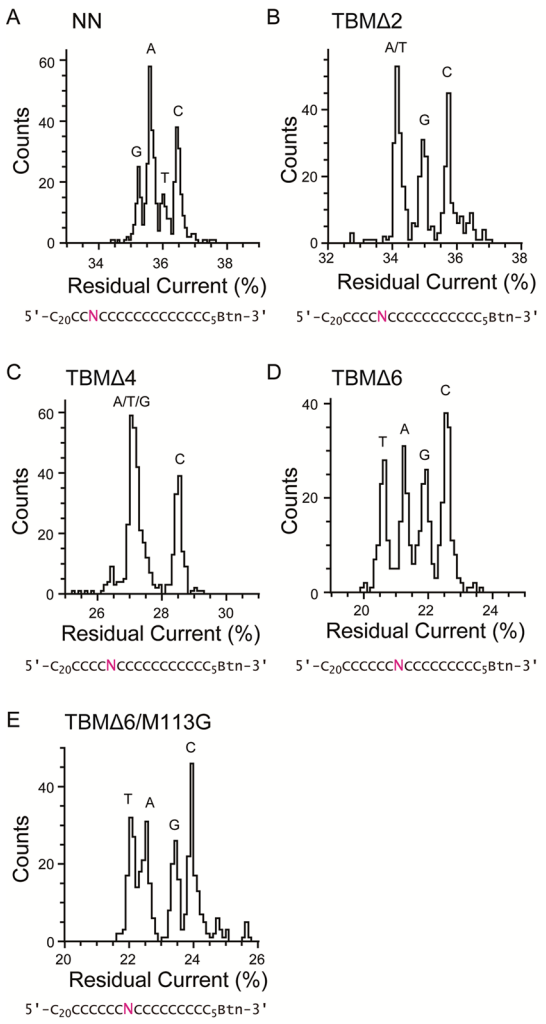


Figure 3. Single nucleobase discrimination in truncated αHL pores. Histograms of the residual current levels for NN, TBMΔ2, Δ4, Δ6 and TBMΔ6/M113G pores are shown. Three sets of four poly(dC) oligonucleotides were used, with each set containing either a single A, T, G or C nucleobase at a specific position. The position of the substituted base, N (purple), was designed to probe recognition site R_3 of each of the proteins. All experiments were conducted at least three times (Table S4), and the results displayed in the figure are from a typical experiment. (A) NN pores were interrogated with four oligonucleotides with the sequences 5′-CCCCCCCCCCCCCCCCCCCCCNCCCCCCCCCC-CCCCCCBtn-3′. TBMΔ2 (panel B) and TBMΔ4 (panel C) pores were interrogated with 5′-CCCCCCCCCCCCCCC-CCCCCCCCCNCCCCCCCCCCCCCCBtn-3′. TBMΔ6 (panel D) and Δ6/M113G (panel E) pores were interrogated with 5′-CCCCCCCCCCCCCCCCCCCCCCCCCNCCCCCCCCCCCC-Btn-3′. The data were collected by using a voltage protocol as described in Experimental Methods. Briefly, +160 mV was applied for 900 ms to drive the negatively charged, DNA into the pore, followed by −140 mV for 50 ms, to eject the immobilized DNA and a final step to 0 mV for 50 ms.

TABLE 1. The overall probability that a nucleobase is detected as itself or another nucleobase as determined from the overlap of the Gaussian fits for each nucleobase probed within NN, TBMΔ2, TBMΔ4, TBMΔ6, TBMΔ6/M113G and TBMΔ6/M113F pores.[a]

NN	G	A	T	C
G	0.89	0.12	0.00	0.00
A	0.12	0.83	0.06	0.00
T	0.00	0.06	0.84	0.11
C	0.00	0.00	0.11	0.90

TBMΔ2	G	A	T	C
G	1.00	0.00	0.00	0.00
A	0.00	0.75	0.25	0.00
T	0.00	0.25	0.75	0.00
C	0.00	0.00	0.00	1.00

TBMΔ4	G	A	T	C
G	0.66	0.17	0.17	0.00
A	0.17	0.66	0.17	0.00
T	0.17	0.17	0.66	0.00
C	0.00	0.00	0.00	1.00

TBMΔ6	G	A	T	C
G	0.85	0.11	0.00	0.05
A	0.11	0.82	0.08	0.00
T	0.00	0.08	0.92	0.00
C	0.05	0.00	0.00	0.96

TBMΔ6/M113G	G	A	T	C
G	0.97	0.00	0.00	0.03
A	0.00	0.79	0.21	0.00
T	0.00	0.21	0.79	0.00
C	0.03	0.00	0.00	0.97

TBMΔ6/M113F	dGMP	dAMP	dCMP	dTMP
dGMP	0.91	0.00	0.10	0.00
dAMP	0.00	0.86	0.14	0.01
dCMP	0.10	0.14	0.77	0.00
dTMP	0.00	0.01	0.00	0.99

TBMΔ6/M113F	rGMP	rAMP	rCMP	rUMP
rGMP	1.00	0.00	0.00	0.00
rAMP	0.00	0.81	0.00	0.20
rUMP	0.00	0.00	1.00	0.00
rCMP	0.00	0.20	0.00	0.81

[a] In order to calculate the overlap of the residual current distributions, each individual nucleobase peak seen in the $I_{RES\%}$ histograms (Figure 3 and 5) was fitted to a single Gaussian. The Gaussians were then normalized so that the probability of detecting a single nucleobase was equal. The overlap between each of the neighboring bases was calculated from half the area of overlap of the normalized Gaussians. Overlaps ranged between 0 (no overlap) and 1.0 (identical distributions).

The main goal of our experiments was to reduce the complexity of nucleobase recognition as it was felt that pores with more than two recognition sites would elicit current signals that would be too complex to be reliably interpreted.

From our initial findings, TBMΔ6 appeared to have very weak nucleobase recognition at the central constriction (R_1). Previous work has suggested that amino acid substitutions at position 113 (which, along with amino acids 111 and 147, comprises the central constriction) influence nucleobase recognition at R_1 with the M113G mutant providing the weakest nucleobase recognition.[19] Given this, the combined TBMΔ6/M113G mutant (Figure S1) was made in an attempt to

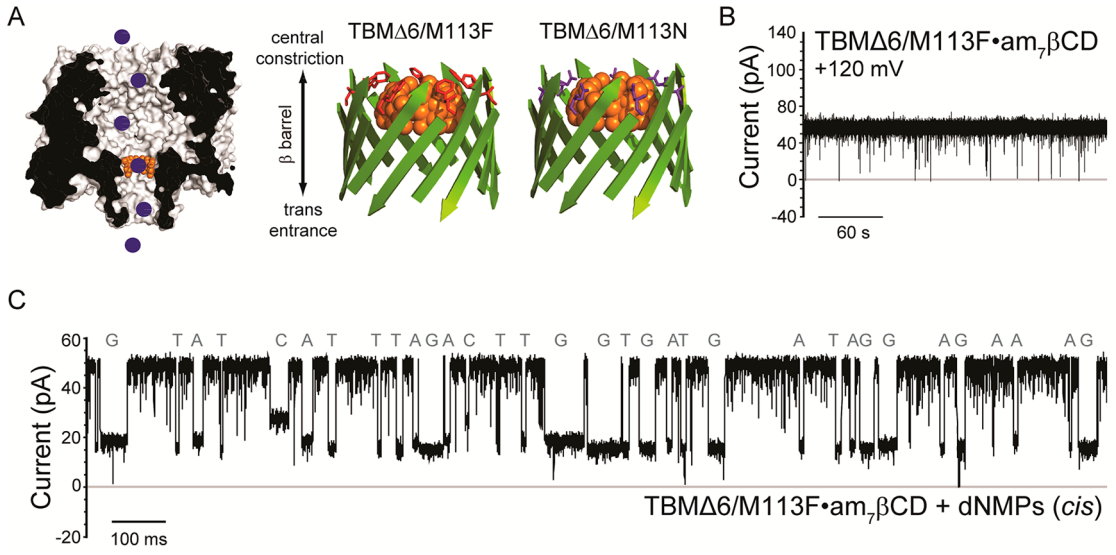


Figure 4. Continuous nucleobase discrimination in a truncated αHL pore with a cyclodextrin adapter. (A) Left: Schematic representation of individual DNA mononucleotides (blue circles), binding inside the TBMΔ6 pore (gray, cross-section) equipped with a cyclodextrin adapter (am$_7$βCD, orange). Right: Cartoon schematics of the TBMΔ6 β-barrel domain, showing the interaction of am$_7$βCD (orange) with the mutants M113F (red) and M113N (purple) as determined for the untruncated pore mutants.[29] (B) Representative single-channel trace of the TBMΔ6/M113F pore, in the presence of 40 μM am$_7$βCD. The recording was made in 1 M KCl, 25 mM Tris·HCl, pH 7.5, at +120 mV. (C) Single-channel recording from the TBMΔ6/M113F·am$_7$βCD pore showing continuous deoxyribonucleoside monophosphate (dNMP) detection (*cis*: 10 μM dGMP, 10 μM dAMP, 10 μM dCMP and 10 μM dTMP). Data were acquired at +120 mV. The amplified signal was low-pass filtered at 5 kHz and sampled at 25 kHz with a computer equipped with a Digidata 1440A digitizer (Molecular Devices).

remove the residual R_1 recognition site from the TBMΔ6 mutant and create a pore with just two recognition sites, TBMΔ6/M113G. It was indeed observed that the R_1 recognition site was removed by this mutation, while R_2 and R_3 remain largely unchanged by comparison with the TBMΔ6 pore (Figure 2E and Table S3).

Probing Recognition Site R_3 of αHL TBM Pores for Four-Base Discrimination. As well as examining the β barrel of each of the TBM mutant pores for the ability to discriminate single adenine bases, the lowermost recognition site, R_3, was also tested for the capability to discriminate all four bases. Sets of four poly(dC) oligonucleotides were used, which contained a single G, T, A or C nucleobase substitution at a position designed to interact with recognition site R_3. The position of the nucleobase substitution for each set differed, and was placed at the peak of recognition site R_3, which differs for each truncated pore (Figure 2). The set used to probe TBMΔ2 and Δ4, had the substitutions at position 16, and the set used to probe TBMΔ6 and TBMΔ6/M113G, had the substitutions at position 13. Although each of the TBM pores contained an R_3 recognition site that provided strong discrimination of adenine *versus* cytosine, only TBMΔ6 (and the Δ6/M113G derivative) retained an R_3 site capable of distinguishing the other nucleobases. Furthermore, the dispersion of the four current levels differs in the full-length pore and the truncated pores (Figure 3, Table 1 and Table S4). This implies that the amino acid deletions toward the *trans* entrance may have an effect on recognition site R_3.

Continuous Four-Base Mononucleotide Discrimination Using a Cyclodextrin Adapter. Successful identification of nucleoside monophosphates has been obtained with engineered αHL pores carrying cyclodextrin adapters, which can be noncovalently bound within the pore[12] or covalently attached for continuous base identification.[13,18] This approach has been proposed as an exosequencing platform, where bases are cleaved from a DNA strand by a processive exonuclease and identified as individual nucleotides by the nanopore. To test the TBMs for this purpose, two further mutants of TBMΔ6, TBM (Met113 → Phe) Δ6/M113F and (Met113 → Asn) Δ6/M113N were prepared to test the ability to detect individual DNA and RNA mononucleotides (NMPs) (Figures 4A, S2). These mutations in the full-length αHL pore have been shown to bind cyclodextrin (CD) adapters strongly (~10^4 times longer than the wild-type (WT) αHL pore).[25,29] Cyclodextrins in turn bind dNMPs and rNMPs, allowing their identification by current recording for potential exosequencing.[12,13,18] In the present study, the cyclodextrin heptakis-(6-deoxy-6-amino)-β-cyclodextrin (am$_7$βCD) was added to the *trans* compartment, and the positively charged amino groups promoted an extended residence time for the CD, at positive applied potentials.[12,13,18] The TBMΔ6 and Δ6/M113N pores released am$_7$βCD quickly (TBMΔ6: $k_{off} = 4600 \pm 300\ s^{-1}$; $K_D = 143 \pm 8$ mM; TBMΔ6/M113N: $k_{off} = 2200 \pm 200\ s^{-1}$; $K_D = 104 \pm 4$ mM) (Figure S3 and Table S5). Upon the addition of all four dNMPs (dGMP, dAMP, dTMP, dCMP) or rNMPs (rGMP, rAMP, rUMP,

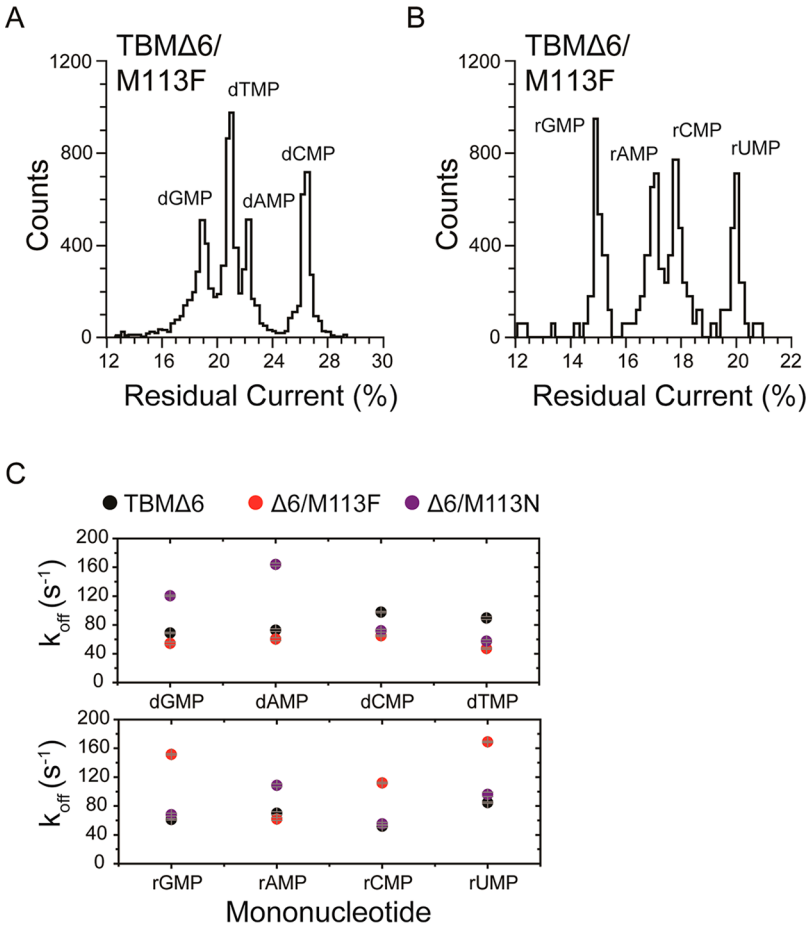


Figure 5. Residual current histogram of nucleotide binding events within the TBMΔ6/M113F·am$_7$βCD pore. Data were acquired in 1 M KCl, 25 mM Tris·HCl, pH 7.5, at +120 mV in the presence of 40 μM am$_7$βCD (*trans*). (A) dNMP results corresponding to the experiment shown in Figure 4B. (B) rNMP results from a typical experiment in the presence of 10 μM rGMP, 10 μM rAMP, 10 μM rCMP and 10 μM rUMP (all *cis*). (C) k_{off} values for each NMP detected with the TBMΔ6·am$_7$βCD, TBMΔ6/M113F·am$_7$βCD and TBMΔ6 M113N·am$_7$βCD pores. Values of k_{off} were determined by using $k_{off} = 1/\tau_{off}$, where τ_{off} is the mean dwell time for each rNMP in the pore. Each experiment was conducted at least three times.

rCMP) to the *cis* compartment, additional current blockades were observed originating from the CD blockade level, which represented the binding of NMPs to am$_7$βCD. The difference in residual current between the two most widely dispersed current peaks were: TBMΔ6, $\Delta I_{RES\%}{}^{OVERALL} = 2.6 \pm 0.4\%$ and Δ6/M113N, $\Delta I_{RES\%}{}^{OVERALL} = 2.8 \pm 0.6\%$ for dNMPs (Table S6).

By contrast, the TBMΔ6/M113F pore showed the remarkable ability to bind am$_7$βCD for over 1.5 h, thereby overcoming the difficult issue of using covalent chemistry to attach CDs (Figure 4B).[12,13,18,25,30,31] In the absence of am$_7$βCD, the Δ6/M113F pore remained open and passed a current, $I_O{}^{TBM\Delta6/M113F} = 113.0 \pm 6.0$ pA (n = 9, independent experiments) at +120 mV, in 1 M KCl, 25 mM Tris·HCl, at pH 6.0. The addition of 40 μM am$_7$βCD to the *trans* compartment produced a blocked level (lasting for >1.5 h) with a residual current level $I_{RES\text{-}am7\beta CD} = 65.0 \pm 4.0$ pA (n = 9, independent experiments) (Figure S3, typical open-states and CD binding traces for TBMΔ6, Δ6/M113N and Δ6/M113F pores).

Upon the addition of dNMPs or rNMPs to the *cis* compartment, additional current blockades were observed originating from the CD-blockade level corresponding to the binding of NMPs (Figure 4C). Both dNMPs and rNMPs could be discriminated in $I_{RES\%}$ histograms (Figure 5A,B and Table 1) with $\Delta I_{RES\%}{}^{OVERALL} = 6.0 \pm 1.4\%$ for dNMPs and $\Delta I_{RES\%}{}^{OVERALL} = 5.5 \pm 1.2\%$ for rNMPs (Table S6). The products of the sequential differences (δ) between each of the four residual current levels in the histograms were used to measure the ability of the different mutant pores to discriminate between the four DNA and RNA nucleotides.[19] A pore that is unable to discriminate between all NMPs has $\delta = 0$ (*i.e.*, the current levels of two or more NMPs overlap). For the experiments shown in Figure 5A and B, $\delta^{dNMP} = 5.4 \pm 0.4$ and $\delta^{rNMP} = 4.4 \pm 0.6$ were obtained. These results represent a significantly improved dispersion of the four standard nucleotides compared to previous studies using the full-length αHL pore.[13,18]

The NMPs also showed variations in mean dwell time (τ_{off}) within the CD adapter, which were used to derive the dissociation rate constants, k_{off} ($1/\tau_{off}$)

(Figure 5C).[13,18,25,31,32] At higher potentials, the binding of the charged nucleotides to the cyclodextrin adapter was promoted, resulting in a general decrease in k_{off} (Figures S4−S5 and Tables S7−S8), which suggests that an optimal potential would be required to obtain k_{off} values suitable for continuous nucleotide identification and accurate base calling to accommodate the rate of cleavage by the exonuclease.

CONCLUSION

Truncated αHL pores were examined for their ability to identify RNA and DNA bases with the view to facilitate the cheap electrical read-out of the sequences of nucleic acid molecules. First, nucleobase discrimination in ssDNA was examined by using 3′-biotinylated oligonucleotides bound to streptavidin.[15] Truncations in the β barrel of the αHL pore reduced the number of bases showing measurable interactions with the pore. The TBMΔ6 protein showed a weakened R_1 recognition site and it was possible to weaken this recognition site still further by additional site-directed mutagenesis to generate the TBMΔ6/M113G mutant pore, which displayed just two recognition sites: R_2 and R_3. Subsequent analysis of the TBMΔ6/M113G pore showed that recognition site R_3 was still capable of four nucleobase discrimination in ssDNA, although the order of the current levels differed from that found with the full-length αHL pore. Such changes in the order of the current levels have been noted previously.[15,16,19]

The ability of a truncated pore to detect mononucleotides was also examined. Such detection is a requisite for single-molecule nanopore exosequencing,[1,4,12,13,18] where bases are cleaved from a DNA or RNA strand by a processive exonuclease and identified as individual nucleotides by the nanopore.[1−3] Identification of all four mononucleotides has been obtained with engineered αHL pores carrying cyclodextrin adapters, which can be noncovalently bound within the pore[12,18] or covalently attached for continuous base identification.[13,18] Remarkably, in the present work, the cyclodextrin adapter am$_7\beta$CD was found to bind essentially irreversibly to the mutant truncated pore TBMΔ6/M113F, allowing all four dNMPs and all four rNMPs to be distinguished without breaks in current recording. This is as effective as working with a covalently attached cyclodextrin, an approach that requires difficult chemistry.[13,18,30]

Early work on base identification in DNA strands used the αHL pore, but it has been shown that the MspA pore gives a wider dispersion of current levels[9,23] suggesting that radical protein engineering (rather than point mutation of known pores) to produce sharper reading heads might improve current peak separation still further. The present work shows that a pore in which three recognition sites have been reduced to two can be quickly developed demonstrating the potential of such an approach, which should be generally applicable to a variety of pore-forming proteins. If further reduction in the length of the αHL pore is required to facilitate base identification, then shorter barrels[11] (*e.g.*, TBMΔ8) that do not form completely stable pores in lipid bilayers, might be inserted into solid-state apertures, to form protein·solid-state nanopore hybrids.[33] An additional advantage to the protein·solid-state hybrid system is that it is amenable to parallelization, which would make the throughput of nanopore sequencing competitive with highly parallel second generation sequencing technologies. Indeed, an array of 10^6 nanopores, each sequencing 100 bases per second could sequence a human genome cheaply in around 10 min, a feat which would make genomic medicine readily available.

EXPERIMENTAL METHODS

Protein Preparation. The αHL truncated barrel mutant (TBM) proteins were produced as described previously.[11] Aliquots of the purified protein were stored at −80 °C.

Mutagenesis. The NN mutant αHL gene was prepared with a kit for site-directed mutagenesis (QuikChange II XL, Catalog no. 200522−5, Stratagene) and verified by sequencing. The αHL truncated barrel mutants (TBMΔ2, Δ4 and Δ6) were generated from the NN gene in a pT7 vector[31,34] by PCR mutagenesis and ligation-free *in vivo* recombination,[35] and their sequences were verified. αHL TBMs NN/M113G/Δ120−125/Δ133−138 (TBMΔ6/M113G), NN/M113N/Δ120−125/Δ133−138 (TBMΔ6/M113N) and NN/M113F/Δ120−125/Δ133−138 (TBMΔ6/M113F) were also prepared by PCR mutagenesis and ligation-free recombination by using the NN/Δ120−125/Δ133−138 gene (TBMΔ6) in pT7 vector as a template (Figure S3).

Planar Bilayer Recording. Electrical recordings were carried out with a planar lipid bilayer apparatus with bilayers of 1,2-diphytanoyl-*sn*-glycero-3-phosphocholine (DPhPC, Avanti Polar Lipids). Bilayers were formed[36] across an aperture (~100 μm in diameter) in a 25-μm thick polytetrafluoroethylene (Teflon) film (Goodfellow, Cambridge, Cat. #FP301200/10), which separated the apparatus into *cis* and *trans* compartments. The transmembrane potential is given as the potential on the *trans* side (*i.e.*, the *trans* potential minus the *cis* potential, which was at ground). A positive current is one in which positive charge moves through the bilayer from the *trans* to *cis* side.

The aperture was first pretreated with hexadecane in *n*-pentane (10 mg mL^{-1}). Electrolyte solution (0.5 mL: 1 M KCl, 25 mM Tris·HCl, 0.1 mM EDTA, pH 8.0) was added to both compartments. Then, DPhPC in *n*-pentane (5 μL, 10 mg mL^{-1}) was added to both sides. The pentane was allowed to evaporate and a bilayer was formed by lowering and raising the electrolyte level past the aperture. All current recordings were performed with a patch clamp amplifier (Axopatch 200B, Molecular Devices). αHL heptameric pores were added to the *cis* compartment.

Single Base Identification with the Streptavidin-Immobilization Technique. ssDNA molecules, with a biotinyl group covalently attached to the 3′ end through a linker (Figure S1 and Table S1), were obtained from Sigma-Aldrich (UK). Solutions of the biotinylated ssDNAs, 100 μM in 1 M KCl, 10 mM Tris·HCl, pH 8.0, 0.1 mM EDTA, were preincubated with equal volumes of 25 μM streptavidin (Sigma-Aldrich) in the same buffer for at least 5 min. Oligonucleotides were added to the *cis* compartment to a final concentration of 400 nM. A voltage protocol was then initiated.

First, +160 mV was applied for 900 ms to drive the negatively charged, DNA into the pore. The capture of a ssDNA molecule by an αHL pore is observed as a stepwise decrease in the open pore current level (I_O) to a lower, stable, current level (I_B). A potential of −140 mV was then applied for 50 ms to eject the immobilized DNA from the pore. The applied potential was then stepped to 0 mV for 50 ms. This one-second sequence was repeated for at least 100 cycles for each ssDNA sequence added. The amplified signal arising from the ionic current was low-pass filtered at 1 kHz and sampled at 5 kHz with a computer equipped with a Digidata 1440A digitizer (Molecular Devices).

Data were analyzed and prepared for presentation with pClamp software (version 10.1, Molecular Devices). Event searches were performed to obtain the average residual current level for each ssDNA blockade (I_{RES}). The mean I_B value for each oligonucleotide (I_{RES}) was determined by a performing a Gaussian fit to a histogram of the I_B values. The percent residual current blockade was $I_{RES\%} = (I_{RES}/I_O) \times 100$. In general, when comparing several oligonucleotide species, a single oligonucleotide species was first added to the *cis* chamber and the current traces required for the determination of I_{RES} recorded. Subsequently, a second, third and then fourth oligonucleotide was added, and additional currents recorded. When such experiments were repeated, the oligonucleotides were added to the chamber in a different order.

Nucleoside Monophosphate Base Identification with a Cyclodextrin Adapter. The mononucleotides, 2-deoxyguanosine 5-monophosphate sodium salt (dGMP, >99%, Acros); 2-deoxycytosine 5-monophosphate disodium salt (dCMP, >95%, Fluka); 2-deoxythymidine 5-monophosphate disodium salt (dTMP, >97%, Fluka); 2-deoxyadenosine 5-monophosphate disodium salt (dAMP, >95%, Fluka); uridine 5-monophosphate disodium salt (rUMP, 99%, Fluka); cytosine 5-monophosphate (rCMP, >98%, Fluka); adenosine 5-monophosphate (rAMP, 99%, Acros); guanosine 5-monophosphate disodium salt (rGMP, 97%, Acros) and heptakis(6-deoxy-6-amino)-β-cyclodextrin·7HCl ($am_7\beta CD$, >99%, Cyclolab) were used without further purification.

TBMΔ6/M113F pores and dNMPs or rNMPs (10 μM) were added to the *cis* compartment, which was connected to ground. Once a single channel was obtained, $am_7\beta CD$ (40 μM) was added to the *trans* compartment. Both compartments contained 0.5 mL of buffer: 1 M KCl, 25 mM Tris·HCl, pH 6.0. After the addition of $am_7\beta CD$ to the *trans* compartment, a permanent drop in the current was observed (~70%). Data were typically acquired at +120 mV. The amplified signal (arising from the ionic current passing through the pore) was low-pass filtered at 5 kHz and sampled at 25 kHz with a computer equipped with a Digidata 1440A digitizer (Molecular Devices). The data were analyzed and presented by using pClamp software (version 10.1, Molecular Devices). Events were detected by using the "Event Detection" feature. The mean residual current value (I_{RES}) for each NMP interacting with $am_7\beta CD$ was determined by performing a Gaussian fit to a histogram of the current values for individual blockades (I_B). The current blockade for each NMP was further described by $I_{RES\%}$, where $I_{RES\%} = (I_{RES}/I_{CD}) \times 100$, where I_{CD} is the current in the presence of the cyclodextrin. The mean dwell time (τ_{off}) for each NMP was determined by fitting an exponential function to a dwell time histogram.

Conflict of Interest: The authors declare the following competing financial interest(s): Hagan Bayley is the Founder, a Director and a share-holder of Oxford Nanopore Technologies, a company engaged in the development of nanopore sequencing technology. Work in the Bayley laboratory at the University of Oxford, including this work, is supported in part by Oxford Nanopore Technologies.

Acknowledgment. This work was supported by grants from the National Institutes of Health and Oxford Nanopore Technologies. D.S. was supported by a BBSRC Doctoral Training Grant. The authors thank Ellina Mikhailova for the preparation of the NN gene and αHL heptamers. All the authors conceived the ideas and designed the experiments. M.A. and D.S. performed the experiments and analyzed the data. All the authors have contributed to and have given approval to the final version of the manuscript.

Supporting Information Available: Details of experimental procedures and the data displayed in Figures 1−5. The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/nn5060317.

REFERENCES AND NOTES

1. Bayley, H.; Cremer, P. S. Stochastic Sensors Inspired by Biology. *Nature* **2001**, *413*, 226–230.
2. Howorka, S.; Siwy, Z. Nanopore Analytics: Sensing of Single Molecules. *Chem. Soc. Rev.* **2009**, *38*, 2360–2384.
3. Majd, S.; Yusko, E. C.; Billeh, Y. N.; Macrae, M. X.; Yang, J.; Mayer, M. Applications of Biological Pores in Nanomedicine, Sensing, and Nanoelectronics. *Curr. Opin. Biotechnol.* **2010**, *21*, 439–476.
4. Bayley, H. Sequencing Single Molecules of DNA. *Curr. Opin. Chem. Biol.* **2006**, *10*, 628–637.
5. Branton, D.; Deamer, D. W.; Marziali, A.; Bayley, H.; Benner, S. A.; Butler, T.; Di Ventra, M.; Garaj, S.; Hibbs, A.; Huang, X.; *et al.* The Potential and Challenges of Nanopore Sequencing. *Nat. Biotechnol.* **2008**, *26*, 1146–1153.
6. Hayden, E. C. Nanopore genome sequencer makes its debut. *Nature* **2012**, DOI: 10.1038/nature.2012.10051.
7. Pennisi, E. Search for Pore-Fection. *Science* **2012**, *336*, 534–537.
8. Cherf, G. M.; Lieberman, K. R.; Rashid, H.; Lam, C. E.; Karplus, K.; Akeson, M. Automated Forward and Reverse Ratcheting of DNA in a Nanopore at 5-Å Precision. *Nat. Biotechnol.* **2012**, *30*, 344–348.
9. Manrao, E. A.; Derrington, I. M.; Laszlo, A. H.; Langford, K. W.; Hopper, M. K.; Gillgren, N.; Pavlenok, M.; Niederweis, M.; Gundlach, J. H. Reading DNA at Single-Nucleotide Resolution with a Mutant MspA Nanopore and Phi29 DNA Polymerase. *Nat. Biotechnol.* **2012**, *30*, 349–353.
10. Laszlo, A. H.; Derrington, I. M.; Ross, B. C.; Brinkerhoff, H.; Adey, A.; Nova, I. C.; Craig, J. M.; Langford, K. W.; Samson, J. M.; Daza, R.; *et al.* Decoding Long Nanopore Sequencing Reads of Natural DNA. *Nat. Biotechnol.* **2014**, *32*, 829–833.
11. Stoddart, D.; Ayub, M.; Höfler, L.; Raychaudhuri, P.; Klingelhoefer, J. W.; Maglia, G.; Heron, A.; Bayley, H. Functional Truncated Membrane Pores. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, 2425.
12. 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.* **2006**, *128*, 1705–1710.
13. Clarke, J.; Wu, H.-C.; Jayasinghe, L.; Patel, A.; Reid, S.; Bayley, H. Continuous Base Identification for Single-Molecule Nanopore DNA Sequencing. *Nat. Nanotechnol.* **2009**, *4*, 265–270.
14. Ashkenasy, N.; Sánchez-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.* **2005**, *44*, 1401–1404.
15. Stoddart, D.; Heron, A. J.; Mikhailova, E.; Maglia, G.; Bayley, H. Single-Nucleotide Discrimination in Immobilized DNA Oligonucleotides with a Biological Nanopore. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 7702–7707.
16. Stoddart, D.; Maglia, G.; Mikhailova, E.; Heron, A. J.; Bayley, H. Multiple Base-Recognition Sites in a Biological Nanopore: Two Heads Are Better Than One. *Angew. Chem., Int. Ed.* **2010**, *49*, 556–559.
17. Ayub, M.; Bayley, H. Individual RNA Base Recognition in Immobilized Oligonucleotides Using a Protein Nanopore. *Nano Lett.* **2012**, *12*, 5637–5643.
18. Ayub, M.; Hardwick, S. W.; Luisi, B. F.; Bayley, H. Nanopore-Based Identification of Individual Nucleotides for Direct RNA Sequencing. *Nano Lett.* **2013**, *13*, 6144–6150.
19. Stoddart, D.; Heron, A. J.; Klingelhoefer, J.; Mikhailova, E.; Maglia, G.; Bayley, H. Nucleobase Recognition in ssDNA at the Central Constriction of the α-Hemolysin Pore. *Nano Lett.* **2010**, *10*, 3633–3637.
20. Ayub, M.; Bayley, H. Single Molecule RNA Base Identification with a Biological Nanopore. *Biophys. J.* **2012**, *102*, 429a–429a.

21. Loman, N. J.; Watson, M. Successful Test Launch for Nanopore Sequencing. *Nat. Methods* **2015**, *12*, 303–304.
22. Jain, M.; Fiddes, I. T.; Miga, K. H.; Olsen, H. E.; Paten, B.; Akeson, M. Improved Data Analysis for the Minion Nanopore Sequencer. *Nat. Methods* **2015**, *12*, 351–356.
23. Manrao, E. A.; Derrington, I. M.; Pavlenok, M.; Niederweis, M.; Gundlach, J. H. Nucleotide Discrimination with DNA Immobilized in the MspA Nanopore. *PLoS One* **2011**, *6*, e25723.
24. 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* **1999**, *398*, 686–690.
25. Gu, L.-Q.; Cheley, S.; Bayley, H. Capture of a Single Molecule in a Nanocavity. *Science* **2001**, *291*, 636–640.
26. Banerjee, A.; Mikhailova, E.; Cheley, S.; Gu, L.-Q.; Montoya, M.; Nagaoka, Y.; Gouaux, E.; Bayley, H. Molecular Bases of Cyclodextrin Adapter Interactions with a Protein Nanopore. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 8165–8170.
27. Li, W.-W.; Claridge, T. D. W.; Li, Q.; Wormald, M. R.; Davis, B. G.; Bayley, H. Tuning the Cavity of Cyclodextrins: Altered Sugar Adaptors in Protein Pores. *J. Am. Chem. Soc.* **2011**, *133*, 1987–2001.
28. Purnell, R. F.; Mehta, K. K.; Schmidt, J. J. Nucleotide Identification and Orientation Discrimination of DNA Homopolymers Immobilized in a Protein Nanopore. *Nano Lett.* **2008**, *8*, 3029–3034.
29. Banerjee, A.; Mikhailova, E.; Cheley, S.; Gu, L.-Q.; Montoya, M.; Nagaoka, Y.; Gouaux, E.; Bayley, H. Molecular Bases of Cyclodextrin Adapter Interactions with Engineered Protein Nanopores. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 8165–8170.
30. Wu, H.-C.; Astier, Y.; Maglia, G.; Mikhailova, E.; Bayley, H. Protein Nanopores with Covalently Attached Molecular Adapters. *J. Am. Chem. Soc.* **2007**, *129*, 16142–16148.
31. Gu, L.-Q.; Cheley, S.; Bayley, H. Prolonged Residence Time of a Noncovalent Molecular Adapter, β-Cyclodextrin, within the Lumen of Mutant α-Hemolysin Pores. *J. Gen. Physiol.* **2001**, *118*, 481–494.
32. Sanchez-Quesada, J.; Ghadiri, M. R.; Bayley, H.; Braha, O. Cyclic Peptides as Molecular Adapters for a Pore-Forming Protein. *J. Am. Chem. Soc.* **2000**, *122*, 11757–11766.
33. Hall, A. R.; Scott, A.; Rotem, D.; Mehta, K. K.; Bayley, H.; Dekker, C. Hybrid Pore Formation by Directed Insertion of α-Haemolysin into Solid-State Nanopores. *Nat. Nanotechnol.* **2010**, *5*, 874–877.
34. Gu, L.-Q.; Dalla Serra, M.; Vincent, J. B.; Vigh, G.; Cheley, S.; Braha, O.; Bayley, H. Reversal of Charge Selectivity in Transmembrane Protein Pores by Using Non-Covalent Molecular Adapters. *Proc. Natl. Acad. Sci. U. S. A.* **2000**, *97*, 3959–3964.
35. Howorka, S.; Bayley, H. Improved Protocol for High-Throughput Cysteine Scanning Mutagenesis. *Biotechniques* **1998**, *25*, 766–772.
36. Montal, M.; Mueller, P. Formation of Bimolecular Membranes from Lipid Monolayers and a Study of Their Electrical Properties. *Proc. Natl. Acad. Sci. U. S. A.* **1972**, *69*, 3561–3566.