Engineering biological nanopores for proteomics study

Huang, Kevin

DOI:
10.33612/diss.102598418

Summary

1. Conclusion

Biological nanopores can be prepared with a high reproducibility and be engineered at the atomic level. Nanopores have been successfully implemented for DNA sequencing. However, nanopore is much less employed for the protein analysis or sequencing due to the high order structure, different charge conditions of protein, which is composed of 20 different amino acids. By contrast, DNA only contains four types of nucleotides and has a uniform negative charge. In this thesis, we proposed a new way to sequence protein with nanopore in a manner similar with traditional mass spectrometry sequencing. The protein could be digested into peptides with enzymes and the peptide mass identified with nanopores instead of mass spectrometry. We named this system a nanopore peptide mass identifier. Comparing with the traditional mass spectrometry detection, nanopore peptide identifier is a single molecule technique which could improve the detection sensitivity, making it possible to detect the ultra-low abundance proteins and post translation modifications (PTMs). To make the identifier feasible, we first worked on the method to capture all peptides with different charge conditions into the nanopores by using Fragaceatoxin C (FraC) nanopores (chapter 2). Wild type FraC nanopores could detect peptides down to 10 AA but the peptide shorter than that translocated too fast to be detected. Therefore, we also focused on engineering the size of FraC nanopores for small pores (chapter 3). More importantly, there should be a correlation between the nanopore ionic current signal and peptides mass if we ought to identify an unknown peptide to make the nanopore peptide mass identifier (chapter 3). Not limited to the protein sequencing endeavors, the two-component pleurotolysin nanopore (PlyAB) with a 5.5 nm constriction site, has been reconstituted and engineered for large folded protein detection (chapter 4). This work has made an important contribution to the nanopore-based proteomics. Several key conclusions from this project are summarized below along with future perspectives.

Conclusion 1: The uniform capture of protein or peptide with different charges was efficiently achieved by tuning the electroosmotic flow and charge repulsion.

The different charge of peptide doesn't allow uniform capture by using electrophoresis. We demonstrated the electroosmotic flow in FraC nanopores can be engineered to allow capturing peptides with different charges. Wild type FraC nanopore contains a negatively charged constriction, hence it is cation selective. At neutral pH, the electroosmotic flow (EOF) generated by FraC

nanopores is not strong enough to capture the negatively charged peptides against the electrophoretic force. By lowering the pH to 4.5, the EOF was reduced due to the loss of negative charge in the constriction, and meanwhile, the negative charge density of the peptide was also diminished. Thus, the charge repulsion was attenuated together with the opposite electrophoretic force for a negatively charged peptide at pH 4.5. In our experiment, the reduced electroosmotic flow became dominating and was able to capture the negatively charged peptide into FraC. In such condition, the electrophoretic force and electroosmotic flow imposed on a positively charged peptide were in the same direction, which enabled positively charged peptides to be captured easily. Therefore, we achieved the uniform capture of peptides with different original charge compositions.

Conclusion 2: The nanopore signal measured by FraC nanopores could be correlated with the peptide mass at pH 3.8 towards a nanopore peptide mass identifier.

Since different charged peptides could be captured into FraC nanopores at pH 4.5, the next question is how to fulfill the identification of an unknown peptide. We looked into the relationship between the ionic current of peptide blockades and the peptide property. In theory, the excluded current should be proportional to the excluded volume of the peptide when it is traversing the pore, and the volume is also related to the mass. We observed the correlation between ionic signal and peptide mass for most of the peptides except two negatively charged peptides at pH 4.5. Furthermore, by lowering the pH to 3.8, we observed a correlation between the excluded current and peptide mass for all peptides we tested. With the correlation available, the mass of an unknown peptide can be identified with the ionic current measured with nanopore. In this way, FarC nanopores can be used as peptide mass identifiers.

Conclusion 3: The size of FraC nanopores could be engineered by manipulation of the interaction between the FraC monomer and lipid bilayer.

One of the disadvantages of biological nanopores is their fixed size. As for wild type FraC, it can only sense the peptides with molecular weight higher than 1.2 kDa. The peptides smaller than that translocate too fast to be properly sampled. Therefore, it is necessary to engineer biological nanopores with a smaller size for the sensing for short peptides. And engineering smaller biological nanopore is a challenge.

In this study, we elucidated a new strategy to change the oligomeric composition of FraC nanopores by weakening the interaction between the monomer and the lipid bilayer. FraC monomers bind to sphingomyelin lipid and assemble into oligomers. Crystal structure of FraC revealed several aromatic amino acids located in the lipid binding interface, which are crucial for the interaction of FraC and the membrane. Therefore, we replaced the tryptophan residues that are important for lipid binding to serine. In this way, the effective concentration of FraC monomer bound with lipid membrane was reduced and rewardingly two types of FraC nanopores with smaller sizes were prepared. The new types of FraC nanopores most likely correspond to different oligomers with fewer subunits. Thus, this method presented a simple way to engineer the pore size and without changing the pore lumen geometry. The type III pore could potentially be the smallest nanopore ever created for electrical sensing with high reproducibility.

Conclusion 4: New biological nanopore with a bigger pore size was reconstituted into the standard lipid bilayer for folded protein detection.

β-barrel biological nanopores usually have a narrow pore size (1~2 nm in diameter), which is ideal for the purpose of sequencing polymers by threading them through the pore. But the small size also makes it difficult to study large folded proteins. Although there are many large protein channels in nature, the concern is that the large membrane channels probably form inhomogeneous pores with variable number of subunits and unpredictable sizes, which is problematic for electrophysiology sensing. Moreover, these pores normally have specific requirements about the chemical composition of lipid membrane to assemble and carry out their functions properly, making it challenging to reconstitute them into the artificial lipid bilayer for electrophysiology sensing.

Here, we successfully reconstituted the two component pleurotolysin nanopores (PlyAB) into a standard lipid bilayer. The PlyAB nanopores is a large protein complex with 39 monomers and a total molecular weight around 1.14 MDa. We observed a very homogeneous channel conductance distribution in electrophysiological experiments. PlyAB nanopore is the largest biological nanopores used for sensing with a diameter of 5.5 nm in its narrowest site, which is approximately twice as big as ClyA nanopores that is the only other biological nanopore capable to accommodate small folded proteins. The properties of PlyAB, such as the pore stability, ion selectivity, protein expression, have been successfully alerted and improved by directed evolution together with rational site-directed mutagenesis. With a bigger size, PlyAB nanopores

were able to detect and discriminate large plasma proteins with molecular weight up to 80 kDa. It provides a new tool for the sensitive detection of folded proteins and characterization of the nanofluidic property of large biological nanopores.

2. Future perspective

Since the correlation between ionic signal and peptide mass has been established, the next task for developing a nanopore-based peptide mass spectrometer is to improve the resolution of peptide discrimination. Currently, we can discriminate between aspartic acid and alanine (44 Da), however identifying smaller difference is still quite challenging. The main factor defining the resolution is the signal homogeneity of peptide. Therefore, engineering the spread of peptide signal to be as narrow as possible is the next priority. Another interesting question is why the correlation between the excluded current and the peptide mass doesn't exist anymore when the pH lower than 3.8. One possibility is the loss of charge interaction between the peptides and the FraC constriction, when aspartic acid in position 10 is protonated when pH lower than 3.8, which causes the shape or conformation change of peptides. The behavior and conformation of a peptide inside the nanopore have not been fully understood yet. Therefore, MD simulations might provide valuable information. Experimentally, the unnatural amino acid with extremely low pKa could be incorporated into the FraC constriction. In such scenario, when pH lower than 3.8, nanopore can still preserve some negative charge to check if the charge interaction is crucial for the correlation.

With the FraC peptide mass identifier available, the final goal of sequencing a protein is to digest the whole protein into peptides and measure with the FraC nanopores. Different proteases could be employed to digest proteins. In proof-of-concept experiments, the digested peptide sequences are predicted and confirmed by mass spectrometry to compare with the signals measured with nanopores. Ideally, the peptide mass extracted from FraC nanopore measurement should be identical to the real peptide mass. It's most likely that more complicated data analysis methods as machine learning is necessary to help to define the digested peptides sequence, considering the sequence complexity and current resolution of peptide separation with nanopores.

Despite the pre-digestion of a protein, a real-time protein sequencer might also be built by anchoring different enzymes such as unfoldase and protease on top of the FraC peptide mass identifier. There are several protein digestion

machines in the cells such as the AAA+ protease ClpXP system, which also assembles into homomeric rings and hence is possible to match the nanopore geometry. The protease might need to be attached chemically or genetically to the nanopores to prevent dissociation. In this way, a protein can be unfolded, digested and the peptide measured by nanopores in real-time.

PlyAB nanopores could be exploited more for other folded proteins identification or function evaluation. Protein-protein interaction (PPI), small molecule detection and single molecule enzymology are all interesting topics for further PlyAB applications. But the trapping or retaining of proteins in the lumen of PlyAB might require further engineering of the internal charge condition and nanofluidic property. For example, an overall positively charged vestibule is needed to allow the dwelling of negatively charged protein inside the pore.