

Supplementary information.

## Polynucleotides differentiation using hybrid solid-state nanopore functionalizing with $\alpha$ hemolysin

Jérémy Bentin<sup>1</sup>, Sébastien Balme<sup>2</sup>, Fabien Picaud<sup>1</sup>

<sup>1</sup>Laboratoire de Nanomédecine, Imagerie et Thérapeutique, Université Bourgogne-Franche-Comté (UFR Sciences et Techniques), Centre Hospitalier Universitaire de Besançon, 16 route de Gray, 25030 Besançon, France

<sup>2</sup>Institut Européen des Membranes, UMR5635 UM ENSCM CNRS, Place Eugène Bataillon, 34095 Montpellier cedex 5, France

Corresponding author: [fabien.picaud@univ-fcomte.fr](mailto:fabien.picaud@univ-fcomte.fr)

Keyword: simulations, polynucleic diffusion, nanopore, biomimetic

- 1) Details on the protocol to setup the system.
  - a) Stabilization of the  $\alpha$ -hemolysin structure

The  $\alpha$ -hemolysin structure was obtained from the Protein Data Bank n° 7AHL. For all missing atoms, we used the psfgen structure building module of NAMD2 as described in Ref <sup>1</sup>. Then, we randomly added 313 water molecules inside the internal channel of the protein and a layer of water of about 3Å width around the entire protein. The system was then aligned such as the protein pore axis was along the z axis.

To study the dynamics of the protein inside a lipid bilayer or a solid-state media, we first preequilibrated a patch of solvated DMPC lipid bilayer or of a solvated neopentane (NEOP) molecules assembled in a nanopore structure. These two media were oriented parallel to the x, y plane in order to allow the pore of the protein to be inserted inside. The  $\alpha$ -hemolysin was then incorporated into the membrane such as the center of mass the hydrophobic belt of the protein was aligned with the center of mass of the lipid bilayer. For the NEOP nanopore, the center of mass of the hydrophobic belt of the protein was placed at the same location as in the lipid media in order to keep the ratio of immersed protein inside the nanopore than in the biological media. When the two systems were built, all lipids and water molecules that overlapped with the protein stem were removed. The entire complexes were then solvated in a rectangular box of water molecules. The size of the system was equal to (11.8x12x16.8 nm<sup>3</sup>). Systems were then ionized with Na<sup>+</sup> and Cl<sup>-</sup> ions at 1M as shown in Fig. S1.

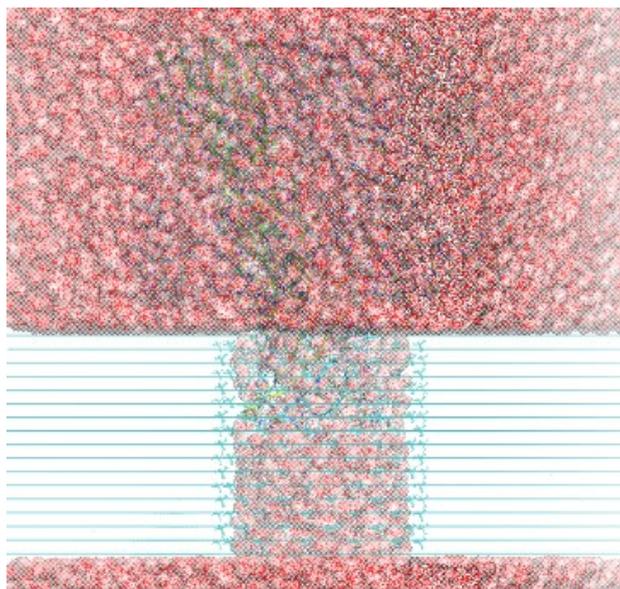


Figure S1: solvated and ionized system in the case of the nanopore of radius 1.9nm (the same kinds of systems were obtained through the larger NEOP nanopore or through the lipid bilayer). Water molecules are shown in red and white area.

All simulations were performed in the framework of the NAMD 2.12 package<sup>2</sup> using periodic boundary conditions, particle Mesh Ewald summation method<sup>3</sup> for the electrostatic contribution. The standard charm 36 parameters<sup>4,5</sup> were used for all atoms and water molecules were described within the TIP3P model.<sup>6</sup> For each simulation, 3200 steps of minimization were first launched before running 5 ns of equilibration of the system at constant pressure (1 bar) and temperature (310K) using the Langevin piston and Langevin thermostat controls (NPT ensemble).<sup>7</sup> During this, the ratio of the unit cell in the x,y plane has been keeping constant. Then production runs were performed for at most 120 ns depending on the system.

We give as an example the complete RMSD of the protein during the whole process of the simulation when the protein is encapsulated inside the small nanopore (Figure S2). The deformation of the protein is clearly progressive during the calculations until reaching a stable structure inside this nanopore.

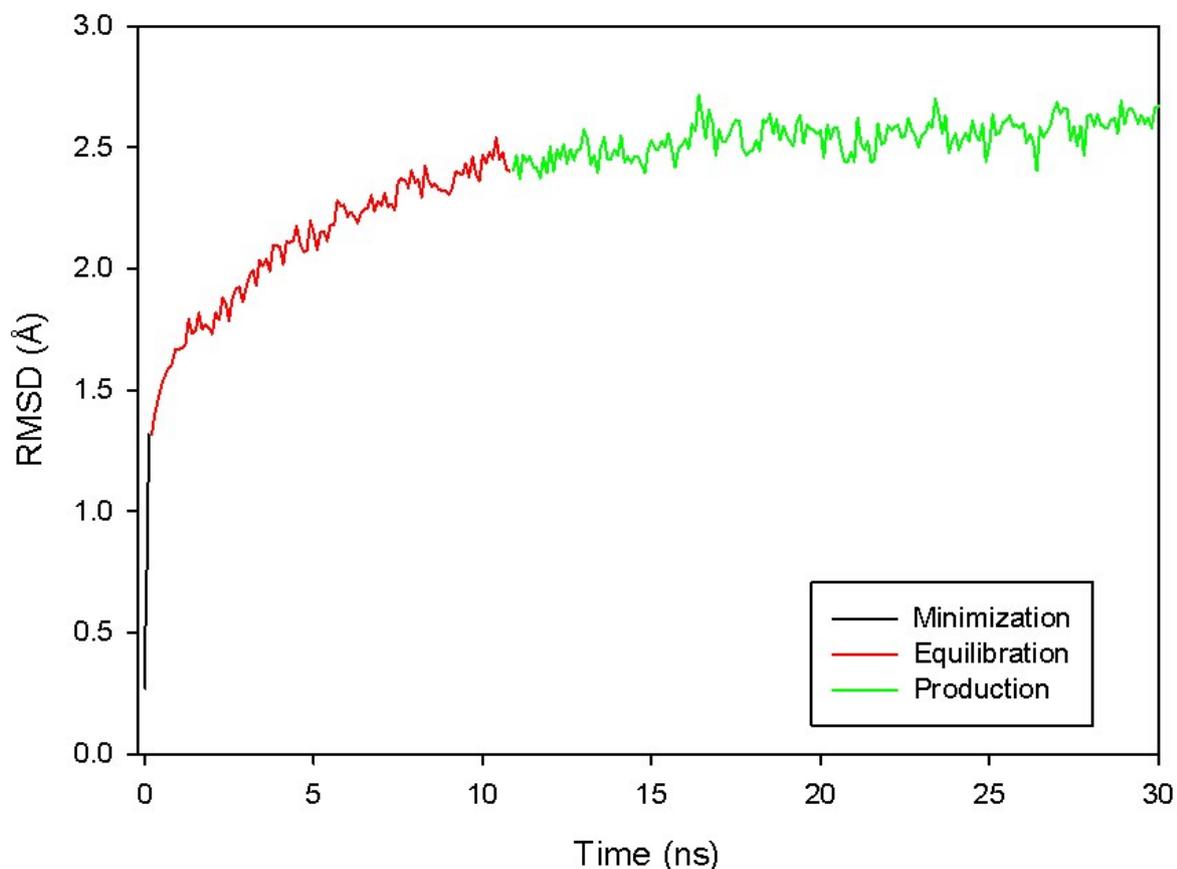


Figure S2: RMSD of the whole protein as a function of the step simulation in order to show the evolution of the protein structure during the different phases of the MD.

To study the ionic current, a constant voltage of 2V was applied to the system (in lipids or in the solid-state nanopore) via a constant electric field imposed in the z direction. To simulate any ionic displacement through the constant voltage application, volume and temperature (NVT ensemble) were kept fixed and these conditions were applied for 100ns at least. The initial configuration of these studies were the final result obtained in the NPT simulation. The value of the ionic current was estimated from the slope of the curve representing the quantity of ions diffusing through the nanopore as a function of time.

#### b) Insertion of DNA

We use single-stranded DNA composed by 37 bases of adenine (noted polyA) or cytosine (noted polyC). These strands were first equilibrated in the solvent, before being unzipped progressively by applying a small force at one extremity of the strand and fixing the other one. The strand was thus stretched progressively in order to obtain a DNA form which could be easily inserted inside the protein. This will accelerate the diffusion process inside the  $\alpha$ -hemolysin structure when simulating the full complex system. At the end of the process (15 ns of NPT simulation in electrolyte solution of 1M of NaCl), the

polyA or polyC strands adopted a more or less straight conformation that was compatible with the shape of the  $\alpha$ -hemolysin pore. We then programmed the insertions of the strands using docking calculations in the framework of the Hex Protein Docking package program.<sup>8</sup> The protocol used in these docking calculations can be described briefly. The protein structures are projected onto a regular 3D cartesian grids by distinguishing the inner volume of the protein from the atom of its surface. Then, rigid molecular docking predict how two or more molecular structures fit together in order to form a stable supramolecular complex. The result of each docked structure is described with a score corresponding to the total free enthalpy taking into account all types of interaction. Hence, this technique can correctly place the structure of a molecule towards a protein. This protocol is quite fast to generate the steric configurations but is fairly limited by the rigidity of the involved molecules. To circumvent this, all the structures should be further optimized using molecular dynamic (MD) simulations to study the role and the influence of the thermal agitation on the complex. Here, when the score of the calculations lead to a strand conformation adsorbed within the protein pore, we added water and ions in the system to complete the model of the full system. All the docked configurations where the strand was outside the protein were not considered. 3200 steps of minimization were launched before running 5 ns of equilibration of the system at constant pressure (1 bar) and temperature (310K) using the Langevin piston and Langevin thermostat controls (NPT ensemble). Then production runs were performed for at most 100 ns depending on the system. The control of the structure in each case was made using the full energy of the system, the pressure and temperature of the system and finally the RMSD of the protein.

- 2) Control of the structure stability
  - a) Protein alone in lipid or nanopore

The stability of the protein inside the lipid bilayer or inside the two nanopores of different diameters was investigated according to the RMSD analysis. Three different RMSDs were considered in each case, the entire protein RMSD (figure S3a), the protein mushroom (part of the protein inside the water) RMSD (figure S3b) and the protein barrel (part of the protein inside the nanopore or the lipid) RMSD (figure S3c).

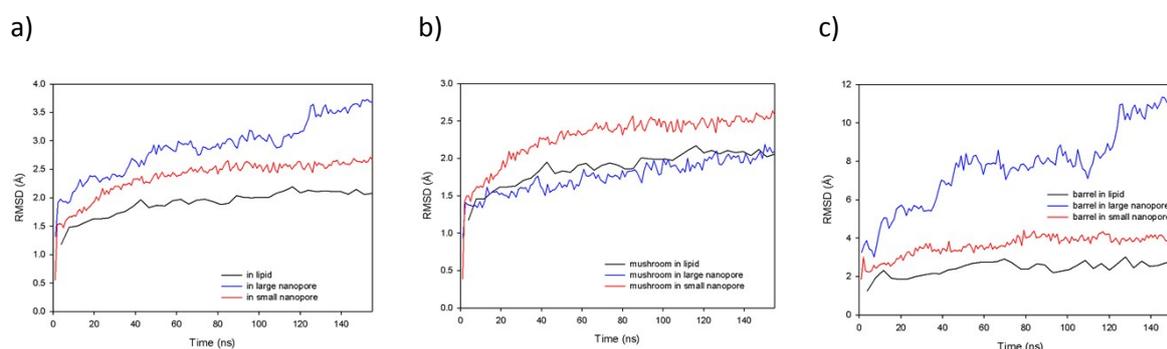


Figure S3: RMSD of the different part of the protein as a function of the step simulation (in lipid for black curve, in large nanopore in blue curve and in small nanopore in red curve). a) The RMSD calculations were performed on the whole structure of the protein. b) The RMSD calculations were performed on the protein mushroom (part of the protein inside the water). c) The RMSD calculations were performed on the protein barrel (part of the protein inserted inside the lipid or the nanopore).

From Figures S3, we can clearly conclude that the protein is almost equally relaxed in the lipid bilayer and in the small nanopore. On the contrary in the larger nanopore, while the mushroom of the protein is not perturbed, the protein barrel progressively denatured inside the nanopore.

To specify the zones that were particularly modified inside the protein during the encapsulation, we performed the sequence alignments of the final protein state in each case (lipid, large and small nanopore in blue, yellow and red representation in Figure S4a).

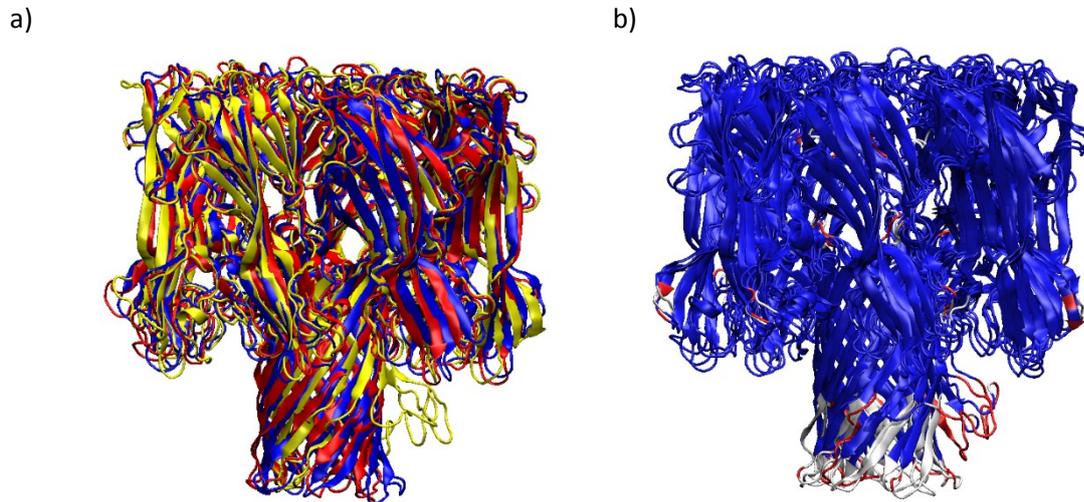


Figure S4: a) Superimposed structures in the final configuration for protein in lipids (blue), large (yellow) and small (red) nanopores. b) Sequence identity modification between the three configurations. In blue, the protein zones that were not affected by the different encapsulations. In white and red respectively, the protein zones that were progressively affected during the encapsulation in lipids to small and large nanopores.

Figure S4a shows mainly the same stable structure unless the barrel for the protein inside the large nanopore that was strongly modified compared to the other encapsulation cases. To precisely identify the protein zones that were strongly affected between the different encapsulation cases, we proceed to the sequence identity scheme where the modified zones of the protein were colored in white to red depending on the protein changes (Figure S4b). As it is clearly shown in Fig. S4b, the main modification of the protein structure is located at the bottom of the barrel.

The RMSD of this zone was calculated in the solid-state nanopore as a function of the protein structure in the lipid bilayer. The first segments of the protein (until segment number 4) were not too affected compared to those obtained in the lipid bilayer. On the contrary, in the large nanopore, we observed a RMSD modification of 13, 19 and 22 Å for the segments 5, 6 and 7 respectively. These zones are denoted in color in Figure S5.

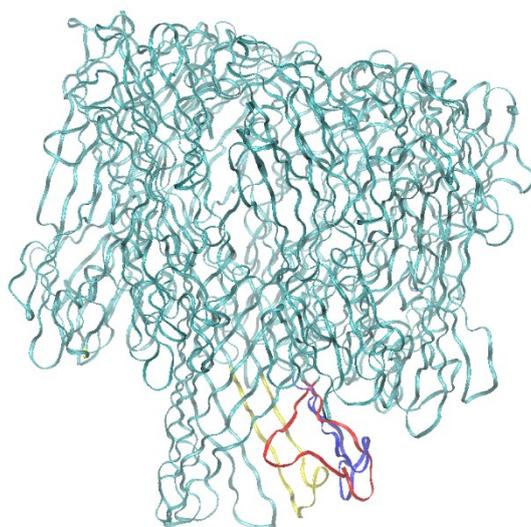
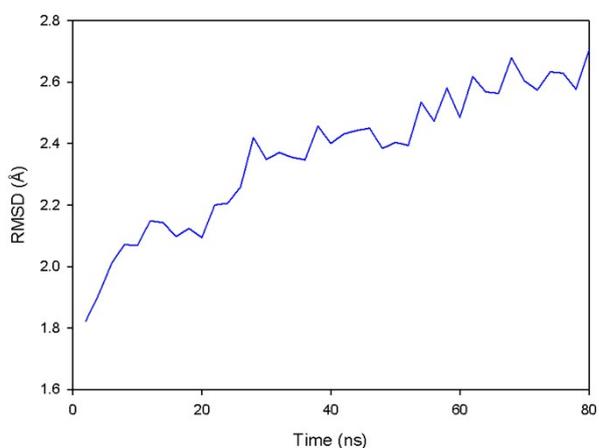


Figure S5: precise zones (shown in yellow, blue and red) that were increasingly affected during the encapsulation of the protein in the large nanopore.

b) Protein and DNA

When the protein is submitted to an electric field and the diffusion of PolyA or PolyC, its structure inside the small nanopore was not damaged. Indeed, Figures S6a and b show the structure modification during the course of the polynucleotides inside the protein. As clearly viewed, the RMSD increases progressively during the displacement of the DNA but does not exceed 3Å for the whole duration of the simulation. This confirms that the protein deforms through the passage of the protein but not being too perturbed by the large applied voltage since the modification of its structure compared to the total number of atoms is low.

a)



b)

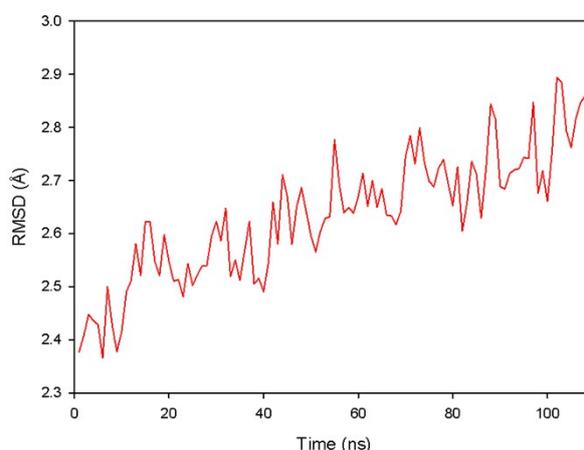
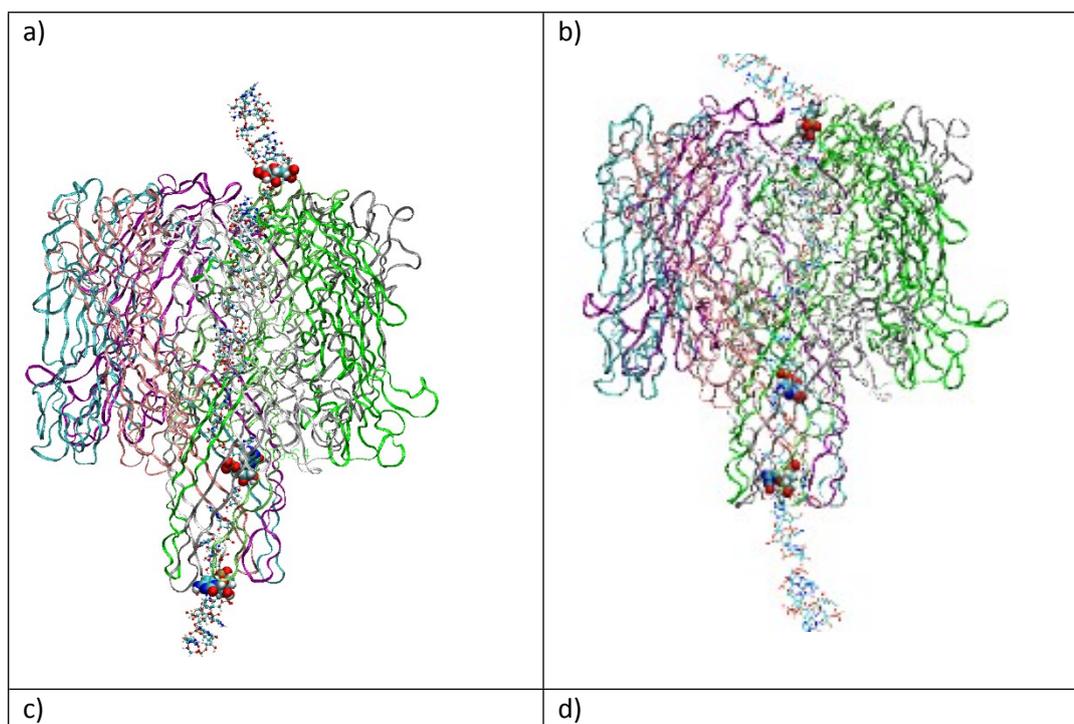


Figure S6a: protein RMSD when PolyA diffuses inside at a voltage equal to 2V. b) The same figure but when PolyC translocates inside the protein

c) Blockage of polyC.

We have observed during 10 ns the blockage of the polyC inside the protein. This blockage is the result of three different zones of the nucleic that modified their interaction with the protein compared to a situation where the nucleic is in displacement. Indeed, as shown in Figures S7, at  $t=25\text{ns}$  and until  $t=35\text{ns}$ , one of the nucleic residues came in interaction with the vestibule entrance while another one is blocked at the barrel entrance. This leads to a disorientation of the polynucleotide inside the barrel (fig. S7a). On the other hand, when the polynucleotide displaces in the protein, we did not observe any residue in special interaction with one part of the protein. These observations can be directly demonstrated using the interaction energy values. Fig.S7c shows the pair interaction between the residue which is blocked at the channel entrance and the protein. From time ranging between 25 and 35 ns, this pair energy is modified in consequence with the blockage of the residue at this position. It involves a small disorientation of the nucleic in the barrel that leads to a slight decrease of the pair energy. All these factors contribute to the blockage of the nucleic during its displacement. Note that these observations were never obtained through the rest of the simulations. It was probably due to the particular morphology of the PolyC during the displacement.



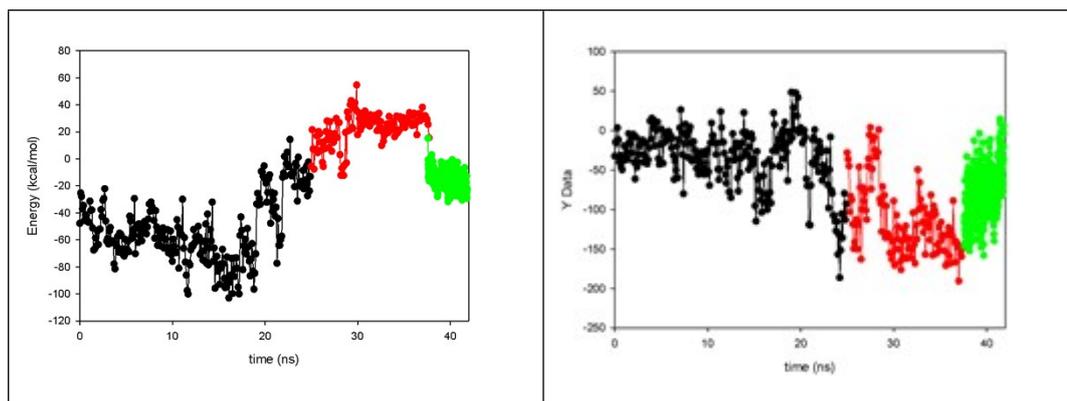


Figure S7a: PolyC blocked inside the protein. Three residues have been shown in van der Waals mode. The one at the entrance of the vestibule (top of the protein) interacts strongly with the protein. This placed at the entrance of the barrel felt an increase of energy. Finally, the nucleic in the barrel is globally closer to the protein when it was blocked compared to a displacement situation. b. PolyC in phase of displacement inside the protein. The PolyC is straight in the barrel. c) Energy variation of the residue in contact with the barrel entrance during the first 42 ns of the simulation. d) Energy variation of the nucleic in the barrel during the first 42 ns of the simulation.

1. A. Aksimentiev and K. Schulten, *Biophysical Journal*, 2005, **88**, 3745-3761.
2. J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kalé and K. Schulten, *Journal of Computational Chemistry*, 2005, **26**, 1781-1802.
3. T. Darden, D. York and L. Pedersen, *The Journal of Chemical Physics*, 1993, **98**, 10089-10092.
4. A. D. Mackerell Jr., M. Feig and C. L. Brooks III, *Journal of Computational Chemistry*, 2004, **25**, 1400-1415.
5. A. D. Mackerell, D. Bashford, M. Bellott, R. L. Dunbrack, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiórkiewicz-Kuczera, D. Yin and M. Karplus, *The Journal of Physical Chemistry B*, 1998, **102**, 3586-3616.
6. W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, *The Journal of Chemical Physics*, 1983, **79**, 926-935.
7. S. E. Feller, Y. Zhang, R. W. Pastor and B. R. Brooks, *The Journal of Chemical Physics*, 1995, **103**, 4613-4621.
8. W. R. David, *Current Protein & Peptide Science*, 2008, **9**, 1-15.