Molecular Recognition

Recognizing a Single Base in an Individual DNA Strand: A Step Toward DNA Sequencing in Nanopores**

*Nurit Ashkenasy, Jorge Sánchez-Quesada, Hagan Bayley, and M. Reza Ghadiri**

There has been an intriguing suggestion that *Staphylococcus aureus* α-hemolysin (α-HL), a stable heptameric transmembrane protein pore,[1] may be of use as the sensor element in a rapid, pore-mediated, single-molecule DNA sequencing process.[2] Although there remain several requirements that must be met before such a process can be realized,[2a,3] the most fundamental concern has been whether α-HL, or any other natural or manmade nanopore structure,[4] is capable of recognizing DNA with single nucleobase resolution. We have probed the nucleobase resolution capacity of α-HL by threading and holding a given strand of single-stranded (ss-) DNA inside the α-HL pore in the form of a single α-HL·DNA pseudorotaxane.[5] 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 on the ion conductance of α-HL. The discovery that α-HL can recognize ss-DNA with single nucleobase resolution strengthens the case for its utility in rapid single-molecule DNA sequencing.

[*] Dr. N. Ashkenasy, Dr. J. Sánchez-Quesada, Prof. Dr. 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
E-mail: ghadiri@scripps.edu
Prof. Dr. H. Bayley
Department of Chemistry
University of Oxford
Chemistry Research Laboratory
Mansfield Road, Oxford, OX1 3TA (UK)

[**] 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 studies.

Supporting information for this article is available on the WWW under http://www.angewandte.org or from the author.

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 α-HL pore by recording the perturbations in the conductance of the ion channel caused by the sequential passage of each nucleotide.[2,6] Toward this goal, in an important early study it was shown that the transient decreases in the conductance of the ion channel could be used to distinguish homopurine from homopyrimidine ss-nucleic acid strands.[2b,c] However, partly 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 α-HL.[2,3] We envisioned that this issue might be resolved by analyzing decreases in the conductance of the ion channel caused by a given strand of ss-DNA while it is captured and held stable inside an α-HL pore in the form of a rotaxane or pseudorotaxane.[5]

The DNA sequences employed in our studies each have a long ss-DNA segment for threading the α-HL pore and a stable terminal hairpin structure for holding the thread in a pseudorotaxane configuration (Figure 1 a).[5b] As the hairpin duplex structure is wider than the α-HL internal lumen, a given DNA strand can only thread the pore with its free single-stranded terminus.[5] Therefore, depending on from which end the α-HL pore is threaded and whether the hairpin structure is placed at the 3′- or 5′-terminus of the DNA thread, four distinct topoisomeric forms of α-HL·DNA pseudorotax-

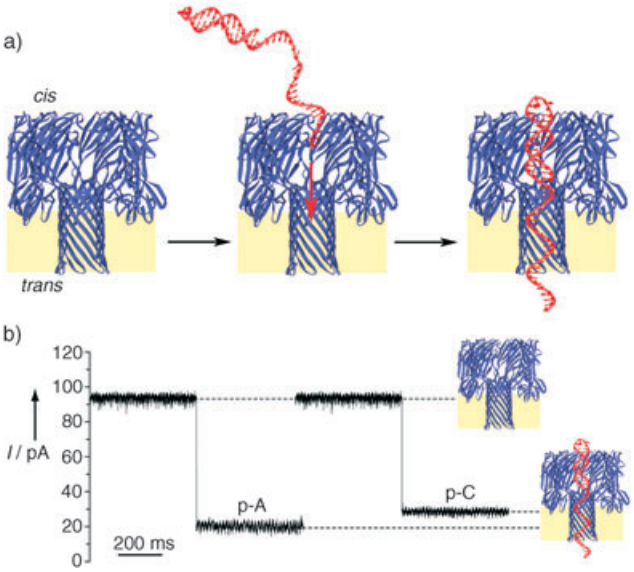

Figure 1. a) Schematic of the formation of single α-HL·DNA pseudorotaxane species. By employing a sequence with a long ss-DNA segment that has a hairpin structure at one end, the ss-DNA can thread the pore upon application of positive potentials and be held stably in a pseudorotaxane configuration by the interactions of the DNA duplex segment at the pore entrance. The heptameric transmembrane pore structure is depicted in a cutaway side view. b) Perturbations in ion current caused by the capture of poly-d(A) (p-A) and poly-d(C) (p-C) single strands (sequences **1** and **2**, respectively). The events shown are typical ion-current traces recorded at 170 mV under symmetrical conditions (KCl (500 mM), MOPS (5 mM, pH 7.5), Bessel-filtered at 5 kHz, sampled at 200 μs) with the thread molecules (1 μM) added to the *cis* chamber.

 DOI: 10.1002/ange.200462114

anes can be formed.[5b] Here we report the characteristics of the topoisomers of α-HL·DNA pseudorotaxanes that are prepared by threading DNA from the *cis* side (Figure 1a). In a typical experiment, the DNA sample was introduced at the *cis* side of a lipid bilayer that contained a single oriented α-HL pore, and the threading process was initiated by applying positive transmembrane holding potentials (see Supporting Information for details). Formation of the α-HL·DNA pseudorotaxane was accompanied by a decrease in the conductance of the ion channel owing to the presence of the ss-DNA inside the pore (Figure 1b).[2,5] The DNA strand can be held stably in the pseudorotaxane configuration as long as a suitable positive transmembrane holding potential is maintained (see Supporting Information).[5b]

Initial experiments were performed with α-HL·DNA pseudorotaxanes of poly-d(A) and poly-d(C) (sequences **1**–**4**, Figure 2) to ascertain whether homopurine and homopyrimidine DNA strands held inside the α-HL pore give rise to distinct decreases in the ion conductance. 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 (see Supporting Information). The residual currents I_R at 170 mV (calculated as a percentage of the current measured for the unoccupied DNA-free α-HL during the same experiment) are $22 \pm 3\%$ and $31 \pm 3\%$ for poly-d(A) (sequence **1**) and poly-d(C) (sequence **2**) pseudorotaxanes, respectively (Figure 2). Interestingly, topoisomeric α-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 **2** versus threading from the 5′-terminus using **3** or **4**) gave similar characteristic "A-type" or "C-type" ion channel blockades, respectively (Figure 2).

As a significant portion of the ss-DNA in an α-HL·DNA pseudorotaxane is held stably inside the protein's pore,[5] we sought to determine which part of the encased DNA thread accounted for the differences observed in the ion conductances associated with poly-d(A) and poly-d(C) oligonucleotides. Accordingly, we prepared single α-HL·DNA pseudorotaxane species from 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–α-HL pseudorotaxanes fall clearly into either "A-type" or "C-type" categories (Figure 2). Therefore, by a simple comparative sequence analysis, it appeared that the nucleobase(s) recognized by α-HL must be located in the region 10 to 20 nucleotides away from the edge of the hairpin structure (Figure 2). Further studies with similar DNA block copolymer threads, which have shorter poly-d(C) and poly-d(A) segments, helped to define the location of the recognition site to around nucleotide 20 (data not shown).

To address whether the α-HL pore structure could recognize DNA with single nucleobase resolution, we designed five poly-d(C) threads (sequences **9**–**13**, Figure 3), each with a single deoxyadenosine at a unique position within the putative recognition site located 18 to 22 nucleotides away

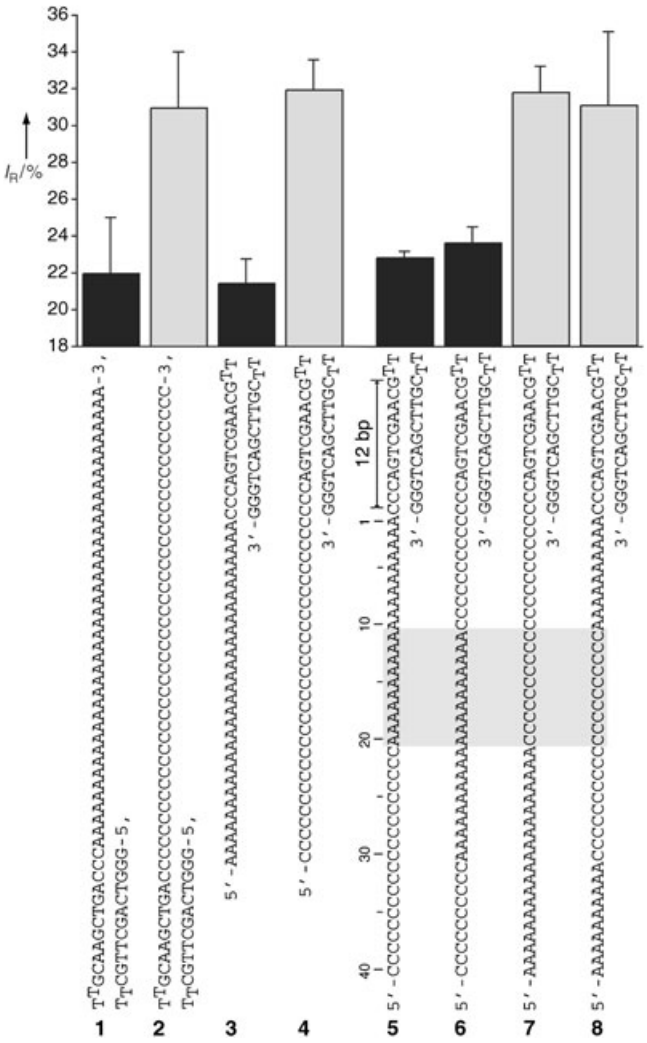

Figure 2. Averaged residual currents (I_R) for various α-HL·DNA pseudorotaxanes measured at 170 mV. Left: Bars **1**–**4** depict the residual currents measured for poly-d(A) and poly-d(C) threads in two topoisomeric α-HL·DNA pseudorotaxane configurations. Note that when the α-HL pore is threaded with a homopolymeric DNA segment either from the free 3′-terminus (**1** or **2**) or from the 5′-terminus (**3** or **4**), similar characteristic A-type or C-type ion-channel blockades are observed. Right: Bars **5**–**8** depict the observed residual currents for α-HL·DNA pseudorotaxanes prepared from four different ss-DNA block copolymer threads (**5**–**8**). A-type and C-type residual currents are shown as black and gray bars, respectively. The region of the ss-DNA segment recognized by the α-HL channel (10 to 20 nucleotides away from the edge of the hairpin segment) is shaded in gray.

from the edge of the 5′-stem–loop structure. Each DNA strand was used to prepare the corresponding single α-HL·DNA pseudorotaxane species. Furthermore, each of the pseudorotaxanes was analyzed several times to obtain statistically significant numbers of events for calculating the individual residual current values, I_R, at 170 mV (see Supporting Information). The measured I_R distributions were roughly bimodal and could be divided into two groups of events corresponding to A-type ($I_R < 27\%$) and C-type ($I_R > 27\%$) values (Figure 3). A given event displayed current characteristics similar to either poly-d(A) or poly-d(C)

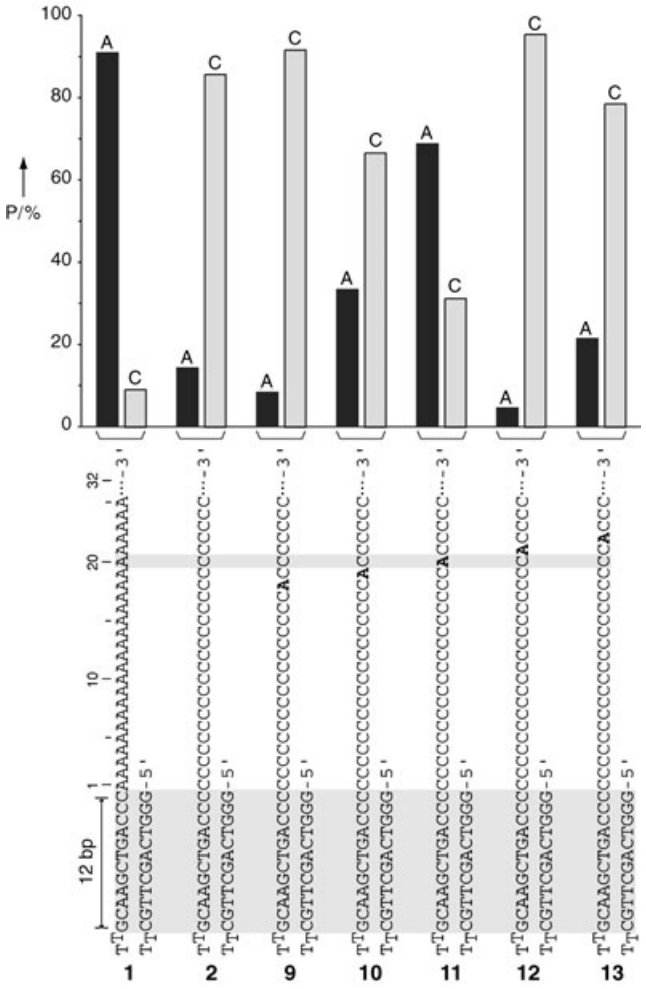

Figure 3. Percentage probability (P) of the A-type ($I_R < 27\%$) and the C-type ($I_R > 27\%$) residual currents for single deoxyadenosine-substituted poly-d(C) DNA strands captured inside the α-HL pore as a pseudorotaxane. Percentage probabilities were calculated by the number of A-type or C-type events 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**, and **13**, respectively). The percentage probabilities of A-type and C-type events for each strand are shown as black and gray bars, respectively.

pseudorotaxanes (Figure 1b) and did not interconvert from one type into the other during the analysis. The data clearly show that when deoxyadenosine is at positions 18, 19, 21, or 22 (sequences **9**, **10**, **12**, or **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**), despite 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 α-HL pore structure can recognize ss-DNA site-specifically with single nucleobase resolution.

The above observations imply that in the α-HL·DNA pseudorotaxanes, the ss-DNA thread must be held in a specific conformation inside the pore to allow the correct position of the deoxyadenosine at a recognition site. Therefore, by considering 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 α-HL pore the nucleobase sensing site is located. As DNA hairpins with stems longer than 8 bp can enter the vestibule and bump up against the internal pore constriction,[7] the 12 bp stem hairpin strands used in our studies are most likely pinned to the α-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 pore structure and dimensions of α-HL.[8] 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 this hypothesis, α-HL·DNA pseudorotaxanes prepared by using hairpin DNA strands that comprise 13 bp (strands **14** and **15**) or 10 bp (strands **16** to **18**) stem segments retained their ability to report a single deoxyadenosine placed 20 nucleotides away from the edge of the hairpin (see Figure 4 and Supporting Information). 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 ds-DNA. ss-DNA, relative to ds-DNA, is highly contractile and at forces below about 8 pN would be shortened compared to B-DNA.[9] Such small forces seem reasonable if we take into account that the effective charge per nucleotide within the α-HL pore is only approximately 0.1 e as suggested by recent studies.[5a,10] Therefore, the above rationale and observations suggest that the nucleobase recognition site is most likely located near the *trans* opening of the α-HL pore (see Figure 1a).

The precision of the single nucleobase recognition is an important factor in single-molecule DNA sequencing. Careful analysis of the data described here indicate that the precision of nucleotide sensing is not sufficiently high in the "native" α-HL pore for use in DNA sequencing. For example, whereas the observed residual currents indicate that an adenine substitution at either positions 18, 21, or 22 (sequences **9**, **12**, or **13**) is not recognized (Figure 3), the discriminating power between A and C at position 19 (by using sequence **10**) is noticeably less pronounced (see Figure 3 and Supporting Information). A number of factors may influence the precision of single nucleobase recognition, including the conformational states and structural dynamics of a single strand of DNA held inside the pore structure under the influence of an electric force[8,11] and sequence-dependent DNA structural features that could affect the precise alignment of the recognized nucleobase with respect to the recognition site. For instance, our data indicate that variations in the hairpin structure seem to shift the alignment of the captured DNA strand with respect to the α-HL sensing site either slightly forward (strands **16** and **17**) or backward (strand **14**) relative to strand **11** (Figures 2 and 3). These small changes in apparent 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**. Factors that concern the precision of recognition arising from the base-dependent length of a ss-DNA in an electric field would be less serious for DNA that is sequenced

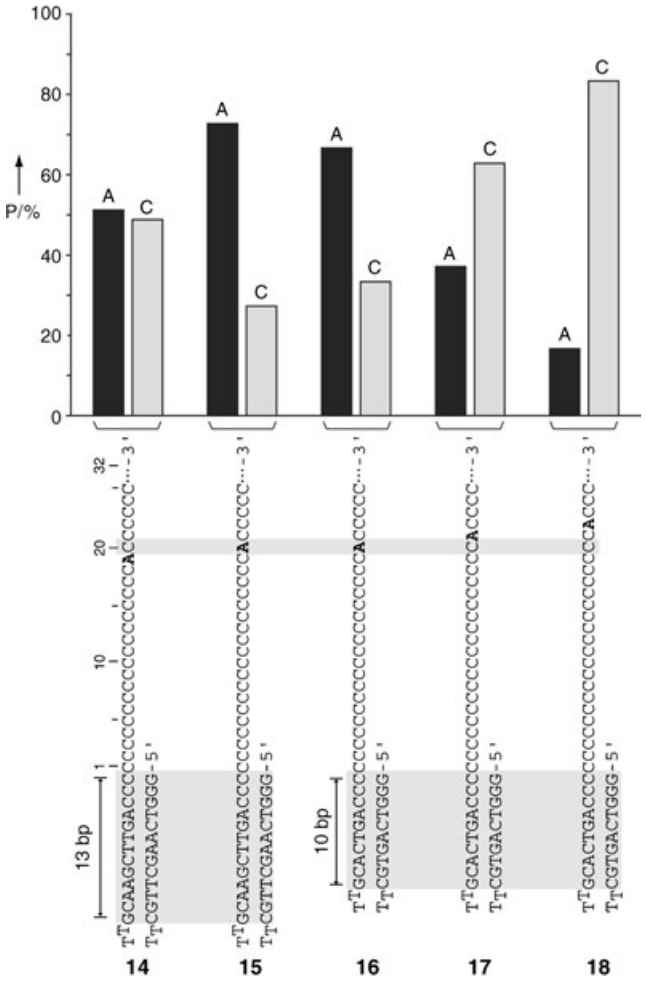

Figure 4. Percentage probability (P) of A-type and C-type events for single deoxyadenosine-substituted poly-d(C) DNA hairpin sequences that have either 13- or 10-base-pair stem duplexes (strands **14**–**15** and **16**–**18**, respectively) measured at 170 mV ($n = 43$, 44, 69, 35, and 66 for **14**–**18**, respectively). The A-type ($I_R < 27\%$) and C-type ($I_R > 27\%$) events are shown as black and gray bars, respectively.

during transit through the pore, in which case the time evolution of the conductance would be monitored, or for stepwise approaches by which the DNA is fed into the pore by an enzyme or a nanoscopic device.[3,6] In these cases, the computer that analyzes the signal would also be programmed in base recognition within the context of neighboring sequences. Approaches based on current amplitude, as demonstrated here, would have advantages over a method based on the mean dwell time of each base at a recognition site.[3,11] 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 distributions of the four dwell times.

In conclusion, our studies support the notion that α-HL-mediated DNA sequencing might be fundamentally feasible. We believe that progress toward this goal would greatly benefit from targeted engineering of pores[12] which could provide the means for improved positional selectivity and signal discrimination between the four DNA nucleobases. Furthermore, the strategies for single α-HL·DNA rotaxane species described recently[5b,c] might also be of use in enhancing the fidelity of single nucleobase recognition through multiple-pass reading.

Received: September 25, 2004
Published online: January 21, 2005

Keywords: DNA · host–guest chemistry · molecular recognition · rotaxanes · single-molecule studies

[1] L. Song, M. R. Hobaugh, C. Shustak, S. Cheley, H. Bayley, J. E. Gouaux, *Science* **1996**, *274*, 1859.
[2] a) J. J. Kasianowicz, E. Brandin, D. Branton, D. W. Deamer, *Proc. Natl. Acad. Sci. USA* **1996**, *93*, 13770; b) M. Akeson, D. Branton, J. J. Kasianowicz, E. Brandin, D. W. Deamer, *Biophys. J.* **1999**, *77*, 3227; c) A. Meller, L. Nivon, E. Brandin, J. Golovchenko, D. Branton, *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 1079; d) S. Howorka, S. Cheley, H. Bayley, *Nat. Biotechnol.* **2001**, *19*, 636.
[3] a) H. Bayley, C. R. Martin, *Chem. Rev.* **2000**, *100*, 2575; b) D. W. Deamer, M. Akeson, *Trends Biotechnol.* **2000**, *18*, 147; c) D. W. Deamer, D. Branton, *Acc. Chem. Res.* **2002**, *35*, 817–825; d) J. J. Nakane, M. Akeson, A. Marziali, *J. Phys. Condens. Matter* **2003**, *15*, R1365.
[4] a) K. B. Jirage, J. C. Hulteen, C. R. Martin, *Science*, **1997**, *278*, 655; b) C. Schmidt, M. Mayer, H. Vogel, *Angew. Chem.* **2000**, *112*, 3267; *Angew. Chem. Int. Ed.* **2000**, *39*, 3137; c) J. Li, D. Stein, C. McMullan, D. Branton, M. J. Aziz, J. A. Golovchenko, *Nature* **2001**, *412*, 166; d) H. Bayley, P. S. Cremer, *Nature* **2001**, *413*, 226; e) P. Chen, T. Mitsui, D. B. Farmer, J. Golovchenko, R. G. Gordon, D. Branton, *Nano Lett.* **2004**, *4*, 1333; f) H. Chang, F. Kosari, G. Andreadakis, M. A. Alam, G. Vasmatzis, R. Bashir, *Nano Lett.* **2004**, *4*, 1551.
[5] a) A. F. Sauer-Budge, J. A. Nyamwanda, D. K. Lubensky, D. Branton, *Phys. Rev. Lett.* **2003**, *90*, 238101; b) J. Sanchez-Quesada, A. Saghatelian, S. Cheley, H. Bayley, M. R. Ghadiri, *Angew. Chem.* **2004**, *116*, 3125; *Angew. Chem. Int. Ed.* **2004**, *43*, 3063; c) J. Nakane, M. Wiggin, A. Marziali, *Biophys. J.* **2004**, *87*, 615.
[6] G. Church, D. W. Deamer, D. Branton, R. Baldarelli, J. Kasianowicz, US Patent 5795782, **1995**, **1998**.
[7] a) W. Vercoutere, S. Winters-Hilt, H. Olsen, D. Deamer, D. Haussler, M. Akeson, *Nat. Biotechnol.* **2001**, *19*, 248; b) S. Winters-Hilt, W. Vercoutere, V. S. DeGuzman, D. Deamer, M. Akeson, D. Haussler, *Biophys. J.* **2003**, *84*, 967; c) W. A. Vercoutere, S. Winters-Hilt, V. S. DeGuzman, D. Deamer, S. E. Ridino, J. T. Rodgers, H. E. Olsen, A. Marziali, M. Akeson, *Nucleic Acids Res.* **2003**, *31*, 1311.
[8] S. Howorka, H. Bayley, *Biophys. J.* **2002**, *83*, 3202.
[9] a) S. B. Smith, Y. Cui, C. Bustamante, *Science* **1996**, *271*, 795; b) C. Bustamante, S. B. Smith, J. Liphardt, D. Smith, *Curr. Opin. Struct. Biol.* **2000**, *10*, 279.
[10] J. Mathé, H. Visram, V. Viasnoff, Y. Rabin, A. Meller, *Biophys. J.* **2004**, *87*, 3205.
[11] a) D. K. Lubensky, D. R. Nelson, *Biophys. J.* **1999**, *77*, 1824; b) M. Bates, M. Burns, A. Meller, *Biophys. J.* **2003**, *84*, 2366.
[12] a) O. Braha, B. Walker, S. Cheley, J. J. Kasianowicz, L. Song, J. E. Gouaux, H. Bayley, *Chem. Biol.* **1997**, *4*, 497; b) J. J. Kasianowicz, D. L. Burden, L. C. Han, S. Cheley, H. Bayley, *Biophys. J.* **1999**, *76*, 837; c) L.-Q. Gu, M. Dalla Serra, J. B. Vincent, G. Vigh, S. Cheley, O. Braha, H. Bayley, *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 3959; d) L. Q. Gu, S. Cheley, H. Bayley, *Science* **2001**, *291*, 636; e) L. Q. Gu, S. Cheley, H. Bayley, *J. Gen. Physiol.* **2001**, *118*, 481; h) S. Cheley, L.-Q. Gu, H. Bayley, *Chem. Biol.* **2002**, *9*, 829.

