Vibrational Sum Frequency Generation spectroscopy of proteins at interfaces.

Max Planck Graduate Center
mit der Johannes Gutenberg-Universität

Lars Schmüser-Steger

Dissertation

zur Erlangung des Grades eines
„Doktor rerum naturalium (Dr. rer. nat.)“
der Fachbereiche:
08 - Physik, Mathematik und Informatik
09 - Chemie, Pharmazie und Geowissenschaften,
10 - Biologie,
Universitätsmedizin
der Johannes-Gutenberg-Universität Mainz

I hereby declare that I wrote this dissertation submitted without any unauthorized external assistance and used only sources acknowledged in the work. All textual passages which are appropriated verbatim or paraphrased from published and unpublished texts as well as all information obtained from oral sources are duly indicated and listed in accordance with bibliographical rules. In carrying out this research, I complied with the rules of standard scientific practice as formulated in the statutes of the Johannes Gutenberg-University Mainz to insure standard scientific practice.

Abstract

This thesis is about Sum Frequency Generation Spectroscopy (SFG) of proteins at different interfaces. The main goal was to obtain detailed structural information of proteins at interfaces gained by this method. To obtain protein structural information is important for the understanding of protein functions. But drawing a molecular picture of proteins at or in cell membranes remains a rarely reached holy grail in protein biophysics. Hence, finding ways for solving structures of those proteins can substantially deepen our understanding of the function of cell membranes.

The results of this thesis are divided into three sections. In the first section, the tilt angle of a membrane binding protein is determined with SFG spectroscopy. The protein IM30 was injected below a lipid monolayer, the subsequent binding was followed by SFG spectroscopy. The protein's tilt angle was then deduced from protein to lipid peak intensity ratios.

The second section deals with the absolute orientation of peptides at interfaces. Since tilt angles of molecules measured with non-phase-resolved SFG exhibit a 180° phase uncertainty, the absolute orientation remains unknown. Phase-resolved SFG provides this missing information. Here, phase-resolved SFG was employed to study the absolute orientation of LK-peptides at the water air interface.

And the last section approaches the holy grail of structural information of membrane proteins with atomic resolution. Here, we describe an approach combining Molecular Dynamic simulations with SFG spectroscopy to predict and verify structures of membrane proteins.

Interestingly, the experimental ease and the structural information gained are working in opposite directions. The more the SFG response of a protein is canceled out due to inversion symmetry, the more information can be deduced from its SFG spectra. IM30, which consists of more than 50 parallel alpha helices, is experimentally easy to study. Yet, the spectral information is limited to a rough estimate of the overall tilt angle. On the other hand, GlpF which mainly consists of anti-parallel alpha helices exhibits an extremely low SFG intensity, but the spectral information suffices to verify MD-simulations.

Abstract

Das Thema dieser Arbeit ist Sum Frequency Generation Spektroskopie (SFG) von Proteinen an verschiedenen Grenzflächen. Das Hauptziel war, detaillierte Strukturinformationen von Proteinen an Grenzflächen zu erhalten. Diese Informationen sind für das Verständnis der jeweiligen Proteinfunktion wichtig. Aber trotz der Relevanz bleibt ein molekulares Bild von Proteinen an oder in Zellmembranen ein selten erreichtes Ziel in der Proteinbiophysik. Das Finden neuer Wege zur Erzeugung von Protein-Strukturinformation könnte demnach das Verständnis der Funktion von Zellmembranen wesentlich bereichern.

Die Ergebnisse dieser Arbeit sind in drei Abschnitte unterteilt. Im ersten Abschnitt wird der Neigungswinkel eines membranbindenden Proteins mit SFG-Spektroskopie bestimmt. Das Protein IM30 wurde unter eine Lipidmonoschicht injiziert, die anschließende Bindung wurde durch SFG-Spektroskopie verfolgt. Der Protein-Neigungswinkel wurde dann aus Protein-zu-Lipid-Peak-Intensitätsverhältnissen abgeleitet.

Der zweite Abschnitt befasst sich mit der absoluten Orientierung von Peptiden an Grenzflächen. Da Neigungswinkel von Molekülen, die mit nicht-phasenaufgelöstem SFG bestimmt werden, eine 180 ° - Phasenunsicherheit aufweisen, bleibt die absolute Orientierung unbekannt. Dieses lässt sich mit Phasen aufgelöster SFG Spektroskopie beheben. Hier wurde phasenaufgelöstes SFG verwendet, um die absolute Orientierung von LK-Peptiden an der Wasser-Luft-Grenzfläche zu untersuchen.

Der letzte Abschnitt beschreibt einen Ansatz, der Strukturinformation von Membranproteinen mit atomarer Auflösung bestimmen kann. Hierbei wird das Ergebnis einer Molecular Dynamics Simulation mit experimenteller und gerechneter SFG-Spektroskopie verglichen, um Strukturen von Membranproteinen vorherzusagen und zu verifizieren.

Interessanterweise sind die experimentelle Durchführbarkeit, und die dabei gewonnenen Strukturinformationen gegenläufige Eigenschaften. Je stärker die SFG-Antwort eines Proteins aufgrund von Inversions-Symmetrie ausgelöscht wird, desto mehr Informationen lassen sich aus seinen SFG-Spektren ableiten. IM30, das aus mehr als 50 parallelen alpha-Helices besteht, ist mit SFG Spektroskopie verhältnismäßig einfach zu studieren. Jedoch ist die gewonnene Information auf eine grobe Abschätzung des Gesamtneigungswinkels beschränkt. Auf der anderen Seite weist GlpF, dass hauptsächlich aus anti-parallelen alpha-Helices besteht, eine extrem niedrige SFG-Intensität auf, aber die Daten reichen aus, um MD-Simulationen zu verifizieren.

Table of content

1. Introduction

Proteins play an important role in many biological processes (all basic biological information can be found in more detail in biochemical textbook like refs. [1-2]). Besides DNA, RNA, lipids and sugars, proteins are the basic fundament of all living organisms. Almost every biological process involves proteins. Important examples are enzymes, which catalyze biological reactions by providing a binding site as a micro environment for chemical reactions. This micro environment facilitates the catalyzed reaction by providing i.e. the perfect local pH, electron donors or acceptors, molecular orientation of the reaction partners or just enhances the reaction speed by keeping the educts close to each other. Other prominent examples of proteins are membrane proteins and the cytoskeleton of cells.

Membranes of all organisms are almost fully covered with protein. Some of them are involved in processes of the immune system i.e. the recognition of healthy cells. Others are responsible for anchoring the cell in the surrounding tissue. Furthermore, membrane transport proteins are the gateway for every large or polar molecule which is inherently not able to cross the membrane.

All living cells exhibit a protein-based cytoskeleton. These highly dynamic structures are involved in all proliferation processes. Malfunction is thus fatal.

All these functions are determined by the protein´s 3D structure. Hence knowing this structure in detail is essential for understanding the protein's function. Knowledge of the structure and function of proteins allows one to target proteins using pharmaceutical drug applications. Drug design would, for instance, be facilitated and quickened by specifically searching for compounds which perfectly match the desired binding site of the protein, which in turn can be inferred from its structure.

1.1 Protein Structures

Proteins are polymers consisting of 21 different amino acids[3]. The amino acids form a linear chain wherein the sequence determines the primary structure (Figure 1a). Depending on the chemical properties of their sidechains the amino acids are divided into several subgroups: electrically charged, polar uncharged and aliphatic. Depending on the environment and the amino acid sequence the protein may fold into a so-called secondary structure which are classified as alpha helices (Figure 1b), beta sheets, turns and unordered random coils. The chemical and steric properties of amino acid sidechains are the driving force for those secondary structures. A model peptide family which demonstrates the importance of the sidechain for its secondary structure is the family of the LK-peptides wherein peptides with repetitions of leucine (L) and lysine (K) form beta sheets at hydrophobic surfaces like the water/air interface. At these interfaces, peptides with LKK repetitions form alpha-helical structures[4]. As soon as the protein consists of more than one secondary structure unit, the protein adopts a tertiary structure (Figure 1c). Assemblies of multiple proteins are regarded as quaternary structures (Figure 1d).

Depending on the environment the protein is embedded in it is further classified as water soluble or a membrane protein.

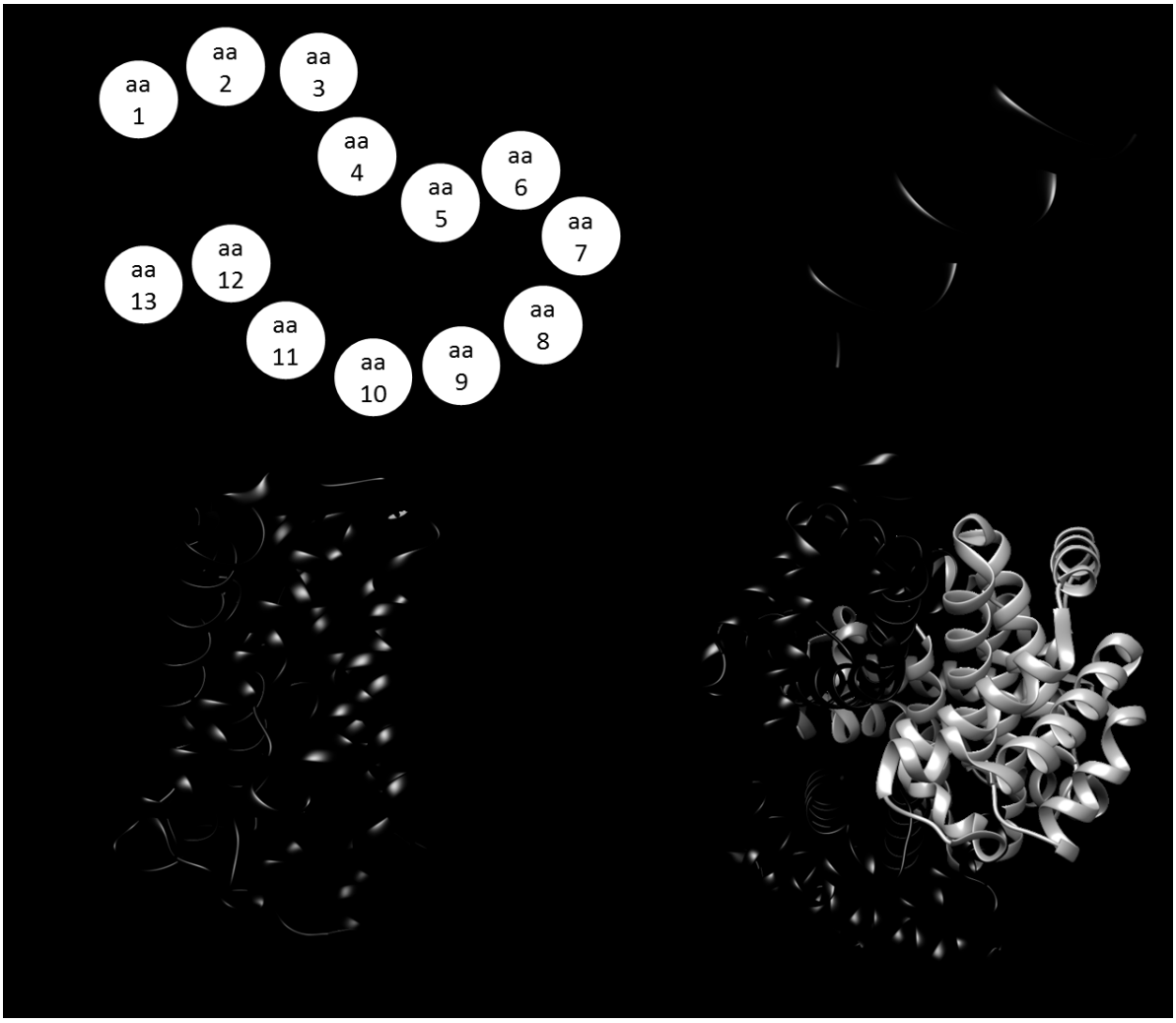


Figure 1: Different instances of protein structure: a) the **primary structure** represents the linear sequence assembled by the 21 different amino acids. b) the **secondary structure** is the basic structure of a protein or peptide. It may be an alpha helix, beta sheet or a random coil. c) the **tertiary structure** represents the entire 3D structure of a protein or peptide which consists in most cases of many different secondary structures. d) the **quaternary structure** (human hemoglobin, black = α-subunit, grey = β-subunit) is an assembly of multiple proteins which are then called subunits. The protein structures have been depicted with UCSF Chimera

Protein structure history and milestones

The first 3D protein structure was derived from crystalized sperm whale myoglobin using X-ray diffraction in 1958[5] for which Sir John Cowdery Kendrew, among others, was awarded with the Nobel Prize in 1962. This has been followed by more than 100,000 protein structures up until now (Figure 2). Due to difficulties crystalizing membrane proteins – mostly caused by their hydrophobic part which is typically embedded in the inner core of the membrane – most of the solved protein structures are from soluble proteins[6] (Figure 2). To overcome the crystallization issues, the membrane proteins were crystalized in detergent containing buffer, which shielded the hydrophobic parts of the proteins. The

first membrane protein was crystalized in 1984 by Deisenhofer *et al.*[7], who was awarded a Nobel Prize in 1988.

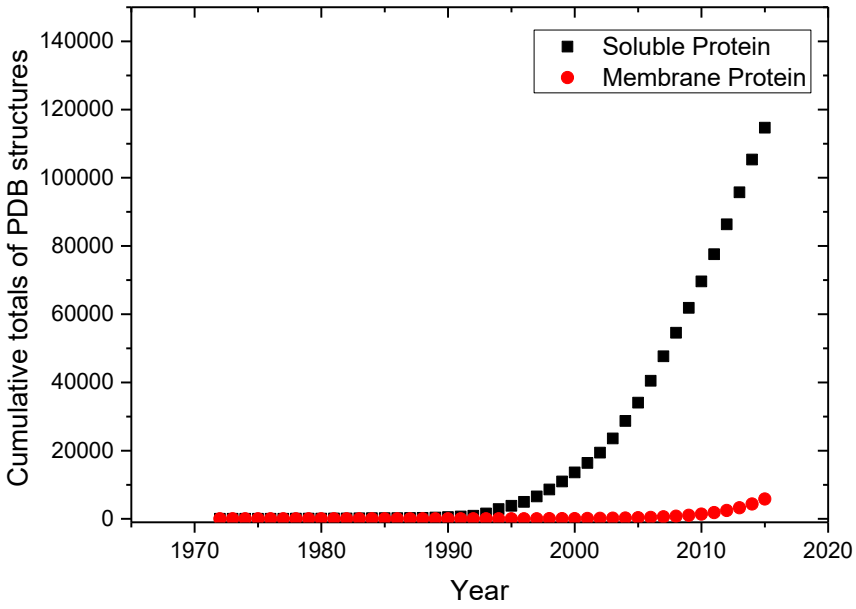


Figure 2 Numerical progression of solved 3D structures of soluble proteins and membrane proteins in the protein data base (PDB[8])

Current Methods

Most protein structures have been solved with X-ray diffraction of protein crystals (Figure 3a). But besides this method there are many other techniques available which can glean information about protein structures. The information gained varies between the different methods. Circular Dichroism (CD)[9], Infrared (IR) spectroscopies[10] and Raman scattering[11] give a rough approximation of the amount and type of secondary structures. Small Angle Neutron Scattering (SANS)[12], Small Angle X-ray Scattering (SAXS)[13] and Electron Microscopy (EM)[14] provide information about the protein shape. Only Nuclear Magnetic Resonance (NMR)[15] and X-ray and Neutron diffraction of protein crystals[7, 16] are able to solve protein structures with atomic resolution. In many cases, only the combination of different techniques yields an atomic picture of a protein[14].

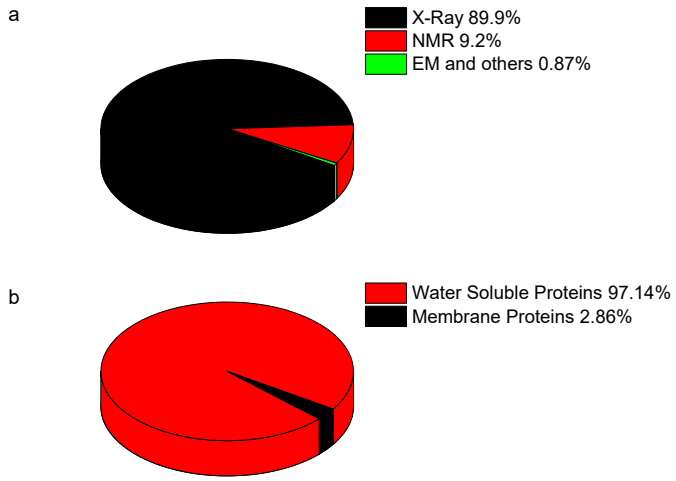


Figure 3 a) methods used for solving protein 3D structures in the protein data base (PDB[8]) b) estimated percentages of water soluble proteins and membrane proteins available in the PDB.

1.2 IR spectroscopy of proteins

Proteins have been widely studied with different infrared techniques[10]. The main focus of IR experiments studying proteins lies in the amide-I region, which is in the frequency range between 1600 cm^{-1} and 1700 cm^{-1}. The IR signals of proteins in this region are mainly originating from their backbone amide C=O stretch vibrations. The strength of each amide bond determines its vibrational frequency which is called eigenmode. The strength of this bond is influenced by the hydrogen bond between the C=O backbone and N-H groups. This is further determined by the type of secondary structure of the protein. The frequencies of protein backbone oscillations are therefore characteristic for its type of secondary structure[10]. Furthermore, the vibrational frequency can be used to confirm proper folding of expressed or isolated proteins. Commonly used IR spectroscopies are transmission Fourier transform infrared (FTIR)[17-18], Attenuated Total Reflectance(ATR)-FTIR[19], Surface Enhanced Infrared Absorption spectroscopy (SEIRAS)[20] and Infrared Reflection-Absorption spectroscopy (IRRAS)[21].

1.3 Proteins at interfaces

Roughly 30% of human proteins are membrane proteins[22]. These membrane proteins, among others, are involved in signaling processes, cell adhesion or cell respiration reactions. Being exposed to the surrounding medium membrane proteins are easily accessible for pharmaceutical drugs[23]. Furthermore, the drug must not cross the membrane barrier to target a membrane protein.

Every component or particle which is incorporated into the human body must interact with human proteins. Those will start to adsorb immediately and nonspecifically after incorporation of the artificial material and form the protein corona, which, from then on, is the particles interaction interface[24], and determines how it interacts with its surroundings. Upon adsorption, the proteins in the adsorbed corona can undergo conformational changes, like clustering or aggregation. But the problem remains that the

predominant number of those interfacial processes are not yet understood. The design of the surface of an artificial particle requires therefore a better understanding of the interfacial processes to purposefully tailor the formed protein corona[25].

Studying those interfacial processes brings certain challenges. In most cases the main part of proteins will be dissolved and 'inactive' in bulk medium and must be differentiated from the 'active' population at the interface. Hence, following protein conformation at interfaces equals looking at an active population, which is hidden behind the vast majority of a similar looking but inactive population in bulk. There are different approaches dealing with this challenge. Removing the bulk solution after incubation which is typically used for methods in ultra-high vacuum like X-ray Photoelectron Spectroscopy (XPS), Secondary Ion Mass Spectrometry (SIMS) or Near Edge X-ray Absorption Fine Structure Spectroscopy (NEXAFS)[26]. Alternative strategies focus on narrowing the sampling depth of the method, to minimize the bulk contribution. This is the typical approach for optical methods like ATR-FTIR, IRRAS, SEIRAS or Sum Frequency Generation spectroscopy (SFG). The latter technique is central to the work described in this thesis, and will be explained in more detail in the following.

Sum Frequency Generation spectroscopy of proteins

Sum frequency generation (SFG) spectroscopy is an inherently interface specific vibrational technique. The interface specificity originates from the fact that the SFG response correlates with the sum over all molecular hyper polarizability vectors. In media, which possesses centrosymmetry or inversion symmetry, this sum will be zero and, as a consequence, so will be the SFG response. Hence, only ordered interfaces will give rise to SFG signals (more details can be found in the theory section)[27]. Due to the selection rules of this second order non-linear optical method, SFG spectra are challenging to analyze. The most basic information one can gain from an SFG peak is that an ordered interfacial layer exists. Understanding changes in peak intensities is less trivial. An increase in intensity of the peak can originate either from a higher degree of order, i.e. an increased degree of the collective orientation of a constant number of molecules, or from a higher density of equally ordered molecules at the interface. Hence, understanding changes in SFG intensities requires a second method which provides information about molecular concentration at interfaces. Such information can be obtained from, for example, measurements of the surface pressure at the water-air interface or by using XPS of solid substrates.

A prominent application of SFG spectroscopy is studying the order of molecules with alkyl chains like i.e. lipids at interfaces. Here, the methyl and methylene groups are ordered in a very characteristic way. The methylene groups possess an inversion symmetry which leads to a decreasing SFG intensity of methylene related peaks with increasing order. On the contrary, the intensity of methyl related peaks is inversely correlated to the alkyl chain order. Therefore, the ratios of methyl to methylene peak intensities can be used to determine the molecular order and orientation of the interfacial layer (Figure 4)[28-29]. A more complex situation arises when doing SFG spectroscopy of proteins in the frequency region corresponding to the amide-I modes. Here, the SFG response originates from hundreds of different

amide bonds with different orientations, and different binding energies. Therefore, an amide-I peak consists of multiple overlapping spectral features of hundreds of different eigenmodes of the interfacial protein layer resulting in a complex pattern of peaks and dips within the spectrum.

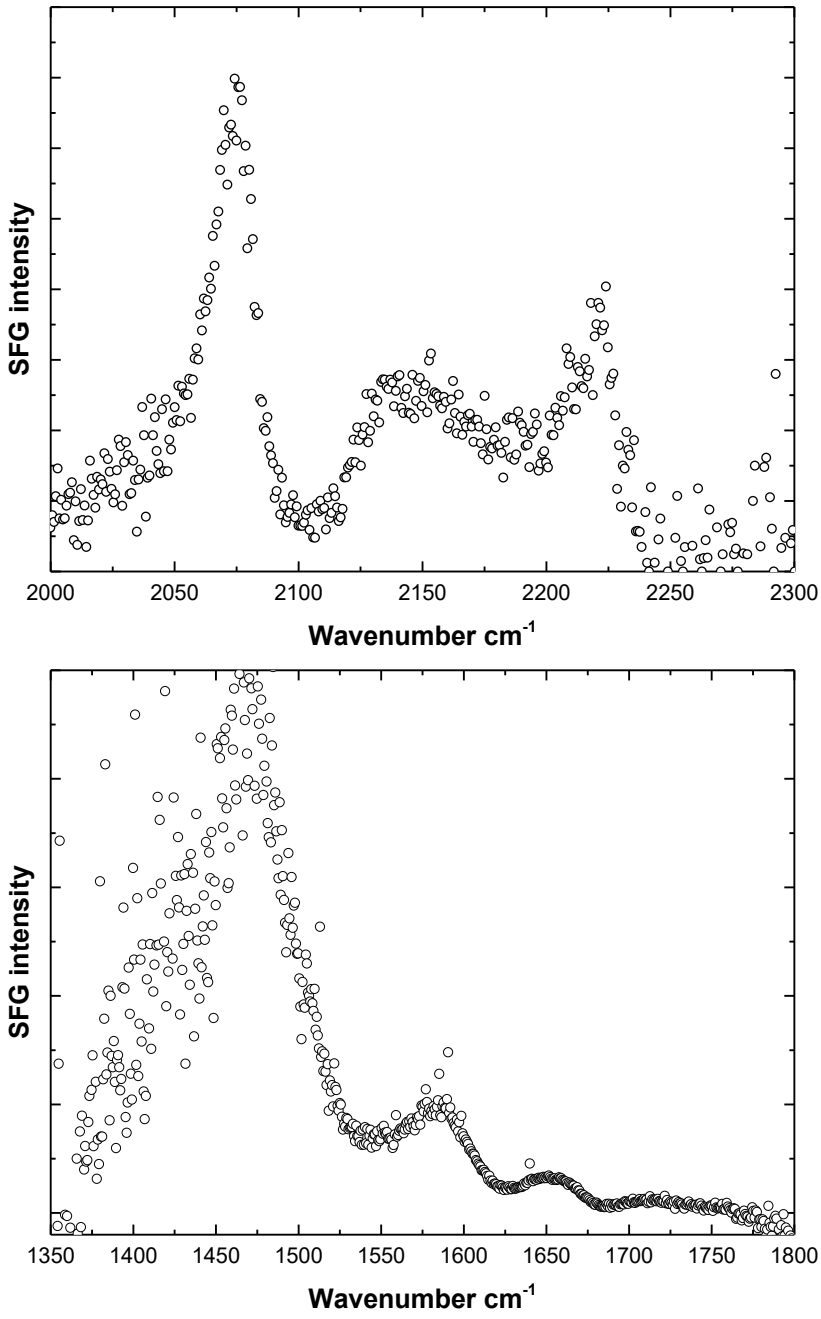


Figure 4 Comparison of deuterated Methyl/Methylene SFG spectra and SFG spectra collected in the amide region. Both spectra were collected at the CaF_2 – water interface of a CaF_2-prism. **A)** SFG spectrum of a hybrid supported DPPC bilayer. One leaflet was protonated; one leaflet was deuterated. The SFG spectrum in SSP polarization of the deuterated leaflet is depicted here. The asymmetric and symmetric CD_3 peaks are clearly separated. **B)** SFG spectrum of the glycerol facilitator protein GlpF in a supported bilayer membrane. CH_3 scissoring modes between 1400 cm^{-1} and 1500 cm^{-1} are dominant making it difficult to differentiate between peaks and dips in the amide I region.

Furthermore, the intensities of those amide-I peaks often do not exceed that of the non-resonant background, leading to many different peak shapes due to interferences of the resonant and non-resonant signals (Figure 4) and interferences between neighboring peaks. In sum: the interpretation of amide-I spectroscopy of proteins with SFG is challenging and, in most cases, delivers only a very rough indication of the secondary structure.

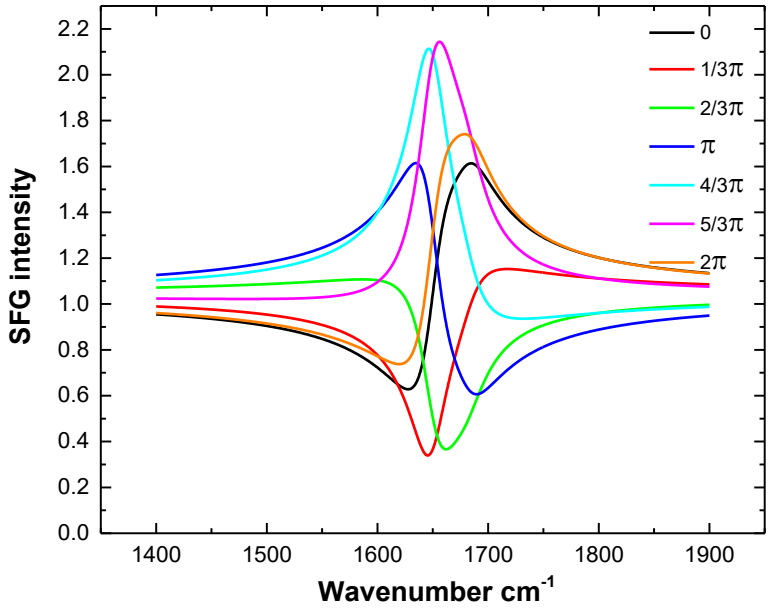


Figure 5 Calculated SFG spectra of a single alpha helix at phase differences from 0 to 2π. Here, the amplitude of the non-resonant part was set to a higher value than the amplitude of the resonant part. The broad variety of peak shapes demonstrates the strong influence of the non-resonant phase to the peak shape and the position of the peak maximum.

The complexity can be reduced by using simplified model systems like small peptides which consist in most cases of only one single secondary structure. Alternatively, modeling the experimental SFG data with calculated SFG data of simulated protein structures helps to disentangle the convoluted information of an amide-I spectrum[30-31]. In a regular SFG experiment, the SFG intensity equals the response squared. Resolving the phase of the SFG response provides the real and the imaginary part of the SFG response. Thus, one can obtain additional information about the SFG response by constituting an experimental SFG spectrum and resolving the phase of the response. All three approaches were employed in this thesis.

1.4 Model membrane systems used for spectroscopies

Lipid Monolayers

A lipid monolayer is a simplified model of a biological membrane. Here, the cell membrane is reduced to one leaflet which is spread at the water air interface. The fluidity of a lipid monolayer depends on the

lateral lipid packing density in addition to the temperature dependent phase of lipid molecules. The lateral lipid packing density is also denoted as surface pressure or surface tension[32].

The main advantages of lipid monolayers are: monolayers are easy to prepare; one only needs a syringe or pipette to add a controlled volume of a diluted lipid solution onto the surface of a water-filled trough. The lipid monolayer then forms automatically. Furthermore, lipid monolayers are very stable: it is possible to use a lipid monolayer even with high concentrations of membrane perturbing reactants like i.e. cell penetrating peptides or detergents[33].

The disadvantages of lipid monolayers are: There is no reasonable steady state: In binding experiments, the binding-molecule will bind to the membrane until an equilibrium surface tension is reached. This equilibrium state or stoichiometry depends on the strength of the binding-molecule / membrane interaction and is an arbitrary value which does not necessarily reflect a physiologically relevant stoichiometry. Using a trough with moving barriers, which makes the surface area adjustable, allows for binding experiments with constant surface pressure. The drawback of this approach is that the concentration of binding-molecule will increase with time until the trough surface area reaches its maximum. Hence, for experiments in which a definite and constant lipid / binding-molecule ratio is crucial, a constant surface pressure approach is inappropriate. Furthermore, the monolayers lipids are completely exposed to the surrounding air and oxygen. Therefore, unsaturated lipids must be used in a nitrogen flushed reaction chamber to avoid oxidation[34].

Planar bilayer

Planar bilayers are the gold standard for protein membrane interaction studies with infrared and laser based techniques[35-37]. They provide a more realistic model for cell membranes than lipid monolayers while still being comparatively easy to handle and prepare. Supported lipid bilayers can be prepared either as supported or as tethered bilayers[38-40]. In supported lipid bilayers, the lipids of the proximal layer are either cushioned by a 10-20 Å water layer or directly in contact with the solid support[41]. In tethered bilayer membranes, tether molecules are covalently linked to the substrate and provide a stable link for the membrane to the substrate. A general drawback of two dimensional membranes is the small number of lipid molecules (one molecule per $\sim 10^{-14}$ cm^{-2} $\approx$ 0.1 nM in a typical trough with 1 cm depth) in comparison to the number of binding-molecules which often makes it impossible to use physiologically relevant concentrations and stoichiometry.

Supported bilayer:

Advantages: With the Langmuir-Blodgett / Langmuir-Schaefer deposition it is possible to prepare heterogeneously supported lipid bilayers which allow for spectroscopical differentiation between both leaflets when one of them is isotopically labeled[38, 42-43].

Disadvantages: The proximal leaflet at the solid substrate is very inflexible. i.e. the diffusion constant for lateral mobility is reduced dramatically[44]. Additionally, supported lipid bilayers possess a low

stability towards membrane perturbing molecules. Therefore, it is difficult to find the right concentration of the interacting-molecule where an impact of the binding is observable and the supported bilayer remains intact[35].

Tethered Bilayer Lipid Membrane (t-BLM)

Tether molecules for t-BLMs are typically thiol or silane functionalized molecules with a spacer region (i.e. PEG spacer) followed by a hydrophilic group to match the lipid head group and then terminated by a hydrophobic tail to match the fatty acids of lipid molecules[40, 45]. Alternatively, tethered membrane proteins can be used as tether molecules too. Then, the system is called protein tethered Bilayer Lipid Membrane (pt-BLM)[46].

The advantages of tethered lipid bilayers over supported lipid bilayers are: A higher stability towards membrane perturbing elements: Due to the strong interaction of the tether molecules with the substrate, a (p)t-BLM tolerates interactions with harsher binding-molecules. Also, the tether molecules increase the distance between proximal leaflet and solid substrate. This gives i.e. integrated membrane proteins more space so that the parts of the proteins which are sticking out of the membrane are not in contact with the solid support.

Disadvantages, especially for SFG spectroscopy: The covalent attachment of tether molecules to the typical substrate for SFG measurements (CaF_2) requires a thin layer of noble metal (i.e. Au, Ag or Pt) or SiO_2. This thin layer modifies the optical parameters of the solid / water interface. A surface Plasmon resonance can be added by the noble metals. Furthermore, noble metals can add an imaginary part to the refractive index of the interface. Both issues are mostly trouble-free and partially even desirable as amplifying factors in linear optics studies, but can cause difficulties while analyzing results of a non-linear spectroscopy.

Vesicles

Lipid vesicles are the model membrane system which is closest to a real cell membrane. The lipid density, lipid mobility, membrane curvature and phase behavior are comparable to those of cell membranes when appropriate lipids are chosen[47]. Depending on their size, lipid vesicles are categorized into small unilamellar vesicles (SUV, radius < 50nm), large unilamellar vesicles (LUV, radius < 1µm) and giant unilamellar vesicles (GUV, radius > 1µm). While SUVs and LUVs are used in spectroscopy, GUVs are mainly used in microscopy experiments.

Advantages: Due to the 3-dimensional nature of vesicles in solution, it is possible to perform experiments with a physiological relevant stoichiometry. Here, an excess of lipids over the binding-molecules can be used which mimics physiologically relevant conditions and prevents steric hindrance due to very high packing of binding molecules.

Disadvantages: IR light is - except for a few open spectral windows - strongly absorbed by $^{1}H_20$ or $^{2}H_2O$. Hence, the penetration depth of IR methods probing lipid vesicles is fairly short and the spectral range is limited to those spectral windows where the IR light is not absorbed by water.

1.5 Goals of this thesis

The aim of the research described in this thesis is to extract structural information gained from complex experimental SFG spectra of proteins. To achieve this, we combined experimental SFG spectra with calculated SFG spectra of modeled peptides and proteins. The main goal was to get a validated or improved 3D model of a membrane protein which was simulated by molecular dynamics simulation in its physiological environment. The motivation behind this question was to establish how reliable protein structures derived from crystalized proteins are in comparison with their structure in physiological environments, to introduce SFG as a structure solving technique and to inspire future projects using this technique for protein folding studies with picosecond time resolution.

1.6 Content of this thesis

This thesis consists of 5 chapters.

The first chapter draws a very basic, general overview about the methods, samples and goals of this thesis.

In chapter 2 the current state of research and theoretical backgrounds of the used methods are explained in detail.

Chapter 3 is about experimental implementation of the experiments explaining in detail which difficulties may occur and how to handle them.

In chapter 4 all results are described and discussed in detail. As all results presented here are either submitted or published in peer reviewed journals this chapter contains a collection of those articles. *

Chapter 5 summarizes the results of this thesis, gives an overview on the goals, the achievements and an outlook of topics for further investigation.

---

*In accordance with the MPGC examination regulations §11.1.

2.Theory

The theoretical background of this chapter is based on reference 48. All vectors in this thesis are represented in bold typeface.

2.1 Nonlinear Optics

When light interacts with matter it can induce a dipole moment (**μ**) which is proportional to the electric field (**E**) and the molecular polarizability tensor (α).

$$\mu = \alpha \mathbf{E} \qquad \text{Eq 1}$$

Macroscopically this is summed up to the polarization of a molecular ensemble (**P**).

$$\mathbf{P} = \epsilon_0 \chi \mathbf{E} \qquad \text{Eq 2}$$

Where ϵ_0 is the permittivity of the free field and χ is the linear susceptibility which describes the macroscopic response of a material to the incoming electric field.

When the strength of an electric field exceeds a certain threshold (typically ~2.5 kW/cm^2) the response becomes nonlinear. In this case Eq 2 must be extended into a power series

$$\mathbf{P} = \epsilon_0 (\chi^{(1)}\mathbf{E} + \chi^{(2)}\mathbf{E^2} + \chi^{(3)}\mathbf{E^3} + \chi^{(4)}\mathbf{E^4} + \cdots) \qquad \text{Eq 3}$$

Where $\chi^{(i)}$ is the nonlinear susceptibility of ith order.

2.2 Second order nonlinear optics

In case of three wave mixing processes – two incoming and one generated wave - $\chi^{(2)}$can give rise to the following five phenomena:

$$\mathbf{P^{(2)}} = \epsilon_0 \chi^{(2)} \begin{bmatrix} 2(\mathbf{E_1 E_1^*}) \\ +\mathbf{E_1^2} exp(-2i\omega_1 t) \\ +\mathbf{E_2^2} exp(-2i\omega_2 t) \\ +2\mathbf{E_1 E_2^*} exp(-i(\omega_1 - \omega_2)t) \\ +2\mathbf{E_1 E_2^*} exp(-i(\omega_1 + \omega_2)t) \end{bmatrix} \qquad \text{Eq 4}$$

The first term corresponds to optical rectification, the second and third term to second harmonic generation (SHG), the fourth term corresponds to difference frequency generation (DFG) and the last term to sum frequency generation (SFG). Only SFG processes are considered in this thesis. They can be specifically separated from other effects by selecting the right angle and optical filters for the reflected light.

The SFG process is considered as a three-wave mixing process. Thus, the susceptibility $\chi^{(2)}$ must be a third order tensor. So, that

$$\mathbf{P_{SF}^{(2)}} = \epsilon_0 \chi^{(2)} \mathbf{E_1 E_2} \quad \text{Eq 5}$$

Can be written as

$$\mathbf{P_{SF}^{(2)}} = \sum_{i}^{x,y,z} \mathbf{P_{i,SF}^{(2)}} = \epsilon_0 \sum_{i}^{x,y,z} \sum_{j}^{x,y,z} \sum_{k}^{x,y,z} \chi_{i,j,k}^{(2)} \mathbf{E_{j,vis} E_{k,IR}} \quad \text{Eq 6}$$

In a system, which possesses centrosymmetry or inversion symmetry, even order nonlinear susceptibilities ($\chi^{(2n)}$) are 0. The Neumann principle states:

"The symmetry elements of any physical property of a crystal must include the symmetry elements of the point group of the crystal." [49]

Therefore, applying the inversion operator (Iop) to equation 6 leads to:

$$\mathrm{Iop}\left(\mathbf{P_{SF}^{(2)}}\right) = -\mathbf{P_{SF}^{(2)}} = -\epsilon_0 (-\chi_{-i,-j,-k}^{(2)} (-\mathbf{E})^2) \quad \text{Eq 7}$$

As $\chi_{i,j,k}^{(2)}$ is a polar tensor, $\chi_{i,j,k}^{(2)} = -\chi_{-i,-j,-k}^{(2)}$ therefore,

$$\mathrm{Iop}(\mathbf{P}) = -\mathbf{P} = -\chi_{i,j,k}^{(2)} = \chi_{-i,-j,-k}^{(2)} = \chi_{i,j,k}^{(2)} = -\chi_{-i,-j,-k}^{(2)} \quad \text{Eq 8}$$

which is only true if $\chi_{i,j,k}^{(2)} = 0$.

2.3 Sum Frequency Generation

Two temporally and spatially overlapped laser pulses of the frequencies $\omega_1$and $\omega_2$impinge on a nonlinear active material, and can generate a Sum Frequency Signal at the frequency $\omega_{SF} = \omega_{vis} + \omega_{IR}$. This process can be described as an infrared absorption of one photon followed by an Anti-Stokes Raman scattering (Figure 6) of a second photon, or in other words an Anti-Stokes Raman scattering process of a vibrationally excited interface. Thus, a vibration must be both Raman and IR active to be SFG active.

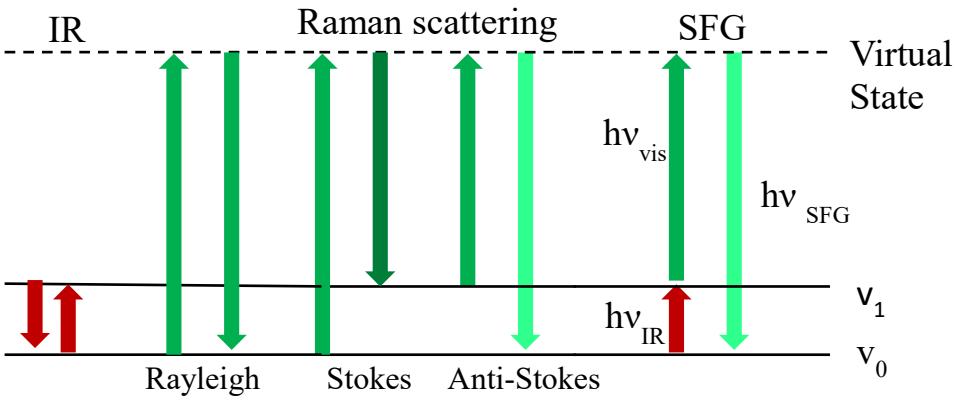


Figure 6 Linear and nonlinear excitation of molecules. Infrared absorption and Rayleigh and Raman scattering processes belong to linear optical methods. Sum Frequency Generation combines the infrared and Raman processes. An infrared pulse generates a vibrational resonant polarization which is subsequently upconverted to a virtual state by a second pulse, so that the emitted photon yields the sum of the energy of both pulses.

The generated SF signal is partially reflected and partially transmitted through the interface. In this thesis, only the reflected part was analyzed.

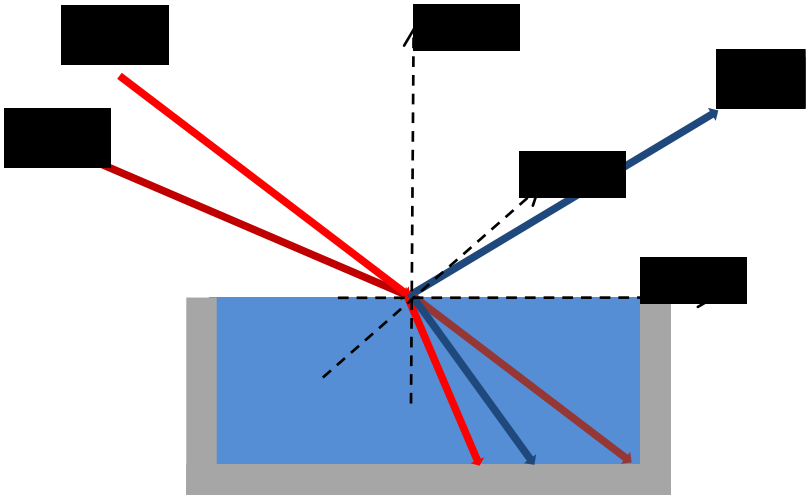


Figure 7 reflection geometry of a standardly used homodyne detected Sum Frequency Generation spectroscopy setup. IR and visible laser pulses are overlapped in time and space to generate the SFG signal.

The intensity of the reflected SF signal can be described as[50]

$$\mathbf{I_{SF}} = \frac{8\pi^3 \omega_{SF}^2 sec^2\theta_{SF}}{c^3 n_1(\omega_{SF}) n_1(\omega_{vis}) n_1(\omega_{IR})} \left|\chi_{eff}^{(2)}\right|^2 \mathbf{I_{vis} I_{IR}} \quad \text{Eq 9}$$

Where c is the speed of light, $n_1(\omega_{SF})n_1(\omega_{vis})n_1(\omega_{IR})$ are the refractive indices of the bulk medium at the frequencies $\omega_{SF}, \omega_{vis}$ *and* ω_{IR} respectively. θ_{SF} is the angle of the SF signal in respect to the surface normal which is defined as:

$$\frac{sin\theta_{SF}}{\lambda_{SF}} = \frac{sin\theta_{vis}}{\lambda_{vis}} + \frac{sin\theta_{IR}}{\lambda_{IR}} \quad \text{Eq 10}$$

The second order effective susceptibility $\chi_{eff}^{(2)}$ must be related to the Fresnel factors $L(\omega)$, the unit electric field vectors $\hat{e}(\omega)$ and the second order susceptibility $\chi^{(2)}$.

$$\chi_{eff}^{(2)} = [\mathrm{L}(\omega_{SF}) \cdot \hat{e}(\omega_{SF})] \cdot \chi^{(2)} : [\mathrm{L}(\omega_{vis}) \cdot \hat{e}(\omega_{vis})] \cdot [\mathrm{L}(\omega_{IR}) \cdot \hat{e}(\omega_{IR})] \quad \text{Eq 11}$$

The third order tensor $\chi_{i,j,k}^{(2)}$ contains 27 different tensor elements. In a non-chiral system, only 7 tensor elements are non-zero. Those can be further reduced to 4 different tensor elements due to symmetry reasons in an azimuthal symmetry. It can be probed by 4 different laser polarization combinations:

$$\chi_{SSP}^{(2)} = \chi_{xxz}^{(2)} = \chi_{yyz}^{(2)} \quad \text{Eq 11}$$

$$\chi_{SPS}^{(2)} = \chi_{yzy}^{(2)} = \chi_{xzx}^{(2)} \quad \text{Eq 12}$$

$$\chi_{PSS}^{(2)} = \chi_{zyy}^{(2)} = \chi_{zxx}^{(2)} \quad \text{Eq 13}$$

$$\chi_{PPP}^{(2)} = -\chi_{xxz}^{(2)} + \chi_{xzx}^{(2)} + \chi_{zxx}^{(2)} - \chi_{zzz}^{(2)} \quad \text{Eq 14}$$

here, the subscript i.e. SSP stands for the laser polarization combination (S-polarized SFG, S-polarized vis and P-polarized IR; S = vertical, P = parallel).

In a chiral system $\chi_{xzz}^{(2)}, \chi_{yzz}^{(2)}, \chi_{zxz}^{(2)}, \chi_{zyz}^{(2)}, \chi_{zzx}^{(2)}$ and $\chi_{zzy}^{(2)}$ are also non-zero. The remaining tensor elements are zero.

The total SFG response $\chi_{eff}^{(2)}$ of an azimuthal, non-chiral system can thus be simplified to

$$\chi_{eff}^{(2)} = \chi_{SSP}^{(2)} + \chi_{SPS}^{(2)} + \chi_{PSS}^{(2)} + \chi_{PPP}^{(2)} \quad \text{Eq 15}$$

Combining Eq 10 - Eq 14 leads to

$$\chi_{SSP}^{(2)} = L_{yy}(\omega_{SFG}) L_{yy}(\omega_{vis}) L_{zz}(\omega_{IR}) \sin\theta_{IR}\, \chi_{yyz}^{(2)} \quad \text{Eq 16}$$

$$\chi_{SPS}^{(2)} = L_{yy}(\omega_{SFG}) L_{zz}(\omega_{vis}) L_{yy}(\omega_{IR}) \sin\theta_{vis}\, \chi_{yzy}^{(2)} \quad \text{Eq 17}$$

$$\chi_{PSS}^{(2)} = L_{zz}(\omega_{SFG}) L_{yy}(\omega_{vis}) L_{yy}(\omega_{IR}) \sin\theta_{SFG}\, \chi_{zyy}^{(2)} \quad \text{Eq 18}$$

$$\begin{aligned} \chi_{PPP}^{(2)} = & -L_{xx}(\omega_{SFG}) L_{xx}(\omega_{vis}) L_{zz}(\omega_{IR}) \cos\theta_{SFG} \cos\theta_{vis} \sin\theta_{IR}\, \chi_{xxz}^{(2)} \\ & - L_{xx}(\omega_{SFG}) L_{zz}(\omega_{vis}) L_{xx}(\omega_{IR}) \cos\theta_{SFG} \sin\theta_{vis} \cos\theta_{IR}\, \chi_{xzx}^{(2)} \\ & + L_{zz}(\omega_{SFG}) L_{xx}(\omega_{vis}) L_{xx}(\omega_{IR}) \sin\theta_{SFG} \cos\theta_{vis} \cos\theta_{IR}\, \chi_{zxx}^{(2)} \\ & + L_{zz}(\omega_{SFG}) L_{zz}(\omega_{vis}) L_{zz}(\omega_{IR}) \sin\theta_{SFG} \sin\theta_{vis} \sin\theta_{IR}\, \chi_{zzz}^{(2)} \end{aligned} \quad \text{Eq 19}$$

Phase resolved SFG

To perform phase resolved SFG experiments some modifications to a regular SFG setup must be done. IR and visible beams are focused to a gold mirror which acts as local oscillator. Subsequently both beams and the generated sum frequency beam are reflected and refocused with a concave mirror to the sample. At that, the SFG beam is delayed with a fused silica plate so that the local oscillator and the sample generate an interference spectrum which contains the phase information of the SFG signal generated at the sample. The phase information can be extracted using the following equations. Here, a transformation from frequency domain into the time domain is necessary to deal with the introduced delay of the SFG response. This transformation is done by an inverse Fourier Transformation.

After inverse Fourier transformation of the electric field into the time domain a conventional SFG signal can be described as

$$\mathbf{I} = |\mathbf{E_{total}}(t)|^2 \qquad \text{Eq 20}$$

The beforehand described modification to a regular SFG setup requires the introduction of the delay T, and the Local oscillator $\mathbf{E_{LO}}(t)$ to Eq 20:

$$\mathbf{I} = |\mathbf{E_{total}}(t)|^2 = \left|\mathbf{E_{sample}}(t-T) + \mathbf{E_{LO}}(t)\right|^2 \qquad \text{Eq 21}$$

According to the Fourier Shift Theorem, $\mathbf{E_{sample}}(t-T)$ can be described as $\mathbf{E_{sample}}\exp(i\omega T)$ in the frequency domain. Hence, Eq 21 can be expressed as

$$\begin{aligned}\mathbf{I} = |\mathbf{E_{tot}}(\omega)|^2 &= \left|\mathbf{E_{sample}}\exp(i\omega T) + \mathbf{E_{LO}}(t)\right|^2 \\ &= \left|\mathbf{E_{sample}}\right|^2 + |\mathbf{E_{LO}}|^2 + \boldsymbol{E_{LO}}\mathbf{E_{sample}}\exp(i\omega T) + \boldsymbol{E_{LO}}\mathbf{E_{sample}}\exp(-i\omega T)\end{aligned} \qquad \text{Eq 22}$$

Now, in a time domain spectrum, $\left|\mathbf{E_{sample}}\right|^2$ and $|\mathbf{E_{LO}}|^2$ are well separated – by roughly 2ps – from their cross terms $\boldsymbol{E_{LO}}\mathbf{E_{sample}}\exp(i\omega T)$ and $+\boldsymbol{E_{LO}}\mathbf{E_{sample}}\exp(-i\omega T)$. Hence, we can select one cross term specifically by applying a filter function which sets the temporally separated$\left|\mathbf{E_{sample}}\right|^2$, $|\mathbf{E_{LO}}|^2$ and $\boldsymbol{E_{LO}}\mathbf{E_{sample}}\exp(-i\omega T)$ to zero. After Fourier transforming the spectrum back to the frequency domain, the remaining signal can be expressed as

$$\mathbf{I} = \tilde{\mathbf{E}}_{\mathbf{sample}}\tilde{\mathbf{E}}^*_{\mathbf{LO}}\exp(i\omega T) \qquad \text{Eq 23}$$

Here, the electric fields $\tilde{\mathbf{E}}_{\mathbf{sample}}$ and $\tilde{\mathbf{E}}^*_{\mathbf{LO}}$ are representing $\mathbf{E_{sample}}$ and $\mathbf{E_{LO}}$ after the filter function. $\tilde{\mathbf{E}}_{\mathbf{sample}}$ and $\tilde{\mathbf{E}}^*_{\mathbf{LO}}$ can be further described as

$$\tilde{\mathbf{E}}_{\mathbf{sample}} = ia_{sample}\chi^{(2)}_{sample}\tilde{\mathbf{E}}_{\mathbf{vis}}\tilde{\mathbf{E}}_{\mathbf{IR}} \quad \text{Eq 24}$$

And

$$\tilde{\mathbf{E}}_{\mathbf{LO}} = a_{LO}\chi^{(2)}_{LO}r^{sample}_{vis}\tilde{\mathbf{E}}_{\mathbf{vis}}r^{sample}_{IR}\tilde{\mathbf{E}}_{\mathbf{IR}} \quad \text{Eq 25}$$

Here, a_{LO} and a_{sample} are real and positive constants and r^{sample}_{vis} and r^{sample}_{IR} are the samples reflectivities for the IR and vis pulses.

Dividing Eq by a gold reference spectrum which was collected using the same settings yields

$$\frac{\tilde{\mathbf{E}}_{\mathbf{sample}}\tilde{\mathbf{E}}^{*}_{\mathbf{LO}}\exp(i\omega T)}{\tilde{\mathbf{E}}_{\mathbf{gold}}\tilde{\mathbf{E}}'_{\mathbf{LO}}\exp(i\omega T)} = \frac{a_{sample}r^{sample}_{vis}r^{sample}_{IR}}{a_{gold}r^{gold}_{vis}r^{gold}_{IR}\chi^{(2)}_{gold}}\chi^{(2)}_{sample} \quad \text{Eq 26}$$

$\chi^{(2)}_{gold}$ is considered to be a real constant because of its non-resonant origin. Hence,

$$\frac{\tilde{\mathbf{E}}_{\mathbf{sample}}\tilde{\mathbf{E}}^{*}_{\mathbf{LO}}\exp(i\omega T)}{\tilde{\mathbf{E}}_{\mathbf{gold}}\tilde{\mathbf{E}}'_{\mathbf{LO}}\exp(i\omega T)} \propto \chi^{(2)}_{sample} \quad \text{Eq 27}$$

Proving that $\chi^{(2)}_{sample}$ can be deduced by normalizing the sample spectra with a non-resonant gold spectrum.

Current state in literature

SFG spectroscopy is a specialized method used for analyzing the interfacial order and orientation of Self Assembled Monolayers (SAMs)[29, 51], lipids[52], proteins[53-54] and many more. It is an important tool to infer molecular structures at interfaces like, for example, ion concentrations at the water air interface or the orientation of molecules at the air water interface[55]. Orientational studies can be further improved by using heterodyne SFG which directly yields the imaginary and real part of the SFG spectrum[56]: specifically, the sign of the imaginary part directly reflects the orientation of the transition dipole. SFG has been used to analyze and to understand interfacial processes like water surface tension or water shearing forces[57-58]. SFG has also obtained some importance in studying biological systems. Here it can be used to study the impact of different effectors on membranes[59]. It can also be used to explore working mechanisms, orientations or conformations of peptides at interfaces[33, 60-61]. More recently SFG has also been employed to study the conformation of larger proteins.[31]

2.4 Calculation of SFG spectra

Theory

The VSFG spectra are calculated using the formalism described in ref. 60.

Protein SFG spectra were calculated based on their structure and orientation, which is given by the input PDB file. In this method, we assume that the amide-I modes originate completely from the protein backbone amide bonds. This allows us to neglect any sidechain contribution. First, we constructed the amide I excitonic Hamiltonian. Here the Hamiltonian has the following form:

$$H = \begin{pmatrix} \hbar\omega_1^0 & K_{12} & K_{13} & K_{14} & \cdots \\ K_{12} & \hbar\omega_2^0 & \beta_{23} & K_{24} & \\ K_{13} & K_{23} & \hbar\omega_3^0 & K_{34} & \\ K_{14} & K_{24} & K_{34} & \hbar\omega_4^0 & \\ \vdots & & & & \ddots \end{pmatrix} \quad \text{Eq 28}$$

Here, ω_i^0 are the frequencies of the local modes, i and K_{ij} are the couplings between the local modes i and j. Couplings between nearest neighbor amide modes were calculated using an ab initio 6-31G +(d) B3LYP- map. Here, the coupling represents a function of the dihedral angle between the amide groups. Non-nearest neighbor couplings were calculated using the transition dipole coupling model which approximates the coupling with a Coulomb-like interaction of the transition-dipole-moments.

The eigenvalues of each amide-I mode depend on the local environment of each amide bond. For example, a downstream neighboring proline residue induces a red-shift of the eigenvalue. Also, hydrogen bonds induce amide-I red-shifts.

Now, eigenvalues ω^{v} and eigenvectors $c^{\sigma v}$ of the delocalized eigenmodes $|v\rangle$ are calculated by diagonalizing the one-exciton Hamiltonian. The eigenmodes can be described as:

$$|v\rangle = \sum_{\sigma} c^{\sigma v} |\sigma\rangle \quad \text{Eq 29}$$

Here, $|\sigma\rangle$ is the localized v = 1 state of each peptide unit's (σ) amide-I mode. The IR and Raman responses, μ_k^v and α_{ij}^v respectively, of each eigenmode are now calculated by replacing the localized eigenmode with either the transition-dipole-moment μ_k^σ or the Raman tensor α_{ij}^σ:

$$\mu_k^v = \sum_{\sigma} c^{\sigma v} \mu_k^\sigma \quad \text{Eq 30}$$

$$\alpha_{ij}^v = \sum_{\sigma} c^{\sigma v} \alpha_{ij}^\sigma \quad \text{Eq 31}$$

Here, i, j, k = x, y, z is the axis of the molecular frame. Now, IR and Raman intensities can be calculated. By assuming homogeneous broadening, each eigenmode is represented with a Lorentzian line shape.

$$I_{IR} \propto \sum_{v} \left| \frac{\mu^{v}}{\omega^{v} - \omega_{IR} - i\Gamma} \right|^{2} \quad \text{Eq 32}$$

$$I_{Raman} \propto \sum_{v} \left| \frac{\alpha^{v}}{\omega^{v} - (\omega_{laser} - \omega_{Stokes} - i\Gamma} \right|^{2} \quad \text{Eq 33}$$

Here, ω_{IR}, ω_{laser} and ω_{Stokes} are the frequency of the IR field, the frequency of the excitation laser and the frequency of the Stokes field respectively. $i\Gamma$ represents the linewidth of the Lorentzian which is equal for all eigenmodes.

Now, the molecular SFG hyper polarizability ${\beta_{ijk}}^{(2)}$ is deduced from the tensor product of the IR and Raman responses

$${\beta_{ijk}}^{(2)v} = \mu_{k}^{v} \otimes \alpha_{ij}^{v} \quad \text{Eq 34}$$

Finally, the effective nonlinear susceptibility ${\chi_{IJK}}^{(2)}$ in the lab frame I, J, K = X, Y, Z can be constructed. To transform the molecular frame x, y, z to the lab frame X, Y, Z an Euler transformation over the orientation distribution of all molecules has to be performed:

$${\chi_{IJK}}^{(2)} = N \sum_{x,y,z} \langle (\hat{X} \cdot \hat{x})(\hat{Y} \cdot \hat{y})(\hat{Z} \cdot \hat{z}) \rangle {\beta_{ijk}}^{(2)} \quad \text{Eq 35}$$

Here, $\hat{X}, \hat{Y}, \hat{Z}$ and $\hat{x}, \hat{y}, \hat{z}$ are the unit vectors in the lab frame and molecular frame respectively.

Now, that we have calculated the frequency dependent, second order nonlinear susceptibility, we can apply the formalism described in equation 11-19 to account for Fresnel factors and different laser polarization combinations.

3.Experimental Methods and Experiences

3.1 Lipid Monolayer Preparation

Lipid monolayers were spread on top of buffer solutions using a Hamilton syringe with a repeating dispenser (PB600-1, Hamilton Company, U.S). Lipids were dissolved in chloroform to a final concentration of 0.1mg/ml and added dropwise until the desired surface tension was reached. The surface tension was measured using a tensiometer (DeltaPi, Kibron Inc., Finland). Protein solutions were injected into the aqueous sub-phase using syringes with steel needles. The latter is important to minimize withdrawing of lipids due to Langmuir deposition of lipids to larger areas of e.g. pipette tips. A permanent, small water inflow into the sub-phase was realized using a syringe pump (NE-500, New Era Pump Systems Inc., U.S.) to compensate for water evaporation. The flow rate was fine-tuned to match exactly the evaporation rate. This step must be done after preparation of the lipid monolayer because the evaporation rate is altered due to the presence of the lipids on the surface. The height was controlled either by following the height of the SFG signal at the CCD camera or by a displacement sensor in case of phase resolved SFG where a constant height level is crucial.

3.2 Lipid Bilayer Preparation

Lipid bilayers were formed at the solid water interface of CaF_2 prisms which were mounted on a flow cell. The leaking in of air bubbles was avoided by keeping the container for the effluent flow at a higher height than the flow cell. Formation of the lipid bilayer was induced by adding lipid vesicles at a final concentration of 12 µM into the water or buffer filled flow cell. According to literature, bilayer spreading and the formation of a proper lipid bilayer takes approximately 30 minutes. After an incubation period of 2 hours the remaining lipid vesicles were removed by rinsing the flow cell with 10ml buffer (D_2O) solution. To allow a complete NH to ND exchange the supported bilayer was exposed to D_2O over night. The latter step is necessary to minimize the H_2O bending mode contribution at the water-CaF_2-interface. It should be noted that the amide-II band shifts towards lower frequencies due to the NH to ND exchange which leads to one large combined peak of CH_2 and CH_3 scissoring and amid-II modes between 1450 cm^{-1} and 1500 cm^{-1}. All lipid bilayer experiments were performed in a flow cell (Figure 8) which was designed as part of this thesis (in cooperation with Johannes Franz).

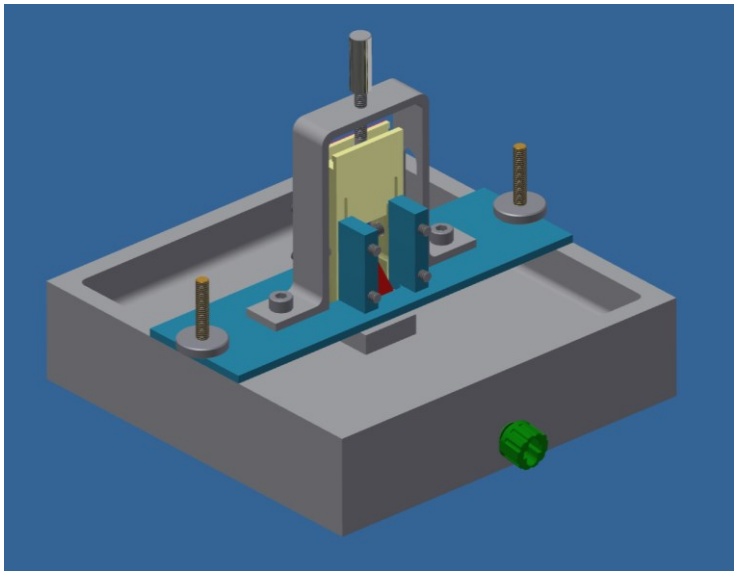


Figure 8 Flow cell for SFG measurements at the solid / water interface of CaF_2-prisms. The mountable bridge (blue and yellow) allowed performing a Langmuir-Schaefer deposition in the trough which helps during the preparation of lipid bilayers with the Langmuir-Blodgett / Langmuir-Schaefer method. The advantage of this trough is that it helps to protect the lipid bilayer from contact with air, which would destroy it.

3.3 SFG setup

The SFG setup used in this thesis (Figure 9) is powered by a femtosecond (fs) regenerative amplifier (Spitfire Ace, Spectra-Physics) which is pumped by a Nd:YLF laser (Empower, Spectra-Physics) and seeded by a Ti:sapphire oscillator (Mai Tai, Spectra-Physics). The resulting broadband laser pulses, centered at 791.8nm, have pulse energies of 5 mJ with 40 fs pulse duration. 1.7 mJ is branched out to pump the optical paramagnetic amplifier (TOPAS-C, Spectra-Physics) which generates the IR pulse by difference frequency generation (DFG) between the signal and idler beams out of the TOPAS. The remaining visible beam is guided through a Fabry-Perot Etalon (SLS Optics Ltd.) which narrows the frequency of the visible pulse to 15 cm^{-1} and reduces the pulse power to 25 µJ. The IR (2 µJ, λ=6.1 µm, FWHM 300 cm^{-1}) and visible (25 µJ, λ=791.8 nm, FWHM 15 cm^{-1}) pulses are focused (focal lengths: IR: 5 cm, vis: 25 cm) to the sample where they are spatially and temporally overlapped to generate the SF light. The generated SF signal is guided through a short-pass filter (SP720), a half-wave plate, a polarizer and finally focused by a lens into a spectrometer (Acton instruments) and recorded with a CCD camera (Newton, Acton instruments).

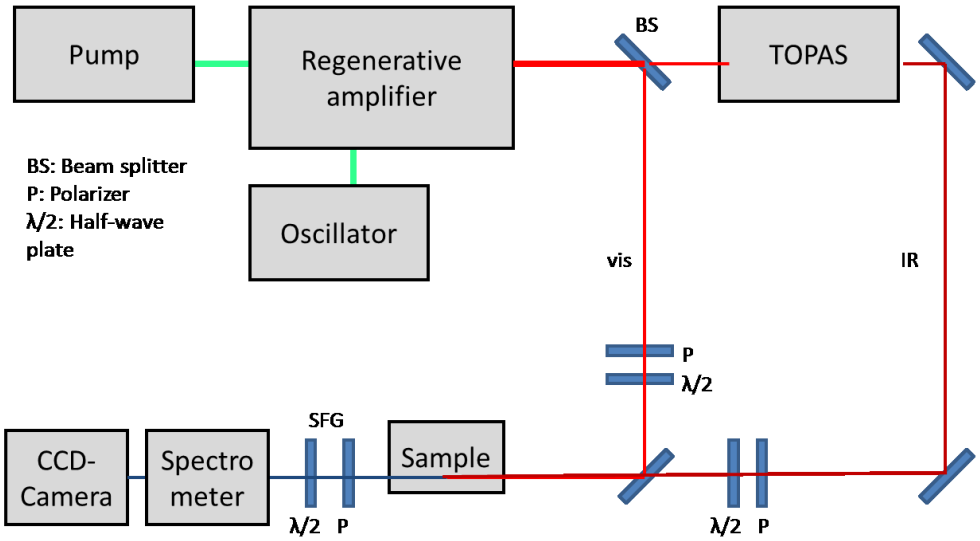


Figure 9: Homodyne SFG Setup: Part of an 800nm visible pulse pumps an optical paramagnetic amplifier (OPA) which generates broadband IR pulses. The remaining visible pulse is narrowed by an Ethalon and subsequently overlapped in space and time with the IR pulse. The generated Sum Frequency signal is spectrally dispersed by a spectrometer and detected by a CCD camera

3.4 SFG spectroscopy in the amide-I region

To acquire SFG spectra in the amide-I region the IR path must be purged with dry air or nitrogen to avoid absorption of the IR beam due to water vapor. Due to the small differences in wavelength between the SFG signal (700 nm-710 nm, this depends on the frequency of the visible pulse and may vary between different SFG setups) and the visible laser pulse (791.8 nm 15 cm^{-1} FWHM) the short-wave pass filter (SP 720) in front of the CCD camera might need to be fine-tuned to cut off all residual visible contributions while completely passing the SFG signal. This can be achieved by rotating the filter around its central axis.

3.5 SFG spectroscopy of prism interfaces at near-total internal reflection

SFG spectroscopy using a CaF_2-prism as solid support requires some adaptations of the SFG setup. It might be necessary to reduce laser intensities to avoid the generation of white light in the prism. Furthermore, it might occur that residual intensities of the signal or idler beams from TOPAS induce bulk four-wave-mixing $\chi^{(3)}$ processes. This can be avoided by placing additional filters in the IR pathway which reduces the intensities of residual signal and idler contributions and separates the IR pulse from residual signal and idler pulses in time. Reference spectra can be obtained either by coating a narrow silver or gold film at one side of the prism or by using a separate reference prism. Employing the latter method brings advantages but also requirements. It must be assured that the reference prism and the experimental prism are mounted at the same position. Using an uncoated prism for the experiments decreases the non-resonant background in PPP polarization significantly. This is probably due to leaked silver or gold atoms even at the uncoated clean spots of CaF_2-prisms with narrow coated stripes.

3.6 Data collection and processing

Spectra were collected for 10 minutes at the water/air interface or, for longer timescales, in kinetic mode for CaF_2 water interface. The kinetic mode (or accumulation) is important for longtime measurements to avoid saturation of the CCD camera which occurs above ~20,000 counts.

For all samples and laser polarization combinations a background must be collected. The IR path was blocked to allow only the visible pulse.

At the beginning of each measurement a reference spectrum for each polarization combination must be acquired. For water air interface measurements, this could be either a clean gold or silver mirror or a z-cut quartz crystal. For CaF_2 water interface a gold or silver coated prism was used for reference spectra.

After subtraction of the background the spectra were divided by the reference spectrum of the according polarization combination.

If needed the data were fitted using Eq 4 in which the second order susceptibility $\chi^{(2)}$ is considered to have a non-resonant and a resonant part which is considered as a set of Lorentzians:

$$\chi^{(2)} = \chi_{NR}^{(2)} e^{i\phi} + \sum_q \frac{A_q}{\omega_{IR} - \omega_q - i\Gamma_q} \qquad \text{Eq 4}$$

Here, ϕ is the phase of the non-resonant background, A_q, ω_q and Γ_q are the qth normal mode's amplitude, frequency and damping factor respectively.

4.Results

4.1 IM30 triggers membrane fusion in cyanobacteria and chloroplasts

The experimental SFG part of the following publication was part of this thesis. This publication was submitted to Nature communications in October 2014 and accepted in March 2015. The work was a cooperation between the Institut für Pharmazie und Biochemie which initiated it, the Institut für Zoologie and the Institut für Biophysik (all at the Johannes Gutenberg University in Mainz), the Max Planck Institute for Polymer Research in Mainz and the Van't Hoff Institute for Molecular Science at the University of Amsterdam.

nature COMMUNICATIONS

ARTICLE

Received 28 Oct 2014 | Accepted 25 Mar 2015 | Published 8 May 2015

DOI: 10.1038/ncomms8018

IM30 triggers membrane fusion in cyanobacteria and chloroplasts

Raoul Hennig[1], Jennifer Heidrich[1], Michael Saur[2], Lars Schmüser[3], Steven J. Roeters[4], Nadja Hellmann[5], Sander Woutersen[4], Mischa Bonn[3], Tobias Weidner[3,6], Jürgen Markl[2] & Dirk Schneider[1]

The thylakoid membrane of chloroplasts and cyanobacteria is a unique internal membrane system harbouring the complexes of the photosynthetic electron transfer chain. Despite their apparent importance, little is known about the biogenesis and maintenance of thylakoid membranes. Although membrane fusion events are essential for the formation of thylakoid membranes, proteins involved in membrane fusion have yet to be identified in photosynthetic cells or organelles. Here we show that IM30, a conserved chloroplast and cyanobacterial protein of approximately 30 kDa binds as an oligomeric ring in a well-defined geometry specifically to membranes containing anionic lipids. Triggered by Mg^{2+}, membrane binding causes destabilization and eventually results in membrane fusion. We propose that IM30 establishes contacts between internal membrane sites and promotes fusion to enable regulated exchange of proteins and/or lipids in cyanobacteria and chloroplasts.

[1] Institut für Pharmazie and Biochemie, Johannes Gutenberg-Universität Mainz, Johann-Joachim-Becher-Weg 30, 55128 Mainz, Germany. [2] Institut für Zoologie, Johannes Gutenberg-Universität Mainz, Johannes-von-Müller-Weg 6, 55128 Mainz, Germany. [3] Max Planck-Institut für Polymerforschung, Ackermannweg 10, 55128 Mainz, Germany. [4] Vant Hoff Institute for Molecular Sciences, University of Amsterdam, Science Park 904, 1098 XH, Amsterdam, The Netherlands. [5] Institut für Molekulare Biophysik, Johannes Gutenberg-Universität Mainz, Jakob-Welder-Weg 26, 55128, Mainz, Germany. [6] Department of Chemical Engineering, University of Washington, Seattle, Washington 98195, USA. Correspondence and requests for materials should be addressed to D.S. (email: Dirk.Schneider@uni-mainz.de).

Engulfment of an early photosynthetic cyanobacterium by a non-photosynthetic cell resulted in the evolution of present chloroplasts, the plant-specific cell organelles where photosynthesis takes place. Both chloroplasts and cyanobacteria contain two internal membranes: the inner envelope (IE) in chloroplasts, which corresponds to the cyanobacterial cytoplasmic membrane (CM), as well as the thylakoid membrane (TM), a unique internal membrane system harbouring the complexes of the photosynthetic electron transfer chain. Despite their apparent importance, remarkably little is known about the biogenesis and maintenance of TMs. Processes involved in TM biogenesis, such as protein, lipid and pigment synthesis, as well as their processing, transport and assembly, have to be controlled and aligned to finally assemble this complex membrane system. Vesicle fission at the IE and vesicle fusion have been implicated in the *de novo* formation of the TM network in greening chloroplasts[1,2]. Synthesis of galactolipids, important TM building blocks, is localized within the chloroplast IE membrane[3], and thus a mechanism ensuring continuous transport to the TM must exist. In cyanobacteria, the biogenesis of certain membrane proteins appears to commence within the cyanobacterial CM but continues in the TM[4,5], and thus here too a transfer mechanism must exist. Regulated material transfer between the IE/CM and the TMs have been proposed to be mediated by vesicular transport or via direct connections between the membrane systems[6,7]. Direct connections between the IE and TM have been observed in chloroplasts[8,9], and TM membrane reorganization is also suggested to involve dynamic fission and fusion events, as found in the membranes of the mitochondria, endoplasmic reticulum and the Golgi apparatus[10]. In the cyanobacterium *Synechocystis* sp. PCC 6803, direct connections between the CM and TM have also been observed[11]. Thus, despite the lack of definitive proof, the existence of direct membrane connections in chloroplasts and cyanobacteria cannot be simply dismissed. Whereas in cold-adapted chloroplasts vesicles appeared, which resemble COPII vesicles found in the cytosol[1,12,13], the existence of vesicles in cyanobacteria remains the subject of debate[14]. However, in recent bioinformatics analyses, chloroplast-localized components of a vesicular transfer system have been identified, and some of the identified proteins are also conserved in cyanobacteria[15,16]. Thus, several recent observations indicate that vesicular transfer processes are potentially involved in TM biogenesis and maintenance. It should be noted that both direct membrane fusion and vesicle transfer likely involve proteins, which mediate membrane fusion. Although fusion proteins have been well characterized for the exocytotic pathway in eukaryotes (for example, SNARE proteins)[17] and viral entry mechanisms[18], a protein-triggering membrane fusion in chloroplasts and/or cyanobacteria has yet to be identified. Interestingly, depletion of the protein IM30, a ~30-kDa protein conserved in chloroplast and cyanobacteria, results in a significant reduction of TMs and loss of cell viability in both chloroplasts and cyanobacteria[19–21], indicating the involvement of IM30 in TM biogenesis and/or maintenance. Despite the lack of a transmembrane domain, IM30 has been suggested to interact with membranes, initially based on the observation that IM30 is in close proximity to membranes in immuno-gold analyses and its co-sedimentation with membranes during ultracentrifugation[21–24]. However, given that IM30 forms oligomeric structures, with molecular masses exceeding 2 MDa (ref. 25), these protein structures will trivially co-sediment with membranes. The closely related phage shock protein A (PspA) of *Escherichia coli* has been demonstrated to bind to the negatively charged lipid phosphatidylglycerol (PG)[26], and thus IM30 binding to the negatively charged PG appears to be likely[27]. Mainly based on analyses of IM30-depleted chloroplasts and cyanobacteria, diverse functions have been suggested for IM30. An *Arabidopsis thaliana* mutant lacking IM30 was not only defective in TM biogenesis but cold-induced vesicles were also no longer observed at the IE, indicating involvement of the protein in vesicle formation[19]. Consequently, the name 'vesicle-inducing protein in plastids 1' (Vipp1) has been suggested. However, further studies have indicated that IM30 is involved in maintaining the structure and stability of internal membranes in chloroplasts and cyanobacteria, similar to its bacterial homologue PspA, and thus has a more protective function[27,28]. In addition, IM30 has been found to enhance protein translocation via the chloroplast Tat pathway[29] and involvement in photosystem biogenesis has been suggested[21,30–32]. Currently, no model combining these diverse IM30 functions has been proposed, and the exact role of IM30 remains enigmatic.

Here we show that IM30 binds as an oligomeric ring in a well-defined geometry specifically to membranes containing anionic lipids. Triggered by Mg^{2+}, membrane binding eventually results in membrane fusion, and IM30-established membrane contacts and fusion eventually allow dynamic remodelling of TM liposomes, which we propose may permit regulated exchange of proteins and/or lipids between internal membranes in cyanobacteria and chloroplasts.

Results

IM30 binds to the negatively charged TM lipids PG and SQDG. As the exact mode of IM30 binding to membranes still is unknown, we first investigated the co-localization of IM30 with TMs more dependably by sucrose density (SD) gradient centrifugation (Fig. 1a). Although the IM30 oligomers were found in the pellet fraction, a significant portion of the protein co-localized with TMs, pointing to a direct interaction of IM30 with cyanobacterial membranes. Similarly, heterologously produced IM30 (Supplementary Fig. 1) co-migrated consistently with urea-washed TMs (Fig. 1b), whereas in the absence of TMs, IM30 was found exclusively in the pellet fraction (Fig. 1c). Clearly, the membrane-binding properties are preserved in the recombinant IM30, so that the IM30-binding affinity to liposomes with different lipid compositions could be readily determined, using this purified IM30.

Binding of IM30 to model membranes derived from natural TM lipids was analysed by centrifugation on 5–50% SD gradients. The fluorescent probe [(2S)-3-[2-aminoethoxy(hydroxy) phosphoryl]oxy-2-[6-[(4-nitro-2,1,3-benzoxadiazol-7-yl)amino] hexanoyloxy]propyl] hexadecanoate (C6-NBD-PE) was always incorporated into the liposomes, to allow optical detection of the liposomes. The main TM lipids are the uncharged galactolipids monogalactosyldiacylglycerol (MGDG) and digalactosyldiacylglycerol (DGDG), which account for 50–60% and ~20% of the lipids, respectively, whereas the two negatively charged lipids sulfoquinovosyldiacylglycerol (SQDG, ~12%) and PG (~8%) constitute minor fractions[33–35]. IM30 alone localized within the lower half of this SD gradient, mainly pelleting, because of its propensity to form high-molecular-mass oligomers (Fig. 1d). However, a fraction of the protein consistently co-localizes with liposomes, containing either solely the anionic lipids SQDG and PG or a combination of neutral and anionic lipids (Fig. 1e, f). In contrast, co-localization of IM30 with liposomes composed of the two neutral galactolipids MGDG and DGDG was not observed. Thus, IM30 interaction with membranes appears to be mediated by the anionic lipids PG and SQDG.

To further elucidate whether co-localization with membranes was caused by a direct membrane interaction, we investigated interaction of IM30 with liposomes by using the fluorescent probe Laurdan, which monitors changes in the membrane lipid order, for example, caused by peripheral binding of a protein (Fig. 1g).

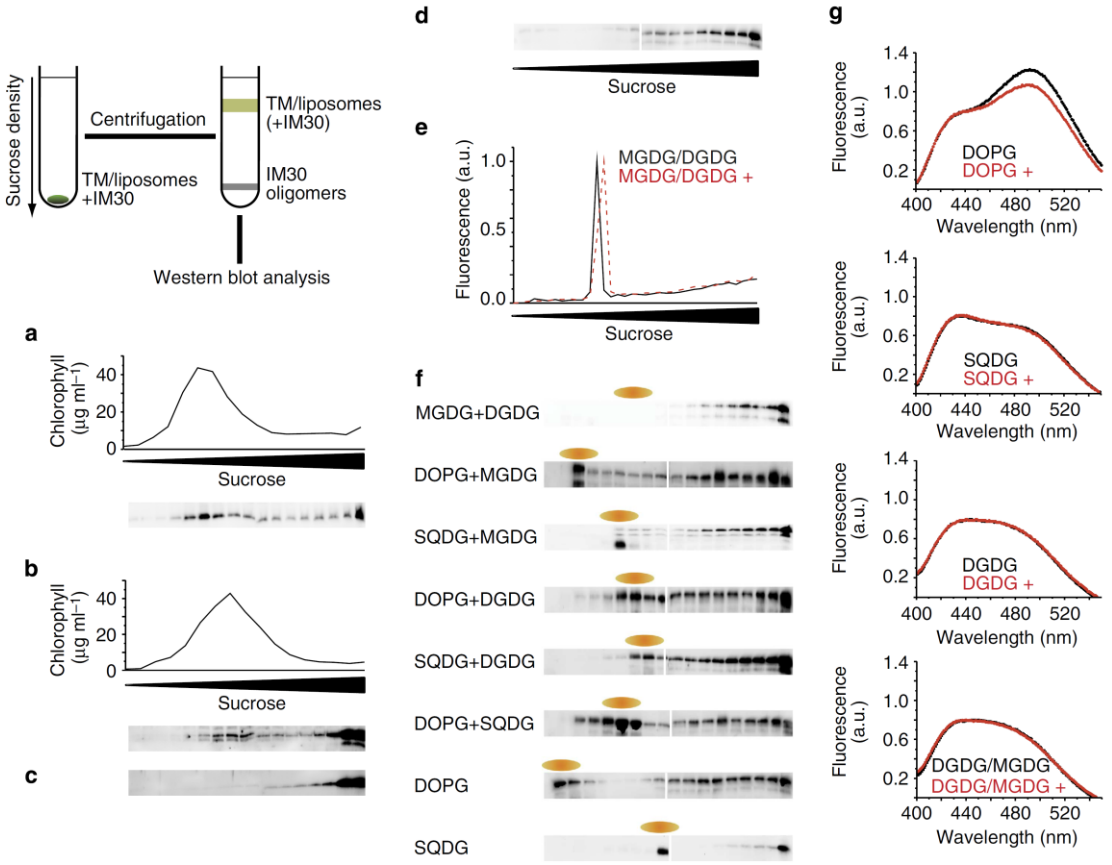


Figure 1 | IM30 interaction with membrane lipids. SD gradient flotation analysis of (**a**) native IM30 attached to *Synechocystis* membranes, (**b**) heterologously expressed IM30 added to urea-washed membranes and (**c**) solely heterologously expressed IM30. Samples were separated on 34–68% SD gradients, fractionated and analysed by SDS-PAGE and western blot analyses, using an anti-IM30 antibody. Localization of the membranes within the gradient was determined by monitoring chlorophyll absorption (shown on top of the western blot in **a**,**b**). (**d**–**f**) Liposomes mixed with IM30 (**e**,**f**) as well as solely IM30 (**d**) were separated on a 5–50% SD gradient prior to analysis. The fluorescent probe C6-NBD-PE was incorporated into the liposomes (1:500) for detection, and the position of the liposomes within the gradient was determined by fluorescence spectroscopy. The localization of the liposomes is shown for MGDG/DGDG-containing liposomes and thereafter only indicated above each western blot. The analyses were repeated three times and representative western blots are shown. Full scans of western blots are shown in Supplementary Fig. 8. (**g**) Peripheral binding of IM30 to liposomes was additionally followed by monitoring changes in the Laurdan fluorescence emission after excitation at 350 nm. Normalized fluorescence emission spectra are shown in the absence (black) and presence of 2.5 µM IM30 (red). Decreased fluorescence emission at 490 nm, that is, increased lipid acyl chain order, indicates peripheral binding of IM30. The experiment was performed four times. See also Supplementary Fig. 2.

The decrease of the Laurdan fluorescence emission at 490 nm indicates that IM30 affects the lipid order by binding to pure PG membranes. In contrast, addition of IM30 did not significantly influence the Laurdan signal monitored in pure SQDG membranes, probably due to the fact that the acyl chain order is already high in SQDG membranes, and binding of IM30 might therefore not induce a further decrease in the 490-nm emission. Nevertheless, when using another (non-physiological) neutral lipid as a background, interaction of IM30 with both, PG- and SQDG-containing membranes, was again observed, but not with MGDG- or DGDG-containing membranes (Supplementary Fig. 2). Together, the SD-gradient and Laurdan experiments strongly indicate that IM30 directly interacts with negatively charged TM lipids, and that membrane-bound IM30 remodels the membrane structure.

IM30 affects the membrane order and promotes liposome fusion. As peripheral binding of IM30 to membrane surfaces is expected to affect surface properties, we monitored liposome–liposome interactions in the absence and presence of IM30, with increasing concentrations of Na^+, K^+, Ca^{2+} or Mg^{2+} (Fig. 2a). Divalent cations severely affect galactolipid-containing liposomes and are known to induce liposome aggregation[36,37]. As expected, addition of increasing Ca^{2+} and Mg^{2+} concentrations resulted in liposome aggregation, starting at 6 and 17.5 mM, respectively. Although addition of IM30 alone had no visible effect, the simultaneous addition of IM30 and Ca^{2+} or Mg^{2+} caused interactions to start at significantly lower ion concentrations. However, Mg^{2+}-induced liposome contacts differed dramatically in the presence of IM30, as the scattering signal decreased at further increasing Mg^{2+} concentrations until it levelled off. In contrast, Na^+ and K^+ had no significant effect on liposome aggregation. Together, these results indicate that IM30 functionally cooperates with Mg^{2+} at membranes. However, liposome aggregation could simply reflect liposome–liposome contacts but could also indicate liposome fusion.

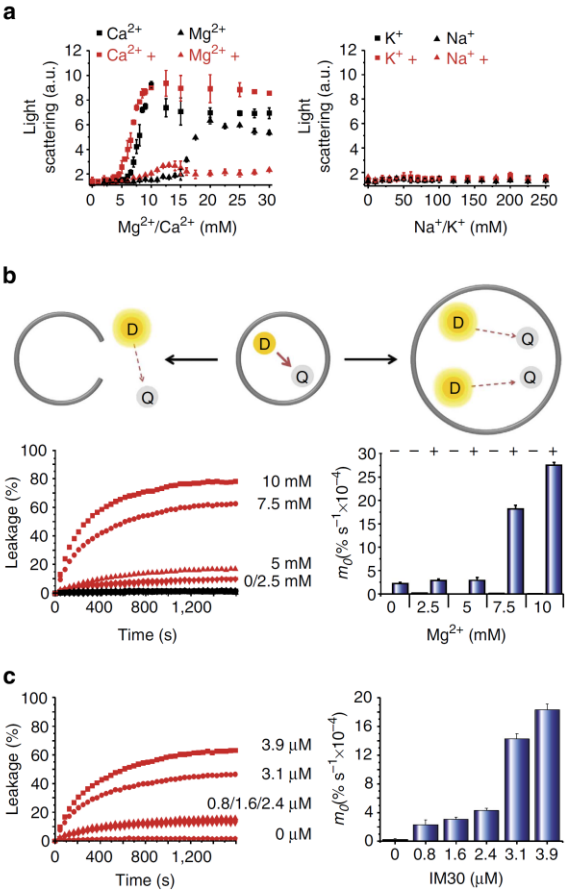


Figure 2 | IM30 influences MGDG/PG liposome aggregation induced by divalent cations. (**a**) MGDG/PG (60:40) liposome aggregation determined in the presence or absence of IM30 and increasing Ca^{2+} and Mg^{2+} (left) or Na^{+} and K^{+} (right) concentrations. Error bars: s.d. ($N=4$). (**b**) Dye-loaded MGDG/PG (60:40) liposomes were mixed with 3.9 μM IM30 and increasing Mg^{2+} concentrations as indicated, and the increase in ANTS fluorescence emission due to content release or vesicle fusion was monitored over 1,600 s (left). In **c**, fluorescence emission was monitored after mixing liposomes with 7.5 mM Mg^{2+} and increasing IM30 concentrations (left). In the right panels of **b** and **c**, the increases in fluorescence emission in the initial phases, calculated by linear regression, are shown. Error bars: s.d. ($N=3$).

To further probe the impact of IM30 on the membrane structure and stability, we next monitored the release of soluble liposome content upon IM30 addition (Fig. 2b). Adding only IM30 to MGDG/PG liposomes loaded with a fluorophore/quencher pair without the addition of ions resulted in membrane binding, but did not significantly destabilize the membrane. Furthermore, over a concentration range of 0–10 mM, the addition of solely Mg^{2+} also did not induce liposome leakage. However, the addition of increasing Mg^{2+} concentrations to IM30-containing liposomes resulted in an increased destabilization of the membrane and in liposome content release. The observed leakage rates strongly depend on both Mg^{2+} and protein concentrations (Fig. 2b,c). Clearly, IM30 and Mg^{2+} destabilize the vesicular membrane structure in a cooperative manner. Sum frequency generation (SFG) vibrational spectroscopy of a PG lipid monolayer further corroborates the IM30-induced perturbation of the lipid layer structure (Supplementary Fig. 3) by showing that the interaction with IM30 decreases the lipid acyl chain ordering and affects the lipid orientation. In summary, the above observations show that IM30 interacts with membranes and together with Mg^{2+} promotes membrane destabilization.

Interaction with and the destabilization of membranes are both essential for an effective trigger of membrane fusion. Thus, to explore a potential fusogenic activity of IM30, we performed a membrane fusion assay using labelled MGDG/PG liposomes. Fusion of labelled with unlabelled liposomes resulted in an increase of the donor fluorescence (Fig. 3). Neither IM30 nor Mg^{2+} alone induced liposome fusion, whereas IM30-mediated membrane fusion was observed above 5 mM Mg^{2+} (Fig. 3a,b). Higher Mg^{2+} concentrations resulted in a dramatic increase of the fusion rates, and fusion efficiencies of ~100% were reached at 10 mM Mg^{2+} and 2.5 μM IM30. Thus, the membrane fusion activity of IM30 clearly depends on Mg^{2+}. It is interesting to note that this membrane fusion activity was not observed for an IM30 mutant, which still forms the oligomeric ring structures at membrane surfaces observed for the wild-type protein (Supplementary Fig. 4).

IM30-triggered liposome fusion was further analysed by single-particle tracking analysis (Fig. 3c). Before addition of Mg^{2+} and IM30, the liposomes had a mean diameter of 130 nm. Addition of Mg^{2+} and 1.3 μM IM30 immediately resulted in a slight increase of the mean vesicle diameter, and with increasing incubation time, the liposome size distribution shifted to larger diameters due to liposome fusion.

IM30 triggers membrane fusion by membrane destabilization. Cryo-electron microscopy revealed that the diameter of the analysed MGDG/PG liposomes is typically less than 200 nm (Fig. 4a), in agreement with the single-particle tracking analysis. Upon addition of either IM30 or Mg^{2+}, the structure of these liposomes did not change significantly (Fig. 4b,c). Despite the observation of loose contacts between different liposomes in the presence of IM30, strong aggregation or fusion was not visible. As depletion of IM30 in *A. thaliana* chloroplasts has caused decoupling of TMs and the IE[28], IM30 potentially mediates a loose contact between neighbouring membranes. Notably, in the presence of high Mg^{2+} concentrations, liposomes were essentially no longer found inside the holes of the holey carbon film, but mostly on the carbon itself. However, upon addition of IM30 and Mg^{2+}, the liposomes were found in large amounts inside the holes (Fig. 4d,e), suggesting that attachment of IM30 affects the surface properties of the liposomes. Furthermore, the size of the liposomes had increased significantly, indicating fusion events, and extended contact surface areas between vesicles were apparent.

Thus far, the observations indicate that IM30 is surface active, perturbs the structure of lipid bilayers and induces membrane fusion. However, although membrane fusion is clearly observed, the assays presented in Figs 2 and 3 cannot distinguish between membrane fusion directly controlled by IM30 and spontaneous fusion induced by membrane destabilization. The formation of large multilamellar liposomes upon addition of IM30 and Mg^{2+}, as observed by cryo-electron microscopy (Fig. 4d,e), indicated that IM30 is not a canonical membrane fusion protein. Liposome fusion mediated by classical membrane fusion proteins would result in unilamellar liposomes with increased size. In contrast, IM30 appears to enable membrane fusion by membrane destabilization and remodelling. To test whether liposome fusion is indeed caused by IM30-triggered membrane destabilization, we

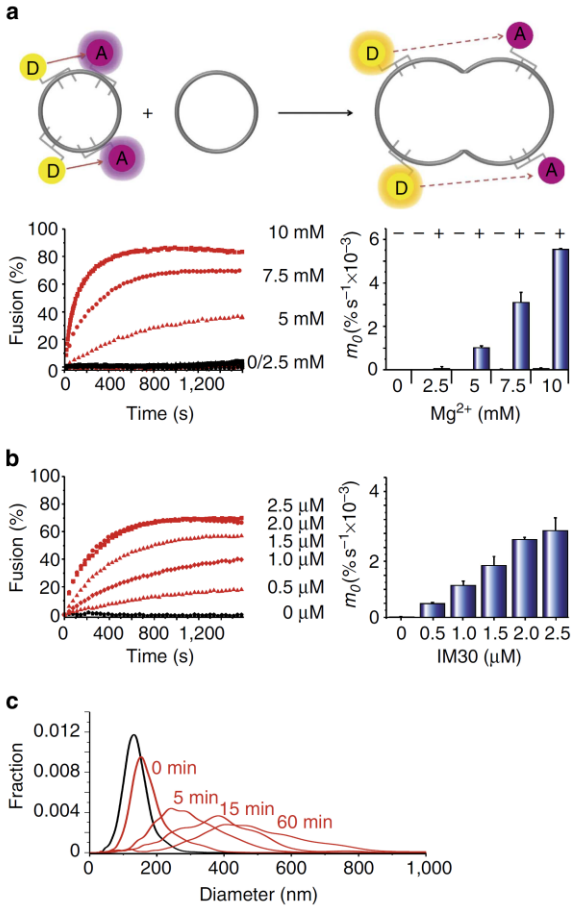


Figure 3 | IM30 and Mg^{2+}-mediated vesicle fusion. (**a**) MGDG/PG liposomes, containing the donor and acceptor dye, were mixed with non-labelled liposomes, and the increase in donor fluorescence emission was monitored over 1,600 s after addition of 2.5 µM IM30 and increasing Mg^{2+} concentrations (left). In **b**, the liposomes were mixed with increasing IM30 concentrations in the presence of 7.5 mM Mg^{2+}. 0 and 100% fusion rates were determined in control experiments to calculate the fusion efficiency for each sample. The right panels of **a** and **b** show the increases in fluorescence emission in the initial phases calculated by linear regression. Error bars: s.d. ($N=3$). (**c**) The size distribution of MGDG/PG liposomes was determined by nanoparticle tracking analysis. The liposomes prepared by extrusion had a mean diameter of 130 nm (black). Addition of Mg^{2+} and 1.3 µM IM30 immediately slightly increased the mean diameter (red, 0 min). With increasing incubation time (red curves), the liposome size distribution shifted to larger diameters. The experiment was repeated two times.

prepared liposomes labelled with NBD in both leaflets of the lipid bilayer. As shown in Fig. 4f, the addition of sodium dithionite to the liposomes selectively quenched the NBD fluorescence in the outer membrane leaflet. In contrast, the NBD probes of the inner leaflet are not quenched by sodium dithionite as the membrane is impermeable for the quencher. In contrast, when the membrane is solubilized by addition of a detergent, the fluorescence is completely quenched. Also, addition of IM30 further reduced the NBD fluorescence, indicating that the fluorescent probes located in the inner membrane leaflet were reached by the quencher. Thus, addition of IM30 disturbs the membrane structure sufficiently to allow sodium dithionite to enter the liposomes. Apparently, IM30-triggered membrane remodelling and destabilization results in spontaneous membrane fusion, which explains the formation of multilamellar liposomes.

IM30 binds as a ring to lipid bilayers. Throughout the samples analysed by cryo-electron microscopy, side views of areas with additional electron density at the membrane surface are visible (arrows in Fig. 4d,e). This extra electron density matches in size and shape oligomeric IM30 rings attached to the membrane surfaces, as shown in Fig. 5a,b. To further corroborate this membrane-binding structure, the structure of IM30 bound to model MGDG/PG lipid monolayers at the air–water interface was studied using SFG vibrational spectroscopy (Fig. 5c). The peak near 1,740 cm^{-1} in the surface-specific vibrational spectra is caused by lipid carbonyls; the peak near 1,650 cm^{-1} by backbone amide vibrations of IM30. The large magnitude of the protein signal with respect to the lipid signal confirms stable and ordered binding of IM30 to the surface of MGDG/PG monolayers, and the 1,650 cm^{-1} peak position indicates a mostly helical secondary structure. The SFG-spectrum can only be reproduced by simulation with the IM30 protein monomers oriented predominantly perpendicular to the membrane surface, indicating that IM30 binds as ring structures to membrane surfaces (Supplementary Fig. 5). In the transmission electron microscope images, superimposition of the extra electron density at the vesicle surface with a three-dimensional (3D) reconstruction of an IM30 ring agrees very well with the concept of perpendicular binding of rings (Fig. 5d). However, in case of the wild-type IM30 protein, we observed IM30 rings only in a few sparse areas of the liposome. In contrast, in case of the fusion-incompetent IM30 mutant (Supplementary Fig. 4b), the liposome surface is more extensively decorated with IM30 rings. This indicates that the oligomeric ring structure is destabilized and disassembles during or after liposome fusion. The role of the Mg^{2+} in the binding, alignment and ring formation of IM30 at the membrane surface cannot be determined with certainty yet. However, the observed decrease (by a factor >3) of the protein SFG-response in the absence of Mg^{2+} (Supplementary Fig. 6) suggests that Mg^{2+} is able to provide the critical membrane-protein link in the surface-binding process.

Discussion

Understanding membrane fusion mechanisms has been an important focus of biochemistry and biophysics research in the past years. Accordingly, proteins involved in fusion processes, such as exocytotic pathways in eukaryotes (SNARE proteins) and viral entry have received substantial interest and are the topics of increasingly dynamic fields of research. Membrane fusion typically not only requires the tethering activity of proteins but also involves significant membrane remodelling.

Membrane fusion/fission processes in the endoplasmic reticulum, the Golgi apparatus or in the mitochondria involve well-defined protein machineries, which help to overcome the energy barrier associated with such processes[38–40]. Although it is clear that fusion proteins are essential for the formation and maintenance of TMs, surprisingly, not a single fusion-inducing protein has been identified in TM-containing organisms yet. Fusion of individual TM sheets and of the CM and TMs has previously been observed in chloroplasts and cyanobacteria[9,11], and in chloroplasts, with changing light conditions the architecture of the TM network rearranges, involving membrane fission and fusion events[10]. As in chloroplasts, also in cyanobacteria, the TM network can be highly interconnected[41], and it is thus likely that the TM system of cyanobacteria is as

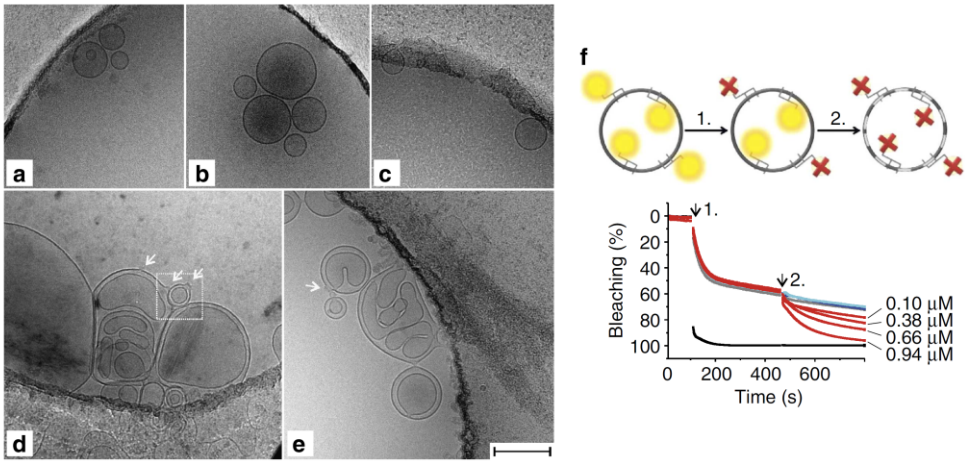


Figure 4 | IM30 disturbs the membrane structure and thereby triggers membrane fusion. (**a**–**e**) Cryo- transmission electron microscope images of pure MGDG/PG liposomes (**a**), of liposomes mixed with solely IM30 (**b**) or Mg^{2+} (**c**), as well as of liposomes mixed with both, IM30 and Mg^{2+} (**d**,**e**). Notably, at high Mg^{2+} concentrations, vesicles were essentially no longer found inside the holes of the holey carbon film, but mostly on the carbon itself. In contrast, with both IM30 and Mg^{2+} present, the vesicles were found in large amounts inside the holes (**d**,**e**), suggesting that attachment of IM30 affects the surface properties of the liposomes. Furthermore, note the large compartmentalized liposomes in **d** and **e**, as well as the extended connection between two liposomes in **d**. Arrows indicate extra electron density found at the liposome surface. The area boxed in **d** is enlarged in Fig. 5. Scale bar, 200 nm. The magnification was identical in **a**–**e**. EM micrographs were taken after 1 min incubation with Mg^{2+} and/or IM30. (**f**) Sodium dithionite was added to NBD-labelled MGDG/PG liposomes (1), resulting in immediate quenching of the NBD fluorescence of the outer membrane leaflet. (2) When nothing further (light grey), solely 0.94 µM IM30 (light blue) or solely 11 mM $MgCl_2$ (dark blue) was added to the liposomes, the fluorescence did not considerably decrease further. After addition of increasing IM30 concentrations at a constant $MgCl_2$ concentration of 7.5 mM (red curves), the NBD-fluorescence decreased further in an IM30 concentration-dependent manner. Addition of Triton X-100, which dissolves the liposomal membranes, results in an immediate quenching of the entire NBD-fluorescence (black). The experiment was repeated two times.

dynamic as in chloroplasts and undergoes remodelling under changing growth conditions or in the course of development[6]. However, proteins involved in this dynamic rearrangement of membranes in chloroplasts have not yet been identified[42]. The recent observation that in maize mesophyll cell chloroplasts, which harbour a rather extended TM network, the amount of IM30 is ~10-fold higher than in maize bundle sheet chloroplasts indicates a crucial involvement of IM30 in TM biogenesis and maintenance[43]. Our results show that IM30 is indeed a potential key player for TM formation, rearrangement and maintenance in chloroplasts and cyanobacteria, by triggering membrane fusion. IM30 is clearly able to bind, assemble and manipulate TMs and thereby triggers fusion of membrane structures. A central role of IM30 in triggering membrane fusion explains the various pleiotropic functions ascribed to IM30 in the past, mainly based on studies using higher plants, algae or cyanobacteria with decreased IM30 contents. Lipid transport[24,44], vesicle budding or fission[19], protein translocation and insertion[29,45,46], membrane modification[28,47], as well as biogenesis of photosystems[21,30,48] have been ascribed to IM30. Our study now provides a unifying explanation for the observations and proposed functions: As its homologue PspA[26], IM30 binds to negatively charged lipid bilayers and affects the membrane structure. Initiated by Mg^{2+}, IM30 can mediate membrane contacts and triggers membrane fusion, and thereby enables regulated exchange of proteins and/or lipids throughout different internal membranes. After membrane attachment, the IM30 ring structures appear to disassemble on the membrane surface during or after the membrane fusion event. The observation that impaired ring disassembly, as seen in case of the IM30 mutant (Supplementary Fig. 4), coincides with inhibited membrane fusion, indicates that ring disassembly is essential for initiating membrane fusion. In our *in vitro* system, membrane fusion is initiated by destabilization of the lipid bilayer structure resulting in spontaneous membrane fusion. However, it appears likely that additional proteins with tethering activity are involved *in vivo* to induce and control selective membrane fusion. These proteins might involve dynamin-related GTPases, such as FZL, or SNARE-like proteins, which have been identified in chloroplasts and cyanobacteria[15,16,49,50]. The notion that the IM30 fusion activity depends on Mg^{2+} agrees well with the observation that Mg^{2+} is generally involved in enzyme regulation within chloroplasts and cyanobacteria[51]. As IM30 is conserved in both chloroplasts and cyanobacteria, the biogenesis and maintenance of TMs may well be controlled by IM30-mediated membrane remodelling.

Methods

IM30 expression and purification. The *Synechocystis* IM30 was heterologously expressed in *E. coli* BL21 (DE3) cells as an N-terminally deca His-tagged protein, using the plasmid pRSET-SynVipp1 (refs 21,25). When the liquid *E. coli* culture had reached an OD_{600} of ~0.8, protein expression was induced by addition of 0.5 mM isopropyl-β-D-thiogalactopyranoside. After 4 h, cells were harvested, resuspended in buffer (300 mM NaCl, 50 mM sodium phosphate, 20 mM imidazol, *pH* 7.6) and broken by ultrasonification. After centrifugation (10,000*g*, 4 °C, 15 min), the supernatant was loaded onto a nickel-nitrilotriacetic acid agarose column (Qiagen). The column was washed three times with buffer containing increasing imidazol concentrations (20, 50 and 100 mM), and the protein was finally eluted with buffer containing 500 mM imidazol. Before subsequent analyses, the protein was dialysed against 20 mM HEPES (pH 7.6). IM30 purity was confirmed on 14% SDS polyacrylamide gels (Supplementary Fig. 1). After purification, the protein was dialysed against 20 mM Hepes-KOH, pH 7.6 (HEPES).

SD gradient centrifugations. SD gradients were produced and fractionated using the Gradient Station (Biocomp). For analyses, all gradients were centrifuged at 100,000*g* for 6 h and immediately fractionated. Individual fractions were analysed by fluorescence spectroscopy and fractionated proteins were separated on 14% SDS–PAGE gels with subsequent immunoblot analysis, using an anti-IM30 antiserum[25] (dilution 1:1,000) and a peroxidase-coupled anti-rabbit secondary antibody (#A0545, Sigma; diluted 1:10,000). The 'ECLTM Prime Western Blotting

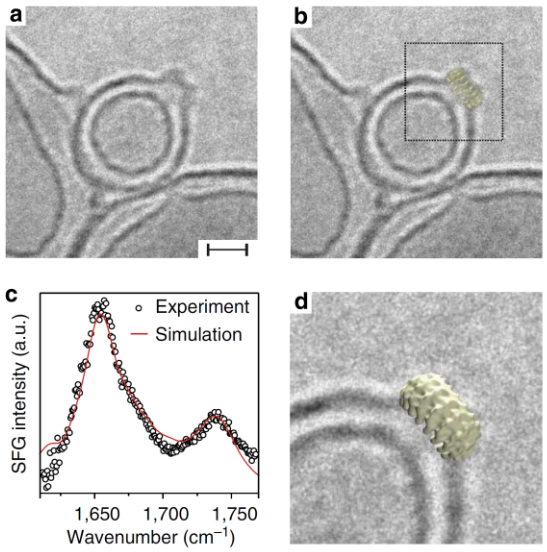


Figure 5 | IM30 rings bind perpendicularly to membrane surfaces. (**a**) Cryo-TEM image showing IM30 as extra electron density attached to a liposome. Scale bar, 20 nm. (**b**) Superposition of the extra density with the side-view of a 3D reconstructed IM30 ring being depicted in projection view (approximately 30 nm in diameter). (**c**) Experimental (circles) and calculated (line) SFG-spectra of IM30 bound to a model MGDG/PG monolayer at the air–water interface using the ssp polarization combination. The experiment was repeated two times. The calculated spectra match the experimental data well for upright monomer orientations. A tilt angle of 0° with respect to the surface normal was used for the calculation shown. (**d**) Magnification of the box in **b**, depicting the 3D reconstruction in an isosurface view. Note that the 3D reconstruction matches the extra electron density in size and shape, indicating perpendicular binding of IM30 to the membrane surface.

Detection Reagent' (GE Healthcare) was used following the manufacturer's instruction, to visualize the secondary antibody using a STELLA imaging system (Raytest).

Membrane purification. *Synechocystis* PCC 6803 cells were cultivated photomixotrophically in liquid BG11 media supplemented with 5 mM glucose[21]. Erlenmeyer flasks were incubated at 30 °C on an orbital shaker (150 r.p.m.) under fluorescent white light at a light intensity of 30 µmol photons per m^2 per s^1 (ref. 21). Cells were harvested in the mid-log phase (OD_{730} of ~1). After cell lyses, the cell lysate was split into two fractions: one was further purified in HEPES buffer and the other one in HEPES buffer containing 4 M urea. Soluble and membrane-associated proteins (MF_H = HEPES sample; MF_U = urea sample) were separated by ultracentrifugation at 100,000*g* for 30 min at 4 °C. The supernatant was discarded and the MF_H pellet was resuspended in HEPES and washed twice. Finally, the pellet was resuspended at a final chlorophyll concentration of 1 mg chlorophyll per ml and mixed with 85% sucrose in HEPES buffer to reach a final sucrose concentration of 68%. Membranes, containing 200 µg chlorophyll, were loaded below a linear 34–68% SD gradient. All gradients were fractionated after centrifugation at 4 °C, followed by absorbance measurements and immunoblot analyses.

The MF_U pellet was resuspended and incubated for 30 min in HEPES buffer containing 4 M urea. Afterwards, the sample was pelleted and washed twice with HEPES buffer as described before. Subsequent sample preparation and analysis were performed as described, except that 3 µg of the purified recombinant IM30 was added to the sample before adjusting the sample to 68% sucrose.

Liposome preparation. The lipids DGDG, DOPC (dioleoylphosphatidylcholine), DOPG (dioleoylphosphatidylglycerol) and MGDG were purchased from Avanti Polar Lipids, Inc. SQDG was from Larodan Fine Chemicals.

The fluorescence dyes NBD-PE (1,2-distearoyl-*sn*-glycero-3-phosphoethanolamine-N-(7-nitro-2-1,3-benzoxadiazol-4-yl)) and LissRhodPE (Lissamin Rhodamin PE; 1,2-Dioleoyl-*sn*-glycero-3-phosphoethanolamin-*N*-(lissamin-rhodamin-B-sulfonyl) were purchased from Avanti Polar Lipids, Inc.

For liposome preparation, lipids were dissolved in organic solvents. The organic solvent was removed under a gentle stream of nitrogen gas and final traces of the solvent were removed by vacuum desiccation overnight. The dried lipid films were hydrated in HEPES buffer, resulting in formation of multi-lamellar liposomes. Unilamellar liposomes were prepared by five cycles of freezing and thawing, followed by 15 times extrusion through a 100-nm filter, using an extruder from Avanti Polar Lipids, Inc.

SD gradient analyses of IM30 binding to liposomes. A measure of 0.4 mg of the individual lipids or of lipid mixtures (50%:50%, w/w) were dissolved in organic solvent together with NBD-PE (1:500 molar ratio) before liposome preparation. After preparation of unilamellar liposomes, the size of the liposomes was adjusted to 50 nm by extrusion. Three-micrograms of protein were added, and the samples were incubated for 10 min prior to loading the sample onto the SD. For flotation analyses, 5–50% SD gradients were run at 25 °C and analysed as mentioned above.

Laurdan fluorescence measurements. To determine the concentration of the non-bilayer forming lipid MGDG tolerated in mixed liposomes, MGDG was mixed with DOPG in increasing concentrations, in 10% increments (compare Supplementary Fig. 7). The fluorescence probe Laurdan (6-Dodecanoyl-*N*, *N*-dimethyl-2-naphthylamine, from Sigma, Taufkirchen, Germany) was added to each sample at a 1:500 molar ratio. The lipid concentration was adjusted to 0.1 mM by addition of HEPES buffer. Laurdan fluorescence emission was monitored on a Horiba Scientific FluoroMax-4 at 25 °C, from 400 to 550 nm after excitation at 350 nm with both the excitation and emission slit-width set at 4 nm.

For the experiments shown in Supplementary Fig. 2, 80% DOPC lipid was mixed with 20% of one of the four thylakoid lipids (w/w). Laurdan was added to each sample at a 1:500 molar ratio. A final sample concentration of 0.1 mM lipid was reached after adding 2.5 µM IM30 or an equal volume of HEPES buffer before the measurement. Generalized polarization (GP), a measure for membrane fluidity/structure, was calculated as described in Parasassi *et al.*[52].

Liposome aggregation monitored by light scattering. As addition of divalent cations to thylakoid lipids results in membrane aggregation and an increased turbidity of the lipid solution[36,37], we have determined whether addition of IM30 influences the turbidity. Liposomes containing 60% MGDG and 40% DOPG were prepared as described. The final lipid concentration was 0.1 mM for each sample. 2.5 µM IM30 was added to each sample or solely HEPES buffer (control). Cations were added to the desired concentration (Mg^{2+}, Ca^{2+}: 0–30 mM; Na^+. K^+: 0–250 mM), the sample was gently mixed by pipetting. After incubation for 5 min at 25 °C, changes in turbidity were monitored at 600 nm.

Liposome content release. ANTS (8-aminonaphtalene-1,3,6-trisulfonic acid) and DPX (*p*-xylene-*bis*-pyridinium bromide) were purchased from Life Technologies.

To determine whether release of soluble liposome content might be induced by addition of Mg^{2+} or IM30 (or both), release of soluble liposome content was monitored following the procedure described in refs 53,54. LissRhodPE was added at a molar ratio of 1:550 to the lipid mixture (MGDG/DOPG (60:40 (w/w)) before evaporation of the organic solvent. The dried lipid film was resuspended in HEPES buffer containing 12.5 mM ANTS (fluorescent donor) and 45 mM DPX (quencher), and 100 nm unilamellar liposomes were prepared by extrusion. Free fluorophore and quencher were removed by gel filtration using a Sephadex-25 column. The lipid concentration was determined by measuring the LissRhodPE absorbance at 568 nm using the extinction coefficient $\varepsilon_{LissRhodPE}$ (88,000 l mol^{-1} cm^{-1}). 0.17 mM lipid was used in the final measurements. ANTS fluorescence at 530 nm was monitored at 25 °C after excitation at 360 nm with both the excitation and emission slit-width set at 8 nm.

Complete leakage was induced by adding 1% (v/v) Triton X-100 to preformed liposomes. The measurements were performed either in the presence of 3.9 µM IM30 and increasing $MgCl_2$ concentrations (0–10 mM from a 1 M stock solution), or in the presence of 7.5 mM $MgCl_2$ and increasing IM30 concentrations (0–3.9 µM from a 15 µM stock solution). The leakage rate was calculated by

$$\text{Leakage}(t) = \frac{I_t - I_{t,0\ \%}}{I_{t,100\ \%} - I_{t,0\ \%}} \cdot 100\ \% \qquad (1)$$

where I_t is the ANTS fluorescence after time t, $I_{t,100\%}$ is the maximal ANTS fluorescence determined using the control sample for complete leakage at time t and $I_{t,0\%}$ is the ANTS fluorescence of the negative control sample at time t.

For the quantification of the leakage rates, the first 100 s of the measurement were fitted by linear regression and the initial slope was calculated.

Liposome fusion. To determine whether IM30-induced liposome fusion depends on the Mg^{2+} concentration, fusion assays were performed and evaluated essentially as described in Meers *et al.*[55]. 0.8 µM LissRhodPE (FRET-donor) and 0.8 µM NBD-PE (FRET-acceptor) were added to 0.1 mM lipids (60% MGDG, 40% DOPG, w/w) and liposomes were prepared as described. Unlabelled liposomes were produced without the fluorescence probes. Labelled and unlabelled liposomes were mixed at a ratio of 1:9 (v/v) to achieve a final lipid concentration of 0.1 mM for

each sample. The measurements were performed either in the presence of 2.5 μM IM30 and increasing $MgCl_2$ concentrations (0–10 mM) or in the presence of 7.5 mM $MgCl_2$ and increasing IM30 concentrations (0–2.5 μM). NBD-PE fluorescence emission at 535 nm was followed at 25 °C after excitation at 460 nm, with both the excitation and emission slit-width set at 5 nm. To determine the FRET value for 100% fusion, 0.1 mM mock-fused liposomes were prepared, containing 0.08 μM of both, LissRhodPE and NBD-PE[55].

Nanoparticle tracking analysis. Unilamellar liposomes (60/40 MGDGD/DOPG with LissRhod-PE added in a molar ratio of 500:1) of 100 nm size were prepared by extrusion in HEPES buffer, as described. A 100-μl reaction stock, containing 0.15 mM lipid, 1.3 μM IM30 and 7.5 mM $MgCl_2$, was prepared. After the desired reaction time, 10 μl from this stock was diluted 1:2,000 into buffer in order to stop the reaction. Particle size distributions were determined by single-particle tracking based on the fluorescence signal (NanoSight LM10, Malvern Instruments, $\lambda_{ex} = 532$ nm, emission filter 565 nm long-pass, 25 °C). For each sample, 30 s of particle movement were recorded ten times and the size distribution was calculated as the average from the distribution for each single measurement, as calculated by the software (NTA2.3). The final distribution for each sample was normalized with respect to the derived particle concentration.

Membrane destabilization followed by NBD fluorescence. For measuring IM30-induced membrane destabilization, unilamellar liposomes (60%/40% MGDG/DOPG (w/w) with NBD-PE added in a molar ration of 500:1) were prepared as described. The measurement was performed at 10 °C using a lipid concentration of 0.15 mM. The time-resolved NBD-PE fluorescence was recorded at 535 nm after excitation at 460 nm with the slit-width set to 4 nm. After 120 s, sodium dithionite was added from a 1 M stock to a final concentration of 10 mM. After 300 s, IM30 and/or $MgCl_2$ were added, respectively, and fluorescence was recorded for further 300 s. As positive control, 1% (v/v) Triton X-100 was added before the measurement to dissolve the membranes. All samples were incubated at 10 °C for 5 min before the measurement.

Transmission electron microscopy. Preparation of samples for cryo-electron microscopy was conducted using 3.5 μl of sample on C-flat CF-2/2-2C-T holey carbon grids (Protochips Inc.). Before sample preparation, the grids were washed for 45 min in chloroform, rinsed in acetone, dried and finally glow discharged for 30 s at 25 mA in an Emitech K100X glow discharge system (Quorum Technologies Ltd). Samples containing $MgCl_2$ were vitrified for either 1 or 30 min after the addition of $MgCl_2$ to the respective samples. Images were taken on a FEI Tecnai 12 transmission electron microscope (acceleration voltage: 120 kV; Cs: 6.3; nominal magnification: ×71,540; nominal underfocus: 1–3 μm) equipped with a TVIPS TemCam-F416 4K CCD camera.

A low-resolution structure of IM30 rings was provided by 3D electron microscopy of negatively stained recombinant protein, as described in Saur *et al.*[56] For negative staining, continuous carbon EM grids (Protochips Inc.) were negatively glow discharged at 25 mA for 30 s in an Emitech K100X glow discharge system (Quorum Technologies Ltd). Five microlitres of the sample were placed onto an EM grid and incubated for 1 min. Subsequently, the sample was washed three times with water by dragging three 20 μl water drops across a strip of Parafilm (Bemis Europe). After side-blotting the grid with Whatman filter paper No. 4 (GE Healthcare Europe GmbH), a 5-μl droplet of 2% uranyl formate was taken up with the grid and side-blotted after 1 min. Subsequently, the sample was imaged in a FEI Tecnai 12 microscope (120 kV, Cs = 6.3, nominal magnification = ×71,540; nominal underfocus: 0.5–1.5 μm; FEI Company) equipped with a TVIPS TemCam-F416 4K CCD camera (TVIPS GmbH). Particles were manually selected from the micrographs and classified in IMAGIC5 (Image Science Software GmbH) using reference-free alignment, multivariate statistical analysis and multiple reference alignment. The heterogeneous data set was initially separated using simulated models of different ring sizes—which were built using stacked 2D class averages—as references in the multirefine procedure of EMAN1.9 (ref. 57). For further separation, one ring monomer of the resulting 3D reconstructions was excised and used in differently sized simulated rings as references. The depicted reconstruction has been calculated in a total of 120 iterations in the EMAN1.9 and IMAGIC5 software packages and symmetrized with a 18-fold cylindrical point-group symmetry. Its resolution is approximately 25 Å according to the 0.5 criterion.

SFG spectroscopy. SFG spectra of IM30 interacting with lipid monolayers were collected at room temperature after 30 min incubation time of 12.6 μM IM30 in HEPES (pH 7.2, 10 mM) with a binary lipid monolayer (60% DOPG, 40% MGDG) in a 25-ml Teflon-coated metal trough. The lipid monolayer was prepared by spreading the lipid mixture, which was dissolved in chloroform, onto a HEPES-buffered solution to a final surface pressure of 14 mN m^{-1}. The surface pressure was measured with a DeltaPi instrument (Kibron inc.). The protein solution was injected beneath the lipid monolayer.

The SFG setup has been described previously in detail[58]. Briefly, a 5-mJ pulse centred at ~800 nm was generated by a Ti:sapphire fs-laser oscillator (MaiTai, Spectra-Physics) and a regenerative amplifier (SpitFire Pro, Spectra-Physics), which was pumped by a Nd:YLF laser (EMPower, Spectra-Physics). One part of the amplifier output (0.5 mJ) was branched off and spectrally narrowed by an etalon to generate a narrow band pulse (~15 cm^{-1}). The second part was used to generate a tunable infrared pulse with an optical parametric amplifier system (TOPAS, Light Conversion). To generate the SFG signal, the infrared and visible beams were spatially and temporally overlapped on the sample. The SFG signal was then focused onto an emCCD camera (Newton, Andor). The sample stage was purged with nitrogen until the relative humidity was lower than 10%. The SFG spectra were normalized using a reference spectrum obtained from silver in *PPP* polarization (P-polarized SFG, visible and infrared radiation, respectively). The energy scale of the spectra was calibrated by using the ~1,737 cm^{-1} lipid C=O stretch peak[59].

VSFG spectra calculations. The VSFG spectra are calculated using the formalism described in ref. 60. The IM30 monomer structural model for which the spectra are calculated is obtained from Bultema *et al.*[61] The intramolecular hydrogen bonds and contacts with surrounding water were calculated using a short, 1 ns molecular dynamics (MD) simulation in the software package GROMACS. In this simulation, the modelled IM30 monomer used for the VSFG spectra simulation was hydrated and relaxed using the TIP3 and CHARMM27 force fields for the waters and the protein, respectively.

We construct the amide-I excitonic Hamiltonian for the amide backbone modes that generate the experimentally observed VSFG response from the protein conformation. The couplings between the nearest neighbour amide-I modes are obtained using an *ab initio* 6-31G ++(d) B3LYP-calculated map that gives the coupling as a function of the dihedral angle between the amide groups[62]. The other couplings, between non-nearest neighbour amide-I modes, are estimated using the transition dipole coupling model[63]. The hydrogen bond-induced frequency shifts of the local amide-I modes are calculated using the empirical model described in Roeters *et al.*[60] We diagonalize the Hamiltonian to obtain the delocalized amide-I normal modes (eigen vectors) and frequencies (eigen values) of the system. After calculating the infrared and Raman response of the normal modes, we determine the VSFG hyperpolarizability by taking the tensor product of the two responses. To account for homogeneous broadening, we convolute the result with a Lorentzian line shape. The response is transformed from the molecular frame to the lab frame by averaging the Euler transformation between the two frames over the orientational distribution of the molecules[64,65], which gives the nonlinear susceptibility $\chi^{(2)}_{\text{protein}}$.

We calculate the C=O stretch VSFG response of the lipids $\chi^{(2)}_{\text{lipid}}$ with the same formalism, except that no hydrogen-bond shifts are taken into account, as this gives the most accurate results for the calculation of the *SSP* and *SPS* VSFG spectra of a pure 60/40% DOPG/MGDG binary lipid monolayer.

The total VSFG intensity is proportional to the modulus square of the weighted sum of the lipid and protein response:

$$I_{\text{SFG,total}} \propto \left| \text{PLR} \cdot \chi^{(2)}_{\text{protein}} + \chi^{(2)}_{\text{lipid}} \right|^2 \qquad (2)$$

with PLR being the protein oscillator-to-lipid oscillator ratio near the interface. Thus, the calculated VSFG response is dependent on the orientation and magnitude of the transition dipole moment $\boldsymbol{\mu}$ and Raman polarizability tensor $\boldsymbol{\alpha}$ of the local modes of the system, the protein and lipid orientation distribution, the line widths of the Lorentzians of the protein and lipid normal modes, $\Gamma_{\text{protein amide-I}}$ and $\Gamma_{\text{lipid ester-C=O}}$, the overall frequency offsets for the amide-I and ester C=O stretch modes, $\nu_{0,\text{ protein amide-I}}$ and $\nu_{0,\text{ lipid ester-C=O}}$, and on the PLR. To fit the model to the experimental data, we include an overall scale factor for each polarization combination.

We estimated the magnitude of the transition-dipole moment $|\boldsymbol{\mu}|$ of the MGDG lipid to be 0.39 D by measuring the FTIR spectrum of the similar lipid DOPC and used the previously determined transition dipole moment of the amide-I mode (0.37 D)[66]. We determine the orientations of the local-mode $\boldsymbol{\mu}$s based on the transition charges in the amide plane[66,67] and the orientations and magnitudes of the $\boldsymbol{\alpha}$s based on previous work of Tsuboi *et al.*[68]

The orientation distribution of the ester groups in the MGDG/PG lipid monolayer layer was determined from a published MD trajectory[69].

IM30 was treated as a cylindrical protein that most likely has an azimuthally isotropic distribution at the lipid monolayer, so that its VSFG response is only dependent on the Euler angle θ (the angle between the helical axis and the surface normal of the monolayer), as the other two Euler angles ϕ and ψ average out in the orientation-distribution averaged Euler transformation from the molecular frame to the lab frame.

By least-squares fitting the calculated VSFG spectra of IM30 and the lipid monolayer to the experimental SSP and SPS data, we determined the optimal values of $\Gamma_{\text{protein amide-I}}$, $\Gamma_{\text{lipid ester-C=O}}$, $\nu_{0,\text{ protein amide-I}}$ and $\nu_{0,\text{ lipid ester-C=O}}$. Assuming that at most a single full protein monolayer coverage is possible, the PLR has an upper limit as a function of θ (see Supplementary Fig. 5). We estimate this limit from the surface density of lipid monolayers, and assume a cylindrical shape for IM30. If there is less than a full monolayer protein coverage (for example, due to an empty space in the middle of a ring structure comprised of IM30 molecules), the PLR decreases, so PLR values below the line in Supplementary Fig. 5b are also physically plausible.

Only calculated *SSP* spectra for $\theta \lessapprox 3°$ and *SPS* spectra for $\theta \lessapprox 15°$ match the experimental spectra well (see Fig. 5c and Supplementary Fig. 5).

The observation that only small values of θ_{IM30} are consistent with the experimental data is because, for both the *SSP* and the *SPS* polarization combination, the VSFG intensity at 1,650 cm^{-1} (helical secondary structure of IM30) decreases monotonously for increasing values of θ. Therefore, the spectra can only be fitted well for small values of θ, for which the protein signal is relatively strong; only for the small angles, the inferred PLR is smaller than the physically plausible upper limit of 10. We determined the orientation distribution solely based on the spectral line shapes of the individual polarization combinations, and not from the intensity ratios between the polarization combinations (analogous to the approach described in Roeters *et al.*[60]). This approach does not require an accurate description of the reflections at the interface, via the so-called the Fresnel factors[64,70]. Therefore, the uncertainty in Fresnel factors (dependent on the intrinsically uncertain interfacial refractive index) does not influence the derived orientation distribution.

References

1. Vothknecht, U. C. & Westhoff, P. Biogenesis and origin of thylakoid membranes. *Biochim. Biophys. Acta* **1541,** 91–101 (2001).
2. von Wettstein, D. Discovery of a protein required for photosynthetic membrane assembly. *Proc. Natl Acad. Sci. USA* **98,** 3633–3635 (2001).
3. Benning, C. A role for lipid trafficking in chloroplast biogenesis. *Prog. Lipid Res.* **47,** 381–389 (2008).
4. Schneider, D. in *The Cell Biology of Cyanobacteria* (eds Flores, E. & Herrero, A.) (Caister Academic, 2014).
5. Zak, E. *et al.* The initial steps of biogenesis of cyanobacterial photosystems occur in plasma membranes. *Proc. Natl Acad. Sci. USA* **98,** 13443–13448 (2001).
6. Liberton, M., Howard Berg, R., Heuser, J., Roth, R. & Pakrasi, H. B. Ultrastructure of the membrane systems in the unicellular cyanobacterium *Synechocystis* sp. strain PCC 6803. *Protoplasma* **227,** 129–138 (2006).
7. Nickelsen, J. *et al.* Biogenesis of the cyanobacterial thylakoid membrane system--an update. *FEMS Microbiol. Lett.* **315,** 1–5 (2010).
8. Charuvi, D. *et al.* Gain and loss of photosynthetic membranes during plastid differentiation in the shoot apex of Arabidopsis. *Plant Cell* **24,** 1143–1157 (2012).
9. Shimoni, E., Rav-Hon, O., Ohad, I., Brumfeld, V. & Reich, Z. Three-dimensional organization of higher-plant chloroplast thylakoid membranes revealed by electron tomography. *Plant Cell* **17,** 2580–2586 (2005).
10. Chuartzman, S. G. *et al.* Thylakoid membrane remodeling during state transitions in Arabidopsis. *Plant Cell* **20,** 1029–1039 (2008).
11. van de Meene, A. M. *et al.* Gross morphological changes in thylakoid membrane structure are associated with photosystem I deletion in Synechocystis sp. PCC 6803. *Biochim. Biophys. Acta* **1818,** 1427–1434 (2012).
12. Westphal, S., Soll, J. & Vothknecht, U. C. A vesicle transport system inside chloroplasts. *FEBS Lett.* **506,** 257–261 (2001).
13. Morre, D. J., Sellden, G., Sundqvist, C. & Sandelius, A. S. Stromal low temperature compartment derived from the inner membrane of the chloroplast envelope. *Plant Physiol.* **97,** 1558–1564 (1991).
14. Schneider, D., Fuhrmann, E., Scholz, I., Hess, W. R. & Graumann, P. L. Fluorescence staining of live cyanobacterial cells suggest non-stringent chromosome segregation and absence of a connection between cytoplasmic and thylakoid membranes. *BMC Cell Biol.* **8,** 39 (2007).
15. Andersson, M. & Sandelius, A. A chloroplast-localized vesicular transport system: a bio-informatics approach. *BMC Genomics* **5,** 40 (2004).
16. Khan, N. Z., Lindquist, E. & Aronsson, H. New putative chloroplast vesicle transport components and cargo proteins revealed using a bioinformatics approach: An Arabidopsis Model. *PLoS ONE* **8,** e59898 (2013).
17. Südhof, T. C. & Rothman, J. E. Membrane fusion: grappling with SNARE and SM proteins. *Science* **323,** 474–477 (2009).
18. Kielian, M. & Rey, F. A. Virus membrane-fusion proteins: more than one way to make a hairpin. *Nat. Rev. Microbiol.* **4,** 67–76 (2006).
19. Kroll, D. *et al.* VIPP1, a nuclear gene of Arabidopsis thaliana essential for thylakoid membrane formation. *Proc. Natl Acad. Sci. USA* **98,** 4238–4242 (2001).
20. Westphal, S., Heins, L., Soll, J. & Vothknecht, U. C. Vipp1 deletion mutant of Synechocystis: a connection between bacterial phage shock and thylakoid biogenesis? *Proc. Natl Acad. Sci. USA* **98,** 4243–4248 (2001).
21. Fuhrmann, E., Gathmann, S., Rupprecht, E., Golecki, J. & Schneider, D. Thylakoid membrane reduction affects the photosystem stoichiometry in the cyanobacterium Synechocystis sp. PCC 6803. *Plant Physiol.* **149,** 735–744 (2009).
22. Aseeva, E. *et al.* Complex formation of Vipp1 depends on its alpha-helical PspA-like domain. *J. Biol. Chem.* **279,** 35535–35541 (2004).
23. Liu, C. *et al.* J-domain protein CDJ2 and HSP70B are a plastidic chaperone pair that interacts with vesicle-inducing protein in plastids 1. *Mol. Biol. Cell* **16,** 1165–1177 (2005).
24. Li, H. M., Kaneko, Y. & Keegstra, K. Molecular cloning of a chloroplastic protein associated with both the envelope and thylakoid membranes. *Plant Mol. Biol.* **25,** 619–632 (1994).
25. Fuhrmann, E. *et al.* The vesicle-inducing protein 1 from Synechocystis sp. PCC 6803 organizes into diverse higher-ordered ring structures. *Mol. Biol. Cell* **20,** 4620–4628 (2009).
26. Kobayashi, R., Suzuki, T. & Yoshida, M. *Escherichia coli* phage-shock protein A (PspA) binds to membrane phospholipids and repairs proton leakage of the damaged membranes. *Mol. Microbiol.* **66,** 100–109 (2007).
27. Vothknecht, U. C., Otters, S., Hennig, R. & Schneider, D. Vipp1: a very important protein in plastids?! *J. Exp. Bot.* **63,** 1699–1712 (2012).
28. Zhang, L., Kato, Y., Otters, S., Vothknecht, U. C. & Sakamoto, W. Essential role of VIPP1 in chloroplast envelope maintenance in arabidopsis. *Plant Cell* **24,** 3695–3707 (2012).
29. Lo, S. M. & Theg, S. M. Role of vesicle-inducing protein in plastids 1 in cpTat transport at the thylakoid. *Plant J.* **71,** 656–668 (2012).
30. Gao, H. & Xu, X. Depletion of Vipp1 in *Synechocystis* sp. PCC 6803 affects photosynthetic activity before the loss of thylakoid membranes. *FEMS Microbiol. Lett.* **292,** 63–70 (2009).
31. Nordhues, A. *et al.* Evidence for a role of VIPP1 in the structural organization of the photosynthetic apparatus in Chlamydomonas. *Plant Cell* **24,** 637–659 (2012).
32. Bryan, S. J. *et al.* Localisation and interactions of the Vipp1 protein in cyanobacteria. *Mol. Microbiol.* **94,** 1179–1195 (2014).
33. Barber, J. & Gounaris, K. What role does sulpholipid play within the thylakoid membrane? *Photosynth. Res.* **9,** 239–249 (1986).
34. Wada, H. & Murata, N. in *Lipids in Photosynthesis: Structure, Function and Genetics* (eds Siegenthaler, P.-A. & Murata, N.) 65–81 (Springer, 1998).
35. Block, M. A., Dorne, A. J., Joyard, J. & Douce, R. Preparation and characterization of membrane fractions enriched in outer and inner envelope membranes from spinach chloroplasts. II. Biochemical characterization. *J. Biol. Chem.* **258,** 13281–13286 (1983).
36. Webb, M. S., Tilcock, C. P. S. & Green, B. R. Salt-mediated interactions between vesicles of the thylakoid lipid digalactosyldiacylglycerol. *Biochim. Biophys. Acta* **938,** 323–333 (1988).
37. Webb, M. S. & Green, B. R. Effects of neutral and anionic lipids on digalactosyldiacylglycerol vesicle aggregation. *Biochim. Biophys. Acta* **1030,** 231–237 (1990).
38. Montessuit, S. *et al.* Membrane remodeling induced by the dynamin-related protein Drp1 stimulates Bax oligomerization. *Cell* **142,** 889–901 (2010).
39. Youle, R. J. & van der Bliek, A. M. Mitochondrial fission, fusion, and stress. *Science* **337,** 1062–1065 (2012).
40. Bonifacino, J. S. & Glick, B. S. The mechanisms of vesicle budding and fusion. *Cell* **116,** 153–166 (2004).
41. Nevo, R. *et al.* Thylakoid membrane perforations and connectivity enable intracellular traffic in cyanobacteria. *EMBO J.* **26,** 1467–1473 (2007).
42. Pribil, M., Labs, M. & Leister, D. Structure and dynamics of thylakoids in land plants. *J. Exp. Bot.* **65,** 1955–1972 (2014).
43. Majeran, W., Cai, Y., Sun, Q. & van Wijk, K. J. Functional differentiation of bundle sheath and mesophyll maize chloroplasts determined by comparative proteomics. *Plant Cell* **17,** 3111–3140 (2005).
44. Benning, C., Xu, C. & Awai, K. Non-vesicular and vesicular lipid trafficking involving plastids. *Curr. Opin. Plant Biol.* **9,** 241–247 (2006).
45. DeLisa, M. P., Lee, P., Palmer, T. & Georgiou, G. Phage shock protein PspA of Escherichia coli relieves saturation of protein export via the Tat pathway. *J. Bacteriol.* **186,** 366–373 (2004).
46. Gohre, V., Ossenbuhl, F., Crevecoeur, M., Eichacker, L. A. & Rochaix, J. D. One of two alb3 proteins is essential for the assembly of the photosystems and for cell survival in Chlamydomonas. *Plant Cell* **18,** 1454–1466 (2006).
47. Zhang, L. & Sakamoto, W. Possible function of VIPP1 in thylakoids: Protection but not formation? *Plant Signal Behav.* **8,** e22860 (2013).
48. Aseeva, E. *et al.* Vipp1 is required for basic thylakoid membrane formation but not for the assembly of thylakoid protein complexes. *Plant Physiol. Biochem.* **45,** 119–128 (2007).
49. Gao, H., Sage, T. L. & Osteryoung, K. W. FZL, an FZO-like protein in plants, is a determinant of thylakoid and chloroplast morphology. *Proc. Natl Acad. Sci. USA* **103,** 6759–6764 (2006).
50. Low, H. H. & Lowe, J. *A bacterial dynamin-like protein* **444,** 766–769 (2006).
51. Buchanan, B. B. Role of light in the regulation of chloroplast enzymes. *Ann. Rev. Plant Physiol.* **31,** 341–374 (1980).
52. Parasassi, T., De Stasio, G., d'Ubaldo, A. & Gratton, E. Phase fluctuation in phospholipid membranes revealed by Laurdan fluorescence. *Biophys. J.* **57,** 1179–1186 (1990).
53. Ellens, H., Bentz, J. & Szoka, F. C. pH-Induced destabilization of phosphatidylethanolamine-containing liposomes: role of bilayer contact. *Biochemistry* **23,** 1532–1538 (1984).

54. Zschörnig, O., Paasche, G., Thieme, C., Korb, N. & Arnold, K. Modulation of lysozyme charge influences interaction with phospholipid vesicles. *Colloids Surf. B Biointerfaces* **42,** 69–78 (2005).
55. Meers, P., Ali, S., Erukulla, R. & Janoff, A. S. Novel inner monolayer fusion assays reveal differential monolayer mixing associated with cation-dependent membrane fusion. *Biochim. Biophys. Acta* **1467,** 227–243 (2000).
56. Saur, M. *et al.* Acetylcholine-binding protein in the hemolymph of the planorbid snail biomphalaria glabrata is a pentagonal dodecahedron (60 Subunits). *PLoS ONE* **7,** e43685 EP - (2012).
57. Ludtke, S. J., Baldwin, P. R. & Chiu, W. EMAN: semiautomated software for high-resolution single-particle reconstructions. *J. Struc. Biol.* **128,** 82–97 (1999).
58. Smits, M. *et al.* Polarization-resolved broad-bandwidth sum-frequency generation spectroscopy of monolayer relaxation. *J. Phys. Chem. C* **111,** 8878–8883 (2007).
59. Bicknell-Brown, E. & Brown, K. G. Raman studies of lipid interactions at the bilayer interface: Phosphatidyl choline-cholesterol. *Biochem. Biopphys. Res. Comm.* **94,** 638–645 (1980).
60. Roeters, S. J. *et al.* Determining in situ protein conformation and orientation from the amide-i sum-frequency generation spectrum: theory and experiment. *J. Phys. Chem. A* **117,** 6311–6322 (2013).
61. Bultema, J. B., Fuhrmann, E., Boekema, E. J. & Schneider, D. Vipp1 and PspA: Related but not twins. *Commun. Integr. Biol.* **3,** 162–165 (2010).
62. 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.* **122,** 224904 (2005).
63. Hamm, P., Lim, M. & Hochstrasser, R. M. Structure of the amide I band of peptides measured by femtosecond nonlinear-infrared spectroscopy. *J. Phys. Chem. B* **102,** 6123–6138 (1998).
64. Zhuang, X., Miranda, P. B., Kim, D. & Shen, Y. R. Mapping molecular orientation and conformation at interfaces by surface nonlinear optics. *Phys. Rev. B* **59,** 12632–12640 (1999).
65. 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* **108,** 3548–3562 (2004).
66. 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. USA* **96,** 2036–2041 (1999).
67. Hamm, P. & Zanni, M. *Concepts and Methods of 2D Infrared Spectroscopy* (Cambridge Univ., 2011).
68. Tsuboi, M., Benevides, J. M. & Thomas, G. J. Raman Tensors and their application in structural studies of biological systems. *Proc. Jpn Acad. B Phys. Biol. Sci.* **85,** 83–97 (2009).
69. Kapla, J., Stevensson, B., Dahlberg, M. & Maliniak, A. Molecular dynamics simulations of membranes composed of glycolipids and phospholipids. *J. Phys. Chem. B* **116,** 244–252 (2011).
70. 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.* **24,** 191–256 (2005).

Acknowledgements

T.W. and S.J.R. thank Jon Kapla and Arnold Maliniak for providing simulation data of MGDG/PG monolayers and Ariana Torres-Knoop for performing MD hydration studies of the IM30 monomer. We thank K. Rusitzka for supplementary EM work, A. Roth for excellent technical assistance and Dr. E. Krämer-Albers for the possibility to perform the NTA experiments. T.W. thanks the Marie Curie Program of the European Union for support of this work (CIG grant #322124). This work was supported by the Max Planck Graduate Center at the Max Planck Institutes and the University of Mainz.

Author contributions

D.S., T.W., N.H. and J.M. designed the experiments and analysed the data. R.H., J.H., M.S., L.S. and N.H. performed the experiments and analysed the data. S.J.R. and S.W. performed the informatics analysis of the SFG-spectra. All authors wrote the paper.

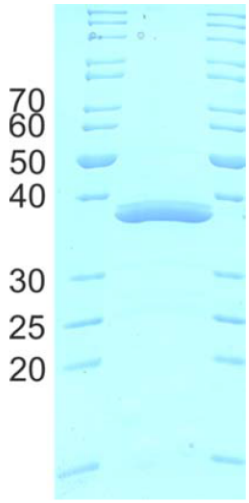
Additional information

Supplementary Information accompanies this paper at http://www.nature.com/naturecommunications

Competing financial interests: The authors declare no competing financial interests.

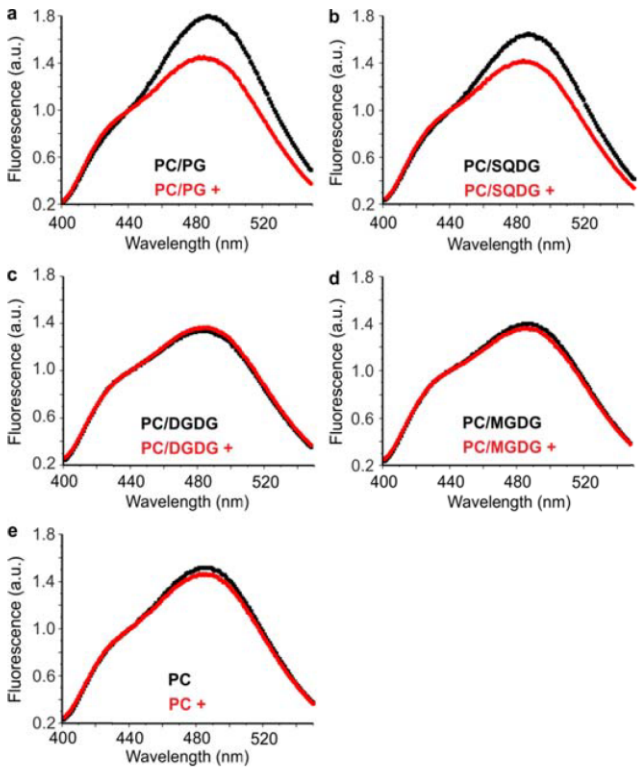
Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/

How to cite this article: Hennig, R. *et al.* IM30 triggers membrane fusion in cyanobacteria and chloroplasts. *Nat. Commun.* 6:7018 doi: 10.1038/ncomms8018 (2015).



Supplementary Fig. 1: Purification of heterologously expressed IM30.

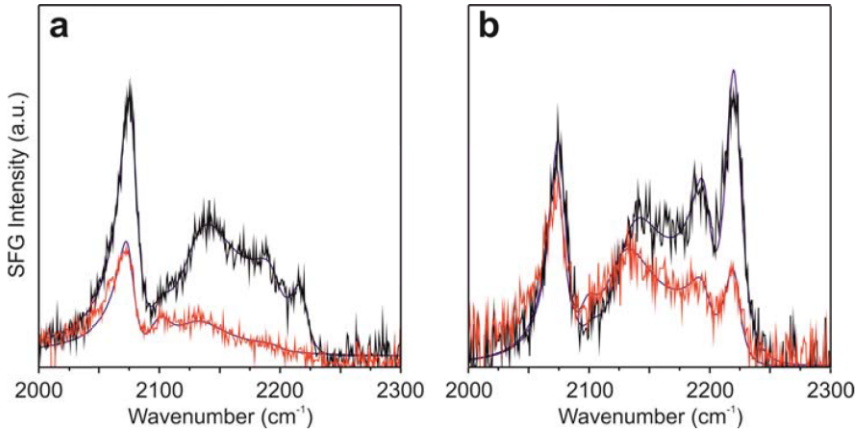
IM30 was expressed and purified as described in Methods. 2 µg of the isolated protein was analyzed by SDS gel electrophoresis on a 12% SDS polyacrylamide gel. The IM30 sample is inserted between a protein standard, and the molecular masses of the protein standards are indicated on the left. Please note that the IM30 protein frequently runs as a double band on SDS-PAGE gels.



Supplementary Fig. 2: IM30 binds to model membranes in presence of negatively charged TM lipids.

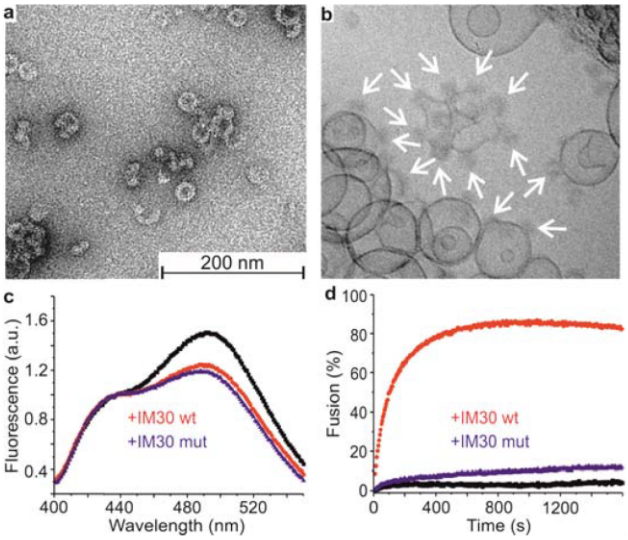
Peripheral binding of IM30 to model membranes was followed by monitoring changes in the Laurdan fluorescence emission after excitation at 350 nm. Fluorescence emission spectra are normalized to the emission at 430 nm and are shown in absence (black) and presence of 2.5 µM IM30 (red). Decreased fluorescence emission at 490 nm indicates an increased lipid acyl chain order, caused by peripheral binding of IM30. IM30 binding was monitored using PC as a neutral background, to which IM30 does not bind (**e**), in presence of PG (**a**), SQDG (**b**), DGDG (**c**) or MGDG (**d**). Only in the case of PG or SQDG-containing membranes, significant changes in Laurdan fluorescence emission were observed. The experiment was performed four times.

2



Supplementary Fig. 3: IM30 interaction with a lipid monolayer decreases the lipid acyl chain ordering and affects the lipid orientation. The blue lines represent fits of the data.

SFG spectra of deuterated DMPG-D54 (Dimyristoyl-d54-phosphoglycerol) lipid monolayers in presence of 7.5 mM Mg^{2+}were recorded in the C–D stretching range, before (black lines) and after (red lines) addition of IM30. Deuterated lipids were chosen to avoid possible contributions to the signal from IM30. The monolayer was prepared at a pressure of 22 mN/m. IM30 was injected to a total concentration of 5 μM in the sub-phase, after which the system was allowed to equilibrate for 60 minutes. (**a**) The ssp spectra show a significant reduction of overall signal, which implies the acyl chains are less ordered and aligned after IM30 adsorption. (**b**) The ppp spectra indicate reorientation of the lipid chains: the ratio of asymmetric methyl stretching modes (near 2220 cm^{-1}) and symmetric methyl modes (near 2070 cm^{-1}) is reduced from 1.3 before to 0.8 after IM30 injection. This indicates a more tilted orientation of the methyl groups in the IM30 covered DMPG-D54 film. Together, ssp and ppp spectra reveal a significant interaction of IM30 with the lipids, causing increased disorder in, and reorientation of the lipid acyl chains.



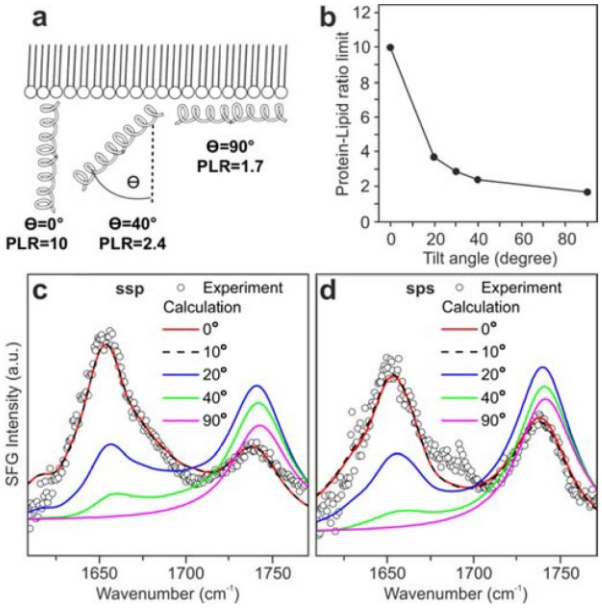
Supplementary Fig. 4: An IM30 mutant that binds to membranes but does not induce membrane fusion.

Three amino acids of the IM30 wild-type sequence have been mutated by side-directed mutagenesis (A76S, A79S, L80A). The mutated protein was expressed and purified as described for the wild-type protein.

(**a**) The IM30 mutant forms large oligomeric ring structures as the wild-type protein, as visible in the EM micrographs. (**b**) Cryo-TEM image of MGDG/PG liposomes in presence of Mg^{2+} and the mutated IM30 taken after 30 minutes of incubation. Arrows indicate extra electron density at the liposome surface, matching in size and shape oligomeric IM30-rings. The magnification was identical in (**a**) and (**b**). (**c**) Peripheral binding of the mutated IM30 (2.5 µM) to PG-containing MGDG membrane surfaces was followed by monitoring changes in the Laurdan fluorescence emission after addition of IM30 (compare Fig. 1g). Addition of the mutated protein (blue) influenced the fluorescence signal, as observed after addition of the wild-type IM30 protein (red),

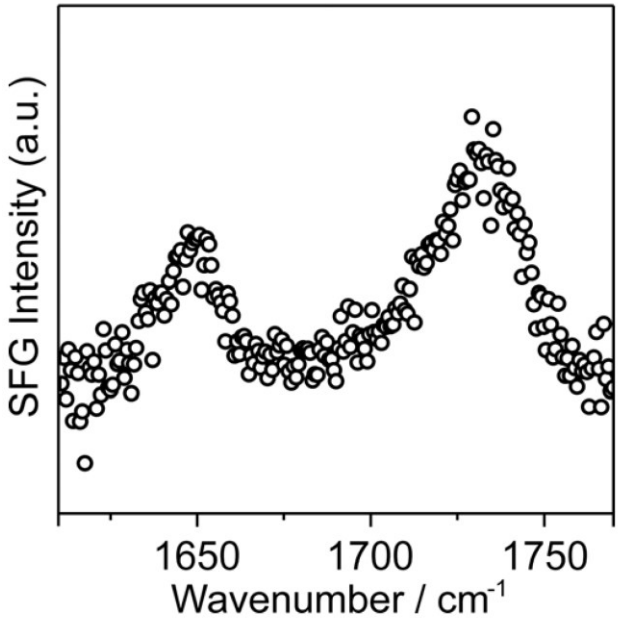
suggesting that both IM30-variants have similar membrane binding properties. For comparison, the Laurdan fluorescence signal of membranes in absence of IM30 is shown (black). The experiment was repeated three times. (**d**) MGDG/PG liposomes, containing a donor and an acceptor dye, were mixed with non-labelled liposomes (1:9), and the increase in donor fluorescence emission was monitored over 1600 s after addition of 2.5 µM IM30 in presence of 7.5 mM Mg^{2+} (compare Fig. 3a, b). 0% and 100% fusion rates were determined in control experiments to calculate the fusion efficiency for each sample (see Methods). While the wild-type protein mediated membrane fusion, the mutated IM30 protein was fusion incompetent. The experiment was repeated three times.

5



Supplementary Fig. 5: IM30 orientation at lipid monolayer surfaces.

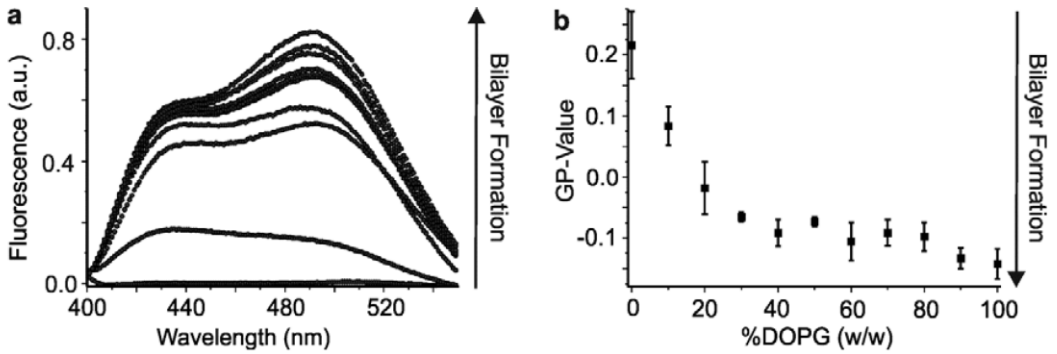
Comparison of experimental and calculated SFG-spectra of IM30 self-assembled onto MGDG/PG (60/40) mixed monolayers for ssp (s polarized SFG, s polarized VIS, p polarized IR) and sps polarization combinations. (**a**) Schematic drawing of the dependency of protein-to-lipid oscillator ratio (PLR) model we used in the calculations. (**b**) PLR upper limits for the tilt angles used in the spectra calculations. (**c**) Comparison of spectra calculated for different orientations of model IM30 units on the monolayers with experimental spectra in ssp-polarization. Calculations, using an average tilt angle of the long helix axis versus the surface normal of 0°-3°, agree well with the data and are in line with an upright orientation of the IM30 monomers on the membrane surface and a ring-type binding geometry. (**d**) Calculations of SFG-spectra collected with sps-polarization combination. Calculations performed with a tilt angle of 0°-15° agree well with the data. The experiments were repeated twice.



Supplementary Fig. 6: Less IM30 is bound at the membrane in absence of Mg^{2+}.

When IM30 is allowed to bind to a lipid monolayer in the absence of Mg^{2+}, a significantly lower IM30 signal is observed in the SFG-spectra (compare Fig. 5c). Thus, Mg^{2+} ions promote the interaction of IM30 with a MGDG/PG monolayer. The experiment was repeated twice.

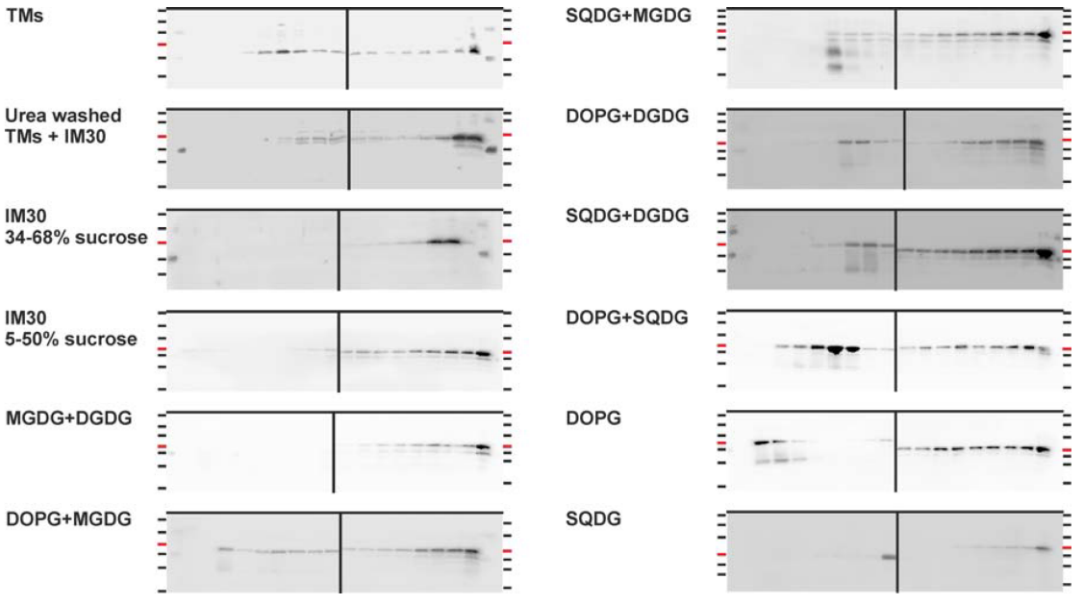
7



Supplementary Fig. 7: Formation of stable MGDG-liposomes by addition of DOPG.

The fluorescent probe Laurdan was added to MGDG-lipids, containing increasing DOPG concentrations, to monitor formation of lamellar lipid bilayer structures. (**a**) The increase in the Laurdan fluorescence emission with increasing DOPG concentrations indicates formation of lipid bilayer structures. In (**b**), the GP-values are summarized for each % DOPG. Error bars: SD (N=3)

Noteworthy, the DOPG content at which lamellar bilayer structures form (here ~40% DOPG) strongly depends on the buffer conditions, and in the glycerol-containing buffer (SD gradients), lamellar structures formed only at ~50% DOPG.



Supplementary Fig. 8: Uncropped membranes of anti-IM30 immunoblots.

Proteins fractionated by SD gradient centrifugation were separated on 14% SDS-PAGE gels with subsequent immunoblot analysis, using an anti-IM30 antiserum and a peroxidase-coupled anti-rabbit secondary antibody. Due to the high number of samples, the fractions of one SD gradient were loaded on two separated SDS-gels. Equal volumes of the respective fractions were loaded, and after SDS-PAGE, the two corresponding gels were blotted in parallel onto a single membrane for subsequent analysis. Full scans of Western blots are shown. The vertical line marks the boundary of the two individual SDS-gels used for separating the individual fractions of a single SD gradient. The position of the molecular weight markers are indicated on the left or right side of each blot, respectively. The molecular weights of the individual molecular weight marker bands are (from the top): 95, 72, 55, 43 (red), 34, 26 and 17 kDa.

4.2 Refining the X-ray crystal structure of aquaporin with sum frequency generation spectroscopy

The experimental and calculated SFG spectra were part of this thesis. This manuscript was submitted to Proceedings of the National Academy of Sciences in October 2016. The proteoliposomes used for generating the supported lipid bilayer were produced by Margareta Trefz and Dirk Schneider at the Institut für Pharmazie und Biochemie (at the Johannes Gutenberg University in Mainz) the Mathematica script and the theory for calculating the SFG spectra were provided by Steven Roeters and Sanders Woutersen from the Van 't Hoff Institute for Molecular Science at the University of Amsterdam and the MD simulation of GlpF in a lipid bilayer membrane was done by Wesley Beckners and Jim Pfaendtner at the Department for Chemical Engineering at the University of Washington.

Refining the X-ray crystal structure of aquaporin with simulations and surface vibrational spectroscopy

Lars Schmüser[a], Margareta Trefz[b], Steven Roeters[c], Wesley Beckner[d], Jim Pfaendtner[d], Sander Woutersen[c], Mischa Bonn[a], Dirk Schneider[b] and Tobias Weidner[1, e]*

Affiliations

[a]Max Planck Institute for Polymer Research, Ackermannweg 10, 55128 Mainz, Germany

[b]Institut für Pharmazie und Biochemie, Universität Mainz, Johann-Joachim-Becher-Weg 30, 55128 Mainz, Germany

[c]Van't Hoff Institute for Molecular Sciences, University of Amsterdam, Science Park 904, 1098 XH, Amsterdam, The Netherlands

[d]Department of Chemical Engineering, University of Washington, Seattle, WA, USA

[e]Department of Chemistry, Aarhus University, Langelandsgade 140, building 1511, 220, 8000 Aarhus C, Denmark

Abstract

High-resolution structural information on membrane proteins is essential for understanding cell biology and for structure-based design of new medical drugs and drug delivery strategies. The foremost technique for determining membrane protein structure has been X-ray diffraction (XRD), which routinely provides Ångstrom-level information about the atomistic structure and conformation of membrane proteins. Yet, for XRD experiments, proteins are removed from their native membrane environment, chemically stabilized and crystalized, all of which can cause conformational changes compared with the native membrane structure. Here, we describe a combination of molecular dynamics (MD) simulations and surface-sensitive vibrational spectroscopy to refine XRD structures and account for the membrane environment. We refine the XRD structure of glycerol facilitator channel (GlpF), an aquaporin membrane channel finely tuned to selectively transport water and glycerol molecules across the membrane barrier. We find that the membrane structure of GlpF differs significantly from the XRD structure PDB 1lda. The combination of spectroscopic and computational techniques described here provides a framework to improve XRD structures of proteins in non-native environments.

Significance Statement

Protein structure dictates protein function. Protein structures solved by X-ray diffraction have been critical for progress in molecular biology and pharmacology. One drawback of X-ray structure analysis of large and fragile membrane proteins is they have to be removed from their native lipid environments, chemically stabilized and crystalized. This treatment can cause significant changes in the structure of a protein and, hence, its function. We report an approach which allows the refinement of X-ray crystal structures of membrane proteins within a native lipid environment.

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 insights into the mechanical properties of cell membranes, the transport of molecules across the membrane barrier, cell sensing and communication. Approximately 30% of the human genome encodes for proteins associated with membranes[1].

By far the most successful methods for experimentally solving membrane protein structures with atomic resolution are nuclear magnetic resonance (NMR)[2] and X-ray diffraction (XRD)[3,4]. NMR can provide high-resolution data for proteins when incorporated in lipid vesicles and lipid stacks, but is limited to smaller (typically about 20 kDa[5]) proteins and peptides. X-ray crystallography can solve the structure of very large proteins with Ångstrom 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 non-physiological environments within crystals by relaxing the protein structure within a more 'native' biological environment.

While MD simulations have provided a wealth of information on membrane proteins with molecular detail, a caveat of the simulations is that the obtained, refined 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 inherently surface sensitive and non-invasive approach, allowing the study of membrane proteins in a hydrated membrane environment.[6,7] SFG makes use of the fact that infrared/Raman vibrations of molecular groups in the proteins depend strongly on both 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 those two incident beams by nonlinear optical frequency mixing. Any vibrational modes in resonance with the IR beam will enhance the signal and lead to distinct spectral features. In analogy with IR and Raman methods, SFG spectra in the amide I region (1500-1800 cm^{-1}) allow the analysis of the secondary and tertiary structure and orientation of proteins. Contrary to IR and Raman, because of the non-linear nature of

SFG, the spectral contributions of individual folding motifs and orientations within SFG spectra will interfere and therefore, result in complex spectra. The challenge to 'disentangle' the spectra and recover the rich structural information contained in the spectra has been the white whale 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 large interfacial proteins.[8,9] Calculated SFG spectra of proteins and peptides have shown good agreement with experimental SFG spectra [9-11]. The method has been used to determine the orientation of proteins[9,10] and to differentiate between different folding states within membrane proteins[8]. However, these studies use crystal structures obtained from proteins in non-native environments as the basis for the

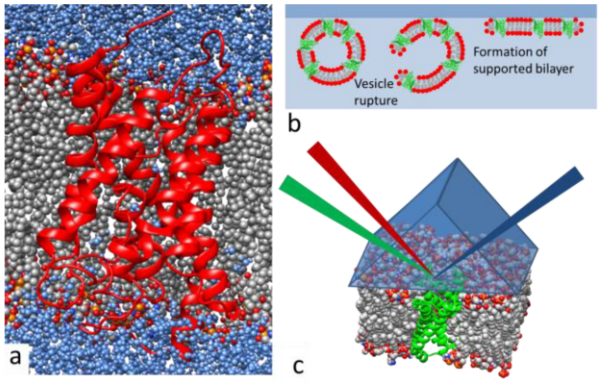


Figure 1 a) The Glycerol facilitator GlpF (1lda, red) embedded in a lipid bilayer (grey) with surrounding water molecules (blue), b) model of the formation of a supported lipid bilayer using liposome spreading, c) experimental setup for the SFG experiments. The lipid bilayer was formed at the solid-water-interface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

spectra calculations, and, thereby, make no use of the potential for in situ studies with SFG.

Here, we combine knowledge of protein structures obtained from crystal XRD with MD simulations of that protein in a membrane environment and SFG measurements of the protein in a model membrane. In brief, we take an XRD structure as the starting point, then relax the protein structure within a membrane environment using MD simulations, calculate SFG spectra from the relaxed structures and then compare with experimental spectra. In other words, we refine the XRD structure by experimentally verified MD simulations.

We chose an aquaporin as a 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 osmoregulation of a variety of tissues in all living organisms[12]. Besides facilitating a 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, *e.g.* 13 aquaporins are expressed in humans where they regulate the water homeostasis according to the individual requirements of the organs and cells[13,14]. Based on their important roles in cell biology, they have great potential in diagnostics and therapeutics[15-18].

All aquaporins form oligomers composed of four identical chains, each with its own channel at the center of the subunit. The channel is filled with a single-file queue of water (or glycerol) molecules passing through the protein. The aquaporin shown in Figure 1a (PDB 1lda), is the well studied glycerol facilitator channel (GlpF) of *E. coli*. A recent very high resolution structure (PDB 3zoj)[19], revealed how sensitive aquaporin function depends on structural fidelity: the exchange of only two amino acids at the entry to the water channel play the deciding role to ensure that water, and not hydronium or hydroxyl ions, can pass the channel. [19] In this study, we refine the XRD structure of the GlpF aquaglyceroporin shown in Fig1a.[20]

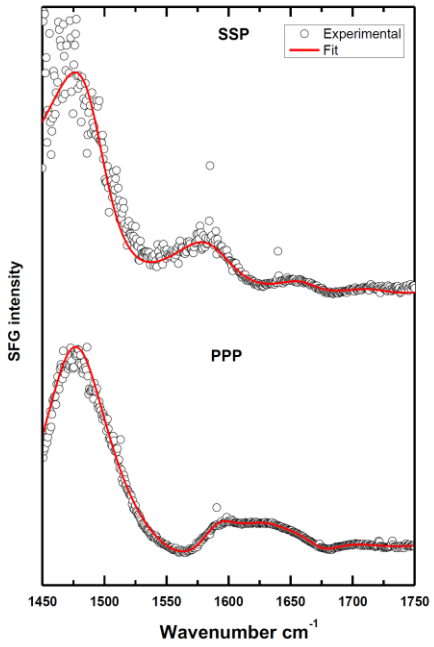


Figure 2: Experimental SFG spectra of a GlpF including lipid bilayer which was formed by proteoliposomes on the solid water interface of a CaF_2-prism mounted on a flow cell. The two different laser polarization combinations SSP and PPP have been measured. The spectra have been fitted (red line) using a Lorentzian line width model (Eq1 and Eq2).

Results and discussion

To collect SFG spectra of GlpF under biologically relevant conditions, we spread GlpF-loaded proteoliposomes, prepared with E. coli polar lipid extract, 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 backside of the prism with SFG in near total internal reflection geometry (Figure 1c). The spectra were corrected for the IR intensity profile using reference silver spectra (also see Methods section). Representative experimental spectra in ssp (s-polarized SFG, s-polarized visible and p-polarized infrared) and ppp polarization are shown in Fig. 2. The spectra could be described very well using a Lorentzian line shape model[21,22] wherein each vibrational resonance is represented by a single Lorentzian:

$$I_{SFG}(\omega_{IR}) \propto \left|\left(\chi^{(2)}\omega_{IR}\right)\right|^2 I_{vis} I_{IR} \quad (1)$$

$$\chi^{(2)}\omega_{IR} = A_0 e^{i\varphi} + \sum_n [A_n / (\omega_{IR} - \omega_n + i\Gamma_n)] \quad (2)$$

Here, $I_{SFG}(\omega_{IR})$, $I_{vis}I_{IR}$, $\chi^{(2)}(\omega_{IR})$, $A_0 e^{i\varphi}$, A_n and $i\Gamma_n$ are the SFG intensity, the intensities of IR and visible laser beams, the second order susceptibility, the phase and amplitude of the non-resonant contribution, the amplitude of resonant contribution and the line width of the resonant contribution respectively.

The spectra are dominated by a prominent peak near 1480 cm^{-1}, which can be attributed to CH_2 and CH_3 scissoring modes[23]. A feature near 1650 cm^{-1} is representative of the α-helical structure of the membrane protein[23]. In addition, there is a smaller peak near 1720 cm^{-1}, which likely originates from carbonyl groups of membrane lipids. The feature observed near 1600 cm^{-} is not well described in literature and might be a contribution that is not discerned in linear spectra because of overlap with neighboring modes and which becomes visible here due to pronounced cancelation of the main vibrational modes in the far field.

The observed spectra are a typical example of the entangled, complex signal expected from large, structurally rich proteins, where direct spectral analysis and peak fitting reach their limits. We have first calculated SFG spectra from GlpF structure files obtained with XRD from protein crystals (PDB 1lda).[24] The calculation followed the procedure described by Roeters et al.[9] The model takes into account all CO, CN, NH and CC bond vibrations along the peptide backbone (see Methods section for more details). The theoretical spectra for the protein crystals are shown in Figure 3 (blue dashed line). To compare with the experimental protein backbone data, we plott the backbone related components of the fit to the experimental data in Figure 3 (solid line) and exclude any non-protein CO, CN, NH and CC modes from the experimental data, for example lipid CO and CH scissoring modes, which could obscure the protein peaks. While the theoretical spectra correctly predict intensity near 1650 cm^{-1} for both ssp and ppp polarization combinations, yet the main feature near 1600 cm^{-1} is not captured correctly by the calculations. For both polarization combinations we observe a 25 cm^{-1} blue-shift for this main spectra feature. This indicates that the structure of GlpF within the stabilized protein crystal used for XRD[24] differs substantially from its structure when incorporated in a lipid bilayer.

To attempt to resolve the discrepancy between the observed SFG response from the protein in a lipid bilayer environment and the response calculated from the measured crystal structure, we investigated the degree to which the protein may change conformation, when going from the crystalline state to the situation in a lipid membrane. Using the XRD structure as the starting point, we ran MD simulations of GlpF within a POPE bilayer until the structure was equilibrated (after ~100 ns, see inset in Figure 3). Indeed, substantial changes in the structure were apparent. Spectra calculated from this 'relaxed' structure yielded theoretical spectra with a substantially improved match with the experimental data (Figure 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 with the spectrum calculated using the crystal structure. The root mean squared deviation (RMSD) of the computed spectra from the experimental data decreased by 55% for the relaxed structures compared with the crystal structure (see SI). The RMSD values between relaxed structures taken at different time points of the simulation varied only by approximately 5-10% (see Figure S4). A mode analysis shows that the modes contributing to the signal are evenly distributed across the protein (see Figure S2) and any changes to the calculated spectra are a result of changes within the

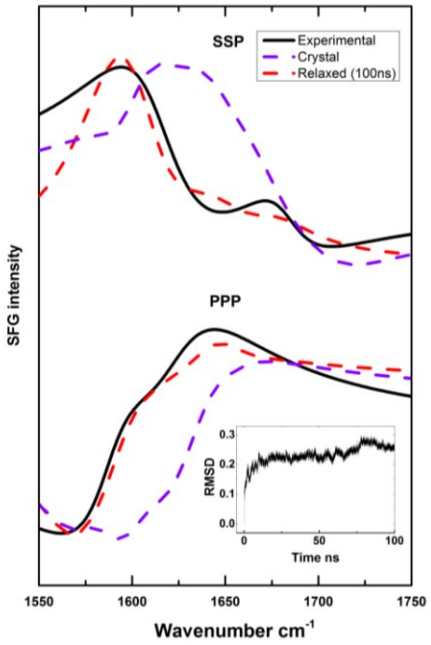


Figure 3: comparison of protein specific SFG spectra extracted from fits of experimental data (solid lines) and calculated SFG spectra of different MD simulation snapshots (dashed lines). The inlet represents the root-mean-square deviation of atomic positions plotted against the simulation time showing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

global protein structure, not only local restructuring of isolated protein sites. Note that, apart from an overall amplitude scaling factor, there are no adjustable parameters when comparing the experimental and theoretical results.

The remaining minor differences between experimental and MD-derived theoretical spectra might be explained by the absence of inter-molecular coupling between neighboring proteins within the theoretical model and the fact that the spectra are calculated from a single snapshot. Future studies will be aimed at including more of these computationally expensive details. Combining all these parameters into the calculation will require significantly higher computational power both for the MD simulation and also for the calculation of SFG spectra but it should further improve the quality of the calculated SFG spectra. However, even the straightforward model used can produce robust, reliable spectra, which match the experimental data very well.

Figure 4 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: both in the loop regions, and particularly in the helix orientations there are subtle but clearly visible variations. It is well known that aquaporin action depends on such minute structural details.[25] Mutations of single amino acids within the transmembrane region can drastically affect the structure and – consequently - the function of aquaporins [26-30]. Also the loop regions are involved in the aquaporin activity and especially in aquaglyceroporins a specific structure of the loop region appears to be crucial for proper protein activity [31]. Knowledge of structural details is therefore crucial for interpreting the changes in the substrate conducting mechanism induced e.g. by mutations in aqua(glycero)porins and in membrane proteins in general. The SFG response is very sensitive to the fairly subtle changes in the protein structure shown in Fig. 4, which can be attributed to changes in the orientation of the helices of the protein.

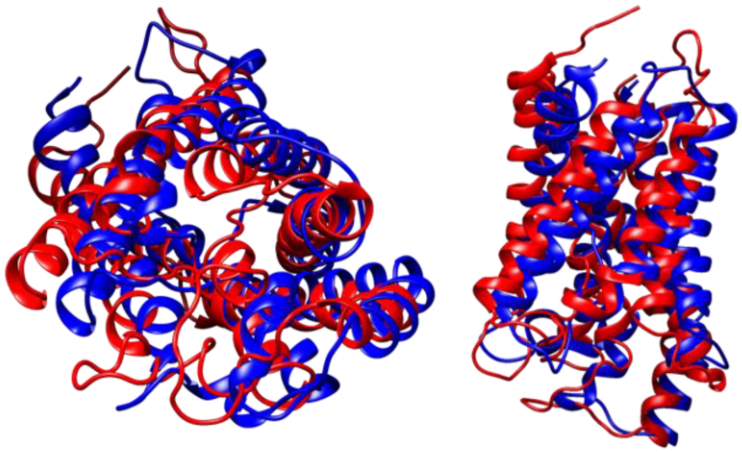


Figure 4: Structural differences between the the crystal structure (1lda, blue) and the refined protein structure (red) The refined structure was obtained from a MD simulation after 100 ns and directly verified by SFG.

The method described here uses the resolving power of MD simulations to study subtle structural details of protein channel structure and the interface specificity and structure sensitivity of SFG to directly experimentally test the simulation results. Such hybrid experimental and theoretical structures allow refining existing XRD protein crystal data to include the influence of the biologically more relevant environment of a hydrated lipid bilayer. In addition to refining the protein structure in contact with a lipid layer, this method can also be used to determine the impact of small molecules or 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.

Methods

Expression and purification of GlpF into proteoliposomes

GlpF expression and purification were performed as described in detail in[26]. The concentration of purified GlpF was determined by absorption measurements at 280 nm, using a calculated extinction coefficient of 37930 M^{-1} cm^{-1} (ExPasy ProtParam tool).

Liposomes prepared from *E. coli* polar lipid extract (EPL: 67% PE, 23.2% PG and 9.8% CL) and 1,2-dioleoyl-sn-glycero-3-phosphocholine (diC18:1-PC, DOPC) were used for GlpF reconstitution[30]. Chloroform lipid solutions were purchased from Avanti Polar Lipids (Alabaster, AL). To remove the organic solvent, 5 mM of the 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 lipid concentration of 12 µM and a molar lipid/GlpF ratio of 400:1. The final sample was adjusted to a volume of 0.5 mL and an OG concentration of 30 mM by 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 MOPS buffer. The dialysis buffer was exchanged three times.

SFG experiment

The SFG setup has been described in detail before.[32] 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 40 fs broadband IR pulse. The remaining 800 nm beam was spectrally narrowed to a 25 µJ 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. The 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 in a nitrogen-flushed chamber to avoid absorption of the IR pulse due to water vapor. The experiments were done in a flow cell with a volume of 1 ml. The flow cell was sealed on one side with an equilateral CaF_2 Prism. 400 µl of proteoliposome solution was injected into the flow cell and incubated for at least 2 h. Remaining proteoliposomes in bulk were rinsed with D_2O followed by an overnight waiting step to allow hydrogen to deuterium exchange. SFG spectra were collected in 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.

Theory of calculation

First we obtain the atom coordinates of the amide groups from the PDB file of the crystal structure or of the MD simulation. We use the coordinates both to calculate the transition dipole moments of the local modes (by determining the transition charge of each atom[33], as this gives more accurate spectra than the conventional approach as well as the Raman polarizabilities similar to refs.[34,35], and to construct the one-exciton Hamiltonian:

$$H = \begin{pmatrix} \hbar\omega^0{}_1 & \kappa_{12} & \kappa_{13} & \kappa_{14} & \mathsf{L} \\ \kappa_{12} & \hbar\omega^0{}_2 & \kappa_{23} & \kappa_{24} & \\ \kappa_{13} & \kappa_{23} & \hbar\omega^0{}_3 & \kappa_{34} & \\ \kappa_{14} & \kappa_{24} & \kappa_{34} & \hbar\omega^0{}_4 & \\ \mathsf{M} & & & & \mathsf{O} \end{pmatrix} \tag{3}$$

with $\omega^0{}_i$ the gas phase frequency of local mode i and κ_{ij} the coupling between local mode i and j.

The diagonal terms are determined with 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[9], comparable to the model used in ref.[36].

For the off-diagonal terms, the couplings between the normal modes, we discriminate between nearest-neighbor and non-nearest-neighbor coupling. As the former is dominated by through bond charge flows we use a parameterization 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[37].

The non- nearest-neighbor couplings are dominated by through-space (Coulomb) interactions, so we estimate these with the Transition Dipole Coupling method[38]:

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

with $\vec{\mu}_i$ the transition dipole moment of local mode i, $\vec{r}_{ij}$ the distance between local mode i and j, and ε_0 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. [9] in which also other details regarding the formalism used here for the spectral calculations can be found.

The non-resonant phase and its amplitude were adapted to yield the best match between the experimental data and the calculated ones. Otherwise, all calculations were done using the same settings.

MD simulation

Simulations of the porin tetramer (1lda[24]) and phosphatidylethanolamine (POPE) lipid bilayer were performed using the GROMACS[39] molecular dynamics (MD) engine and explicitly solvated using the TIP3P[40] water model. Force fields for the lipid bilayer were taken from Tieleman and Berendsen[41] and adapted into the GROMOS96 53A6 force field[42], which is extended to include Berger lipid parameters and has been verified to perform as well or better than previous versions for protein simulations[43]. 2.0 fs time steps in the MD simulation were integrated using a leap-frog algorithm[44]. 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 particle mesh Ewald (PME) summation. Bond lengths between Hydrogen and heavy atoms were fixed with the LINCS linear constraint solver[45].

The tetramer and lipid bilayer were centered in an overall charge-neutral system with periodic x, y, 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[46] 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[47] 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[48] with a 0.4 ps coupling constant to more realistically capture temperature fluctuations[49]. 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 time point 100 ns.

Data availability

The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

Acknowledgements

This work was supported by the Max Planck Graduate Center with the Johannes Gutenberg-Universität Mainz (MPGC). LS and TW thank the European Commission (CIG grant #322124) and the Deutsche Forschungsgemeinschaft (WE4478/4-1) for financial support. Molecular graphics and analyses were performed with the UCSF Chimera package. Chimera is developed by the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco (supported by NIGMS P41-GM103311). Partial financial support to JP and WB provided by NSF (CBET-1264459).

References

1 Liu, W., Wacker, D., Wang, C., Abola, E. & Cherezov, V. Femtosecond crystallography of membrane proteins in the lipidic cubic phase. *Philos T R Soc B* **369**, (2014).

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

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

4 Moraes, I., Evans, G., Sanchez-Weatherby, J., Newstead, S. & Stewart, P. D. S. Membrane protein structure determination The next generation. *Bba-Biomembranes* **1838**, 78-87, (2014).

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* **23**, 734-739, (2013).

6 Ding, B. *et al.* Physiologically-Relevant Modes of Membrane Interactions by the Human Antimicrobial Peptide, LL-37, Revealed by SFG Experiments. *Sci Rep-Uk* **3**, (2013).

7 Nguyen, K. T., Le Clair, S. V., Ye, S. & Chen, Z. Molecular Interactions between Magainin 2 and Model Membranes in Situ. *The Journal of Physical Chemistry B* **113**, 12358-12363, (2009).

8 Liang, C. W., Louhivuori, M., Marrink, S. J., Jansen, T. L. C. & Knoester, J. Vibrational Spectra of a Mechanosensitive Channel. *J Phys Chem Lett* **4**, 448-452, (2013).

9 Roeters, S. J. *et al.* Determining In Situ Protein Conformation and Orientation from the Amide-I Sum-Frequency Generation Spectrum: Theory and Experiment. *J Phys Chem A* **117**, 6311-6322, (2013).

10 Hennig, R. *et al.* IM30 triggers membrane fusion in cyanobacteria and chloroplasts. *Nat Commun* **6**, (2015).

11 Bellucci, L. *et al.* The interaction with gold suppresses fiber-like conformations of the amyloid beta (16-22) peptide. *Nanoscale* **8**, 8737-8748, (2016).

12 King, L. S., Kozono, D. & Agre, P. From structure to disease: the evolving tale of aquaporin biology. *Nat Rev Mol Cell Biol* **5**, 687-698, (2004).

13 Ishibashi, K., Hara, S. & Kondo, S. Aquaporin water channels in mammals. *Clin Exp Nephrol* **13**, 107-117, (2009).

14 Rojek, A., Praetorius, J., Frokiaer, J., Nielsen, S. & Fenton, R. A. A current view of the mammalian aquaglyceroporins. *Annu Rev Physiol* **70**, 301-327, (2008).

15 Fruhbeck, G., Catalan, V., Gomez-Ambrosi, J. & Rodriguez, A. Aquaporin-7 and glycerol permeability as novel obesity drug-target pathways. *Trends Pharmacol Sci* **27**, 345-347, (2006).

16 Verkman, A. S., Anderson, M. O. & Papadopoulos, M. C. Aquaporins: important but elusive drug targets. *Nat Rev Drug Discov* **13**, 259-277, (2014).

17 Badaut, J., Ashwal, S. & Obenaus, A. Aquaporins in cerebrovascular disease: a target for treatment of brain edema? *Cerebrovasc Dis* **31**, 521-531, (2011).

18 Son, M., Kim, D., Park, K. S., Hong, S. & Park, T. H. Detection of aquaporin-4 antibody using aquaporin-4 extracellular loop-based carbon nanotube biosensor for the diagnosis of neuromyelitis optica. *Biosens Bioelectron* **78**, 87-91, (2016).

19 Eriksson, U. K. *et al.* Subangstrom Resolution X-Ray Structure Details Aquaporin-Water Interactions. *Science* **340**, 1346-1349, (2013).

20 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* **21**, 171-181, (2004).

21 El Haitami, A., Backus, E. H. G. & Cantin, S. Synthesis at the Air-Water Interface of a Two-Dimensional Semi-Interpenetrating Network Based on

Poly(dimethylsiloxane) and Cellulose Acetate Butyrate. *Langmuir* **30**, 11919-11927, (2014).

22 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* **129**, 10056-+, (2007).

23 Barth, A. Infrared spectroscopy of proteins. *Bba-Bioenergetics* **1767**, 1073-1101, (2007).

24 Tajkhorshid, E. *et al.* Control of the selectivity of the aquaporin water channel family by global orientational tuning. *Science* **296**, 525-530, (2002).

25 Wang, Y., Schulten, K. & Tajkhorshid, E. What Makes an Aquaporin a Glycerol Channel? A Comparative Study of AqpZ and GlpF. *Structure* **13**, 1107-1118, (2005).

26 Cymer, F. & Schneider, D. A single glutamate residue controls the oligomerization, function, and stability of the aquaglyceroporin GlpF. *Biochemistry-Us* **49**, 279-286, (2010).

27 Canfield, M. C., Tamarappoo, B. K., Moses, A. M., Verkman, A. S. & Holtzman, E. J. Identification and characterization of aquaporin-2 water channel mutations causing nephrogenic diabetes insipidus with partial vasopressin response. *Hum Mol Genet* **6**, 1865-1871, (1997).

28 Liu, K. *et al.* Conversion of aquaporin 6 from an anion channel to a water-selective channel by a single amino acid substitution. *Proc Natl Acad Sci U S A* **102**, 2192-2197, (2005).

29 Marr, N. *et al.* Cell-biologic and functional analyses of five new Aquaporin-2 missense mutations that cause recessive nephrogenic diabetes insipidus. *J Am Soc Nephrol* **13**, 2267-2277, (2002).

30 Klein, N., Hellmann, N. & Schneider, D. Anionic Lipids Modulate the Activity of the Aquaglyceroporin GlpF. *Biophys J* **109**, 722-731, (2015).

31 Uzcategui, N. L. *et al.* Alteration in glycerol and metalloid permeability by a single mutation in the extracellular C-loop of Leishmania major aquaglyceroporin LmAQP1. *Mol Microbiol* **70**, 1477-1486, (2008).

32 Pandey, R. *et al.* Ice-nucleating bacteria control the order and dynamics of interfacial water. *Sci Adv* **2**, e1501630, (2016).

33 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. *P Natl Acad Sci USA* **96**, 2036-2041, (1999).

34 Tsuboi, M., Ikeda, T. & Ueda, T. Raman Microscopy of a Small Uniaxial Crystal - Tetragonal Aspartame. *J Raman Spectrosc* **22**, 619-626, (1991).

35 Pajcini, V. *et al.* Glycylglycine pi->pi* acid charge transfer transition moment orientations: Near-resonance Raman single-crystal measurements. *J Am Chem Soc* **118**, 9716-9726, (1996).

36 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* **102**, 6123-6138, (1998).

37 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* **122**, (2005).

38 Krimm, S. & Abe, Y. Intermolecular Interaction Effects in Amide I Vibrations of Beta Polypeptides. *P Natl Acad Sci USA* **69**, 2788-2792, (1972).

39 Berendsen, H. J. C., Vanderspoel, D. & Vandrunen, R. Gromacs - a Message-Passing Parallel Molecular-Dynamics Implementation. *Comput Phys Commun* **91**, 43-56, (1995).

40 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* **79**, 926-935, (1983).

41 Tieleman, D. P. & Berendsen, H. J. C. A molecular dynamics study of the pores formed by E-coli OmpF porin in a fully hydrated pope bilayer. *Biophys J* **74**, A392-A392, (1998).
42 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* **72**, 2002-2013, (1997).
43 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 Biophy* **34**, 273-284, (2005).
44 Lindahl, E., Hess, B. & van der Spoel, D. GROMACS 3.0: a package for molecular simulation and trajectory analysis. *J Mol Model* **7**, 306-317, (2001).
45 Hess, B., Bekker, H., Berendsen, H. J. C. & Fraaije, J. G. E. M. LINCS: A linear constraint solver for molecular simulations. *J Comput Chem* **18**, 1463-1472, (1997).
46 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-Us* **37**, 17554-17561, (1998).
47 Bussi, G., Donadio, D. & Parrinello, M. Canonical sampling through velocity rescaling. *J Chem Phys* **126**, (2007).
48 Nose, S. A Unified Formulation of the Constant Temperature Molecular-Dynamics Methods. *J Chem Phys* **81**, 511-519, (1984).
49 Basconi, J. E. & Shirts, M. R. Effects of Temperature Control Algorithms on Transport Properties and Kinetics in Molecular Dynamics Simulations. *J Chem Theory Comput* **9**, 2887-2899, (2013).

Supporting information:
Refining the X-ray crystal structure of aquaporin with simulations and surface vibrational spectroscopy

Lars Schmüser[a], Margareta Trefz[b], Steven Roeters[c], Wesley Beckner[d], Jim Pfaendtner[d], Sander Woutersen[c], Mischa Bonn[a], Dirk Schneider[b] and Tobias Weidner[a, e]*

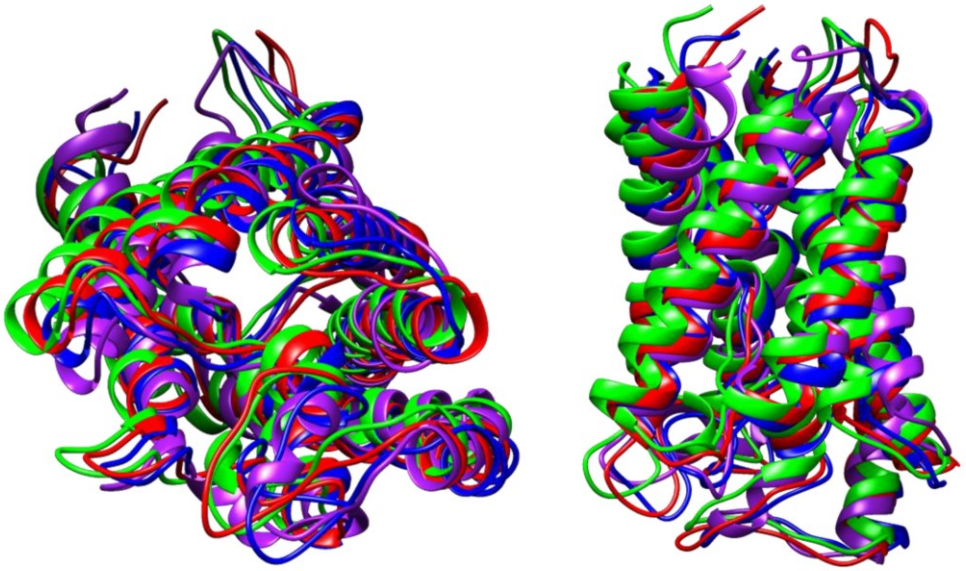


Figure S5 Comparison between the crystal structure (1lda, purple) and MD snapshots after 100ns (red), 75ns (green) and 50ns (blue). The structural overlays were created using the chimera tool matchmaker. Relaxing the protein with MD simulation changed slightly the tilt angles of its alpha helical elements.

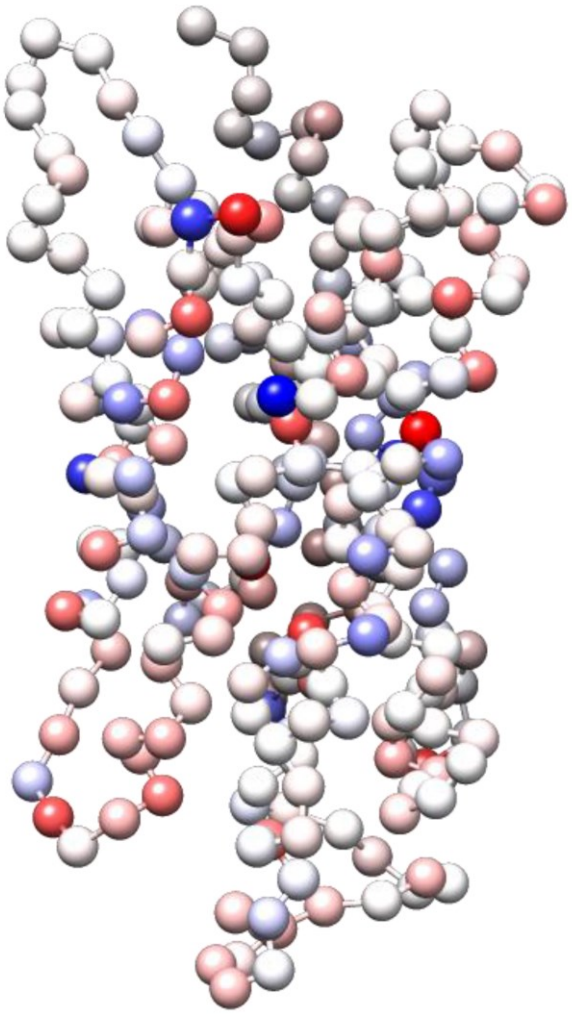


Figure S6 Averaged contributions of amino acids to the eigenmodes around 1600 cm^{-1} proving that the origin of the main peak at 1600 cm^{-1} is broadly distributed over the whole protein. Blue and red spheres are indicating a high contribution with a negative (blue) or positive (red) amplitude. This figure is plotted using Chimera (UCSF). The amplitudes for each eigenmode are derived from the Mathematica script, which calculates the SFG spectra.

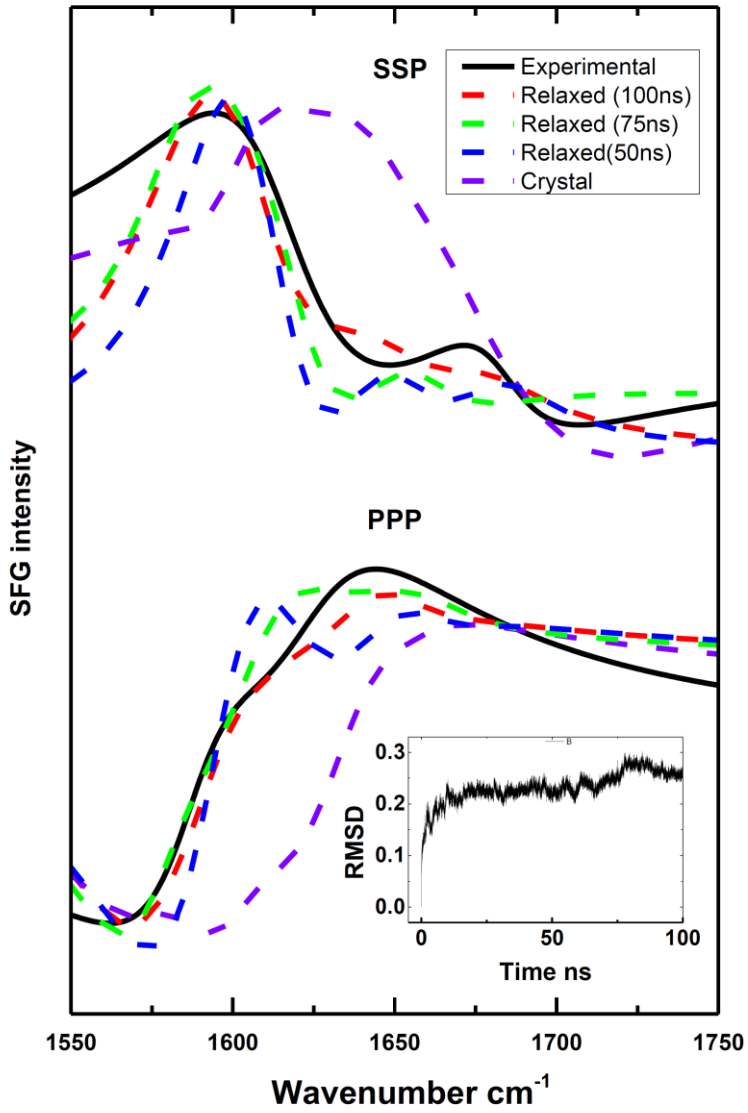


Figure S7 Comparison of the experimental SFG spectra (black solid line) with calculated SFG spectra of the crystal structure (dashed purple line) and of different snapshots of the MD simulation (green 100 ns, red 75 ns, and blue 50 ns, see S1). The inlet represents the root mean square deviation of atomic distances between the starting structure (the crystal structure) and structures of the MD simulation, plotted against the simulation time.

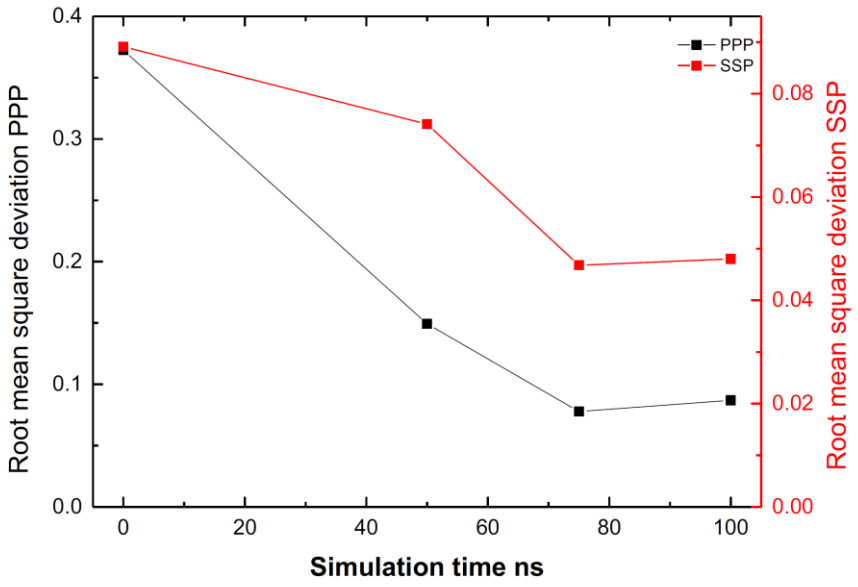


Figure S8 Quantification of the structural agreement between the aquaporin GlpF in physiological condition with the X-ray crystallography structure (PDB: 1lda) and with the structures derived from MD simulations. The differences between the experimental SFG spectra and those which were calculated from the PDB file 1lda and the MD simulations were determined using the root mean square deviations (RMSD) between experimental and calculated SFG spectra. The RMSD was calculated using the following equation:

$$RMSD = \sqrt{\sum_{\omega=1550}^{1750} \frac{(I\,exp_{\omega} - I\,calc_{\omega})^2}{200}}$$

Here, ω is the wavenumber and $I\,exp_{\omega}$ and $I\,calc_{\omega}$ are the experimental and calculated SFG intensities respectively. The calculated SFG spectra in Fig S3 were scaled to match the experimental spectra.

4.3 Orientation analysis of α-helices at interfaces using heterodyne detected SFG spectroscopy

The following manuscript was submitted to The Journal of Physical Chemistry Letters in November 2016. Here, the experimental and the calculated SFG spectra were part of this thesis. The MD simulations of the peptides at the water / air interface have been done by Helmut Lutz at the Max Planck Institute for Polymer Research in Mainz.

Determination of absolute orientation of α-helices at interfaces using phase resolved SFG spectroscopy

*Lars Schmüser,[1] Helmut Lutz,[1] Steven Roeters,[2] Sander Woutersen,[2] Mischa Bonn,[1] Tobias Weidner[1,3]**

[1]Molecular Spectroscopy Department, Max Planck Institute for Polymer Research, 55128 Mainz, Germany.

[2]Van 't Hoff Institute for Molecular Sciences, University of Amsterdam, 1098 EP, Amsterdam, The Netherlands

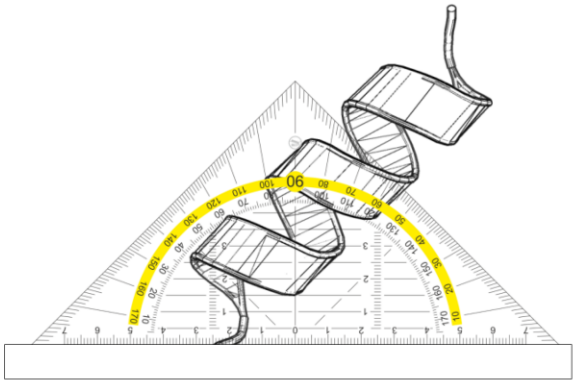
[3]Department of Chemistry, Aarhus University, 8000 Aarhus C, Denmark

AUTHOR INFORMATION

Corresponding Author

*Correspondence should be addressed to Tobias Weidner: weidner@chem.au.dk

TOC GRAPHICS



KEYWORDS. Sum frequency generation, protein adsorption, air water interface, leucine, lysine, surface

Abstract.

Understanding the structure of proteins at surfaces is key in fields such as biomaterials research, biosensor design, membrane biophysics and drug design. A particularly important factor is the orientation of proteins when bound to a particular surface. The orientation of the active site of enzymes or protein sensors, the availability of binding pockets within membrane proteins – these are important design parameters for engineers developing new sensors, surfaces and drugs. Recently developed methods to probe protein orientation, including immunoessays and mass spectrometry, either lack structural resolution or require harsh experimental conditions. We here report a new method to track the absolute orientation of interfacial proteins using phase resolved sum frequency generation spectroscopy in combination with molecular dynamics simulations and theoretical spectra calculations. As a model system we have determine the orientation a helical lysine-leucine peptide at the air-water interface. The data show that the absolute orientation of the helix can be reliably determined even for orientations almost parallel to the surface.

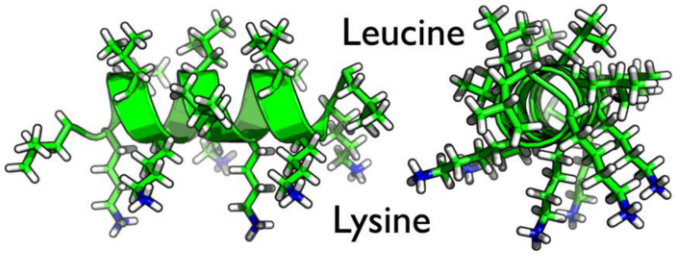
A molecular understanding of the interaction of proteins or peptides with surfaces is essential for the rational design of biomaterials for use in implants and sensors but also for comprehending the interfacial biology at lipid membranes and mineralized tissue[1-3]. An important question for biologists, bioengineers and pharmacologists is the orientation of proteins when bound to surfaces: Is the enzyme active site within a protein chip pointing towards the solute for effective detection?[4] Which parts of a membrane protein are accessible for interaction with therapeutic drugs?[3] Which sites are in contact with minerals and control the nucleation of hard tissue?[5] Despite the importance, the experimental observation of protein orientation is still challenging. The most established methods are based on enzyme-linked immunoassays,[6-7] where the accessibility of antibody binding sites is observed. While this method provides important practical information for bioengineers, the detail of structural information obtained is limited. Time of flight secondary ion mass spectrometry (ToF-SIMS) has recently been developed to probe protein orientation using its extremely shallow probing depth to detect asymmetries in the amino acid distribution within ordered protein films to determine the molecular orientation[8]. A drawback of ToF-SIMS is that the experiments require ultra-high vacuum conditions, which makes sample preparation difficult and can alter or destabilize protein structures. Nuclear magnetic resonance[9] and electron paramagnetic relaxation[10] can determine the orientation of proteins but need series of isotope and spin labeled proteins, which are time consuming, often difficult to produce and limit the size of proteins that can be studied.

Over the past decade sum frequency generation (SFG) spectroscopy has been developed into a reliable method to probe proteins of arbitrary size at interfaces in situ and in real time without the need for labels. In an SFG experiment, infrared and visible laser beams are overlapped in space and time at an interface to generate sum frequency photons by optical frequency mixing.[11] When the IR light is in resonance with a vibrational mode of molecules at the interface, this leads to an enhancement of the SFG process and corresponding features in the resulting spectra. In analogy to IR and Raman spectroscopy, SFG in the amide I region (1500-1800 cm^{-1}) allows to determine the secondary and tertiary structure of proteins. SFG has been used to determine the structure of proteins on membranes, polymers, self-assembled monolayers and inorganic surfaces[12-15]. Reliable protocols have been developed to determine secondary structure and tilt angle of entire proteins or individual protein sites[16] with respect to the surface.

However, published methods for determining the orientation of protein backbones are based on homodyne SFG intensity spectra[17-19] , and as such provide tilt angles modulo 180°, i.e. are unable to tell the *absolute* orientation of the protein, pointing 'up' or 'down', relative to the surface. As discussed above, for applications the absolute orientation is often a key factor. The inability to infer absolute orientation from SFG intensity spectra, originates from the fact that the intensity is proportional to the square of the SFG response of the protein ensemble. The SFG response is a complex variable, and in particular the sign of the imaginary part, or, equivalently, the phase of the SFG response directly reflects the absolute orientation. Oppositely oriented moieties are precisely 180° out of phase. The phase of a signal can be determined by interfering the SFG signal with a reference signal of known phase.

Phase-resolved SFG has been used to determine the chirality, but not orientation of interfacial proteins[24-27]. It has been used to infer the absolute orientation of small molecules at interfaces, which requires, however, knowledge of the orientation of the transition dipole moment relative to the molecular axis; for methyl, carbonyl and water molecules[20-23], this relation is well known. However, for proteins, the transition dipole moment across all backbone carbonyls relative to the orientation of the protein is unknown. In other words, for protein systems, even knowing the phase of the signal does not directly provide absolute orientation information. This requires, in addition, knowledge of the directionality of the transition dipole moment in the protein molecular frame. Here, we show that for a model protein, combining heterodyne SFG with transition dipole moment calculations, we can determine the absolute orientation of proteins at an interface.

The experiments were performed with the well-established model peptide LKα14, which is based only on leucine (L) lysine (K) side chains and an acetylated N-terminus (Ac-LKKLLKLLKKLLKL) and has been designed to adopt an α-helical secondary structure at the air-water interface (see Scheme 1)[28-29].



Scheme 1. *Model leucine (L) lysine (K) peptides LKα14 are designed to adopt α-helical secondary structures at interfaces. LKα14 peptides are amphiphilic with hydrophobic leucines and hydrophilic lysines orienting to opposite sides of the helix.*

To determine the amide transition dipole moment LKα1 at the air-water interface, we have first performed molecular dynamics simulations (MD) of the assembly of the peptides at the air-water interface. A snapshot of the MD simulation, taken at 100 ns, is shown in Figures 1a,b. The snapshot illustrates that the peptides assume a mostly α-helical secondary structure at the interface. The hydrophobic leucines point out of the water phase and the lysines are in contact with the water. This general conformation is in agreement with previously published experimental and theoretical studies of LK peptides at water surfaces and other hydrophobic interfaces.[cite PNAS, Weidner Langmuir, Sayar, Pfaendtner, Marmut2006]

The simulation indicates the peptides are oriented largely parallel to the water surface, in good agreement with previous experimental and theoretical results. However, a more quantitative analysis of the orientation of the peptide backbone orientation and the average amide transition dipole moment (TDM) orientation shown in Figure 1c show that the peptides are not oriented entirely parallel to the water surface (see SI for more details about the analysis). The amide TDM and the long peptide helix axis are tilted, with angles of 80° and 82°, respectively, with respect to the surface normal and the N-termini pointing towards the water phase.

Phase-resolved SFG spectra for a monolayer of LKα14 at the air-water interface are shown in Figure 2: Panels a and b display the imaginary and real SFG spectra, respectively. The imaginary spectrum exhibits a pronounced peak near 1630 cm^{-1}. The real spectrum shows the related zero crossing close to 1630 cm^{-1}.

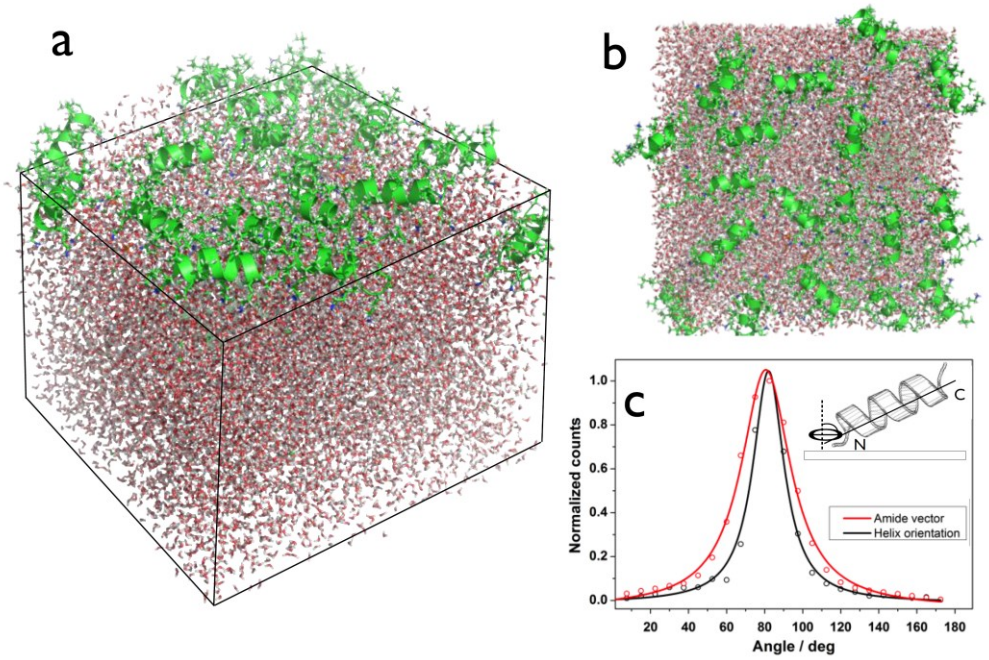


Figure 1. *Snapshot of the MD simulation of LKα14 at the air-water interface at time point 100 ns. (a) Side view and (b) top view of the simulation box. The LKα14 peptides predominantly assume α-helical secondary structures. While the leucine side chains are oriented away from the water surface, the lysine side chains are interacting with the water. (c) Orientation analysis if the amide bonds and the helix axes of the peptides.*

For a close comparison of the experimental SFG data with the simulation, we calculated theoretical imaginary and real SFG spectra from the 100 ns MD snapshot shown in Figure 1. The results are included in Figure 2 (red lines). Clearly, the calculated spectra capture the spectral shape and polarity of both the imaginary and the real experimental spectra well. The sensitivity to the peptide absolute orientation, even at a 'flat' angle of 82° becomes apparent when calculating the related spectra for a 180° rotated snapshot as a reference. The spectra of the inverted snapshot (black line) neither match the spectra original simulation nor the experimental spectra.

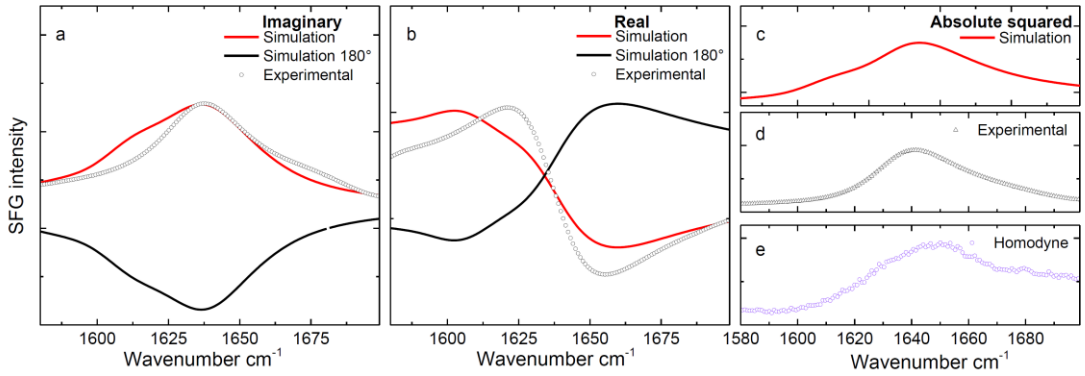
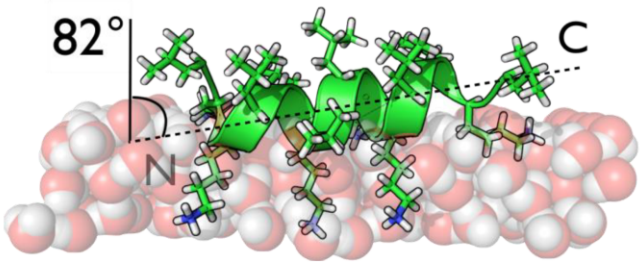


Figure 2. *Experimental and theoretical SFG spectra for LKα14 at the air water interface. (a) Experimental (dots) imaginary spectrum along spectra calculated for the MD snapshot (red) and the snapshot rotated by 180° (black). (b) Calculated real SFG spectra. (c) Absolute squared spectrum calculated from the simulation spectra (d) Absolute squared spectrum constructed from the experimental heterodyne spectra. (e) Homodyne detected SFG spectrum as reference.*

To verify the consistency of the measured phase resolved spectra with the related and well-known homodyne detected spectra we calculated the absolute square of the imaginary and real spectra parts and compared with experimental homodyne spectra and homodyne spectra calculated from the simulation. The results are summarized in Figure 2c. The heterodyne and homodyne spectra as well as the calculations are in excellent agreement and exhibit a peak near 1645 cm^{-1}, which agrees well with literature values.

In summary we have determined the absolute orientation of the model peptide LKα14 at the air water interface using a combination of phase-resolved SFG data, MD simulations and spectra calculations. We found, that the widely studied model peptide LKα14 on average adopts a rather planar orientation at water surfaces and is slightly tilted (82°) with the N-terminus oriented towards the water phase (Scheme 2). Orientation measurements of water molecules, hydrocarbons and protein components have been successfully used for years to answer questions in physical chemistry. Knowledge of the absolute orientation of proteins is essential for protein-based biosensors, drug design and the development of new biocompatible coatings. The procedure presented here allows direct assessment of absolute protein orientation by a comparison of simulated structures with SFG spectroscopic data.



Scheme 2. *Illustration of the average orientation of the LK peptides at the air water interface as determined from the MD simulations and phase resolved SFG experiments. The N-terminus is pointing into the water phase.*

Experimental Methods

Details of the phase specific SFG setup have been described before. Briefly, a 10 mJ laser pulse with a pulse duration of 40 fs is generated by a Ti:Sa amplified laser (Spitfire Ace, centred at 800). 2 mJ of the amplifier output is branched out to pump an optical paramagnetic amplifier (TOPAS, light conversion) which generates 2 µJ broadband pulses centered at 6.1 µm with a bandwidth of roughly 300cm^{-1}. Another 1mJ is guided through a Fabry-Perot Etalon (SLS Optics Ltd.) to narrow the broadband 800 nm pulses to a bandwidth of 25 cm^{-1} with 3 µJ pulse energy. IR and visible pulse are focused on a gold mirror with a 100 cm (visible) and a 5 cm (IR) focal length lens and overlapped in space and time to generate the local oscillator. Local oscillator, IR and visible pulses are refocused to the sample using a concave mirror. The local oscillator is either delayed by a 1mm thick fused silica plate for phase specific heterodyne detected SFG spectra or blocked with a metal plate for homodyne detected SFG spectra. The local oscillator and the samples SFG pulses are collimated and focused to a spectrometer and finally detected by an EMCCD camera. Subsequently, the imaginary and real parts are deduced from the interference spectrum by Fourier transformation. The formalism explaining this procedure has been described in detail before[22, 30].

The spectra are calculated with an amide-I exciton model for the backbone amide groups, based on the formalism described in ref.[17]. Briefly, the Hamiltonian is constructed with (I) the local mode gas phase frequencies modulated by a hydrogen-bond induced red-shift on the diagonal,

(II) for the nearest neighbors, a through-bond coupling model based on parameterized quantum-chemical calculations with the 6-31G+(d) basis set and B3LYP-functional that correlates the dihedral angles between the neighboring amide groups to the interaction strength[31], and (III) for the non-nearest neighbors, the Transition Dipole Coupling (TDC) model is used that calculates the through-space coupling between to amide groups based on their relative orientation and distance[32]. By diagonalizing the Hamiltonian, the eigenvalues and eigenmodes of the normal modes are obtained, from which the IR, Raman and SFG responses are calculated, and the heterodyned spectra using:

$$\chi_{ijk,RE} = Re\left(\sum_{\nu} \frac{\chi_{ijk,\nu}\sqrt{\Gamma_{Exc}+\Gamma_{\nu}}}{\omega_{\nu}+\omega-i(\Gamma_{Exc}+\Gamma_{\nu})}\right)$$

$$\chi_{ijk,Im} = Im\left(\sum_{\nu} \frac{\chi_{ijk,\nu}\sqrt{\Gamma_{Exc}+\Gamma_{\nu}}}{\omega_{\nu}+\omega-i(\Gamma_{Exc}+\Gamma_{\nu})}\right)$$

, with $\chi_{ijk,\nu}$, Γ_{ν} and ω_{ν} the hyperpolarizability, linewidth and frequency of normal mode ν. We assume all Γ_{ν} are 5 cm^{-1} and set Γ_{Exc}, the width of the visible pulse, to 7.5 cm^{-1}.

Molecular Dynamics Simulations: Simulations were carried out using the GROMACS 4.6 simulation package.[33] Peptide parameters are defined in the AMBER99SB-ildn force field,[34] parameters for phosphate ions were adopted from ref.[35] and the TIP3P Water model was used. The initial state was packed using Packmol,[36] with 23 peptides at the vacuum-water interface of 8 x 8 x 7 nm of vacuum and 8 x 8 x 6.8 nm of water, containing 10 phosphate anions and sufficient chloride anions to neutralize the simulation box. Thereby the system can be described as 6.8 nm slabs of water separated by 7 nm of vacuum. Topology and coordinate files were generated with Tleap and subsequently converted to GROMACS-type files using Acpype.[37-38] Constraining all bonds with the LINCS algorithm the simulations were run at a 2 fs timestep with periodic boundary conditions in all three dimensions for 100 ns, using the particle mesh Ewald (PME) method for long range Coulomb interactions.[39] The distance for Lennard-Jones potential cut-off was set to 1 nm and the temperature was maintained at 300 K using velocity rescaling with a stochastic term.[40] The tilt of the peptide chain at the vacuum-water interface was

calculated using a python implementation of a method developed by Kahn, based on the alpha carbon atoms along the peptide chain, disregarding the first and last 2 amino acids.[41]

AUTHOR INFORMATION

ACKNOWLEDGMENT

This work was supported by the Max Planck Graduate Center with the Johannes Gutenberg-Universität Mainz (MPGC). LS and TW thank the Deutsche Forschungsgemeinschaft (WE4478/4-1) for financial support. We thank Dr. Ellen Backus for helpful discussions.

REFERENCES

1. Lu, H.; Hood, M. A.; Mauri, S.; Baio, J. E.; Bonn, M.; Munoz-Espi, R.; Weidner, T., Biomimetic Vaterite Formation at Surfaces Structurally Templated by Oligo(Glutamic Acid) Peptides. *Chemical Communications* **2015**, *51*, 15902-15905.

2. Ding, B.; Jasensky, J.; Li, Y. X.; Chen, Z., Engineering and Characterization of Peptides and Proteins at Surfaces and Interfaces: A Case Study in Surface-Sensitive Vibrational Spectroscopy. *Accounts Chem Res* **2016**, *49*, 1149-1157.

3. Yin, H.; Flynn, A. D., Drugging Membrane Protein Interactions. *Annu Rev Biomed Eng* **2016**, *18*, 51-76.

4. Brady, D.; Jordaan, J., Advances in Enzyme Immobilisation. *Biotechnol Lett* **2009**, *31*, 1639-1650.

5. Lutz, H.; Jaeger, V.; Berger, R.; Bonn, M.; Pfaendtner, J.; Weidner, T., Biomimetic Growth of Ultrathin Silica Sheets Using Artificial Amphiphilic Peptides. *Adv Mater Interfaces* **2015**, *2*.

6. Rossier, J. S.; Girault, H. H., Enzyme Linked Immunosorbent Assay on a Microchip with Electrochemical Detection. *Lab on a Chip* **2001**, *1*, 153-157.

7. Castner, D. G., Surface Science: View from the Edge. *Nature* **2003**, *422*, 129-130.

8. Baio, J. E.; Weidner, T.; Ramey, D.; Pruzinsky, L.; Castner, D. G., Probing the Orientation of Electrostatically Immobilized Cytochrome C by Time of Flight Secondary Ion Mass Spectrometry and Sum Frequency Generation Spectroscopy. *Biointerphases* **2013**, *8*.

9. Strandberg, E.; Wadhwani, P.; Tremouilhac, P.; Dürr, U. H. N.; Ulrich, A. S., Solid-State Nmr Analysis of the Pgla Peptide Orientation in Dmpc Bilayers: Structural Fidelity of 2h-Labels Versus High Sensitivity of 19f-Nmr. *Biophysical Journal* **2006**, *90*, 1676-1686.

10. Stevens, M. A.; McKay, J. E.; Robinson, J. L. S.; El Mkami, H.; Smith, G. M.; Norman, D. G., The Use of the Rx Spin Label in Orientation Measurement on Proteins, by Epr. *Physical Chemistry Chemical Physics* **2016**, *18*, 5799-5806.

11. Shen, Y. R., *The Principles of Nonlinear Optics*, 1 ed.; John Wiley & Sons: New York, 1984.

12. Zhang, C.; Myers, J. N.; Chen, Z., Elucidation of Molecular Structures at Buried Polymer Interfaces and Biological Interfaces Using Sum Frequency Generation Vibrational Spectroscopy. *Soft Matter* **2013**, *9*, 4738-4761.

13. Weidner, T.; Castner, D. G., Sfg Analysis of Surface Bound Proteins: A Route Towards Structure Determination. *Phys Chem Chem Phys* **2013**, *15*, 12516-24.

14. Hennig, R., et al., Im30 Triggers Membrane Fusion in Cyanobacteria and Chloroplasts. *Nat Commun* **2015**, *6*.

15. Baio, J. E.; Zane, A.; Jaeger, V.; Roehrich, A. M.; Lutz, H.; Pfaendtner, J.; Drobny, G. P.; Weidner, T., Diatom Mimics: Directing the Formation of Biosilica Nanoparticles by Controlled Folding of Lysine-Leucine Peptides. *J Am Chem Soc* **2014**, *136*, 15134-7.

16. Ding, B.; Panahi, A.; Ho, J. J.; Laaser, J. E.; Brooks, C. L.; Zanni, M. T.; Chen, Z., Probing Site-Specific Structural Information of Peptides at Model Membrane Interface in Situ. *J Am Chem Soc* **2015**, *137*, 10190-10198.

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*, 6311-6322.

18. Nguyen, K. T.; Le Clair, S. V.; Ye, S. J.; Chen, Z., Orientation Determination of Protein Helical Secondary Structures Using Linear and Nonlinear Vibrational Spectroscopy. *J Phys Chem B* **2009**, *113*, 12169-12180.

19. Nguyen, K. T.; King, J. T.; Chen, Z., Orientation Determination of Interfacial Beta-Sheet Structures in Situ. *J Phys Chem B* **2010**, *114*, 8291-8300.

20. 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*.

21. Pool, R. E.; Versluis, J.; Backus, E. H. G.; Bonn, M., Comparative Study of Direct and Phase-Specific Vibrational Sum-Frequency Generation Spectroscopy: Advantages and Limitations. *J Phys Chem B* **2011**, *115*, 15362-15369.

22. Nihonyanagi, S.; Yamaguchi, S.; Tahara, T., Direct Evidence for Orientational Flip-Flop of Water Molecules at Charged Interfaces: A Heterodyne-Detected Vibrational Sum Frequency Generation Study. *J Chem Phys* **2009**, *130*.

23. 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.

24. Lotze, S.; Versluis, J.; Olijve, L. L.; van Schijndel, L.; Milroy, L. G.; Voets, I. K.; Bakker, H. J., Communication: Probing the Absolute Configuration of Chiral Molecules at Aqueous Interfaces. *J Chem Phys* **2015**, *143*, 201101.

25. Strazdaite, S.; Meister, K.; Bakker, H. J., Orientation of Polar Molecules near Charged Protein Interfaces. *Phys Chem Chem Phys* **2016**, *18*, 7414-8.

26. Okuno, M.; Ishibashi, T.-a., Heterodyne-Detected Achiral and Chiral Vibrational Sum Frequency Generation of Proteins at Air/Water Interface. *The Journal of Physical Chemistry C* **2015**, *119*, 9947-9954.

27. Okuno, M.; Ishibashi, T. A., Chirality Discriminated by Heterodyne-Detected Vibrational Sum Frequency Generation. *J Phys Chem Lett* **2014**, *5*, 2874-8.

28. Degrado, W. F.; Wasserman, Z. R.; Lear, J. D., Protein Design, a Minimalist Approach. *Science* **1989**, *243*, 622-628.

29. 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.

30. 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 Edit* **2014**, *53*, 8146-8149.

31. 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*.

32. Krimm, S.; Abe, Y., Intermolecular Interaction Effects in Amide I Vibrations of Beta Polypeptides. *P Natl Acad Sci USA* **1972**, *69*, 2788-2792.

33. Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E., Gromacs 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *J Chem Theory Comput* **2008**, *4*, 435-47.

34. Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.; Shaw, D. E., Improved Side-Chain Torsion Potentials for the Amber Ff99sb Protein Force Field. *Proteins* **2010**, *78*, 1950-8.

35. Lutz, H.; Jaeger, V.; Berger, R.; Bonn, M.; Pfaendtner, J.; Weidner, T., Biomimetic Growth of Ultrathin Silica Sheets Using Artificial Amphiphilic Peptides. *Advanced Materials Interfaces* **2015**, *2*, n/a-n/a.

36. Martinez, L.; Andrade, R.; Birgin, E. G.; Martinez, J. M., Packmol: A Package for Building Initial Configurations for Molecular Dynamics Simulations. *J Comput Chem* **2009**, *30*, 2157-64.

37. Case, D.; Babin, V.; Berryman, J.; Betz, R.; Cai, Q.; Cerutti, D.; Cheatham Iii, T.; Darden, T.; Duke, R.; Gohlke, H., Amber 14. **2014**.

38. Sousa da Silva, A. W.; Vranken, W. F., Acpype - Antechamber Python Parser Interface. *BMC Res Notes* **2012**, *5*, 367.

39. 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.

40. Bussi, G.; Donadio, D.; Parrinello, M., Canonical Sampling through Velocity Rescaling. *J. Chem. Phys.* **2007**, *126*.

41. Kahn, P. C., Defining the Axis of a Helix. *Comput. Chem.* **1989**, *13*, 185-189.

5.Conclusion and Outlook

The goal of the work presented in this thesis was to establish a way to maximize the information gained from complex SFG spectra of proteins in the amide-I region. The comparison of the experimental SFG spectra with calculated spectra helped with the interpretation of the experimental SFG spectra. Furthermore, phase resolved SFG was employed to obtain information on the absolute orientation of proteins at interphases. Both approaches were employed successfully and have led to publications or submissions to peer reviewed journals.

For the future, these methods could be further improved by extending the MD simulations to multiple proteins to account for intermolecular couplings and further intermolecular interactions. This would require significantly higher computational power for the MD simulation and for calculating the SFG spectra. This is mainly because of the slow data processing of an interpreter language like the currently used Mathematica. Translating the current script to a compiler language like C, C++ or Fortran would increase the calculation speed dramatically.

In this thesis, the used lipid model systems evolved from a very basic lipid monolayer system to a more advanced supported lipid bilayer. However, both membrane model systems cannot be considered as a perfect choice for protein membrane interactions since in both cases only one lipid monolayer exhibits the full flexibility of a physiological membrane. Furthermore, the extremely low number of lipid molecules in a two-dimensional membrane renders it impossible to work in a physiological relevant stoichiometry. Over time, all binding sites will be occupied by proteins or the membrane will be even partially replaced. A completely covered membrane could possibly prevent refolding or reorientation events of the bound proteins due to steric hindrance[62]. Therefore, a three-dimensional model membrane system like vesicles, in which a physiological protein to lipid ratio can be achieved, would be the next logical step towards more physiological conditions.

A structural picture of a protein says more than a thousand words – imagine what a movie could do. The Holy Grail remains a molecular movie of protein folding, refolding or membrane interaction confirmed with time resolved SFG spectroscopy with ps or even sub-ps time resolution. This must be achieved with pump probe SFG setup, where an arrested state of a protein is optically activated (pump) followed by a probe pulse with an adjustable time delay. A possible intermediate step for protein membrane interactions could be the implementation of a stopped-flow mechanism which would allow for sub-ms time resolutions.

6.References

1. Berg, J. M.; Tymoczko, J. L.; Gatto, G. J.; Stryer, L., *Biochemistry*, 2015.
2. Voet, D.; Voet, J. G.; Pratt, C. W., *Fundamentals of Biochemistry : Life at the Molecular Level*; Wiley: Hoboken, NJ, 2013.
3. Bock, A.; Forchhammer, K.; Heider, J.; Leinfelder, W.; Sawers, G.; Veprek, B.; Zinoni, F., Selenocysteine - the 21st Amino-Acid. *Mol Microbiol* **1991**, *5*, 515-520.
4. Weidner, T.; Apte, J. S.; Gamble, L. J.; Castner, D. G., Probing the Orientation and Conformation of Alpha-Helix and Beta-Strand Model Peptides on Self-Assembled Monolayers Using Sum Frequency Generation and Nexafs Spectroscopy. *Langmuir* **2010**, *26*, 3433-3440.
5. Kendrew, J. C.; Bodo, G.; Dintzis, H. M.; Parrish, R. G.; Wyckoff, H.; Phillips, D. C., A Three-Dimensional Model of the Myoglobin Molecule Obtained by X-Ray Analysis. *Nature* **1958**, *181*, 662-6.
6. Carpenter, E. P.; Beis, K.; Cameron, A. D.; Iwata, S., Overcoming the Challenges of Membrane Protein Crystallography. *Curr Opin Struc Biol* **2008**, *18*, 581-586.
7. Deisenhofer, J.; Epp, O.; Miki, K.; Huber, R.; Michel, H., X-Ray Structure Analysis of a Membrane Protein Complex. Electron Density Map at 3 a Resolution and a Model of the Chromophores of the Photosynthetic Reaction Center from Rhodopseudomonas Viridis. *J Mol Biol* **1984**, *180*, 385-98.
8. Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E., The Protein Data Bank. *Nucleic Acids Res* **2000**, *28*, 235-242.
9. Greenfield, N. J., Using Circular Dichroism Spectra to Estimate Protein Secondary Structure. *Nature Protocols* **2006**, *1*, 2876-2890.
10. Barth, A., Infrared Spectroscopy of Proteins. *Bba-Bioenergetics* **2007**, *1767*, 1073-1101.
11. Tuma, R., Raman Spectroscopy of Proteins: From Peptides to Large Assemblies. *J Raman Spectrosc* **2005**, *36*, 307-319.
12. Yong, W.; Lomakin, A.; Kirkitadze, M. D.; Teplow, D. B.; Chen, S. H.; Benedek, G. B., Structure Determination of Micelle-Like Intermediates in Amyloid Beta-Protein Fibril Assembly by Using Small Angle Neutron Scattering. *P Natl Acad Sci USA* **2002**, *99*, 150-154.
13. Meesters, C.; Brack, A.; Hellmann, N.; Decker, H., Structural Characterization of the Alpha-Hemolysin Monomer from Staphylococcus Aureus. *Proteins-Structure Function and Bioinformatics* **2009**, *75*, 118-126.
14. Gatsogiannis, C.; Markl, J., Keyhole Limpet Hemocyanin: 9-Angstrom Cryoem Structure and Molecular Model of the Klh1 Didecamer Reveal the Interfaces and Intricate Topology of the 160 Functional Units. *J Mol Biol* **2009**, *385*, 963-983.
15. Opella, S. J.; Marassi, F. M., Structure Determination of Membrane Proteins by Nmr Spectroscopy. *Chem Rev* **2004**, *104*, 3587-3606.
16. O'Dell, W. B.; Bodenheimer, A. M.; Meilleur, F., Neutron Protein Crystallography: A Complementary Tool for Locating Hydrogens in Proteins. *Arch Biochem Biophys* **2016**, *602*, 48-60.
17. Doglia, S. M.; Ami, D.; Natalello, A.; Gatti-Lafranconi, P.; Lotti, M., Fourier Transform Infrared Spectroscopy Analysis of the Conformational Quality of Recombinant Proteins within Inclusion Bodies. *Biotechnol J* **2008**, *3*, 193-201.
18. Kandori, H.; Yamazaki, Y.; Hatanaka, M.; Needleman, R.; Brown, L. S.; Richter, H. T.; Lanyi, J. K.; Maeda, A., Time-Resolved Fourier Transform Infrared Study of Structural Changes in the Last Steps of the Photocycles of Glu-204 and Leu-93 Mutants of Bacteriorhodopsin. *Biochemistry-Us* **1997**, *36*, 5134-41.
19. Chittur, K. K., Ftir/Atr for Protein Adsorption to Biomaterial Surfaces. *Biomaterials* **1998**, *19*, 357-69.
20. Ataka, K.; Heberle, J., Biochemical Applications of Surface-Enhanced Infrared Absorption Spectroscopy. *Anal Bioanal Chem* **2007**, *388*, 47-54.
21. Mendelsohn, R.; Mao, G.; Flach, C. R., Infrared Reflection-Absorption Spectroscopy: Principles and Applications to Lipid-Protein Interaction in Langmuir Films. *Biochim Biophys Acta* **2010**, *1798*, 788-800.
22. Wallin, E.; von Heijne, G., Genome-Wide Analysis of Integral Membrane Proteins from Eubacterial, Archaean, and Eukaryotic Organisms. *Protein Sci* **1998**, *7*, 1029-38.

23. Yildirim, M. A.; Goh, K. I.; Cusick, M. E.; Barabasi, A. L.; Vidal, M., Drug-Target Network. *Nat Biotechnol* **2007**, *25*, 1119-26.
24. Cedervall, T.; Lynch, I.; Lindman, S.; Berggard, T.; Thulin, E.; Nilsson, H.; Dawson, K. A.; Linse, S., Understanding the Nanoparticle-Protein Corona Using Methods to Quantify Exchange Rates and Affinities of Proteins for Nanoparticles. *Proc Natl Acad Sci U S A* **2007**, *104*, 2050-5.
25. Lundqvist, M.; Stigler, J.; Elia, G.; Lynch, I.; Cedervall, T.; Dawson, K. A., Nanoparticle Size and Surface Properties Determine the Protein Corona with Possible Implications for Biological Impacts. *Proc Natl Acad Sci U S A* **2008**, *105*, 14265-70.
26. Franz, J.; Graham, D. J.; Schmuser, L.; Baio, J. E.; Lelle, M.; Peneva, K.; Mullen, K.; Castner, D. G.; Bonn, M.; Weidner, T., Full Membrane Spanning Self-Assembled Monolayers as Model Systems for Uhv-Based Studies of Cell-Penetrating Peptides. *Biointerphases* **2015**, *10*.
27. Shen, Y. R., Surface-Properties Probed by 2nd-Harmonic and Sum-Frequency Generation. *Nature* **1989**, *337*, 519-525.
28. Wang, J.; Chen, C. Y.; Buck, S. M.; Chen, Z., Molecular Chemical Structure on Poly(Methyl Methacrylate) (Pmma) Surface Studied by Sum Frequency Generation (Sfg) Vibrational Spectroscopy. *J Phys Chem B* **2001**, *105*, 12118-12125.
29. Cecchet, F.; Lis, D.; Guthmuller, J.; Champagne, B.; Caudano, Y.; Silien, C.; Mani, A. A.; Thiry, P. A.; Peremans, A., Orientational Analysis of Dodecanethiol and P-Nitrothiophenol Sams on Metals with Polarisation-Dependent Sfg Spectroscopy. *Chemphyschem* **2010**, *11*, 607-615.
30. Hennig, R., et al., Im30 Triggers Membrane Fusion in Cyanobacteria and Chloroplasts. *Nat Commun* **2015**, *6*.
31. 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*, 6311-6322.
32. Macdonald, R. C.; Simon, S. A., Lipid Monolayer States and Their Relationships to Bilayers. *P Natl Acad Sci USA* **1987**, *84*, 4089-4093.
33. Schach, D. K.; Rock, W.; Franz, J.; Bonn, M.; Parekh, S. H.; Weidner, T., Reversible Activation of a Cell-Penetrating Peptide in a Membrane Environment. *J Am Chem Soc* **2015**, *137*, 12199-12202.
34. Liljeblad, J. F. D.; Bulone, V.; Tyrode, E.; Rutland, M. W.; Johnson, C. M., Phospholipid Monolayers Probed by Vibrational Sum Frequency Spectroscopy: Instability of Unsaturated Phospholipids. *Biophys J* **2010**, *98*, L50-L52.
35. Avery, C. W.; Palermo, E. F.; McLaughin, A.; Kuroda, K.; Chen, Z., Investigations of the Interactions between Synthetic Antimicrobial Polymers and Substrate-Supported Lipid Bilayers Using Sum Frequency Generation Vibrational Spectroscopy. *Anal Chem* **2011**, *83*, 1342-1349.
36. Ding, B.; Soblosky, L.; Nguyen, K.; Geng, J. Q.; Yu, X. L.; Ramamoorthy, A.; Chen, Z., Physiologically-Relevant Modes of Membrane Interactions by the Human Antimicrobial Peptide, Ll-37, Revealed by Sfg Experiments. *Sci Rep-Uk* **2013**, *3*.
37. Tamm, L. K.; Tatulian, S. A., Infrared Spectroscopy of Proteins and Peptides in Lipid Bilayers. *Q Rev Biophys* **1997**, *30*, 365-429.
38. Liu, J.; Conboy, J. C., Structure of a Gel Phase Lipid Bilayer Prepared by the Langmuir-Blodgett/Langmuir-Schaefer Method Characterized by Sum-Frequency Vibrational Spectroscopy. *Langmuir* **2005**, *21*, 9091-9097.
39. Raguse, B.; Braach-Maksvytis, V.; Cornell, B. A.; King, L. G.; Osman, P. D. J.; Pace, R. J.; Wieczorek, L., Tethered Lipid Bilayer Membranes: Formation and Ionic Reservoir Characterization. *Langmuir* **1998**, *14*, 648-659.
40. Knoll, W.; Frank, C. W.; Heibel, C.; Naumann, R.; Offenhausser, A.; Ruhe, J.; Schmidt, E. K.; Shen, W. W.; Sinner, A., Functional Tethered Lipid Bilayers. *J Biotechnol* **2000**, *74*, 137-58.
41. Johnson, S. J.; Bayerl, T. M.; McDermott, D. C.; Adam, G. W.; Rennie, A. R.; Thomas, R. K.; Sackmann, E., Structure of an Adsorbed Dimyristoylphosphatidylcholine Bilayer Measured with Specular Reflection of Neutrons. *Biophys J* **1991**, *59*, 289-94.
42. Soblosky, L.; Ramamoorthy, A.; Chen, Z., Membrane Interaction of Antimicrobial Peptides Using E-Coli Lipid Extract as Model Bacterial Cell Membranes and Sfg Spectroscopy. *Chem Phys Lipids* **2015**, *187*, 20-33.
43. Liu, J.; Conboy, J. C., 1,2-Diacyl-Phosphatidylcholine Flip-Flop Measured Directly by Sum-Frequency Vibrational Spectroscopy. *Biophys J* **2005**, *89*, 2522-2532.

44. Scomparin, C.; Lecuyer, S.; Ferreira, M.; Charitat, T.; Tinland, B., Diffusion in Supported Lipid Bilayers: Influence of Substrate and Preparation Technique on the Internal Dynamics. *Eur Phys J E Soft Matter* **2009**, *28*, 211-20.
45. Wagner, M. L.; Tamm, L. K., Tethered Polymer-Supported Planar Lipid Bilayers for Reconstitution of Integral Membrane Proteins: Silane-Polyethyleneglycol-Lipid as a Cushion and Covalent Linker. *Biophys J* **2000**, *79*, 1400-1414.
46. Giess, F.; Friedrich, M. G.; Heberle, J.; Naumann, R. L.; Knoll, W., The Protein-Tethered Lipid Bilayer: A Novel Mimic of the Biological Membrane. *Biophys J* **2004**, *87*, 3213-3220.
47. Kahya, N.; Scherfeld, D.; Bacia, K.; Schwille, P., Lipid Domain Formation and Dynamics in Giant Unilamellar Vesicles Explored by Fluorescence Correlation Spectroscopy. *J Struct Biol* **2004**, *147*, 77-89.
48. Boyd, R. W., Chapter 1 - the Nonlinear Optical Susceptibility. In *Nonlinear Optics (Third Edition)*, Academic Press: Burlington, 2008; pp 1-67.
49. Nye, J. F., *Physical Properties of Crystals : Their Representation by Tensors and Matrices / by J.F. Nye*; Clarendon: Oxford, 1957.
50. Zhuang, X.; Miranda, P. B.; Kim, D.; Shen, Y. R., Mapping Molecular Orientation and Conformation at Interfaces by Surface Nonlinear Optics. *Phys Rev B* **1999**, *59*, 12632-12640.
51. Baio, J. E.; Weidner, T.; Brison, J.; Graham, D. J.; Gamble, L. J.; Castner, D. G., Amine Terminated Sams: Investigating Why Oxygen Is Present in These Films. *J Electron Spectrosc* **2009**, *172*, 2-8.
52. Rzeznicka, I. I.; Sovago, M.; Backus, E. H. G.; Bonn, M.; Yamada, T.; Kobayashi, T.; Kawai, M., Duramycin-Induced Destabilization of a Phosphatidylethanolamine Monolayer at the Air-Water Interface Observed by Vibrational Sum-Frequency Generation Spectroscopy. *Langmuir* **2010**, *26*, 16055-16062.
53. Baio, J. E., et al., Reversible Activation of Ph-Sensitive Cell Penetrating Peptides Attached to Gold Surfaces. *Chem Commun* **2015**, *51*, 273-275.
54. Chen, X. Y.; Clarke, M. L.; Wang, J.; Chen, Z., Sum Frequency Generation Vibrational Spectroscopy Studies on Molecular Conformation and Orientation of Biological Molecules at Interfaces. *Int J Mod Phys B* **2005**, *19*, 691-713.
55. Carrier, O.; Backus, E. H. G.; Shahidzadeh, N.; Franz, J.; Wagner, M.; Nagata, Y.; Bonn, M.; Bonn, D., Oppositely Charged Ions at Water-Air and Water-Oil Interfaces: Contrasting the Molecular Picture with Thermodynamics. *J Phys Chem Lett* **2016**, *7*, 825-830.
56. 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*, 10056-+.
57. Nagata, Y.; Ohto, T.; Backus, E. H. G.; Bonn, M., Molecular Modeling of Water Interfaces: From Molecular Spectroscopy to Thermodynamics. *J Phys Chem B* **2016**, *120*, 3785-3796.
58. Lis, D.; Backus, E. H. G.; Hunger, J.; Parekh, S. H.; Bonn, M., Liquid Flow Along a Solid Surface Reversibly Alters Interfacial Chemistry. *Science* **2014**, *344*, 1138-1142.
59. Chen, X. Y.; Chen, Z., Sfg Studies on Interactions between Antimicrobial Peptides and Supported Lipid Bilayers. *Bba-Biomembranes* **2006**, *1758*, 1257-1273.
60. Chen, X. Y.; Wang, J.; Sniadecki, J. J.; Even, M. A.; Chen, Z., Probing Alpha-Helical and Beta-Sheet Structures of Peptides at Solid/Liquid Interfaces with Sfg. *Langmuir* **2005**, *21*, 2662-2664.
61. Lutz, H.; Jaeger, V.; Berger, R.; Bonn, M.; Pfaendtner, J.; Weidner, T., Biomimetic Growth of Ultrathin Silica Sheets Using Artificial Amphiphilic Peptides. *Advanced Materials Interfaces* **2015**, *2*, n/a-n/a.
62. Nakanishi, K.; Sakiyama, T.; Imamura, K., On the Adsorption of Proteins on Solid Surfaces, a Common but Very Complicated Phenomenon. *J Biosci Bioeng* **2001**, *91*, 233-44.

7.Acronyms

ATR-FTIR	Attenuated total internal reflection Fourier transform infrared spectroscopy
CD	Circular Dichroism
DNA	Deoxyribonucleic acid
DFG	Difference frequency generation
DPPC	Dipalmitoylphosphatidylcholine
FWHM	Full width at half maximum
fs	Femtosecond
GUV	Giant Unilammellar Vesicle
IR	Infrared
IRRAS	Infrared reflection absorption spectroscopy
K	Lysine
L	Leucine
LO	Local oscillator
LUV	Large Unilammellar Vesicle
MD-Simulation	Molecular dynamics simulation
ms	Millisecond
MPGC	Max Planck Graduate Center
NEXAFS	Near edge x-ray absorption fine structure spectroscopy
ps	Picosecond
RNA	Ribonucleic acid
SAM	Self-Assembled Monolayer
SANS	Small Angle Neutron Scattering
SAXS	Small Angle X-ray Scattering
SEIRAS	Surface enhanced infrared absorption spectroscopy
SFG	Sum frequency generation
SHG	Second harmonic generation
SUV	Small Unilammellar Vesicle
t-BLM	Tethered Bilayer Membrane
XPS	X-ray photoelectron spectroscopy

8.Acknowledgements

I want to thank all those who have contributed to this thesis.

Molecular graphics and analyses were performed with the UCSF Chimera package. Chimera is developed by the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco (supported by NIGMS P41-GM103311).

9.List of scientific contributions

9.1 Conference contributions

Talks

L. Schmüser, M. Paven, N. Encinas, D.J. Graham, D.G. Castner, D. Vollmer, H.J. Butt, T. Weidner, *"Liquid Repelling Surfaces Based on Candle Soot are Non-Fouling"*, AVS 62nd International Symposium & Exhibition, San Jose, USA, 2015.

L. Schmüser, N. Hellmann, M. Bonn, T. Weidner, *"Structural changes of proteins at interfaces"*, 79. Jahrestagung der DPG und DPG-Frühjahrstagung, Berlin, Germany, 2015.

Posters

L. Schmüser, M. Schwiering, S. Weisenburger, N. Hellmann, "*Role of cholesterol and lipid headgroup charge for a-toxin binding"*, Annual Meeting of the German Biophysical Society, Göttingen, Germany, 2012.

L. Schmüser, M. Marx, W. Zhang, L. Mammen, D. Vollmer, H. J. Räder, M. Bonn, J. Baio and T. Weidner, ”The *importance of surface chemistry for the super-hydrophobicity of springtail cuticle”,* Annual Meeting of the German Biophysical Society, Lübeck, Germany, 2015.

9.2 Publications

Baio, J. E., Schach, D., Fuchs, A. V., Schmuser, L., Billecke, N., Bubeck, C., Landfester, K., Bonn, M., Bruns, M., Weiss, C. K., & Weidner, T. 2015. Reversible activation of pH-sensitive cell penetrating peptides attached to gold surfaces. Chemical Communications, 51(2): 273-275.

Franz, J., Graham, D. J., Schmuser, L., Baio, J. E., Lelle, M., Peneva, K., Mullen, K., Castner, D. G., Bonn, M., & Weidner, T. 2015. Full membrane spanning self-assembled monolayers as model systems for UHV-based studies of cell-penetrating peptides. Biointerphases, 10(1).

Frisch, H., Fritz, E. C., Stricker, F., Schmuser, L., Spitzer, D., Weidner, T., Ravoo, B. J., & Besenius, P. 2016. Kinetically Controlled Sequential Growth of Surface-Grafted Chiral Supramolecular Copolymers. Angew Chem Int Ed Engl, 55(25): 7242-7246.

Hennig, R., Heidrich, J., Saur, M., Schmuser, L., Roeters, S. J., Hellmann, N., Woutersen, S., Bonn, M., Weidner, T., Markl, J., & Schneider, D. 2015. IM30 triggers membrane fusion in cyanobacteria and chloroplasts. Nature Communications, 6.

Johnson, R., Nawroth, T., Khoshakhlagh, P., Langguth, P., Schmueser, L., Hellmann, N., Decker, H., & Szekely, N. K. 2014. Amphotericin B microparticles 'AmbiShell' from phospholipid and gelatin: Development and investigation by combined DLS and SANS resolves the core-shell structure. European Journal of Lipid Science and Technology, 116(9): 1167-1173.

Khoshakhlagh, P., Johnson, R., Langguth, P., Nawroth, T., Schmueser, L., Hellmann, N., Decker, H., & Szekely, N. K. 2015. Fasted-State Simulated Intestinal Fluid "FaSSIF-C", a Cholesterol Containing Intestinal Model Medium for In Vitro Drug Delivery Development. Journal of Pharmaceutical Sciences, 104(7): 2213-2224.

Khoshakhlagh, P., Johnson, R., Nawroth, T., Langguth, P., Schmueser, L., Hellmann, N., Decker, H., & Szekely, N. K. 2014. Nanoparticle structure development in the gastro-intestinal model fluid FaSSIF(mod6.5) from several phospholipids at various water content relevant for oral drug administration. European Journal of Lipid Science and Technology, 116(9): 1155-1166.

Lutz, H., Jaeger, V., Schmüser, L., S., Bonn, M., Pfaendtner, J., Weidner, T. 2016. Biomineralization to Cause Structural Change in a Silaffin-derived Peptide. Angewandte Chemie, submitted

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. 2016. Ice-nucleating bacteria control the order and dynamics of interfacial water. Sci Adv, 2(4): e1501630.

Schmüser, L., Encinas, N., Paven, M., Graham, D. J., Castner, D. G., Vollmer, D., Butt, H. J., & Weidner, T. 2016. Candle soot-based super-amphiphobic coatings resist protein adsorption. Biointerphases, 11(3): 031007.

Schmüser, L., Lutz, H., Roeters, S., Woutersen, S., Bonn, M., Weidner, T. 2016. Orientation analysis of α-helices at interfaces using heterodyne detected SFG spectroscopy. Journal, submitted

Schmüser, L., Trefz, M., Roeters, S., Beckner, W., Pfaendtner, J., Woutersen, S., Bonn, M., Schneider, D. and Weidner, T. 2016. Refining the X-ray crystal structure of aquaporin with sum frequency generation spectroscopy. Nature Methods, submitted