Heterogeneous Preferential Solvation of Water and Trifluoroethanol in Homologous Lysozyme Systems: Supplementary Information

Evan J. Arthur¹, John T. King¹, Kevin J. Kubarych¹, and Charles L. Brooks III¹,²,*

¹Department of Chemistry, University of Michigan, 930 N. University Ave, Ann Arbor, MI 48109-1055
²Biophysics Program, University of Michigan, 930 N. University Ave, Ann Arbor, MI 48109-1055
*Corresponding author: Phone: 734-647-6682; Email: brooksccl@umich.edu

Content

p. 2  SI table 1: number of solvent molecules in each simulation
p. 2  SI figure 1: root mean square deviation (RMSD) and circular dichroism (CD) of lysozymes
p. 3 - 4  SI figure 2: radial distribution functions of solvent from protein atoms
p. 5 - 8  SI figure 3: convergence of 3-dimensional solvent distribution versus simulation time
p. 9  SI table 2: trajectory-averaged solvent-accessible surface area of each amino acid
p. 10  SI figure 4 and table 3: comparison of results to previous X-ray studies¹,²
p. 11  SI figure 5: probability distribution of empty voxels near H15 on HEWL and H78 on humLys
p. 12  SI figure 6: cross-sections of solvent density around HEWL
p. 13  references for Supplementary Information
| Percent TFE | TFE/H₂O molecules | Percent TFE | TFE/H₂O molecules |
|------------|-------------------|------------|-------------------|
| 0          | 0 / 15,194        | 0          | 0 / 15,749        |
| 1          | 39 / 14,975       | 1          | 39 / 15,539       |
| 5          | 184 / 14,258      | 5          | 178 / 14,789      |
| 10         | 349 / 13,466      | 10         | 337 / 13,959      |
| 15         | 497 / 12,773      | 15         | 481 / 12,139      |
| 20         | 631 / 12,150      | 20         | 611 / 11,173      |

**SI table 1: Number of solvent molecules in each simulation.** Listed above are the exact numbers of solvent molecules included in each simulation. The percents of TFE v/v are approximate, and non-ideal volume effects and preferential solvation at the protein-solvent interfaces were not included when choosing the number of TFE molecules.

**SI figure 1: Root mean square deviation (RMSD) and circular dichroism (CD) of lysozymes.** In red is the CD for all simulations, which showed an ellipticity of 10.1 ± 0.8 degrees. In green is the RMSD for all simulations of protein backbone atoms from the initial structure, which was 1.0 ± 0.1 Å. Circles represent data for hen egg white lysozymes, and triangles represent data for human lysozymes. The error bars represent the standard deviation among each of the three 20 ns simulations at each concentration. The algorithm for calculating the CD ellipticity was developed by Hirst and Brooks.
SI figure 2a: Radial distribution functions of solvent from protein atoms of hen egg white lysozyme. The subfigures above are all of the solvent distribution functions as a radius from protein atoms. Red signifies data from trifluoroethanol (TFE), blue signifies data for water, and all plots were normalized to the bulk concentration of their respective solvent. The data was not averaged over the three trajectories at each concentration in order to show the variations in solvent distribution. Notice that the distribution of water remains almost the same during each simulation. The peak in the TFE $g(r)$ function is proportionately reduced as the bulk density of TFE increases. As the protein’s surface becomes saturated with TFE, a second peak begins to form between 5 and 9 Å from protein atoms. Beyond 9 Å almost no perturbation in solvent radial distribution was seen.
**SI figure 2b:** Radial distribution functions of solvent from protein atoms of human lysozyme. The sub-figures above are all of the solvent distribution functions as a radius from protein atoms. Red signifies data from trifluoroethanol (TFE), blue signifies data for water, and all plots were normalized to the bulk concentration of their respective solvent. Refer to SI figure 2a for a complete description.
SI figure 3a: Convergence of 3-dimensional solvent distribution versus simulation time of hen egg white lysozyme. The subfigures above are the averaged Pearson correlation coefficients between environments surrounding surface lying residues separated by increments of 1 ns. Each residue’s local environment was the atomic density data within a 7 Å sphere around the center of geometry of the backbone atoms, and each sphere was compared to a homologous sphere 1 ns previous from the same trajectory. Accordingly, each concentration of TFE has three separate 20 ns trajectories, each of which provides a map of protein, water, and TFE atomic occupancies. Pearson correlation coefficients were generated using eq. 4 from the main manuscript. These data indicate that atomic density from the simulations converged to one consistent distribution over the course of 20 ns: the slowest convergence to an average correlation of greater than 0.95 for TFE was 14 ns, for water was 2 ns, and for the protein was 2 ns. Note that the minimum y-axis value was set to 0.6 to enlarge the minute details of the data. The reason data oscillates with increased time is up to speculation, but it may be due to each system’s resting in a local minimum that is dissimilar from previously explored minima.
SI figure 3b: Convergence of 3-dimensional solvent distribution versus simulation time of human lysozyme. The subfigures above are the averaged Pearson correlation coefficients between environments around homologous surface lying residues separated by increments of 1 ns. Refer to SI figure 3a for a complete description of the plots.
SI figure 3c: Comparison of 3-dimensional solvent distribution versus averaged solvent distribution of hen egg white lysozyme. The subfigures above are the averaged Pearson correlation coefficients between environments around homologous surface lying residues. The environment around each residue was compared to the average atomic density at each corresponding concentration. Refer to SI figure 3a for a complete description of calculating the correlation coefficients from solvent density data. These data explain whether or not the independent trajectories converge to a common distribution of solvent density. Since each trajectory explored a different configurational pathway, the comparison of each trajectory to the average is expected to be significantly less than 1.00. Interestingly, in many simulations the protein atomic density was the most poorly correlative with the average of a concentration. This may be due to the lysozyme’s long timescales of fluctuations and the subsequent reduction of overlap of atomic density.
SI figure 3d: Convergence of 3-dimensional solvent distribution versus simulation time of human lysozyme with averaged solvent density. The subfigures above are the averaged Pearson correlation coefficients of environments around homologous surface lying residues. The environment around each residue was compared to the average atomic density at each corresponding concentration. Refer to SI figures 3a and 3c for a complete description of calculating the correlation coefficients from solvent density data. Interestingly, the protein atomic density remains consistent among the simulations in comparison to HEWL (in SI figure 3c). This observation indicates that human lysozyme was well-equilibrated relative to hen egg white lysozyme, and did not explore conformations significantly different from its initial structure.
### SI table 2: Solvent-accessible surface area of each residue

SASA was calculated using the double cubic lattice method\(^4\) within the \textit{g}\_\textit{sas} utility of GROMACS. The SASA values were averaged for all saved protein structures. Data for residue 43 on humLys was ignored due to its lack of a homologous counterpart on HEWL. Secondary structure assignments were made using the Smith-Waterman Sequence Alignment method on PDB.org. \(^5, 6\)

| Residue | Hen egg white lysozyme | Human lysozyme |
|---------|------------------------|----------------|
|         | SASA (Å\(^2\)) | SASA (Å\(^2\)) |         | SASA (Å\(^2\)) | SASA (Å\(^2\)) |         | SASA (Å\(^2\)) | SASA (Å\(^2\)) |         | SASA (Å\(^2\)) | SASA (Å\(^2\)) |         | SASA (Å\(^2\)) | SASA (Å\(^2\)) |
| 1       | 101 | unstructured | 117 | unstructured | 65 | beta bridge | 100 | beta bridge |
| 92      | 92  | beta bridge | 103 | beta bridge | 41 | unstructured | 42  | unstructured |
| 18      | 87  | unstructured | 25  | unstructured | 80  | unstructured | 92  | unstructured |
| 30      | 55  | unstructured | 120 | unstructured | 140 | unstructured | 176 | unstructured |
| 5       | 83  | alpha-helix | 92  | alpha-helix | 70  | unstructured | 129 | unstructured |
| 38      | 77  | unstructured | 29  | unstructured | 93  | unstructured | 92  | unstructured |
| 0       | 2   | turn | 3  | turn | 50 | unstructured | 58  | unstructured |
| 5       | 55  | unstructured | 96  | unstructured | 96  | unstructured | 98  | unstructured |
| 10      | 35  | turn | 107 | turn | 75 | unstructured | 42  | turn |
| 15      | 35  | turn | 113 | turn | 77 | turn | 77  | turn |
| 3       | 3   | turn | 189 | turn | 80 | turn | 84  | turn |
| 127     | 103 | turn | 42  | turn | 22 | 3/10-helix | 22  | 3/10-helix |
| 202     | 25  | turn | 22  | turn | 4  | 3/10-helix | 17  | 3/10-helix |
| 15      | 55  | turn | 76  | turn | 80 | turn | 22  | turn |
| 32      | 32  | turn | 0  | turn | 102 | turn | 29  | turn |
| 1       | 16  | beta bridge | 53  | beta bridge | 16 | beta bridge | 49  | beta bridge |
| 67      | 80  | unstructured | 7  | unstructured | 5  | unstructured | 9  | unstructured |
| 88      | 66  | unstructured | 30  | unstructured | 92  | unstructured | 118 | unstructured |
| 25      | 13  | alpha-helix | 14  | alpha-helix | 50 | alpha-helix | 52  | alpha-helix |
| 1       | 3   | alpha-helix | 17  | alpha-helix | 4  | alpha-helix | 5  | alpha-helix |
| 17      | 3   | alpha-helix | 19  | alpha-helix | 7  | alpha-helix | 7  | alpha-helix |
| 0       | 0   | alpha-helix | 19  | alpha-helix | 4  | alpha-helix | 7  | alpha-helix |
| 95      | 90  | alpha-helix | 17  | alpha-helix | 3  | alpha-helix | 7  | alpha-helix |
| 30      | 1   | alpha-helix | 19  | alpha-helix | 1  | alpha-helix | 0  | alpha-helix |
| 2       | 2   | alpha-helix | 19  | alpha-helix | 95 | alpha-helix | 3  | alpha-helix |
| 65      | 84  | alpha-helix | 97  | alpha-helix | 18  | alpha-helix | 10  | alpha-helix |
| 73      | 97  | alpha-helix | 97  | alpha-helix | 2  | alpha-helix | 2  | alpha-helix |
| 35      | 57  | turn | 14  | turn | 100 | unstructured | 144 | unstructured |
| 12      | 19  | turn | 42  | turn | 103 | unstructured | 62  | unstructured |
| 86      | 17  | turn | 17  | turn | 92  | unstructured | 131 | unstructured |
| 19      | 17  | turn | 40  | turn | 108 | unstructured | 103 | unstructured |
| 45      | 45  | turn | 2  | turn | 105 | turn | 9  | turn |
| 4       | 4   | turn | 2  | turn | 105 | turn | 9  | turn |
| 41      | 41  | turn | 150 | turn | 105 | turn | 9  | turn |
| 26      | 19  | turn | 19  | turn | 105 | turn | 9  | turn |
| 123     | 73  | unstructured | 73  | unstructured | 103 | unstructured | 103 | unstructured |
| 43      | 92  | beta strand | 55  | beta strand | 20  | unstructured | 28  | unstructured |
| 44      | 92  | beta strand | 127 | beta strand | 67  | unstructured | 55  | unstructured |
| 45      | 171 | unstructured | 36  | unstructured | 110 | unstructured | 55  | unstructured |
| 69      | 74  | unstructured | 16  | unstructured | 17  | unstructured | 17  | unstructured |
| 173     | 79  | turn | 18  | turn | 15  | unstructured | 15  | unstructured |
| 111     | 50  | turn | 147 | turn | 128 | unstructured | 128 | unstructured |
| 25      | 166 | unstructured | 115 | unstructured | 95  | unstructured | 95  | unstructured |
| 50      | 17  | beta strand | 0  | beta strand | 149 | unstructured | 109 | unstructured |
| 14      | 16  | beta strand | 115 | beta strand | 149 | unstructured | 109 | unstructured |
| 42      | 10  | beta strand | 120 | beta strand | 98  | unstructured | 132 | unstructured |
| 14      | 20  | beta strand | 122 | beta strand | 122 | unstructured | 98  | unstructured |
| 53      | 77  | beta strand | 122 | beta strand | 122 | unstructured | 98  | unstructured |
| 2       | 0   | turn | 2  | turn | 98  | unstructured | 98  | unstructured |
| 55      | 3   | turn | 120 | turn | 120 | turn | 98  | turn |
| 6       | 6   | turn | 120 | turn | 120 | turn | 98  | turn |
| 38      | 25  | turn | 120 | turn | 120 | turn | 98  | turn |
| 48      | 13  | turn | 120 | turn | 120 | turn | 98  | turn |
| 60      | 2   | turn | 122 | turn | 122 | turn | 98  | turn |
| 22      | 2   | turn | 122 | turn | 122 | turn | 98  | turn |
maps of local concentration of TFE on hen egg white lysozyme

crystal contact map of TFE
local concentration of TFE v/v

| Residue Number | local % TFE v/v | crystal contacts1,2 |
|----------------|-----------------|---------------------|
| 1              | 5.2  +          | 36                  |
| 6.5            | 8.2  +          | 71                  |
| 7.9            | 8.1  +          | 106                 |
| 9.1            | 8.3  +          |                      |
| 5              | 13.5 +          | 40                  |
| 13.6           | 11.9 +          | 75                  |
| 10             | 21.2 +          | 45                  |
| 13.0           | 8.2  +          | 80                  |
| 15             | 8.1  +          | 50                  |
| 6.8            | 8.2  +          | 85                  |
| 5.7            | 19.6 +          | 120                 |
| 7.6            | 8.4  +          |                    |
| 20             | 11.0 +          | 10.3 +              |
| 8.6            | 10.3 +          | 10.3 +              |
| 9.0            | 19.6 +          | 8.2 +               |
| 25             | 7.6  +          | 55                  |
| 9.8            | 13.9 +          | 20.5 +              |
| 60             | 16.3 +          | 125                 |
| 14.9           | 18.3 +          | 10.1 +              |
| 30             | 65              | 129                 |
| 11.1           | 14.3 +          |                    |
| 15.3           | 11.5 +          |                    |
| 35             | 17.1 +          |                    |
| 19.2           | 17.1 +          |                    |
| 10.7           | 9.7  +          |                    |
| 9.8            | 9.9  +          |                    |
| 8.5            | 8.5  +          |                    |
| 10             | 10.9 +          |                    |
| 11.1           | 11.1 +          |                    |
| 7.6            | 10.7 +          |                    |
| 6.8            | 10.7 +          |                    |
| 5.7            | 10.7 +          |                    |
| 4.9            | 10.7 +          |                    |
| 3.8            | 10.7 +          |                    |
| 2.9            | 10.7 +          |                    |

SI figure 4 and table 3: comparison of results to previous X-ray studies1,4 All local concentrations in this section were calculated from simulation data averaged over the three trajectories at 10 % TFE v/v. Figure 4 shows HEWL colored by crystal contacts from X-ray diffraction studies (left) and calculated local concentrations of TFE (right). Buried residues were colored grey. Table 3 provides an explicit map of the concentration of TFE at each residue, and whether or not it was a crystal contact. Buried residues were omitted from these calculations. The red highlight indicates which residues had a high local concentration of TFE, and whether or not they matched a crystal contact. The 50 highlighted residues (56% of the 88 surface residues) located 18 crystal contacts (50% of the 36 surface residue contacts).
SI figure 5: probability distribution of empty voxels near H15 on HEWL and H78 on humLys. During each timestep the number of voxels without water were counted from the local volume around the two histidines. For details on defining the voxels and local volumes around a residue, please refer to the “Local Percent of TFE by Volume” section of the main paper. The resulting data are plotted above as probability distributions of finding a particular number empty voxels near the histidines, with H15 on HEWL in Figure 5a, and H78 on humLys in Figure 5b. The distributions widen as more TFE is introduced into the bulk solution, which indicates a dehydration of the local environment around the histidines. In Figure 5c, the standard distribution from the Gaussian fits of Figures 5a and 5b are plotted and compared between the proteins. The error bars show the 95% confidence interval of the fit. We see that in experimentally-comparable concentrations of 10% TFE and below, the H15 on HEWL is more likely to experience a completely evacuated local volume than H78 on humLys during any given timestep, and that of the two residues, H15 is significantly less hydrated at 10% TFE.
SI figure 6: cross-sections of solvent density around HEWL. For each concentration of 1, 5, 10, and 15 % TFE v/v, the solvent density of TFE is plotted as colored isosurfaces around HEWL. All data is normalized to the bulk density of TFE. Blue, green, and red isosurfaces encompass 3, 7, and 11 times the bulk density of TFE. Since each subsequent isosurface is inside another, the data was sliced open to expose both the protein and to reveal how one isosurface is nested in another. As the bulk concentration of TFE increases, the corresponding solvent hot spots reduce in size.
References

1. Buck, M.; Radford, S. E.; Dobson, C. M., A Partially Folded State of Hen Egg-White Lysozyme in Trifluoroethanol - Structural Characterization and Implications for Protein Folding. *Biochemistry* **1993**, *32*, 669-678.

2. Lehmann, M. S.; Mason, S. A.; McIntyre, G. J., Study of Ethanol-Lysozyme Interactions Using Neutron Diffraction. *Biochemistry* **1985**, *24*, 5862-5869.

3. Hirst, J. D.; Brooks III, C. L., Helicity, Circular Dichroism and Molecular Dynamics of Proteins. *J. Mol. Biol.* **1994**, *243*, 173-178.

4. Eisenhaber, F.; Lijnzaad, P.; Argos, P.; Sander, C.; Scharf, M., The double cubic lattice method: Efficient approaches to numerical integration of surface area and volume and to dot surface contouring of molecular assemblies. *J. Comput. Chem.* **1995**, *16*, 273-284.

5. Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E., The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235-42.

6. Smith, T. F.; Waterman, M. S., Identification of common molecular subsequences. *J. Mol. Biol.* **1981**, *147*, 195-197.