Electronic Supplementary Material (ESI) for Physical Chemistry Chemical Physics.

Proton transfer network that generates deprotonated tyrosine is a key to producing reactive oxygen species in phototoxic KillerRed protein

Electronic Supplementary Information

Wook Lee[a], Inkoo Kim[b] and Young Min Rhee[a]*

[a] Department of Chemistry, Korea Advanced Institute of Science and Technology,

291 Daehak-ro, Yuseong-gu, Daejeon 34141, Korea

[b] Samsung Advanced Institute of Technology, Samsung Electronics, 130 Samsung-ro,

Yeongtong-gu, Suwon 16678, Korea

* E-mail: ymrhee@kaist.ac.kr

Computational Details for MD simulations and umbrella sampling

MD simulations. After assigning proper protonation states of titratable amino acids, the structure from protein data bank was solvated in a truncated octahedral box filled with TIP3P water molecules[1] and neutralized by adding Na^+ ions using tleap of AMBER tool.[2] The force field parameters for the chromophore and deprotonated tyrosine were taken from AMBER ff14SB[3] and general AMBER force field (GAFF)[4] parameter sets, together with their atomic point charges computed with AM1-BCC.[5,6] All obtained parameters for these non-standard residues were assigned with the antechamber module.[7] For other standard residues, AMBER ff14SB parameter set was used. The solvated structure was then energy-minimized to avoid possible steric clashes and gradually heated up until the temperature reached 300 K with positional restraints on the solute. Subsequently, these restraints were gradually removed, and the structure was equilibrated for 6 ns prior to actual production runs of 20 ns durations. The root mean square deviation (RMSD) curves to check whether equilibration is achieved are also shown in Fig. S8. All MD simulations were carried out using AMBER software suite with version 15.[2] The Langevin thermostat with a collision frequency of 5 ps^{-1} was used to keep the temperature constant, and the pressure was maintained at 1 bar with Berendsen's weak-coupling algorithm.[8] Periodic boundary conditions were applied with a cutoff distance of 8 Å for nonbonded interaction. Particle mesh Ewald (PME) summation was employed for long range electrostatic interactions,[9] and long-range van der Waals interactions were computed by continuum model.[2] The SHAKE algorithm was applied to allow a time step of 2 fs.[10] When assigning hydrogen bonds for the trajectory snapshots, we adopted geometric criteria with the angle O-H-O larger than 110 deg and the nonbonded distance H-O shorter than 3.0 Å.[11]

Umbrella sampling. Because of the ambiguity of using energetics from QM/MM optimizations as discussed in a previous part, the necessity of statistically treating the proton transfer from Tyr110 to Glu68 arose. Therefore, we adopted the umbrella sampling method[12] toward computing the potential of mean force (PMF) profile described by QM/MM for direct and water-mediated proton transfers. For this, we first used the semi-empirical method Recife model 1[13] for conformational sampling, by

employing the reaction coordinates described in Fig. 2. For the direct proton transfer (Fig. 2C), 15 equally spaced umbrella potentials of the form

$$V = \frac{1}{2}k(x - x_0)^2$$

(* MERGEFORMAT 1)

with k = 200 kcal mol^{-1} Å^{-2} were adopted toward generating 1 ns of biased MD trajectory in each umbrella. For each trajectory, the initial 200 ps data were discarded as equilibration, and the remaining 800 ps data were used as an input to the weighted histogram analysis method (WHAM). For the water-mediated proton transfer (Fig. 2D), 80 umbrella potentials with the same functional form were placed in the (RC1, RC2) space. We originally placed 64 umbrellas in the space with an equally spaced square grid with a spacing of 0.2 Å, and manually added 16 more umbrellas after inspecting the overlaps of the two-dimensional histograms in space. The final locations of the umbrella centers are listed in Table S2. All the simulation conditions were the same as the unbiased pure MD simulations except for the time step of 1 fs. The reduction in the time step was to properly handle the relatively fast motion of the transferring proton. For the histograms, the bin size of 0.028 Å was adopted in all cases.

This RM1 sampling was followed by corrections at the B3LYP/def2-TZVPP level of theory, following the recently developed free energy correction scheme.[14–16] Due to the exponential difficulty associated with the correction scheme, the raw data were noisy. Thus, toward generating Fig. 2D, we adopted the simple moving averages as represented by

$$y_i = \frac{1}{M}\sum_{j=-(M-1)/2}^{(M-1)/2} x_{i+j}$$

(* MERGEFORMAT 2)

in the one-dimensional case. Here x represents the raw data from the free energy correction and y represents the smoothened data displayed as in Fig. 2D, with i and j denoting the histogram bin index. Of course, M represents the degree of averaging, and M = 5 was adopted in this work. This smoothing scheme can be trivially extended to the two-dimensional case.

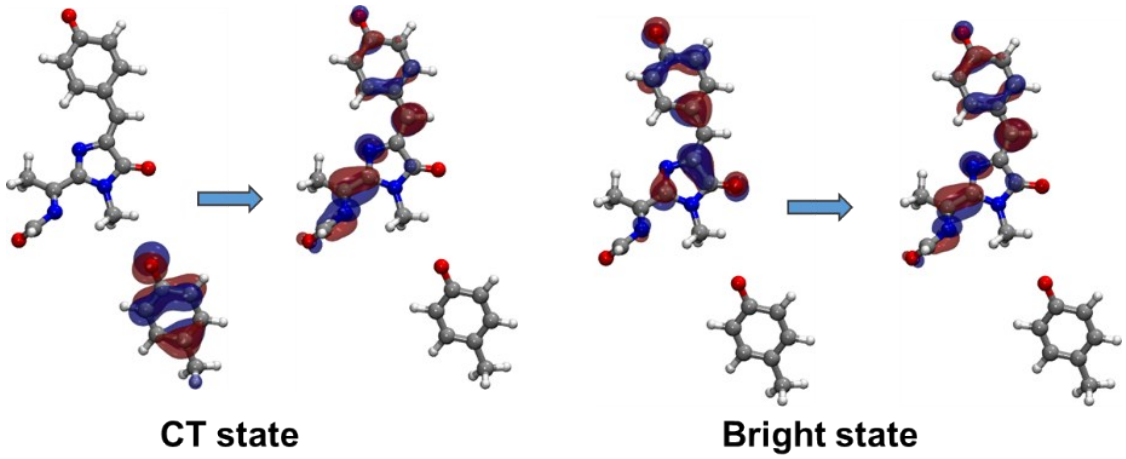

Fig. S1 Natural transition orbitals for the CT and bright states in KillerRed.

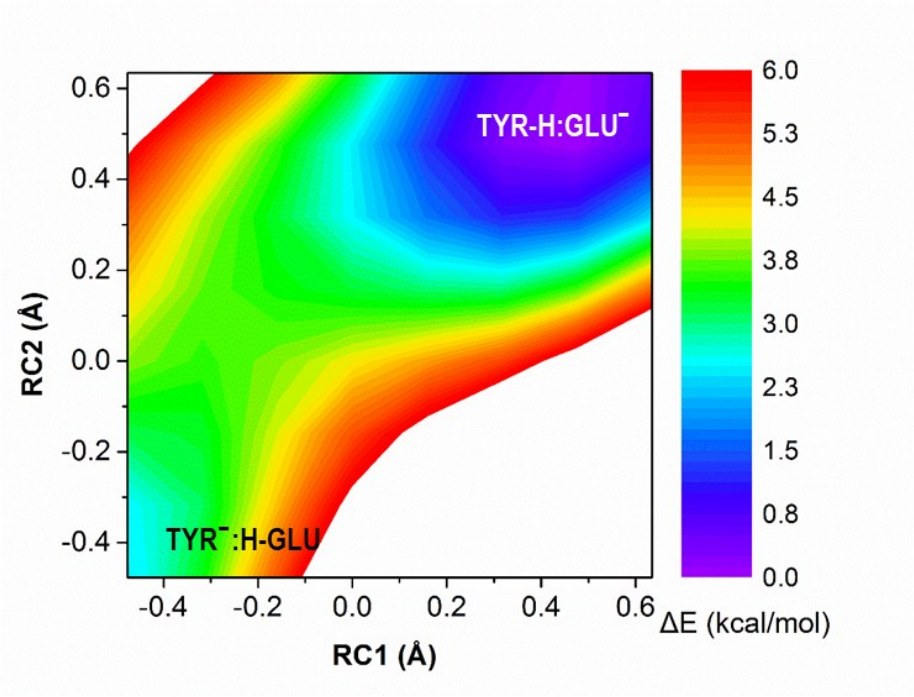

Fig. S2 Potential energy surface for the proton transfer between Glu68 and Tyr110 from QM/MM geometry optimizations, initiated differently from Fig. 2D. The definitions of RC1 and RC2 are the same as in Fig. 2.

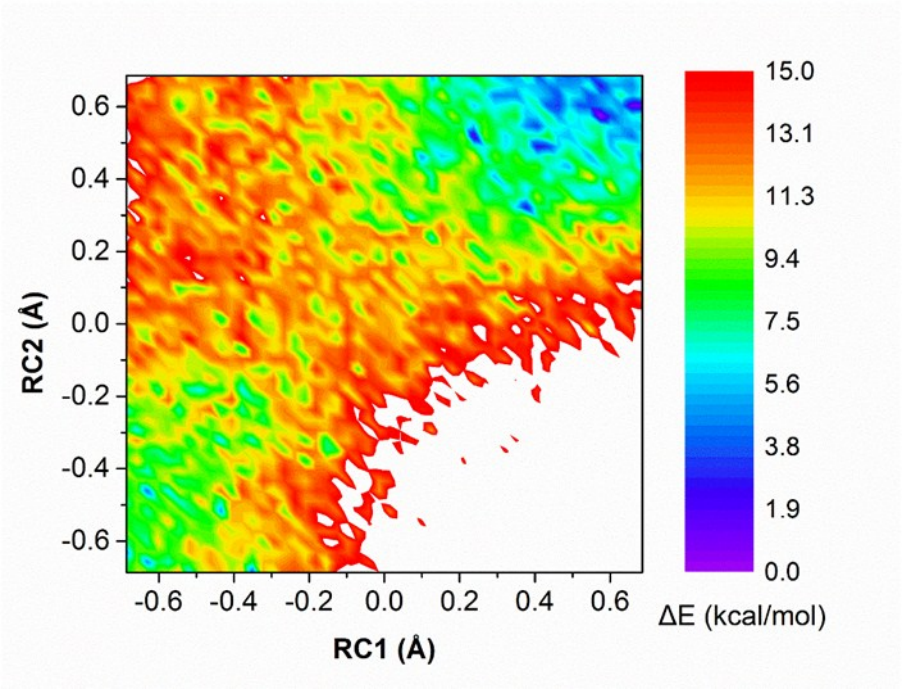

Fig. S3 Raw free energy data before smoothing for the water-mediated proton transfer. Compare against Fig. 4 in the main text.

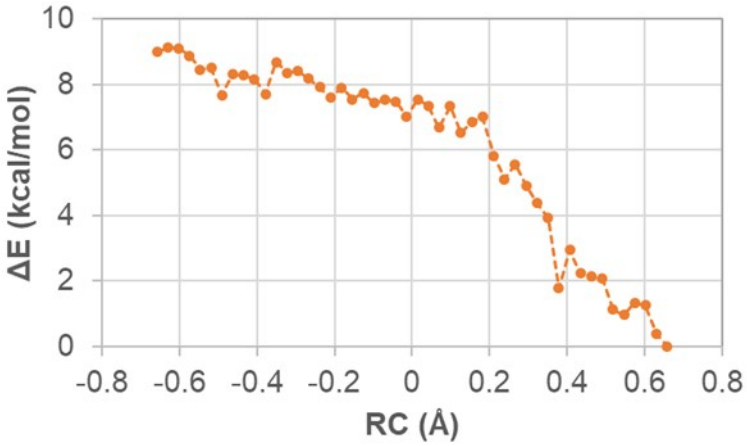

Fig. S4 Free energy profile for direct proton transfer obtained at the B3LYP/def2-TZVPP level of theory after correcting sampling from semi-empirical RM1 level simulations. The definition of RC is the same as in Fig. 2.

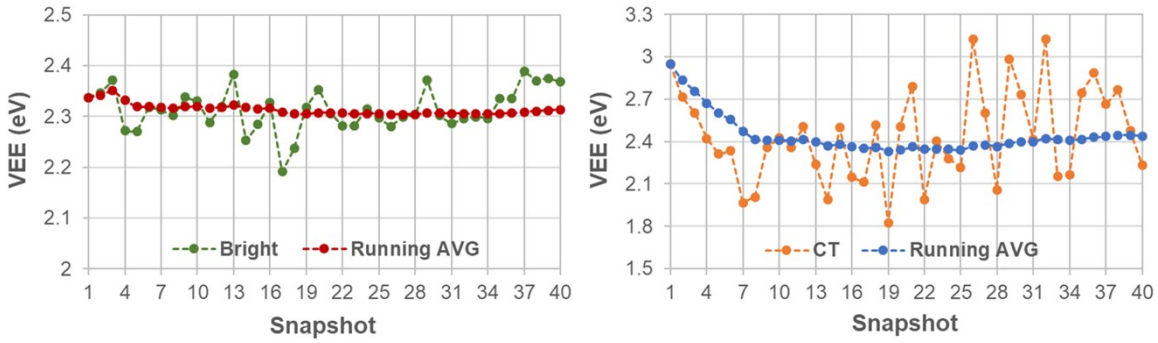

Fig. S5 Convergences of the energies of the bright and the CT states. Running averages are plotted to check whether the average energy of each state is converging.

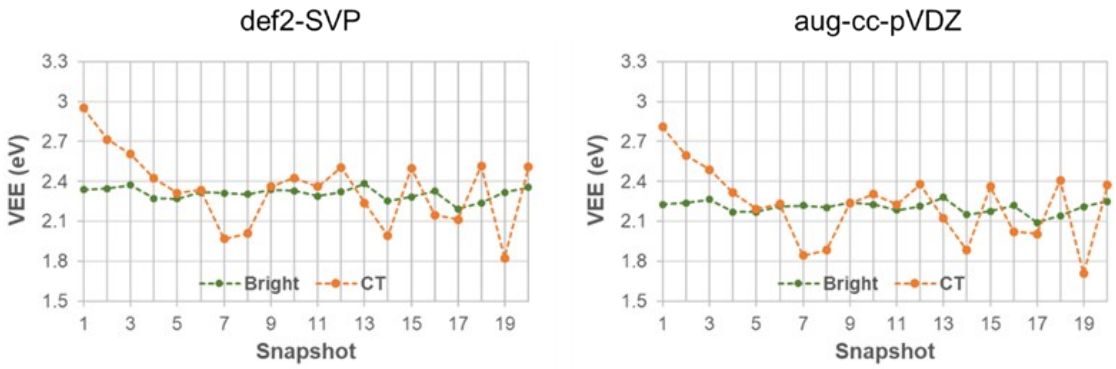

Fig. S6 Adequacy of the adopted basis set, def2-SVP. The relative positioning between the bright and the charge-transferred states is not affected in comparison to the result with a larger basis set with diffuse functions, aug-cc-pVDZ. Comparisons were performed by using the first 20 snapshots from Fig. 5.

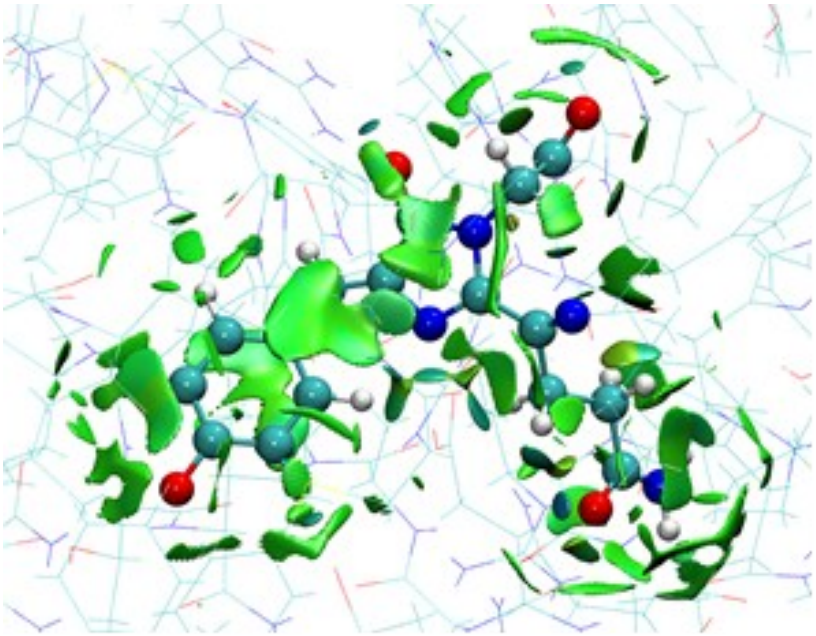

Fig. S7 Non-covalent bond interaction between the chromophore and protein, plotted by NCIPLOT. The interactions are colored on a blue-green-red scale according to values of sign(λ_2)ρ in the range of [–0.07, 0.07] au, where ρ is the promolecular density and λ_2 is the second eigenvalue of the density Hessian. There is no blue or red region with intense non-covalent interaction that must be covered within the QM region.

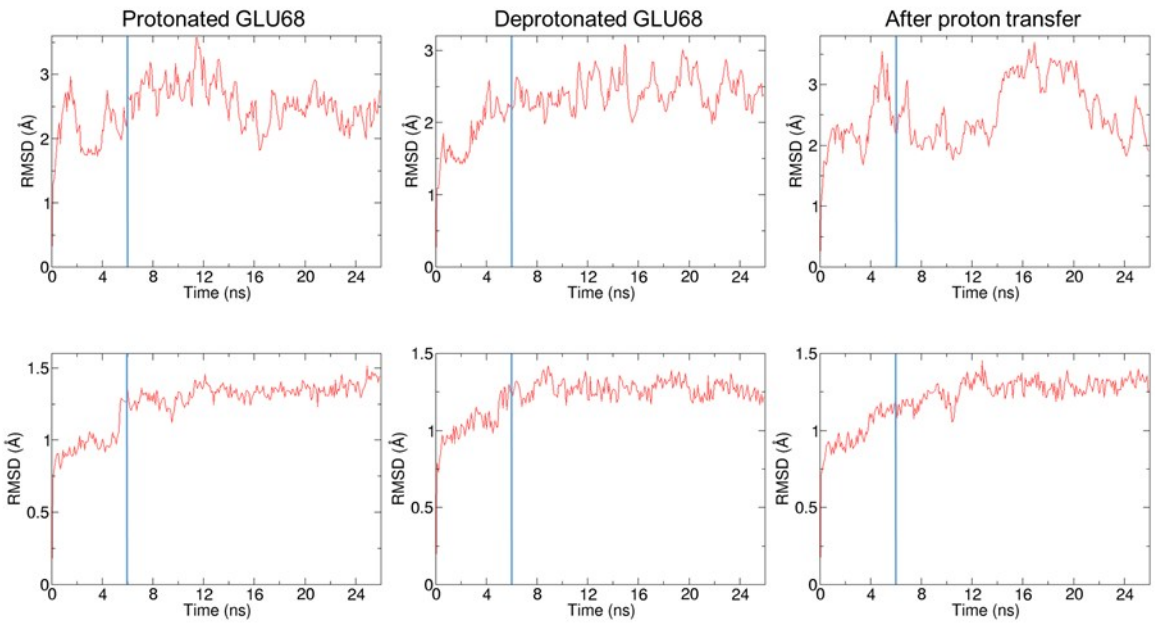

Fig. S8 RMSD curves monitored during MD simulations of the three model systems with different protonation states: RMSD for the whole protein (top) and RMSD for the region within 10 Å of the chromophore (bottom). The time zero denotes a point after MM geometry optimization followed by heating simulations up to 300 K. Based on these, we chose 6 ns (blue lines) as the length of equilibration and adopted the ensuing 20 ns as the production MD run period.

Table S1 The bright and the CT state energies at the optimized bright state minima for seven tested snapshots.

	1	2	3	4	5	6	7
Bright (eV)	1.6900	1.6466	1.6889	1.6351	1.6437	1.3946	1.3071
CT (eV)	1.6819	1.6422	1.6859	1.6349	1.6417	1.3929	1.3040

Table S2 Locations of umbrella windows and the corresponding force constants.[a]

X	*Y*	*k*	*X*	*Y*	*k*	*X*	*Y*	*k*	*X*	*Y*	*k*
0.7	0.7	200	0.3	–0.1	200	–0.3	0.7	200	–0.7	–0.1	200
0.7	0.5	200	0.3	–0.3	200	–0.3	0.5	200	–0.7	–0.3	200
0.7	0.3	200	0.3	–0.5	200	–0.3	0.3	200	–0.7	–0.5	200
0.7	0.1	200	0.3	–0.7	200	–0.3	0.1	200	–0.7	–0.7	200
0.7	–0.1	200	0.1	0.7	200	–0.3	–0.1	200	0.7	–0.6	300
0.7	–0.3	200	0.1	0.5	200	–0.3	–0.3	200	0.6	–0.7	300
0.7	–0.5	200	0.1	0.3	200	–0.3	–0.5	200	0.4	–0.6	300
0.7	–0.7	200	0.1	0.1	200	–0.3	–0.7	200	0.4	0.0	300
0.5	0.7	200	0.1	–0.1	200	–0.5	0.7	200	0.2	–0.2	300
0.5	0.5	200	0.1	–0.3	200	–0.5	0.5	200	0.1	–0.6	300
0.5	0.3	200	0.1	–0.5	200	–0.5	0.3	200	0.0	0.0	300
0.5	0.1	200	0.1	–0.7	200	–0.5	0.1	200	0.0	–0.2	300
0.5	–0.1	200	–0.1	0.7	200	–0.5	–0.1	200	–0.2	–0.2	300
0.5	–0.3	200	–0.1	0.5	200	–0.5	–0.3	200	–0.3	0.6	300
0.5	–0.5	200	–0.1	0.3	200	–0.5	–0.5	200	–0.4	0.7	300
0.5	–0.7	200	–0.1	0.1	200	–0.5	–0.7	200	–0.5	0.4	300
0.3	0.7	200	–0.1	–0.1	200	–0.7	0.7	200	–0.5	0.6	300
0.3	0.5	200	–0.1	–0.3	200	–0.7	0.5	200	–0.6	0.5	300
0.3	0.3	200	–0.1	–0.5	200	–0.7	0.3	200	–0.7	0.4	300
0.3	0.1	200	–0.1	–0.7	200	–0.7	0.1	200	–0.7	0.6	300

[a] Units: *X* and *Y*, in Å; *k*, in kcal mol^{-1} Å^{-2}.

References

1 W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, Comparison of Simple Potential Functions for Simulating Liquid Water, *J. Chem. Phys.*, 1983, **79**, 926.

2 D. Case, J. Berryman, R. Betz, D. Cerutti, T. Cheatham III, T. Darden, R. Duke, T. Giese, H. Gohlke, A. Goetz, N. Homeyer, S. Izadi, P. Janowski, J. Kaus, A. Kovalenko, T. Lee, S. LeGrand, P. Li, T. Luchko, R. Luo, B. Madej, K. Merz, G. Monard, P. Needham, H. Nguyen, H. Nguyen, I. Omelyan, A. Onufriev, D. Roe, A. Roitberg, R. Solomon-Ferrer, C. Simmerling, W. Smith, J. Swails, R. Walker, J. Wang, R. Wolf, X. Wu, D. York and P. Kollman, *AMBER 2015*, University of California, San Francisco, 2015.

3 J. A. Maier, C. Martinez, K. Kasavajhala, L. Wickstrom, K. E. Hauser and C. Simmerling, ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB, *J. Chem. Theory Comput.*, 2015, **11**, 3696–3713.

4 J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman and D. A. Case, Development and Testing of a General Amber Force Field, *J. Comput. Chem.*, 2004, **25**, 1157–1174.

5 A. Jakalian, B. L. Bush, D. B. Jack and C. I. Bayly, Fast, efficient generation of high-quality atomic charges. AM1-BCC model: I. Method, *J. Comput. Chem.*, 2000, **21**, 132–146.

6 A. Jakalian, D. B. Jack and C. I. Bayly, Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and validation, *J. Comput. Chem.*, 2002, **23**, 1623–1641.

7 J. Wang, W. Wang, P. A. Kollman and D. A. Case, Automatic Atom Type and Bond Type Perception in Molecular Mechanical Calculations, *J. Mol. Graph. Model.*, 2006, **25**, 247–260.

8 H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola and J. R. Haak, Molecular dynamics with coupling to an external bath, *J. Chem. Phys.*, 1984, **81**, 3684–3690.

9 U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee and L. G. Pedersen, A smooth particle mesh Ewald method, *J. Chem. Phys.*, 1995, **103**, 8577–8593.

10 J.-P. Ryckaert, G. Ciccotti and H. J. C. Berendsen, Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes, *J. Comput. Phys.*, 1977, **23**, 327–341.

11 T. Steiner, The Hydrogen Bond in the Solid State, *Angew. Chem. Int. Ed.*, 2002, **41**, 48–76.

12 G. M. Torrie and J. P. Valleau, Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling, *J. Comput. Phys.*, 1977, **23**, 187–199.

13 G. B. Rocha, R. O. Freire, A. M. Simas and J. J. P. Stewart, RM1: a reparameterization of AM1 for H, C, N, O, P, S, F, Cl, Br, and I, *J. Comput. Chem.*, 2006, **27**, 1101–1111.

14 S. Kumar, J. M. Rosenberg, D. Bouzida, R. H. Swendsen and P. A. Kollman, The Weighted Histogram Analysis Method for Free-energy Calculations on Biomolecules. I. The Method, *J. Comput. Chem.*, 1992, **13**, 1011–1021.

15 A. Grossfield, "WHAM: The Weighted Histogram Analysis Method", version 2.0.9, http://membrane.urmc.rochester.edu/content/wham, (accessed May 27, 2014).

16 M. R. Shirts and J. D. Chodera, Statistically optimal analysis of samples from multiple equilibrium states, *J. Chem. Phys.*, , DOI:10.1063/1.2978177.