Control of subunit stoichiometry in single-chain MspA nanopores

Mikhail Pavlenok¹, Luning Yu², Dominik Herrmann¹, Meni Wanunu^{2,3,4}, and
Michael Niederweis^{1*}

1 Department of Microbiology, University of Alabama at Birmingham, 845 19th Street South,
Birmingham, AL 35294, USA

2 Department of Physics, Northeastern University, Boston, MA 02115

3 Department of Bioengineering, Northeastern University, Boston, MA 02115

4 Department of Chemistry and Chemical Biology, Northeastern University, Boston, MA 02115

Running title: Single-chain MspA nanopores

KEYWORDS

biosensor; channel; pore protein; DNA sequencing; stochastic sensing; analyte; nucleotide recognition

***Corresponding author:** Michael Niederweis, mnieder@uab.edu

845 19th Street South, BBRB 609

Birmingham, AL 35294 USA

Phone: 1-205-975-4390

ABSTRACT

Transmembrane protein channels enable fast and highly sensitive electrical detection of single molecules. Nanopore sequencing of DNA was achieved using an engineered *Mycobacterium smegmatis* porin A (MspA) in combination with a motor enzyme. Due to its favorable channel geometry, the octameric MspA pore exhibits the highest current level as compared to other pore proteins. To date, MspA is the only protein nanopore with a published record of DNA sequencing. While widely used in commercial devices, nanopore sequencing of DNA suffers from significant base-calling errors due to stochastic events of the complex DNA-motor-pore combination and the contribution of up to five nucleotides to the signal at each position. Asymmetric mutations within subunits of the channel protein offer an enormous potential to improve nucleotide resolution and sequencing accuracy. However, random subunit assembly does not allow control of the channel composition of MspA and other oligomeric protein pores. In this study, we showed that it is feasible to convert octameric MspA into a single-chain pore by connecting eight subunits using peptide linkers. We constructed single-chain MspA trimers, pentamers, hexamers and heptamers to demonstrate that it is feasible to alter the subunit stoichiometry and the MspA pore diameter. All single-chain MspA proteins formed functional channels in lipid bilayer experiments. Importantly, we demonstrated that single-chain MspA discriminated all four nucleotides identical to MspA produced from monomers. Thus, single-chain MspA constitutes a new milestone in its development and adaptation as a biosensor for DNA sequencing and many other applications.

STATEMENT OF SIGNIFICANCE

Nanopore sequencing of DNA is a fast and cheap technology that uniquely delivers multi-kilobase reads. It is currently used world-wide in many applications such as genome sequencing, epigenetics, and surveillance of viral and bacterial pathogens and has started to revolutionize human lives in medicine, agriculture and environmental studies. However, the high base-calling error rates prevent nanopore DNA sequencing from reaching its full potential. In this study, we converted octameric MspA into a single-chain pore enabling asymmetric mutations to fine-tune the pore geometry and chemistry and address the shortcomings of nanopores. Thus, single-chain MspA constitutes a new milestone in its development and adaptation as a biosensor for DNA sequencing and many other applications.

INTRODUCTION

Protein pores have attracted an enormous interest as biosensors as they enable fast and highly sensitive detection of analytes by recording current traces. This concept of stochastic sensing was pioneered using bacterial and fungal pores (1,2) and has been applied to a wide variety of small molecules and polymers (3). In particular, DNA sequencing with pore proteins is a promising technology due to its low cost, fast processing, minimal sample amount and lack of sample amplification (4). While the concept of nanopore DNA sequencing was developed using the staphylococcal α -hemolysin pore (5), it was demonstrated to be a feasible technology using the *Mycobacterium smegmatis* porin A (MspA) (6). MspA is the most stable of the known pore proteins resisting denaturation by 2% SDS at 100 °C (7) and is functional in many different membrane-like environments (8-10), an important feature for technical applications (3). We first engineered MspA to translocate DNA (11), showing that MspA can distinguish the four nucleobases (12) and then achieved nanopore sequencing of DNA in combination with a motor enzyme (13). Several other channel proteins have also been used for nucleotide sensing including α -hemolysin (14), CsgG (15), ClyA (16), FraC (17) and aerolysin (18). MspA exhibits the highest current level compared to α -hemolysin, CsgG and aerolysin, likely due to its short channel constriction and its wide vestibule at the pore entrance (19), increasing the signal-to-noise ratio. MspA is the only protein nanopore with a published record of DNA sequencing (13,20). Mutants of the CsgG nanopore are apparently being used in the commercially available MinION devices (21), but no molecular details of the performance of CsgG in DNA sequencing experiments have been published to our knowledge.

One of the main problems of nanopore sequencing is the very fast translocation of DNA (22). Control of the DNA translocation rate is currently achieved by DNA-processing enzymes, but this adds complexity and stochastic signals from the motor protein, decreasing the signal-to-noise ratio (13,23). In addition, the residual current of single-stranded DNA passing through the pore is determined by four to five nucleotides at each position (13). These and other limitations result in raw base calling errors up to 12% for MspA (24). To our knowledge, raw base calling errors are not published for any other nanopore. Many of these challenges could, in principle, be addressed by protein engineering. However, the oligomeric nature of MspA and of other protein nanopores comprising seven to twelve subunits, severely limits these efforts (25). Asymmetric mutations are required to efficiently improve nucleotide recognition, to better control DNA capture and translocation rates and to control the path of DNA translocation. While first steps in this direction have been made by electrophoretically separating permutations of a mixture of two distinct hemolysin subunits after chemical modifications (26), this process is tedious, difficult to scale and is not usable for a combination of mutations in different subunits. This can only be effectively achieved by synthesizing one gene which encodes all subunits. We demonstrated that this is feasible for MspA because its C-terminus is close to the N-terminus of the next subunit (Fig. 1A) and showed that two covalently connected subunits indeed assemble to functional MspA pores (25). However, obtaining pure single-chain MspA, in which all eight subunits are linked (Fig. 1B), has been elusive for almost a decade.

In this study, we show that it is feasible to convert octameric MspA into a monomeric pore enabling asymmetric mutations to optimize the pore properties for specific applications. As proof-of-principle, we also constructed and purified single-chain MspA trimers, pentamers, hexamers, and heptamers, highlighting the possibility to construct pores with different channel

diameters by controlling their subunit stoichiometry. All single-chain MspA proteins formed functional channels in lipid bilayer experiments. Importantly, we demonstrated that full-length single-chain MspA discriminated all four nucleotides in a manner identical to MspA produced from monomers. Single-chain MspA constitutes a new milestone in its development and adaptation as a nanosensor for DNA sequencing and many other applications.

MATERIAL AND METHODS

Chemicals and Reagents, bacterial strains, media and growth conditions and standard procedures are described in the **SI Appendix**.

Plasmid construction. Standard molecular biology methods were used for plasmid construction. Single-chain MspA genes were ordered from GenScript (Piscataway, NJ). Description of construction of single-chain MspA M2 and single-chain MspA PN1 plasmids are in the **SI Appendix**.

Purification of MspA octamers. Purification of octameric MspA from *M.smegmatis* was performed as described elsewhere (27) and is summarized in the **SI Appendix**.

Expression and purification of single-chain MspA proteins. Single-chain MspA protein were purified from *E.coli* BL21(DE3)omp8 strain (28). After induction with 1 mM IPTG, cells were harvested and inclusion bodies were purified and solubilized in 8 M urea. Affinity purification or gel excision were used to further purify single-chain proteins. The proteins were then refolded by dialysis against buffer containing 0.5% OPOE detergent. For detailed purification procedures of single-chain MspA proteins see **SI Appendix**.

Lipid bilayer experiments and single-channel hairpin experiments. Lipid bilayer experiments to check pore formation by MspA proteins were performed in diphytanoylphosphatidylcholine artificial membranes in a custom made bilayer apparatus as previously described (27) and summarized in **SI Appendix**. Single-channel recordings and hairpin experiments were done in SU-8 wedge-on-pillar aperture and poly(1,2-butadiene)-b-poly(ethylene oxide) copolymer was used to form a bilayer as described (10). Data were collected at 250 kHz sampling rate and digitized. For details see **SI Appendix**.

RESULTS

Purification and characterization of single-chain MspA.

Eight MspA subunits assemble in the outer membrane of *Mycobacterium smegmatis* to produce a central water-filled channel (Fig. 1A). We have shown previously that the C-terminus of a monomer can be connected to the N-terminus of a neighboring subunit by a 17-mer peptide linker due to the close proximity of the termini (25). Here, we used the same approach to synthesize a single-chain *mspA m2* (*scmspA m2*) gene where eight *mspA m2* genes encoding the mutations D90N/D91N/D93N/D118R/D134K/D139R, which enable efficient DNA capture and translocation (11), are connected by DNA fragments encoding (GGGGS)₃ peptide linkers (Fig. 1B; Table S1). Each gene is flanked by unique restriction sites to enable specific modifications of each MspA subunit (Table S1). We first tried to produce scMspA M2 protein in the porin quadruple mutant *M. smegmatis* ML712, which lacks all four *msp* genes (29). However, all *scmspA m2* genes in clones producing MspA protein were scrambled and protein quantities were very low (not shown). Then, we used *E. coli* to produce scMspA M2 protein. To reduce the probability of homologous recombination between the individual *mspA* genes, we altered every second codon in the *mspA* subunit genes 2 through 8 to generate DNA sequence differences without altering the amino acid sequence. In addition, the *scmspA m2* DNA sequence was altered for optimal expression in *E. coli* and the signal peptide was removed for cytoplasmic expression of MspA. We expressed the synthetic *scmspA m2* gene under the control of the T7 promoter using the plasmid pML3216 (Fig. S1) in the BL21(DE3)omp8 strain lacking the three major porins of *E. coli* (28) to avoid contamination with endogenous porins. MspA protein (166 kDa theoretical mass) was purified from inclusion bodies by gel electrophoresis and excision of a protein band of 170 kDa and refolded in a buffer containing the detergent octyl polyoxyethylene (OPOE) (Fig. 1D). The protein yield was 12 µg of purified sc₈MspA M2 per single preparation from four polyacrylamide gels (approximately 190 µg per liter of *E. coli* culture).

Octameric MspA is an extremely stable protein that does not denature after boiling for 10 min in 2% SDS and other harsh denaturing conditions (7). To dissociate octameric MspA into its subunits we boiled the purified proteins in 80% (v/v) dimethylsulfoxide (DMSO) (25). Only octameric MspA M2 dissociated into monomers, while sc₈MspA M2 was stable demonstrating that all eight subunits are covalently linked and that we purified full-length scMspA (Fig. 1E). To examine whether sc₈MspA M2 forms functional channels we performed lipid bilayer experiments in diphytanoyl phosphatidylcholine (DPhPC) membranes of 1 mm in diameter in a Montal-Mueller setup (25). Addition of both octameric MspA M2 purified from *M. smegmatis* (Fig. 1G) and recombinant sc₈MspA M2 protein to the membranes resulted in a step-wise current increase indicative of channel insertions (Fig. 1F), while no channel activity was observed when only detergent-containing buffer was added (not shown). Analysis of the single channel insertions revealed similar conductance values ranging from 0.5 to 6 nS for octameric MspA M2 and from 0.2 to 4.5 nS for single chain MspA (Fig. S4). These wide ranges of single channel conductances were also observed for other octameric MspA proteins such as wt MspA and MspA M1 (Fig. S4) and seem to be an intrinsic feature of the MspA pore. Taken together, these experiments demonstrate that sc₈MspA M2 forms functional pores and show that it is feasible to convert an oligomeric pore protein into a functional monomeric protein enabling asymmetric mutations to improve channel properties.

Single-chain MspA variants with altered subunit stoichiometries form functional channels.

To highlight the advantages of scMspA in tailoring the pore for specific applications, we aimed at changing its channel diameter as one of the most important properties determining the interactions of the pore constriction with the analyte. Previously, the constriction diameter could only be altered by mutations of individual amino acids in the protein monomer, which also changes the chemical nature of the interactions with the analyte. A monomeric pore consisting of identical subunits enables for the first time to vary the channel diameter by altering the subunit stoichiometry. To this end, we designed scMspA variants with three, five, six, seven, and eight subunits based on MspA M2 with an additional P97F mutation (Table S5). This additional mutation was chosen because octameric MspA PN1 has a well defined channel distribution with a peak at 1.2 nS (Fig. S5A) in contrast to all other examined octameric MspA pores (Fig. S4). The corresponding proteins were named sc_xMspA PN1 with “x” numbers of covalently linked subunits and were purified from *E. coli* BL21(DE3)omp8 as described above. The target proteins accounted for approximately 10% of proteins in the whole-cell lysate as determined by quantitative image analysis of protein gels (Fig. S3). The apparent molecular masses for sc₃MspA PN1, sc₅MspA PN1, sc₆MspA PN1, and sc₇MspA PN1 proteins were 60 kDa, 110 kDa, 120 kDa, and 140 kDa, respectively (Fig. 2A), and were in a good agreement with the theoretical molecular masses. To examine whether the scMspA PN1 subunits are covalently linked, the proteins were denatured by boiling in 80% (v/v) DMSO. DMSO treatment dissociated octameric MspA M2 protein into its 20 kDa monomers, while all scMspA PN1 proteins were stable under these conditions (Fig. 2A) demonstrating that the subunits of these proteins are indeed covalently linked. To examine the channel-forming properties of scMspA PN1 pores we performed planar lipid bilayer experiments as previously described (29). In these experiments addition of 70 – 840 ng of refolded proteins to the membrane resulted in a step-

wise current increase indicative of channel insertions into lipid membranes (Fig. 2B-F). The channel insertion frequencies were similar for all scMspA PN1 proteins. The scMspA PN1 pores showed conductance values ranging between 0.5 to 5 nS (Fig. S5) similar to sc₈MspA M2 and octameric MspA proteins (Fig. S4). Interestingly, the distribution peaks shifted to larger single channel conductances for the sc₃MspA PN1, sc₅MspA PN1 and sc₇MspA PN1 variants (Figs. S5), indicating the presence of pores with more than eight subunits. Taken together, these experiments demonstrated that it is feasible to construct MspA pores with altered stoichiometries using covalently linked subunits. Thus, the single-chain concept enables the design of MspA pores with constriction zone diameters optimized for translocating different substrates, e.g. in nucleic acid and polypeptide sequencing.

Improvement of the purification of recombinant single-chain MspA protein.

While the above experiments demonstrated that single-chain MspA produces functional pores, the purification based on gel extraction is labor intensive and inefficient resulting in low yields of approximately 10 µg per preparation. We estimated that ~98% of the initial amount of scMspA proteins in the inclusion bodies were lost during this process. One of the main challenges was to separate full-length sc₈MspA M2 from its many degradation products which are only marginally smaller, i.e. these degradation products may have cleaved linker peptides which appear to be very susceptible to proteolysis (Fig. S6). To address this issue we exploited the fact that both termini of MspA are accessible on the outside of the MspA channel (Fig. 1B). Thus, we added a His₈-tag at the N-terminus and a Twin-Strep II tag at the C-terminus flanking the *sc₈mspA m2* gene (Fig. 3A). We refer to this protein as double-tagged single-chain MspA M2 (sc₈MspA_{dt} M2). The plasmid pML4170 carrying *sc₈mspA_{dt} m2* was transformed into *E. coli* BL21(DE3)Omp8 for protein production and purification. Sc₈MspA_{dt} M2 was purified under denaturing conditions using subsequent Ni(II)- and streptavidin affinity chromatography to isolate full-length protein (Figs. 3B,C, S7). Refolding in the detergent OPOE and a final size exclusion (Fig. S7) yielded 0.2 mg per liter of culture per single preparation of apparently pure protein (Fig. 3 B, C). Importantly, we did not observe contaminating degradation products (Fig. S6) in the refolded

Fig. 2. Channel activity of MspA pores with different subunit stoichiometries.

(A) Coomassie-stained 8% polyacrylamide gel with different single-chain MspA PN1 constructs. Lanes: wt – octameric MspA M2; sc-3, -5, -6, -7, -8 – single-chain MspA PN1 with three, five, six, seven, eight covalently linked subunits. Octameric MspA M2 and proteins with different subunit composition were boiled in 80% DMSO (+) and compared to untreated protein (-). Note that the 20 kDa band is monomeric subunit of MspA. Equal amounts (2 µg) of each sample were loaded on the gel. (B - F) Current traces of single-channel conductances of scMspA PN1 constructs. Membrane currents were recorded in 1 M KCl, 10 mM HEPES, pH 7.4 electrolyte at -10 mV applied potential. Diameter of the diphytanoyl phosphatidylcholine (DPhPC) bilayer was approximately 1 µm. (B) sc₃MspA PN1: 95 insertion events from 4 membranes; (C) sc₅MspA PN1: 221 insertions from 7 membranes; (D) sc₆MspA PN1: 463 insertion events from 7 different membranes; (E) sc₇MspA PN1: 62 insertions from 5 membranes; (F) sc₈MspA PN1: 150 insertions from 6 membranes.

protein sample as evident from western blot on Figure 3C. As a quality control we performed denaturing experiments with 80% DMSO. Only octameric MspA M2 dissociated into its monomeric subunits of 20 kDa while sc_8MspA_{dt} M2 was stable indicating that all eight subunits are covalently linked in the purified refolded protein (Fig. 3D). Overall, the combination of N- and C-terminal affinity tags enabled us to purify full-length $scMspA$ protein with a 20-fold increase in protein yield per single preparation in comparison to the previous gel excision method.

Channel-forming properties of single-chain MspA in large membranes.

To examine whether sc_8MspA_{dt} M2 forms functional channels we performed lipid bilayer experiments in a Montal-Mueller setup using cuvettes with an aperture diameter of 1 mm. As expected, refolded sc_8MspA_{dt} M2 had channel-forming activity in DPhPC membranes as shown by the stepwise current increase after addition of the protein (Fig. 3E). Analysis of 432 insertions from 8 different membranes showed predominant conductance peaks of 1.3 nS, 1.9 nS, and 3.0 nS (Fig. S4E). Notably, sc_8MspA_{dt} M2 had a broader conductance distribution than sc_8MspA PN1 (Fig. S5F). These differences may be the result of different purification methods, the absence of the P97F mutation near the constriction zone of sc_8MspA_{dt} M2 compared to sc_8MspA PN1, and/or the presence of purification tags on the termini of sc_8MspA_{dt} M2.

We recently demonstrated that octameric MspA M2 forms channels in bilayers composed of poly(1,2-butadiene)-b-poly(ethylene oxide) (PBD-PEO) polymer (10). PBD-PEO bilayers have a two- to three-fold increased lifetime and are more robust towards chemicals and high voltages than membranes made from biological lipids, and can be used in nanopore sequencing experiments (10). Thus, we also examined the channel properties of $scMspA$ insertions in PBD-PEO bilayers. Similarly to DPhPC bilayers, we observed insertions of sc_8MspA_{dt} M2 into PBD-PEO bilayers (Fig. 3F). Analysis of 207 insertions from 28 membranes showed broad distribution of single-channel conductances (Fig. S4F) in PBD-PEO bilayers similar to those obtained for sc_8MspA_{dt} M2 in DPhPC bilayers. It should be noted that insertion frequency of sc_8MspA_{dt} M2 was reduced in PBD-PEO bilayers in comparison to DPhPC membranes. Taken

together, these experiments showed that sc_8MspA_{dt} M2 forms functional channels in both DPhPC and PBD-PEO bilayers.

Hairpin translocation through the single-chain MspA pore.

To examine whether the nucleotide recognition capability of MspA is preserved in $scMspA$, we used sc_8MspA_{dt} M2 in DNA hairpins experiments as shown previously (12). This assay is based on distinct residual currents when the single-stranded homopolymer tail is located inside the constriction zone (Table S4), while the double-stranded region of the DNA hairpins is stalled in the lumen of the MspA pore and temporarily prevents translocation (Fig. 4A). Single sc_8MspA_{dt} M2 pores were inserted into poly(1,2-butadiene)-b-poly(ethylene oxide) (PBD₁₁-PEO₈) block-copolymer bilayer membranes with a diameter of 100 μ m and bathed in a buffer containing 10 mM Tris (pH 7.5), 1 M KCl and 2 M guanidinium chloride (GdmCl). Based on previous data (10) we used GdmCl in the electrolyte solution because it induces a semi-melting state of DNA hairpins enabling smooth DNA passage through the nanopore at low applied voltage resulting in increased DNA capture rate and decreased trapping time.

The current-voltage curves and power spectral density plots reveal that the conductance and trace noise are almost identical for $sc_8\text{MspA}_{dt}$ M2 and octameric MspA M2 (Fig. 4B). Both proteins exhibited gating when negative voltage was applied (Fig. S8A). Since the gating conductance varies for the same MspA pore even at the same voltage, we used the ungated current values for the IV curves resulting in identical symmetric IV curves for both $sc_8\text{MspA}_{dt}$ M2 and MspA M2 (Fig. 4B). Uncorrected IV curves for the same raw data are shown in Figure S8B. Interestingly, the protein concentration required to observe channel insertions was approximately 60-fold higher for $sc_8\text{MspA}_{dt}$ M2 (4.2 nM) than for octameric MspA M2 (0.063 nM). Consistent with previous results in large aperture bilayers (Fig. S4F), a wide conductance distribution was observed for individual $sc_8\text{MspA}_{dt}$ M2 channel ranging from 1 nS to 5.5 nS. However, only the pore with conductance of around 4.9 nS resulted in clean and deep DNA translocation signals. This appears to be a general property of the MspA pore as populations of pores with different conductance levels were previously observed in hairpin DNA experiments with octameric MspA isolated from *M. smegmatis* (25).

The current traces after addition of the poly-dT hairpin to $sc_8\text{MspA}_{dt}$ M2 and MspA M2 show current blockades resulting from translocation of the DNA hairpin through the pore in GdmCl-containing buffer (Fig. 4C, D). Similar current blockades were observed with the poly-dT hairpin and $sc_8\text{MspA}_{dt}$ M2 when no GdmCl was present in the electrolyte buffer (Fig. 4E). However, a higher voltage of 140 mV instead of 75 mV was necessary to overcome the energy barrier for DNA translocation in accordance with previous hairpin experiments with octameric MspA where 140 to 180 mV was used for hairpin DNA translocation (12). Our data further indicate that GdmCl improves both capture and translocation of hairpin DNA through the pore.

Nucleotide recognition by single-chain MspA.

To examine nucleotide recognition by $sc_8\text{MspA}_{dt}$ M2 we used DNA hairpins with a duplex region of 14 nucleotides and a homopolymer tail of 50 nt (hp14-dT50 (poly-dT), hp14-dA50 (poly-dA), hp14-dC27 (poly-dC), and hp14-dG3dA47 (poly-dG) (Table S3, Fig. S9). It should be noted that we used only three dG nucleotides in the “poly”dG hairpin tail in otherwise dA background to avoid G-tetrad formation as in our previous experiments (12). Histograms of the current change ΔI (current I in the presence of a hairpin minus the baseline current level I_0) are shown for the four individual DNA hairpins and mixture of four hairpins with octameric MspA M2 (Fig. 4F) and $sc_8\text{MspA}_{dt}$ M2 (Fig. 4G), respectively. Fitted Gaussians for each individual hairpin with a homopolymer tail were well-resolved and separated, and looked almost identical for both MspA M2 and $sc_8\text{MspA}_{dt}$ M2 (Fig. 4F, 4G, top panel). The partial overlap between purines was more pronounced for $sc_8\text{MspA}_{dt}$ M2, nonetheless, individual peaks for poly-dA and poly-dG were still distinguished (Fig. 4G, top panels). Remarkably, the histograms of DNA hairpin mixture clearly show four separate peaks representing the four different nucleotides, which match the current blockades for individual hairpin peaks (Fig. 4F, 4G, bottom panels). In this experiment $sc_8\text{MspA}_{dt}$ M2 had better resolving properties than octameric MspA M2. In conclusion, our DNA hairpin experiments showed that $sc_8\text{MspA}_{dt}$ M2 and octameric MspA M2 have very similar capabilities in distinguishing all four DNA nucleotides. These results indicate that covalently linking of all eight MspA subunits preserves the nucleotide recognition properties of the MspA pore and open many avenues to tailor the channel properties of the MspA pore for specific sensing applications.

DISCUSSION

Properties and potential applications of single-chain MspA nanopores.

In this study we showed that it is feasible to convert multimeric MspA into a functional monomeric nanopore with identical electrophysiological properties. This achievement is a major improvement over previous approaches based on self-assembly of wt and mutated subunits, which required the physical separation of pores with different subunit combinations lowering the yield dramatically, and suffered from different permutations with identical physical properties (3). By contrast, single-chain MspA can be produced as a full-length protein with many asymmetric mutations opening numerous avenues to improve the performance of MspA in DNA sequencing. For example: (i) Distinct amino acids in the MspA constriction zone will enable different chemical interactions with the DNA nucleobases. Previously, mutations that enable DNA translocation were present in all eight MspA monomers (11), making asymmetric interactions with DNA impossible. Asymmetric mutations open numerous avenues to increase the specificity of the current blockade for each nucleotide and to increase the dwell time in the construction zone. Both of these effects are likely to reduce the contribution of neighboring nucleotides to the current blockade, which is currently 4–5 nucleotides (30), and concomitantly reduce the high raw data error rates in nanopore sequencing (31). (ii) Single-chain MspA can be used to slightly alter the diameter of the pore to modulate the interactions between amino acids in the constriction zone and nucleotides. This can be achieved changing the subunit stoichiometry as shown in this study (Fig. 2). In a simple geometrical model adding or removing a subunit alters the diameter of the MspA pore by $1/8^{\text{th}}$, which is equivalent to approximately 1.5 Å, a reasonable range for significantly changing chemical interactions. Both the hexameric and heptameric pores are functional demonstrating that this approach is feasible. (iii) Motor proteins such as DNA polymerases employed to control the translocation rate of DNA through the pore have been instrumental for sequencing (13) and have become standard in nanopore sequencing devices (32). However, the long distance between the motor proteins and the MspA constriction zone decreases the positional precision and increases error due to thermal motions of the flexible DNA strand (33) and the motor protein. Random positioning of the motor protein on top of the MspA pore might enhance this error. A single cysteine in one of the surface loops of single-chain MspA would enable the covalent attachment of the motor protein, eliminating the random positioning of motor and assembly with the MspA pore, and thereby reducing the system variability. (iv) It might be possible to create a specific path for single-stranded DNA inside of the MspA pore, thereby decreasing the translocation velocity and eliminating the need for a motor protein altogether. While the possibilities to tailor the single-chain MspA pore for DNA sequencing by combining asymmetric mutations are almost endless, it would require an efficient assay to test these combinations. However, the currently used lipid bilayer systems are not high-throughput. By contrast, the single molecule measurements are tedious and usually require the analysis of several pores before a pore suitable for sequencing is identified. Hence, the recent developments of computer modelling for structure predictions (34,35) and virtual drug screening (36) could provide a valuable alternative to test and identify beneficial combinations of mutations of the single-chain MspA pore *in silico*.

Heterogeneous channel conductances of MspA pores.

Our study showed a wide variety of channel conductances for our single-chain MspA constructs ranging from 0.5 nS to 4.5 nS for sc₈MspA M2 without tags (Fig. S4D). This heterogeneity makes identification of single-chain MspA pores suitable for sequencing more difficult. However, this is not a consequence of the covalent linkers between the subunits of single-chain MspA as the channel conductances of the unmodified octameric MspA pore isolated from *M. smegmatis* also vary between 2 and 5 nS (Fig. S4A). These measurements are consistent with early measurements (6,37) indicating that the conductance variability is an intrinsic property of the MspA pore. The mutations introduced in MspA M1 and MspA M2 to enable DNA translocation (11) appear to increase the intrinsic conductance variability but shift the major conductance peak from 4.5 - 5 nS to app. 1.5 nS (Figs. S4B, C), indicating a significant contribution of the constriction zone to the overall channel conductance of MspA. It is plausible that neutralizing the constriction zone by mutating the negatively charged aspartates D90 and D91 to asparagines leads to a smaller and more flexible constriction zone due to the lack of electrostatic repulsion explaining both the shift to smaller channel conductances and their wider range. While we cannot exclude that the harsh procedures used to extract MspA from *M. smegmatis* such as organic solvents (6) or boiling in detergents (27) contribute to pore heterogeneity, it is striking that the recombinant single-chain MspA M2 pores, which were isolated from *E. coli*, showed a similar conductance profile as octameric MspA M2 isolated from *M. smegmatis*. Thus, it appears that pore conductance heterogeneity is largely an intrinsic property of MspA, driven mainly by the properties of the constriction zone. The hypothesis that the flexible constriction zone of the MspA variants M1 and M2 used in DNA sequencing experiments and in the corresponding single-chain variant sc₈MspA M2 is major driver of the conductance heterogeneity of these pores is supported by an additional experimental observation. The MspA PN1 variant with a phenylalanine 97 at the tip of the loop 6, which connects the constriction zone with the subsequent β -strand (38,39), has a much smaller conductance distribution (Fig. S5A), indicating that putative hydrophobic interactions of neighboring phenylalanines stabilize the constriction zone. Conductance heterogeneity is common in general diffusion pores from other bacteria. For example, OmpF, the main porin of *E. coli*, and porins of *Salmonella* and *Pseudomonas aeruginosa* show a broad range of channel conductances (40-43). By contrast, the channel conductances are much better defined for specific porins, which have specific substrate binding sites, and the much smaller ion channels, which are located in the cytoplasmic membrane. For example, LamB, a maltodextrine-specific porin of *E. coli* (44-46), OprO, a phosphate-specific porin of *P.aeruginosa* (47), and the bacterial amyloid secretion channel CsgG (48) have narrow conductance distributions.

Single-chain MspA pores with different subunits stoichiometries.

In this study we constructed scMspA pores with different subunit stoichiometries to demonstrate the feasibility of changing the channel diameter, an important feat for nanopore sensing applications. All four single-chain MspA constructs with less than eight subunits formed functional pore proteins despite the likely limited tolerance of the MspA structure, in particular of the β -barrel, for expansions or reductions of the number of subunits. This is probably due to the self-assembly of the scMspA constructs with three and five subunits considering that the efficiency of the assembly of covalently linked subunits to a functional pore probably decreases with larger deviations from the octameric pore, which was found previously to be the most

dominant form in a self-assembly process with the purified MspA monomer (38). Interestingly, the sc₈MspA PN1 pore had a much broader channel distribution than octameric MspA PN1 indicating that the purification and refolding procedure of recombinant scMspA introduced channel heterogeneity, which is not observed for octameric MspA PN1 purified from *M. smegmatis*. Thus, we expected that sc₃MspA PN1 assembles mostly to a hexameric structure. This is consistent with the conductance profile obtained for sc₃MspA PN1, which resembles that of sc₆MspA PN1 (Fig. S5B, D). The few additional larger channels for sc₃MspA PN1 might have been from a nonameric MspA pore. The conductance distribution of sc₅MspA PN1 is shifted to larger values compared to sc₈MspA PN1 indicating the presence of mainly decameric MspA pores. The conductance profile of sc₆MspA PN1 is similar to that of sc₈MspA PN1, while the conductances of sc₇MspA are shifted to larger values indicating assembly to larger pores (Fig. S5). Structural analysis of the single-chain variants by electron microscopy will be necessary to obtain quantitative results about different scMspA assembly forms for each subunit composition and exact constriction zone diameters of the scMspA pores.

Conclusions.

This study demonstrated the feasibility of producing single-chain MspA pores and their use in DNA sequencing. The unique adaptability of the single-chain MspA pore as a biosensor by making asymmetric mutations has enormous potential to improve the DNA sequencing capability of MspA and for many other applications such as detection of RNA (49), proteins (50) and small molecules (51). The single-chain concept might be applicable to other oligomeric protein pores such as CsgG(52), ClyA (53), FraC(54) and α -HL(55), depending on the proximity of the C-terminus of one monomer to the N-terminus of the next monomer in the pore structure.

ACKNOWLEDGEMENTS

This work was supported by the National Institutes of Health grant R21 HG010543 to M.N. The funding source had no involvement in preparation of the article, study design, in the collection, analysis and interpretation of data; in the writing of the report; and in the decision to submit the article for publication.

This article contains supporting information.

REFERENCES

1. Bezrukov, S. M., and I. Vodyanoy. 1993. Probing alamethicin channels with water-soluble polymers. Effect on conductance of channel states. *Biophys J.* 64(1):16-25.
2. Nekolla, S., C. Andersen, and R. Benz. 1994. Noise analysis of ion current through the open and the sugar-induced closed state of the LamB channel of Escherichia coli outer membrane: evaluation of the sugar binding kinetics to the channel interior. *Biophys J.* 66(5):1388-1397.
3. Bayley, H., and C. R. Martin. 2000. Resistive-Pulse Sensing-From Microbes to Molecules. *Chem Rev.* 100(7):2575-2594.
4. Deamer, D., M. Akeson, and D. Branton. 2016. Three decades of nanopore sequencing. *Nat Biotechnol.* 34(5):518-524.
5. Kasianowicz, J. J., E. Brandin, D. Branton, and D. W. Deamer. 1996. Characterization of individual polynucleotide molecules using a membrane channel. *Proc Natl Acad Sci U S A.* 93(24):13770-13773.
6. Niederweis, M., S. Ehrt, C. Heinz, U. Klöcker, S. Karosi, K. M. Swiderek, L. W. Riley, and R. Benz. 1999. Cloning of the *mspA* gene encoding a porin from *Mycobacterium smegmatis*. *Mol. Microbiol.* 33(5):933-945.
7. Heinz, C., H. Engelhardt, and M. Niederweis. 2003. The core of the tetrameric mycobacterial porin MspA is an extremely stable beta-sheet domain. *J. Biol. Chem.* 278:8678-8685.
8. Niederweis, M., C. Heinz, K. Janik, and S. H. Bossmann. 2001. Nanostructuring of carbon surfaces by deposition of a channel-forming protein and subsequent polymerization of methyl-methacrylate-prepolymers. *Nano Letters.* 1:169-174.
9. Niederweis, M., C. Heinz, K. Janik, and S. H. Bossmann. 2002. Nanostructuring by deposition of protein channels formed on carbon-surfaces. *Nano Letters.* 2(11):1263-1268.
10. Yu, L., X. Kang, M. A. Alibakhshi, M. Pavlenok, M. Niederweis, and M. Wanunu. 2021. Stable polymer bilayers for protein channel recordings at high guanidinium chloride concentrations. *Biophys J.* 120:1-5.
11. Butler, T. Z., M. Pavlenok, I. M. Derrington, M. Niederweis, and J. H. Gundlach. 2008. Single-molecule DNA detection with an engineered MspA protein nanopore. *Proc Natl Acad Sci U S A.* 105(52):20647-20652.
12. Derrington, I. M., T. Z. Butler, M. D. Collins, E. Manrao, M. Pavlenok, M. Niederweis, and J. H. Gundlach. 2010. Nanopore DNA sequencing with MspA. *Proc Natl Acad Sci U S A.* 107(37):16060-16065.
13. Manrao, E. A., I. M. Derrington, A. H. Laszlo, K. W. Langford, M. K. Hopper, N. Gillgren, M. Pavlenok, M. Niederweis, and J. H. Gundlach. 2012. Reading DNA at single-nucleotide resolution with a mutant MspA nanopore and phi29 DNA polymerase. *Nat Biotechnol.* 30(4):349-353.

14. Cherf, G. M., K. R. Lieberman, H. Rashid, C. E. Lam, K. Karplus, and M. Akeson. 2012. Automated forward and reverse ratcheting of DNA in a nanopore at 5-A precision. *Nat Biotechnol.* 30(4):344-348.
15. Brown, C. G., and J. Clarke. 2016. Nanopore development at Oxford Nanopore. *Nat Biotechnol.* 34(8):810-811.
16. Franceschini, L., T. Brouns, K. Willems, E. Carlon, and G. Maglia. 2016. DNA Translocation through Nanopores at Physiological Ionic Strengths Requires Precise Nanoscale Engineering. *ACS Nano.* 10(9):8394-8402.
17. Wloka, C., N. L. Mutter, M. Soskine, and G. Maglia. 2016. Alpha-Helical Fragaceatoxin C Nanopore Engineered for Double-Stranded and Single-Stranded Nucleic Acid Analysis. *Angew Chem Int Ed Engl.* 55(40):12494-12498.
18. Cao, C., Y. L. Ying, Z. L. Hu, D. F. Liao, H. Tian, and Y. T. Long. 2016. Discrimination of oligonucleotides of different lengths with a wild-type aerolysin nanopore. *Nat Nanotechnol.* 11(8):713-718.
19. Zhou, W., H. Qiu, Y. Guo, and W. Guo. 2020. Molecular Insights into Distinct Detection Properties of alpha-Hemolysin, MspA, CsgG, and Aerolysin Nanopore Sensors. *J Phys Chem B.* 124(9):1611-1618.
20. Noakes, M. T., H. Brinkerhoff, A. H. Laszlo, I. M. Derrington, K. W. Langford, J. W. Mount, J. L. Bowman, K. S. Baker, K. M. Doering, B. I. Tickman, and J. H. Gundlach. 2019. Increasing the accuracy of nanopore DNA sequencing using a time-varying cross membrane voltage. *Nat Biotechnol.* 37(6):651-656.
21. Carter, J. M., and S. Hussain. 2017. Robust long-read native DNA sequencing using the ONT CsgG Nanopore system. *Wellcome Open Res.* 2:23.
22. Agah, S., M. Zheng, M. Pasquali, and A. B. Kolomeisky. 2016. DNA sequencing by nanopores: advances and challenges. *J Phys D Appl Phys.* 49:413001.
23. Derrington, I. M., J. M. Craig, E. Stava, A. H. Laszlo, B. C. Ross, H. Brinkerhoff, I. C. Nova, K. Doering, B. I. Tickman, M. Ronaghi, J. G. Mandell, K. L. Gunderson, and J. H. Gundlach. 2015. Subangstrom single-molecule measurements of motor proteins using a nanopore. *Nat Biotechnol.* 33(10):1073-1075.
24. Schreiber, J., Z. L. Wescoe, R. Abu-Shumays, J. T. Vivian, B. Baatar, K. Karplus, and M. Akeson. 2013. Error rates for nanopore discrimination among cytosine, methylcytosine, and hydroxymethylcytosine along individual DNA strands. *Proc Natl Acad Sci U S A.* 110(47):18910-18915.
25. Pavlenok, M., I. M. Derrington, J. H. Gundlach, and M. Niederweis. 2012. MspA nanopores from subunit dimers. *PLoS One.* 7(6):e38726.
26. Distefano, M. D., H. Kuang, D. Qi, and A. Mazhary. 1998. The design of protein-based catalysts using semisynthetic methods. *Curr Opin Struct Biol.* 8(4):459-465.
27. Heinz, C., and M. Niederweis. 2000. Selective extraction and purification of a mycobacterial outer membrane protein. *Anal. Biochem.* 285(1):113-120.

28. Prilipov, A., P. S. Phale, P. Van Gelder, J. P. Rosenbusch, and R. Koebnik. 1998. Coupling site-directed mutagenesis with high-level expression: large scale production of mutant porins from *E. coli*. *FEMS Microbiol Lett.* 163(1):65-72.
29. Pavlenok, M., and M. Niederweis. 2016. Hetero-oligomeric MspA pores in *Mycobacterium smegmatis*. *FEMS Microbiol Lett.* 363(7).
30. Manrao, E. A., I. M. Derrington, M. Pavlenok, M. Niederweis, and J. H. Gundlach. 2011. Nucleotide discrimination with DNA immobilized in the MspA nanopore. *PLoS One.* 6(10):e25723.
31. Dohm, J. C., P. Peters, N. Stralis-Pavese, and H. Himmelbauer. 2020. Benchmarking of long-read correction methods. *NAR Genom Bioinform.* 2(2):lqaa037.
32. Jain, M., H. E. Olsen, B. Paten, and M. Akeson. 2016. The Oxford Nanopore MinION: delivery of nanopore sequencing to the genomics community. *Genome Biol.* 17(1):239.
33. Lu, B., S. Fleming, T. Szalay, and J. Golovchenko. 2015. Thermal Motion of DNA in an MspA Pore. *Biophys J.* 109(7):1439-1445.
34. Yang, J., I. Anishchenko, H. Park, Z. Peng, S. Ovchinnikov, and D. Baker. 2020. Improved protein structure prediction using predicted interresidue orientations. *Proc Natl Acad Sci U S A.* 117(3):1496-1503.
35. Senior, A. W., R. Evans, J. Jumper, J. Kirkpatrick, L. Sifre, T. Green, C. Qin, A. Zidek, A. W. R. Nelson, A. Bridgland, H. Penedones, S. Petersen, K. Simonyan, S. Crossan, P. Kohli, D. T. Jones, D. Silver, K. Kavukcuoglu, and D. Hassabis. 2020. Improved protein structure prediction using potentials from deep learning. *Nature.* 577(7792):706-710.
36. Pereira, J. C., E. R. Caffarena, and C. N. Dos Santos. 2016. Boosting Docking-Based Virtual Screening with Deep Learning. *J Chem Inf Model.* 56(12):2495-2506.
37. Stahl, C., S. Kubetzko, I. Kaps, S. Seeber, H. Engelhardt, and M. Niederweis. 2001. MspA provides the main hydrophilic pathway through the cell wall of *Mycobacterium smegmatis*. *Mol. Microbiol.* 40(2):451-464 (Authors' correction appeared in *Mol. Microbiol.* **457**, 1509).
38. Faller, M., M. Niederweis, and G. E. Schulz. 2004. The structure of a mycobacterial outer-membrane channel. *Science.* 303(5661):1189-1192.
39. Huff, J., M. Pavlenok, S. Sukumaran, and M. Niederweis. 2009. Functions of the periplasmic loop of the porin MspA from *Mycobacterium smegmatis*. *J Biol Chem.* 284(15):10223-10231.
40. Benz, R., K. Janko, W. Boos, and P. Lauger. 1978. Formation of large, ion-permeable membrane channels by the matrix protein (porin) of *Escherichia coli*. *Biochim Biophys Acta.* 511(3):305-319.
41. Benz, R., J. Ishii, and T. Nakae. 1980. Determination of ion permeability through the channels made of porins from the outer membrane of *Salmonella typhimurium* in lipid bilayer membranes. *J. Membr. Biol.* 56(1):19-29.
42. Benz, R., K. Janko, and P. Lauger. 1980. Pore formation by the matrix protein (porin) of *Escherichia coli* in planar bilayer membranes. *Ann N Y Acad Sci.* 358:13-24.

43. Benz, R., and R. E. Hancock. 1981. Properties of the large ion-permeable pores formed from protein F of *Pseudomonas aeruginosa* in lipid bilayer membranes. *Biochim. Biophys. Acta.* 646(2):298-308.
44. Nakae, T. 1979. A porin activity of purified lambda-receptor protein from *Escherichia coli* in reconstituted vesicle membranes. *Biochem. Biophys. Res. Commun.* 88(3):774-781.
45. Nakae, T., and J. Ishii. 1980. Permeability properties of *Escherichia coli* outer membrane containing, pore-forming proteins: comparison between lambda receptor protein and porin for saccharide permeation. *J. Bacteriol.* 142(3):735-740.
46. Klebba, P. E. 2002. Mechanism of maltodextrin transport through LamB. *Res Microbiol.* 153(7):417-424.
47. Hancock, R. E., C. Egli, R. Benz, and R. J. Siehnel. 1992. Overexpression in *Escherichia coli* and functional analysis of a novel PPI-selective porin, OprO, from *Pseudomonas aeruginosa*. *J Bacteriol.* 174(2):471-476.
48. Goyal, P., P. V. Krasteva, N. Van Gerven, F. Gubellini, I. Van den Broeck, A. Troupiotis-Tsailaki, W. Jonckheere, G. Pehau-Arnaudet, J. S. Pinkner, M. R. Chapman, S. J. Hultgren, S. Howorka, R. Fronzes, and H. Remaut. 2014. Structural and mechanistic insights into the bacterial amyloid secretion channel CsgG. *Nature.* 516(7530):250-253.
49. Garalde, D. R., E. A. Snell, D. Jachimowicz, B. Sipos, J. H. Lloyd, M. Bruce, N. Pantic, T. Admassu, P. James, A. Warland, M. Jordan, J. Ciccone, S. Serra, J. Keenan, S. Martin, L. McNeill, E. J. Wallace, L. Jayasinghe, C. Wright, J. Blasco, S. Young, D. Brocklebank, S. Juul, J. Clarke, A. J. Heron, and D. J. Turner. 2018. Highly parallel direct RNA sequencing on an array of nanopores. *Nat Methods.* 15(3):201-206.
50. Nivala, J., L. Mulrone, G. Li, J. Schreiber, and M. Akeson. 2014. Discrimination among protein variants using an unfoldase-coupled nanopore. *ACS Nano.* 8(12):12365-12375.
51. Guan, X., L. Q. Gu, S. Cheley, O. Braha, and H. Bayley. 2005. Stochastic sensing of TNT with a genetically engineered pore. *ChemBiochem.* 6(10):1875-1881.
52. Cao, B., Y. Zhao, Y. Kou, D. Ni, X. C. Zhang, and Y. Huang. 2014. Structure of the nonameric bacterial amyloid secretion channel. *Proc Natl Acad Sci U S A.* 111(50):E5439-5444.
53. Mueller, M., U. Grauschopf, T. Maier, R. Glockshuber, and N. Ban. 2009. The structure of a cytolytic alpha-helical toxin pore reveals its assembly mechanism. *Nature.* 459(7247):726-730.
54. Tanaka, K., J. M. Caaveiro, K. Morante, J. M. Gonzalez-Manas, and K. Tsumoto. 2015. Structural basis for self-assembly of a cytolytic pore lined by protein and lipid. *Nat Commun.* 6:6337.
55. Song, L., M. R. Hobaugh, C. Shustak, S. Cheley, H. Bayley, and J. E. Gouaux. 1996. Structure of staphylococcal alpha-hemolysin, a heptameric transmembrane pore. *Science.* 274(5294):1859-1866.