Accurate and Precise Predictions of the Influence of Salt Concentration on the Conformational Stability and Membrane-Binding Modes of Multifunctional DNA Nanopores using Ensemble-Based Coarse-Grained Molecular Dynamics

Katya Ahmad[1,2], Stefan Howorka[2] and Peter V. Coveney [1,3]

[1]Centre for Computational Science, Department of Chemistry, University College London, London WC1H 0AJ, United Kingdom

[2]Institute of Structural Molecular Biology, Department of Biological Sciences, Birkbeck, University of London, Malet Street, London WC1E 7HX, United Kingdom

[3]Institute of Informatics, University of Amsterdam, 1098 XH Amsterdam, Netherlands

1. Introduction

Pore-forming protein analogues have been fabricated from triethylene glycol-cholesterol modified DNA sequences, which hybridize to form cholesterol anchored DNA nanopores (TEG-C NP's). These versatile nanopores can be chemically tuned to exhibit an array of functionalities with a broad range of potential applications in biomedicine e.g. novel ligand-controlled and light-controlled drug delivery systems[1,2,3]. The interactions between TEG-C NP's and membrane lipids are pivotal to their function, but these interactions remain poorly understood. Here we use an ensemble-based, coarse-grained molecular dynamics (CG-MD) protocol to gather detailed, reproducible data on the structure and dynamics of TEG-C NP's at two experimentally relevant ionic concentrations, allowing us to calculate reliable pore dimensions and perform comprehensive fluctuation analyses on membrane-spanning TEG-C NP's, as well as TEG-C NP's in free solution. Thus we can confidently characterise the influence of ionic concentration and membrane encapsulation on the dimensions, structural and mechanical properties of TEG-C NP's, and pinpoint areas of constriction, strain and stability within their structure. Collecting ensembles of micro-second long trajectories of a membrane-spanning TEG-C NP allows us to observe a comprehensive spread of large-scale motions available to the TEG-C NP at these timescales and draw parallels with what is observed in experiment.

2. Simulation Protocols

The reduced complexity of CG methods increases the feasibility of running larger ensembles of potentially lengthy simulations at a reasonable cost in a systematic and reproducible fashion, using a comparatively small amount of wall clock time. Results published in most biomolecular MD studies (both all-atom (AA) and CG) are usually based on macroscopic properties that are calculated by averaging over some number of repeat simulations. Individual MD trajectories are highly sensitive to their starting conditions, and the temporal evolution of an individual trajectory is stochastic, meaning that neighbouring trajectories with different initial velocities diverge quickly. One cannot assume that the conformational space has been sufficiently sampled from a handful of repeat simulations without performing appropriate error analyses

on the computed data to determine whether or not the computed average is representative of the true ensemble average. Therefore, the reliability of the results produced in such studies is uncertain. In this work, we address this issue by employing an ensemble-based protocol, according to which a set of N concurrent "replicas" are run, producing a stable ensemble average and associated fluctuations such that running N+1 would not alter the behaviour significantly. Running ensembles of computationally inexpensive CG simulations enhances the sampling of conformational space, giving reliable results in a rapid and reproducible way. For uncertainty quantification, we have employed the bootstrap method for calculating the standard error associated with ensemble-averaged macroscopic properties of the CG TEG-C NP systems; namely the pore height (for all four models), and the bilayer thickness (for the membrane models). This statistical method has been utilized successfully in many ensemble-based all-atom MD studies [4][5], where the typical simulation duration is fairly short (4 – 20 ns), but its use in CG studies at longer timescales has not been reported previously. In this work, we have achieved similar success to previous AA studies, in terms of error control and reproducibility. For this investigation, a total of four TEG-C NP models were built to represent four sets of experimental conditions; a free solvated TEG-C NP in 0.3 M NaCl solution (low salt) and another in 1.0 M NaCl (high salt), and a POPC membrane-spanning TEG-C NP in 0.3 M NaCl solution, and another in 1 M NaCl (high-salt). The membrane models were built by inserting the equilibrated CG TEG-C NP model into the centre of a pre-equilibrated CG POPC bilayer patch, which was solvated within a box of polarized MARTINI water molecules with the *insane.py* script. All CG production simulations were performed with GROMACS v5.1.4. Temperature control was achieved with the use of the velocity-rescale thermostat, and the Parrinnelo-Rahman barostat was used for pressure control. Replica production simulations were run by generating separate run files with different random velocity seeds. Production simulations were then run for up to 1 μs, with temperature and pressure held at 300K and 1.013 bar, respectively. All simulations were performed on tier-1 supercomputers, namely ARCHER (UK) and Cartesius (Netherlands).

3. Results and Discussion

Overall, the TEG-C NP adopts a highly asymmetrical, bloated conformation in solution; with regions of constriction at the pore termini where the inter-duplex polyT crossovers are located. Increasing the salt concentration from 0.3 M to 1.0 M had little effect on the average pore dimensions; in both cases the average pore height remained stable at around 7.6 nm, and the pore width at around 7.5 nm. These values are in fairly close agreement with the predicted dimensions (9 nm by 5 nm) and with experimental AFM measurements. [2] It should be noted that predicted values do not account for the effects of solvent/salt concentration, while AFM measurements of these delicate DNA nanostructures are prone to tip convolution effects. Minimum pore width values were calculated for both models and, in both sets of simulations, the minimum values were clustered at the pore termini; specifically, the R8 and R1 regions (illustrated in Fig.1), which were identified as sites of significant constriction. In the low-salt membrane simulations, there was a high tendency for the TEG-C NP to pop out of the bilayer and bind to the bilayer surface. This occurred in ten out of fifteen trajectories, the expulsion

process typically commencing after around 400 ns of simulation time. In experimental studies, TEG-C NPs have been shown to exhibit a strong preference for curved membranes over planar membranes, and insertion into planar membranes in similar conditions (0.3 M KCl buffer) has been reported to be very slow. [1,2] This is possibly indicative of a high energy barrier associated with insertion of the NP into planar membranes. While we cannot comment on the thermodynamics associated with membrane insertion, the expulsion of the NP in these simulations suggests that the artificial starting configuration imposed on the membrane-spanning NP is unfavourable in low salt conditions. The TEG-C NP exits the bilayer in a series of steps (Fig. 2) during which hydrophobic matching effects (i.e. NP twisting, tilting and breathing motions) are observed. The expulsion process appears to be the time reverse of the proposed two-step insertion mechanism, where the initial side-on binding is fast and favourable, while the transmembrane insertion is rate-limiting. [1] This behaviour is not observed in the 1.0 M NaCl model, suggesting that the initial transmembrane binding mode is more favourable at higher salt concentrations. The stark contrast between the RMSF-per-residue plots for two membrane models provides some evidence to support this theory. The incidence of extreme RMSF values in the high-salt model is very low compared to the low-salt model, and even the most extreme value (0.35 nm) is around 0.1 nm lower than the median value in the 0.3 M plot. This indicates that the NP is far less mobile within the bilayer, as the midsection of the NP experiences compression and immobilization within the membrane. This compression in the membrane model is made possible by the enhanced electrostatic screening of the 1.0 M NaCl solution, which lessens the repulsion between adjacent DNA helices, allowing them to be pushed together to form a compact, slightly ellipsoidal lumen. It can be deduced that 0.3 M NaCl does not provide sufficient electrostatic screening, so inter-helix repulsion within the TEG-C NP is too large to withstand the compressive force of the membrane, causing it to bind to the bilayer side-on instead. This is reflected in the calculated minimum and maximum pore width values (Fig. 1). In the high-salt simulations, the width of the R8 constriction is significantly narrower than in the low-salt simulations, and the same is true for the maximum width R4.

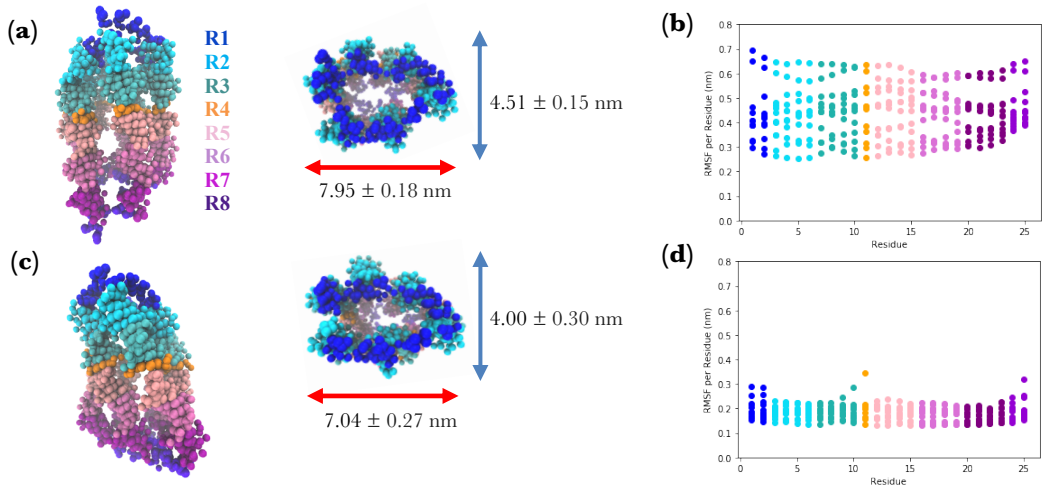

Figure 1. *(a) Representative structure of the membrane-spanning TEG-C NP in 0.3 M NaCl. (b) Corresponding RMSF per residue plot. (c) Representative structure of the membrane-spanning TEG-C NP in 1.0 M NaCl (d) Corresponding RMSF per residue plot. The NP was split into eight distinct regions during fluctuation analysis, indicated by colour coding. Maximum pore width values are denoted by red arrows, and minimum values are denoted by blue arrows.*

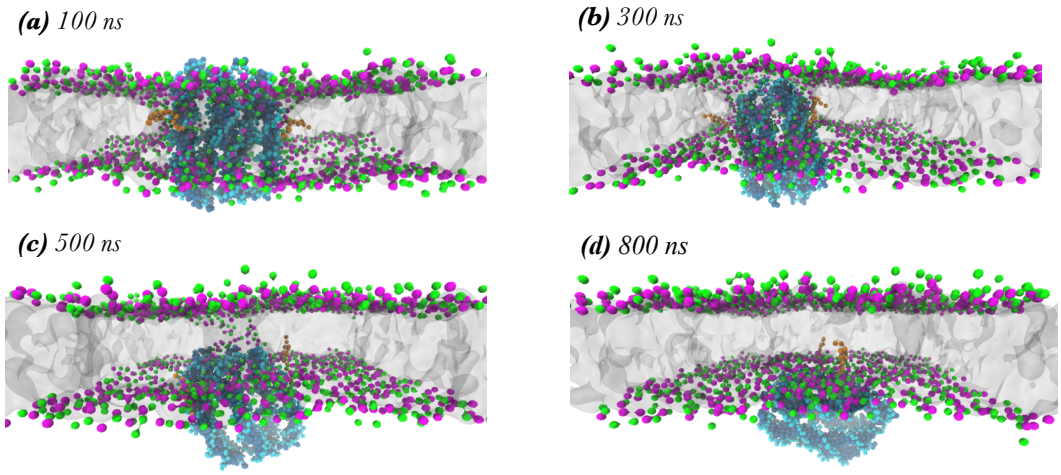

Figure 2. *Snapshots taken from a simulation in which the TEG-C NP departs from the bilayer in a multi-step process. (a) The NP spans the membrane normally. (b) Membrane lipids reorganise, and the NP descends from the centre of the bilayer. (c) The NP tilts and rotates, moving further out of the membrane. (d) The DNA backbone aligns with the POPC headgroups, and the TEG-C NP is held on the lower surface of the bilayer by two cholesterol anchors.*

4. Conclusions

In conclusion, our ensemble-based CG MD protocol has allowed us to accurately predict the dimensions, structural properties, and membrane-binding modes of an archetypal DNA nanopore in different salt conditions. The computed dimensions are in very good agreement with experimentally-derived values, and the collated trajectory data from our CG simulations are reliable and reproducible. Ongoing ionic current simulations will provide average nanopore conductance values for comparison against existing experimental values. We can explain the influence of salt concentration on the structural and mechanical properties of multifunctional DNA nanopores, and how these properties may dictate the behaviour of these nanopores in the presence of membranes. Many different DNA nanopores have been designed with different intended applications, but experimental observations are often difficult to interpret due to the small size and high mechanosensitivity of DNA nanostructures, and the limited precision of imaging techniques. Our MD protocol provides a powerful tool to address these issues, furthering our understanding of this emerging branch of nanotechnology.

5. Key References

[1] O. Birkholz, J. R. Burns, C. P. Richter, S. Howorka, and J. Piehler, "Multi-functional DNA nanostructures that puncture and remodel lipid membranes into hybrid materials," *Nat. Commun.*, no. 2018, pp. 1–26.

[2] J. R. Burns, A. Seifert, N. Fertig, and S. Howorka, "A biomimetic DNA-based channel for the ligand-controlled transport of charged molecular cargo across a biological membrane," *Nat. Nanotechnol.*, vol. 11, no. 2, pp. 152–156, 2016.

[3] V. Maingi, J. R. Burns, J. J. Uusitalo, S. Howorka, S. J. Marrink, and M. S. P. Sansom, "Stability and dynamics of membrane-spanning DNA nanopores," *Nat. Commun.*, vol. 8, p. 14784, 2017.

[4] A. P. Bhati, S. Wan, D. W. Wright, and P. V. Coveney, "Rapid, accurate, precise, and reliable relative free energy prediction using ensemble based thermodynamic integration," *J. Chem. Theory Comput.*, vol. 13, no. 1, pp. 210–222, 2017.

[5] S. Wan, A. P. Bhati, S. J. Zasada, I. Wall, D. Green, P. Bamborough, and P. V. Coveney, "Rapid and Reliable Binding Affinity Prediction of Bromodomain Inhibitors: A Computational Study," *J. Chem. Theory Comput.*, vol. 13, no. 2, pp. 784–795, 2017