Supporting Information

Full Protein Flexibility is Essential for Proper Hot-Spot Mapping

Katrina W. Lexa, Heather A. Carlson

Department of Medicinal Chemistry, University of Michigan, Ann Arbor, MI 48109-1065

Molecular Dynamics Simulation

The stability of the protein core was verified on the basis of low backbone rmsd over the course of the trajectory, as measured with *ptraj*. The presence of mixed solvent did not destabilize the unrestrained protein on the timescale examined.

Five independent trajectories of 10 ns each were performed, but for the analysis, we combined the last 2 ns of each simulation to provide the most "equilibrated" 10 ns of the available 50 ns of simulation time. To calculate the solvent grids, the central C2 atom of acetonitrile and the oxygen atoms for water were binned into 0.5 Å × 0.5 Å × 0.5 Å volume elements, which spanned the entire box. The electron density from the initial crystal structure and the grids from MD simulation were compared in Chimera[1] to examine the ability of MixMD to recover known hydration sites and binding sites of CCN.

The CCP4i suite[2] was used to produce electron density grids for the solvent probe density in the original crystallographic study. The *ptraj* function in AMBERTOOLS1.2 was used to generate grids of solvent density from simulation data. Prior to grid output, each trajectory was centered, imaged, and aligned to the protein core backbone from the 2LYO crystal structure.

Convergence in the simulations was examined with several measures. The first test compared the grids between the five independent simulations (the last 2 ns from each that were used to make the 10 ns of sampling noted above were calculated separately). Difference grids were calculated to determine differences in the density data. The second test compared the position of maximal occupancies for each of the five separate simulations. The 99% maximum occupancy calculation yields the grid points with the highest occupied solvent density.

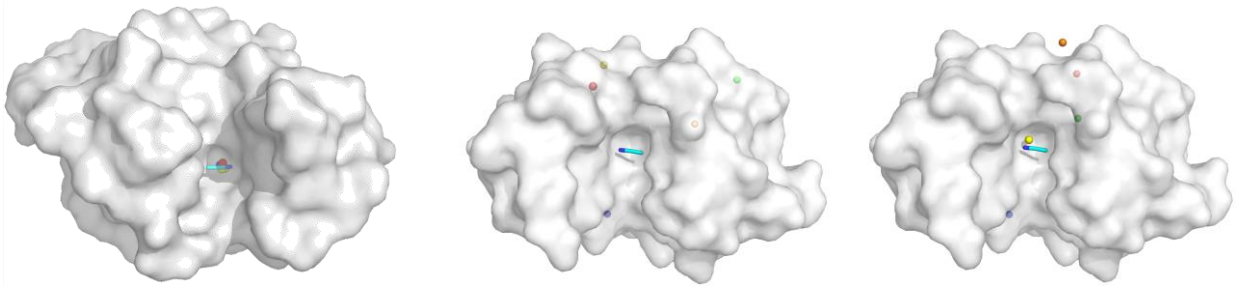
Third, the influence of the protein should be minimal at the box edges. Therefore, the ratio of probe to water on the distal points of the occupancy grids should simply approach the ratio of the total number of probes to water in the simulation. Occupancy grids within 5 Å of each box boundary were examined. The ratio of probe occupancy to water occupancy was compared to the ratio of probes and water in the initial system set up. Occupancy within 5 Å from the box boundary for each plane was used to calculate this ratio. Correspondence between the two different ratios implies adequate sampling has occurred (values near 1.0). The calculation of this ratio was performed through a python script in Chimera.

CCN parameters[3]

Atom Type	NY	CY	CT	HC
q (e)	-0.533	0.481	-0.479	0.177

Convergence of Simulations of All-Atom Rigid, Backbone Fixed, and Fully Flexible HEWL in Pre-equilibrated Solvent

Figure S1: The maximal occupancy positions over the last 2 ns for each independent simulation of HEWL in pre-equilibrated 50% w/w CCN and water. Only the fully flexible system shows convergence of the five simulations in agreement with experiment. Individual runs are indicated by color (red, orange, yellow, green, and blue).



Fully Flexible | Backbone Restrained | All-Atom Fixed

Table S1: The extent of convergence for HEWL in MixMD, as determined by the ratio of co-solvent to water molecules at the box edges. Values of Obs/Exp near 1.0 indicate complete and unbiased sampling at the edges of the box.

	Pre-equil 50% Fully Flexible	Pre-equil 50% Backbone Restrained	Pre-equil 50% All Atoms Fixed
$(Np/Nw)_{Obs}$	0.4501	0.4664	0.4709
$(Np/Nw)_{Exp}$	0.4490	0.4490	0.4490
Obs/Exp	1.0024	1.0388	1.0488

Figure S2: Combined results from the last 2 ns of all 5 simulations (10 ns of sampling) with flexible HEWL (white surface), comparing the layered solvent to the pre-equilibrated protocol. The probe density is shown in blue mesh for 50% w/w pre-equilibrated solvent and in orange mesh for 50% w/w layered water and CCN. The highest sampled densities overlap exactly and agree well with the position of CCN in the 2LYO crystal structure.

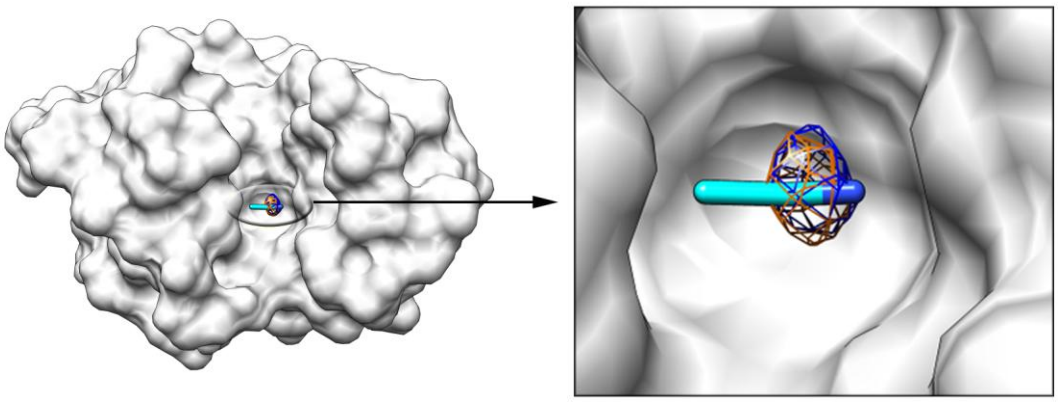


Figure S3: The positions of maximal occupancy on the CCN grid were calculated over the last 2 ns for each independent simulation around a fully flexible HEWL (white). **A**) Results for the individual simulations based on pre-equilibrated solvent are shown. All five positions are in excellent agreement with one another and the CCN molecule in the 2LYO structure. The second (orange) and fourth (green) run results lie beneath the first (red) and fifth (blue) run results. **B**) Results for the individual simulations based on layered solvent are shown, with the orientation skewed slightly from A to show the green site (run 3) that does not agree. Four of the five positions are in agreement with one another and the crystallographic CCN. The first run result (red) lies beneath the fifth run result (blue).

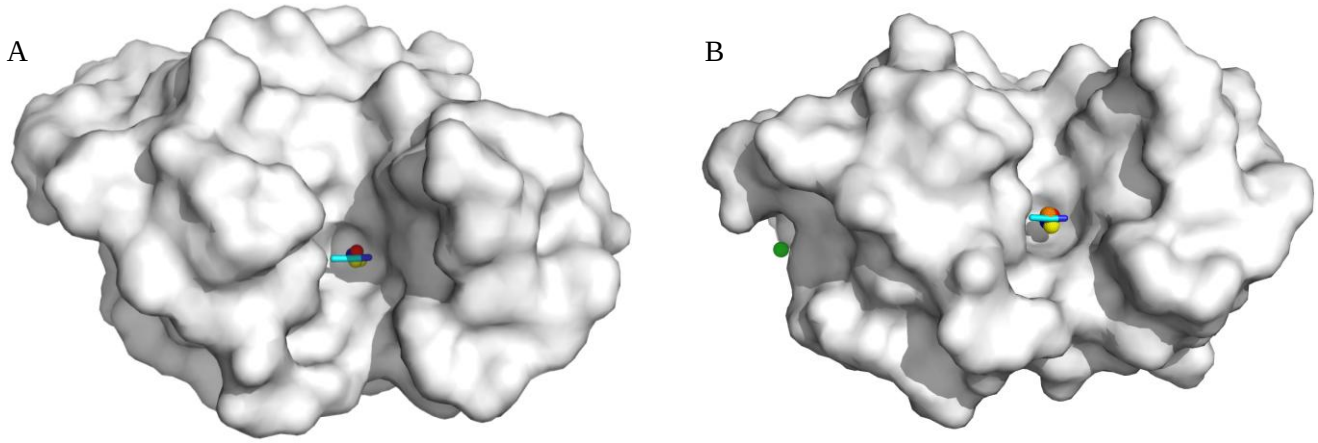


Figure S4: Results from the three different solvation protocols with fully flexible HEWL and CCN as the organic probe. The probe density calculated by combining the last 2 ns of the five independent simulations (to give 10 ns of sampling) is shown in purple mesh for 90% w/w CCN, in blue for 50%, and in green for 10%. Results from all three simulations are nearly identical and reproduce the crystallographic position of CCN.

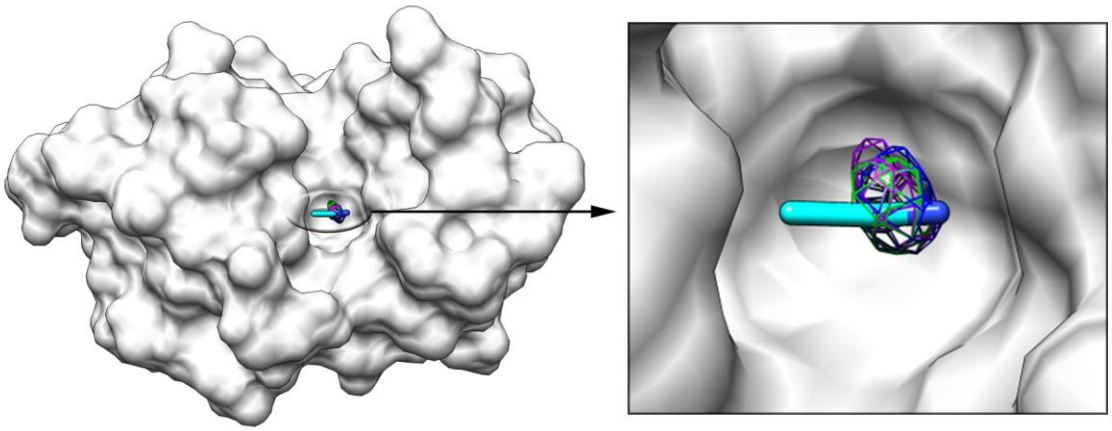
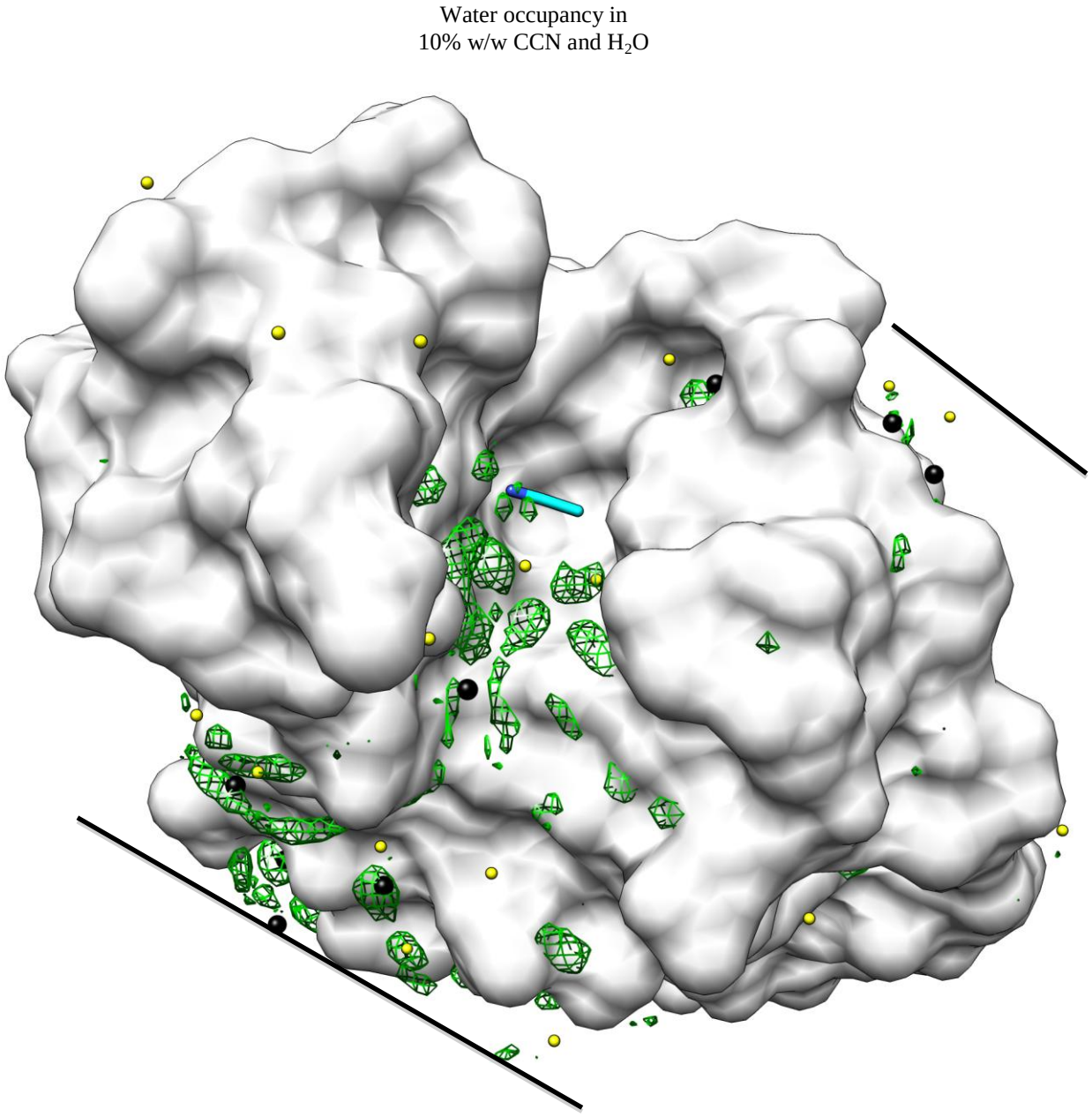


Table S2: The extent of convergence for fully flexible HEWL in MixMD, as determined by the ratio of CCN to water at the box edges. Simulations of 90% w/w CCN were problematic, and the behavior at the edges of the box show that they do not demonstrate even, unbiased sampling (the Obs/Exp value is much larger than 1.0).

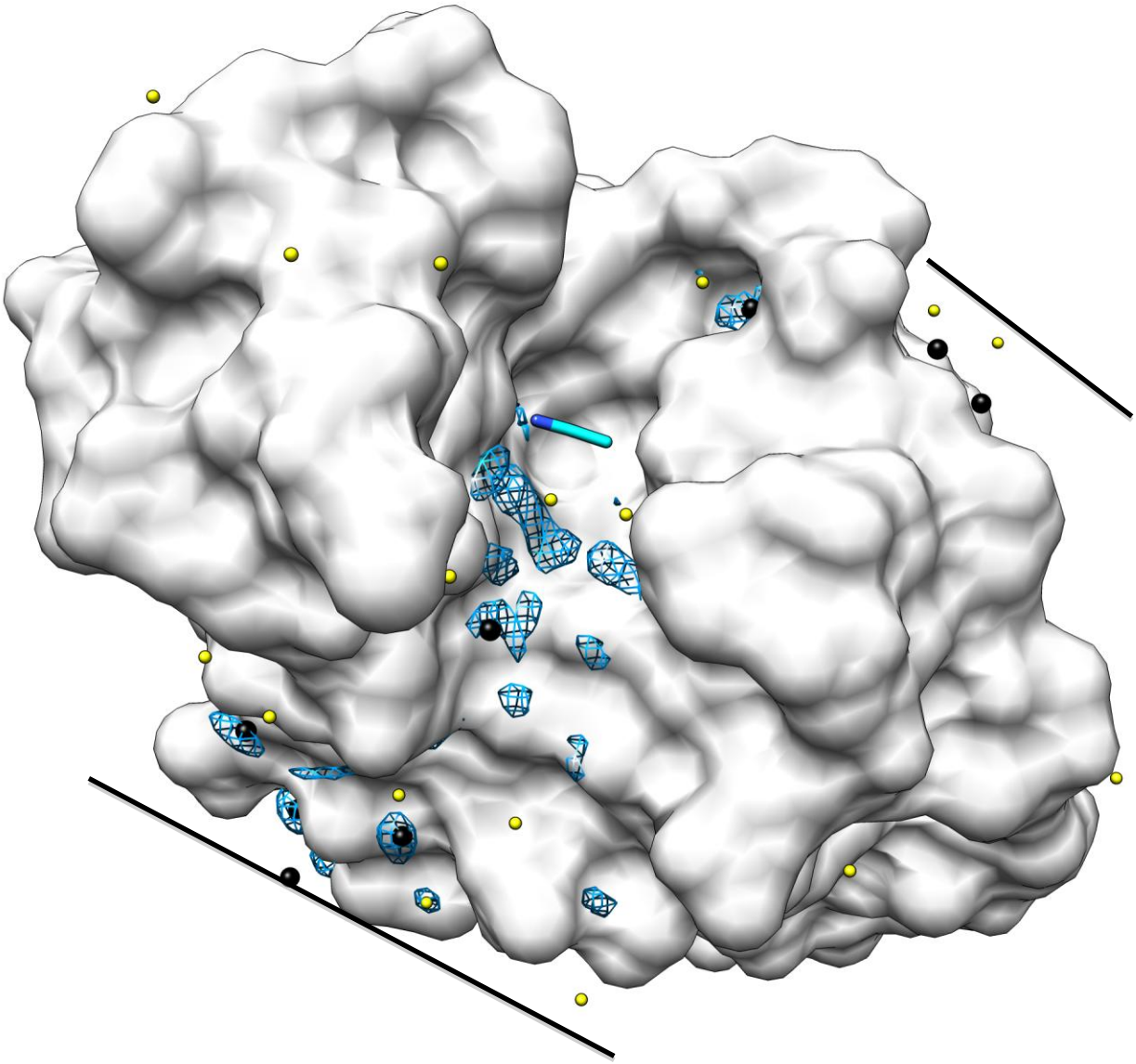
	Pre-equil 50% w/w	Layered 50% w/w	Layered 10% w/w	Layered 90% w/w
$(Np/Nw)_{Obs}$	0.4501	0.4619	0.0477	4.7453
$(Np/Nw)_{Exp}$	0.4490	0.4314*	0.0489	3.6563
Obs/Exp	1.0024	1.0707	0.9755	1.2978

* The layered 50% box contained a few more water molecules than the pre-mixed box, giving it a slightly different expected (Np/Nw).

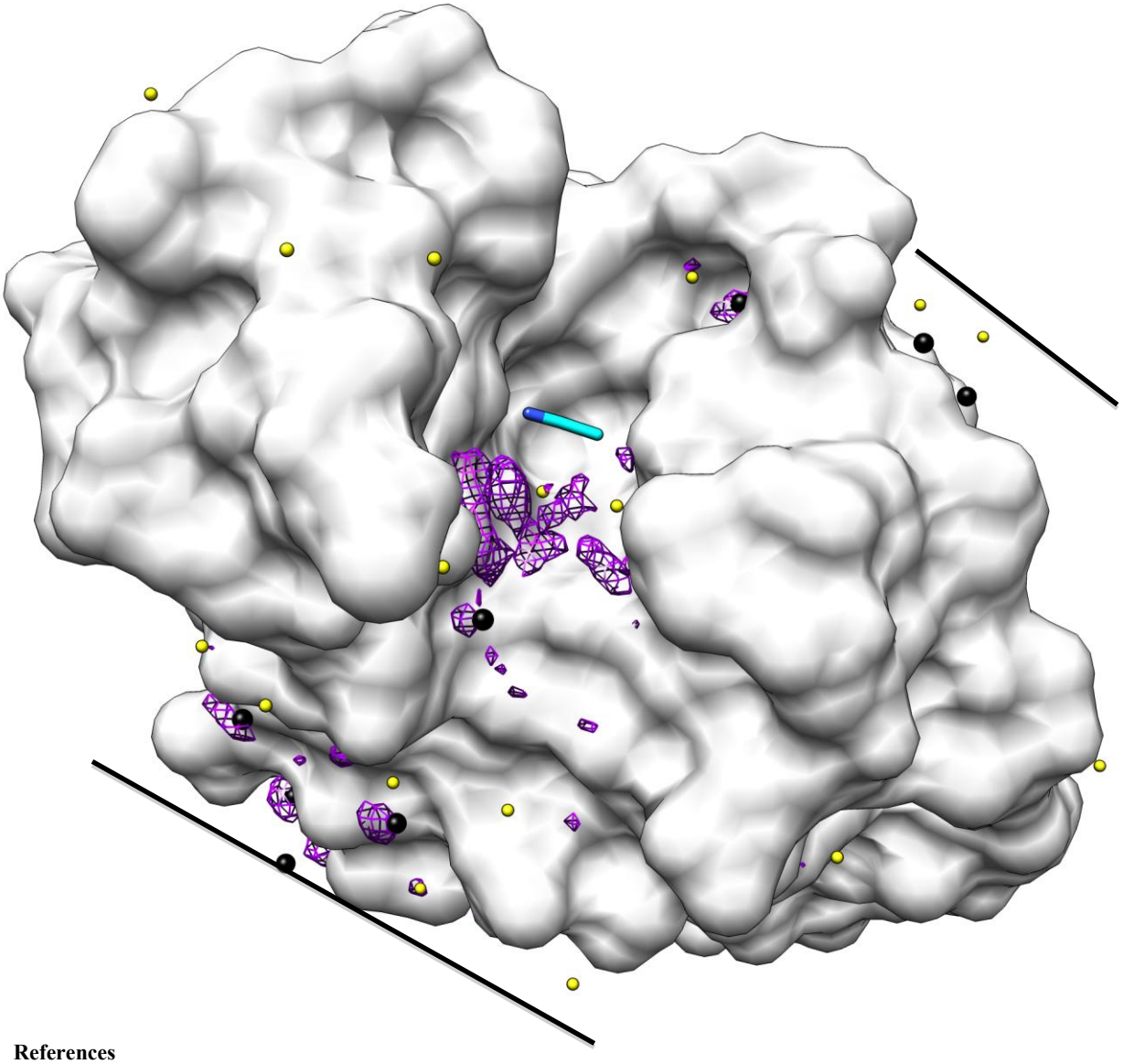
Figure S5: The water density calculated by combining the last 2 ns of the five independent simulations (giving 10 ns of sampling) is shown. The water density is shown in green mesh for 10% w/w CCN (A), in blue for 50% (B), and in purple for 90% (C). Crystallographic waters are colored black for B-factors below 33 Å and yellow above 33 Å. Although the 10% and 90% environments reproduce the CCN position, they do not give appropriate mapping for the water. In A, too many equivalent positions for water are seen without identifying the water with higher B-factor (yellow). In C, there are too few water locations.



Water occupancy in
50% w/w CCN and H_2O



Water occupancy in
90% w/w CCN and H_2O



References

(1) Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. *J. Comp. Chem.* **2004**, *25*, 1605-1612.
(2) Potterton, E.; Briggs, P.; Turkenburg, M.; Dodson , E. *Acta. Cryst.* **2003**, *D59*, 1131-1137.
(3) Grabuleda, X.; Jaime, C.; Kollman, P. A. *J. Comp. Chem.* **2000**, *21*, 901-908.

Complete Reference 17

Case, D. A.; Darden, T. A.; Cheatham, T.E., III; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Crowley, M.; Walker, R. C.; Zhang, W.; Merz, K. M.; Wang, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Kolossváry, I.; Wong, K. F.; Paesani, F.; Vanicek, J.; Wu, X.; Brozell, S. R.; Steinbrecher, T.; Gohlke, H.; Yang, L.; Tan, C.; Mongan, J.; Hornak, V.; Cui, G.; Mathews, D. H.; Seetin, M. G.; Sagui, C.; Babin, V.; Kollman, P. A. *Amber 10*; University of California: San Francisco, 2008.