Quantitative analysis of water dynamics in and near proteins

Oliver Beckstein[1,2*], Naveen Michaud-Agrawal[1], and Thomas B. Woolf[1†]

[1] Johns Hopkins University, School of Medicine. Dept. of Physiology. Baltimore 21205, MD.

[2] University of Oxford. Dept. of Biochemistry. Structural Bioinformatics and Computational Biochemistry Unit. Oxford OX1 3QU. UK

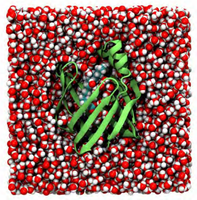
Merton College, Oxford

*oliver.beckstein@bioch.ox.ac.uk †twoolf@jhmi.edu

Protein solvation and MD simulations

"Types" of water in biology

1. *solvent* ("bulk-like")
2. *associated* with proteins (e.g. involved in ligand binding, epitope-antibody contacts, catalysis); exchanges over 10 ps – 100 ps with bulk
3. *structural* (buried); as conserved as amino acids, exchanges over 10 ns – 0.01 s



Molecular dynamics (MD) simulations of solvation

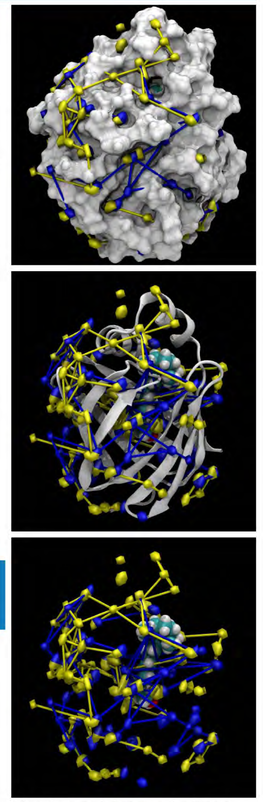
- Trajectories of individual water molecules
- Full atomistic detail

... but thousands of particles! Individual waters are not interesting but how waters behave relative to a protein: Analyze trajectories in terms of "hydration sites" in the spirit of X-ray hydration sites [1].

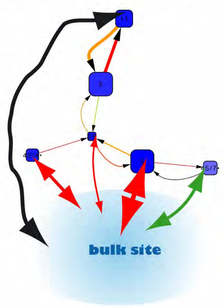
Hopping analysis of water molecules

Analyze water movement in terms of "hops" between regions in space ("sites") [2]

- Align MD trajectory to a common frame of reference (e.g. RMS-fitting to the protein backbone).
- Define *hydration sites* (e.g. where water density is larger than the bulk water density by a cutoff of 2.72).
- Define a *bulk site* (e.g. where water density of water farther away from protein than 4.5 Å has a density > 0.64 of bulk).
- Translate water movement in 3D to 1D movement between sites ("hopping trajectory"), keeping track of site entry and exit times.
- Generate graph from the 1D connectivity.

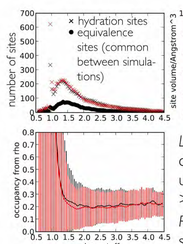


Hopping graph between hydration sites in CRBPII (cellular retinol binding protein II) from 20 ns of MD simulations. Blue: **apo** protein, yellow: **holo** complex (bound retinol shown).



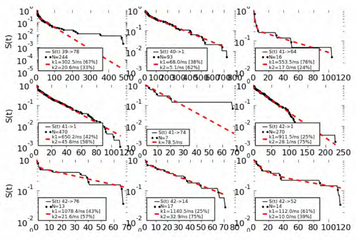
Observables

- Exchange **rates** between sites: calculate as exponential fits to the site survival function of the waiting times for the exchange $i \to j$, $S_{ji} = \langle \Theta(t_{ij} - t) \rangle$.
- **Occupancy** (average number of waters in the site volume) is calculated directly from MD and rates (both estimates agree).



Left: Sensitivity to the density cutoff (MD I-FABP 20ns). No unique definition exists but >2 is robust.

Right: Examples of fits to the site survival function $S(t)$.

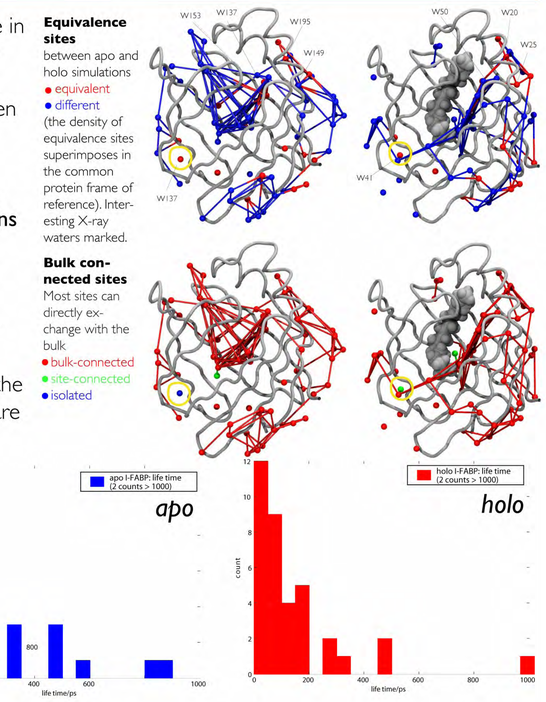


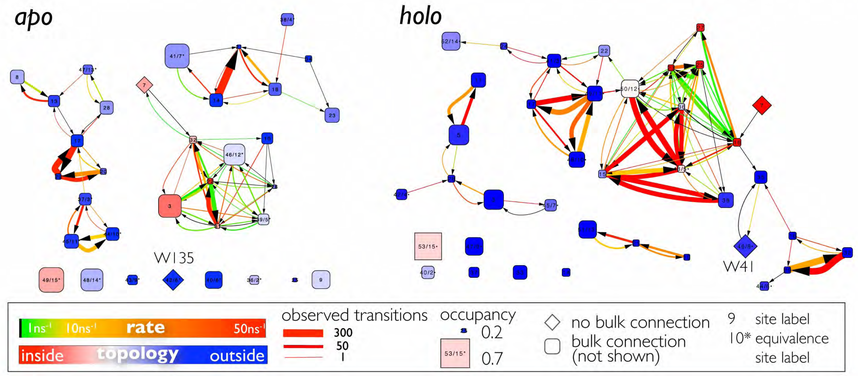
Application I: Water network in apo vs holo fatty acid binding protein

I-FABP binds fatty acids such as palmitate in a central cavity. Between 20 to 30 water molecules are seen in the cavity, both in X-ray structures and MD simulations, even with the ligand bound.

Water network in apo (1IFC) and holo (2IFB) I-FABP from 20 ns MD simulations

- 18 equivalent sites between apo and holo simulations including the structural water molecule W135/W41
- Apo simulation shows 10 of the X-ray hydration sites, holo shows 11.
- Clear water access pathways through the "portal" region and D-E gap; most sites are also directly accessible from the bulk
- Life times of sites: 2–3 long lived (>7ns) and 15–20 "ordered" (0.1 ns<t<1ns), in agreement with experiments [3] and simulations [4].
- Network representation integrates many observables and summarizes solvation dynamics.

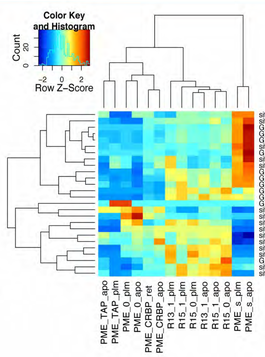




Application II: Comparison of simulation methods

Graph properties can be used to characterize water behavior at a global level and provide a "fingerprint" of a whole simulation. With long and converged samples from two or more methods we can use clustering to assess the degree of overlap. In principle answers should be insensitive to boundary conditions or system size. By applying a range of different methods to the same calculation we can cluster methods that give similar answers and those that are dramatically different. This then might be a good way to decide whether some approximations are simply too approximate to be useful.

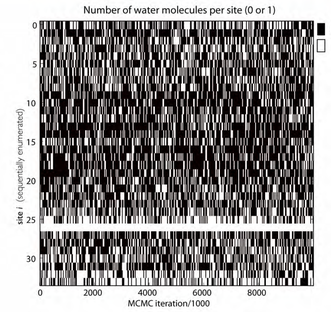
Here we compare 20 ns of apo and holo simulations with periodic boundary conditions and Ewald electrostatics (*PME*) with simulations in a solvent droplet around the FABP cavity, using the Generalized Solvent Boundary Potential in Charmm with radii R=13 Å and 15 Å. We look at multiple runs (*Rxx_0* and *Rxx_1*), short PME (1ns, *PME_s*), and a different method to calculate hydration site densities [2] (*PME_TAP*).

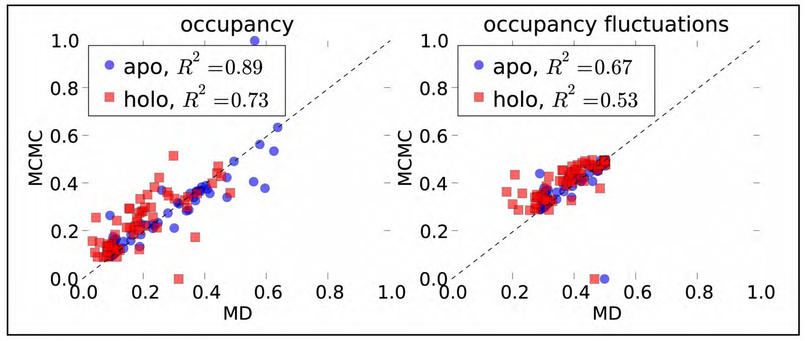


Application III: Generating solvent configurations from the hopping graph via Markov Chain Monte Carlo sampling

The hopping graph (with rates) provides a *coarse grained description* of solvation. If the MD trajectory sampled sufficiently then this description implicitly contains the probability distribution $\pi(\boldsymbol{n})$ of all the site occupancies $\boldsymbol{n}=(n_1,n_2,\ldots n_N)$.

Generating new, independent, and equilibrium-distributed solvent configurations $\boldsymbol{n}$ rapidly can be useful for free energy perturbation calculations in water accessible cavities with slow exchange or possibly for docking/screening protocols where MD or Grand Canonical Monte Carlo is too slow.

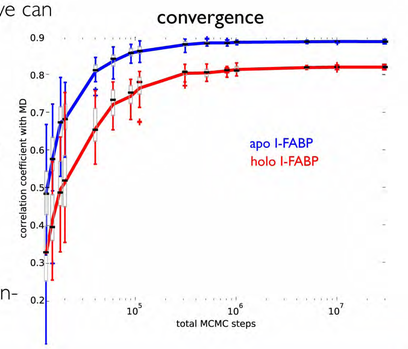




Using a Markov Chain Monte Carlo algorithm we can generate configurations in minutes:

- reproduces MD site *occupancies* (average number of particles per site)
- reproduces *fluctuations in the occupancies*
- converges after ~1 Mio steps (or ~5 Min run time on a dated PowerBook G4 laptop)
- can use any sensible definition of "site" to define regions (e.g. Voronoi volumes)

Note: This is a method to *enhance sampling* of states that have already been explored in conventional MD not to simulate solvation de novo.



Summary

- Hopping analysis captures an integrated picture of solvation dynamics.
- Embedding of the hopping graph in 3D space around the protein allows analysis of permeation pathways trivially.
- Graph properties can be used as a fingerprint to characterize solvation.
- MCMC sampling on the graph opens up the possibility of enhanced sampling of solvation states

Acknowledgements & References

[1] V Lounnas and B M Pettitt. A connected-cluster of hydration around myoglobin: correlation between molecular dynamics simulations and experiment. *Proteins*, **18** (1994),133–147.

[2] R. H. Henchman and J. A. McCammon. Extracting hydration sites around proteins from explicit water simulations. *J. Comp. Chem.*, **23** (2002), 861–869.

[3] S. Wiesner, E. Kurian, F. G. Prendergast, and B. Halle. Water molecules in the binding cavity of intestinal fatty acid binding protein: Dynamic characterization by water O^{17} and H^2 magnetic relaxation dispersion. *J. Mol. Biol.*, **286** (1999), 233–246.

[4] D. Bakowies and W. F. Van Gunsteren. Simulations of apo and holo-fatty acid binding protein: Structure and dynamics of protein, ligand and internal water. *J. Mol. Biol.*, **315** (2002), 713–736.

CRBPII simulations provided by *Hirsh Nanda*. Analysis is implemented with the MDANALYSIS python library mdanalysis.googlecode.com and NETWORKX networkx.lanl.gov. *John Stone* provided a customized version of VMD www.ks.uiuc.edu/Research/vmd for 3D network images (rendered with TACHYON). 2D networks were drawn with CYTOSCAPE www.cytoscape.org.

Funding: JHMI Department of Physiology; OB was also supported by a Junior Research Fellowship from Merton College, Oxford.