Published in final edited form as:
Nano Lett. 2012 November 14; 12(11): 5637–5643. doi:10.1021/nl3027873.

Individual RNA Base Recognition in Immobilized Oligonucleotides using a Protein Nanopore

Mariam Ayub and **Hagan Bayley***
Department of Chemistry, University of Oxford, Oxford, OX1 3TA, United Kingdom

Abstract

Protein nanopores are under investigation as key components of rapid, low-cost platforms to sequence DNA molecules. Previously, it has been shown that the α-hemolysin (αHL) nanopore contains three recognition sites, capable of discriminating between individual DNA bases when oligonucleotides are immobilized within the nanopore. However, the direct sequencing of RNA is also of critical importance. Here, we achieve sharply defined current distributions that enable clear discrimination of the four nucleobases, guanine, cytosine, adenine and uracil, in RNA. Further, the modified bases, inosine, N^6-methyladenosine and N^5-methylcytosine, can be distinguished.

Keywords

α-hemolysin; nanopore; RNA sequencing; modified bases; protein engineering; single-nucleotide discrimination

Single-molecule nanopore technology is under development for ultra-rapid, low-cost sequencing of DNA and RNA molecules. Two types of nanopore are being investigated: solid state pores[1] and protein pores, such as the heptameric α-hemolysin (αHL) pore. Protein pores have spearheaded the approach as they can be precisely manipulated by chemical and genetic engineering[2], which facilitates the determination of sequence in individual DNA strands through base-dependent transitions in ionic current flow[3]. By immobilizing and stretching DNA strands within protein nanopores, the four canonical DNA bases and epigenetically modified bases have been individually identified by ionic current recording[4–10]. Individual oxidized bases and abasic sites can also be identified after chemical modification[11, 12]. However, ssDNA moves through protein nanopores at remarkably high velocities (e.g. ~1–3 μs per nucleobase)[13], which provides insufficient signal-to-noise for individual bases to be identified. Therefore, enzymes have been used to ratchet DNA through pores[14–17]. By using approaches related to these published methods, Oxford Nanopore Technologies have demonstrated nanopore sequencing, achieving kilobase reads[18]. Additional studies of nanopore sequencing are appearing in the open literature[19, 20].

Nanopore sequencing of RNA has received less attention. The direct high-throughput sequencing of RNAs (mRNA, miRNA etc) will allow the rapid identification and quantitation of functional elements of the genome and reveal important splicing patterns and post-transcriptional modifications[21–23]. RNA sequencing will be valuable in medical

*Corresponding Author: hagan.bayley@chem.ox.ac.uk.

ASSOCIATED CONTENT

Supporting Information. Details of experimental procedures, oligonucleotide sequences, and the data displayed in Figures 1–5. This material is available free of charge via the Internet at http://pubs.acs.org.

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.

diagnosis, the selection of therapies and prognosis. For example, transcriptome sequencing has already been used to detect gene fusions in cancer[24, 25]. The ability to sequence extracellular RNA from plasma enhances the power of such approaches[22, 26, 27]. The genomes of RNA viruses and viral RNA transcripts are also accessible. RNA sequencing might also be used to monitor the levels and turnover of therapeutic RNAs[28–30]. Further, nanopores can directly identify modified bases, which are prevalent in RNA[9, 31].

Short ssRNA homopolymer molecules have been distinguished on the basis of differences in residual currents (I_{RES}) recorded while the RNAs are translocating through the wild-type (WT) αHL pore in an applied potential[32–35]. The transition between two homopolymer runs oligo(rA) and oligo(rC) within a single translocating RNA strand has also been observed, and may be accentuated by differences in the helical structures of the two regions[32, 33]. However, as in the case of DNA, individual RNA bases have not been identified in moving strands. Therefore, we have examined nucleobase identification in RNA strands by capturing and immobilizing them within the αHL pore with the biotin-streptavidin approach used previously for DNA base identification[5, 6]. The 5 nm-long β barrel of the αHL nanopore contains three recognition sites, R_1, R_2 and R_3, capable of recognizing individual bases in ssDNA (Figure 1a). We investigated RNA base substitutions at position 9 of synthetic oligonucleotides (bases numbered from the 3' end) to probe the most promising recognition site R_1 (which is located at the central constriction of the pore, comprising residues Lys-147, Glu-111 and Met-113, Figure 1b). Modification of the charge distribution within R_1 has a powerful effect on I_{RES} for DNA oligonucleotides[5, 8]. Therefore, we employed homoheptameric mutant pores made from αHL E111N/K147N (NN) and αHL E111N/K147N/M113Y (NNY) subunits.

Single nucleotide detection in homopolymeric strands

RNA base discrimination was first tested in homopolymer oligonucleotides, consisting of 30 nucleotides (the sequences of the oligonucleotides used in this paper are in Table S1). ssRNA oligonucleotides with biotin tags at the 3' end (Figure S1) were allowed to form complexes with streptavidin (SI Methods). In this state, the strands were captured and immobilized by αHL pores in an applied potential, but they were not translocated into the *trans* compartment (Figure 1a, Figure S2 automated voltage protocol). The one-second capture sequence was repeated for at least 400 cycles for each ssRNA added, with >90% of the cycles giving current blockades. The extended residence times of the oligonucleotides within the pore allowed reduction of the current noise by stringent filtration, thereby improving the signal-to-noise ratio and the precision of the measurements.

Once captured, the immobilized RNA molecules caused a sequence-dependent decrease in the ionic current through the pore. At +200 mV in 1 M KCl, 25 mM Tris-HCl, pH 7.5, containing 100 μM EDTA, WT αHL pores have a mean open pore current level (I_O^{WT}) of 199 ± 6 pA (n = 12), while the pores formed from NN and NNY, the mutants used in this work, gave I_O^{NN} = 214 ± 7 pA (n = 8) and I_O^{NNY} = 210 ± 8 pA (n = 8), respectively. Although the open pore currents carried by the WT pore, NN and NNY are similar (Figure S3), the residual currents in the presence of ssDNA[5, 36, 37] and ssRNA (this work) are ~50% higher in the mutant pores (Figure 1c–e) owing to the increased cross section of the lumen after mutation. Immobilized $oligo(rA)_{30}$ blocked NNY pores to a greater extent ($I_{RES\%}^{oligo(rA)}$ = 32.6 ± 0.2%) than $oligo(rC)_{30}$ ($I_{RES\%}^{oligo(rC)}$ = 33.5 ± 0.2%) and $oligo(rU)_{30}$ ($I_{RES\%}^{oligo(rU)}$ = 34.2 ± 0.2%) (Figure 1e, $I_{RES\%} = I_{RES}/I_O$ X 100). The residual current difference between the $oligo(rC)_{30}$ and the $oligo(rA)_{30}$ oligonucleotide blockades ($\Delta I_{RES\%} = I_{RES\%}^{oligo(rC)} - I_{RES\%}^{oligo(rA)}$) is +0.9 ± 0.2% and the $\Delta I_{RES\%}$ between $oligo(rU)_{30}$ and the $oligo(rA)_{30}$ oligonucleotide ($\Delta I_{RES\%} = I_{RES\%}^{oligo(rU)} - I_{RES\%}^{oligo(rA)}$) is +1.6 ± 0.2%. Furthermore, the difference in residual current between the two most widely

dispersed current peaks, ($\Delta I_{RES\%}{}^{OVERALL}$) between the three homopolymers also increases in the mutant pores compared to WT, providing improved levels of discrimination: αHL WT, $\Delta I_{RES\%}{}^{OVERALL}$ = 1.5 ± 0.4%; αHL NN, $\Delta I_{RES\%}{}^{OVERALL}$ = 2.4 ± 0.2% and αHL NNY, $\Delta I_{RES\%}{}^{OVERALL}$ = 2.8 ± 0.4% (n = 3 for each pore). There is also a change in the pattern of the current blocks with oligo(rC) producing higher $I_{RES\%}$ values than oligo(rA) in the NN and NNY pores, compared to WT (Figure 1c–e).

The $I_{RES\%}$ values of the immobilized RNA homopolymers increased with applied positive potential for WT, NN and NNY pores (Figure S4, Table S2). A likely explanation is stretching of the RNA molecules inside the pore. Single-stranded homopolymers are highly flexible with a persistence length of ~1 nm[38, 39]. Therefore, the molecules will be elongated by confinement within the β barrel and further elongated by the applied potential, which generates a force of ~10 pN on the molecule (SI Methods)[40–42]. At higher potentials, the RNAs will become more fully extended within the pore, and blockade of the ionic current will be correspondingly reduced, resulting in higher $I_{RES\%}$ values. The overall dispersion, $\Delta I_{RES\%}{}^{OVERALL}$, is also affected by increased potentials, which result in a tighter distribution of blockades by the three homopolymers, e.g. for αHL NNY we observed a ~64% decrease in dispersion, from $\Delta I_{RES\%}{}^{OVERALL}$ = 4.4 ± 0.4% at +100 mV to 2.8 ± 0.4% at +200 mV (Table S2). The product of the sequential differences (δ) between each of the three residual current levels in the histograms, can also be determined to gauge the ability of the pore to discriminate between the three homopolymers (Table S2, SI Methods). All three pores are able to discriminate between oligo(rA), oligo(rC) and oligo(rU) within the potential range (+100 to +200 mV, δ > 0). With αHL NNY displaying the best discrimination at the highest applied potential (+200 mV), δ^{NNY} = 1.8 ± 0.2%. These investigations of voltage dependence aid in the optimization of I_{RES} overlap.

Discrimination of all four ribobases in single-stranded oligonucleotides

We examined the ability of WT, NN and NNY αHL pores to distinguish between rG, rA, rU, and rC within individual nucleic strands. A first set of four oligonucleotides comprised oligo$(dC)_{30}$ with a ribonucleotide at position-9 relative to the biotinylated 3' end. A second set of four oligonucleotides comprised oligo$(rA)_{30}$ with a ribonucleotide at position-9. The oligo(dC) and oligo(rA) backgrounds were used, as both sequences are thought to have minimal secondary structure[32, 43]. A lack of secondary structure in our model system is advantageous, because any current differences observed between oligonucleotides can be attributed to the nucleobase sequences, rather than structural differences.

For the oligo$(dC)_{30}$ oligos, WT αHL pores showed weak discrimination between rC, rA, rU and rG (in order of increasing $I_{RES\%}$, Figure 2c). The αHL NN pores displayed improved discrimination, clearly separating the bases in the order rC, rA, rG and rU (in order of increasing $I_{RES\%}$, Figure 2d). The overall dispersion of current levels was far greater for NN pores (NN-$\Delta I_{RES\%}{}^{OVERALL}$ = 2.2 ± 0.4% and δ^{NN} = 0.3 ± 0.02%) than it was for WT pores (WT-$\Delta I_{RES\%}{}^{OVERALL}$ = 1.2 ± 0.2% and δ^{WT} = 0.1 ± 0.02%). The αHL NNY pores displayed a twofold improvement in the dispersion of $I_{RES\%}$ for the four nucleobases, compared to NN: NNY, $\Delta I_{RES\%}{}^{OVERALL}$ = 4.5 ± 0.4%, δ^{NNY} = 2.4 ± 0.2%, along with a different order of increasing $I_{RES\%}$ (rG<rA<rC<rU, Figures 2a and 2e). In the NNY pore, the tyrosines at position 113 may provide enhanced hydrogen bonding and aromatic stacking interactions with the immobilized bases[7, 44, 45].

Similar results were obtained for the second set of oligos in the oligo$(rA)_{30}$ background. Again, the WT αHL pore showed the weakest discrimination between rC, rA, rU and rG (in order of increasing $I_{RES\%}$, Figure 2f), with a narrow dispersion between the residual current levels: $\Delta I_{RES\%}{}^{OVERALL}$ = 1.9 ± 0.1% and δ^{WT} = 0.2 ± 0.01%. The αHL NN pore displayed

discrimination that differed from that of the WT, with the increasing order of $I_{RES\%}$ being rG, rU, rA and rC (Figure 2g), and the dispersion was again improved, $\Delta I_{RES\%}{}^{OVERALL}$ = 2.7 ± 0.2% and δ^{NN} = 0.5 ± 0.01%. The αHL NNY pores gave the best dispersion in $I_{RES\%}$ with a similar pattern between the four nucleobases as seen for the oligo(dC) background (rG, rA, rC and rU, in order of increasing $I_{RES\%}$, Figures 2b and 2h).

The voltage dependencies of the current blocks caused by the four RNA bases within the $oligo(dC)_{30}$ and $oligo(rA)_{30}$ chains were also examined (Figures S5, S6 and S7; Tables S3 and S4). We observed a similar trend to that seen in the homopolymer data, in which an increased applied positive potential gave higher $I_{RES\%}$ levels for the DNA and RNA oligos across the three pores (Figure S5). For the NN and NNY pores, the four bases within the RNA oligo(rA) gave a higher $\Delta I_{RES\%}{}^{OVERALL}$ (between +100 and +140 mV) than that observed for the DNA oligo(dC): eg. at +120 mV, within the oligo(rA) chain, NNY-$\Delta I_{RES\%}{}^{OVERALL}$ = 4.9 ± 0.3%, δ^{NNY} = 4.2 ± 0.2% (n = 3); at +120 mV, within the oligo(dC) chain, NNY-$\Delta I_{RES\%}{}^{OVERALL}$ = 2.8 ± 0.3%, δ^{NNY} = 0.7 ± 0.02% (n = 3) (Table S4). In addition, at each applied potential, we consistently observed lower $I_{RES\%}$ levels for the four bases within the RNA background $oligo(rA)_{30}$ in comparison to the DNA background $oligo(dC)_{30}$.

Examining modified ribobases in ssDNA and ssRNA oligonucleotides

Cellular RNAs contain more than a hundred different base modifications at thousands of sites. These RNA modifications are dynamic and have critical regulatory roles[31]. The internal N^6-methyladenosine (m^6A) modification in messenger RNA (mRNA) is one of the most abundant modifications in higher eukaryotes, and is present at 3 to 5 sites on average per mRNA. The inability to carry out this modification leads to apoptosis. 5-Methylcytosine (m^5C) is also widespread in cellular RNAs. While epigenetic DNA methylation has been extensively studied, the precise location of m^5C in RNA remains to be elucidated. The deamination of adenosine (rA) to inosine (rI) is another important modification, which occurs in the editing of mRNA[46, 47]. rI base is read as rG by the ribosome, leading to amino acid substitution with functional implications[48].

We therefore attempted to distinguish the riboforms of I, m^6A and m^5C from rG, rA, rC and rU at position-9 in ssDNA and ssRNA oligonucleotides by using the αHL NNY pore. The NNY pore had given superior discrimination between the standard bases (rG, rA, rC and rU) at recognition site R_1 based on the increased $\Delta I_{RES\%}{}^{OVERALL}$ values observed in the homopolymeric strands oligo(dC) and oligo(rA). The addition of the rI oligonucleotide to the standard mixture of four gave: $\Delta I_{RES\%}{}^{OVERALL}$ = 5.8 ± 0.4%, δ = 2.3 ± 0.2% in the oligo(dC) background; and $\Delta I_{RES\%}{}^{OVERALL}$ = 2.6 ± 0.2% and δ = 0.2 ± 0.04% in the oligo(rA) background. The difference in residual current levels between the rI oligo and oligo(rA), $\Delta I_{RES\%}{}^{rI\text{-}rA}$, was +0.9 ± 0.1% (n = 3) (Figure 3a–c, Tables S5 and S6). These results were striking, given the small chemical difference between the two bases.

m^6A oligonucleotides also produced distinct current blocks (Figure 3a,d). The $\Delta I_{RES\%}{}^{m6A\text{-}rA}$ between m^6A and rA was −0.5 ± 0.1% (n = 3) in oligo(rA) with a similar dispersion of $\Delta I_{RES\%}{}^{OVERALL}$ to that observed with rI: $\Delta I_{RES\%}{}^{OVERALL}$ = 2.5 ± 0.1% and δ = 0.1 ± 0.04%. m^5C could also be identified in the same manner (Figure 3a,e, Table S6): in oligo(rA), $\Delta I_{RES\%}{}^{m5C\text{-}rA}$ = +1.4 ± 0.4% (n = 3), $\Delta I_{RES\%}{}^{OVERALL}$ = 3.2 ± 0.2% and δ = 0.3 ± 0.02%. While the molecular bases for the small differences in $I_{RES\%}$ levels for the modified bases are unclear, the observed current levels are distinct from all other bases and each other, demonstrating the possibility of using the αHL nanopore for mapping modified bases in the transcriptome. The $\Delta I_{RES\%}$ patterns for the WT, NN and NNY αHL pores (Figures 4a and 4b; Tables S5 and S6) demonstrate the potential to identify all seven bases

within the oligo(dC) and oligo(rA) backgrounds. The αHL NNY pore displays a large dispersion of current levels for the seven bases in oligo(dC) and a more modest dispersion in oligo(rA), showing that base identification is modulated by the background[7].

Recognition of individual nucleotides in a heteropolymeric background

The ability to sequence ssRNA would require the recognition of individual nucleotides in heteropolymeric backgrounds. To examine discrimination within a heteropolymer, we tested the R_1 site in the NNY pore. The sequence we used (Figure 5) did not contain secondary structures, such as hairpins, as predicted by the mfold algorithm[49, 50]. At +200 mV, all four bases at position 9 were recognized with the same order of $I_{RES\%}$ (rG, rA, rC and rU, Figure 5a), as seen in the homopolymeric backgrounds (Figure 2 and 3). The dispersion of the current levels in the histogram ($\Delta I_{RES\%}{}^{OVERALL} = 1.7 \pm 0.3\%$, $\delta = 0.2 \pm 0.06\%$) were also similar to those seen in the homopolymeric backgrounds, suggesting that the different current levels were indeed directly due to the nucleobases changes, rather than changes in secondary structure. The voltage dependence of the currents arising from the four bases again displayed similar characteristics to the results obtained in the homopolymer backgrounds, with higher positive potentials resulting in higher $I_{RES\%}$ levels (Figure 5b, Table S7). The span between rC and rA (Figure 5c) in the residual current histogram ($\Delta I_{RES\%}{}^{rC\text{-}rA} = +0.7 \pm 0.1\%$) was similar to that seen in the oligo(rA) homopolymeric background, $\Delta I_{RES\%}{}^{rC\text{-}rA} = +0.8 \pm 0.2\%$ (Figure 2h, Table S4).

Conclusions

We have shown that individual RNA bases can be identified in immobilized DNA and RNA strands. By using the αHL NNY mutant pore, which has superior nucleobase discrimination properties, we were able to distinguish between the standard bases (rG, rA, rC, and rU) and the modified bases rI, m^6A and m^5C. The ultra-rapid sequencing of RNA will ultimately require an active process to control movement of the nucleic acid through the pore[14, 15, 17, 19, 20, 51]. Hence, it will be necessary to combine αHL or an alternative protein nanopore with a processive RNA translocating enzyme, such as a exoribonuclease[52] or a reverse transcriptase[53], to ratchet RNA through the pore at a speed at which base identification is feasible. For a functional device, a ~10 ms measurement per base is feasible, and would reduce the sequencing time for a human genome to less than a day with a 104-pore array device. The data acquisition rate and signal filtering required for a 10 ms measurement time would allow the separation of the currents levels seen with NN and NNY in the present work. Additionally, integration of the protein nanopore with a processing enzyme will support the unfolding of translocating RNA molecules, thereby overcoming a major concern in the analysis of native RNA, which can form complex structures mediated by the base-pairing[54, 55]. For high throughput sequencing, thousands of pores must be incorporated into arrays and viable approaches towards this end are under development[1, 56].

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

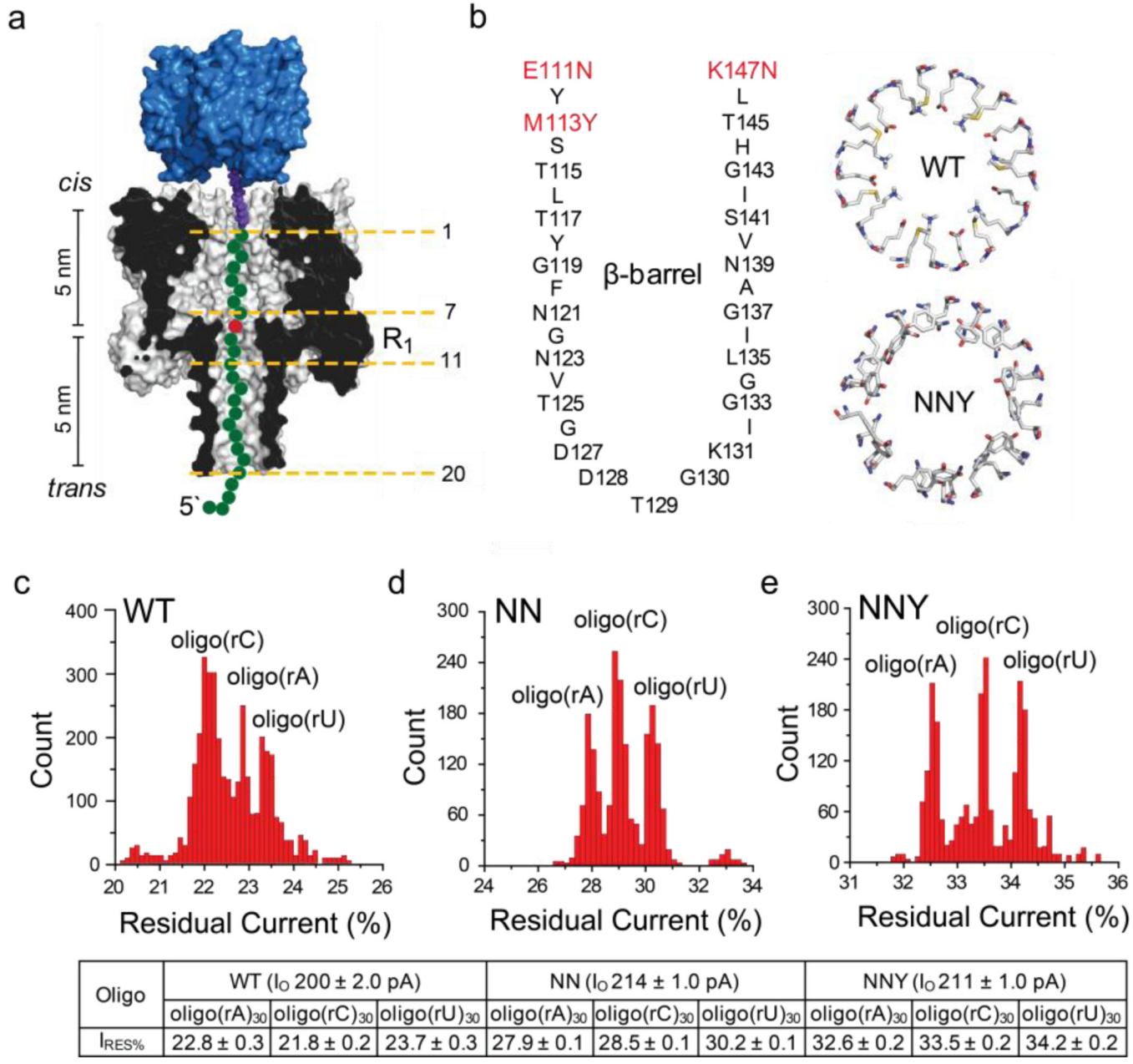
This work was supported by grants from the National Institutes of Health and Oxford Nanopore Technologies. The authors thank E. Mikhailova for the preparation of proteins, and D. Stoddart and A. Heron for useful discussions.

REFERENCES

1. Dekker C. Nat Nanotechnol. 2007; 2(4):209–215. [PubMed: 18654264]

2. Bayley H, Cremer PS. Nature. 2001; 413(6852):226–230. [PubMed: 11557992]
3. Kasianowicz JJ, Brandin E, Branton D, Deamer DW. Proc Natl Acad Sci U S A. 1996; 93(24): 13770–13773. [PubMed: 8943010]
4. Ashkenasy N, Sanchez-Quesada J, Bayley H, Ghadiri MR. Angew Chem Int Ed Engl. 2005; 44(9): 1401–1404. [PubMed: 15666419]
5. Stoddart D, Heron AJ, Mikhailova E, Maglia G, Bayley H. Proc Natl Acad Sci U S A. 2009; 106(19):7702–7707. [PubMed: 19380741]
6. Purnell RF, Schmidt JJ. ACS Nano. 2009; 3(9):2533–2538. [PubMed: 19694456]
7. Stoddart D, Maglia G, Mikhailova E, Heron AJ, Bayley H. Angew Chem Int Ed Engl. 2010; 49(3): 556–559. [PubMed: 20014084]
8. Stoddart D, Heron AJ, Klingelhoefer J, Mikhailova E, Maglia G, Bayley H. Nano Lett. 2010; 10(9): 3633–3637. [PubMed: 20704324]
9. Wallace EV, Stoddart D, Heron AJ, Mikhailova E, Maglia G, Donohoe TJ, Bayley H. Chem Commun (Camb). 2010; 46(43):8195–8197. [PubMed: 20927439]
10. Manrao EA, Derrington IM, Pavlenok M, Niederweis M, Gundlach JH. PLoS One. 2011; 6(10):e25723. [PubMed: 21991340]
11. Schibel AEP, An N, Jin Q, Fleming AM, Burrows CJ, White HS. J Am Chem Soc. 2010; 132(51): 17992–17995. [PubMed: 21138270]
12. An N, Fleming AM, White HS, Burrows CJ. Proceedings of the National Academy of Sciences. 2012; 109(29):11504–11509.
13. Meller A, Nivon L, Brandin E, Golovchenko J, Branton D. Proc Natl Acad Sci U S A. 2000; 97(3): 1079–1084. [PubMed: 10655487]
14. Cockroft SL, Chu J, Amorin M, Ghadiri MR. J Am Chem Soc. 2008; 130(3):818–820. [PubMed: 18166054]
15. Lieberman KR, Cherf GM, Doody MJ, Olasagasti F, Kolodji Y, Akeson M. J Am Chem Soc. 2010; 132(50):17961–17972. [PubMed: 21121604]
16. Olasagasti F, Lieberman KR, Benner S, Cherf GM, Dahl JM, Deamer DW, Akeson M. Nature Nanotechnology. 2010; 5(11):798–806.
17. Chu J, Gonzalez-Lopez M, Cockroft SL, Amorin M, Ghadiri MR. Angew Chem Int Ed Engl. 2010; 49(52):10106–10109. [PubMed: 21105031]
18. Hayden EC. Nature. 2012
19. Cherf GM, Lieberman KR, Rashid H, Lam CE, Karplus K, Akeson M. Nat Biotechnol. 2012; 30(4):344–348. [PubMed: 22334048]
20. Manrao EA, Derrington IM, Laszlo AH, Langford KW, Hopper MK, Gillgren N, Pavlenok M, Niederweis M, Gundlach JH. Nat Biotechnol. 2012; 30(4):349–353. [PubMed: 22446694]
21. Nilsen TW, Graveley BR. Nature. 2010; 463(7280):457–463. [PubMed: 20110989]
22. Ozsolak F, Milos PM. Nat Rev Genet. 2011; 12(2):87–98. [PubMed: 21191423]
23. Pritchard CC, Cheng HH, Tewari M. Nat Rev Genet. 2012; 13(5):358–369. [PubMed: 22510765]
24. Mitelman F, Johansson B, Mertens F. Nat Rev Cancer. 2007; 7(4):233–245. [PubMed: 17361217]
25. Maher CA, Kumar-Sinha C, Cao X, Kalyana-Sundaram S, Han B, Jing X, Sam L, Barrette T, Palanisamy N, Chinnaiyan AM. Nature. 2009; 458(7234):97–101. [PubMed: 19136943]
26. Metzker ML. Nat Rev Genet. 2010; 11(1):31–46. [PubMed: 19997069]
27. Wang Z, Gerstein M, Snyder M. Nat Rev Genet. 2009; 10(1):57–63. [PubMed: 19015660]
28. Ghildiyal M, Zamore PD. Nat Rev Genet. 2009; 10(2):94–108. [PubMed: 19148191]
29. Davidson BL, McCray PB Jr. Nat Rev Genet. 2011; 12(5):329–340. [PubMed: 21499294]
30. Matzke M, Matzke AJ, Kooter JM. Science. 2001; 293(5532):1080–1083. [PubMed: 11498576]
31. He C. Nat Chem Biol. 2010; 6(12):863–865. [PubMed: 21079590]
32. Akeson M, Branton D, Kasianowicz JJ, Brandin E, Deamer DW. Biophys J. 1999; 77(6):3227–3233. [PubMed: 10585944]
33. Butler TZ, Gundlach JH, Troll MA. Biophys J. 2006; 90(1):190–199. [PubMed: 16214857]
34. Butler TZ, Gundlach JH, Troll M. Biophys J. 2007; 93(9):3229–3240. [PubMed: 17675346]

35. Lin J, Kolomeisky A, Meller A. Phys Rev Lett. 2010; 104(15):158101. [PubMed: 20482020]
36. Maglia G, Restrepo MR, Mikhailova E, Bayley H. Proc Natl Acad Sci U S A. 2008; 105(50): 19720–19725. [PubMed: 19060213]
37. Rincon-Restrepo M, Mikhailova E, Bayley H, Maglia G. Nano Lett. 2011; 11(2):746–750. [PubMed: 21222450]
38. Seol Y, Skinner GM, Visscher K, Buhot A, Halperin A. Physical Review Letters. 2007; 98(15): 158103. [PubMed: 17501388]
39. Seol Y, Skinner GM, Visscher K. Physical Review Letters. 2004; 93(11):118102. [PubMed: 15447383]
40. Sauer-Budge AF, Nyamwanda JA, Lubensky DK, Branton D. Physical Review Letters. 2003; 90(23):238101. [PubMed: 12857290]
41. Mathé J, Visram H, Viasnoff V, Rabin Y, Meller A. Biophys J. 2004; 87(5):3205–3212. [PubMed: 15347593]
42. Keyser UF, Koeleman BN, Van Dorp S, Krapf D, Smeets RMM, Lemay SG, Dekker NH, Dekker C. Nature Physics. 2006; 2(7):473–477.
43. Adler BA, Grossman L, Fasman GD. Proc Natl Acad Sci U S A. 1967; 57(2):423–430. [PubMed: 16591487]
44. Jones S, Daley DT, Luscombe NM, Berman HM, Thornton JM. Nucleic Acids Res. 2001; 29(4): 943–954. [PubMed: 11160927]
45. Mackay JP, Font J, Segal DJ. Nat Struct Mol Biol. 2011; 18(3):256–261. [PubMed: 21358629]
46. Gallo A, Galardi S. RNA Biol. 2008; 5(3):135–139. [PubMed: 18758244]
47. Kawahara Y, Zinshteyn B, Sethupathy P, Iizasa H, Hatzigeorgiou AG, Nishikura K. Science. 2007; 315(5815):1137–1140. [PubMed: 17322061]
48. Bass BL. Annu Rev Biochem. 2002; 71:817–846. [PubMed: 12045112]
49. Zuker M. Nucleic Acids Res. 2003; 31(13):3406–3415. [PubMed: 12824337]
50. Ying X, Luo H, Luo J, Li W. Nucleic Acids Res. 2004; 32(Web Server issue):W150–W153. [PubMed: 15215369]
51. Olasagasti F, Lieberman KR, Benner S, Cherf GM, Dahl JM, Deamer DW, Akeson M. Nat Nanotechnol. 2010; 5(11):798–806. [PubMed: 20871614]
52. Raines RT. Chem Rev. 1998; 98(3):1045–1065. [PubMed: 11848924]
53. Kohlstaedt LA, Wang J, Friedman JM, Rice PA, Steitz TA. Science. 1992; 256(5065):1783–1790. [PubMed: 1377403]
54. Berman HM, Olson WK, Beveridge DL, Westbrook J, Gelbin A, Demeny T, Hsieh SH, Srinivasan AR, Schneider B. Biophys J. 1992; 63(3):751–759. [PubMed: 1384741]
55. Andronescu M, Bereg V, Hoos HH, Condon A. Bmc Bioinformatics. 2008; 9
56. Hall AR, Scott A, Rotem D, Mehta KK, Bayley H, Dekker C. Nat Nanotechnol. 2010; 5(12):874–877. [PubMed: 21113160]

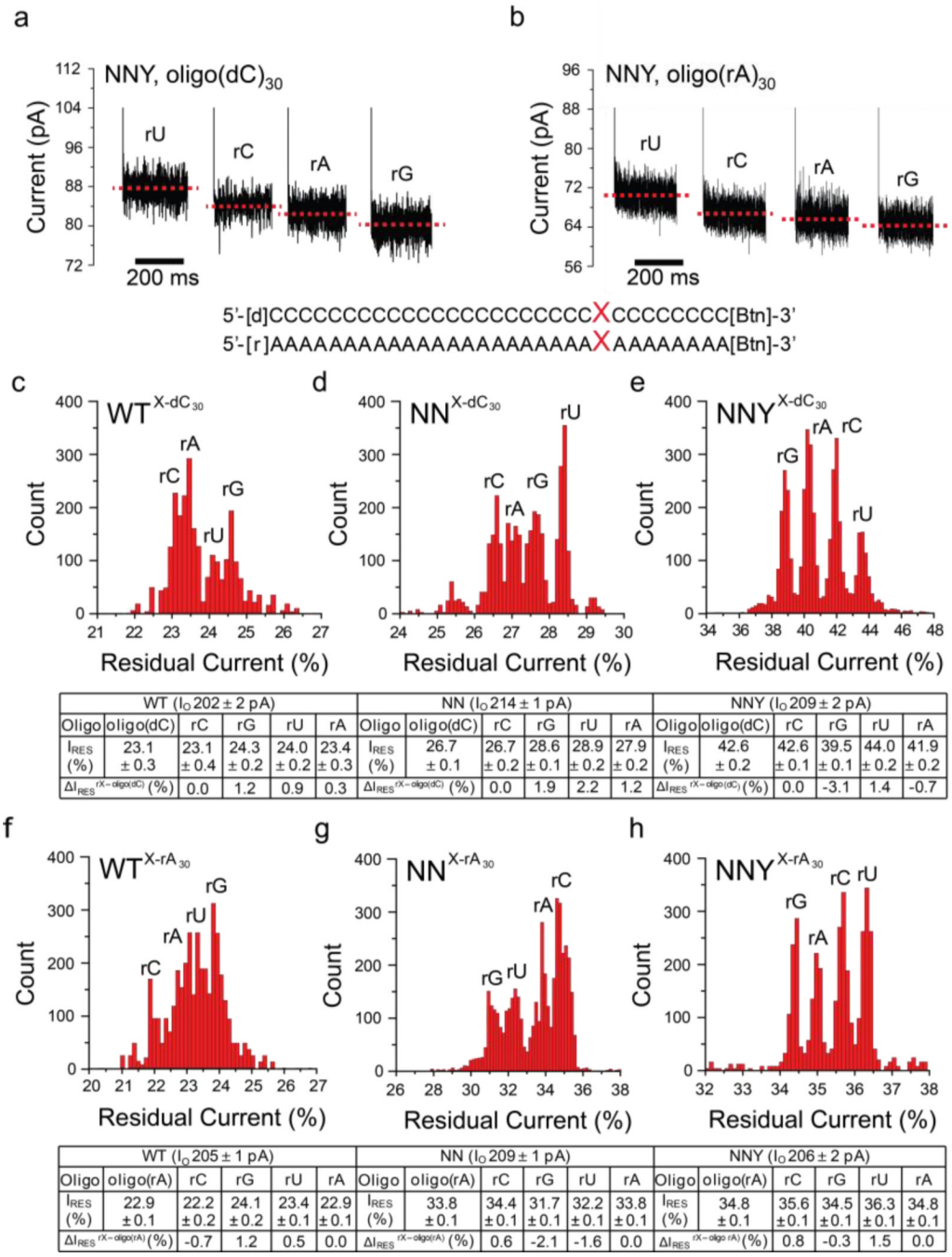


Oligo	WT (I_O 200 ± 2.0 pA)			NN (I_O 214 ± 1.0 pA)			NNY (I_O 211 ± 1.0 pA)		
	$oligo(rA)_{30}$	$oligo(rC)_{30}$	$oligo(rU)_{30}$	$oligo(rA)_{30}$	$oligo(rC)_{30}$	$oligo(rU)_{30}$	$oligo(rA)_{30}$	$oligo(rC)_{30}$	$oligo(rU)_{30}$
$I_{RES\%}$	22.8 ± 0.3	21.8 ± 0.2	23.7 ± 0.3	27.9 ± 0.1	28.5 ± 0.1	30.2 ± 0.1	32.6 ± 0.2	33.5 ± 0.2	34.2 ± 0.2

Figure 1. Interaction of homopolymeric RNA strands with αHL pores

(a) Schematic representation of a homopolymeric RNA oligonucleotide (green circles), with a single ribonucleotide substituted at position 9 (*red circle*), immobilized inside an αHL pore (*grey, cross-section*) by using a 3'-biotin-TEG (*purple*)•streptavidin (*blue*) complex. The structure of the biotin linker is provided in Figure S1. The αHL pore can be divided into 2 parts, each 5 nm in length; an upper vestibule located between the *cis* entrance and the central constriction, and a 14-stranded, transmembrane, antiparallel β barrel, located between the central constriction and *trans* entrance. The oligonucleotide bases are numbered relative to the 3'-biotin tag and position 9 was chosen for interrogation by recognition site R_1 (at the constriction). (b) The amino acid sequence of the transmembrane β barrel. The mutated residues at the top of the barrel are highlighted (*red*); these mutations enlarge the diameter of the pore. (Right, top) A view of the WT residues Glu-111, Lys-147 and Met-113

from the *cis* side of the pore. These residues form a constriction with diameter of ~13 Å. (Right, bottom) A view of the NNY residues Asn-111, Asn-147 and Tyr-113 from the *cis* side of the pore. These residues form a constriction with diameter of ~18 Å. The images were generated in Pymol. (c) Histogram of the residual current levels for RNA homopolymers oligo$(rA)_{30}$, oligo$(rC)_{30}$ and oligo$(rU)_{30}$ immobilized in the WT αHL pore. (d) Histogram of the residual current levels for the homoheptameric αHL pore formed from the mutant E111N/K147N (NN). (e) Histogram of the residual current levels for the homoheptameric αHL pore formed from the mutant E111N/K147N/M113Y (NNY). Gaussian fits were performed for each peak, and the mean value of the open pore (I_O), the residual currents ($I_{RES\%}$) and the standard deviation for each oligonucleotide are displayed in the table below the histograms.



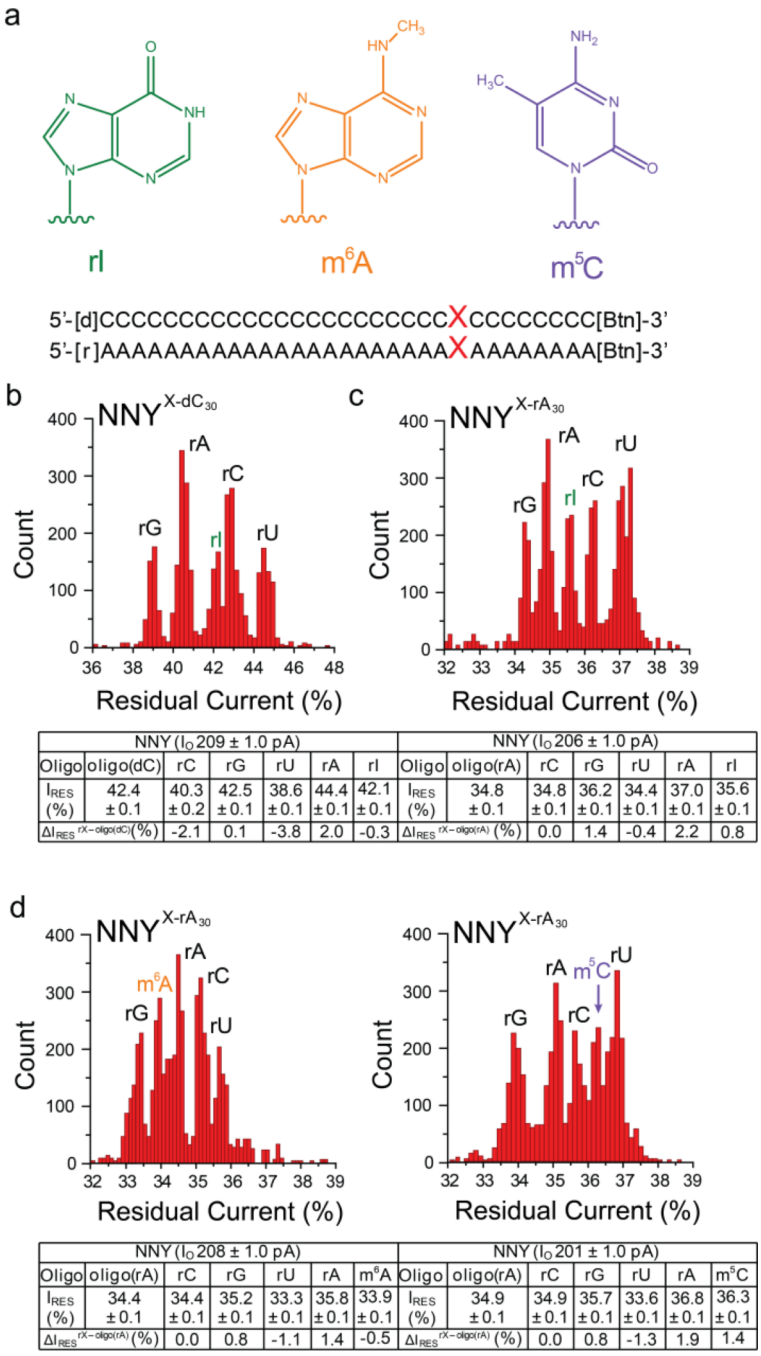
WT (I_O 202 ± 2 pA)						NN (I_O 214 ± 1 pA)						NNY (I_O 209 ± 2 pA)					
Oligo	oligo(dC)	rC	rG	rU	rA	Oligo	oligo(dC)	rC	rG	rU	rA	Oligo	oligo(dC)	rC	rG	rU	rA
I_{RES} (%)	23.1 ± 0.3	23.1 ± 0.4	24.3 ± 0.2	24.0 ± 0.2	23.4 ± 0.3	I_{RES} (%)	26.7 ± 0.1	26.7 ± 0.2	28.6 ± 0.1	28.9 ± 0.2	27.9 ± 0.2	I_{RES} (%)	42.6 ± 0.2	42.6 ± 0.1	39.5 ± 0.1	44.0 ± 0.2	41.9 ± 0.2
ΔI_{RES} rX−oligo(dC) (%)		0.0	1.2	0.9	0.3	ΔI_{RES} rX−oligo(dC) (%)		0.0	1.9	2.2	1.2	ΔI_{RES} rX−oligo(dC) (%)		0.0	-3.1	1.4	-0.7

WT (I_O 205 ± 1 pA)						NN (I_O 209 ± 1 pA)						NNY (I_O 206 ± 2 pA)					
Oligo	oligo(rA)	rC	rG	rU	rA	Oligo	oligo(rA)	rC	rG	rU	rA	Oligo	oligo(rA)	rC	rG	rU	rA
I_{RES} (%)	22.9 ± 0.1	22.2 ± 0.2	24.1 ± 0.2	23.4 ± 0.1	22.9 ± 0.1	I_{RES} (%)	33.8 ± 0.1	34.4 ± 0.1	31.7 ± 0.1	32.2 ± 0.1	33.8 ± 0.1	I_{RES} (%)	34.8 ± 0.1	35.6 ± 0.1	34.5 ± 0.1	36.3 ± 0.1	34.8 ± 0.1
ΔI_{RES} rX−oligo(rA) (%)		-0.7	1.2	0.5	0.0	ΔI_{RES} rX−oligo(rA) (%)		0.6	-2.1	-1.6	0.0	ΔI_{RES} rX−oligo rA) (%)		0.8	-0.3	1.5	0.0

Figure 2. Discrimination of individual ribobases in ssDNA and ssRNA with the WT, NN and NNY αHL pores

(a) Current traces for immobilized ssDNAs and (b) ssRNAs in the αHL NNY pore at +200 mV. (*Below*) Sequences of the oligonucleotides, biotinylated at the 3' ends. Two sets of four oligonucleotides were used, based on oligo(dC)$_{30}$ and oligo(rA)$_{30}$. Each set contained rG, rA, rC, or rU at position 9 (represented by X) relative to the biotin tag. Residual current ($I_{RES\%}$) histograms were compiled for oligonucleotide sets examined with the three αHL pores: (c) oligo(dC)$_{30}$ oligonucleotides examined with the WT pore. (d) oligo(dC)$_{30}$ oligonucleotides examined with the NN pore. (e) oligo(dC)$_{30}$ oligonucleotides examined with the NNY pore. (f) oligo(rA)$_{30}$ oligonucleotides examined with the WT pore. (g)

oligo(rA)$_{30}$ oligonucleotides examined with the NN pore. (h) oligo(rA)$_{30}$ oligonucleotides examined with the NNY pore. Experiments were conducted at least 3 times, and the results displayed in each histogram are from a typical experiment. Gaussian fits were performed for each peak, and the mean values of the open pore currents (I_O, pA), the normalized residual currents ($I_{RES\%}$) and the differences in the normalized residual currents ($\Delta I_{RES\%}$) are displayed in the tables below the histograms with their standard deviations (± S.D). $\Delta I_{RES\%}$ is defined as the difference in residual current between an rX (X= rG, rC, rA or rU) oligonucleotide and either oligo(dC)$_{30}$ or oligo(rA)$_{30}$. $\Delta I_{RES\%}{}^{rX\text{-}oligo(dC)} = I_{RES}$ for the rX oligonucleotide - I_{RES} for oligo(dC)$_{30}$ or $\Delta I_{RES\%}{}^{rX\text{-}oligo(rA)} = I_{RES\%}$ for the rX oligonucleotide - $I_{RES\%}$ for oligo(rA)$_{30}$.



NNY (I_O 209 ± 1.0 pA)							NNY (I_O 206 ± 1.0 pA)						
Oligo	oligo(dC)	rC	rG	rU	rA	rI	Oligo	oligo(rA)	rC	rG	rU	rA	rI
I_{RES} (%)	42.4 ± 0.1	40.3 ± 0.2	42.5 ± 0.1	38.6 ± 0.1	44.4 ± 0.1	42.1 ± 0.1	I_{RES} (%)	34.8 ± 0.1	34.8 ± 0.1	36.2 ± 0.1	34.4 ± 0.1	37.0 ± 0.1	35.6 ± 0.1
ΔI_{RES} rX−oligo(dC) (%)		-2.1	0.1	-3.8	2.0	-0.3	ΔI_{RES} rX−oligo(rA) (%)		0.0	1.4	-0.4	2.2	0.8

NNY (I_O 208 ± 1.0 pA)							NNY (I_O 201 ± 1.0 pA)						
Oligo	oligo(rA)	rC	rG	rU	rA	m^6A	Oligo	oligo(rA)	rC	rG	rU	rA	m^5C
I_{RES} (%)	34.4 ± 0.1	34.4 ± 0.1	35.2 ± 0.1	33.3 ± 0.1	35.8 ± 0.1	33.9 ± 0.1	I_{RES} (%)	34.9 ± 0.1	34.9 ± 0.1	35.7 ± 0.1	33.6 ± 0.1	36.8 ± 0.1	36.3 ± 0.1
ΔI_{RES} rX−oligo(rA) (%)		0.0	0.8	-1.1	1.4	-0.5	ΔI_{RES} rX−oligo(rA) (%)		0.0	0.8	-1.3	1.9	1.4

Figure 3. Discrimination of individual modified ribobases in ssDNA and ssRNA with the NNY αHL pore

(a) Chemical structures of the modified bases: rI, inosine; m^6A, N^6-methyladenosine; m^5C, 5-methylcytosine. *(Below)* Sequences of oligonucleotides biotinylated at the 3' end. Three sets of five oligo(dC)$_{30}$ or five oligo(rA)$_{30}$ oligonucleotides were used. Each set contained oligonucleotides with the four standard bases, rG, rA, rC, or rU at position 9 (represented by X) relative to the biotin tag, as well as an oligonucleotide with a modified base, one of rI, m^6A or m^5C. Histograms of the residual currents ($I_{RES\%}$) from oligonucleotide sets examined with the NNY pore were constructed: (b) the oligo(dC)$_{30}$ set containing rI; (c) the oligo(rA)$_{30}$ set containing rI; (d) the oligo(rA)$_{30}$ set containing m^6A; (e) the oligo(rA)$_{30}$ set

containing m^5C oligo$(rA)_{30}$. Experiments were conducted at least 3 times, and the results displayed in each panel are from a typical experiment. Gaussian fits were performed for each peak, and the mean values of the open pore currents (I_O, pA), the normalized residual currents ($I_{RES\%}$) and the differences between the normalized residual currents ($\Delta I_{RES\%}$) are displayed in the table below the histograms with their standard deviations (± S.D). $\Delta I_{RES\%}$ is defined as the residual current between an rX (X= rG, rC, rA, rU, rI, m^6A or m^5C) oligonucleotide and either oligo$(dC)_{30}$ or oligo$(rA)_{30}$. $\Delta I_{RES\%}{}^{rX\text{-}oligo(dC)}$ = $I_{RES\%}$ of the rX oligonucleotide - $I_{RES\%}$ of oligo$(dC)_{30}$ and $\Delta I_{RES\%}{}^{rX\text{-}oligo(rA)}$ = $I_{RES\%}$ of the rX oligonucleotide - $I_{RES\%}$ of oligo$(rA)_{30}$. The experiments were conducted at +200 mV.

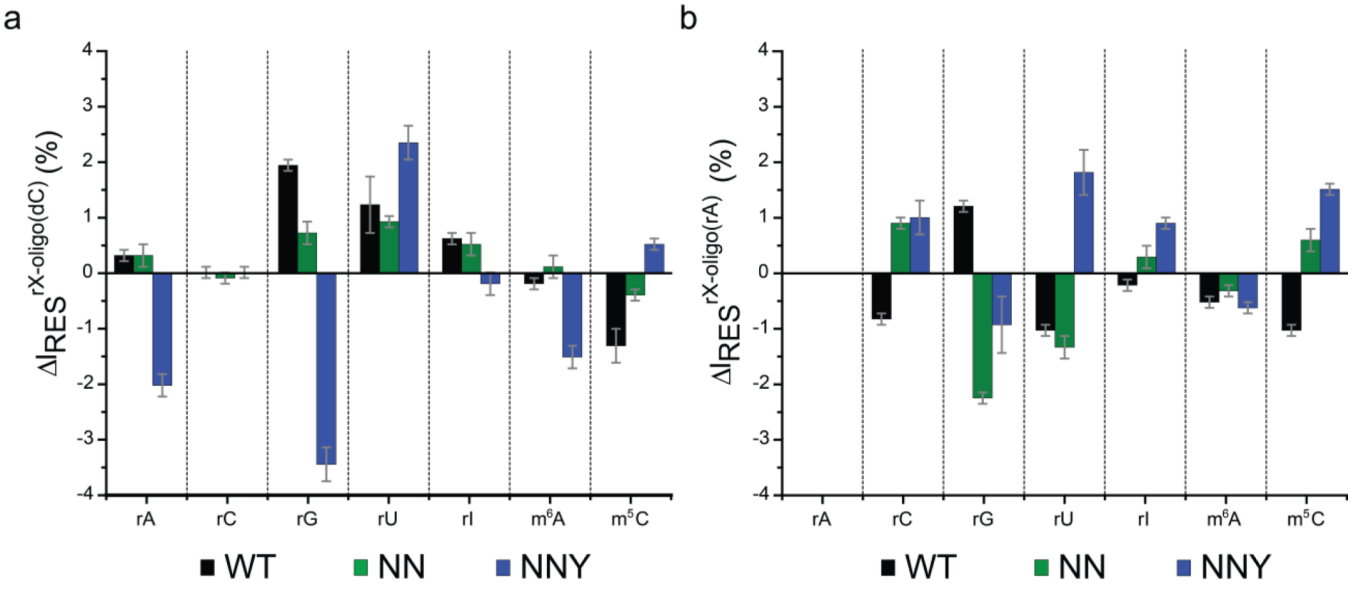
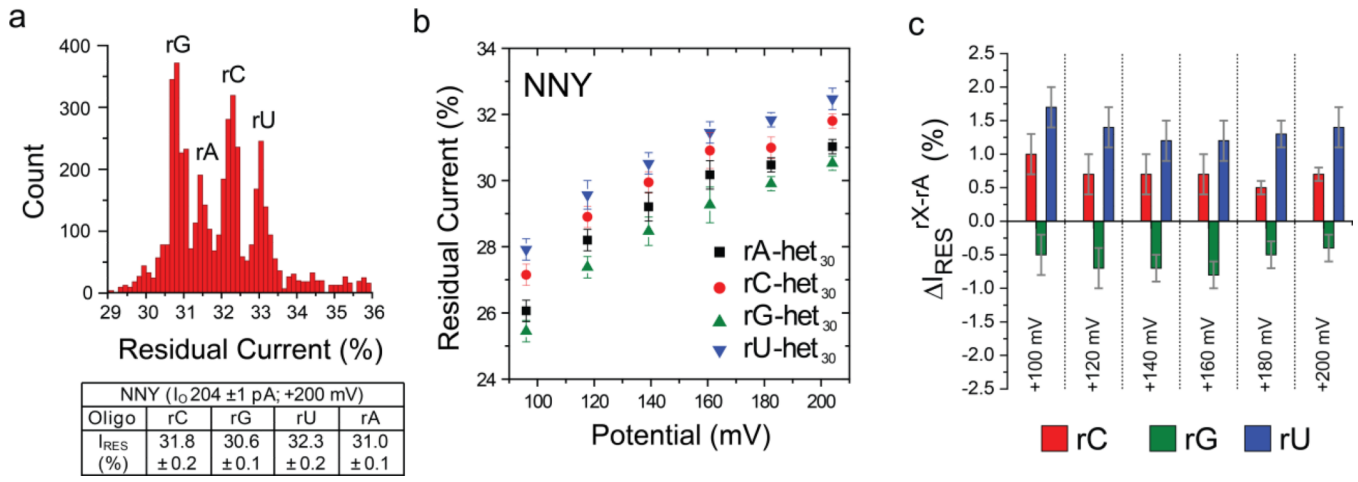


Figure 4. Plots of residual current difference for all seven bases identified in ssDNA and ssRNA oligonucleotides

(a) $\Delta I_{RES\%}$ for each base substituted at position 9 of oligo(dC)$_{30}$ for WT (*black bars*), NN (*green bars*) and NNY (*blue bars*) αHL pores. $\Delta I_{RES\%}{}^{rX\text{-}oligo(dC)} = I_{RES\%}$ of the rX oligonucleotide - $I_{RES\%}$ of oligo(dC)$_{30}$. X = rG, rA, rC, rU, rI, m^{6}A or m^{5}C. (b) $\Delta I_{RES\%}$ for each base substituted at position 9 of oligo(rA)$_{30}$ for WT (*black bars*), NN (*green bars*) and NNY (*blue bars*) αHL pores. $\Delta I_{RES\%}{}^{rX\text{-}oligo(rA)} = I_{RES\%}$ of the rX oligonucleotide - $I_{RES\%}$ of oligo(rA)$_{30}$. X = rG, rA, rC, rU, rI, m^{6}A or m^{5}C. The errors given are standard deviations from three experiments (Table S5 and S6).



NNY (I_O 204 ±1 pA; +200 mV)				
Oligo	rC	rG	rU	rA
I_{RES} (%)	31.8 ±0.2	30.6 ±0.1	32.3 ±0.2	31.0 ±0.1

Figure 5. Discrimination of individual ribobases in heteropolymeric ssRNA with the NNY αHL pore

(a) Histogram of the residual currents ($I_{RES\%}$) from a heteropolymeric ssRNA set (sequence 5'-[r]UAGCUAAACCGAUAGCUUCAGXCAUGUAAC[Btn]-3') examined with the NNY pore. The four 3'-biotinylated RNA strands differed only at position 9 as shown. Experiments were conducted at least 3 times, and the panel displays the results from a typical experiment. Gaussian fits were performed for each peak, and the mean values of the open pore currents (I_O, pA) and the normalized residual currents ($I_{RES\%}$) are displayed in the table below the histogram. (b) Plots of $I_{RES\%}$ versus the applied potential for the bases at position 9: rG, rA, rC and rU. (c) Plots of residual current difference, $\Delta I_{RES\%}$, for each base. $\Delta I_{RES\%}{}^{rX\text{-}rA}$ = $I_{RES\%}$ of the rX-het$_{30}$ oligonucleotide - $I_{RES\%}$ of rA-het$_{30}$ oligonucleotide. The errors given are standard deviations from three independent experiments (Table S7).