Sampling the protonation states: pH-dependent UV absorption spectrum of a polypeptide dyad

Electronic Supporting Information

Elisa Pieri,* Vincent Ledentu, Miquel Huix-Rotllant, and Nicolas Ferré*

Aix-Marseille Univ, CNRS, Institut de Chimie Radicalaire, Marseille, France

E-mail: elisa.pieri@univ-amu.fr; nicolas.ferre@univ-amu.fr

Contents

List of Figures

1 Peptide M Simulations Setting

Here the needed information for basic data reproducibility is provided. The initial PDB structure has been provided by Pagba *et al.* (see ref. 16 in the main article) as the average structure created from 20 NMR experimental structures.

For the system initial building the LEaP Amber tool has been used. The peptide has been solvated with 7751 water molecules (TIP3P model), resulting in a cubic-type box of approximately 60x65x75 Å; this result has been reached by setting the minimum distance between any atom originally present in the peptide and the edge of the periodic box to 20 Å. The system has been neutralized by adding 4 chloride ions, using the Coulombic potential on a grid. Finally, the parameters for the system have been written loading a special version of the *ff14SB* Amber force field that has been customized to allow Constant pH Molecular Dynamics.

The system energy has therefore been minimized for 1000 cycles using the steepest descent method and 4000 additional cycles using the conjugated gradient method. During the minimization, the peptide backbone was positionally restrained with a weight (in kcal/molÅ) of 10.0. During this and the subsequent steps, the nonbonded cutoff was 8.0 Å.

The system has been slowly heated for 400 ps at constant volume from 0 to 300 K using the Langevin Thermostat (collision frequency = 5.0), constraining bonds involving hydrogen (SHAKE algorithm).

The equilibration has been made by running a classical molecular dynamics in the NPT ensemble (isotropic position scaling) for 4 ns at 300K. The final structure resulting from the procedure has been used to start CpHMD at the desired pH values using pH-REMD, the 500 ns long classical molecular dynamics and t-REMD at the various temperatures.

2 Production setting

The 500 ns long classical molecular dynamics has been calculated with isotropic position scaling at 300 K keeping the standard protonation states for the aspartic acid and tyrosine residues (deprotonated and protonated, respectively), using a nonbonded cutoff was 8.0 Å. The CpHMD simulations have been carried out for 40 ns in explicit solvent, with protonation state change attempts performed each 0.1 ps. We executed the calculations using the pH-REMD technique at pH 3 to 6 and 9 to 12 (1 pH unit of interval), for a total of 8 replicas. Exchanging attempts between replicas have been performed each 0.2 ps.

The T-REMD simulations have been carried out at temperatures 260 to 360 K (equally spaced, with 20 K interval) for 40 ns for the five most populated protonation microstates; exchanges between replicas have been attempted each 0.2 ps.

3 Timings

The minimization, heating and equilibration steps have been performed using the PMEMD CUDA version in the Amber software package; the respective walltime consumptions are nearly 40 seconds, 20 minutes and 4.5 hours (on a single GPU card). The pH-REMD CpHMD simulations have been performed using the MPI version of the program; each replica required 360 hours on 144 cores for 40 ns long trajectories. In comparison, the subsequent steps are much less demanding: the electrostatic potential calculation required 2 seconds on each frame (on a single core), and each vertical excitation energy (TDDFT B3LYP/6-31G* with Tamm-Dancoff approximation) on the phenolic ring demanded nearly 3 minutes on 4 cores.

4 Tyrosine in water Simulation Setting

The procedure followed to set up the calculations on tyrosine in water is exactly the same as the one described in the previous section. Capping groups (NME and ACE) have been added using LEaP. The number of water molecule is 3828; the box is cubic, with an edge equal to $\sim$55Å; the total number of atoms is 11517. Minimization, heating, equilibration and production have been performed using the same parameters and techniques mentioned above: the CpHMD simulation replicas pH values are 5, 9, 10 and 11, their length is 10 ns.

5 Convergence and statistics

The accuracy of the proposed simulation protocol ultimately depends on the quality of the underlying statistics, i.e. the production of a sufficiently large number of uncorrelated snapshots extracted from the CpHMD trajectories. We first investigated the dependence of the pK_a predicted values with the length of the trajectories, using four 10 ns and two 20 ns windows extracted from the available 40 ns trajectories at each pH value and compared them with the pK_a values obtained from the 40 ns trajectories. As reported in Figure 1, the D3 pK_a value does not change much, converging to 4.11. However, D11 pK_a value is less stable, ranging from 4.66 to 3.80 if only 10 ns of trajectory are used (Figure 2). Nevertheless, the pK_a value obtained from the 40 ns trajectories is 4.27.

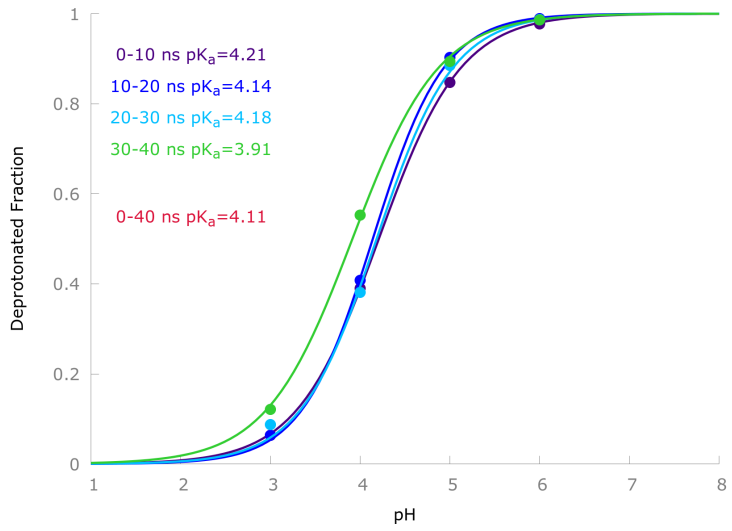

Figure 1: Evolution of D3 pK_a along the 40 ns trajectory on 10 ns sub-intervals, obtained using the deprotonated fraction of the population to fit the Hill equation.

We then determined the minimum time step between two consecutive uncorrelated snapshots extracted from the CpHMD trajectories. This was achieved by analyzing the autocorrelation function of the QM/MM vertical excitation energies (ground to the first 4 excited

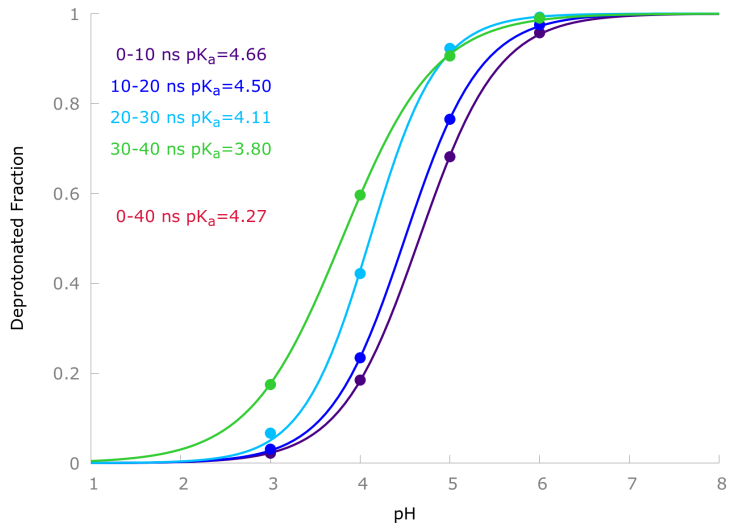

Figure 2: Evolution of D11 pK_a along the 40 ns trajectory on 10 ns sub-intervals, obtained using the deprotonated fraction of the population to fit the Hill equation.

states) computed from 10000 snapshots separated by 1 fs. As apparent in Figure 3, two consecutive snapshots are uncorrelated if they are separated by about 700 fs. Because the QM/MM calculations are somehow expensive, we decided to sample the CpHMD trajectories each ps in the following.

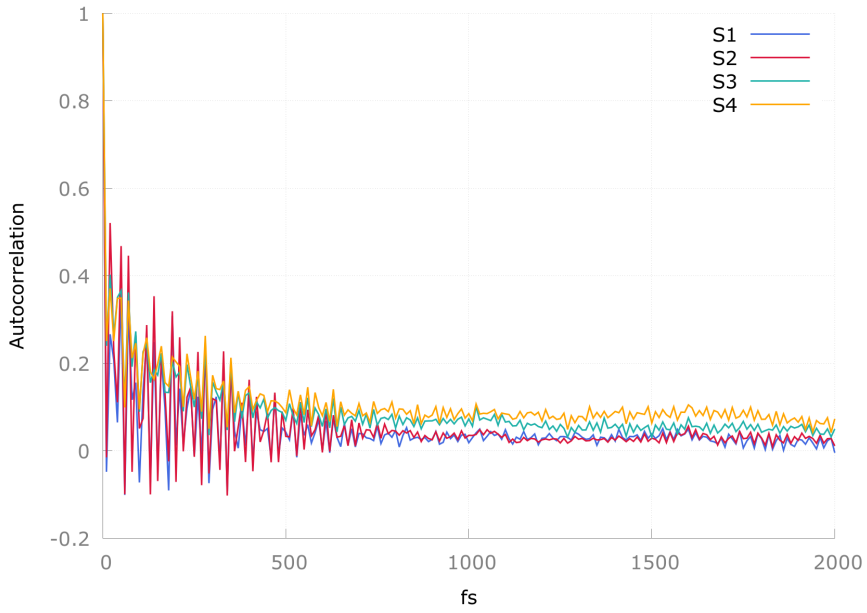

Figure 3: Auto Correlation Function calculated on the excitation energy of the first four excited states of the Tyrosine side-chain embedded in the Peptide M + Solvent electrostatic environment at the TD-DFT B3LYP/6-311G* level of theory.