UNIVERSITY OF AMSTERDAM

UvA-DARE (Digital Academic Repository)

# Membrane Structure of Aquaporin Observed with Combined Experimental and Theoretical Sum Frequency Generation Spectroscopy

Schmüser, L.; Trefz, M.; Roeters, S.J.; Beckner, W.; Pfaendtner, J.; Otzen, D.; Woutersen, S.; Bonn, M.; Schneider, D.; Weidner, T.

**DOI**
10.1021/acs.langmuir.1c02206

**Publication date**
2021

**Document Version**
Final published version

**Published in**
Langmuir

**License**
Article 25fa Dutch Copyright Act (https://www.openaccess.nl/en/policies/open-access-in-dutch-copyright-law-taverne-amendment)

Link to publication

**Citation for published version (APA):**
Schmüser, L., Trefz, M., Roeters, S. J., Beckner, W., Pfaendtner, J., Otzen, D., Woutersen, S., Bonn, M., Schneider, D., & Weidner, T. (2021). Membrane Structure of Aquaporin Observed with Combined Experimental and Theoretical Sum Frequency Generation Spectroscopy. *Langmuir*, *37*(45), 13452-13459. https://doi.org/10.1021/acs.langmuir.1c02206

**General rights**
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

**Disclaimer/Complaints regulations**
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: https://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, P.O. Box 19185, 1000 GD Amsterdam, The Netherlands. You will be contacted as soon as possible.

*UvA-DARE is a service provided by the library of the University of Amsterdam (https://dare.uva.nl)*

Download date:11 10月 2026

LANGMUIR

pubs.acs.org/Langmuir Article

# Membrane Structure of Aquaporin Observed with Combined Experimental and Theoretical Sum Frequency Generation Spectroscopy

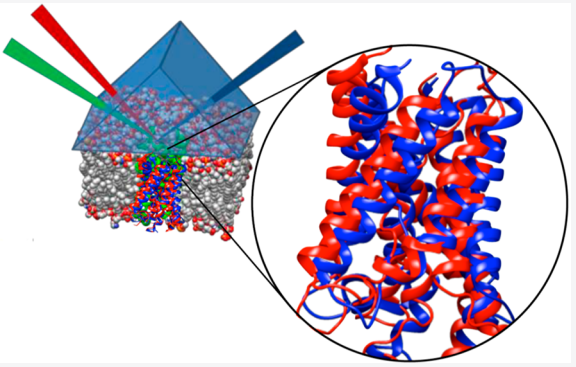
L. Schmüser, M. Trefz, S. J. Roeters, W. Beckner, J. Pfaendtner, D. Otzen, S. Woutersen, M. Bonn, D. Schneider, and T. Weidner*

Cite This: *Langmuir* 2021, 37, 13452−13459

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

**ABSTRACT:** High-resolution structural information on membrane proteins is essential for understanding cell biology and for the structure-based design of new medical drugs and drug delivery strategies. X-ray diffraction (XRD) can provide angstrom-level information about the structure of membrane proteins, yet for XRD experiments, proteins are removed from their native membrane environment, chemically stabilized, and crystallized, all of which can compromise the conformation. Here, we describe how a combination of surface-sensitive vibrational spectroscopy and molecular dynamics simulations can account for the native membrane environment. We observe the structure of a glycerol facilitator channel (GlpF), an aquaporin membrane channel finely tuned to selectively transport water and glycerol molecules across the membrane barrier. We find subtle but significant differences between the XRD structure and the inferred *in situ* structure of GlpF.



Downloaded via UNIV AMSTERDAM on April 17, 2022 at 09:06:35 (UTC).
See https://pubs.acs.org/sharingguidelines for options on how to legitimately share published articles.

## 1. INTRODUCTION

Technologies for elucidating the atomic-level structure of membrane proteins are essential for advancing our understanding of cell biology: knowledge of membrane protein structure provides information about the transport of molecules across the membrane barrier, cell sensing, and communication and also helps to engineer proteins with designed functions.[1]

By far the most successful methods for experimentally solving membrane protein structures with atomic resolution are nuclear magnetic resonance (NMR),[2] cryo electron microscopy (cryo EM),[3] and X-ray diffraction (XRD).[4] NMR can provide high-resolution data for proteins when incorporated into lipid vesicles and lipid stacks but is limited to smaller (typically about 20 kDa[5]) proteins and peptides. Cryo EM, on the other hand, is restricted to larger proteins with a lower size limit of 40−50 kDa. X-ray crystallography can solve the structure of very large proteins with angstrom resolution but requires high-quality crystals in which proteins are removed from their native and hydrated membrane environment and artificially stabilized. Molecular dynamics (MD) simulations can take the XRD-determined structures as starting points and compensate for the nonphysiological environments within crystals by relaxing the protein structure within a more "native" biological environment.[6] The caveat with simulations has been that the obtained structures have not been directly experimentally verified. Feedback from experimental data would be important not only to test the validity of the obtained results but also to provide feedback to improve force fields, water models, and sampling methods.

Vibrational sum frequency generation (SFG) spectroscopy is an ideal tool for directly testing molecular simulations with experimental data. SFG is an inherently surface-sensitive and noninvasive method, allowing the study of membrane proteins in a hydrated membrane environment.[7−9] The SFG signal depends strongly on both the protein structure and orientation. For an SFG experiment, infrared and visible laser beams are overlapped in time and space at an interface to produce sum-frequency photons of these two incident beams by nonlinear optical frequency mixing.[10] Vibrational modes in resonance with the IR beam will give rise to an enhanced signal and lead to distinct spectral features. The selection rules of SFG dictate that the signal is generated only when inversion symmetry is broken. Since this is always the case at an interface, SFG is in most cases surface-specific. Vibrations of ordered species at interfaces contribute most strongly to SFG

Received: August 18, 2021
Revised: October 8, 2021
Published: November 3, 2021

© 2021 American Chemical Society

https://doi.org/10.1021/acs.langmuir.1c02206
*Langmuir* 2021, 37, 13452−13459

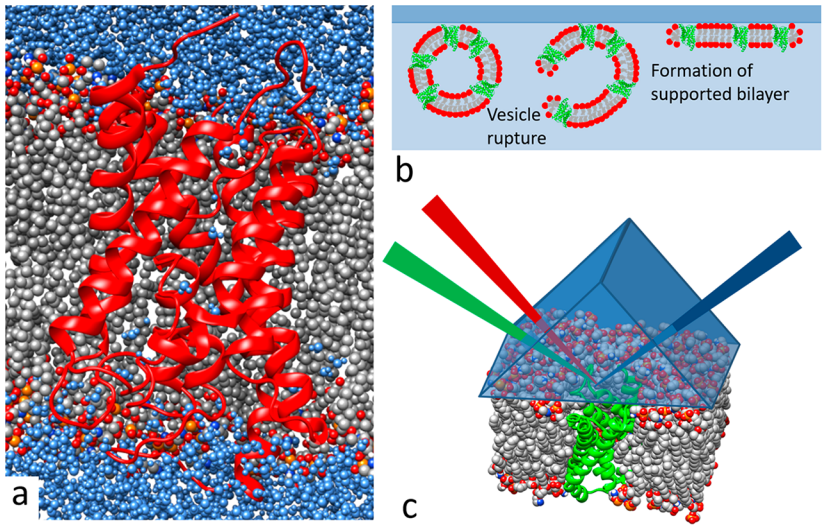



**Figure 1.** (a) Glycerol facilitator GlpF (PDB 1lda, red) embedded in a lipid bilayer (gray) with surrounding water molecules (blue), (b) model of the formation of a supported lipid bilayer using liposome spreading, and (c) experimental setup for the SFG experiments. The lipid bilayer was formed at the solid−water interface of an equilateral $CaF_2$ prism.

spectra. The surface specificity is a distinct advantage over linear spectroscopies such as IR and Raman, where the sampling depth is much greater and also modes of unbound molecules in solution close to the surface are recorded.

SFG spectra in the amide-I region allow the analysis of the secondary and tertiary structure and orientation of proteins.[11−16] Because of the nonlinear nature of SFG, spectral contributions of individual folding motifs and orientations within SFG spectra will interfere, resulting in complex spectra. Even though line shapes within SFG amide-I spectra are often broad, interference between neighboring modes still leads to spectral shapes, which cannot be analyzed by inspection. In addition, the orientation of the protein as a whole can significantly affect the spectra and the relative intensity of the different resonances, which means that spectral changes by orientation and by conformation are entangled, complicating direct spectral interpretation.

The challenge to "disentangle" and recover the rich structural information contained in the spectra has been a major challenge in the SFG community for years and has recently led to protocols to calculate theoretical SFG spectra from structure files, which can be used to interpret SFG data of interfacial proteins.[16−19] Calculated SFG spectra of proteins and peptides have shown good agreement with experimental SFG spectra.[20−27] However, the structural analysis of a protein the size of a membrane protein has not been reported.

Here, we take that step and observe the secondary structure of the large membrane protein aquaporin within a membrane environment by combining calculated and experimental spectra with MD simulation. In other words, we obtain the aquaporin *in situ* structure by experimentally testing and verifying MD simulations.

We chose an aquaporin model system because these membrane channels form an important family of membrane proteins. Water is a major component of life, and aquaporins are involved in the osmoregulation of a variety of tissues in all living organisms.[28] Besides facilitating transmembrane water flux, the subfamily of aquaglyceroporins additionally facilitates the flux of glycerol and other small, polar solutes. Aquaporins are found in all domains of life. For example, 13 aquaporins are expressed in humans, where they regulate the water homeostasis according to the individual requirements of the organs and cells.[29] On the basis of their important roles in cell biology, they have great potential in diagnostics and therapeutics.[30] All aquaporins form oligomers composed of four identical chains, each with one central channel.

The structure of aquaglyceroporins indicates that water (or glycerol) molecules pass through the protein channel in single file. The aquaporin shown in Figure 1a (PDB 1lda crystallized in 28% (w/v) poly(ethylene glycol) 2000, 100 mM bicine, 15% (v/v) glycerol, 35 mM *n*-octyl-*β*-D-glucoside, 300 mM $MgCl_2$, and 5 mM dithiothreitol (DTT) (pH 8.9))[31] is the well-studied glycerol facilitator channel (GlpF) of *E. coli*. A recent very high resolution study[32] revealed how sensitively aquaporin function depends on structural fidelity: the exchange of only two amino acids at the entry to the water channel plays the deciding role in ensuring that water and not hydronium or hydroxyl ions can pass the channel.[32] In our study, we now observe how the XRD-determined structure of the GlpF aquaglyceroporin shown in Figure 1a changes when exposed to a lipid environment.[33]

## 2. METHODS

**2.1. Expression and Purification of GlpF into Proteoliposomes.** GlpF expression and purification were performed as described in detail in ref 34. The concentration of purified GlpF was determined by absorption measurements at 280 nm using a calculated extinction coefficient of 37 930 $M^{-1}$ $cm^{-1}$ (online ExPasy ProtParam tool).

Liposomes prepared from *E. coli* polar lipid extract (EPL: 67% PE, 23.2% PG, and 9.8% CL) or 1,2-dioleoyl-*sn*-glycero-3-phosphocholine (DOPC) were used for GlpF reconstitution.[35] Chloroform lipid solutions were purchased from Avanti Polar Lipids (Alabaster, AL). To remove the organic solvent, 5 mM dissolved lipid was set under a stream of nitrogen gas. Remaining organic solvent was removed by overnight vacuum desiccation. The resulting lipid film was rehydrated in 30 mM *n*-octyl *β*-D-glycopyranoside (OG; Roth, Karlsruhe, Germany), 50 mM MOPS pH 7.5 (Sigma-Aldrich, Munich, Germany), 150 mM *N*-methyl D-glucamine (Acros Organics, Morris Plains, NJ), and 50 mM NaCl (Roth, Karlsruhe, Germany) at 37 °C for 45 min. Purified GlpF was added to the rehydrated lipids to reach a final protein concentration of 12 *μ*M and a molar lipid/GlpF ratio of


https://doi.org/10.1021/acs.langmuir.1c02206
Langmuir 2021, 37, 13452−13459

400:1. The final sample was adjusted to a volume of 0.5 mL and an OG concentration of 30 mM by the addition of buffer (50 mM MOPS, pH 7.5, 150 mM *N*-methyl D-glucamine, 50 mM NaCl). Subsequently, the sample was dialyzed for 48 h at 4 °C against 500 mL of MOPS buffer. The dialysis buffer was exchanged three times.

**2.2. SFG Experiment.** The SFG setup has been described in detail before.[36] Briefly a 40 fs, 5 mJ, 800 nm visible pulse was generated by a regenerative amplifier (Spitfire ACE, Spectra Physics) using a Nd:YLF pulse laser (Empower, Spectra Physics) and a Ti:sapphire seed laser (MaiTai, Spectra Physics). One part of the 800 nm beam was branched out to pump an optical parametric amplifier (TOPAS, Spectra Physics), which generates a broadband IR pulse. The remaining 800 nm beam was spectrally narrowed to a 25 $\mu$J 15 $cm^{-1}$ fwhm pulse using a Fabry–Perot Etalon (SLS Optics Ltd.) and temporally and spatially overlapped with the IR pulse. IR and visible laser pulses were both focused on the sample. Laser polarization combinations SSP (S-polarized SFG, S-polarized visible, and P-polarized IR) and PPP were obtained using polarizers and half-wave plates in each beam path. The generated SFG signal was collimated using lenses and separated from the visible light using low-pass filters. The focused SFG signal was directed onto a spectrograph (Acton Instruments) and finally detected by a CCD camera (Newton, Andor Technologies).

The SFG experiments were performed at 20 °C in a nitrogen-flushed chamber to avoid absorption of the IR pulse due to water vapor. All reported spectra have been recorded and reproduced at least three times. The experiments were performed in a flow cell with a volume of 1 mL. The flow cell was sealed on one side with an equilateral $CaF_2$ prism. Following established protocols for surface vesicle fusion into solid-supported bilayers,[37] 400 $\mu$L of the proteoliposome solution was injected into the flow cell and incubated for at least 2 h. Remaining proteoliposomes in the bulk were rinsed with $D_2O$. According to previous studies, lipid bilayers prepared from lipid vesicle fusion are separated from the solid support by a 1 nm layer of water, which protects inserted proteins from direct contact with the calcium fluoride surface.[38] To ensure that the interfacial water is exchanged with $D_2O$, the sample is allowed to undergo deuterium exchange overnight in $D_2O$.

SFG spectra were collected in an SSP and PPP polarization combination and normalized using reference spectra of a $CaF_2$ prism, which was coated with a 100 nm silver film at the $CaF_2$ water interface.

**2.3. Theory of the Calculations.** First, we obtain the atom coordinates of the backbone amide groups from a PDB file based on the crystal structure or a frame of the MD simulation. We then use the coordinates to calculate the transition-dipole moments of the local modes by determining the transition charge of each atom.[39] This calculation is based on the difference in the charge density for the stretched and compressed amide-I mode that can be determined with the *Gaussian* electronic-structure software package. The local-mode Raman tensors are obtained from ref 40, the magnitudes and orientations of which are corroborated by ref 41. The next step is to use the atomic coordinates to construct the one-exciton Hamiltonian[42]

$$\begin{pmatrix} \hbar\omega_1^0 & \kappa_{12} & \kappa_{13} & \kappa_{14} & \cdots \\ \kappa_{12} & \hbar\omega_2^0 & \kappa_{23} & \kappa_{24} & \cdots \\ \kappa_{13} & \kappa_{23} & \hbar\omega_3^0 & \kappa_{34} & \cdots \\ \kappa_{14} & \kappa_{24} & \kappa_{34} & \hbar\omega_4^0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix} \quad (1)$$

with $\omega_i^0$ being the gas-phase frequency of local mode $i$ and $k_{ij}$ being the coupling between local modes $i$ and $j$.

The diagonal terms are determined using an empirical model that gives the local-mode frequency as a function of the strength of the three possible hydrogen bonds that each amide group can form,[17] comparable to the model used in ref 43.

For the off-diagonal terms and the couplings between local modes, we discriminate between nearest-neighbor and non-nearest-neighbor coupling. As the former is dominated by through-bond couplings,[42] we use a parametrization of the coupling as a function of the two dihedral angles between the two neighboring amide groups calculated for the "glycine dipeptide" (Ac-Gly-NHCH3), using the 6-31G+(d) basis set and B3LYP functional.[44]

The non-nearest-neighbor couplings are dominated by through-space (Coulomb) interactions, so we estimate these with the transition dipole coupling method[45]

$$K_{ij} = \frac{1}{4\pi\varepsilon_0}\left(\frac{\vec{\mu}_i\cdot\vec{\mu}_j}{|\vec{r}_{ij}|^3} - 3\frac{(\vec{r}_{ij}\cdot\vec{\mu}_i)(\vec{r}_{ij}\cdot\vec{\mu}_j)}{|\vec{r}_{ij}|^5}\right) \quad (2)$$

with $\vec{\mu}_i$ being the transition dipole moment of local mode $i$, $\vec{r}_{ij}$ being the distance between local modes $i$ and $j$, and $\varepsilon_0$ being the dielectric constant.

Subsequently, the Hamiltonian is diagonalized to obtain the normal mode eigenvalues and eigenvectors, from which the IR, Raman, and VSFG responses are calculated, according to ref 17 in which other details regarding the formalism used here for the spectral calculations can also be found.

The nonresonant phase and its amplitude were adapted to yield the best match between the experimental data and the calculated ones. Otherwise, all calculations were made using the same settings (see SI).

**2.4. MD Simulation.** Simulations of the porin tetramer (1lda[31]) and the phosphatidylethanolamine (POPE) lipid bilayer were performed using the GROMACS[46] molecular dynamics (MD) engine and explicitly solvated using the TIP3P[47] water model. Force fields for the lipid bilayer were taken from Tieleman and Berendsen[48] and adapted into the GROMOS96 53A6 force field,[49] which is extended to include Berger lipid parameters and has been verified to perform as well as or better than previous versions for protein simulations.[50] Time steps of 2.0 fs in the MD simulation were integrated using a leapfrog algorithm.[51] For distances exceeding 1.2 nm, van der Waals interactions were shifted to 0 with a switching function applied at 1.0 nm, and electrostatic forces were treated with a particle mesh Ewald (PME) summation. Bond lengths between hydrogen and heavy atoms were fixed with the LINCS linear constraint solver.[52]

The tetramer and lipid bilayer were centered in an overall charge-neutral system with periodic $x$, $y$, and $z$ dimensions of 11.49, 11.38, and 10.39 nm, respectively. The system was restricted from lateral diffusion of the membranes by restraining the relative motion of the protein and bilayer to the solvent. Following energy minimization, a 100 ps NVT simulation was conducted at a temperature above the phase-transition temperature ($T$ = 298 K) of the lipid membrane[53] to allow the equilibration of water and ions. Protein, lipid, and solvent/ions were temperature coupled independently at 315 K using a stochastic global thermostat[54] and a 0.1 ps coupling constant.

A 1 ns NPT equilibration step was conducted after NVT equilibration. The thermostat was switched to Nose-Hoover[55] with a 0.4 ps coupling constant to more realistically capture temperature fluctuations.[56] Semi-isotropic pressure coupling was used to allow the membrane to deform in the $xy$ plane independently of $z$. Following NVT and NPT equilibration, position restraints on the tetramer and lipid bilayer were relaxed and the system underwent 100 ns of production MD in the NPT ensemble. Coordinates of the resulting structure were used for SFG analysis from a time point of 100 ns.

## 3. RESULTS AND DISCUSSION

We spread GlpF-loaded proteoliposomes, prepared with *E. coli* polar lipid extract (SI), on one side of a $CaF_2$ prism (Figure 1b). The liposomes spread at the surface to form a supported lipid bilayer, which was then probed through the back side of the prism with SFG in near total internal reflection geometry (Figure 1c). Representative experimental spectra in SSP (S-polarized SFG, S-polarized visible, and P-polarized infrared) and PPP polarizations are shown in Figure 2. The observed


https://doi.org/10.1021/acs.langmuir.1c02206
Langmuir 2021, 37, 13452−13459

spectra are typical for the convoluted, complex signal expected from large protein systems.

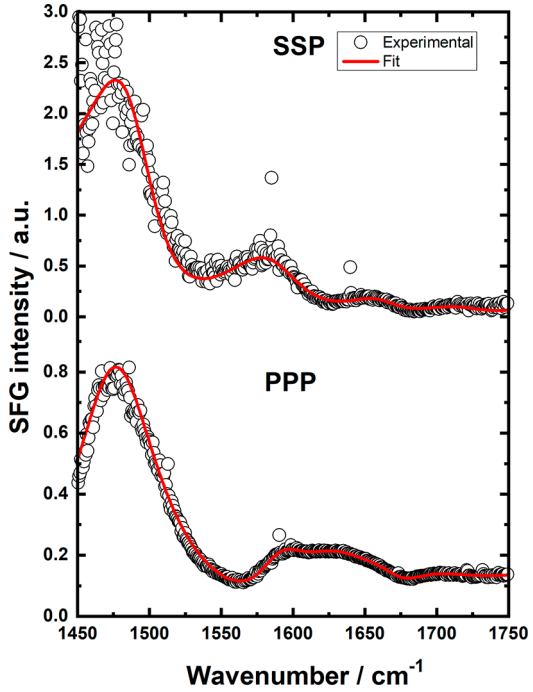



**Figure 2.** Experimental SFG spectra of a GlpF incorporated in a lipid bilayer, which was formed by proteoliposomes on the solid water interface of a $CaF_2$ prism mounted on a flow cell. The two different laser polarization combinations, SSP and PPP, have been measured. The spectra have been fitted (red line) using a Lorentzian line shape model (eqs 3 and 4).

While an assignment of all modes is difficult without calculations, we can identify several major contributors to the spectra: very strong peaks near 1480 $cm^{-1}$ can be attributed to $CH_2$ and $CH_3$ scissoring modes of both lipids and proteins.[57] A feature near 1650 $cm^{-1}$ is representative of the $\alpha$-helical structure of the membrane protein.[57] In addition, there is a smaller peak near 1720 $cm^{-1}$ that originates from ester groups of membrane lipids. The feature observed near 1600 $cm^{-1}$ is not well described in the literature and may represent a new spectral feature. We speculate that this mode is typically overwhelmed by helical and sheet-type modes in IR and Raman spectra but becomes visible in the SFG spectrum because of significant cancelation of the antiparallel helix and turn sequences.

To directly relate the SFG measurements to protein structure, we first calculate the SFG spectra from GlpF structure files obtained with XRD from protein crystals (PDB 1lda).[31] The calculation followed the procedure described by Roeters et al.[17] The model takes into account all amide-I vibrations along with the peptide backbone. The theoretical spectra for the protein crystals are shown in Figure 3 (purple dashed line). Since only backbone-related modes are calculated, it is important to compare the theoretical spectra directly with the experimental SFG backbone modes. Ideally also side-chain modes would be included in the analysis, and several side chains can have signals close enough to the amide-I region to affect the spectral shape. At the same time, SFG arises from ordered IR- and Raman-active modes. While including side-chain modes will be an important step in future

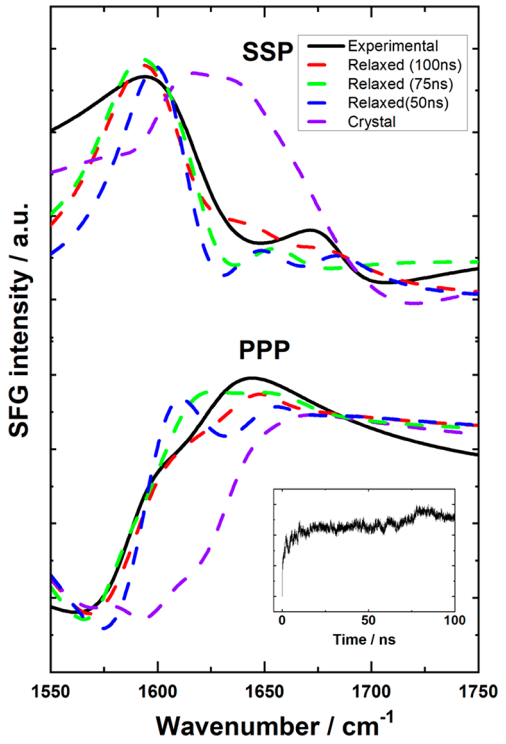



**Figure 3.** Comparison of protein-specific SFG spectra extracted from fits of experimental data (solid lines) and calculated SFG spectra of the MD simulation at different time points (dashed lines). The inset represents the root-mean-square deviation of atomic positions plotted against the simulation time showing that a stable conformation was already reached after a few nanoseconds.

improvements of the spectral calculations, typically the side-chain modes are less aligned and are distributed all over the protein in different orientations. It is conceivable that in a large protein such as GlpF the arrangement would be quasicentrosymmetric and would result in a relatively weak SFG signal.

We fit the SFG spectra to determine the contributions of the different modes and focus only on the protein-related peaks. The spectrum can be fitted using the following standard equation describing the SFG response from an interface:[58]

$$I_{SFG}(\omega_{IR}) \propto \left|(\chi^{(2)}\omega_{IR})\right|^2 I_{vis} I_{IR} \tag{3}$$

$$\chi^{(2)}\omega_{IR} = A_0 e^{i\varphi} + \sum_n \frac{A_n}{\omega_{IR} - \omega_n + i\Gamma_n} \tag{4}$$

Here, $I_{SFG}$ $(\omega_{IR})$, $I_{vis}$ and $I_{IR}$, $\chi^{(2)}$ $\omega_{IR}$, $A_0$, $\varphi$, $\omega_n$ and $i\Gamma_n$ are the SFG intensity, the intensities of IR and visible laser beams, the second-order susceptibility, the amplitude and phase of the nonresonant background and the frequency, and full width at half-maximum of the $n$th vibrational mode, respectively.

Using the fit results (parameters in the SI), we plot all protein backbone amide-I components of the fit spectrum in Figure 3 (solid lines): the modes near 1600 and 1650 $cm^{-1}$ while excluding any non-amide-I modes and side-chain modes from the new amide-I-only spectrum. In SSP polarization, the mode near 1600 $cm^{-1}$ dominates the derived spectra while the feature near 1650 $cm^{-1}$ is very strong in the PPP data.

The calculated spectra based on the XRD structure predict the intensity near 1650 $cm^{-1}$ for both SSP and PPP polarization combinations. However, the feature near 1600 $cm^{-1}$ is not captured correctly by the calculations. The


https://doi.org/10.1021/acs.langmuir.1c02206
Langmuir 2021, 37, 13452−13459

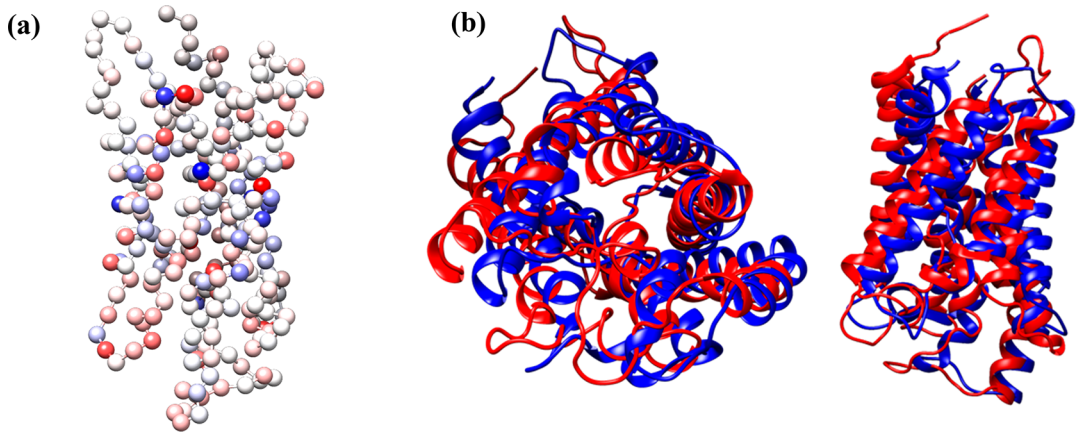



**Figure 4.** (a) Averaged contributions of amide-I local modes to the 5 eigenmodes closest to 1600 $cm^{-1}$. The distribution of the contributions over the entire sequence shows that the origin of the main peak at 1600 $cm^{-1}$ is delocalized over the whole protein and is not the result of a local feature. Blue and red spheres indicate the magnitude of the contributions with a negative (blue) or positive (red) amplitude. The amplitudes for each eigenmode are derived from the SFG spectral calculations. (b) Structural differences between the crystal structure (PDB 1lda, blue) and the refined protein structure (red). The refined structure was obtained from an MD simulation after 100 ns and directly validated by SFG.

discrepancy between the calculated and observed spectra indicates that the structure of GlpF within the stabilized protein crystal used for XRD differs somewhat from its structure when incorporated in a lipid bilayer.

This discrepancy between the observed and expected SFG response can be resolved when going from the crystalline state represented by the XRD structure to the native lipid membrane state. Using the XRD structure as the starting point, we ran MD simulations of GlpF within a 1-palmitoyl-2-oleoyl-*sn*-glycero-3-phosphoethanolamine (POPE) bilayer until the structure was equilibrated (~100 ns, Figure 3). Indeed, subtle changes in the structure were apparent (Figure 4).

Spectra calculated from this "relaxed" structure yielded theoretical spectra with a substantially improved match to the measured SFG spectra (Figures 3 and S3, red dashed line). The spectra based on the simulated model capture the spectral shape of the experimental data very well. The 1600 $cm^{-1}$ peak and also the intensity at higher energies now match much better compared to the spectrum calculated using the crystal structure. The root-mean-square deviation (RMSD) of the computed spectra from the experimental data decreased by 55% for the relaxed structures compared with the crystal structure (SI). The RMSD values between relaxed structures taken at different time points in the simulation varied only by approximately 5−10% (Figure S3). Clearly, the *in silico* hydration and relaxation within a membrane environment capture the experimental data significantly better than does the original crystal structure.

Note that, apart from an overall amplitude scaling factor, there are no adjustable parameters when comparing the experimental and theoretical results.

To test our approach to remove the non-amide-I modes from the spectra and then calculate the amide-I modes, we also used a second, slightly more complex approach by including the non-amide-I modes as fitted peaks directly in the calculations. Details of this approach can be found in the SI. We found that the conclusion remains that with increasing simulation time the mismatch between theory and experiment decreases (Figures S4 and S5). This clearly shows that the protein state simulated for a properly hydrated environment matches the experimental data considerably better compared with the protein crystal structure.

The following question arises: is the improved spectral match related to the more native global tertiary structure within the membrane, or are the spectral changes dominated by changes in local features, such as the flexible N- and C-termini and the relatively mobile loop structures? This is particularly important to answer here since the 1600 $cm^{-1}$ feature we see in the spectra has not previously been described, and it should be excluded that the mode comes from a localized artifact within the protein sequence. To answer this question, we have performed a normal-mode analysis in which the contributions of the local modes to the normal modes (given by the eigenvectors, obtained by diagonalizing the Hamiltonian) are revealed. In Figure 4(a), the phase (blue = negative, red = positive) and magnitude of each of these contributions is mapped onto the protein sequence. Performing this analysis demonstrates that the 1600 $cm^{-1}$ feature is indeed a protein mode. It is also evident from the plot that the modes contributing to the signal are not very localized but that they are evenly distributed across the protein and any changes to the calculated spectra are a result of changes within the global protein structure and are not caused by localized orientation changes in the terminal regions. A mode analysis of other protein-related signals yielded very similar results.

Figure 4b shows the crystal-based GlpF structure along with the refined structure that is consistent with the SFG results. There are marked differences between the two structures, with variations in the loop regions and particularly in the helix orientations. A detailed computational study of the impact of these structural variations on protein function is beyond the scope of this study. However, it may be worthwhile for future studies to look into the impact of the structural differences: it is well known that aquaporin action depends on minute structural details.[59] Also, the loop regions appear to be crucial for proper activity.[60] Thus, a knowledge of experimentally validated structural details in a simulated native environment could potentially provide valuable information for interpreting the changes in the substrate-conducting mechanism induced by mutations in aqua(glycero)porins and in membrane proteins in general.


https://doi.org/10.1021/acs.langmuir.1c02206
*Langmuir* 2021, 37, 13452−13459

## 4. CONCLUSIONS

Importantly, the data highlights how sensitive SFG spectra can be to fairly subtle changes in the protein structure shown in Figure 4. This is likely explained by changes in the arrangement of the helices, which are of opposite orientation and therefore create destructively interfering signals. Any change in the relative angles within the assembly of helices can have a major impact on the overall spectral shape. Infrared and Raman spectra tend to become more crowded and more difficult to interpret as proteins become larger. This study could be an indication that SFG spectra become more complex and, with a theoretical framework to analyze the spectra, more descriptive and unique as the structure becomes larger and more diverse.

The presented methods could be promising for future hybrid experimental and theoretical structural studies. Combining SFG and MD to study membrane proteins could allow the fine-tuning of existing XRD protein crystal data to include the influence of the hydrated lipid bilayer but also the impact of other biophysical parameters such as temperature, pH, and ionic strength, which can easily be adjusted in simulations and SFG experiments. Moreover, this method has the potential to determine the impact of therapeutic drugs on aquaporins *in situ*. As such, we expect the combination of MD simulations and SFG to be an asset to future studies of selective molecular transport within aquaporins and the function of a broad variety of membrane proteins.

## ■ ASSOCIATED CONTENT

### Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.langmuir.1c02206.

Further details of the SFG experiments and data analysis as well as the simulations and the spectra calculations (PDF)

## ■ AUTHOR INFORMATION


### Corresponding Author

**T. Weidner** − *Department of Chemistry, Aarhus University, 8000 Aarhus C, Denmark; Department of Chemical Engineering, University of Washington, Seattle, Washington 98195-1750, United States;* orcid.org/0000-0002-7083-7004; Email: weidner@chem.au.dk

### Authors

**L. Schmüser** − *Department of Molecular Spectroscopy, Max Planck Institute for Polymer Research, Mainz 55128, Germany; Department of Chemistry, Aarhus University, 8000 Aarhus C, Denmark*

**M. Trefz** − *Department of Chemistry−Biochemistry, University of Mainz, 55128 Mainz, Germany*

**S. J. Roeters** − *Department of Chemistry, Aarhus University, 8000 Aarhus C, Denmark; Van't Hoff Institute for Molecular Sciences, University of Amsterdam, 1098 XH Amsterdam, The Netherlands*

**W. Beckner** − *Department of Chemical Engineering, University of Washington, Seattle, Washington 98195-1750, United States*

**J. Pfaendtner** − *Department of Chemical Engineering, University of Washington, Seattle, Washington 98195-1750, United States;* orcid.org/0000-0001-6727-2957

**D. Otzen** − *iNANO, Aarhus University, 8000 Aarhus C, Denmark;* orcid.org/0000-0002-2918-8989

**S. Woutersen** − *Van't Hoff Institute for Molecular Sciences, University of Amsterdam, 1098 XH Amsterdam, The Netherlands;* orcid.org/0000-0003-4661-7738

**M. Bonn** − *Department of Molecular Spectroscopy, Max Planck Institute for Polymer Research, Mainz 55128, Germany;* orcid.org/0000-0001-6851-8453

**D. Schneider** − *Department of Chemistry−Biochemistry, University of Mainz, 55128 Mainz, Germany;* orcid.org/0000-0003-4517-6387

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.langmuir.1c02206


### Author Contributions

L.S., T.W., J.P., and D.S. designed the experiment. L.S. performed the SFG experiments. M.T. isolated and prepared the proteoliposomes. S.J.R and W.B. performed the calculations and simulation. J.P., S.W., M.B., D.O., D.S., and T.W. analyzed the data. All authors discussed the results and wrote the article.

### Notes

The authors declare no competing financial interest.

## ■ ACKNOWLEDGMENTS

This work was supported by the Max Planck Graduate Center with the Johannes Gutenberg-Universität Mainz. S.J.R. and T.W. thank the Villum Foundation (experiment grant 22956) for financial support. Partial financial support to J.P. and W.B. was provided by NSF (CBET-1264459). L.S. and S.J.R. thank the Lundbeck Foundation for postdoctoral fellowships.

## ■ REFERENCES


(1) Morsbach, S.; Gonella, G.; Mailander, V.; Wegner, S.; Wu, S.; Weidner, T.; Berger, R.; Koynov, K.; Vollmer, D.; Encinas, N.; Kuan, S. L.; Bereau, T.; Kremer, K.; Weil, T.; Bonn, M.; Butt, H. J.; Landfester, K. Engineering Proteins at Interfaces: From Complementary Characterization to Material Surfaces with Designed Functions. *Angew. Chem., Int. Ed.* **2018**, *57* (39), 12626−12648.

(2) Wuthrich, K. Protein-Structure Determination in Solution by Nmr-Spectroscopy. *J. Biol. Chem.* **1990**, *265* (36), 22059−22062.

(3) Sun, C.; Gennis, R. B. Single-particle cryo-EM studies of transmembrane proteins in SMA copolymer nanodiscs. *Chem. Phys. Lipids* **2019**, *221*, 114−119.

(4) Glaeser, R. M. Electron Crystallography of Biological Macromolecules. *Annu. Rev. Phys. Chem.* **1985**, *36*, 243−275.

(5) Frueh, D. P.; Goodrich, A. C.; Mishra, S. H.; Nichols, S. R. NMR methods for structural studies of large monomeric and multimeric proteins. *Curr. Opin. Struct. Biol.* **2013**, *23* (5), 734−9.

(6) Snow, C. D.; Nguyen, H.; Pande, V. S.; Gruebele, M. Absolute comparison of simulated and experimental protein-folding dynamics. *Nature* **2002**, *420*, 102.

(7) Nguyen, K. T.; Le Clair, S. V.; Ye, S.; Chen, Z. Molecular Interactions between Magainin 2 and Model Membranes in Situ. *J. Phys. Chem. B* **2009**, *113* (36), 12358−12363.

(8) 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* (7), 2769−2785.

(9) Volpati, D.; Aoki, P. H. B.; Alessio, P.; Pavinatto, F. J.; Miranda, P. B.; Constantino, C. J. L.; Oliveira, O. N. Vibrational spectroscopy for probing molecular-level interactions in organic films mimicking biointerfaces. *Adv. Colloid Interface Sci.* **2014**, *207*, 199−215.

(10) Zhu, X. D.; Suhr, H.; Shen, Y. R. Surface vibrational spectroscopy by infrared-visible sum frequency generation. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1987**, *35* (6), 3047−3050.

(11) Mauri, S.; Weidner, T.; Arnolds, H. The structure of insulin at the air/water interface: monomers or dimers? *Phys. Chem. Chem. Phys.* **2014**, *16* (48), 26722−26724.
(12) Wang, Z. G.; Morales-Acosta, M. D.; Li, S. H.; Liu, W.; Kanai, T.; Liu, Y. T.; Chen, Y. N.; Walker, F. J.; Ahn, C. H.; Leblanc, R. M.; Yan, E. C. Y. A narrow amide I vibrational band observed by sum frequency generation spectroscopy reveals highly ordered structures of a biofilm protein at the air/water interface (vol 52, pg 2956, 2016). *Chem. Commun.* **2016**, *52* (68), 10440−10441.
(13) Nguyen, K. T.; Soong, R.; Im, S. C.; Waskell, L.; Ramamoorthy, A.; Chen, Z. Probing the Spontaneous Membrane Insertion of a Tail-Anchored Membrane Protein by Sum Frequency Generation Spectroscopy. *J. Am. Chem. Soc.* **2010**, *132* (43), 15112−15115.
(14) Meister, K.; Lotze, S.; Olijve, L. L. C.; DeVries, A. L.; Duman, J. G.; Voets, I. K.; Bakker, H. J. Investigation of the Ice-Binding Site of an Insect Antifreeze Protein Using Sum-Frequency Generation Spectroscopy. *J. Phys. Chem. Lett.* **2015**, *6* (7), 1162−1167.
(15) 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* (18), 9947−9954.
(16) 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* (7), 3420−3465.
(17) Roeters, S. J.; van Dijk, C. N.; Torres-Knoop, A.; Backus, E. H. G.; Campen, R. K.; Bonn, M.; Woutersen, S. Determining In Situ Protein Conformation and Orientation from the Amide-I Sum-Frequency Generation Spectrum: Theory and Experiment. *J. Phys. Chem. A* **2013**, *117* (29), 6311−6322.
(18) Harrison, E. T.; Weidner, T.; Castner, D. G.; Interlandi, G. Predicting the orientation of protein G B1 on hydrophobic surfaces using Monte Carlo simulations. *Biointerphases* **2017**, *12* (2), 02D401.
(19) Weidner, T.; Castner, D. G. Developments and Ongoing Challenges for Analysis of Surface-Bound Proteins. *Annu. Rev. Anal. Chem.* **2021**, *14* (1), 389−412.
(20) Bellucci, L.; Ardevol, A.; Parrinello, M.; Lutz, H.; Lu, H.; Weidner, T.; Corni, S. The interaction with gold suppresses fiber-like conformations of the amyloid beta (16−22) peptide. *Nanoscale* **2016**, *8* (16), 8737−8748.
(21) Hennig, R.; Heidrich, J.; Saur, M.; Schmuser, L.; Roeters, S. J.; Hellmann, N.; Woutersen, S.; Bonn, M.; Weidner, T.; Markl, J.; Schneider, D. IM30 triggers membrane fusion in cyanobacteria and chloroplasts. *Nat. Commun.* **2015**, *6*, 7018.
(22) Lu, H.; Lutz, H.; Roeters, S. J.; Hood, M. A.; Schafer, A.; Munoz-Espi, R.; Berger, R.; Bonn, M.; Weidner, T. Calcium-Induced Molecular Rearrangement of Peptide Folds Enables Biomineralization of Vaterite Calcium Carbonate. *J. Am. Chem. Soc.* **2018**, *140* (8), 2793−2796.
(23) Lutz, H.; Jaeger, V.; Schmuser, L.; Bonn, M.; Pfaendtner, J.; Weidner, T. The Structure of the Diatom Silaffin Peptide R5 within Freestanding Two-Dimensional Biosilica Sheets. *Angew. Chem., Int. Ed.* **2017**, *56* (28), 8277−8280.
(24) Schmuser, L.; Roeters, S.; Lutz, H.; Woutersen, S.; Bonn, M.; Weidner, T. Determination of Absolute Orientation of Protein alpha-Helices at Interfaces Using Phase-Resolved Sum Frequency Generation Spectroscopy. *J. Phys. Chem. Lett.* **2017**, *8* (13), 3101−3105.
(25) Golbek, T. W.; Padmanarayana, M.; Roeters, S. J.; Weidner, T.; Johnson, C. P.; Baio, J. E. Otoferlin C2F Domain-Induced Changes in Membrane Structure Observed by Sum Frequency Generation. *Biophys. J.* **2019**, *117*, 1820.
(26) Lu, H.; Schafer, A.; Lutz, H.; Roeters, S. J.; Lieberwirth, I.; Munoz-Espi, R.; Hood, M. A.; Bonn, M.; Weidner, T. Peptide-Controlled Assembly of Macroscopic Calcium Oxalate Nanosheets. *J. Phys. Chem. Lett.* **2019**, *10* (9), 2170−2174.
(27) Lutz, H.; Jaeger, V.; Weidner, T.; de Groot, B. L. Interpretation of Interfacial Protein Spectra with Enhanced Molecular Simulation Ensembles. *J. Chem. Theory Comput.* **2019**, *15* (1), 698−707.
(28) King, L. S.; Kozono, D.; Agre, P. From structure to disease: the evolving tale of aquaporin biology. *Nat. Rev. Mol. Cell Biol.* **2004**, *5* (9), 687−98.
(29) Rojek, A.; Praetorius, J.; Frokiaer, J.; Nielsen, S.; Fenton, R. A. A current view of the mammalian aquaglyceroporins. *Annu. Rev. Physiol.* **2008**, *70*, 301−27.
(30) Verkman, A. S.; Anderson, M. O.; Papadopoulos, M. C. Aquaporins: important but elusive drug targets. *Nat. Rev. Drug Discovery* **2014**, *13* (4), 259−77.
(31) Tajkhorshid, E.; Nollert, P.; Jensen, M. O.; Miercke, L. J. W.; O'Connell, J.; Stroud, R. M.; Schulten, K. Control of the selectivity of the aquaporin water channel family by global orientational tuning. *Science* **2002**, *296* (5567), 525−530.
(32) Eriksson, U. K.; Fischer, G.; Friemann, R.; Enkavi, G.; Tajkhorshid, E.; Neutze, R. Subangstrom Resolution X-Ray Structure Details Aquaporin-Water Interactions. *Science* **2013**, *340* (6138), 1346−1349.
(33) Chang, A. B.; Lin, R.; Studley, W. K.; Tran, C. V.; Saier, M. H. Phylogeny as a guide to structure and function of membrane transport proteins (Review). *Mol. Membr. Biol.* **2004**, *21* (3), 171−181.
(34) Cymer, F.; Schneider, D. A single glutamate residue controls the oligomerization, function, and stability of the aquaglyceroporin GlpF. *Biochemistry* **2010**, *49* (2), 279−86.
(35) Klein, N.; Hellmann, N.; Schneider, D. Anionic Lipids Modulate the Activity of the Aquaglyceroporin GlpF. *Biophys. J.* **2015**, *109* (4), 722−731.
(36) Pandey, R.; Usui, K.; Livingstone, R. A.; Fischer, S. A.; Pfaendtner, J.; Backus, E. H.; Nagata, Y.; Frohlich-Nowoisky, J.; Schmuser, L.; Mauri, S.; Scheel, J. F.; Knopf, D. A.; Poschl, U.; Bonn, M.; Weidner, T. Ice-nucleating bacteria control the order and dynamics of interfacial water. *Sci. Adv.* **2016**, *2*(4), DOI: 10.1126/sciadv.1501630.
(37) Leonenko, Z. V.; Carnini, A.; Cramb, D. T. Supported planar bilayer formation by vesicle fusion: the interaction of phospholipid vesicles with surfaces and the effect of gramicidin on bilayer properties using atomic force microscopy. *Biochim. Biophys. Acta, Biomembr.* **2000**, *1509* (1), 131−147.
(38) Kim, J.; Kim, G.; Cremer, P. S. Investigations of Water Structure at the Solid/Liquid Interface in the Presence of Supported Lipid Bilayers by Vibrational Sum Frequency Spectroscopy. *Langmuir* **2001**, *17* (23), 7255−7260.
(39) Hamm, P.; Lim, M.; DeGrado, W. F.; Hochstrasser, R. M. The two-dimensional IR nonlinear spectroscopy of a cyclic penta-peptide in relation to its three-dimensional structure. *Proc. Natl. Acad. Sci. U. S. A.* **1999**, *96* (5), 2036−2041.
(40) Tsuboi, M.; Ikeda, T.; Ueda, T. Raman Microscopy of a Small Uniaxial Crystal - Tetragonal Aspartame. *J. Raman Spectrosc.* **1991**, *22* (11), 619−626.
(41) Pajcini, V.; Chen, X. G.; Bormett, R. W.; Geib, S. J.; Li, P. S.; Asher, S. A.; Lidiak, E. G. Glycylglycine pi->pi* acid charge transfer transition moment orientations: Near-resonance Raman single-crystal measurements. *J. Am. Chem. Soc.* **1996**, *118* (40), 9716−9726.
(42) Hamm, P.; Zanni, M. *Concepts and Methods of 2D Infrared Spectroscopy*; Cambridge University Press: 2011.
(43) Hamm, P.; Lim, M. H.; Hochstrasser, R. M. Structure of the amide I band of peptides measured by femtosecond nonlinear-infrared spectroscopy. *J. Phys. Chem. B* **1998**, *102* (31), 6123−6138.
(44) Gorbunov, R. D.; Kosov, D. S.; Stock, G. Ab initio-based exciton model of amide I vibrations in peptides: Definition, conformational dependence, and transferability. *J. Chem. Phys.* **2005**, *122* (22), 224904.
(45) Krimm, S.; Abe, Y. Intermolecular Interaction Effects in Amide I Vibrations of Beta Polypeptides. *Proc. Natl. Acad. Sci. U. S. A.* **1972**, *69* (10), 2788−2792.
(46) Berendsen, H. J. C.; Vanderspoel, D.; Vandrunen, R. Gromacs - a Message-Passing Parallel Molecular-Dynamics Implementation. *Comput. Phys. Commun.* **1995**, *91* (1−3), 43−56.

(47) 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* (2), 926−935.
(48) Tieleman, D. P.; Berendsen, H. J. C. A molecular dynamics study of the pores formed by E-coli OmpF porin in a fully hydrated Palmitoyloleoylphosphatidylcholine bilayer. *Biophys. J.* **1998**, *74* (6), 2786.
(49) Berger, O.; Edholm, O.; Jahnig, F. Molecular dynamics simulations of a fluid bilayer of dipalmitoylphosphatidylcholine at full hydration, constant pressure, and constant temperature. *Biophys. J.* **1997**, *72* (5), 2002−2013.
(50) Oostenbrink, C.; Soares, T. A.; van der Vegt, N. F. A.; van Gunsteren, W. F. Validation of the 53A6 GROMOS force field. *Eur. Biophys. J.* **2005**, *34* (4), 273−284.
(51) Lindahl, E.; Hess, B.; van der Spoel, D. GROMACS 3.0: a package for molecular simulation and trajectory analysis. *J. Mol. Model.* **2001**, *7* (8), 306−317.
(52) Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. LINCS: A linear constraint solver for molecular simulations. *J. Comput. Chem.* **1997**, *18* (12), 1463−1472.
(53) Tieleman, D. P.; Forrest, L. R.; Sansom, M. S. P.; Berendsen, H. J. C. Lipid properties and the orientation of aromatic residues in OmpF, influenza M2, and alamethicin systems: Molecular dynamics simulations. *Biochemistry* **1998**, *37* (50), 17554−17561.
(54) Bussi, G.; Donadio, D.; Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* **2007**, *126* (1), 014101.
(55) Nose, S. A Unified Formulation of the Constant Temperature Molecular-Dynamics Methods. *J. Chem. Phys.* **1984**, *81* (1), 511−519.
(56) Basconi, J. E.; Shirts, M. R. Effects of Temperature Control Algorithms on Transport Properties and Kinetics in Molecular Dynamics Simulations. *J. Chem. Theory Comput.* **2013**, *9* (7), 2887−2899.
(57) Barth, A. Infrared spectroscopy of proteins. *Biochim. Biophys. Acta, Bioenerg.* **2007**, *1767* (9), 1073−1101.
(58) Ji, N.; Ostroverkhov, V.; Chen, C. Y.; Shen, Y. R. Phase-sensitive sum-frequency vibrational spectroscopy and its application to studies of interfacial alkyl chains. *J. Am. Chem. Soc.* **2007**, *129* (33), 10056.
(59) Wang, Y.; Schulten, K.; Tajkhorshid, E. What Makes an Aquaporin a Glycerol Channel? A Comparative Study of AqpZ and GlpF. *Structure* **2005**, *13* (8), 1107−1118.
(60) Uzcategui, N. L.; Zhou, Y.; Figarella, K.; Ye, J.; Mukhopadhyay, R.; Bhattacharjee, H. Alteration in glycerol and metalloid permeability by a single mutation in the extracellular C-loop of Leishmania major aquaglyceroporin LmAQP1. *Mol. Microbiol.* **2008**, *70* (6), 1477−86.


https://doi.org/10.1021/acs.langmuir.1c02206
*Langmuir* 2021, 37, 13452−13459