Nonequilibrium Molecular Dynamics

Student 1118161

King's College London

August 22, 2018

1 Simulating the HIV-1 capsid

The human immunodeficiency virus type 1 (HIV-1) infection is highly dependent on its capsid. All-atom simulations of the capsid may reveal molecular properties that can be targeted to treat the infection.

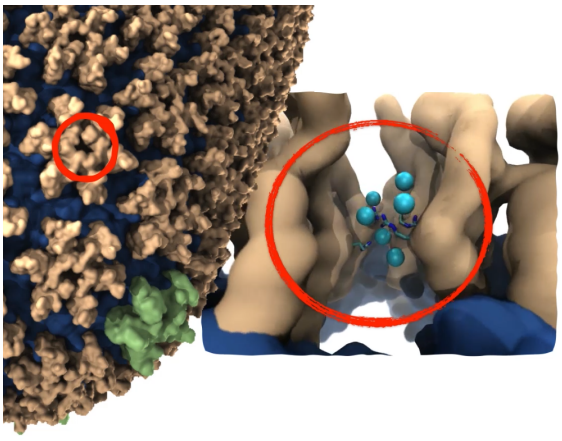

Figure 1: Chloride ion translocation through hexamers that form the HIV-1 capsid

Simulations where ran [1], numbers appeared, statements were published. Blah.

2 Unfolding proteins with umbrella sampling

2.1 Structure and unfolding of α-helix and β-hairpin

Umbrella sampling is a technique in molecular dynamics with which one can efficiently estimate the free energy surface $E(r)$ of a molecule — referred to as the *potential of mean force* along a chosen reaction coordinate r. The reaction coordinate r is typically a nonlinear function of a large configuration space $x \in \mathbb{R}^M$.

To enhance sampling in the reaction coordinate r, an artifical *umbrella potential* $V(r)$ is added to the native classical force fields $U(x)$ to bias the molecule configuration x towards a fixed value of the reaction coordinate $r = f(x)$. The molecule is allowed to explore the configuration space x in the vicinity of this value and the *potential of mean force* is estimated from the resulting trajectory, which is treated as an ergodic ensemble. For finite temperature β

$$E(r) = -\frac{1}{\beta} \ln \left\langle \delta \left(r - f(x) \right) \right\rangle_{U(x)+V \circ f(x)} - V(r) + \text{const} \tag{2.1}$$

Empirically the expectation $\langle \delta \left(r - f(x) \right) \rangle_{U(x)+V \circ f(x)}$ is nothing but the histogram of r over a trajectoy biased by a fixed $V(r)$. This locally estimates $E(r)$. For a sufficiently highly resolved grid in r these local estimates can be vertically aligned to form a global estimate of $E(r)$. Highly resolved r-grids require more molecular dynamics runs and thus are computationally expensive. The weighted histogram analysis method (WHAM) is used to estimate intermediate values of $E(r)$ given a lower resolution grid, whos r histograms still sufficiently overlap.

In the case of unfolding of the α-helix and β-hairpin the reaction coordinate is the end-to-end distance of the protein. A harmonic potential is used to keep the ends of the protein a fixed distance apart for each ensemble. We expect the breaking of bonds and deformation of structure as r increases. Figure 2 reveals the structure of typical conformations of both proteins for ensembles along the reaction coordinate and Figure 3 quantifies unfolding in terms of the histograms of intra-protein hydrogen bonds $n(r)$ with respect to local histograms of r. Hydorgen bonds were counted using `MDAnalysis.analysis.hbonds.HydrogenBondAnalysis` code in Python with a distance cutoff of 3Å. End-to-end distance was calculated with `MDAnalysis.analysis.distances.dist` between the N- and C-terminus; relevant code is avaialble on Github [2].

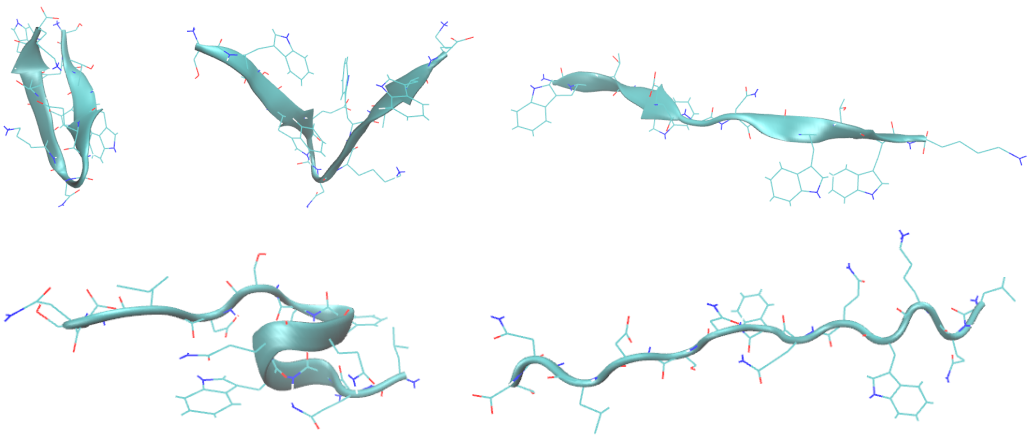

Figure 2: Structures of β-hairpin (top) and α-helix (bottom) from ensembles of increasing reaction coordinate from left to right

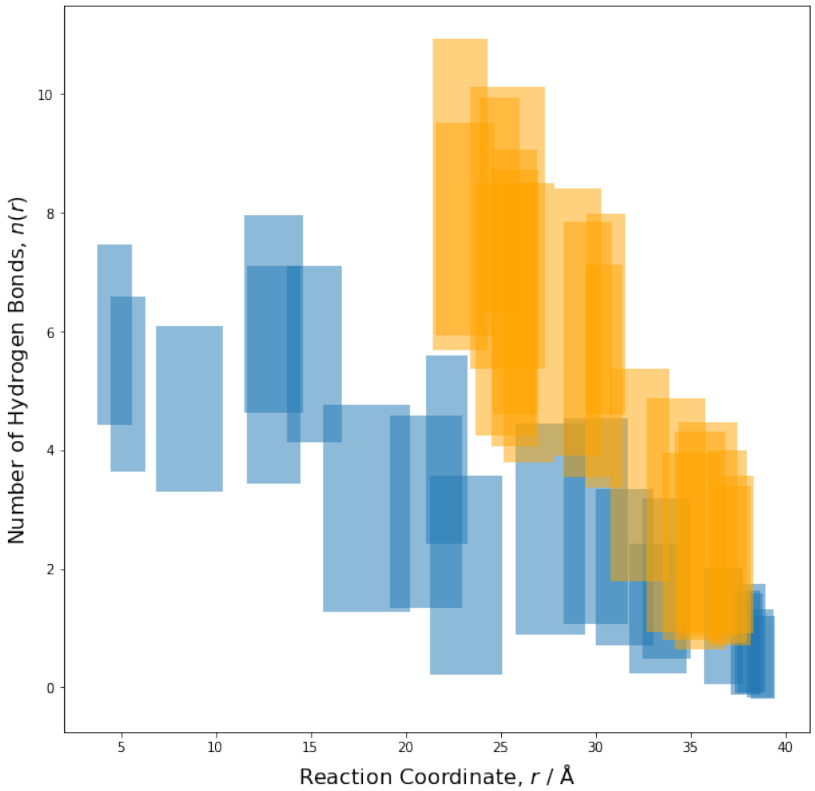

Figure 3: Boxes represent biased ensembles of $\beta-$hairpin and $\alpha-$helix centered at the means and bounded by standard deviations of $n(r)$ and r

Figure 3 reveals the patchy overlap for β-hairpin and good overlap for α-helix between adjacent ensembles the reaction coordinate r and its negative correlation with number of hydrogen bonds $n(r)$. The transient increase in hydrogen bonds in the β-hairpin at around 12Å may indicate a metastable fold state. While the breaking of a hydrogen bond does contribute to the increase $E(r)$ by $1-3\text{kcal}\,\text{mol}^{-1}$ and may account for an increase of up to $\sim 21\text{kcal}\,\text{mol}^{-1}$ as seen in Figure 6, the large uncertainties do not allow us to confidently say those are the only energetic contributions. The Ramachandran plots in Figure 4 reveal contributions from the deformation of non-bonded interactions between residues. The two islands of stability that define β-hairpin's loop dissappear, while the larger island defining the β-sheet remains largely intact. The island defining the α-helix dissappears too, leaving the protein in a β-sheet-like confirmational state.

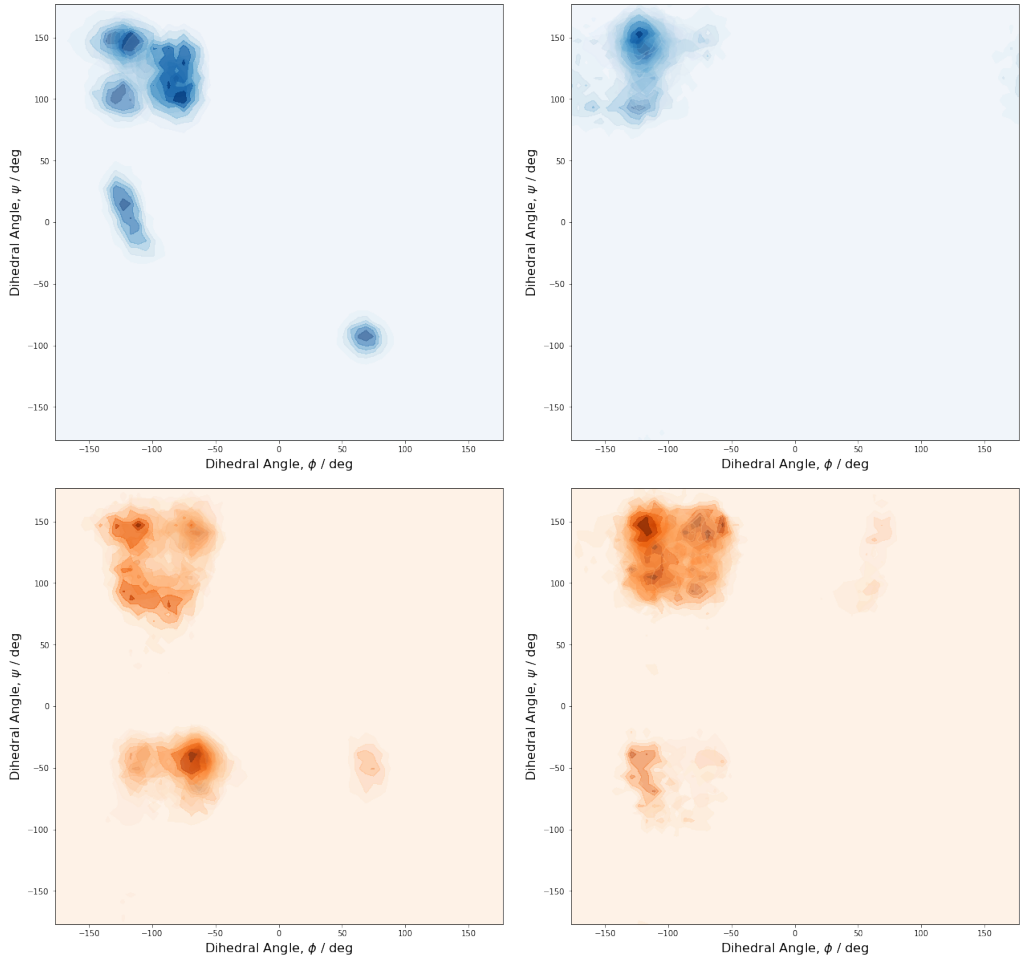

Figure 4: Folded (left) and unfolded (right) conformations of $\beta-$hairpin and $\alpha-$helix revealling the disappearance of islands that define the secondary structure

2.2 Potential mean force and bootstraping

Bootstrapping is a resampling method that quantifies statistical uncertainty by dividing the data into N subsets, hence without additional measruements uncertainty is obtained by *pulling the data up by its own bootstraps*. By default the `g_wham` code forms subsets with replacement over the complete histograms along the reaction coordinate [3], leaving us with N as a hyperparameter.

This hyperparameter is subject to the bias-variance trade-off. Chooseing a large N results in relatively unbiased and uncertain estimates, while a small number of subsets give rise to more certain yet biased estimates. The former is preferred but comes with a computational cost. According to Figure 5 it appears that the uncertainty bounds converge to a fixed value after around $N \geq 20$ so there is little sense in setting N at any other value than $N = 20$.

Figure 6 shows the increasing potential of mean force $E(r)$ with platauing regions that coincide with the transient increases in hydrogen bonds for the β-hairpin in Figure 3. This confirms the existence of metastable states for β-hairpin, and lack of them in the α-helix. The sharp increase in $E(r)$ for the β-hairpin at ~ 7Å corresponds to the unpinning of the hairpin, followed by steady deformations leading to a metastable state, which is followed by further deformations. Finally at ~ 35Å the potential becomes harmonic which corresponding to the stretching of covalent bonds. An artifact at ~ 38Å for the α-helix can be seen, possibly due to a jump across the periodic boundary conditions.

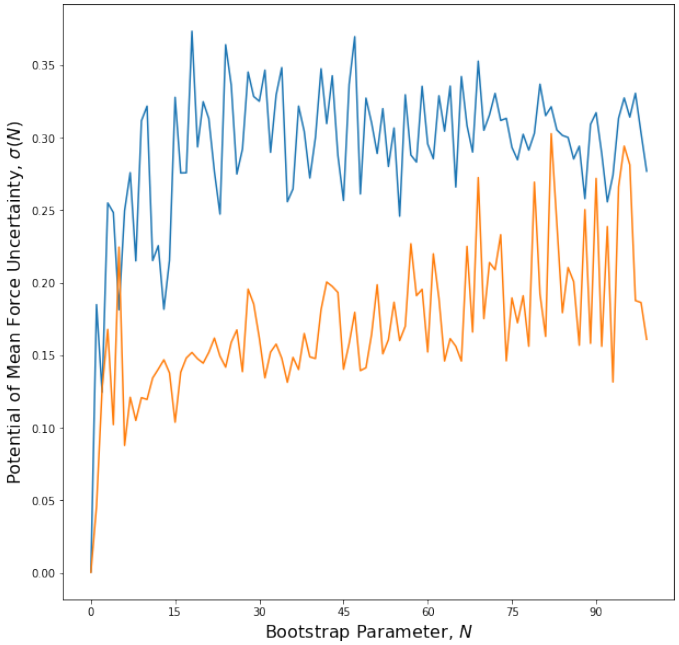

Figure 5: Standard deviation of potential of mean force $\sigma(N)$ as a function of bootstrap parameter N for the $\beta-$hairpin and $\alpha-$helix

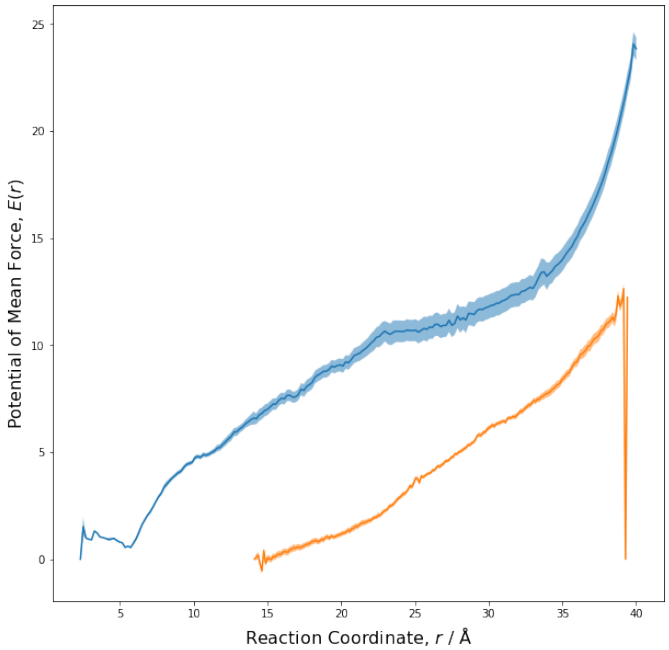

Figure 6: Potential mean force $E(r)$ estimated by WHAM with bootstrap parameter $N = 20$ as a function of reaction coordinate r for the $\beta-$hairpin and $\alpha-$helix

3 Conductance of a porin

3.1 Structure and function

Porins are β-barrel transmembrane proteins that act as a pore through which molecules can diffuse. Outer membrane protein F (ompF) is a nonspecific weakly cationic selective porin found in *E. Coli* [4]. It is nonspecific in the sense that there is no preference for a particular chemical species, however it does exhibit weak selective permeability for cations.

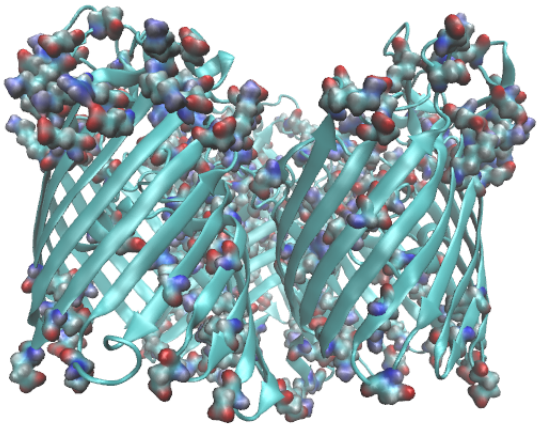

Figure 7: Side-view of ompF: β-barrel homotrimer forms transmembrane channel. Highlighted are polar residues, which dominanate the outer membrane side.

Figure 7 shows the structure of ompF: antiparallel β-strands forming three barrels that interact locally with neighboring β-strands, a long loop inserts into each barrel resulting in channel narrowing, charged residues form clusters at the entrance to a pore on the outer membrane side as well as within the narrow channel zone.

The narrowing geometry prevents larger potentially harmful molecules such as detergents from diffusing across the membrane, while allowing small hydrophilic molecules involved in bacterial metabolism to pass. Charged residues in the narrowing region of each barrel influence the ion selectivity and permeability of the channel. Two clusters of oppositely charged residues as shown in Figure 8 generate a screw-like electric field twisted along the channel axis with a strong transverse component in the pore narrowing region [4]. This field favors cations to pass through the channel.

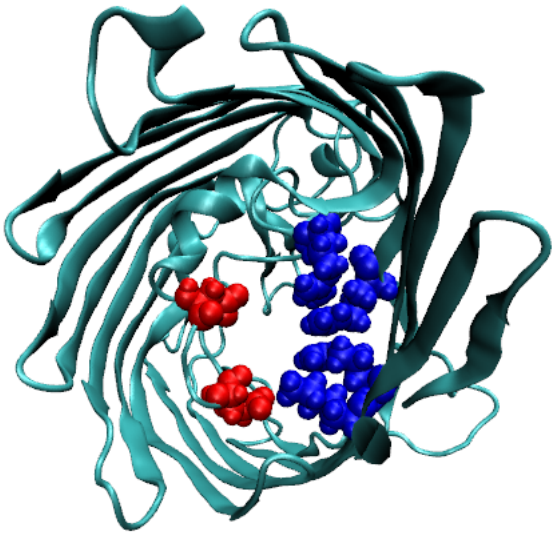

Figure 8: ompF β-barrel from inner membrane side having residues with positive COOH and negative NH_2 partial charged groups that facilitate cationic selectivity

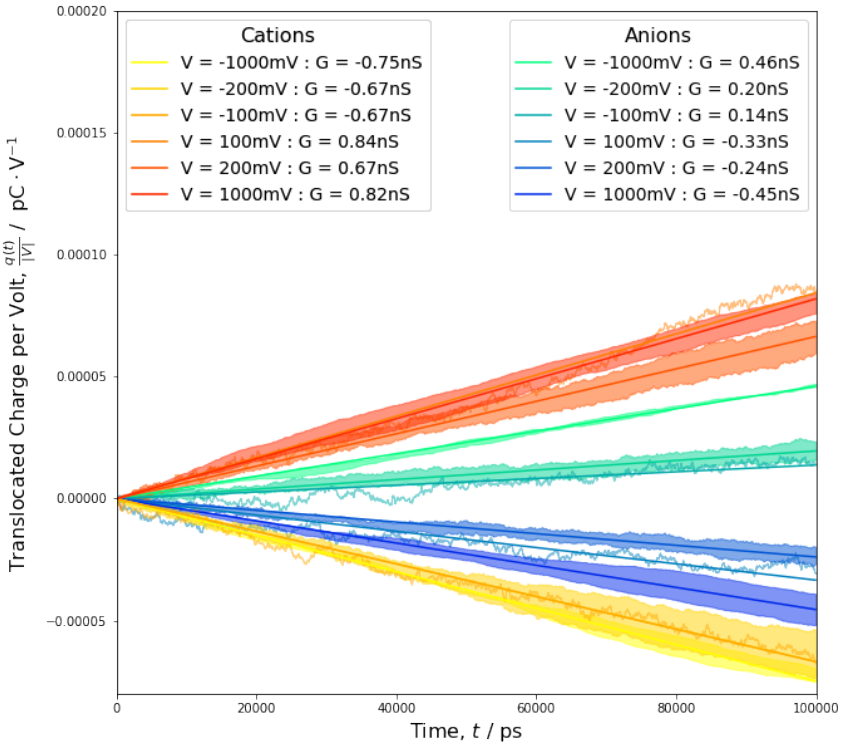

Figure 9: Conductance estimates G from ensemble trajectories for anions and cations across ompF under various applied voltages V

3.2 Conductance estimates from molecular dynamics

Classical force field molecular dynamics with an extental electric field was used to estimate the conductance of ompF. The structure of ompF was embedded in a lipid bilayer, surrounded by a water, Potassium K^+ and Chloride Cl^- ions. Simulations were run for up to 100ns with external fields given by $V = \pm 100\text{mV}, \pm 200\text{mV}, \pm 1\text{V}$. The translocated anionic and cationic charge $q(t)$ was recorded as a function of time. Several trajectories for a given paramerer set where run to obtain a mean and uncertainty. The cationic/anionic conductance $G_\pm$ was estimated via least squares according to the formula

$$q(t) = G_\pm |V| t \tag{3.1}$$

Results summarised in Figure 9 show mean and standard deviation for each trajectory ensemble. The claim of cationic selectivity [5] is supported as $|G_+| > |G_-|$ for all estimates and applied voltages. Estimated conductances also share the same order of magnitude as those reported in literature, but do not coincide precisely [5].

References

[1] J. R. Perilla and K. Schulten, "Physical properties of the HIV-1 capsid from all-atom molecular dynamics simulations," *Nature Communications*, vol. 8, p. 15959, jul 2017.

[2] G. Szep, "Molecular dynamics." `https://github.com/gszep/molecular-dynamics`, 2018.

[3] J. S. Hub, B. L. de Groot, and D. van der Spoel, "g_wham—A Free Weighted Histogram Analysis Implementation Including Robust Error and Autocorrelation Estimates," *Journal of Chemical Theory and Computation*, vol. 6, pp. 3713–3720, dec 2010.

[4] O. D. Novikova and T. F. Solovyeva, "Nonspecific porins of the outer membrane of Gram-negative bacteria: Structure and functions," *Biochemistry (Moscow) Supplement Series A: Membrane and Cell Biology*, vol. 3, pp. 3–15, mar 2009.

[5] R. Benz, A. Schmid, and R. E. Hancock, "Ion selectivity of gram-negative bacterial porins.," *Journal of bacteriology*, vol. 162, pp. 722–7, may 1985.