



Published in final edited form as:

*Angew Chem Int Ed Engl.* 2005 February 18; 44(9): 1401–1404.

## Recognizing a Single Base in an Individual DNA Strand: A Step Toward Nanopore DNA Sequencing\*\*

N. Ashkenasy, J. Sánchez-Quesada, and M. R. Ghadiri\*

Departments of Chemistry and Molecular Biology and The Skaggs Institute for Chemical Biology, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, CA 92037, USA, Fax: (+1)858-784-2798

H. Bayley

Department of Chemistry University of Oxford, Chemistry Research Laboratory, Mansfield Road, Oxford, OX1 3TA, England, UK

### Abstract

Functional supramolecular chemistry at the single-molecule level. Single strands of DNA can be captured inside  $\alpha$ -hemolysin transmembrane pore protein to form single-species  $\alpha$ -HL-DNA pseudorotaxanes. This process can be used to identify a single adenine nucleotide at a specific location on a strand of DNA by the characteristic reductions in the  $\alpha$ -HL ion conductance. This study suggests that  $\alpha$ -HL-mediated single-molecule DNA sequencing might be fundamentally feasible.

### Keywords

Single-molecule DNA sequencing; transmembrane pore; pseudorotaxanes; supramolecular chemistry; channel conductance

There has been an intriguing suggestion that *Staphylococcus aureus*  $\alpha$ -hemolysin ( $\alpha$ -HL), a stable heptameric transmembrane protein pore<sup>[1]</sup>, might be of use as the sensor element in a rapid pore-mediated single-molecule DNA sequencing process.<sup>[2]</sup> Although there remain several unmet requirements before such a process can be realized<sup>[2a,3]</sup>, the most fundamental concern has been whether  $\alpha$ -HL, or for that matter any other natural or man-made nanopore structure<sup>[4]</sup>, is capable of recognizing DNA with single nucleobase resolution. In this study, we have probed the nucleobase resolution capacity of  $\alpha$ -HL by threading and holding stably a given strand of ss-DNA inside the  $\alpha$ -HL pore in the form of a single-species  $\alpha$ -HL-DNA pseudorotaxane.<sup>[5]</sup> By using block copolymers of DNA and homopolymeric strands with position-specific single-nucleotide substitutions, we have found that a single adenine nucleotide at a specific location on a strand of poly d(C) can be distinguished by its characteristic effect in reducing the  $\alpha$ -HL ion conductance. The discovery that  $\alpha$ -HL can recognize ss-DNA with single nucleobase resolution strengthens the case for its utility in rapid single-molecule DNA sequencing.

It has been proposed that it might be possible to rapidly establish the sequence of an individual strand of ss-DNA as it traverses through an  $\alpha$ -HL pore by recording the distinct ion channel conductance perturbations caused by the sequential passage of each nucleotide.<sup>[2,6]</sup> Toward this goal, in an important early study, it was shown that the transient reductions in ion channel

\*\*This work was supported by a grant from the Office of Naval Research (Grant no. MURI-99, N000149910717). We thank W. S. Horne for his generous assistance in molecular modeling.

\*E-mail: ghadiri@scripps.edu.

conductance could be used to distinguish homopurine from homopyrimidine ss-nucleic acid strands.<sup>[2b,c]</sup> However, in part because of the experimental limitations imposed by the rapid rates of pore-mediated ss-nucleic acid transport, it has not yet been possible to use transient DNA transport measurements to establish the nucleobase resolution capacity of  $\alpha$ -HL.<sup>[2,3]</sup> We envisioned that this issue might be resolved by analyzing ion channel conductance reductions caused by a given strand of ss-DNA while it is captured and held stably inside an  $\alpha$ -HL pore in the form of a rotaxane or pseudorotaxane.<sup>[5]</sup>

The DNA sequences employed in our studies each have a long ss-DNA segment for threading the  $\alpha$ -HL pore and a stable terminal hairpin structure for holding the thread in a pseudorotaxane configuration (Figure 1a).<sup>[5b]</sup> Because the hairpin duplex structure is wider than the  $\alpha$ -HL internal lumen, a given DNA strand can only thread the pore with its free single-stranded terminus.<sup>[5]</sup> Therefore, depending on from which end an  $\alpha$ -HL pore is threaded and whether the hairpin structure is placed on the 3' or 5' terminus of the DNA thread, four distinct topoisomeric forms of  $\alpha$ -HL-DNA pseudorotaxanes can be formed.<sup>[5b]</sup> In this study we report on the characteristics of the topoisomers of  $\alpha$ -HL-DNA pseudorotaxanes that are prepared by threading DNA from the *cis* side (Figure 1a). In a typical experiment, the DNA sample is introduced at the *cis* side of a lipid bilayer containing a single oriented  $\alpha$ -HL pore and the threading process is initiated by applying positive transmembrane holding potentials (see supporting information for details). Formation of the  $\alpha$ -HL-DNA pseudorotaxane is signified by the reduction in ion channel conductance caused by the presence of the ss-DNA inside the pore (Figure 1b).<sup>[2,5]</sup> The DNA strand can be held stably in the pseudorotaxane configuration as long as suitable positive transmembrane holding potential is maintained (see supporting information).<sup>[5b]</sup>

Initial experiments were performed with  $\alpha$ -HL-DNA pseudorotaxanes of Poly d(A) and Poly d(C) (sequences **1-4**) to ascertain whether homopurine and homopyrimidine DNA strands held inside the  $\alpha$ -HL pore can give rise to distinct ion conductance blockades. Current versus voltage curves, measured from +10 to +170 mV, indicated that the homopurine and homopyrimidine pseudorotaxanes could be readily distinguished at applied potentials of greater than 100 mV (Figure S1). The residual current  $I_R$  at +170 mV (calculated as a percentage of the DNA free  $\alpha$ -HL current measured during the same experiment) is  $22\pm 3\%$  and  $31\pm 3\%$  for poly d(A) (sequence **1**) and poly d(C) (sequence **2**) pseudorotaxanes, respectively (Figure 2). Interestingly, topoisomeric  $\alpha$ -HL-DNA pseudorotaxanes of poly d(A) and poly d(C) that differ only in the orientation of the captured DNA strand (threading from the 3'-terminus using **1** or **3** versus threading from the 5'-terminus using **2** or **4**) gave similar characteristic "A-type" or "C-type" ion channel blockades, respectively (Figure 2).

Given that in an  $\alpha$ -HL-DNA pseudorotaxane a significant portion of the ss-DNA is held stably inside the protein's pore,<sup>[5]</sup> we sought to determine what part of the encased DNA thread could factor in the production of the observed ion conductance differences between poly d(A) and poly d(C) oligonucleotides. Accordingly, we prepared single species  $\alpha$ -HL-DNA pseudorotaxanes using four different DNA block copolymers (sequences **5-8**) each designed to place stretches of poly d(A) and poly d(C) at discrete locations along the DNA thread (Figure 2). The measured residual currents at +170 mV for each of the DNA block copolymer/ $\alpha$ -HL pseudorotaxanes fall clearly into either "A-type" or "C-type" signal categories (Figure 2). Therefore, by a simple comparative sequence analysis, it can be concluded that the nucleobase (s) recognized by  $\alpha$ -HL must be located in the region 10 to 20 nucleotides away from the edge of the hairpin structure (Figure 2). Additional studies with similar DNA block copolymer threads, but having shorter poly d(C) and poly d(A) segments, further refined the recognition site to around nucleotide 20 (data not shown).

To address whether the  $\alpha$ -HL pore structure is capable of recognizing DNA with single nucleobase resolution, we designed five poly d(C) threads (sequences **9-13**, Figure 3), each having a single deoxyadenosine at a unique position within the putative recognition site located 18 to 22 nucleotides away from the edge of the 5'-stem-loop structure. Each DNA strand was used to prepare the corresponding single species  $\alpha$ -HL-DNA pseudorotaxanes. Furthermore, each of the pseudorotaxanes was analyzed several times to obtain statistically significant numbers of events for calculating the residual current ( $I_R$  at +170 mV, Figure S2). The measured  $I_R$  distribution was roughly bimodal and could be classified into two groups corresponding to the A-type ( $I_R < 27\%$ ) and the C-type ( $I_R > 27\%$ ) events (Figure 3). A given event had current characteristics similar to either poly d(A) or poly d(C) pseudorotaxanes (Figure 1b) and did not interconvert from one type to the other during the analysis period. The data clearly show that when deoxyadenosine is at positions 18, 19, 21, or 22 (sequences **9, 10, 12, 13**) the majority of the ion conductance blockades correspond to the C-type signal. However, when a single deoxyadenosine is positioned 20 nucleotides away from the edge of the stem-loop segment (sequence **11**) in spite of being flanked on either side by long stretches of poly d(C), most observed events corresponded to the A-type signal (Figure 3). Therefore, these studies demonstrate that the  $\alpha$ -HL pore structure can recognize ss-DNA site-specifically with a single nucleobase resolution.

The above observations imply that in the  $\alpha$ -HL-DNA pseudorotaxanes the ss-DNA thread must be held in a specific or closely related conformations inside the pore to allow correct positioning of the recognized nucleotide at the putative  $\alpha$ -HL sensing element. Therefore, by estimating how the threaded DNA hairpin is pinned to the protein and the conformation of the ss-DNA inside the pore barrel it should be possible to propose where on the  $\alpha$ -HL pore the nucleobase sensing site might be located. Considering that DNA hairpins with stems longer than 8-bp can enter the vestibule and bump up against the internal pore constriction,<sup>[8a]</sup> the 12-bp stem hairpin strands used in our studies are most likely pinned to the  $\alpha$ -HL pore constriction by the edge of their stem duplex structure (Figure 1a). This view is also consistent with a study in which covalently tethered DNA molecules were used to probe the internal  $\alpha$ -HL pore structure and dimensions.<sup>[9]</sup> Therefore, in such a configuration the position of the recognized nucleotide (counting from the edge of the stem structure) should be independent of the number of stem base pairs (if  $> 8$ -bp). Consistent with the above hypothesis,  $\alpha$ -HL-DNA pseudorotaxanes prepared using hairpin DNA strands having 13-bp (strands **14** and **15**) or 10-bp (strands **16** to **18**) stem segments retained their ability to report on a single deoxyadenosine placed 20 nucleotides away from the edge of the hairpin (Figures 4, S3). Furthermore, we speculate that the estimated length of the single DNA strand within the barrel is somewhat shorter than a DNA strand with the same number of bases within B-form dsDNA. ssDNA, by comparison with dsDNA, is highly contractile and at forces below  $\sim 8$  pN would be shortened compared to B-DNA.<sup>[10]</sup> Such small forces seem reasonable taking into account that the effective charge per nucleotide within the  $\alpha$ -HL pore is only  $\sim 0.1e$  as suggested by recent studies.<sup>[11]</sup> Therefore, the above rationale and observations suggest that the  $\alpha$ -HL recognition site is most likely located near its trans opening (Figure 1a).<sup>[11]</sup>

The precision of the single nucleobase recognition is an important factor in single-molecule DNA sequencing. Careful analysis of the data describe above indicate that the precision of nucleotide sensing is not sufficiently high in the "native"  $\alpha$ -HL pore for use in DNA sequencing. For example, while the observed residual currents indicate that an adenine substitution at either positions 18, 21, and 22 (sequences **9, 12, 13**) is not recognized (Figure 3), the discriminating power between A and C at position 19 (using sequence **10**) is noticeably less pronounced (Figures 3, S2). There are a number of factors that might influence the precision of the single nucleobase recognition. These include the conformational states and structural dynamics of a single strand of DNA held inside the pore structure under the influence of an electric force<sup>[9,12]</sup>, and potential sequence-dependant DNA structural features that could effect the

precise alignment of the recognized nucleobase with respect to the  $\alpha$ -HL recognition. For instance, our data indicate that variations in the hairpin structure seem to somewhat shift the alignment of the captured DNA strand with respect to the  $\alpha$ -HL sensing site either slightly forward (strands **16** and **17**) or backward (strand **14**) as compared to strand **11** (Figures 2, 3). These apparently small changes in relative strand positioning could be responsible for decreased discrimination between A and C in the case of strand **14** and diminished selectivity for the C-type events in strand **17**. We note that issues concerning the precision of recognition arising from the base-dependent length of a ssDNA in an electric field would be less serious for DNA sequenced during transit through the pore, where the time-evolution of the conductance would be monitored, or for stepwise approaches where the DNA is fed into the pore by an enzyme or a nanoscopic device.<sup>[3,6]</sup> In these cases, the computer analyzing the signal would also be trained in base recognition within the context of neighboring sequences. Approaches based on current amplitude could also have advantages over a method based on the mean dwell time of each base at a recognition site.<sup>[3,12]</sup> Even if the mean dwell time for each of the four bases differed by an order of magnitude from the others, there would be a problematical overlap between the four dwell-time distributions.

In conclusion, our studies seem to support the notion that  $\alpha$ -HL-mediated DNA sequencing might be fundamentally feasible. We believe that progress toward that goal would greatly benefit from targeted engineering of pores<sup>[7]</sup> which could provide the means for improved positional selectivity and signal discrimination between the four DNA nucleobases. Furthermore, the single-species  $\alpha$ -HL-DNA rotaxane strategies described recently<sup>[5b,c]</sup> might also be of use in enhancing the fidelity of single nucleobase recognition via multiple-pass reading.

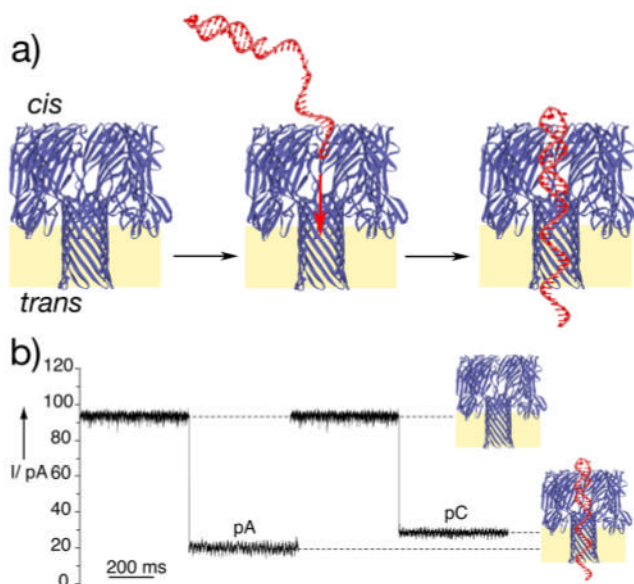
## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## References

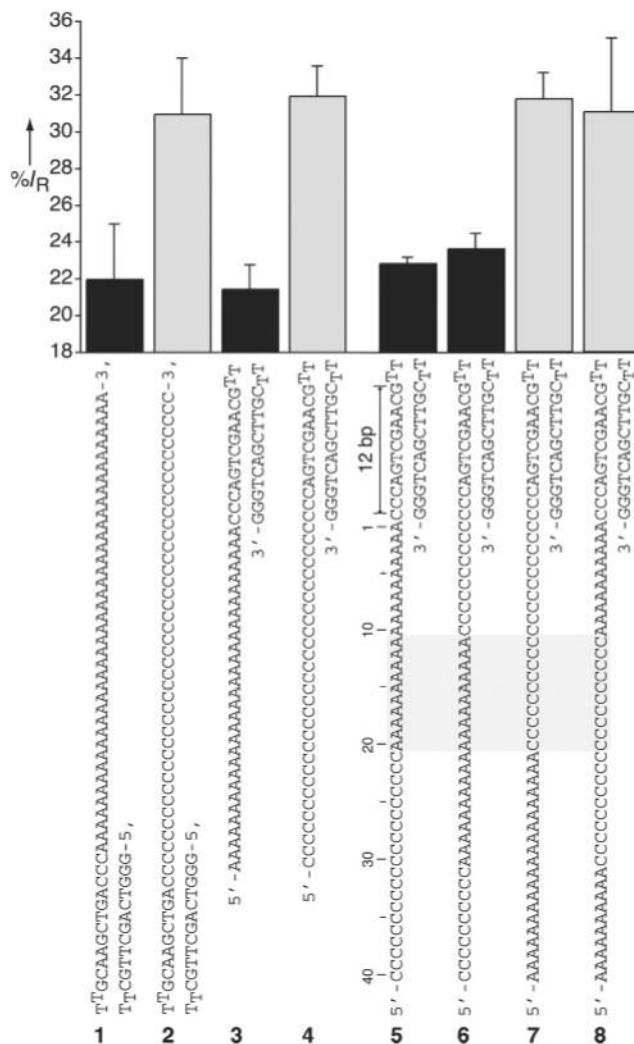
1. Song L, Hobaugh MR, Shustak C, Cheley S, Bayley H, Gouaux JE. *Science* 1996;274:1859. [PubMed: 8943190]
2. a) Kasianowicz JJ, Brandin E, Branton D, Deamer DW. *Proc Natl Acad Sci USA* 1996;93:13770. [PubMed: 8943010] b) Akesson M, Branton D, Kasianowicz JJ, Brandin E, Deamer DW. *Biophys J* 1999;77:3227. [PubMed: 10585944] c) Meller A, Nivon L, Brandin E, Golovchenko J, Branton D. *Proc Natl Acad Sci USA* 2000;97:1079. [PubMed: 10655487] (d) Howorka S, Cheley S, Bayley H. *Nat Biotechnol* 2001;19:636. [PubMed: 11433274]
3. a) Bayley H, Martin CR. *Chem Rev* 2000;100:2575. [PubMed: 11749296] b) Deamer DW, Akesson M. *TIBTECH* 2000;18:147. c) Deamer DW, Branton D. *Acc Chem Res* 2002;35:817–825. [PubMed: 12379134] (d) Nakane JJ, Akesson M, Marziali A. *J Phys: Condens Matter* 2003;15:R1365.
4. a) Jirage KB, Hulteen JC, Martin CR. *Science* 1997;278:655. b) Schmidt C, Mayer M, Vogel H. *Angew Chem* 2000;112:3267. *Angew. Chem. Int. Ed.* 2000, 39, 3137 c) Li J, Stein D, McMullan C, Branton D, Aziz MJ, Golovchenko JA. *Nature* 2001;412:166. [PubMed: 11449268] d) Bayley H, Cremer PS. *Nature* 2001;413:226. [PubMed: 11557992] e) Chen P, Mitsui T, Farmer DB, Golovchenko J, Gordon RG, Branton D. *Nano Lett* 2004;4:1333. f) Chang H, Kosari F, Andreadakis G, Alam MA, Vasmatazis G, Bashir R. *Nano Lett* 2004;4:1551.
5. a) Sauer-Budge AF, Nyamwanda JA, Lubensky DK, Branton D. *Phys Rev Lett* 2003;90:238101. [PubMed: 12857290] b) Sanchez-Quesada J, Saghatelian A, Cheley S, Bayley H, Ghadiri MR. *Angew Chem Int Ed* 2004;43:3063. *Angew. Chem.* 2004, 116, 3125 c) Nakane J, Wiggin M, Marziali A. *Biophys J* 2004;87:615. [PubMed: 15240494]
6. Church, G.; Deamer, DW.; Branton, D.; Baldarelli, R.; Kasianowicz, J. U.S. Patent. 5,795,782. 1995 1998.

7. a) Cheley S, Braha O, Lu X, Conlan S, Bayley H. *Protein Sci* 1999;8:1257. [PubMed: 10386875] b) Kasianowicz JJ, Burden DL, Han LC, Cheley S, Bayley H. *Biophys J* 1999;76:837. [PubMed: 9929485] c) Gu L-Q, Dalla Serra M, Vincent JB, Vigh G, Cheley S, Braha O, Bayley H. *Proc Natl Acad Sci U S A* 2000;97:3959. [PubMed: 10760267] d) Gu LQ, Cheley S, Bayley H. *Science* 2001;291:636. [PubMed: 11158673] e) Gu LQ, Cheley S, Bayley H. *J Gen Physiol* 2001;118:481. [PubMed: 11696607] h) Cheley S, Gu LQ, Bayley H. *Chem Biol* 2002;9:829. [PubMed: 12144927]
8. a) Vercoutere W, Winters-Hilt S, Olsen H, Deamer D, Haussler D, Akeson M. *Nat Biotechnol* 2001;19:248. [PubMed: 11231558] b) Winters-Hilt S, Vercoutere W, DeGuzman VS, Deamer D, Akeson M, Haussler D. *Biophys J* 2003;84:967. [PubMed: 12547778] c) Vercoutere WA, Winters-Hilt S, DeGuzman VS, Deamer D, Ridino SE, Rodgers JT, Olsen HE, Marziali A, Akeson M. *Nucleic Acids Res* 2003;31:1311. [PubMed: 12582251]
9. Howorka S, Bayley H. *Biophys J* 2002;83:3202. [PubMed: 12496089]
10. a) Smith SB, Cui Y, Bustamante C. *Science* 1996;271:795. [PubMed: 8628994] b) Bustamante C, Smith SB, Liphardt J, Smith D. *Curr Op Struct Biol* 2000;10:279–285.
11. a) Sauer-Budge AF, Nyamwanda JA, Lubensky DK, Branton D. *Phys Rev Lett* 2003;90:238101-1–238101-4. [PubMed: 12857290] b) Mathé J, Visram H, Viasnoff V, Rabin Y, Meller A. *Biophys J*. in press
12. a) Lubensky DK, Nelson DR. *Biophys J* 1999;77:1824. [PubMed: 10512806] b) Bates M, Burns M, Meller A. *Biophys J* 2003;84:2366. [PubMed: 12668445]



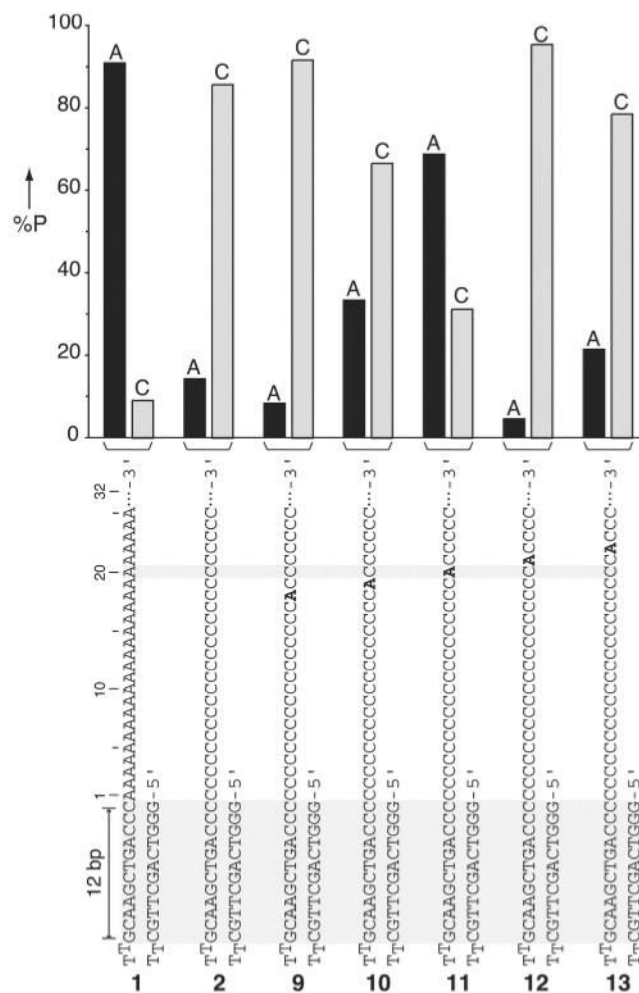
**Figure 1.**

a) Schematic representation of the single-species  $\alpha$ -HL-DNA pseudorotaxane formation. By employing sequences with a long ss-DNA segment having a hairpin structure at one end, the SS-DNA can thread the pore under positive applied potentials and held stably in a pseudorotaxane configuration by the interactions of the DNA duplex segment at the pore entrance. The heptameric transmembrane pore structure is depicted in a cutaway side view. b) Ion conductance blockades caused by capture of poly d(A) and poly d(C) single strands (sequences **1** and **2**, respectively). The typical events shown are from ion conductance traces recorded at +170 mV when the thread molecules (1  $\mu$ M) were added to the *cis* chamber. Traces were recorded under symmetrical conditions (500 mM KCl, 5 mM MOPS pH 7.5), Bessel filtered at 5 kHz, and sampled at 200  $\mu$ s.



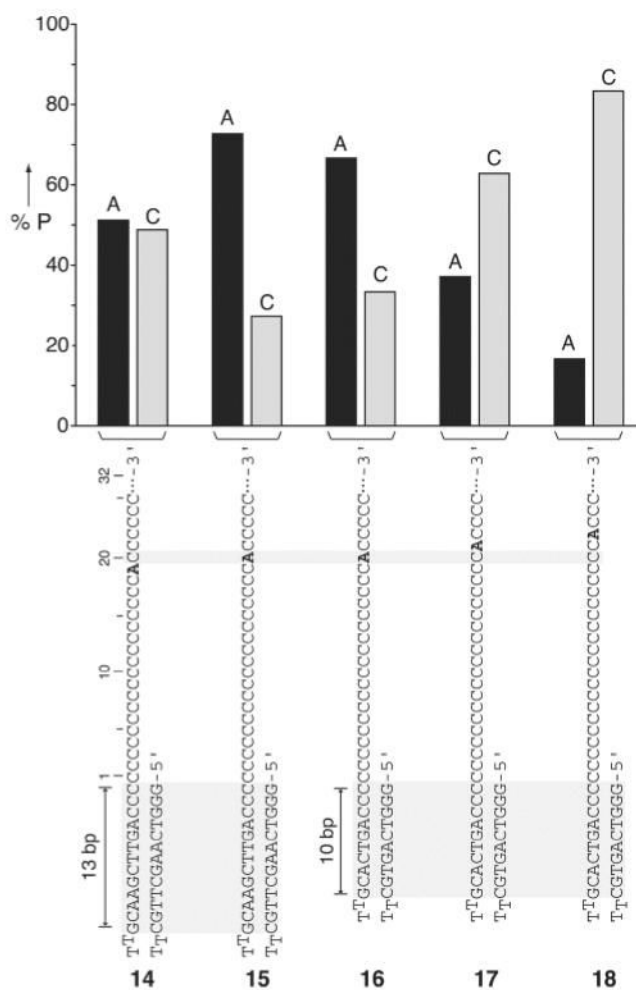
**Figure 2.**

The averaged residual currents for various  $\alpha$ -HL-DNA pseudorotaxanes measured at +170 mV. The left panel depicts the percent residual currents measured for poly d(A) and poly d(C) threads in two topoisomeric  $\alpha$ -HL-DNA pseudorotaxane configurations. Note that when the  $\alpha$ -HL pore is threaded with a homopolymeric DNA segment either from the free 3'-terminus (1 or 2) or from 5'-terminus (3 or 4), similar characteristic A-type or C-type ion channel blockades are observed. The right panel depicts the observed residual currents for  $\alpha$ -HL-DNA pseudorotaxanes prepared using four different ss-DNA block copolymer threads (5–6). The region of the ss-DNA segment recognized by the  $\alpha$ -HL channel (10 to 20 nucleotides away from the edge of the hairpin segment) is marked by the gray rectangle. For clarity A-type and C-type residual currents are shown as black and gray bars, respectively.



**Figure 3.** Percent probability of the A-type ( $I_R < 27\%$ ) and the C-type ( $I_R > 27\%$ ) residual currents for single deoxyadenosine-substituted poly d(C) DNA strands captured inside the  $\alpha$ -HL pore as a pseudorotaxane. Percent probability was calculated by the number of A-type or C-type events measured at +170 mV divided by the total number of events ( $n$ ) recorded for a given strand ( $n = 11, 21, 22, 32, 22,$  and  $14$  for strands **1, 2, 9, 10, 11, 12, 13**, respectively). The percent probability of A-type and C-type events for each strand are given as black and gray bars respectively.





**Figure 4.** Percent probability of the A-type and the C-type events for single deoxyadenosine-substituted poly d(C) DNA hairpins sequences having either 13- or 10-basepair stem duplexes (strands **14–15**, and **16–18**, respectively) measured at +170 mV ( $n = 43, 44, 69, 35, 66$  for **14–18**, respectively). For clarity the A-type ( $I_R < 27\%$ ) and C-type ( $I_R > 27\%$ ) events are shown as black and grey bars, respectively.