

Supporting Information

Detecting first hydration shell structure around biomolecules at interfaces

Daniel Konstantinovsky^{1,2†}, Ethan A. Perets^{1†}, Ty Santiago^{3,‡}, Luis Velarde³, Sharon Hammes-Schiffer^{1*}, and Elsa C. Y. Yan^{1*}

¹Department of Chemistry, Yale University, New Haven, CT 06520, USA

²Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT 06520, USA

³Department of Chemistry, University at Buffalo, Buffalo, NY 14260, USA

[‡]Current address: Department of Chemistry, Yale University, New Haven, CT 06520, USA

[†] These authors contributed equally

*Corresponding Authors

Email Addresses: sharon.hammes-schiffer@yale.edu, elsa.yan@yale.edu

Contents

Description of the SFG spectrometer.....	S2
Experimental spectra: Sewn spectra	S3
Experimental spectra: Fits.....	S4
Quartz-only spectrum.....	S7
MD equilibration details	S7
Selection methods for water subsets	S8
Construction of the Hamiltonian for calculation of the SFG spectrum	S10
Illustration of the lack of a reflection plane in the first hydration shell	S11
References.....	S12

Description of the SFG spectrometer

A detailed description and characterization of the custom SFG spectrometer utilized in the studies can be found in Ma, *et al.*¹ Briefly, a 5 kHz ultrafast regenerative amplifier (Spitfire, Spectra-Physics) is pumped by two Nd:YLF lasers (Empower, Spectra-Physics) and seeded with a Ti:Sapphire laser (MaiTai, Spectra-Physics). The amplifier creates 120 femtosecond pulses with a total power of 6 W. A 50/50 beamsplitter is used to divide the amplifier output along two paths. The first path (3 W) is passed through a custom pulse-shaper yielding narrow-band pulses centered at 800 nm. The second path (3 W) is used to pump a tunable optical parametric amplifier (TOPAS, Spectra-Physics/Light Conversion) to generate broad-band infrared pulses. The resultant visible and infrared beams are spatially and temporally overlapped at the sample surface. The SFG signal reflected from the sample interface is dispersed with a monochromator, and the spectrum is collected with a CCD. The polarization of the SFG, visible, and infrared beams for all experiments reported in the main text was *p*-polarized, *s*-polarized, and *p*-polarized, respectively.²

Experimental spectra: Sewn spectra

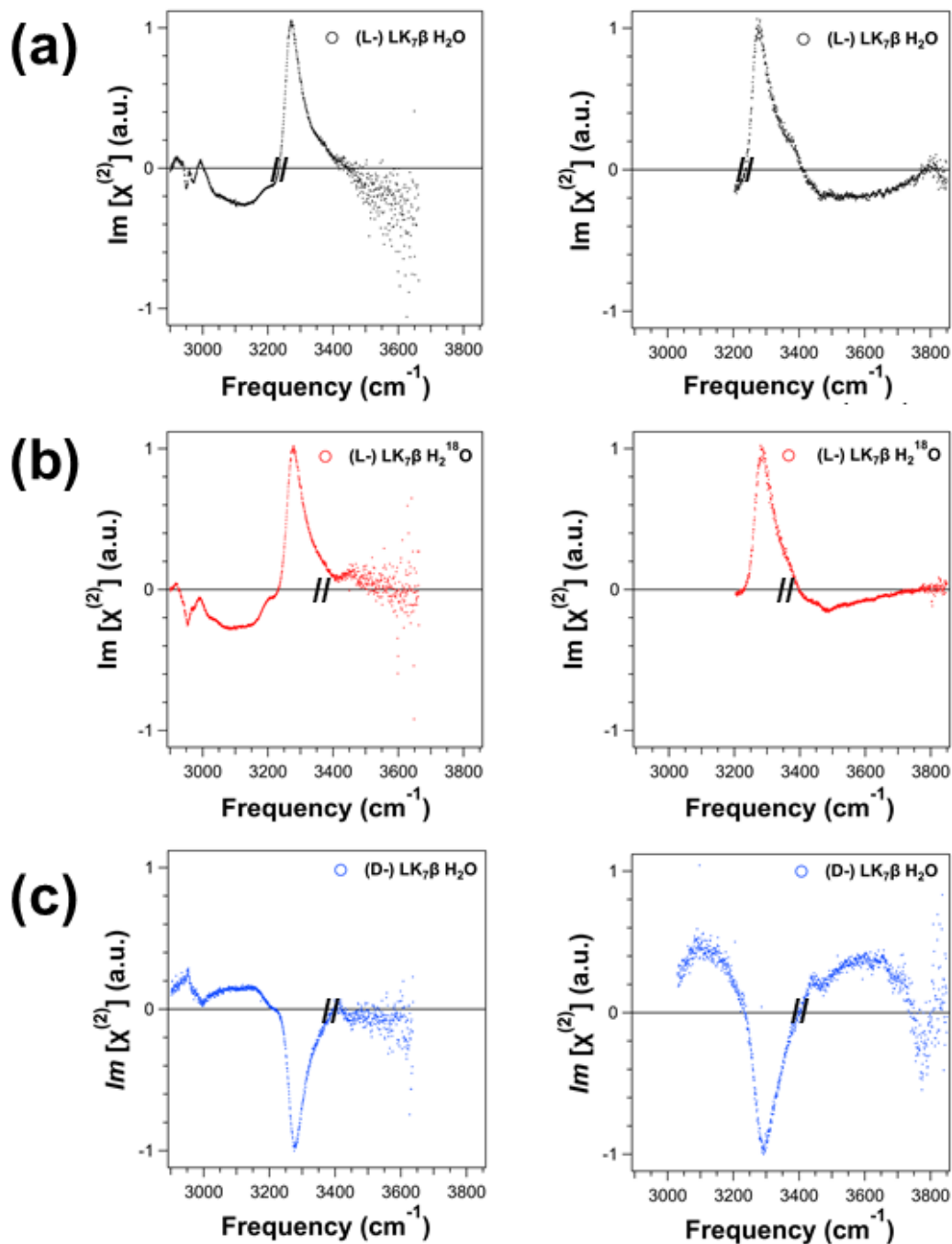


Figure S1. The two spectral windows collected for each spectrum presented in Figure 2 in the main text. The two individual spectra were overlaid and normalized to the N-H stretching peak of $\text{LK}_7\beta$ at $\sim 3270 \text{ cm}^{-1}$ and stitched together at the point indicated with hashmarks. The noisy regions that appear at the high frequency ends of the spectra are due to normalization of the broadband spectra by clean quartz. a.u.: arbitrary units.

Experimental spectra: Fits

Description of Fitting Experimental Spectra.

In the spectral region 2900 – 3800 cm⁻¹, C-H and N-H stretching of the LK7β protein, O-H stretching of water, and possibly additional signals due to Fermi resonance of water can contribute to the vibrational spectra (for more information on vibrational analyses of proteins, see Barth and Zscherp).³ The experimental spectra in Figure 2b were fit with Lorentzians according to:

$$\text{Im}[\chi^{(2)}] \propto \text{Im}\left[\sum_q \frac{A_q}{\omega_{IR} - \omega_q - i\Gamma_q}\right] \quad (\text{S1})$$

where ω_q is the resonant frequency of the q^{th} vibrational mode, A_q is the amplitude, Γ_q is the half-width half-maximum of the q^{th} vibrational mode, and ω_{IR} is the frequency of the incident infrared. Therefore, each vibrational peak contributes three parameters (ω_q, A_q, Γ_q). The minimum number of vibrational peaks needed to fit the experimental data (Figure 2b) is determined by residual analyses as shown in Figure S2. Eleven peaks are used to fit the spectra, and they are tentatively assigned as shown in Table S1.

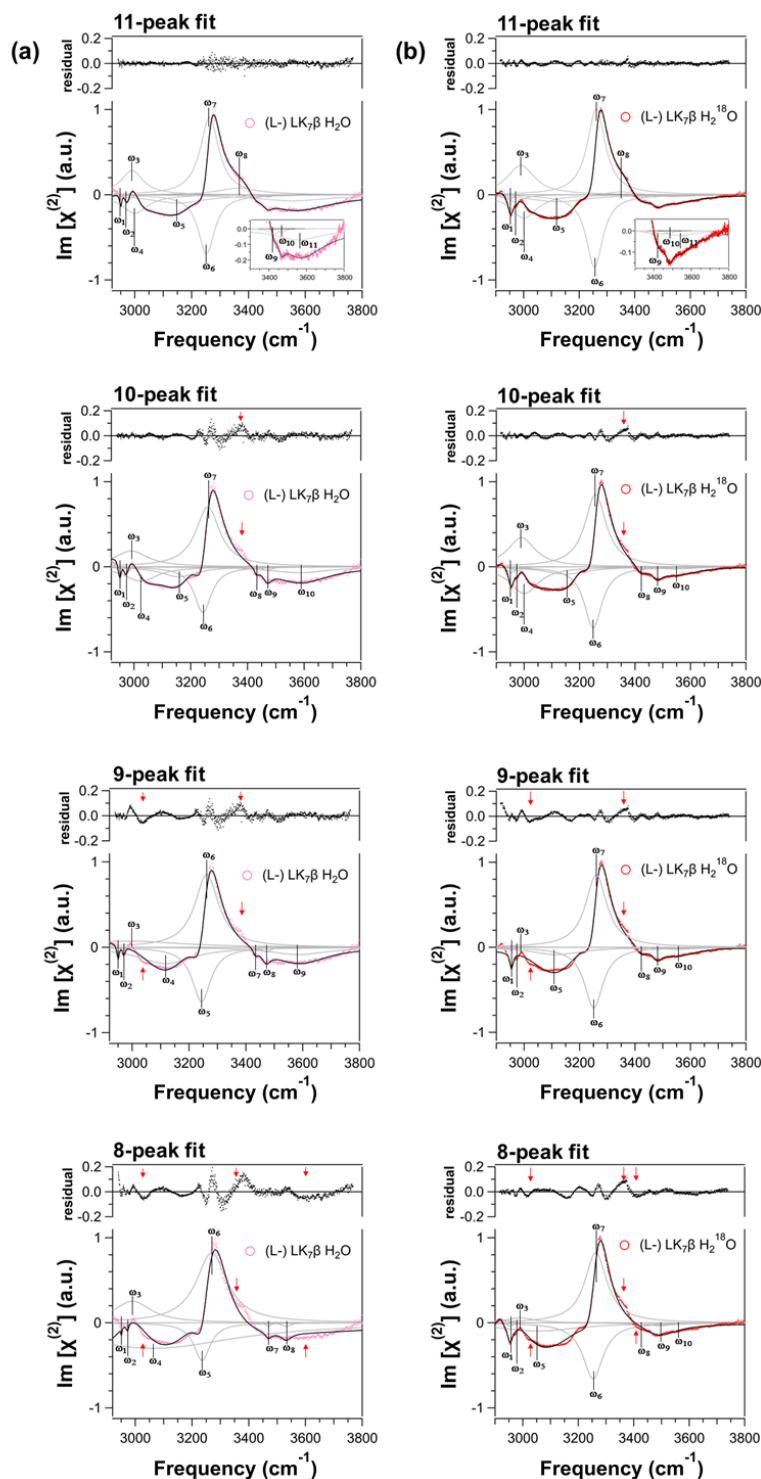


Figure S2. Fitting and residual analyses of the fitting to the experimental spectra of (a) LK₇β-H₂O and (b) LK₇β-H₂¹⁸O presented in Figure 2b in the main text with (from top to bottom) 11, 10, 9, and 8 vibrational peaks. Fitting parameters for the 11-peak fitting (top row) are reported in Table S1. Individual vibrational peaks of each fitting are shown as gray lines (these have been scaled to fit within the viewing window). Residuals between the experimental data and the fitting are included above each spectrum. The residual analyses suggest that 11 component vibrational resonances are needed to model the experimental spectra in Figure 2b of the main text.

Parameter	Figure 2b (L-) LK7 β H ₂ O	Figure 2b (L-) LK7 β H ₂ ¹⁸ O	Tentative Assignments
ω_1	2950 \pm 1	2952 \pm 1	C-H (LK7 β)
A ₁	-0.7 \pm 0.2	-2.8 \pm 0.9	
Γ_1	4.2 \pm 0.7	12.5 \pm 1.6	
ω_2	2973 \pm 1	2970 \pm 2	C-H (LK7 β)
A ₂	-3.1 \pm 0.6	-13.1 \pm 12.2	
Γ_2	13.1 \pm 1.3	34.3 \pm 9.1	
ω_3	2990 \pm 0*	2990 \pm 0*	C-H (LK7 β)
A ₃	40.7 \pm 0*	94.1 \pm 35.6	
Γ_3	48.0 \pm 2.6	60.7 \pm 0.0*	
ω_4	3000 \pm 2	3001 \pm 7	C-H or N-H (LK7 β)
A ₄	-36.7 \pm 3.2	-71.1 \pm 19.7	
Γ_4	56.6 \pm 4.4	58.0 \pm 1.5	
ω_5	3151 \pm 5	3125 \pm 4	O-H (H ₂ O) and/or F.R.** with H ₂ O bend overtone
A ₅	-35.0 \pm 7.4	-46.3 \pm 10.3	
Γ_5	119.5 \pm 16.8	125.7 \pm 17.6	
ω_6	3252 \pm 1	3255 \pm 1	N-H (LK7 β)
A ₆	-74.1 \pm 2.9	-139.6 \pm 2.3	
Γ_6	35.9 \pm 0.6	36.6 \pm 0.4	
ω_7	3263 \pm 1	3261 \pm 1	N-H (LK7 β)
A ₇	108.8 \pm 0.0*	179.1 \pm 0.0*	
Γ_7	40.0 \pm 0.7	39.6 \pm 0.5	
ω_8	3366 \pm 6	3350 \pm 2	O-H (H ₂ O)
A ₈	20.3 \pm 4.0	7.8 \pm 2.1	
Γ_8	85.9 \pm 10.0	50.1 \pm 6.6	
ω_9	3426 \pm 2	3420 \pm 10	N-H (LK7 β)
A ₉	-3.2 \pm 1.6	-4.5 \pm 0.8	
Γ_9	31.2 \pm 7.6	40.9 \pm 7.8	
ω_{10}	3468 \pm 2	3483 \pm 1	N-H (LK7 β)
A ₁₀	-1.6 \pm 0.2	-1.4 \pm 0.2	
Γ_{10}	20.0 \pm 0.0*	20.0 \pm 0.0*	
ω_{11}	3555 \pm 10	3538 \pm 6	O-H (H ₂ O)
A ₁₁	-34.4 \pm 1.8	-9.5 \pm 1.5	
Γ_{11}	161.1 \pm 12.2	95.6 \pm 7.5	

*Parameters held during fitting.

**F.R. = Fermi Resonance

Table S1. The fitting parameters (see equation S1, above) used for the 11-peak fitting shown in Figure S2. Water contributions are suggested based on H₂¹⁸O substitution.

Quartz-only spectrum

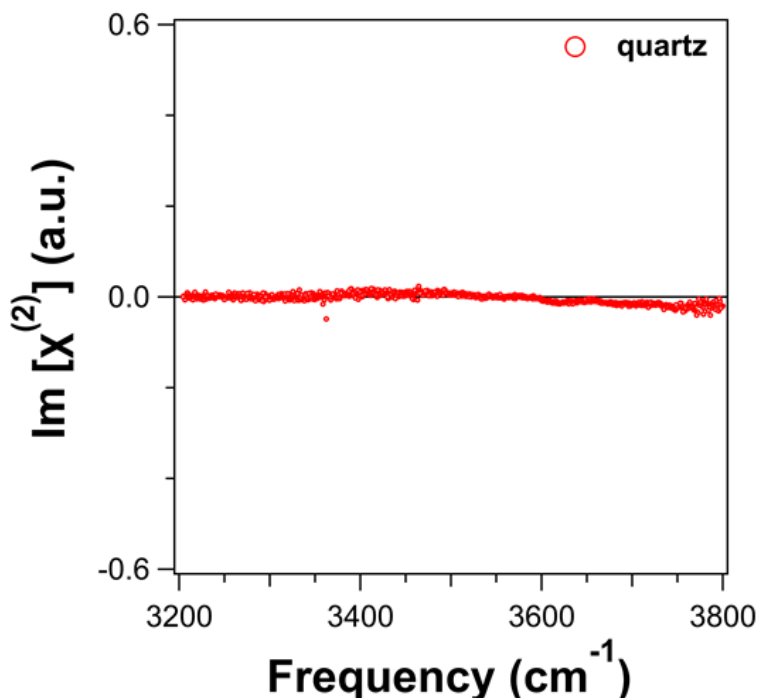


Figure S3. Chiral SFG spectrum of clean quartz substrate in the absence of LK7 β . The spectrum is normalized by GaAs. The spectrum shows no vibrational resonances. 2 mins. acquisition time. a.u.: arbitrary units.

MD equilibration details

See the Methods section in the main text for more information about the MD simulations. The equilibration was initiated with an energy minimization of the solvent. This was followed by a 500 ps NVT equilibration of the solvent with a restrained protein (force constant of 500 kcal mol⁻¹ Å⁻²). Then the positions of the hydrogen atoms of the protein were minimized. Subsequently, the system was minimized three times, gradually reducing the restraints on the protein heavy atoms (100, then 50, then 10 kcal mol⁻¹ Å⁻²), followed by energy minimization of the entire system. This minimization was followed by simulated annealing over 360 ps from 0 K to 298 K under NVT conditions, followed by 6 ns of NVT equilibration at 298 K. No NPT equilibration was done because of the vacuum-water interface.

Selection methods for water subsets

The subsets in Figure 4 were selected using a combination of MDAnalysis selection logic,⁴ in-house code, and the Voronoi tessellation library `freud`,⁵ which uses the `voro++` engine⁶ to efficiently generate Voronoi diagrams. All selections were restricted to water molecules in the first hydration shell. All hydrogen bonds were identified if the donor-acceptor heavy atom distance was less than 3.5 Å and the donor-hydrogen-acceptor angle was greater than 135°. This cutoff distance is larger than typically used for defining hydrogen bonds because we wanted to ensure that the non-hydrogen-bonded subsets were truly free from hydrogen-bonding interactions with the protein. Periodic boundary conditions were taken into account for all water selections, including the Voronoi tessellations and hydrogen bonding analyses.

Figure 4a (entire first hydration shell)

The `freud` library was used to generate a Voronoi diagram with points corresponding to all atoms in the system. The neighbor list feature was then used to identify water molecules with at least one atom's Voronoi cell bordering the Voronoi cell of at least one protein atom. All atoms in the water molecules were considered as possible neighbors to the protein.

Figure 4b (backbone)

Backbone-associated water molecules were selected using the following MDAnalysis selection logic: *byres (rename HOH and (around 4.5 ((not rename HOH) and (name N H CA HA C O HNI HN2))) and (not around 3.0 ((rename LEU) and not name N H CA HA C O HNI HN2) or ((rename LYS) and not name N H CA HA C O HNI HN2)))*

This selection was fairly liberal in terms of backbone-associated water molecules but ensured that no water molecules closely associating with the sidechains were selected.

Figure 4c (sidechains)

Sidechain-associated water molecules were selected using the following MDAnalysis selection logic: *byres (rename HOH and (around 4.5 ((rename LYS) and not name N H CA HA C O HNI HN2) or ((rename LEU) and not name N H CA HA C O HNI HN2)) and (not around 3.0 (not rename HOH) and (name N H CA HA C O HNI HN2)))*

This selection was fairly liberal in terms of sidechain-associated water molecules but ensured that no water molecules closely associating with the backbone were selected.

Figure 4d (backbone – not hydrogen bonded to protein)

First, molecules were preselected using the backbone MDAnalysis selection logic. Then, all water molecules hydrogen bonded to the C=O, NH, or NH₃⁺ groups on the protein were identified and excluded, and the remaining water molecules not hydrogen bonded to the protein were considered in this selection. Both acceptor and donor scenarios were considered for hydrogen bonds with NH, although water donor configurations that met our hydrogen bonding definition were rare. The terminal NH₂ groups were also considered as NH groups.

Figure 4e: (backbone – hydrogen bonded to C=O)

First, molecules were preselected with the following MDAnalysis selection logic to reduce the number of water molecules needing to be processed by our hydrogen bond identification code (to improve performance with no effect on the result): *byres ((rename HOH) and (around 3.5 (not rename HOH NA CL Na Cl)))*. Then, water molecules donating a hydrogen bond to a C=O group were identified.

Figure 4f: (backbone – hydrogen bonded to NH)

The same procedure was followed as above with C=O groups, except that water molecules *accepting* hydrogen bonds from NH groups were considered. The water molecule donor scenario was not considered after it was found that these configurations are rare and produce a negligible SFG signal.

Figure 4g: (backbone, hydrogen bonded to C=O with very short hydrogen bonds)

First, water molecules hydrogen bonded to C=O were identified. Then, the following MDAnalysis selection was used to identify water molecules forming very short hydrogen bonds to the C=O groups: *byres ((rename HOH) and (around 1.6 (not rename HOH NA CL Na Cl)))*. No water molecules were found to form such short hydrogen bonds to any other functional group on the protein.

Figure 4h (sidechains – not hydrogen bonded to protein)

The same procedure was followed as for the backbone-associated not hydrogen bond water molecules, except the sidechain selection logic was used.

Figure 4i (sidechains – hydrogen bonded to lysine NH₃⁺)

The same procedure was followed as for the C=O groups and NH groups, except water molecules accepting hydrogen bonds from NH₃⁺ groups were identified.

Construction of the Hamiltonian for calculation of the SFG spectrum

The exciton Hamiltonian for the SFG spectrum calculation was constructed as illustrated in Figure S4. The main idea is to treat only the selected water molecules in the exciton Hamiltonian and to regard those water molecules, the other water molecules, and the protein as point charges influencing the local electric field, which in turn leads to transition dipoles and polarizabilities for each selected O-H group. All water selections began by selection of the first hydration shell (see Methods). The number of water molecules in the selection could fluctuate slightly between frames without negatively impacting the calculation because the calculation used the inhomogeneous limit approximation, where the signal is conceived as a simple average of individual frames.

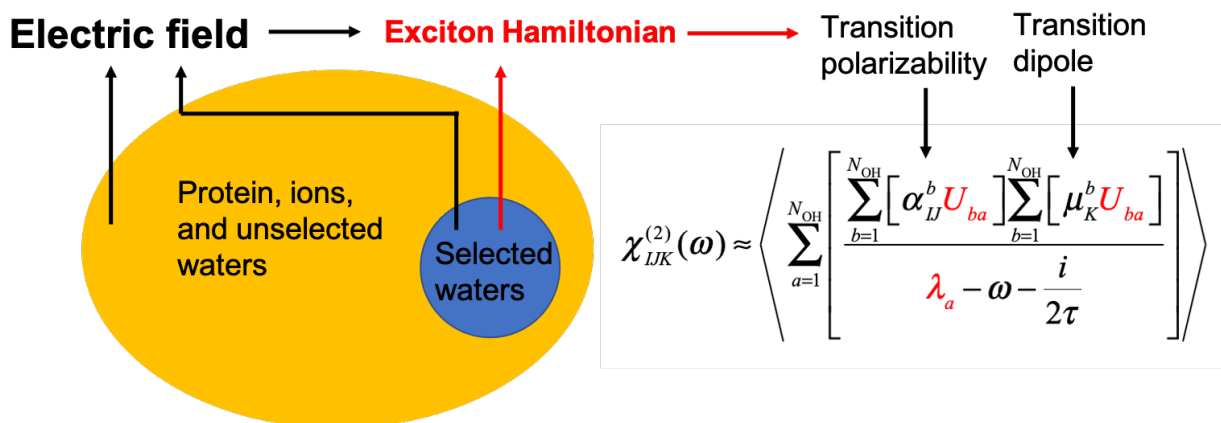


Figure S4. A schematic of our adaptation of Skinner's electric field mapping approach to calculate SFG spectra arising from subsets of the water molecules in the system. Here, α is the transition polarizability tensor, μ is the transition dipole vector, U are the eigenvectors of the Hamiltonian, and λ are the eigenvalues. τ is the vibrational lifetime (1.3 ps here).

Illustration of the lack of a reflection plane in the first hydration shell

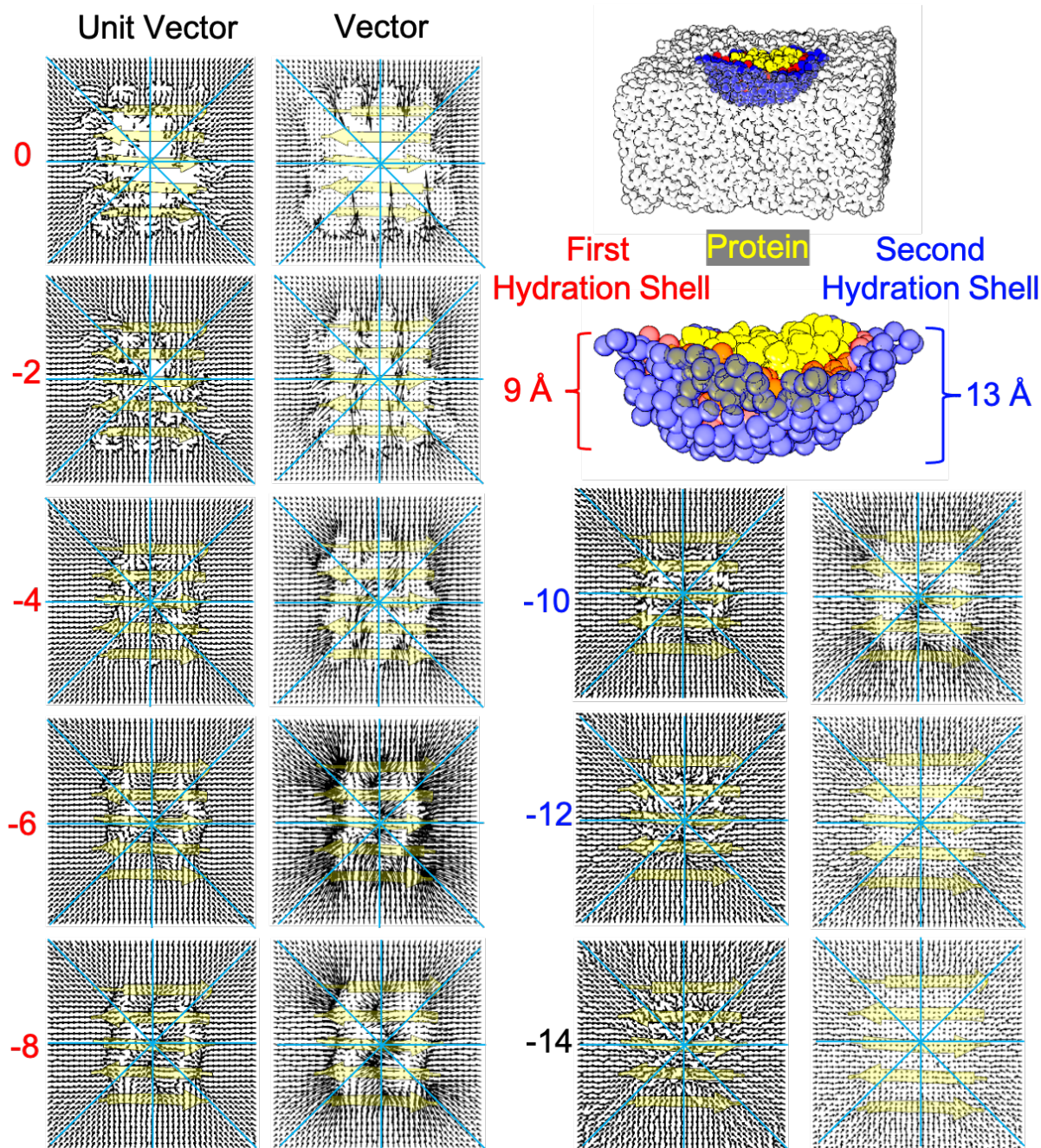


Figure S5. Water dipole moment vector directions at various depths (in Å) from the vacuum-water interface. The LK7 β protein is shown in yellow. The blue lines are a guide for the eye. This figure shows that the lack of reflection plane perpendicular to the vacuum-water interface seen in Figure 6 in the main text extends to the bottom of the first hydration shell (~ -8 Å). The water dipole pattern becomes completely symmetric only at -12 to -14 Å, but the greatest asymmetry is between 0 and -8 Å. Both unit vector and vector with magnitude representations are shown.

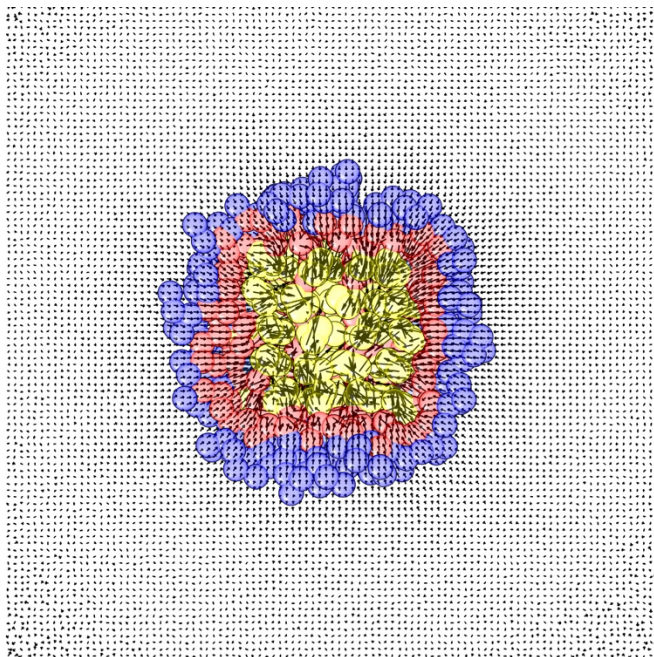


Figure S6. Unnormalized relative magnitudes of the dipole moment vectors at various locations around LK₇β. The ordering of water dipoles past the second hydration shell is minimal relative to the ordering near the protein (see Figure 6 in the main text for comparison).

References

1. Ma, G.; Liu, J.; Fu, L.; Yan, E. C. Y. Probing Water and Biomolecules at the Air—Water Interface with a Broad Bandwidth Vibrational Sum Frequency Generation Spectrometer from 3800 to 900 cm⁻¹. *Appl. Spectrosc.* **2009**, *63*, 528-537.
2. Yan, E. C. Y.; Fu, L.; Wang, Z.; Liu, W. Biological Macromolecules at Interfaces Probed by Chiral Vibrational Sum Frequency Generation Spectroscopy. *Chem. Rev.* **2014**, *114*, 8471-8498.
3. Barth, A.; Zscherp, C. What vibrations tell about proteins. *Quarterly Reviews of Biophysics* **2002**, *35*, 369-430.
4. Michaud-Agrawal, N.; Denning, E. J.; Woolf, T. B.; Beckstein, O. MDAAnalysis: A toolkit for the analysis of molecular dynamics simulations. *J. Comput. Chem.* **2011**, *32*, 2319-2327.
5. Ramasubramani, V.; Dice, B. D.; Harper, E. S.; Spellings, M. P.; Anderson, J. A.; Glotzer, S. C. freud: A software suite for high throughput analysis of particle simulation data. *Comput. Phys. Commun.* **2020**, *254*, 107275.
6. Rycroft, C. H. VORO++: A three-dimensional Voronoi cell library in C++. *Chaos* **2009**, *19*, 041111.