Abstract

Probing Diverse Biological Systems and their Hydration Shells by Modeling Chiral Sum Frequency Generation Spectroscopy

Daniel Konstantinovsky

2023

All biological processes occur in aqueous solution. Water dictates biomolecular structure, is responsible for a large portion of the free energy of ligand binding, and participates structurally in protein-protein interactions. However, few experimental techniques can probe the biomolecular hydration shell and its diverse environments. Chirality-sensitive vibrational sum frequency generation spectroscopy (chiral SFG) is a powerful technique that probes hydration shell water. The work presented here clarifies the conditions required for a chiral SFG signal to emerge, identifies the water immediately neighboring a biomolecule as the source of the aqueous chiral SFG signal, and investigates the interaction between biomolecules and their hydration shell by modeling the responses of the water and biomolecule and the vibrational coupling between them. In the process, this work extends the much-used electrostatic frequency mapping method to the NH stretch of proteins and develops Voronoi tessellation methods for computationally dissecting molecular systems. Taken together, this work helps to put chiral SFG on a firm theoretical and computational footing, extending the pioneering work of Simpson, Skinner, Corcelli, Knoester, Jansen, and Morita.

Probing Diverse Biological Systems and their Hydration Shells by Modeling Chiral Sum Frequency Generation Spectroscopy

A Dissertation
Presented to the Faculty of the Graduate School
of
Yale University
In Candidacy for the Degree of
Doctor of Philosophy

by

Daniel Konstantinovsky

Dissertation Directors: Prof. Sharon Hammes-Schiffer and Prof. Elsa C.Y. Yan
December 2023

Table of Contents

ABSTRACT ..... **I**

LIST OF FIGURES ..... **VIII**

LIST OF TABLES ..... **XIII**

LIST OF ABBREVIATIONS ..... **XIII**

ACKNOWLEDGEMENTS ..... **XIV**

DEDICATION ..... **XV**

CHAPTER 1: INTRODUCTION **1**

The biomolecular hydration shell **2**

Vibrational spectroscopy **3**

The second-order spectroscopic response **4**

Surface selectivity **5**

Heterodyne SFG experiments **7**

General theory of the second-order response **8**

Vibrational SFG in the frequency and time domains **14**

Chiral SFG of achiral molecules in chiral configurations **18**

Calculating the chiral SFG signal of a uniaxial system **21**

The meaning of the sign of the chiral (*psp*) response and achiral (*ssp*) response **22**

Preview of Chapters **26**

CHAPTER 2: MIRROR-IMAGE ANTIPARALLEL B-SHEETS ORGANIZE WATER MOLECULES INTO SUPERSTRUCTURES OF OPPOSITE CHIRALITY **31**

Abstract **32**

Introduction **33**

Methods **35**
Quartz Calibration 35
Sample Preparation 36
Phase-Resolved Chiral Sum Frequency Generation Experiments 36
Molecular Dynamics 37

Results **38**
Homodyne Chiral Sum Frequency Generation of (L-) and (D-) $LK_7\beta$ 38
Heterodyne Phase-Resolved Chiral Sum Frequency Generation of (L-) and (D-) $LK_7\beta$ Prepared in H_2O, $H_2{}^{18}O$, D_2O, and $D_2{}^{18}O$ 40
Molecular Dynamics and Chiral Sum Frequency Generation Simulation of $LK_7\beta$ Hydration Structure 41

Discussion **44**

Conclusions **49**

APPENDIX TO CHAPTER 2 ..... **62**

Supplementary Computational Methods **62**
System preparation 62
Equilibration and Production Molecular Dynamics 62
Trajectory Analysis – Hydrogen Bonding 63
Trajectory Analysis –Water Orientation 63
Average Infrared Response Frequency 64
Chiral Sum Frequency Generation Response 66
Chiral (psp) Signal from an MD Trajectory 67
Assumptions of the Computational Model 69

Theoretical Background of Phase-Resolved Chiral Sum Frequency Generation **70**

References **83**

CHAPTER 3: SIMULATION OF THE CHIRAL SFG RESPONSE OF SUPRAMOLECULAR STRUCTURES REQUIRES VIBRATIONAL COUPLINGS **85**

Abstract **86**

Introduction **87**

Symmetry-based chiral SFG theory **90**

Computational Methods **94**
MD simulations 94
SFG calculations of the aqueous protein system 95
Extracting chiral SFG signal from MD trajectories 97
SFG calculation of model systems 98

Results **99**

Discussion **102**

Conclusion **108**

Summary of Appendix to Chapter 3 ..... **110**

Acknowledgments and Funding Sources **110**

APPENDIX TO CHAPTER 3 ..... **121**

Combining components of $\chi^{(2)}$ to form the effective *psp* SFG signal **121**

Calculating $\chi^{(2)}_{psp}$ from a molecular dynamics trajectory **122**

Molecular dynamics (MD) equilibration details **123**

Trajectory wrapping details **124**

Impact of bond polarizability on the chiral SFG response of water **124**

Achiral (ssp) SFG response of water in terms of hyperpolarizability elements **125**

Connection between point group, chiral SFG, and vibrational coupling **126**

CHAPTER 4: DETECTING THE FIRST HYDRATION SHELL STRUCTURE AROUND BIOMOLECULES AT INTERFACES **135**

Abstract **136**

Introduction **137**

Results and Discussion **139**
Comparison of experimental and computational spectra 139
Chiral SFG signal of water is primarily generated from the first hydration shell 142
Chiral SFG response of water reveals local interactions in the first hydration shell 143
Chiral SFG intensity and water mobility are correlated for water H-bonded to protein 147
Chiral SFG theory can explain the selectivity of chiral SFG to the first hydration shell 148

Conclusions **150**

Methods **153**
Sample Preparation 153
Phase-resolved Vibrational Chiral SFG 153
MD simulations 154
Calculation of SFG spectra of subsets of water molecule 155
Using Voronoi tessellation to identify water molecule neighbors 156
Calculation of local water dipoles on a grid 157
Analysis of retention time 157

Summary of Appendix to Chapter 4 ..... **157**

APPENDIX TO CHAPTER 4 ..... **169**

Description of the SFG spectrometer **169**

Description of Fitting Experimental Spectra **169**

MD equilibration details **170**

Selection methods for water subsets **171**
Figure 4.4a (entire first hydration shell) 171
Figure 4.4b (backbone) 171
Figure 4.4c (sidechains) 172
Figure 4.4d (backbone – not hydrogen bonded to protein) 172
Figure 4.4e: (backbone – hydrogen bonded to C=O) 173
Figure 4.4f: (backbone – hydrogen bonded to NH) 173
Figure 4.4g: (backbone, hydrogen bonded to C=O with very short hydrogen bonds) 173
Figure 4.4h (sidechains – not hydrogen bonded to protein) 174
Figure 4.4i (sidechains – hydrogen bonded to lysine NH_3^+) 174

Construction of the Hamiltonian for calculation of the SFG spectrum **174**

CHAPTER 5: DESIGN OF AN ELECTROSTATIC FREQUENCY MAP FOR THE NH STRETCH OF THE PROTEIN BACKBONE AND APPLICATION TO CHIRAL SUM FREQUENCY GENERATION SPECTROSCOPY **183**

Abstract **184**

Introduction **185**

Computational methods **189**
Designing the NH stretch electrostatic map 189
Estimating the relative intensities of the NH and OH vibrational response. 194
Intermediate configurations 195
Calculating chiral SFG spectra using MD trajectories with electrostatic maps 196

Experimental methods **197**

Results and discussion **198**
Model protein systems 198
Infrared (linear) response 199
Comparing computational and experimental results 200
Theory matches experiments at the air-water interface 203
The role of vibrational couplings 205
Treatment of hydrogen bonds of intermediate length 207
The NH group as a useful vibrational probe 207

Conclusion **208**

Correction to Chapter 5 ..... **210**

Summary of Appendix to Chapter 5 ..... **210**

Acknowledgements **210**

APPENDIX TO CHAPTER 5 ..... **225**

Methods details **225**

MD simulation 225
Fourier smoothing of spectra 226
Applying the NH map to (D-) enantiomeric proteins 227
Intermolecular couplings 227
Vibrational lifetime 228
Metadynamics 228
Details of chiral SFG spectrum calculation 229
Description of the random forest regression model 230
Fitting x_{01} (see equation 1) vs frequency (linear regression) 230
Experimental procedure of Fourier-transform infrared (FTIR) spectroscopy 231

SFG experimental details 231
Broadband SFG Spectrometer 231
Phase Resolved Chiral SFG Measurements 232
Air-water interface measurements 233

CHAPTER 6: CHARACTERIZING INTERFACES BY VORONOI TESSELLATION 248

Abstract 249

Summary of Appendix to Chapter 6 259

Acknowledgements 259

APPENDIX TO CHAPTER 6 269

Computational methods 269
Molecular dynamics (MD) simulation of protein 269
MD simulation of the air-water interface 270
Voronoi tessellation and neighbor identification 271
Layer-by-layer breakdown of the air-water interface 271
Hydrogen bonding analysis 271
Calculation of SFG spectra 272
Voronoi tessellation code 272

List of Figures

Figure 1.1: Vibrational SFG experimental setup

Figure 1.2: Contributions of different areas of the hydration shell to the chiral signal

Figure 2.1: Homodyne and phase-resolved chiral SFG spectra of chiral water superstructures around LK7β antiparallel β-sheets

Figure 2.2: The phase-resolved chiral SFG spectra of chiral water superstructures around (L-) LK7β

Figure 2.3: The phase-resolved chiral SFG spectra of chiral water superstructures around (D-) LK7β

Figure 2.4: Hydrogen-bond and water structure analyses of the (L-) $LK_7\beta$ homopentamer at the vacuum–water interface obtained from 1 μs of MD sampling using the AMBER ff14SB force field for the protein and the TIP4P-Ew water potential

Figure 2.5: The simulated chiral phase-resolved SFG response from the O-H stretch of water surrounding LK7β

Figure A2.1: Quartz frame relative to laboratory frame

Figure A2.2: Residual analyses of fitting phase-resolved chiral SFG spectra with three peaks

Figure A2.3: High resolution phase-resolved chiral SFG in O-H/N-H spectral region of (L-) $LK_7\beta$ and (D-) $LK_7\beta$

Figure A2.4: Analysis of molecular dynamics of (L-) $LK_7\beta$

Figure A2.5: Orientation of vectors parallel to backbone C=O bonds of residues in (L-) $LK_7\beta$ homopentamer backbone

Figure A2.6: Average *z*-components of water dipole moment vectors for a box of pure water

Figure A2.7: Average z-components of water dipole moment vectors around the (L-) $LK_7\beta$ homopentamer obtained from 1 μs of MD sampling using AMBER ff14SB for the protein and the TIP4P-Ew water potential (slices along *y*-dimension)

Figure A2.8: Average z-components of water dipole moment vectors around the (L-) $LK_7\beta$ homopentamer obtained from 1 μs of MD sampling using AMBER ff14SB for the protein and the TIP4P-Ew water potential (slices along *z*-dimension)

Figure A2.9: Infrared response of the O-H stretching mode of water within 3 Å of the backbone of strands 2, 3, and 4

Figure A2.10: Average z-components of water dipole moment vectors around the (L-) $LK_7\beta$ homopentamer obtained from 1 μs of MD sampling using AMBER ff14SB for the protein and the TIP4P-Ew water potential (four slices on each side of the protein)

Figure 3.1: The $LK_7\beta$ (LKLKLKL) peptide model system

Figure 3.2: Schematic depiction of a vibrational SFG experiment with the *psp* polarization setting

Figure 3.3: Imaginary and real components of the *psp* SFG response of the O-H stretch of water around a pentamer of (L-) $LK_7\beta$ (red) and its (D-) enantiomer (blue) at the vacuum-water interface

Figure 3.4: The role of vibrational couplings in achiral and chiral SFG

Figure 3.5: Effect of couplings on model dimer systems

Figure 3.6: Schematic showing how including vibrational couplings effectively changes the geometry (and symmetry) of a molecular system

Figure A3.1: Imaginary component of the *psp* SFG response for the water O-H stretch around a pentamer of (L-) LK7β at the vacuum-water interface computed using the time-averaging approximation

Figure A3.2: Imaginary component of the *psp* SFG response for a chiral water dimer configuration that could arise near a biomacromolecule. Each curve represents the SFG response calculated with a different $\alpha'_{\parallel} / \alpha'_{\perp}$ ratio

Figure A3.3: Imaginary component of the *psp* response for a chiral water dimer model system with different resonance frequencies

Figure A3.4: Convergence of the imaginary component of the *psp* chiral (black) and *ssp* achiral (red) responses

Figure A3.5: Chiral SFG responses of two independent trajectories for a pure-water interface, showing that the observed signal is noise that most likely originates from imperfect sampling

Figure 4.1: Schematic of $LK_7\beta$ and its antiparallel β-sheet structure

Figure 4.2: Experimental and computational chiral SFG spectra containing spectral contributions due to water

Figure 4.3: The OH stretch's chiral SFG response originates from the first hydration shell

Figure 4.4: Chiral SFG lineshapes of different subsets of water molecules around (L-) $LK_7\beta$

Figure 4.5: Relationship between water mobility and chiral SFG signal generation

Figure 4.6: Symmetrical and asymmetrical orientation of water molecular dipole moments outside and inside the first hydration shell of the $LK_7\beta$ protein, respectively, at the vacuum-water interface. Water dipoles bisect the H-O-H angle and point toward the water oxygen

Figure A4.1: The two spectral windows collected for each spectrum presented in Figure 4.2

Figure A4.2. Fitting and residual analyses of the fitting to the experimental spectra of $LK_7\beta$-H_2O and $LK_7\beta$-$H_2{}^{18}O$

Figure A4.3: Chiral SFG spectrum of clean quartz substrate

Figure A4.4: 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

Figure A4.5: Water dipole moment vector directions at various depths (in Å) from the vacuum-water interface

Figure A4.6: Unnormalized relative magnitudes of the dipole moment vectors at various locations around $LK_7\beta$.

Figure 5.1: A folded tetraalanine peptide mimics secondary structures from the point of view of an NH group

Figure 5.2: Schematic showing the design and use of the NH electrostatic map

Figure 5.3: Schematic showing the structure of $LK_7\beta$

Figure 5.4 (corrected): the NH response contributes little to the IR spectrum of aqueous protein

Figure 5.5 (corrected): Comparison of calculated chiral SFG spectra with heterodyne experimental results for $LK_7\beta$ and $LE_7\beta$

Figure 5.6 (corrected): Comparison of calculated homodyne spectra and experimental homodyne results for $LK_7\beta$ and $LE_7\beta$

Figure 5.7 (corrected): Inner strands of $LK_7\beta$ have lower frequencies

Figure 5.8 (corrected): The effect of vibrational coupling between NH and OH groups on the calculated chiral SFG spectra and comparison with experimental spectra.

Figure A5.1: Points of electric field sampling for the NH map

Figure A5.2: Comparison of the chiral SFG spectrum for $LK_7\beta$ using two different definitions of an H-bond

Figure A5.3: Influence of the exponential decay time for the Fourier smoothing on the chiral SFG spectrum for $LK_7\beta$.

Figure A5.4: The effect of transition dipole location on the NH signal

Figure A5.5: IR spectra of the NH stretch in $LK_7\beta$ using a whole-molecule or atom-by-atom distance cutoff for calculating the electric fields

Figure A5.6: The effect of including Fermi resonance of the OH stretch with the water bend according to the method of Kananenka and Skinner

Figure A5.7: Fourier-transform infrared (FTIR) spectra of $LK_7\beta$ and $LE_7\beta$ films on glass in the amide I region

Figure A5.8: Spectra corresponding to the two spectral regions used to record the chiral heterodyne SFG spectra of $LE_7\beta$ presented in Figure 5.5

Figure A5.9: Spectrum of quartz alone

Figure A5.10: Homodyne spectra of $LE_7\beta$ stitched together

Figure A5.11: Homodyne spectra of $LK_7\beta$ stitched together

Figure A5.12: Chiral SFG spectra of $LK_7\beta$ at the air-water interface

Figure 6.1: Illustration of Voronoi tessellation and neighbor analysis

Figure 6.2: Analysis of air-water interface separated into layers

Figure 6.3: Illustration of $LK_7\beta$ and $LE_7\beta$ as modeled computationally

Figure 6.4: Voronoi-tessellation-based dissection of the hydration shell around $LK_7\beta$ and $LE_7\beta$

Figure 6.5: First-hydration-shell approximation in NH/OH SFG calculations

List of Tables

Table 1.1: Character table for the C_2v point group.

Table 2.1: Fitting parameters and assignments for the spectra in Figure 2.2

Table 2.2: Fitting parameters and assignments for the spectra in Figure 2.3

Table A2.1: Fitting parameters and assignments for the spectra in Figure A2.3

Table A4.1: The fitting parameters (see equation 1, above) used for the 11-peak fitting shown in Figure A4.2.

List of Abbreviations

DFT: density functional theory

IR: infrared

MD: molecular dynamics

NMR: nuclear magnetic resonance

SFG: sum frequency generation

Acknowledgements

I would like to first thank my computational collaborators and friends Dr. Clorice Reinhardt, Dr. Pablo Videla, and Matthew Tremblay for teaching me what you know and listening to my crazy ideas, and my experimental collaborators and friends Dr. Ethan A. Perets, Kristian Olesen, Dr. Ty Santiago, Steve Wang, Hannah Castillo, Erica Bocanegra, and Raibat Sarker for giving me the scientific motivation to do my computational work and for countless fruitful conversations both about science and about everything outside.

I would also like to thank my advisors Prof. Sharon Hammes-Schiffer and Prof. Elsa C.Y. Yan and for encouraging me to work on computational prediction of chiral SFG even at the time when it seemed a poorly defined project with no results and a really stubborn bug, teaching me how to write compelling papers through thick and thin, and helping to give me confidence in my work and in myself as a scientist. I also thank Prof. Garth Simpson and Dr. Alexander Soudackov for their key insights.

Finally, I thank my family, my parents Katya and Vadim, my sister Anya, my grandparents Sergey and Ella, my aunt and uncle Zhenya and Vova, and Betty and Levin for being behind me one hundred percent all the time and reminding me that I am a competent scientist and person even when it did not feel that way and when things were not going so well, as happens during a PhD.

Dedication

for my grandfather, Sergey,
who sparked my interest in science

and

for Clorice and Ethan,
better friends and colleagues than I could possibly have wished for

Chapter 1: Introduction

The biomolecular hydration shell

All biological processes take place in water, and yet water has never been a central player in structural biology. One can see why. X-ray crystallography may be able to resolve very slow-moving ("structural") water molecules, but beyond that the biomolecular hydration shell is too dynamic to be considered part of the structure of a macromolecule. This is very unfortunate because water is often a critical part of binding complexes[1-3] and is the main driver of protein folding via the hydrophobic effect. So dominant are the interactions of water with proteins that protein structure is frequently stabilized by a free energy equivalent to a few hydrogen bonds. Water can be used as an "entropic currency" by biomolecules binding various ligands.[4] Releasing water molecules increases the system's entropy, which helps push the free energy of binding down to a more negative value. When "structural" water molecules are identified in complexes such as antibody-antigen pairs, one wonders what other water molecules may be involved in the complex's formation and endurance. Very few experimental techniques are capable of seeing these more dynamic water molecules. Advanced nuclear magnetic resonance (NMR) techniques may able to measure the dynamics of this water,[5-7] but are generally unable to provide a structural picture of both the protein and its hydration shell on the timescale necessary to resolve the full heterogeneity of the hydration shell.

Vibrational spectroscopy

Vibrational spectroscopy, which reports directly on the motion of molecules on a picosecond timescale, seems like a reasonable approach for studying the hydration shell, but conventional vibrational techniques such as infrared (IR) and Raman spectroscopy are unable to resolve macromolecule-associated water molecules from bulk water. Bulk water overwhelms the signal. One may try to glean information about hydration water from the protein's vibrational signal, but these signals are often hidden behind the huge water response (the OH stretch overwhelms the NH stretch of the protein and the water bend overwhelms the amide I mode signal from the protein). Isotopic labeling (for example, use of D_2O for amide I experiments) can mitigate this problem in some cases (in the case of NH isotopic labeling with heavy hydrogen is useless due to proton exchange), but the resulting measurement is still plagued by a lack of selectivity for hydration shell water. An ideal technique for looking at hydration shell water would selectively probe water near a biomolecule *regardless of its dynamics, and only due to its structure*. This structure need not be fixed in place. The individual water molecules making up the structure can exchange with bulk solvent. This is analogous to the Greek myth of Theseus's ship, where although every part of the ship is exchanged one by one, the ship itself remains due to the overarching structure being preserved. Such a technique should be vibrational in nature to be able to comment on the local environment of different water molecules, which affects the frequency of vibrational modes such as the OH stretch and bend.

The second-order spectroscopic response

To study any given object, one must isolate it. Isolation depends on the identification of some property that is unique to the object being studied. What is unique to the hydration shell of a biomolecule that is not shared by bulk water? One answer is its utter lack of symmetry, or specifically its *chirality*. The water molecules near a biomolecule are imprinted with the shape of the biomolecule, and thus form chiral superstructures that span the hydration shell. Many of these water molecules interact to varying degrees with the biomolecule and assume a particular chiral shape. As mentioned before, although the individual water molecules may exchange quite rapidly, the structure is preserved. Second-order nonlinear vibrational responses, including sum frequency generation (SFG),[8-11] second harmonic generation (SHG),[8, 12-15] and difference frequency generation (DFG), carry information about the symmetry of a molecular system. The reason for this is hidden in the mathematical object that describes a system's nonlinear response, the second-order response tensor ($\chi^{(2)}$). This is a 3 by 3 by 3 tensor (in general n^{th} order spectroscopy is described by a n+1-order tensor; the linear response tensor is of order 2). Because $\chi^{(2)}$ is a third-order tensor, it can capture the three-dimensional symmetry of a molecular system. Specifically, components of $\chi^{(2)}$ that have three different indices, such as $\chi^{(2)}_{zyx}$, are only nonzero for systems that are chiral. These components are also called "orthogonal". Thus SFG spectroscopy can be sensitive to chirality, including the induced chirality of the hydration shell (see below).[16, 17]

As a second-order process, SFG occurs when two photons (two electric fields) impinge on a sample and induce a nonlinear response photon (electric field) with the sum

frequency (difference frequency photons are also released, but are not measured in SFG spectroscopy). The second-order response is the third term in the following expansion

$$P = P_0 + \chi^{(1)} E_1 + \chi^{(2)} : E_1 E_2 + \ldots \quad (1)$$

where P is the system's polarization and P_0 is the polarization in the absence of perturbation. The response tensor directs the response in the following way,[18]

$$E_i \propto \sum_{jk} \chi^{(2)}_{ijk} E_j E_k \quad (2)$$

where E_i is the output field and E_j and E_k are the input fields. Equation 2 implies that different components of $\chi^{(2)}$ can be isolated by polarizing the input electric fields and the output field (detector). For example, polarizing the inputs and output in the *y* direction isolates $\chi^{(2)}_{yyy}$. Usually, polarization in the *y*-direction is taken to be parallel to the laser table (this is called "*s*" polarization), while polarization in *x* and *z* is achieved by polarization perpendicular to the laser table (this is called "*p*" polarization). Polarization of the detector in *p*, one of the beams in *s*, and the other beam in *p*, isolates the tensor components $\chi^{(2)}_{zyx}$, $\chi^{(2)}_{xyz}$, $\chi^{(2)}_{zyz}$, and $\chi^{(2)}_{xyx}$. Two of these elements are chiral and two are achiral. It so happens that only $\chi^{(2)}_{zyx}$ is nonzero for a uniaxial system far from electronic resonance (see below).[18] Therefore the *psp* polarization setting (detector *p*, beam 1 *s*, beam 2 *p*) is sensitive to chirality.

Surface selectivity

In addition to offering sensitivity to chirality, all second-order spectroscopy is inherently surface sensitive. This can be easily shown as follows.[19] Suppose that a system

is *not* interfacial and thus has inversion symmetry. In this case, $\chi^{(2)}$ will not change sign if the system is inverted, but the electric fields and resulting nonlinear polarization ($P^{(2)}$) will, so one obtains

$$\begin{aligned} P^{(2)} &= \chi^{(2)} E_{in,1} E_{in,2} \\ -P^{(2)} &= \chi^{(2)}(-E_{in,1})(-E_{in,2}) = \chi^{(2)} E_{in,1} E_{in,2} \\ -P^{(2)} &= P^{(2)} \\ P^{(2)} &= 0 \end{aligned} \tag{3}$$

The same is true for any even-ordered spectroscopy for obvious reasons. SFG's surface and chirality sensitivity makes it an ideal technique for probing chirality at interfaces.

Like chirality, interfaces are ubiquitous in biology. Processes like the mitochondrial electron transport chain, parts of photosynthesis, and even T-cell signaling[20] involve interfaces and phase separation. SFG is thus a perfect approach to probe the structure of biological interfaces, including associated hydration water.

Vibrational SFG is a technique where one of the input beams is an IR pulse and the other is a visible pulse far from electronic resonance. The sum frequency response carries complementary information to IR and Raman spectroscopy, but with interface (and possibly chiral) selectivity. As will be shown below, vibrational SFG can be used to probe the local environments of molecules in chiral water superstructures around biomolecules selectively. In vibrational SFG, the polarization notation is detector, visible, IR, so in a *psp* experiment, the IR beam has *p* (*x*, *z*) polarization. Most SFG experiments performed since the 1980s, when the technique was introduced by the Shen group,[10] use *ssp* polarization, which isolates $\chi^{(2)}_{yyz}$ and $\chi^{(2)}_{yyx}$ (only $\chi^{(2)}_{yyz}$ survives in a uniaxial system), and study achiral systems such as the air-water interface. This polarization is used because it is *not* chirality-selective but carries information about the orientation of relevant dipoles.[9] For example, in

an air-water interface spectrum in the OH stretch region (3000-4000 cm^{-1}), the sign of the imaginary component of $\chi^{(2)}_{yyz}$ reports on the *z*-direction of OH dipoles.[21-24]

Heterodyne SFG experiments

Isolating the imaginary, absorptive, component of an SFG signal requires *heterodyne detection.*[9, 24, 25] There are several ways to accomplish this. The Yan group uses the *internal heterodyne* method,[26] where the experiment is done at the air-quartz interface and the mixing of the non-resonant response of the quartz with the resonant response of the sample allows for extraction of the imaginary component of the response. Intuitively, the *imaginary* component represents absorption because absorption corresponds to the production of electromagnetic fields that have a 90° phase shift from the incoming fields, and thus lead to attenuation of the incoming field. This 90° phase shift is expressed by a factor of *i*. More information can be found in later chapters. "Homodyne" SFG experiments that measure $\left|\chi^{(2)}\right|^2$ also contain useful information and are as selective as heterodyne experiments, though without the enhanced peak resolution and orientational information provided by the imaginary response. Note that homodyne spectra can be estimated from heterodyne spectra by calculating the real component of the response from the imaginary component using the Kramers-Kronig relation,[27, 28]

$$\mathrm{Re}[\chi^{(2)}(\omega)] = \frac{2}{\pi} P\int_0^\infty \frac{\omega' \,\mathrm{Im}[\chi^{(2)}(\omega')]}{\omega'^2 - \omega^2} d\omega' \tag{4}$$

where *P* indicates the principal value of the integral (to deal with the singularity), and then using

$$|\chi^{(2)}|^2 = \mathrm{Re}[\chi^{(2)}]^2 + \mathrm{Im}[\chi^{(2)}]^2 \tag{5}$$

All theory and experiments discussed in this thesis also rely on the electric dipole approximation, which is the approximation where magnetic dipole and higher-order electric field effects (quadrupole, etc.) are neglected. This results in a great simplification of the nonlinear response functions, and is an excellent approximation in most biological contexts. Note that SFG is only fully surface-selective under the electric dipole approximation.

Figure 1.1 shows a typical experimental setup for vibrational SFG. Experimental details can be found in Chapters 2, 3, and 4.

General theory of the second-order response

The second-order response can be derived from time-dependent perturbation theory. Generally, the term "perturbation theory" refers to a class of methods where, assuming a "small" perturbation, the n+1th-order correction to the system's response can be derived directly from the nth-order correction. Thus, the first-order correction is derived from the unperturbed system (using the perturbed Hamiltonian). The following discussion closely follows the treatments of Mukamel and Tuckerman.[19, 29]

The time-evolution of a system is described by the time-dependent Schrödinger equation

$$\frac{\partial \psi(t)}{\partial t} = -\frac{i}{\hbar}\hat{H}\psi(t) \tag{6}$$

where $\psi(t)$ is the system wave function and $\hat{H}$ is the system Hamiltonian. The solution to the Schrödinger equation is described by the quantum *propagator* $U(t,t_0)$, as in

$$\psi(t) = U(t,t_0)\psi(t_0) \tag{7}$$

To find an approximate expression for the propagator, we start with a modified Schrödinger equation

$$\frac{\partial U(t,t_0)}{\partial t} = -\frac{i}{\hbar}\hat{H}U(t,t_0) \tag{8}$$

which has the formal solution

$$U(t,t_0) = \hat{I} - \frac{i}{\hbar}\int_{t_0}^{t} d\tau \hat{H}(\tau)U(\tau,t_0) \tag{9}$$

This solution obviously is not sufficient because the propagator is on both sides. In perturbation theory, the Hamiltonian is divided into an unperturbed component H_0 and a small disturbance $H'(t)$. The laws of quantum mechanics can be expressed in different *pictures*. In the Schrödinger picture, the wave function moves while the operators do not. In the Heisenberg picture, the operators move and the wave function does not. In the *interaction* picture, the movement of the wave function due to the unperturbed Hamiltonian is transferred to the operators but movement due to the perturbation remains with the wave function. This leads to the following expansion for the propagator due to the perturbation, which comes from plugging in the propagator repeatedly and recursively into equation 8

$$\begin{aligned} U(t,t_0) = U_0(t,t_0) + \sum_{n=1}^{\infty}\left(-\frac{i}{\hbar}\right)^n \int_{t_0}^{t} d\tau_n \int_{t_0}^{\tau_n} d\tau_{n-1} \ldots \int_{t_0}^{\tau_2} d\tau_1 \\ U_0(t_0,\tau_n)H'(\tau_n)U_0(\tau_n,\tau_{n-1})H'(\tau_{n-1})\ldots U_0(\tau_2,\tau_1)H'(\tau_1)U_0(\tau_1,t_0) \end{aligned} \tag{10}$$

The derivation of this expansion from the interaction picture can be found in Mukamel and Tuckerman's books. This equation may look complicated, but it has a very simple interpretation. Moving from the right, the operators describe

1. Free evolution from t_0 to τ_1
2. Interaction with the perturbation at τ_1

3. Free evolution from τ_1 to τ_2

4. Etc.

Note that because the operators do not commute with each other, the order of the operators matters *and* the order of the time arguments matters. In other words, $\tau_2 > \tau_1 > t_0$ must always be enforced when determining the order of operators. This expansion is called a "time-ordered exponential" because it can be seen as a Taylor expansion of a kind of matrix exponential. The Fourier transform of an exponential is a Lorentzian, and is important when formulating the frequency-domain version of the propagator's effect (see Mukamel Chapters 2 and 3). To enforce causality, a "Green function" is defined, which I call a "causal propagator" in this context. It is given in the time domain by

$$G(t-t_0) = \theta(t-t_0)U(t,t_0) \tag{11}$$

Where $\theta(t)$ is a Heavyside function that prevents propagation backwards in time. The Green function is given in the frequency (energy) domain by

$$G(E) = \frac{1}{E - H + i\Gamma} \tag{12}$$

where Γ occurs here just to avoid the singularity, but in more detailed treatments can represent dephasing and relaxation of the system (see Mukamel Chapter 6).

To fully understand the nonlinear response, we need to define the system's state in a more detailed or descriptive way using the *density matrix*, which is defined for a pure state by the outer product of the system wave function

$$\hat{\rho} = |\psi\rangle\langle\psi| \tag{13}$$

The average value of an operator's corresponding observable is given by

$$\langle A \rangle = Tr[\hat{\rho}\hat{A}] = Tr[\hat{A}\hat{\rho}] \tag{14}$$

The Liouville-von Neumann equation describes the density matrix's time evolution:

$$\frac{\partial \rho}{\partial t} = -\frac{i}{\hbar}[\hat{H}, \hat{\rho}] \tag{15}$$

This equation of motion can be derived from the Schrödinger equation by combining the original Schrödinger equation with its complex conjugate. An analogous expansion to equation 9 using the Liouville-von Neumann equation produces a bewildering set of nested commutators that are difficult to write down and reason about directly. To avoid this difficulty, Mukamel developed the concept of *Liouville space*, which defined as the space similar to Hilbert space where the density matrix and other operators are vectors rather than matrices. This allows one to write down the Liouville-von Neumann equation as a direct analogue to the Schrödinger equation:

$$\frac{\partial \rho}{\partial t} = -\frac{i}{\hbar}\mathfrak{L}\rho \tag{16}$$

where $\mathfrak{L}$ is the Liouvillian superoperator that represents the commutator in equation 15 applied to the density *vector* in Liouville space. This equation looks the same as the Schrödinger equation, which is useful. In particular, it allows the definition of a Liouville space propagator superoperator $\mathfrak{A}$, which represents the following operator combination in Hilbert space ($[\bullet]$ represents the expression being acted upon)

$$\mathfrak{A}(t, t_0) = U(t, t_0)[\bullet]U^{\dagger}(t, t_0) \tag{17}$$

Just like this, one can define superoperators in Liouville space corresponding to the dipole operator and Green function (i.e. causality-aware propagator) (V → $\mathfrak{V}$, G → $\mathfrak{G}$, etc.). Replacing the absolute times in equation 9 with time *intervals*, and noting that under the electric dipole approximation

$$\mathfrak{L}'(t) = -E(t)\mathfrak{V} \tag{18}$$

where E(t) is the time-varying electric field and $\mathfrak{V}$ is the Liouville space dipole superoperator, one can write the system's polarization to n[th] order as

$$P^{(n)}(t) = \int_0^\infty dt_n \int_0^\infty dt_{n-1} \ldots \int_0^\infty dt_1 S^{(n)}(t_n, t_{n-1}, \ldots, t_1) E(t - t_n) E(t - t_n - t_{n-1}) \ldots E(t - t_n \ldots - t_1) \tag{19}$$

where $S^{(n)}$ is the "response function" of the system

$$S^{(n)}(t_n, t_{n-1}, \ldots, t_1) = \left(\frac{i}{\hbar}\right)^n \langle\langle V \,|\, \mathfrak{G}(t_n)\mathfrak{V}\mathfrak{G}(t_{n-1}) \ldots \mathfrak{G}(t_1)\mathfrak{V}\rho(-\infty)\rangle\rangle \tag{20}$$

Note that this expansion similar to the Hilbert space wave function-based expansion in equation 9. The double brackets represent the trace, which becomes an inner product in Liouville space. The unpleasant nested commutators implied by equation 19 do not have to be written at this point. Of course, in actual calculations, Hilbert space must be reintroduced. For example, one way of writing the second-order response using equation 19 in Hilbert space is

$$S^{(2)}(t_2, t_1) = \left(\frac{i}{\hbar}\right)^2 \theta(t_1)\theta(t_2)\langle V(t_2 + t_1) \,|\, [V(t_1)[V(t_0, \rho]]\rangle \tag{21}$$

Here the propagators (Green functions) have been absorbed into the dipole operators' time arguments. This response function contains many terms, one for each combination of operations in the commutators. Each term is called a "Liouville space pathway" and represents a certain sequence of events that happen to the density matrix. Sometimes it gets hit from the left and sometimes from the right, all at different times. Each sequence of events (pathway) can be represented with a double-sided Feynman diagram (see Boyd *Nonlinear Optics*[30] and Mukamel). In second-order spectroscopy, sum frequency generation is one group of pathways. Difference-frequency generation is another. Note that

the leftmost dipole operator is not part of the nested commutator because it is the dipole operator being combined with the density matrix in the expression

$$P^{(n)} = Tr[V\rho^{(n)}] \tag{22}$$

The easiest way to think about what happens the density matrix is to remember that the (transition) dipole operator takes the wave function from eigenstate *a* to eigenstate *b*. The density matrix represents the system in more complete way, and it has two wave function components, the "ket" ($|\psi\rangle$) and the "bra" ($\langle\psi|$). In the nested commutator in equation 21, some events affect the ket, flipping it from *a* to *b*, and some affect the bra, doing the same. This is what is captured in double-sided Feynman diagrams. The presence of a particular ket and bra at a given timepoint is called a *coherence*, because it represents the relationship between two different eigenstates. A second-order Liouville space pathway could be something like *aa*→*ba*→*ca*, where the first letter represents the ket and the second represents the bra (remember that the density operator is a ket and then a bra). In this case, only the ket is acted upon. Another pathway could be *aa*→*ac*→*bc*, where the ket and bra are both acted upon once. In the frequency domain, the second-order response function can be written as

$$\begin{aligned} S^{(2)}(\omega_2,\omega_1) \propto \langle a|\hat{\mu}|b\rangle\langle b|\hat{\mu}|c\rangle\langle c|\hat{\mu}|a\rangle[&I_{ca}(\omega_1+\omega_2)I_{ba}(\omega_1) \\ &-I_{bc}(\omega_1+\omega_2)I_{ba}(\omega_1)+I_{ab}(\omega_1+\omega_2)I_{ac}(\omega_1)-I_{bc}(\omega_1+\omega_2)I_{ac}(\omega_1)] \end{aligned} \tag{23}$$

where

$$I_{ab}(\omega) = \frac{1}{\omega - \omega_{ab} + i\Gamma_{ab}} \tag{24}$$

where ω_{ab} is the frequency of the a→*b* transition and Γ_{ab} is the dephasing/relaxation rate associated with the transition *a*→*b*. Note that the dipole terms are vectors, so the response

function is a third-order tensor, as expected. The full second-order response equations in the frequency domain with tensor components and the resulting permutations can be found in refs 30 and 16. Note also that the reason the frequency domain representation is a set of Lorentzians is because a Lorentzian is the Fourier transform of a (decaying) exponential, and the time-domain propagators are decaying exponentials. The decaying aspect is only accessible in the density matrix formulation, as wave functions alone contain no mechanism to express decoherence due to a random bath. As expected, each term in equation 22 represents a sequence of events happening to the density matrix (ket or bra). The sign of the term is determined by the number of bra interactions (odd means negative). $S^{(2)}$ is directly proportional to the susceptibility $\chi^{(2)}$. The time- and frequency-domain approaches to the nonlinear response are completely equivalent, and deciding which to focus on is purely a matter of convenience. One can always perform a Fourier transform.

Vibrational SFG in the frequency and time domains

Very often, only a few terms implied by equation 22 are relevant because of resonance enhancement. For example, in vibrational SFG spectroscopy, only terms involving ω_2 (the IR frequency) are significant.[16] This makes calculation of the SFG response much simpler. The vibrational SFG response is given by a Liouville space pathway that corresponds to 1. Excitation from the ground state to a vibrationally excited state, and 2. A Raman transition includes virtual states high above the ground state. The polarizability follows from the full expression for β found in ref [21]. Assuming only one ground state *a*, the response becomes[16, 31, 32]

$$\beta_{ijk}(\omega) = \sum_b \frac{\langle a | \hat{\alpha}_{ij} | b \rangle \langle b | \hat{\mu}_k | a \rangle}{\omega - \omega_{ab} + i\Gamma_{ab}} \tag{25}$$

in the frequency domain and

$$\beta_{ijk}(\omega) = i\int_0^\infty dt\, e^{-i\omega t} \sum_b e^{-\Gamma_{ab}t} \langle a | e^{\frac{i}{\hbar}\omega_a t} \hat{\alpha}_{ij} e^{-\frac{i}{\hbar}\omega_b t} | b \rangle \langle b | \hat{\mu}_k | a \rangle \tag{26}$$

in the time domain. Because

$$\hat{H} | a \rangle = \hbar\omega_a | a \rangle \tag{27}$$

equation 26 can be rewritten as

$$\beta_{ijk}(\omega) = i\int_0^\infty dt\, e^{-i\omega t} \sum_b e^{-\Gamma_{ab}t} \langle a | e^{\frac{i}{\hbar}\hat{H}t} \hat{\alpha}_{ij} e^{-\frac{i}{\hbar}\hat{H}t} | b \rangle \langle b | \hat{\mu}_k | a \rangle \tag{28}$$

At this point two distinct (semi)classical approximation can be made. The first one assumes that the dipole and polarizability operators can be replaced by their classical equivalents, yielding

$$\beta_{ijk} = i\int_0^\infty dt\, e^{-i\omega t} \langle \alpha_{ij}(t) \mu_k(0) \rangle \tag{29}$$

This is the Morita-Hynes equation for the vibrational SFG response.[21] It is the Fourier transform of the dipole-polarizability time correlation function. An alternative approximation, made by Skinner[33, 34] and independently by Jansen,[35] is to replace the matrix operators $\hat{\alpha}_{ij}$ and $\hat{\mu}_k$ with vectors with one entry for each vibrational chromophore and to absorb the time evolution of $e^{\frac{i}{\hbar}\hat{H}t} \hat{\alpha}_{ij} e^{-\frac{i}{\hbar}\hat{H}t}$ on the right into $\alpha_{ij}(t)$, leaving just one complex exponential yielding, formally,

$$\beta_{ijk}(\omega) = i\int_0^\infty dt\, e^{-i\omega t} e^{-\Gamma t} \sum_{ab} \alpha_{ij}^a(t) e^{-\frac{i}{\hbar}\hat{H}t} \mu_k^b(0) \tag{30}$$

where here the *a* and *b* indices represent vibrational chromophores and the Hamiltonian is in the local mode basis, with vibrational frequencies on the diagonal and vibrational couplings between modes on the off-diagonal. The propagator in equation 31 is solved for by integrating the time-dependent Schrödinger equation for the propagator. This method is more expensive that the Morita-Hynes method[21, 36] due to the explicit integration, but can be more accurate if the Hamiltonian, dipole, and polarizability are obtained from electrostatic density functional theory maps accurately linking electric field to these quantities (see below). By contrast, obtaining accurate transition dipoles and polarizabilities from classical simulations can be difficult, even with polarizable force fields. The highest-quality studies using the Morita-Hynes approach use *ab initio* molecular dynamics trajectories.[36-39] The SFG calculation is rapid but the simulations are very expensive. Thus, it can be advantageous to use electrostatic maps and run cheap classical simulations to obtain copious sampling. Crucially, the propagator in equation 31 provides the oscillation needed for the correlation function to yield the correct spectrum, so the vibrational modes in question can be safely constrained using the SHAKE algorithm,[40] etc. in the classical simulation to increase the integration timestep. In this case the fluctuating dipole and polarizability are merely the envelopes of the "true" values that oscillate rapidly as the bond vibrates. In the Morita-Hynes approach the oscillations must come directly from the dynamics.

The Skinner-Jansen approach scales as N^3 with the number of vibrational chromophores, which can be expensive for large systems. A cheaper approach that also scales as N^3 but requires far fewer N^3 operations is the *inhomogeneous limit approximation*. In this approach, the dynamical aspect of the line shape is neglected and the line shape is

assumed to be the simple average of configurations over a trajectory. In the absence of dynamics, for each configuration the semiclassical propagator is given by[41]

$$U(t) = e^{-i\omega_0 t} \tag{31}$$

where ω_0 is the current Hamiltonian. Inserting this into equation 30 and diagonalizing the Hamiltonian, we obtain

$$\beta_{ijk}(\omega) = i\int_0^\infty dt\, e^{-i\omega t} \sum_a \sum_b e^{-\Gamma_{ab} t} \alpha_{ij}^a(t) [\mathbf{U} e^{-i\lambda t} \mathbf{U}^\dagger]_{ab} \mu_k^b(0) \tag{32}$$

where U is the eigenvector matrix and λ is the diagonal eigenvalue matrix. Expressing the matrix product as a set of sums, assuming a constant Γ, and rearranging and reindexing significantly, we perform the following series of operations

$$\begin{aligned}
\beta_{ijk}(\omega) &= i\int_0^\infty dt\, e^{-i\omega t} e^{-\Gamma t} \sum_a \sum_b \alpha_{ij}^a(t) \mu_k^b(0) \sum_{a'b'} U_{a'a} U_{b'b} [e^{-i\lambda t}]_{a'b'} \\
&= i\int_0^\infty dt\, e^{-i\omega t} e^{-\Gamma t} \sum_a \sum_b \alpha_{ij}^a(t) (\sum_{a'} U_{aa'} U_{ba'} [e^{-i\lambda t}]_{a'a'}) \mu_k^b(0) \\
&= i\int_0^\infty dt\, e^{-i\omega t} e^{-\Gamma t} \sum_{a'} [e^{-i\lambda t}]_{a'} \sum_a \alpha_{ij}^a(t) (U_{aa'}) \sum_b U_{a'b} \mu_k^b(0) \\
&= i\int_0^\infty dt\, e^{-i\omega t} e^{-\Gamma t} \sum_{a'} [e^{-i\lambda t}]_{a'} (\sum_a \alpha_{ij}^a(t) (U_{aa'})) (\sum_b U_{ba'} \mu_k^b(0)) \\
&= i\int_0^\infty dt\, e^{-i\omega t} e^{-\Gamma t} \sum_{a'} [e^{-i\lambda t}]_{a'} (\sum_b \alpha_{ij}^b(t) (U_{ba'})) (\sum_b U_{ba'} \mu_k^b(0)) \\
&= i\int_0^\infty dt\, e^{-i\omega t} e^{-\Gamma t} \sum_a [e^{-i\lambda t}]_a (\sum_b \alpha_{ij}^b(t) (U_{ba})) (\sum_b U_{ba} \mu_k^b(0))
\end{aligned} \tag{33}$$

Performing the Fourier transform, we obtain Skinner's expression for the inhomogeneous limit in Lorentzian form

$$\beta_{ijk}(\omega) = \left\langle \sum_a \frac{\sum_b U_{ba} \alpha_{ij}^b(0) \sum_b U_{ba} \mu_k^b(0)}{\lambda_a - \omega - i\Gamma} \right\rangle \tag{34}$$

Note that the denominator can be expressed in other conventions with different signs. This is the convention used by Skinner. The average is over all configurations in a trajectory. Note also that if equation 34 is used for an entire molecular system, β is replaced with $\chi^{(2)}$.

Chiral SFG of achiral molecules in chiral configurations

Most of the work presented here concerns the chiral SFG response of the hydration shell. The hydration shell is chiral *as an assembly*, but it is not immediately obvious that achiral molecules should produce chiral SFG signals, because achiral molecules lack any nonzero orthogonal β elements (elements like β_{xyz} where all three indices are different). Naively, it seems that these elements alone should give rise to the orthogonal (chiral) elements of $\chi^{(2)}$. The reality is more complicated. $\chi^{(2)}$ arises from β according to[16]

$$\chi^{(2)}_{IJK}(\phi,\theta,\psi) = \sum_{ijk} \langle R_{Ii}(\phi,\theta,\psi) R_{Jj}(\phi,\theta,\psi) R_{Kk}(\phi,\theta,\psi) \rangle \beta_{ijk} \tag{35}$$

where R is an Euler angle matrix describing the transformation of the molecular frame into the laboratory frame. This discussion will assume the *zyz* Euler rotation convention,[18] where the three Euler angles represent rotation about the original *z*-axis (ϕ), tilting of the *z* axis toward a new *z'* axis about the *y*-axis (θ), and then subsequent rotation about the new *z'* axis (ψ). There are many other, equivalent Euler angle conventions that allow rotation to any conformation in three-dimensional space that differ slightly in the role of the angles. In uniaxial systems (i.e. systems with C_∞ symmetry), $\chi^{(2)}$ is given by

$$\chi^{(2)}_{IJK}(\theta,\psi) = \frac{1}{2\pi}\int_0^{2\pi} d\phi \sum_{ijk} \langle R_{Ii}(\phi,\theta,\psi) R_{Jj}(\phi,\theta,\psi) R_{Kk}(\phi,\theta,\psi)\rangle \beta_{ijk} \tag{36}$$

where the integral is over all initial rotations about the *z*-axis. All the systems mentioned in this work have uniaxial symmetry. Crystals, for example, do not, and equation 36 does not apply to them.

To determine which elements of β survive in a given vibrational normal mode of a given molecule, we turn to group theory, the theory of (molecular) symmetry. Each molecular point group contains a set of symmetry operations that leave the molecule unchanged. The point group contains a number of *irreducible representations*, which for the purposes of this discussion correspond to the possible normal modes. Note that not every irreducible representation necessarily has a corresponding normal mode. The symmetry operations and irreducible representations are summarized in a *character table*, where the figures ("characters") represent the effect of the symmetry operation on each irreducible representation. A character table for the point group C_2v is found in Table 1.1. In this table, a "1" represents no change upon the given transformation, while a -1 indicates inversion under that transformation. The character table can be used directly to calculate the remaining elements of β using *projection operators*,[42] which determine which components of β are insensitive to the symmetry operations of the point group. These objects "project" the effect of belonging in each irreducible representation onto an operator β. Note that because vibrational SFG essentially relies on an infrared transition followed by a Raman transition, a normal mode must be both IR- and Raman-active to be SFG-active, as indicated by

$$\beta_{ijk} = \frac{\partial \alpha_{ij}}{\partial Q} \frac{\partial \mu_k}{\partial Q} \tag{37}$$

where Q is a normal mode coordinate and the two derivatives represent the transition dipole and transition polarizability. To obtain the surviving elements of β in a given point group and in various irreducible representations within that point group, we first determine the surviving elements of α and μ and then take the tensor product to determine the remaining elements of β, as in

$$\beta = \alpha \otimes \mu \tag{38}$$

The projection operators are given by[42]

$$\begin{aligned} \mu^{\Gamma} &= \frac{f_{\Gamma}}{|G|} \sum_{g \in \mathbf{G}} \chi_{\Gamma}(g)^{*} \mathbf{O_g} \mu \\ \alpha^{\Gamma} &= \frac{f_{\Gamma}}{|G|} \sum_{g \in \mathbf{G}} \chi_{\Gamma}(g)^{*} \mathbf{O_g} \alpha [\mathbf{O_g}]^{-1} \end{aligned} \tag{39}$$

where μ^{Γ} and α^{Γ} are the remaining dipole and polarizability vector and matrix, respectively, G is the point group, |G| is the number of symmetry operations in the point group, $\chi_{\Gamma}(g)$ is the character of the current symmetry operation (g) within the current irreducible representation (Γ), $\mathbf{O_g}$ is the matrix representing the symmetry operation g, and f_{Γ} is the dimensionality of the irreducible representation Γ (often 1, though not always).

The water molecule belongs to the point group C_{2v}, for which the remaining elements of β are β_{zzz}, β_{xxz}, and β_{yyz} (symmetric stretch, irreducible representation A_1) and β_{xzx}, β_{zxx}, β_{yzy}, and β_{zyy} (asymmetric stretch, irreducible representations B_1 or B_2).

Using equation 36 and these β elements, the element $\chi^{(2)}_{ZYX}$, probed by the *psp* polarization setup, is given by[16]

$$\chi^{(2)}_{ZYX}(\theta,\psi) \propto \frac{1}{2}\sin^2(\theta)\sin(2\psi)(-\beta_{xxz} + \beta_{yyz} - \beta_{yzy} + \beta_{xzx}) \tag{40}$$

where *XYZ* refers to the laboratory frame and *xyz* to the molecular frame. Note that none of the included β elements are chiral (orthogonal). Instead, achiral elements add up to a chiral component of the susceptibility. But, $\beta_{xxz} \neq \beta_{yyz}$ and $\beta_{xzx} \neq \beta_{yzy}$ only in a system that has no symmetry. This allows a chiral SFG signal to form from the chiral hydration shell. Molecules with some higher symmetries like C_{3v} cannot form chiral SFG signals even in a chiral configuration because not enough β elements remain. However, this is not completely true in real systems, as asymmetric vibrational couplings can completely break molecular symmetry and allow chiral assemblies of any point group to produce a chiral SFG signal (see Chapter 3).

C_2v	E	C_2 (z)	σ_{xz}	σ_{yz}	modes
A_1	1	1	1	1	symmetric stretch, bend
A_2	1	1	-1	-1	none
B_1	1	-1	1	-1	asymmetric stretch
B_2	1	-1	-1	1	

Table 1.1 Character table for the C_2v point group. E represents the identity operation, C_2 rotation about the principal axis, σ_{xz} reflection across the *xz* plane, and σ_{yz} reflection across the *yz* plane.

Calculating the chiral SFG signal of a uniaxial system

The experimental systems mentioned in this thesis are all uniaxial (they have C_∞ symmetry). However, the MD trajectories do not have this property. To simulate the effect

of this property it is necessary to convert from the original system to the uniaxial system. This can be done by taking the integral in equation 36, effectively rotating the system into every possible configuration. When that is done, it is found that the element $\chi^{(2)}_{ZYX}$ in the rotated system corresponds to $\chi^{(2)}_{Z'Y'X'} - \chi^{(2)}_{Z'X'Y'}$ in the original MD trajectory. For this reason, the chiral SFG (*psp*) signal is calculated as $\chi^{(2)}_{ZYX} - \chi^{(2)}_{ZXY}$ in practice. See ref 43 for more information.

The meaning of the sign of the chiral (*psp*) response and achiral (*ssp*) response

As mentioned earlier, heterodyne detection allows the isolation of the imaginary, absorptive, component of $\chi^{(2)}$. The imaginary component's sign carries information about the orientation of the molecule. In the case of *ssp* polarization, the sign of the response tracks with the *z*-component of the relevant transition dipole. In *psp* polarization, although the sign of the lineshape does change depending on molecular orientation, the absolute sign of any one conformation is very hard to predict without explicitly modeling the response, although it also depends on orientation. This section shows why this is, starting from equation 34, from which it is easiest to see the meaning of the phase. The discussion uses a simplified system of one water molecule with identical OH groups. To simplify matters further, we invoke the bond polarizability approximation,[44, 45] which models the polarizability as belonging to an individual vibrating bond without interplay between bonds. The transition polarizability α is then proportional to[45]

$$\alpha_{ij} \propto (\alpha_{\parallel}' - \alpha_{\perp}')u_i u_j + \alpha_{\perp}'(\hat{i} \cdot \hat{j}) \tag{41}$$

where $\alpha_{\parallel}'$ and $\alpha_{\perp}'$ are the components of the transition polarizability derivative parallel and perpendicular to the bond, respectively, $\hat{i}$ and $\hat{j}$ are unit vectors for the *i* and *j* axes, and **u** is the bond unit vector. In chiral SFG, the relevant components of the polarizability are α_{zy} and α_{zx} (see previous section), so $\hat{i} \cdot \hat{j} = 0$. The chiral polarizabilities are then proportional to

$$\begin{aligned} \alpha_{zy} &\propto (\alpha_{\parallel}' - \alpha_{\perp}')u_z u_y \\ \alpha_{zx} &\propto (\alpha_{\parallel}' - \alpha_{\perp}')u_z u_x \end{aligned} \tag{42}$$

and the achiral polarizability is given by

$$\alpha_{yy} \propto (\alpha_{\parallel}' - \alpha_{\perp}')u_y^2 + \alpha_{\perp}' \tag{43}$$

Putting it all together, and assuming for simplicity that the two OH groups have the same frequency and so the eigenvector matrix is approximately given by

$$\mathbf{U} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \tag{44}$$

where the first column corresponds to the symmetric stretch and the second to the asymmetric stretch. We also assume that the transition dipole is given by

$$\mu_k \propto u_k \tag{45}$$

Finally, for the chiral response, we obtain (see previous section)

$$\begin{aligned} \beta_{chiral} = \beta_{zyx} - \beta_{zxy} \propto & \frac{r[u_z^{(1)}u_y^{(1)} + u_z^{(2)}u_y^{(2)}][u_x^{(1)} + u_x^{(2)}] - r[u_z^{(1)}u_x^{(1)} + u_z^{(2)}u_x^{(2)}][u_y^{(1)} + u_y^{(2)}]}{\lambda_1 - \omega - i\Gamma} \\ + & \frac{r[u_z^{(1)}u_y^{(1)} - u_z^{(2)}u_y^{(2)}][u_x^{(1)} - u_x^{(2)}] - r[u_z^{(1)}u_x^{(1)} - u_z^{(2)}u_x^{(2)}][u_y^{(1)} - u_y^{(2)}]}{\lambda_2 - \omega - i\Gamma} \end{aligned} \tag{46}$$

where *r* represents $\alpha_{\parallel}' - \alpha_{\perp}'$. This expands and simplifies to

$$\beta_{chiral} = \frac{r(u_z^{(1)}u_y^{(1)}u_x^{(2)} + u_z^{(2)}u_y^{(2)}u_x^{(1)} - u_z^{(1)}u_x^{(1)}u_y^{(2)} - u_z^{(2)}u_x^{(2)}u_y^{(1)})}{\lambda_1 - \omega - i\Gamma} + \frac{r(-u_z^{(1)}u_y^{(1)}u_x^{(2)} - u_z^{(2)}u_y^{(2)}u_x^{(1)} + u_z^{(1)}u_x^{(1)}u_y^{(2)} + u_z^{(2)}u_x^{(2)}u_y^{(1)})}{\lambda_2 - \omega - i\Gamma} \tag{47}$$

Note that only cross-terms remain. If there is no coupling between the OH groups, on average $\lambda_1 = \lambda_2$, so there is no signal because the terms all approximately cancel. Equation 47 illustrates why the sign of the chiral response is so hard to predict – it depends on *products* of the components of the bond unit vectors, each of which can be positive or negative. By contrast, the achiral response for single water molecule is given by

$$\begin{aligned}\beta_{achiral} = \beta_{yyz} &= \frac{(\alpha_{\parallel}' - \alpha_{\perp}')(u_y^{2(1)} + u_y^{2(2)} + 2\alpha_{\perp}')(u_z^{(1)} + u_z^{(2)})}{\lambda_1 - \omega - i\Gamma} \\ &- \frac{(\alpha_{\parallel}' - \alpha_{\perp}')(u_y^{2(1)} - u_y^{2(2)})(u_z^{(1)} + u_z^{(2)})}{\lambda_2 - \omega - i\Gamma} \\ &= \frac{(+)(u_z^{(1)} + u_z^{(2)})}{\lambda_1 - \omega - i\Gamma} - \frac{(+)(u_y^{2(1)} - u_y^{2(2)})(u_z^{(1)} + u_z^{(2)})}{\lambda_2 - \omega - i\Gamma}\end{aligned} \tag{48}$$

where $(+)$ represents collective quantities that are always positive and thus do not influence the phase contribution of each term. This expression shows that while the phase of the achiral response is not simply a function of the *z*-component of the dipole as is often assumed (at least when intramolecular couplings are engaged), the first term has a very simple behavior and the second term is approximately zero when the water molecule faces with one OH group into the interface and the other into the water, and it is these water molecules that contribute most to the achiral SFG signal from the air-water interface. The second term explains the minor shift in the response due to inclusion of intramolecular couplings (see Chapter 3). Overall, it can be seen that the chiral SFG phase is more complicated than the achiral phase. For this reason, the chiral phase can be thought of as a

fingerprint of a molecular system in a particular orientation. Figure 1.2 shows the complexity of the chiral phase distribution by breaking down different regions of water around two interfacial protein systems by their contributions to the phase (see future Chapters for a description of these types of systems).

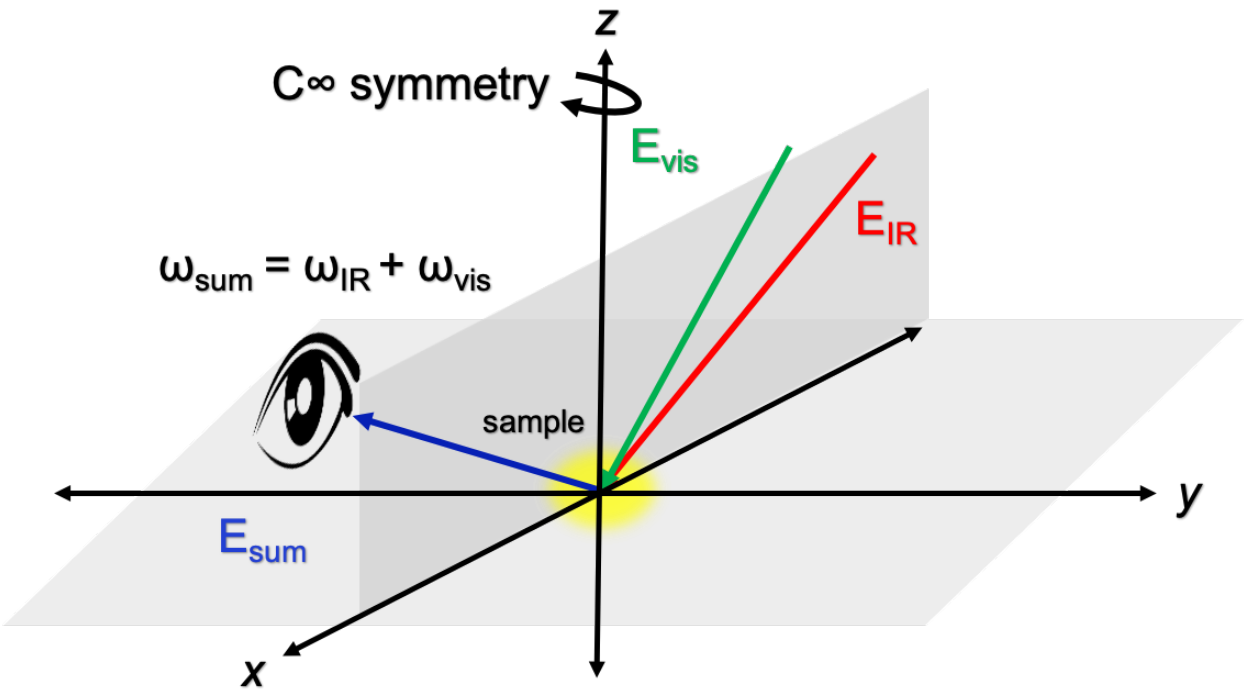

Figure 1.1. A typical vibrational SFG setup, with IR and visible beams overlapped in space and time on a sample. The eye represents the sum-frequency detector.

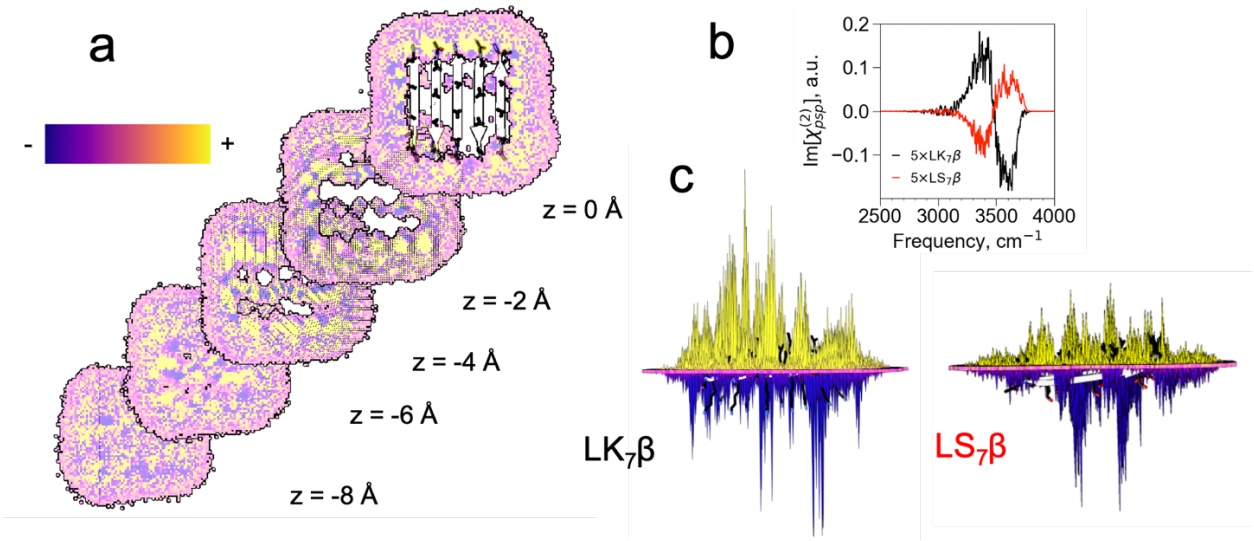

Figure 1.2. a) Contributions of various regions of water around the protein LK$_7\beta$ to the chiral phase b) chiral SFG spectra of LK$_7\beta$ and a related peptide, LS$_7\beta$, that has a nearly opposite chiral SFG response. c) The total chiral phase summed along the *z*-direction for the two systems, showing the dominance of positive contributions (yellow) for LK$_7\beta$ and negative contributions (blue) for LS$_7\beta$.

Preview of Chapters

I began my studies of chiral SFG by implementing and adapting Skinner and Corcelli's electrostatic mapping method to chiral SFG and modeling the chiral SFG response of water (Chapter 2). I then investigated the role of vibrational couplings between OH groups in the chiral SFG response (Chapter 3). I found that vibrational couplings are crucial to the production of a chiral SFG response in water. This can be explained using symmetry arguments. I then investigated the source of the chiral SFG signal of water (Chapter 4). I found that only the first hydration shell, as defined by Voronoi tessellation and neighbor analysis, contributes significantly to the chiral SFG signal. This can be explained by noting that the water molecule superstructure around a protein only lacks symmetry when very close to the protein. This study depended on the rarely used but very powerful Voronoi tessellation method. Although I had successfully modeled the chiral SFG

response of water *around* protein, I lacked a way to model the response of the protein itself, and the protein-water interaction. The relevant vibrational mode in the OH stretch region is the protein backbone NH stretch, so I designed an electrostatic map to relate local electric field around the NH group to vibrational frequencies and transition dipoles (Chapter 5). The key idea in this study was to use an internal hydrogen bond in a tetraalanine peptide to model the secondary structure of proteins in a computationally tractable way. The resulting map correctly predicts peak location trends between different systems and allows for much more accurate modeling of the experimental chiral SFG spectrum of protein at the air-water interface. I discovered that chiral SFG is quite sensitive to the vibrational coupling between NH and OH groups, which makes it an excellent method to probe protein-water interactions. I later returned to Voronoi tessellation and demonstrated its usefulness in a variety of molecular contexts, with an emphasis on SFG modeling (Chapter 6).

References

1. Bhat, T. N.; Bentley, G. A.; Boulot, G.; Greene, M. I.; Tello, D.; Dall'Acqua, W.; Souchon, H.; Schwarz, F. P.; Mariuzza, R. A.; Poljak, R. J. Bound water molecules and conformational stabilization help mediate an antigen-antibody association. *Proc. Natl. Acad. Sci. USA* **1994,** *91*, 1089-1093.
2. Dong, J.; Zost, S. J.; Greaney, A. J.; Starr, T. N.; Dingens, A. S.; Chen, E. C.; Chen, R. E.; Case, J. B.; Sutton, R. E.; Gilchuk, P., et al. Genetic and structural basis for SARS-CoV-2 variant neutralization by a two-antibody cocktail. *Nature Microbiology* **2021,** *6*, 1233-1244.
3. Petrone, P. M.; Garcia, A. E. MHC–Peptide Binding is Assisted by Bound Water Molecules. *J. Mol. Biol.* **2004,** *338*, 419-435.
4. Ahmad, M.; Kalinina, O.; Lengauer, T. Entropy gain due to water release upon ligand binding. *Journal of Cheminformatics* **2014,** *6*, P35.
5. Halle, B. Protein hydration dynamics in solution: a critical survey. *Philos Trans R Soc Lond B Biol Sci* **2004,** *359*, 1207-23; discussion 1223-4, 1323-8.
6. Nucci, N. V.; Pometun, M. S.; Wand, A. J. Site-resolved measurement of water-protein interactions by solution NMR. *Nat. Struct. Mol. Biol.* **2011,** *18*, 245-249.
7. Otting, G. NMR studies of water bound to biological molecules. *Prog. Nucl. Magn. Reson. Spectrosc.* **1997,** *31*, 259-285.
8. Eisenthal, K. B. Liquid Interfaces Probed by Second-Harmonic and Sum-Frequency Spectroscopy. *Chem. Rev.* **1996,** *96*, 1343-1360.
9. Superfine, R.; Huang, J. Y.; Shen, Y. R. Phase measurement for surface infrared–visible sum-frequency generation. *Opt. Lett.* **1990,** *15*, 1276-1278.
10. Zhu, X. D.; Suhr, H.; Shen, Y. R. Surface vibrational spectroscopy by infrared-visible sum frequency generation. *Phys. Rev. B* **1987,** *35*, 3047-3050.
11. Hunt, J. H.; Guyot-Sionnest, P.; Shen, Y. R. Observation of C-H stretch vibrations of monolayers of molecules optical sum-frequency generation. *Chem. Phys. Lett.* **1987,** *133*, 189-192.
12. Simpson, G. J.; Perry, J. M.; Moad, A. J.; Wampler, R. D. Uncoupled oscillator model for interpreting second harmonic generation measurements of oriented chiral systems. *Chem. Phys. Lett.* **2004,** *399*, 26-32.
13. Burke, B. J.; Moad, A. J.; Polizzi, M. A.; Simpson, G. J. Experimental confirmation of the importance of orientation in the anomalous chiral sensitivity of second harmonic generation. *J. Am. Chem. Soc.* **2003,** *125*, 9111-5.
14. Simpson, G. J. Structural origins of circular dichroism in surface second harmonic generation. *J. Chem. Phys.* **2002,** *117*, 3398-3410.
15. Shen, Y. R. Surface properties probed by second-harmonic and sum-frequency generation. *Nature* **1989,** *337*, 519-525.
16. Moad, A. J.; Simpson, G. J. A Unified Treatment of Selection Rules and Symmetry Relations for Sum-Frequency and Second Harmonic Spectroscopies. *J. Phys. Chem. B* **2004,** *108*, 3548-3562.
17. Simpson, G. J. Molecular Origins of the Remarkable Chiral Sensitivity of Second-Order Nonlinear Optics. *ChemPhysChem* **2004,** *5*, 1301-1310.

18. 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.
19. Mukamel, S., *Principles of Nonlinear Optical Spectroscopy*. Oxford University Press: 1999.
20. Su, X.; Ditlev, J. A.; Hui, E.; Xing, W.; Banjade, S.; Okrut, J.; King, D. S.; Taunton, J.; Rosen, M. K.; Vale, R. D. Phase separation of signaling molecules promotes T cell receptor signal transduction. *Science* **2016,** *352*, 595-599.
21. Morita, A.; Hynes, J. T. A Theoretical Analysis of the Sum Frequency Generation Spectrum of the Water Surface. II. Time-Dependent Approach. *J. Phys. Chem. B* **2002,** *106*, 673-685.
22. Morita, A. Improved Computation of Sum Frequency Generation Spectrum of the Surface of Water. *J. Phys. Chem. B* **2006,** *110*, 3158-3163.
23. Morita, A.; Ishiyama, T. Recent progress in theoretical analysis of vibrational sum frequency generation spectroscopy. *Phys. Chem. Chem. Phys.* **2008,** *10*, 5801-5816.
24. Ostroverkhov, V.; Waychunas, G. A.; Shen, Y. R. New Information on Water Interfacial Structure Revealed by Phase-Sensitive Surface Spectroscopy. *Phys. Rev. Lett.* **2005,** *94*, 046102.
25. Stiopkin, I. V.; Jayathilake, H. D.; Bordenyuk, A. N.; Benderskii, A. V. Heterodyne-detected vibrational sum frequency generation spectroscopy. *J. Am. Chem. Soc.* **2008,** *130*, 2271-2275.
26. Fu, L.; Chen, S.-L.; Wang, H.-F. Validation of Spectra and Phase in Sub-1 cm–1 Resolution Sum-Frequency Generation Vibrational Spectroscopy through Internal Heterodyne Phase-Resolved Measurement. *J. Phys. Chem. B* **2016,** *120*, 1579-1589.
27. de L. Kronig, R. On the Theory of Dispersion of X-Rays. *J. Opt. Soc. Am.* **1926,** *12*, 547-557.
28. Kramers H, A. La diffusion de la lumiere par les atomes. *Atti Cong. Intern. Fisica (Transactions of Volta Centenary Congress) Como* **1927,** *2*, 545-557.
29. Tuckerman, M. E., *Statistical mechanics: theory and molecular simulation*. Oxford University Press: New York, 2010.
30. Boyd, R. W., *Nonlinear Optics, Third Edition*. Academic Press, Inc.: 2008.
31. Morita, A.; Hynes, J. T. A theoretical analysis of the sum frequency generation spectrum of the water surface. *Chem. Phys.* **2000,** *258*, 371-390.
32. Morita, A.; Kato, S. The Charge Response Kernel with Modified Electrostatic Potential Charge Model. *J. Phys. Chem. A* **2002,** *106*, 3909-3916.
33. Corcelli, S. A.; Lawrence, C. P.; Skinner, J. L. Combined electronic structure/molecular dynamics approach for ultrafast infrared spectroscopy of dilute HOD in liquid H2O and D2O. *J. Chem. Phys.* **2004,** *120*, 8107-8117.
34. Auer, B. M.; Skinner, J. L. Dynamical effects in line shapes for coupled chromophores: Time-averaging approximation. *J. Chem. Phys.* **2007,** *127*, 104105.
35. Jansen, T. l. C.; Knoester, J. Nonadiabatic Effects in the Two-Dimensional Infrared Spectra of Peptides: Application to Alanine Dipeptide. *J. Phys. Chem. B* **2006,** *110*, 22910-22916.
36. Ohto, T.; Usui, K.; Hasegawa, T.; Bonn, M.; Nagata, Y. Toward ab initio molecular dynamics modeling for sum-frequency generation spectra; an efficient

algorithm based on surface-specific velocity-velocity correlation function. *J. Chem. Phys.* **2015,** *143*, 124702.
37. Ohto, T.; Dodia, M.; Xu, J.; Imoto, S.; Tang, F.; Zysk, F.; Kühne, T. D.; Shigeta, Y.; Bonn, M.; Wu, X., et al. Accessing the Accuracy of Density Functional Theory through Structure and Dynamics of the Water–Air Interface. *J. Phys. Chem. Lett.* **2019,** *10*, 4914-4919.
38. Ohto, T.; Dodia, M.; Imoto, S.; Nagata, Y. Structure and Dynamics of Water at the Water–Air Interface Using First-Principles Molecular Dynamics Simulations within Generalized Gradient Approximation. *J. Chem. Theory Comput.* **2019,** *15*, 595-602.
39. Khatib, R.; Sulpizi, M. Sum Frequency Generation Spectra from Velocity–Velocity Correlation Functions. *J. Phys. Chem. Lett.* **2017,** *8*, 1310-1314.
40. Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C. 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.
41. Auer, B. M.; Skinner, J. L. Dynamical effects in line shapes for coupled chromophores: Time-averaging approximation. *J. Chem. Phys.* **2007,** *127*, 104105.
42. Lee, S.-H.; Wang, J.; Krimm, S.; Chen, Z. Irreducible Representation and Projection Operator Application to Understanding Nonlinear Optical Phenomena: Hyper-Raman, Sum Frequency Generation, and Four-Wave Mixing Spectroscopy. *J. Phys. Chem. A* **2006,** *110*, 7035-7044.
43. Carr, J. K.; Wang, L.; Roy, S.; Skinner, J. L. Theoretical Sum Frequency Generation Spectroscopy of Peptides. *J. Phys. Chem. B* **2015,** *119*, 8969-8983.
44. Long, D. A.; Bell, R. P. Intensities in Raman spectra I. A bond polarizability theory. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences* **1953,** *217*, 203-221.
45. Auer, B. M.; Skinner, J. L. IR and Raman spectra of liquid water: Theory and interpretation. *J. Chem. Phys.* **2008,** *128*, 224511.

Chapter 2: Mirror-image antiparallel β-sheets organize water molecules into superstructures of opposite chirality

Adapted from the following reference with permission:

Perets, E. A.; Konstantinovsky, D.; Fu, L.; Chen, J.; Wang, H.-F.; Hammes-Schiffer, S.;

Yan, E.C.Y.

(co-first-author publication)

Proc. Natl. Acad. Sci. USA 2020, *117*, 32902-32909.

DOI: 10.1073/pnas.2015567117

Abstract

Biomolecular hydration is fundamental to biological functions. Using phase-resolved chiral sum-frequency generation spectroscopy (SFG), we probe molecular architectures and interactions of water molecules around a self-assembling antiparallel β-sheet protein. We find that the phase of the chiroptical response from the O-H stretching vibrational modes of water flips with the absolute chirality of the (L-) or (D-) antiparallel β-sheet. Therefore, we can conclude that the (D-) antiparallel β-sheet organizes water solvent into a chiral supermolecular structure with opposite handedness relative to that of the (L-) antiparallel β-sheet. We use molecular dynamics to characterize the chiral water superstructure at atomic resolution. The results show that the macroscopic chirality of antiparallel β-sheets breaks the symmetry of assemblies of surrounding water molecules. We also calculate the chiral SFG response of water surrounding (L-) and (D-) LK7β to confirm the presence of chiral water structures. Our results offer a different perspective as well as introduce experimental and computational methodologies for elucidating hydration of biomacromolecules. The findings imply potentially important but largely unexplored roles of water solvent in chiral selectivity of biomolecular interactions and the molecular origins of homochirality in the biological world.

Introduction

Biological functions at the molecular level require changes in the structures and dynamics of biomacromolecules. These changes are strongly influenced by water hydration shells.[1] Probing the structures and dynamics of water molecules in hydration shells is fundamental for understanding molecular mechanisms of biological functions. Our group demonstrated that a water superstructure assembles around an antiparallel β-sheet protein and gives a chiroptical response using vibrational sum frequency generation (SFG) spectroscopy.[2] This result was informed by Petersen and coworkers' observation of chiral vibrational SFG signals from water molecules in the minor groove of double helix DNA.[3, 4] Since water molecules are achiral, these two studies suggest that water molecules are arranged in chiral supermolecular structures. Those results align with Simpson's theory that achiral molecular entities incorporated into a chiral architecture can generate chiral second-order optical responses.[5, 6] These chiral SFG studies have led to a model in which biomacromolecules act as a template to organize the surrounding water solvent into chiral superstructures. Nonetheless, many questions remain. For example, what are the architectures of water molecules in the superstructures? What are the molecular interactions that drive water self-assembly?

Here, we probe the chiral water superstructure around an antiparallel β-sheet secondary structure using heterodyne phase-resolved chiral SFG vibrational spectroscopy. We study the amphiphilic $LK_7β$ peptide[2, 7-12] composed of (L-) or (D-) amino acids. These (L-) or (D-) $LK_7β$ peptides self-assemble into mirror-image antiparallel β-sheets.[11-14] Designed by DeGrado and Lear in 1985,[7] the $LK_7β$ peptide with the sequence LKLKLKL is strongly amphiphilic, with the hydrophobic leucine sidechains orienting toward the

hydrophobic phase and the polar lysine sidechains orienting toward the hydrophilic phase.[11] This sidechain orientation forces the peptide backbone to adopt Ramachandran angles standard for antiparallel β-sheets.[11] Hence, it is energetically favorable for the peptide to self-assemble into antiparallel β-sheet structures. Indeed, the $LK_7\beta$ peptide and its variants in native (L-) form have been commonly used as a model system in surface studies.[8-12, 15-17]

We obtain the phase-resolved chiral SFG vibrational spectra of (L-) and (D-) $LK_7\beta$ in the spectral regions of the O-H/N-H stretching at ~3,200 cm^{-1} using H_2O as solvent and the O-D/N-D stretching at ~2,400 cm^{-1} using D_2O. We also perform the experiments using $H_2{}^{18}O$ and $D_2{}^{18}O$. The ^{18}O-isotopic substitution allows for the unambiguous assignment of two red-shifted vibrational bands to the water O-H (or O-D) stretching modes and two unaffected vibrational bands to the peptide backbone N-H (or N-D) stretching modes. Remarkably, when the chirality of the antiparallel β-sheets changes from (L-) to (D-), not only does the phase of the two N-H (N-D) stretching modes reverse but also the phase of the two O-H (O-D) stretching modes.

We also use molecular dynamics (MD) to provide a qualitative description of the average structure of the chiral water envelope around the (L-) antiparallel β-sheet. The results show that water molecules are strongly oriented in the presence of positively charged polar sidechains, both near the protein backbone at the vacuum–water interface and at least 10 Å into bulk water. We identify aspects of the water superstructure that possess a chiral topology. Additional hydrogen-bond analyses elucidate the relationship between the macroscopic chirality of antiparallel β-sheets and the organization of a water

superstructure. Finally, we calculate a chiral SFG response of water molecules in the presence of the protein.

The observation of a direct correlation between chiroptical responses of achiral water solvent and the absolute chirality of a biomacromolecule has an important implication: mirror-image biomacromolecules can organize water solvent into chiral supermolecular structures with opposite handedness. Since water O-H stretching modes are highly sensitive to hydrogen-bonding environments, this work offers a different perspective as well as introduce experimental and computational methods for investigating water structures and hydrogen-bonding interactions in hydration shells of biomacromolecules. Our findings imply that chiral biomacromolecules can be enveloped by chiral water superstructures. This observation promises greater understanding of the chiral selectivity of biomolecular interactions and potentially suggests a key role for water in the origins of homochirality in the biological world.

Methods

Quartz Calibration

Right-handed z-cut α-quartz crystal was cleaned with high-performance liquid chromatography-grade methanol and then ddH_2O, dried with nitrogen, and plasma-cleaned (Harrick Plasma; PDC-32G) on "low" for 15 min. Adsorbed 1-palmitoyl-2-oleoyl-glycero-3-phosphocholine (Avanti Polar Lipids; catalog number 850457C) was employed to define the quartz crystal coordinate system as described in the literature.[18]

Sample Preparation

The (L-) $LK_7\beta$ peptide (GL Biochem Ltd., Shanghai) and the (D-) $LK_7\beta$ (AnaSpec, Inc.) were obtained as lyophilized powders. The sample preparation was previously described.[2, 11] Briefly, $LK_7\beta$ (Ac-Leu-Lys-Leu-Lys-Leu-Lys-Leu-NH_2) was synthesized with all (L-) or all (D-) amino acids. The lyophilized powder was prepared at a concentration of 1 mM in H_2O, $H_2{}^{18}O$, D_2O, or $D_2{}^{18}O$. The solution was aliquoted, flash frozen in liquid nitrogen, and transferred to −80 °C for storage. Once thawed for SFG experiments, unused material was discarded to avoid additional freeze–thaw cycles.

For each experiment, 10 to 20 μL of $LK_7\beta$ aqueous solution was dried directly on the cleaned quartz surface. The solution was dried under nitrogen flow in a sealed container where desiccant was present to prevent exchange with ambient humidity, resulting in a hydrated thin film of peptide. The adsorbed film was immediately measured.

Phase-Resolved Chiral Sum Frequency Generation Experiments

The femtosecond broadband SFG spectrophotometer has been previously described.[19] The quartz crystal with $LK_7\beta$ hydrated thin-film was probed at φ = 90° (+y axis) and 270° (−y axis), where φ is the azimuthal angle of the quartz axis relative to the +x axis in the laboratory frame (Figure A2.1). All chiral SFG spectra were obtained using the psp polarization (p-polarized sum frequency, s-polarized visible, p-polarized IR).[18, 20] For each sample, 10 to 12 spectra (2 min each) were acquired along the +y and −y axes. The spectra were cleaned for cosmic rays. The averages of the measurements along the +y and −y axes were normalized by $\chi^{(2)}_{quartz,\phi=90°}$ and then subtracted according to the following:[18, 20, 21]

$$\mathrm{Im}[\chi^{(2)}] \propto \frac{I_{\phi=90^\circ} - I_{\phi=270^\circ}}{4} \quad (1)$$

The IR frequencies were calibrated according to a polystyrene standard (Buck Scientific; 0.05-mm film). Finally, the vibrational spectra were fit with Lorentzians according to the following:

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

where ω_{IR} is the frequency of the incident IR, A_q is the amplitude, ω_q is the frequency of the vibrational resonance, and Γ_q is the damping coefficient of the q^{th} vibrational mode.

Molecular Dynamics

A starting structure of five antiparallel $LK_7\beta$ strands was placed into a 125 × 125 × 125 Å box, half of which was filled with TIP4P-Ew[22] water and Cl^- ions to neutralize the system to create a vacuum–water interface. The protein, which was described with the AMBER ff14SB force field,[23] was placed at the interface with the lysine sidechains pointing into the water and the leucine sidechains facing the vacuum. The N termini were acetylated and the C termini were amidated to be consistent with the experimental system. The structure was minimized and then equilibrated for 6 ns at 298 K and constant volume to preserve the vacuum. Langevin dynamics was propagated at 298 K with a friction coefficient of 1 ps^{-1}. The long-range electrostatics were treated with the particle-mesh Ewald method.[24] Multiple trajectories for a total of 1 μs sampling were propagated, and the configurations were saved every 10 ps. For the chiral SFG calculation, a single trajectory was propagated for 100 ns, saving configurations every 10 fs. The simulations were

performed with OpenMM 7.4 and CUDA 9.2.[25] See Appendix for more details on the MD trajectory analysis as well as the IR and chiral SFG calculations.

Results

Homodyne Chiral Sum Frequency Generation of (L-) and (D-) LK$_7$β

We prepared hydrated thin films of the amphiphilic LK7β peptide on the surface of a right-handed α-quartz crystal. Using the psp polarization setting (see Methods), we first obtained homodyne chiral SFG spectra of the (L-) LK7β (Figure 2.1, top) as control experiments. We then used the quartz signal as an internal phase reference to obtain the phase-resolved chiral SFG spectra of the antiparallel β-sheets formed by the (L-) and (D-) LK7β peptides (Figure 2.1, bottom). Compared to conventional homodyne SFG, the heterodyne phase-resolved SFG method reveals the absolute phases of each vibrational band relative to the phase of the quartz crystal, providing information about orientation and chirality of the molecular systems (see "Theoretical Background of Phase-Resolved Chiral Sum Frequency Generation" in the Appendix).[18, 20, 21]

Figure 2.1a shows the chiral SFG spectra in the amide I/amide II spectral region. The amide I vibrational modes are mainly the C=O stretching of the amide group coupled to the bending of the N-H bond, while the amide II vibrational modes are the C-N stretching vibrations in combination with N-H bending.[26, 27] They are sensitive to protein secondary structures.[26-28] The homodyne amide I band (Figure 2.1a, top) of (L-) LK$_7$β is centered at 1,620 cm^{-1}, indicating the formation of antiparallel β-sheet structures. Previously, we showed that the homodyne chiral SFG signal vanishes when the pH of the sample is lowered to ~2, which denatures the antiparallel β-sheets, or when the LK$_7$β peptide is

composed of (L-) leucine and (D-) lysine, which disrupts β-sheet formation.[11] These experiments demonstrated that the chiral SFG response originates from the folded β-sheet secondary structure. Moreover, the phase-resolved amide I and amide II chiral SFG responses (Figure 2.1a, bottom) for the (L-) $LK_7\beta$ (red) and (D-) $LK_7\beta$ (blue) are mirror images, demonstrating the capacity of phase-resolved chiral SFG spectroscopy to reveal the handedness of the enantiomeric (L-) versus (D-) antiparallel β-sheets.

Figures 2.1b and 2.1c (top, gray) show the homodyne chiral SFG spectra of the (L-) $LK_7\beta$ samples on a glass surface in the spectral region of O-H/N-H and O-D/N-D stretching for the β-sheet prepared in H_2O and D_2O. Without phase resolution, the spectra are very similar to the previously reported homodyne spectra of the β-sheets on a glass surface or at the air–water interface, suggesting that the structures of $LK_7\beta$ are likely conserved regardless of whether the peptide is at the air–water interface or on the solid substrates.[2, 10] In the previous report, the major and minor bands in the (L-) $LK_7\beta$-H_2O spectrum were assigned to the peptide N-H stretching and water O-H stretching, respectively, while both the major and minor vibrational bands in the (L-) $LK_7\beta$-D_2O spectrum were assigned to chiral water O-D stretching modes.[2] Although these previous assignments were supported by $H_2{}^{18}O$ and $D_2{}^{18}O$ isotopic shifts, they give rise to the question, why is the N-D stretching band of the peptide backbone in the (L-) $LK_7\beta$-D_2O spectrum missing?[2] We address this question in the next section using our heterodyne phase-resolved spectra.

Heterodyne Phase-Resolved Chiral Sum Frequency Generation of (L-) and (D-) $LK_7\beta$ Prepared in H_2O, $H_2{}^{18}O$, D_2O, and $D_2{}^{18}O$

The phase-resolved chiral SFG spectra of (L-) $LK_7\beta$ in the O-H/N-H and O-D/N-D stretching vibrational regions are shown on the bottom row of Figures 2.1b and 2.1c (red). These phase-resolved spectra reveal spectral features that are hidden in the homodyne spectra of (L-) $LK_7\beta$ (Figures 2.1b and 2.1c, top, gray). The same figures also show the phase-resolved spectra for the enantiomeric (D-) $LK_7\beta$ antiparallel β-sheet (blue), displaying a mirror-image spectral response. Residual analyses of these spectra indicate that at least four vibrational bands are needed to fit the spectra (Figures A2.2 and S2.3). Since both the peptide N-H (N-D) and water O-H (O-D) stretches can contribute to these vibrational bands, we performed ^{18}O substitution using $H_2{}^{18}O$ and $D_2{}^{18}O$ to distinguish their contributions. The ^{18}O substitution is expected not to alter the N-H or N-D stretching frequency but to redshift the O-H or O-D stretching frequency by roughly 12 cm^{-1}.[29] If a vibrational band undergoes such a redshift upon ^{18}O substitution, it is assigned to water O-H (or O-D) stretching modes; otherwise, it is assigned to peptide N-H (or N-D) stretching modes. We present the results of ^{18}O substitution for (L-) $LK_7\beta$ in Figure 2.2 and (D-) $LK_7\beta$ in Figure 2.3.

Figure 2.2 shows phase-resolved chiral SFG spectra of the (L-) $LK_7\beta$ β-sheet hydrated in H_2O, $H_2{}^{18}O$, D_2O, and $D_2{}^{18}O$. For the β-sheet prepared using H_2O (Figure 2.2a), the spectrum is resolved into four vibrational bands. The ones at the lowest (ω_1) and highest (ω_4) frequencies are assigned to the water O-H stretching modes because the corresponding vibrational bands in the $H_2{}^{18}O$ spectrum (Figure 2.2b) are redshifted by 11 to 13 cm^{-1} (Table 2.1). The frequencies of the two vibrational bands in the middle (ω_2 and ω_3)

are not significantly affected by ^{18}O substitution and thus are assigned to the peptide N-H stretching modes. The spectra of the β-sheet prepared using D_2O (Figure 2.2c) and $D_2{}^{18}O$ (Figure 2.2d) are also resolved into four vibrational bands. Similarly, the vibrational bands at the lowest (ω_1) and highest (ω_4) frequencies are redshifted (>10 cm^{-1}) upon ^{18}O substitution (Table 2.1) and are assigned to the water O-D stretching modes, while the two vibrational bands in the middle (ω_2 and ω_3) are assigned to the peptide N-D stretching modes. The peptide N-D stretching modes buried in the homodyne chiral SFG spectra of (L-) β-sheet (Figure 2.1c, top) are revealed in the phase-resolved chiral SFG spectrum.

We repeated the ^{18}O-substitution experiments with the enantiomeric (D-) $LK_7\beta$. Figure 2.3 presents the phase-resolved chiral SFG spectra of the (D-) β-sheet hydrated in H_2O, $H_2{}^{18}O$, D_2O, and $D_2{}^{18}O$. Each spectrum is fit to four vibrational bands, yielding fitting parameters (Table 2.2) that are comparable to those obtained for the (L-) $LK_7\beta$ (Table 2.1) but with opposite phase. The vibrational bands at the lowest (ω_1) and highest (ω_4) frequencies are redshifted (>10 cm^{-1}) upon ^{18}O substitution with $H_2{}^{18}O$ ($D_2{}^{18}O$) (Table 2.2) and thus are assigned to the water O-H (O-D) stretching modes, while the other two vibrational bands (ω_2 and ω_3) are assigned to the peptide N-H (N-D) stretching modes.

Molecular Dynamics and Chiral Sum Frequency Generation Simulation of $LK_7\beta$ Hydration Structure

To gain further insights about the hydration structure around $LK_7\beta$, we performed MD simulations and hydrogen-bond analyses (see Methods and Figure A2.3). $LK_7\beta$ was modeled as a protein composed of five β-strands at the vacuum–water interface, as previously reported.[11] The MD simulations confirmed that the amphiphilic $LK_7\beta$ adopts a

stable pleated, antiparallel β-sheet structure (Figures A2.4 and A2.5). Figure 2.4a shows the backbone structure of the (L-) $LK_7\beta$ homopentamer at the atomic level as well as the average number of hydrogen-bonding interactions between the protein backbone and surrounding water molecules. A major difference between the two exterior strands 1 and 5 is that strand 1 has four N-H groups exposed to water, whereas strand 5 has only three N-H groups exposed to water. This difference in hydrogen-bonding capacity is one aspect that defines the macroscopic chirality of antiparallel β-sheets, that is, a lack of a reflection plane along the peptide strands (Figure 2.4a, right).

Another source of chirality is revealed by analyses of the average z-component (normal to the vacuum–water interface) of the water dipole vectors around the (L-) $LK_7\beta$ homopentamer (Figure 2.4b). In the absence of $LK_7\beta$, bulk water in the simulation box adopts an isotropic orientation, except at the interface where water molecules preferentially orient oxygen either toward or away from the water phase. We observe that this behavior persists up to ∼5 Å below the interface (Figure A2.6).[30-32] In contrast, the presence of $LK_7\beta$ strongly perturbs the water structure and orientation. This effect is clearly illustrated by the nonisotropic orientation of water at least ∼10 Å below the vacuum–water interface (Figure 2.4b, top) and extending at least ∼15 Å on either side from the center of mass of the (L-) $LK_7\beta$ homopentamer (Figures 2.4b, bottom and S2.7). In the vicinity of lysine residues and close to the vacuum–water interface, water tends to orient with its oxygen away from the interface toward the positively charged lysine sidechain (dark purple at $z = -2$ Å in Figure 2.4b, top). In contrast, in the interstices near the protein backbone between lysine residues, water tends to orient its oxygen toward the interface. The resulting truncated checkerboard-like pattern of alternating up/down hydration water dipole orientations ∼2 to 6 Å below

the protein surface (Figure 2.4b, top) is a chiral superstructure. This pattern was stable over the course of the trajectory and was reproduced with multiple independent trajectories. The pattern appeared with both the SPC/E and TIP4P-Ew water models (Figure A2.8).[22, 33]

We also calculated the infrared vibrational response of the O-H stretch of water in our MD simulations (Appendix). We selectively probed the infrared (IR) response of water molecules within 3 Å of the $LK_7\beta$ backbone. We then identified water molecules for which the average z-component (μ_z) of the molecular dipole moment was oriented up or down relative to the vacuum–water interface. Figure 2.4c (and Figure A2.9) shows that the calculated mean IR frequency of the O-H stretch is around 3,400 cm^{-1} (black) for upward-oriented waters and around 3,300 cm^{-1} (red) for downward-oriented waters. Therefore, the orientation of water molecules near the (L-) $LK_7\beta$ homopentamer peptide backbone significantly influences the water O-H stretching frequency.

Finally, we calculated the phase-resolved (heterodyne) chiral SFG response of the water molecules surrounding the (L-) $LK_7\beta$ homopentamer and its enantiomer, (D-) $LK_7\beta$ (Appendix). The enantiomers produced identical SFG responses but with opposite phase, consistent with the (D-) protein templating an oppositely chiral water structure compared to the (L-) protein (Figure 2.5a). The (L-) $LK_7\beta$ spectrum contains a prominent peak around 3,357 cm^{-1}, which possibly corresponds to the positive peak at 3,359 cm^{-1} (ω_4) in the experimental heterodyne spectrum (Figure 2.2a). The computational result did not show the additional negative-phase 3,157 cm^{-1} peak (ω_1) observed experimentally. Given how redshifted this peak is from the gas-phase O-H stretching frequency (~3,700 cm^{-1}), it may arise from a rare population of water molecules forming very strong hydrogen bonds. These bonds may not be captured by the rigid TIP4P-Ew water model used in these calculations.

This peak may also simply require more sampling to be resolved. Our simulations also predict an additional negative peak at a higher frequency that was beyond the spectral range probed experimentally. In addition, we calculated the chiral SFG response from consecutive water layers of widths of 5 Å starting at the air–water interface (Figure 2.5b). The results confirm that water forms chiral assemblies at least as deep as the lysine sidechains can extend into the solution (∼10 Å) and possibly deeper (Figure 2.4b). The spectra start to become noisier below 10 Å, but the opposite phases of the (L-) and (D-) $LK_7\beta$ spectra exhibited at all depths sampled suggest that some chiral water structures exist at least 15 Å below the interface.

Discussion

A comparison of the phase-resolved chiral spectra of the (L-) and (D-) β-sheets (Figures 2.2 and 2.3) reveals a remarkable observation. Although all of the vibrational bands observed under the same solvent conditions (H_2O, D_2O, $H_2{}^{18}O$, and $D_2{}^{18}O$) are comparable for the (L-) and (D-) β-sheets in terms of peak positions (ω), peak widths (Γ), and absolute amplitudes (A) (Tables 2.1 and 2.2), the phases of both the peptide N-H (N-D) stretching modes and the water O-H (O-D) stretching modes are reversed with the (L-) and (D-) handedness of the $LK_7\beta$ antiparallel β-sheets. It is expected that the phase flips for the N-H (N-D) stretching vibrational bands because the vibrational modes originate from the chiral peptide backbone.[12, 34] However, the accompanying phase reversal of the two water-stretching bands has an important implication. Enantiomers of a biomacromolecule can stamp their intrinsic handedness onto nearby water solvent molecules and organize water into superstructures with opposite chirality. This observation

provides a new perspective on the molecular architectures of water in the hydration shells of biomacromolecules.

Since 2011 when vibrational chiral SFG signatures of N-H stretching modes of peptide backbones were first reported for characterizing protein secondary structures,[35] the interpretation of the chiral SFG spectra of $LK_7\beta$ has continued to evolve. On the one hand, Fu et al.[10] reported the homodyne chiral SFG spectra of $LK_7\beta$ at the air-H_2O and air-D_2O interfaces. Each of these spectra exhibits a major band and a minor band. Based on *ab initio* calculations, the major band was assigned to the peptide backbone and the minor band was assigned to combinational bands of various peptide vibrational modes. On the other hand, Perets et al.[2] obtained the homodyne chiral SFG spectra of $LK_7\beta$ hydrated in H_2O and D_2O on glass slides. Based on the ^{18}O-substitution analyses, the major and minor bands in the $LK_7\beta$-H_2O spectrum were assigned to the peptide N-H stretching and the water O-H stretching modes, respectively, and both the major and minor bands in the $LK_7\beta$-D_2O spectrum were assigned to the water O-D stretching modes.[2] Paradoxically, the peptide N-D stretching signal seemingly disappears in the $LK_7\beta$-D_2O spectrum. In this study, the paradox is resolved by heterodyne phase-resolved SFG spectroscopy. The capacity to reveal phases of vibrational bands allows deconvolution of the two apparent vibrational bands in the homodyne spectra into four vibrational bands. Two bands are assigned to peptide N-H stretching modes and two are assigned to water O-H stretching modes (Figures 2.2 and 2.3). Since both N-H and O-H stretching modes are highly sensitive to hydrogen-bonding environments, heterodyne phase-resolved chiral SFG spectroscopy in conjunction with computational studies is a promising approach for revealing molecular details in the chiral assembly of water molecules around the antiparallel β-sheet.

The two O-H (or O-D) stretches at ω_1 and ω_4 in each spectrum of Figures 2.2 and 2.3 can inform hydrogen-bonding interactions of water molecules in the chiral superstructures. As shown in Table 2.1, the two water O-H stretching bands are at 3,157 ± 1 cm^{-1} and 3,359 ± 1 cm^{-1} ($\Delta\nu$ = 202 ± 1 cm^{-1}), while the two ^{18}O-H stretching bands are at 3,146 ± 3 cm^{-1} and 3,346 ± 2 cm^{-1} ($\Delta\nu$ = 200 ± 3 cm^{-1}). Correspondingly, the two water O-D stretching bands are at 2,342 ± 2 cm^{-1} and 2,499 ± 6 cm^{-1} ($\Delta\nu$ = 157 ± 6 cm^{-1}), while the two ^{18}O-D stretching bands are at 2,331 ± 4 cm^{-1} and 2,488 ± 3 cm^{-1} ($\Delta\nu$ = 157 ± 4 cm^{-1}). When an O-H oscillator forms a hydrogen bond with a proton acceptor, the O-H bond is weakened, and thus the O-H stretching frequency is lowered. Therefore, the stronger the hydrogen-bonding interaction, the lower the O-H stretching frequency. Consequently, we assign the low-frequency (ω_1) band to strongly hydrogen-bonded O-H and the high-frequency (ω_4) band to weakly hydrogen-bonded O-H.

The 200 cm^{-1} vibrational energy difference of the two water O-H stretching bands (ω_1 and ω_4) suggests their hydrogen-bond strength is substantially different.[36] To understand the origins of this vibrational energy difference, we calculated the IR vibrational response of the O-H stretch of water molecules in our MD simulations. Our calculations show that for water molecules less than 3 Å from the protein backbone, the orientation significantly influences the mean O-H stretching frequency (Figure 2.4c). If the oxygen atom orients upward at the vacuum–water interface, the O-H stretching frequency is calculated to be ∼100 cm^{-1} greater than if the oxygen orients downward. As discussed earlier, the stronger the hydrogen-bonding interaction, the lower the O-H stretching frequency. Hence, water molecules pointing toward the interface may form hydrogen bonds that are weaker than those formed by water molecules pointing into the bulk. This

model aligns with our experimental observation of the two water O-H stretching bands (ω_1 and ω_4) at vibrational frequencies different by 200 cm^{-1}.

The analysis of the water dipole vectors shows not only a truncated checkerboard-like pattern with alternating up/down water orientations (Figure 2.4b) but also variations of water orientation patterns at the four edges of the $LK_7\beta$ homopentamer (Figure A2.10). These results indicate a lack of a reflection plane perpendicular to the β-sheet, suggesting that the water assemblies around (L-) $LK_7\beta$ are chiral. This suggestion is confirmed by our chiral SFG calculations (Figure 2.5). This behavior may be compared with the chiral "spine of hydration" around DNA,[37] which computational studies suggest requires steric interactions to order water molecules in the DNA minor groove.[38, 39] However, in the case of $LK_7\beta$, the driving force to assemble water molecules into a chiral superstructure around the antiparallel β-sheet appears to be hydrogen-bonding interactions. Our computational results suggest that these hydrogen-bonding interactions may significantly perturb the topology of water molecules at least several solvation shells beyond the protein.

We also identified two peptide N-H stretching bands with an opposite phase at 3,247 and 3,270 cm^{-1} (Figure 2.2a), consistent with a prior report,[12] and two peptide N-D stretching bands with an opposite phase at 2,391 and 2,445 cm^{-1} (Figure 2.2c). To explain the opposite phases of the two N-H stretching bands, Hu et al.[12] proposed that adjacent backbone nitrogen atoms along the β-strand are chiral centers with opposite handedness (i.e., pro-*S* and pro-*R*). This proposal would require a substantial sp^3-hybridized character of the nitrogen atoms. Researchers in the broader fields of structural biology and computational biology have long considered peptide bonds to be in planar configuration .[40-42] In a survey of ultra-high-resolution (1 Å or better) protein structures, larger deviations

from planarity (>20°) were observed in only 0.5% of residues.[43] However, small deviations (±6.3°) of the amide bond from planarity are more common.[42, 43] Determining whether such small, dynamic deviations from planarity can confine the nitrogen atoms into pro-*S* and pro-*R* chiral centers requires further investigation.

The two N-H (or N-D) stretching bands at different frequencies ($\Delta\nu = \sim 23$ cm^{-1}) likely arise due to backbone amines that experience distinct chemical environments in the antiparallel β-sheet. For instance, the hydrophobic leucine and hydrophilic lysine side chains are oriented toward opposite sides of the β-sheet, and the pleated antiparallel β-sheet backbone (Figure A2.4a) constrains the orientation of the backbone amine groups. Indeed, the orientations of backbone amines relative to the plane of the β-sheet alternate up/down across adjacent residues (Figures A2.4b and A2.4c). The leucine or lysine backbone amines could give chiral SFG responses at different frequencies because of orientational differences that lead to changes in hydrogen bonding or proximity to the hydrophobic or hydrophilic sidechains. Alternatively, hydrogen bonding of amine groups is likely different in the interior and exterior of β-sheet domains.[44] The interior N-H groups form hydrogen bonds with the C=O groups of the peptide backbone in the neighboring β-strands, while the exterior N-H groups at the edge of the β-sheet form hydrogen bonds with water solvent. The interior N-H groups presumably form stronger hydrogen bonds (C=O···H-N) at lower stretching frequencies, and the exterior N-H groups form weaker hydrogen bonds (H_2O···H-N) at higher stretching frequencies.[45-47]

Finally, as the chirality of water superstructures is dictated by the chirality of the protein, there is room for speculation about the role chiral water architectures might have played in the origins of homochirality in the biological world.[48] With few exceptions,[49-51]

native proteins are made of (L-) amino acids, and DNA and RNA are composed of (D-) nucleotides. The origins of enantiomeric enrichment evolving from a prebiotic state remains a matter of fascination and controversy. One proposal is that physical or chemical processes (e.g., autocatalytic Soai reactions) that establish an energy difference between chiral enantiomers on the order of 10^{-7} kJ/mol could introduce a bias toward homochirality.[52-57] Another proposal notes that the incorporation of achiral solvent molecules (such as chloroform and fumaric acid) in a racemic crystal of amino acids lowers the crystal's solubility; when the crystal is grown from a solution of small enantiomeric excess, this leads to significant chiral enrichment in the solution phase.[58, 59] One might plausibly extend this concept from achiral organic solvents to water itself. Here, we show that the description of chirality of biomacromolecules is incomplete without accounting for the water solvent. The presence of a chiral hydration shell around chiral biomolecules is potentially one of the missing links in the search for the molecular origins of biological homochirality.

Conclusions

Chiral vibrational SFG spectroscopy revealed that achiral water solvent can form a chiral hydration shell around a chiral biomacromolecular solute.[2-4] In this work, phase-sensitive chiral SFG reveals that the phase of such chiroptical response is inverted with the absolute handedness of the chiral biomacromolecule. The MD and hydrogen-bonding analyses show that the intrinsic macroscopic chirality of a protein can impact the surrounding water structure. Analysis of the MD trajectories reveals a chiral water topology around the antiparallel β-sheet in terms of the water orientations and hydrogen-bonding

interactions. These experimental and computational results suggest that chiral macromolecules can imprint their intrinsic chirality onto surrounding water solvent molecules and organize them into chiral supermolecular structures. A chiral hydration shell wrapping around a biomacromolecule is expected to change not only the structural and dynamical properties of the biomacromolecule but also chiral selectivity toward biomolecular interactions.

The two water O-H stretching bands and two peptide N-H stretching bands contain information about vibrational structures and hydrogen-bonding interactions of the chiral water assembly around the $LK_7\beta$ antiparallel β-sheet. These vibrational resonances can be used as molecular probes for future ultrafast dynamic studies of vibrational energy exchange between water solvent and protein. The phase, peak position, and line shape of the four vibrational bands under various solvent conditions (H_2O, D_2O, $H_2{}^{18}O$, and $D_2{}^{18}O$) will serve as quantitative benchmarks in building computational models for elucidating detailed structures and interactions in the water chiral superstructures around the β-sheet. Heterodyne phase-sensitive chiral SFG in conjunction with computational studies constitutes a powerful approach for studying the molecular details of structures and dynamics of biomacromolecular hydration. This study has established a molecular picture: Mirror-image biomacromolecules are enveloped in hydration shells with opposite chirality. From here, researchers can start to question what roles water might have played in the origins of homochirality in the biological world.

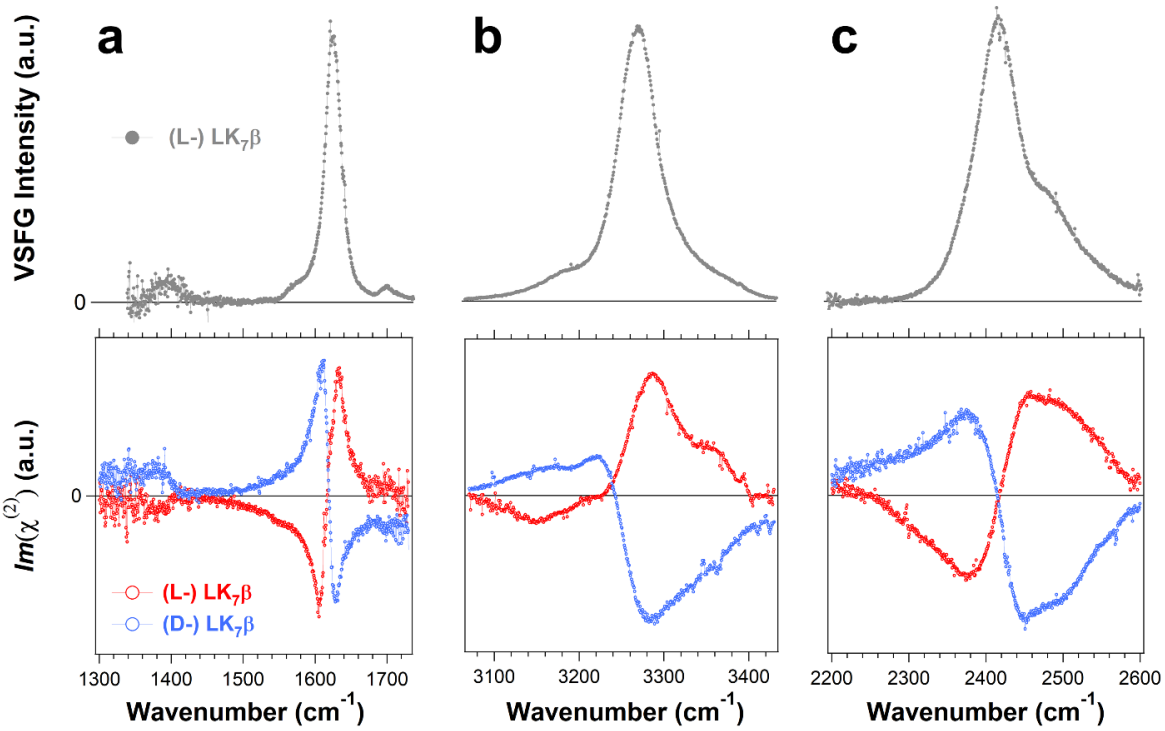

Figure 2.1. Homodyne (top) and phase-resolved chiral (bottom) SFG spectra of chiral water superstructures around $LK_7\beta$ antiparallel β-sheets (VSFG: vibrational sum frequency generation; a.u.: arbitrary units). (a) The amide I/amide II spectral region of (L-) $LK_7\beta$ (gray and red) and (D-) $LK_7\beta$ (blue) prepared in H_2O. (b) The O-H/N-H spectral region of $LK_7\beta$ prepared in H_2O. (c) The O-D/N-D spectral region of $LK_7\beta$ prepared in D_2O. All spectra were collected with the chiral *psp* polarization.

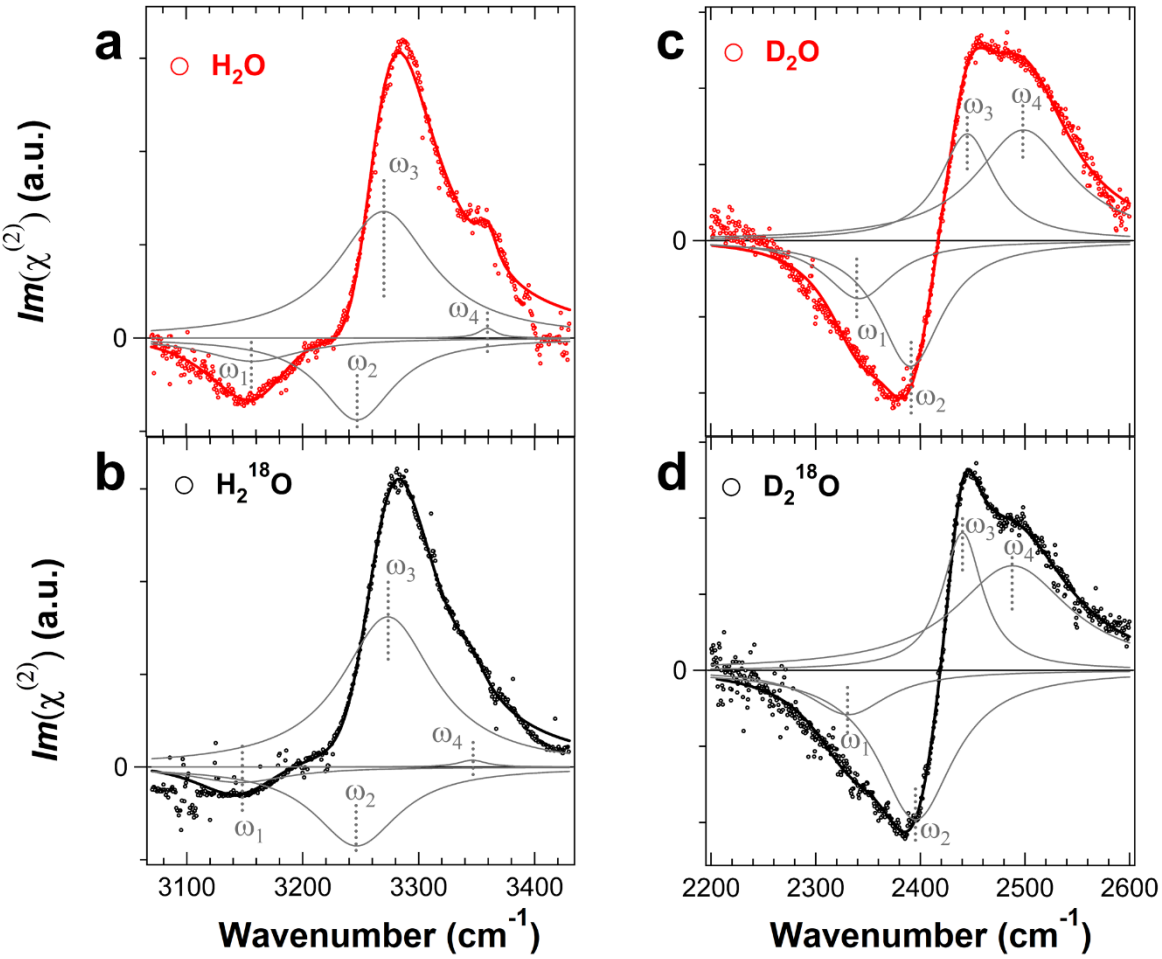

Figure 2.2. The phase-resolved chiral SFG spectra of chiral water superstructures around (L-) LK$_7$β. (Left) O-H/N-H spectral region of (L-) LK$_7$β prepared in (a) H_2O and (b) $H_2{}^{18}O$. (Right) O-D/N-D spectral region of (L-) LK$_7$β prepared in (c) D_2O and (d) $D_2{}^{18}O$. The spectra are fit to equation 2 (**Methods**). The component peaks based on the fitting parameters are shown in gray, and the sum of these peaks are shown as red or black solid lines. The labels of vibrational modes correspond to the numbering in Table 2.1. All spectra were collected with the chiral *psp* polarization.

Parameter	(Figure 2.2a) (L-) $LK_7\beta$-H_2O	(Figure 2.2b) (L-) $LK_7\beta$-$H_2{}^{18}O$	Assignments
ω_1 (cm^{-1})	3157 ± 1	3146 ± 3	O-H (H_2O)
A_1 (a.u.)	-10.1 ± 0.5	-5.1 ± 1.3	
Γ_1 (cm^{-1})	50.7 ± 1.8	50.9 ± 9.6	
ω_2 (cm^{-1})	3247 ± 2	3247 ± 6	N-H (L-$LK_7\beta$)
A_2 (a.u.)	-27.1 ± 1.7	-22.7 ± 13.5	
Γ_2 (cm^{-1})	38.6 ± 0.9	44.5 ± 5.5	
ω_3 (cm^{-1})	3270 ± 1	3274 ± 3	N-H (L-$LK_7\beta$)
A_3 (a.u.)	54.0 ± 1.6	47.6 ± 12.9	
Γ_3 (cm^{-1})	49.7 ± 0.7	49.0 ± 1.2	
ω_4 (cm^{-1})	3359 ± 1	3346 ± 2	O-H (H_2O)
A_4 (a.u.)	0.8 ± 0.2	0.6 ± 0.2	
Γ_4 (cm^{-1})	9.5 ± 2.4	14.7 ± 4.4	
Parameter	(Figure 2.2c) (L-) $LK_7\beta$-D_2O	(Figure 2.2d) (L-) $LK_7\beta$-$D_2{}^{18}O$	Assignments
ω_1 (cm^{-1})	2342 ± 2	2331 ± 4	O-D (D_2O)
A_1 (a.u.)	-9.0 ± 3.4	-8.0 ± 2.2	
Γ_1 (cm^{-1})	39.5 ± 3.4	45.7 ± 4.7	
ω_2 (cm^{-1})	2391 ± 1	2396 ± 1	N-D (L-$LK_7\beta$)
A_2 (a.u.)	-18.2 ± 3.1	-26.0 ± 4.5	
Γ_2 (cm^{-1})	36.8 ± 3.6	44.1 ± 4.5	
ω_3 (cm^{-1})	2445 ± 10	2440 ± 1	N-D (L-$LK_7\beta$)
A_3 (a.u.)	13.4 ± 5.6	12.7 ± 1.9	
Γ_3 (cm^{-1})	31.8 ± 3.5	23.4 ± 1.6	
ω_4 (cm^{-1})	2499 ± 6	2488 ± 3	O-D (D_2O)
A_4 (a.u.)	24.2 ± 5.7	27.5 ± 1.7	
Γ_4 (cm^{-1})	55.5 ± 1.7	66.8 ± 2.0	

Table 2.1. Fitting parameters and assignments for the spectra in Figure 2.2.

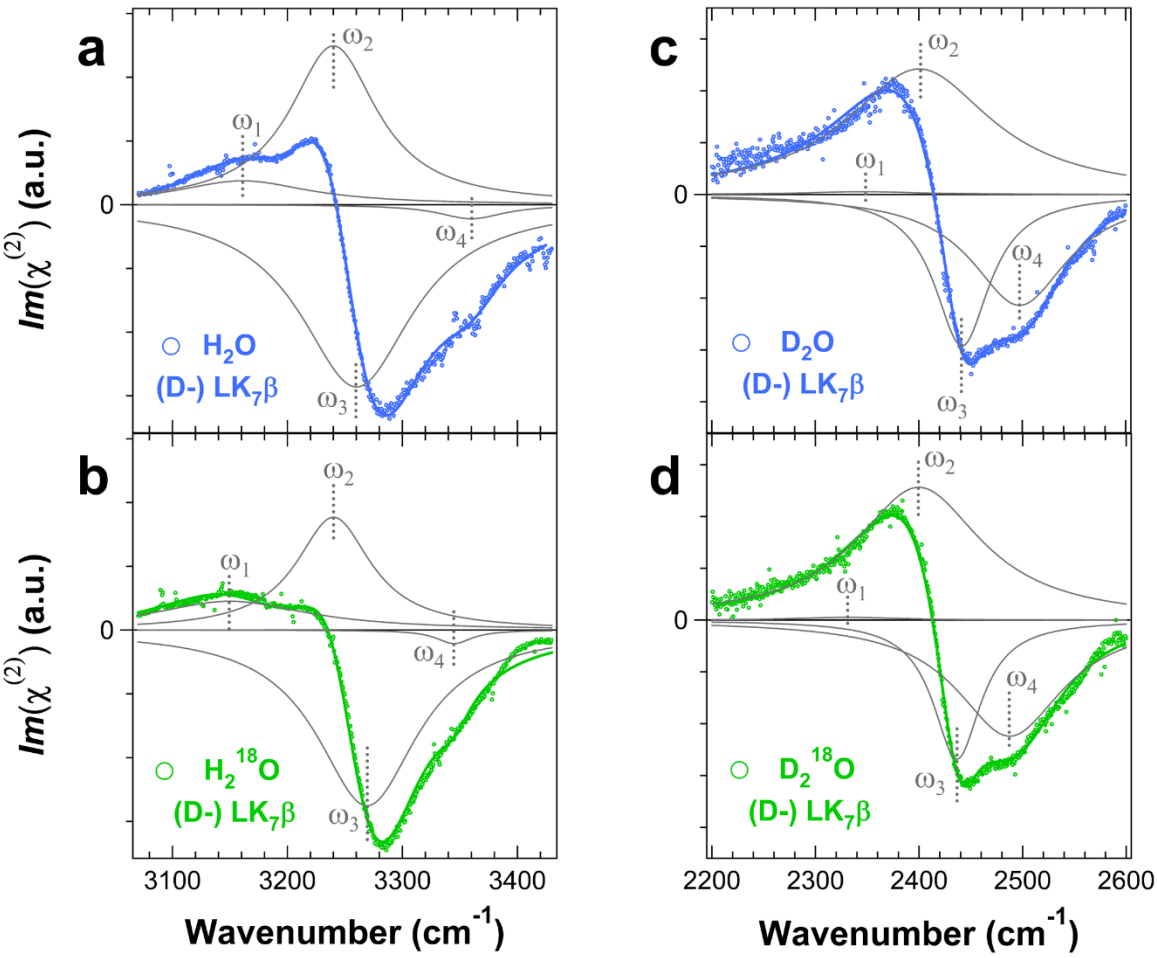

Figure 2.3. The phase-resolved chiral SFG spectra of chiral water superstructures around (D-) $LK_7\beta$. (Left) O-H/N-H spectral region of (D-) $LK_7\beta$ prepared in (a) H_2O and (b) $H_2{}^{18}O$. (Right) The O-D/N-D spectral region of (D-) $LK_7\beta$ prepared in (c) D_2O and (d) $D_2{}^{18}O$. The component peaks based on the fitting parameters are shown in gray, and the sum of these peaks are shown as blue or green solid lines. The labels of vibrational modes correspond to the numbering in Table 2.2. All spectra were collected with the chiral *psp* polarization.

Parameter	(Figure 2.3a) (D-) $LK_7\beta$-H_2O	(Figure 2.3b) (D-) $LK_7\beta$-$H_2{}^{18}O$	Assignments
ω_1 (cm^{-1})	3161 ± 4	3150 ± 5	O-H (H_2O)
A_1 (a.u.)	16.0 ± 3.6	15.0 ± 2.0	
Γ_1 (cm^{-1})	71.3 ± 2.3	82.7 ± 6.2	
ω_2 (cm^{-1})	3240 ± 6	3240 ± 1	N-H (D-$LK_7\beta$)
A_2 (a.u.)	69.1 ± 3.5	28.0 ± 2.4	
Γ_2 (cm^{-1})	46.2 ± 1.1	39.6 ± 1.7	
ω_3 (cm^{-1})	3260 ± 6	3269 ± 1	N-H (D-$LK_7\beta$)
A_3 (a.u.)	-102.9 ± 2.9	-59.1 ± 1.4	
Γ_3 (cm^{-1})	60.0 ± 1.7	53.7 ± 1.3	
ω_4 (cm^{-1})	3360 ± 4	3345 ± 1	O-H (H_2O)
A_4 (a.u.)	-4.6 ± 1.2	-1.6 ± 0.4	
Γ_4 (cm^{-1})	35.0 ± 4.6	19.0 ± 3.2	
Parameter	(Figure 2.3c) (D-) $LK_7\beta$-D_2O	(Figure 2.3d) (D-) $LK_7\beta$-$D_2{}^{18}O$	Assignments
ω_1 (cm^{-1})	2346 ± 0*	2331 ± 0*	O-D (D_2O)
A_1 (a.u.)	1.0 ± 0*	1.0 ± 0*	
Γ_1 (cm^{-1})	62.0 ± 0*	62.0 ± 0*	
ω_2 (cm^{-1})	2401 ± 4	2400 ± 2	N-D (D-$LK_7\beta$)
A_2 (a.u.)	61.9 ± 3.8	57.2 ± 2.2	
Γ_2 (cm^{-1})	85.3 ± 1.2	74.4 ± 0.9	
ω_3 (cm^{-1})	2442 ± 1	2437 ± 4	N-D (D-$LK_7\beta$)
A_3 (a.u.)	-28.5 ± 2.3	-22.7 ± 2.3	
Γ_3 (cm^{-1})	32.5 ± 1.2	28.2 ± 1.2	
ω_4 (cm^{-1})	2497 ± 4	2487 ± 1	O-D (D_2O)
A_4 (a.u.)	-36.4 ± 4.4	-42.2 ± 1.8	
Γ_4 (cm^{-1})	56.7 ± 1.4	62.8 ± 1.5	

Table 2.2. Fitting parameters and assignments for the spectra in Figure 2.3. *Parameter fixed during fitting.

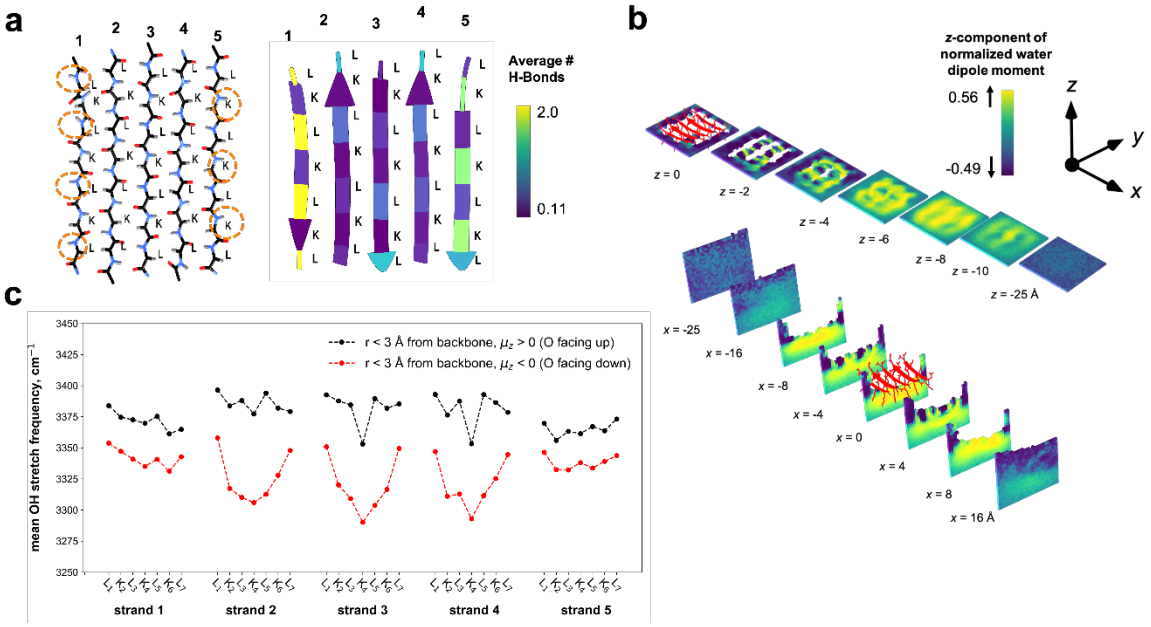

Figure 2.4. Hydrogen-bond and water structure analyses of the (L-) $LK_7\beta$ homopentamer at the vacuum–water interface obtained from 1 μs of MD sampling using the AMBER ff14SB force field for the protein and the TIP4P-Ew water potential. (a, left) $LK_7\beta$ backbone atoms show the asymmetry of the N-H and C = O groups on opposite sides of the β-sheet. (a, right) The average number of hydrogen-bonding interactions between the protein and water molecules depicted by residue. (b) The average z-components of water dipole moment vectors extending below the protein surface (top) and from side to side (bottom) with a grid resolution of 1 Å. (c) The infrared response of O-H stretch of water within 3 Å of the peptide backbone for every residue. The IR response was calculated for water molecules where the average z-component of the dipole was oriented either up (black) or down (red) relative to the vacuum–water interface.

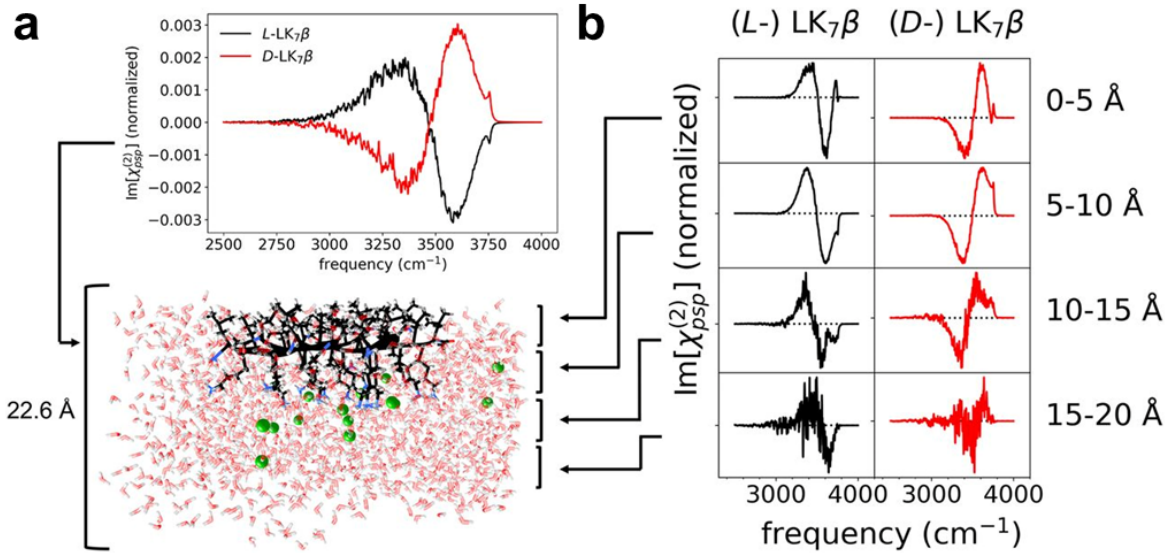

Figure 2.5. The simulated chiral phase-resolved SFG response from the O-H stretch of water surrounding $LK_7\beta$. (A) The *psp* response of water surrounding (L-) $LK_7\beta$ (black) and its enantiomer, (D-) $LK_7\beta$ (red), showing the effect of chirality on the response phase. (B) The *psp* response of consecutive layers of water starting at the air–vacuum interface, again showing the effect of protein chirality on the water SFG response. The chiral signal persists at least as far into the water as the lysine residues, and possibly beyond. The SFG response was calculated from 10,000,000 configurations obtained from 100 ns of MD sampling. The green spheres correspond to chloride ions.

References

1. Bellissent-Funel, M.-C.; Hassanali, A.; Havenith, M.; Henchman, R.; Pohl, P.; Sterpone, F.; van der Spoel, D.; Xu, Y.; Garcia, A. E. Water Determines the Structure and Dynamics of Proteins. *Chem. Rev.* **2016,** *116*, 7673-7697.
2. Perets, E. A.; Yan, E. C. Y. Chiral Water Superstructures around Antiparallel β-Sheets Observed by Chiral Vibrational Sum Frequency Generation Spectroscopy. *J. Phys. Chem. Lett.* **2019,** *10*, 3395-3401.
3. McDermott, M. L.; Vanselous, H.; Corcelli, S. A.; Petersen, P. B. DNA's Chiral Spine of Hydration. *ACS Cent. Sci.* **2017,** *3*, 708-714.
4. Perets, E. A.; Yan, E. C. Y. The H2O Helix: The Chiral Water Superstructure Surrounding DNA. *ACS Cent. Sci.* **2017,** *3*, 683-685.
5. Simpson, G. J. Molecular Origins of the Remarkable Chiral Sensitivity of Second-Order Nonlinear Optics. *ChemPhysChem* **2004,** *5*, 1301-1310.
6. Simpson, G. J., *Nonlinear Optical Polarization Analysis in Chemistry and Biology*. Cambridge University Press: Cambridge, 2017.
7. DeGrado, W. F.; Lear, J. D. Induction of peptide conformation at apolar water interfaces. 1. A study with model peptides of defined hydrophobic periodicity. *J. Am. Chem. Soc.* **1985,** *107*, 7684-7689.
8. Phillips, D. C.; York, R. L.; Mermut, O.; McCrea, K. R.; Ward, R. S.; Somorjai, G. A. Side Chain, Chain Length, and Sequence Effects on Amphiphilic Peptide Adsorption at Hydrophobic and Hydrophilic Surfaces Studied by Sum-Frequency Generation Vibrational Spectroscopy and Quartz Crystal Microbalance. *J. Phys. Chem. C* **2007,** *111*, 255-261.
9. Wang, Z.; Fu, L.; Yan, E. C. Y. C–H Stretch for Probing Kinetics of Self-Assembly into Macromolecular Chiral Structures at Interfaces by Chiral Sum Frequency Generation Spectroscopy. *Langmuir* **2013,** *29*, 4077-4083.
10. Fu, L.; Xiao, D.; Wang, Z.; Batista, V. S.; Yan, E. C. Y. Chiral Sum Frequency Generation for In Situ Probing Proton Exchange in Antiparallel β-Sheets at Interfaces. *J. Am. Chem. Soc.* **2013,** *135*, 3592-3598.
11. Perets, E. A.; Videla, P. E.; Yan, E. C. Y.; Batista, V. S. Chiral inversion of amino acids in antiparallel beta sheets at interfaces probed by vibrational sum frequency generation spectroscopy. *J. Phys. Chem. B* **2019,** *123*, 5769−5781.
12. Hu, X.-H.; Fu, L.; Hou, J.; Zhang, Y.-N.; Zhang, Z.; Wang, H.-F. N–H Chirality in Folded Peptide LK7β Is Governed by the Cα–H Chirality. *J. Phys. Chem. Lett.* **2020,** *11*, 1282-1290.
13. Milton, R. C. d.; Milton, S. C. F.; Kent, S. B. H. Total Chemical Synthesis of a D-Enzyme: The Enantiomers of HIV-1 Protease Show Reciprocal Chiral Substrate Specificity. *Science* **1992,** *256*, 1445-1448.
14. Kent, S. B. H. Novel protein science enabled by total chemical synthesis. *Protein Sci.* **2019,** *28*, 313-328.
15. Breen, N. F.; Weidner, T.; Li, K.; Castner, D. G.; Drobny, G. P. A Solid-State Deuterium NMR and Sum-Frequency Generation Study of the Side-Chain Dynamics of Peptides Adsorbed onto Surfaces. *J. Am. Chem. Soc.* **2009,** *131*, 14148-14149.

16. Weidner, T.; Breen, N. F.; Li, K.; Drobny, G. P.; Castner, D. G. Sum frequency generation and solid-state NMR study of the structure, orientation, and dynamics of polystyrene-adsorbed peptides. *Proc. Natl. Acad. Sci. USA* **2010,** *107*, 13288-13293.
17. Donovan, M. A.; Lutz, H.; Yimer, Y. Y.; Pfaendtner, J.; Bonn, M.; Weidner, T. LK peptide side chain dynamics at interfaces are independent of secondary structure. *Phys. Chem. Chem. Phys.* **2017,** *19*, 28507-28511.
18. Hu, X.-H.; Wei, F.; Wang, H.; Wang, H.-F. α-Quartz Crystal as Absolute Intensity and Phase Standard in Sum-Frequency Generation Vibrational Spectroscopy. *J. Phys. Chem. C* **2019,** *123*, 15071-15086.
19. 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.
20. Fu, L.; Chen, S.-L.; Wang, H.-F. Validation of Spectra and Phase in Sub-1 cm–1 Resolution Sum-Frequency Generation Vibrational Spectroscopy through Internal Heterodyne Phase-Resolved Measurement. *J. Phys. Chem. B* **2016,** *120*, 1579-1589.
21. Chen, S.-L.; Fu, L.; Gan, W.; Wang, H.-F. Homogeneous and inhomogeneous broadenings and the Voigt line shapes in the phase-resolved and intensity sum-frequency generation vibrational spectroscopy. *J. Chem. Phys.* **2016,** *144*, 034704.
22. Horn, H. W.; Swope, W. C.; Pitera, J. W.; Madura, J. D.; Dick, T. J.; Hura, G. L.; Head-Gordon, T. Development of an improved four-site water model for biomolecular simulations: TIP4P-Ew. *J. Chem. Phys.* **2004,** *120*, 9665-9678.
23. Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theory Comput.* **2015,** *11*, 3696-3713.
24. Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. *J. Chem. Phys.* **1993,** *98*, 10089-10092.
25. Eastman, P.; Swails, J.; Chodera, J. D.; McGibbon, R. T.; Zhao, Y.; Beauchamp, K. A.; Wang, L.-P.; Simmonett, A. C.; Harrigan, M. P.; Stern, C. D., et al. OpenMM 7: Rapid development of high performance algorithms for molecular dynamics. *PLoS Comput. Biol.* **2017,** *13*, e1005659.
26. Barth, A.; Zscherp, C. What vibrations tell about proteins. *Quat. Rev. Biophys.* **2002,** *35*, 369-430.
27. Barth, A. Infrared spectroscopy of proteins. *Biochim. Biophys. Acta BBA - Bioenerg.* **2007,** *1767*, 1073-1101.
28. Tan, J.; Zhang, J.; Li, C.; Luo, Y.; Ye, S. Ultrafast energy relaxation dynamics of amide I vibrations coupled with protein-bound water molecules. *Nat. Commun.* **2019,** *10*, 1010.
29. Jung, S.-Y.; Lim, S.-M.; Albertorio, F.; Kim, G.; Gurau, M. C.; Yang, R. D.; Holden, M. A.; Cremer, P. S. The Vroman Effect: A Molecular Level Description of Fibrinogen Displacement. *J. Am. Chem. Soc.* **2003,** *125*, 12782-12786.
30. Khatib, R.; Backus, E. H. G.; Bonn, M.; Perez-Haro, M.-J.; Gaigeot, M.-P.; Sulpizi, M. Water orientation and hydrogen-bond structure at the fluorite/water interface. *Sci. Reports* **2016,** *6*, 24287.
31. Pezzotti, S.; Serva, A.; Gaigeot, M.-P. 2D-HB-Network at the air-water interface: A structural and dynamical characterization by means of ab initio and classical molecular dynamics simulations. *J. Chem. Phys.* **2018,** *148*, 174701.

32. Heyden, M. Heterogeneity of water structure and dynamics at the protein-water interface. *J. Chem. Phys.* **2019,** *150*, 094701.
33. Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. The missing term in effective pair potentials. *J. Phys. Chem. A* **1987,** *91*, 6269-6271.
34. Hosseinpour, S.; Roeters, S. J.; Bonn, M.; Peukert, W.; Woutersen, S.; Weidner, T. Structure and Dynamics of Interfacial Peptides and Proteins from Vibrational Sum-Frequency Generation Spectroscopy. *Chem. Rev.* **2020,** *120*, 3420-3465.
35. Fu, L.; Liu, J.; Yan, E. C. Y. Chiral Sum Frequency Generation Spectroscopy for Characterizing Protein Secondary Structures at Interfaces. *J. Am. Chem. Soc.* **2011,** *133*, 8094-8097.
36. Ishiyama, T.; Morita, A. Nuclear Quantum Effect on the χ(2) Band Shape of Vibrational Sum Frequency Generation Spectra of Normal and Deuterated Water Surfaces. *J. Phys. Chem. Lett.* **2019,** *10*, 5070-5075.
37. Kopka, M. L.; Fratini, A. V.; Drew, H. R.; Dickerson, R. E. Ordered water structure around a B-DNA dodecamer: A quantitative study. *J. Mol. Biol.* **1983,** *163*, 129-146.
38. Duboué-Dijon, E.; Fogarty, A. C.; Hynes, J. T.; Laage, D. Dynamical Disorder in the DNA Hydration Shell. *J. Am. Chem. Soc.* **2016,** *138*, 7610-7620.
39. Laage, D.; Elsaesser, T.; Hynes, J. T. Water Dynamics in the Hydration Shells of Biomolecules. *Chem. Rev.* **2017,** *117*, 10694-10725.
40. Pauling, L., *The Nature of the Chemical Bond*. Cornell University Press: Ithaca, NY, 1960; Vol. 260.
41. Kuriyan, J.; Konforti, B.; Wemmer, D., *The Molecules of Life: Physical and Chemical Principles*. WW Norton & Company: New York, 2012.
42. Matthews, B. W. How planar are planar peptide bonds? *Protein Sci.* **2016,** *25*, 776-777.
43. Berkholz, D. S.; Driggers, C. M.; Shapovalov, M. V.; Dunbrack, R. L.; Karplus, P. A. Nonplanar peptide bonds in proteins are common and conserved but not biased toward active sites. *Proc. Natl. Acad. Sci. USA* **2012,** *109*, 449-453.
44. Carr, J. K.; Wang, L.; Roy, S.; Skinner, J. L. Theoretical Sum Frequency Generation Spectroscopy of Peptides. *J. Phys. Chem. B* **2015,** *119*, 8969-8983.
45. Fersht, A. R.; Shi, J.-P.; Knill-Jones, J.; Lowe, D. M.; Wilkinson, A. J.; Blow, D. M.; Brick, P.; Carter, P.; Waye, M. M. Y.; Winter, G. Hydrogen bonding and biological specificity analysed by protein engineering. *Nature* **1985,** *314*, 235-238.
46. Sheu, S.-Y.; Yang, D.-Y.; Selzle, H. L.; Schlag, E. W. Energetics of hydrogen bonds in peptides. *Proc. Natl. Acad. Sci. USA* **2003,** *100*, 12683-12687.
47. Fleming, P. J.; Rose, G. D. Do all backbone polar groups in proteins form hydrogen bonds? *Protein Sci.* **2005,** *14*, 1911-1917.
48. Blackmond, D. G. The origin of biological homochirality. *Cold Spring Harb. Perspect. Biol.* **2010,** *2*, a002147.
49. Ketchem, R. R.; Hu, W.; Cross, T. A. High-Resolution Conformation of Gramicidin A in a Lipid Bilayer by Solid-State NMR. *Science* **1993,** *261*, 1457-1460.
50. Finking, R.; Marahiel, M. A. Biosynthesis of Nonribosomal Peptides. *Ann. Rev. Microbiol.* **2004,** *58*, 453-488.
51. Durani, S. Protein Design with l- and d-α-Amino Acid Structures as the Alphabet. *Acc. Chem. Res.* **2008,** *41*, 1301-1308.

52. Yamagata, Y. A hypothesis for the asymmetric appearance of biomolecules on earth. *J. Theor. Biol.* **1966,** *11*, 495-498.
53. Soai, K.; Shibata, T.; Morioka, H.; Choji, K. Asymmetric autocatalysis and amplification of enantiomeric excess of a chiral molecule. *Nature* **1995,** *378*, 767-768.
54. Quack, M. How Important is Parity Violation for Molecular and Biomolecular Chirality? *Angew. Chem. Int. Ed.* **2002,** *41*, 4618-4630.
55. McBride, J. M.; Tully, J. C. Did life grind to a start? *Nature* **2008,** *452*, 161-162.
56. Noorduin, W. L.; Izumi, T.; Millemaggi, A.; Leeman, M.; Meekes, H.; Van Enckevort, W. J. P.; Kellogg, R. M.; Kaptein, B.; Vlieg, E.; Blackmond, D. G. Emergence of a Single Solid Chiral State from a Nearly Racemic Amino Acid Derivative. *J. Am. Chem. Soc.* **2008,** *130*, 1158-1159.
57. Hawbaker, N. A.; Blackmond, D. G. Energy threshold for chiral symmetry breaking in molecular self-replication. *Nat. Chem.* **2019,** *11*, 957-962.
58. Klussmann, M.; White, A. J. P.; Armstrong, A.; Blackmond, D. G. Rationalization and Prediction of Solution Enantiomeric Excess in Ternary Phase Systems. *Angew. Chem. Int. Ed.* **2006,** *45*, 7985-7989.
59. Klussmann, M.; Izumi, T.; White, A. J. P.; Armstrong, A.; Blackmond, D. G. Emergence of Solution-Phase Homochirality via Crystal Engineering of Amino Acids. *J. Am. Chem. Soc.* **2007,** *129*, 7657-7660.

Appendix to Chapter 2

Supplementary Computational Methods

System preparation

A single $LK_7\beta$ strand with an N-terminal acetyl group and a C-terminal amide (-NH_2) was generated using UCSF Chimera. VMD was used to repeatedly copy and translate the strand to generate a five-strand antiparallel β-sheet, with the strands roughly 3.8 Å apart. The β-sheet was aligned in the *xy* plane with all leucine residues pointing in the $+z$ direction. This system was solvated with TIP4P-Ew water molecules.[1] Subsequently, water molecules with a *z*-coordinate greater than zero were removed to create a vacuum-water interface with leucine residues pointed into the vacuum and lysine residues pointed into the solvent. Then Cl^- ions were added to the water to neutralize the system. For all simulations, the protein and ions were modelled with the AMBER ff14sb force field.[2] A large box was used with at least ~50 Å of solvent on each side of the protein (a 125×125×125 Å box, half filled, half vacuum) to minimize edge artifacts in the grid-based calculations of water orientation.

Equilibration and Production Molecular Dynamics

The protein was restrained with force constants of 200 kcal/mol • $Å^{-2}$, and the solvent and ions were energy-minimized and then equilibrated for 500 ps in an NVT ensemble at 300 K. Next, the energy of the entire system was minimized. Then the system was gradually heated from 0 K to 298 K in seven cycles, where each cycle contained 10 ps

of temperature ramping and 50 ps of constant temperature MD in an NVT ensemble. The system was then equilibrated in an NVT ensemble for 6 ns at 298 K. All NVT equilibration steps used a Langevin integrator with a friction coefficient of 1 ps^{-1}. Long-range electrostatics were treated with a particle-mesh Ewald approach, and Lennard-Jones interactions were cut off at 14 Å. The water molecule geometry was kept rigid with the SETTLE algorithm,[3] and the protein covalent bonds involving hydrogen were constrained with the SHAKE algorithm.[4] A 1 fs time step was used for the MD. These same methods and parameters were used for all production runs. NPT equilibration was not performed as it would have severely disrupted the vacuum-water interface and the structure of the amphiphilic β-sheet.

Trajectory Analysis – Hydrogen Bonding

The HydrogenBondAnalysis tool in the MDAnalysis library[5, 6] was used to identify all hydrogen bonds involving the $LK_7\beta$ backbone. Hydrogen-bonding interactions were defined to have a donor-acceptor distance of less than 3.2 Å and a donor-hydrogen-acceptor bond angle greater than 135°. The average number of hydrogen bonds between the protein backbone and water was determined for each residue.

Trajectory Analysis –Water Orientation

For each configuration, the protein was centered in the unit cell, and then the simulation box was translated and rotated in a way that minimized the root-mean-square deviation of the protein with the initial protein structure serving as the reference. For each configuration, the volume around the protein (30 Å each way) was divided into a three-

dimensional grid with a resolution of 1 Å. The water dipole moment vector and the number of water molecules, as defined by the oxygen position, at each grid point were averaged over all configurations sampled. The grid points with the average number of water molecules below 10% of the median were removed from the analysis. The results were plotted along with an average LK$_7$β structure using UCSF Chimera's BILD drawing language. The images were rendered with POV-Ray 3.6. All analysis code was written in Python using the MDAnalysis, MDTraj, Numba, Matplotlib, and Pandas libraries.[5-10]

Average Infrared Response Frequency

IR spectra were calculated by adapting a semiclassical electric field mapping method developed by Skinner and coworkers.[11, 12] In this approach, IR spectra are derived from a modified Fourier transform of a dipole autocorrelation function:

$$\chi^{(2)}_{IR(k)} \approx \left\langle \sum_{a}^{N_{OH}} \frac{\left(\sum_{b}^{N_{OH}} \mu_{bk} U_{ba} \right)^2}{\lambda_a - \omega - \frac{i}{2\tau}} \right\rangle \tag{3}$$

Here k is the spatial dimension (x, y, or z, which are all equivalent for IR), the indices a and b denote a particular O-H bond, and λ_a and U_{ba} and are the eigenvalues and elements of the eigenvectors, respectively, of the coupling matrix, which is defined as:

$$\Omega_{ij} = \begin{cases} \omega_{\text{stretch}}(E) & i = j \\ \omega_{\text{intramolecular}}(E) & i - j = 1,\ j \text{ even or } j - i = 1,\ i \text{ even} \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

Each O-H group's vibrational transition dipole derivative (μ', defined along the OH stretch), bond stretch associated with vibrational excitation (Δx), vibrational stretch frequency (ω_{stretch}), and intramolecular coupling ($\omega_{\text{intramolecular}}$) were obtained from the local electric field at each hydrogen atom via empirical maps developed by the Skinner group for TIP4P water.[13] These maps link these quantities to the component of the electric field along the O-H bond at each hydrogen atom. The dipole derivative, bond length stretch, and bond unit vector $\hat{u}$ were used to approximate the dipole vector for each O-H bond:

$$\mu = \mu'(E)\Delta x_{OH}(E)\hat{u} \tag{5}$$

Point charges up to 7.831 Å from each O-H hydrogen were included in the electric field calculation, but the current molecule itself was not included.[12] The decay constant τ was set to 1.3 ps to approximate the lifetime broadening of the O-H stretch. In the expression for the susceptibility given in equation 1, the angular brackets indicate averaging over conformations, and the dipole vectors for the O-H bonds as well as the coupling matrix depend on the conformation.

To obtain the IR response of water molecules close to a particular residue and with a certain orientation, the appropriate molecules were selected using a combination of the MDAnalysis selection language and additional code.[5, 6] These molecules formed the coupling matrix $\mathbf{\Omega}$, and only these molecules' dipoles were calculated for a given frame. The rest of the system, whether protein, water, or ions, was treated as a set of point charges contributing to the local electric field. All water molecules in the selection also contributed to the local field, except the molecule at which the field was evaluated. The IR responses calculated from 100,000 configurations obtained from 1 μs of total MD sampling were averaged to obtain the final IR spectra with a frequency resolution of 1 cm^{-1}. The mean IR

response frequency was determined by a simple average of the frequency weighted by peak height.

Chiral Sum Frequency Generation Response

Components of the SFG response tensor were also calculated with a similar method as for the IR spectra (previous section).[11-13] Components of the second-order response tensor were calculated using:

$$\chi^{(2)}_{ijk}(\omega) \approx \left\langle \sum_{a}^{N_{OH}} \frac{\left(\sum_{b}^{N_{OH}} \alpha_{b(ij)} U_{ba} \right)\left(\sum_{b}^{N_{OH}} \mu_{b(k)} U_{ba} \right)}{\lambda_a - \omega - \frac{i}{2\tau}} \right\rangle \tag{6}$$

where α is the polarizability tensor, *i*, *j*, and *k* are spatial dimensions, and all other quantities are as defined in the previous section (equation 3). The polarizability for each O-H bond in all water molecules was treated as an uncoupled response. The polarizability was adjusted based on the OH bonds' orientations using:

$$\alpha_{ij} = \Delta x_{OH}((\alpha_{\parallel}' - \alpha_{\perp}')(u_i u_j) + \alpha_{\perp}'(\hat{i} \bullet \hat{j})) \tag{7}$$

where u_i and u_j are the i^{th} and j^{th} components of the bond unit vector and $\hat{i}$ and $\hat{j}$ are unit vectors representing the *x*, *y*, or *z* direction. Δx_{OH} is the change in bond length due to vibrational excitation and is obtained from electric field mappings. $\alpha_{\parallel}'$ is the transition polarizability derivative parallel to the O-H stretch, while $\alpha_{\perp}'$ is the perpendicular polarizability derivative. The $\alpha_{\parallel}'$ and $\alpha_{\perp}'$ quantities are assumed to be independent of the

field E, and the ratio $\alpha'_{\parallel}/\alpha'_{\perp}$ is assumed to be 5.6.[12] Frequencies and intramolecular couplings were obtained as above, while intermolecular couplings were calculated using a transition dipole coupling approximation[12]:

$$\omega_{ij} = \Delta x_i \Delta x_j \frac{\mu'_i \mu'_j ((\hat{u}_i \bullet \hat{u}_j) - 3(\hat{u}_i \bullet \hat{r}_{ij})(\hat{u}_j \bullet \hat{r}_{ij}))}{r_{ij}^3} \tag{8}$$

Here $\hat{u}$ is an O-H bond unit vector and $\hat{r}$ is the unit vector in the direction between the two OH groups. The effective location of the O-H dipole is defined as 0.67 Å along the OH bond toward the H.[13]

Intermolecular couplings were not included in the calculations on subsets of the solution (Figure 2.5b) because the precise set of water molecules in the selection changed for every frame, and only water molecules in the selection can be part of the coupling matrix (equation 2). In any case, intermolecular couplings have a very minor effect on the spectra. For reference, the typical O-H frequency is around 3400 cm^{-1}, the typical intramolecular couplings have magnitudes of around 10 cm^{-1}, and the typical intermolecular couplings are around 1 cm^{-1}.

The MDAnalysis library was used to select subsets of the system. Water molecules not part of a selection still contributed to the electric field experienced by the selected waters (see the section **Average Infrared Response Frequency**).

Chiral (psp) Signal from an MD Trajectory

The theoretical foundation of chiral SFG assumes that systems are isotropic upon rotation about the *z*-axis,[14] but even quite long simulations may not reach this limit. The

Skinner group introduced an elegant approach for addressing this issue with a rotating reference frame.[15]

The observed SFG response is a sum of multiple response tensor elements[16]:

$$\chi_{eff}^{(2)} = \sum_{i=1}^{3}\sum_{j=1}^{3}\sum_{k=1}^{3}(\hat{u}_i L_{ii}) \bullet \chi_{ijk}^{(2)} \bullet (\hat{v}_j L_{jj})(\hat{w}_k L_{kk}) \tag{9}$$

where $\hat{u}$, $\hat{v}$, and $\hat{w}$ are unit vectors describing the orientation of the input and output beams, and the components L_{ij} are Fresnel factors. Assuming the input beams move from right to left, for a *psp* setup:

$$\begin{aligned} \hat{u} &= [\ -\cos(\theta) \quad 0 \quad \sin(\theta)\ ] \\ \hat{v} &= [\ 0 \quad 1 \quad 0\ ] \\ \hat{w} &= [\ \cos(\theta) \quad 0 \quad \sin(\theta)\ ] \end{aligned} \tag{10}$$

and:

$$\begin{aligned} \chi_{eff}^{(2)} = & -L_{xx}^2 L_{yy} \cos^2(\theta)\chi_{xyx}^{(2)} \\ & - L_{xx} L_{yy} L_{zz} \cos(\theta)\sin(\theta)\chi_{xyz}^{(2)} \\ & + L_{xx} L_{yy} L_{zz} \cos(\theta)\sin(\theta)\chi_{zyx}^{(2)} \\ & + L_{yy} L_{zz}^2 \sin^2(\theta)\chi_{zyz}^{(2)} \end{aligned} \tag{11}$$

Eliminating terms equal to zero due to surface rotational symmetry[14] and transforming from an imaginary rotating lab frame to the simulation box frame leads to the following expression[15]:

$$\chi_{psp}^{(2)} = \frac{1}{2} L_{xx} L_{yy} L_{zz} (-\chi_{xyz}^{(2)} + \chi_{yxz}^{(2)} - \chi_{zxy}^{(2)} + \chi_{zyx}^{(2)}) \cos(\theta)\sin(\theta) \tag{12}$$

If the OH polarizability tensor is assumed to be symmetric, which is a reasonable approximation given the cylindrical bond symmetry, then $\chi^{(2)}_{xyz} = \chi^{(2)}_{yxz}$. In this case, the expression simplifies to:

$$\chi^{(2)}_{psp} \propto \chi^{(2)}_{zyx} - \chi^{(2)}_{zxy} \tag{13}$$

This approach removes the effects of *xy*-plane anisotropy in the simulation to give the correct chiral SFG signal.

Assumptions of the Computational Model

In applying the MD results to help interpret the experimental observations, we assume that the $LK_7\beta$ structure and the water structure are similar at the air-water interface and on glass or quartz surfaces. This assumption is supported by the experimental results that homodyne chiral SFG spectra obtained at the air-water interface previously[17] and on a glass surface in the current study (**Figure 2.1**, top) are comparable in the amide I, O-H/N-H stretching, and O-D/N-D stretching regions. In addition, the assumption also aligns with the simulated chiral SFG spectra. The simulated spectra show that only a few layers of water molecules can account for the chiral SFG response and these few layers of water molecules are likely present in the hydrated samples on quartz surfaces.

Theoretical Background of Phase-Resolved Chiral Sum Frequency Generation

In the chiral SFG measurements,[18] infrared and visible beams at frequencies, ω_{IR} and ω_{vis}, were overlapped in space and time to generate a sum frequency optical response at $\omega_{IR} + \omega_{vis} = \omega_{SFG}$. The chiral spectra were obtained using the *psp* configuration (*p*-polarized sum frequency, *s*-polarized visible, *p*-polarized IR). For the *psp* polarization combination, the right-handed α-quartz surface gives maximum non-resonant intensity along the +*y* ($\varphi = 90^\circ$) and -*y* axes ($\varphi + 180^\circ = 270^\circ$). The +*y* and -*y* axes were established relative to the right-handed α-quartz crystal +*x* axis ($\varphi = 0^\circ$) and the laboratory frame (Figure A2.1), following definitions in the literature,[19] and were experimentally determined here using an adsorbed layer of POPC lipid (see Methods in Chapter 2).

Under the dipole approximation, the chiral SFG signal is described by the twenty-seven element (3 × 3 × 3) second-order susceptibility tensor $\chi^{(2)}$. In the phase-resolved chiral SFG measurement of a hydrated thin-film of the antiparallel β-sheet $LK_7\beta$ (Figure 2.1, bottom), signal intensity is proportional to the square modulus of $\chi^{(2)}$:

$$I_{SFG} \propto |\chi^{(2)}|^2 + 2\chi^{(2)}_{quartz} \bullet \chi^{(2)}_{LK7\beta,\mathrm{Im}} + |\chi^{(2)}_{LK7\beta}|^2 \qquad (14)$$

where I_{SFG} is the intensity of the phase-resolved chiral SFG measurement, $\chi^{(2)}_{quartz}$ is the non-resonant optical response of the quartz surface, and $\chi^{(2)}_{LK7\beta}$ is the resonant optical response of the adsorbed β-sheet. The $\chi^{(2)}_{quartz}$ is real because it is off-resonance, whereas the resonant optical response of the sample $\chi^{(2)}_{LK7\beta}$ has both real and imaginary components. Moreover, due to the symmetry of the right-handed α-quartz crystal, $\chi^{(2)}_{quartz,\varphi} =$

$-\chi^{(2)}_{quartz,\varphi+180°}$. The SFG intensity of $LK_7\beta$ on quartz (measured at φ and φ + 180°) is normalized by the non-resonant intensity of the clean quartz crystal (see Experimental Methods) measured at φ = 90°, $\left|\chi^{(2)}_{quartz,\varphi}\right|^2$.

$$I_{\phi} = \frac{|\chi^{(2)}_{\phi}|^2}{|\chi^{(2)}_{quartz}|^2} = 1 + 2 \bullet \mathrm{Im}[\chi^{(2)}_{LK7\beta}] + \frac{|\chi^{(2)}_{LK7\beta}|^2}{|\chi^{(2)}_{quartz,\phi}|^2} \quad (15)$$

and,

$$I_{\phi+180°} = \frac{|\chi^{(2)}_{\phi+180°}|^2}{|\chi^{(2)}_{quartz}|^2} = 1 - 2 \bullet \mathrm{Im}[\chi^{(2)}_{LK7\beta}] + \frac{|\chi^{(2)}_{LK7\beta}|^2}{|\chi^{(2)}_{quartz,\phi}|^2} \quad (16)$$

Thus, we measured $I_{\varphi=90°}$ and $I_{\varphi+180°=270°}$ in the *psp* polarization configuration and obtained the phase-sensitive chiral SFG optical response:

$$\mathrm{Im}[\chi^{(2)}_{LK7\beta}] = \frac{I_{\phi} - \mathrm{I}_{\phi+180°}}{4} \quad (17)$$

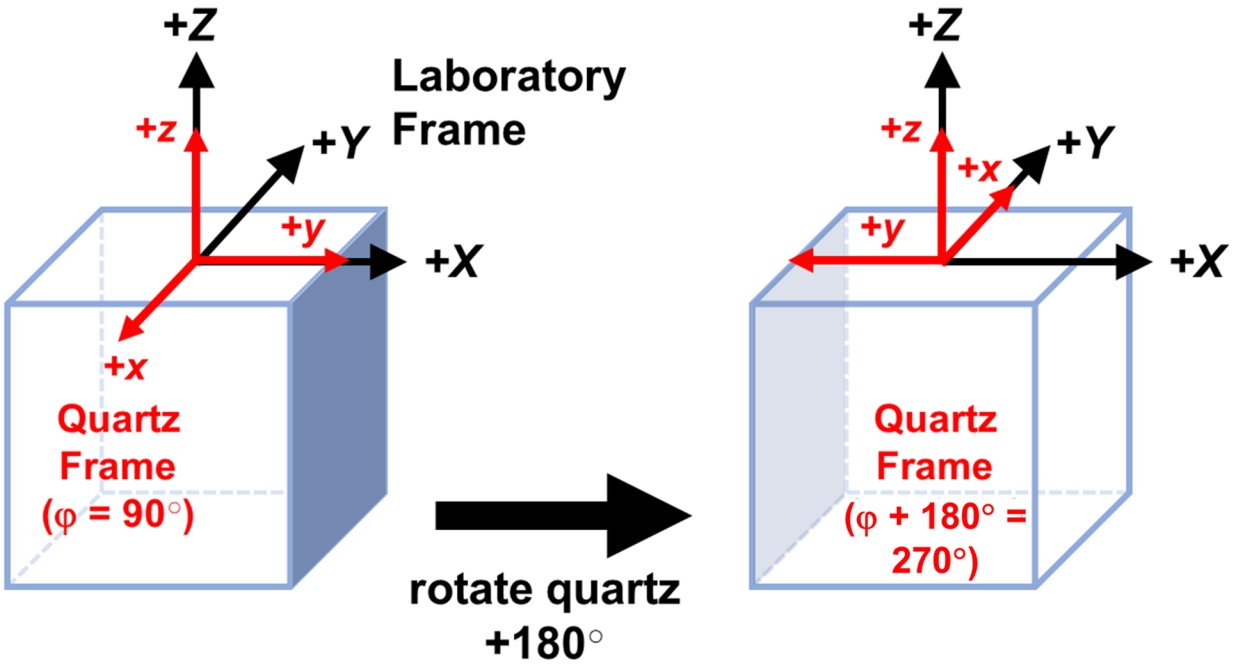

Figure A2.1. Quartz frame relative to laboratory frame. For phase-resolved chiral SFG measurements using the *psp* configuration, sample measurements are acquired at the φ = 90° and φ + 180° = 270° configurations. The SFG signal propagates from the quartz surface in the +*X*/*Z* laboratory frame for our measurements. Figure adapted based on Hu, *et al*.[20]

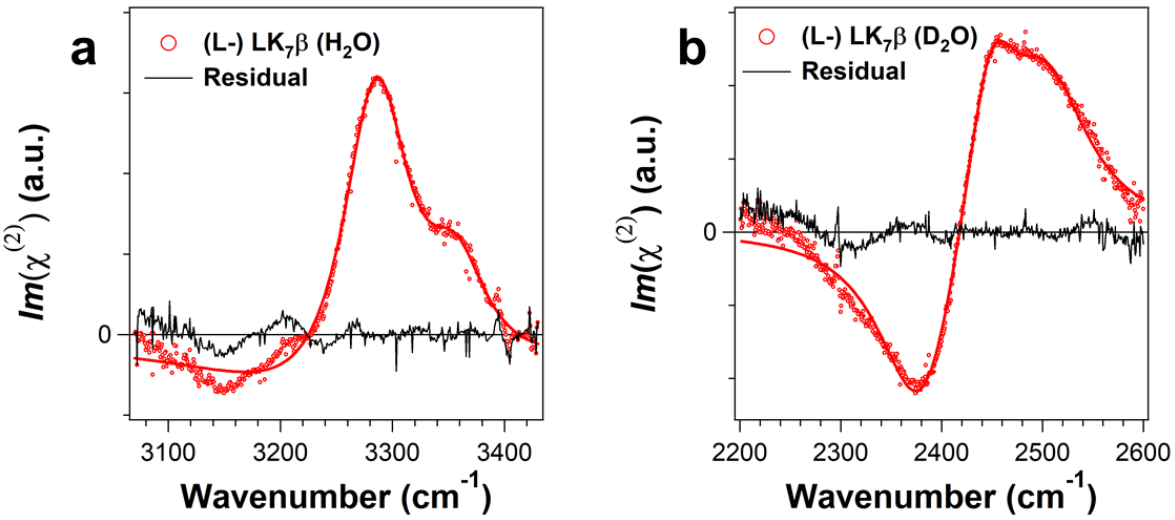

Figure A2.2. Residual analyses of fitting phase-resolved chiral SFG spectra with three peaks. (a) O-H/N-H spectral region of (L-) $LK_7\beta$ prepared in H_2O. (b) O-D/N-D spectral region of (L-) $LK_7\beta$ prepared in D_2O. The spectra are fit to equation 2 (**Methods** in **Chapter 2**). Both fittings fail to accurately capture features towards redder frequencies (see also **Figure A2.3**).

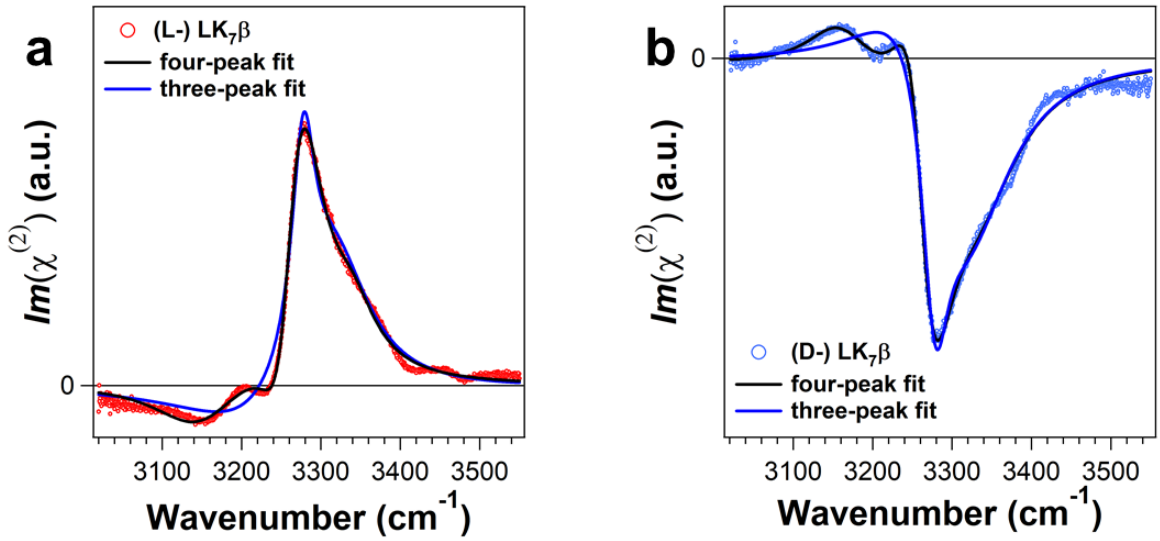

Figure A2.3. High resolution phase-resolved chiral SFG in O-H/N-H spectral region of (L-) ≈ (a) and (D-) $LK_7\beta$ (b) prepared in H_2O. A SFG setup with picosecond pulsed lasers was used,[21] in contrast to the femtosecond system[22] used for the results presented in Chapter 4.

Parameter	(Figure A2.3a) (L-) $LK_7\beta$-H_2O	(Figure A2.3b) (D-) $LK_7\beta$-H_2O	Assignments
ω_1 (cm^{-1})	3152 ± 2	3161 ± 1	O-H (H_2O)
A_1 (a.u.)	-19.3 ± 1.3	22.4 ± 0.9	
Γ_1 (cm^{-1})	76.1 ± 2.9	61.7 ± 2.1	
ω_2 (cm^{-1})	3251 ± 2	3252 ± 1	N-H (L-$LK_7\beta$)
A_2 (a.u.)	-39.7 ± 1.8	43.2 ± 2.0	
Γ_2 (cm^{-1})	29.3 ± 2.4	28.7 ± 0.4	
ω_3 (cm^{-1})	3270 ± 3	3272 ± 1	N-H (L-$LK_7\beta$)
A_3 (a.u.)	79.0 ± 2.1	-71.7 ± 2.3	
Γ_3 (cm^{-1})	36.6 ± 1.7	32.8 ± 0.6	
ω_4 (cm^{-1})	3336 ± 2	3330 ± 2	O-H (H_2O)
A_4 (a.u.)	17.0 ± 2.2	-51.7 ± 3.1	
Γ_4 (cm^{-1})	44.9 ± 2.2	67.1 ± 1.5	

Table A2.1. Fitting parameters and assignments for the spectra in **Figure A2.3**.

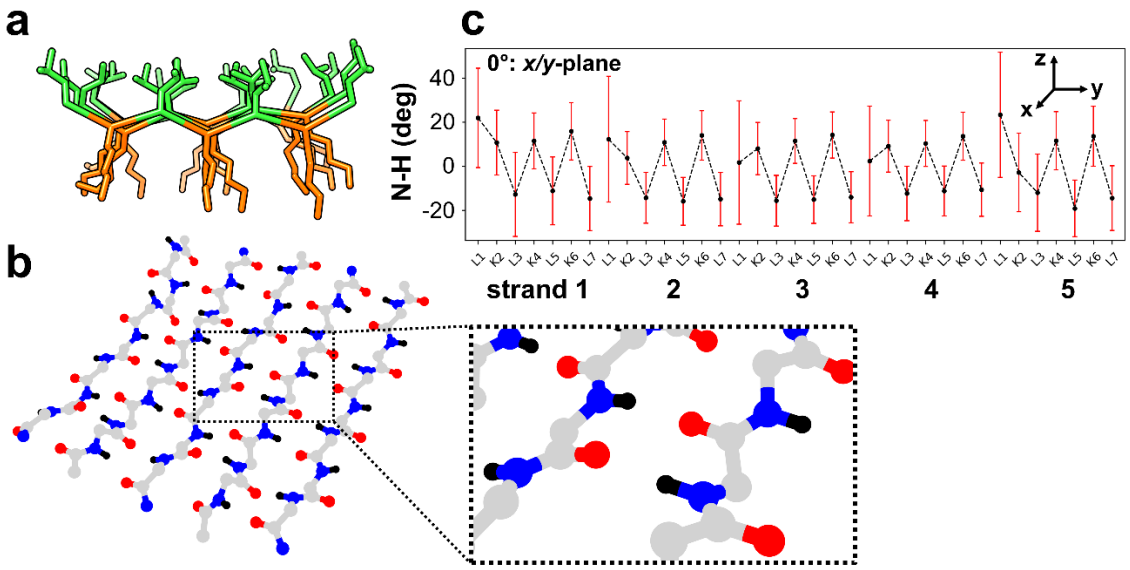

Figure A2.4. Molecular dynamics of (L-) $LK_7\beta$ show (a) leucine (green) and lysine (orange) sidechains align on opposite sides of the antiparallel β-sheet. The characteristic pleated structure of the backbone is apparent. (b) The up/down orientations of backbone N-H groups (nitrogen atoms in blue and hydrogen atoms in black) alternate across adjacent residues. (c) Orientation of vectors parallel to backbone N-H bonds of residues. Angles are relative to plane of protein backbone, averaged over 100,000 configurations from 1 μs of MD sampling, with error bars in red. Dashed lines are guides for the eye.

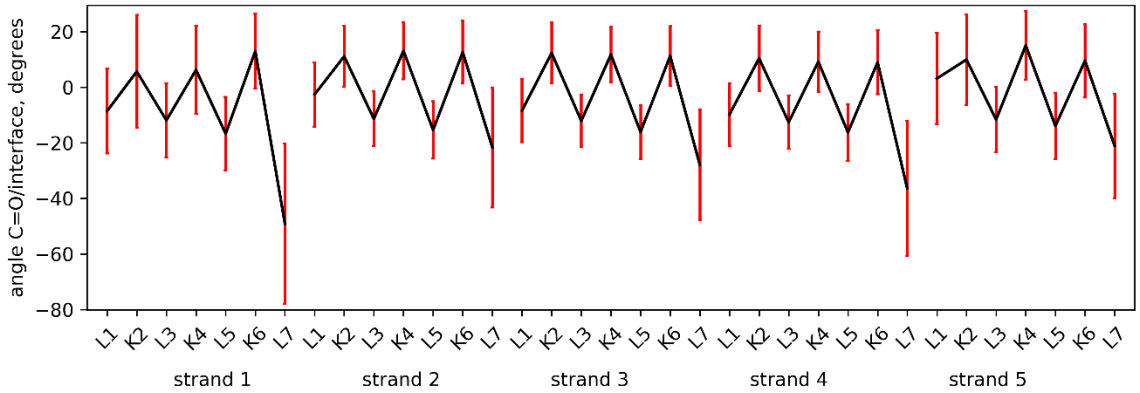

Figure A2.5. Orientation of vectors parallel to backbone C=O bonds of residues in (L-) $LK_7\beta$ homopentamer backbone. Angles relative to plane of vacuum-water interface, averaged over 100,000 configurations from 1 μs of MD sampling, with error bars in red. A positively-valued angle indicates that the bond points toward the vacuum phase, whereas a negatively-valued angle indicates that the bond points toward the water phase. Solid black lines are guides for the eye: the pleating of the β-sheet backbone is readily apparent.

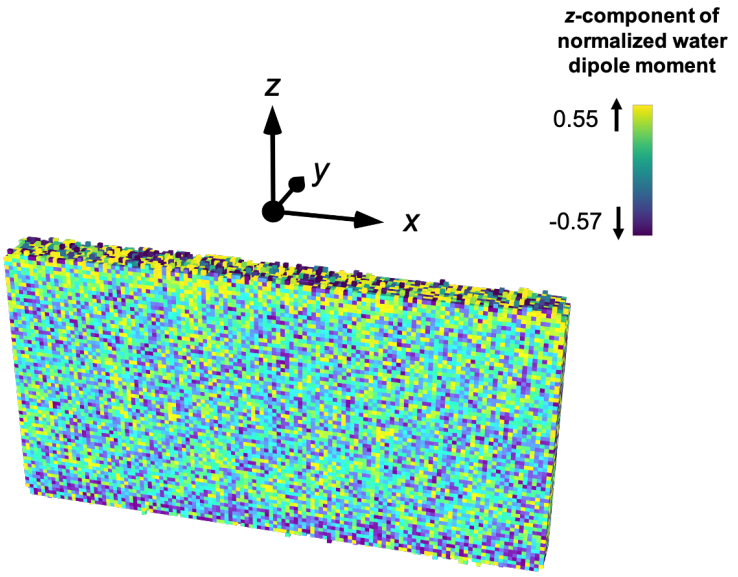

Figure A2.6. Average *z*-components of water dipole moment vectors for a box of pure water (a slice is shown), as obtained from 1 μs of MD sampling with the TIP4P-Ew water potential using a grid resolution of 1 Å. The orientation of water near the vacuum-water interface becomes isotropic ~5 Å below the interface.

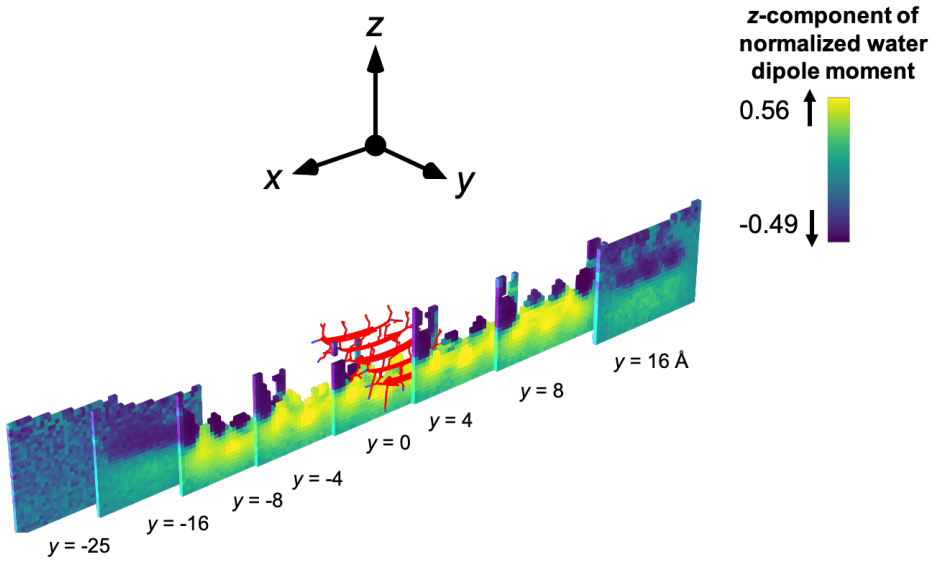

Figure A2.7. Average *z*-components of water dipole moment vectors around the (L-) $LK_7\beta$ homopentamer obtained from 1 μs of MD sampling using AMBER ff14SB for the protein and the TIP4P-Ew water potential. The slices shown extend along the *y*-dimension, perpendicular to the β-strand axis.

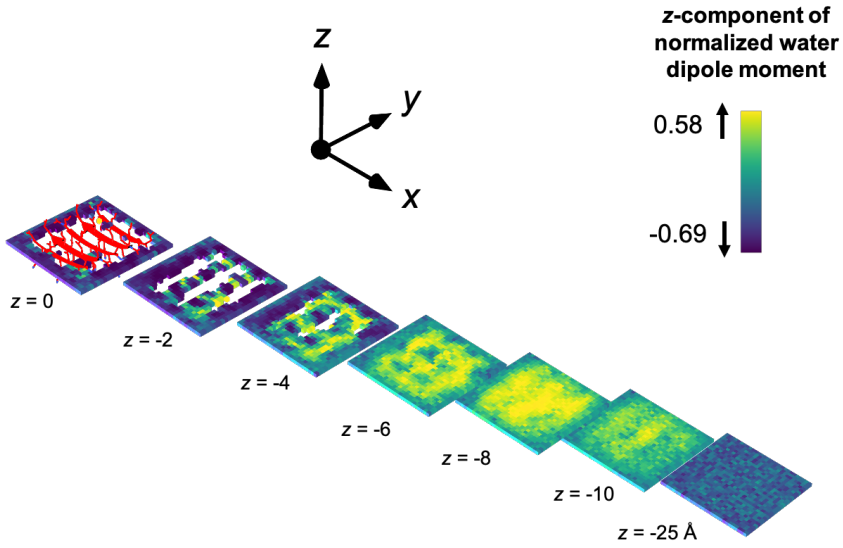

Figure A2.8. Average *z*-components of water dipole moment vectors around the (L-) $LK_7\beta$ homopentamer obtained from 60 ns of MD sampling using AMBER ff14SB for the protein and the SPC/E water potential. The slices shown extend along the *z*-dimension, which runs perpendicular to the interface. These calculations illustrate that the same qualitative behavior is reproduced for two different water potentials.

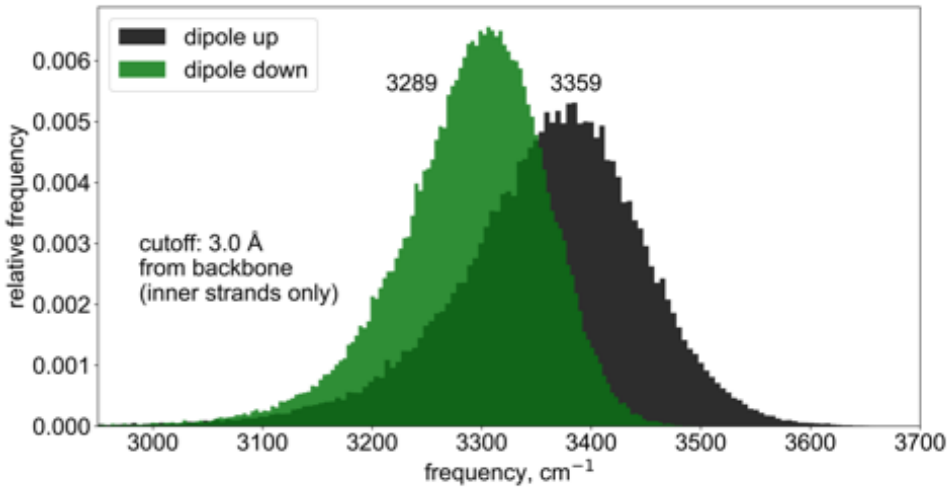

Figure A2.9. Infrared response of the O-H stretching mode of water within 3 Å of the backbone of strands 2, 3, and 4 (for numbering of strands, see Figure 2.4a in Chapter 2). The IR frequency was calculated for water molecules with the average *z*-component of the dipole moment oriented either up (black) or down (green) relative to the vacuum-water interface.

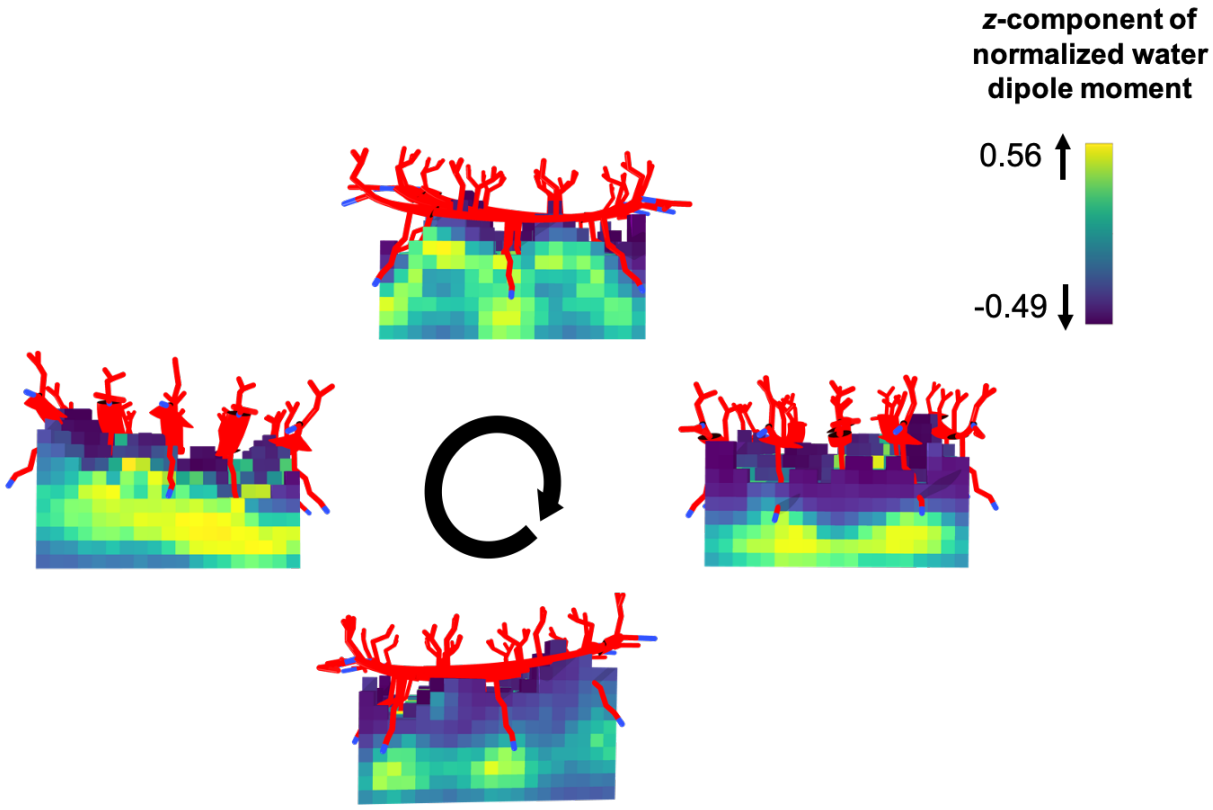

Figure A2.10. Average *z*-components of water dipole moment vectors around the (L-) $LK_7\beta$ homopentamer obtained from 1 μs of MD sampling using AMBER ff14SB for the protein and the TIP4P-Ew water potential. The slices at four edges of the β-sheet are 10 Å from the center of the β-sheet in both directions, parallel and perpendicular to the β-sheet axis.

References

1. Horn, H. W.; Swope, W. C.; Pitera, J. W.; Madura, J. D.; Dick, T. J.; Hura, G. L.; Head-Gordon, T. Development of an improved four-site water model for biomolecular simulations: TIP4P-Ew. *J. Chem. Phys.* **2004,** *120*, 9665-9678.
2. Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theory Comput.* **2015,** *11*, 3696-3713.
3. Miyamoto, S.; Kollman, P. A. Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models. *J. Comput. Chem.* **1992,** *13*, 952-962.
4. Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C. 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.
5. Michaud-Agrawal, N.; Denning, E. J.; Woolf, T. B.; Beckstein, O. MDAnalysis: A toolkit for the analysis of molecular dynamics simulations. *J. Comput. Chem.* **2011,** *32*, 2319-2327.
6. Gowers, R. J.; Linke, M.; Barnoud, J.; Reddy, T. J. E.; Melo, M. N.; Seyler, S. L.; Domanski, J.; Dotson, D. L.; Buchoux, S.; Kenney, I. M. *MDAnalysis: a Python package for the rapid analysis of molecular dynamics simulations*; 2575-9752; Los Alamos National Lab.(LANL), Los Alamos, NM (United States): 2019.
7. Hunter, J. Matplotlib: A 2D Graphics Environment. *Comput. Sci. Eng.* **2007,** *9*, 90-95.
8. McKinney, W. In *Data structures for statistical computing in python*, Proceedings of the 9th Python in Science Conference, Austin, TX: 2010; pp 51-56.
9. Lam, S.; Pitrou, A.; Seibert, S., Proceedings of the Second Workshop on the LLVM Compiler Infrastructure in HPC. LLVM'15. ACM New York, NY, USA: 2015.
10. McGibbon, Robert T.; Beauchamp, Kyle A.; Harrigan, Matthew P.; Klein, C.; Swails, Jason M.; Hernández, Carlos X.; Schwantes, Christian R.; Wang, L.-P.; Lane, Thomas J.; Pande, Vijay S. MDTraj: A Modern Open Library for the Analysis of Molecular Dynamics Trajectories. *Biophys. J.* **2015,** *109*, 1528-1532.
11. Auer, B. M.; Skinner, J. L. Dynamical effects in line shapes for coupled chromophores: Time-averaging approximation. *J. Chem. Phys.* **2007,** *127*, 104105.
12. Auer, B. M.; Skinner, J. L. IR and Raman spectra of liquid water: Theory and interpretation. *J. Chem. Phys.* **2008,** *128*, 224511.
13. Pieniazek, P. A.; Tainter, C. J.; Skinner, J. L. Interpretation of the water surface vibrational sum-frequency spectrum. *J. Chem. Phys.* **2011,** *135*, 044701.
14. Moad, A. J.; Simpson, G. J. A Unified Treatment of Selection Rules and Symmetry Relations for Sum-Frequency and Second Harmonic Spectroscopies. *J. Phys. Chem. B* **2004,** *108*, 3548-3562.
15. Carr, J. K.; Wang, L.; Roy, S.; Skinner, J. L. Theoretical Sum Frequency Generation Spectroscopy of Peptides. *J. Phys. Chem. B* **2015,** *119*, 8969-8983.
16. Wang, H.-F.; Gan, W.; Lu, R.; Rao, Y.; Wu, B.-H. Quantitative spectral and orientational analysis in surface sum frequency generation vibrational spectroscopy (SFG-VS). *Int. Rev. Phys. Chem.* **2005,** *24*, 191-256.

17. Fu, L.; Xiao, D.; Wang, Z.; Batista, V. S.; Yan, E. C. Y. Chiral Sum Frequency Generation for In Situ Probing Proton Exchange in Antiparallel β-Sheets at Interfaces. *J. Am. Chem. Soc.* **2013,** *135*, 3592-3598.
18. 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.
19. Hu, X.-H.; Wei, F.; Wang, H.; Wang, H.-F. α-Quartz Crystal as Absolute Intensity and Phase Standard in Sum-Frequency Generation Vibrational Spectroscopy. *J. Phys. Chem. C* **2019,** *123*, 15071-15086.
20. Hu, X.-H.; Fu, L.; Hou, J.; Zhang, Y.-N.; Zhang, Z.; Wang, H.-F. N–H Chirality in Folded Peptide LK7β Is Governed by the Cα–H Chirality. *J. Phys. Chem. Lett.* **2020,** *11*, 1282-1290.
21. Velarde, L.; Zhang, X.-y.; Lu, Z.; Joly, A. G.; Wang, Z.; Wang, H.-f. Communication: Spectroscopic phase and lineshapes in high-resolution broadband sum frequency vibrational spectroscopy: Resolving interfacial inhomogeneities of "identical" molecular groups. *J. Chem. Phys.* **2011,** *135*, 241102.
22. 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.

Chapter 3: Simulation of the chiral SFG response of supramolecular structures requires vibrational couplings

Adapted from the following reference with permission:

Konstantinovsky, D.; Perets, E. A.; Yan, E. C. Y.; Hammes-Schiffer, S.

J. Chem. Phys. B 2021, *125* (43), 12072-12081.

DOI: 10.1021/acs.jpcb.1c06360

Abstract

Chiral vibrational sum frequency generation (SFG) spectroscopy probes the structure of the solvation shell around chiral macromolecules. The dominant theoretical framework for understanding the origin of chiral SFG signals is based on analysis of molecular symmetry, which assumes no interaction between molecules. However, water contains strong intermolecular interactions that significantly affect its properties. Here, the role of intermolecular vibrational coupling in the chiral SFG response of the O-H stretch of water surrounding an antiparallel β-sheet at the vacuum-water interface is investigated. Both intramolecular and intermolecular coupling between O-H groups is required to simulate a chiral SFG signal. This dependence is also observed for a chiral water dimer, illustrating that this phenomenon is not specific to larger systems. We also find that a dimer of C_{3v} molecules predicted to be chiral SFG inactive by the symmetry-based theory can generate a chiral SFG signal when intermolecular couplings are considered, suggesting that even highly symmetric solvent molecules may produce chiral SFG signals when interacting with a chiral solute. The consideration of intermolecular couplings extends the prevailing theory of the chiral SFG response to structures larger than individual molecules and provides guidelines for future modeling.

Introduction

Biological processes occur in aqueous solution or at an interface involving water, and this water does not always behave as a passive medium. In particular, water molecules can bind to biomacromolecules and modulate their structure and function. These water molecules can interact with a solute tightly enough to adopt large-scale structure, but loosely enough to form a structural ensemble and exchange rapidly with bulk solvent. A complete understanding of the behavior of biomacromolecules requires a thorough description of these water molecules. Chiral vibrational sum frequency generation (SFG) spectroscopy is an emergent technique to probe the solvation shell of biomacromolecules.[23-25] As a second-order nonlinear chirality-sensitive technique, chiral SFG can suppress background signal from water at non-centrosymmetric interfacial environments and isotropic bulk water. Thus, it can selectively detect solvent molecules forming chiral supramolecular structures around chiral macromolecules, such as DNA and protein.[23, 24] Vibrational spectroscopy can resolve subpopulations exchanging on the picosecond timescale and does not require molecular labels that perturb chemical structures. Importantly, chiral SFG can detect chiral solvent structure induced by interactions with biomacromolecules even if the individual solvent molecules in that structure exchange rapidly.[23, 24, 26, 27]

The last two decades have seen extensive applications of chiral SFG to biological systems.[18, 28, 29] In 2003, a chiral SFG spectrum of the amide I response of a protein at the air-water interface was published.[28] Our group extended the application to probe various protein secondary structures using the amide I, C-H, and N-H stretch modes.[18, 30, 31] Achiral spectra of DNA in the O-D and C-H stretch regions showed that SFG can detect the effects

of solvation in D_2O.[32] Chiral SFG spectra of the C-H stretch in DNA were reported.[33, 34] Others have applied femtosecond time-resolved SFG to monitor the relaxation of the amide I vibrational mode in the presence of water solvent.[35] Reports of the chiral SFG response of water surrounding DNA and antiparallel β-sheet protein demonstrated that even highly dynamic solvation structures of a biological system can be probed with SFG.[23, 24, 27, 31] We recently used the internal heterodyne phase-sensitive chiral SFG method[36] and observed that the phase of the water O-H stretching bands is correlated with the absolute handedness of (L-) and (D-) proteins, further establishing the method for probing hydration structures of biomacromolecules.[27, 31]

Simpson's chiral SFG theory[14] has served as a foundation for predicting and interpreting chiral SFG studies of interfaces. This theory takes intramolecular couplings into account by considering molecular units larger than single bonds when assigning the symmetry of vibrational modes.[14] Because this symmetry-based theory has been derived and applied mainly for chiral macromolecules (e.g., various protein secondary structures), intermolecular coupling is not relevant and thus not considered. However, the hydrogen-bonding interactions between water molecules are paramount. Thus, the recent reports of chiral SFG signals from water solvent surrounding biomacromolecules[23, 24, 27] call for examination of intermolecular couplings.

In this paper, we examine the impact of intermolecular coupling as well as intramolecular coupling on chiral SFG signals[14] using an electric field mapping method.[13, 37, 38] We show that intramolecular or intermolecular vibrational coupling is necessary to produce nonzero chiral SFG signals in the simulation and that the signal intensity contributed by these two couplings are comparable. In contrast, neither of these couplings

is needed to generate an achiral SFG response, and the achiral SFG response is only mildly affected when such couplings are included. We demonstrate these effects using the antiparallel β-sheet system formed by the $LK_7\beta$ (LKLKLKL) peptide at the vacuum-water interface (Figure 3.1).[39] To examine the fundamental origin of the coupling effect, we also study a small model system composed of a dimer of water frozen in a chiral configuration. We find that couplings are necessary to observe a chiral SFG signal in this minimal system as well. Similarly, we examine a dimer of ammonia with C_{3v} symmetry in contrast to the C_{2v} symmetry of water to further illustrate the interplay of molecular symmetry and vibrational couplings for generating chiral SFG responses. Finally, we examine a dimer of methane to show that the trends seen with the ammonia system translate to other highly symmetric point groups. We discuss our results in the context of Simpson's symmetry-based theory.[14]

This study will be helpful to experimentalists considering SFG spectra of isotopically labeled water molecules in biological systems (e.g., HOD largely lacks intramolecular coupling but still has intermolecular coupling) and solvent systems other than water for chiral polymers and macromolecules. The results will also guide theorists in proposing approximations to the dipole-polarizability time correlation function formalism[40] for calculating SFG responses. In particular, although couplings can be omitted to save computational time for calculating achiral SFG spectra, they must be included to model chiral SFG responses.[41] We recognize that couplings have not always been reported to be essential for a chiral SFG signal,[15] and we address this discrepancy in the Discussion.

Symmetry-based chiral SFG theory

We interpret the simulated SFG spectra of water in this study based on Simpson's symmetry-based chiral SFG theory.[14] Herein, (*x*, *y*, *z*) will represent the lab frame coordinates, and (*a*, *b*, *c*) will represent the molecular frame coordinates. The indices *I*, *J*, and *K* will be used to indicate *x*, *y*, or *z*, whereas the indices *i*, *j*, and *k* will be used to indicate *a*, *b*, or *c*.

Chiral SFG signals can be generated from a chiral interface, which is defined as a molecular system with C_∞ symmetry with the rotational axis aligned with the surface normal. Within the electric dipole approximation, SFG signals arise from the second-order nonlinear response of materials

$$\mathbf{P} = \mathbf{P}_0 + \chi^{(1)}\mathbf{E}_1 + \chi^{(2)}\mathbf{E}_1\mathbf{E}_2 + \ldots \quad (1)$$

where P is the material's polarization, E_1 and E_2 are incident electric fields, and $\chi^{(n)}$ is the n^{th}-order susceptibility tensor.[42] In vibrational SFG, one beam is in the mid-infrared (IR) range and the other is in the visible range but typically far off electronic resonance.[43-45] The $\chi^{(2)}$ tensor contains 27 elements $\chi^{(2)}_{IJK}$ as in

$$E^I_{SFG} \propto \sum_{JK} \chi^{(2)}_{IJK} E^J_{vis} E^K_{IR} \quad (2)$$

where E^I_{SFG}, E^J_{vis}, and E^K_{IR} are the optical fields of the SFG, visible, and IR beams, respectively, with the indices *I*, *J*, and *K* specifying the components of these optical fields along a Cartesian direction *(x*, *y*, or *z*) in the lab reference frame (Figure 3.2). Manipulating the polarization of the IR and visible beams and the detector of the SFG signal isolates different elements of the $\chi^{(2)}$ tensor. Among the 27 elements, those with all three indices

that are different ($I \neq J \neq K$) are called orthogonal. Orthogonal $\chi^{(2)}$ elements for a uniaxial assembly can be nonzero only for a chiral interface.

To perform a chiral SFG experiment, the *psp* polarization setup can be used (Figure 3.2), where *s* polarization is taken to lie along the *y*-axis, and *p* polarization lies along *x* and *z* on the plane of incidence. Figure 3.2 shows that the effective second-order susceptibility, $\chi^{(2)}_{psp}$, depends only on $\chi^{(2)}_{xyx}$, $\chi^{(2)}_{zyz}$, $\chi^{(2)}_{xyz}$, and $\chi^{(2)}_{zyx}$. The first two elements vanish due to isotropy in the *xy* plane, as is found at a non-centrosymmetric interface.[14] The last two elements are orthogonal and $\chi^{(2)}_{xyz} = 0$ under the condition of electronic nonresonance due to the symmetric polarizability tensor.[18,14] Hence, the experimentally measurable chiral SFG response, $\chi^{(2)}_{psp}$, is only related to $\chi^{(2)}_{zyx}$:

$$\chi^{(2)}_{psp} \propto \chi^{(2)}_{zyx} - \chi^{(2)}_{xyz} = \chi^{(2)}_{zyx} \qquad (3)$$

In the symmetry-based theory, the system of interest has C_∞ symmetry. Although this is true for many experimental systems, it is not true for molecular dynamics (MD) trajectories with limited sampling. To avoid this issue, the $\chi^{(2)}_{xyz}$ term in equation 3 is kept in our calculations for extracting chiral SFG signal from MD trajectories (see equations 9-10 and 4 in the Appendix). The current study assumes a *psp* setup, but the methods and results are directly applicable to the other two polarization setups (*spp* and *pps*) for chiral SFG experiments.

The macroscopic chiral SFG response, $\chi^{(2)}_{psp}$, is the ensemble average of microscopic responses given by molecular hyperpolarizabilities (β).[44] The hyperpolarizability arises from the transition dipole and polarizability derivatives as in

$$\beta_{ijk} \propto \left(\frac{\partial \alpha_{ij}}{\partial Q}\right)\left(\frac{\partial \mu_{k}}{\partial Q}\right) \tag{4}$$

where $\boldsymbol{\alpha}$ is the transition polarizability matrix, $\boldsymbol{\mu}$ is the transition dipole vector, and Q is a normal mode coordinate. The ensemble average can be expressed as

$$\chi^{(2)}_{IJK} \propto \sum_{i,j,k} \left\langle R_{Ii}(\psi) R_{Jj}(\theta) R_{Kk}(\phi) \right\rangle \beta_{ijk} \tag{5}$$

where R is an Euler rotation matrix, ϕ, θ, and ψ are Euler rotation angles, (*i*, *j*, *k*) represents the molecular frame, (*I, J, K*) represents the laboratory frame, and the brackets indicate averaging over molecular orientations. Applying equation 5 and integrating over all rotations in the *xy*-plane (ϕ) leads to

$$\begin{aligned}
\chi^{(2)}_{zyx} \propto \frac{1}{2}\{&(-\beta_{cab}+\beta_{cba})\cos^2(\theta) \\
&+\cos(\theta)\sin(\theta)[(\beta_{aab}-\beta_{aba}+\beta_{cbc}-\beta_{ccb})\cos(\psi) \\
&+(-\beta_{bab}+\beta_{bba}+\beta_{cac}-\beta_{cca})\sin(\psi)] \\
&+\sin^2(\theta)((-\beta_{abc}+\beta_{acb})\cos^2(\psi) \\
&-(\beta_{aac}-\beta_{aca}-\beta_{bbc}+\beta_{bcb})\cos(\psi)\sin(\psi)+(\beta_{bac}-\beta_{bca})\sin^2(\psi)\}
\end{aligned} \tag{6}$$

The *c*-axis points along the primary rotation axis of the molecule. Although all chiral molecular β_{ijk} ($i \neq j \neq k$) elements are zero for water molecules with C_{2v} symmetry, the achiral β elements can combine to give rise to chiral $\chi^{(2)}$ elements if the water molecules are arranged in a chiral configuration at an interface. Eliminating all zero-valued β elements in equation 6[14, 16] based on symmetry arguments results in the expression of the chiral $\chi^{(2)}_{zyx}$ element as

$$\chi^{(2)}_{zyx} \propto \frac{1}{4}\sin^2(\theta)\sin(2\psi)(\beta_{aac}-\beta_{bbc}+\beta_{bcb}-\beta_{aca}) \tag{7}$$

Therefore, chiral SFG can be sensitive to water when the water molecules adopt a chiral supramolecular structure in the solvation shells of biomacromolecules (e.g., protein and DNA) at interfaces. We are not the first to notice this. In fact, Simpson theorized the appearance of chiral spectra from achiral chromophores with certain symmetries as early as 2004,[14] and in 2012 Zanni published a theory of chiral 2D-SFG signals arising from pairs of achiral but coupled chromophores.[46]

Derivation of Equation 7 from Equation 6 suggests that water is quite a unique solvent with respect to generating the chiral SFG response. Its C_{2v} symmetry is high enough (with two reflection planes) to eliminate all microscopic chiral (orthogonal) β elements, but low enough to preserve some achiral β elements that can still contribute to the macroscopic chiral $\chi^{(2)}_{zyx}$, as shown in equation 7. Hence, the symmetry-based theory predicts that chiral supramolecular assemblies of water situated in an interfacial environment with C_∞ symmetry are chiral SFG active. However, when the molecule becomes just slightly more symmetric (such as C_{3v}), the prediction can be drastically different because the higher the symmetry, the more β elements become zero. In particular, all β elements in equation 6 are zero or cancel out for a C_{3v} molecular system. Thus, the symmetry-based theory predicts that a C_{3v} molecule (e.g., ammonia, NH_3) cannot generate chiral signals (e.g., N-H stretching in NH_3) even when the molecules are assembled into chiral supramolecular structures at an interface.[14, 16] Below, we will discuss how introducing intermolecular vibrational couplings can alter this prediction, leading to a chiral SFG response in simulated spectra.

Computational Methods

MD simulations

A starting structure of five antiparallel $LK_7\beta$ (LKLKLKL) strands (Figure 3.1A) was placed into a box with at least 10 Å of TIP4P-Ew water (a modification of the original TIP4P model that is optimized for Ewald summation treatment of long-range interactions)[23] next to each of the four edges of the protein and above and below it.[27] Half of the water was removed and Cl^- ions were added to the aqueous half to neutralize the system, creating a vacuum above the neutral solvated protein system.[1, 24, 27] The protein was placed at the interface with the lysine sidechains pointing into the water and the leucine side chains facing the vacuum (Figure 3.1B) and was modeled with the AMBER ff14SB force field.[2] The N-termini were acetylated, and the C-termini were amidated ($-NH_2$). The energy of the system was minimized, followed by equilibration in an NVT (constant number of particles, volume, and temperature) ensemble for 6 ns at 298 K. Langevin dynamics was propagated with a friction coefficient of 1 ps^{-1} and a timestep of 1 fs. Long-range electrostatic interactions were treated with the particle-mesh Ewald method.[47] Water molecules were kept rigid with the SETTLE algorithm, and bonds involving protein hydrogens were constrained with the SHAKE method.[3, 4] Configurations were saved every 10 fs. For the calculation of achiral SFG spectra, trajectories were wrapped in the *z*-direction to consolidate the slab into a single unit, and spectral calculations took periodic boundary conditions in the *x*- and *y*- directions into account. Wrapping was not necessary for chiral SFG calculations (see SI for explanation). The simulations were performed with the OpenMM 7.4 library on a Tesla V100 GPU using CUDA 9.2.[48] See the SI for full equilibration details.

SFG calculations of the aqueous protein system

The chiral SFG response was calculated using the dipole-polarizability time correlation function approach with electric field mappings.[13, 37, 38, 40, 49] This approach uses empirical mappings between the electric field measured at the water hydrogens in the direction along the O-H bond and quantities such as the dipole derivative, the vibrational frequency, and the vibrational couplings. The inhomogeneous limit approximation was used as in the expression

$$\chi^{(2)}_{IJK}(\omega) \approx \left\langle \sum_{a=1}^{N_{\mathrm{OH}}} \frac{\sum_{b=1}^{N_{\mathrm{OH}}} \alpha^{b}_{IJ} U_{ba} \sum_{b=1}^{N_{\mathrm{OH}}} \mu^{b}_{K} U_{ba}}{\lambda_a - \omega - \frac{i}{2\tau}} \right\rangle \tag{8}$$

where α is the transition polarizability tensor of each O-H bond computed with a bond-polarizability approach, μ is the O-H stretch transition dipole moment, U is the eigenvector matrix of the array of vibrational frequencies and couplings, λ is the corresponding set of eigenvalues of the Hamiltonian matrix, and τ is the vibrational lifetime, which is assumed to be 1.3 ps.[37, 38, 50] The indices *a* and *b* refer to particular O-H bonds, while the indices *I*, *J*, and *K* each refer to the dimensions *x*, *y*, *z* making up a component of $\chi^{(2)}$.

In the inhomogeneous limit, the response is a simple average over conformations with no explicit time dependence.[37] This approach was used to maximize the number of frames averaged during the procedure for an optimal signal-to-noise ratio. The computationally less expensive alternative, the time-averaging approximation, correlates the dipole and polarizability across a short time interval, with the frequencies and couplings

averaged over that interval. As shown in Figure A3.1, this approach did not significantly change the spectrum other than increasing the noise.[37] A fully time-dependent correlation-function approach such as that introduced previously[13] is not feasible for a system of this size and a signal that requires long trajectories to converge.

The calculation of intermolecular couplings using this approach requires a dipole-dipole distance. To find this distance, each O-H bond was assigned an effective dipole location 0.67 Å from the oxygen along the bond.[13] For achiral SFG, the contribution of the bottom half of the system was suppressed with a sigmoidal damping function to avoid contributions from a second surface opposite the surface containing the protein.[13] The *z*-component of the effective dipole location of each O-H bond was used to assign the bond's position along the *z*-axis for the damping function. No damping was necessary for chiral SFG calculations because the protein stayed on the top surface.

The polarizability is assumed to be localized at each O-H bond, and the ratio of the component of the polarizability derivative along the bond to the components perpendicular to the bond is assumed to be constant.[49, 51] We use a ratio of 5.6, as the Skinner group has in the past.[49] The value of this constant affects the magnitude and sign of the chiral SFG response but does not affect the position of the peaks (see Figure A3.2).

For each SFG calculation, 100 ns of MD data was analyzed, corresponding to 10,000,000 frames. The electric fields needed for the frequency mapping were calculated using the previously optimized charges for the atoms of TIP4P water,[19] while AMBER ff14sb charges were used for the protein and ions.[2, 13] The protein was assumed not to have any vibrational coupling with water molecules and therefore influenced the system only geometrically and as a source of electric field. The frequency resolution was 1 cm^{-1}. The

calculations were implemented in Python and Cython, using the MDAnalysis and NumPy libraries.[5, 52]

Extracting chiral SFG signal from MD trajectories

The chiral SFG theory is derived based on the assumption that the system of interest has C_∞ symmetry. This is not true for MD trajectories with limited sampling. However, it is not necessary to sample every orientation equally.[15] Instead, one can simply assume that each frame of the trajectory is rotated by a random angle about the *z*-axis and then integrate over all values of the angle. In the rotated frame of reference denoted by the prime symbol, each component of $\chi^{(2)}$ is transformed according to

$$\begin{aligned} \chi^{(2)}_{z'y'x'} &\rightarrow \chi^{(2)}_{zyx} - \chi^{(2)}_{zxy} \\ \chi^{(2)}_{x'y'z'} &\rightarrow \chi^{(2)}_{xyz} - \chi^{(2)}_{yxz} \end{aligned} \tag{9}$$

and therefore, according to equation 3, the *psp* response is given by

$$\chi^{(2)}_{psp} \propto \chi^{(2)}_{zyx} - \chi^{(2)}_{zxy} + \chi^{(2)}_{yxz} - \chi^{(2)}_{xyz} \tag{10}$$

where $\chi^{(2)}_{yxz} - \chi^{(2)}_{xyz}$ in the original frame, which comes from $\chi^{(2)}_{x'y'z'}$, is approximately equal to zero because the Raman polarizability tensor, which corresponds to the first two indices in $\chi^{(2)}$, is approximately symmetric when the incident beam frequencies are far from an electronic resonance.[15] This leaves

$$\chi^{(2)}_{psp} \propto \chi^{(2)}_{zyx} - \chi^{(2)}_{zxy} \tag{11}$$

SFG calculation of model systems

To calculate the SFG spectra of the model systems, the approach described above (equation 8) was adapted and implemented in a Mathematica notebook. Unlike our SFG calculations of the protein system, electric fields were ignored in the model system calculations, and the gas phase values of mapped quantities for the OH bond were used for all systems.[49] The water dimer structure coordinates were taken obtained from a pair of water molecules taken from near $LK_7\beta$ in a frame of an MD trajectory. The C_{3v} (i.e., ammonia) molecule was built as a perfectly trigonal pyramidal structure (109.5° between adjacent bonds), and the dimer was built using simple geometric transformations that created a segment of a chiral helix. The same procedure was used to create a dimer of methane molecules. The chiral SFG spectrum of each model structure was calculated by averaging 200 conformations rotating around the *z*-axis, and $\chi^{(2)}_{psp} = \chi^{(2)}_{zyx} - \chi^{(2)}_{xyz} + \chi^{(2)}_{zyz} - \chi^{(2)}_{xyx}$ (see Figure 3.2), with the disappearance of the $\chi^{(2)}_{zyz}$ and $\chi^{(2)}_{xyx}$ components monitored to ensure adequate rotational averaging (full integration over all rotations was impractical). The negative signs arise from the vectors specifying the orientations of the incoming and outgoing beams.[15] Unless otherwise stated, the O-H/N-H/C-H groups were treated as equivalent in terms of dipole and polarizability magnitudes, intramolecular and intermolecular couplings, and vibrational resonance frequency. Natural fluctuations in these quantities were not sufficient to create a chiral SFG signal in the absence of couplings in the protein-containing system (data not shown), so variation in these quantities was not considered in the model systems. The O-H/N-H stretch frequencies were set to 3400 cm^{-1}, roughly reflecting IR spectra of the O-H and N-H stretches, while the C-H frequency of methane was set to 2900, roughly reflecting IR spectra of liquid alkanes.[49, 53] Note that the

precise value of the resonance frequency does not affect the predicted spectra except by shifting them (see Figure A3.3). The vibrational lifetime (i.e., Lorentzian damping coefficient) was set to 0.01 seconds for the simple model dimer systems to achieve a Lorentzian peak broadness similar to that observed in the real system. If the lifetime is assumed to be in a more realistic range (~1 picosecond), the peaks become extremely sharp and less convenient to plot. The conclusions presented here are not affected by the choice of damping coefficient. The frequency resolution was 10 cm^{-1}. The polarizability was modeled as above, with a parallel-to-perpendicular polarizability derivative ratio of 5.6 assumed for both water and ammonia.[49, 54]

Results

Biomacromolecules have been observed to organize surrounding water into dynamic chiral superstructures.[23, 24, 27, 55] Here, we use a model system consisting of the antiparallel β-sheet formed by the short $LK_7β$ peptide that is stabilized by the vacuum-water interface (Figure 3.1). Switching the chirality of the protein amino acids from (L-) to (D-) causes the experimental chiral SFG response to flip.[27] This effect is clearly reproduced in our calculations (Figure 3.3). The chiral (*psp*) signal is around two orders of magnitude weaker than the achiral (*ssp*) signal and converges about ten times slower (Figure A3.4). These observations are likely the result of the complexity and variety of the chiral water supramolecular structures, the relatively slow dynamics of the protein scaffold compared to water, and the relatively small number of water molecules actively participating in the solvation shell. An interface containing pure water produces no significant chiral SFG signal (Figure 3.3). The small residual signal for pure water may be

due to imperfect MD sampling, which is limited by the high computational cost of the SFG spectrum calculations for systems of this size (~1350 water molecules). Figure A3.5 shows that two independent simulations of a pure water interface system produce extremely noisy and completely different signals, confirming that the chiral SFG response of a pure water surface is effectively zero. The small difference in intensity between (L-) and (D-) $LK_7\beta$ signals is most likely also a result of limited sampling due to the high computational cost.

Understanding the role of vibrational couplings in chiral SFG signals is important for the interpretation of experimental data. We find that neglecting both intermolecular and intramolecular vibrational couplings completely removes the chiral SFG signal from water around $LK_7\beta$ (Figure 3.4A), whereas the achiral SFG signal only blue-shifts slightly when both couplings are absent (Figure 3.4B). Note that we neglect couplings only in the SFG calculation and not in the MD simulation, where all interactions specified by the force field are preserved. Although both intermolecular and intramolecular couplings contribute to the chiral signal, either one is enough to make it appear. Intermolecular couplings cause a small red-shift in both the chiral and achiral SFG responses.[56] In real systems, this shift is likely caused by the vibrational delocalization of the O-H stretch over many molecules. Water molecules that interact with their neighbors have less electron density to devote to covalent bonds, which weakens the bonds and thus lowers the vibrational frequencies. In our model, the red shift is caused by the shift in the eigenspectrum of the coupling matrix due to significant off-diagonal components, thereby approximating the aforementioned physical effect. Because of the method by which these couplings were calculated, couplings between the O-H stretch and protein modes, for example the N-H stretch in the protein

backbone, were not included.[49] However, the protein still contributed to the spectrum by modulating the local electric field experienced by the water molecules.[2]

Figure 3.4C shows the distribution of vibrational couplings calculated from a representative frame of an MD simulation plotted against the dipole-dipole distance. The average value of the intramolecular coupling (-17 cm^{-1}) is much larger in magnitude than the average value of the intermolecular coupling (-0.025 cm^{-1}), but a few very short-distance (< 3 Å) intermolecular couplings are larger in magnitude (> 100 cm^{-1}) than any intramolecular coupling. The standard deviation of the intermolecular coupling magnitudes (1.1 cm^{-1}) indicates that these larger couplings are extremely rare.[49] In Figure 3.4D, intramolecular couplings are included fully, but intermolecular couplings are only included up to a distance cutoff. This plot reveals that most of the contribution to the spectral intensity from the intermolecular couplings arises from interaction distances within 3 Å, that is, from the immediate neighbors of each molecule.

Although we have shown that short-range interactions produce most of the chiral SFG signal, we have not addressed the length scale at which these interactions originate. It may be that the observed effect of couplings is additive and only valid for large macromolecular systems and not for smaller chiral assemblies. To test this possibility, we calculated the chiral SFG response of a minimal water dimer system in a chiral configuration at an interface with the macroscopic symmetry of C_∞ (Figure 3.5A). Figure 3.5A (top) shows a pair of the enantiomeric water dimers. For simplicity we assumed that each O-H group has the same dipole moment and polarizability magnitude. Intramolecular couplings were calculated using gas-phase values[49] for mapped quantities such as the dipole derivative. Intermolecular couplings were calculated based on distance by using gas-

phase values for the transition dipoles, as described in the Methods. We assumed a complex Lorentzian response with a damping factor chosen to produce peaks with a similar width as in the protein-water spectra, although the peak width does not affect our conclusions. As shown in the simulated spectra (Figure 3.5A) from top to bottom, when all couplings are included, the mirror-image water dimers produce signals with opposite phase. When only intramolecular couplings are considered, the intensity of the signals does not change significantly. Including just intermolecular couplings (as would be the case for a system such as HOD) produces a small but reliable signal as well. When there is no vibrational coupling, the signals disappear completely.

To further illustrate the interplay of vibrational couplings and molecular symmetry in determining the chiral SFG response, we performed the same calculation for two model systems with a high degree of symmetry: a dimer of ammonia molecules and a dimer of methane molecules, both arranged in a chiral configuration, shown as a pair of enantiomers in Figure 3.5B-C (top). We observed coupling effects similar to those of the water dimers but with a major difference – intramolecular couplings on their own produced no signal. This result aligns with Simpson's prediction of no chiral signal from an ideal C_{3v} or T_d (tetrahedral) molecule.[14, 16] In the next section, we will discuss how intermolecular couplings can break symmetry and thereby produce chiral SFG signals in any molecule.

Discussion

Water has a uniquely strong and extensive network of hydrogen bonds. These bonds significantly delocalize the O-H stretch mode and therefore change the symmetry of vibrational chromophores. To illustrate this concept, Figure 3.6 schematically shows how

couplings change the fundamental geometry of a water dimer that is frozen in an asymmetric configuration. When O-H stretches are uncoupled, the system reduces to a set of four O-H bonds. When intramolecular couplings are added, each pair of bonds is unified, and the smallest indivisible unit of the system becomes a whole water molecule with C_{2v} symmetry. When intermolecular couplings are considered, the entire system is interconnected, and the fundamental geometry behind the spectral response becomes the (chiral) geometry of the entire water superstructure interacting with the chiral protein or other macromolecules.

Based on the concept that vibrational couplings change the symmetry of vibrational chromophores (Figure 3.6), we can discuss the observations in the simulated chiral SFG spectra (Figures 3.4-5) in the context of Simpson's symmetry-based theory. First, for Figure 3.4A, in the case of no vibrational coupling, each O-H bond in the protein-water system is an isolated linear oscillator, and the O-H vibrational mode belongs to the $C_{\infty v}$ symmetry group. For such a system, all tensor elements of the molecular hyperpolarizability are zero except β_{ccc}, β_{aac}, and β_{bbc}, where the *c*-axis points along the O-H bond, and all three elements are either proportional or equal to each other.[16] These three elements do not contribute to the chiral $\chi^{(2)}$ response according to equation 6 because $\beta_{aac} = \beta_{bbc}$. Thus, chiral SFG is forbidden[14] (purple, Figure 3.4A). When only intramolecular couplings are considered (red, Figure 3.4A), the symmetry of the vibrational chromophores becomes C_{2v}. This lower symmetry leads to additional nonzero β elements (β_{aca},β_{bcb}, β_{caa}, and β_{cbb}), where the *c*-axis is defined as the C_2 rotation axis of the molecule) which can contribute to chiral SFG as described by equation 7. Thus, the chiral SFG signal can be nonzero (red, Figure 3.4A). In the case of intermolecular couplings

alone, each O-H group of individual water molecules communicates with O-H groups of other water molecules. This cross-talk further reduces the symmetry and results in additional nonzero β elements if the solvent surrounds a chiral template. In the extreme case, the symmetry is reduced to C_1 (with no symmetry elements), and thereby all β hyperpolarizability tensor elements are nonzero. In this case, β is associated with the delocalized vibrational chromophore due to intermolecular coupling, not individual molecules, such that β effectively becomes $\chi^{(2)}$ if the molecular frame is chosen to align with the laboratory frame. We note that β refers to the hyperpolarizability tensor of the smallest independent molecular unit of the system, which is larger than a single molecule if all couplings are considered. As long as not all β elements in equation 6 are zero or cancel out, $\chi^{(2)}_{zyx}$ does not vanish. Hence, chiral SFG is allowed (blue, Figure 3.4A). Finally, when both intramolecular and intermolecular couplings contribute, the net effect is a combination of the second and third cases, yielding larger chiral SFG signals (black, Figure 3.4A). We emphasize that both types of coupling are necessary to obtain the full lineshape, although we cannot generally predict the ratio of intramolecular vs. intermolecular contributions in all systems.

The above arguments can also be applied to elucidate the simulated achiral spectra shown in Figure 3.4B, where the signals do not significantly depend on vibrational coupling. For achiral SFG experiments performed using the *ssp* polarization response, the effective $\chi^{(2)}_{ssp}$ arises from $\chi^{(2)}_{xxz}$ and $\chi^{(2)}_{yyz}$ according to the definition of coordinates presented in Figure 3.2, where $\chi^{(2)}_{xxz} = \chi^{(2)}_{yyz}$ due to isotropy in the *xy* plane. Thus, $\chi^{(2)}_{ssp}$ can be expressed in terms of β elements as in equation 11 that contains β_{ccc}. This element does not vanish

even for linear structures when β is averaged over all rotations about the *z*-axis.[16] Hence, in the case of no couplings, where the vibrational chromophore has $C_{\infty v}$ symmetry, the nonzero element β_{ccc} can contribute to the achiral signal (purple, Figure 3.4B). When intramolecular and/or intermolecular couplings are considered, the vibrational chromophores have less symmetry, which leads to more nonzero β elements that can contribute to $\chi_{ssp}^{(2)}$ as described in equation 11.

The differing sensitivity to couplings between chiral responses (Figure 3.4A) and achiral responses (Figure 3.4B) may also be due to the fundamentally different origins of the signals. Achiral signals are related to the magnitude of the *z*-component of the total system dipole, resolved by vibrational frequency, while chiral responses result from imperfect cancellation of hyperpolarizability terms and have no direct dependence on the overall dipole. Hence, achiral SFG signal intensity does not depend strongly on couplings or any other symmetry-breaking features unless those features increase the orientational anisotropy of the system as a whole along the *z*-axis.

Although the coupling effects observed for the protein-water system (Figure 3.4A) generally agree with those observed for the water dimer system (Figure 3.5A), intermolecular couplings alone produce a larger effect in the protein-water system than in the model water dimer. It seems that the large effect of intermolecular couplings shown in Figure 3.4A may require water assemblies larger than a dimer or a shorter inter-water distance than assumed in the model system. The intramolecular couplings may also be somewhat inflated relative to the intermolecular couplings in this simple model because gas-phase values of mapped quantities are used. The intramolecular coupling in the gas phase is around -55 cm^{-1}, whereas the average value in bulk water is approximately -30

cm^{-1}.[49] We decided to use gas-phase values for the model system (i.e., ignore the electric field) to demonstrate that the effect of couplings is mainly a geometric or topological effect, rather than the result of emergent chirality in local fields. Local fields may still play a role in the real system, but they are not necessary for the coupling effect.

Intramolecular and intermolecular couplings can have synergistic effects. For the C_{3v} dimer system, intermolecular coupling alone produces a weaker signal than when all couplings are included, even though intramolecular coupling alone produces no signal (Figure 3.5B). This observation suggests that intramolecular couplings contribute to the signal, but only if allowed to do so by a symmetry-breaking feature such as intermolecular couplings. This effect is not seen in the methane (T_d) dimer (Figure 3.5C), suggesting that the synergistic effect of intramolecular and intermolecular couplings is system-dependent. Other symmetry-breaking features such as varying local electric fields may also be able to unlock chiral signals. However, all of our results suggest that intermolecular vibrational couplings alone are sufficient to unlock chiral signals in the simulated spectra.

The C_{2v} point group is effectively just asymmetric enough that the surviving hyperpolarizability components can give rise to a chiral SFG signal (equation 7). This is under the condition that the C_{2v} molecules are assembled in chiral supramolecular assemblies at an interface with C_∞ symmetry. Molecules of the C_{3v} point group, for example, have relatively higher symmetry, so all components of the hyperpolarizability that can give rise to chiral signals disappear.[14] The same is true for other highly symmetric solvents. Although chiral SFG studies have focused on water interacting with biomolecules, our work suggests that any solvent surrounding a chiral solute at an interface with the C_∞ symmetry may be able to generate a chiral SFG signal, provided that some

symmetry-breaking features are present. Our results show that intermolecular vibrational couplings can be one of these features and generate chiral signals in simulated SFG spectra. This is consistent with theory developed by Zanni in 2012, which predicts nonzero chiral (2D-)SFG signals in pairs of achiral but coupled chromophores.[46]

The results in Figures 3.5B and 3.5C have significant implications for the study of the chiral SFG responses of non-aqueous solvents, some of which are highly symmetric molecules or contain highly symmetric functional groups, such the $-CCl_3$ group in chloroform and methyl group in methanol. One of the most common chemical groups in organic solvents is the methyl group, which has C_{3v} symmetry. Figure 3.5B shows that a C_{3v} molecule (i.e., ammonia) can produce chiral SFG signals if heterogeneous intermolecular couplings are present. As the calculations in Figure 3.5 assume little about each chemical system except geometry and the presence of couplings, the results for one C_{3v} molecule translate qualitatively to all C_{3v} molecules and functional groups. Similarly, the results for methane translate qualitatively to tetrahedral solvents such as carbon tetrachloride. Accordingly, the symmetry-breaking feature of intermolecular (or more precisely, inter-functional group) couplings may explain the prominent chiral peak assigned to the methyl asymmetric stretch observed by Wang et al. in the antiparallel β-sheet $LK_7\beta$ peptides at the air-water interface.[57] Moreover, Geiger and coworkers observed a chiral SFG signal for methyl groups in DNA.[14, 33] This observation has been a puzzle in the field because C_{3v} entities are forbidden to produce chiral SFG signals based on symmetry arguments. However, our results show that vibrational coupling or other environmental factors (e.g., local molecular interactions or electric field) can potentially break the symmetry and induce chiral SFG signals.

In this paper, we have discussed coupling between individual O-H vibrational modes with linear symmetry and found that coupling of some kind is essential to break the symmetry and allow for a chiral SFG signal. There are many vibrational modes, however, that involve symmetry-breaking coupling within the mode itself. This includes the amide I vibrational mode, which consists of both a carbonyl stretch and an N-H bend, so it is not surprising that Skinner reports only a modest effect of couplings on the chiral SFG spectrum in the amide I region.[15]

Both the symmetric and asymmetric vibrational modes of water involve both O-H groups, but in isotopically labeled variants such as HOD, intramolecular coupling is minimal due to the difference in vibrational frequency of O-H (~3750 cm^{-1}) and O-D (~2750 cm^{-1}).[13, 58] In this case, intramolecular coupling is ignored, and the symmetry-based theory predicts no chiral SFG signal. However, our results indicate that intermolecular couplings contribute substantially, and they alone can produce a basic lineshape that is similar to that for the combination of intramolecular and intermolecular couplings (Figure 3.4A). Hence, our simulations imply that the O-H stretch in HOD systems surrounding chiral solutes will produce a chiral response. This is in agreement with isotope dilution experiments by Perets et al.[24]

Conclusion

This work examines the origins of the chiral SFG signal of the O-H stretch from aqueous interfaces containing chiral biomacromolecular solutes. Although vibrational couplings change the achiral SFG spectrum only slightly,[56] they are absolutely essential to the simulation of a chiral response. Our analysis indicates that vibrational couplings change

the effective geometry of vibrational chromophores and introduce crucial asymmetry to induce chiral SFG responses. Our calculations show that the dependence on coupling arises even for a simple model system composed of two water molecules in a fixed geometry. Therefore, the effect of couplings on chiral SFG is not a phenomenon arising only in complex macromolecular systems.

Our findings have implications for theorists modeling SFG responses. Importantly, researchers using any method that omits vibrational couplings between O-H groups (or other bonds) must exercise caution when modeling chiral SFG signals. Intramolecular and intermolecular couplings are found to have comparable contributions to the chiral SFG response of water. Including intermolecular couplings as well as intramolecular couplings is not critical to qualitatively capturing a chiral response, although it is necessary to obtain the full signal. In contrast, for molecules with higher symmetry than water, the symmetry-breaking effect of intermolecular couplings is absolutely required for the appearance of chiral SFG signals in simulated spectra. Nonetheless, experimental detection of such signals depends on many factors (e.g., signal-to-noise level) and thus remains to be examined. Other symmetry-breaking effects, such as local molecular interactions and applications of external electric fields, may also be effective in inducing chiral SFG signals, and these remain to be explored with experiments and simulations. This work provides the foundation for further investigations that may examine the chiral templating of non-aqueous solvents surrounding non-biological macromolecules.

Summary of Appendix to Chapter 3

Contains additional discussion of some of the symmetry arguments discussed in this paper, details of calculation of chiral SFG spectra from MD simulation trajectories, MD equilibration details, achiral (*ssp*) SFG response of water in terms of hyperpolarizability elements, details related to the model system calculations, and chiral SFG signal convergence information.

Acknowledgments and Funding Sources

The authors thank Dr. Pablo E. Videla and Valentina Rangel Angarita at Yale University for helpful discussions. This work was supported by the National Institutes of Health Grant R35 GM139449 (S.H.-S.) and the NSF Grant CHE-2108690 (E.C.Y.Y.). D.K. was also supported by the NIH (5T32GM008283-32). E.A.P. was supported by the NIH (5T32GM008283-31) and a John C. Tully Chemistry Research Fellowship. L.V. acknowledges support from the NSF (CHE-1753207).

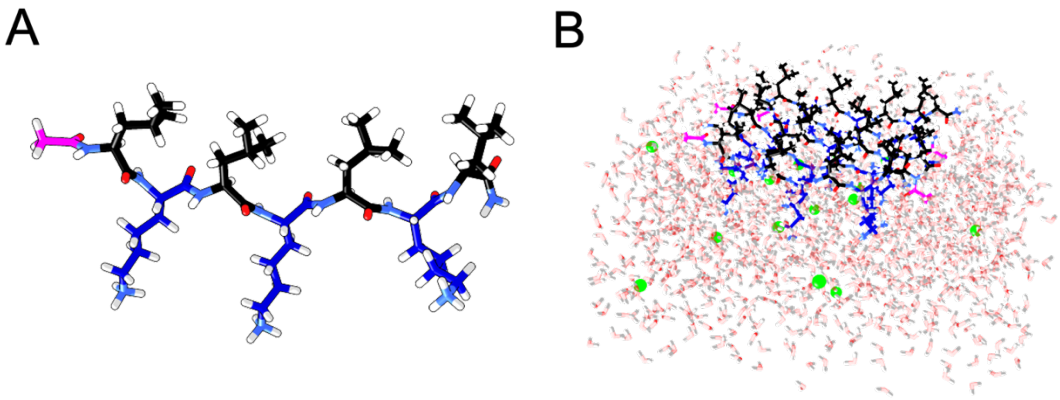

Figure 3.1. The $LK_7\beta$ (LKLKLKL) peptide model system. (A) Each $LK_7\beta$ peptide consists of alternating leucine (black) and lysine (blue) residues and is capped with an acetyl group (magenta) on the N-terminus and an NH_2 group at the C-terminus. (B) An antiparallel β-sheet is constructed at the vacuum-water interface from five $LK_7\beta$ peptides arranged such that leucine sidechains point toward the vacuum and lysine sidechains point into the water. The system is neutralized with Cl^- ions (green). Approximately 1350 water molecules surround the peptide. The simulation box is roughly 45 × 45 × 45 Å. Half of the box is composed of the system and the other half is vacuum.

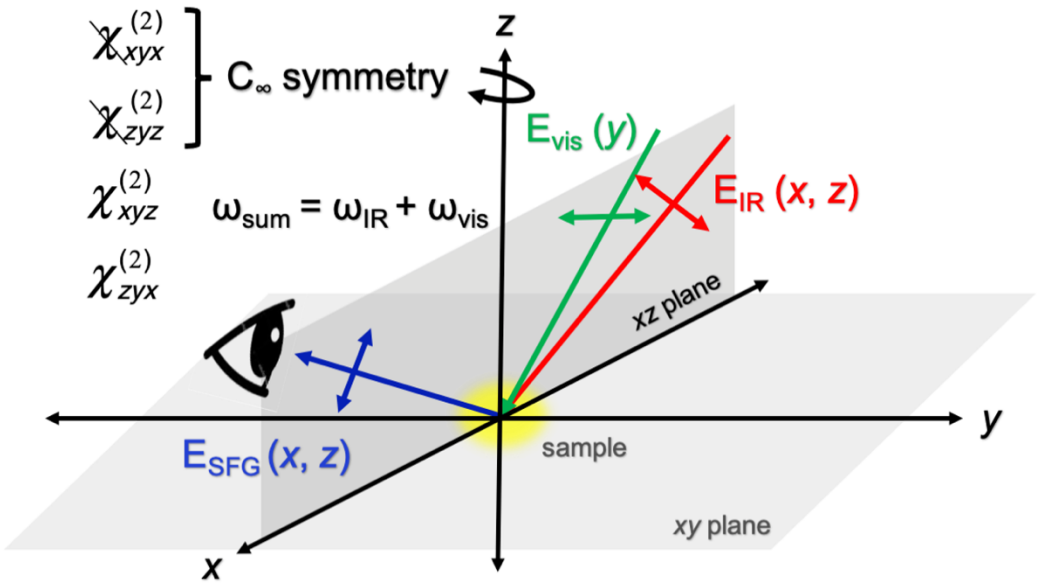

Figure 3.2. Schematic depiction of a vibrational SFG experiment with the *psp* polarization setting, where *p* indicates polarization on the plane of incidence (the *xz* plane) and *s* indicates polarization perpendicular to that plane (i.e., along *y*). The polarization is specified in the order of sum frequency (detector), visible, and infrared. The *psp* polarization setting isolates four elements of the second order susceptibility tensor ($\chi^{(2)}$). Among these four elements, $\chi^{(2)}_{xyx} = \chi^{(2)}_{zyz} = 0$ due to the C_∞ symmetry of a non-centrosymmetric interface (*xy*-plane) about the *z*-axis and $\chi^{(2)}_{xyz} = 0$ given the symmetry of the Raman polarizability tensor in the absence of electronic resonance. Ignoring Fresnel factors, the chiral response is given by $\chi^{(2)}_{psp} \propto \chi^{(2)}_{zyx}$.

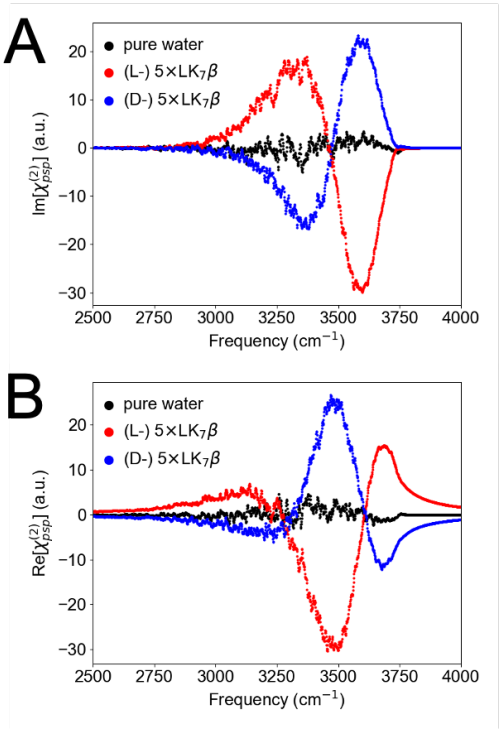

Figure 3.3. (A) Imaginary and (B) real components of the *psp* SFG response of the O-H stretch of water around a pentamer of (L-) LK$_7$β (red) and its (D-) enantiomer (blue) at the vacuum-water interface (see Figure 3.1 for structure) as well as a pure-water interface with the same simulation box size (black). Spectra are averaged over 10,000,000 frames from 100 ns of MD simulation. Spectra are unnormalized.

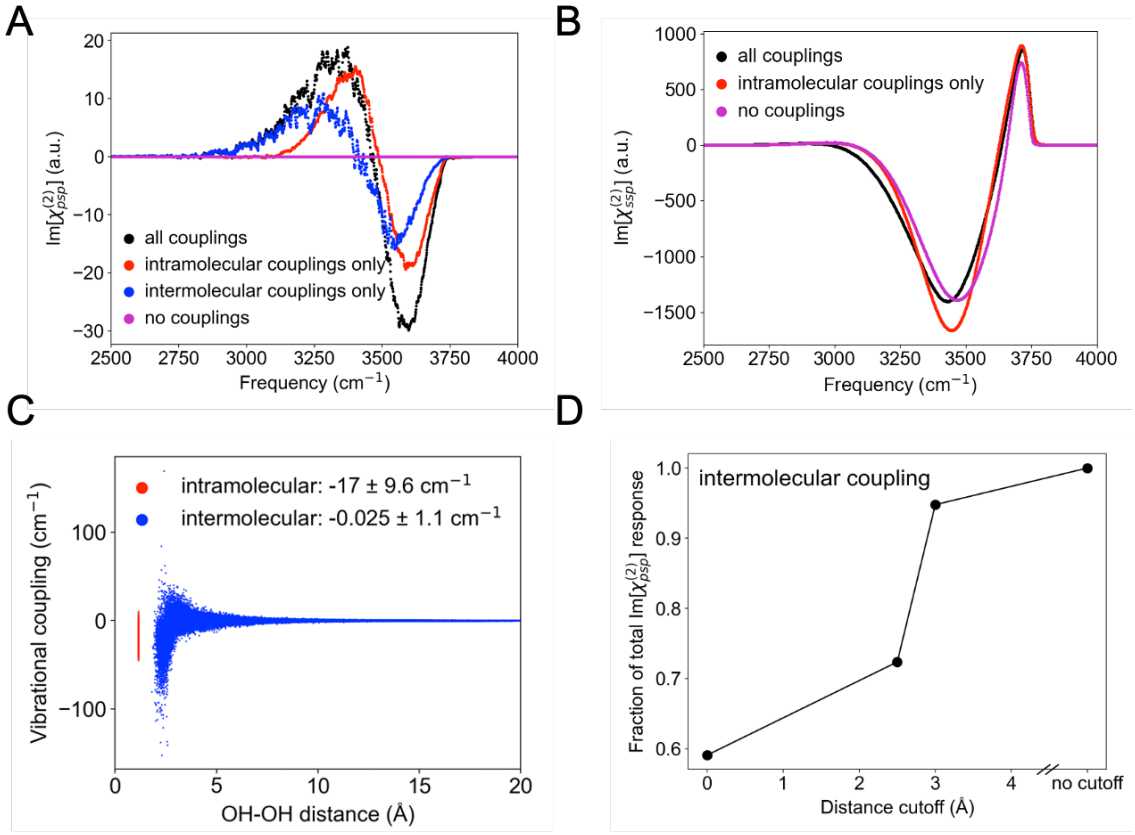

Figure 3.4. (A) Imaginary chiral (*psp*) SFG response of an (L-) LK$_7$β pentamer at the vacuum-water interface with different treatments of O-H/O-H vibrational coupling. (B) Imaginary achiral (*ssp*) SFG response of the same (L-) LK$_7$β system. (C) Intramolecular and intermolecular coupling magnitudes for the chiral SFG response between O-H groups among waters in the LK$_7$β system as a function of dipole-dipole distance, where each dot represents coupling between two OH groups for one configuration obtained from an MD trajectory. The mean and standard deviation are given for each type of coupling. The intramolecular couplings (red) are given for a single distance (1.14 Å) and thus correspond to a vertical line. (D) Fraction of total chiral SFG response (the sum of the absolute values of the responses for all frequencies) versus intermolecular coupling distance cutoff, where the distances were computed using effective dipole locations for each O-H bond.

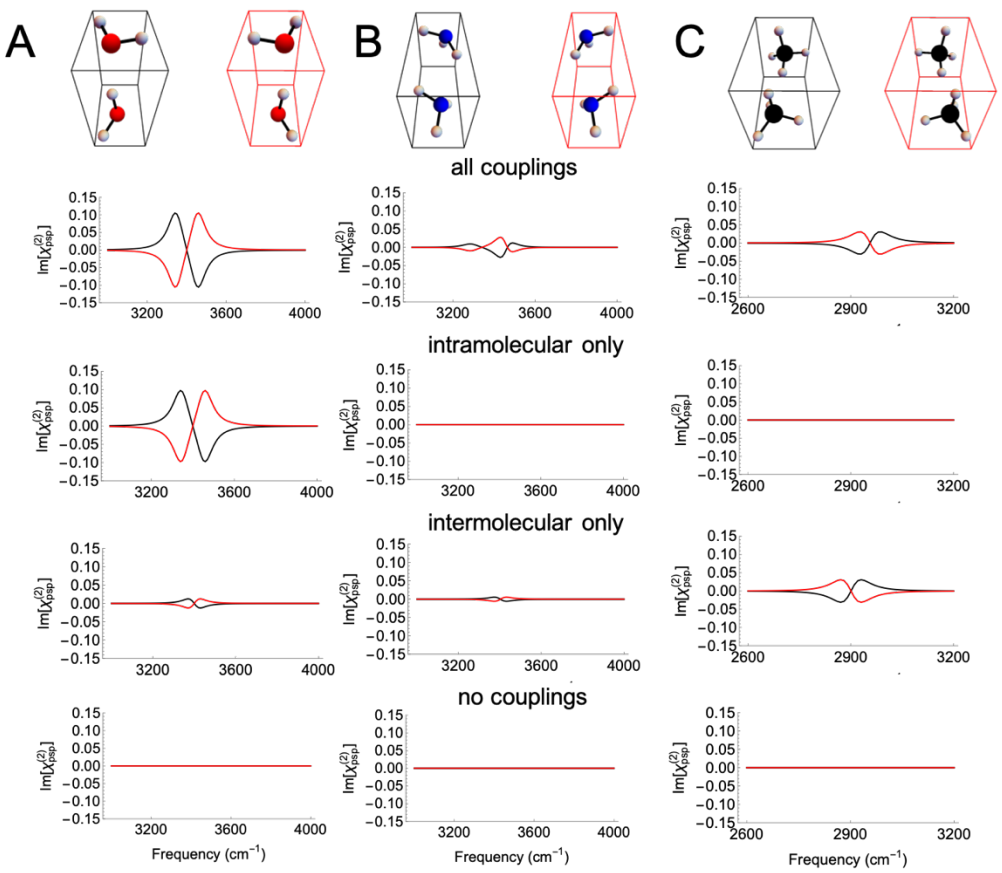

Figure 3.5. Effect of coupling on model dimer systems: (A) water dimer, (B) ammonia dimer, (C) methane dimer. Top: Enantiomers of the dimers. Bottom: simulated chiral SFG spectra of the dimers in mirror images (black and red) for various combinations of vibrational couplings. The peak was set at 3400 cm^{-1} for water and ammonia and 2900 cm^{-1} for methane and the couplings were calculated using Skinner's approach assuming no electric field, where the couplings still depend on dipole alignment and distance between dipoles (see Methods).

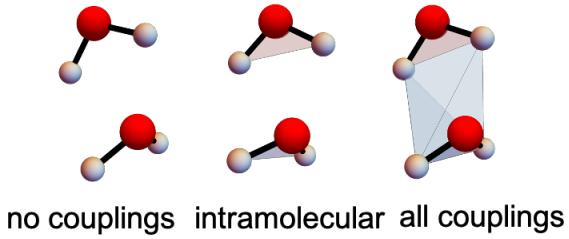

Figure 3.6. Schematic showing how including vibrational couplings effectively changes the geometry (and symmetry) of a molecular system.

References

1. Horn, H. W.; Swope, W. C.; Pitera, J. W.; Madura, J. D.; Dick, T. J.; Hura, G. L.; Head-Gordon, T. Development of an improved four-site water model for biomolecular simulations: TIP4P-Ew. *J. Chem. Phys.* **2004,** *120*, 9665-9678.
2. Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theory Comput.* **2015,** *11*, 3696-3713.
3. Miyamoto, S.; Kollman, P. A. Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models. *J. Comput. Chem.* **1992,** *13*, 952-962.
4. Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C. 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.
5. Michaud-Agrawal, N.; Denning, E. J.; Woolf, T. B.; Beckstein, O. MDAnalysis: A toolkit for the analysis of molecular dynamics simulations. *J. Comput. Chem.* **2011,** *32*, 2319-2327.
6. Gowers, R. J.; Linke, M.; Barnoud, J.; Reddy, T. J. E.; Melo, M. N.; Seyler, S. L.; Domanski, J.; Dotson, D. L.; Buchoux, S.; Kenney, I. M. *MDAnalysis: a Python package for the rapid analysis of molecular dynamics simulations*; 2575-9752; Los Alamos National Lab.(LANL), Los Alamos, NM (United States): 2019.
7. Hunter, J. Matplotlib: A 2D Graphics Environment. *Comput. Sci. Eng.* **2007,** *9*, 90-95.
8. McKinney, W. In *Data structures for statistical computing in python*, Proceedings of the 9th Python in Science Conference, Austin, TX: 2010; pp 51-56.
9. Lam, S.; Pitrou, A.; Seibert, S., Proceedings of the Second Workshop on the LLVM Compiler Infrastructure in HPC. LLVM'15. ACM New York, NY, USA: 2015.
10. McGibbon, Robert T.; Beauchamp, Kyle A.; Harrigan, Matthew P.; Klein, C.; Swails, Jason M.; Hernández, Carlos X.; Schwantes, Christian R.; Wang, L.-P.; Lane, Thomas J.; Pande, Vijay S. MDTraj: A Modern Open Library for the Analysis of Molecular Dynamics Trajectories. *Biophys. J.* **2015,** *109*, 1528-1532.
11. Auer, B. M.; Skinner, J. L. Dynamical effects in line shapes for coupled chromophores: Time-averaging approximation. *J. Chem. Phys.* **2007,** *127*, 104105.
12. Auer, B. M.; Skinner, J. L. IR and Raman spectra of liquid water: Theory and interpretation. *J. Chem. Phys.* **2008,** *128*, 224511.
13. Pieniazek, P. A.; Tainter, C. J.; Skinner, J. L. Interpretation of the water surface vibrational sum-frequency spectrum. *J. Chem. Phys.* **2011,** *135*, 044701.
14. Moad, A. J.; Simpson, G. J. A Unified Treatment of Selection Rules and Symmetry Relations for Sum-Frequency and Second Harmonic Spectroscopies. *J. Phys. Chem. B* **2004,** *108*, 3548-3562.
15. Carr, J. K.; Wang, L.; Roy, S.; Skinner, J. L. Theoretical Sum Frequency Generation Spectroscopy of Peptides. *J. Phys. Chem. B* **2015,** *119*, 8969-8983.
16. Wang, H.-F.; Gan, W.; Lu, R.; Rao, Y.; Wu, B.-H. Quantitative spectral and orientational analysis in surface sum frequency generation vibrational spectroscopy (SFG-VS). *Int. Rev. Phys. Chem.* **2005,** *24*, 191-256.

17. Fu, L.; Xiao, D.; Wang, Z.; Batista, V. S.; Yan, E. C. Y. Chiral Sum Frequency Generation for In Situ Probing Proton Exchange in Antiparallel β-Sheets at Interfaces. *J. Am. Chem. Soc.* **2013,** *135*, 3592-3598.
18. 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.
19. Hu, X.-H.; Wei, F.; Wang, H.; Wang, H.-F. α-Quartz Crystal as Absolute Intensity and Phase Standard in Sum-Frequency Generation Vibrational Spectroscopy. *J. Phys. Chem. C* **2019,** *123*, 15071-15086.
20. Hu, X.-H.; Fu, L.; Hou, J.; Zhang, Y.-N.; Zhang, Z.; Wang, H.-F. N–H Chirality in Folded Peptide LK7β Is Governed by the Cα–H Chirality. *J. Phys. Chem. Lett.* **2020,** *11*, 1282-1290.
21. Velarde, L.; Zhang, X.-y.; Lu, Z.; Joly, A. G.; Wang, Z.; Wang, H.-f. Communication: Spectroscopic phase and lineshapes in high-resolution broadband sum frequency vibrational spectroscopy: Resolving interfacial inhomogeneities of "identical" molecular groups. *J. Chem. Phys.* **2011,** *135*, 241102.
22. 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.
23. McDermott, M. L.; Vanselous, H.; Corcelli, S. A.; Petersen, P. B. DNA's Chiral Spine of Hydration. *ACS Cent. Sci.* **2017,** *3*, 708-714.
24. Perets, E. A.; Yan, E. C. Y. Chiral Water Superstructures around Antiparallel β-Sheets Observed by Chiral Vibrational Sum Frequency Generation Spectroscopy. *J. Phys. Chem. Lett.* **2019,** *10*, 3395-3401.
25. Perets, E. A.; Yan, E. C. Y. The H2O Helix: The Chiral Water Superstructure Surrounding DNA. *ACS Cent. Sci.* **2017,** *3*, 683-685.
26. Vanselous, H.; Petersen, P. B. Extending the capabilities of heterodyne-detected sum-frequency generation spectroscopy: probing any Interface in any polarization combination. *J. Phys. Chem. C* **2016,** *120*, 8175−8184.
27. Perets, E. A.; Konstantinovsky, D.; Fu, L.; Chen, J.; Wang, H.-F.; Hammes-Schiffer, S.; Yan, E. C. Y. Mirror-image antiparallel β-sheets organize water molecules into superstructures of opposite chirality. *Proc. Natl. Acad. Sci. USA* **2020,** *117*, 32902-32909.
28. Wang, J.; Even, M. A.; Chen, X.; Schmaier, A. H.; Waite, J. H.; Chen, Z. Detection of Amide I Signals of Interfacial Proteins in Situ Using SFG. *J. Am. Chem. Soc.* **2003,** *125*, 9914-9915.
29. Hosseinpour, S.; Roeters, S. J.; Bonn, M.; Peukert, W.; Woutersen, S.; Weidner, T. Structure and Dynamics of Interfacial Peptides and Proteins from Vibrational Sum-Frequency Generation Spectroscopy. *Chem. Rev.* **2020,** *120*, 3420-3465.
30. Fu, L.; Liu, J.; Yan, E. C. Y. Chiral Sum Frequency Generation Spectroscopy for Characterizing Protein Secondary Structures at Interfaces. *J. Am. Chem. Soc.* **2011,** *133*, 8094-8097.
31. Perets, E. A.; Videla, P. E.; Yan, E. C. Y.; Batista, V. S. Chiral inversion of amino acids in antiparallel beta sheets at interfaces probed by vibrational sum frequency generation spectroscopy. *J. Phys. Chem. B* **2019,** *123*, 5769−5781.

32. Wurpel, G. W. H.; Sovago, M.; Bonn, M. Sensitive Probing of DNA Binding to a Cationic Lipid Monolayer. *J. Am. Chem. Soc.* **2007,** *129*, 8420-8421.
33. Stokes, G. Y.; Gibbs-Davis, J. M.; Boman, F. C.; Stepp, B. R.; Condie, A. G.; Nguyen, S. T.; Geiger, F. M. Making "Sense" of DNA. *J. Am. Chem. Soc.* **2007,** *129*, 7492-7493.
34. Walter, S. R.; Geiger, F. M. DNA on Stage: Showcasing Oligonucleotides at Surfaces and Interfaces with Second Harmonic and Vibrational Sum Frequency Generation. *J. Phys. Chem. Lett.* **2010,** *1*, 9-15.
35. Tan, J.; Zhang, J.; Li, C.; Luo, Y.; Ye, S. Ultrafast energy relaxation dynamics of amide I vibrations coupled with protein-bound water molecules. *Nat. Commun.* **2019,** *10*.
36. Fu, L.; Chen, S.-L.; Wang, H.-F. Validation of Spectra and Phase in Sub-1 cm–1 Resolution Sum-Frequency Generation Vibrational Spectroscopy through Internal Heterodyne Phase-Resolved Measurement. *J. Phys. Chem. B* **2016,** *120*, 1579-1589.
37. Auer, B. M.; Skinner, J. L. Dynamical effects in line shapes for coupled chromophores: Time-averaging approximation. *J. Chem. Phys.* **2007,** *127*, 104105.
38. Auer, B. M.; Skinner, J. L. Vibrational Sum-Frequency Spectroscopy of the Water Liquid/Vapor Interface. *J. Phys. Chem. B* **2009,** *113*, 4125-4130.
39. DeGrado, W. F.; Lear, J. D. Induction of peptide conformation at apolar water interfaces. 1. A study with model peptides of defined hydrophobic periodicity. *J. Am. Chem. Soc.* **1985,** *107*, 7684-7689.
40. Morita, A.; Hynes, J. T. A Theoretical Analysis of the Sum Frequency Generation Spectrum of the Water Surface. II. Time-Dependent Approach. *J. Phys. Chem. B* **2002,** *106*, 673-685.
41. Ohto, T.; Usui, K.; Hasegawa, T.; Bonn, M.; Nagata, Y. Toward ab initio molecular dynamics modeling for sum-frequency generation spectra; an efficient algorithm based on surface-specific velocity-velocity correlation function. *J. Chem. Phys.* **2015,** *143*, 124702.
42. Boyd, R. W., Chapter 1 - The Nonlinear Optical Susceptibility. In *Nonlinear Optics (Fourth Edition)*, Boyd, R. W., Ed. Academic Press: 2020; pp 1-64.
43. Eisenthal, K. B. Liquid Interfaces Probed by Second-Harmonic and Sum-Frequency Spectroscopy. *Chem. Rev.* **1996,** *96*, 1343-1360.
44. Richmond, G. L. Molecular Bonding and Interactions at Aqueous Surfaces as Probed by Vibrational Sum Frequency Spectroscopy. *Chem. Rev.* **2002,** *102*, 2693-2724.
45. Shen, Y. R.; Ostroverkhov, V. Sum-Frequency Vibrational Spectroscopy on Water Interfaces: Polar Orientation of Water Molecules at Interfaces. *Chem. Rev.* **2006,** *106*, 1140-1154.
46. Laaser, J. E.; Zanni, M. T. Extracting Structural Information from the Polarization Dependence of One- and Two-Dimensional Sum Frequency Generation Spectra. *J. Phys. Chem. A* **2013,** *117*, 5875-5890.
47. Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. *J. Chem. Phys.* **1993,** *98*, 10089-10092.
48. Eastman, P.; Swails, J.; Chodera, J. D.; McGibbon, R. T.; Zhao, Y.; Beauchamp, K. A.; Wang, L.-P.; Simmonett, A. C.; Harrigan, M. P.; Stern, C. D., et al. OpenMM 7: Rapid development of high performance algorithms for molecular dynamics. *PLoS Comput. Biol.* **2017,** *13*, e1005659.

49. Auer, B. M.; Skinner, J. L. IR and Raman spectra of liquid water: Theory and interpretation. *J. Chem. Phys.* **2008,** *128*, 224511.
50. McGuire, J. A.; Shen, Y. R. Ultrafast Vibrational Dynamics at Water Interfaces. *Science* **2006,** *313*, 1945-1948.
51. Belch, A. C.; Rice, S. A. The OH stretching spectrum of liquid water: A random network model interpretation. *J. Chem. Phys.* **1983,** *78*, 4817-4823.
52. Harris, C. R.; Millman, K. J.; van der Walt, S. J.; Gommers, R.; Virtanen, P.; Cournapeau, D.; Wieser, E.; Taylor, J.; Berg, S.; Smith, N. J., et al. Array programming with NumPy. *Nature* **2020,** *585*, 357-362.
53. Max, J.-J.; Chapados, C. Infrared spectroscopy of acetone-hexane liquid mixtures. *J. Chem. Phys.* **2007,** *126*, 154511.
54. Scherer, J. R.; Snyder, R. G. Raman intensities of single crystal ice Ih. *J. Chem. Phys.* **1977,** *67*, 4794-4811.
55. Kocsis, I.; Sorci, M.; Vanselous, H.; Murail, S.; Sanders, S. E.; Licsandru, E.; Legrand, Y.-M.; van der Lee, A.; Baaden, M.; Petersen, P. B., et al. Oriented chiral water wires in artificial transmembrane channels. *Sci. Adv.* **2018,** *4*, eaao5603.
56. Kaliannan, N. K.; Henao Aristizabal, A.; Wiebeler, H.; Zysk, F.; Ohto, T.; Nagata, Y.; Kühne, T. D. Impact of intermolecular vibrational coupling effects on the sum-frequency generation spectra of the water/air interface. *Mol. Phys.* **2020,** *118*, 1620358.
57. Wang, Z.; Fu, L.; Yan, E. C. Y. C–H Stretch for Probing Kinetics of Self-Assembly into Macromolecular Chiral Structures at Interfaces by Chiral Sum Frequency Generation Spectroscopy. *Langmuir* **2013,** *29*, 4077-4083.
58. Wiafe-Akenten, J.; Bansil, R. Intermolecular coupling in HOD solutions. *J. Chem. Phys.* **1983,** *78*, 7132-7137.

Appendix to Chapter 3

Combining components of $\chi^{(2)}$ to form the effective *psp* SFG signal

Polarization settings in an SFG experiment isolate linear combinations of particular components of $\chi^{(2)}$. To see which are isolated, it is necessary to define the experimental setup. For a given polarization setup, the effective response is given by[1]

$$\chi^{(2)}_{\text{eff}} = \sum_i \sum_j \sum_k u_i L_{ii,\text{sum}} \cdot \chi^{(2)}_{ijk} \cdot (v_j L_{jj,\text{vis}})(w_k L_{kk,\text{IR}}) \tag{1}$$

where the components L_{ij} are Fresnel factors and $\hat{\mathbf{u}}$, $\hat{\mathbf{v}}$, and $\hat{\mathbf{w}}$ are unit vectors representing the polarization of the optical field of the sum-frequency, visible, and IR beams, respectively. Assuming a *psp* setup and an angle α between the laser table and each beam, these unit vectors are given by

$$\begin{aligned} \hat{\mathrm{u}} &= (-\cos(\alpha), 0, \sin(\alpha)) \\ \hat{\mathrm{v}} &= (0,1,0) \\ \hat{\mathrm{w}} &= (\cos(\alpha), 0, \sin(\alpha)) \end{aligned} \tag{2}$$

which, assuming common Fresnel factors for each beam and each dimension and ignoring angle terms, gives an effective response of

$$\chi^{(2)}_{psp} \propto \chi^{(2)}_{zyx} - \chi^{(2)}_{xyz} + \chi^{(2)}_{zyz} - \chi^{(2)}_{xyx} \tag{3}$$

We effectively neglect the Fresnel factors because calculating them would involve assumptions about the experimental setup that are not necessarily justified. Moreover, it is customary in the SFG simulation $_{\text{field}}$ to neglect the Fresnel factors.[1, 2] For systems with C_∞

symmetry, only components that survive the following integration contribute to SFG signals.

$$\chi^{(2)}_{IJK,\text{transformed}} = \int_0^{2\pi} d\theta \sum_{lmn} R_{il}(\theta) R_{jm}(\theta) R_{kn}(\theta) \chi^{(2)}_{lmn} \tag{4}$$

where **R** is a matrix expressing a rotation of the tensor $\chi^{(2)}$ by the angle θ about the *z*-axis. When this transformation is performed, the last two terms in equation 3 vanish, and the remaining two terms map onto the following in the lab reference frame:

$$\chi^{(2)}_{psp} \propto \chi^{(2)}_{zyx} - \chi^{(2)}_{xyz} \tag{5}$$

Equation 5 is used to extract the *psp* signal from MD simulations, effectively assuming random rotation about the *z*-axis. Please see the next section for details.

Calculating $\chi^{(2)}_{psp}$ from a molecular dynamics trajectory

The previous section assumed that the system of interest has C_∞ symmetry. Although this is true for many experimental systems, it is not true for molecular dynamics trajectories with limited sampling. To avoid this issue, it is not necessary to sample every orientation equally. Instead, one simply assumes that each frame of the trajectory is rotated by a random angle τ about the *z*-axis. In this rotated frame of reference, each component of $\chi^{(2)}$ is transformed according to equation 4. Under this transformation, the component $\chi^{(2)}_{z'y'x'}$ in the rotated frame becomes $\chi^{(2)}_{zyx} - \chi^{(2)}_{zxy}$ in the original frame, and the component $\chi^{(2)}_{x'y'z'}$ becomes $\chi^{(2)}_{yxz} - \chi^{(2)}_{xyz}$. Thus $\chi^{(2)}_{psp}$ is given by

$$\chi^{(2)}_{psp} \propto \chi^{(2)}_{z'y'x'} - \chi^{(2)}_{x'y'z'} = \chi^{(2)}_{zyx} - \chi^{(2)}_{zxy} + \chi^{(2)}_{yxz} - \chi^{(2)}_{xyz} \tag{6}$$

where the rotating frame is identified by primed indices. The hyperpolarizability tensor arises from the outer product of the polarizability and dipole derivatives associated with vibrational normal mode *Q,*

$$\beta^{(2)}_{ijk} \propto \left(\frac{\partial \alpha_{ij}}{\partial Q} \right) \left(\frac{\partial \mu_k}{\partial Q} \right) \tag{7}$$

and the SFG response arises from the hyperpolarizability as in equation 3.5 in Chapter 3. Thus, the last two terms in equation 6 vanish due to symmetry of the Raman tensor far from electronic resonance, leaving

$$\chi^{(2)}_{psp} \propto \chi^{(2)}_{zyx} - \chi^{(2)}_{zxy} \tag{8}$$

Equation 8 allows one to calculate the response of a simulated system *as if* it were freely rotating about *z*, so this equation is relevant only for computing chiral SFG signals from MD simulations. See ref. 1 for more details.

Molecular dynamics (MD) equilibration details

First the energy of the system composed of the protein, solvent, and ions was minimized while keeping the protein restrained by harmonic restraints with force constants 200 kcal/mol• $Å^{-2}$. Next, the solvent and ions were equilibrated with MD in an NVT ensemble for 500 ps at 298 K. Then the hydrogen atoms in the protein were relaxed by energy minimization, followed by three rounds of energy minimization of sidechains, reducing the restraints on the backbone atoms from 100 kcal/mol•$Å^{-2}$, to 50 kcal/mol•$Å^{-2}$, and finally to 10 kcal/mol•$Å^{-2}$. The energy of the entire system was then minimized with no restraints. Following minimization, the system was heated at constant volume from 0 K

to 298 K in seven steps, each containing 10 ps of temperature ramping and 50 ps of constant temperature MD. The entire system was equilibrated for 6 ns in an NVT ensemble. NPT equilibration was not performed to avoid damaging the vacuum-water interface. Lennard-Jones interactions were cut off at 14 Å with a switching function starting at 12.6 Å to smooth the transition. Long-range electrostatic interactions were treated with the particle-mesh Ewald method.[3] The water molecule geometry was constrained with the SETTLE algorithm,[4] while protein bonds involving hydrogen were constrained with the SHAKE algorithm.[5] The timestep for all equilibration and production MD was 1 fs.

Trajectory wrapping details

For simplicity, we used 3D periodic boundary conditions for our MD simulations. These boundary conditions led to an inconvenient wrapping of water molecules from the "bottom" of the slab up to the top of the system, where they created a fictitious interface above the vacuum. Our wrapping code simply returned these water molecules to the bottom of the slab. This wrapping was required only for achiral SFG calculations, where the second (bottom) surface had to be damped out (see Chapter 3).

Impact of bond polarizability on the chiral SFG response of water

This work uses the bond polarizability model[6] to describe the polarizability of water molecules. In the bond polarizability model, the polarizability of a molecule is separated into polarizabilities associated with each bond. In the bond's frame of reference, assuming the bond points along the *x*-axis, the polarizability derivative matrix associated with a vibrational transition is given by

$$\begin{pmatrix} \alpha'_{\parallel} & 0 & 0 \\ 0 & \alpha'_{\perp} & 0 \\ 0 & 0 & \alpha'_{\perp} \end{pmatrix} \tag{9}$$

where $\alpha'_{\parallel}$ is the component of the polarizability derivative associated with the vibrational transition along the bond and $\alpha'_{\perp}$ is the component perpendicular to the bond. In the laboratory frame of reference, the components of the transition polarizability are given by the following expression[7]

$$\alpha'_{ij} = x_{01}((\alpha'_{\parallel} - \alpha'_{\perp})u_i u_j + \alpha'_{\perp}\delta_{ij}) \tag{10}$$

where x_{01} is the change in bond length associated with vibrational excitation, δ_{ij} is the Kronecker delta function, and $\hat{\mathbf{u}}$ is the bond unit vector. The ratio $\alpha'_{\parallel} / \alpha'_{\perp}$ is a key quantity in this expression and is assumed to be 5.6 for the O-H bonds of water in this work.[7] If the ratio is equal to 1 ($\alpha'_{\parallel} = \alpha'_{\perp}$), the polarizability cannot have off-diagonal terms in the laboratory frame, and thus all components of $\chi^{(2)}$ except $\chi^{(2)}_{xxz}$, $\chi^{(2)}_{yyz}$, and $\chi^{(2)}_{zzz}$ become zero, and no chiral SFG signal is possible within the bond polarizability approximation. For a simple molecular system such as a water dimer, the value of $\alpha'_{\parallel} / \alpha'_{\perp}$ affects the sign of the peaks but not their location (Figure A3.2).

Achiral (ssp) SFG response of water in terms of hyperpolarizability elements

Water belongs to the C_{2v} point group, for which only some hyperpolarizability elements survive (see the Symmetry-based chiral SFG theory section in Chapter 3).[8] These elements give rise to SFG signals in a water-containing system. The chiral SFG signal is given by equations 3.6 and 3.7 in Chapter 3. The achiral (*ssp*) signal is captured by the $\chi^{(2)}$ element $\chi^{(2)}_{xxz}$, which is given by the following equation

$$\chi^{(2)}_{xxz} = \frac{1}{2}\cos(\theta)\{\beta_{ccc}\sin^2(\theta) + \cos^2(\psi)(\beta_{aac} + \beta_{bbc}\cos^2(\theta) - 2\beta_{bcb}\sin^2(\theta)) + (\beta_{bbc} + \beta_{aac}\cos^2(\theta) - 2\beta_{aca}\sin^2(\theta))\sin^2(\psi)\} \quad (11)$$

where θ and ψ are Euler rotation angles mapping the molecular frame to the lab frame, and the *c*-axis aligns with the primary (C_2) rotation axis of the water molecule. This result is averaged over all values of the first Euler angle ϕ to reflect the C_∞ symmetry of the system about the *z*-axis.

Connection between point group, chiral SFG, and vibrational coupling

Vibrational SFG signals ultimately arise from the hyperpolarizability tensors of the molecules in the system, as in equations 6 and 7 in Chapter 3. The point group of the constituent molecules determines which elements of the hyperpolarizability tensor survive. For more information and references to further reading on the subject, see ref. 9, Appendix A.

As mentioned in Chapter 3, when vibrational couplings between O-H groups in an aqueous system are ignored, the system is effectively modeled as a collection of linear (O-

H) oscillators that happen to be arranged into the shape of molecules of water. Since there is no cross-talk between O-H bonds, the molecular entity of interest is a linear structure with the point group $C_{\infty v}$. In molecules of the $C_{\infty v}$ point group, only the elements β_{ccc}, β_{aac}, and β_{bbc} survive, where $\beta_{aac} = \beta_{bbc} = k\beta_{ccc}$, where k is a constant, a, b, and c denote axes in the molecular frame, where the z-axis points along the bond.[9] In a system with C_{∞} symmetry about the z-axis, components of $\chi^{(2)}$ are given by

$$\chi^{(2)}_{IJK} \propto \frac{1}{2\pi}\int_0^{2\pi} d\phi \sum_{i,j,k} \left\langle R_{Ii}(\psi) R_{Jj}(\theta) R_{Kk}(\phi) \right\rangle \beta_{ijk} \tag{12}$$

and

$$\chi^{(2)}_{zyx} = 0 \tag{13}$$

The same is true for other chiral components of $\chi^{(2)}$. Thus, no SFG signal is possible when couplings are ignored. When intramolecular couplings are considered, the minimal molecular unit in the system becomes an entire water molecule, which has point group C_{2v} and can give rise to chiral signals, as discussed in Chapter 3 (equations 5-7).

Molecules with the C_{3v} point group cannot produce chiral signals without intermolecular coupling. In this point group, 11 components of β survive, of which four are independent when the Raman tensor is symmetric (i.e., there is no electronic resonance[9]), but these are not enough for chiral components of $\chi^{(2)}$ to survive the transformation in equation 12. However, when intermolecular couplings are included, the fundamental molecular unit of the system becomes, in effect, the entire system due to vibrational delocalization, and if a chiral solute is present, all components of β survive, and β becomes effectively equivalent to $\chi^{(2)}$, as discussed in Chapter 3.

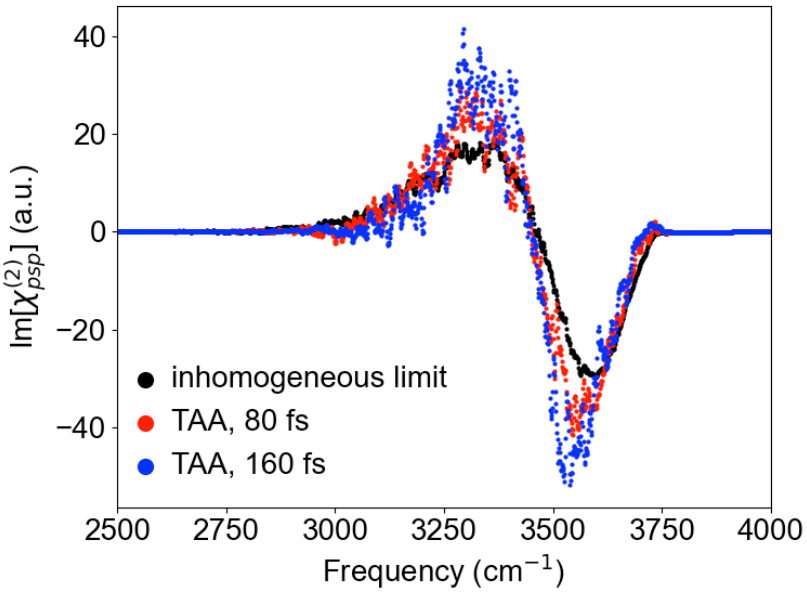

Figure A3.1. Imaginary component of the *psp* SFG response for the water O-H stretch around a pentamer of (L-) LK7β at the vacuum-water interface computed using the time-averaging approximation (TAA)[10] with 80 fs (red) or 160 fs (blue) averaging time intervals. These signals are noisier than the signal obtained from the calculation using the inhomogeneous limit (i.e., a simple average of conformations, shown in black), but show a similar lineshape.

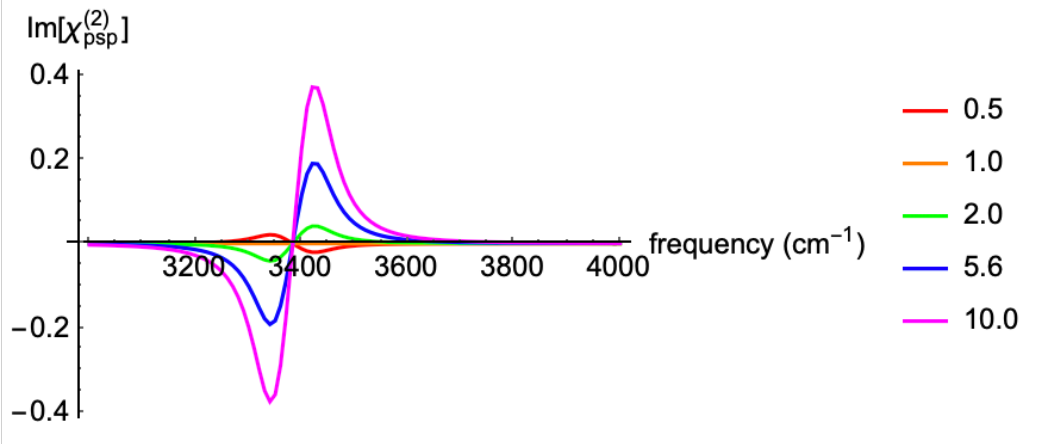

Figure A3.2. Imaginary component of the *psp* SFG response for a chiral water dimer configuration that could arise near a biomacromolecule. Each curve represents the SFG response calculated with a different $\alpha'_{\parallel}/\alpha'_{\perp}$ ratio (see legend), where $\alpha'_{\parallel}$ and $\alpha'_{\perp}$ are polarizability derivatives associated with vibrational excitation of an O-H bond in water (see text). This ratio affects the sign and magnitude of the chiral SFG signal for a water dimer model system but has no effect on peak frequencies. A ratio of one causes the chiral signal to disappear.

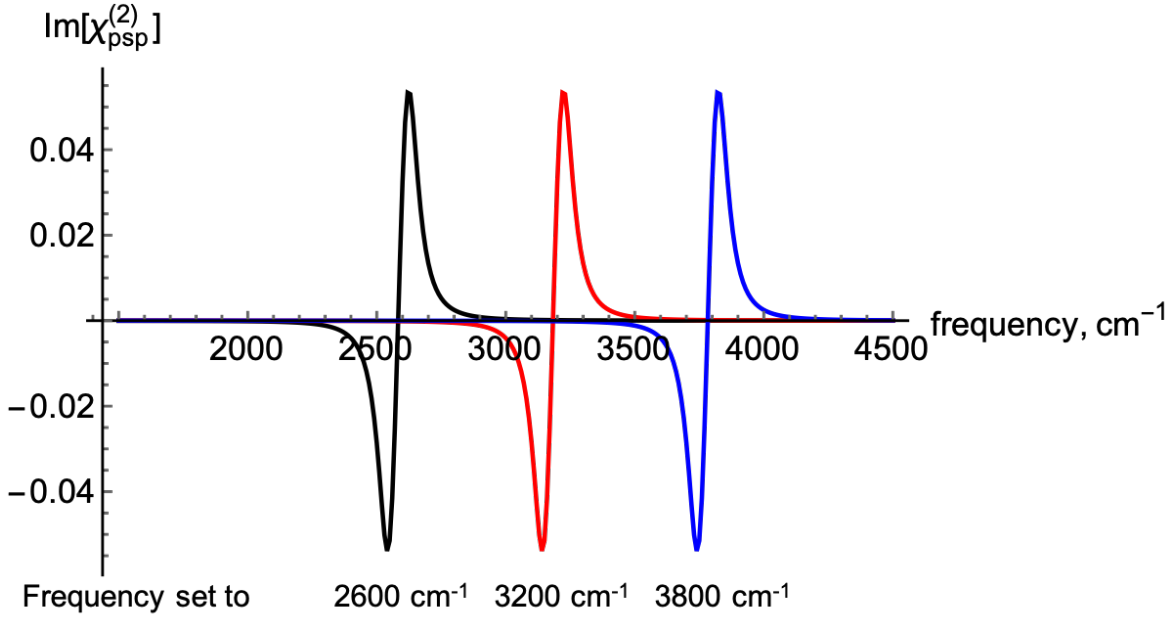

Figure A3.3. Imaginary component of the *psp* response for a chiral water dimer model system with different resonance frequencies. Varying the vibrational resonance frequency does not affect the chiral SFG lineshape, except by shifting it, so the choice of O-H stretch frequency for the water dimer model system is unimportant.

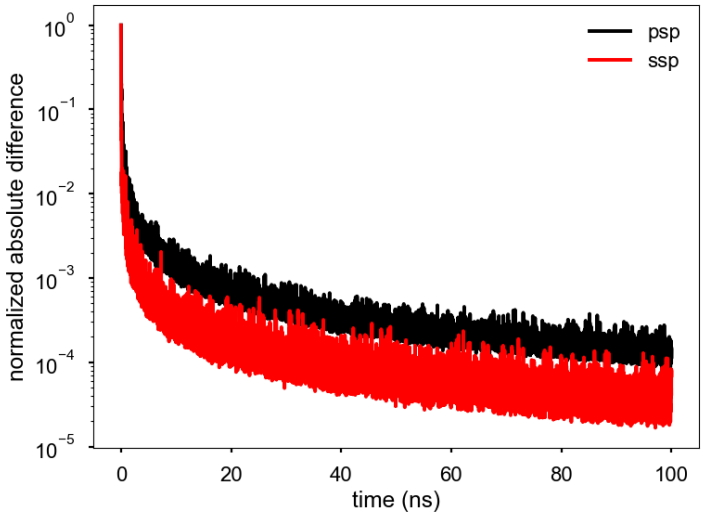

Figure A3.4. Convergence of the imaginary component of the *psp* chiral (black) and *ssp* achiral (red) responses. The normalized absolute difference is calculated by the following steps: (1) calculate the average signal at each timestep for each frequency, (2) subtract the result from the average calculated at the previous timestep, (3) take the absolute value of the differences, (4) sum the result over all frequencies, and (5) normalize to the initial absolute difference. The chiral SFG spectrum converges about an order of magnitude more slowly than the achiral spectrum.

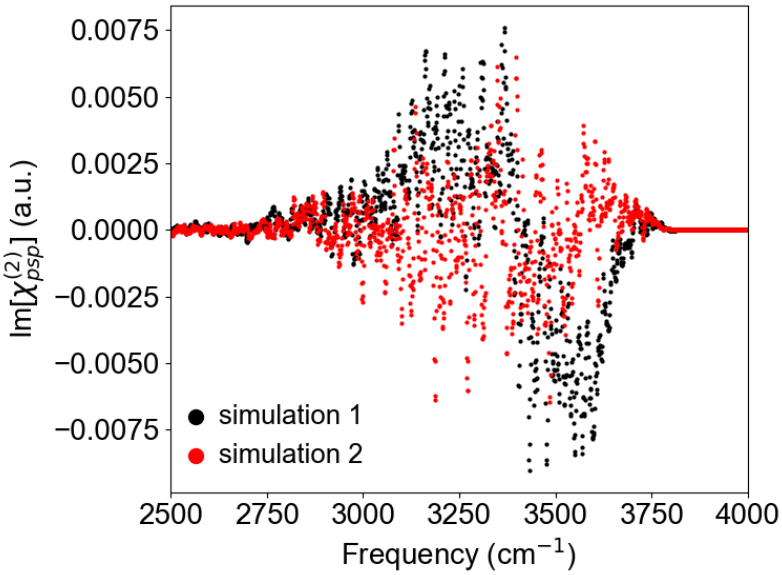

Figure A3.5. Chiral SFG responses of two independent trajectories for a pure-water interface, showing that the observed signal is noise that most likely originates from imperfect sampling. These spectra are averaged over 10,000,000 frames from 100 ns of sampling. The size of the system (20×20×10 Å) was reduced compared to the water-only control in Chapter 3 (45×45×22.5 Å) for computational efficiency.

References

1. Carr, J. K.; Wang, L.; Roy, S.; Skinner, J. L. Theoretical Sum Frequency Generation Spectroscopy of Peptides. *J. Phys. Chem. B* **2015,** *119*, 8969-8983.
2. Morita, A. Improved Computation of Sum Frequency Generation Spectrum of the Surface of Water. *J. Phys. Chem. B* **2006,** *110*, 3158-3163.
3. Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. *J. Chem. Phys.* **1993,** *98*, 10089-10092.
4. Miyamoto, S.; Kollman, P. A. Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models. *J. Comput. Chem.* **1992,** *13*, 952-962.
5. Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C. 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.
6. Long, D. A.; Bell, R. P. Intensities in Raman spectra I. A bond polarizability theory. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences* **1953,** *217*, 203-221.
7. Auer, B. M.; Skinner, J. L. IR and Raman spectra of liquid water: Theory and interpretation. *J. Chem. Phys.* **2008,** *128*, 224511.
8. Moad, A. J.; Simpson, G. J. A Unified Treatment of Selection Rules and Symmetry Relations for Sum-Frequency and Second Harmonic Spectroscopies. *J. Phys. Chem. B* **2004,** *108*, 3548-3562.
9. Wang, H.-F.; Gan, W.; Lu, R.; Rao, Y.; Wu, B.-H. Quantitative spectral and orientational analysis in surface sum frequency generation vibrational spectroscopy (SFG-VS). *Int. Rev. Phys. Chem.* **2005,** *24*, 191-256.
10. Auer, B. M.; Skinner, J. L. Dynamical effects in line shapes for coupled chromophores: Time-averaging approximation. *J. Chem. Phys.* **2007,** *127*, 104105.

Chapter 4: Detecting the First Hydration Shell Structure around Biomolecules at Interfaces

Adapted from the following reference with permission:

Konstantinovsky, D.; Perets, E. A.; Santiago, T.; Velarde, L.; Hammes-Schiffer, S.; Yan, E. C. Y.

(co-first-author publication)

ACS Cent. Sci. 2022, 8, 1404-1414.

DOI: 10.1021/acscentsci.2c00702

Abstract

Understanding the role of water in biological processes remains a central challenge in the life sciences. Water structures in hydration shells of biomolecules are difficult to study *in situ* due to overwhelming background from aqueous environments. Biological interfaces introduce additional complexity because biomolecular hydration differs at interfaces compared to bulk solution. Here, we perform experimental and computational studies of chiral sum frequency generation (chiral SFG) spectroscopy, probing chirality transfer from a protein to the surrounding water molecules. This work reveals that chiral SFG probes almost exclusively the first hydration shell around the protein. We explain the selectivity to the first hydration shell in terms of asymmetry induced by the protein structure and specific protein-water hydrogen-bonding interactions. This work establishes chiral SFG as a powerful technique for studying hydration shell structures around biomolecules at interfaces, presenting new possibilities to address grand research challenges in biology, including the molecular origins of life.

Introduction

Water, interfaces, and homochirality are chemical hallmarks of biological life. However, most water molecules in the hydration shells of biomolecules are missing from structural data using standard techniques such as x-ray crystallography and electron microscopy. Hydration shell water molecules are very difficult to study due to their rapid dynamics and exchange with bulk solvent. Nevertheless, the motion and arrangement of these water molecules can significantly modulate the structures and functions of biomolecules.[1-5] Moreover, the chemical behavior of water and biomolecules also changes near interfaces.[6] Aqueous interfaces are ubiquitous in biological systems and are chemically distinct from bulk aqueous environments because of the presence of chemical asymmetries, notably the termination of bulk water hydrogen bonding (H-bonding) networks. Hydration shell structures around biomolecules will therefore change dramatically as the biomolecule approaches an interface. Although NMR,[7-9] neutron scattering,[4, 10-15] terahertz spectroscopy,[16, 17] vibrational circular dichroism,[18] and Raman optical activity[19] are useful in probing aspects of structure and dynamics of hydration of biomolecules, none of these methods is interface specific. Thus, studying local water interactions in hydration shells of biomolecules *in situ* and under ambient conditions at interfaces remains a significant experimental challenge.

Here, we show that chiral sum frequency generation (chiral SFG) spectroscopy can tackle this challenge by offering remarkable selectivity to probe water structures in the first hydration shell of a biomolecule. We demonstrate that the method can suppress signals of water molecules outside the first hydration shell at interfaces and in the bulk solution. Chiral SFG reveals the O-H stretching vibrational frequencies of water, thus enabling

studies of local H-bonding interactions between water and biomolecules. This extraordinary capability of probing local H-bonding structures of water in the first hydration shell of biomolecules *in situ* at interfaces will support the pursuit of a central challenge in the life sciences – elucidating the role of water in biomolecular function.

Chiral SFG is an interface-specific, chiral-selective vibrational spectroscopy. For over a decade, chiral SFG has been developed into a tool for characterization of biomolecules in chiral secondary and higher-order structures at interfaces, such as on lipid membranes and polymer surfaces.[20-28] In 2017, Petersen and coworkers first detected chiral SFG signals from a chiral assembly of water molecules hydrating DNA.[29] Our group later showed that chiral assemblies of water molecules also form around proteins, and that the chirality of the water assemblies correlate with the chirality of the protein.[23, 30-32] These studies show the promise of chiral SFG to probe water structures in hydration shells of biomolecules at interfaces. Nonetheless, one outstanding question remains. What are the structural and dynamical properties of the chiral supramolecular assemblies of achiral water molecules that generate a chiral SFG response? This question remains to be answered before chiral SFG can be fully utilized to study hydration of biomolecules at the fundamental level.

We seek to answer this outstanding question using a model system, $LK_7\beta$ (Ac-LKLKLKL-NH_2), which is an amphiphilic peptide that assembles into antiparallel β-sheets at interfaces (Figure 4.1).[23, 30] We perform experiments to measure the chiral SFG response from water that hydrates $LK_7\beta$. We then build atomistic models of the antiparallel β-sheet at an aqueous interface and simulate the chiral SFG response of water with molecular dynamics (MD). Based on these simulations, we dissect the overall chiral SFG signals into

those generated by subsets of water molecules in the system. Such computational investigations reveal correlations between local H-bonding interactions, the O-H stretching frequency, and the intensity of the chiral SFG response. More importantly, chiral SFG is demonstrated to possess exclusive selectivity for water structures in the first hydration shell. We further analyze how the mobility of water molecules is related to their effectiveness in producing chiral SFG signals and how the orientations of water dipoles arrange around the protein at the interface. These analyses uncover the molecular origin of the remarkable selectivity of the chiral SFG method to the first hydration shell based on established symmetry-based chiral SFG theory.[33, 34] We discuss the implications of these findings in the context of recent progress in studying biomolecular hydration using various methods in the field. We conclude that chiral SFG offers unique capabilities to probe water vibrations in hydration structures of biomolecules at interfaces with selectivity to the first hydration shell.

Results and Discussion

Comparison of experimental and computational spectra

We start our investigations by obtaining experimental chiral SFG spectra of water around the antiparallel β-sheet $LK_7β$ (Figure 4.1) and comparing the experimental spectra with the computational spectra simulated with MD (Figure 4.2). Our experimental setup consists of a hydrated film of (L-) or (D-) $LK_7β$ deposited on a quartz surface for heterodyne detection, as previously described.[31] Compared with our previous studies,[31] the experimental chiral SFG spectrum extends the spectral region beyond 3400 cm^{-1} (Figures 4.2a and 4.2b) for direct comparison with the simulated spectra (Figure 4.2c). Figure 4.2a

shows that the phase of the chiral SFG signal flips when (L-) $LK_7\beta$ is replaced with (D-) $LK_7\beta$. The experimental spectra consist of N-H and C-H stretching peaks from protein as well as O-H stretching peaks from water. The C-H stretching modes in $LK_7\beta$ only appear below 3000 cm^{-1}. Because N-H and O-H stretching peaks are overlapping at frequencies above 3000 cm^{-1}, we identify the water O-H stretching contributions using $H_2{}^{18}O$ labeling (Figure 4.2b). Isotopic substitution with $H_2{}^{18}O$ shifts O-H stretching frequencies to the red by roughly 12 cm^{-1} (Figure A4.2 and Table A4.1).[35] This identifies peaks at 3151 cm^{-1}, 3366 cm^{-1}, and 3555 cm^{-1} as coming from the O-H stretch (see Table A4.1). Other spectral features (> 3000 cm^{-1}) do not redshift upon $H_2{}^{18}O$ substitution and therefore are assigned to the N-H stretching modes of $LK_7\beta$ (Figure A4.2 and Table A4.1). The $H_2{}^{18}O$ labeling reveals broad spectral features of water that span three spectral regions: 3050-3150 cm^{-1}, 3300-3400 cm^{-1}, and 3500-3800 cm^{-1}. The features appearing around 3050-3150 cm^{-1} and 3300-3400 cm^{-1} were previously observed and assigned to the O-H stretch of water solvating $LK_7\beta$.[30, 31] This experiment therefore identifies additional contributions of water in the high-frequency region of 3500-3800 cm^{-1}.

We calculate the chiral SFG spectrum of the water around (L-) $LK_7\beta$ (Figure 4.2c) using MD simulations. The model system consists of five antiparallel strands of the $LK_7\beta$ protein surrounded by ~1340 water molecules and placed at the vacuum-water interface, where the hydrophilic lysine side chains point toward the aqueous phase and the hydrophobic leucine side chains point toward the vacuum (Figure 4.1). Our calculations of the vibrational spectra rely on Skinner's mapping approach, which is based on density functional theory (DFT) calculations linking local electric fields to O-H vibrational frequencies and transition dipoles.[36-40] The computational spectra (Figure 4.2c) therefore

report water O-H peaks but not protein N-H peaks. To calculate the chiral SFG signal, we calculate an orthogonal ("chiral") second order susceptibility tensor component that is nonzero only for a chiral surface at an interface. The more commonly detected and simulated SFG response from achiral interfaces originates from non-orthogonal components, such as those calculated in the Skinner group's work[36, 38-40] (see Methods for technical details). As shown in our previous study,[31] the phase of the computational chiral SFG spectrum flips when the chirality of the protein is switched from the (L-) form to the (D-) form (Figure 4.2c), agreeing with the experimental observation (Figure 4.2a). Each spectrum contains two major features: a positive peak at around 3400 cm^{-1} and a negative peak at around 3600 cm^{-1} for (L-) $LK_7\beta$, and the same pair of peaks with opposite phase for (D-) $LK_7\beta$. These two peaks appear to agree with the experimentally observed water contribution in the two high-frequency regions of 3300-3400 cm^{-1} and 3500-3800 cm^{-1} (Figure 4.2a).

The computational spectrum seemingly does not show any low-frequency spectral features at 3050-3150 cm^{-1} to match the experimental observation. Nonetheless, careful examination of the computational spectrum (inset, Figure 4.2c) reveals a small feature at around 3000 cm^{-1} just above the noise level but statistically significant ($p=0.037$ in a two-sided *t*-test when the signal between 2850 and 3050 cm^{-1} is compared to the baseline between 3800 and 4000 cm^{-1}). Further analysis of the chiral SFG response shows that O-H stretches at this low frequency can originate from water forming strong H-bonds with $LK_7\beta$, which will be discussed in detail below. It is important to note that our calculations did not take Fermi resonances into account, and Fermi resonance with the first overtone of the water bending mode may contribute to the low-frequency peak.[41] Overall, the

computational spectra (Figure 4.2c) capture the most important features of the experimental spectra (Figure 4.2a), showing the chiral SFG response of the O-H stretching modes of water molecules hydrating $LK_7\beta$.

Chiral SFG signal of water is primarily generated from the first hydration shell

Our current study centers on dissecting the overall chiral SFG responses of water molecules into those interacting with various parts of $LK_7\beta$, e.g., the first and second hydration shells. Key to the analysis is the use of the Voronoi tessellation algorithm (see Methods). This algorithm divides the entire simulation system into non-overlapping three-dimensional compartments, where each compartment is exclusively occupied by one atom. The boundary of each compartment is delineated using Voronoi tessellation. These boundaries can then be used to identify all immediate neighbors of each atom in the simulation system.[42-44] Using this algorithm, we define the first hydration shell to contain water molecules with at least one atom neighboring the protein, and the second hydration shell to contain water molecules neighboring those in the first hydration shell. Figure 4.3a shows the $LK_7\beta$ protein (yellow) surrounded by water in the first (red) and second (blue) hydration shells, as well as all water molecules in the simulation system (black). We can then calculate the chiral SFG spectra of the first and second hydration shells, shown respectively as red and blue spectra in Figure 4.3b, where the spectrum of all water (black, same as in Figure 4.2c) is also presented for comparison. Remarkably, the spectrum of the first hydration shell almost overlaps with the spectrum arising from all water molecules in the molecular model. The area under the curve of the first hydration shell spectrum accounts for ~92% of the analogous area for the spectrum arising from all water molecules.

By contrast, the second hydration shell produces almost no signal. Hence, nearly all the chiral SFG response arises from water molecules in the first hydration shell. These results indicate that chiral SFG spectroscopy is sensitive to the water structures in the first hydration shell of the protein at the air-water interface.

Chiral SFG response of water reveals local interactions in the first hydration shell

We further analyze the chiral SFG signals generated from the first hydration shell. We aim to identify the water molecules that generate the chiral SFG responses corresponding to the three spectral regions (~3050-3150 cm^{-1}, ~3300-3400 cm^{-1}, and ~3500-3800 cm^{-1}) revealed from experiments (Figure 4.2a). We then focus on the correlation between the O-H stretching frequencies and the local H-bonding interactions between water and the protein. We examine this correlation because water-water and water-protein H-bonding interactions modulate the O-H stretching frequency, giving rise to chiral SFG responses across the three spectral regions. Specifically, stronger H-bonds with water acting as the H-bond donor lead to lower O-H stretching frequencies. To perform this analysis, we divide the first hydration shell into various subsets of water molecules and compute the chiral SFG spectra of these subsets (Figure 4.4). Figure 4.4a shows the spectrum of all water molecules in the first hydration shell. These water molecules are divided into those close to the protein backbone (Figure 4.4b) and sidechains (Figure 4.4c) (see the Appendix for the selection details). Next, the backbone subset is subdivided into *not* H-bonded to the protein (Figure 4.4d), and H-bonded to the C=O (Figure 4.4e) and N-H (Figure 4.4f) groups. Similarly, the sidechain subset (Figure 4c) is grouped into *not* H-bonded to the protein (Figure 4.4h) and H-bonded to the positively

charged lysine sidechains (Figure 4.4i). To explore the origin of the low-frequency experimental signals at 3050-3150 cm^{-1} (Figure 4.2a), we further calculate the spectrum of water molecules forming strong (short) H-bonds with the backbone C=O groups (Figure 4.4g). All these spectra are plotted with absolute signals (black spectra) on the left-hand axis and signals normalized to the number of water molecules (blue spectra) on the right-hand axis in Figure 4.4.

We first focus on the spectra in Figure 4.4 that share the same line shape as the spectrum of the first hydration shell (Figure 4.4a): a positive peak at ~3400 cm^{-1} and a negative peak at ~3600 cm^{-1}. These spectra include Figures 4.4b-e and h. The full hydration spectrum (Figure 4.4a) is the sum of the backbone (Figure 4.4b) and sidechain (Figure 4.4c) spectra. The absolute contributions of these two subset spectra are on the same order of magnitude. Hence, both the protein backbone and sidechains can template water molecules into chiral water superstructures and generate the overall chiral SFG signals. Within the backbone subset, water molecules that are not H-bonded to the backbone (Figure 4.4d) and H-bonded to the backbone C=O groups (Figure 4.4e) dominate the spectral response and contribute almost equally to the overall backbone spectrum (Figure 4.4b). In contrast, within the sidechain subset, water molecules not engaging in H-bonding interactions (Figure 4.4h) are responsible for almost all the signal in the overall sidechain spectrum (Figure 4.4c). Hence, there are three major contributors to the overall chiral SFG signal of water O-H stretches. They are water molecules near but not H-bonded to the backbone (Figure 4.4d), H-bonded to C=O on the backbone (Figure 4.4e), and near but not H-bonded to the sidechains (Figure 4.4h). These subsets of water are likely responsible for the

experimental signals of water O-H stretches in the middle and high frequency regions of ~3300-3400 cm^{-1} and ~3500-3800 cm^{-1} (Figure 4.2a).

It may be surprising that subsets of water molecules with different molecular interactions with the protein give similar spectral line shapes. Importantly, water molecules in the subsets that are not H-bonded to the protein still participate in H-bonding networks with neighboring water molecules. Indeed, we find that the average number of H-bonds formed by first hydration shell water molecules is 3.2 for those near but not H-bonded to the backbone, 3.2 for those H-bonded to C=O on the backbone, and 3.1 for those near but not H-bonded to the sidechains. These numbers are slightly lower than but comparable to an average of 3.4 H-bonds per water molecule considering all water molecules in the entire system. For water molecules in these H-bonding environments, the overall line shape appears to be simply the signature of water experiencing electric fields from the protein, and it does not matter whether the H-bond acceptor is a carbonyl group oxygen or a water oxygen because the electric fields experienced by the O-H groups are similar. This typical H-bonding environment for most water molecules around the antiparallel β-sheet likely accounts for the general line shape of water molecules in the first hydration shell of $LK_7\beta$.

Three spectra in Figure 4.4 deviate from the overall two-peak line shape. Among them, two exhibit similar spectral features: the backbone NH-bound (Figure 4.4f) and the sidechain NH_3^+-bound (Figure 4.4i) spectra. These spectra show not only the general feature with a positive peak at ~3400 cm^{-1} and a negative peak red shifted to around ~ 3550 cm^{-1} but also an additional positive peak at ~3700 cm^{-1}. A water O-H stretch at that high frequency often suggests a weak or free H-bonding environment. Indeed, the sharp peak at high frequency in Figure 4.4f resembles the prototypical dangling O-H pointing toward the

air at the air-water interface as first reported by Shen and coworkers.[45] Analysis of our MD trajectories indicates that about 15% of the water molecules that are H-bonded to backbone N-H have OH groups pointing into the vacuum. These OH groups in H-bond free environments are expected to contribute to the high-frequency signal. In contrast, the positive peak at ~3700 cm^{-1} in the sidechain NH_3^+-bound spectrum (Figure 4.4i) likely arises from a different source. The NH_3^+-bound water molecules form only 2.3 H-bonds on average (versus 3.4 for all water molecules in the system), suggesting that the H-bonding structure of water near the positive charge of the lysine is severely disrupted. This disruption was previously proposed to give rise to a high-frequency peak corresponding to a water molecule containing O-H groups that are not engaged in H-bonds.[45]

The last spectrum that shows a line shape different from the general one is Figure 4.4g. This spectrum is generated from water molecules forming the shortest (strongest) H-bonds with the backbone C=O (1.6 Å or less between the water hydrogen and protein oxygen). The spectrum contains three peaks. Two of them resemble the general positive peak (3400 cm^{-1}) and negative peak (3600 cm^{-1}) of the first-hydration shell spectrum (Figure 4.4a) but are red shifted to 3250 cm^{-1} and 3550 cm^{-1}, respectively. These red shifts are often associated with stronger H-bonds. Uniquely, the spectrum also shows a small but significant third peak at ~3000 cm^{-1}. A peak at this low frequency is also present in the spectrum of all water molecules in the system (inset, Figure 4.2c). Such low-frequency signals are seen in the $H_2{}^{18}O$ isotope substitution experiments (Figure 4.2b). Hence, the low-frequency signals could potentially be due to water molecules strongly H-bonded to carbonyl groups on the protein backbone. Nonetheless, it is important to note that Fermi resonance with the first overtone of the water bending mode can also generate a water

vibrational response at ~3200 cm^{-1}.[41, 46] Because our current calculations do not include Fermi resonance, further investigation is needed to better understand the experimentally observed signals in the low-frequency region of ~3050-3150 cm^{-1}.

Chiral SFG intensity and water mobility are correlated for water H-bonded to protein

We further investigate how water mobility regulates the chiral SFG intensity. We hypothesized that the more static the water molecules in the first hydration shell the higher the chiral O-H stretching signals produced per water molecule in that subset. To test this hypothesis, we calculate the retention times (i.e., the average time spent within a subset) of the water molecules in each subset described in Figure 4.4.[47] Figure 4.5 plots the chiral SFG signal per water molecule versus the retention time. The overall correlation is noticeably poor. However, there is a positive correlation between the retention time and the chiral SFG signal production efficiency per water molecule for water interacting with the protein through H-bonds. These water molecules belong to the subsets that are hydrogen bonded to backbone C=O, backbone N-H, and sidechain NH_3^+ groups (i.e., "HB to C=O", "HB to N-H", and "HB to NH_3^+" in Figure 4.4). The above correlation analyses reveal two classes of water molecules in the first hydration shell that generate chiral SFG signal generation. The first class is characterized by water molecules that form H-bonds with the protein. In this case, the more a water molecule's mobility is reduced by H-bonding interactions with the protein, the more it produces chiral signal (Figures 4.4e, f, and i and 4.5). The second class is characterized by water molecules that do not form H-bonds with the protein (Figures 4.4d, 4.4h, and 4.5). According to this analysis, the water

molecules that form more persistent H-bonds with the protein (i.e., H-bonds that are retained for longer times prior to exchange) contribute the most to the chiral signal.

Chiral SFG theory can explain the selectivity of chiral SFG to the first hydration shell

We then seek to elucidate the origin of the selectivity of chiral SFG to the first hydration shell. We ask whether the ordering of water dipoles by the $LK_7\beta$ protein extends beyond the first hydration shell, and if it does, why this extended ordering does not generate chiral SFG signals. Figure 4.6a shows the top view of the simulation system composed of $LK_7\beta$ at the vacuum-water interface, where the protein (yellow) is surrounded by the first hydration shell (red) and second hydration shell (blue). The average water molecular dipole moments in the interfacial plane are calculated with a resolution of 1 Å and presented as unit vectors in Figures 4.6a, b, c, and d. The relative magnitude of each vector is not depicted in these figures to emphasize even a slight bias in the dipole direction. Figure A4.6 provides a depiction with the actual vector magnitudes, illustrating that the ordering of water dipoles past the second hydration shell is minimal compared to the ordering near the protein.

Figure 4.6a shows that the water dipoles are pointing toward the protein because $LK_7\beta$ contains positively charged lysine sidechains. Clearly, $LK_7\beta$ changes the ordering of water molecules past the first hydration shell. The region of effect extends 30–40 Å away from the protein (red box, Figure 4.6c), and gradually fades into random arrangements beyond 40 Å (blue box, Figure 4.6b). Focusing on the water molecules closest to the protein, Figure 4.6d (orange box) depicts the orientations of water dipoles inside the first hydration shell, and Figure 4.6e depicts the average water dipole in this region with both

direction and magnitude. The magnitude at a given gridpoint can be less than the dipole moment of one water molecule due to averaging over different orientations and occupation. The arrangement of water dipoles within the first hydration shell appears to be less symmetrical than the arrangement of water dipoles beyond the first hydration shell, mainly due to direct interactions between the water molecules and the chiral protein.

The ordering pattern shown in Figure 4.6 can explain the remarkable sensitivity of chiral SFG to the first hydration shell of $LK_7\beta$ in terms of symmetry-based chiral SFG theory.[33] The theory developed by Simpson in 2004 dictates that a surface-specific chiral SFG response is allowed from an interface of uniaxial C_∞ symmetry. In this context, the lack of a reflection plane perpendicular to the interface is the hallmark of a chiral interface, making it distinct from an achiral interface with $C_{\infty v}$ symmetry. The absence of a reflection plane perpendicular to the interface is key to distinguishing chiral SFG from conventional (achiral) SFG methods in providing selectivity to interfacial chirality. Figure 4.6a shows that the ordering beyond the first hydration shell is striking but highly symmetrical. At least two reflection planes of the water dipoles can be identified, $\sigma_{//}$ aligning with the β-strands and $\sigma_\perp$ perpendicular to the β-strands (Figure 4.6a). Hence, water molecules outside the first hydration shell are not in C_∞ symmetry and thus are not chiral SFG active. In contrast, the ordering of water dipoles in the first hydration shell is much less symmetrical. No reflection plane in any direction can be identified, which is more evident in the plot of the magnitudes of the water dipole moments (Figure 4.6e).

An alternative illustration of the lack of symmetry in the first hydration shell is presented in Figure 4.6f. This plot shows the magnitude of the sum of the water dipole moment vectors within each 1-Å annulus between concentric circles in the *xy* plane

originating at the center point. A nonzero magnitude of this sum is a good indication of asymmetry (chirality). The yellow, red, and blue regions roughly correspond to the area of the protein, the first hydration shell, and the second hydration shell, respectively. The region composed of the protein and the first hydration shell shows significant magnitude of the dipole vector sum. The region beyond the first hydration shell shows little or no magnitude, indicating radial symmetry across the two reflection planes ($\sigma_{\parallel}$ and $\sigma_{\perp}$) and thus the water arrangement is achiral in $C_{\infty v}$ symmetry. However, the region within the first hydration clearly shows non-zero magnitude, supporting that water arrangement is chiral in C_{∞} symmetry.

We performed the same analysis for water dipoles below the water surface up to 14 Å (Figure A4.5) and made the same observation. Water molecules in the first hydration shell form a supramolecular structure with the uniaxial C_{∞} symmetry and thus are chiral SFG active. Water molecules beyond the first hydration shell, however, are organized in the $C_{\infty v}$ symmetry and hence are chiral SFG inactive. Consequently, the selectivity of chiral SFG to the first hydration shell can be elucidated using the chiral SFG theory.[33]

Conclusions

This study establishes the label-free selectivity of chiral SFG for probing vibrational structures of water in the first hydration shell of proteins. The selectivity of chiral SFG is based on chirality transfer from a protein to the first hydration shell. However, proteins also impact water orientation and presumably dynamics beyond the first hydration shell (Figure 4.6). Many label-free physical methods that can detect hydration of biomolecules report on water well beyond the first hydration shell. For example,

intermolecular Nuclear Overhauser Effect NMR can report on water molecules 4-5 Å from the surface of biomolecules, but the signal can be dominated by couplings between protein and water molecules far outside the first hydration shell.[48] Terahertz spectroscopy measures collective vibrations of water molecules averaged over water molecules both near and far from the protein surface.[49, 50] Although 2D-infrared spectroscopy is more sensitive to localized protein vibrations and can provide site-specific insights on short-range protein-water interactions, the technique lacks selectivity for the O-H stretch of water molecules in protein hydration shells.[51] X-ray crystallography and electron microscopy provide a high level of site specificity of water interaction with biomolecules, but they are sensitive to the most strongly associated water molecules and lack reliable data about more dynamic water populations.[52] Neutron scattering is selective for the first hydration shell; however, the method requires powder samples or highly concentrated protein solutions,[51] restricting applications for which *in situ* or interfacial measurements are desired. Conventional (*achiral*) SFG has been used to probe hydration of protein and biomolecules at interfaces. Nonetheless, this method detects all water molecules oriented by biomolecules (see Figure 4.6), lacking selectivity to the first hydration shell.[53-58]

The higher selectivity of *chiral* SFG for the first hydration shell will be beneficial in answering many outstanding questions in molecular biology concerning both intramolecular and intermolecular interactions of biomolecules. For instance, the structural rearrangements and dynamical fluctuations of biomolecules that enable all molecular functions[5, 59, 60] rely on the presence of water in ways that are still not understood. Most importantly, the chemical details underlying the role of the first hydration shell in the hydrophobic effect, protein denaturation, and protein folding remain enigmatic. With chiral

SFG, the architectures and dynamics of hydration shells involved in protein-protein, protein-ligand, protein-nucleic acid, and protein-lipid interactions, including transmembrane proteins in amphiphilic lipid environments, can be investigated. To tackle such important biophysical problems, chiral SFG methods rely on the synergy between chiral SFG experiments and computational analysis, as demonstrated in this study. This is especially true for extraction of chiral SFG vibrational lineshapes that are hidden in the experimental spectrum (Figure 4.4). Such analyses will prove useful for relating the chiral SFG responses of water to specific water-protein or protein-protein interactions.

Chirality transfer from small molecules to solvent has been observed in prior studies,[61] but the exceptional sensitivity of chiral SFG to the chirality of the first hydration shell around biomolecules at interfaces presents unique opportunities for exploring questions relevant to the origin of life. Hydration, homochirality, and interfaces are defining features of life. Understanding these aspects in the context of the origin of life[62-65] demands a technique that can detect chirality at interfaces. Water may have played a role in the emergence of homochirality through the creation of extended chiral molecular structures beyond the immediate borders of chiral molecules.[31, 63, 66] Furthermore, prebiotic chemistry such as formation of peptide bonds may have taken place in microdroplets[67] at the interface formed between water and air or between water and an organic phase. Our discovery that chiral SFG exclusively selects for the first hydration shell around biomolecules at interfaces positions chiral SFG to tackle new questions and produce new insights about the emergence of biological life.

Methods

Sample Preparation

(L-) $LK_7\beta$ (GL Biochem Ltd., Shanghai, China) and (D-) $LK_7\beta$ (AnaSpec, Inc, Fremont, CA, USA) were obtained as lyophilized powders and dissolved in H_2O (or $H_2{}^{18}O$) at 1 mM. The peptide solutions (10 μL) were applied to the right-handed z-cut α-quartz surface. The solution was dried in a desiccator and under nitrogen flow to prevent exchange of ambient humidity with $H_2{}^{18}O$ in the hydrated protein film. Spectra were collected immediately.

Phase-resolved Vibrational Chiral SFG

The phase-resolved chiral SFG spectra reported in this study were performed using a broadband SFG spectrophotometer, previously described.[68] Chiral SFG recordings utilized the *psp* polarization (*p*-polarized sum frequency, *s*-polarized visible, *p*-polarized infrared). Spectra were collected along the +*y* and -*y* axes of the quartz calibrated to the laboratory frame (see Appendix).[31] For all reported spectra, 10–12 spectra (2 minutes each) were acquired along both the +*y* and -*y* axes and averaged. The average spectra along the +*y* and -*y* directions were normalized by the signal from the clean quartz surface and subtracted:[69, 70]

$$\mathrm{Im}[\chi^{(2)}] = \frac{I_{+y} - I_{-y}}{4} \qquad (1)$$

Obtaining the phase-resolved vibrational chiral SFG spectrum of $LK_7\beta$ from 2900–3800 cm^{-1} with our femtosecond broadband SFG spectrophotometer required collection in two spectral windows (Figure A4.1). The frequencies of the SFG spectra in both spectral windows were calibrated using a polystyrene standard (Buck Scientific; 0.05 mm film).

The individual spectra were overlaid, normalized to the N-H stretching peak of $LK_7\beta$ (~3270 cm^{-1}), and stitched to produce the overall spectrum (see Appendix).

MD simulations

Five $LK_7\beta$ (LKLKLKL) strands were arranged into an antiparallel β-sheet that was solvated in TIP4P-Ew water[71] (a version of the TIP4P model adapted for Ewald summation), with at least 1 nm of solvent on each side of the protein. The upper half of the water was then removed, and chloride ions were added to neutralize the system, leaving a slab with two vacuum-water interfaces, one of which contained the protein. The *z*-dimension of the system was extended to create a vacuum region equal in size to the slab. Periodic boundary conditions were applied in three dimensions. Acetyl groups were added to the N-termini of the peptides, and amide ($-NH_2$) groups were created at the C-termini to stabilize the β-sheet. After an energy minimization, the slab was equilibrated in the NVT ensemble for 6 ns at 298 K using Langevin dynamics (friction coefficient 1/ps, timestep 1 fs). The ff14SB force field was used for the protein.[72] The particle-mesh Ewald method[73] was used to treat long range electrostatic interactions. The SETTLE algorithm[74] was used to maintain rigid water molecules, and the SHAKE method[75] was used to constrain bonds involving hydrogens. For the production run, Langevin dynamics was propagated for 100 ns, storing a configuration every 10 fs. OpenMM 7.4 tools[76] were used for equilibration and production runs as well as for part of the system setup, and VMD scripts were used to arrange the antiparallel β-sheet. An NVIDIA Tesla V100 GPU and CUDA 9.2 were used to run the MD simulations. The Appendix contains details regarding equilibration. The

trajectory was not wrapped to put the protein in the center for the SFG calculations, as the calculation takes periodic boundary conditions into account.

Calculation of SFG spectra of subsets of water molecule

The inhomogeneous limit version of Skinner's electric field mapping approach[36, 38-40] was used to calculate spectra, as in our previous work.[31, 32] SFG signals derive from the second-order susceptibility ($\chi^{(2)}$) of a system. Chiral SFG is distinct from the more commonly reported achiral SFG because the two techniques probe different elements of $\chi^{(2)}$ by manipulating the polarization of the incoming radiation and the detector in different ways. The form of chiral SFG presented here (*psp*, see above) detects the orthogonal tensor component $\chi^{(2)}_{zyx}$, which is only nonzero for chiral systems at the interface, whereas achiral SFG (e.g. *ssp*) detects nonorthogonal components such as $\chi^{(2)}_{yyz}$, which can be nonzero at any interface.[21] In the calculations, we control the output element of $\chi^{(2)}$ by analyzing particular components of the dipole and polarizability.[77] This is analogous to changing the polarization of the incoming radiation and detector in experiments. Previous work by Simpson[33] and Konstantinovsky et al.[32] has provided a quantitative description of the ability of certain *achiral* molecules (such as water) to produce *chiral* SFG signals when in a chiral superstructure.

In the construction of the exciton Hamiltonian, intermolecular couplings between O-H groups on different molecules were neglected to allow clear separation of water molecule subsets, but intramolecular couplings between O-H groups on the same molecule were included.[32] All water subset selections were done using in-house code. The

MDAnalysis tool select_atoms was used extensively for simple selections.[78] See Figure A4.4 for a graphical illustration of our approach.

Using Voronoi tessellation to identify water molecule neighbors

The first hydration shell was defined using Voronoi tessellation. A Voronoi tessellation of a set of points consists of cells containing the regions closer to each point than to any other point in the set. We break down the atoms of the system into Voronoi cells. Water neighbors of the protein are defined as water molecules for which at least one atom's Voronoi cell is in contact with a Voronoi cell belonging to a protein atom.[79] The freud library[79] was used to access the Voronoi tessellation engine voro++[80] from Python to efficiently produce Voronoi tessellations and neighbor lists. Periodic boundary conditions were taken into account by all water selections, including Voronoi tessellation. Within the Voronoi diagram, the water molecules bordering a protein cell were considered the "first hydration shell", while the water molecules bordering the first hydration shell were considered the "second hydration shell". The "backbone" and "sidechain" subsets were deliberately made to overlap somewhat (~46 water molecules shared between them in a typical selection) to ensure that all water molecules in the first Voronoi shell were considered. As a result, the numbers of water molecules in the two subsets add up to slightly more than the number of water molecules in the first hydration shell. The hydrogen-bonding criteria used in this work are a distance between heavy atoms less than 3.5 Å and a hydrogen-bonding angle greater than 135°. The distance cutoff was chosen to be larger than is typically used to ensure that water molecules *not* hydrogen bonded to the protein were highly unlikely to be engaging in any form of specific interaction with the protein.

Calculation of local water dipoles on a grid

The average water dipole moments were obtained by adding the dipoles of the water molecules at each grid point (based on the location of the oxygen) over the trajectory to determine the total dipole at each grid point and subsequently normalizing this vector to obtain the unit dipole vector. The grid point calculation was done as described previously.[31] A 100 ns trajectory was used to generate the images in Figure 4.4, with frames considered every 1 ps for a total of 100,000 frames analyzed.

Analysis of retention time

A 100 ns trajectory was used to generate these data, with frames saved every 1 ps for a total of 100,000 frames analyzed. The retention time plotted is the average time from the arrival of a water molecule into a given subset (selection) to its exit from that subset with a time resolution of 1 ps.

Summary of Appendix to Chapter 4

Description of the SFG spectrometer, fitting of experimental spectra, quartz only spectrum (experimental), details of MD equilibration, details of selection methods for water subsets, information about the construction of the exciton Hamiltonian for the calculation of SFG spectra, illustration of the lack of a reflection plane in the first hydration shell at various distances from the interface, unnormalized (non-unit) dipole moment vectors around $LK_7\beta$.

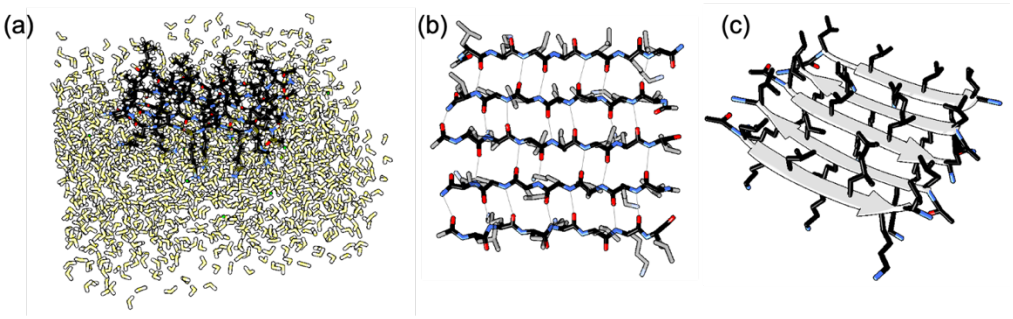

Figure 4.1. The molecular system used for computational analysis of the chiral SFG response of water around a protein. (a) An $LK_7\beta$ pentamer (black) at the vacuum-water interface forming an antiparallel β-sheet. The protein is surrounded by ~1340 water molecules (yellow). (b) A top view of the $LK_7\beta$ pentamer with the backbone highlighted in colors and the sidechains in gray. The aqueous phase is below this structure. (c) A side view of the same pentamer with sidechains highlighted in colors and the backbone in gray. The polar lysine residues are directed into the water, and the nonpolar leucine residues are directed into the vacuum. Hydrogen atoms have been removed for clarity in (b) and (c).

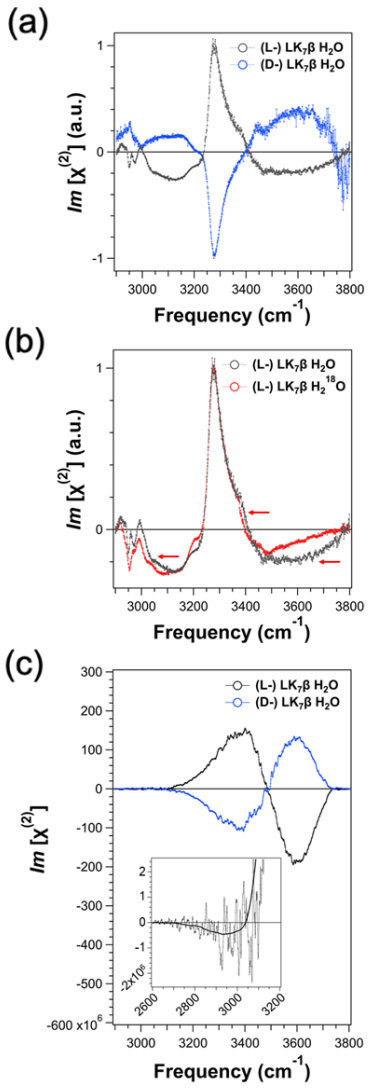

Figure 4.2. Experimental and computational chiral SFG spectra containing spectral contributions due to water. (a) Full experimental spectrum in the N-H/O-H region of (L-) $LK_7\beta$ and (D-) $LK_7\beta$. (b) Experimental H_2O and $H_2{}^{18}O$ spectra of the N-H/O-H region of a hydrated film of $LK_7\beta$, revealing peaks due to water (red arrows). (c) Calculated spectra of O-H stretch response for (L-) $LK_7\beta$ and (D-) $LK_7\beta$. Inset shows small low-frequency peak in the (L-) $LK_7\beta$ spectrum (raw data and smoothed curve). The low-frequency peak is statistically significant compared to the spectral baseline noise ($p=0.037$, two-sided *t*-test).

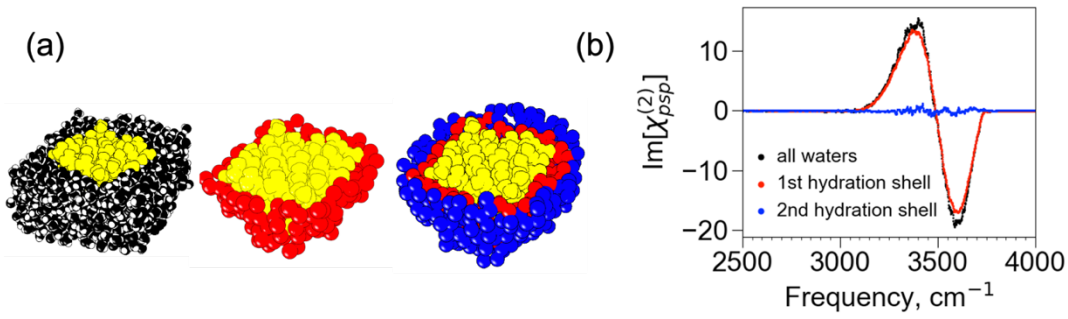

Figure 4.3. Almost all the chiral SFG response of water originates from the first hydration shell around the protein. (a) The model systems used for this analysis contain the LK$_7$β pentamer protein (yellow) together with different subsets of water molecules: all water (black), the first hydration shell (red), and the second hydration shell (blue). (b) Calculated spectra of the O-H stretch response from all water molecules (black), the first hydration shell (red), and the second hydration shell (blue).

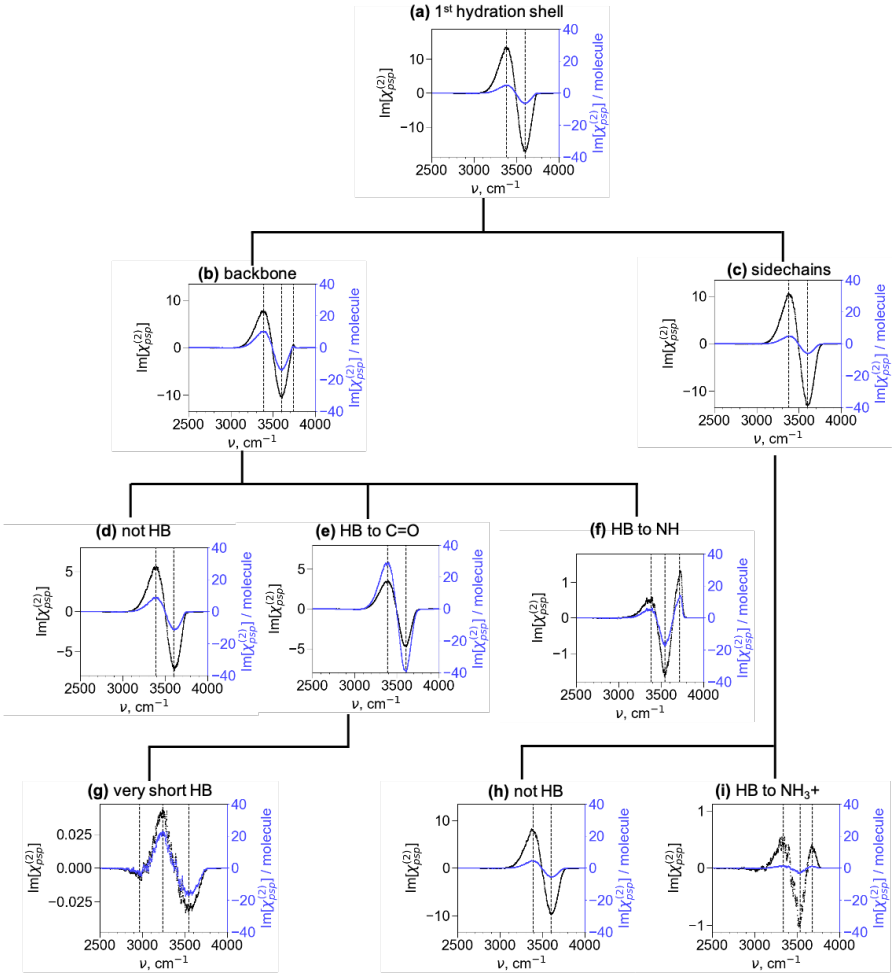

Figure 4.4. Chiral SFG lineshapes of different subsets of water molecules around (L-) LK$_7$β. The names of the plots for (b) and (c) indicate the region around the protein from which water molecules were taken for each subset. For (d) through (i), HB stands for "hydrogen bonded". "HB to NH" refers to the water accepting a hydrogen bond from an amide N-H group. "HB to C=O" refers to the water donating a hydrogen bond to an amide carbonyl group. The left-hand *y*-axes and the black spectra indicate the absolute signals to show the relative intensities of the various spectra. The right-hand *y*-axes and the blue spectra indicate the chiral SFG signal *per water molecule* for each subset.

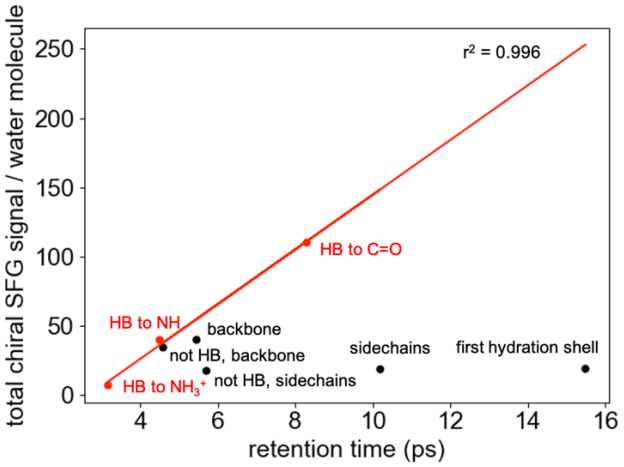

Figure 4.5. Relationship between water mobility and chiral SFG signal generation. This plot depicts the chiral SFG signal per water molecule versus average retention time of a water molecule in each subset shown in Figure 4.4.

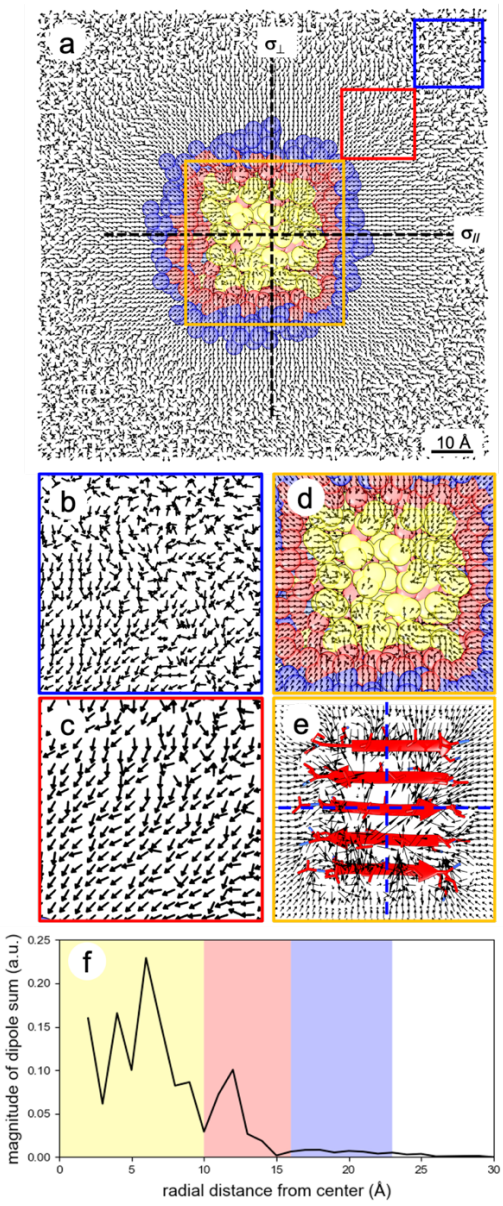

Figure 4.6. Symmetrical and asymmetrical orientation of water molecular dipole moments outside and inside the first hydration shell of the $LK_7\beta$ protein, respectively, at the vacuum-water interface. Water dipoles bisect the H-O-H angle and point toward the water oxygen. (a) Top view of water dipoles at the vacuum-water interface. The protein structure of $LK_7\beta$ (yellow) is the average conformation over the course of the MD trajectory. Red and blue regions, respectively, indicate the first and second hydration shells. Water dipoles point toward the protein due to the positive charges of the lysine residues. (b), (c), and (d) Zoom-in details of water dipoles in the blue, red, and orange boxes, respectively. (e) Zoom-in details of water dipoles in the orange box with the magnitude of water dipoles represented by the relative length of the arrows. For water dipoles outside the first hydration shell but still within ~40 Å of the protein, two reflection planes perpendicular to the interface, $\sigma_{//}$ aligning with and $\sigma_{\perp}$ perpendicular to the β-strands, can be identified. For water dipoles inside the first hydration shell, no such reflection planes can be identified. (f) The magnitude of the vector sum of the water dipoles is plotted as a function of radial distance from the center of Figure 4.6a.

References

1. Kauzmann, W. Some factors in the interpretation of protein denaturation. *Adv. Protein Chem.* **1959,** *14*, 1-63.
2. Chen, D.; Li, Y.; Zhao, M.; Tan, W.; Li, X.; Savidge, T.; Guo, W.; Fan, X. Effective lead optimization targeting the displacement of bridging receptor–ligand water molecules. *Phys. Chem. Chem. Phys.* **2018,** *20*, 24399-24407.
3. Matricon, P.; Suresh, R. R.; Gao, Z.-G.; Panel, N.; Jacobson, K. A.; Carlsson, J. Ligand design by targeting a binding site water. *Chem. Sci.* **2021,** *12*, 960-968.
4. Fenimore, P. W.; Frauenfelder, H.; McMahon, B. H.; Parak, F. G. Slaving: Solvent fluctuations dominate protein dynamics and functions. *Proc. Natl. Acad. Sci. USA* **2002,** *99*, 16047.
5. Kempf, J. G.; Loria, J. P. Protein dynamics from solution NMR. *Cell Biochem. Biophys.* **2002,** *37*, 187-211.
6. D'Imprima, E.; Floris, D.; Joppe, M.; Sánchez, R.; Grininger, M.; Kühlbrandt, W. Protein denaturation at the air-water interface and how to prevent it. *Elife* **2019,** *8*.
7. Nucci, N. V.; Pometun, M. S.; Wand, A. J. Site-resolved measurement of water-protein interactions by solution NMR. *Nat. Struct. Mol. Biol.* **2011,** *18*, 245-249.
8. Otting, G. NMR studies of water bound to biological molecules. *Prog. Nucl. Magn. Reson. Spectrosc.* **1997,** *31*, 259-285.
9. Halle, B. Protein hydration dynamics in solution: a critical survey. *Philos Trans R Soc Lond B Biol Sci* **2004,** *359*, 1207-23; discussion 1223-4, 1323-8.
10. Doster, W.; Settles, M. Protein–water displacement distributions. *Biochim. Biophys. Acta BBA - Proteins Proteom.* **2005,** *1749*, 173-186.
11. Roh, J. H.; Curtis, J. E.; Azzam, S.; Novikov, V. N.; Peral, I.; Chowdhuri, Z.; Gregory, R. B.; Sokolov, A. P. Influence of Hydration on the Dynamics of Lysozyme. *Biophys. J.* **2006,** *91*, 2573-2588.
12. Wood, K.; Frölich, A.; Paciaroni, A.; Moulin, M.; Härtlein, M.; Zaccai, G.; Tobias, D. J.; Weik, M. Coincidence of Dynamical Transitions in a Soluble Protein and Its Hydration Water: Direct Measurements by Neutron Scattering and MD Simulations. *J. Am. Chem. Soc.* **2008,** *130*, 4586-4587.
13. Smith, J. C.; Tan, P.; Petridis, L.; Hong, L. Dynamic Neutron Scattering by Biological Systems. *Annu. Rev. Biophys.* **2018,** *47*, 335-354.
14. Chen, S. H.; Liu, L.; Fratini, E.; Baglioni, P.; Faraone, A.; Mamontov, E. Observation of fragile-to-strong dynamic crossover in protein hydration water. *Proc. Natl. Acad. Sci. USA* **2006,** *103*, 9012.
15. Tarek, M.; Tobias, D. J. Role of Protein-Water Hydrogen Bond Dynamics in the Protein Dynamical Transition. *Phys. Rev. Lett.* **2002,** *88*, 138101.
16. Novelli, F.; Ostovar Pour, S.; Tollerud, J.; Roozbeh, A.; Appadoo, D. R. T.; Blanch, E. W.; Davis, J. A. Time-Domain THz Spectroscopy Reveals Coupled Protein–Hydration Dielectric Response in Solutions of Native and Fibrils of Human Lysozyme. *J. Phys. Chem. B* **2017,** *121*, 4810-4816.
17. Heyden, M.; Ebbinghaus, S.; Havenith, M., Terahertz Spectroscopy as a Tool to Study Hydration Dynamics. In *Encyclopedia of Analytical Chemistry*.

18. Yang, G.; Xu, Y., Vibrational Circular Dichroism Spectroscopy of Chiral Molecules. In *Electronic and Magnetic Properties of Chiral Molecules and Supramolecular Architectures*, Naaman, R.; Beratan, D. N.; Waldeck, D., Eds. Springer Berlin Heidelberg: Berlin, Heidelberg, 2011; pp 189-236.
19. Barron, L. D.; Hecht, L.; McColl, I. H.; Blanch, E. W. Raman optical activity comes of age. *Mol. Phys.* **2004,** *102*, 731-744.
20. Wang, Z.; Fu, L.; Yan, E. C. Y. C–H Stretch for Probing Kinetics of Self-Assembly into Macromolecular Chiral Structures at Interfaces by Chiral Sum Frequency Generation Spectroscopy. *Langmuir* **2013,** *29*, 4077-4083.
21. 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.
22. Liu, W.; Fu, L.; Wang, Z.; Sohrabpour, Z.; Li, X.; Liu, Y.; Wang, H.-f.; Yan, E. C. Y. Two dimensional crowding effects on protein folding at interfaces observed by chiral vibrational sum frequency generation spectroscopy. *Phys. Chem. Chem. Phys.* **2018,** *20*, 22421-22426.
23. Perets, E. A.; Videla, P. E.; Yan, E. C. Y.; Batista, V. S. Chiral inversion of amino acids in antiparallel beta sheets at interfaces probed by vibrational sum frequency generation spectroscopy. *J. Phys. Chem. B* **2019,** *123*, 5769−5781.
24. Stokes, G. Y.; Gibbs-Davis, J. M.; Boman, F. C.; Stepp, B. R.; Condie, A. G.; Nguyen, S. T.; Geiger, F. M. Making "Sense" of DNA. *J. Am. Chem. Soc.* **2007,** *129*, 7492-7493.
25. Okuno, M.; Ishibashi, T.-a. Heterodyne-Detected Achiral and Chiral Vibrational Sum Frequency Generation of Proteins at Air/Water Interface. *J. Phys. Chem. C* **2015,** *119*, 9947-9954.
26. Wang, J.; Chen, X.; Clarke, M. L.; Chen, Z. Detection of chiral sum frequency generation vibrational spectra of proteins and peptides at interfaces in situ. *Proc. Natl. Acad. Sci. USA* **2005,** *102*, 4978-4983.
27. Tan, J.; Zhang, J.; Luo, Y.; Ye, S. Misfolding of a Human Islet Amyloid Polypeptide at the Lipid Membrane Populates through β-Sheet Conformers without Involving α-Helical Intermediates. *J. Am. Chem. Soc.* **2019,** *141*, 1941-1948.
28. Perets, E. A.; Olesen, K. B.; Yan, E. C. Y. Chiral Sum Frequency Generation Spectroscopy Detects Double-Helix DNA at Interfaces. *Langmuir* **2022,** *38*, 5765-5778.
29. McDermott, M. L.; Vanselous, H.; Corcelli, S. A.; Petersen, P. B. DNA's Chiral Spine of Hydration. *ACS Cent. Sci.* **2017,** *3*, 708-714.
30. Perets, E. A.; Yan, E. C. Y. Chiral Water Superstructures around Antiparallel β-Sheets Observed by Chiral Vibrational Sum Frequency Generation Spectroscopy. *J. Phys. Chem. Lett.* **2019,** *10*, 3395-3401.
31. Perets, E. A.; Konstantinovsky, D.; Fu, L.; Chen, J.; Wang, H.-F.; Hammes-Schiffer, S.; Yan, E. C. Y. Mirror-image antiparallel β-sheets organize water molecules into superstructures of opposite chirality. *Proc. Natl. Acad. Sci. USA* **2020,** *117*, 32902-32909.
32. Konstantinovsky, D.; Perets, E. A.; Yan, E. C. Y.; Hammes-Schiffer, S. Simulation of the Chiral Sum Frequency Generation Response of Supramolecular Structures Requires Vibrational Couplings. *J. Phys. Chem. B* **2021,** *125*, 12072–12081.

33. Moad, A. J.; Simpson, G. J. A Unified Treatment of Selection Rules and Symmetry Relations for Sum-Frequency and Second Harmonic Spectroscopies. *J. Phys. Chem. B* **2004,** *108*, 3548-3562.
34. Perry, J. M.; Moad, A. J.; Begue, N. J.; Wampler, R. D.; Simpson, G. J. Electronic and Vibrational Second-Order Nonlinear Optical Properties of Protein Secondary Structural Motifs. *J. Phys. Chem. B* **2005,** *109*, 20009-20026.
35. Jung, S.-Y.; Lim, S.-M.; Albertorio, F.; Kim, G.; Gurau, M. C.; Yang, R. D.; Holden, M. A.; Cremer, P. S. The Vroman Effect: A Molecular Level Description of Fibrinogen Displacement. *J. Am. Chem. Soc.* **2003,** *125*, 12782-12786.
36. Corcelli, S. A.; Lawrence, C. P.; Skinner, J. L. Combined electronic structure/molecular dynamics approach for ultrafast infrared spectroscopy of dilute HOD in liquid H2O and D2O. *J. Chem. Phys.* **2004,** *120*, 8107-8117.
37. Corcelli, S. A.; Skinner, J. L. Infrared and Raman Line Shapes of Dilute HOD in Liquid H2O and D2O from 10 to 90 °C. *J. Phys. Chem. A* **2005,** *109*, 6154-6165.
38. Auer, B. M.; Skinner, J. L. IR and Raman spectra of liquid water: Theory and interpretation. *J. Chem. Phys.* **2008,** *128*, 224511.
39. Auer, B. M.; Skinner, J. L. Vibrational Sum-Frequency Spectroscopy of the Water Liquid/Vapor Interface. *J. Phys. Chem. B* **2009,** *113*, 4125-4130.
40. Pieniazek, P. A.; Tainter, C. J.; Skinner, J. L. Interpretation of the water surface vibrational sum-frequency spectrum. *J. Chem. Phys.* **2011,** *135*, 044701.
41. Kananenka, A. A.; Skinner, J. L. Fermi resonance in OH-stretch vibrational spectroscopy of liquid water and the water hexamer. *J. Chem. Phys.* **2018,** *148*, 244107.
42. Voronoi, G. Nouvelles applications des paramètres continus à la théorie des formes quadratiques. Deuxième mémoire. Recherches sur les parallélloèdres primitifs. *J. Reine Angew. Math.* **1908,** *1908*, 198-287.
43. David, E. E.; David, C. W. Voronoi polyhedra as a tool for studying solvation structure. *J. Chem. Phys.* **1982,** *76*, 4611-4614.
44. Neumayr, G.; Rudas, T.; Steinhauser, O. Global and local Voronoi analysis of solvation shells of proteins. *J. Chem. Phys.* **2010,** *133*, 084108.
45. Du, Q.; Superfine, R.; Freysz, E.; Shen, Y. R. Vibrational spectroscopy of water at the vapor/water interface. *Phys. Rev. Lett.* **1993,** *70*, 2313-2316.
46. Medders, G. R.; Paesani, F. Infrared and Raman Spectroscopy of Liquid Water through "First-Principles" Many-Body Molecular Dynamics. *J. Chem. Theory Comput.* **2015,** *11*, 1145-1154.
47. Duboué-Dijon, E.; Fogarty, A. C.; Hynes, J. T.; Laage, D. Dynamical Disorder in the DNA Hydration Shell. *J. Am. Chem. Soc.* **2016,** *138*, 7610-7620.
48. Halle, B. Cross-relaxation between macromolecular and solvent spins: The role of long-range dipole couplings. *J. Chem. Phys.* **2003,** *119*, 12372-12385.
49. Bellissent-Funel, M.-C.; Hassanali, A.; Havenith, M.; Henchman, R.; Pohl, P.; Sterpone, F.; van der Spoel, D.; Xu, Y.; Garcia, A. E. Water Determines the Structure and Dynamics of Proteins. *Chem. Rev.* **2016,** *116*, 7673-7697.
50. Meister, K.; Ebbinghaus, S.; Xu, Y.; Duman John, G.; DeVries, A.; Gruebele, M.; Leitner David, M.; Havenith, M. Long-range protein–water dynamics in hyperactive insect antifreeze proteins. *Proc. Natl. Acad. Sci. USA* **2013,** *110*, 1617-1622.
51. Laage, D.; Elsaesser, T.; Hynes, J. T. Water Dynamics in the Hydration Shells of Biomolecules. *Chem. Rev.* **2017,** *117*, 10694-10725.

52. Lee, J.; Kim, S.-H. Water polygons in high-resolution protein crystal structures. *Protein Sci.* **2009,** *18*, 1370-1376.
53. Bonn, M.; Bakker, H. J.; Tong, Y.; Backus, E. H. G. No Ice-Like Water at Aqueous Biological Interfaces. *Biointerphases* **2012,** *7*, 20.
54. Chen, X.; Flores, S. C.; Lim, S.-M.; Zhang, Y.; Yang, T.; Kherb, J.; Cremer, P. S. Specific Anion Effects on Water Structure Adjacent to Protein Monolayers. *Langmuir* **2010,** *26*, 16447-16454.
55. Okur, H. I.; Hladílková, J.; Rembert, K. B.; Cho, Y.; Heyda, J.; Dzubiella, J.; Cremer, P. S.; Jungwirth, P. Beyond the Hofmeister Series: Ion-Specific Effects on Proteins and Their Biological Functions. *J. Phys. Chem. B* **2017,** *121*, 1997-2014.
56. Strazdaite, S.; Meister, K.; Bakker, H. J. Orientation of polar molecules near charged protein interfaces. *Phys. Chem. Chem. Phys.* **2016,** *18*, 7414-7418.
57. Devineau, S.; Inoue, K.-i.; Kusaka, R.; Urashima, S.-h.; Nihonyanagi, S.; Baigl, D.; Tsuneshige, A.; Tahara, T. Change of the isoelectric point of hemoglobin at the air/water interface probed by the orientational flip-flop of water molecules. *Phys. Chem. Chem. Phys.* **2017,** *19*, 10292-10300.
58. Nihonyanagi, S.; Yamaguchi, S.; Tahara, T. Water hydrogen bond structure near highly charged interfaces is not like ice. *J. Am. Chem. Soc.* **2010,** *132*, 6867-9.
59. Mittermaier, A.; Kay Lewis, E. New Tools Provide New Insights in NMR Studies of Protein Dynamics. *Science* **2006,** *312*, 224-228.
60. Hub, J. S.; de Groot, B. L. Detection of Functional Modes in Protein Dynamics. *PLoS Comput. Biol.* **2009,** *5*, e1000480.
61. Yang, G.; Xu, Y. Probing chiral solute-water hydrogen bonding networks by chirality transfer effects: A vibrational circular dichroism study of glycidol in water. *J. Chem. Phys.* **2009,** *130*, 164506.
62. Rapf, R. J.; Vaida, V. Sunlight as an energetic driver in the synthesis of molecules necessary for life. *Phys. Chem. Chem. Phys.* **2016,** *18*, 20067-20084.
63. Yamagata, Y. A hypothesis for the asymmetric appearance of biomolecules on earth. *J. Theor. Biol.* **1966,** *11*, 495-8.
64. Noorduin, W. L.; Izumi, T.; Millemaggi, A.; Leeman, M.; Meekes, H.; Van Enckevort, W. J. P.; Kellogg, R. M.; Kaptein, B.; Vlieg, E.; Blackmond, D. G. Emergence of a Single Solid Chiral State from a Nearly Racemic Amino Acid Derivative. *J. Am. Chem. Soc.* **2008,** *130*, 1158-1159.
65. McBride, J. M.; Tully, J. C. Did life grind to a start? *Nature* **2008,** *452*, 161-162.
66. Blackmond, D. G. The origin of biological homochirality. *Cold Spring Harb. Perspect. Biol.* **2010,** *2*, a002147-a002147.
67. Griffith Elizabeth, C.; Vaida, V. In situ observation of peptide bond formation at the water–air interface. *Proc. Natl. Acad. Sci. USA* **2012,** *109*, 15697-15701.
68. 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.
69. Chen, S.-L.; Fu, L.; Gan, W.; Wang, H.-F. Homogeneous and inhomogeneous broadenings and the Voigt line shapes in the phase-resolved and intensity sum-frequency generation vibrational spectroscopy. *J. Chem. Phys.* **2016,** *144*, 034704.

70. Ostroverkhov, V.; Waychunas, G. A.; Shen, Y. R. New Information on Water Interfacial Structure Revealed by Phase-Sensitive Surface Spectroscopy. *Phys. Rev. Lett.* **2005,** *94*, 046102.
71. Horn, H. W.; Swope, W. C.; Pitera, J. W.; Madura, J. D.; Dick, T. J.; Hura, G. L.; Head-Gordon, T. Development of an improved four-site water model for biomolecular simulations: TIP4P-Ew. *J. Chem. Phys.* **2004,** *120*, 9665-9678.
72. Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theory Comput.* **2015,** *11*, 3696-3713.
73. Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. *J. Chem. Phys.* **1993,** *98*, 10089-10092.
74. Miyamoto, S.; Kollman, P. A. Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models. *J. Comput. Chem.* **1992,** *13*, 952-962.
75. Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C. 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.
76. Eastman, P.; Swails, J.; Chodera, J. D.; McGibbon, R. T.; Zhao, Y.; Beauchamp, K. A.; Wang, L.-P.; Simmonett, A. C.; Harrigan, M. P.; Stern, C. D., et al. OpenMM 7: Rapid development of high performance algorithms for molecular dynamics. *PLoS Comput. Biol.* **2017,** *13*, e1005659.
77. Carr, J. K.; Wang, L.; Roy, S.; Skinner, J. L. Theoretical Sum Frequency Generation Spectroscopy of Peptides. *J. Phys. Chem. B* **2015,** *119*, 8969-8983.
78. Michaud-Agrawal, N.; Denning, E. J.; Woolf, T. B.; Beckstein, O. MDAnalysis: A toolkit for the analysis of molecular dynamics simulations. *J. Comput. Chem.* **2011,** *32*, 2319-2327.
79. 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.
80. Rycroft, C. H. VORO++: A three-dimensional Voronoi cell library in C++. *Chaos* **2009,** *19*, 041111.

Appendix to Chapter 4

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 Chapter 4 was *p*-polarized, *s*-polarized, and *p*-polarized, respectively.[2]

Description of Fitting Experimental Spectra

In the spectral region 2900 – 3800 cm^{-1}, C-H and N-H stretching of the $LK_7\beta$ 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 4.2b were fit with Lorentzians according to:

$$\mathrm{Im}[\chi^{(2)}] \propto \mathrm{Im}[\sum_{q} \frac{A_q}{\omega_{IR} - \omega_q - i\Gamma_q}] \quad (1)$$

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 4.2b) is determined by residual analyses as shown in Figure A4.2. Eleven peaks are used to fit the spectra, and they are tentatively assigned as shown in Table A4.1.

MD equilibration details

See the Methods section in Chapter 4 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 4.4 were selected using a combination of MDAnalysis selection logic,[4] in-house code, and the Voronoi tessellation library freud,[5] which uses the voro++ engine[6] 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 4.4a (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 4.4b (backbone)

Backbone-associated water molecules were selected using the following MDAnalysis selection logic: *byres (resname HOH and (around 4.5 ((not resname HOH) and (name N*

H CA HA C O HN1 HN2))) and (not around 3.0 ((resname LEU) and not name N H CA HA C O HN1 HN2) or ((resname LYS) and not name N H CA HA C O HN1 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 4.4c (sidechains)

Sidechain-associated water molecules were selected using the following MDAnalysis selection logic: *byres (resname HOH and (around 4.5 ((resname LYS) and not name N H CA HA C O HN1 HN2) or ((resname LEU) and not name N H CA HA C O HN1 HN2)) and (not around 3.0 (not resname HOH) and (name N H CA HA C O HN1 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 4.4d (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 NH_3^+ 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 NH_2 groups were also considered as NH groups.

Figure 4.4e: (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 ((resname HOH) and (around 3.5 (not resname HOH NA CL Na Cl)))*. Then, water molecules donating a hydrogen bond to a C=O group were identified.

Figure 4.4f: (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 4.4g: (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 ((resname HOH) and (around 1.6 (not resname 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 4.4h (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 4.4i (sidechains – hydrogen bonded to lysine NH_3^+)

The same procedure was followed as for the C=O groups and NH groups, except water molecules accepting hydrogen bonds from NH_3^+ 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 A4.4. 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.

Parameter	Figure 4.2b (L-) $LK_7\beta$ H_2O	Figure 4.2b (L-) $LK_7\beta$ $H_2{}^{18}O$	Tentative Assignments
ω_1	2950 ± 1	2952 ± 1	C-H ($LK_7\beta$)
A_1	-0.7 ± 0.2	-2.8 ± 0.9	
Γ_1	4.2 ± 0.7	12.5 ± 1.6	
ω_2	2973 ± 1	2970 ± 2	C-H ($LK_7\beta$)
A_2	-3.1 ± 0.6	-13.1 ± 12.2	
Γ_2	13.1 ± 1.3	34.3 ± 9.1	
ω_3	2990 ± 0*	2990 ± 0*	C-H ($LK_7\beta$)
A_3	40.7 ± 0*	94.1 ± 35.6	
Γ_3	48.0 ± 2.6	60.7 ± 0.0*	
ω_4	3000 ± 2	3001 ± 7	C-H or N-H ($LK_7\beta$)
A_4	-36.7 ± 3.2	-71.1 ± 19.7	
Γ_4	56.6 ± 4.4	58.0 ± 1.5	
ω_5	3151 ± 5	3125 ± 4	O-H (H_2O) and/or F.R.** with H_2O bend overtone
A_5	-35.0 ± 7.4	-46.3 ± 10.3	
Γ_5	119.5 ± 16.8	125.7 ± 17.6	
ω_6	3252 ± 1	3255 ± 1	N-H ($LK_7\beta$)
A_6	-74.1 ± 2.9	-139.6 ± 2.3	
Γ_6	35.9 ± 0.6	36.6 ± 0.4	
ω_7	3263 ± 1	3261 ± 1	N-H ($LK_7\beta$)
A_7	108.8 ± 0.0*	179.1 ± 0.0*	
Γ_7	40.0 ± 0.7	39.6 ± 0.5	
ω_8	3366 ± 6	3350 ± 2	O-H (H_2O)
A_8	20.3 ± 4.0	7.8 ± 2.1	
Γ_8	85.9 ± 10.0	50.1 ± 6.6	
ω_9	3426 ± 2	3420 ± 10	N-H ($LK_7\beta$)
A_9	-3.2 ± 1.6	-4.5 ± 0.8	
Γ_9	31.2 ± 7.6	40.9 ± 7.8	
ω_{10}	3468 ± 2	3483 ± 1	N-H ($LK_7\beta$)
A_{10}	-1.6 ± 0.2	-1.4 ± 0.2	
Γ_{10}	20.0 ± 0.0*	20.0 ± 0.0*	
ω_{11}	3555 ± 10	3538 ± 6	O-H (H_2O)
A_{11}	-34.4 ± 1.8	-9.5 ± 1.5	
Γ_{11}	161.1 ± 12.2	95.6 ± 7.5	

*Parameters held during fitting.
**F.R. = Fermi Resonance

Table A4.1. The fitting parameters (see equation 1, above) used for the 11-peak fitting shown in Figure A4.2. Water contributions are suggested based on $H_2{}^{18}O$ substitution.

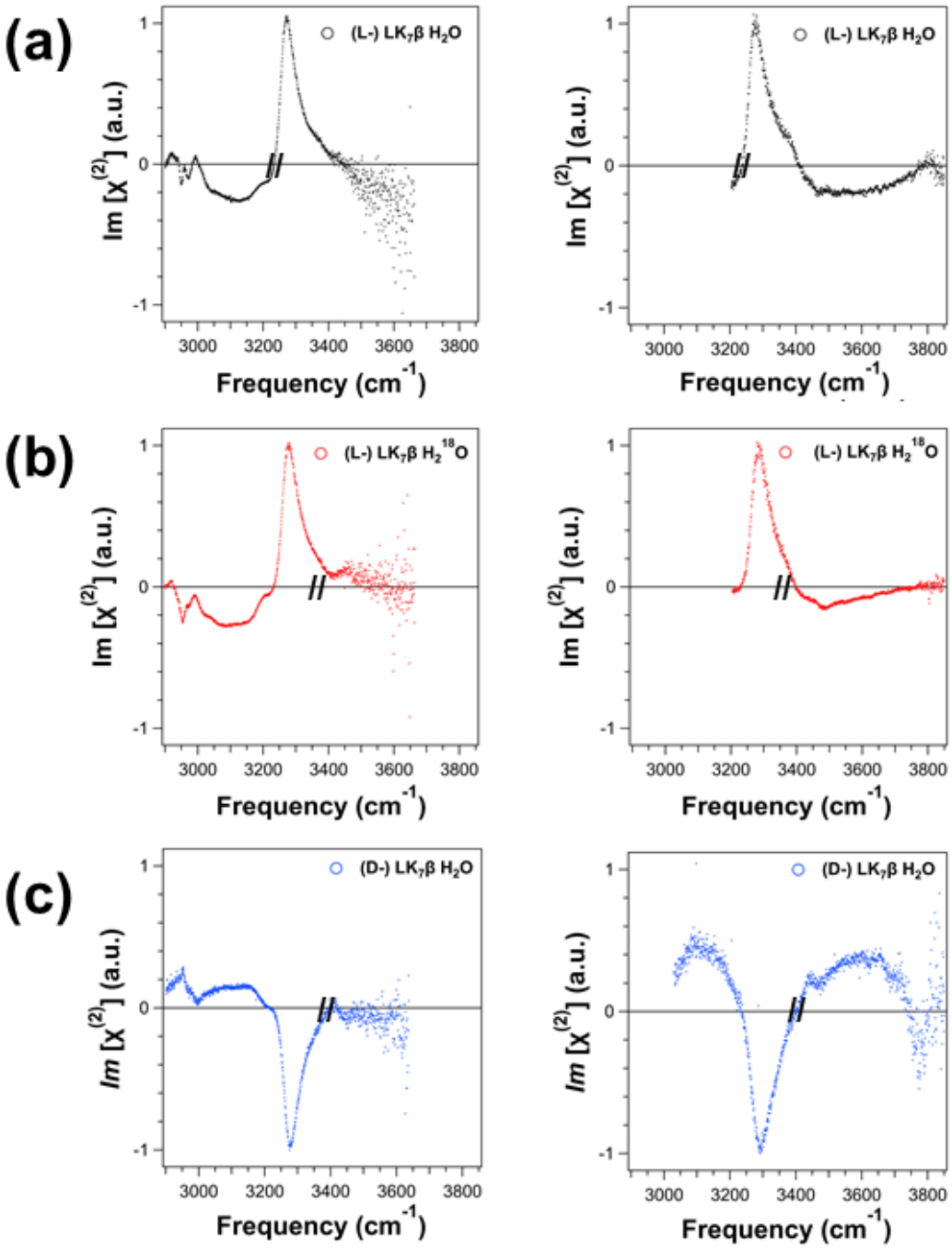

Figure A4.1. The two spectral windows collected for each spectrum presented in Figure 4.2 in Chapter 4. The two individual spectra were overlaid and normalized to the N-H stretching peak of $LK_7\beta$ at ~3270 cm^{-1} and stitched together at the point indicated with hashmarks. The noisy regions that appear at the high frequency ends of the spectra are due to normalization of the broadband spectra by clean quartz. a.u.: arbitrary units.

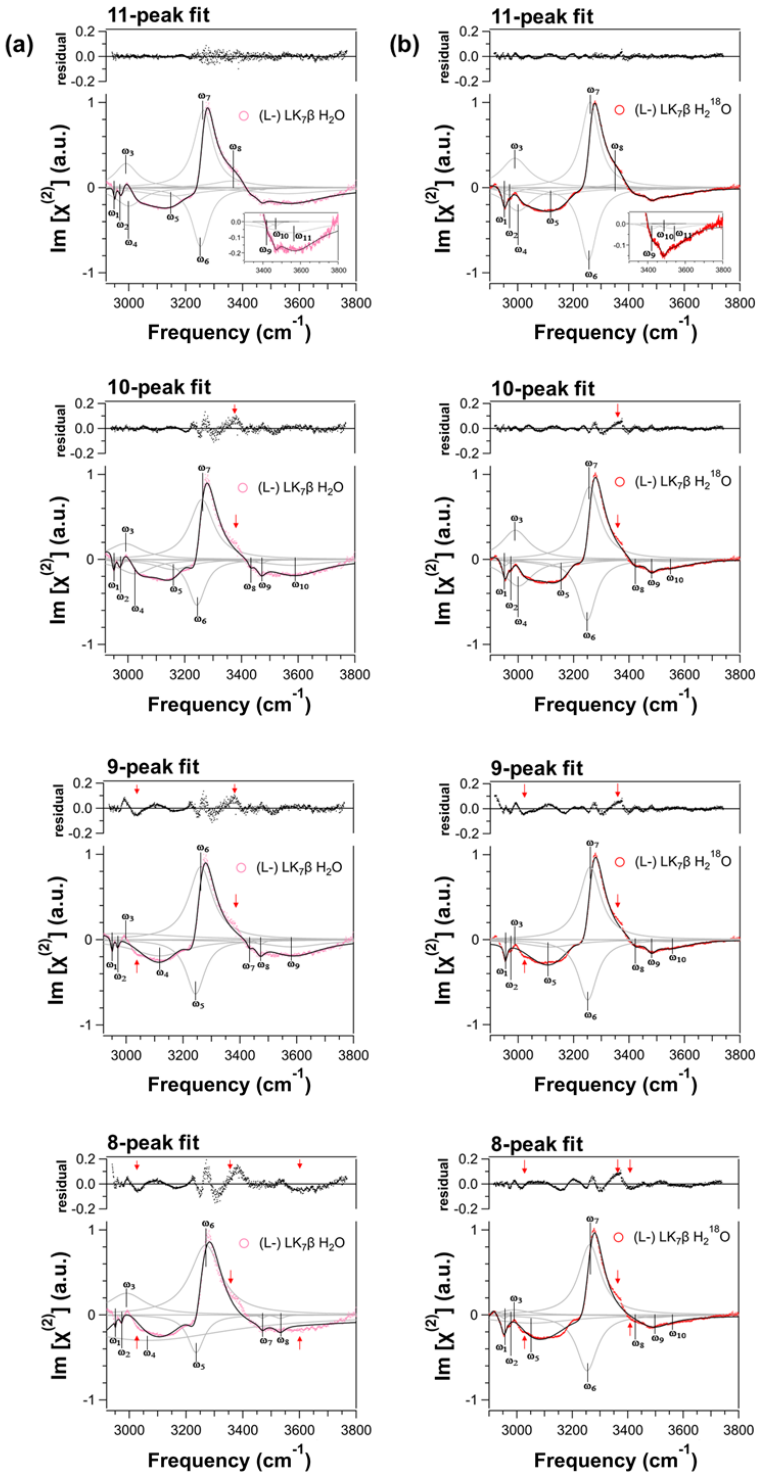

Figure A4.2. Fitting and residual analyses of the fitting to the experimental spectra of (a) $LK_7\beta$-H_2O and (b) $LK_7\beta$-$H_2{}^{18}O$ presented in Figure 4.2b in Chapter 4 with (from top to bottom) 11, 10, 9, and 8 vibrational peaks. Fitting parameters for the 11-peak fitting (top row) are reported in Table A4.1. 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 4.2b of Chapter 4.

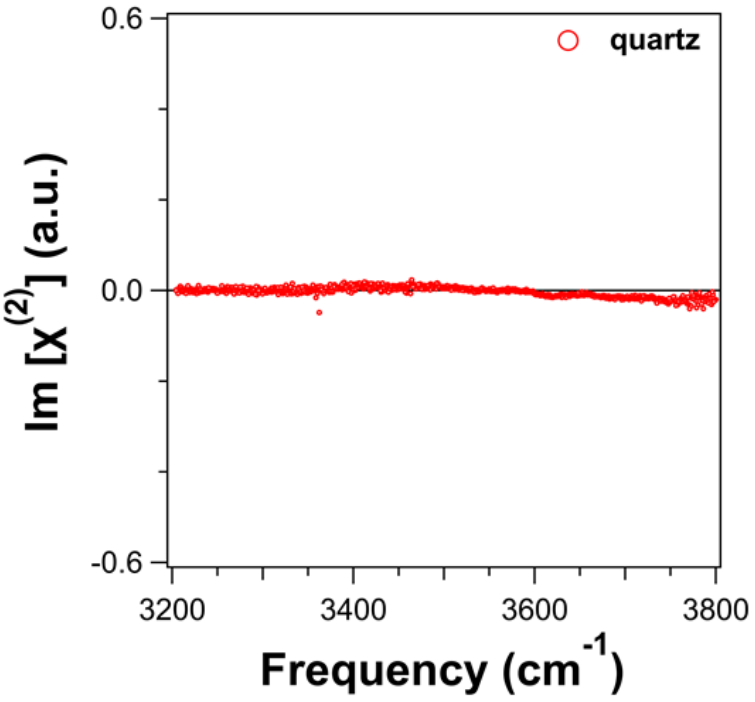

Figure A4.3. Chiral SFG spectrum of clean quartz substrate in the absence of LK$_7$β. The spectrum is normalized by GaAs. The spectrum shows no vibrational resonances. 2 mins. acquisition time. a.u.: arbitrary units.

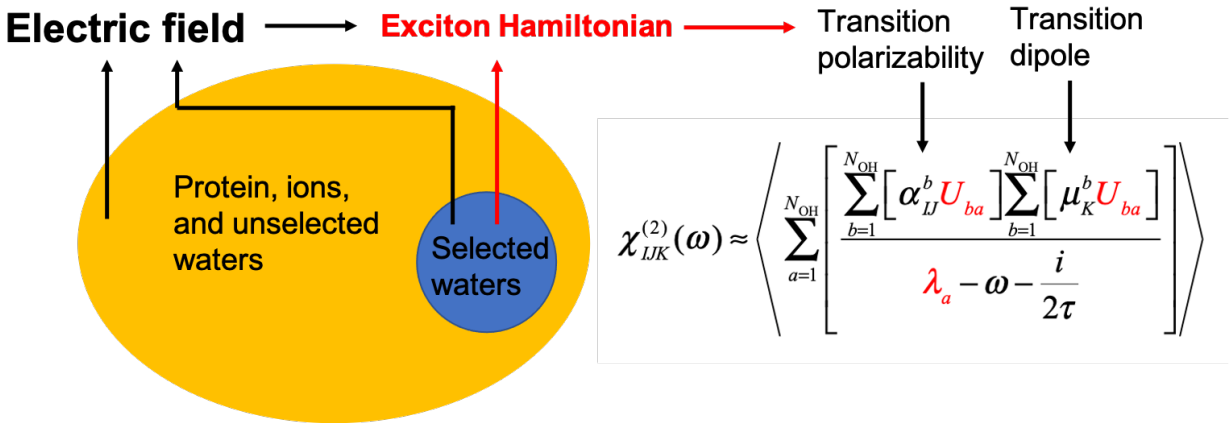

Figure A4.4. A schematic of our adaptation of Skinner's electric field mapping approach to calculate SFG spectra arising from subsets of the water molecules in the system. Here, $\boldsymbol{\alpha}$ is the transition polarizability tensor, $\boldsymbol{\mu}$ is the transition dipole vector, $\mathbf{U}$ are the eigenvectors of the Hamiltonian, and $\boldsymbol{\lambda}$ are the eigenvalues. τ is the vibrational lifetime (1.3 ps here).

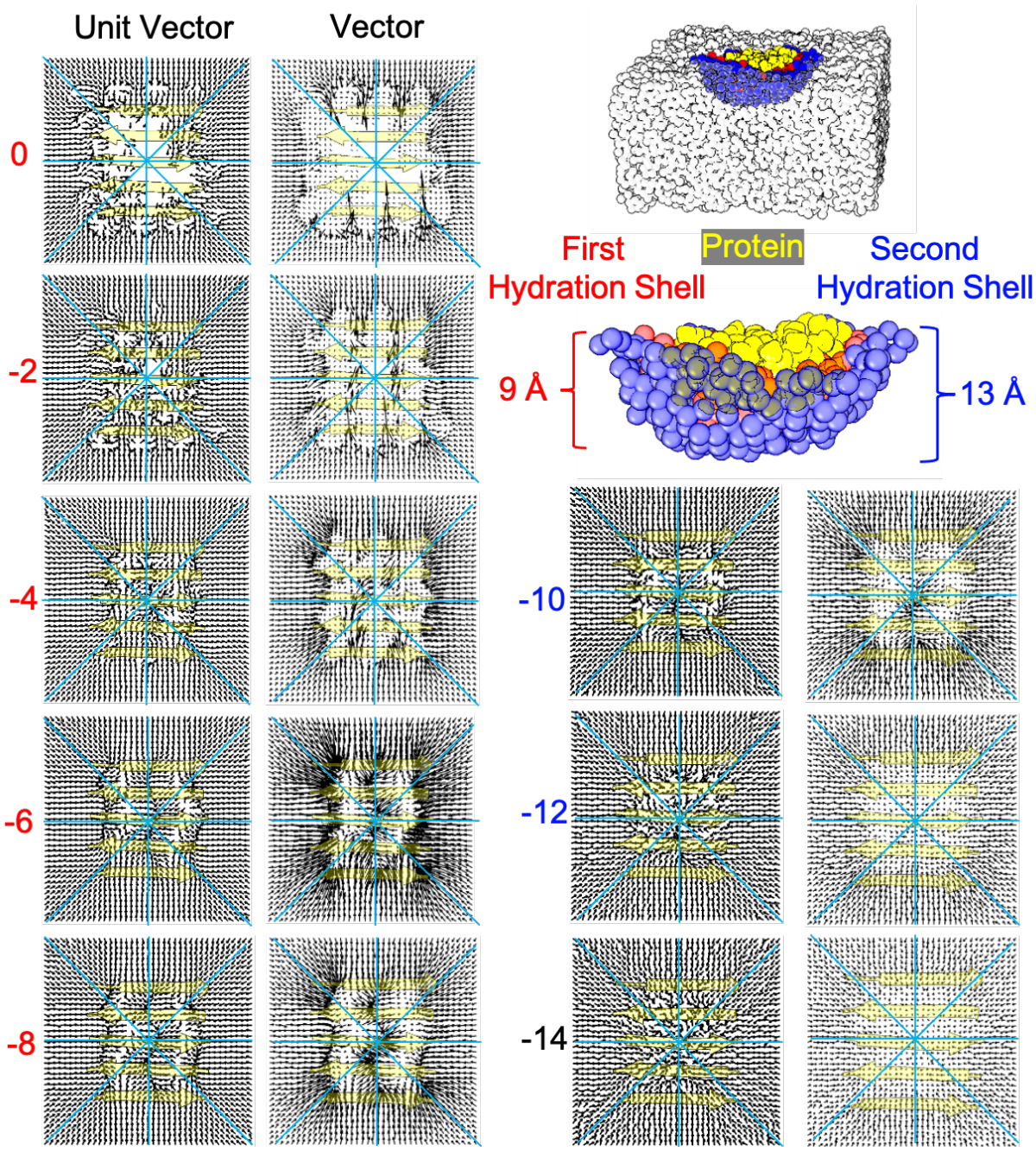

Figure A4.5. Water dipole moment vector directions at various depths (in Å) from the vacuum-water interface. The $LK_7\beta$ protein is shown in yellow. The blue lines are a guide for the eye. This Figure Ahows that the lack of reflection plane perpendicular to the vacuum-water interface seen in Figure 4.6 in Chapter 4 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.

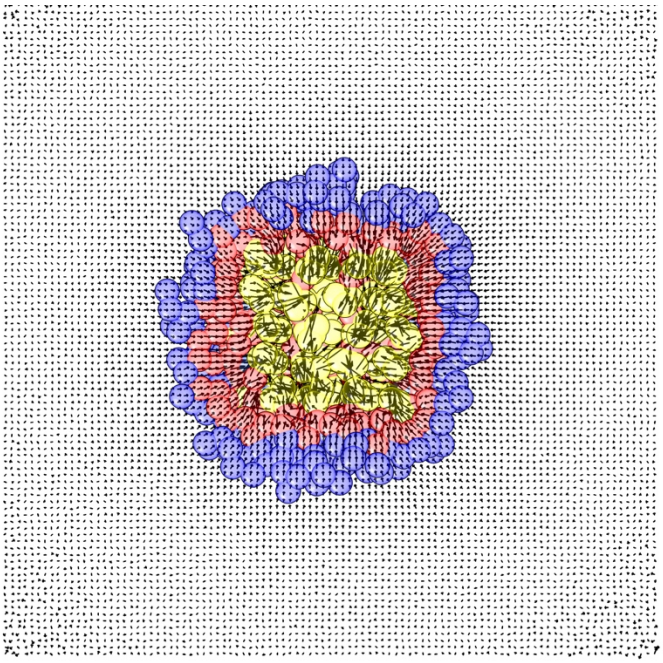

Figure A4.6. Unnormalized relative magnitudes of the dipole moment vectors at various locations around $LK_7\beta$. The ordering of water dipoles past the second hydration shell is minimal relative to the ordering near the protein (see Figure 4.6 in Chapter 4 for comparison).

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. *Quat. Rev. Biophys.* **2002,** *35*, 369-430.
4. Michaud-Agrawal, N.; Denning, E. J.; Woolf, T. B.; Beckstein, O. MDAnalysis: 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.

Chapter 5: Design of an electrostatic frequency map for the NH stretch of the protein backbone and application to chiral sum frequency generation spectroscopy

Adapted from the following reference with permission:

Konstantinovsky, D.; Perets, E. A.; Santiago, T.; Olesen, K.; Wang, Z.;

Soudackov, A. V.; Yan, E. C. Y.; Hammes-Schiffer, S.

J. Phys. Chem. B 2023, *127*, 2418-2429.

DOI: 10.1021/acs.jpcb.3c00217

Abstract

We develop an electrostatic map for the vibrational NH stretch (amide A) of the protein backbone with a focus on vibrational chiral sum frequency generation spectroscopy (chiral SFG). Chiral SFG has been used to characterize protein secondary structure at interfaces using the NH stretch and to investigate chiral water superstructures around proteins using the OH stretch. Interpretation of spectra has been complicated because the NH stretch and OH stretch overlap spectrally. Although an electrostatic map for water OH developed by Skinner and coworkers was used previously to calculate the chiral SFG response of water structures around proteins, a map for protein NH that is directly responsive to biological complexity has yet to be developed. Here, we develop such a map, linking the local electric field to vibrational frequencies and transition dipoles. We apply the map to two protein systems and achieve much better agreement with experiment than was possible in our previous studies. We show that couplings between NH and OH vibrations are crucial to the lineshape, which informs the interpretation of chiral SFG spectra, and that the chiral NH stretch response is sensitive to small differences in structure. This work increases the utility of the NH stretch in biomolecular spectroscopy.

Introduction

Of the four key vibrational normal modes of the protein amide group, the NH stretch mode (amide A) is perhaps the least studied. This is most likely because the NH group and the OH group of water vibrate in the same frequency range (3000–4000 cm^{-1}), and in most vibrational spectra of biological systems, the water signal overwhelms the signal from the NH stretch. However, because the NH group participates in hydrogen bonding interactions with both other parts of the protein during the formation of secondary structures as well as the water around the protein, the NH stretch contains valuable information about the local environment. Unfortunately, this information is hidden in conventional vibrational spectroscopy, and it cannot be brought out with the use of D_2O to shift the water response because the hydrogen atom of the NH group exchanges with water and shifts in the same way.

The NH stretch can be isolated in higher-order spectroscopies. Vibrational chiral-selective sum frequency generation spectroscopy (chiral SFG) with homodyne detection can detect the NH stretch and help characterize protein secondary structures.[1, 2] Ultrafast methods such as two-dimensional infrared spectroscopy (2D-IR) can pick up the NH stretch through its interaction with other modes as a cross-peak.[3, 4] Methods to model 2D-IR and chiral SFG of large systems are closely related,[5, 6] so being able to model one method computationally corresponds to being able to model the other method as well. The key task is to find a way to relate local environments seen in molecular dynamics (MD) simulations to a fluctuating exciton Hamiltonian, which contains vibrational frequencies and couplings,[5, 7-15] and thereby extract structural and dynamic information about molecular systems from spectra.

In 2004, Simpson and coworkers established symmetry-based chiral SFG theory to describe vibrational chiral SFG in the absence of electronic resonance as surface-selective in uniaxial (C_∞) systems.[16, 17] Over the past decade, chiral SFG has developed into a powerful technique for probing the secondary structure of chiral biomacromolecules at interfaces and the solvent around them.[1, 18-25] There are aspects of chiral SFG that are not yet understood. For example, we do not have a straightforward mapping from chiral phase (signal sign) to dipole orientation, unlike in conventional (achiral) SFG. However, chiral SFG response of the NH stretch can provide vibrational signatures for distinguishing secondary structure of protein,[2] and the phase of the response correlates with the absolute chirality of protein.[20] Hence, being able to simulate the chiral SFG response of the protein NH stretch can help interpret chiral SFG spectra of protein and extract information about protein structure.

In addition, Petersen's group used chiral SFG to demonstrate the existence of a chiral spine of hydration in the minor groove of DNA.[26] Later, we showed that chiral SFG can also detect hydration structures around protein.[20-22] In particular, we demonstrated that this technique is specifically sensitive to the first hydration shell of water around a protein.[25] However, the chiral OH signals produced by water overlap with the chiral NH signals produced by the protein.[8, 27] Although signal at 3600-3800 cm^{-1} can be unmistakably assigned to the OH stretch, signal at lower frequencies may be due to the NH or OH stretch. Previously, we showed that isotope labeling with $H_2{}^{18}O$ allows identification of OH stretch peaks by red-shifting water peaks by roughly 12 cm^{-1}.[20, 22, 25] However, NH and OH interact with each other. Thus, the spectral perturbation induced by the ^{18}O substitution may also be due to a delocalized mode involving OH and NH. Furthermore,

the NH stretch peaks provide useful structural information as well as information about hydrogen-bonding interactions within the protein and between the protein and water. Hence, the ability to model the NH stretch spectra will reveal information about not only protein structures but also local interactions of water with the protein.

Vibrational spectra can be modeled computationally using MD simulations with an electrostatic mapping approach or with a direct dipole-polarizability time-correlation function approach.[6, 8-12, 14, 15, 28-38] Recently we used the Skinner group's OH stretch electrostatic map[6, 8-10, 39-41] to predict chiral SFG signals produced by water around proteins. This map links the electric field along the OH bond at the OH hydrogen to the vibrational frequency, transition dipole, and transition polarizability. Although the map was developed for pure water,[6, 9, 10, 42, 43] it can be extended to aqueous systems that include biomolecules and their atomic partial charges, as shown in our previous work.[19, 20, 25] Of particular interest is $LK_7\beta$, which forms a stable antiparallel β-sheet at the air-water interface. This system is an ideal model system for SFG, which is inherently surface selective.[44-48] Although qualitative agreement was achieved between our computational and experimental spectra of water hydrating $LK_7\beta$ at interfaces,[19, 20, 25] the lack of NH peaks in the computational spectra made comparison between experiment and theory challenging. Any insights hidden in the NH response remained elusive.

The Mukamel group has developed maps for the vibrational modes of the peptide bond group, including the NH stretch using N-methyl acetamide (NMA).[49, 50] Although NMA contains all the relevant vibrational modes, it is very different from an actual protein environment. The frequencies extracted from this map range from ~3400 to 3500 cm^{-1}. This range does not match the full range of frequencies exhibited by the NH stretch in

biological systems.[24, 27] An alternative approach is needed to model lower frequencies and understand complex biological environments.

Here, we report the development of an electrostatic map for the NH stretch of the protein backbone. Previous work has demonstrated that the NH signal produced by proteins is primarily due to the backbone.[1, 24] Our map applies to all NH groups in the backbone except C-terminal amide (NH_2) capping groups that most likely contribute little to the total NH response. This map, together with previously developed maps for the OH stretch, enables computational modeling of the full chiral SFG spectra, including both OH and NH modes and the coupling between them. These full spectra contain information about the local hydrogen-bonding environment and the protein-solvent interactions. In the future, our map may allow simulation of 2D-IR and related spectra that directly probe energy exchange between a protein and the surrounding solvent. Our map is validated using a combination of previously published and new experimental data. We focus on two closely related protein systems – $LK_7\beta$ (Ac-LKLKLKL-NH_2) and $LE_7\beta$ (Ac-LELELEL-NH_2). Each is a seven-residue peptide consisting of alternating nonpolar and polar residues. The nonpolar residue is leucine and the polar residue is positively charged lysine in $LK_7\beta$ or negatively charged glutamate in $LE_7\beta$. These protein systems form antiparallel β-sheets at the air-water interface. We find that the map reproduces the experimental chiral SFG spectra of these systems quite well. Moreover, analysis of the spectra provides insights into subtle differences in hydrogen bonding between the two apparently similar protein systems. This analysis demonstrates that the NH group is a useful probe for the local environment in biological systems when heterodyne chiral SFG attenuates the overwhelming OH stretch background.

Computational methods

Designing the NH stretch electrostatic map

Our NH stretch electrostatic map was based on Skinner's method of mapping electric fields around water molecules to spectroscopic quantities (i.e., vibrational frequency, transition dipole, transition polarizability, vibrational couplings). In brief, Skinner's method[6] consists of running an MD simulation of bulk water using a water model such as TIP4P,[51] extracting small water clusters from the trajectory, performing density functional theory (DFT) calculations to determine local vibrational frequencies and other quantities, and correlating these properties to the local electric field (as calculated from local point charges – this is distinct from the long range electric field that is typically calculated using the Ewald method). Importantly, rather than optimizing the geometry of the water cluster and using the Hessian to compute the harmonic frequencies, Skinner's group stretched the O–H bond and solved the vibrational Schrödinger equation for the resulting one-dimensional potential energy curve.[6] This procedure results in more accurate vibrational frequencies because it accounts for anharmonic effects and does not involve relaxing (i.e., changing) the configuration for which the electric field is measured.

The first step of our development of an NH stretch electrostatic map was to generate configurations of a relatively small protein that was amenable to DFT calculations but also captured the essence of protein secondary structure. As both α-helices and β-sheets, the two most common secondary structures in proteins, involve hydrogen bonds (H-bonds) between NH groups and carbonyl groups, we focused on reproducing this interaction. We selected a tetraalanine peptide that is able to fold on itself to produce an intramolecular NH/carbonyl H-bond between the first (C=O) and fourth (NH) alanine residue (Figure 5.1).

Because the NH stretch is highly local,[52] the "folded" structure could effectively simulate NH interactions that would occur in both α-helices and β-sheets. Capping groups (acetyl on the N-terminal, $–NH_2$ on the C-terminal) were used to avoid charged terminus effects. To overcome the significant thermodynamic barriers to this kind of folding, we conducted a metadynamics enhanced sampling simulation with the distance between the NH nitrogen and carbonyl oxygen as the collective variable and collected configurations sampling the H-bonded state.[53] The simulation was performed in explicit TIP4P-Ew water[54] with ~1000 water molecules. See the SI for details about the metadynamics simulation.

We also extracted configurations by sampling the non-H-bonded state and created a separate electrostatic map for that case. In total, we extracted 400 "H-bonded" and 400 "free" configurations. In the free configurations, the NH group may still donate a H-bond to a water molecule. (Note that NH as an acceptor from water is vanishingly rare in this type of system.) We defined a H-bond using a heavy atom donor−acceptor distance cutoff of 3.2 Å and an angle greater than 135°. Because most free configurations of interest would involve NH groups on the outer edge of β-sheets or in unstructured loops, to ensure that the free configurations were not close to the H-bonded ones, we only considered free configurations where the donor−acceptor distance between the carbonyl oxygen and the amide nitrogen was greater than 4.5 Å, corresponding mostly to extended conformations of tetraalanine. Intermediate configurations (3.2–4.5 Å) were not considered for the maps.

After gathering these configurations, we extracted the entire peptide and the three closest water molecules to the fourth residue's NH hydrogen. We then stretched the NH bond about the NH center of mass, keeping all other atoms frozen. The center of mass remained stationary during the stretch, so each atom moved in inverse proportion to its

mass. The NH distance ranged from 0.58 Å to 1.70 Å over 15 evenly spaced configurations. We then performed single-point energy calculations on each configuration at the B3LYP/6-311++G(d,p)[55-59] level of theory using Gaussian 16[60] with default convergence criteria.[6] We also included implicit water using the conductor-like polarizable continuum model (C-PCM) to account for polarization effects due to long range interactions with the bulk solvent.[61-64] The inclusion of implicit water was necessary to bring the frequencies down to the experimental range (~3200-3500 cm^{-1}). We then solved the one-dimensional Schrödinger equation for the resulting potential energy curve using the Fourier grid Hamiltonian method[65] using 1,500 quadratically interpolated grid points for high accuracy. We used the reduced mass of NH as the effective mass because both atoms were moving about their center of mass. We used a scaling factor of 1.0006 for the frequency based on comparing an analogous calculation with NMA in vacuum to the experimental value cited by Mukamel and coworkers.[49, 66] To estimate the NH dipole derivative, we also performed a constrained geometry optimization for each configuration, allowing only the hydrogen of the NH group to relax, followed by a dipole derivative calculation for the NH stretch mode.[8]

We then calculated the electric field at various points around the NH group. We found that computing the electric field at points slightly past the hydrogen atom along the NH bond led to better correlations. To maximize the accuracy of the map in complex environments, we sampled the electric field at 51 (3 × 17) points, with one group of 17 points centered at 1.0 Å past the N toward the H along the NH bond, a second group centered at 1.5 Å, and a third group centered at 2.0 Å (See Figure A5.1 for details). For each central point, the group of 17 points included 16 points located 0.1 and 0.2 Å in all

four directions from the central point along y and z. Moreover, the electric field was sampled in three directions for a total of 153 electric field components. The directions were defined by the NH bond (unit vector x), the vector perpendicular to the NH bond vector and pointing toward the adjacent residue's carbonyl carbon (unit vector y), and the cross product of these vectors, $\mathbf{z} = \mathbf{x} \times \mathbf{y}$. We used these directions to set up our 51 points at each center.

For both the design and application of this model, the electric fields were calculated from the solvent charges (i.e., water molecules and ions) and the charges of the bound carbonyl carbon and oxygen for the H-bonded configurations and from only the solvent charges for the free configurations. TIP4P-Ew[54] charges were used for the water, standard parameters associated with the TIP4P-Ew model were used for the ions when present,[67, 68] and the AMBER ff14SB force field[69] was used for the carbonyl group charges. Solvent charges within 10 Å of the NH sampling points were included. The distance cutoff was applied to individual atoms. We found that including whole water molecules (if any atom was within 10 Å of the sampling point) produced essentially the same distribution of frequencies and transition dipoles as an individual atom cutoff (Figure A5.5).

During the model design stage, the calculated electric fields, as well as the frequencies and transition dipole derivatives, were used to train a random-forest regression model. The input is the 153 electric field components described above, and the output is the frequency and transition dipole derivative for the specified NH group. We chose this nonlinear model both because it offered better inference about the data than linear regression (measured by r^2) and because it tends to give more reasonable predictions for rare cases where the input data falls outside the bounds of the training data. Such a model

is robust to unexpected and complex environments. Random forest models tend to respond to an extreme input with an output close to the edge of that produced by the training data, which is a reasonable response for spectroscopy, where peaks do not generally shift past a certain point regardless of the extremes of the environment. In contrast, linear regression can produce wildly inaccurate predictions for even a small deviation from the training data bounds. However, the 153-point model was designed to capture as much local complexity as possible to avoid such edge cases. All r^2 values from the model were above 0.90 (see the SI for details).

In addition, a linear regression model was designed to determine the transition, off-diagonal, position matrix element defined as

$$x_{01} = \langle \psi_0 | \hat{x} | \psi_1 \rangle \tag{1}$$

where ψ_0 is the ground state vibrational wave function and ψ_1 is the first excited state vibrational wave function. In the model, the input is the frequency of the specified NH mode, and the output is the transition position matrix element. The transition dipole could then be approximated to first order by multiplying the transition dipole derivative by the transition position matrix element. The combination of the random forest model and the linear regression model constitutes the NH mapping model developed herein. The SI provides more details about this model, including binaries that enable its direct use with Python.

To apply this model, an MD trajectory of the system is propagated using the same force fields as those used to train the model. For each configuration sampled, the electric fields induced by the solvent and ion charges, as well as the H-bonded carbonyl group charges for the H-bonded cases, are calculated for each NH group. The random forest

regression map is used to determine the frequency and transition dipole derivatives from these electric fields, and the linear regression model is used to obtain the transition position matrix element from the frequency. These quantities can be used to compute IR, Raman, or SFG spectra. Note that the same map can also be used to compute higher-dimensional spectra from correlation functions, including 2D-IR. The development of the NH map is summarized in Figure 5.2.

Estimating the relative intensities of the NH and OH vibrational response.

When spectra due to both NH and OH responses are calculated, the relative magnitudes of quantities, such as the transition dipole and transition polarizability, are important. The polarizability was approximated as a bond polarizability, so that the transition polarizability is given by[8, 70]

$$\alpha_{01,ij} = x_{01}((\alpha'_{\|} - \alpha'_{\perp})u_i u_j + \alpha'_{\perp}(\hat{i} \bullet \hat{j})) \qquad (2)$$

where α_{01} is the transition polarizability tensor, x_{01} is given by equation 1, $\alpha'_{\|}$ is the component of the bond polarizability parallel to the bond axis, $\alpha'_{\perp}$ is the component of the bond polarizability perpendicular to the bond axis, u is the bond unit vector, and *i* and *j* are indices or unit vectors representing the *x*, *y*, and *z* directions. The second term is zero for chiral SFG, where the relevant polarizability component is either *zy* or *zx*,[12] so only the magnitude of the first term needs to be evaluated. For OH groups, Skinner and coworkers used an $\alpha'_{\|}/\alpha'_{\perp}$ ratio of 5.6.[8] The precise value of this ratio only controls the magnitude of the chiral spectrum and does not change the lineshape, so for NH we chose to use the same ratio. The ratio of the magnitudes of the polarizability derivative ($\alpha'_{\text{iso,NH}}/\alpha'_{\text{iso,OH}}$) was

estimated by calculating the transition polarizability derivative matrix associated with the OH stretch for water and the NH stretch for NMA in vacuum to obtain the isotropic transition polarizability derivative magnitude using[43]

$$\alpha'_{\mathrm{iso}} = \frac{1}{3}(\alpha'_{xx} + \alpha'_{yy} + \alpha'_{zz}) \tag{3}$$

where α'_{xx}, α'_{yy}, and α'_{zz} are the diagonal components of the resulting tensor. This ratio (1.23/0.82) was then applied to the transition polarizability of NH as a prefactor in equation 2 to estimate the ratio of its magnitude to that of OH. Note that this ratio is quite close to unity, so the relative prominence of NH peaks in the combined NH and OH spectrum does *not* reflect a substantially larger transition dipole or polarizability, but rather the greater chiral ordering of the NH groups relative to the water and the contribution of only the first hydration shell to the water signal.[25] If we had simply assumed the transition polarizability derivatives were the same, the predicted spectra would have been very similar.

Intermediate configurations

The design of the map considered only groups that were very far from a recipient carbonyl group to be free NH. In real systems, however, occasionally there *is* a nearby carbonyl, but the H-bond between it and an NH group does not quite conform to our definition of a H-bond (i.e., the heavy atom donor−acceptor distance is greater than 3.2 Å). In this case, we assigned the NH group in question to the H-bonded map as long as the heavy atom donor−acceptor distance was less than 3.4 Å.

Calculating chiral SFG spectra using MD trajectories with electrostatic maps

Chiral SFG spectra were calculated from MD trajectories as previously reported by our group,[19, 20, 25] using a method originally developed by Skinner and coworkers.[41] In brief, the spectra were calculated using an inhomogeneous limit approximation, which neglects dynamical effects but is much more computationally efficient than the time-dependent approach.[8, 35] In particular, the absorptive (imaginary) component of the second-order response tensor component $\chi^{(2)}_{IJK}$ is approximated as[41]

$$\mathrm{Im}[\chi^{(2)}_{IJK}(\omega)] \approx \mathrm{Im}\left[\left\langle \sum_{a}^{N_{OH/NH}} \frac{\left(\sum_{b}^{N_{OH/NH}} U_{ba}\alpha_{01,b}{}^{IJ}\right)\left(\sum_{b}^{N_{OH/NH}} U_{ba}\mu_{01,b}{}^{K}\right)}{\lambda_a - \omega - \frac{i}{2\tau}} \right\rangle\right] \quad (4)$$

where ω is the response frequency, U is the eigenvector matrix of the exciton Hamiltonian, $\boldsymbol{\lambda}$ is the eigenvalue vector of the exciton Hamiltonian, $\boldsymbol{\alpha}_{01}$ is the transition polarizability tensor, $\boldsymbol{\mu}_{01}$ is the transition dipole vector, τ is the vibrational lifetime, and the indices *a* and *b* run over each OH and each NH group. The average is over all configurations sampled in the MD trajectory. The exciton Hamiltonian matrix in the local-mode basis contains vibrational frequencies on the diagonal and vibrational couplings on the off-diagonal. The technical details for the calculation of the couplings have been described in previous work.[6, 8, 71] The transition dipole, transition polarizability, frequencies, and couplings are obtained from the electrostatic maps described above and in previous work.[6, 8, 10, 19, 20, 25, 41, 43] To calculate chiral SFG spectra, the component $\chi^{(2)}_{ZYX}$ must be obtained from a uniaxial system (i.e., a system with C_∞ symmetry). To simulate the effect of a uniaxial system from

an MD trajectory, the quantity $\chi^{(2)}_{ZYX}$ was obtained as $\chi^{(2)}_{ZYX} - \chi^{(2)}_{ZXY}$, as described in prior work.[12]

Experimental methods

The (L-) $LK_7\beta$ (Ac-Leu-Lys-Leu-Lys-Leu-Lys-Leu-NH_2) (98.39% purity) and (L-) $LE_7\beta$ (Ac-Leu-Glu-Leu-Glu-Leu-Glu-Leu-NH_2) (98.51% purity) peptides were obtained as lyophilized powders (GL Biochem Ltd., Shanghai). For heterodyne experiments on quartz substrate, stock solutions of (L-) $LK_7\beta$ at 1.0 mM and (L-) $LE_7\beta$ at 0.5 mM were made by dissolving the powders in H_2O (Millipore Synergy UV-R system, 18.2 MΩ), or in $H_2{}^{18}O$ (97%, SIGMA-ALDRICH Co., St. Louis, MO). The quartz substrate used in these experiments was a right-handed z-cut α-quartz crystal (Conex Systems Technology nc., San Jose, CA), which was rinsed with copious H_2O, dried with nitrogen, and plasma-cleaned (Harrick Plasma; PDC-32G) on "low" for 15 min. Samples were prepared by applying 5 – 10 μL of peptide solution directly onto the clean quartz substrate, which was subsequently dried in a sealed container, with desiccant. The quartz substrate was used as phase reference to obtain internal heterodyne spectra using a broad-bandwidth SFG spectrometer[72] following the procedure described previously.[20, 25, 47] For the experiments at the air-water interface, 1 mM stock solutions of (L-) $LK_7\beta$ and (L-) $LE_7\beta$ were prepared. (L-) $LE_7\beta$ was dissolved in phosphate buffer (10 mM pH = 7.4) and (L-) $LK_7\beta$ was dissolved in H_2O. Then, 100 μL of the stock solutions were added to 3.9 mL of 10 mM phosphate buffer (pH = 7.4) in a Teflon dish using a Hamilton syringe to give a final peptide concentration of 25 μM. The intensity of the main NH peak was monitored to determine

when the system reached equilibrium, roughly two hours after addition of peptide. Homodyne spectra were then obtained (see the SI). All stock solutions of the peptides were prepared in batches and separated into aliquots, which were then flash frozen in liquid nitrogen and stored at –80 °C. Aliquots were thawed on the day of the experiment and discarded after use.

Results and discussion

Model protein systems

After developing our NH map, we applied it to two related protein systems, $LK_7\beta$ and $LE_7\beta$. $LK_7\beta$ is a seven-residue peptide consisting of alternating lysine residues (polar, hydrophilic) and leucine residues (nonpolar, hydrophobic).[73] $LK_7\beta$ forms antiparallel β-sheets at the air-water interface, where the leucine residues point into the air and the lysine residues point into the water (Figure 5.3). $LK_7\beta$ is positively charged due to the lysines and has been used as a model system for SFG studies.[44,74] $LE_7\beta$ is a closely related peptide that has negatively charged glutamate residues instead of lysines. $LE_7\beta$ is negatively charged but also forms β-sheets at the air-water interface. Folding at the interface makes $LK_7\beta$ and $LE_7\beta$ ideal model systems for chiral SFG, which is interface-selective. Computationally, we model $LK_7\beta$ and $LE_7\beta$ as pentamers consisting of antiparallel β-sheets surrounded by ~1350 water molecules at the vacuum-water interface. We also added 15 Cl^- ions to the $LK_7\beta$ system and 15 Na^+ ions to the $LE_7\beta$ system to neutralize the charged residues. We have previously published both experimental spectra and calculations of the OH stretch response for $LK_7\beta$,[20, 25] whereas $LE_7\beta$ is a new system. We chose $LE_7\beta$ as a second test

system because although it is similar to $LK_7\beta$, it has a different polar sidechain with an opposite charge. Our aim was to determine if we could detect any differences in the antiparallel β-sheet structure between the two systems experimentally and computationally. By so doing, we could generate experimental benchmarks to test our computational methods. We have achieved qualitative agreement with experiment for $LK_7\beta$ in the past.[19, 20, 25] For example, the phase of our calculated spectra of water OH flips when the (D-) enantiomer is used instead of the native (L-) enantiomer, and the OH peaks appear approximately at the correct frequencies. However, the quantitative agreement was limited due to our inability to model the NH response of the protein backbone, which dominates the total chiral SFG response. Here, we test our new NH map with the $LK_7\beta$ and $LE_7\beta$ systems and compare to our experimental results.

Infrared (linear) response

Although the NH response dominates chiral SFG, it is but a small component of the total vibrational response of these systems. Figure 5.4 shows the calculated IR spectrum of the $LK_7\beta$ system, for water alone (black) and for the whole system (NH and OH), including vibrational couplings between and among NH and OH groups (red). The two spectra are almost identical. Therefore, the NH stretch contributes very little to the IR signal. This is the central problem with using the NH group as a probe of local hydrogen bonding structure – it is obscured by the water response. Chiral SFG solves this problem because the water signal is heavily suppressed due to the small number of water molecules in the first hydration shell. In addition, water molecules are not intrinsically chiral, and their chirality is a higher-order phenomenon and therefore yields a more subtle effect.[25] In

contrast, the NH groups on the protein backbone are directly bonded to a chiral center, the α-carbon. Heterodyne detection further resolves NH and OH peaks from each other.

Comparing computational and experimental results

After calculating the IR response with a combination of Skinner's OH map and our new NH map, we moved on to chiral SFG. Figure 5.5 shows the results of our calculations for the heterodyne response of $LK_7\beta$ and $LE_7\beta$. The calculated spectrum for $LK_7\beta$ (Figure 5.5a, i) consists of a positive, main peak at ~3336 cm^{-1}, a negative peak at ~3575 cm^{-1}, a small but distinct negative feature at ~3457 cm^{-1}, and a positive shoulder at ~3387 cm^{-1}. The calculated $LE_7\beta$ spectrum (Figure 5.5a, ii) consists of a small low-frequency negative peak at ~3243 cm^{-1}, a shoulder at ~3303 cm^{-1}, a main peak at ~3346 cm^{-1}, a distinct but small negative feature at ~3467 cm^{-1}, and a high-frequency negative feature at ~3610 cm^{-1}. Comparison to heterodyne SFG experiments (Figure 5.5a, iii-iv) illustrates that we have achieved quite good agreement between experiment and theory for both systems, with both the main peaks and the subtler features. Almost every peak in the calculated spectra has a corresponding peak of the same phase and description in the experimental spectra. For example, the "main" positive peak in the $LK_7\beta$ spectra at 3336 cm^{-1} (Figure 5.5a, i) corresponds to a similar peak in the experiment at 3277 cm^{-1} (Figure 5.5a, iii). Similarly, the calculated spectra capture the subtler features, for example the small negative feature at 3457 cm^{-1} in the $LK_7\beta$ calculated spectra (Figure 5.5a, i) corresponds to a very similar feature at 3463 cm^{-1} in the experiment (Figure 5.5a, iii). The peak at ~3336 cm^{-1} in the $LK_7\beta$ spectrum is assigned to NH because it does not shift in the presence of $H_2{}^{18}O$.[25] The shoulder to the blue of the main peak in the $LK_7\beta$ spectrum and the small negative, kink-

like, feature in both the $LK_7\beta$ and $LE_7\beta$ spectra are due to complex interactions between a broad OH peak and the main NH peak, as revealed in Figure 5.5b, i-iv. The high-frequency negative peak is due mostly to the OH stretch in both systems. The shoulder and the smaller, kink-like feature reveal how crucial it is to calculate the NH response as well as the OH response even if one is not interested in the protein response – some OH-related peaks appear only due to an NH/OH interaction. It is important to note that while the sign of a chiral SFG peak may depend in some way on orientation, there is currently no clear way to infer orientation from the sign of a peak, as is possible with some forms of conventional (achiral) SFG (for example, the *ssp* polarization setting).

The only major features missing in the calculated spectra are the low-frequency experimental signals at around 3125 cm^{-1} in the $LK_7\beta$ spectrum (Figure 5.5a, iii) and 3100 and 3175 cm^{-1} in the $LE_7\beta$ spectrum (Figure 5.5a, iv). This overall feature is common to both systems in the experiment and is perturbed when $H_2{}^{18}O$ is used as solvent, indicating a contribution from water.[25] This feature may be due to a Fermi resonance or other interaction not considered in the computational model. However, it is unlikely to originate from the Fermi resonance between the OH stretch and the water bend, because the resonance usually leads to features with higher frequency (around 3200 cm^{-1}).[75] Indeed, we tried including this Fermi resonance phenomenologically as a vibrational coupling with the OH stretch following Skinner's recent work.[75] However, we could not computationally reproduce these low-frequency peaks (Figure A5.6). Our group continues to investigate the complex origins of these features.

The main peak in the theoretical and experimental heterodyne spectra (Figure 5.5), which is mostly due to the NH stretch, as revealed by comparison to the NH-only and OH-

only peaks in Figure 5.5b, shifts by 10−15 cm^{-1} between $LK_7\beta$ and $LE_7\beta$ in both the experimental and calculated spectra. In the experiment, it shifts from 3277 cm^{-1} for $LK_7\beta$ to 3290 cm^{-1} for $LE_7\beta$. In the computed spectra, the shift is from 3336 cm^{-1} to 3346 cm^{-1}. This suggests that at least some NH groups in $LE_7\beta$ may experience weaker or fewer hydrogen bonds than the NH groups of $LK_7\beta$. This is somewhat unexpected because both systems form stable antiparallel β-sheet structures as revealed by Fourier-transform infrared (FTIR) characterization (Figure A5.7) and MD simulation. However, we analyzed the average number of backbone H-bonds in the two simulated systems, and found that an $LK_7\beta$ pentamer typically contains ~20 "strong" H-bonds ($r < 3.4$ Å, angle > 135°) out of a possible ~24, whereas an $LE_7\beta$ pentamer contains only ~17 on average. There are roughly 24 possible H-bonds between the strands because there are roughly six possible H-bonds between each pair of strands and there are four inter-strand spaces in a pentamer. The average "strong" H-bond length in $LE_7\beta$ is also slightly longer (only by 0.005 Å, although this is statistically significant, $p < 0.001$ from a two-sided t-test). It seems that the lysine residues are a greater force for stabilizing the antiparallel b-sheet structure than are glutamate residues. Alternatively, the shift may be due to a more complicated interaction between the organizing forces of the nonpolar and polar residues. In any case, this finding illustrates how the NH stretch can be used as a sensitive probe of hydrogen bonding environments in proteins when its response is unmasked from the OH water stretches, even when the proteins are very closely related. The results demonstrated that it is possible to extract local structural information from the chiral SFG response of the protein NH stretch.

The chiral SFG spectra suggest that $LE_7\beta$ has a weaker antiparallel β-sheet structure than $LK_7\beta$, but the differences do not stop there. Figure 5.5b shows that in the calculated

OH-only spectra, the peaks due to the OH stretch are significantly *red*-shifted in the $LE_7\beta$ spectrum (Figure 5.5b, iv) compared to the $LK_7\beta$ spectrum (Figure 5.5b, iii) (from 3328 cm^{-1} to 3290 cm^{-1} for the positive peak, and from 3590 cm^{-1} to 3560 cm^{-1} for the negative peak). This suggests, perhaps paradoxically, that although $LE_7\beta$ has a weaker b-sheet structure, its interactions with and chirality transfer to water may be somewhat stronger than the analogous $LK_7\beta$ interactions. $LE_7\beta$ possibly gives up some of its inter-strand interactions in favor of interactions with water molecules. In any case, it is clear how, together, the OH and NH stretch responses reveal a detailed picture of the structural similarities and differences of the two related biological systems.

Theory matches experiments at the air-water interface

The heterodyne experiments were done with a quartz surface as a phase reference. This is somewhat of a different environment compared to the computational model, where the antiparallel β-sheet pentamers of $LK_7\beta$ and $LE_7\beta$ are at the vacuum-water interface. We wanted to see if a true air-water interface experiment would agree with our theoretical predictions as to the difference between $LK_7\beta$ and $LE_7\beta$. We obtained the homodyne chiral SFG spectra at the air-water interface with both $LK_7\beta$ and $LE_7\beta$ in the spectral region of 3000–3800 cm^{-1} and compared to the homodyne chiral SFG spectra predicted by the OH and NH maps (Figure 5.6). We find that the computational results match the experimental blue shift of the main peak for $LE_7\beta$ relative to $LK_7\beta$ (from 3340 to 3350 in the calculated spectra and from 3280 to 3305 in the experimental spectra) and correctly predict the lower-frequency peak to be by far the dominant feature. In the experiment the high-frequency response is very small, especially for $LK_7\beta$, although it is not zero (see inset, Figure 5.6).

In the calculated spectrum, the high-frequency features are larger relative to the main peak, although they are still much smaller than the main peak. Overall, the calculated and experimental results exhibit good qualitative agreement.

Although our map was designed to be robust and sensitive to complex biological environments, there are features that may affect its accuracy. The map was designed to work primarily with solvent-adjacent secondary structures. As a result, deeply buried structures, for which water is far away, may not perform as well as structures closer to the solvent. On the other hand, the electric field in the H-bonded configurations was dominated by the H-bonded carbonyl group, so buried NH groups interacting with carbonyl bonds could be expected to perform well using this map. Our observation of a disagreement between experiment and theory for the lower-frequency peaks in the homodyne spectra (Figure 5.6) as well as in the heterodyne spectra (Figure 5.5) suggests that the differences are due to a deficiency of the map rather than a difference in experimental and computational setups. The map was derived using a small, computationally tractable peptide surrounded by water molecules and implicit *water* solvent. It is possible that this environment was too aqueous and did not faithfully reproduce the environment experienced by secondary structures with more hydrophobic surroundings. Future applications of the NH map to larger proteins with true hydrophobic cores will test this assertion further. It may be necessary to develop an extended NH map using less polar implicit solvents to mimic hydrophobic cores. Another possibility is that the tetraalanine peptide used to develop the map does not adequately sample strongly hydrogen-bonded configurations as much as an actual protein backbone samples such configurations. Future refinements of the map may weight strongly bound configurations more in the training of

the model to capture low frequencies more faithfully. In any case, the agreement with experiment found in the present work represents a significant step forward in the modeling of chiral SFG. This map may also be extendable to model other kinds of vibrational spectroscopy, including higher-order techniques such as 2D-IR.

It should be noted that although the lower frequencies are less accurate than the higher frequencies using the current NH map, the map correctly predicts the qualitative effect of a strong H-bonding environment. Figure 5.7 shows the average NH stretch frequency predicted by the current map for each NH group in $LK_7\beta$. NH groups directly facing the solvent are shown as blue bars, and those facing inward are shown as yellow bars. It is clear, with a few exceptions, that the NH groups facing inward have a lower frequency than the NH groups facing the solvent. This is a result of a more stable H-bonding environment "inside" the structure compared to the exterior surface of the protein. The lowest average frequencies are found in the very center of the structure, as expected. Note that there are actually two different maps being used here (see Computational Methods), one for NH groups experiencing H-bonds with the protein and one for those *not* experiencing such H-bonds. Future studies using this map might reproduce Figure 5.7 for globular proteins and other structures with highly buried NH groups.

The role of vibrational couplings

One of the central advantages of being able to model the NH stretch response is the newfound ability to model interactions between the NH and OH stretches, here modeled as transition dipole vibrational couplings.[42] The data in Figure 5.8 shows the unexpected importance of these interactions for both systems we studied. The NH-only spectra for both

systems look very little like the NH-derived components of the full spectra and exhibit poor agreement with experimental spectra. In particular, the NH-only spectra have a large negative peak at ~3352 cm^{-1} (Figure 5.5b,i-ii and Figure 5.8) that is not present in the full spectra. Similarly, the OH-only spectra have a large negative peak at ~3590 cm^{-1} ($LK_7\beta$) and ~3560 cm^{-1} ($LE_7\beta$) (Figure 5.5b,iii-iv and Figure 5.8) that is severely attenuated and reshaped in the full spectra, even though there is no corresponding positive NH peak that could attenuate these peaks by simple addition. In fact, adding the NH-only and OH-only responses produces "total" spectra that are much farther from the experimental spectra than the modeled spectra that include vibrational couplings (Figure 5.8, uncoupled spectra). This suggests that vibrational couplings between NH and OH groups are a key component of the chiral SFG response, which is perhaps not surprising given that vibrational couplings play a formative role in the chiral SFG response of water by itself for fundamental, symmetry-related, reasons.[19] Although these findings do not negate the assignment of chiral SFG peaks as "NH" or "OH" (the couplings are much smaller in magnitude than the vibrational frequencies), they do suggest that an OH peak may contain the *influence* of an NH oscillation, and vice versa. This may not be surprising, as we previously found that the majority of the chiral SFG signals of water arises from the first hydration shell of the protein.[25] In general, the true picture of these condensed-phase systems is one of highly delocalized vibrational excitons including many different molecular species. These findings show that vibrational chiral SFG is more than a way to enumerate molecular environments experienced by OH and NH groups. It is a way to study and document the complex interactions between these species. Thus, interactions that would produce a slight

frequency shift (if that) in linear (e.g., IR) spectroscopy radically reshape chiral SFG spectra and are more visible as a result.

Treatment of hydrogen bonds of intermediate length

When we gathered tetraalanine configurations for the H-bonded and free NH maps, we used a H-bond cutoff of 3.2 Å between the heavy atom donor and acceptor. This cutoff was chosen to ensure that the H-bonded configurations accurately represented the typical situation of an NH group inside a β-sheet or α-helix. Free configurations were gathered exclusively from unfolded tetraalanine molecules for the same reason. In real systems, however, H-bonds within a secondary structure are occasionally slightly longer but in all other ways resemble the H-bonded tetraalanine configurations more than the free configurations. These situations were treated using the H-bonded map rather than the free map, and the H-bond distance cutoff used for the calculation of the SFG spectra was 3.4 Å to capture these edge cases. We found that assigning the edge cases to the H-bonded versus free map did not significantly change the chiral SFG spectrum, except by making the OH-related shoulder at 3380 cm^{-1} more pronounced in the 3.4 Å cutoff case (Figure A5.2). This similarity also shows that the H-bonded and free maps generally agree quite well on the edge cases, where an inter-residue H-bond is nearly broken. This analysis suggests that the two maps cover the entire range of likely environments around an NH group with some degree of overlap.

The NH group as a useful vibrational probe

Although a great deal of effort has been directed toward developing electrostatic maps for water and the amide modes,[6, 8, 11, 12, 14, 15, 71] few groups have pursued the NH

stretch.[49] The main reason for the smaller number of models developed for the NH stretch is that this mode is difficult to observe experimentally due to the overwhelming OH stretch of water. This problem is largely solved by chiral SFG, where the NH signal is prominent due to the chiral nature of proteins and the enhanced peak resolution afforded by heterodyne detection. In contrast, the water signal is relatively small because only the water molecules in the first hydration shell contribute significantly to the chiral SFG signal from OH groups.[25] Thus, chiral SFG is an ideal technique to study the environments experienced by NH groups in complex biological systems with the caveat that "NH" peaks may contain solvent contributions as well if they are not buried in a hydrophobic pocket. The presence of NH-like groups on amino acid sidechains is unlikely to cause problems for chiral SFG. The fact that we were able to reproduce almost all spectral features computationally for $LK_7\beta$ while completely ignoring the lysine NH_3^+ groups suggests that the high degree of chiral ordering of the backbone eclipses any signal from the flexible sidechains. We have also shown previously that the water molecules near the NH_3^+ groups produce minimal chiral SFG signal, possibly due to the ability of positive charge to disrupt water structure.[25]

Conclusion

The NH stretch map developed in this paper enables the simulation of full chiral SFG spectra of aqueous proteins in the 3000–4000 cm^{-1} range. The NH maps developed herein have distinct advantages over previously developed maps for modeling vibrational responses. First, they were designed to model the protein environment around the NH group using a tetraalanine peptide. Second, the design of these maps included the effect of solvent both explicitly (through the inclusion of nearby water molecules in the tetraalanine

configurations) and implicitly (through C-PCM implicit solvation in the DFT calculations of the NH frequencies), thereby driving the NH frequency down to the experimental range. Third, the map is based on a complete survey of the local environment through the electric field at 51 points around the NH, including points beyond the NH bond. Finally, the dependence of the transition dipole moment on the local environment is included by mapping the transition dipole to the local electric field. However, one limitation of the current map is that it was created with the AMBER ff14SB force field[69] and TIP4P-Ew[54] water and therefore must be used with these force fields to ensure accuracy.

Most importantly, this work has revealed the NH stretch to be a potent reservoir of information about the local environment of the amide group, on par with the more-studied amide I mode. Chiral SFG is ideally suited to probing this mode, as the problems of OH interference almost disappear with this technique. One aspect of the OH stretch response that does not disappear, however, is the coupling of this response to the NH response. We show that coupling between the OH and NH stretch modes is vital to the creation of the chiral SFG lineshape. This finding has implications for the ability of researchers to label peaks as "OH" or "NH" via fitting or even with $H_2{}^{18}O$ labeling. This work paves the way for the NH stretch to become a more useful, informative probe in the biological spectroscopy community. Although it can be probed easily by chiral SFG, the NH stretch can also serve as a probe for other vibrational spectroscopy methods such as 2D-IR, which can be modeled with the same type of NH electrostatic maps. Thus, the NH maps developed in this work have implications for a wide range of vibrational spectroscopies applied to diverse chemical and biological systems.

Correction to Chapter 5

We recently discovered a minor technical error in the mapping procedure used to compute the chiral sum frequency generation (chiral SFG) spectra in this paper. Due to an error in the input files, some of the electric fields used in the electrostatic frequency map for the NH stretch were obtained using an incorrect coordinate system, slightly altering the frequencies for some NH groups. The effect of the error on the simulated chiral SFG spectra was subtle. Corrected versions of Figures 5.4 – 5.8 are provided below, and the Appendix contains corrected figurrees. The agreement between the simulated and experimental spectra is essentially the same, and this correction does not impact any of the conclusions of the paper.

Summary of Appendix to Chapter 5

Computational and experimental methods details, the effect of H-bond distance cutoff on the chiral SFG spectrum, information about the Fourier smoothing process, effect of transition dipole placement on chiral SFG, infrared amide I spectra of $LK_7\beta$ and $LE_7\beta$, the effect of including Fermi resonance with the bending mode. Python scripts that show the correct ordering of the electric fields for the random forest regression models and binaries of the trained random forest regression models ready for use in Python can be found at https://github.com/dkonstan/NH_map.

Acknowledgements

The authors thank Matthew Tremblay and Dr. Pablo E. Videla at Yale University and Prof. Thomas la Cour Jansen at the University of Groningen for helpful

discussions. This work was supported by the National Institutes of Health Grant R35GM139449 (S.H.-S.) and the NSF Grant CHE-2108690 (E.C.Y.Y.). D.K. was supported by these grants. T.S. was supported by CHE-2108690 (to E.C.Y.Y.). E.A.P. was supported by the NIH (5T32GM008283-31) and a John C. Tully Chemistry Research Fellowship from Yale University. K.O. was supported by a Donald Crothers Chemistry Research Fellowship from Yale University.

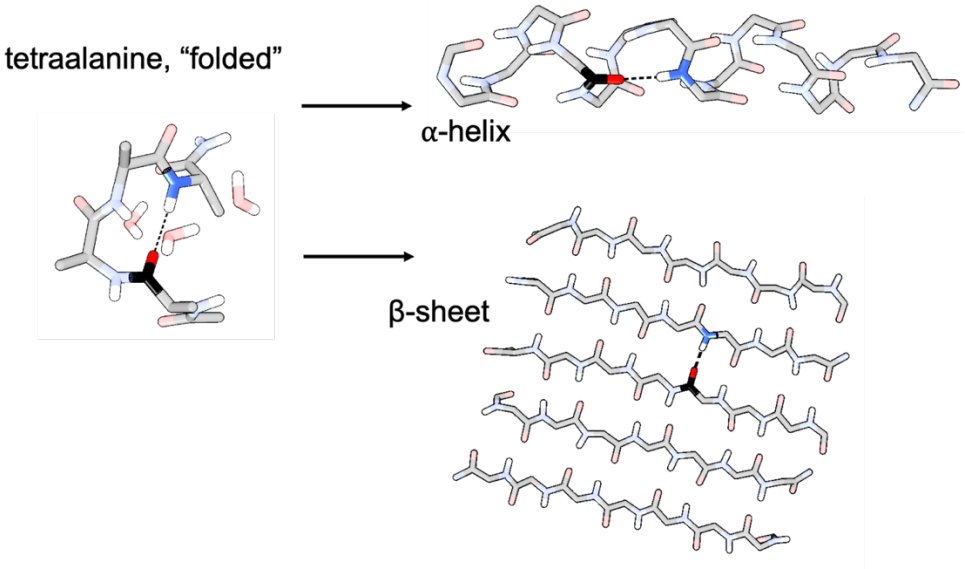

Figure 5.1. The mapping procedure relied on a small molecule (tetraalanine) mimic of the protein secondary structure environment, which is a H-bond between a backbone NH group and a backbone carbonyl group.

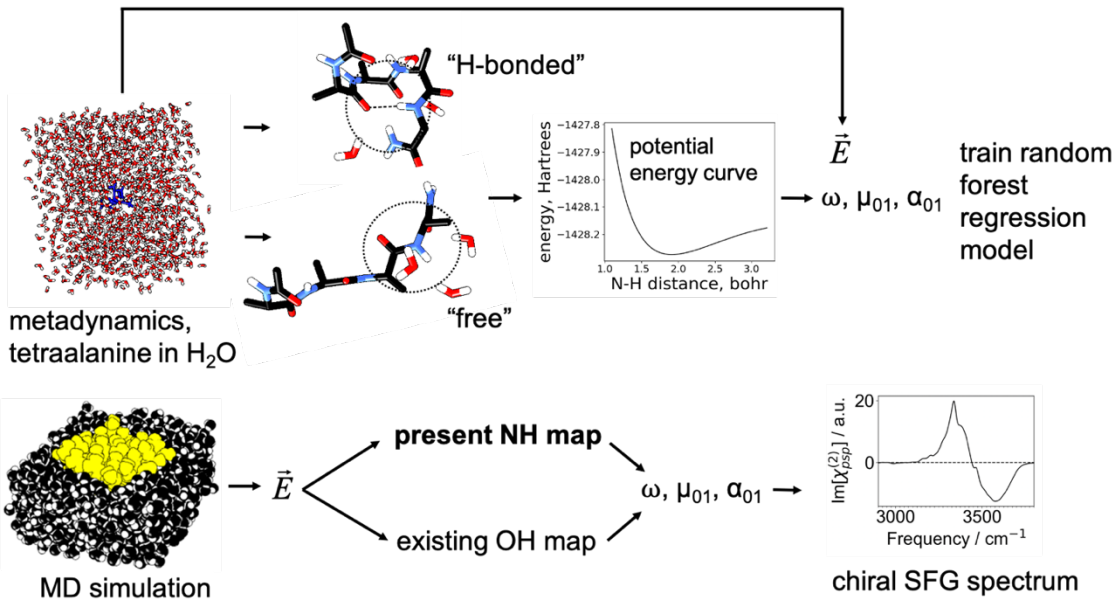

Figure 5.2. Top: H-bonded and free structures (400 each) were collected from a metadynamics simulation of tetraalanine in water, and each type of structure was used to produce a map for the corresponding type of environment. For a small cluster extracted from the system, the NH bond was stretched to map out a potential energy curve, and an anharmonic vibrational frequency was calculated. The transition dipole derivatives were estimated for a partially minimized structure of this cluster, and the transition dipole was approximated as the product of the dipole derivative and the transition position matrix element given in equation 1. The electric field was calculated near the NH group (Figure A5.1) from the MD simulation. The vibrational frequency and transition dipole were mapped to the electric field by a random forest regression model combined with a linear regression model. The transition polarizability was inferred partially from the transition position matrix element (equation 2). Bottom: The map was then used to estimate these quantities from an MD simulation of the protein (yellow) in aqueous solution (black), and a chiral SFG spectrum was generated.

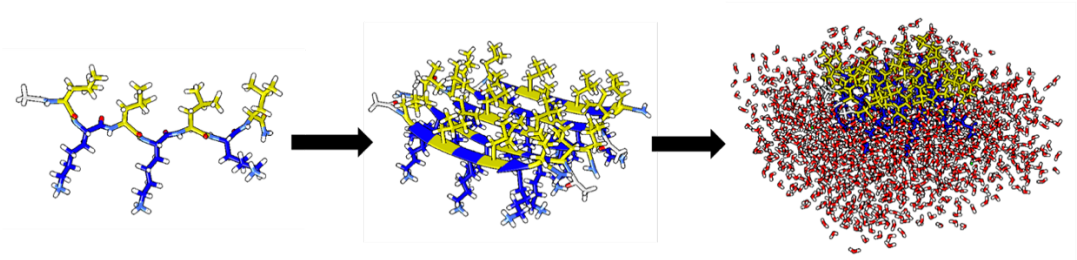

Figure 5.3. $LK_7\beta$ is a peptide (Ac-LKLKLKL-NH_2) that assembles into antiparallel β-sheets at the air-water interface. Leucine residues are in yellow, lysine residues are in blue, and water molecules in the computational model are red. The polar lysine residues face into the aqueous phase, while the hydrophobic leucine residues face into the vacuum. $LK_7\beta$ is modeled as a five-strand β-sheet with ~1350 water molecules around it. $LE_7\beta$ forms the same structure (Figure A5.7), but with negatively charged glutamate residues facing the solvent instead of positively charged lysines.

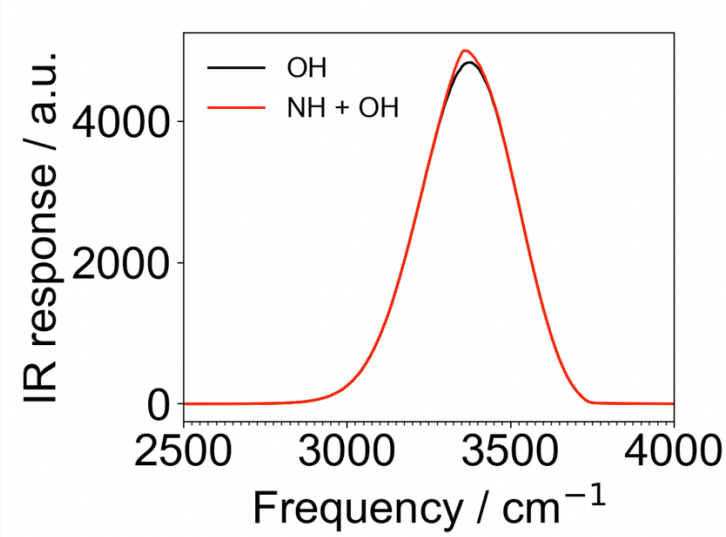

Figure 5.4 (corrected). Computational IR spectra of the $LK_7\beta$ system for water alone (OH, black) and for NH + OH (red), including all vibrational couplings between and within each type of stretch. The NH stretch contributes very little to the total IR spectrum. Spectra were obtained from 1 ns of sampling (100,000 configurations) using Skinner's time averaging approximation[71] to achieve a realistic peak width. All vibrational couplings were included. a.u. = arbitrary units.

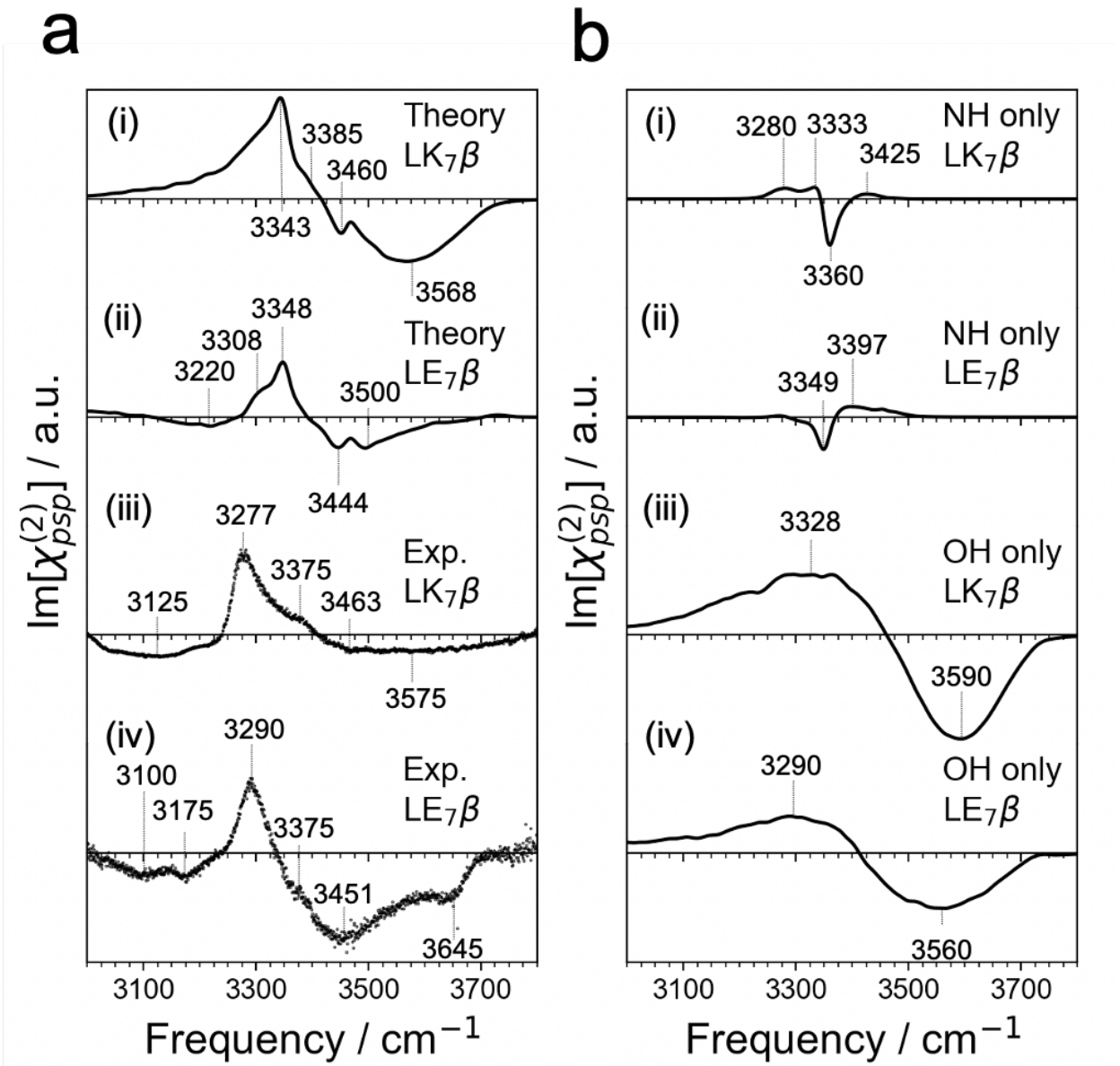

Figure 5.5 (corrected). (a) Comparison of calculated chiral SFG spectra with heterodyne experimental results for $LK_7\beta$ and $LE_7\beta$. Note the blue-shift of the main N−H peak in $LE_7\beta$ vs $LK_7\beta$ in both the experiment and the theory. All possible vibrational couplings are included. (b) NH-only and OH-only spectra, including only vibrational couplings within these groups. All peaks are labeled. Calculated spectra are averaged over 10,000,000 configurations spanning 100 ns. Data for $LK_7\beta$ in panel (a, iii) is reproduced with permission from ref 25. Copyright 2022 American Chemical Society. a.u. = arbitrary units.

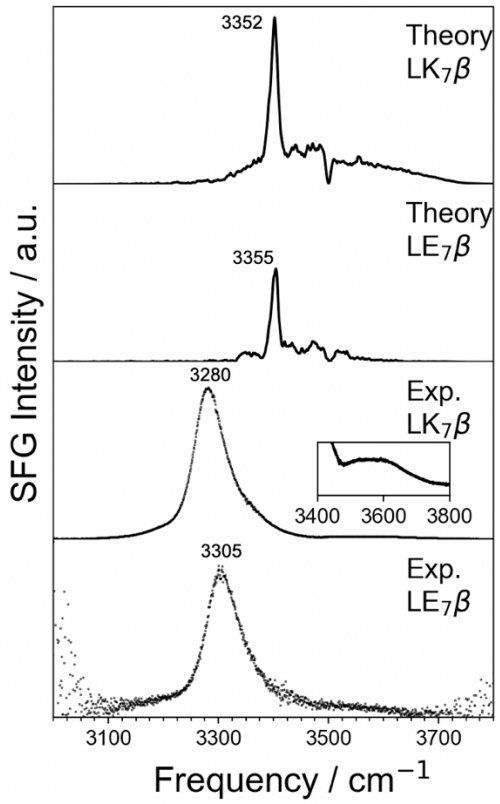

Figure 5.6 (corrected). Comparison of calculated homodyne spectra and experimental homodyne results for $LK_7\beta$ and $LE_7\beta$. Note the blue-shift in both theory and experiment between $LK_7\beta$ and $LE_7\beta$. The inset in the second plot from the bottom shows a detailed view of the high-frequency peak in the spectrum of $LK_7\beta$ that is relatively low in intensity but undergoes some fluctuation (see Figure A5.12). Key peaks are labeled. The computational spectra are not smoothed because the key peaks are very sharp. Calculated spectra are averaged over 10,000,000 configurations spanning 100 ns. a.u. = arbitrary units.

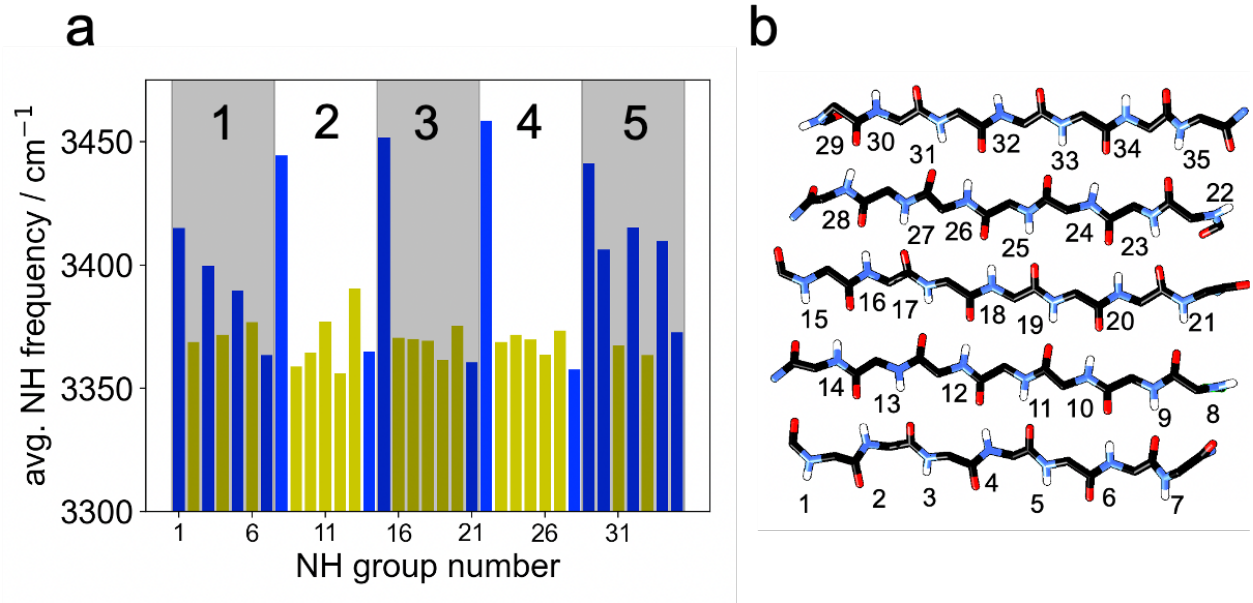

Figure 5.7 (corrected). (a) Average NH frequencies in LK$_7$β by NH group number computed using our map. The shading highlights the different β-strands, which are numbered at the top. The blue bars correspond to NH groups pointing into the solvent, and the yellow bars correspond to NH groups on the inside of the β-sheet. The average frequencies were based on analysis of 100,000 configurations spanning 100 ns. For the most part, the NH groups pointing toward the core of the LK$_7$β pentamer (yellow) experience stronger H-bonding interactions than those facing the solvent (blue), corresponding to lower vibrational frequencies. (b) Illustration of the numbering scheme.

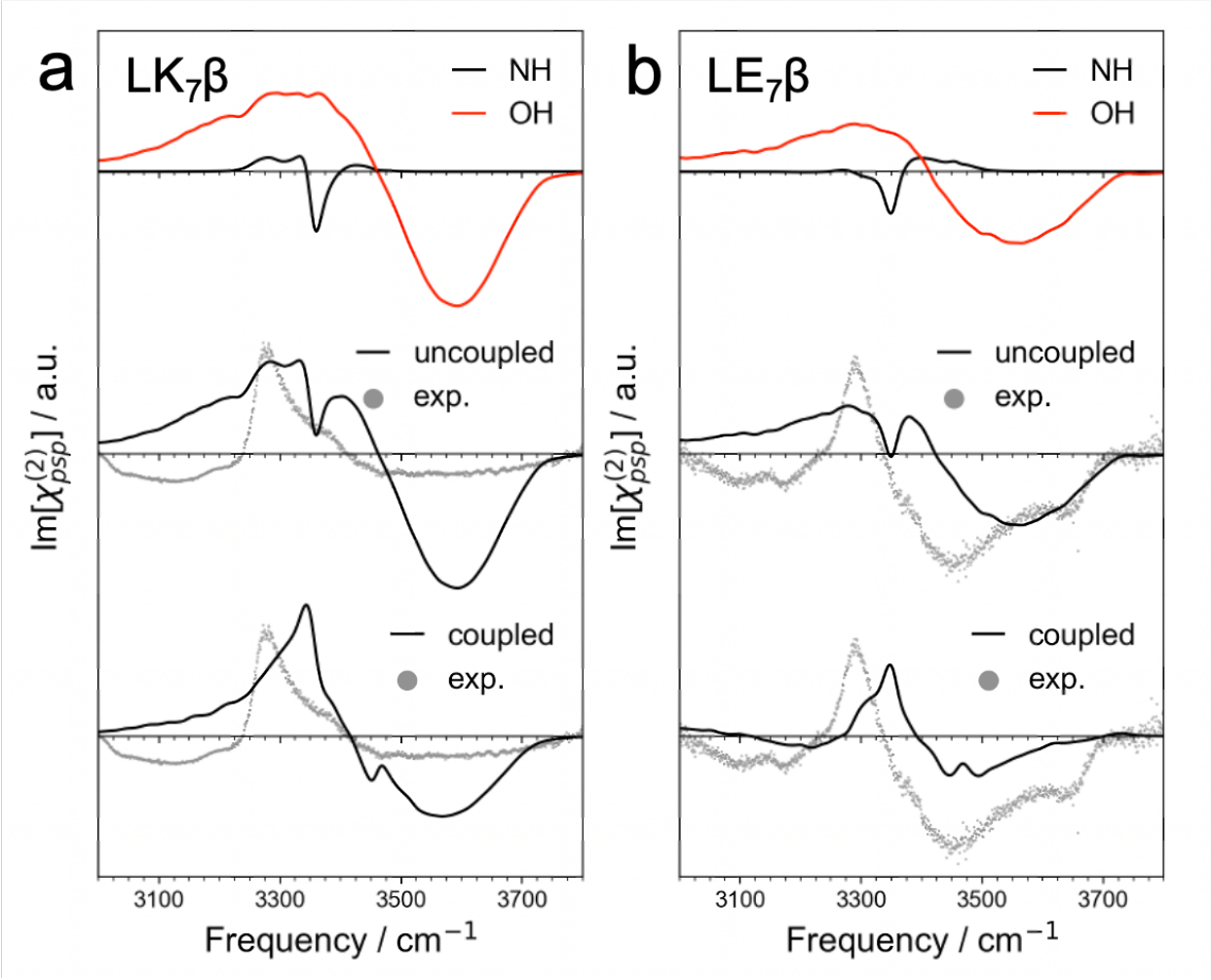

Figure 5.8 (corrected). Effect of vibrational coupling between NH and OH groups on the calculated chiral SFG spectra and comparison with experimental spectra. (a) Spectra for $LK_7\beta$ and (b) spectra for $LE_7\beta$. Experimental data for $LK_7\beta$ in (a) was obtained from ref 25. a.u. = arbitrary units.

References

1. 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.
2. Fu, L.; Liu, J.; Yan, E. C. Y. Chiral Sum Frequency Generation Spectroscopy for Characterizing Protein Secondary Structures at Interfaces. *J. Am. Chem. Soc.* **2011,** *133*, 8094-8097.
3. Rubtsov, I. V.; Wang, J.; Hochstrasser, R. M. Dual frequency 2D-IR of peptide amide-A and amide-I modes. *J. Chem. Phys.* **2003,** *118*, 7733-7736.
4. Rubtsov, I. V.; Wang, J.; Hochstrasser, R. M. Vibrational Coupling between Amide-I and Amide-A Modes Revealed by Femtosecond Two Color Infrared Spectroscopy. *J. Phys. Chem. A* **2003,** *107*, 3384-3396.
5. Jansen, T. l. C.; Knoester, J. Nonadiabatic Effects in the Two-Dimensional Infrared Spectra of Peptides: Application to Alanine Dipeptide. *J. Phys. Chem. B* **2006,** *110*, 22910-22916.
6. Corcelli, S. A.; Lawrence, C. P.; Skinner, J. L. Combined electronic structure/molecular dynamics approach for ultrafast infrared spectroscopy of dilute HOD in liquid H2O and D2O. *J. Chem. Phys.* **2004,** *120*, 8107-8117.
7. Schmidt, J. R.; Corcelli, S. A.; Skinner, J. L. Ultrafast vibrational spectroscopy of water and aqueous N-methylacetamide: Comparison of different electronic structure/molecular dynamics approaches. *J. Chem. Phys.* **2004,** *121*, 8887.
8. Auer, B. M.; Skinner, J. L. IR and Raman spectra of liquid water: Theory and interpretation. *J. Chem. Phys.* **2008,** *128*, 224511.
9. Auer, B. M.; Skinner, J. L. Vibrational Sum-Frequency Spectroscopy of the Water Liquid/Vapor Interface. *J. Phys. Chem. B* **2009,** *113*, 4125-4130.
10. Pieniazek, P. A.; Tainter, C. J.; Skinner, J. L. Interpretation of the water surface vibrational sum-frequency spectrum. *J. Chem. Phys.* **2011,** *135*, 044701.
11. Wang, L.; Middleton, C. T.; Zanni, M. T.; Skinner, J. L. Development and Validation of Transferable Amide I Vibrational Frequency Maps for Peptides. *J. Phys. Chem. B* **2011,** *115*, 3713-3724.
12. Carr, J. K.; Wang, L.; Roy, S.; Skinner, J. L. Theoretical Sum Frequency Generation Spectroscopy of Peptides. *J. Phys. Chem. B* **2015,** *119*, 8969-8983.
13. Hayashi, T.; la Cour Jansen, T.; Zhuang, W.; Mukamel, S. Collective Solvent Coordinates for the Infrared Spectrum of HOD in D2O Based on an ab Initio Electrostatic Map. *J. Phys. Chem. A* **2005,** *109*, 64-82.
14. Jansen, T. l. C.; Knoester, J. A transferable electrostatic map for solvation effects on amide I vibrations and its application to linear and two-dimensional spectroscopy. *J. Chem. Phys.* **2006,** *124*, 044502.
15. Jansen, T. l. C.; Dijkstra, A. G.; Watson, T. M.; Hirst, J. D.; Knoester, J. Modeling the amide I bands of small peptides. *J. Chem. Phys.* **2006,** *125*, 044312.
16. Moad, A. J.; Simpson, G. J. A Unified Treatment of Selection Rules and Symmetry Relations for Sum-Frequency and Second Harmonic Spectroscopies. *J. Phys. Chem. B* **2004,** *108*, 3548-3562.

17. Simpson, G. J. Molecular Origins of the Remarkable Chiral Sensitivity of Second-Order Nonlinear Optics. *ChemPhysChem* **2004,** *5*, 1301-1310.
18. Perets, E. A.; Olesen, K. B.; Yan, E. C. Y. Chiral Sum Frequency Generation Spectroscopy Detects Double-Helix DNA at Interfaces. *Langmuir* **2022,** *38*, 5765-5778.
19. Konstantinovsky, D.; Perets, E. A.; Yan, E. C. Y.; Hammes-Schiffer, S. Simulation of the Chiral Sum Frequency Generation Response of Supramolecular Structures Requires Vibrational Couplings. *J. Phys. Chem. B* **2021,** *125*, 12072–12081.
20. Perets, E. A.; Konstantinovsky, D.; Fu, L.; Chen, J.; Wang, H.-F.; Hammes-Schiffer, S.; Yan, E. C. Y. Mirror-image antiparallel β-sheets organize water molecules into superstructures of opposite chirality. *Proc. Natl. Acad. Sci. USA* **2020,** *117*, 32902-32909.
21. Perets, E. A.; Videla, P. E.; Yan, E. C. Y.; Batista, V. S. Chiral inversion of amino acids in antiparallel beta sheets at interfaces probed by vibrational sum frequency generation spectroscopy. *J. Phys. Chem. B* **2019,** *123*, 5769−5781.
22. Perets, E. A.; Yan, E. C. Y. Chiral Water Superstructures around Antiparallel β-Sheets Observed by Chiral Vibrational Sum Frequency Generation Spectroscopy. *J. Phys. Chem. Lett.* **2019,** *10*, 3395-3401.
23. Yan, E. C. Y.; Wang, Z.; Fu, L. Proteins at Interfaces Probed by Chiral Vibrational Sum Frequency Generation Spectroscopy. *J. Phys. Chem. B* **2015,** *119*, 2769-2785.
24. Fu, L.; Wang, Z.; Yan, E. C. Y. N-H Stretching Modes Around 3300 Wavenumber From Peptide Backbones Observed by Chiral Sum Frequency Generation Vibrational Spectroscopy. *Chirality* **2014,** *26*, 521-524.
25. Konstantinovsky, D.; Perets, E. A.; Santiago, T.; Velarde, L.; Hammes-Schiffer, S.; Yan, E. C. Y. Detecting the First Hydration Shell Structure around Biomolecules at Interfaces. *ACS Cent. Sci.* **2022,** *8*, 1404-1414.
26. McDermott, M. L.; Vanselous, H.; Corcelli, S. A.; Petersen, P. B. DNA's Chiral Spine of Hydration. *ACS Cent. Sci.* **2017,** *3*, 708-714.
27. Lee, S.-H.; Mirkin, N. G.; Krimm, S. A quantitative anharmonic analysis of the amide A band in α-helical poly (L-alanine). *Biopolymers* **1999,** *49*, 195-207.
28. Nagata, Y.; Mukamel, S. Vibrational Sum-Frequency Generation Spectroscopy at the Water/Lipid Interface: Molecular Dynamics Simulation Study. *J. Am. Chem. Soc.* **2010,** *132*, 6434-6442.
29. Nagata, Y.; Hsieh, C.-S.; Hasegawa, T.; Voll, J.; Backus, E. H. G.; Bonn, M. Water Bending Mode at the Water–Vapor Interface Probed by Sum-Frequency Generation Spectroscopy: A Combined Molecular Dynamics Simulation and Experimental Study. *J. Phys. Chem. Lett.* **2013,** *4*, 1872-1877.
30. Ohto, T.; Usui, K.; Hasegawa, T.; Bonn, M.; Nagata, Y. Toward ab initio molecular dynamics modeling for sum-frequency generation spectra; an efficient algorithm based on surface-specific velocity-velocity correlation function. *J. Chem. Phys.* **2015,** *143*, 124702.
31. Strunge, K.; Madzharova, F.; Jensen, F.; Weidner, T.; Nagata, Y. Theoretical Sum Frequency Generation Spectra of Protein Amide with Surface-Specific Velocity–Velocity Correlation Functions. *J. Phys. Chem. B* **2022,** *126*, 8571–8578.

32. Khatib, R.; Backus, E. H. G.; Bonn, M.; Perez-Haro, M.-J.; Gaigeot, M.-P.; Sulpizi, M. Water orientation and hydrogen-bond structure at the fluorite/water interface. *Sci. Reports* **2016,** *6*, 24287.
33. Khatib, R.; Hasegawa, T.; Sulpizi, M.; Backus, E. H. G.; Bonn, M.; Nagata, Y. Molecular Dynamics Simulations of SFG Librational Modes Spectra of Water at the Water–Air Interface. *J. Phys. Chem. C* **2016,** *120*, 18665-18673.
34. Khatib, R.; Sulpizi, M. Sum Frequency Generation Spectra from Velocity–Velocity Correlation Functions. *J. Phys. Chem. Lett.* **2017,** *8*, 1310-1314.
35. Morita, A.; Hynes, J. T. A Theoretical Analysis of the Sum Frequency Generation Spectrum of the Water Surface. II. Time-Dependent Approach. *J. Phys. Chem. B* **2002,** *106*, 673-685.
36. Nihonyanagi, S.; Ishiyama, T.; Lee, T.-k.; Yamaguchi, S.; Bonn, M.; Morita, A.; Tahara, T. Unified Molecular View of the Air/Water Interface Based on Experimental and Theoretical χ(2) Spectra of an Isotopically Diluted Water Surface. *J. Am. Chem. Soc.* **2011,** *133*, 16875-16880.
37. Ishiyama, T.; Takahashi, H.; Morita, A. Molecular dynamics simulations of surface-specific bonding of the hydrogen network of water: A solution to the low sum-frequency spectra. *Phys. Rev. B* **2012,** *86*, 035408.
38. Ishiyama, T.; Morita, A. Nuclear Quantum Effect on the χ(2) Band Shape of Vibrational Sum Frequency Generation Spectra of Normal and Deuterated Water Surfaces. *J. Phys. Chem. Lett.* **2019,** *10*, 5070-5075.
39. Gruenbaum, S. M.; Tainter, C. J.; Shi, L.; Ni, Y.; Skinner, J. L. Robustness of Frequency, Transition Dipole, and Coupling Maps for Water Vibrational Spectroscopy. *J. Chem. Theory Comput.* **2013,** *9*, 3109-3117.
40. Pieniazek, P. A.; Tainter, C. J.; Skinner, J. L. Surface of Liquid Water: Three-Body Interactions and Vibrational Sum-Frequency Spectroscopy. *J. Am. Chem. Soc.* **2011,** *133*, 10360-10363.
41. Auer, B. M.; Skinner, J. L. Dynamical effects in line shapes for coupled chromophores: Time-averaging approximation. *J. Chem. Phys.* **2007,** *127*, 104105.
42. Auer, B. M.; Skinner, J. L. IR and Raman spectra of liquid water: Theory and interpretation. *J. Chem. Phys.* **2008,** *128*, 224511.
43. Corcelli, S. A.; Skinner, J. L. Infrared and Raman Line Shapes of Dilute HOD in Liquid H2O and D2O from 10 to 90 °C. *J. Phys. Chem. A* **2005,** *109*, 6154-6165.
44. Phillips, D. C.; York, R. L.; Mermut, O.; McCrea, K. R.; Ward, R. S.; Somorjai, G. A. Side Chain, Chain Length, and Sequence Effects on Amphiphilic Peptide Adsorption at Hydrophobic and Hydrophilic Surfaces Studied by Sum-Frequency Generation Vibrational Spectroscopy and Quartz Crystal Microbalance. *J. Phys. Chem. C* **2007,** *111*, 255-261.
45. Mermut, O.; Phillips, D. C.; York, R. L.; McCrea, K. R.; Ward, R. S.; Somorjai, G. A. In Situ Adsorption Studies of a 14-Amino Acid Leucine-Lysine Peptide onto Hydrophobic Polystyrene and Hydrophilic Silica Surfaces Using Quartz Crystal Microbalance, Atomic Force Microscopy, and Sum Frequency Generation Vibrational Spectroscopy. *J. Am. Chem. Soc.* **2006,** *128*, 3598-3607.
46. Wang, Z.; Fu, L.; Yan, E. C. Y. C–H Stretch for Probing Kinetics of Self-Assembly into Macromolecular Chiral Structures at Interfaces by Chiral Sum Frequency Generation Spectroscopy. *Langmuir* **2013,** *29*, 4077-4083.

47. Fu, L.; Chen, S.-L.; Wang, H.-F. Validation of Spectra and Phase in Sub-1 cm–1 Resolution Sum-Frequency Generation Vibrational Spectroscopy through Internal Heterodyne Phase-Resolved Measurement. *J. Phys. Chem. B* **2016,** *120*, 1579-1589.
48. York, R. L.; Browne, W. K.; Geissler, P. L.; Somorjai, G. A. Peptides Adsorbed on Hydrophobic Surfaces—A Sum Frequency Generation Vibrational Spectroscopy and Modeling Study. *Isr. J. Chem.* **2007,** *47*, 51-58.
49. Hayashi, T.; Zhuang, W.; Mukamel, S. Electrostatic DFT Map for the Complete Vibrational Amide Band of NMA. *J. Phys. Chem. A* **2005,** *109*, 9747-9759.
50. Hayashi, T.; Mukamel, S. Vibrational−Exciton Couplings for the Amide I, II, III, and A Modes of Peptides. *J. Phys. Chem. B* **2007,** *111*, 11032-11046.
51. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983,** *79*, 926-935.
52. Wang, J.; Hochstrasser, R. M. Anharmonicity of Amide Modes. *J. Phys. Chem. B* **2006,** *110*, 3798-3807.
53. Barducci, A.; Bussi, G.; Parrinello, M. Well-Tempered Metadynamics: A Smoothly Converging and Tunable Free-Energy Method. *Phys. Rev. Lett.* **2008,** *100*, 020603.
54. Horn, H. W.; Swope, W. C.; Pitera, J. W.; Madura, J. D.; Dick, T. J.; Hura, G. L.; Head-Gordon, T. Development of an improved four-site water model for biomolecular simulations: TIP4P-Ew. *J. Chem. Phys.* **2004,** *120*, 9665-9678.
55. Becke, A. D. Density-functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.* **1993,** *98*, 5648-5652.
56. Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J. Ab Initio Calculation of Vibrational Absorption and Circular Dichroism Spectra Using Density Functional Force Fields. *J. Phys. Chem.* **1994,** *98*, 11623-11627.
57. Lee, C.; Yang, W.; Parr, R. G. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B* **1988,** *37*, 785-789.
58. McLean, A. D.; Chandler, G. S. Contracted Gaussian basis sets for molecular calculations. I. Second row atoms, Z=11–18. *J. Chem. Phys.* **1980,** *72*, 5639-5648.
59. Krishnan, R.; Binkley, J. S.; Seeger, R.; Pople, J. A. Self-consistent molecular orbital methods. XX. A basis set for correlated wave functions. *J. Chem. Phys.* **1980,** *72*, 650-654.
60. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Petersson, G. A.; Nakatsuji, H., et al. *Gaussian 16 Rev. C.01*, Wallingford, CT, 2016.
61. Klamt, A.; Schüürmann, G. COSMO: a new approach to dielectric screening in solvents with explicit expressions for the screening energy and its gradient. *J. Chem. Soc., Perkin trans.* **1993,** 799-805.
62. Andzelm, J.; Kölmel, C.; Klamt, A. Incorporation of solvent effects into density functional calculations of molecular energies and geometries. *J. Chem. Phys.* **1995,** *103*, 9312-9320.
63. Barone, V.; Cossi, M. Quantum Calculation of Molecular Energies and Energy Gradients in Solution by a Conductor Solvent Model. *J. Phys. Chem. A* **1998,** *102*, 1995-2001.

64. Cossi, M.; Rega, N.; Scalmani, G.; Barone, V. Energies, structures, and electronic properties of molecules in solution with the C-PCM solvation model. *J. Comput. Chem.* **2003,** *24*, 669-681.
65. Marston, C. C.; Balint-Kurti, G. G. The Fourier grid Hamiltonian method for bound state eigenvalues and eigenfunctions. *J. Chem. Phys.* **1989,** *91*, 3571-3576.
66. Mayne, L. C.; Hudson, B. Resonance Raman spectroscopy of N-methylacetamide: overtones and combinations of the carbon-nitrogen stretch (amide II') and effect of solvation on the carbon-oxygen double-bond stretch (amide I) intensity. *J. Phys. Chem.* **1991,** *95*, 2962-2967.
67. Joung, I. S.; Cheatham, T. E., III Determination of Alkali and Halide Monovalent Ion Parameters for Use in Explicitly Solvated Biomolecular Simulations. *J. Phys. Chem. B* **2008,** *112*, 9020-9041.
68. Joung, I. S.; Cheatham, T. E., III Molecular Dynamics Simulations of the Dynamic and Energetic Properties of Alkali and Halide Ions Using Water-Model-Specific Ion Parameters. *J. Phys. Chem. B* **2009,** *113*, 13279-13290.
69. Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theory Comput.* **2015,** *11*, 3696-3713.
70. Belch, A. C.; Rice, S. A. The OH stretching spectrum of liquid water: A random network model interpretation. *J. Chem. Phys.* **1983,** *78*, 4817-4823.
71. Auer, B. M.; Skinner, J. L. Dynamical effects in line shapes for coupled chromophores: Time-averaging approximation. *J. Chem. Phys.* **2007,** *127*, 104105.
72. 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.
73. DeGrado, W. F.; Lear, J. D. Induction of peptide conformation at apolar water interfaces. 1. A study with model peptides of defined hydrophobic periodicity. *J. Am. Chem. Soc.* **1985,** *107*, 7684-7689.
74. Weidner, T.; Breen, N. F.; Drobny, G. P.; Castner, D. G. Amide or Amine: Determining the Origin of the 3300 cm−1 NH Mode in Protein SFG Spectra Using 15N Isotope Labels. *J. Phys. Chem. B* **2009,** *113*, 15423-15426.
75. Kananenka, A. A.; Skinner, J. L. Fermi resonance in OH-stretch vibrational spectroscopy of liquid water and the water hexamer. *J. Chem. Phys.* **2018,** *148*, 244107.

Appendix to Chapter 5

Methods details

MD simulation

Five strands with the sequence Ac-LKLKLKL-NH_2 were arranged into an approximate antiparallel β-sheet pattern using VMD tcl scripts. The $LK_7β$ system was solvated with TIP4P-Ew water[1] using OpenMM tools,[2] ensuring that there was at least 10 Å of solvent on each side of the antiparallel β-sheet. The protein was solvated with the leucine residues facing up and the lysine residues facing down. The top half of the solvent was then removed without changing the *z*-dimension of the box to create a vacuum-water interface. At this point, the system was neutralized with 15 chloride ions. The preparation for $LE_7β$ was the same except 15 sodium ions were used instead of chloride to neutralize the system.

The equilibration protocol started with a minimization of the solvent using steepest descent until convergence (within 10 kJ/mol), followed by 500 ns of solvent dynamics with restrained protein. After this initial step, the hydrogen atoms were energy minimized with frozen heavy atoms. Then the sidechains were minimized with a frozen backbone in three steps with gradually reduced restraints on the backbone (100 kcal/mol·$Å^{-2}$, 50 kcal/mol·$Å^{-2}$, and 10 kcal/mol·$Å^{-2}$). Following these steps, the entire system was minimized without restraints. After the minimization, the system was heated in the NVT ensemble from 0 to 298 K over ~350 ps. Finally, the system was equilibrated for 6 ns in the NVT ensemble. No NPT equilibration was performed to preserve the interface. The production run was performed in the NVT ensemble for 100 ns, saving configurations every 10 fs.

The AMBER ff14SB force field[3] was used for the protein, the TIP4P-Ew model[1] was used for the water, and the standard parameters associated with the TIP4P-Ew model were used for the ions.[4, 5] The particle-mesh Ewald method was used for long-range electrostatic interactions.[6] Van der Waals interactions were cut off at 14 Å with a switching function that began working at 12 Å. A Langevin integrator was used with a coupling constant of 1/ps and a timestep of 1 fs at a temperature of 298 K. Since our spectrum modeling code is aware of periodic boundary conditions, the trajectories were not wrapped in any way, although each water molecule was confirmed to be intact. OpenMM tools[2] were used for the equilibration and the production trajectories.

Fourier smoothing of spectra

Although the NH stretch signals converge relatively quickly, OH stretch chiral SFG signals are relatively weak and do not converge to a smooth spectrum even when 10,000,000 configurations are considered, especially when intermolecular couplings are included. To improve presentation, spectra were smoothed by Fourier transforming the frequency-dependent complex SFG signal into the time domain, applying a decay exponential function with a decay constant of 500 fs, and transforming back into the frequency domain. This method resembles smoothing performed by Nagata,[7] Sulpizi,[8] and others on the time correlation functions they used to calculate SFG spectra from MD trajectories. The decay time was chosen to maximize ease of reading the spectrum without removing any significant features (Figure A5.4).

Applying the NH map to (D-) enantiomeric proteins

The **x**, **y**, and **z** directions used to sample the local electric field (Figure A5.1) are defined in terms of the (L-) amino acid structure. In particular, the **z** direction is defined by a right-handed cross product of the **x** and **y** unit vectors. Correspondingly, for the (D-) amino acid structures, the cross product must be left-handed to sample the equivalent local structure. This was accomplished by flipping the sign of the **z** component of the electric fields for each sampling point for the (D-) amino acid structures. This adjustment is required for the chiral SFG spectrum of the (D-) structures to be equivalent to (i.e., a mirror image of) the spectrum of the (L-) structures. This problem does not exist for the Skinner OH map because water is achiral and the electric field is defined along each OH bond.

Intermolecular couplings

Intermolecular couplings between OH groups, between NH groups, and between NH and OH groups were approximated as transition dipole couplings. The effective location of the OH and NH dipole (for the purposes of calculating the dipole-dipole distance) was 0.67 Å along the bond starting from the O or N toward the H.[9] This distance was used by Skinner with TIP4P-like water based on a study of water dimers and was found to provide the best fit to *ab initio* data. Since there is no equivalent simple system for which to calculate couplings for NH, the strategy for finding the optimal dipole location is not clearly defined. For this reason, we decided to simply use the same distance as Skinner did for OH in our calculations for both OH and NH. We found that 0.67 Å is a reasonable distance to use based on the resulting phase of the NH peaks compared to experiment. In particular, the lower-frequency peaks are positive for this distance but negative for certain

other distances (Figure A5.5). The higher-frequency peaks are only mildly affected by the dipole location. Note that Figure A5.5 shows the NH-only spectrum, which has an extra negative peak in the middle of the spectral range that disappears when NH/OH couplings and the OH contributions are considered. Future work may involve a more rigorous optimization of the transition dipole location.

Vibrational lifetime

We used 1.3 ps as the vibrational lifetime in our calculations for both the NH and OH groups, following Skinner's value for OH.[10, 11] This decision has no significant implications because the peak broadening produced by reasonable vibrational lifetimes is insignificant compared to the broadening caused by environmental heterogeneity.

Metadynamics

A tetraalanine peptide (Ac-AAAA-NH_2) was generated using the AMBER tool tleap.[12] This structure was solvated in TIP4P-Ew[1] water ensuring that there was at least 10 Å of solvent on each side of the tetraalanine peptide (~1000 water molecules total). The equilibration procedure was the same as that described above, except that an additional 5 ns NPT step was added because there was no interface for this system. A collective variable was defined as the distance between the amide nitrogen of the fourth alanine and the carbonyl oxygen of the first alanine. The NH group of the fourth alanine was used in the derivation of the map. Metadynamics was performed using an OpenMM[2] biasVariable object with a bias width of 0.82 Å, as determined from the standard deviation of the N—O distance in a separate, standard MD trajectory. The initial height of the Gaussians added to

the potential was 1 kJ/mol, and the frequency of adding Gaussians was every 100 timesteps. The bias factor was set to 10, which corresponds approximately to sampling the collective variable as if the temperature were 10 times higher. The metadynamics simulation was performed at 300 K for 500 ns using the Metadynamics tool of OpenMM, but only a fraction of this trajectory was needed to extract configurations for the map. The configurations used for the mapping procedure were obtained from frames at least 100 ps apart.

Details of chiral SFG spectrum calculation

The spectra were calculated as reported previously using the method developed by Skinner and coworkers.[9, 10, 13-15] The inhomogeneous limit approximation was used for all spectra. One effect of this approximation is to slightly widen peaks due to neglecting dynamical effects such as motional narrowing.[16] The inhomogeneous limit approximation was used because for biological systems of this size, using the polarizability-dipole time correlation function is too expensive. The exciton Hamiltonian was constructed as reported previously, except that NH stretch frequencies and couplings were used in addition to OH stretch frequencies and couplings. The NH/OH couplings were calculated with the transition dipole coupling scheme using the transition dipoles of the NH and OH stretches calculated from the maps. The details of the NH map are described in Chapter 5, and the details of the OH map can be found in the Skinner group's foundational papers.[14, 15, 17, 18] The spectral simulation code was written in Python, with performance-intensive parts in optimized Cython. NumPy[19] functions were used throughout, except for matrix diagonalization. The bottleneck step for large systems (i.e., the NH/OH spectra) is

diagonalization of the Hamiltonian. This step was accomplished with the Google jax library[20] eigh routine. This calculation was performed in a massively parallel manner by assigning a small number of simulation configurations to each of 1,000−10,000 CPUs.

Description of the random forest regression model

The random forest regression model provided by the scikit-learn Python library was used to develop the NH map. The default parameters were used, as indicated in the documentation at https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestRegressor.html. The fit qualities (r^2 values) were 0.95 for the H-bonded frequency map, 0.96 for the H-bonded transition dipole derivative map, 0.94 for the free frequency map, and 0.97 for the free transition dipole derivative map. The NH maps are provided in the pickled binary files regressionHbonded_frequency.pkl, regressionHbonded_dipolederivative.pkl regressionFree_frequency.pkl, and regressionFree_dipolederivative.pkl. The map assembly script is also provided.

Fitting x_{01} (see equation 1) vs frequency (linear regression)

H-bonded: intercept: 0.1975, slope: -1.8228, r^2: 0.9978.

Free: intercept: 0.19799, slope: -1.8361, r^2: 0.9985.

Experimental procedure of Fourier-transform infrared (FTIR) spectroscopy

Solutions of $LK_7\beta$ (1 mM) and $LE_7\beta$ (0.5 mM) were made by dissolving $LK_7\beta$ and $LE_7\beta$ in pure water. Then, 2 μL of the solutions were added to glass slides. The solutions were dried by gently purging with nitrogen gas and kept in a desiccator for 50 minutes. The FTIR spectra were taken by an attenuated total reflectance-FTIR (ATR-FTIR) spectrometer (FTIR Shimazu IRTracer-100) with the protein samples on the glass slides facing down on the sample stage with 256 scans, resolution of 4 cm^{-1}, and scanning range of 400-4000 cm^{-1}.

SFG experimental details

Broadband SFG Spectrometer

The broadband SFG spectrometer contained a Ti:Sapphire regenerative amplifier (Astrella F, Coherent, CA). The amplifier was seeded by a 40 fs 575 mW Ti:Sapphire oscillator (Vitara S, Coherent, CA) that was pumped by an optically pumped semiconductor laser (Verdi-G, Coherent, CA). A Nd:YLF laser (Revolution, Coherent, CA) with an output of ~33 W pumps the amplifier with a final output of 9 W with a repetition rate of 5 kHz and 100 fs pulse width. The output was split by a beam splitter with 2.3 W passing through a pulse shaper to produce a narrowband 800-nm visible pulse. The remaining 6.7 W pumped an optical parametric amplifier (TOPAS-PRIME, Coherent, CA) which was coupled to a noncollinear difference frequency generator (NDFG, Coherent, CA) which generated broadband infrared pulses tunable at 2.4 – 11 μm (900 - 4000 cm^{-1}). The 800-nm visible and the mid infrared beams are then spatially and temporally overlapped at our

sample surface to generate the SFG signal, which is collected by a collimating lens and focused into a spectrograph and ultimately detected by a cryogenically cooled CCD camera. The infrared source is inherently p-polarized upon exiting the NDFG, while the 800-nm visible and SFG polarizations are selected using a waveplate and polarizer combination. A complete description of the sample stage and detection optics is provided in detail by Ma et al.[21]

Phase Resolved Chiral SFG Measurements

The chiral SFG spectra of hydrated films of peptides were obtained following the procedure reported in our previous studies, using an internal heterodyne approach where the sample substrate is also the phase reference.[22, 23] The sample was probed such that the quartz substrate was positioned with an azimuthal angle at f = 90°(+y-axis) and then f =270°(–y-axis) with respect to the X-axis of the laboratory frame, where the Z axis is perpendicular to the optical table and the XZ plane is the incident plane. To achieve chiral specificity of the SFG probe, the spectra were recorded using the combination of p-polarized detected sum frequency, s-polarized visible probe, and p-polarized infrared-probe (i.e., psp-polarization). Spectra were an average of 10-12 spectra (2 minutes each) for $LK_7\beta$ and 20 spectra (2.5 minutes each) for $LE_7\beta$. The cosmic rays in the spectra were then removed. The intensity of the averaged spectra (I_f) was normalized to the intensity spectrum of clean quartz recorded at f = 90°(+y-axis), i.e., $I_{quartz,\ 90}$. The heterodyne spectra were then obtained by[24]

$$\mathrm{Im}[\chi^{(2)}] = \frac{I_{\phi=90} - I_{\phi=270}}{4I_{quartz,90}} \quad (1)$$

Air-water interface measurements

For the measurements at the air-water interface, 3.9 mL of 10 mM pH = 7.4 phosphate buffer was added to a Teflon beaker. Then, 100 μL of a 1 mM stock solution of either $LK_7\beta$ or $LE_7\beta$ was added dropwise to the interface using a Hamilton syringe to give a final peptide concentration of 25 μM. After equilibration (~2 hours), spectra were acquired in two windows and stitched together to cover the spectral region of 3000 – 3800 cm^{-1} (see Figures A5.10 and S5.11). Spectra of $LE_7\beta$ were acquired for 20 min with a width of the narrowband visible of ~26.7 cm^{-1}, while $LK_7\beta$ spectra were acquired for 5 min with a width of ~20.3 cm^{-1}. The spectra were processed following the procedure described in detail by Ma et al.[21] For each spectral window, five spectra and five background measurements were averaged. The background was subtracted from each spectrum. The resulting spectra were then normalized to the non-resonant SFG spectrum of GaAs in the corresponding window.

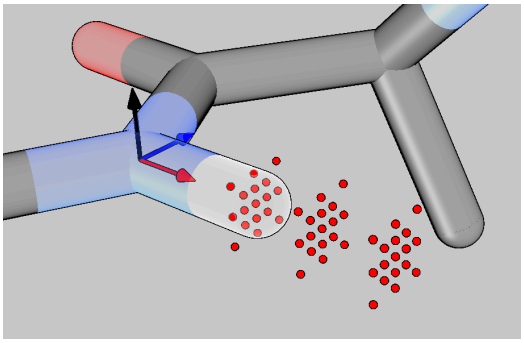

Figure A5.1. The electric field was sampled at 51 (3×17) positions around the NH bond, in three perpendicular directions for each position, for a total of 153 electric field components. The amide nitrogen atom is in blue, the hydrogen atom is in white, the carbon atoms are in black, and the oxygen atom is in red. The **x**-direction (red arrow) is along the NH bond. The **y**-direction (blue arrow) is toward the amide carbon but perpendicular to **x**. The **z**-direction (black arrow) is perpendicular to the peptide bond plane. The script eField_ordering.py gives the ordering of positions for the electric field vector that is input to the random forest regression models. This script also contains the precise locations of each point relative to the NH group.

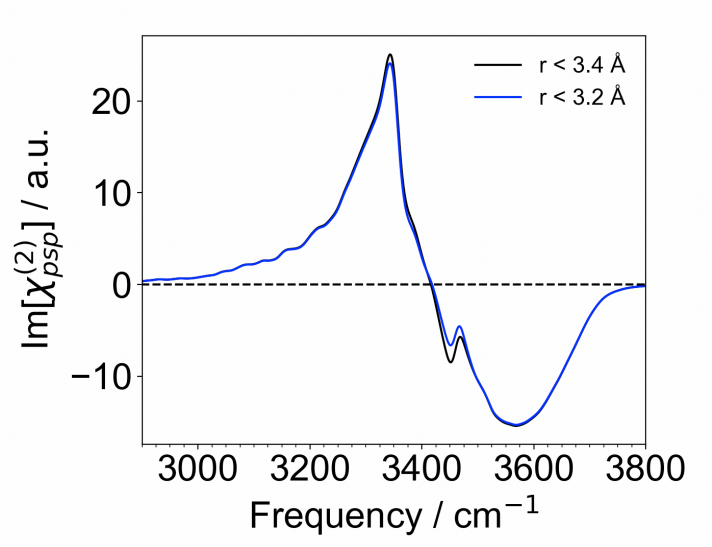

Figure A5.2. Comparison of the chiral SFG spectrum for $LK_7\beta$ using two different definitions of an H-bond, with a donor−acceptor distance cutoff of 3.4 Å (black line) or 3.2 Å (blue line). The effect of hydrogen bond distance cutoff on the chiral SFG spectrum of $LK_7\beta$ is modest. a.u. = arbitrary units.

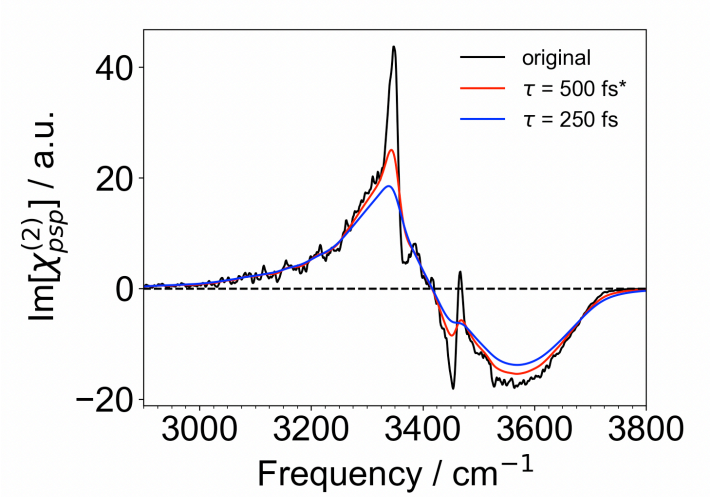

Figure A5.3. Influence of the exponential decay time for the Fourier smoothing on the chiral SFG spectrum for LK$_7$β. The decay time τ = 500 fs was chosen for presenting the heterodyne data in this work because it is a reasonable compromise between readability and reproduction of the original data. a.u. = arbitrary units.

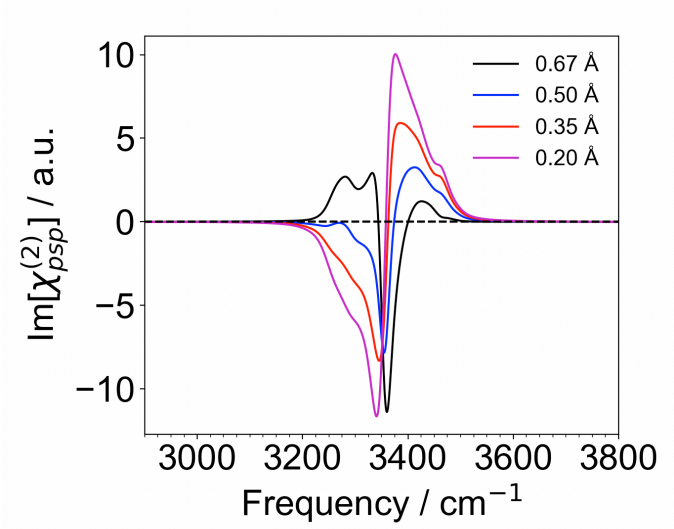

Figure A5.4. The effect of transition dipole placement (starting at the nitrogen atom and moving toward the H along the NH bond) on the NH-only chiral SFG spectrum of $LK_7\beta$. Note that the large negative peak is absent in the NH/OH coupled spectrum. Using the original value of 0.67 Å results in lower-frequency peaks with a phase that matches experiment. The higher-frequency peaks are less affected by the placement of the transition dipole. a.u. = arbitrary units.

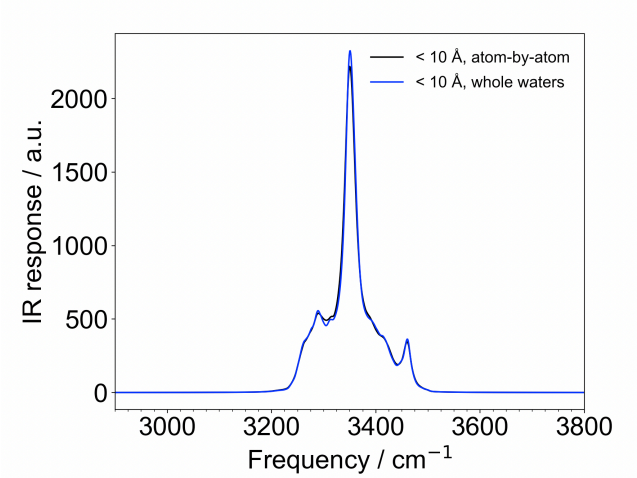

Figure A5.5. IR spectra of only the NH stretch in $LK_7\beta$ (1 ns, Δt 10 fs) using an individual atom distance cutoff (10 Å) for the electric field vs. including the entire water molecule if any atom is within this distance relative to the sampling point (see Figure A5.1). The IR spectrum is nearly identical when whole water molecules within 10 Å are included or when an individual atom cutoff is used. For simplicity in implementation, an individual atom cutoff is used in this paper. a.u. = arbitrary units.

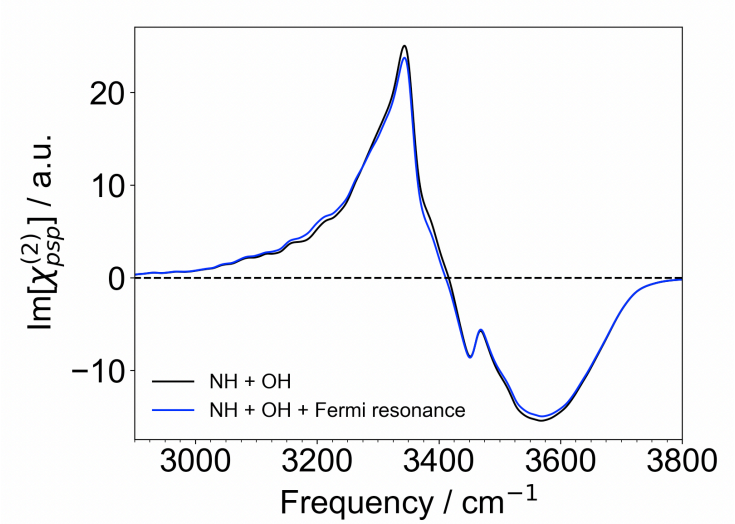

Figure A5.6. The effect of including Fermi resonance of the OH stretch with the water bend according to the method of Kananenka and Skinner.[1] Spectra are for the $LK_7\beta$ system. a.u. = arbitrary units.

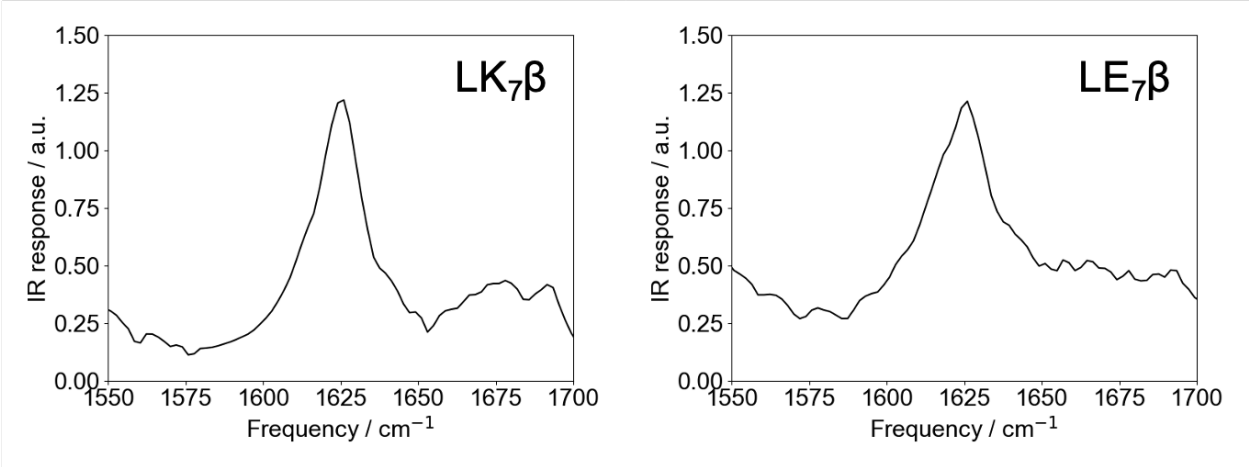

Figure A5.7. Fourier-transform infrared (FTIR) spectra of $LK_7\beta$ and $LE_7\beta$ films on glass in the amide I region. Peaks around 1625 cm^{-1} indicate a primarily β-sheet structure.

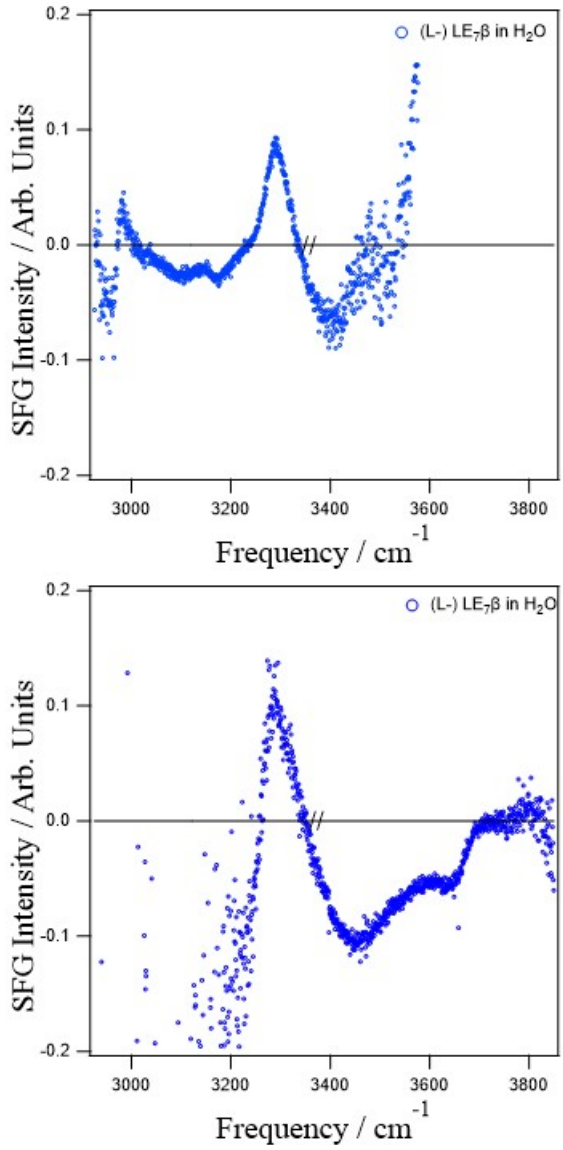

Figure A5.8. Spectra corresponding to the two spectral regions used to record the chiral heterodyne SFG spectra of LE7β presented in Figure 5.5. The noise on the edges was amplified due to normalization to the infrared profile. The centers of the infrared profiles used were ~3075 cm^{-1} (left) and ~3550 cm^{-1} (right).

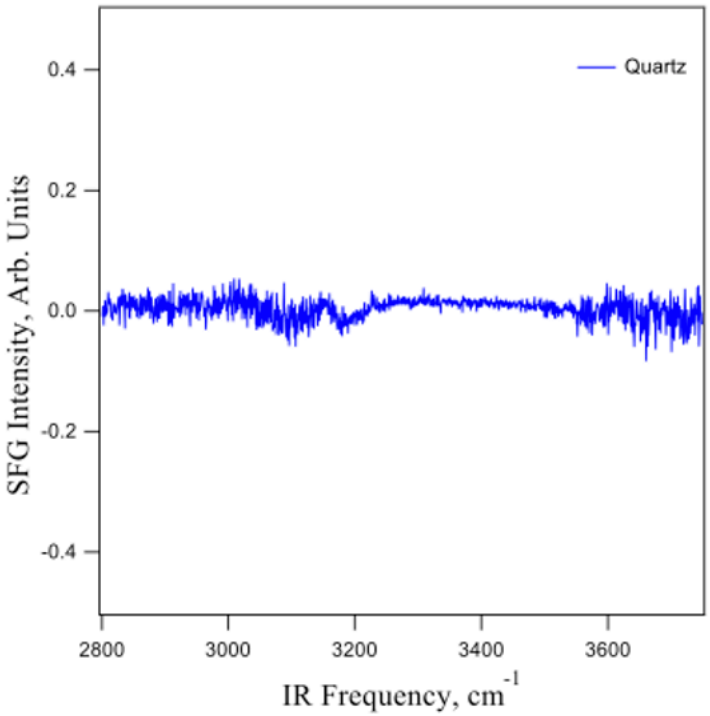

Figure A5.9. Spectrum of quartz alone.

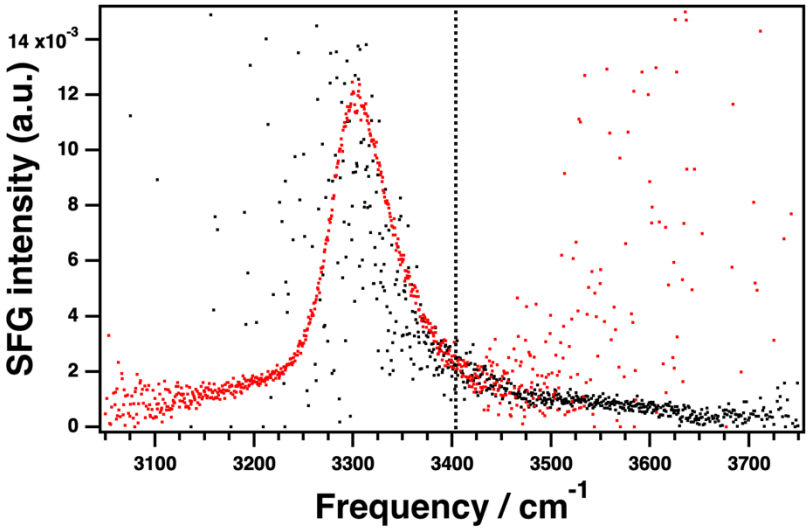

Figure A5.10. Homodyne spectra obtained in two spectral windows. They were stitched together to produce the spectrum of $LE_7\beta$ at the air-buffer interface (Figure 5.6). The spectra in the low-frequency (red) and high-frequency (black) windows were obtained using infrared profiles centered at 2900 cm^{-1} and 3200 cm^{-1}, respectively. The noise level at the of each spectrum was amplified due to normalization to the infrared profile. The black spectrum was scaled so that the high frequency tail of the red spectrum overlapped with the black spectrum. The two spectra were stitched at the position indicated by the dashed line.

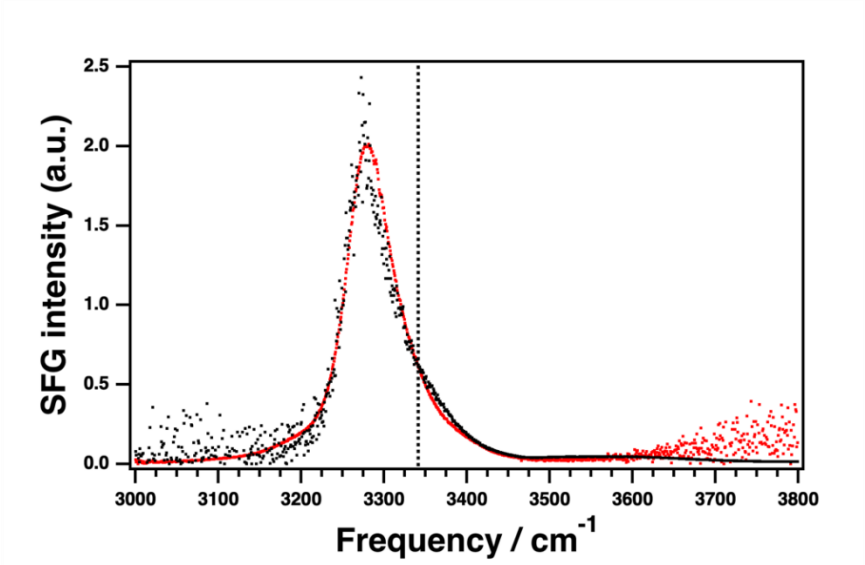

Figure A5.11. The two spectral windows that were stitched together to produce the homodyne spectrum of $LK_7\beta$ at the air-buffer interface (Figure 5.6). The spectra in the low-frequency (red) and high-frequency (black) windows were obtained using infrared profiles centered at 2875 cm^{-1} and 3175 cm^{-1}, respectively. The black spectrum was scaled so that the main feature at 3280 cm^{-1} overlapped with the corresponding feature in the red spectrum. The two spectra were stitched at the position indicated by the dashed line.

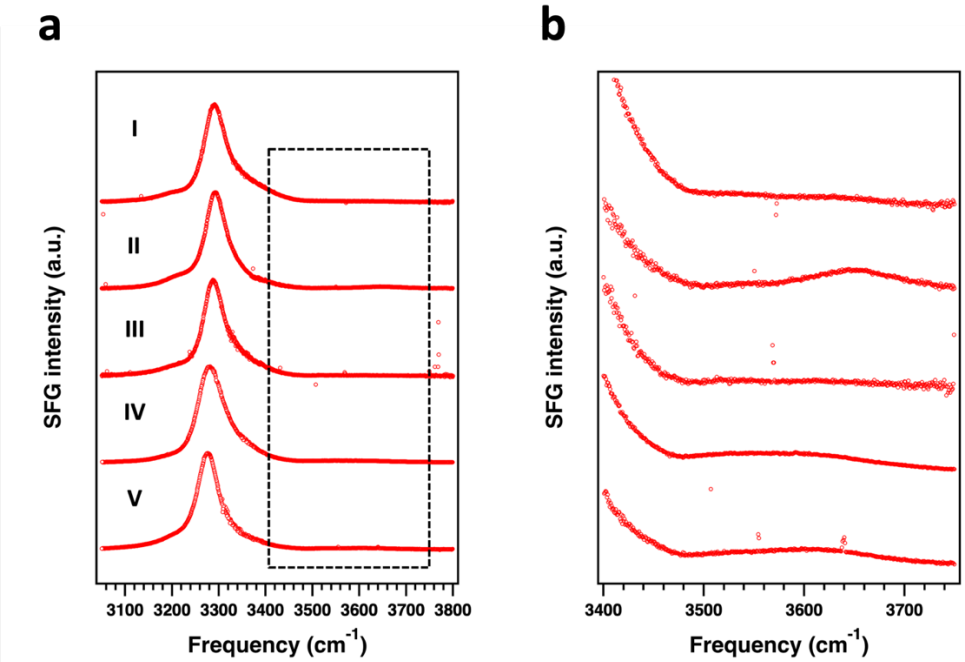

Figure A5.12. Chiral SFG spectra of $LK_7\beta$ at the air-water interface (a) in the 3000–3800 cm^{-1} region and (b) an enlargement of the dashed window covering the lower intensity part of the spectra in the region of 3400–3750 cm^{-1}. Some fluctuations of the very low SFG intensity at the high-frequency region were observed, and their origin is still under investigation. The spectra, I-V, were acquired on separate days. On each day, a frozen aliquot from the same stock solution of $LK_7\beta$ (1 mM) was freshly thawed and added onto the buffer with a Hamilton syringe for the measurements.

References

1. Horn, H. W.; Swope, W. C.; Pitera, J. W.; Madura, J. D.; Dick, T. J.; Hura, G. L.; Head-Gordon, T. Development of an improved four-site water model for biomolecular simulations: TIP4P-Ew. *J. Chem. Phys.* **2004,** *120*, 9665-9678.
2. Eastman, P.; Swails, J.; Chodera, J. D.; McGibbon, R. T.; Zhao, Y.; Beauchamp, K. A.; Wang, L.-P.; Simmonett, A. C.; Harrigan, M. P.; Stern, C. D., et al. OpenMM 7: Rapid development of high performance algorithms for molecular dynamics. *PLoS Comput. Biol.* **2017,** *13*, e1005659.
3. Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theory Comput.* **2015,** *11*, 3696-3713.
4. Joung, I. S.; Cheatham, T. E., III Determination of Alkali and Halide Monovalent Ion Parameters for Use in Explicitly Solvated Biomolecular Simulations. *J. Phys. Chem. B* **2008,** *112*, 9020-9041.
5. Joung, I. S.; Cheatham, T. E., III Molecular Dynamics Simulations of the Dynamic and Energetic Properties of Alkali and Halide Ions Using Water-Model-Specific Ion Parameters. *J. Phys. Chem. B* **2009,** *113*, 13279-13290.
6. Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. *J. Chem. Phys.* **1993,** *98*, 10089-10092.
7. Ohto, T.; Usui, K.; Hasegawa, T.; Bonn, M.; Nagata, Y. Toward ab initio molecular dynamics modeling for sum-frequency generation spectra; an efficient algorithm based on surface-specific velocity-velocity correlation function. *J. Chem. Phys.* **2015,** *143*, 124702.
8. Khatib, R.; Sulpizi, M. Sum Frequency Generation Spectra from Velocity–Velocity Correlation Functions. *J. Phys. Chem. Lett.* **2017,** *8*, 1310-1314.
9. Pieniazek, P. A.; Tainter, C. J.; Skinner, J. L. Interpretation of the water surface vibrational sum-frequency spectrum. *J. Chem. Phys.* **2011,** *135*, 044701.
10. Auer, B. M.; Skinner, J. L. Vibrational Sum-Frequency Spectroscopy of the Water Liquid/Vapor Interface. *J. Phys. Chem. B* **2009,** *113*, 4125-4130.
11. McGuire, J. A.; Shen, Y. R. Ultrafast Vibrational Dynamics at Water Interfaces. *Science* **2006,** *313*, 1945-1948.
12. University of California, S. F., Amber20. D.A. Case, H. M. A., K. Belfon, I.Y. Ben-Shalom, S.R. Brozell, D.S. Cerutti, T.E. Cheatham, III, G.A. Cisneros, V.W.D. Cruzeiro, T.A. Darden, R.E. Duke, G. Giambasu, M.K. Gilson, H. Gohlke, A.W. Goetz, R. Harris, S. Izadi, S.A. Izmailov, C. Jin, K. Kasavajhala, M.C. Kaymak, E. King, A. Kovalenko, T. Kurtzman, T.S. Lee, S. LeGrand, P. Li, C. Lin, J. Liu, T. Luchko, R. Luo, M. Machado, V. Man, M. Manathunga, K.M. Merz, Y. Miao, O. Mikhailovskii, G. Monard, H. Nguyen, K.A. O'Hearn, A. Onufriev, F. Pan, S. Pantano, R. Qi, A. Rahnamoun, D.R. Roe, A. Roitberg, C. Sagui, S. Schott-Verdugo, J. Shen, C.L. Simmerling, N.R. Skrynnikov, J. Smith, J. Swails, R.C. Walker, J. Wang, H. Wei, R.M. Wolf, X. Wu, Y. Xue, D.M. York, S. Zhao, and P.A. Kollman, Ed. University of California, San Francisco., 2021.
13. Schmidt, J. R.; Corcelli, S. A.; Skinner, J. L. Ultrafast vibrational spectroscopy of water and aqueous N-methylacetamide: Comparison of different electronic structure/molecular dynamics approaches. *J. Chem. Phys.* **2004,** *121*, 8887.

14. Auer, B. M.; Skinner, J. L. Dynamical effects in line shapes for coupled chromophores: Time-averaging approximation. *J. Chem. Phys.* **2007,** *127*, 104105.
15. Auer, B. M.; Skinner, J. L. IR and Raman spectra of liquid water: Theory and interpretation. *J. Chem. Phys.* **2008,** *128*, 224511.
16. Auer, B. M.; Skinner, J. L. Dynamical effects in line shapes for coupled chromophores: Time-averaging approximation. *J. Chem. Phys.* **2007,** *127*, 104105.
17. Corcelli, S. A.; Lawrence, C. P.; Skinner, J. L. Combined electronic structure/molecular dynamics approach for ultrafast infrared spectroscopy of dilute HOD in liquid H2O and D2O. *J. Chem. Phys.* **2004,** *120*, 8107-8117.
18. Corcelli, S. A.; Skinner, J. L. Infrared and Raman Line Shapes of Dilute HOD in Liquid H2O and D2O from 10 to 90 °C. *J. Phys. Chem. A* **2005,** *109*, 6154-6165.
19. Harris, C. R.; Millman, K. J.; van der Walt, S. J.; Gommers, R.; Virtanen, P.; Cournapeau, D.; Wieser, E.; Taylor, J.; Berg, S.; Smith, N. J., et al. Array programming with NumPy. *Nature* **2020,** *585*, 357-362.
20. Frostig, R.; Johnson, M.; Leary, C. In *Compiling machine learning programs via high-level tracing*, 2018.
21. Ma, G.; Liu, J.; Fu, L.; Yan, E. C. Probing water and biomolecules at the air–water interface with a broad bandwidth vibrational sum frequency generation spectrometer from 3800 to 900 cm− 1. *Applied spectroscopy* **2009,** *63*, 528-537.
22. Perets, E. A.; Konstantinovsky, D.; Fu, L.; Chen, J.; Wang, H.-F.; Hammes-Schiffer, S.; Yan, E. C. Mirror-image antiparallel β-sheets organize water molecules into superstructures of opposite chirality. *Proceedings of the National Academy of Sciences* **2020,** *117*, 32902-32909.
23. Konstantinovsky, D.; Perets, E. A.; Santiago, T.; Velarde, L.; Hammes-Schiffer, S.; Yan, E. C. Detecting the First Hydration Shell Structure around Biomolecules at Interfaces. *ACS Central Science* **2022**.
24. Fu, L.; Chen, S.-L.; Wang, H.-F. Validation of Spectra and Phase in Sub-1 cm–1 Resolution Sum-Frequency Generation Vibrational Spectroscopy through Internal Heterodyne Phase-Resolved Measurement. *J. Phys. Chem. B* **2016,** *120*, 1579-1589.

Chapter 6: Characterizing Interfaces by Voronoi Tessellation

Adapted from the following reference with permission:

Konstantinovsky, D.; Yan, E. C. Y.; Hammes-Schiffer, S.

J. Phys. Chem. Lett. 2023 (Accepted).

Abstract

The chemistry of interfaces differs markedly from that of the bulk. Calculation of interfacial properties depends strongly on the definition of the interface, which can lead to ambiguous results that vary between studies. There is a need for a method that can explicitly define the interfaces and boundaries in molecular systems. Voronoi tessellation offers an attractive solution to this problem through its ability to determine neighbors among specified groups of atoms. Here we discuss three cases where Voronoi tessellation combined with modeling of vibrational sum frequency generation (SFG) spectroscopy yields relevant insights: the breakdown of the air-water interface into clear and intuitive molecular layers, the study of the hydration shell in biological systems, and the acceleration of difficult spectral calculations where intermolecular vibrational couplings dominate. The utility of Voronoi tessellation has broad applications that extend beyond any single type of spectroscopy or system.

Interfaces play a central role in biological, chemical, and industrial processes. Most biological processes occur in aqueous environments, and the interface between water and the protein or nucleic acid often plays an important structural and functional role.[1-3] Moreover, electrochemical processes critical to energy conversion, such as carbon dioxide reduction and the oxygen evolution and oxygen reduction reactions, occur at metal-electrolyte interfaces.[4-6] The air-water interface is particularly important in atmospheric chemistry that occurs on the surface of water droplets.[7, 8] Thus, developing both experimental and computational methods for probing interfaces has widespread applications.

One of the central challenges in vibrational spectroscopic approaches aimed at probing interfaces is identifying the sources of the vibrational signals, which is often achieved through simulation. Analysis of the simulations typically requires the dissection of the system into interfacial regions and boundaries. Although distance-based cutoffs are commonly used for this purpose, the cutoff definition can be arbitrary and therefore yield inconsistent results. Voronoi tessellation defines interfaces and boundaries unambiguously by dividing a space containing a set of points into cells (Figure 6.1a) and determining the neighbors of each cell (Figure 6.1b).[9] A Voronoi cell consists of the area that is closer to its central point than to any other point. Once a Voronoi tessellation is computed, it becomes trivial to identify nearest neighbors of any set of points. Voronoi tessellation offers a computational scalpel that allows the atomic-level dissection of complex systems to help identify the sources of spectroscopic features.

Vibrational sum frequency generation (SFG) spectroscopy is a second-order nonlinear spectroscopy that is surface-selective and thus is an excellent technique for

probing both achiral and chiral interfacial systems.[10-28] This technique involves overlapping infrared (IR) and visible pulses in space and time on the sample and measuring the sum frequency response. By manipulating the polarization of the IR and visible beams and that of the detector, individual elements or a subset of the 27 elements of the susceptibility tensor $\chi^{(2)}_{IJK}$ can be isolated. Our group performs chirality-selective SFG (chiral SFG) using the polarization that isolates the element $\chi^{(2)}_{ZYX}$, which is sensitive to chirality at interfaces, including solvent chirality induced by a chiral solute.[29-31] Our group also performs conventional SFG (achiral SFG) using the polarization that isolates the element $\chi^{(2)}_{YYZ}$, which is sensitive to interfaces but not chirality. The SFG process is coherent, which allows for the extraction of the imaginary component of the response with heterodyne detection.[15, 16, 24, 26, 27, 32, 33] In achiral SFG, the sign of the peaks corresponds to the *z*-direction orientation of corresponding dipoles, whereas in chiral SFG, the sign of the phase has no clear physical meaning. Our group has used chiral SFG, in conjunction with molecular dynamics (MD) simulations and spectral calculations using electrostatic frequency maps,[34] to guide the interpretation of these spectra and study various biological systems.[24, 26, 27, 35, 36]

Herein, we combine Voronoi tessellation with computational modeling of SFG to reveal significant atomic-level properties of interfacial systems. We illustrate the power of this strategy for three diverse applications. First, we show how Voronoi tessellation can be used to divide the water at an air-water interface into well-defined layers, which then allows the calculation of the SFG response from each individual layer. This type of analysis provides insight into the hydrogen-bonding interactions between layers at the interface. Second, we show how Voronoi tessellation can be used to define the first hydration shell

around a protein at an air-water interface, and further dissect the first hydration shell into water molecules near the protein backbone versus those near the sidechains. Simulation of the chiral SFG response from each subset of water provides insight into the different hydrogen-bonding interactions between the protein and water and assists in the interpretation of the experimental data. Third, we show how the identification of the first hydration shell around a biomolecule can be used to speed up simulations of chiral SFG spectra without a significant loss of accuracy.

Although the pure air-water interface has been studied extensively, it is challenging to divide the interface into well-defined layers to calculate the vibrational response of each layer.[10, 12, 17, 19, 37-47] In some cases, SFG calculations of the air-water interface contradict each other due to subtle differences in definitions of interfacial layers.[45, 48] The popular instantaneous liquid interface method[49] relies on coarse-graining of the atomic density of the system, which involves arbitrary parameters such as coarse-graining length and grid size. By contrast, Voronoi tessellation defines layers unambiguously regardless of their composition. We use this approach to computationally divide the interfacial water into well-defined molecular layers (see Figure 6.2a and Appendix). In brief, we first add virtual points above the surface, construct a Voronoi tessellation with a cell for each atom and virtual point, and then identify the atom neighbors of the virtual points. If any one of a water molecule's atoms is a neighbor, the whole molecule is included in the first layer. We repeat the procedure for subsequent layers using the previous layer's atoms instead of the virtual points. We then calculate the achiral SFG response of each layer (Figure 6.2b) using two different water models, SPC/E and TIP4P-Ew.[50,51] Future layer-by-layer SFG simulations might use a higher level of theory (i.e., *ab initio* MD), following previous

work,[37, 52, 53] but this will most likely only allow convergence of the top few layers' spectra due to sampling limitations (i.e., *ab initio* MD trajectories cannot be propagated for enough time to converge very weak signals). To obtain the necessary copious sampling, we use classical force fields for this study.

We find that the top layer (layer 1, Figure 6.2a, red) produces most of the observed signal (Figure 6.2b, red), as expected from previous analyses.[45] There is a positive peak at high frequency (~3700 cm^{-1}) corresponding to free OH groups pointing up into the vacuum and a negative peak at lower frequency (~3450 cm^{-1}) corresponding to OH groups pointing down and interacting with the lower layer of water. Because red-shifting of the OH stretch away from ~3700 cm^{-1} is associated with stronger hydrogen bonding, this negative peak corresponds to hydrogen bonded OH groups. However, the SFG response of layer 2 (Figure 6.2a, blue) is less obvious. Its lineshape is inverted compared to that of layer 1 (Figure 6.2b, red versus blue), with the strong hydrogen bonds being donated *up* into layer 1 rather than down into layer 3 (Figure 6.2a, orange). This indicates that layer 2 and layer 1 have a particularly strong attraction compared to those of other pairs of layers. Analysis of hydrogen bonds between the layers (Figure 6.2c) reveals that for both water models, the hydrogen bond length is 2.88 Å between all layers except that the water molecules donating hydrogen bonds from layer 2 into layer 1 have shorter hydrogen bonds of length 2.86 Å with a hydrogen bond angle slightly closer to the ideal 180°. This stronger hydrogen bonding corresponds to the water molecules pointing up from layer 2 into layer 1. This analysis demonstrates the exquisite sensitivity of SFG to the substructure of the interface as well as the utility of Voronoi tessellation in unraveling subtle molecular features when combined with SFG modeling. We note that our SFG results for the first few layers are

similar to those obtained by Kaliannan et al. using a probe-based layer identification method.[38] An advantage of Voronoi tessellation is that once the tessellation is done for a given configuration, identifying any number of layers is computationally trivial, as there is no need to probe each layer individually.

Although the SFG responses of layer 1 and layer 2 (i.e., the top ~5 Å) are significant, those of the lower layers are very small (Figure 6.2b). However, layer 3 (Figure 6.2a, orange) has a very well-converged SFG response that is similar to that of layer 6.2 – molecules with OH groups pointing up have stronger hydrogen-bonding interactions, based on the relative frequencies of the up and down peaks. This layer extends ~10 Å below the surface. Although our hydrogen bond analysis (Figure 6.2c) was unable to pick up any significant differences between molecules pointing up and down in this layer, the SFG response indicates a subtle asymmetry in this layer. This suggests that the orientations of water molecules in layer 3 are still influenced by the interface, however slightly. Even though we averaged over 10,000,000 frames and 100 ns of simulation, the layer 4 spectra (Figure 6.2b, magenta) are poorly converged. However, both water models show a lineshape curiously similar to that of layer 1 – molecules facing up have very slightly weaker hydrogen bonds than molecules facing down. Together all these results suggest a model where truly bulk behavior is not reached even ~16 Å below the surface, contrary to past suggestions that only the first several Å are important.[54-57]

Although the SFG signal mostly originates within the first few water layers,[45] some properties converge much less quickly. In particular, we calculated the average number of water molecules that move between adjacent layers (normalized per water molecule) over 1 ps and found that the value does not converge to the bulk value even four layers down

(Figure 6.2c). The rate of interlayer water transfer from a given layer to the layer above it increases with each subsequent layer, suggesting that the interface has slower diffusion in the *z*-direction than the bulk and that this property does not rapidly approach a bulk value. This implies that water molecules that are affected by the surface may not only include the top few layers but may also include deeper regions as well. Voronoi tessellation has enabled us to precisely define molecular layers and uncover hidden properties of the interface and reveals that the definition of the interface is dependent on the property used to define it.

In addition to defining molecular layers at interfaces, Voronoi tessellation can also unambiguously define the hydration shell of a biomolecule.[58] Our group has used Voronoi tessellation to define the first and second hydration shells around the model system $LK_7\beta$, an amphiphilic peptide that folds into antiparallel β-sheets at the air-water interface, which makes it an ideal benchmark system for SFG studies.[27, 35] This system consists of leucine residues that point into the air and positively charged lysine residues that point into the water (Figure 6.3). $LE_7\beta$ contains negatively charged glutamate residues instead of lysine residues, but it is otherwise very similar to $LK_7\beta$. We model each protein as a five-stranded antiparallel β-sheet at the vacuum-water interface. Together, Voronoi tessellation and chiral SFG reveal surprising differences between the hydration shells of the two similar proteins.

In a previous study, we used Voronoi tessellation to identify the first solvation shell but then used distance cutoffs to separate water molecules interacting with the backbone from those interacting with the sidechains. This involved rather convoluted selection criteria (see ref [27], SI) that ultimately did not perfectly remove the overlap between subsets,

complicating spectral interpretation of the backbone- and sidechain-associated water molecule contributions. Here, we greatly simplify the selection by defining "backbone water" as those molecules that are direct neighbors of backbone atoms and "sidechain water" as those that are direct neighbors of the ends of the sidechains (the NH_3^+ group in $LK_7\beta$; the COO^- group in $LE_7\beta$) and are not also neighbors of the backbone (Figure 6.4a). Voronoi tessellation enables precise yet simple partitioning of the first hydration shell into components. The chiral SFG response of these subsets can then be calculated to reveal the components of the total SFG spectrum.

We have previously shown that the main NH stretch peak in the chiral SFG spectrum of $LE_7\beta$ is blue-shifted compared to the analogous peak in $LK_7\beta$. We argued that this shift is due to significantly fewer backbone hydrogen bonds in $LE_7\beta$ compared to $LK_7\beta$, which indicates weaker inter-strand hydrogen-bonding interactions in $LE_7\beta$, but we did not analyze the OH stretch components of the spectrum.[35] Figures 6.4b and 6.4c show that the first hydration shell and backbone water responses of $LE_7\beta$ are very different from those of $LK_7\beta$, even though the systems appear similar. The OH stretch response of $LE_7\beta$ is both much smaller and somewhat blue-shifted compared to that of $LK_7\beta$. The blue shift might indicate that the hydration shell of $LE_7\beta$ is less strongly bound compared to the hydration shell of $LK_7\beta$. However, hydrogen bond analysis reveals that this is not the case. Instead, protein-water hydrogen bonds are actually somewhat stronger in $LE_7\beta$ compared to $LK_7\beta$ (donor-acceptor distances of 2.74 ± 0.0002 Å vs. 2.87 ± 0.0002 Å and angles of $164.58 \pm 0.0083°$ vs. $162.56 \pm 0.0110°$, mean ± standard error). This analysis suggests that the key variable in determining the chiral OH stretch response and the integrity of the first hydration shell is the stability of the protein structure rather than the strength of the

interaction between the protein and the surrounding water. $LK_7\beta$ forms a more stable antiparallel β-sheet structure,[35] and therefore its hydration shell maintains a more rigid chiral structure. The blue shift and smaller response of $LE_7\beta$ compared to $LK_7\beta$ illustrates differential chiral induction – a different degree of imprinting of chirality on achiral solvent. Comparison of the two systems suggests that hydration water spectra can be used to infer relative biomolecule stability in some cases. In addition, we find that the backbone-associated water molecules contribute most of the lineshape in both systems (Figure 6.4c), and the sidechain responses (Figure 6.4d) are quite small. This observation is consistent with our previous work, where we found that water molecules hydrogen bonded to carbonyl groups on the backbone were significant contributors to the chiral SFG response of the first hydration shell, especially on a per-water-molecule basis.[27]

In addition to revealing the structure of molecular systems, Voronoi tessellation can also speed up the calculation of chiral SFG spectra. The SFG calculations presented thus far have taken advantage of an approximation where intermolecular couplings between OH groups are neglected. This approximation allows simple separation of water molecule subsets and makes the exciton Hamiltonian matrix block-diagonal and thus trivial to diagonalize.[27,59-62] However, in certain cases, intermolecular couplings are crucial to chiral SFG spectra, and these make the Hamiltonian expensive to diagonalize. In particular, our previous work showed that intermolecular coupling is critical to the modeling of the combined NH/OH stretch chiral SFG response of $LK_7\beta$.[35] If intermolecular coupling is neglected, a downward-pointing peak (indicated by a dip) appears that is not found in the experimental spectrum (Figure 6.5a). Our previous work also showed that almost all of the OH stretch signal and the most relevant NH/OH couplings arise between the biomolecule

and the first hydration shell.[27] Here we obtain the first hydration shell by Voronoi tessellation, as discussed above, and then only consider the protein and the first hydration shell in the exciton Hamiltonian to speed up the calculation.[35]

Figure 6.5a shows the result of the first-hydration-shell approximation along with the full-system calculation for $LK_7\beta$. All features found in the full-system spectrum are present in the first-hydration-shell spectrum. The agreement with experiment is good whether or not the approximation is applied, except for the low-frequency negative peak at ~3150 cm^{-1}, which is missing in all the calculated spectra and is a focus of ongoing efforts by our group.[35]

Figure 6.5b shows the increasing speedup of including only the first hydration shell as a function of system size. Using Voronoi tessellation to unambiguously define the first hydration shell can significantly accelerate prohibitively expensive spectral calculations and will enable modeling of vibrational spectra of large biological systems in the future without a significant sacrifice in accuracy. Modeling chiral SFG spectra of large systems will enable us to probe water-protein interactions of amyloids, antibodies, and other biologically relevant systems and learn how surrounding water molecules shape, and are shaped by, these systems.

We have shown a diverse set of three situations where Voronoi tessellation combined with SFG modeling yields insights. First, we demonstrated that Voronoi tessellation allows the dissection of the air-water interface at a high level of detail, yielding new findings about interface structure and thickness. We then showed that Voronoi tessellation enables detailed analysis of the differences between the first hydration shells of closely related protein systems. Finally, we used Voronoi tessellation and the knowledge

that chiral SFG is largely selective to the first hydration shell[27] to greatly simplify and speed up the calculation of vibrational spectra of highly coupled condensed-phase systems. Although we restricted our study to SFG, the Voronoi tessellation approaches developed here can potentially be applicable in a variety of molecular contexts, for example in the analysis of surface-selective attenuated total reflectance IR spectroscopy (ATR-IR) experiments, protein-protein interactions, and biomolecule-lipid interactions. These approaches will also be applicable to the SFG study of charged aqueous interfaces relevant to electrocatalysis. Altogether, our findings demonstrate the remarkable power of combining Voronoi tessellation with modeling of vibrational spectroscopy.

Summary of Appendix to Chapter 6

The Appendix contains a detailed description of computational methods used in this study. The Voronoi tessellation analysis code used in this study can be found at https://github.com/dkonstan/VoronoiSelection along with an example script.

Acknowledgements

The authors thank Dr. Pablo E. Videla at Yale University for helpful discussions and for sharing VMD Tcl scripts for setting up the $LK_7\beta$ system. We also thank Dr. Alexander Soudackov, Matthew Tremblay, Kristian Olesen, and Dr. Ty Santiago at Yale University and Dr. Ethan A. Perets at Harvard University for helpful suggestions. This work was supported by the National Institutes of Health Grant R35 GM139449 (S.H.-S.) and the NSF Grant CHE-2108690 (E.C.Y.Y.). D.K. was supported by these grants.

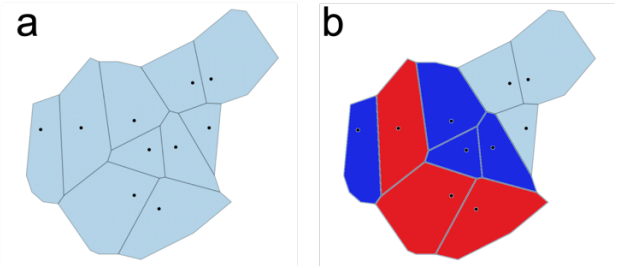

Figure 6.1. a) A two-dimensional Voronoi tessellation, where each cell consists of the area closer to the central point than to any other points. b) Finding neighbors of a set of points is straightforward once the Voronoi tessellation has been made. Here, the dark blue cells are the neighbors of the red cells.

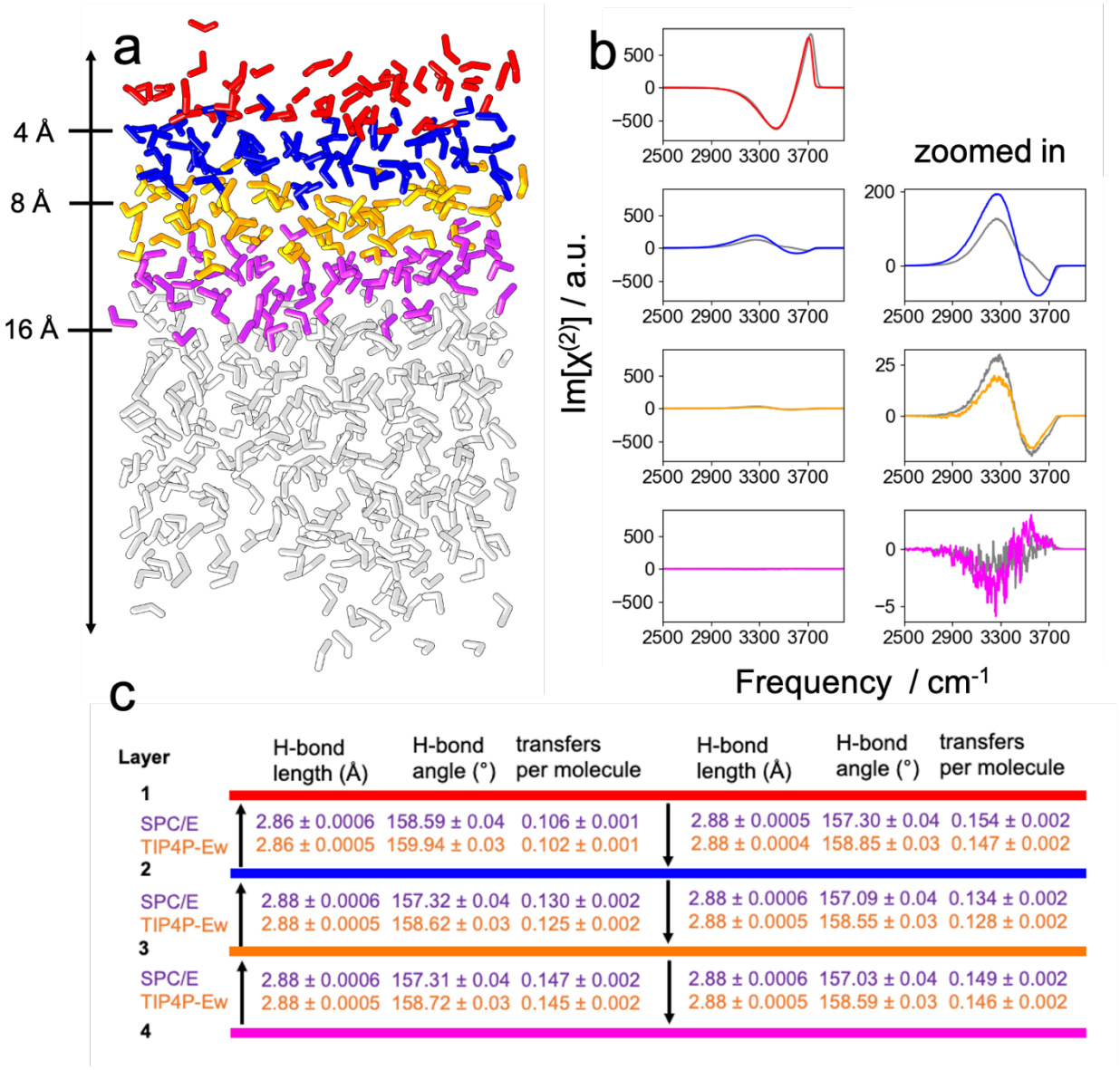

Figure 6.2. a) Water surface separated into layers (layer 1 – red, layer 2 – blue, layer 3 – orange, layer 4 – magenta) using Voronoi tessellation and neighbor analysis. b) Computed achiral SFG spectra using two different water models, SPC/E (gray) and TIP4P-Ew (colors corresponding to those in part a). c) Analysis of hydrogen bonds and molecule diffusion between water layers for SPC/E (purple) and TIP4P-Ew (orange) systems. The arrows indicate either donation of a hydrogen bond from one layer to the adjacent layer or the transfer of water molecules from one layer to the adjacent layer over 1 ps. Values are shown as mean ± standard error.

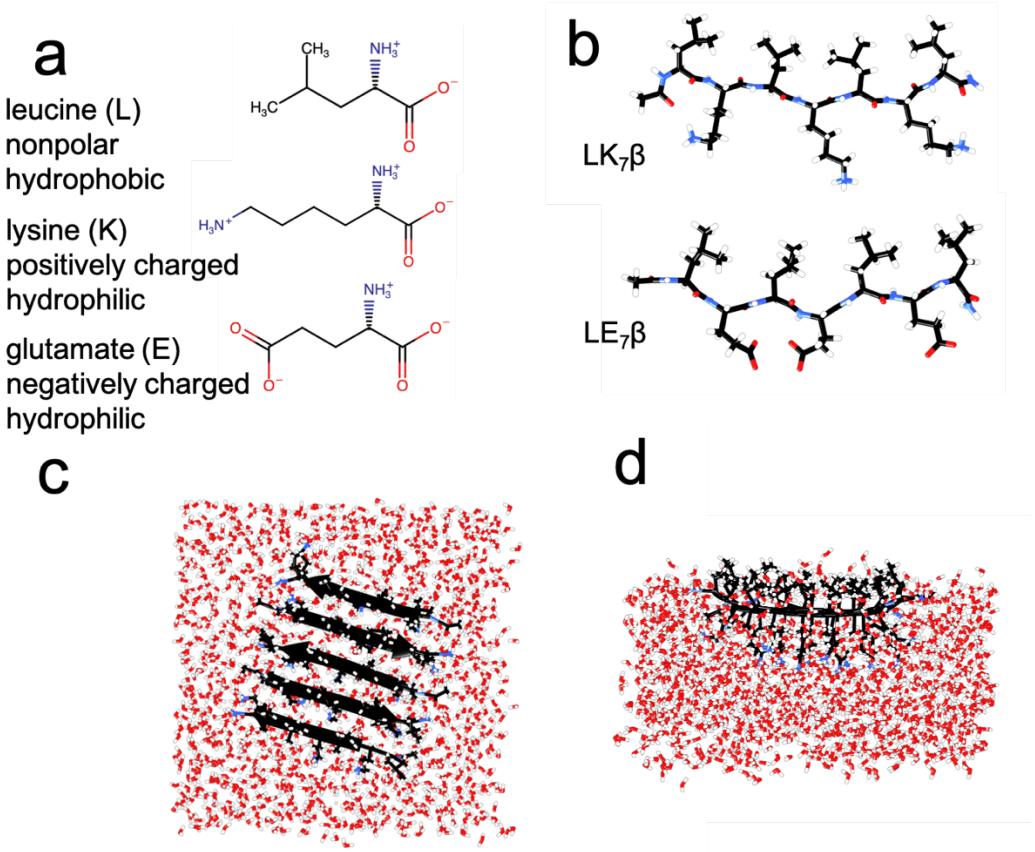

Figure 6.3. a) Structure and chemical properties of leucine (L), lysine (K), and glutamate (E). b) $LK_7\beta$ and $LE_7\beta$ peptides. c) These peptides form antiparallel β-sheets at the air-water interface, where polar residues lysine (K) or glutamate (E) point into the solvent while the hydrophobic residue leucine (L) points into the air; top view of aqueous $LK_7\beta$. d) and side view of aqueous $LK_7\beta$.

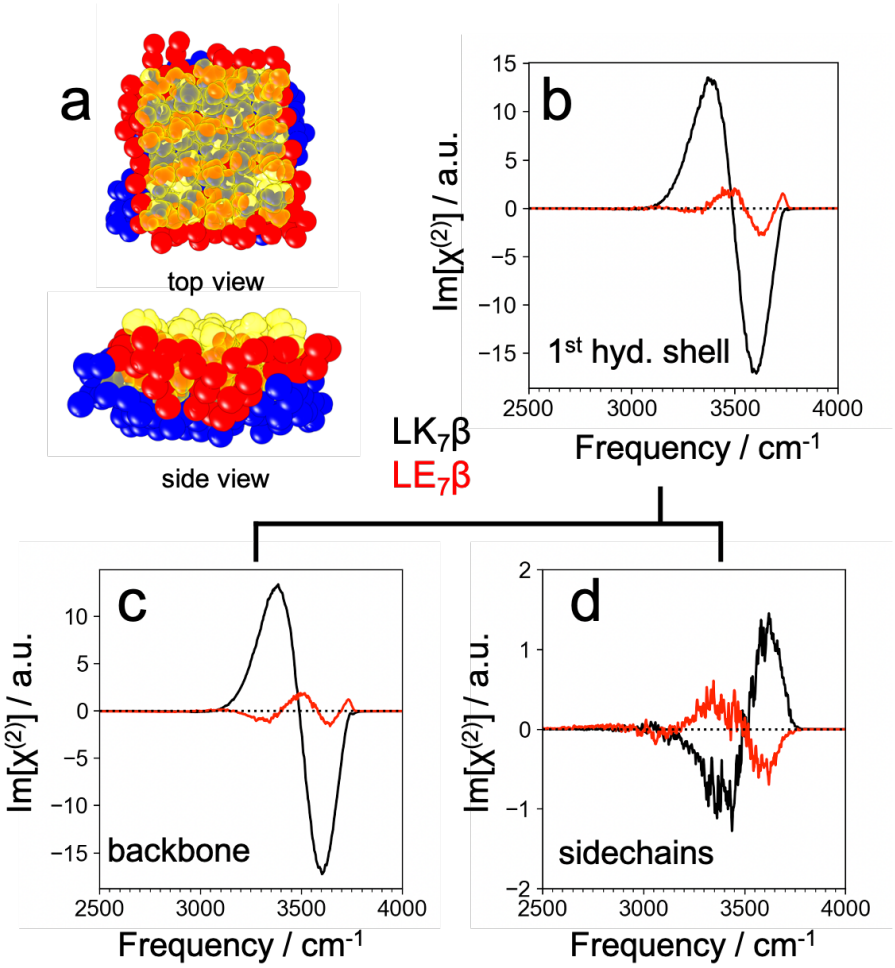

Figure 6.4. (a) Voronoi tessellation-based selections of water molecules near the backbone (red) and near sidechains (blue) of $LK_7\beta$ (transparent yellow). (b) Computed first hydration shell OH-stretch response for $LK_7\beta$ (black) and $LE_7\beta$ (red). (c) Computed backbone-associated water OH-stretch response for $LK_7\beta$ and $LE_7\beta$. (d) Computed sidechain-associated water OH-stretch response for $LK_7\beta$ and $LE_7\beta$.

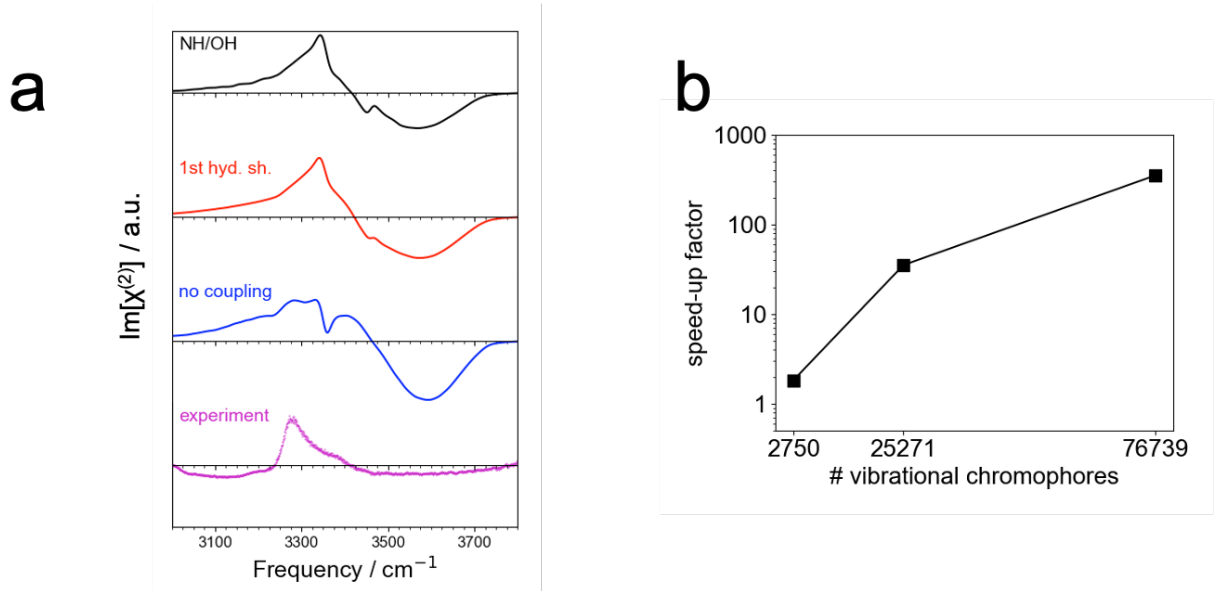

Figure 6.5. a) NH stretch / OH stretch chiral SFG spectra of $LK_7\beta$ and surrounding water, showing the modest effect of the first-hydration-shell approximation and strong effect of excluding intermolecular couplings between NH and OH groups. b) Speedup factor for the first-hydration-shell method over full-system calculation for molecular systems of various sizes.

References

1. Bellissent-Funel, M.-C.; Hassanali, A.; Havenith, M.; Henchman, R.; Pohl, P.; Sterpone, F.; van der Spoel, D.; Xu, Y.; Garcia, A. E. Water Determines the Structure and Dynamics of Proteins. *Chem. Rev.* **2016,** *116*, 7673-7697.
2. Bhat, T. N.; Bentley, G. A.; Boulot, G.; Greene, M. I.; Tello, D.; Dall'Acqua, W.; Souchon, H.; Schwarz, F. P.; Mariuzza, R. A.; Poljak, R. J. Bound water molecules and conformational stabilization help mediate an antigen-antibody association. *Proc. Natl. Acad. Sci. USA* **1994,** *91*, 1089-1093.
3. Privalov, P. L.; Crane-Robinson, C. Role of water in the formation of macromolecular structures. *Eur. Biophys. J.* **2017,** *46*, 203-224.
4. Todorova, T. K.; Schreiber, M. W.; Fontecave, M. Mechanistic Understanding of CO2 Reduction Reaction (CO2RR) Toward Multicarbon Products by Heterogeneous Copper-Based Catalysts. *ACS Catal.* **2020,** *10*, 1754-1768.
5. Yang, Y.; Peltier, C. R.; Zeng, R.; Schimmenti, R.; Li, Q.; Huang, X.; Yan, Z.; Potsi, G.; Selhorst, R.; Lu, X., et al. Electrocatalysis in Alkaline Media and Alkaline Membrane-Based Energy Technologies. *Chem. Rev.* **2022,** *122*, 6117-6321.
6. Steinmann, S. N.; Michel, C. How to Gain Atomistic Insights on Reactions at the Water/Solid Interface? *ACS Catal.* **2022,** *12*, 6294-6301.
7. Rapf, R. J.; Vaida, V. Sunlight as an energetic driver in the synthesis of molecules necessary for life. *Phys. Chem. Chem. Phys.* **2016,** *18*, 20067-20084.
8. Griffith Elizabeth, C.; Vaida, V. In situ observation of peptide bond formation at the water–air interface. *Proc. Natl. Acad. Sci. USA* **2012,** *109*, 15697-15701.
9. Voronoi, G. Nouvelles applications des paramètres continus à la théorie des formes quadratiques. Deuxième mémoire. Recherches sur les parallélloèdres primitifs. *J. Reine Angew. Math.* **1908,** *1908*, 198-287.
10. Du, Q.; Superfine, R.; Freysz, E.; Shen, Y. R. Vibrational spectroscopy of water at the vapor/water interface. *Phys. Rev. Lett.* **1993,** *70*, 2313-2316.
11. Zhu, X. D.; Suhr, H.; Shen, Y. R. Surface vibrational spectroscopy by infrared-visible sum frequency generation. *Phys. Rev. B* **1987,** *35*, 3047-3050.
12. Morita, A.; Hynes, J. T. A Theoretical Analysis of the Sum Frequency Generation Spectrum of the Water Surface. II. Time-Dependent Approach. *J. Phys. Chem. B* **2002,** *106*, 673-685.
13. Morita, A. Improved Computation of Sum Frequency Generation Spectrum of the Surface of Water. *J. Phys. Chem. B* **2006,** *110*, 3158-3163.
14. Morita, A.; Ishiyama, T. Recent progress in theoretical analysis of vibrational sum frequency generation spectroscopy. *Phys. Chem. Chem. Phys.* **2008,** *10*, 5801-5816.
15. Nihonyanagi, S.; Ishiyama, T.; Lee, T.-k.; Yamaguchi, S.; Bonn, M.; Morita, A.; Tahara, T. Unified Molecular View of the Air/Water Interface Based on Experimental and Theoretical $\chi(2)$ Spectra of an Isotopically Diluted Water Surface. *J. Am. Chem. Soc.* **2011,** *133*, 16875-16880.
16. Kundu, A.; Tanaka, S.; Ishiyama, T.; Ahmed, M.; Inoue, K.-i.; Nihonyanagi, S.; Sawai, H.; Yamaguchi, S.; Morita, A.; Tahara, T. Bend Vibration of Surface Water Investigated by Heterodyne-Detected Sum Frequency Generation and Theoretical Study: Dominant Role of Quadrupole. *J. Phys. Chem. Lett.* **2016,** *7*, 2597-2601.

17. Joutsuka, T.; Hirano, T.; Sprik, M.; Morita, A. Effects of third-order susceptibility in sum frequency generation spectra: a molecular dynamics study in liquid water. *Phys. Chem. Chem. Phys.* **2018,** *20*, 3040-3053.
18. Ishiyama, T.; Morita, A. Nuclear Quantum Effect on the χ(2) Band Shape of Vibrational Sum Frequency Generation Spectra of Normal and Deuterated Water Surfaces. *J. Phys. Chem. Lett.* **2019,** *10*, 5070-5075.
19. 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.
20. Fu, L.; Liu, J.; Yan, E. C. Y. Chiral Sum Frequency Generation Spectroscopy for Characterizing Protein Secondary Structures at Interfaces. *J. Am. Chem. Soc.* **2011,** *133*, 8094-8097.
21. Fu, L.; Wang, Z.; Yan, E. C. Y. Chiral Vibrational Structures of Proteins at Interfaces Probed by Sum Frequency Generation Spectroscopy. *Int. J. Mol. Sci.* **2011,** *12*, 9404-9425.
22. Wang, Z.; Fu, L.; Yan, E. C. Y. C–H Stretch for Probing Kinetics of Self-Assembly into Macromolecular Chiral Structures at Interfaces by Chiral Sum Frequency Generation Spectroscopy. *Langmuir* **2013,** *29*, 4077-4083.
23. Yan, E. C. Y.; Wang, Z.; Fu, L. Proteins at Interfaces Probed by Chiral Vibrational Sum Frequency Generation Spectroscopy. *J. Phys. Chem. B* **2015,** *119*, 2769-2785.
24. Perets, E. A.; Videla, P. E.; Yan, E. C. Y.; Batista, V. S. Chiral inversion of amino acids in antiparallel beta sheets at interfaces probed by vibrational sum frequency generation spectroscopy. *J. Phys. Chem. B* **2019,** *123*, 5769−5781.
25. Perets, E. A.; Yan, E. C. Y. Chiral Water Superstructures around Antiparallel β-Sheets Observed by Chiral Vibrational Sum Frequency Generation Spectroscopy. *J. Phys. Chem. Lett.* **2019,** *10*, 3395-3401.
26. Perets, E. A.; Konstantinovsky, D.; Fu, L.; Chen, J.; Wang, H.-F.; Hammes-Schiffer, S.; Yan, E. C. Y. Mirror-image antiparallel β-sheets organize water molecules into superstructures of opposite chirality. *Proc. Natl. Acad. Sci. USA* **2020,** *117*, 32902-32909.
27. Konstantinovsky, D.; Perets, E. A.; Santiago, T.; Velarde, L.; Hammes-Schiffer, S.; Yan, E. C. Y. Detecting the First Hydration Shell Structure around Biomolecules at Interfaces. *ACS Cent. Sci.* **2022,** *8*, 1404-1414.
28. Anglin, T. C.; Massari, A. M. Polarization-multiplexed vibrational sum frequency generation for comprehensive simultaneous characterization of interfaces. *Opt. Lett.* **2012,** *37*, 1754-1756.
29. Moad, A. J.; Simpson, G. J. A Unified Treatment of Selection Rules and Symmetry Relations for Sum-Frequency and Second Harmonic Spectroscopies. *J. Phys. Chem. B* **2004,** *108*, 3548-3562.
30. Simpson, G. J. Molecular Origins of the Remarkable Chiral Sensitivity of Second-Order Nonlinear Optics. *ChemPhysChem* **2004,** *5*, 1301-1310.
31. 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.

32. Fu, L.; Chen, S.-L.; Wang, H.-F. Validation of Spectra and Phase in Sub-1 cm–1 Resolution Sum-Frequency Generation Vibrational Spectroscopy through Internal Heterodyne Phase-Resolved Measurement. *J. Phys. Chem. B* **2016,** *120*, 1579-1589.
33. Stiopkin, I. V.; Jayathilake, H. D.; Bordenyuk, A. N.; Benderskii, A. V. Heterodyne-detected vibrational sum frequency generation spectroscopy. *J. Am. Chem. Soc.* **2008,** *130*, 2271-2275.
34. Baiz, C. R.; Błasiak, B.; Bredenbeck, J.; Cho, M.; Choi, J.-H.; Corcelli, S. A.; Dijkstra, A. G.; Feng, C.-J.; Garrett-Roe, S.; Ge, N.-H., et al. Vibrational Spectroscopic Map, Vibrational Spectroscopy, and Intermolecular Interaction. *Chem. Rev.* **2020,** *120*, 7152-7218.
35. Konstantinovsky, D.; Perets, E. A.; Santiago, T.; Olesen, K.; Wang, Z.; Soudackov, A. V.; Yan, E. C. Y.; Hammes-Schiffer, S. Design of an Electrostatic Frequency Map for the NH Stretch of the Protein Backbone and Application to Chiral Sum Frequency Generation Spectroscopy. *J. Phys. Chem. B* **2023**.
36. Konstantinovsky, D.; Perets, E. A.; Santiago, T.; Olesen, K.; Wang, Z.; Soudackov, A. V.; Yan, E. C. Y.; Hammes-Schiffer, S. Design of an Electrostatic Frequency Map for the NH Stretch of the Protein Backbone and Application to Chiral Sum Frequency Generation Spectroscopy. *J. Phys. Chem. B* **2023,** *127*, 2418-2429.
37. Ishiyama, T.; Morita, A. Analysis of anisotropic local field in sum frequency generation spectroscopy with the charge response kernel water model. *J. Chem. Phys.* **2009,** *131*, 244714.
38. Kaliannan, N. K.; Henao Aristizabal, A.; Wiebeler, H.; Zysk, F.; Ohto, T.; Nagata, Y.; Kühne, T. D. Impact of intermolecular vibrational coupling effects on the sum-frequency generation spectra of the water/air interface. *Mol. Phys.* **2020,** *118*, 1620358.
39. Tang, F.; Ohto, T.; Sun, S.; Rouxel, J. R.; Imoto, S.; Backus, E. H. G.; Mukamel, S.; Bonn, M.; Nagata, Y. Molecular Structure and Modeling of Water–Air and Ice–Air Interfaces Monitored by Sum-Frequency Generation. *Chem. Rev.* **2020,** *120*, 3633-3667.
40. Ohto, T.; Dodia, M.; Imoto, S.; Nagata, Y. Structure and Dynamics of Water at the Water–Air Interface Using First-Principles Molecular Dynamics Simulations within Generalized Gradient Approximation. *J. Chem. Theory Comput.* **2019,** *15*, 595-602.
41. Khatib, R.; Hasegawa, T.; Sulpizi, M.; Backus, E. H. G.; Bonn, M.; Nagata, Y. Molecular Dynamics Simulations of SFG Librational Modes Spectra of Water at the Water–Air Interface. *J. Phys. Chem. C* **2016,** *120*, 18665-18673.
42. Ni, Y.; Skinner, J. L. IR and SFG vibrational spectroscopy of the water bend in the bulk liquid and at the liquid-vapor interface, respectively. *J. Chem. Phys.* **2015,** *143*, 014502.
43. Auer, B. M.; Skinner, J. L. Vibrational Sum-Frequency Spectroscopy of the Water Liquid/Vapor Interface. *J. Phys. Chem. B* **2009,** *113*, 4125-4130.
44. Shen, Y. R.; Ostroverkhov, V. Sum-Frequency Vibrational Spectroscopy on Water Interfaces: Polar Orientation of Water Molecules at Interfaces. *Chem. Rev.* **2006,** *106*, 1140-1154.
45. Pezzotti, S.; Galimberti, D. R.; Gaigeot, M.-P. 2D H-Bond Network as the Topmost Skin to the Air–Water Interface. *J. Phys. Chem. Lett.* **2017,** *8*, 3133-3141.

46. Perakis, F.; De Marco, L.; Shalit, A.; Tang, F.; Kann, Z. R.; Kühne, T. D.; Torre, R.; Bonn, M.; Nagata, Y. Vibrational Spectroscopy and Dynamics of Water. *Chem. Rev.* **2016,** *116*, 7590-7607.
47. Eisenthal, K. B. Liquid Interfaces Probed by Second-Harmonic and Sum-Frequency Spectroscopy. *Chem. Rev.* **1996,** *96*, 1343-1360.
48. Liang, C.; Jeon, J.; Cho, M. Ab initio Modeling of the Vibrational Sum-Frequency Generation Spectrum of Interfacial Water. *J. Phys. Chem. Lett.* **2019,** *10*, 1153-1158.
49. Willard, A. P.; Chandler, D. Instantaneous Liquid Interfaces. *J. Phys. Chem. B* **2010,** *114*, 1954-1958.
50. Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. The missing term in effective pair potentials. *J. Phys. Chem. A* **1987,** *91*, 6269-6271.
51. Horn, H. W.; Swope, W. C.; Pitera, J. W.; Madura, J. D.; Dick, T. J.; Hura, G. L.; Head-Gordon, T. Development of an improved four-site water model for biomolecular simulations: TIP4P-Ew. *J. Chem. Phys.* **2004,** *120*, 9665-9678.
52. Ohto, T.; Usui, K.; Hasegawa, T.; Bonn, M.; Nagata, Y. Toward ab initio molecular dynamics modeling for sum-frequency generation spectra; an efficient algorithm based on surface-specific velocity-velocity correlation function. *J. Chem. Phys.* **2015,** *143*, 124702.
53. Khatib, R.; Sulpizi, M. Sum Frequency Generation Spectra from Velocity–Velocity Correlation Functions. *J. Phys. Chem. Lett.* **2017,** *8*, 1310-1314.
54. Bonn, M.; Nagata, Y.; Backus, E. H. G. Molecular Structure and Dynamics of Water at the Water–Air Interface Studied with Surface-Specific Vibrational Spectroscopy. *Angew. Chem. Int. Ed.* **2015,** *54*, 5560-5576.
55. Hsieh, C.-S.; Okuno, M.; Hunger, J.; Backus, E. H. G.; Nagata, Y.; Bonn, M. Aqueous Heterogeneity at the Air/Water Interface Revealed by 2D-HD-SFG Spectroscopy. *Angew. Chem. Int. Ed.* **2014,** *53*, 8146-8149.
56. Sun, Q.; Guo, Y. Vibrational sum frequency generation spectroscopy of the air/water interface. *J. Mol. Liq.* **2016,** *213*, 28-32.
57. Reddy, S. K.; Thiraux, R.; Wellen Rudd, B. A.; Lin, L.; Adel, T.; Joutsuka, T.; Geiger, F. M.; Allen, H. C.; Morita, A.; Paesani, F. Bulk Contributions Modulate the Sum-Frequency Generation Spectra of Water on Model Sea-Spray Aerosols. *Chem* **2018,** *4*, 1629-1644.
58. David, E. E.; David, C. W. Voronoi polyhedra as a tool for studying solvation structure. *J. Chem. Phys.* **1982,** *76*, 4611-4614.
59. Corcelli, S. A.; Lawrence, C. P.; Skinner, J. L. Combined electronic structure/molecular dynamics approach for ultrafast infrared spectroscopy of dilute HOD in liquid H2O and D2O. *J. Chem. Phys.* **2004,** *120*, 8107-8117.
60. Auer, B. M.; Skinner, J. L. IR and Raman spectra of liquid water: Theory and interpretation. *J. Chem. Phys.* **2008,** *128*, 224511.
61. Jansen, T. l. C.; Knoester, J. A transferable electrostatic map for solvation effects on amide I vibrations and its application to linear and two-dimensional spectroscopy. *J. Chem. Phys.* **2006,** *124*, 044502.
62. Jansen, T. l. C.; Dijkstra, A. G.; Watson, T. M.; Hirst, J. D.; Knoester, J. Modeling the amide I bands of small peptides. *J. Chem. Phys.* **2006,** *125*, 044312.

Appendix to Chapter 6

Computational methods

Molecular dynamics (MD) simulation of protein

$LK_7\beta$ (Ac-LKLKLKL-NH_2)[1-4] and $LE_7\beta$ (Ac-LELELEL-NH_2)[4] were modeled as five-stranded antiparallel β-sheets. The strands were arranged in antiparallel form using VMD tcl scripts. The structure was solvated by at least 10 Å of TIP4P-Ew water[5] on all sides. Half of the water was removed to create an interface. The leucine residues faced toward the vacuum in both systems. The amino acid charges were then neutralized with 15 chloride ions ($LK_7\beta$) or 15 sodium ions ($LE_7\beta$). The structure was then equilibrated.

The equilibration began with solvent energy minimization followed by 500 ps of solvent equilibration with restrained protein (200 kcal $mol^{-1}Å^{-2}$). The protein hydrogen atoms were then minimized, followed by minimizing the energy of the entire system with gradually diminishing restraints on the protein backbone atoms (100 kcal $mol^{-1}Å^{-2}$, 50 kcal $mol^{-1}Å^{-2}$, 10 kcal $mol^{-1}Å^{-2}$). The entire structure was then minimized. Then, the system was heated over ~360 ps from 0 K to 298 K by performing molecular dynamics (MD) under constant volume. This was followed by 5 ns of MD equilibration in the NVT ensemble. NPT equilibration was not performed to avoid collapsing the vacuum region and damaging the protein structure.

The production MD trajectory was propagated in the NVT ensemble for 100 ns, saving configurations every 10 fs for a total of 10,000,000 frames. Hydrogen-containing bonds in the protein were kept rigid with the SHAKE algorithm,[6] and water molecules were kept rigid with the analytical SETTLE algorithm.[7] Particle mesh Ewald summation[8] was

used to calculate long-range electrostatics, and a nonbonded cutoff of 14 Å was used with a switching function that began at 12.6 Å. OpenMM 7.6 tools[9] were used to conduct the equilibration and production simulations on GPUs on a Yale University cluster. Trajectories were wrapped to ensure the slab was in one piece. This is necessary for the new generalized implementation of the Voronoi tessellation analysis and was not required in our previous studies. A Langevin integrator was used for all MD with a temperature of 298 K, a friction coefficient of 1 ps^{-1}, and a timestep of 1 fs.

MD simulation of the air-water interface

Interfacial water simulations were performed twice using both the TIP4P-Ew[5] and SPC/E[10] models. The systems contained roughly 560 water molecules each. For each simulation, the equilibration consisted of energy minimization, heating over 360 ps using Berendsen pressure coupling[11] (1 atm), an NPT equilibration for 1 ns, and an NVT equilibration for 1 ns. The *z*-dimension was then increased from ~32 Å to 100 Å to create a slab with two interfaces. The resulting system was equilibrated in the NVT ensemble for 1 ns. This was followed by a 100 ns production trajectory, saving configurations every 10 fs. A nonbonded cutoff of 8 Å was used because the systems were small. All other aspects of the equilibration and simulations were similar to the protocol for proteins described above, except the simulations were performed using Amber 20[12] and the systems were built with *tleap*.

Voronoi tessellation and neighbor identification

For each configuration of the system, a Voronoi tessellation was created using the freud library,[13] which uses the voro++ code[14] as a backend. Periodic boundary conditions were considered. For performance, Python sets were used throughout when searching lists for values. The Voronoi tessellation was performed only once per configuration regardless of subsequent analyses. Two layers of virtual points above the top surface of the slab were added to the end of the list of atom positions before calculation of the Voronoi tessellation. This ensured that Voronoi tessellation did not accidentally select atoms from the bottom of the slab as neighbors of atoms on the top surface due to periodic boundary conditions. The virtual points were first laid out on a 1 Å grid above the top surface, and then the positions were shifted randomly (~0.1 Å) in each direction to prevent point regularity, which can crash voro++. All selections were conducted by residue – if one atom of a water molecule was a neighbor, the entire molecule was selected.

Layer-by-layer breakdown of the air-water interface

Virtual points were placed above the top surface, and the neighbors of the bottom virtual point layer were considered as layer 1. The not-yet-selected neighbors of layer 1 were considered layer 2, and so on. All selections were conducted by residue as above.

Hydrogen bonding analysis

Hydrogen bonds between proteins and water and within proteins were analyzed using the MDAnalysis HydrogenBondAnalysis tool.[15] Hydrogen bonds between layers at the air-water interface were analyzed manually. A donor-acceptor distance cutoff of <3.2

Å and an angle cutoff of >135° were used for all hydrogen bond analyses. A total of 10,000 frames over 100 ns of simulation were analyzed in all cases.

Calculation of SFG spectra

Chiral SFG spectra were calculated as previously described using the Skinner group's OH stretch electrostatic map[16-19] and the inhomogeneous limit approximation. Achiral SFG spectra were calculated in a similar way. No damping of the bottom surface signal[20] was needed for the achiral SFG calculation because the bottom surface was not considered in the Voronoi tessellation selections. In calculations that include intermolecular couplings, the solvent outside the first hydration shell was treated as point charges influencing the electric field experienced by the protein and hydration shell. (Note that the protein and hydration shell themselves also influence the electric field.) From the electric field we obtained spectroscopic quantities as described in earlier studies.[2-4, 21] All other spectra included only intramolecular couplings. A total of 10,000,000 frames over 100 ns of simulation were analyzed.

Voronoi tessellation code

Voronoi tessellation and by-residue neighbor analysis are facilitated by the Python class VoronoiSelection, which requires the NumPy, MDAnalysis, and freud libraries. Use of the class ensures that the Voronoi diagram is only calculated once per configuration and offers a simple interface for calculating layers, hydration shells, and so forth. The code and an example script can be found at https://github.com/dkonstan/VoronoiSelection.

References

1. DeGrado, W. F.; Lear, J. D. Induction of peptide conformation at apolar water interfaces. 1. A study with model peptides of defined hydrophobic periodicity. *J. Am. Chem. Soc.* **1985,** *107*, 7684-7689.
2. Perets, E. A.; Konstantinovsky, D.; Fu, L.; Chen, J.; Wang, H.-F.; Hammes-Schiffer, S.; Yan, E. C. Y. Mirror-image antiparallel β-sheets organize water molecules into superstructures of opposite chirality. *Proc. Natl. Acad. Sci. USA* **2020,** *117*, 32902-32909.
3. Konstantinovsky, D.; Perets, E. A.; Santiago, T.; Velarde, L.; Hammes-Schiffer, S.; Yan, E. C. Y. Detecting the First Hydration Shell Structure around Biomolecules at Interfaces. *ACS Cent. Sci.* **2022,** *8*, 1404-1414.
4. Konstantinovsky, D.; Perets, E. A.; Santiago, T.; Olesen, K.; Wang, Z.; Soudackov, A. V.; Yan, E. C. Y.; Hammes-Schiffer, S. Design of an Electrostatic Frequency Map for the NH Stretch of the Protein Backbone and Application to Chiral Sum Frequency Generation Spectroscopy. *J. Phys. Chem. B* **2023**.
5. Horn, H. W.; Swope, W. C.; Pitera, J. W.; Madura, J. D.; Dick, T. J.; Hura, G. L.; Head-Gordon, T. Development of an improved four-site water model for biomolecular simulations: TIP4P-Ew. *J. Chem. Phys.* **2004,** *120*, 9665-9678.
6. Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C. 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.
7. Miyamoto, S.; Kollman, P. A. Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models. *J. Comput. Chem.* **1992,** *13*, 952-962.
8. Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. *J. Chem. Phys.* **1993,** *98*, 10089-10092.
9. Eastman, P.; Swails, J.; Chodera, J. D.; McGibbon, R. T.; Zhao, Y.; Beauchamp, K. A.; Wang, L.-P.; Simmonett, A. C.; Harrigan, M. P.; Stern, C. D., et al. OpenMM 7: Rapid development of high performance algorithms for molecular dynamics. *PLoS Comput. Biol.* **2017,** *13*, e1005659.
10. Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. The missing term in effective pair potentials. *J. Phys. Chem. A* **1987,** *91*, 6269-6271.
11. Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **1984,** *81*, 3684-3690.
12. University of California, S. F., Amber20. D.A. Case, H. M. A., K. Belfon, I.Y. Ben-Shalom, S.R. Brozell, D.S. Cerutti, T.E. Cheatham, III, G.A. Cisneros, V.W.D. Cruzeiro, T.A. Darden, R.E. Duke, G. Giambasu, M.K. Gilson, H. Gohlke, A.W. Goetz, R. Harris, S. Izadi, S.A. Izmailov, C. Jin, K. Kasavajhala, M.C. Kaymak, E. King, A. Kovalenko, T. Kurtzman, T.S. Lee, S. LeGrand, P. Li, C. Lin, J. Liu, T. Luchko, R. Luo, M. Machado, V. Man, M. Manathunga, K.M. Merz, Y. Miao, O. Mikhailovskii, G. Monard, H. Nguyen, K.A. O'Hearn, A. Onufriev, F. Pan, S. Pantano, R. Qi, A. Rahnamoun, D.R. Roe, A. Roitberg, C. Sagui, S. Schott-Verdugo, J. Shen, C.L. Simmerling, N.R. Skrynnikov, J. Smith, J. Swails, R.C. Walker, J. Wang, H. Wei, R.M. Wolf, X. Wu, Y. Xue, D.M. York, S. Zhao, and P.A. Kollman, Ed. University of California, San Francisco., 2021.

13. 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.
14. Rycroft, C. H. VORO++: A three-dimensional Voronoi cell library in C++. *Chaos* **2009,** *19*, 041111.
15. Michaud-Agrawal, N.; Denning, E. J.; Woolf, T. B.; Beckstein, O. MDAnalysis: A toolkit for the analysis of molecular dynamics simulations. *J. Comput. Chem.* **2011,** *32*, 2319-2327.
16. Corcelli, S. A.; Lawrence, C. P.; Skinner, J. L. Combined electronic structure/molecular dynamics approach for ultrafast infrared spectroscopy of dilute HOD in liquid H2O and D2O. *J. Chem. Phys.* **2004,** *120*, 8107-8117.
17. Auer, B. M.; Skinner, J. L. IR and Raman spectra of liquid water: Theory and interpretation. *J. Chem. Phys.* **2008,** *128*, 224511.
18. Auer, B. M.; Skinner, J. L. Vibrational Sum-Frequency Spectroscopy of the Water Liquid/Vapor Interface. *J. Phys. Chem. B* **2009,** *113*, 4125-4130.
19. Pieniazek, P. A.; Tainter, C. J.; Skinner, J. L. Surface of Liquid Water: Three-Body Interactions and Vibrational Sum-Frequency Spectroscopy. *J. Am. Chem. Soc.* **2011,** *133*, 10360-10363.
20. Pieniazek, P. A.; Tainter, C. J.; Skinner, J. L. Interpretation of the water surface vibrational sum-frequency spectrum. *J. Chem. Phys.* **2011,** *135*, 044701.
21. Konstantinovsky, D.; Perets, E. A.; Yan, E. C. Y.; Hammes-Schiffer, S. Simulation of the Chiral Sum Frequency Generation Response of Supramolecular Structures Requires Vibrational Couplings. *J. Phys. Chem. B* **2021,** *125*, 12072–12081.