Published in final edited form as:
*Nano Lett*. 2010 September 8; 10(9): 3633–3637. doi:10.1021/nl101955a.

# Nucleobase recognition in ssDNA at the central constriction of the αhemolysin pore

**David Stoddart**[†], **Andrew J. Heron**[†], **Jochen Klingelhoefer**[‡], **Ellina Mikhailova**[†], **Giovanni Maglia**[†], and **Hagan Bayley**[†,*]

[†]Department of Chemistry, University of Oxford, Oxford, OX1 3TA, United Kingdom

[‡]Department of Biochemistry, University of Oxford, Oxford, OX1 3QU, United Kingdom

## Abstract

Nanopores are under investigation for single-molecule DNA sequencing. The α-hemolysin (αHL) protein nanopore contains three recognition points capable of nucleobase discrimination in individual immobilized ssDNA molecules. We have modified the recognition point $R_1$ by extensive mutagenesis of residue 113. Amino acids that provide an energy barrier to ion flow (e.g. bulky or hydrophobic residues) strengthen base identification, while amino acids that lower the barrier weaken it. Amino acids with related side chains produce similar patterns of nucleobase recognition providing a rationale for the redesign of recognition points.

### Keywords

DNA sequencing; nanopore; protein engineering; single-nucleobase discrimination

Single-molecule nanopore sequencing is envisioned as a means for the cheap and rapid electrical readout of unamplified DNA sequences, which will surpass the specifications of the "$1000 genome".1 In one proposed approach to nanopore sequencing, an applied potential is used to drive both a DNA molecule and an ionic current through a pore. As a DNA base passes a recognition point within the pore, the ionic current is modulated in a base-dependent manner, and the sequence is read from the current signature.2 A $10^4$-nanopore device operating at 100 bases per second per pore (see below) could determine a human genome sequence with high-coverage in under a day, while a $10^6$-nanopore device could offer a 15 min genome.3 The sequencing of unamplified genomic DNA would allow the direct detection of epigenetic markers such as methylated cytosine bases.4

Both protein nanopores and solid-state nanopores are being investigated for nanopore sequencing.1 Protein nanopores such as the heptameric α-hemolysin (αHL) pore are advantageous because they can be precisely manipulated by genetic and chemical modification.5 Under the applied potentials required to thread ssDNA and translocate it through the wild-type (WT)-αHL pore, the DNA moves too quickly for individual bases to be identified, and means are being investigated to reduce the translocation speed. For example, a polymerase can be used to ratchet DNA through the αHL pore,6 and might ultimately operate at up to ~100 bases per second. In a stopgap approach, nucleotide

[*]To whom correspondence may be addressed: hagan.bayley@chem.ox.ac.uk, Hagan Bayley, PhD, Department of Chemistry, University of Oxford, Chemistry Research Laboratory, Mansfield Road, Oxford, OX1 3TA, England, UK, Tel: +44 1865 285101, Fax: +44 1865 275708.

**Supporting Information Available**. Details of experimental procedures, oligonucleotide sequences and data displayed in Figures 2, 3 and 4 are available free of charge via the internet at http://pubs.acs.org.

recognition is being studied in DNA strands immobilized within the αHL pore through a terminal hairpin7 or a biotin•streptavidin linkage.8–10 This approach mimics the pauses between enzymatic base additions, during which the DNA strand is held within the pore lumen, stretched in the applied potential.

The immobilization of a DNA strand within the αHL pore is characterized by a step decrease in the open-pore current level ($I_O$) to a steady blocked-pore level ($I_B$). In this study, we quote the blocked-pore level obtained from many $I_B$ values compiled into a histogram as the residual current $I_{RES}$, expressed as a percentage of the open-pore current: $I_{RES} = (I_B/I_O) \times 100$. The prolonged observation time provided by immobilization, allows small differences in $I_{RES}$ associated with the four DNA nucleobases to be detected, because noise in the signal can be reduced by stringent filtration.

The WT-αHL pore contains three nucleobase recognition points, termed $R_1$, $R_2$ and $R_3$ and located in the 5 nm-long β barrel, which are capable of discriminating between individual bases in immobilized ssDNA molecules.8–10 It may be advantageous to use a pore with two recognition points for DNA sequencing, because they can provide additional information about each base.10 However, more than two reading heads would be impracticable, as it is unlikely that the 64 current levels that would be produced by a three reading-head pore could be resolved above the electrical noise in the system.10 Therefore, for superior base recognition it would be advantageous to augment one or two recognition points in the αHL pore and remove superfluous recognition points.

In the present work, we focus on $R_1$ and use site-directed mutagenesis to define the amino acid side chains that modulate base recognition in homoheptameric pores. We find that different classes of side chains (basic, hydrophobic and aromatic) exhibit distinct patterns of base recognition. Importantly, as desired, recognition can be either eliminated or enhanced by this approach.

## Effects of amino acid substitutions at position 113 on residual current levels

Recognition point $R_1$ is located at the central constriction of the αHL pore, and in the WT pore recognizes bases at positions 8 to 11 in an immobilized DNA strand with a peak at position 9 (Figure 1a). $R_2$ is located near the middle of the β barrel and recognizes bases 13 to 15 (peak at position 14) and $R_3$ is located near the *trans* entrance of the pore and recognizes bases 16 to 20 (peak at position 17) (Figure 1a).9 To alter the base recognition properties of $R_1$, amino acid substitutions were made at position 113, which along with residues 111 and 147, comprises the central constriction of the pore.11 It had been observed previously that “opening” the central constriction by replacing the charged Glu-111 and Lys-147 residues with smaller, neutral asparagines (Figure 1bc) increases the residual current ($I_{RES}$) during blockades by poly(dC) oligonucleotides immobilized in αHL pores by a 3' biotin•streptavidin linkage.9 Although the E111N/K147N (NN) mutation increases $I_{RES}$, under the conditions used for DNA capture (+160 mV), the $I_O$ level is similar to that of the WT pore.

By using the same immobilization technique, we have now tested fifteen E111N/K147N/M113X (NNX) mutant homoheptameric pores with poly(dC) oligonucleotides that contain single nucleobase substitutions at position 9 relative to the 3' biotin-TEG tag (Figure 1d, Figure S1), a position which is known to occupy $R_1$ when oligonucleotides are immobilized in the pore.9 We found that an increase in the volume of the amino acid side chain at position 113 does not have a predictable effect on $I_O$, under the conditions used for DNA capture (Figure 2a, Table S1). However, with the exception of the pronounced effects of Lys

and Arg, large amino acids produced a modest decrease in $I_{RES}$ during poly(dC) blockades (Figure 2b, Table S1). Presumably, in the presence of ssDNA, an increase in the volume of the side chains at position 113 significantly reduces the volume available for hydrated ions at the central constriction, and thereby reduces ion flow. It has also been proposed that the presence of DNA within the αHL pore heightens the electrostatic self-energy barrier for ion movement,12 and this effect might be accentuated by bulky side chains. It should be noted that $I_{RES}$ values for ssDNA immobilized with streptavidin are significantly greater than $I_{RES}$ values for freely translocating DNA, suggesting that the DNA is more stretched in the former case.

Positively charged residues at position 113 (Arg or Lys) have a dramatic effect on $I_{RES}$ (Figure 2b). It is possible that the side chains of Arg or Lys extend and clamp onto the DNA, severely reducing the solvent volume and hence the current flow at the constriction.13 Conceivably, partial collapse of the barrel onto the DNA strand occurs in this situation, which would also reduce $I_{RES}$.

## Effects of amino acid substitutions at position 113 on single nucleotide recognition

We next investigated whether altering the volume of the amino acid side chains at position 113 altered discrimination between single nucleobases at $R_1$. We tested the base recognition properties of $R_1$, for each of the mutant pores, with four oligonucleotides (Figure 1d). $I_{RES}$ values for each oligonucleotide were determined from histograms of individual oligonucleotide current blockades (Table S1, Figure S2). To determine how well a particular mutant could discriminate between single nucleobases, two parameters were used. First, the overall dispersion was measured (the difference between the most widely dispersed residual current levels in the histogram: $\Delta I_{RES}^{OVERALL}$, Figure S3) and plotted against the volume of the amino acid at 113 (Figure 3a, Table S1). In general, larger side chains at position 113 produce an increase in $\Delta I_{RES}^{OVERALL}$ and therefore improve nucleobase discrimination. Second, the product of the three sequential differences between each of the four residual current levels (one level for each oligonucleotide) was calculated (δ, Figure S3), as a second measure of the ability of a pore to discriminate between the four bases. The δ values were also plotted against the volume of the amino acid at position 113 (Figure 3b, Table S1). An αHL pore that is unable to discriminate between all four nucleobases (i.e. the current levels from two or more oligonucleotides overlap) has $\delta = 0$. However, pores that are able to discriminate between all four bases have $\delta > 0$, as was the case for NN pores with Val, Leu, Ile, Tyr, Trp or Arg at position 113. The hydrophobic (Val, Leu and Ile) and bulky (Tyr and Trp) side chains provide steric barriers to ion flow, which improve nucleobase discrimination at $R_1$. The positively charged Arg residues, which might clamp onto the DNA or cause collapse of the barrel, provide strong discrimination despite the low values of $I_B$ (Figure S2).

It had previously been observed that the WT αHL pore discriminates between single nucleobases at position 9 in the order T/G, C, A, in order of increasing $I_{RES}$ (where the current levels for T and G overlap).9 Therefore, it is unlikely that nucleobase size alone determines the observed current levels (e.g. T is smaller than A, yet it produces a larger current blockade when immobilized at $R_1$, see also examples in Figure S2). An increased energy barrier at the central constriction may promote interactions between the flowing ions and the immobilized bases at $R_1$, augmenting nucleobase identification. Glycine has no side chain and the E111N/K147N/M113G (NNG) pore presents a low energy barrier to ion flow (NNG pores have high $I_{RES}$ values) and exhibits the weakest base identification properties (both $\Delta I_{RES}^{OVERALL}$ and $\delta = 0$). Interactions between the amino acids at position 113 and

the ssDNA could also influence nucleobase discrimination by causing each of the four bases to adopt different average orientations, which could in turn modulate $I_{RES}$.

The amino acid side chains were categorized according to their properties and the residual current differences between the oligonucleotides containing A, T and G (at position 9) and the oligonucleotide containing C (at position 9) were assigned to them (Figure 3c, Table S1): $\Delta I_{RES}^{X-C} = I_{RES}^{X} - I_{RES}^{C}$ (where X = A, T or G). Striking patterns are readily apparent. First, amino acids with small or uncharged hydrophilic side chains discriminate between bases poorly, Gly not at all. By contrast, amino acids with hydrophobic, basic or aromatic side chains show superior base discrimination, which varies in character between the amino acid classes: for the hydrophobic amino acids (Val, Leu, Ile): A < G < C < T (order of $I_{RES}$ value); basic amino acids (Lys, Arg): G < T < A ≤ C; aromatic amino acids (Tyr, Trp): T < G < A < C.

All three of these classes of amino acids have documented interactions with nucleobases. Hydrophobic residues are capable of forming van der Waals interactions with DNA and RNA bases,[14, 15] as supported by computational analysis.[16] Basic side chains can also interact with nucleobases, in this case by hydrogen bonding and cation-π interactions as documented for Arg[17–20] and Lys.[21] Tyr and Trp residues have the potential to form hydrogen bonds with nucleobases, via the hydroxyl group of the phenolic ring or the pyrrole NH of the indole ring, respectively.[19] They are also capable of forming stacking interactions with heteroaromatic bases in extended single-stranded polynucleotides.[22, 23] Interestingly, amino acids that are capable of forming only either hydrogen bonds (Asn, Gln, Ser and Thr) or stacking interactions with bases (Phe)[24] are poor at discriminating between nucleobases at $R_1$ (Figure 3bc. Table S1).

## Effects of amino acid substitutions on base discrimination at distal recognition points

The mutant that gave the poorest base discrimination at $R_1$ was the NNG pore (both $\Delta I_{RES}^{OVERALL}$ and δ = 0) and the mutant that gave the best recognition was NNY (largest δ value and second largest $\Delta I_{RES}^{OVERALL}$) (Figure 3). The differences in the ability to recognize nucleobases exhibited by these two pores are remarkable, given that there is only a single amino acid change, and provide evidence that it is possible to either greatly weaken or significantly improve nucleobase recognition by mutagenesis. However, it is important to know whether amino acid substitutions at one recognition point (in this case $R_1$) alter nucleobase recognition at additional recognition points (here $R_2$ and $R_3$). Therefore, we probed the entire length of the β barrel for its ability to recognize single nucleobases in NN,[9] NNY and NNG by using fourteen oligonucleotides in which a single adenine base is moved through a poly(dC) background (Table S2, Figure 4a). The difference in residual current between each adenine-containing oligonucleotide ($A_n$, where n gives the position of the adenine relative to the 3' biotin-tag) and poly(dC) was plotted against the position of the adenine base (Figure 4bc, Table S3). The pores exhibit different nucleobase discrimination properties at $R_1$. The NNY pore shows strong recognition compared to the NN pore, while the NNG pore shows weak recognition. By contrast, the $\Delta I_{RES}$ values at $R_2$ and $R_3$ for the NNG and NNY pores were similar to those seen with the NN pore. These data suggest that the effect of making amino acid substitutions at the central constriction on nucleobase discrimination is limited largely to $R_1$, i.e. it is possible to target one recognition point by site-directed mutagenesis, without grossly affecting other recognition points.

## Conclusions

Previous studies have shown that the β barrel of the αHL pore contains three nucleobase recognition points.9 To improve base identification, it will be necessary to enhance recognition at one or two of the points and eliminate the other(s).10 We have shown here that it is indeed possible to weaken (NNG) or strengthen (NNY) a recognition point ($R_1$) by targeted mutagenesis. Similar investigations could be done to alter the base discrimination power of the remaining recognition points. Our studies have been limited to the use of genetically encoded amino acids. Given the enhanced recognition properties of NNY and NNW, unnatural amino acids, especially those with aromatic side chains,25, 26 will be valuable in future attempts to improve nucleobase recognition.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

This work was supported by grants from the NIH and the European Commission's seventh Framework Programme (FP7) READNA Consortium. D.S. was supported by a BBSRC Doctoral Training Grant. J.K. was supported by the EPSRC. The authors thank Mark Sansom for his guidance with molecular modeling.

## REFERENCES

1. Branton D, Deamer DW, Marziali A, Bayley H, Benner SA, Butler T, Di Ventra M, Garaj S, Hibbs A, Huang X, Jovanovich SB, Krstic PS, Lindsay S, Ling XS, Mastrangelo CH, Meller A, Oliver JS, Pershin YV, Ramsey JM, Riehn R, Soni GV, Tabard-Cossa V, Wanunu M, Wiggin M, Schloss JA. Nature Biotechnology. 2008; 26:1146–1153.
2. Kasianowicz JJ, Brandin E, Branton D, Deamer DW. Proc.Natl.Acad.Sci.USA. 1996; 93:13770–13773. [PubMed: 8943010]
3. Bayley H. Curr Opin Chem Biol. 2006; 10:628–637. [PubMed: 17113816]
4. Clarke J, Wu H, Jayasinghe L, Patel A, Reid S, Bayley H. Nature Nanotechnology. 2009; 4:265–270.
5. Bayley H, Cremer PS. Nature. 2001; 413:226–230. [PubMed: 11557992]
6. Cockroft SL, Chu J, Amorin M, Ghadiri MR. J Am Chem Soc. 2008; 130(3):818–820. [PubMed: 18166054]
7. Ashkenasy N, Sánchez-Quesada J, Bayley H, Ghadiri MR. Angew.Chem.Int.Ed.Engl. 2005; 44:1401–1404. [PubMed: 15666419]
8. Purnell RF, Schmidt JJ. ACS Nano. 2009; 3:2533–2538. [PubMed: 19694456]
9. Stoddart D, Heron A, Mikhailova E, Maglia G, Bayley H. Proc. Natl. Acad. Sci. USA. 2009; 106:7702–7707. [PubMed: 19380741]
10. Stoddart D, Maglia G, Mikhailova E, Heron A, Bayley H. Angew Chem Int Ed. 2010; 49:556–559.
11. Song L, Hobaugh MR, Shustak C, Cheley S, Bayley H, Gouaux JE. Science. 1996; 274:1859–1865. [PubMed: 8943190]
12. Bonthuis DJ, Zhang J, Hornblower B, Mathe J, Shklovskii BI, Meller A. Phys Rev Lett. 2006; 97(12):128104. [PubMed: 17026003]
13. Cheley S, Gu L-Q, Bayley H. Chem.Biol. 2002; 9:829–838. [PubMed: 12144927]
14. Horvath MP, Schweiker VL, Bevilacqua JM, Ruggles JA, Schultz SC. Cell. 1998; 95(7):963–974. [PubMed: 9875850]
15. Liu Z, Luyten I, Bottomley MJ, Messias AC, Houngninou-Molango S, Sprangers R, Zanier K, Kramer A, Sattler M. Science. 2001; 294(5544):1098–1102. [PubMed: 11691992]
16. Jones S, Daley DT, Luscombe NM, Berman HM, Thornton JM. Nucleic Acids Res. 2001; 29(4): 943–954. [PubMed: 11160927]

17. Bochkarev A, Pfuetzner RA, Edwards AM, Frappier L. Nature. 1997; 385(6612):176–181. [PubMed: 8990123]
18. Gallivan JP, Dougherty DA. Proc Natl Acad Sci U S A. 1999; 96(17):9459–9464. [PubMed: 10449714]
19. Allers J, Shamoo Y. J Mol Biol. 2001; 311(1):75–86. [PubMed: 11469858]
20. Wang X, McLachlan J, Zamore PD, Hall TM. Cell. 2002; 110(4):501–512. [PubMed: 12202039]
21. Antson AA, Dodson EJ, Dodson G, Greaves RB, Chen X, Gollnick P. Nature. 1999; 401(6750): 235–242. [PubMed: 10499579]
22. Hu G, Gershon PD, Hodel AE, Quiocho FA. Proc.Natl.Acad.Sci.USA. 1999; 96:7149–7154. [PubMed: 10377383]
23. Raghunathan S, Kozlov AG, Lohman TM, Waksman G. Nat Struct Biol. 2000; 7(8):648–652. [PubMed: 10932248]
24. Lei M, Podell ER, Cech TR. Nat Struct Mol Biol. 2004; 11(12):1223–1229. [PubMed: 15558049]
25. Wang L, Xie J, Schultz PG. Annu Rev Biophys Biomol Struct. 2006; 35:225–249. [PubMed: 16689635]
26. Banerjee A, Mikhailova E, Cheley S, Gu L-Q, Montoya M, Nagaoka Y, Gouaux E, Bayley H. Proc. Natl. Acad. Sci. USA. 2010; 107:8165–8165-8170. [PubMed: 20400691]
27. Zamyatnin AA. Prog.Biophys.Mol.Biol. 1972; 24:107–123. [PubMed: 4566650]
28. Marti-Renom MA, Stuart AC, Fiser A, Sanchez R, Melo F, Sali A. Annu Rev Biophys Biomol Struct. 2000; 29:291–325. [PubMed: 10940251]

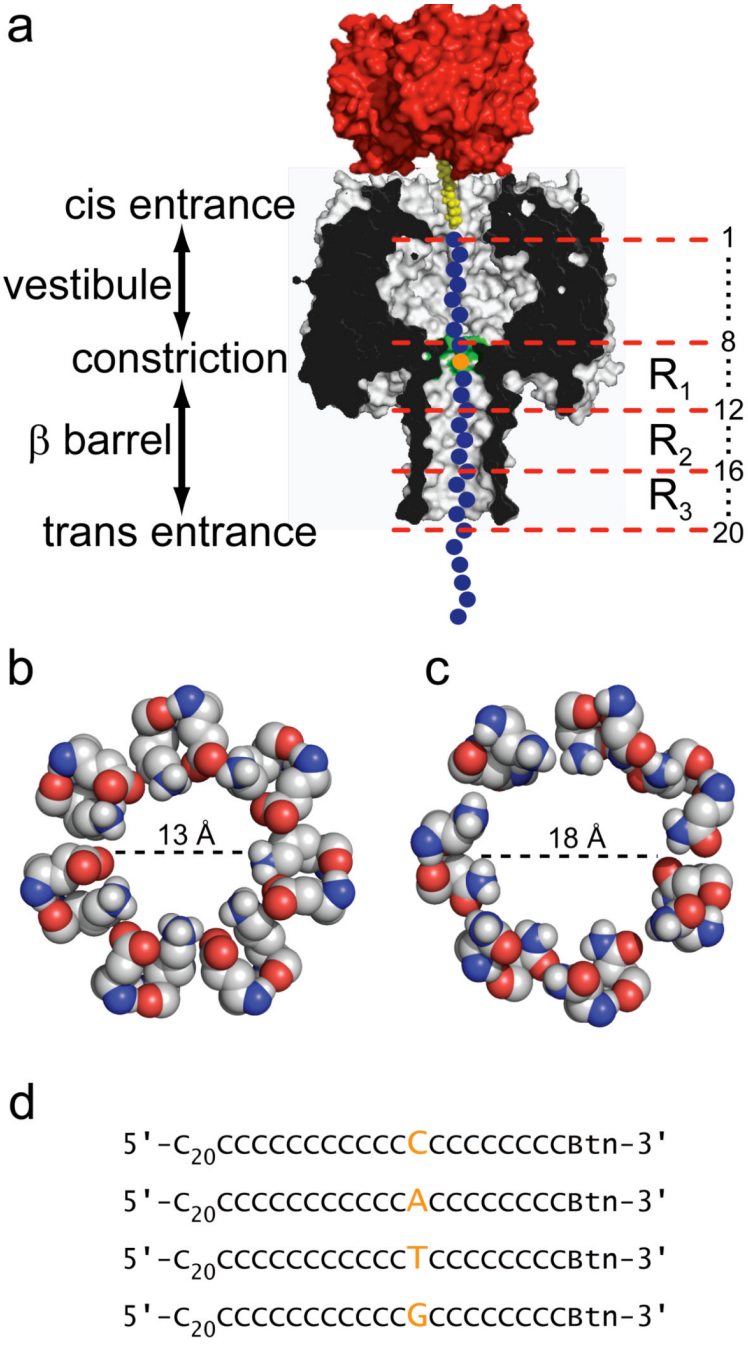


**Figure 1. Immobilizing oligonucleotides inside the αHL pore**

**a**, Schematic representation of a homopolymeric DNA oligonucleotide (blue circles), with a single nucleotide substituted at position 9 (orange circle) immobilized inside an αHL pore (grey, cross-section) through the use of a 3' biotin-TEG (yellow)•streptavidin (red) complex (Figure S1). The bases are numbered relative to the 3' biotin tag. The αHL pore can be divided into two halves, each approximately 5 nm in length: an upper cap domain containing a large vestibule located between the *cis* entrance and the constriction, and a fourteen-stranded, transmembrane, antiparallel β barrel, located between the constriction and the *trans* entrance. The constriction is formed by the Glu-111, Lys-147 and Met-113 side chains contributed by all seven subunits (green). The three nucleobase recognition points, $R_1$, $R_2$

and $R_3$, are indicated. **b**, A view of the WT residues Glu-111 and Lys-147 (atoms shown as spheres) from the cis side of the pore. These residues form a constriction with diameter of ~13 Å. **c**, The Glu-111 and Lys-147 residues were changed to Asn with MODELLER.28 The atoms in the Asn-111 and Asn-147 residues are viewed as spheres, from the *cis* side. The mutated constriction has a wider diameter, compared to that of the WT pore, which is estimated to be ~18 Å. See Supporting Information for details. **d**, The sequences of the four poly(dC) oligonucleotides, with single nucleobase substitutions at position 9, that were used to probe $R_1$.

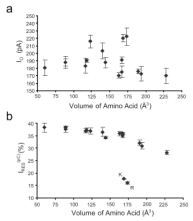


**Figure 2. The effect of amino acid substitutions at position 113 on the open pore and residual current levels**

**a**, The effect of the volume of the amino acid at position 113 on $I_O$. **b**, The effect of the volume of the amino acid at position 113 on $I_{RES}$ for a poly(dC) oligonucleotide.

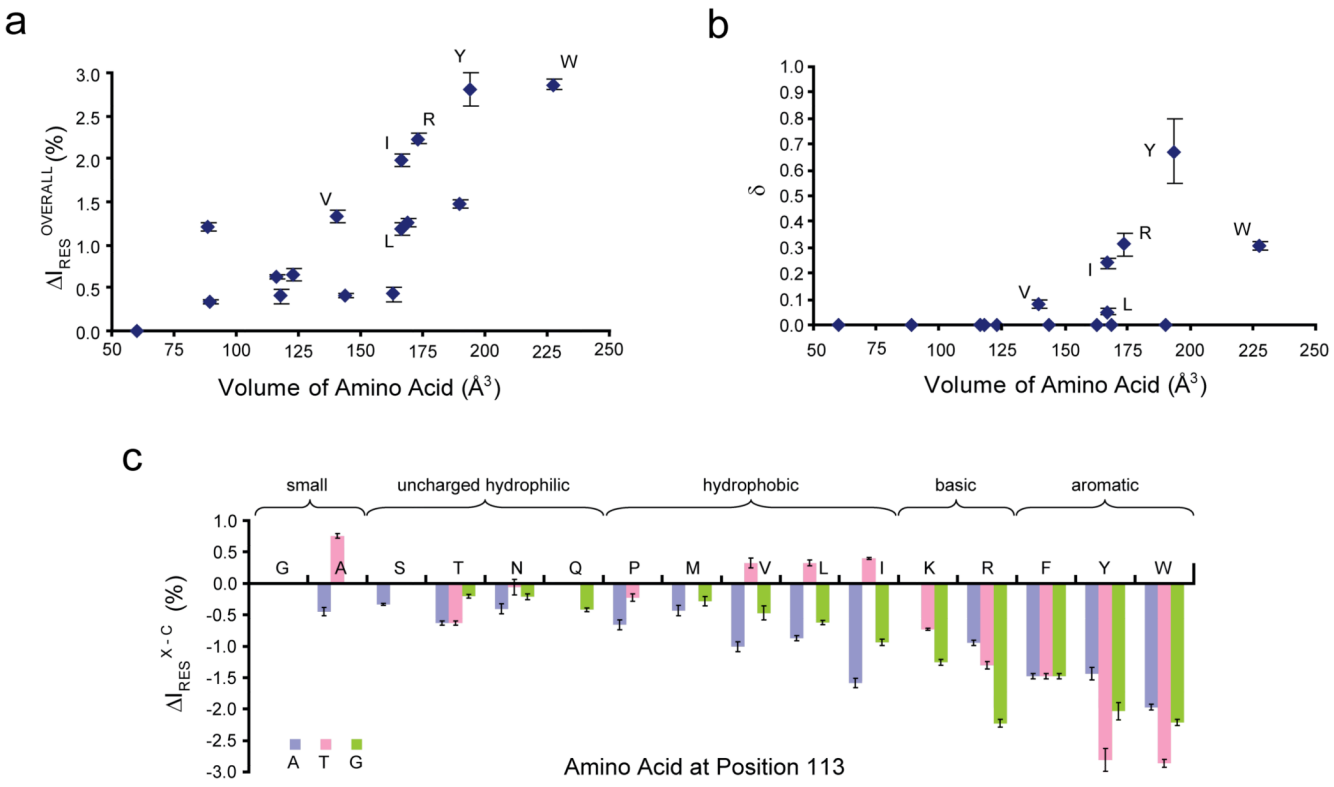


**Figure 3. The effect of amino acid substitutions at position 113 on the residual current levels for poly(dC) oligonucleotides with single nucleotide substitutions at position 9**

**a**, The relationship between the volume of the amino acid at position 113 and the difference between the two most widely dispersed current levels ($\Delta I_{RES}^{OVERALL}$, Figure S3). The mutants that discriminate between all four nucleobases are indicated. **b**, The relationship between the volume of the amino acid at position 113 and the product of the three successive differences between the four current levels in a histogram of residual currents (δ, Figure S3). The mutants that discriminate between all four oligonucleotides are indicated. **c**, The differences in residual current between the oligonucleotides containing A, T and G at position 9 and the oligonucleotide containing a C at position 9 ($\Delta I_{RES}^{X-C} = I_{RES}^{X} - I_{RES}^{C}$) are plotted for each amino acid substituted into position 113. Amino acids are categorized based on the properties of their side chains.

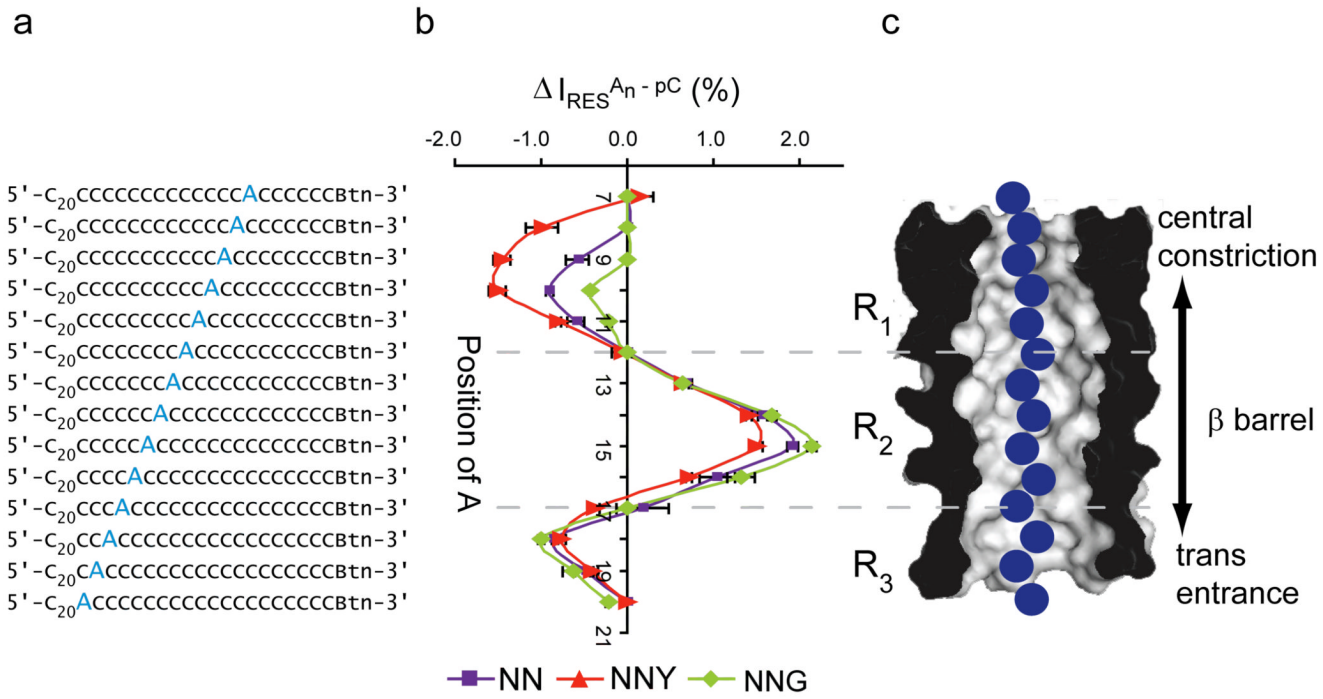


**Figure 4. The effect of amino acid substitutions on adenine recognition along the length of the β barrel**

**a**, Poly(dC) oligonucleotides containing a single adenine base (Table S2). **b**, The difference in residual current ($\Delta I_{RES}$) between blockades caused by each of the oligonucleotides shown in 'a' and poly(dC) for the αHL pores E111N/K147N (NN, purple),9 E111N/K147N/M113Y (NNY, red) and E111N/K147N/M113G (NNG, green) (see Table S3). **c**, A cross-section of the β barrel domain of the αHL pore, filled with DNA indicating the positions of recognition sites $R_1$, $R_2$ and $R_3$.