Structure and Dynamics of Interfacial Peptides and Proteins from Vibrational Sum-Frequency Generation Spectroscopy

Saman Hosseinpour, Steven J. Roeters, Mischa Bonn, Wolfgang Peukert, Sander Woutersen, and Tobias Weidner

†Institute of Particle Technology (LFG), Friedrich-Alexander-University Erlangen-Nürnberg (FAU), 91058 Erlangen, Germany
‡Department of Chemistry, Aarhus University, 8000 Aarhus C, Denmark
§Molecular Spectroscopy Department, Max Planck Institute for Polymer Research, 55128 Mainz, Germany
∥Van’t Hoff Institute for Molecular Sciences, University of Amsterdam, 1098 EP Amsterdam, The Netherlands

ABSTRACT: Proteins at interfaces play important roles in cell biology, immunology, bioengineering, and biomimetic material design. Many biological processes are based on interfacial protein action, ranging from cellular communication to immune responses and the protein-driven mineralization of bone. Despite the importance of interfacial proteins, comparatively little is known about their structure. The standard methods for studying crystalline or solution-phase proteins (X-ray diffraction and NMR spectroscopy) are not well-suited for studying proteins at interfaces, and for these proteins we still lack a corresponding technique that can provide the same level of structural resolution. This is not surprising in view of the challenges involved in probing the structure of proteins within monomolecular films assembled at a very thin interface in situ. Vibrational sum-frequency generation (SFG) spectroscopy has the potential to overcome this challenge and investigate the structure and dynamics of proteins at interfaces at the molecular level with subpicosecond time resolution. While SFG studies were initially limited to simple model peptides, the past decade has seen a dramatic advancement of experimental techniques and data analysis methods that has made it possible to also study interfacial proteins and their folding, binding, orientation, hydration, and dynamics. In this review, we first explain the principles of SFG spectroscopy and the experimental and theoretical methods to measure and analyze protein SFG spectra. Then we give an extensive overview of the interfacial proteins studied to date with SFG. We highlight representative examples to demonstrate recent advances in probing the structure of proteins at the interfaces of liquids, membranes, minerals, and synthetic materials.

CONTENTS
1. Introduction 3421
2. Methods 3422
2.1. Theory 3422
2.1.1. Vibrational Sum-Frequency Generation 3422
2.1.2. SFG Polarization Combinations 3423
2.1.3. SFG Spectroscopy of Amide Modes 3425
2.2. Experimental Methods 3427
2.2.1. Experimental Setups 3427
2.2.2. Sample Geometries Used To Study Peptides and Proteins with SFG 3429
2.3. Calculation of Vibrational Amide SFG Spectra of Peptides and Proteins 3430
2.3.1. Ideal Secondary Structure Elements 3431
2.3.2. Ab Initio-Based Spectral Calculations 3431
2.3.3. Exciton Models 3432
2.3.4. Spectral Calculations Based on Molecular Dynamics Simulations 3434
3. Vibrational SFG Studies of Peptides and Proteins 3434
3.1. Model Systems 3434
3.1.1. Model Peptides 3434
3.1.2. Model Proteins 3439
3.2. Antimicrobial Peptides 3443
3.2.1. WLBU2 3443
3.2.2. Cecropin P1 3443
3.2.3. Surface-Immobilized Antimicrobial Peptides 3443
3.3. Membrane-Associated Proteins 3443
3.4. Fibrils and Fibers 3444
3.4.1. Amyloids 3444
3.4.2. Other Fibrous Proteins 3446
3.5. Biomineralization 3447
3.5.1. Biosilica 3447
3.5.2. Calcium-Based Minerals 3448
3.6. Ice-Binding Proteins 3449
3.6.1. Ice-Nucleating Proteins 3449
3.6.2. Antifreezing Proteins 3449
3.7. Proteins in Surface Bioengineering 3449
3.7.1. Albumin and Fibrinogen at PEGylated Surfaces 3449

Special Issue: Bond Specific Spectroscopy of Peptides and Proteins

Received: June 28, 2019
Published: January 15, 2020
1. INTRODUCTION

Proteins are nature’s machinery: they drive key biological processes, including sensing, immune response, molecular transport, biogenesis of soft and hard tissue, and manipulation of DNA. Many of these processes take place at interfaces, such as membranes, biomaterials, and artificial materials, with the active proteins located at the interface. Interfacial conditions such as concentration, pH, chemical potential, and steric requirements can differ considerably from the bulk, and therefore, the protein structure at surfaces often differs from the structure in solution. At the same time, understanding such interfacial processes will be key in diagnosing diseases, discovering new treatments, delivering pharmaceutical compounds efficiently and with limited side effects, and even in developing novel biomaterials and biosensors. An enormous amount of research has therefore been dedicated to studying protein function in vivo and in vitro, but the complexity of these systems and the lack of specificity of the analytical tools are often limiting factors in these studies. While there has been a strong push over the past decades to determine the structure of proteins with atomic detail using X-ray crystallography, nuclear magnetic resonance (NMR) spectroscopy, and cryo-electron microscopy (cryo-EM), protein structure at interfaces remains poorly understood. Large electronic archives such as the Protein Data Bank (PDB) contain more than 100,000 resolved protein structures. However, very few structures have been solved for surface-bound proteins.

This lack of information about interfacial protein structure is not surprising given the difficulties involved in examining extremely thin (often monomolecular) layers at the very interface between two material phases. Since an interface is a physicochemical environment that is different from the homogeneous bulk phase, it is likely that the folding and molecular structure (and thus the function) of proteins will differ significantly from the crystal or solution-state structures provided by X-ray crystallography, cryo-EM, and NMR spectroscopy. Some interfaces, for instance artificial material surfaces, can even deactivate proteins by denaturation. The presence of interfaces can also lead to folding of a protein into a functional state, as is typical for membrane and biomimetic proteins.

Over the past years, several techniques have emerged that can capture protein structure at surfaces. Neutron reflectometry, solid-state NMR spectroscopy, surface circular dichroism and computer simulations have demonstrated their usefulness as probes of protein structure at hydrated interfaces. Recently, vibrational sum-frequency generation (SFG) spectroscopy, which is inherently surface-sensitive, has evolved into a versatile tool for studying protein structure and dynamics at surfaces and biologically relevant interfaces. Substantial advances have been made in the field over the past years, which have been reviewed previously for certain types of interfaces and SFG techniques. This review aims to provide an overview of the fundamentals and recent advances in studying proteins at all types of interfaces with all relevant forms.
of SFG spectroscopy. The progress in the protein SFG field up to 2009 has been reviewed by Chen in refs 20 and 21, and his group’s valuable contributions to the field up to 2016 have been described in ref 22. Figure 1 provides an overview of the different interfaces and types of proteins for which SFG spectroscopy has been employed.

This review is organized as follows. Section 2 provides a theoretical and experimental description of SFG. In section 3, we discuss the SFG literature on peptides and proteins, which is listed in Table 2. It should be noted that while in this table we tried to provide a complete list of all proteins studied with SFG spectroscopy to date, the more detailed description of studies provided in this section had to be limited to a representative selection. We tried to select studies that illustrate the potential and application of various SFG methods and the recent advances in different fields of interfacial molecular biology. We conclude the review with a summary and outlook in section 4.

2. METHODS

This section provides background information to readers who are new to the field by discussing the theory of vibrational SFG in general (section 2.1.1) and the polarization combinations (section 2.1.2) and normal modes (section 2.1.3) that are most frequently used in vibrational SFG spectroscopy of peptides and proteins. After that, we discuss the experimental implementation of emerging and well-established forms of SFG spectroscopy (section 2.2.1) and the analysis of protein SFG spectra through spectral calculations (section 2.3).

2.1. Theory

2.1.1. Vibrational Sum-Frequency Generation. Vibrational SFG is an interface-sensitive second-order nonlinear spectroscopic technique in which two pulsed laser beams with visible ($\omega_{\text{VIS}}$) and infrared ($\omega_{\text{IR}}$) frequencies are overlapped in the sample both in time and space, generating a third beam that has the sum frequency of the two incoming beams ($\omega_{\text{SF}} = \omega_{\text{VIS}} + \omega_{\text{IR}}$). Although the abbreviation “SFG” can be used to describe all types of sum-frequency generation (i.e., also those that do not involve an IR beam that excites vibrational modes, such as electronic SFG$^{24-26}$), throughout this review we will use “SFG” to refer to vibrational SFG. According to the transition selection rule of SFG, only vibrational modes that are both IR- and Raman-active give rise to a vibrational SFG signal.$^{30-32}$ The intensity of the SFG field ($I_{\text{SFG}}$) depends on the intensity of the incoming fields ($I_{\text{IR}}$ and $I_{\text{VIS}}$) and is strongly enhanced when the IR beam is in resonance with an IR- and Raman-active vibration, which is quantitatively described by the second-order susceptibility $\chi^{(2)}$:

$$I_{\text{SFG}} \propto |\chi^{(2)}|^2 I_{\text{VIS}} I_{\text{IR}}$$

(1)

where $\chi^{(2)}$ is the coherent sum of a resonant part and a nonresonant part:

$$\chi^{(2)} = \chi_{\text{NR}}^{(2)} + \sum_q \frac{\chi_{\text{RI}}^{(2)q}}{\omega_q - \omega_{\text{IR}} + i\Gamma_q}$$

(2)

in which $\chi_{\text{RI}}^{(2)q} = A_{\text{NR}} e^{i\varphi_{\text{RI}}} e^{i\varphi_{\text{IR}}} A_{\text{NR}} e^{i\varphi_{\text{IR}}} \varphi_{\text{IR}}$ is the nonresonant susceptibility, with a nonresonant amplitude $A_{\text{NR}}$ and phase $\varphi_{\text{RI}}$, $\varphi_{\text{IR}}$ and $\Gamma_q$ are the resonant susceptibility, frequency, and Lorentzian width of the $q$th vibrational mode, respectively. To reveal the vibrational signatures of the interfacial molecules, SFG spectra are commonly plotted as a function of the incident IR wavenumber (in cm$^{-1}$). Unlike for metal surfaces, where the nonresonant susceptibility is very large, $\chi_{\text{RI}}^{(2)}$ is typically very small in most biologically relevant experiments.

The interface sensitivity of SFG stems from the fact that under the dipole approximation (i.e., only the electric-dipole coupling with the light is considered), sum-frequency generation is forbidden (i.e., $\chi_{\text{RI}}^{(2)q} = 0$) in centrosymmetric media. Therefore, an SFG signal cannot be generated in most bulk materials (gases, bulk liquids, and crystals with centrosymmetry). However, when the centrosymmetry is broken, as is always the case at interfaces, an SFG signal can be generated, which thus gives molecular information without interference from a bulk signal (provided that the bulk material is isotropic). This selection rule follows from the fact that the measured signal is a collective response of the molecules in the region where the IR and the visible beams overlap and thus is a function of the average molecular orientation. The nonlinear susceptibility can be described as:
\[ \chi^{(2),4}_{jk} = N \sum_{i,j,k=1}^{3} \langle (\hat{X} \cdot \hat{x})(\hat{Y} \cdot \hat{y})(\hat{Z} \cdot \hat{z}) \rangle \beta^{(2),q}_{jk} \]

where \( I, J, K = X, Y, Z \) for \( i, j, k = x, y, z \); \( \hat{X}, \hat{Y}, \hat{Z} \) and \( \hat{x}, \hat{y}, \hat{z} \) are the unit vectors in the lab frame and molecular frame, respectively; \( N \) is the number of molecules in the lab frame and molecular frame, respectively; \( \langle \cdots \rangle \) indicates the averaging over the molecular orientation distribution; and \( \beta^{(2),q}_{jk} \) is the molecular hyperpolarizability tensor in the molecular frame.

The resonance between interfacial vibrational modes and the frequency of the IR beam employed in vibrational SFG provides molecular-level information on the interface through the molecular hyperpolarizability \( \beta^{(2),q}_{jk} \), which is a rank-3 tensor whose 27 elements (given by the molecular-frame Cartesian coordinates \( i, j, k = x, y, z \) for the polarizations of each of three field interactions) are direct products of the underlying IR and Raman processes:

\[ \beta^{(2),q}_{jk} = \mu_{\parallel}^{q} \alpha_{q}^{j} \]

where \( \mu_{\parallel}^{q} \) and \( \alpha_{q}^{j} \) are the transition dipole moment and Raman polarizability of mode \( q \), respectively. Figure 2 illustrates the SFG process and provides example spectra of two types of SFG-active modes.

The two selection rules of SFG make it an ideal tool for studying peptides and proteins at interfaces, as demonstrated by the large number of groundbreaking biophysical SFG studies (see section 3). More detailed theoretical treatments of the physical principles behind SFG spectroscopy can be found elsewhere.23,35–37

2.1.2. SFG Polarization Combinations. In general, SFG spectra are recorded with the three beams in linear S or P polarization (see Figure 3), but mixed polarization settings, often termed \( m \)-polarized, are also used.38–42 There are eight (\( \approx 2^3 \)) possible SFG polarization combinations with pure S or P polarizations. The intensity of the SSS polarization combination is zero for azimuthally isotropic interfaces because integrating over the Euler angle \( \phi \) renders \( \chi^{(2),4}_{ijk} \), the macroscopic susceptibility element that it probes,43 zero. Of the remaining seven polarization combinations, four (PPP, SSP, SPS, and PSS) probe achiral macroscopic susceptibility elements, and three (PSP, SSP, and PPS) probe chiral macroscopic susceptibility elements. The different polarization combinations are often measured sequentially, although they can also be recorded simultaneously.34,43

2.1.2.1. Achiral Polarization Combinations. Until 2004, only achiral polarizations were used in the study of proteins with SFG, and also since then achiral intensity ratios and spectral lineshapes are often used to determine the intensity and orientation of protein vibrational modes and thereby the protein orientation and structure at interfaces. The details of protein orientation analysis based on intensity ratios obtained from these polarization combinations are discussed in section 2.3.

For vibrational sum-frequency generation, the achiral polarization combinations are related to the achiral macroscopic susceptibilities as follows (see Figure 3 for the definitions of the angles \( \rho_{\parallel} \)).

\[ \chi^{(2),q}_{ijk} = L_{yy}(\omega_{\text{VIS}}) \cdot L_{yz}(\omega_{\text{VIS}}) \cdot L_{zz}(\omega_{\text{IR}}) \cdot \sin \rho_{\parallel} \chi^{(2),q}_{YZZ} \]

\[ \chi^{(2),q}_{ijk} = L_{xx}(\omega_{\text{VIS}}) \cdot L_{zz}(\omega_{\text{VIS}}) \cdot L_{zz}(\omega_{\text{IR}}) \cdot \sin \rho_{\parallel} \chi^{(2),q}_{XXZ} \]

\[ \chi^{(2),q}_{ijk} = L_{zz}(\omega_{\text{VIS}}) \cdot L_{yz}(\omega_{\text{VIS}}) \cdot L_{zz}(\omega_{\text{IR}}) \cdot \sin \rho_{\parallel} \chi^{(2),q}_{YZZ} \]

\[ \chi^{(2),q}_{ijk} = L_{xx}(\omega_{\text{VIS}}) \cdot L_{xx}(\omega_{\text{VIS}}) \cdot L_{zz}(\omega_{\text{IR}}) \cdot \cos \rho_{\parallel} \cos \rho_{\parallel} \sin \rho_{\parallel} \chi^{(2),q}_{XZZ} \]

\[ \chi^{(2),q}_{ijk} = L_{zz}(\omega_{\text{VIS}}) \cdot L_{xx}(\omega_{\text{VIS}}) \cdot L_{zz}(\omega_{\text{IR}}) \cdot \cos \rho_{\parallel} \sin \rho_{\parallel} \sin \rho_{\parallel} \chi^{(2),q}_{XZZ} \]

\[ \chi^{(2),q}_{ijk} = L_{xx}(\omega_{\text{VIS}}) \cdot L_{xx}(\omega_{\text{VIS}}) \cdot L_{zz}(\omega_{\text{IR}}) \cdot \sin \rho_{\parallel} \sin \rho_{\parallel} \sin \rho_{\parallel} \chi^{(2),q}_{XZZ} \]

in which \( \omega_{\text{VIS}}, \omega_{\text{VIS}}, \omega_{\text{IR}} \) and \( \rho_{\parallel} \) are the sum, visible, and IR frequencies, respectively; \( \rho_{\parallel} \) and \( \rho_{\parallel} \) are the angles of incidence of the visible and IR beams, respectively; \( \rho_{\parallel} \) is the angle at which the sum-frequency beam is generated (see Figure 3), given by

\[ L_{XX}(\alpha) = \frac{2n_{1}(\alpha) \cdot \cos \gamma_{j}}{n_{1}(\alpha) \cdot \cos \gamma_{j} + n_{2}(\alpha) \cdot \cos \rho_{j}} \]

\[ L_{YY}(\alpha) = \frac{2n_{1}(\alpha) \cdot \cos \rho_{j}}{n_{1}(\alpha) \cdot \cos \rho_{j} + n_{2}(\alpha) \cdot \cos \gamma_{j}} \]

\[ L_{ZZ}(\alpha) = \frac{2n_{1}(\alpha) \cdot \cos \rho_{j}}{n_{1}(\alpha) \cdot \cos \gamma_{j} + n_{2}(\alpha) \cdot \cos \rho_{j}} \left[ \frac{n_{1}(\alpha) \cdot \cos \rho_{j}}{n_{1}(\alpha) \cdot \cos \gamma_{j} + n_{2}(\alpha) \cdot \cos \rho_{j}} \right]^{2} \]

in which \( n_{1}, n_{2}, \) and \( n' \) are the indices of refraction of the dispersive media and \( \gamma_{j} \) is the refracted angle, given by

\[ \sin \gamma_{j} = \frac{n_{1}}{n_{2}} \sin \rho_{j}, \text{ and } j = \text{SF, VIS, or IR}. \]

In the amide-I region, the achiral signals from \( \alpha \)-helical and disordered structures often overlap strongly, and discerning these is crucial for the interpretation of SFG experiments. The details of protein orientation analysis based on intensity ratios obtained from these polarization combinations are discussed in section 2.3.

2.1.2.2. Chiral Polarization Combinations. To overcome the difficulties mentioned above, in recent years chiral SFG spectroscopy has gained popularity for determination of the conformations of peptides and proteins at interfaces. The discussion below is based on previous treatments of the theory of chiral SFG,18,32,47,48 and here only the basic principles and
The intensity of the chiral SFG signals is probably dependent on the presence of macroscopic chiral structures, as was shown for LK-β, whose antiparallel β-sheets formed at pH 7 exhibit a strong chiral amide-I signal, while its similarly structured but isolated β-strand structures formed at pH 2 do not generate a chiral SFG signal. Besides in the amide-I and N–H stretching regions, chiral SFG was also applied for the detection of C–H stretching modes to monitor the self-assembly of amphiphilic LK-β peptide into chiral antiparallel β-sheets.

As mentioned earlier, combining chiral SFG with other measurements, such as achiral and heterodyned SFG, surface tension, and X-ray studies, can give a more complete picture of protein structure and dynamics. Also, recent technical improvements like electronically enhanced chiral SFG have made the collection of chiral SFG spectra more accurate.

The accuracy of chiral SFG spectroscopy in determining the structure of proteins is highly dependent on the proper setup alignment and “polarization purity” of the incoming and outgoing beams. For instance, leakage of a strong achiral signal through a polarizer might be interpreted as a weak chiral signal. The twin polarization angle (TPA) approach for chiral SFG has proven to be a very accurate method for conducting chiral SFG measurement and analysis. In the TPA approach, the polarization angles of the incoming visible and infrared beams are varied simultaneously while the polarization of the incoming infrared beam is kept fixed (either at S or P). This allows accurate quantification of the chiral and achiral signal contributions. However, TPA is technically complex, and the amount of data that needs to be collected increases dramatically compared with single-polarization SFG measurements. McDermott and Petersen recently presented a simple and cheap yet robust self-referencing method for detection of chiral SFG signals that can be implemented in most SFG spectrometers by introducing a birefringent calcite beam displacer in the detection path and recording the interference between the chiral and achiral signals by selecting mixed polarizations. In a benchmarking study, the chiral and achiral C–H stretch spectra of chiral samples (adsorbed ubiquitin and drop-casted collagen films) deposited on achiral layers (a self-assembled monolayer of octadecyltrichlorosilane (OTS) and spin-coated poly(methyl methacrylate) (PMMA)) were recorded simultaneously in a chirally resolved manner, demonstrating the high discriminative power of this methodology. Although the C–H stretching region used in their study is insensitive to the secondary structure of the protein, it is straightforward to apply this method to the amide regions. Using the same method, McDermott et al. subsequently observed the O–H stretch signal from a chiral superstructure of water (a chiral spine of hydration) surrounding DNA molecules under physiological conditions. The authors found that this chiral superstructure of water imprinted by the DNA strands is not sequence-specific, as a chiral water signal was observed for different DNA duplexes. The water molecules within the chiral superstructure are more strongly hydrogen-bonded than the achiral solvation shell of DNA, as the chiral water O–H stretch signal is red-shifted relative to the achiral water signal. This phenomenon—the generation of strong second-order nonlinear signals by achiral molecules arranged in a chiral macroscopic structure—was reviewed by Simpson in 2004. Recently, Perets and Yan demonstrated that LKβ peptides at the air–water interface also give rise to a chiral hydration shell that can be

As mentioned earlier, combining chiral SFG with other measurements, such as achiral and heterodyned SFG, surface tension, and X-ray studies, can give a more complete picture of protein structure and dynamics. Also, recent technical improvements like electronically enhanced chiral SFG have made the collection of chiral SFG spectra more accurate.

The accuracy of chiral SFG spectroscopy in determining the structure of proteins is highly dependent on the proper setup alignment and “polarization purity” of the incoming and outgoing beams. For instance, leakage of a strong achiral signal through a polarizer might be interpreted as a weak chiral signal. The twin polarization angle (TPA) approach for chiral SFG has proven to be a very accurate method for conducting chiral SFG measurement and analysis. In the TPA approach, the polarization angles of the incoming visible and infrared beams are varied simultaneously while the polarization of the incoming infrared beam is kept fixed (either at S or P). This allows accurate quantification of the chiral and achiral signal contributions. However, TPA is technically complex, and the amount of data that needs to be collected increases dramatically compared with single-polarization SFG measurements. McDermott and Petersen recently presented a simple and cheap yet robust self-referencing method for detection of chiral SFG signals that can be implemented in most SFG spectrometers by introducing a birefringent calcite beam displacer in the detection path and recording the interference between the chiral and achiral signals by selecting mixed polarizations. In a benchmarking study, the chiral and achiral C–H stretch spectra of chiral samples (adsorbed ubiquitin and drop-casted collagen films) deposited on achiral layers (a self-assembled monolayer of octadecyltrichlorosilane (OTS) and spin-coated poly(methyl methacrylate) (PMMA)) were recorded simultaneously in a chirally resolved manner, demonstrating the high discriminative power of this methodology. Although the C–H stretching region used in their study is insensitive to the secondary structure of the protein, it is straightforward to apply this method to the amide regions. Using the same method, McDermott et al. subsequently observed the O–H stretch signal from a chiral superstructure of water (a chiral spine of hydration) surrounding DNA molecules under physiological conditions. The authors found that this chiral superstructure of water imprinted by the DNA strands is not sequence-specific, as a chiral water signal was observed for different DNA duplexes. The water molecules within the chiral superstructure are more strongly hydrogen-bonded than the achiral solvation shell of DNA, as the chiral water O–H stretch signal is red-shifted relative to the achiral water signal. This phenomenon—the generation of strong second-order nonlinear signals by achiral molecules arranged in a chiral macroscopic structure—was reviewed by Simpson in 2004. Recently, Perets and Yan demonstrated that LKβ peptides at the air–water interface also give rise to a chiral hydration shell that can be

As mentioned earlier, combining chiral SFG with other measurements, such as achiral and heterodyned SFG, surface tension, and X-ray studies, can give a more complete picture of protein structure and dynamics. Also, recent technical improvements like electronically enhanced chiral SFG have made the collection of chiral SFG spectra more accurate.

The accuracy of chiral SFG spectroscopy in determining the structure of proteins is highly dependent on the proper setup alignment and “polarization purity” of the incoming and outgoing beams. For instance, leakage of a strong achiral signal through a polarizer might be interpreted as a weak chiral signal. The twin polarization angle (TPA) approach for chiral SFG has proven to be a very accurate method for conducting chiral SFG measurement and analysis. In the TPA approach, the polarization angles of the incoming visible and infrared beams are varied simultaneously while the polarization of the incoming infrared beam is kept fixed (either at S or P). This allows accurate quantification of the chiral and achiral signal contributions. However, TPA is technically complex, and the amount of data that needs to be collected increases dramatically compared with single-polarization SFG measurements. McDermott and Petersen recently presented a simple and cheap yet robust self-referencing method for detection of chiral SFG signals that can be implemented in most SFG spectrometers by introducing a birefringent calcite beam displacer in the detection path and recording the interference between the chiral and achiral signals by selecting mixed polarizations. In a benchmarking study, the chiral and achiral C–H stretch spectra of chiral samples (adsorbed ubiquitin and drop-casted collagen films) deposited on achiral layers (a self-assembled monolayer of octadecyltrichlorosilane (OTS) and spin-coated poly(methyl methacrylate) (PMMA)) were recorded simultaneously in a chirally resolved manner, demonstrating the high discriminative power of this methodology. Although the C–H stretching region used in their study is insensitive to the secondary structure of the protein, it is straightforward to apply this method to the amide regions. Using the same method, McDermott et al. subsequently observed the O–H stretch signal from a chiral superstructure of water (a chiral spine of hydration) surrounding DNA molecules under physiological conditions. The authors found that this chiral superstructure of water imprinted by the DNA strands is not sequence-specific, as a chiral water signal was observed for different DNA duplexes. The water molecules within the chiral superstructure are more strongly hydrogen-bonded than the achiral solvation shell of DNA, as the chiral water O–H stretch signal is red-shifted relative to the achiral water signal. This phenomenon—the generation of strong second-order nonlinear signals by achiral molecules arranged in a chiral macroscopic structure—was reviewed by Simpson in 2004. Recently, Perets and Yan demonstrated that LKβ peptides at the air–water interface also give rise to a chiral hydration shell that can be
detected with chiral SFG spectroscopy. Additional recent developments in chiral SFG spectroscopy, such as phase-resolved chiral SFG measurements and the application of high-resolution broad-band setups, are discussed in section 2.2.1.2.

2.1.3. SFG Spectroscopy of Amide Modes. In the study of peptides and proteins, the amide vibrations are particularly interesting since the amide group is omnipresent in these molecules as peptide bonds between the amino acids. Especially the amide-I band (\(\sim 1600\)−\(1700\) cm\(^{-1}\)), which arises mainly from the C==O stretching vibration with minor contributions from the out-of-phase CN stretching vibration, CCN deformation, and NH in-plane bend\(^59\) (see Figure 5), is very sensitive to the protein structure because of its large transition dipole moment, which leads to strong coupling between the local amide-I modes. As a result, amide-I spectra are sensitive reporters of the structure and environment of proteins.\(^34,60−62\)

Other amide modes that can provide similar information, albeit with smaller transition dipole moments, are the amide-II (\(\sim 1550\) cm\(^{-1}\)), -III (\(\sim 1200\)−\(1400\) cm\(^{-1}\)), and -A (\(\sim 3300\) cm\(^{-1}\)) modes.\(^59,63\)

Figure 6 demonstrates the potential of amide-I SFG as a probe for protein structure at interfaces: while the calculated amide-I IR and Raman spectra of lysozyme in two different conformations at the air−water interface are very similar, the calculated SFG spectra are very different. The two structures are snapshots taken at different time points of a molecular dynamics (MD) simulation. The linear IR and Raman spectra differ very little for the two different states, probably because the two structures contain similar amounts of the different secondary structure elements. However, SFG is a coherent technique, and

Figure 4. (A) Schematic structures of seven proteins that have been investigated with chiral SFG spectroscopy by the Yan group, with the associated chiral (B) amide-I and (C) N−H stretch spectra. It should be noted that the absence of a chiral N−H signal for hIAPP was later explained by orientational effects at the DPPG surface and is not a general feature of parallel β-sheets. Reproduced from ref 17. Copyright 2015 American Chemical Society.

Figure 5. (a) Atomic motion in four different amide local modes, or normal modes,\(^64\) and (b–d) amide-I eigenmodes of proteins calculated with an excitonic amide-I Hamiltonian (see section 2.3.3) and depicted using the software package Chimera.\(^65\) The strongest eigenmode of each of the structures is shown: the A mode of an ideal α-helix created with Chimera, a B2-type mode of an antiparallel β-sheet (based on a fragment of PDB entry 2OMQ) and a B-type mode of a parallel β-sheet (based on a fragment of PDB entry 1DAB).\(^66\) The gray arrows indicate the local modes, the blue/white/red spheres indicate their phases within the depicted normal mode, and the black arrows indicate the normal mode transition dipole moments.\(^67\)
therefore, sum-frequency photons from different parts of the protein interfere with each other in complex ways depending on the relative orientation and frequency of the associated normal modes. This leads to the rich spectra shown in Figure 6 for different polarization combinations. The spectra differ appreciably for the two different states of lysozyme at the water surface, even though their secondary structure contents are quite similar. Amide spectroscopy thus provides insights into the folding behavior and conformation of proteins, which determine their biological function.68 NMR,69 circular dichroism (CD),70 fluorescence,71 and two-dimensional infrared (2D-IR) spectroscopy72,73 are typically used to determine the secondary structure of proteins in bulk solutions. However, the information obtained through these methods generally cannot be unambiguously extrapolated to protein structures at different surfaces and interfaces. The different environments at interfaces compared with bulk phases can lead to drastic changes in protein conformation.

Despite the relatively strong SFG signal in the amide-I region, the achiral amide-I spectra of α-helical and random coil structures of proteins overlap strongly, making it difficult to distinguish them. As an alternative (or in addition) to probing the chiral amide-I spectrum (see section 2.1.2.2), the amide-III region can be probed to unambiguously identify the secondary structure of proteins because the characteristic spectral features of α-helices, β-sheets, β-turns, and random coils are well-separated in this spectral region.74,75 However, the SFG signal intensity in the amide-III region is relatively weak, and technical improvements are often required to obtain SFG signals with a reasonable signal-to-noise ratio. These improvements include increasing the sensitivity of the detector, employing thick difference-frequency generation crystals to generate IR pulses with more energy, and using the near-total internal reflection geometry (see section 2.2.2.2).76,77

Other spectral regions used to study peptides and proteins with SFG include the C−H and N−H stretching region (2800–3500 cm−1) to observe protein–surface interactions and the O−H stretching region (3100–3750 cm−1) to study the hydration of protein layers.78–80

Figure 6. Structure sensitivity of SFG amide-I spectra. Shown are calculated IR, Raman, and SFG spectra for lysozyme at the air–water interface for two different snapshots of an MD simulation, normalized to the maximum intensity. While the IR and Raman spectra are very similar for the two structural states, the SFG spectra are very different.

Figure 7. Schematic representation of a femtosecond homodyne SFG setup in which an etalon is used to narrow the spectral width of the visible beam and in which the light is dispersed with a Czerny–Turner spectrograph.67
2.2. Experimental Methods

2.2.1. Experimental Setups. 2.2.1.1. Homodyne Sum-Frequency Generation. The most commonly applied method to record SFG spectra is homodyne-detected (also called “conventional” or “intensity”) SFG spectroscopy. An example of such a setup is given in Figure 7. The pico- or femtosecond output of a regenerative amplifier is split into two beams: (i) the first (“visible”) arm, which is spectrally narrowed in broad-band femtosecond systems to achieve good spectral resolution, and (ii) the second arm, which pumps an optical parametric amplifier (OPA) and difference-frequency mixing stage to generate tunable mid-IR pulses. Each beam is passed through a half-wave plate and a polarizer and focused (with a focal length and spot size typically on the order of 10 cm and 500 μm, respectively) so as to impinge on the sample overlapping in time (using a delay stage) and space. The sum-frequency light emitted by the sample is collimated, and the remaining visible light is filtered out. In the case of a picosecond setup, the SF light is detected with a photomultiplier/gated integrator system, and in the case of a femtosecond setup, the beam is dispersed in a spectrograph and imaged using a charge-coupled device (CCD) camera. By rotation of the half-wave plates, the various polarization combinations are measured. The spectra are first background-corrected (using a spectrum recorded in absence of the IR beam) and subsequently normalized with a reference SFG spectrum (typically measured from a gold sample or a z-cut quartz crystal).

2.2.1.2. Phase-Resolved Sum-Frequency Generation. As can be seen from eq 1, in homodyne SFG measurements, the absolute square of the effective second-order nonlinear susceptibility, \( \chi_{\text{eff}}^{(2)} \), is detected. Therefore, the sign of the \( \chi_{\text{eff}}^{(2)} \), which carries information about the absolute polar orientation of the molecules, is lost in homodyne SFG measurements. Moreover, interference between different overlapping vibrational modes with opposite phases or interference between the resonant and nonresonant parts of the SFG signal complicates the interpretation of homodyne SFG spectra. For instance, it has been shown that nonunique sets of fitting parameters can be used to describe a conventional SFG spectrum. This nonuniqueness of fitting parameters is often caused by the interference of the resonant and nonresonant signal contributions to the SFG spectrum. In analyzing the N–H stretching vibration of proteins using conventional SFG measurements, special attention should be paid to the interference between the N–H signal of the protein and the broad OH stretching vibration of water molecules. This interference can easily lead to misinterpretation of the spectra. Similar interferences also typically occur between the C–H vibration modes of proteins or surfactants and the low-frequency tail of the stretching modes of water molecules, which can significantly alter the SFG spectral shape. The introduction of interferometric techniques, such as heterodyne-detected SFG (also known as “phase-sensitive” or “phase-resolved” SFG), enabled direct measurement of the complex \( \chi_{\text{eff}}^{(2)} \) spectrum, where the sign of the absorptive component (\( \text{Im}[\chi_{\text{eff}}^{(2)}] \)) reflects the net orientation of the transition dipole moment of the probed normal mode. Phase-resolved measurements also make it easier to unambiguously interpret SFG data. For example, heterodyned SFG measurements on hydrophobins (also see section 3.7.4) at the air–water interface showed that the strong pH dependence of the homodyne-detected N–H stretching mode signal was the result of a change in the interference with reorienting OH stretching modes of water, while the phase of the N–H stretch mode did not change as a function of pH. Therefore, phase-resolved SFG measurements have become increasingly popular to study the absolute orientation of molecules at different interfaces. The majority of phase-resolved SFG measurements are performed at the air–liquid interface. However, a few examples exist where the interface between a solid and a liquid phase has been probed using phase-resolved SFG. In what follows, we discuss the different approaches and experimental geometries for phase-resolved SFG measurements and their theoretical description.

Since the introduction of phase-resolved SFG measurements, different approaches and experimental geometries have been adopted that are based on transmission or reflection for narrow-band picosecond scanning infrared, femtosecond broad-band infrared, or high-resolution broad-band infrared collinear or noncollinear setups. Schematics of these different approaches are presented in Figure 8. The underlying theory behind phase-resolved measurements has been reviewed by Shen. A common aspect of all of these approaches is that the SFG signal from the sample (SFGsample) interferes with the SFG signal from a local oscillator (SFGLO) that is generated before or after the sample. Introducing a delay between the two SFG signals through a delay plate or a delay stage results in interference between the two signals at the grating. When SFGLO is generated before the sample (Figure 8d), the electric field strengths at the local oscillator (ELO) and the sample (Esample) can be written as:

\[
E_{\text{LO}} \propto \chi_{\text{LO}}^{(2)} E_{\text{VIS}} E_{\text{IR}}
\]

\[
E_{\text{sample}} \propto \chi_{\text{sample}}^{(2)} E_{\text{VIS}} E_{\text{IR}} R_{\text{LO,ωVIS}} r_{\text{LO,ωIR}}
\]

with \( r_{\text{LO}} \) the reflection coefficient of the local oscillator. At a time delay \( \Delta t \), the detected field strength (Edet) in the frequency domain is:

\[
E_{\text{det}}(ω) = E_{\text{sample}}(ω) + r_{\text{sample}} E_{\text{LO}}(ω)e^{-iωΔt}
\]

where \( r_{\text{sample}} \) is the reflection coefficient of the sample surface. Therefore, the interfering signal at the detector can be written as:

\[
E_{\text{det}}(ω) = \left| E_{\text{sample}}(ω) \right|^2 + \left| r_{\text{sample}} E_{\text{LO}}(ω) \right|^2 + \left| r_{\text{sample}} E_{\text{LO}}(ω) e^{-iωΔt} \right|^2 + \left| E_{\text{sample}}(ω) e^{-iωΔt} \right|^2
\]

The last two terms in eq 11 can be selected by Fourier transformation, as explained in detail in ref 93. With a proper correction by the Fresnel factors and normalization to the frequency dependence of the broad-band IR spectrum, \( \text{Im}[\chi_{\text{eff}}^{(2)}] \) and \( \text{Re}[\chi_{\text{eff}}^{(2)}] \) can be extracted.

2.2.1.3. Time-Resolved Sum-Frequency Generation. Time-resolved SFG measurements can directly resolve the subpicosecond dynamics of the proteins residing at a surface or interface. Performing time-resolved SFG measurements requires an additional laser beam, often called the pump pulse, that spatially overlaps with the SFG probe beams (i.e., the IR and visible beams). By changing the time delay between the pump and probe pulses, transient spectra can be recorded. An example of a pulse sequence scheme for time-resolved SFG is presented in Figure 9.

2.2.1.4. Two-Dimensional Sum-Frequency Generation. Two-dimensional surface-specific vibrational SFG spectroscopy
(2D-SFG) was first introduced in 2008 to study molecular monolayers at the air–water interface.96,97 Multiple experimental setups to record 2D-SFG spectra are conceivable,98 but to date only a setup in which a pulse shaper is used to create the IR pump pulse has been applied to peptides99 and proteins,100 so we will discuss only that methodology in detail here. Other experimental layouts that could be used to do 2D-SFG studies of proteins could involve a frequency-scanned picosecond beam, a Fabry–Perot interferometer–(etalon)-narrowed femtosecond beam, or a Fourier transform Michelson interferometer-based setup to generate the IR pump axis.

Ghosh et al. split the mid-infrared (mIR) output of their OPA and difference-frequency generator (DFG) into two arms, of which one is sent into a pulse shaper that interferes with itself with a delay $t_1$ (see Figures 10B and 11), thus creating a pump frequency axis that is reconstructed through Fourier transformation. This beam is then overlapped with the earlier split-off mIR and visible SFG beam pair at the sample. The authors heterodyne the 2D-SFG signal with a local oscillator (in the case of refs 99 and 100 using a gold subphase). Although Backus et al.101 showed that the homodyne and heterodyne 2D-SFG signals of aqueous interfaces contain essentially the same information, Ghosh et al. reported that the homodyne $\chi^{(4)}$ signal is affected by phase distortions of the 2D-SFG spectrum. Fourier filtering is used to remove the local oscillator signal, and phase cycling through the phase of the pump pulses with the pulse shaper (see the Figure 11 inset) allows the $E^{(4)}E_{LO}^{(4)}$ and $E^{(2)}E^{(4)}$ signals to be extracted from the other cross-terms. Because $E_{LO}^{(4)} \gg E^{(2)}$, the resulting signal can be directly related to the desired
fourth-order electric field by isolating the $E_4^{(4)}E_{LO}$ term, as the nonresonant $E_{LO}$ of gold is temporally short and has a well-defined phase with respect to the resonant signal.

2.2.1.5. Scattering Sum-Frequency Generation Spectroscopy. In cases where studying interfacial properties of proteins and their ordered structure in solution is desired, SFG measurements may be performed in a scattering mode. Sum-frequency scattering (SFS) was first developed by Roke et al.\textsuperscript{102,103} to study interfacial properties of submicron particles in suspension and has been used as a complementary technique to dynamic light scattering (DLS), second harmonic generation (SHG), and SFG.\textsuperscript{104} There are several interesting examples where SFS has been employed in studying cationic vesicles\textsuperscript{112} and collagen fibers in solution\textsuperscript{113} (see section 3.9.3), but it seems that the potential of SFS for protein studies has not yet been fully realized.

2.2.1.6. Sum-Frequency Generation Microscopy. SFG studies are typically based on a wide-field excitation, where an area of several hundred square micrometers of the sample is illuminated by the incoming beams, and the generated signal is collected and directed to the detector. Although detailed information can be obtained using wide-field excitation SFG methods, they fail to provide any information on the spatial heterogeneity of the system. However, even well-prepared model surfaces may contain angstrom to millimeter length scale inhomogeneities. Therefore, the realization of a surface-sensitive spectroscopic method with lateral resolution on this length scale is desirable, and multiple attempts have been made to extend the application of SFG spectroscopy to microscopy and imaging using either a point or wide-field illumination configuration.\textsuperscript{114–119} There are multiple methods for SFG microscopy. For instance, point-illumination polarization-sensitive SFG imaging, as described in ref 120, is based on a collinear nonlinear optical microscope (see Figure 12). In this setup, the output beams of a picosecond optical parametric oscillator are combined collinearly on a dichroic mirror and focused onto a microscope frame with a 0.65 numerical aperture (NA) reflective objective. The generated SFG signal is captured in transmission by a refractive condenser lens and detected by a photomultiplier tube. To generate an image, the sample is mounted on a two-dimensional piezo stage and raster-scanned relative to the focal spot of the incoming beam. Using an appropriate bandpass filter allows detection of the SHG signal with the same setup. In this system, the spectral resolution is limited by the spectral width of the pulses ($~6$ cm$^{-1}$). Inoue et al.\textsuperscript{117} used objective lenses with large NA, which enabled imaging of a single cell with lateral resolution significantly smaller than the IR diffraction limit. With a moderate focus of the visible and IR beams on a spot size of $~100$ μm, the authors were able to obtain the whole image without scanning. With a decade of developments, the Baldelli group\textsuperscript{121,122} combined SFG imaging with compressive sensing techniques, which improved the pixel density beyond the resolution achieved by the raster scanning approach, but their approach has not yet been used to study biologically relevant systems.

2.2.2. Sample Geometries Used to Study Peptides and Proteins with SFG. 2.2.2.1. Air–Water/Air–Lipid Interface. A variety of SFG studies have focused on air–water or air–lipid interfaces, which are appealing because of their relative simplicity. The air–water interface is a model system for hydrophobic surfaces and very straightforward to prepare. In addition, surface tension measurements provide a simple way of tracking changes in interfacial binding in real time during the SFG measurements. Lipid monolayers have been employed as mimics of lipid bilayers to follow binding to phospholipids. The method is of limited use in scenarios where proteins penetrate into the hydrophobic core of the lipid layer and when the propensity toward the air–water interface is greater for the protein under study than for the lipid molecules. In such cases, the protein structure can be determined by hydrophobic contacts with the air–water interface. A challenge when probing proteins at lipid monolayers can be beam damage and local heating of the samples. Backus et al. showed that in the case of low-density lipid monolayers, local heating can complicate the experiment by convective flow within the interfacial water region.\textsuperscript{123} This problem can be solved by rotating the sample cell. Franz et al. developed a Langmuir trough that can be
scanned in the $xy$ plane to avoid local heating and beam damage for experiments with constant area and pressure (Figure 13).\textsuperscript{124}

2.2.2.2. Near-Total Internal Reflection Setup. Passing the beams through a high refractive index (e.g., $\text{CaF}_2$) prism in a near total-internal reflection configuration on top of the sample interface\textsuperscript{76,77} has multiple advantages for SFG studies on proteins and peptides: (i) the signal intensity can go up by as much as a factor 1000 as a result of increased Fresnel factors (see eqs 1 and 6); (ii) the PPP polarization combination is mainly composed of the lab-frame ZZZ component; (iii) supported lipid bilayers can be adsorbed onto the prism, which can greatly enhance the biological relevance of peptide and protein experiments (especially those performed with polypeptides that insert into lipid bilayers); (iv) spin-coating techniques allow the study of additional types of surface chemistry\textsuperscript{125}.

In 2010, Verreault et al. realized a setup using a hemicylinder with a curved top, by which a large range of incoming beam angles can be employed.\textsuperscript{126} In a follow-up paper, the authors demonstrated that the vibrational signals of a monolayer of cells could be probed, as well as the self-assembled monolayers below the cellular layer.\textsuperscript{127}

2.3. Calculation of Vibrational Amide SFG Spectra of Peptides and Proteins

Soon after the first SFG spectra of proteins at interfaces were collected by Shen, Somorjai, Cremer, Chen, and their co-workers almost 20 years ago,\textsuperscript{85,128–131} researchers in the field realized that SFG amide-I spectra of large proteins are too complex to be analyzed using peak assignments based on tabulated IR and Raman frequencies. This is due to (i) the coherent nature of sum-frequency generation, which leads to constructive and destructive interference between signals from various normal modes (often delocalized over various secondary structure elements) as a function of their relative phase, and (ii) the orientation sensitivity of SFG, which in some cases can render a certain secondary structure element invisible. There have been several approaches to overcome this issue: first, simple model peptides and proteins with just a single secondary structure element (a particular type of helix or strand) were designed and studied with SFG. Such spectra can be analyzed well by comparing them with spectral calculations based on the symmetry of isolated secondary structure elements (see section 2.3.1). While such calculations were initially performed only for the amide-I mode,\textsuperscript{132,133} in more recent years similar approaches have been developed for other amide modes.\textsuperscript{74} In 2012, ab initio calculations of small parts of an amyloid fibril structure were combined in order to calculate SFG spectra of the intermolecular $\beta$-sheet part of an amyloid aggregate\textsuperscript{34} (see section 2.3.2). At around the same time, excitonic amide-I Hamiltonian models that had previously been developed to calculate IR and Raman spectra were adapted for the calculation of SFG spectra, allowing the calculation of SFG for arbitrary-sized proteins based on the coupling between amide-I modes\textsuperscript{86,135} (see section 2.3.3). In more recent years, methods have been developed to calculate protein SFG spectra directly from MD trajectories, thereby including the spectral effects of conformational fluctuations on the pico- and nanosecond time scales into the calculations\textsuperscript{136} (see section 2.3.4).

These four types of methods all calculate the molecular hyperpolarizability tensor of the protein on the basis of an assumption for the protein conformation. In all spectral SFG calculations of proteins, the following steps are taken:

1. Obtain an assumption of the protein structure (from theory, the Protein Data Bank (PDB), or an MD trajectory).
2. Calculate the hyperpolarizability tensor of each normal mode in the molecular frame (based on one of the four methods described in sections 2.3.1, 2.3.2, 2.3.3, and 2.3.4).
3. Determine the second-order susceptibility tensor of each normal mode in the lab frame, $\chi^{(2)}_{\text{lab}}$ (see discussion below).
4. Sum over all normal modes and incorporate the potential nonresonant contribution (see eq 2).
5. Calculate the effective second-order susceptibilities $\chi^{(2)}_{\text{eff}}$ for the different polarization combinations (see discussion below).
6. Determine the SFG intensity using eq 1, in which the product of $I_{Y_{3S}}$ and $I_{Y_{3R}}$ is replaced by an overall scaling factor when the absolute intensity of the signal is not of interest.

In step 3 (vide supra), the second-order susceptibility tensor in the lab frame, $\chi^{(2)}_{\text{lab}}$, is computed with eq 3. In practice, this is often done using Euler transformations,\textsuperscript{82} for example as follows (although other rotation matrices can also be used for the same purpose; see, e.g., refs 53, 134, and 137):

$$\chi^{(2)}_{ijk} = \sum_{i,j,k=x,y,z} (R_{ij}R_{jk}R_{kl})\chi^{(2)}_{kl}$$

(12)

in which the rotation matrix $R$ is given by:

![Figure 13. Langmuir trough for SFG experiments developed by Franz et al. The trough can be used to set surface tension and prepare lipid monolayers with a well-defined thermodynamic phase. A surface tensiometer is used to track surface tension throughout the SFG experiment. The trough and surface tensiometer can be scanned in the xy plane to avoid local heating and beam damage. Legend: (1) syringe incision to the subphase; (2) Langmuir trough with heat sink; (3) external circulating path to/from chiller; (6) vertical level stage. Adapted with permission from ref 124. Copyright 2017 AIP Publishing.](Image)
In step 5 (vide supra), the effective second-order susceptibilities $\chi^{(2)}_{\text{eff}}$ for the different polarization combinations are calculated as:

$$
\chi^{(2)}_{\text{eff}} = \frac{[\hat{e}(\omega_{\text{IR}}) \cdot \mathbf{L}(\omega_{\text{IR}})]}{[\hat{e}(\omega_{\text{SF}}) \cdot \mathbf{L}(\omega_{\text{SF}})]} \left[ \exp(i \Omega \cdot \mathbf{L}(\omega_{\text{SF}})) \right]
$$

(14)

in which $\hat{e}(\Omega)$ and $\mathbf{L}(\Omega)$ are the unit polarization vectors and Fresnel factors (see eq 6) at frequency $\Omega$, respectively. The solutions of this equation are given in eqs 5 and 7.

![Figure 14. Definition of the molecular-frame axes $x$, $y$, $z$, the lab-frame axes $X$, $Y$, $Z$, and the Euler angles $\theta$, $\varphi$, and $\psi$, using the homopentameric cholera toxin B subunit as a model protein, with the $z$-axis aligned with its symmetry axis. Reproduced from ref 46. Copyright 2013 American Chemical Society.](image)

All of the methods described below will eventually lead to calculated SFG spectra that can be compared with experimental data. The calculated spectra (especially for homodyne SFG spectra) may not in all cases be a unique fit to the associated experimental spectra. As opposed to other techniques such as X-ray diffraction and NMR spectroscopy, it is not yet possible to derive the protein conformation directly from SFG spectra. However, especially when using more advanced techniques like phase-resolved or two-dimensional SFG or when studying multiple polarization combinations and/or frequency regions, one can optimize the discriminative value of the data and employ statistical methods to quantify the uniqueness of the match between the calculated and experimental spectra.

### 2.3.1. Ideal Secondary Structure Elements

In cases where only single secondary structure elements are studied, one can interpret the spectra at the molecular level by considering the symmetry properties of the peptide or protein. Ye, Chen, and co-workers derived the amide-I hyperpolarizability tensors of several perfect secondary structure elements, including the $\alpha$-helix, $\beta$-sheet, and the antiparallel $\beta$-sheet. To compare the results with those for random coils, the authors also derived the orientation dependence of the response of a free C=O group in the lab frame. In all cases, the hyperpolarizability tensor of the secondary structure element was first calculated by multiplying the polarization-dependent IR and Raman responses (according to eq 4), using a bond additivity model that follows the methodology of Higgs and validating the relative magnitudes of the IR transition dipole moments and Raman tensors with experimental values from previous experimental studies. Subsequently, the authors applied eq 3 to perform the Euler transformation of the various derived modes from the molecular frame into the lab frame. Because of their symmetry, the solutions simplify substantially, especially with the assumption of azimuthal symmetry (i.e., averaging over the Euler angle $\varphi$; see Figure 14) and uniaxial symmetry (i.e., averaging over the Euler angle $\psi$), which, for the example of a perfect $\alpha$-helix, results in the $\theta$ dependencies plotted in Figure 15.

The Chen group combined these spectral calculations with SFG measurements through a CaF$_2$ prism in a near-total internal reflection geometry (see section 2.2.2.2), allowing them to use the ratios of all of the polarization combinations to determine the orientational distribution (after taking the Fresnel factors into consideration), most notably the PPP polarization combination, which is nearly completely determined by $\chi_{zzz}^{(2)}$ in this case. By using other spectroscopies (e.g., attenuated total reflectance (ATR) IR spectroscopy) on the same system, the authors obtained an additional parameter that also allowed them to resolve the width of the distribution.

The underlying principles of such spectral calculations can be easily applied to other (amide) modes, as was shown for the amide-III mode by the Ye group.

### 2.3.2. Ab Initio-Based Spectral Calculations

The SFG spectra of smaller peptides and small parts of a protein can also be calculated with ab initio methods. For example, in a recent study of the peptide LK-$\beta$, the achiral and chiral C-H stretch SFG spectra were calculated to reveal the influence of the handedness of the amino acids on the different spectra. Even for such a small system (a seven amino acid peptide), computations based on second-order perturbation theory (VPT2) or vibrational self-consistent field (VSCF) methods, which include overtones and combination bands, are infeasible because of the dimensions. However, the authors found that an anharmonic analysis comprising Fermi resonances on single leucine and lysine residues and couplings between

$$
\mathbf{R} = \begin{bmatrix}
-sin(\psi) \cdot sin(\varphi) + cos(\theta) \cdot cos(\psi) \cdot cos(\varphi) & -cos(\psi) \cdot sin(\varphi) - cos(\theta) \cdot sin(\psi) \cdot cos(\varphi) & sin(\theta) \cdot cos(\varphi) \\
-sin(\theta) \cdot cos(\varphi) & cos(\theta) \cdot cos(\psi) & sin(\theta) \cdot sin(\varphi) \\
sin(\psi) \cdot cos(\varphi) + cos(\theta) \cdot sin(\psi) \cdot sin(\varphi) & cos(\psi) \cdot cos(\theta) & sin(\theta) \cdot sin(\varphi).
\end{bmatrix}
$$

(13)
modes was sufficient for understanding the C–H stretch spectral region.

If it is assumed that longer-range interactions can be neglected, ab initio methods can be combined with a so-called divide-and-conquer approach to determine the hyperpolarizabilities of larger proteins, as was shown by Xiao et al., who estimated the amide-I SFG response of an hIAPP fibril. The authors computed the hyperpolarizability of the fibril based on the amyloid parallel β-sheet structure proposed by Luca et al. (Figure 16a). In the spectral calculation, the structure was limited to two monomers in the fibril direction (Figure 16b). Subsequently, the structure was divided into 16 partially overlapped tripeptide segments (Figure 16c,d), whose geometries were optimized (subject to fixed dihedral angles), after which energy minimization and normal-mode analysis were performed using density functional theory (DFT) with the B3LYP functional and the 6-31G* basis set. The 96 normal modes obtained were then categorized as parallel β-sheet A and B modes on the basis of their orientation and frequency-scaled on the main peak of the experimental PSP spectrum (1622 cm⁻¹) with the B-mode frequency of the calculation. After the normal modes were summed in order to obtain the total hyperpolarizability, the response was transformed to the lab frame using eqs 12 and 13. With the assumption of a perfect parallel β-sheet structure (i.e., C₂ symmetry), a relatively simple dependence on the Euler angles θ and ψ was found, enabling them to determine the A/B intensity ratios for various polarization combinations as a function of the orientation. This also allowed the authors to compare the calculated chiral hyperpolarizability elements of the A and B modes to the experimental PSP spectrum after consideration of the Fresnel factors (eq 7). By combining their results with previously reported infrared reflection absorption spectroscopy (IRRAS) data, the authors found a unique orientation that matches the experimental data best (see section 3.4.1 for further details about the hIAPP studies by the Yan group).

2.3.3. Exciton Models. For peptides and proteins that cannot be approximated with ideal secondary structures or that are too large and/or complex to be simulated with ab initio methods, one can apply excitonic Hamiltonian models to...
calculate amide spectra.\textsuperscript{46,72,142} In such models, a Hamiltonian of the system is constructed that contains all of the local amide mode frequencies as well as the couplings between them. Solving the time-independent Schrödinger equation $\hat{H}\Psi = E\Psi$ determines the vibrational eigenmodes, from which spectroscopic responses (e.g., the molecular hyperpolarizability) can be derived. Many different models for estimating the local mode frequencies and couplings of amide-I modes have been published, but the general approach described here is the same for all.

This approach is based on various theoretical frameworks published in preceding years that used excitonic Hamiltonian models to calculate amide-I IR and Raman spectra of proteins. Linear (steady-state) amide-I IR spectra of proteins were calculated with excitonic coupling models in refs 142–147; Huang and Schweitzer-Stenner\textsuperscript{148} and Tsuboi et al.\textsuperscript{149} performed such calculations to predict the Raman spectra of proteins, and Gorbunov et al.,\textsuperscript{150} Falvo et al.,\textsuperscript{151} Ganin et al.,\textsuperscript{72} and Hamm et al.\textsuperscript{152,153} created similar excitonic models for interpreting nonlinear amide-I 2D-IR protein spectra.

Shortly after this, excitonic models were also applied to calculate SFG spectra.\textsuperscript{46,135,154,155} In a computational study, Liang et al.\textsuperscript{135} calculated SFG spectra of a mechanosensitive channel. The calculated spectra were different for an intermediate state compared with the initial and final states, while the states were indiscernible in calculated IR spectra. Roeters et al. demonstrated how spectra calculated with a single-exciton Hamiltonian model can be fitted to experimental SFG spectra in order to determine the orientation of proteins\textsuperscript{86} (also see section 3.3). Ding et al. used an excitonic Hamiltonian calculation to show that the coupling between the amide-I band and an isotopic label has a strong influence on the line shape.\textsuperscript{155} Such single-exciton Hamiltonians can be straightforwardly expanded to two-exciton Hamiltonians, from which 2D-SFG spectra can be calculated, as shown by the Zanni group.\textsuperscript{73,99,100,154,156} Since then, the conformation and orientation of several peptides and proteins at interfaces have been determined with this methodology.\textsuperscript{73,99,100,154–167}

2.3.3.1. Single-Exciton Amide-I Hamiltonian. The excitonic amide-I Hamiltonians are constructed from the local modes and the couplings between them, which are calculated from the atomic coordinates of the hypothetical protein structure. In the local mode basis, the single-exciton Hamiltonian matrix has the following form:

\[
H^{(1)} = \begin{bmatrix}
\omega_1^0 & \kappa_{12} & \kappa_{13} & \kappa_{14} & \cdots \\
\kappa_{12} & \omega_2^0 & \kappa_{23} & \kappa_{24} & \cdots \\
\kappa_{13} & \kappa_{23} & \omega_3^0 & \kappa_{34} & \cdots \\
\kappa_{14} & \kappa_{24} & \kappa_{34} & \omega_4^0 & \cdots \\
\vdots & \vdots & \vdots & \vdots & \ddots \\
\end{bmatrix}
\] (15)

in which $\omega_i^0$ is the frequency of local mode $i$ (see section 2.3.3.2) and $\kappa_{ij}$ is the coupling between local modes $i$ and $j$ (see section 2.3.3.3). Often the numerical implementation relies on the PDB format, so that spectra can be calculated for any structure obtained by X-ray crystallography, NMR spectroscopy, or molecular dynamics (MD) simulation that has been uploaded to a protein structure database as well as for theoretically predicted structures that might be expected for a particular experimental situation. The amide-I vibrations of proteins are mainly determined by the secondary structure and are not strongly affected by the nature of the side chains.\textsuperscript{59,66} This approximation generally works well, but more sophisticated local mode frequency approximations may be applied.\textsuperscript{168} Furthermore, it is often assumed that the amide-I modes are spectrally isolated from other modes, so that the backbone modes couple only with each other and not with other modes. This enables calculation of the amide-I spectrum using only the amide-I Hamiltonian. Including other modes (e.g., from side chains or water) is a straightforward extension of this formalism.

Once all of the elements of the Hamiltonian have been estimated with the models discussed below (sections 2.3.3.2 and 2.3.3.3), the delocalized vibrational eigenmodes are obtained by solving the time-independent Schrödinger equation, from which the IR, Raman, and vibrational SFG spectral responses (see section 2.1.1), as well as their higher-dimensional variants, can be calculated.\textsuperscript{86,99}

2.3.3.2. Local Mode Frequencies. In several SFG studies, the local mode frequencies were assumed to be the same for all backbone amide-I modes\textsuperscript{99,161–164} (which works particularly well for repetitive protein structures or when a similar hydrogen-bonding configuration is expected for all amide groups), but in general, more complicated models have been applied. Hydrogen bonds, side chains, and the valency of the amide part (secondary vs tertiary, as in the case of proline residues) all affect the uncoupled frequencies of the modes to a certain degree. These frequencies can be estimated with varying levels of accuracy ranging from simple empirical models (e.g., only taking into account the intramolecular hydrogen-bond-induced shifts with a linear model\textsuperscript{155}) to highly complex models (e.g., mapping the effect of the electric field and its gradients at the C, O, N, and D atoms of deuterated NMA on the vibrational frequency via ab initio calculations\textsuperscript{155,165}) to hybrid forms (e.g., also including hydrogen bonds from solvent molecules and to the N–D part of the deuterated backbone amide groups\textsuperscript{46} or exploiting the influence of each of the three hydrogen bonds in which an amide group can participate on the C=O bond length,\textsuperscript{170} which in turn has a linear relationship with the vibrational frequency according to another DFT study\textsuperscript{171}).

2.3.3.3. Couplings. Also, the couplings can be estimated at varying levels of complexity. The earliest model developed for the calculation of IR and Raman spectra is the transition dipole coupling (TDC) model,\textsuperscript{172} which has been used in many publications since (e.g., refs 72 and 142–151). In this model, the coupling between transition dipole moments is given by a Coulomb-like model based on the relative orientation and distance between the transition dipole moments (see Figure 17). An alternative coupling model that is also often used in excitonic spectral calculations is the transition charge coupling model, in which the derivatives of the Mulliken charges with respect to the amide-I mode are calculated and a more detailed version of the TDC equation is used that incorporates the interactions between all of the charges involved in the coupling.\textsuperscript{152,169,173} More complex models of the electrostatic interaction have also been applied.\textsuperscript{169,174} The through-space couplings can be estimated well with such models, but the couplings between adjacent amide groups in tightly folded conformations can be overestimated by this model.\textsuperscript{164} Because these are dominated by through-bond effects, the nearest-neighbor couplings are often estimated using parametrized maps of ab initio calculations that give the coupling as a function of the Ramachandran angles between the neighboring amide groups,\textsuperscript{86,135,150,169}
2.3.4. Spectral Calculations Based on Molecular Dynamics Simulations. Harrison et al. recently developed a method to determine the orientation of proteins at surfaces using a combination of Monte Carlo simulations and SFG. The authors simulated the orientation of the B1 domain of protein G on polystyrene using Monte Carlo methods and then compared calculated spectra based on the two most probable (i.e., lowest-energy) states with experimental results. It turned out that neither of the two low-energy states matched the experimental data. Only a combination of the two structures yielded spectra with a good match. This result underlines that one has to be careful about assuming that there is one well-defined protein structure at the surface. Rather, an ensemble of structures with varying probabilities based on the binding energy and the folding state is likely to be present. It is thus important to capture the variability of protein structure at interfaces to fully understand interfacial protein action. To do this, one can calculate SFG spectra for molecular dynamics (MD) trajectories, as has been done by the Jansen group, the Zanni group, and the Weidner group. In 2015, Carr et al. demonstrated how MD trajectories can be combined with line shape theory to examine the importance of dynamical and coupling effects on calculated homo- and heterodyne SFG spectra. Also, snapshots from metadynamics MD simulations can be used for such an approach, as shown by Bellucci et al., allowing a more exhaustive sampling of the protein conformation. Spectral calculations like the ones presented in this subsection can help with the notoriously difficult development of force fields and sampling methods for proteins at interfaces by providing a direct connection between simulation and experiment.

3. VIBRATIONAL SFG STUDIES OF PEPTIDES AND PROTEINS

In this section, we review the SFG literature in which peptides and proteins at interfaces have been studied. An overview of the reviewed studies is given in Table 2. In the selection of studies discussed in more detail in the various subsections of this section, we tried to highlight cases where either a new insight into the studied system was gained or a new approach was taken to complement conventional SFG studies. Excellent review articles exist that cover the main findings for biological systems using nonlinear spectroscopy. Table 2 summarizes the variety of interfaces and proteins studied with SFG. While the table lists the earliest citations for the research topics, most of the proteins are still under study, and several are discussed further within this review. In the following, this review focuses on articles published in recent years.

3.1. Model Systems

The first SFG experiments investigating the surface binding of proteins were reported by the Cremer group and the Chen group. In these studies, SFG was used to probe the adsorption of proteins such as bovine serum albumin, lysozyme, and fibrinogen onto hydrophobic polymer surfaces as well as polar surfaces such as calcium fluoride and silicon dioxide. These early experiments provided mostly qualitative results, but they demonstrated the potential of SFG as a reporter for biointerface science and laid the groundwork for more quantitative SFG methods. In those early days of protein SFG, the immense complexity and redundant occurrence of amino acid side chains made it almost impossible to interpret SFG spectra of large proteins in terms of backbone folding or side-chain structure. For this reason, many early experiments were performed using short synthetic model peptides, which allowed a detailed analysis of protein–surface interactions on the molecular level in terms of orientation, folding, and side-chain structure. The early model peptide studies established methods to determine protein orientation by analyzing the phases of side-chain methyl and amide modes. The interaction of individual side chains with surfaces is often key to understanding the specific binding of proteins with target surfaces. Holinga et al. and Watry et al. made the first efforts to understand the interactions of single amino acids with surfaces by probing the structure of a series of amino acids at hydrophilic and hydrophobic interfaces. In addition to providing valuable reference data, the authors showed that at hydrophobic surfaces most of the amino acids investigated exhibited C−H vibrational modes, whereas the SFG spectra obtained at hydrophilic surfaces were featureless. This indicated that hydrophobic surfaces induced an ordered conformation, whereas at hydrophilic surfaces the amino acids remained unordered. The interactions of individual amino acids with surfaces likely differ significantly from the interactions of amino acid side chains.

3.1.1. Model Peptides

3.1.1.1. Leucine−Lysine Peptides.

Synthetic leucine−lysine (LK) peptides have served as prominent model systems for years. Samuel was the first to use LK peptides as model systems for SFG studies. It turns out that LK peptides are especially useful to study the role of interfacial folding because (i) they are amphiphilic and therefore bound to many different types of surfaces and (ii) depending on the hydrophobic periodicity of the amino acid sequence, surface binding enforces α-helix, 310-helix, or β-strand structures. LK peptides have been studied on a variety of surfaces, such as self-assembled monolayers (SAMs), polymers, and liquid surfaces, and also at silica interfaces. Weidner et al. used LK peptides to develop methods to probe the structure of specific side chains at surfaces using isotope labels. By labeling the individual leucine sites within an α-helical LK peptide, the authors determined the orientation of all hydrophobic side chains at a polystyrene interface (Figure 18). York et al. extended the basic concept of LK peptides to also study lysine−phenylalanine- and lysine−alanine-based peptides at surfaces. Peres et al. recently studied LK/β peptides composed of solely L- or D-amino acids or mixed D/L-amino acids to explore the impact of handedness on chiral SFG signals. Protein chains that contain all D- or mixed D/L-amino acids are promising as antimicrobial, antiviral, and
Table 2. Overview of the Reviewed Literature

<table>
<thead>
<tr>
<th>Peptides and Proteins</th>
<th>Year</th>
<th>Ref(s)</th>
<th>Technique</th>
<th>Region(s)</th>
<th>Type</th>
<th>Interface</th>
</tr>
</thead>
<tbody>
<tr>
<td>15-mer oligonucleotides</td>
<td>2007</td>
<td>40</td>
<td>conventional</td>
<td>CH</td>
<td>DNA or RNA oligomer</td>
<td>solid–air, on a functionalized glass substrate</td>
</tr>
<tr>
<td>α-chymotrypsin</td>
<td>2015</td>
<td>178</td>
<td>chiral/achiral/heterodyne</td>
<td>amide-I/CH/OH</td>
<td>enzyme</td>
<td>liquid–air</td>
</tr>
<tr>
<td>α-synuclein</td>
<td>2009</td>
<td>179</td>
<td>conventional</td>
<td>amide-I</td>
<td>model protein</td>
<td>liquid–air</td>
</tr>
<tr>
<td>α-lactalbumin</td>
<td>2016</td>
<td>180</td>
<td>phase-resolved</td>
<td>CH/OH</td>
<td>milk protein</td>
<td>liquid–air</td>
</tr>
<tr>
<td>AFP-III</td>
<td>2014</td>
<td>181</td>
<td>conventional</td>
<td>CH/OH</td>
<td>antifreeze protein</td>
<td>liquid–air</td>
</tr>
<tr>
<td>AHP peptide</td>
<td>2014</td>
<td>99</td>
<td>2D-SFG</td>
<td>amide-I</td>
<td>model peptide</td>
<td>solid–liquid, on a gold substrate</td>
</tr>
<tr>
<td>AK14</td>
<td>2007</td>
<td>182, 183</td>
<td>conventional</td>
<td>CH/OH</td>
<td>model peptide</td>
<td>solid–liquid</td>
</tr>
<tr>
<td>AK7</td>
<td>2007</td>
<td>182, 183</td>
<td>conventional</td>
<td>CH/OH</td>
<td>model peptide</td>
<td>solid–liquid</td>
</tr>
<tr>
<td>alamethicin</td>
<td>2010</td>
<td>184–186</td>
<td>conventional</td>
<td>amide-I</td>
<td>antimicrobial peptide, model for large channel proteins</td>
<td>lipid bilayer</td>
</tr>
<tr>
<td>albumin</td>
<td>2006</td>
<td>187</td>
<td>conventional</td>
<td>CD</td>
<td>model protein</td>
<td>polymer–solution</td>
</tr>
<tr>
<td>amyloid-β peptide (Aβ16–22)</td>
<td>2016</td>
<td>176</td>
<td>conventional</td>
<td>amide-I</td>
<td>amyloid peptide, model protein for aggregation</td>
<td>liquid–solid (gold)</td>
</tr>
<tr>
<td>AR14</td>
<td>2007</td>
<td>182, 183</td>
<td>conventional</td>
<td>CH/OH</td>
<td>model peptide</td>
<td>solid–liquid</td>
</tr>
<tr>
<td>AR5</td>
<td>2007</td>
<td>182, 183</td>
<td>conventional</td>
<td>CH/OH</td>
<td>model peptide</td>
<td>solid–liquid</td>
</tr>
<tr>
<td>avidin</td>
<td>2004</td>
<td>188, 189</td>
<td>conventional</td>
<td>CH/OH</td>
<td>biotin-binding protein</td>
<td>gold nanoparticles</td>
</tr>
<tr>
<td>β-galactosidase</td>
<td>2013</td>
<td>190</td>
<td>conventional</td>
<td>amide-I</td>
<td>model enzyme</td>
<td>immobilized on a SAM on gold</td>
</tr>
<tr>
<td>biocytin</td>
<td>2004</td>
<td>188</td>
<td>conventional</td>
<td>CH/OH</td>
<td>derived vitamin</td>
<td>solid–liquid</td>
</tr>
<tr>
<td>bisphenol A</td>
<td>2003</td>
<td>191</td>
<td>conventional</td>
<td>CH</td>
<td>endocrine disrupter</td>
<td>polymer</td>
</tr>
<tr>
<td>BLG</td>
<td>2013</td>
<td>78</td>
<td>conventional</td>
<td>CH/OH</td>
<td>milk protein</td>
<td>liquid–air</td>
</tr>
<tr>
<td>bovine insulin</td>
<td>2015</td>
<td>192</td>
<td>conventional</td>
<td>amide-I</td>
<td>model protein</td>
<td>lipid monolayer</td>
</tr>
<tr>
<td>BdA</td>
<td>2016</td>
<td>205</td>
<td>chiral/achiral</td>
<td>amide-I</td>
<td>bacterial biofilm protein</td>
<td>liquid–air</td>
</tr>
<tr>
<td>catalase</td>
<td>2019</td>
<td>206</td>
<td>conventional</td>
<td>CH/OH</td>
<td>enzyme</td>
<td>liquid–air</td>
</tr>
<tr>
<td>cepacillin P1 (CP1)</td>
<td>2011</td>
<td>74, 207–209</td>
<td>conventional</td>
<td>amide-I/amide-III/CH</td>
<td>antimicrobial peptide</td>
<td>CsF prism and lipid bilayer, liquid–air</td>
</tr>
<tr>
<td>SAP(E)</td>
<td>2016</td>
<td>210</td>
<td>conventional</td>
<td>CH/amide-I</td>
<td>cell-penetrating peptide</td>
<td>liquid–air</td>
</tr>
<tr>
<td>cellulose Iα and Iβ crystals</td>
<td>2013</td>
<td>211, 212</td>
<td>conventional</td>
<td>CH/OH/fingerprint region</td>
<td>polysaccharide</td>
<td>crystalline cellulose</td>
</tr>
<tr>
<td>chitin</td>
<td>2016</td>
<td>213</td>
<td>conventional</td>
<td>CH/CH/OH</td>
<td>glucose derivative</td>
<td>solid–air, film on glass substrate</td>
</tr>
<tr>
<td>cholera toxin B</td>
<td>2013</td>
<td>46</td>
<td>conventional</td>
<td>CH/OH</td>
<td>enzyme</td>
<td>lipid monolayer</td>
</tr>
<tr>
<td>chymotrypsinogen A</td>
<td>2019</td>
<td>206</td>
<td>conventional</td>
<td>CH/OH</td>
<td>enzyme</td>
<td>liquid–air</td>
</tr>
<tr>
<td>concanavalin A</td>
<td>2015</td>
<td>178</td>
<td>chiral/achiral/heterodyne</td>
<td>amide-I/CH/OH</td>
<td>model protein</td>
<td>liquid–air</td>
</tr>
<tr>
<td>cytochrome b₅</td>
<td>2010</td>
<td>214</td>
<td>conventional</td>
<td>amide-I</td>
<td>membrane-penetrating protein</td>
<td>lipid bilayer</td>
</tr>
<tr>
<td>cytochrome c</td>
<td>2008</td>
<td>18,21,2,16</td>
<td>electronic SFG</td>
<td>350–450 nm</td>
<td>heme protein</td>
<td>liquid–air</td>
</tr>
<tr>
<td>DAFP-1</td>
<td>2015</td>
<td>217</td>
<td>conventional</td>
<td>CH/OH</td>
<td>antifreeze protein</td>
<td>liquid–air</td>
</tr>
<tr>
<td>D-magainin 2</td>
<td>2006</td>
<td>119, 218</td>
<td>conventional/polarization-sensitive</td>
<td>amide</td>
<td>model polymeric amino acid</td>
<td>solid–liquid</td>
</tr>
<tr>
<td>DOPA-inspired polymer</td>
<td>2013</td>
<td>219</td>
<td>conventional</td>
<td>amide-I/CH</td>
<td>antimicrobial peptide</td>
<td>liquid–solid</td>
</tr>
<tr>
<td>duramycin</td>
<td>2010</td>
<td>220</td>
<td>conventional</td>
<td>amide-I/CH</td>
<td>cell-binding peptide</td>
<td>lipid monolayer</td>
</tr>
<tr>
<td>elastin-like polypeptide (ELP)</td>
<td>2010</td>
<td>198</td>
<td>conventional</td>
<td>CH/OH</td>
<td>elastin-like model polypeptide</td>
<td>liquid–air</td>
</tr>
<tr>
<td>Peptides and proteins</td>
<td>Year</td>
<td>Ref(s)</td>
<td>Technique</td>
<td>Region(s)</td>
<td>Type</td>
<td>Interface</td>
</tr>
<tr>
<td>-----------------------</td>
<td>------</td>
<td>--------</td>
<td>-----------</td>
<td>-----------</td>
<td>------</td>
<td>-----------</td>
</tr>
<tr>
<td>Factor XII</td>
<td>2005</td>
<td>20, 131, 221</td>
<td>Conventional</td>
<td>CH/OH</td>
<td>Model protein</td>
<td>Solid-liquid</td>
</tr>
<tr>
<td>FGAIL</td>
<td>2015</td>
<td>100, 166</td>
<td>2D-SFG</td>
<td>Amide-I</td>
<td>Amyloid-forming peptide</td>
<td>Solid-liquid, on gold coated with MMB and MBA monolayers</td>
</tr>
<tr>
<td>FGF-1</td>
<td>2004</td>
<td>222</td>
<td>Conventional</td>
<td>CH</td>
<td>Regulator of embryogenesis</td>
<td>Lipid bilayer</td>
</tr>
<tr>
<td>Fibrils in guinea pig spinal cord</td>
<td>2007</td>
<td>29</td>
<td>Imaging</td>
<td>CH</td>
<td>Helical fibrils</td>
<td>Biological tissue</td>
</tr>
<tr>
<td>Fibrinogen</td>
<td>2003</td>
<td>20, 50, 128, 139, 187, 193, 223-225</td>
<td>Chiral/achiral</td>
<td>CH/Amide-I</td>
<td>Glycoprotein</td>
<td>PS-liquid</td>
</tr>
<tr>
<td>FGAIL</td>
<td>2015</td>
<td>100, 166</td>
<td>2D-SFG</td>
<td>Amide-I</td>
<td>Amyloid-forming peptide</td>
<td>Solid-liquid, on gold coated with MMB and MBA monolayers</td>
</tr>
<tr>
<td>FGF-1</td>
<td>2004</td>
<td>222</td>
<td>Conventional</td>
<td>CH</td>
<td>Regulator of embryogenesis</td>
<td>Lipid bilayer</td>
</tr>
<tr>
<td>G protein</td>
<td>2004</td>
<td>122</td>
<td>Conventional</td>
<td>CH</td>
<td>Regulator of embryogenesis</td>
<td>Lipid bilayer</td>
</tr>
<tr>
<td>G protein βγ</td>
<td>2007</td>
<td>223, 227</td>
<td>Conventional</td>
<td>Amide</td>
<td>Nucleotide binding protein</td>
<td>Lipid bilayer</td>
</tr>
<tr>
<td>GALA</td>
<td>2015</td>
<td>228, 229</td>
<td>Conventional</td>
<td>Amide-I</td>
<td>Cell-penetrating peptide</td>
<td>Liquid-air</td>
</tr>
<tr>
<td>GB1 (V21C and T11C)</td>
<td>2010</td>
<td>33, 230</td>
<td>Conventional</td>
<td>Amide-I</td>
<td>Model immunoglobulin binding protein</td>
<td>Maleimide on gold-solution, immobilized</td>
</tr>
<tr>
<td>GFPmut2</td>
<td>2004</td>
<td>231</td>
<td>Conventional</td>
<td>CH and 350-550 nm</td>
<td>Green fluorescent protein</td>
<td>Pt(111)-solution</td>
</tr>
<tr>
<td>Glu5</td>
<td>2015</td>
<td>232</td>
<td>Conventional</td>
<td>Amide-I</td>
<td>Model peptide</td>
<td>Solid-liquid, gold substrate</td>
</tr>
<tr>
<td>Glycoprotein from frog mucus</td>
<td>2018</td>
<td>232</td>
<td>Conventional</td>
<td>CH</td>
<td>Glycoprotein</td>
<td>Liquid-air</td>
</tr>
<tr>
<td>Gramicidin A</td>
<td>2003</td>
<td>234, 235</td>
<td>Conventional</td>
<td>CH</td>
<td>Model peptide</td>
<td>Liquid-air, DMPC monolayer</td>
</tr>
<tr>
<td>GRK 5</td>
<td>2013</td>
<td>226</td>
<td>Conventional</td>
<td>Amide-I</td>
<td>Membrane protein</td>
<td>Lipid bilayer</td>
</tr>
<tr>
<td>Gβ1/2 and Gαβ1γ2</td>
<td>2013</td>
<td>136</td>
<td>Conventional</td>
<td>Amide-I</td>
<td>Heterotrimeric G protein complexes</td>
<td>Lipid bilayer</td>
</tr>
<tr>
<td>Hemoglobin</td>
<td>2017</td>
<td>237</td>
<td>Heterodyme</td>
<td>CH/OH</td>
<td>Oxygen transport metalloprotein</td>
<td>Liquid-air</td>
</tr>
<tr>
<td>Human islet amyloid polypeptide (hIAPP) and other IAPPs</td>
<td>2011</td>
<td>52, 61, 68, 238, 239</td>
<td>Chiral</td>
<td>NH/Amide-I</td>
<td>Insulin</td>
<td>Liquid-air</td>
</tr>
<tr>
<td>Human insulin</td>
<td>2014</td>
<td>240</td>
<td>Conventional</td>
<td>Amide-I</td>
<td>Human plasma fibrinogen</td>
<td>Silica-liquid</td>
</tr>
<tr>
<td>Human plasma fibrinogen (HPF)</td>
<td>2003</td>
<td>241</td>
<td>Conventional</td>
<td>CH/OH</td>
<td>Human serum albumin</td>
<td>Solid-liquid, PEO film</td>
</tr>
<tr>
<td>Human serum albumin</td>
<td>2012</td>
<td>242</td>
<td>Conventional</td>
<td>Amide-I/Fingerprint region</td>
<td>Cysteine-rich protein</td>
<td>Liquid-air</td>
</tr>
<tr>
<td>Hydrophobin (HFB1)</td>
<td>2017</td>
<td>160, 243</td>
<td>Conventional</td>
<td>CH</td>
<td>Membrane protein</td>
<td>Lipid monolayer</td>
</tr>
<tr>
<td>IM 30</td>
<td>2015</td>
<td>158</td>
<td>Conventional</td>
<td>Amide-I</td>
<td>Antibody</td>
<td>Solid-liquid, solid-air</td>
</tr>
<tr>
<td>Immunoglobulin G (IgG)</td>
<td>2005</td>
<td>177, 244</td>
<td>Conventional</td>
<td>CH</td>
<td>Antibody</td>
<td>Solid-liquid, solid-air</td>
</tr>
<tr>
<td>InaZ</td>
<td>2016</td>
<td>245</td>
<td>Conventional/time-resolved</td>
<td>CD/OD</td>
<td>Ice-nucleating protein</td>
<td>Liquid-air, on the outer membrane of P. syringae bacteria</td>
</tr>
<tr>
<td>Isoniazide</td>
<td>2010</td>
<td>246</td>
<td>Conventional</td>
<td>CH</td>
<td>Bactericidal/hydrolytic therapeutic agent</td>
<td>Lipid monolayer</td>
</tr>
<tr>
<td>KALP23</td>
<td>2019</td>
<td>68, 247</td>
<td>Pump/probe</td>
<td>NH/Amide-I</td>
<td>Model peptide</td>
<td>Liquid-air</td>
</tr>
<tr>
<td>L 1 (NCAM L1)</td>
<td>2011</td>
<td>248</td>
<td>Conventional</td>
<td>CH/OH</td>
<td>Cell-adhesion protein</td>
<td>Solid-liquid, bound to a silver substrate</td>
</tr>
<tr>
<td>Limonene</td>
<td>2014</td>
<td>56, 249</td>
<td>Chiral/heterodyne/prochiral</td>
<td>CH</td>
<td>Chirality model</td>
<td>Liquid-air</td>
</tr>
<tr>
<td>l-Isoleucine</td>
<td>2002</td>
<td>250</td>
<td>Conventional</td>
<td>CH</td>
<td>Model peptide</td>
<td>Solid-air</td>
</tr>
<tr>
<td>LKα</td>
<td>2007</td>
<td>18, 53, 54, 58, 182, 183, 251, 252</td>
<td>Conventional</td>
<td>CH/OH</td>
<td>Model peptide</td>
<td>Solid-liquid</td>
</tr>
<tr>
<td>LKR14</td>
<td>2011</td>
<td>238</td>
<td>Chiral</td>
<td>NH/Amide-I</td>
<td>Model peptide</td>
<td>Liquid-air</td>
</tr>
<tr>
<td>LKer14</td>
<td>2009</td>
<td>10, 34, 74, 161, 183, 253, 254</td>
<td>Conventional</td>
<td>NH/Amide/CH</td>
<td>Model peptide</td>
<td>Solid-liquid, polystyrene (PS)</td>
</tr>
<tr>
<td>LKφ15</td>
<td>2010</td>
<td>183, 254</td>
<td>Conventional</td>
<td>CH</td>
<td>Model peptide</td>
<td>Solid-liquid</td>
</tr>
<tr>
<td>Peptides and proteins</td>
<td>Year(s)</td>
<td>Ref(s)</td>
<td>Technique</td>
<td>Region(s)</td>
<td>Type</td>
<td>Interface</td>
</tr>
<tr>
<td>--------------------------------------</td>
<td>---------</td>
<td>--------</td>
<td>----------------------------</td>
<td>-----------------------------</td>
<td>--------------------------</td>
<td>-----------------------------</td>
</tr>
<tr>
<td>LL-37</td>
<td>2013</td>
<td>257</td>
<td>Conventional</td>
<td>Amide-I/CH/OH</td>
<td>Human antimicrobial peptide</td>
<td>Lipid bilayer</td>
</tr>
<tr>
<td>t-leucine</td>
<td>2002</td>
<td>250, 258</td>
<td>Conventional</td>
<td>CH</td>
<td>Model amino acid</td>
<td>Oil–air, liquid–air</td>
</tr>
<tr>
<td>l-lysine monohydrate</td>
<td>2002</td>
<td>250, 259</td>
<td>Conventional</td>
<td>CH–OH</td>
<td>Model amino acid</td>
<td>Oil–air, liquid–PS</td>
</tr>
<tr>
<td>l-mandelic acid</td>
<td>2002</td>
<td>250</td>
<td>Conventional</td>
<td>CH</td>
<td>Model amino acid</td>
<td>Oil–air</td>
</tr>
<tr>
<td>l-methionine</td>
<td>2002</td>
<td>250</td>
<td>Conventional</td>
<td>CH</td>
<td>Model amino acid</td>
<td>Oil–air</td>
</tr>
<tr>
<td>l-phenylalanine</td>
<td>2002</td>
<td>250, 260</td>
<td>Conventional</td>
<td>CH</td>
<td>Model amino acid, antimicrobial</td>
<td>Oil–air, glassy carbon</td>
</tr>
<tr>
<td>l-proline</td>
<td>2009</td>
<td>259</td>
<td>Conventional</td>
<td>CH/OH</td>
<td>Model amino acid</td>
<td>Liquid–PS</td>
</tr>
<tr>
<td>L-tryptophan</td>
<td>2002</td>
<td>250</td>
<td>Conventional</td>
<td>CH</td>
<td>Model amino acid</td>
<td>Oil–air</td>
</tr>
<tr>
<td>Lysozyme</td>
<td>2002</td>
<td>250, 258, 259</td>
<td>Conventional</td>
<td>CH</td>
<td>Model amino acid</td>
<td>Oil–air</td>
</tr>
<tr>
<td>mastoparan and mastoparan X (MP-X)</td>
<td>2013</td>
<td>74, 262–264</td>
<td>Conventional</td>
<td>Amide-I/Amide-II</td>
<td>G-protein-activating peptide</td>
<td>Liquid–air, lipid bilayer</td>
</tr>
<tr>
<td>matrix protein 2</td>
<td>2018</td>
<td>265</td>
<td>Conventional</td>
<td>Amide-I/Amide-II</td>
<td>Influenza A virus</td>
<td>Lipid bilayer</td>
</tr>
<tr>
<td>Mefp-2</td>
<td>2003</td>
<td>131</td>
<td>Conventional</td>
<td>Amide-I/Amide-II</td>
<td>Mussel adhesive protein</td>
<td>Solid–liquid</td>
</tr>
<tr>
<td>Mefp-3</td>
<td>2008</td>
<td>266</td>
<td>Conventional</td>
<td>Amide-I/Amide-II</td>
<td>Mussel adhesive protein</td>
<td>Solid–liquid</td>
</tr>
<tr>
<td>Melittin</td>
<td>2007</td>
<td>60, 267, 268</td>
<td>Conventional</td>
<td>Amide-I/Amide-II</td>
<td>Model peptide</td>
<td>Lipid bilayer</td>
</tr>
<tr>
<td>Msi-367</td>
<td>2010</td>
<td>269</td>
<td>Conventional</td>
<td>Amide-I/Amide-II</td>
<td>Synthetic peptide</td>
<td>Lipid bilayer</td>
</tr>
<tr>
<td>Msi-594</td>
<td>2005</td>
<td>218</td>
<td>Conventional</td>
<td>Amide-I/Amide-II</td>
<td>Model peptide</td>
<td>Solid–liquid, PS</td>
</tr>
<tr>
<td>Msi-78 or Pexiganan</td>
<td>2011</td>
<td>270</td>
<td>Conventional</td>
<td>Amide-I/Amide-II</td>
<td>Antimicrobial peptide</td>
<td>Lipid bilayer</td>
</tr>
<tr>
<td>Msi-843</td>
<td>2006</td>
<td>218</td>
<td>Conventional</td>
<td>Amide-I/Amide-II</td>
<td>Model peptide</td>
<td>Solid–liquid</td>
</tr>
<tr>
<td>Myoglobin</td>
<td>2017</td>
<td>237</td>
<td>Heterodyne</td>
<td>Ch/OH</td>
<td>Iron- and oxygen-binding protein</td>
<td>Liquid–air</td>
</tr>
<tr>
<td>Nitroreductase (NfsB)</td>
<td>2014</td>
<td>271–273</td>
<td>Conventional</td>
<td>Amide-I</td>
<td>Enzyme</td>
<td>Solid–liquid, immobilized on a SAM on gold</td>
</tr>
<tr>
<td>Oleosin</td>
<td>2013</td>
<td>274</td>
<td>Conventional</td>
<td>Amide-I</td>
<td>Structural protein</td>
<td>Liquid–air</td>
</tr>
<tr>
<td>Ovalbumin</td>
<td>2019</td>
<td>206</td>
<td>Conventional</td>
<td>CH/OH</td>
<td>Egg white protein</td>
<td>Liquid–air</td>
</tr>
<tr>
<td>Ovisprin-1</td>
<td>2013</td>
<td>155, 275</td>
<td>Conventional</td>
<td>Amide-I</td>
<td>Polypeptide</td>
<td>Solid–liquid</td>
</tr>
<tr>
<td>Pardaxin</td>
<td>2014</td>
<td></td>
<td>Conventional</td>
<td>Amide-I/Amide-II</td>
<td>Model peptide</td>
<td>Liquid–air</td>
</tr>
<tr>
<td>PHLG</td>
<td>2004</td>
<td>276</td>
<td>Conventional</td>
<td>Amide-I/CH/OH</td>
<td>Homopolypeptide</td>
<td>Solid–air, thin film</td>
</tr>
<tr>
<td>Pentapeptide FGAIL (part of amylin/blAPP)</td>
<td>2018</td>
<td>166</td>
<td>2D-SPG</td>
<td>Amide-I/Carboxylic acid/enter</td>
<td>Part of human islet amyloid polypeptide</td>
<td>Solid–air, immobilized on a SAM on gold</td>
</tr>
<tr>
<td>Pep-1</td>
<td>2012</td>
<td>277</td>
<td>Conventional</td>
<td>Amide-I/CH/OH</td>
<td>Cell-penetrating peptide</td>
<td>Lipid bilayer</td>
</tr>
<tr>
<td>Peptin</td>
<td>2015</td>
<td>178</td>
<td>Chiral/achiral/heterodyne</td>
<td>Amide-I/CH/OH</td>
<td>Model protein</td>
<td>Liquid–air</td>
</tr>
<tr>
<td>Peptide R5</td>
<td>2017</td>
<td>278</td>
<td>Conventional</td>
<td>Amide-I/CH/OH</td>
<td>Biomineralization</td>
<td>Liquid–air</td>
</tr>
<tr>
<td>PHILIP</td>
<td>2011</td>
<td>238</td>
<td>Chiral</td>
<td>NH/Amide-I</td>
<td>Low-pH insertion peptide</td>
<td>Liquid–air</td>
</tr>
<tr>
<td>Phospholipase A2 (PLA2)</td>
<td>2010</td>
<td>279</td>
<td>Conventional</td>
<td>CH/CD</td>
<td>Enzyme</td>
<td>Lipid bilayer</td>
</tr>
<tr>
<td>Polyaldehyde B</td>
<td>2004</td>
<td>280</td>
<td>Conventional</td>
<td>CH</td>
<td>Antibiotic</td>
<td>DPPC monolayer</td>
</tr>
<tr>
<td>Prion protein fragment (PrP118–135)</td>
<td>2012</td>
<td>68, 281</td>
<td>Conventional</td>
<td>Amide-I/Amide-II/CH/OH</td>
<td>Prion protein fragment</td>
<td>Lipid bilayer</td>
</tr>
<tr>
<td>Rhodopsin</td>
<td>2011</td>
<td>238</td>
<td>Chiral</td>
<td>NH/Amide-I</td>
<td>Receptor protein</td>
<td>Liquid–air</td>
</tr>
<tr>
<td>rAPPLE</td>
<td>2011</td>
<td>238</td>
<td>Chiral</td>
<td>NH/Amide-I</td>
<td>Rat islet amyloid polypeptide</td>
<td>Liquid–air</td>
</tr>
<tr>
<td>RNase A</td>
<td>2019</td>
<td>206</td>
<td>Conventional</td>
<td>CH/OH</td>
<td>Enzyme</td>
<td>Liquid–air</td>
</tr>
<tr>
<td>Sp1 zinc finger (ZnF)</td>
<td>2018</td>
<td>282</td>
<td>Chiral</td>
<td>Amide-I/CH</td>
<td>Model protein</td>
<td>Liquid–air</td>
</tr>
<tr>
<td>Tachypleis I</td>
<td>2007</td>
<td>50, 133, 218, 223, 232, 283</td>
<td>Chiral and conventional</td>
<td>Amide-I/CH</td>
<td>Model peptide</td>
<td>Solid–liquid, PS</td>
</tr>
<tr>
<td>Thermomyces lanuginosus lipase (TLL)</td>
<td>2008</td>
<td>284</td>
<td>Conventional</td>
<td>CH</td>
<td>Enzyme</td>
<td>Lipid layer, liquid–air</td>
</tr>
<tr>
<td>Ubiquitin</td>
<td>2005</td>
<td>20, 131, 187</td>
<td>Conventional</td>
<td>CH/OH</td>
<td>Regulatory protein</td>
<td>Solid–liquid</td>
</tr>
</tbody>
</table>
anticancer agents. By combining achiral and chiral SFG measurements with molecular dynamics simulations and the type of spectral calculations discussed in section 2.3.2, the authors found that chiral inversion affects the backbone and side-chain CH orientation, that the structure of the peptides is determined by the side chains (i.e., not by the chirality of the backbone peptide groups), and that the presence of the chiral amide-I signal is dependent on the formation of extended β-sheets (also see section 2.1.2.2).

Lu et al. recently showed that when the positively charged lysine is replaced by negatively charged glutamic acids, secondary structures similar to LK peptides can be enforced at interfaces.162,170 An important role of LK peptides has also been to provide insight into the use of SFG as a tool to interrogate interfacial folding using amide-I spectra. These studies are well-summarized in previous reviews16,18 and will not be discussed here. Since LK peptides are well-controlled and understood model systems, they are often used to benchmark new spectroscopic methods. Ye et al. pioneered the use of other amide regions for a more accurate assignment of protein backbone modes.74 The authors reported the first (very weak) SFG spectra measured in the amide-III region on peptides in a model cell membrane in a study of the structure of LKα14 in a negatively charged POPG lipid bilayer. The single peak in the amide-III region (at ∼1275 cm$^{-1}$) was assigned to the α-helical structure, and from its frequency a Ramachandran dihedral angle ($\psi$) of $-50^\circ$ was deduced. Although the amide-I signal was stronger than the amide-III signal, the formation of a random coil could not be ruled out solely on the basis of the spectra in the amide-I region. The authors concluded from the absence of a chiral amide-I signal (see section 2.1.2.2) that the structure did not contain any β-sheets.

Schm{"u}ser et al. used LK peptides at the air−water interface to develop methods to determine absolute protein orientation by combining phase-resolved amide-I SFG spectroscopy with spectral calculations and MD simulations. The authors found that LK peptides are oriented almost parallel to the air−water interface, with a tilt angle of the long peptide axis of $82^\circ$ with respect to the surface normal. The results show that phase-resolved SFG can reliably differentiate absolute orientations even when the transition dipole moment is almost parallel to the surface.61 It is worth noting that rotating the molecular frame by $180^\circ$ would result in a conventional SFG spectrum that is similar to the nonrotated one, while only the nonrotated frame having the N-terminus pointing into the water reproduced the experimentally measured phase-resolved spectrum.

3.1.1.2. Model Peptides for Amide-III Spectroscopy. To demonstrate the utility of amide-III SFG signals in identifying the secondary structures of peptides, Ye et al. investigated peptides with varying helicity (as deduced from independent NMR experiments reported in the literature), including mastoparan X, cecropin P1, pardaxin, and melittin.74 The amide-III spectral region of these peptides features two peaks. The authors suggested that the ratio of these two peaks can be used as an indicator of the ratio of helical to random structure in peptides. It must be noted that if other structures besides random coil and α-helix exist for a given peptide, this analysis fails to provide a useful quantitative ratio between the random coil and α-helix. The authors utilized the observed linear relationship to determine the interfacial conformational transition in situ by studying the interactions between pardaxin and different DMPC/DMPG lipid bilayers. By mixing the

<table>
<thead>
<tr>
<th>Year</th>
<th>Ref(s)</th>
<th>Technique</th>
<th>Region(s)</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>2018</td>
<td>163</td>
<td>conventional amide-I</td>
<td>amide-I/amide-A</td>
<td>model peptide lipid bilayer</td>
</tr>
<tr>
<td>2009</td>
<td>268</td>
<td>conventional CH</td>
<td>amide-I</td>
<td>model peptide lipid bilayer</td>
</tr>
<tr>
<td>2018</td>
<td>286−287</td>
<td>conventional amide-I/C$\equiv$O</td>
<td>amide-I/C$\equiv$O</td>
<td>antimicrobial peptide model membrane</td>
</tr>
<tr>
<td>2017</td>
<td>288</td>
<td>conventional</td>
<td>amide-I</td>
<td>model peptide lipid bilayer</td>
</tr>
</tbody>
</table>

*a In the case of multiple studies, the year of the earliest publication is listed.*
negatively and positively charged lipids in different ratios, the authors followed the interfacial conformational transition of pardaxin using its SFG signal in the amide-III region in situ. Their results indicated that pardaxin undergoes a conformational change from the random coil structure to the α-helical structure as the fraction of negatively charged lipids in the mixed bilayer is increased. This methodology has for example also been applied to determine the α-helix/random coil ratio of the 25 residue long M2 transmembrane domain of the influenza A virus. In that study, Liu et al. used the amide-III peak ratio to assess the amount of α-helical and random coil structure and the amide-I signal to assess the peptide orientation. The results revealed the pH-dependent structural plasticity of the peptide when it interacts with different types of lipid membranes. The authors hypothesized that the variation in the observed responses to different lipid compositions might provide important clues into the gating mechanism and for the design of new anti-influenza drugs.

3.1.1.3. Model Peptides for Isotope Labeling. After an initial isotope labeling SFG study using deuterium-labeled proteins by the group of Chen,99 his group, in collaboration with the Zanni group, published an SFG study of a site-specifically $^{13}$C=16O-labeled peptide, which exhibited a much stronger isotope-induced frequency shift.155 Using the formalisms described in sections 2.3.1 and 2.3.3, the authors showed that the orientation of the predominantly α-helical 18 amino acid model peptide (and antibiotic) ovispirin-1 could not be fully determined using either the one isotope label or the unlabeled amide-I band alone but that the combination of the two spectra allowed the determination of the tilt and twist angle of the ovispirin helix at a solid–liquid interface. In a follow-up study, Ding et al. $^{13}$C=16O-labeled 10 different backbone amide groups and obtained a data set that provides valuable information via both the site-specific spectral width (which reports on the local hydration and thus the location of the labeled amino acid within or outside the lipid bilayer to which it is adsorbed) and the interferences between the labeled and unlabeled modes (which allow for a very precise orientation determination).

3.1.1.4. GALA Peptide. Regulating the insertion of peptides and proteins into cell membranes by adjusting the pH is a promising method for drug delivery, e.g., to cancer cells, as these cells typically maintain lower pH values than healthy cells,228 or to transport drugs through endosomal membranes into the cytosol, as endosomes also have a lower pH.157,297 For example, the EALA-repetition peptide GALA adopts an α-helical structure at pH < 6 that can insert into cell membranes, while at higher pH values the helical structure is destabilized because of deprotonation of the glutamic acid side chains. To assess whether this pH dependence of the GALA conformation is affected by the membrane geometry, Schach et al. studied the interaction of GALA with giant unilamellar vesicles.228 On the basis of a time-lapse leakage test of a fluorescent dye through vesicles at different pH values, the authors concluded that the curvature does not affect the functionality of GALA. When GALA was injected into the subphase of the DPPC monolayer at pH < 5, an immediate increase in the surface pressure was detected that indicated the surface activity of GALA and its incorporation into the lipid monolayer. Simultaneous SFG spectra in the amide-I region showed a peak at 1650 cm$^{-1}$ corresponding to the α-helical state of GALA. A subsequent increase in the pH resulted in the disappearance of the SFG signal in the amide-I region due to the destabilization of the helical structure. A comparable membrane-disrupting pH-dependent behavior of GALA was observed with SFG and fluorescence assays when it was cysteine-linked to a gold surface, although the high-pH structure appeared to contain β-sheet structure instead of random coils.97 Near the air–water interface, however, a comparison of experimental and calculated spectra based on MD simulations using the methodology described in section 2.3.4 showed that GALA remains in a mainly α-helical state irrespective of the pH.157

A recent systematic study of the pH response of GALA in lipid bilayers with different hydrophobic lengths and lipid headgroups by Hu et al. revealed that the reversible activation of pH-responsive GALA in a model cell membrane relies sensitively on the nature of the lipid with which it interacts.129 By recording amide-I and amide-III SFG spectra, the authors were able to differentiate between random coil and α-helical structures (see section 3.1.1.2 for details on this methodology).

3.1.2. Model Proteins. Initial studies performed on model proteins have been excellently reviewed in refs 20–22 and 177. In 2014, Engelhardt et al. presented the SFG spectra of bovine serum albumin (BSA), β-lactoglobulin (BLG), and lysozyme (LSZ) for different solution pH values over a wide range of frequencies (950 to 3800 cm$^{-1}$).79 As can be seen in Figure 19, this wide spectral range is divided into smaller frequency ranges: (a) fingerprint bands, (b) carbohydrate (RCOO$^{-}$) symmetric stretching vibrations, (c) amide-I bands, (d) pure nonresonant contributions, and (e) CH and OH stretching bands. As highlighted in Figure 19e, the intensities and shapes of the SFG spectra in the OH stretching region of pure water are very different from those of the various investigated proteins at the air–liquid interface. The sharp peak at ~3740 cm$^{-1}$, which is
attributed to non-hydrogen-bonded free OH, almost completely disappears when the proteins are adsorbed. Also, the intensity of the hydrogen-bonded water band (3100 to 3600 cm\(^{-1}\)) strongly varies depending on the net charge at the charged air–liquid interface. A higher net charge results in a stronger ordering of the interfacial water molecules, with their dipole moments pointing toward air (in the case of negatively charged interfaces) or toward bulk water (in the case of positively charged interfaces). Therefore, the intensity of the SFG signal in the OH stretching region can be attributed to the amount of charge, which has the minimum value at the isoelectric point of the surface (IEPS). The IEPSs were determined to be 5.5, 5.0, and 10.0 for BSA, BLG, and LSZ, respectively, which is in agreement with the \(\zeta\) potential values measured for the associated bulk protein solutions. It should be noted that higher-order nonlinear effects, such as \(\chi^{(3)}\), can also affect the spectral shape at charged surfaces.

To further investigate the deterministic role of electrostatic interactions in the functionality, colloidal stability, and adsorption of proteins on different surfaces and interfaces, Guckeisen et al. studied the isoelectric points of proteins at the air–liquid interface in comparison to those in the bulk solution for a wider range of proteins with different sizes, hydrophobicities, and structural elements using SFG, ellipsometry, and \(\zeta\) potential measurements in a complementary fashion. The pH-dependent results obtained from ellipsometry and \(\zeta\) potential measurements are measures of the thickness of the protein layer at the air–liquid interface and the electrophoretic mobility of the investigated proteins, respectively. When the net charge of the proteins approaches the minimum, the repulsive electrostatic forces diminish, resulting in the formation of a thicker protein film at the air–liquid interface. By comparing the results of SFG, ellipsometry, and \(\zeta\) potential measurements, the authors concluded that the isoelectric points of the investigated proteins at the air–liquid interface do not differ from those in the bulk solution, irrespective of the absolute isoelectric point of the proteins, as depicted in Figure 20. Nevertheless, Devineau et al. found that the isoelectric point of hemoglobin at the air–liquid interface is significantly lower (by about one pH unit) than in the bulk. The authors hypothesized that the structural change of the protein upon adsorption is likely to be responsible for the alteration of the charge state of hemoglobin at the air–liquid interface. The application of phase-resolved SFG (the subject of section 2.2.1.2) allowed them to have a higher sensitivity toward the net orientation of water molecules and thus to determine the interfacial isoelectric point of hemoglobin with improved accuracy.

Besides pH variations, different mixtures of oppositely charged proteins and surfactants with different ratios can also alter the net charge at the air–liquid interface. For instance, as shown in Figure 21, at a CTAB/BLG ratio below 0.07 at pH 8.7, the surface charge is dominated by negatively charged BLG. When the CTAB/BLG ratio approaches 0.33, the SFG signal intensity in the OH stretching region is minimal, indicating charge cancellation between negatively charged BLG and the cationic CTAB surfactant. With an increased fraction of CTAB in the mixture, the surface charge becomes increasingly positive.

The SFG signal at around 3070 cm\(^{-1}\) in Figure 21 originates from an aromatic mode of the protein. The sign of this signal changes from negative to positive as the CTAB/BLG ratio is increased. This inversion of the sign of the peak is due to the destructive and constructive interference, respectively, between the aromatic CH vibrations of the protein and the OH stretching vibrations of the water molecules for CTAB/BLG ratios below and above 0.33. Although determining the absolute orientation of the molecules using SFG often requires a maximum entropy analysis or phase-resolved SFG measurements (see section 2.2.1.2), in particular cases such as the CTAB/BLG mixture, induced changes in the sign of the aromatic mode are a clear indication of a relative reorientation of the water molecules with respect to the proteins. Similarly, the interaction of BLG with an anionic surfactant, sodium dodecyl sulfate (SDS), was investigated at two different pH values, 3.8 and 6.7. Changing the pH from 3.8 to 6.7 allows tuning of the electrostatic interactions from an attractive to a repulsive regime. At these pH values, BLG is positively and negatively charged, respectively. Therefore, at low pH, the total amount of charge at the air–liquid interface can be regulated by changing the ratio between
the protein and surfactant. Complementary isothermal titration calorimetry and ellipsometry measurements were performed to determine the complex formation in the bulk and the interfacial layer thickness, respectively. In this case, SFG spectra acquired over a wide frequency range, including the fingerprint, amide-I, CH, and OH stretching regions allowed an assessment of the degree of ordering of the proteins as well as the electrostatic charges at the air−liquid interface as the ratio between the surfactant and protein varies.299

Besides the electrostatic forces, interfacial water layers can also regulate protein adsorption. Nagasawa et al. investigated the role of interfacial water molecules on polymer brushes with different structures using SFG and quartz-crystal microbalance with energy dissipation monitoring (QCM-D).300 The authors fabricated surfaces with various physicochemical properties using high-density polymer brushes with a well-defined structure and quantified the mass of the adsorbed proteins and their hydration state using QCM-D. As expected, protein adsorption on zwitterionic brush polymers with nearly zero surface charge was strongly suppressed because of the weak electrostatic interactions between the brush polymers and the proteins. In the case of a neutral buffer solution with a cationic brush polymer (poly(AEMA)) and an anionic brush polymer (poly(MPS)), adsorption of negatively charged BSA and positively charged LSZ was dominant, respectively. On all of the zwitterionic polymer brushes, the observed SFG intensities in the OH-stretching region were similar, whereas the oppositely charged polymer brushes exhibited very different SFG spectra since the surface charges affect the interfacial water structure. More importantly, the position of the OH stretching peak, an indication of the strength of hydrogen bonding between interfacial water molecules and the surrounding molecules,301 showed large variations depending on the charge of the polymer brushes. Zwitterionic polymers exhibited the lowest OH stretching frequency, indicating the strongest hydrogen bonds, while interfacial water molecules on poly(AEMA) showed the highest OH stretching frequency, indicating weak hydrogen bonding. The relationship between the amount of adsorbed proteins, measured with QCM-D, and the OH stretching peak position, measured with SFG, indicated that for zwitterionic brush polymers, the stronger hydrogen bonding could be related to the smaller amount of adsorbed protein. A more strongly hydrogen-bonded network of interfacial water was therefore hypothesized to act as a barrier toward the approaching proteins. For charged brush polymers, where the electrostatic interactions between oppositely charged proteins and polymers are the dominant adsorption mechanism, no relationship between the position of the OH stretching peak and the amount of adsorbed protein was found. It should be pointed out that the OH stretching peak positions in this study were measured in the absence of proteins in solution which may affect the hydrogen-bonding strength within the interfacial water network.

Okuno and Ishibashi178,249 extended the application of phase-resolved SFG measurements from peptides161 to proteins by measuring chiral and achiral SFG spectra at the air−protein-solution interface for model proteins with different secondary structures. The authors examined the relationship between the secondary structure of proteins and their stability and denaturation at the air−liquid interface in a setup comparable to the one depicted in Figure 8b. For BSA, pepsin, and concanavalin, negative signs of $\text{Im}[\chi^{(2)}]$ are observed for the

Figure 21. (a) SFG spectra in the SSP polarization configuration of CTAB/BLG mixtures at different molar ratios. The pH of the solution was set to 8.7, where BLG is negatively charged. (b) Aromatic stretching band at ~3070 cm$^{-1}$, which interferes destructively or constructively with the OH stretching signal and appears as a negative or positive peak. The sign of this peak can be used to determine the orientation of water molecules adjacent to the protein layer with respect to the proteins. Reproduced with permission from ref 79. Copyright 2014 Elsevier.

Figure 22. (left) $\text{Im}[\chi^{(2)}]$ spectra of $\alpha$-lactalbumin for pH values ranging from 3 to 9 at the air−liquid interface. The $\text{Im}[\chi^{(2)}]$ spectrum of water is shown in orange as a reference. (right) Schematic representation of the water orientation adjacent to differently charged proteins at different solution pH values. Reproduced with permission from ref 180. Copyright 2016 Royal Society of Chemistry.
symmetric CH$_3$ stretching, Fermi resonance, and aromatic CH stretching modes, accompanied by a broad positive signal from the OH stretching vibrations. These observations suggest that unlike the bulk phase, where the proteins have a tertiary structure with the methyl groups folded inward, at the air−liquid interface the majority of the methyl groups in all of these proteins, despite their different secondary structures, are oriented with the protons toward the air. Also, the water molecules at the air−liquid interface adopt the proton-up orientation, as the sign of Im[χ(2)] for the OH stretching mode is positive. Similarly, the phase of the achiral signal in the amide-I region for all of the investigated proteins showed a positive sign of Im[χ(2)]. However, distinct differences were observed in the phase-resolved chiral amide-I and N−H stretching SFG signals of the proteins. Because of the increased signal-to-noise ratio of phase-resolved measurements, the authors also observed in situ the interface-induced conformational change and denaturation of α-chymotrypsin at the air−liquid interface as a function of time. The authors proposed that thermal instability and reorientation of water molecules around the protein molecules at the air−liquid interface are responsible for the structural change that α-chymotrypsin undergoes at the air−liquid interface.

Strazdaite et al. studied the denaturation mechanism of α-lactalbumin by comparing conventional SFG spectra with phase-resolved SFG spectra measured in a setup similar to that shown in Figure 8d. The authors showed that replacement of the water molecules surrounding proteins with urea, a small polarizable molecule with similar properties as water, results in the denaturation of α-lactalbumin. As depicted in Figure 22, the net orientation of water molecules follows the net charge of the α-lactalbumin at different pH values, where negative and positive signs of the Im[χ(2)] signal in the OH stretching region indicate downward and upward orientation of water molecules, respectively. The negative sign of Im[χ(2)] for the methyl symmetric stretching mode (CH$_3$sym at ~2880 cm$^{-1}$) and Fermi-resonant mode (CH$_3$FR at ~2950 cm$^{-1}$) and the positive sign of Im[χ(2)] for the methyl antisymmetric stretching mode (CH$_3$asym at ~2990 cm$^{-1}$) indicate that the methyl groups are oriented toward the air. A similar orientation of the methyl groups was observed for the ice-binding site of the antifreezing protein DAFP-1 at the air−liquid interface. For an additional peak from the NH-stretch vibrations of urea, the

Figure 23. SFG spectra in the SSP configuration of WLBU2 interacting with (A) a bacteria-like (negatively charged) DMPG lipid monolayer and (B) a (zwitterionic) mammalian-like DPPC lipid monolayer. WLBU2 adopts α-helical and β-strand structures upon interacting with DMPG and DPPC, respectively, as depicted at the bottom. Adapted with permission from ref 288. Copyright 2017 American Vacuum Society.
Im[χ(2)] spectra of α-lactalbumin in urea solutions follow the same trend as depicted in Figure 22, which agrees with the expected dipolar orientation. Phase-resolved amide-I spectra show that changing the pH of the solution affects only the orientation of water or urea molecules, while the amide groups of the protein’s backbone have an overall C==O orientation that points into the bulk independent of the net charge of the surface. Therefore, the authors hypothesized that protein denaturation by urea does not require hydrogen bonding between urea and the protein’s backbone and that a urea-induced change in the water structure causes protein denaturation.

3.2. Antimicrobial Peptides

As mentioned above, minor variations in the sequences of peptides and proteins strongly influence their interactions with cell membranes. The exact mechanism of how peptides and proteins target and bind to cell membranes is not very well understood yet.

3.2.1. WLBU2. Golbek et al. identified the mechanism behind the selectivity of the engineered antimicrobial peptide WLBU2 for zwitterionic DPPC lipid monolayers (a model for mammalian cell membranes) versus anionic DMPG lipid monolayers (a model for bacterial cell membranes). In solutions with physiological pH values, WLBU2 is a cationic amphiphilic peptide with a high positive net charge. Therefore, its interaction with differently charged lipid layers may result in different secondary structures. Surface-pressure measurements showed that the interaction of WLBU2 with DMPG is stronger than that with DPPC. Subsequently, the authors used SFG spectroscopy in the amide-I region to probe the peptide structure and in the ester C==O stretch (~1735 cm⁻¹) and C–D stretch (2000–2300 cm⁻¹) regions to probe the lipid conformation. As depicted in Figure 23, the amide-I region contains a single peak at 1651 cm⁻¹ that is consistent with α-helical structure. In the presence of bacterium-like membranes, and on the basis of the (absence of) changes in the C–D and C==O stretch regions, it appears that the peptide inserts itself into the membrane upon adsorption without strongly affecting the lipid order. However, in the presence of mammal-like membranes, the amide-I spectrum is composed of two peaks at 1642 and 1678 cm⁻¹, which are typically assigned to β-sheet structure, and the spectral changes in the other two regions indicate that in this case WLBU2 adsorbs superficially to the lipid monolayer, thereby decreasing the lipid headgroup hydration, which results in increased order of the lipid chains. The authors note that WLBU2 is a promising peptide that might aid in the treatment of bacterial sepsis, in which it is desirable to capture endotoxins intact and increase their size by binding to them rather than to disrupt them into smaller fragments that are more difficult to clear from the bloodstream, e.g., by a hemoperfusion device.

3.2.2. Cecropin P1. The role of surface charges in the interaction of antimicrobial peptides with lipids was also reported by Han et al. The authors observed that, governed by electrostatic interactions, the cationic cecropin P1 peptide inserts into a negatively charged phosphatidylglycerol (PG) bilayer with an α-helical secondary structure, while no significant interactions were observed with a mammalian cell-membrane model based on zwitterionic 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC) bilayers. Upon surface immobilization of cecropin P1 (also see section 3.2.3 for other surface-immobilized antimicrobial peptides) and when in contact with lipid bilayers, the peptides changed their orientation from perpendicular to parallel to the interface and/or denatured in the presence of bacterium-like model membranes, losing the ability to insert into the membrane, while again no interaction was observed for mammal-like model membranes.

3.2.3. Surface-Immobilized Antimicrobial Peptides. Xiao et al. investigated the secondary structure and antimicrobial activity of chemically immobilized cecropin–melittin hybrid peptides on a chemical vapor deposition-prepared polymer surface and on a silane-based SAM surface. The authors found that the peptides adopt an α-helical structure on both surfaces and that the stability of the polymer-based antimicrobial peptide coating is better than the SAM-based coating. The antibacterial activity of the peptides was associated with the charge of their orientation upon interaction with bacteria, i.e., the bacterium-induced peptide bending inflicts damage to the bacteria. This behavior differs from the surface-immobilized behavior of the cecropin P1 peptide (see section 3.2.2).

3.3. Membrane-Associated Proteins

The surface sensitivity of SFG has also been used to probe the interactions of proteins and peptides with lipid membrane surfaces. The Chen lab has been pioneering the use of SFG to probe the structure of antimicrobials and membrane proteins at supported bilayer surfaces. As this work has already been the subject of several excellent reviews, we focus here on recent additions to the field.

Mauri et al. compared bovine and human insulin adsorption at a 1,2-dipalmitoyl-sn-glycero-3-phosphoglycerol (DPPG) model lipid layer by examining their SFG spectra in the amide-I region. In their studies, the SFG peak at 1730 cm⁻¹ related to lipid C==O modes indicated the presence of an ordered DPPG monolayer. Upon injection of bovine insulin, a new peak at 1653 cm⁻¹ appeared that was assigned to α-helical domains of insulin. The intensity of this peak increased over time after injection and reached a maximum level after 20 min. With the addition of HCl to the subphase and the consequent decrease in the pH of the solution from 4 to 2, no observable change in the intensity of the 1653 cm⁻¹ peak was detected. At pH 2, insulin molecules in bulk solution are expected to be mainly monomeric, while, at higher pH, dimers, trimers, and hexamers are formed. The absence of a change in the SFG spectrum upon lowering the pH from 4 to 2 thus seems to indicate that this structural transition is absent at the lipid surface, probably due to stabilizing interactions with the DPPG monolayer. The minor mutation in the structure of insulin in going from bovine insulin to human insulin resulted in the disappearance of the insulin SFG signal in the amide-I region. Complementary X-ray photoelectron spectroscopy measurements on human insulin deposited on gold using the Langmuir–Schaefer method confirmed the presence of insulin at the lipid membrane. Therefore, the absence of an SFG signal for human insulin was attributed to its disordered binding geometry, which leads to SFG signal cancellation at the lipid–water interface.

Nguyen employed electronically enhanced chiral SFG to study the interaction between the positively charged cytochrome c and negatively charged cardiolipin under physiological conditions. Electronically enhanced chiral SFG is specifically beneficial when studying nonaggregated proteins at interfaces with low interfacial protein populations and under circumstances that result in small Fresnel factors for chiral polarization combinations. Cytochrome c is a heme protein and exhibits a characteristic absorption band at 530 nm, and thus, with 532 nm as the visible beam the chiral SFG signal can be
electronically enhanced. Native cytochrome c does not generate any chiral signal, while upon interaction with cardiolipin, a strong chiral signal due to the formation of β-sheets was observed. Thus, the newly formed β-sheets should be associated with the amino acid residues that are coordinated with the heme structure. A reversible shift in the relative intensities of the B1 and B2 modes was observed in the chiral amide-I signal as the pH of the solution was changed. This pH dependence of the signal was associated with changes in the β-sheet orientation.

Roeters et al. demonstrated that spectral calculations based on an excitonic Hamiltonian (see section 2.3.3) can be used to determine the orientation of cholera toxin B subunit, a 55 kDa homopentameric protein, at the lipid−water interface, as depicted in Figure 24.60 The B subunit binds strongly to ganglioside lipids and was previously shown to do so with a small angle between its central axis and the surface normal when injected in the subphase under a ganglioside-containing lipid monolayer.310,311 The authors measured IR, Raman, and SFG spectra and compared them with the calculated spectra. The IR and Raman spectra indicated that the structure was similar to the X-ray-resolved structure, and from a least-squares fit to the SFG spectra, the angle between the protein symmetry axis and the surface normal was found to be 6 ± 17°.

This benchmarking study was followed by an investigation of the structure of a thylakoid membrane fusion protein in chloroplasts and cyanobacteria.158 In this study, the authors reveal the binding mode and function of a protein called IM30 that is involved in the biogenesis of the thylakoid membrane network, which hosts the photosynthetic machinery in plants. The thylakoid membrane is very dynamic since plant cells constantly adapt the extent of photosynthesis to factors such as light intensity and metabolic factors, and IM30 was found to play an important role in these processes. SFG spectra were used to determine the orientation of IM30 on a model thylakoid membrane. The authors calculated SFG spectra for two scenarios that were plausible on the basis of experience with large membrane fusion proteins: a carpet state and a ring state. Figure 25 shows the two configurations along with the experimental SFG spectrum and spectra calculated for the two scenarios. It is clear that the ring model fits the experimental data better than the carpet model.

3.4. Fibrils and Fibers

3.4.1. Amyloids. Aggregation of various (often bulk-disordered) proteins into amyloid fibrils has been associated with over 50 neurodegenerative diseases, such as Alzheimer’s disease, Parkinson’s disease, Huntington’s disease, prion diseases, and type-II diabetes. In many cases, interactions with lipid membranes and hydrophobic surfaces are thought to catalyze amyloid formation. Therefore, recording the transformation from the native form to the aggregated form of proteins near such interfaces is key in understanding the molecular mechanism of these protein-induced diseases. SFG is well-suited for this task, as demonstrated by the studies discussed below. Another feature of amyloid studies indicated by the studies discussed below is that it is challenging to reproduce the kinetics of the concentrations and structures of the different aggregation species if one is not very careful about using the exact same experimental conditions (notably, the protein concentration,312−314 salt concentration,315−318 pH,317 temperature,319−321 solvent,322−324 and type of interface in

![Figure 24](image-url)  
**Figure 24.** (left) SFG spectrum of 43 nM CTB under a 1:9 GM1/d75-DPPC mixed monolayer with a surface pressure of 30 mN/m at pH 7.4 in the SSP polarization combination. The calculated spectrum (in blue) shows a good agreement with the experimentally measured spectrum (in red) for θ = 6 ± 17°. (right) Definition of the molecular axis of CTB, where the z axis of the protein is defined as the symmetry axis of the homopentamer. Reproduced from ref 46. Copyright 2013 American Chemical Society.

![Figure 25](image-url)  
**Figure 25.** Experimental (circles) SFG spectrum of IM30 bound to a monolayer of mixed MGDG/PG at the air−water interface with the SSP polarization combination. The simulated scenarios where the large membrane fusion protein adopts a ring or carpet state are shown alongside their corresponding calculated SFG spectra (lines). Adapted with permission from ref 158. Copyright 2015 Macmillan Publishers Limited.
contact with the protein solution \(^{321,325-327}\). Even when strictly the same experimental conditions are used, the kinetics observed in (unseeded) amyloid aggregation experiments are subject to considerable variation because amyloid aggregation is a nucleation-based phenomenon. However, it is expected that the intermediates and aggregates that are formed during the process are reproducible, and the functional form of the aggregation curves can be analyzed with kinetic models, which provides molecular information on the processes.\(^{325,328,329}\) A 2D-SFG study of the amyloid aggregation of the FGAIL segment of hIAPP\(^{100}\) is discussed in section 3.9.2.

3.4.1. Amylin. The amyloid aggregation of amylin (also known as islet amyloid polypeptide, IAPP) is thought to be related to type-II diabetes. It is the amyloid-forming protein that has been studied most frequently with SFG spectroscopy.

In 2010, Fu et al. compared the chiral and achiral SFG spectra of human IAPP (hIAPP) and rat IAPP (rIAPP) in the presence and absence of a monolayer of the negatively charged phospholipid dipalmitoylphosphoglycerol (DPPG).\(^{51}\) The chiral (PSP) spectra that the authors recorded in the amide-I region (Figure 26) showed that only when DPPG and hIAPP were incubated for 10 h a chiral peak around 1620 cm\(^{-1}\) appeared (assigned to amyloid fibrils), while no chiral amide-I signal was observed for incubation with rIAPP. This chiral signal was also absent when hIAPP was incubated in the absence of the DPPG monolayer at the air–water interface. When the authors recorded the SSP and PSP time dependences of the phospholipid-induced aggregation of hIAPP, Fu et al. observed a shift in the amide-I peak position of the achiral spectrum from \(\sim 1645\) to \(\sim 1660\) cm\(^{-1}\) and an increase in the chiral amide-I signal intensity after \(\sim 7\) h of incubation with DPPG.

In a follow-up study, the Yan group also measured the chiral N–H signal (\(\sim 3100–3400\) cm\(^{-1}\)) of hIAPP during the same process and observed that it increases up to \(\sim 3\) h after the addition of the lipid monolayer and then disappears after \(\sim 10\) h.\(^{238}\) A first empirical interpretation of the data based on

![Figure 26. Overview of SFG studies of aggregating hIAPP. (a, b) Achiral (SSP) and chiral (PSP) amide-I SFG spectra of hIAPP at the DPPG–D\(_2\)O interface, which are thought to indicate the appearance of amyloid structure at the interface. Reproduced from ref 61. Copyright 2010 American Chemical Society. (c, d) Chiral (PSP) amide-I and N–H stretch spectra of hIAPP at the DPPG–water and air–glass interfaces that are thought to indicate the strong dependence of SFG signals on the protein’s orientation. Reproduced from ref 239. Copyright 2015 American Chemical Society. (e) Achiral (SSP) and chiral (PSP) amide-I and II SFG spectra of the amyloid aggregation of hIAPP on DPPG lipids and the derived hypothesis in which no \(\alpha\)-helical intermediate is present,\(^{68}\) as opposed to what was hypothesized in previous papers.\(^{32,61,238}\) Potentially the different hypotheses stem from different experimental circumstances, as these are known to heavily influence the structural kinetics of amyloid aggregation. Reproduced from ref 68. Copyright 2019 American Chemical Society.](https://dx.doi.org/10.1021/acs.chemrev.9b00410)
reference measurements with model proteins and peptides led to the conclusion that only helical folds produce chiral N-H signals. By comparing the time evolution of the chiral spectral features in the amide-I and N-H stretching spectral regions, the authors correlated the observed chiral N-H signal to the orientation of α-helical intermediates (observed with the chiral N-H stretch signal) that subsequently convert into amyloid parallel β-sheets (observed with the chiral amide-I signal). In a later study, the same group analyzed the chiral SFG data with ab initio quantum-chemistry calculations (see section 2.3.2 for details on the employed computational method) to determine the orientation of the hIAPP aggregates at the lipid-water interface on the basis of the amide-I SFG spectrum and found that the experimental data could be reproduced with an angle of 48° between the surface and the axis parallel to the β-strands of a fibrillar structure composed of two intermolecularly hydrogen-bonded strands in the fibril direction. An MD study following the orientational and structural developments of trimers and tetramers inserted in DPPG monolayers yielded structures and orientations (40° between the surface and the axis parallel to the β-strands) that could also reproduce the chiral amide-I data. In 2015, Fu et al. performed a comparable analysis of the N-H stretching region, which suggested that the chiral SFG response of parallel β-sheets is highly influenced by the orientation at the surface and that specific orientations of fibrillar hIAPP also lead to a strong chiral signal in the N-H stretching region, assuming that the observed signal of hIAPP on a glass slide is only from its parallel β-sheet form. On the basis of this new observation, the evolution of the chiral N-H signal could also be explained by a changing orientation of the β-sheet structures, rather than by a transient helical state. Nevertheless, this observation does not rule out the original explanation presented by Fu et al., who assigned the intermediate spectra to α-helical species. This interpretation was substantiated in 2015 by a tip-enhanced Raman spectroscopy study revealing that the hIAPP fibril surface formed near a DPPG monolayer is highly heterogeneous and mainly composed of α-helical and/or disordered structure, while the core consists of parallel β-sheets.

In 2019, Tan et al. used a new SFG spectrometer with fast detection capability to follow hIAPP aggregation at membrane surfaces using chiral and achiral SFG spectra in the amide-I, -II, and -III regions and found no evidence for helical intermediates but rather for direct assembly of disordered solution-state hIAPP into β-strands and fibrils. Possibly the observed differences with respect to the previous studies by the Yan group are related to differences in the experimental circumstances (the Yan group used 10 mM pH 7.4 phosphate buffer, a hIAPP bulk concentration of 4 μg/mL, and a monolayer of DPPG lipids, while the Ye group performed the measurements without a buffer, with a pH of approximately 6.2, at an hIAPP bulk concentration of 0.1 mg/mL and a bilayer of DPPG lipids), as ionic strength, pH, protein concentration, and differences in the interface are known to strongly influence amyloid aggregation processes (see introduction of section 3.4.1 for references).

The interaction between aggregating hIAPP and potential drugs has also been studied with SFG. In 2012, Engel et al. managed to reproduce the amide-I SSP spectra of the Yan group and to show—by combining the SFG results with Fourier transform infrared spectroscopy (FTIR) and atomic force microscopy (AFM) measurements—that the amyloid inhibitor EGCG, which is very potent in the bulk, is much less effective in inhibiting amyloid growth near the phospholipid interface and that it is unable to disaggregate the fibrils formed at the interface, as opposed to its behavior in bulk. Already during this study the researchers observed heterogeneity, which they revealed at a microscopic level using tip-enhanced Raman spectroscopy in a follow-up paper. It is therefore important to account for sample heterogeneity when studying amyloid formation with SFG, e.g., by rotating the sample during the acquisition (see also section 2.2.2.1).

### 3.4.1.2. Amyloid-β

Bellucci et al. combined SFG with metadynamics MD simulations to determine the structure of the 16–22 segment (KLVFFAE) of amyloid-β at gold surfaces in solution, as it is known that inorganic surfaces can both inhibit and accelerate amyloid aggregation. The authors extracted snapshots from the simulation, calculated SFG spectra using an excitonic Hamiltonian model (see section 2.3.3), and compared them with experimental SFG amide-I data, revealing an inhibition mechanism that involves details of the microscopic peptide–surface interaction rather than generic properties such as peptide sequestration from the solution.

### 3.4.1.3. β-Lactoglobulin

In 2011, vandenAkker et al. studied the amyloid aggregation of the model protein β-lactoglobulin, which is known to form amyloid fibrils at low pH and high temperature. The authors prepared both long, straight and short, wormlike β-lactoglobulin amyloid fibrils by varying the protein concentration and incubation time and determined the morphology and persistence length by AFM and the molecular conformation using amide-I SFG spectroscopy. They found that long fibrils with near-100% β-sheet content have a 40-times higher persistence length than short, wormlike fibrils with a β-sheet content below 80%.

### 3.4.1.4. Prion Protein

The amyloid aggregation of prion proteins is related to Creutzfeldt–Jakob disease, during which the protein converts from a mainly α-helical form to a β-sheet form. Li et al. studied the 118–135 segment of the prion protein, which shares homology with the C-terminal domain of the Alzheimer’s β-amyloid peptide, by measuring achiral and chiral amide-I SFG spectra. They found a concentration-dependent structure when the fragment was incubated near a POPG bilayer deposited on a CaF₂ prism in solution (see section 2.2.2.2), going from purely α-helical at low concentrations to a structure that contains approximately 50% β-sheets at higher concentrations. The orientations of the structures were determined using the methodology described in section 2.3.1, and the structural assignment was confirmed by recording N-H spectra.

In a later study, the same authors reported achiral and chiral amide-I and -II spectra of this segment as well as of the 106–126 segment, which is known to form amyloid oligomers and fibrils near POPG bilayers. The structural difference between the two structures has a pronounced effect on the achiral spectra, with the main difference in the amide-II region, where the peak intensity is comparable in the amide-I region, but an amide-II peak is present in the spectrum of the 106–126 segment that is absent in the spectrum of the 118–135 segment.

### 3.4.2. Other Fibrous Proteins

SFG can also be employed particularly well to study other fibrous proteins like collagen, the helical fibrils that are the main component of spinal-cord white matter, and spider-web proteins. Rocha-Mendoza et al. studied type I collagen, the predominant member of the collagen superfamily of extracellular matrix proteins. These proteins play a crucial role throughout the lifespan of many life forms, from early growth and development to the maintenance of homeostasis. All collagens are composed of three polypeptide chains that fold
Chemical Reviews

3.5. Biomineralization

Control of the precipitation and growth of inorganic materials by peptides and proteins has drawn significant attention over the past decade.\textsuperscript{334–337} For designing biomimetic approaches to control the growth of nanomaterials and biomedical coatings as well as understanding the molecular basis for genetic disorders that lead to deficient growth of teeth, it is important to understand the action of biomineral proteins at mineral surfaces.\textsuperscript{338–344} SFG has been used to probe the processes involved at mineral surfaces. The focus has been on systems intended for biomimetic growth of new materials and on understanding hard tissue biogenesis in humans. Most studies have been focused on determining the folding and orientation of the proteins involved using amide-I spectra as well as the role of side-chain structure for targeted interactions with specific mineral facets.

3.5.1. Biosilica. 3.5.1.1. LK Peptides. Diatoms use intricate nano- to microscale architectures at their surfaces with extraordinary properties in terms of toughness, strength, and weight and therefore have tremendous potential for applications in photonics and microelectronics. The interaction of proteins with silica has been one of the earliest protein adsorption processes studied. Already in the early 2000s, in the first days of protein SFG, the Chen and Somorjai laboratories could follow the binding of BSA\textsuperscript{85} and other model proteins\textsuperscript{199} to silica surfaces. The Castner and Somorjai groups followed how LK peptides bind and orient at silica surfaces.\textsuperscript{295} The authors studied the relationship between peptide ordering and silica morphology using the LK platform with different hydrophobic periodicities. The sequences allowed different helical and strandlike secondary structures to be specifically induced without otherwise changing the chemical composition of the peptides. Baio et al. hypothesized that differences in silica morphologies can be attributed to different aggregation behavior resulting from the peptide’s characteristic hydrophobic/hydrophilic sequence.\textsuperscript{345} By combining SFG data about interfacial folding with coarse-grained MD simulations, the authors showed that the morphology of silica is influenced by the overall hydrophobic/hydrophilic pattern of the peptide sequence and their solution-state assembly. Surprisingly, small changes of the system such as N-terminal acetylation of an LK peptide can lead to completely different silica morphologies, even though their secondary structure in solution and at the interface appears to be identical.\textsuperscript{346} It appears that silica bioprecipitation is highly sensitive to even small changes in the peptide system.

To follow biosilification at interfaces directly, Lutz et al. studied LK peptide biosilica formation at the air–water interface with SFG.\textsuperscript{253,255,291} They showed that helical and \(\beta\)-strand LK peptides mineralize different silica morphologies at the air–water interface. This difference is driven by the nature of silica interactions, where helical peptides engage with silica predominantly via their side chains while \(\beta\)-strand LK peptides contact silica mostly via backbone moieties.

3.5.1.2. R5 Diatom Peptides. R5 peptides are based on a repeat unit within the diatom silica protein silaffin.\textsuperscript{1} The protein is found in the diatom silica cell wall and is of great interest for biomimetic molecular electronics, drug delivery, and photonics.\textsuperscript{344} The structure and mode of interaction of R5 with silica have been the subject of some debate in the literature. Solution-state NMR studies have reported that R5 remains unstructured in contact with silica and interacts with all side chains.\textsuperscript{345} Solid-
state NMR studies, on the other hand, have determined that R5 likely forms structured micellelike assemblies and interacts mostly through side chains exposed at the micelle periphery.\(^{348}\) To understand how R5 mineralizes silica at an extended interface, Lutz at al. assembled R5 peptides at the air–water interface.\(^{278}\) Upon injection of tetramethyl orthosilicate (TMOS) into the subphase of the trough, the authors observed the growth of a nanometer-thin and stable biosilica sheet that could be lifted off for ex situ analysis with electron microscopy (Figure 27a,b). Using SFG spectroscopy and MD simulations in combination with spectral calculations using the methodology described in section 2.3.3 (Figure 27c), Lutz et al. observed that R5 assembled into ordered layers at the air–water interface with a strandlike secondary structure and that the order and folding were preserved during silica precipitation, a result that is supported by the solid-state NMR results. At the same time, the silica interaction was mediated through all side chains, which is closer to the results of the solution state NMR studies. Oftentimes, silica formation at the surface uses elements of both NMR-based models, providing a good example of how different biological processes can commence at interfaces.

### 3.5.2. Calcium-Based Minerals

Nature uses proteins to generate the organic matrix for hard tissue engineering. Calcium-based minerals such as oxalates, carbonates, and phosphates play a tremendous role in biology. There have been numerous studies on the role of proteins in the formation of these materials. Understanding the molecular principles of how proteins dictate the growth of different phases of calcium-based minerals would spark new approaches in both medicine and nanotechnology. Here we review recent SFG studies of protein action at calcium mineral surfaces.

#### 3.5.2.1. Calcium Oxalate

Urinary proteins are potent inhibitors of calcium oxalate, a major component of kidney stones. The protection mechanisms are based on specific binding and passivation of calcium oxalate nuclei. The effective inhibition is based on facet-specific binding of calcium oxalate monohydrate and therefore relies on a subtle interplay of several structural factors such as charge distribution, secondary folding structure, and phosphorylation. An extensive literature attests to the importance of direct interactions between proteins and calcium oxalate surfaces for prevention of kidney stones.\(^{349,350}\) Lu et al. used a combination of SFG and MD to study specific recognition mechanisms governing the formation of calcium oxalate.\(^{170}\) Using peptides with the amino acid sequence E(LE)\(_9\) inspired by the human mineral protein osteopontin, the authors followed the precipitation of calcium oxalate dihydrate. The results showed that the strandlike LE peptides refolded upon interaction with calcium ions and formed a turn-rich structure (Figure 28). This rearrangement allowed bidentate binding of calcium ions by the glutamic acid side chains. In this conformation, the calcium ions were arranged at a distance that matched the crystal structure of calcium oxalate dihydrate. Analysis of the oxalate films showed that the dihydrate phase had indeed been formed at the interface.

#### 3.5.2.2. Calcium Carbonate

Proteins can control the growth of calcium carbonate and selectively steer the precipitation pathway toward the calcite, aragonite, or vaterite phases. Since the templating of the different phases must occur predominantly at the protein–carbonate interface, Lu et al. studied the role of peptide folding and structural flexibility in the formation of stable phases of calcium carbonate.\(^{162,235}\) These studies showed that LE peptides based on glutamic acid and leucine with \(\beta\)-strand and \(\alpha\)-helical structures both precipitate calcium carbonate but that only the strand folding leads to a well-defined vaterite-phase calcium carbonate. In addition, the strandlike sequences lead to the formation of stable, self-supported composite sheets of vaterite interspersed with LE peptides. A combined SFG and MD analysis showed that similar to the mineralization of calcium oxalate monohydrate, LE peptides rearrange after contact with calcium ions, which leads to the effective templating of vaterite by structural mimicry of the (001) crystal plane. A helical fold provides less freedom for restructuring and hinders the rearrangement of the side chains in response to the calcium interaction.

#### 3.5.2.3. Hydroxyapatite

The human salivary protein statherin (44 residues) can control the growth of hydroxyapatite (HAP), the principal component of teeth and bones, and prevents the buildup of excess HAP by inhibiting spontaneous calcium phosphate growth. This protein therefore plays an important role in the regulation of hard tissue formation in humans. Because of this importance for human health, statherin has been investigated at HAP surfaces with a variety of methods, including simulations and solid-state NMR spectroscopy.\(^{12,331–335}\) The focus has been on understanding the binding motif within statherin’s N-terminal binding site, a 15 amino acid peptide called SN15. Since the protein–surface interactions are known to involve specific binding of key amino acids with the surface, Weidner et al. studied SN15 on HAP using a combination of SFG and near-edge X-ray absorption fine structure (NEXAFS) spectroscopy.\(^{338}\) Fluorine labeling allowed quantification of the orientation of individual side chains within SN15. The orientations of the phenyl rings of phenylalanines F7 and F14 were determined using SFG analysis and verified with

---

**Figure 28.** (a) Transmission electron microscopy image of self-supported vaterite-phase calcium carbonate sheets precipitated by LE peptides at the air–water interface. (b) Experimental SFG spectra along with those calculated from the snapshots of the LE peptides (c) before and (d) after mineralization. The top-view snapshots of the MD simulation box show that the peptides fold back on themselves upon mineral interactions. This refolding brings the glutamic acid sites into an ideal distance for bidentate binding of calcium ions—an important prerequisite for phase-selective mineralization. Adapted from ref 170. Copyright 2019 American Chemical Society.
NEXAFS spectroscopy. The ring structure indicates that the phenylalanine groups are likely involved in peptide–peptide interactions, a potential mechanism for the facet-specific binding observed for statherin.

3.6. Ice-Binding Proteins

3.6.1. Ice-Nucleating Proteins. Specific bacteria use ice-nucleating proteins to grow ice crystals at their outer surface at high subzero temperatures. Ice-nucleating bacteria are ubiquitous and cause massive frost damage to plants. In addition, ice-nucleating bacteria have been observed to nucleate ice crystals and induce rainfall when airborne in the atmosphere. Ice-nucleating bacteria are the most effective ice nucleators known and can potentially impact global precipitation patterns and atmospheric cloud formation. Pandey et al. studied the temperature dependence of SFG spectra from water molecules interacting with the ice-nucleating protein inaZ. For practical reasons, the authors used D$_2$O instead of H$_2$O in their studies. A progressive increase in the SFG intensity of strongly hydrogen-bonded water was observed as the temperature decreased from room temperature to just above the freezing point of D$_2$O (4 °C) (Figure 29), while such a temperature response was not observed for ice-inactive substances. By comparison with the SFG results from MD simulations of the inaZ ice-active site, the alternating hydrophilic–hydrophobic patterns in inaZ were concluded to be responsible for templating water molecules at low temperatures and promoting freezing of supercooled water. Moreover, the particular alignment of the interfacial hydrogen-bonding network can effectively funnel heat from the interface and enhance the ice nucleation. This energy transfer mechanism was experimentally observed using time-resolved SFG spectroscopy, which will be discussed in more detail in section 3.9.1

3.6.2. Antifreezing Proteins. As can be seen in Figure 30, the signature of an ice-like layer was also observed at the ice-binding site of antifreezing protein type III (AFP-III) in aqueous solutions above the freezing temperature by Meister et al. In this figure, the broad spectral features from the neat air–water interface were replaced by a relatively narrow peak at 3200 cm$^{-1}$, which corresponds to a highly ordered ice-like structure of water. The disappearance of the non-hydrogen-bonded water peak at 3740 cm$^{-1}$ as well as the appearance of CH vibrational modes at <3000 cm$^{-1}$ indicated the presence of the protein at the air–liquid interface. The shape and intensity of the SFG signal in the CH and OH region were dependent on the concentration of the protein, and a critical concentration of 10 μM was found, below which the spectrum is dominated by the air–liquid water signatures and above which the highly ordered ice-like water signature dominates the SFG signal. It was shown that even a single point mutation in the ice-binding site can eliminate the antifreezing activity of the protein, as evidenced by the absence of the ice-like signature in the SFG spectrum of the T18N mutant. It was therefore concluded that not the protein conformation but rather the preordered ice-like water layers are mainly responsible for the recognition and binding to ice. At the same time, a recent study found that the AFP-III ice-binding site is oriented away from the water surface and is mostly decoupled from the bulk water. In view of these observations, it is unclear what type of water was observed in the mutant study. As opposed to the moderate AFP-III, which acts by promoting the formation of a well-ordered ice-like layer of water, in the antifreezing protein DAFP-1, the highly ordered array of threonine residues directly binds to the ice surface without ordering the hydrogen-bonded water molecules. The antifreezing mechanism of DAFP-1 was studied by surface tension measurements and conventional SFG measurements as well as chiral SFG (the subject of section 2.1.2.2) and phase-resolved SFG (the subject of section 2.2.1.2). An aqueous solution of DAFP-1 shows significantly lower surface tension than pure water and intense SFG bands at $\sim$3300 and $\sim$3485 cm$^{-1}$. These results point toward the surface activity of DAFP-1, while chiral SFG helps with the proper assignment of the peak at $\sim$3300 cm$^{-1}$ originating from the N–H vibrations of the helical protein backbone. The peak at higher frequency was also assigned to the O–H vibration of threonine residues or adsorbed water molecules at the protein surface. The relatively narrow band of the peak at $\sim$3485 cm$^{-1}$ (with a width of $\sim$95 cm$^{-1}$) compared with the liquid water peak (with a width of $\sim$300 cm$^{-1}$) indicates well-defined strength of the hydrogen bonds donated by the threonine residues. Lowering the temperature from 20 to 0 °C leads to only a minor reduction in the intensity and a small red shift in the SFG spectra from the protein at the air–liquid interface. Comparison of these results with those obtained for AFP-III shows that the water molecules adsorbed to the ice-binding sites of these proteins are very different in their character.

3.7. Proteins in Surface Bioengineering

3.7.1. Albumin and Fibrinogen at PEGylated Surfaces. Bernhard et al. studied BSA and fibrinogen at poly(ethylene glycol) (PEG)-coated surfaces. PEG coating is a strategy to reduce nonspecific protein adsorption to material and nanoparticle surfaces. Since protein binding still takes place, even if at a reduced rate, it is important to study the binding of proteins to PEG surfaces. Surprisingly, Bernhard et al. found that BSA and fibrinogen become oriented when they interact with PEG-coated surfaces. The authors observed that the proteins remained in their native folding states and that a combination of
attraction by charges embedded in the surface coatings and friction caused by the PEG layer had an orienting effect on the proteins. The spectra were interpreted by comparison with calculated spectra based on crystal structures available in the Protein Data Bank.

3.7.2. Von Willebrand Factor. The clotting protein von Willebrand factor (VWF) is an important element in the design of cardiovascular implants and biomaterials used in blood-contacting applications. VWF binds to the platelet receptor glycoprotein Ibα (GPIbα) when activated by binding to surfaces. In its active state, VWF can lead to failure of cardiovascular implants and dialysis devices. Tronic et al. determined that activation of VWF is dictated by surface chemistry and that when VWF is bound to hydrophobic polystyrene, its A1 domain remains accessible for GPIbα binding. Roeters et al. resolved these features by studying the system with SFG spectroscopy. According to their analysis, the VWF A1 domain maintains a native fold when binding to polystyrene. By comparing experimental and calculated SFG spectra, the authors found that the A1 domain assumes an orientation in which the GPIbα binding domain is exposed to the solution phase. The availability of such orientation information about functional blood proteins will support the design of novel blood-contacting biomaterials and wound-healing applications.

3.7.3. Peptides and Proteins on 2D Materials. The interaction of proteins and peptides with artificial functional materials such as graphene, molybdenum disulfide (MoS2), and tungsten diselenide (WSe2) has recently become an important topic. Tailoring protein binding to such materials will be key for in vivo applications.

Zou et al. recently reported on the role of the peptide sequence in binding to graphene surfaces by studying the binding of cecropin P1 and MSI-78-C1 with a combination of SFG and MD simulations. The orientation of the helical peptides was determined with the methodology described in section 2.3.1, from the peak ratios of the amide-I helical bands near 1650 cm−1, assuming ideal helical folds. The authors found that planar side chains and charged sites dictate the orientation of peptides on graphene (see Figure 31). Cecropin P1 and MSI-78-C1 were found to bind differently to graphene. While cecropin P1, with an unbalanced distribution of planar and charged sites, adopts an upright orientation, MSI-78-C1 adopts a parallel orientation with respect to the surface because it has an even distribution of phenylalanine residues and charged residues. A mutation within MSI-78-C1 that replaced two phenylalanine residues with alanine (F13A, F17A), breaking the even distribution of planar side chains, led to an upright orientation of MSI-78-C1.

In a comparable study by the same group, Xiao et al. investigated the interaction between several de novo-designed α-helical model peptides (based on a hybrid sequence of cecropin and melittin) and a single-layer MoS2 surface. The authors found that charged amino acids in the peptide sequence play a vital role in ensuring a standing-up configuration and that—in contrast to graphene surfaces—the aromatic residues do not interact strongly with the MoS2 surface. In a follow-up paper, the Chen group investigated the interaction of proteins with MoS2 and another monolayer transition metal dichalcogenide, WSe2, which has also been widely studied because of its potential application in micro- or nanoelectronic biosensors. The authors showed that out of the three model proteins studied (β-Gal-D308C, β-Glu, and HLD-A141C), only the protein that exposed a large hydrophobic area in a relatively flat region of the protein’s surface (HLD-A141C) adsorbed to these surfaces. Furthermore, Xiao et al. found that none of the proteins formed disulfide bonds with the surfaces, even though this was expected from the fact that such bonds were observed previously between MoS2 and polyethyleneimine (PEI) and PEG and from the presence of cysteine residues in the investigated proteins.

Harrison et al. determined the orientation of model LK peptides and the B1 domain of protein G on graphene by combining Monte Carlo simulations with SFG amide spectra through spectral calculations (see section 2.3.4 for details on the applied methodology). Their results indicate that it is important to include structural variability into surface protein structure analysis to capture the thermodynamics of interactions with artificial surfaces.

3.7.4. Hydrophobins. Hydrophobins are a group of proteins that are expressed by filamentous fungi and bacteria. They self-assemble into monolayers at hydrophobic–hydrophilic interfaces, such as the air–water interface, e.g., to aid the growth of fungi in humid environments. Because these monolayers are water-repellent, they are also of interest for technological applications such as foam stabilization, hydrophobic substance dispersal, and recombinant protein purification.

Figure 30. (left) Schematic of AFP-III with the ice-binding site marked in red. (right) SFG spectra of AFP-III at different concentrations at the air–water interface. The relatively narrow peak at ~3200 cm−1 is assigned to the water molecules contained in the ice-like layer bound to the ice-binding site of AFP-III. Adapted from ref 181. Published by the National Academy of Sciences.
Both class I and class II hydrophobins form a rigid and well-ordered β-barrel- and α-helix-containing structure at the air–water interface. Meister et al. combined chiral and achiral SFG in the amide-I, C–H, and O–H stretching regions to study the self-assembly mechanism of both classes of hydrophobins at the air–liquid interface. Observation of a chiral signal suggests that the proteins remain folded upon adsorption to the surface. However, time-dependent measurements clearly showed a gradual buildup of the β-sheet structure during surface-driven self-assembly for class I hydrophobin, while class II hydrophobin spectra showed a negligible structural change during this period. In this study, the application of a Rochon polarizer allowed simultaneous chiral and achiral measurements, which ensured monitoring of identical self-assembly processes in all measurements. When the solution pH is changed, HFBI (a type II hydrophobin) undergoes a reversible change in its tilt angle due to the interaction of the hydrophilic part of the hydrophobin with water molecules, following the pH-induced change in the

![Figure 31. SFG and MD study of (a) cecropin P1 and (b) MSI-78-C1 on graphene surfaces. Cecropin P1 adopts an upright orientation, while MSI-78-C1 is oriented flat on the surface. Zou et al. explained this effect in terms of the distribution of planar and charged residues along the peptide chain. (c) The mutant of MSI-78-C1, which has fewer aromatic sites, is oriented upright on the graphene surface. Adapted from ref 365. Copyright 2017 American Chemical Society.](image-url)
charge state of amino acid residues. Essentially, protonation and deprotonation of the charged residues of the protein at low and high pH values, respectively, modify the net charge of the protein, triggering protein reorientation at the air–liquid interface. This change in the hydrophobin’s tilt angle (θ) and rotation angle (ψ) with changing pH is manifested by a shift in the peak positions of the achiral SFG signal and a change in the peak ratio in the chiral SFG signal. As can be seen in Figure 32, the experimental spectra match well with spectra calculated with the formalism described in section 2.3.3 for both achiral (SSP) and chiral (PSP) polarization combinations, pointing toward rotation angle (θ) in the water interface. Schloss et al. showed that the SFG intensity of the water interface was characterized by SFG.205

BslA was the first hydrophobin identified in bacteria, and its structure at the air–water interface was determined by SFG.205 Like its fungal counterparts, BslA shows extreme surface activity. Wang et al. found a strong dependence of the spectral shape on the solution pH and protein concentration, with very narrow features in the achiral spectra recorded at neutral pH and high protein concentration. The authors hypothesized that under these circumstances, BslA self-assembles into extremely ordered structures at the air–water interface. Decreasing the pH to 1.2 resulted in a loss of the chiral SFG signal, ascribed to denaturation of the protein. This was corroborated by a loss of the sharp features in the achiral spectra. By combining these results with surface compression isotherm, AFM, and X-ray reflectivity measurements, the authors speculated that the α-helices and β-sheets of BslA, might protrude from the biofilm formed at the air–water interface. Schloss et al. showed that the protein can be used as a surface anchor in combination with a molecular biology method (SpyCatcher/SpyTag) to create modularly functionalizable microcapsules. The authors used SFG spectroscopy in the amide-I and CH regions to show that the fused SpyTag does not significantly affect the BslA secondary structure or orientation within the interfacial hydrophobin film, although some structural differences were revealed in the CH spectra, probably resulting from interactions between the N-terminal SpyTag and the methyl groups in the hydrophobic region of BslA. The successful formation of the microcapsules was confirmed by the observation of a fluorescent shell formed by SpyTagged BslA capsules covalently linked to SpyCatcher–green fluorescent protein (GFP) fusion proteins.

3.8. Surface-Bound Enzymes

The activity and stability of enzymes can be greatly affected by surfaces. Therefore, controlling the orientation of chemically immobilized enzymes and proteins on surfaces is desirable. While SFG has not yet been used to determine the structure of surface-bound enzymes in detail, studies to determine the orientation and presence of certain expected secondary structure motifs have been performed. Shen et al. used ATR-FTIR and SFG to demonstrate how the mode of enzyme attachment, determined by cysteine mutations at different positions, to maleimide-terminated SAMs on glass substrates affects the enzyme’s orientation. Along the same lines, Badieyan et al. engineered surface-immobilized enzymes that exhibited a much higher catalytic activity in air compared with that of the lyophilized enzyme powder. The high activity of the tethered enzyme was achieved by providing stabilized hydrogen-bonding interactions through site-specific covalent tethering of the enzyme to the surface. An amide-I SFG signal was observed from haloalkane dehalogenase (HLD) in the presence of poly-(sorbitol methacrylate) (PSMA) in air. As the entrainment of water in PSMA was ruled out by complementary QCM-D measurements, the authors concluded that in the absence of bulk water, PSMA preserves the secondary structure of the protein by replacing the stabilizing interactions that exist in the presence of solvent.
3.9. Emerging Methods

3.9.1. Time-Dependent SFG. Proteins are dynamic systems, and a wide range of a protein’s properties and functions are directly influenced by its dynamics. These properties include, but are not limited to, ligand binding, protein folding, protein hydration, enzymatic activities, protein–protein interactions, and protein transitional and torsional motion. A wide range of multidimensional ultrafast spectroscopic techniques have been developed recently to reveal a wealth of structural, energetic, and dynamical information about molecular, biological, and nanomaterial systems. The dynamics of protein side chains in a bulk solution are often measured using NMR spectroscopy. However, the side-chain dynamics for proteins at different surfaces and interfaces might be different from those in the bulk. Time-resolved SFG measurements can directly resolve the subpicosecond dynamics of proteins residing at a surface or interface. Performing time-resolved SFG measurements requires an additional laser beam, also known as the pump pulse, that is spatially overlapped with the SFG probe beams (i.e., the IR and visible beams). Transient spectra are obtained by changing the time delay between the pump and probe pulses (see section 2.2.1.3 for more details about the experimental technique).

With this technique, time-resolved SFG measurements have been employed to determine the interfacial dynamics of water molecules at the air–water interface and water molecules interacting with surfactants, peptides, and proteins. The OH stretch vibration of water near a film of ice-nucleating Pseudomonas syringae bacteria was studied in this fashion by Pandey et al., who found that the film induces an efficient energy transfer to lower energies within the water network. This energy transfer was significantly lower in the case of ice-inactive proteins.

Donovan et al. assessed for the first time the reorientation of the isopropyl methyl groups of L-leucine at the air–water interface and showed that in-plane and out-of-plane methyl residues have slightly different reorientation rates. These side-chain dynamics can play a critical role in molecular functions at surfaces. Following the reorientation dynamics of leucine side chains in an LK model peptide at the air–liquid interface for three representative peptide folds, i.e., α-helix, 310 helix, and β-strand, the spectra revealed that the peptides exhibit remarkably similar subpicosecond reorientation dynamics regardless of their secondary structure.

The Ye group investigated the impact of hydration water on the vibrational energy relaxation in an interfacial protein. Tan et al. probed the vibrational dynamics of the amide-I mode of the CALP23 peptide at the H2O and D2O interface. The vibrational relaxation times showed a strong solvent dependence that the authors attributed to resonant vibrational energy transfer from the amide-I mode to water bending modes. The study underscores the importance of backbone hydration for protein dynamics.

Vibrational energy transfer from the amide-A mode to the amide-I mode of membrane-bound peptides was also studied with time-resolved SFG spectroscopy. Tan et al. found that the vibrational energy relaxation of transmembrane peptide WALP23 (which forms continuous α-helices in DPPG lipid bilayers) and fusion peptide GP41 (which forms mixed α-helical and β-sheet structures in such bilayers) from the amide-A mode to amide-I mode goes through two pathways: (a) direct NH–CO coupling and (b) coupling through intermediate states, where the NH–O=C hydrogen-bonding strength determines the ratio of these two pathways to a large extent. If there are strong NH–O=C hydrogen bonds between a peptide’s backbone amide groups, the pathway through the intermediate states is favored.

3.9.2. Two-Dimensional SFG. Just as in the time-resolved experiments, in 2D-SFG techniques additional pump beams are overlapped at the interface—here to create an extra frequency dimension (see section 2.2.1.4 for experimental details). The potential of 2D-SFG for determining protein structure and orientation was first illustrated by Laaser et al., who showed that the technique can probe interfacial peptide dynamics and that it can function as a fourth method (besides recording: the amide-III spectrum, see section 2.1.3; the chiral N-H stretch spectrum, see section 2.1.2.2; and the ratio between various achiral polarization combinations, see section 2.3.1) for discriminating between random coil and α-helical structures. By comparing the time dependences of the nodal slopes of a 20 amino acid peptide in bulk solution (observed with 2D-IR) and tethered to a gold surface (observed with 2D-SFG), the authors showed that the tethered peptides are less accessible to water molecules. Interestingly, the presence of random coil structure, which is invisible in the linear SFG spectrum and on the diagonal of the 2D-SFG spectrum, is indicated via a cross-peak in the 2D-SFG spectrum.

Later, the same technique was used by Ghosh et al. to monitor the amyloid formation of the FGAIL pentapeptide. Because this middle segment of the otherwise highly conserved islet amyloid polypeptide (also see section 3.4.1.1) has at least one mutation in all species that do not contract type II diabetes, it is thought to play an important role in the development of the disease. It also forms well-ordered amyloid fibrils under many conditions, for example near a SAM of methyl 4-mercapto-benzoate (MMB) on gold. On the basis of a comparison of calculated 2D-SFG spectra with preliminary experimental results, it seems likely that MMB monolayers align adsorbed FGAIL peptides parallel to the surface, thereby stimulating amyloid growth. Even though their spectral calculation model did not fully reproduce the potentially highly informative cross-peak between the MMB ester peak and the peptide’s fibrillar modes, Ghosh et al. did provide the first molecular insight into surface-catalyzed amyloid fibril formation.

In a follow-up study, the Zanni group used the same technique to show that the FGAIL peptide forms a heterogeneous mixture of amyloid fibrils (containing both antiparallel and parallel β-sheets) near monolayers of hydrogen-bond-promoting mercaptobenzoic acid (MBA) on gold, while they form homogeneous amyloid parallel β-sheets when they are chemically ligated to the MBA layer. Using the formalism from section 2.3.3, with a new approach to speed up the computationally exhaustive 2D-SFG calculations, the authors were able to reproduce the phase-resolved 2D-SFG spectra of the two types of samples with either a mixture of different antiparallel and parallel β-sheets (without covalent linkage) or with an almost purely parallel β-sheets (when the FGAIL peptides are tethered to the MBA monolayer).

3.9.3. Scattering SFG. SFG scattering (SFS) holds great potential to study the structure of proteins on micro- and nanoscopic surfaces. In general, surfaces dominate the physical and chemical properties of these nanomaterials, which are of interest for biocatalysis, biomaterials, nanomedicine, targeted drug delivery, chromatography, and medical imaging. To control the biological response to coated and/or functionalized nanoparticles (NPs), one must understand at the molecular...
level the surface structures of the surface-modified NPs and their interactions with proteins from the surrounding biological fluids. However, compared with flat surfaces, it is extremely challenging to obtain molecular-level information about the surface structure of NPs and their coatings. SFG scattering spectroscopy can potentially provide this information, as it has been proven to be a reliable tool to probe nanosurfaces.

The first step toward protein spectroscopy with SFG scattering was reported by Johansson and Koelsch, who investigated the structure of collagen fibers in solution. Since protein fibers are involved in many diseases, it will be important to provide information about the three-dimensional structure of such molecules. Johansson and Koelsch demonstrated that SFG scattering can report back on the N–H stretching, C–H stretching, and amide-I regions of collagen fibers. Their study exploited not only the scattering spectra but also the scattering angles of the SFG photons, demonstrating that SFS can provide information about the relative orientations of the side-chain and backbone groups. Such information may be useful in the design of chemical modification strategies for synthetic scaffolds in tissue engineering and for the development of medical drugs to counter fibril-related diseases.

3.9.4. SFG Microscopy. SFG imaging and microscopy can provide valuable information on chemically and structurally inhomogeneous surfaces and interfaces. However, the application of SFG imaging is technically challenging and to date has mainly been restricted to the study of model systems. Improvements in the spatial resolution of the technique, the acquisition time, and the image-processing algorithms will further expand the utility of SFG imaging to include biologically relevant systems.

Hanninen et al. used point-scanning polarization-sensitive SFG microscopy to boost the contrast in images of collagen-rich tissues with respect to SHG imaging. The authors were able to directly resolve the orientation of collagen fibers in rat tail tendon without the need to change the beam polarizations. Other molecules studied in the Potma group for which the molecular arrangement was determined with SFG imaging are cellulose fibers and cholesterol microcrystals (see Figure 33).

SFG microscopy has also been used to map starch granules in rice grains and living plants. Mizutani et al. used scattered SFG intensity images to distinguish four different saccharide species (i.e., D-glucose, amylopectin, β-cyclodextrin, and amylose), illustrating the usefulness of the chemical selectivity of SFG imaging. Using a nonscanning system, Inoue et al. obtained SFG images of onion root cells at two different IR wavelengths (2880 and 2703 cm\(^{-1}\)).

4. SUMMARY AND OUTLOOK

The Protein Data Bank contains more than 100 000 protein structures, but only a few have been resolved for proteins at their native interface. With a growing realization of the importance of interfacial structural biology, we also realize that protein structure determination at interfaces is still at an early stage in comparison with what is already standard practice for crystal- and solution-state protein studies based on methods such as crystallography, electron microscopy, and NMR spectroscopy. This review has highlighted the potential of SFG to fill the “resolution gap” between the bulk and interfaces and to repeat the success of structural biology at interfaces. Recently, we have seen a sharp rise in experimental methods, theoretical frameworks, and creative combinations of SFG with other surface-sensitive tools. The advent of spectral simulations now allows the assessment of large and complex protein structures. Future inclusion of calculations of spectral regions such as additional amide modes, water, and C–H resonances will allow a holistic picture of almost all molecular components driving interfacial protein folding. Two-dimensional SFG methods, phase-resolved detection, and high-resolution methods such as nanosecond SFG can potentially play an important role for the unique spectral assignments and data interpretation required for the analysis of complex and/or heterogeneous systems. The combination of SFG with metadynamics simulation data sets can potentially include the structural diversity at biological interfaces. At the same time, proteins are highly dynamic, and SFG is only beginning to play out its strength as an ultrafast optical method to investigate protein dynamics at interfaces. Pump–probe techniques can provide information about protein motion at all relevant time scales from the subpicosecond range to the millisecond range. Here, SFG can provide information that is not available with any other current method. Although it is in the very early stages of development, SFG microscopy methods could in the future be developed into tools to follow the spatiotemporal evolution of biological surfaces at the molecular level.

AUTHOR INFORMATION

Corresponding Authors
*E-mail: weidner@chem.au.dk
*E-mail: saman.hosseinpour@fau.de
*E-mail: s.j.roeters@chem.au.dk

ORCID
Saman Hosseinpour: 0000-0001-8861-1217
Steven J. Roeters: 0000-0003-3238-2181
Mischa Bonn: 0000-0001-6851-8453
Wolfgang Peukert: 0000-0002-2647-107X
Sander Woutersen: 0000-0003-4661-7738
Tobias Weidner: 0000-0002-7083-7004

Author Contributions
S.H. and S.J.R. contributed equally.

Notes
The authors declare no competing financial interest.

Biographies
Saman Hosseinpour is the surfaces and interfaces group leader at the Institute of Particle Technology at Friedrich-Alexander University Erlangen-Nuremberg. He received his B.A. and M.A. degrees in material science and engineering with a focus on surface science in Iran.
and later was promoted as a faculty member of the Department of Material Science at Azad University. In 2009 he started his Ph.D. studies at the Royal Institute of Technology (KTH) in Sweden to explore the effectiveness of self-assembled organic molecules as corrosion inhibitors. In 2016 he received the prestigious Morris Cohen Graduate Award for the best doctoral thesis in corrosion science and electrochemistry worldwide. He did his postdoctoral studies with Mischa Bonn and Ellen Backus at Max Planck Institute for Polymer Research (MPIP) in Mainz, Germany, and studied photocatalytic water splitting on TiO$_2$ using time-resolved SFG spectroscopy.

Steven J. Roeters is a postdoctoral researcher in Tobias Weidner’s group at Aarhus University in Denmark. He received his Ph.D. at the University of Amsterdam in 2018 under the supervision of Sander Woutersen. His thesis work focused on advanced vibrational spectroscopies such as SFG and 2D IR to study large biomolecular systems such as membrane proteins and amyloidic peptides. Ever since his thesis work, an important aspect of his research has been theoretical modeling of complex spectroscopic datasets. Funded by a Lundbeck Foundation Postdoctoral Fellowship, his research interests are currently expanding into combining computer simulations with cryo-electron microscopy and SFG in order to elucidate the structure of membrane-associated proteins.

Mischa Bonn is a research director at MPIP. He works on label-free (ultrafast) vibrational spectroscopy and microscopy of biomolecular systems and water in such systems. He received his Ph.D. in 1996 from the University of Eindhoven for research performed at the FOM Institute for Atomic and Molecular Physics (AMOLF) in Amsterdam. After postdoctoral research at the Fritz Haber Institute in Berlin and Columbia University in New York, he worked at Leiden University as an assistant professor starting in 1999 and then as an associate professor starting in 2002. In 2004 he became a group leader at AMOLF, and in 2011 he joined MPIP. His research interests are the structure and dynamics of molecules at interfaces and electron transfer across interfaces. He has won several prizes and awards and has published over 400 research articles.

Wolfgang Peukert is director of the Institute of Particle Technology at the Friedrich-Alexander University Erlangen-Nuremberg, Germany. He studied chemical engineering and received his Ph.D. in 1990 from the University of Karlsruhe. After 7 years in industry, he returned to studies at the Technical University of Munich. He did his postdoctoral studies with Michael Zhamikov and the University of Washington with David Castner. Before coming to Aarhus he was a group leader at MPIP. His research is focused on biological surfaces, including biominalerization, membrane proteins, ice proteins, and artificial biomaterials. He uses ultrafast SFG vibrational spectroscopy in combination with computer simulations to determine the structure and orientation of surface-bound biomolecules and develops methods to follow molecular motion at interfaces.

**ACKNOWLEDGMENTS**

We thank Lars Schmüser and Thaddeus W. Golbeck for valuable discussions. S.H. thanks Friedman-Alexander University Erlangen-Nuremberg for supporting his research through Emerging Talents Initiative (ETI) Grant 2018/2_Tech_06. German Research Foundation is grateful for financial support via the DFG Priority Program DiSPBiotech (project PE427/32-2 within SPP 1934). T.W. and S.J.R. acknowledge funding from the Villum Foundation (Experiment Grant 22956). S.J.R. thanks the Lundbeck Foundation for a postdoctoral fellowship.

**REFERENCES**


(74) Ye, S.; Li, H.; Yang, W.; Luo, Y. Accurate Determination of Interfacial Protein Secondary Structure by Combining Interfacial-


(143) Huang, Q.; Schweitzer-Stenner, R. Conformational Analysis of Tetrapeptides by Exploiting the Excitonic Coupling between Amide I Modes. J. Raman Spectrosc. 2004, 35, 586–591.


(327) Morse, J. W.; Arvidson, R. S.; Lutte, A. Calcium Carbonate Formation and Dissolution. Chem. Rev. 2007, 107, 342–381.


3464


