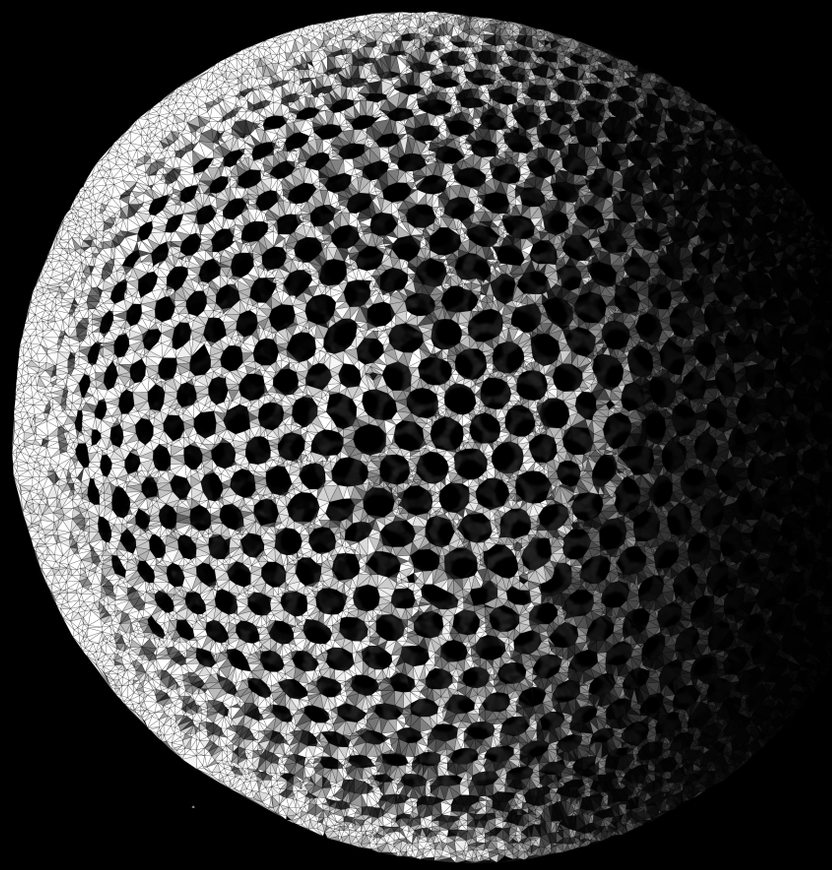
Shedding light on peptide controlled silica mineralization

Lutz, Helmut

Shedding Light on Peptide Controlled Silica Mineralization

Helmut Lutz



Shedding Light on Peptide Controlled Silica Mineralization

Helmut Lutz

Shedding Light on Peptide Controlled Silica Mineralization

ISBN 978-3-95638-851-4

Cover art: "Crucible" by Helmut Lutz. A derivative work (created with permission of ZEISS Microscopy) based on an electron microscopy image of a diatom, courtesy of ZEISS Microscopy.

About the cover: The cover is an abstract illustration of piecing together our knowledge of biomineralization and using it to create something of our own imagination.

Shedding Light on Peptide Controlled Silica Mineralization

Academisch Proefschrift

ter verkrijging van de graad van doctor
aan de Universiteit van Amsterdam
op gezag van de Rector Magnificus
prof. dr. ir. K.I.J. Maex
ten overstaan van een door het College voor Promoties
ingestelde commissie,
in het openbaar te verdedigen in de Agnietenkapel
op woensdag 13 september 2017, te 12:00 uur

door

Helmut Lutz

geboren te Boekarest, Roemenië

Promotiecommissie

Promotoren:	prof. dr. M. Bonn	Universiteit van Amsterdam
	prof. dr. T. Weidner	Aarhus Universitet
Overige leden:	prof. dr. H. Birkedal	Aarhus Universitet
	prof. dr. W.J. Buma	Universiteit van Amsterdam
	prof. dr. G.P. Drobny	University of Washington
	prof. dr. E.J. Meijer	Universiteit van Amsterdam
	prof. dr. S. Woutersen	Universiteit van Amsterdam

Faculteit der Natuurwetenschappen, Wiskunde en Informatica

The research described in this thesis was conducted at the *Max Planck Institute for Polymer Research* (MPIP), Ackermannweg 10, 55128 Mainz, Germany. This work was financially supported by the Max Planck Society, the European Commission (CIG grant #322124) and the Deutsche Forschungsgemeinschaft (WE4478/2-1).

To my wife Sarah.

Contents

This thesis is based on the following publications:

Biomimetic growth of ultrathin silica sheets using artificial amphiphilic peptides,
H. Lutz, V. Jaeger, R. Berger, M. Bonn, J. Pfaendtner and T. Weidner, *Adv. Mater. Interfaces* **2015**, *2*, 1500282.

Acetylation dictates the morphology of nanophase biosilica precipitated by a 14-amino acid leucine–lysine peptide,
H. Lutz, V. Jaeger, M. Bonn, J. Pfaendtner and T. Weidner,
J. Pept. Sci. **2016**, *23*, 141-147.

The structure of the diatom silaffin peptide R5 within freestanding two-dimensional biosilica sheets,
H. Lutz, V. Jaeger, L. Schmüser, M. Bonn, J. Pfaendtner and T. Weidner,
Angew. Chem. Int. Ed. **2017**, *56*, 8277-8280.

How well do force fields reproduce protein secondary structure and orientation at the air-water interface?
H. Lutz, V. Jaeger, M. Bonn, B.L. de Groot and T. Weidner,
to be published.

Other publications by the author:

Diatom mimics: directing the formation of biosilica nanoparticles by controlled folding of lysine-leucine peptides,
J.E. Baio, A. Zane, V. Jaeger, A.M. Roehrich, H. Lutz, J. Pfaendtner, G.P. Drobny and T. Weidner,
J. Am. Chem. Soc. **2014**, *136*, 15134-15137.

The interaction with gold suppresses fiber-like conformations of the amyloid β (16-22) peptide,
L. Bellucci, A. Ardevol, M. Parrinello, H. Lutz, H. Lu, T. Weidner and S. Corni,
Nanoscale **2016**, *8*, 8737-8748.

Determination of absolute orientation of α-helices at interfaces using phase resolved SFG spectroscopy,
L. Schmüser, H. Lutz, S. Roeters, S. Woutersen, M. Bonn and T. Weidner,
to be published.

Side chain reorientational dynamics at interface insensitive to protein secondary structure,
M.A. Donovan, Y.Y. Yimer, H. Lutz, J. Pfaendtner, M. Bonn and T. Weidner,
to be published.

1 Introduction

1.1 Overview

Since the early days of human history society has always searched for better materials for tools and in other areas of everyday life. However, in the last decades materials have come to face extreme conditions. Unprecedented situations like ballistic impact, atmospheric re-entry, or deep sea exploration are pushing the limits of material stability. Putting extreme examples aside, there is a constant drive for improvement of materials used in infrastructure and technology. Engineers and scientists alike are seeking ways to increase the strength as well as the toughness of materials – two properties that rarely go hand in hand. Combining these properties usually requires complex production methods, involving high temperatures and aggressive chemicals. With conventional methods at their limit, the materials sciences turned to nature where we find lightweight materials with a high degree of fracture resistance and strength. Over hundreds of millions of years these materials have evolved, driven by the need of living organisms to adapt to predators and hostile environments.

By a process known as *biomineralization* natural composites are generated from soft organic and hard inorganic matter. An intriguing feature not only of biominerals but of natural materials as a whole is that they regularly outperform the sum of their components.[1] This is in no small part due to a gradient of complex architectures spanning the micro- to nanometer scale. In biominerals proteins and other organic molecules interact with the mineral phase on a molecular scale. On this scale, nucleation or inhibition of mineral growth takes place. On a macromolecular scale the organic matter can serve as glue contributing to biomineral toughness by viscoelastic energy dissipation.[1] This thesis is dedicated to the question of how organic molecules like proteins and peptides interact with the mineral phase and whether it is possible to obtain a molecular picture of the interface between organic and inorganic matter.

We have chosen to study the biomineralization of silica ($SiO_2 \cdot nH_2O$), most notably occurring in unicellular algae called diatoms (introduced in Section 1.2). The silicified cell walls of certain diatom species were shown to possess remarkable strength due to the architecture and toughness of the material. The pressures resisted were equivalent to 100-700 tonnes m^{-2}.[2]

The precipitation of silica by diatoms is still regarded as one of the best studied model systems to understand biomineralization. At the same time one has to bear in mind that science is far away from mimicking *biosilicification* to the level seen in the beautifully intricate shells of diatoms (Figure 1.1).

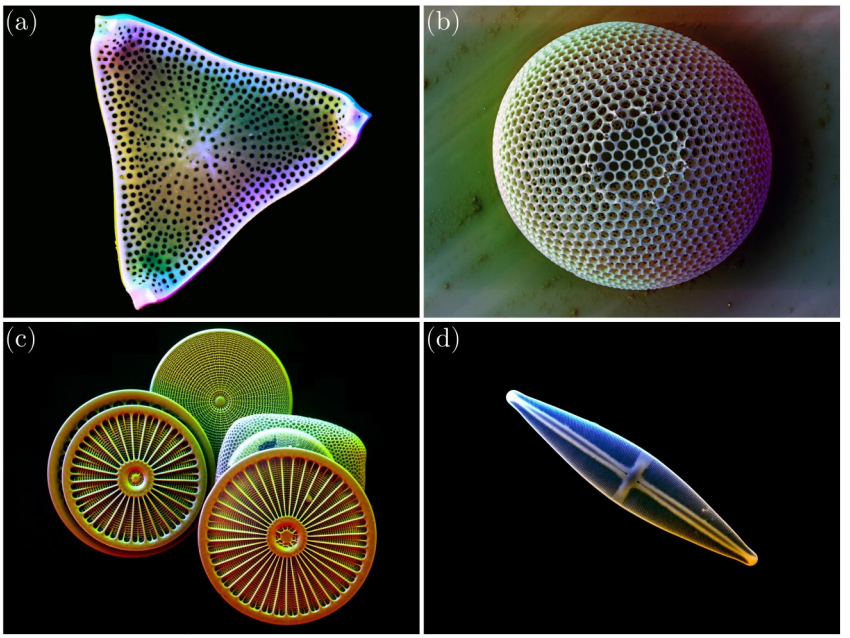


Figure 1.1: Illustrations of the diversity of diatom cell walls. Scale bars were not provided. Diatoms are generally 2-200 μm in size.[3] Images (a), (b), (c) and (d) are showing colored electron microscopy images of diatoms (made with ZEISS EVO SEM). Image courtesy of ZEISS Microscopy.

The investigation of organic-mineral interfaces is not a trivial undertaking since few scientific methods are suitable for this purpose. Of these methods even fewer can be used to study *in situ* systems, i.e. biomineralization in solution. A method meeting both of these requirements is *vibrational sum-frequency generation* (VSFG) spectroscopy.

VSFG spectroscopy is a specialized form of vibration spectroscopy that uses laser pulses to probe interfacial molecular vibrations. The interface specificity is related to the physical selection rules of the three-field process introduced in Section 1.3. In Chapters 2-5 the information about interfacial molecular vibrations will be frequently compared to molecular dynamics simulation results. The fundamentals of molecular dynamics simulations are introduced in Section 1.4. The results presented in this thesis have produced the following conclusions about biomineralization, specifically biosilicification by proteins:

(a) We have established a model system to study interfacial biomineralization on the scale of a few nanometers. Molecular order can be monitored with VSFG spectroscopy.
(b) The secondary structure of biomineralization-active peptides is a determining factor for the morphology of silica which is mineralized at an interface.
(c) It is difficult to predict the morphological outcome of a biomineralization process. Minimal changes to the peptide (like a protecting group) can result in substantially different outcomes.
(d) The diatom biosilicification peptide R5 grows self-supported thin silica sheets at the air-water interface. We observed that the peptide refolds when interacting with silica.

At the end of this chapter, I will give a brief outline of the scientific content of this thesis (Chapters 2-5).

1.2 Silica Mineralization in Nature

Silica-based materials are used in a manifold of industrially relevant areas like food, polymers composites, adhesives, detergents, and catalysts. In the medical sector, inorganic/organic hybrid materials have been developed for bone augmentation and repair.[4] Another example are mesoporous silica nanoparticles for drug delivery.[5] Despite many advances however, we are still unable to control the structure of silica at the nano-scale level in contrast to primitive single- or multicellular organisms.

Silica mineralization is a widespread phenomenon in nature. It is found in marine sponges dating back 525 million years,[6] diatoms (100 million years)[7] and higher plants. Amongst these organisms, the process of biosilicification is particularly well studied in diatoms, a major group of algae, due to their

facile cultivation. Diatoms can accumulate silica and produce intricately shaped silica cell walls called *frustules*. The frustule is composed of two parts, the *epitheca* and the *hypotheca*, overlapping like the two halves of a petri dish (Figure 1.2).

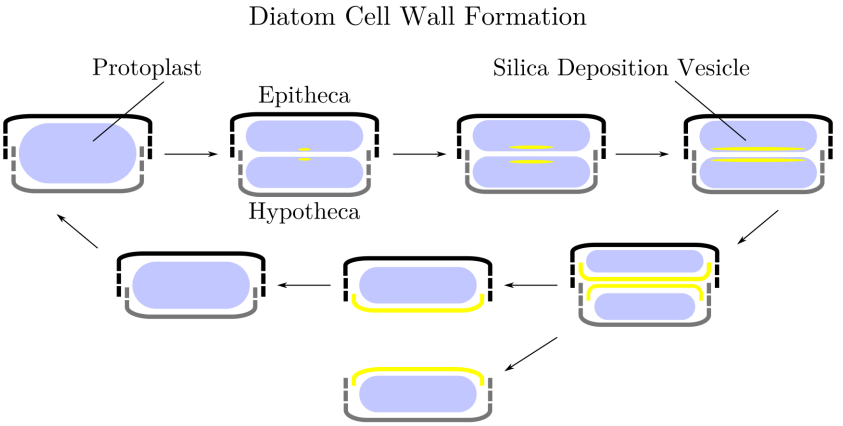


Figure 1.2: Schematic illustration of diatom cell wall formation and cell division (modified from reference [6]). The new epitheca and hypotheca (highlighted in yellow) are produced by the silica deposition vesicle inside the parent frustule.

Cell division can be pictured as the opening of this petri dish. Since one side of each daughter cell would be unprotected, a new epitheca and hypotheca must be formed prior to cell division in the *silica deposition vesicle* (SDV). It has been shown by fluorescent labeling experiments on the diatom *Navicula salinarum* that the newly formed *thecae* reach full two-dimensional size within 15 to 20 minutes.[8] The thickness of the thecae increases continuously over the next 4 h until the cells separate.[8]

Microscopically the frustule appears to be assembled from silica particles roughly the size of 5 nm.[9] These silica particles are embedded in an organic matrix of polypeptides, polyamines and carbohydrates.[10] Specifically, polypeptides and polyamines have been shown to precipitate silica *in vitro*. The chemistry of these two components will be discussed in the following.

Polyamines are dominating the composition of the organic matrix that is associated with silica.[11] Polyamines are linear, oligomeric chains based on propyleneimine units. In the literature, polyamines found in diatoms are commonly referred to as long-chain polyamines (LCPAs). Depending on the diatom species their exact chemical structure can vary.[11-12] It has been shown

that the amino groups present in polyamines catalyze the silica polycondensation at neutral pH by S_N2 substitution.[13] The chemical structure of typical polyamines found in *Cylindrotheca fusiformis* is depicted in Figure 1.3a.

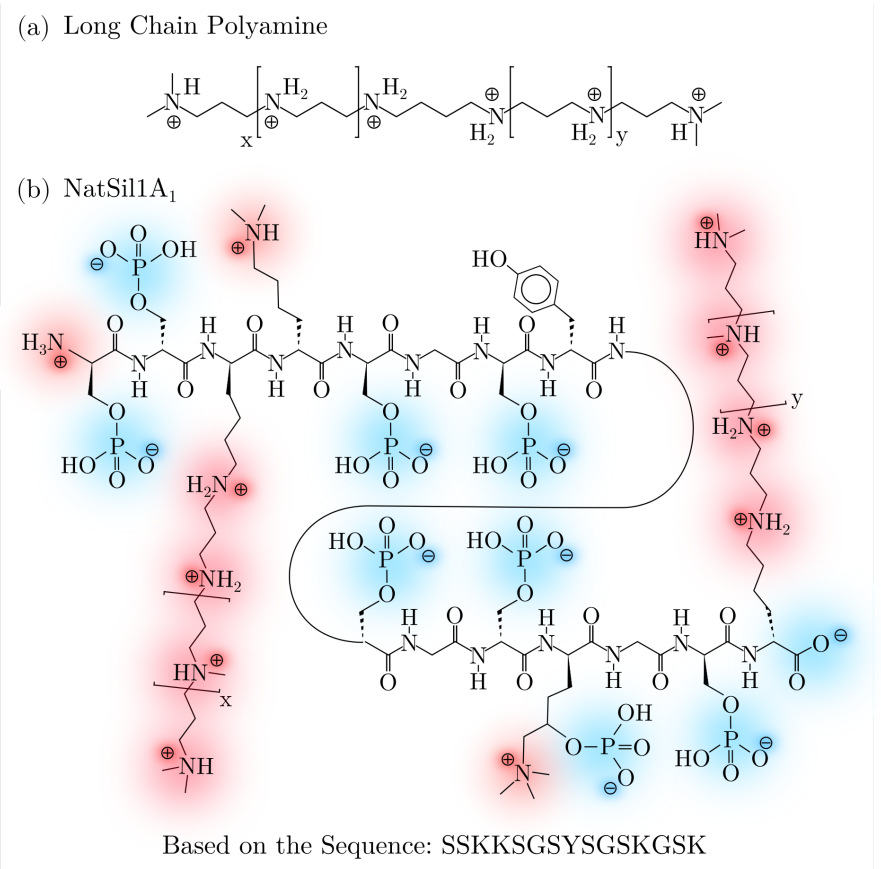


Figure 1.3: (a) The chemical structure of a long-chain polyamine found in *C. fusiformis* (x = 2-7, y = 5-8) and (b) the primary structure of silaffin-1A$_1$. The charges on silaffin-1A$_1$ are assigned according to a pH of 5. This pH was assumed due to the fact that the silica deposition vesicle was found to be at least mildly acidic. Negatively charged parts of the molecule are colored blue, positive parts are colored red.

To explain the intricate patterning found in natural diatom frustules, Sumper et al. have proposed a phase-separation model.[14] This model was later supported by experimental results.[12c, 15] Within this model polyamines

self-assemble in solution, forming micro- and nanodroplets (i.e. an organic template). At the interface of these droplets amino groups catalyze the polycondensation of silicic acid contained in the aqueous phase. Thus, we hypothesize that molecular and macromolecular structure at the interface play an important role in the patterning of silica interfaces. Changing the structure of polyamines represents one route to tailor the surface properties of silica.[16]

The other major class of molecules known to precipitate silica is composed of polypeptides. Polypeptides are chains of amino acids linked by peptide bonds. In general, polypeptides with less than ~50 amino acids are termed peptides. Polypeptides composed of more amino acids or assemblies of polypeptides with biochemical functionality are known as proteins.

Studies on the diatom *C. fusiformis* have revealed that the second most abundant class of molecules in the diatom frustule are *silaffin* peptides.[17] These peptides carry a variety of post-translational modifications. Among others, serines are phosphorylated and lysines are methylated or modified by the addition of polyamines.[18] A particularly well characterized member of the silaffin protein family is *silaffin-1A₁* (a type-1 silaffin). The primary structure of silaffin-$1A_1$ is shown in Figure 1.3b.

Silaffin-$1A_1$ contains many positively as well as negatively charged groups. It has been shown that if silaffin-$1A_1$ lacks the phosphate modifications, additional phosphate ions are required to precipitate silica *in-vitro*. Based on the amino acid sequence of silaffin-$1A_1$ a peptide known as R5 has been synthesized.[19] The primary structure of R5 is shown in Figure 1.4.

R5 has been shown to produce similar silica precipitates *in vitro* as silaffin-$1A_1$.[19-20] The amino acid sequence of R5 includes four additional C-terminal amino acids (RRIL). This four amino acid motif is absent in the mature silaffin-$1A_1$ peptide due to proteolytic processing. A previous study has found that the presence of lysines as well as the four amino acid motif RRIL in unmodified R5 is critical for silica formation.[21] Thus, similar to secondary and tertiary amines in LCPAs, the R5 mediated silica polycondensation is presumably catalyzed by amino groups of the lysine side chains of the peptide. Furthermore, solid-state nuclear magnetic resonance (ssNMR) measurements of R5-silica precipitates indicate that the four amino acid motif RRIL is more likely to be involved in peptide-peptide interactions.[22]

The peptide-directed silica mineralization shares many similarities with polyamine-directed silica mineralization: E.g. similar silica morphologies are obtained and both systems are phosphate dependent.

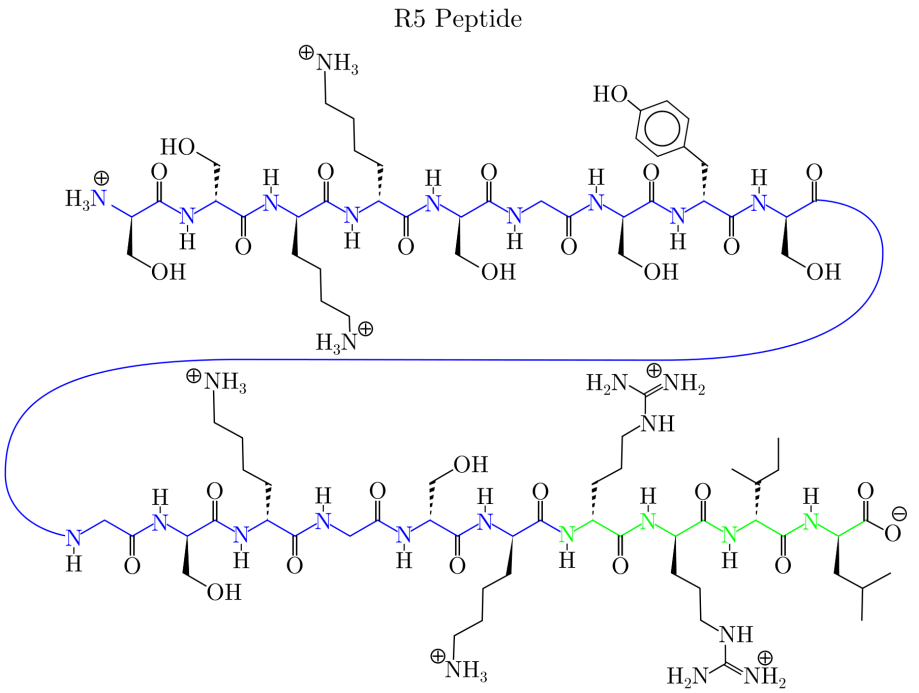


Figure 1.4: The primary structure of unmodified R5 peptide with *in vitro* silica precipitation activity. The backbone of R5 is colored in blue to indicate the sequence identity with silaffin-1A$_1$ shown in Figure 1.3b. The C-terminal RRIL motif (green) is absent in silaffin-1A$_1$.

Thus, it is likely that the peptides self-assemble in solution according to the phase-separation model which has been proposed in reference [14]. As discussed in the context of polyamines, the morphology of silica may depend on the peptide structure and ordering at the surface of a peptide template.

Based upon this hypothesis, we need technique that enables us to obtain information about peptides at surfaces. However, most spectroscopic techniques won't allow one to distinguish interfacial signals from the bulk signal which is typically orders of magnitude larger. In the following section I introduce a spectroscopic method that allows us to overcome this limitation and to study this model interface *in situ*. The second experimental challenge is the formation of a peptide-enriched phase. For the studies presented in this thesis, the peptide-enriched organic phase is approximated by a layer of

peptides adsorbed to the air-water interface. The aqueous solution below contains the silicic acid for the silica formation at the peptide layer.

1.3 Sum Frequency Generation Spectroscopy

1.3.1 Theoretical Background

Many details about the world of molecules and atoms can be derived from their interaction with electromagnetic waves. The electromagnetic spectrum covers a broad range of frequencies with the visible light representing only a small portion of it. Molecules can absorb energy from incident electromagnetic waves if their photon energy equals the energy difference between two molecular quantum states (resonance). There are many degrees of freedom in a molecule (e.g. vibrations, rotations, electronic states, etc.) that can be excited with waves of certain energy. An overview is given in Figure 1.5.

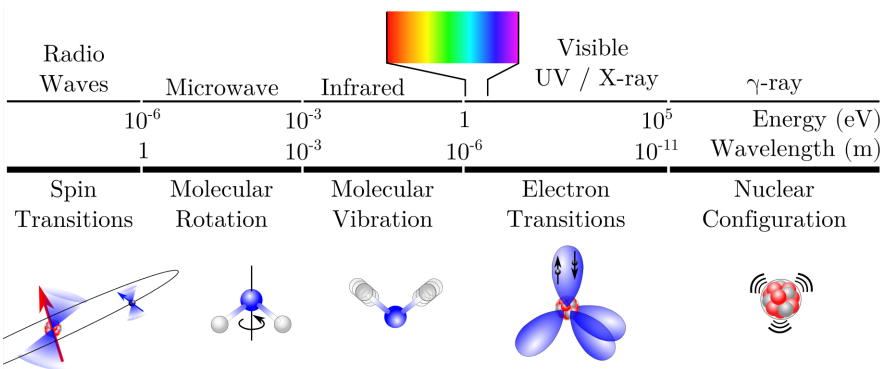


Figure 1.5: Overview of the electromagnetic spectrum in terms of energy and wavelength and corresponding atomic and molecular interactions.[23]

The aim of this thesis is to characterize the structures of peptides at aqueous silica *interfaces*. The structure of a peptide can be probed by measuring its vibrational resonances. The intrinsic *bulk* sensitivity of spectroscopic methods based on *linear* optics makes it very difficult to obtain interfacial information about bulk systems. Therefore, we used the *nonlinear* optical phenomenon of sum-frequency generation (SFG), which allows us to probe selectively interfacial layers of molecules. In this section I introduce the

basics of SFG spectroscopy, in particular vibrational SFG (VSFG). I discuss different aspects of VSFG spectroscopy such as molecular selection rules and interpretation of the spectra. The experimental setup will be explained in the next section.

In all types of spectroscopy the signal is generated by the interaction between matter and incident electromagnetic waves. The electric field of these waves (E-field) polarizes molecules in a given sample. These microscopic molecular polarizations superimpose to produce a macroscopic polarization of the sample. The macroscopic polarization acts as source of new electromagnetic waves.[24] The generated waves which are measured by a detector contain information on molecular properties of the sample. Thus, we have to understand the relation of the macroscopic polarization $\tilde{\boldsymbol{P}}(t)$ to the molecular properties of the sample.* The polarization is dipole moment per unit volume and can be expressed as a power series:[24]

$$\tilde{\boldsymbol{P}}(t) = \varepsilon_0\left(\overleftrightarrow{\chi}^{(1)}\tilde{\boldsymbol{E}}(t) + \overleftrightarrow{\chi}^{(2)}\tilde{\boldsymbol{E}}^2(t) + \overleftrightarrow{\chi}^{(3)}\tilde{\boldsymbol{E}}^3(t) + \cdots\right)$$

or 1.1

$$\tilde{\boldsymbol{P}}(t) = \tilde{\boldsymbol{P}}^{(1)}(t) + \tilde{\boldsymbol{P}}^{(2)}(t) + \tilde{\boldsymbol{P}}^{(3)}(t) + \cdots,$$

where $\tilde{\boldsymbol{E}}(t)$ is the E-field of the incident radiation and the constant ε_0 is the permittivity of vacuum. $\overleftrightarrow{\chi}^{(1)}$ is known as the *linear* first-order optical susceptibility; $\overleftrightarrow{\chi}^{(2)}$ and $\overleftrightarrow{\chi}^{(3)}$ denote the second- and third-order *nonlinear* susceptibilities, respectively.† Equation 1.1 assumes that there is an instantaneous effect of the field strength on the polarization.[24]

The origin of sum-frequency generation can be understood if we consider two electromagnetic waves which are overlapped in space. The resulting electric field can be written as:

$$\tilde{\boldsymbol{E}}(t) = \tilde{\boldsymbol{E}}_1(\boldsymbol{r}, t) + \tilde{\boldsymbol{E}}_2(\boldsymbol{r}, t) \quad 1.2$$

For the sake of simplicity we will consider monochromatic waves propagating in space along $\boldsymbol{r}$. We note that such simplification does not lead to the loss of generality because any arbitrary wave can be represented as linear

* Bold letters are used to denote vectors while regular letters denote scalars. The tilde symbol (~) is used to indicate that a quantity is rapidly varying in time. The tilde is omitted when a quantity remains constant in time.

† The accent denotes a tensor. $\overleftrightarrow{\chi}^{(1)}$ is a two-dimensional array of scalars, $\overleftrightarrow{\chi}^{(2)}$ is a three-dimensional array of scalars, etc.

combination of monochromatic waves. The corresponding electric field can be represented mathematically as a cosine function. Equation 1.2 becomes

$$\tilde{\boldsymbol{E}}(\boldsymbol{r},t) = \boldsymbol{E}_1(\boldsymbol{r})\cdot\cos(\omega_1 t) + \boldsymbol{E}_2(\boldsymbol{r})\cdot\cos(\omega_2 t). \tag{1.3}$$

With ω_1 and ω_2 denoting the frequencies of the waves. $\boldsymbol{E}_1$ and $\boldsymbol{E}_2$ denote the field amplitudes. We recall from Equation 1.1 that the second-order polarization $\tilde{\boldsymbol{P}}^{(2)} = \overleftrightarrow{\boldsymbol{\chi}}^{(2)}\tilde{\boldsymbol{E}}^2(t)$. If we now substitute $\tilde{\boldsymbol{E}}(t)$ from Equation 1.3 we obtain

$$\begin{aligned}\tilde{\boldsymbol{P}}^{(2)} = \varepsilon_0\overleftrightarrow{\boldsymbol{\chi}}^{(2)}[&\boldsymbol{E}_1^2 cos^2(\omega_1 t) + \boldsymbol{E}_2^2 cos^2(\omega_2 t)\\ &+ 2\boldsymbol{E}_1\boldsymbol{E}_2\cos(\omega_1 t)\cdot\cos(\omega_2 t)].\end{aligned} \tag{1.4}$$

Trigonometric rearrangement gives:

$$\begin{aligned}\tilde{\boldsymbol{P}}^{(2)} = \frac{1}{2}\varepsilon_0\overleftrightarrow{\boldsymbol{\chi}}^{(2)}[&(\boldsymbol{E}_1^2 + \boldsymbol{E}_2^2) + \boldsymbol{E}_1^2\cos(2\omega_1 t) + \boldsymbol{E}_2^2\cos(2\omega_2 t)\\ &+ 2\boldsymbol{E}_1\boldsymbol{E}_2\cos\big((\omega_1+\omega_2)t\big)\\ &+ 2\boldsymbol{E}_1\boldsymbol{E}_2\cos\big((\omega_1-\omega_2)t\big)].\end{aligned} \tag{1.5}$$

In Equation 1.5 the first term is time-independent and describes the DC field, known as optical rectification. The last two terms, which oscillate at the frequencies $\omega_1+\omega_2$ and $\omega_1-\omega_2$, reflect sum and difference frequency generation, respectively. The remaining terms which oscillate at the frequency of $2\omega_1$ and $2\omega_2$, respectively, reflect second harmonic generation. The amplitude of the sum-frequency term reads:

$$\boldsymbol{P}_{\mathrm{SF}}^{(2)} = \varepsilon_0\overleftrightarrow{\boldsymbol{\chi}}^{(2)}\boldsymbol{E}_1\boldsymbol{E}_2. \tag{1.6}$$

This polarization oscillating at the sum of both frequencies generates an electromagnetic wave with the electric field $\tilde{\boldsymbol{E}}_{\mathrm{SF}}(t)$. The intensity associated with this wave is the signal which is measured in the experiment. In the experiments of this thesis we use square-law detectors, which measure the intensity of light. The intensity of emitted sum-frequency signal is proportional to the squared amplitude of the electric field ($\boldsymbol{E}_{\mathrm{SF}}$). The amplitude of the electric field is proportional to the amplitude of the polarization (Equation 1.6). Thus, the signal intensity measured by the detector can be expressed as

$$I_{\mathrm{SF}} \propto |\boldsymbol{E}_{\mathrm{SF}}|^2 \propto \left|\overleftrightarrow{\boldsymbol{\chi}}^{(2)}\right|^2. \tag{1.7}$$

Based on Equation 1.7 we can find an explanation why the sum-frequency signal is surface specific, as mentioned earlier. $\overleftrightarrow{\boldsymbol{\chi}}^{(2)}$ is a tensor with 27

components. Each of these tensor components, denoted as $\chi^{(2)}_{ijk}$, mixes a certain combination of Cartesian components of the electric fields $\boldsymbol{E}_{1,j}$ and $\boldsymbol{E}_{2,k}$ to generate the signal field component i. The inherent surface specificity of sum-frequency generation spectroscopy stems from the fact that $\chi^{(2)}_{ijk}$ must change sign under the dipole approximation if i, j and k are inverted:

$$\chi^{(2)}_{ijk} = -\chi^{(2)}_{-i-j-k}. \tag{1.8}$$

In centrosymmetric media, the second-order susceptibility should be invariant under inversion. Thus it can only be zero. But in non-centrosymmetric media the following components of the second-order susceptibility tensor satisfy Equation 1.8 and, thus, can be non-zero: $\chi^{(2)}_{zxx}$ $(= \chi^{(2)}_{zyy})$, $\chi^{(2)}_{xxz}$ $(= \chi^{(2)}_{yyz})$, $\chi^{(2)}_{xzx}$ $(= \chi^{(2)}_{yzy})$ and $\chi^{(2)}_{zzz}$.[25] Strictly speaking, it has been shown that under certain circumstances, e.g. molecular or macroscopic chirality of chromophores, other components of $\vec{\vec{\chi}}^{(2)}$ can be non-zero as well.[26]

In the case of a centrosymmetric medium, e.g. a solution of molecules, the inversion symmetry is broken for a thin region due to the presence of an interface. In the experiments described in this thesis sum frequency generation is typically used to probe the air-water interface of a peptide containing solution. Peptides with amphiphilic character tend to orient at the air-water interface due to the hydrophobic effect. Therefore, such an ordered layer of peptides is non-centrosymmetric and in principle allows sum frequency generation. A schematic representation of sum frequency generation by such a layer of molecules is shown in Figure 1.6.

The incident electromagnetic waves are in the same incidence plane. The Cartesian coordinate system is defined with respect to the sample interface in Figure 1.6. We recall that $\chi^{(2)}_{ijk}$ generates a sum frequency electric field component i (along x, y or z) by mixing x, y and z components of incident electric fields present at the interface. By rational choice of polarizations for incident waves it is possible to probe single components of the second-order susceptibility tensor. The electric field of incident waves is typically chosen to be polarized parallel (p) or perpendicular (s, from the German equivalent "senkrecht") to the plane of incidence. At the interface the p-polarized field gives rise to a local field with x and z component. The s-polarized field gives rise to a local field with y component, only.

In the experiments described in this thesis I obtain information about peptide structure or side chain orientation at the air-water interface. To this end, I measure vibrational transitions with the incident electromagnetic waves.

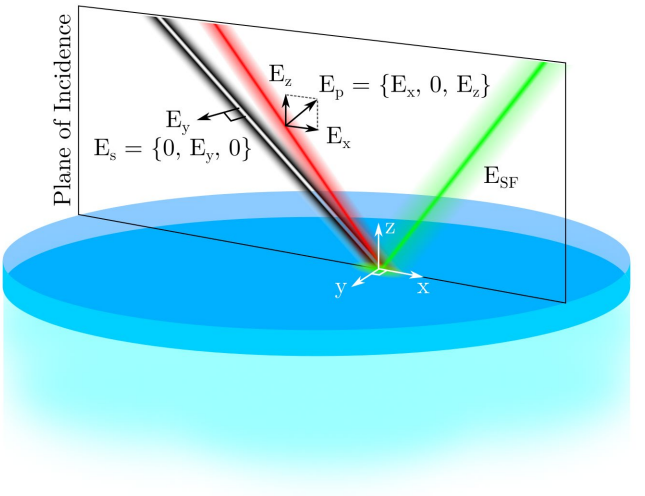


Figure 1.6: Scheme of sum frequency generation in reflection at an air-water interface. A non-centrosymmetric layer of molecules is depicted in dark blue. Centrosymmetric medium below is depicted in light blue. The polarization of the two incident electromagnetic waves was chosen arbitrarily.

Vibrations can be excited with infrared light. Thus, for sum frequency generation, the frequency of one of the electric fields should be in the infrared region of the electromagnetic spectrum. For the experiments conducted in this thesis I use a broadband IR laser pulse which can excite several vibrational transitions at once. The second wave was chosen at a non-resonant, fixed frequency in the visible (VIS) region of the electromagnetic spectrum. In this respect, by convention, SFG, VIS and IR electric field components are denoted by the indices i, j and k in $\chi^{(2)}_{ijk}$, respectively. Thus, in order to probe $\chi^{(2)}_{xxz}\left(=\chi^{(2)}_{yyz}\right)$, we have to detect s-polarized sum frequency signal, using an s-polarized VIS and a p-polarized IR beam.

To derive the molecular structure of an interface we need to understand how $\vec{\vec{\chi}}^{(2)}$ depends on the molecular conformation. $\vec{\vec{\chi}}^{(2)}$ stems from the molecular hyperpolarizability $\vec{\vec{\beta}}$. In the surface-bound coordinate system the components of $\vec{\vec{\beta}}$ (β_{ijk}) depend on the orientation of the molecule. The macroscopic quantity $\chi^{(2)}_{ijk}$ is obtained by averaging β_{ijk} over all molecules:[27]

$$\chi^{(2)}_{ijk} \propto \langle \beta_{ijk} \rangle, \qquad \text{where } \langle \beta_{ijk} \rangle = \frac{1}{2\hbar} \frac{\langle M_{ij} A_k \rangle}{\omega_\nu - \omega_{\mathrm{IR}} - \mathrm{i}\Gamma} \,. \tag{1.9}$$

Averaging is denoted by $\langle\ \rangle$ brackets and depends on the orientation distribution of molecules. The expression for β_{ijk} is derived with second-order perturbation theory in reference [28]. Γ is the dipole dephasing rate, which includes relaxation of the excited state population and dephasing due to processes, not associated with population transfer. ω_ν is the frequency of a resonant molecular vibration and ω_{IR} is the frequency of the incident infrared light. Since β_{ijk} is directly proportional to the Raman- (M_{ij}) and infrared (A_k) transition dipole moments the first selection rule for VSFG is that only Raman- and infrared-active molecular vibrations can appear in VSFG spectra. Furthermore, the denominator in Equation 1.9 gives rise to intensity maxima in the sum-frequency spectrum when $\omega_{\mathrm{IR}} = \omega_\nu$.

Apart from the resonant (R) response of interfacial molecules it is possible to observe a non-resonant (NR) sum-frequency response with little dependence on frequency. As a consequence the susceptibility can be expressed as

$$\chi^{(2)}_{ijk} = \chi^{(2)}_{R,ijk} + \chi^{(2)}_{NR,ijk}. \tag{1.10}$$

For the signal intensity of a single resonance we substitute $\vec{\boldsymbol{\chi}}^{(2)}$ in Equation 1.7 by Equation 1.10 and obtain

$$I_{\mathrm{SF}} \propto \left|\chi^{(2)}_{R,ijk} + \chi^{(2)}_{NR,ijk}\right|^2. \tag{1.11}$$

Now we use Equation 1.9 for $\chi^{(2)}_{R,ijk}$ and the constant $C = \langle M_{ij} A_k \rangle / 2\hbar$. By doing so we obtain

$$\boldsymbol{I}_{\mathrm{SF}} \propto \left|\tfrac{C}{\omega_\nu - \omega_{\mathrm{IR}} - \mathrm{i}\Gamma} + \left|\chi^{(2)}_{NR,ijk}\right| e^{\mathrm{i}\phi}\right|^2. \tag{1.12}$$

By convention, $\chi^{(2)}_{NR,ijk}$ – being a complex quantity – is expressed by polar coordinates in complex space, i.e. $\chi^{(2)}_{NR,ijk} = \left|\chi^{(2)}_{NR,ijk}\right| e^{\mathrm{i}\phi}$. With Equation 1.12 we can fit an experimental spectrum with a single resonant absorption. To fit multiple peaks the following expression is used throughout the thesis:

$$\boldsymbol{I}_{\mathrm{SF}} \propto \left|\sum_i \frac{C_i}{\omega_i - \omega_{\mathrm{IR}} - \mathrm{i}\Gamma_i} + \left|\vec{\boldsymbol{\chi}}^{(2)}_{\mathrm{NR}}\right| e^{\mathrm{i}\phi}\right|^2 \tag{1.13}$$

1.3.2 The Experimental Setup

Nonlinear optical effects like sum-frequency generation can only be detected with highly intense external electromagnetic fields. The following paragraphs describe how these fields are generated at the sample surface and how the

sum-frequency emission from the sample is detected and analyzed. A scheme of the experimental setup is shown in Figure 1.7.

The setup is based on a regenerative amplifier (*Spitfire Ace PA*, Spectra-Physics, Santa Clara, CA, USA) which produces 800 nm pulses with an energy of 9 mJ per pulse at 1 kHz repetition rate. An oscillator (*Mai Tai*, Spectra Physics) generates the seed pulse for chirped pulse amplification. At first the seed pulse is stretched in time to prevent damage to the amplifier crystal. The seed pulses are amplified by exciting an amplifier crystal with a pump pulse (*Empower*, Spectra Physics) shortly before each seed pulse arrives at the crystal to cause stimulated emission. The amplified pulse is again amplified in a second crystal and subsequently compressed to around 40 fs. 2 mJ of the total output is used for our experiments. Out of this total 2 mJ, 1 mJ is directed through an etalon to produce a narrow-band laser beam (VIS) with a full width at half maximum (FWHM) of 20 cm^{-1}. The other 1mJ is directed into a parametric amplifier (*TOPAS/NDFG*, Light Conversion, Vilnius, Lithuania). The output of the parametric amplifier is an infrared laser beam (IR) with a FWHM of 150 cm^{-1} and 700 cm^{-1} when tuned to 1650 cm^{-1} and 3200 cm^{-1}, respectively.
We control the polarization of both the IR and the VIS beam at the sample by using a polarizer and a half-wave plate. Overlapping the IR and VIS beam spatially and temporally on the sample surface at angles of incidence of 60 and 55° respectively produces a sum-frequency response. The sum-frequency field emitted by a sample is collimated with a lens. After passing through a half-wave plate and a polarizer the sum-frequency signal is focused with a lens onto the entrance slit of the spectrometer (Shamrock, Andor, Belfast, UK) coupled to a CCD camera (*Newton 970*, Andor). During the measurement, the IR beam was kept under an atmosphere of dry nitrogen along its path to the sample.

It is possible for a fraction of scattered VIS light to reach the detector. This scattered light generates a background signal. To derive correct molecular spectra this background needs to be subtracted. To this end we measure a background spectrum with IR beam blocked and only the VIS beam incident on the sample. The background spectrum is then subtracted from the sum frequency generation measurement with both IR and VIS beam incident on the sample.

After the background is subtracted we need to take into account that the intensity of the broad-band IR laser pulse varies with frequency. Thus, the measured sum frequency signal has to be normalized with the intensity profile of the IR pulse.

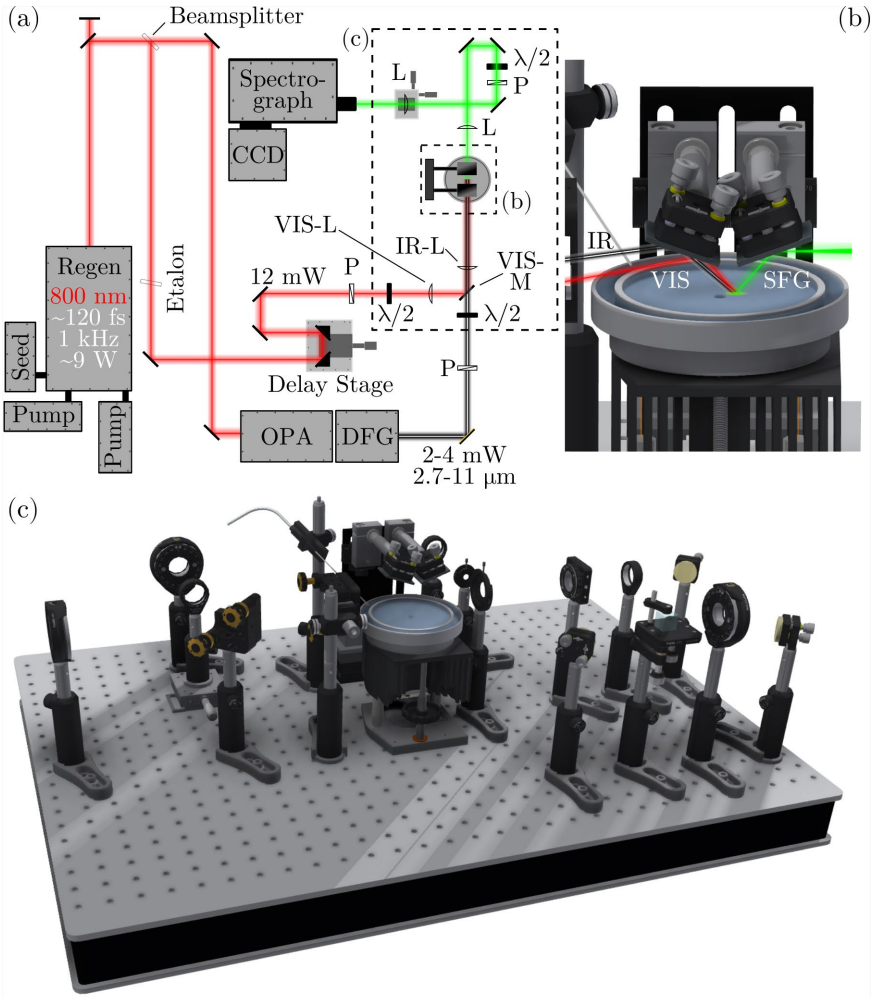


Figure 1.7: (a) Overview of the VSFG setup. The VIS mirror, where the VIS and IR beam appear to overlap on the scheme, is cut so that the VIS beam (red) is reflected while the IR beam (black/white) passes over it. The IR lens is cut as well so that only the IR beam is focused and the VIS beam can pass under the lens. The SFG beam is depicted in green. (b) Close up front view of the sample stage. (c) 3D image of part of the optical setup where the sample is located. The rendered images (b) and (c) were kindly provided by Marc-Jan van Zadel.

This normalization is done by first measuring the sum-frequency response of a z-cut quartz crystal. The sum frequency data obtained from a sample is then divided by the background-corrected quartz spectrum.

By convention the sum-frequency intensity is plotted versus the corresponding IR wavenumber on the x-axis. The x-axis is calibrated by measuring the VSFG response of a z-cut quartz crystal in ambient air in the spectral range of water vapor IR absorption (near 1650 cm^{-1}). Due to the modest humidity of ambient air, IR light is absorbed at certain frequencies by vibrational transitions of water. This absorption results in minima of the VSFG spectrum at these frequencies. These minima can be calibrated using the IR absorption spectrum of water vapor. The same calibration is then used for the peptide samples.

1.4 Molecular Modeling

1.4.1 Theoretical Background

Molecular modeling can significantly facilitate the interpretation of experimental data such as the VSFG spectra presented in this thesis. However, the biophysical processes addressed in this thesis can only be observed in simulations of complex systems over long periods of time (100-1000 ns). Thus, certain approximations have to be made to effectively use the computational resources at hand.

In classical *molecular dynamics* (MD) we make two main approximations: 1) Electron dynamics are neglected, i.e. electrons are assumed to adjust instantaneously to the motion of nuclei and remain in the ground state. 2) The motion of nuclei is described by classical mechanics.[29] In essence, a classical MD simulation calculates the forces ($\boldsymbol{F}$) on any of the N atoms contained in a virtual box. This calculation is based on a given set of velocities ($\boldsymbol{v}$), atomic positions ($\boldsymbol{x}$) and the potential interaction function $V(x_1, x_2 \ldots x_N)$:[30]

$$\boldsymbol{F}_i = -\frac{\partial V}{\partial \boldsymbol{x}_i}, \text{with } i = 1 \ldots N \text{ atoms.} \tag{1.14}$$

The atom positions in space are updated by numerically integrating the equations of motion for each time step, given by

$$\frac{\partial \boldsymbol{x}_i}{\partial t} = \boldsymbol{v}_i \text{ and } \frac{\partial \boldsymbol{v}_i}{\partial t} = \frac{\boldsymbol{F}_i}{m_i}\,. \tag{1.15}$$

The mass of atom i is denoted as m_i, the atom coordinate is denoted as $\boldsymbol{x}$ and the force that acts on the mass is $\boldsymbol{F}_i$. A record of the coordinates of all atoms at each integration time step is called a *trajectory*.

As is seen in Equations 1.14 and 1.15, the molecular motions are dependent on the potential interaction function V. In a classical MD simulation V is the sum of several energy terms describing the interaction potential in a many-body system as a function of distance or angles between atoms. Thus we can write:

$$V = (V_s + V_\theta + V_\varphi + V_\omega)_b + (V_{el} + V_{VdW})_{Nb}. \quad 1.16$$

Equation 1.16 includes bonded as well as non-bonded interactions. The bonded interactions are: the bond stretching potential (V_s), the bond angle potential (V_θ), the dihedral angle potential (V_φ) and the improper torsion potential (V_ω). The non-bonded interactions are: the electrostatic potential (V_{el}) and the Van der Waals potential (V_{VdW}). Each of the aforementioned potentials is defined by an equation containing constants, which are commonly referred to as parameters. The ensemble of functions and parameter sets is commonly known as the *force field*.

Equation 1.16 can also include constraining potentials on bond lengths and angles. These potentials serve to prevent large atom displacements during minimization and allow for larger integration time steps in Equation 1.15. The integration time step can be increased further by removing additional fast degrees of freedom and by reducing the number of particles in the system using coarse-grained force fields. In contrast to atomistic force fields, several atoms are combined into one coarse-grained "bead" that encapsulates the forces exerted by the several atoms.

For most systems studied in this thesis however, an atomistic force field model was used. Atomistic force field implementations optimized for the simulation of peptide behavior are CHARMM, AMBER and OPLS/AA.[31] A detailed account on the parameterization of these force fields is given in reference [32]. In reference [33] these force fields are benchmarked with solution-state NMR data.

Force field development is an ongoing process, and improvements are made continuously to the major force field families discussed here. Although the major force fields produce comparable results, there may be differences in performance when an interface is introduced into the system. In this thesis, the problem of choosing a force field for systems containing an air-water interface will be addressed. Specifically, MD will be used to generate peptide structural ensembles using different force fields in Chapter 5. Theoretical VSFG spectra will be generated using those ensembles, and the results will be compared to experiment.

1.4.2 Enhanced Sampling

In an ideal (infinitely long) simulation it should be possible to visit all points in *phase space*.* Such a simulation is termed ergodic: The time average of states in the trajectory would equal the ensemble average of states in phase space. In turn this would allow calculating the free energy of a state from the probability of how often this state is visited. However, the phase space of systems modeled in this thesis is enormous and an ergodic simulation cannot be attained.[34] It is very likely that most of phase space will remain unexplored. This limitation can be overcome partly by methods of *enhanced sampling*. Two of these methods are employed in Chapters 4 and 5: *Parallel Tempering* and *Metadynamics*.[35]

In metadynamics,[35c] we define a number of *collective variables* (CVs) that characterize interesting areas of phase space. For example the radius of gyration of a peptide chain can be used to define such an area of phase space. The simulation explores these areas of phase-space by adding a cumulative bias potential to already visited points in the energy landscape of the respective CV space. The total potential reads:

$$V(\boldsymbol{x}) = V_0(\boldsymbol{x}) + \Delta V(\boldsymbol{x}, t). \tag{1.17}$$

$V_0(\boldsymbol{x})$ is the potential given by the force field model and $\Delta V(\boldsymbol{x}, t)$ is the history-dependent bias potential. As a consequence of the time-dependence, the bias grows progressively and prevents the system from staying in one location in CV space. The bias potential at time t can be expressed as a sum of Gaussian functions with a deposition stride of τ, d collective variable functions S of the coordinates $\boldsymbol{x}$, a Gaussian width of σ_i and a Gaussian height of W:

$$V[\boldsymbol{S}, t] = \sum_{k=1}^{\frac{t}{\tau}-1} W e^{-\sum_{i=1}^{d} \frac{(S_i - S_i[\boldsymbol{x}(k\tau)])^2}{2\sigma_i^2}} \tag{1.18}$$

In the long run the bias converges to

$$V(\boldsymbol{S}, t \to \infty) = F(\boldsymbol{S}) + C, \tag{1.19}$$

where $F(\boldsymbol{S})$ is the negative free energy and C is a constant.

* Phase space is a $6N$-dimensional space, where N is the number of atoms and 6 is the number of degrees of freedom (the x, y and z coordinates and three momentum components).

In Equation 1.18 we successively add Gaussians of constant height, thus, the free energy estimate $F(\boldsymbol{S})$ oscillates around its real value and therefore has a certain error related to the bias growth rate and the CV diffusion coefficient. The error can be reduced by decreasing the height of the deposited Gaussians over time.[36] The invariant W in Equation 1.18 is replaced by

$$W(k\tau) = W_0 e^{-\frac{V(\boldsymbol{S}[\boldsymbol{x}(k\tau)],k\tau)}{k_B \Delta T}}, \tag{1.20}$$

with the initial Gaussian height W_0, the input parameter ΔT in units of temperature and the Boltzmann constant k_B. This approach is commonly referred to as well-tempered metadynamics. Once converged, the CV of interest is explored as if it were at an elevated temperature ($T + \Delta T$), thus overcoming free energy barriers preventing important structural transitions. The potential changes to

$$V(\boldsymbol{S}, t \to \infty) = -\frac{\Delta T}{T + \Delta T} F(\boldsymbol{S}) + C, \tag{1.21}$$

where T is the system temperature. The biasfactor $\gamma = (T + \Delta T)/T$ is typically used in the experimental description of a well-tempered metadynamics simulation (e.g. in Chapter 5).

Parallel Tempering is a different method of enhanced sampling which attempts to accelerate the sampling of phase space by evolving several replicas of a simulation in parallel at a range of temperatures.[35a, 35b] At certain intervals the program attempts to exchange molecular coordinates of neighboring temperatures with the probability $\alpha_{i \leftrightarrow j}$ according the Metropolis criterion:[34]

$$\alpha_{i \leftrightarrow j} = \min\{1, e^{\Delta^{PT}}\}, \tag{1.22}$$

$$\text{where } \Delta^{PT} = \left(\frac{1}{k_B T_i} - \frac{1}{k_B T_j}\right) [U(\boldsymbol{x}_i) - U(\boldsymbol{x}_j)]. \tag{1.23}$$

$\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ denote collective coordinates at the temperatures T_i and T_j with the internal energies U_i and U_j. The coordinate swap among the temperature replicas allows the system to overcome energy barriers at high temperature while maintaining Canonical sampling at lower, more relevant temperatures. Equation 1.22 and 1.23 suggests that an efficient exploration of phase space, i.e. a high exchange probability requires a good overlap of energy distributions of two replicas. It has been shown that a high exchange probability increases the efficiency with which the ensemble is explored.[37] However, with larger systems this overlap can be difficult to attain. The gap

in potential energy between two adjacent replicas increases with the number of particles in the system (N) while the fluctuation of the potential energy increases as the square root of N.[38]

The problem of scalability can be overcome by combining parallel tempering with a form of metadynamics known as the *Well-Tempered Ensemble* (WTE).[39] In the WTE, we bias the potential energy of a system using well-tempered metadynamics. Thus, the potential energy fluctuations of the system increase. In combination with parallel tempering (PT-WTE) the energy overlap between replicas increases, and therefore the exchange probability is enhanced.[39] To account for the additional bias potentials V on the ith and jth replicas Δ^{PT} has to be replaced with

$$\Delta^{PTMetaD} = \Delta^{PT} + \frac{V^i\big(S(\boldsymbol{x}_i,t)\big) - V^i\left(S(\boldsymbol{x}_j,t)\right)}{k_B T_i} + \frac{V^j\big(S(\boldsymbol{x}_j,t)\big) - V^j\big(S(\boldsymbol{x}_i,t)\big)}{k_B T_j}. \tag{1.24}$$

The average energy of the system remains near that of the Canonical Ensemble.[39] Thus, average ensemble properties are similar in the Canonical Ensemble and in the WTE. The method of PT-WTE was applied in Section 3.4 to ensure the sampling of a diverse peptide structural ensemble.

1.4.3 Considerations for Implementation

The initial state of a system is defined with two files: A coordinate file containing a list of all atoms and their positions in Cartesian coordinates and a topology file containing the force field parameters and defining atomic bonds. However, the edges of a box introduce inhomogeneity into the system, i.e. sudden jumps in the intermolecular interaction potential where the edge of the system experiences a vacuum. This problem is circumvented to a large extent by applying *periodic boundary conditions*: We can imagine this mathematical trick as surrounding the original system with infinitely many translated copies of itself. With periodic boundary conditions molecules not only "feel" the image of other molecules through the box boundary but can also travel through the boundary and end up on the opposite side of the box.[29] With increasing system size this tends to introduce a much smaller error than a discontinuity introduced by a fixed boundary would.

For the simulations in this thesis, the *minimum image convention* is applied: Non-bonded short-range interaction potentials only take into account the nearest image of a particle. As a consequence, the cut-off distance for

non-bonded forces should be shorter than half the length of the shortest box edge. Thus, a molecule cannot "feel" both sides of a neighboring molecule through the respective neighbors' image.

The Coulombic interaction decays slowly in space since it is inversely proportional to the distance between two charged particles. For periodic systems it is thus computationally expensive to calculate the electrostatic potential accurately. In the so called smooth *particle-mesh Ewald* (PME) method,[40] the electrostatic potential is split into a short-ranged term, summed over all particles in real space and a long-ranged term, summed within a Fourier transformation. Both terms converge rapidly and therefore enable an accurate computationally inexpensive calculation of the electrostatic forces. The charge density part of the long-ranged term is transformed using the *fast Fourier transform*, where charge densities are assigned to a grid (mesh) using cardinal B-spline interpolation.

Most of the simulations presented in this thesis aim to accurately represent the behavior of peptides at the air-water interface. To obtain an air-water interface (in simulations: vacuum-water interface), the box z-dimension can be simply extended in the coordinate file, thus creating two vacuum-water interfaces. With periodic boundary conditions, the system can be pictured as infinitely many stacked sheets of water, extending in the xy-plane and separated in the z-dimension by the additional vacuous space.

In general, force fields were not intended to be used for this so called *slab*-geometry. Therefore we have to critically consider the implications of classical force fields on the interaction of molecules in such systems. The cut-off radius for non-bonded interactions is typically smaller than the size of the box in the z-dimension by one order of magnitude. Therefore, only the electrostatic interaction potential could introduce an error in the force calculations. To minimize these artifacts, the size of the vacuous space in between the infinite slabs is made large enough that only a very small electrostatic force is experienced between the top of one slab and the bottom of the next. The error arising from this approximation is likely much smaller than the errors introduced by imperfections in the original assumption of a fixed point charge electrostatic model or in the parameterization of the force field using the HF/6-31G* level of theory.

1.5 Outline of this thesis

The following Chapters 2, 3, and 4 present the results of three studies related to the biomineralization of silica: In Chapter 2 we present a strategy to mimic interfacial silica biomineralization with artificial peptides of different folding at the air-water interface. The work has been published in the journal *Advanced Materials Interfaces*. In Chapter 3 we examine the effects that even small changes to a peptide can have on bulk silica mineralization. We published this research in the *Journal of Peptide Science*. Taking one step further towards natural biosilicification, Chapter 4 presents a structural investigation of the R5 peptide – a derivative of the diatom peptide Silaffin. Within the model system of peptide induced biomineralization at the air-water interface we studied structural changes and interactions of R5 with silica. This research has been published by the journal *Angewandte Chemie*. Based on these studies it became clear that we have to critically question the model parameters used for atomistic molecular simulations of peptides at the air-water interface. Chapter 5 presents the results of a work in progress where the aim is to benchmark the simulation parameters to make the choice of computational parameters in current research less arbitrary.

2 Interfacial Silica Biomineralization by Artificial Peptides

Acknowledgement: Dr. Vance Jaeger's contribution was in the design, conducting and analysis of the molecular dynamics simulations. SFM data was acquired and interpreted by Helma Burg and Dr. Rüdiger Berger. Gunnar Glasser and Katrin Kirchhoff acquired and interpreted SEM and TEM images, respectively. XPS was measured and interpreted by Prof. Tobias Weidner and Dr. Hao Lu. My contribution to this work was the acquisition of VSFG spectroscopy data and the analysis thereof. In addition, I synthesized the peptides together with Sabine Pütz and assisted Dr. Vance Jaeger in conducting the molecular dynamics simulations.

2.1 Introduction

Silica based materials are typically deposited and structured using extreme pH conditions and temperatures.[41] Biological organisms, on the other hand, can precisely control the composition and morphology of minerals like silica in ambient conditions.[18-19] Fascinating and well-studied silica mineralizing species are diatoms, marine unicellular organisms capable of growing intricately shaped silica cell walls.[19, 42] A general introduction to silica biomineralization in diatoms was given in Section 1.2. Biochemical studies of diatom cell walls identified two classes of molecules that precipitate silica *in vitro*: Polyamines and peptides. These molecules likely self-assemble in solution which results in a phase-separation of organic and aqueous phase.

The surface-bound polycondensation of silica is presumably dictated by the interface of this biphasic system. This interface is approximated in the following by a layer of molecules adsorbed at the air-water interface.

In this chapter we chose peptides to study the effect of different molecular conformations on the interfacial biomineralization of silica. The diatom peptide R5, introduced in Section 1.2, appears to be an ideal candidate based on its silica mineralizing activity. However, the conformation of R5 in solution or within silica precipitates does not appear to fall into the classical categories of helix or beta strand.[22, 43] In fact, it was characterized to adopt a random coil conformation in solution.[43] Thus, the conformation of R5 would introduce unnecessary complexity to a study of the influence of peptide conformation on interfacial silica mineralization.

Recent research has focused on artificial model peptides inspired by diatom biosilicification for *in vitro* silica formation.[44] A particularly simple mineralization model system is based on amphiphilic peptides composed of hydrophobic leucine (L) and hydrophilic lysine (K).[44a] These leucine-lysine peptides (LKs) can adopt a variety of secondary structures at the air-water interface depending on the hydrophobic periodicity of the amino acid sequence.[45] Recent work has shown that LKs of different folds can form different aggregate shapes, which allows the controlled precipitation of nanoscale silica particles with shapes ranging from spheres to rods and wire-like structures.[44a, 46]

LK peptides have been designed to bind and fold into specific structural motifs at the air-water interface and have been studied at surfaces extensively.[45, 47] Here we demonstrate that LK peptides can template nanometer thin silica films at the air-water interface. We compare film morphologies obtained from peptides with two different conformations – α-helical LKα14 (Ac-LKKLLKLLKKLLKL-COOH) and β-strand LKβ15 (Ac-LKLKLKLKLKLKLKL-COOH), shown in Figure 2.1.[45, 47c] The abbreviation "Ac-" denotes an acetylation of the N-terminal amino group of the peptides.

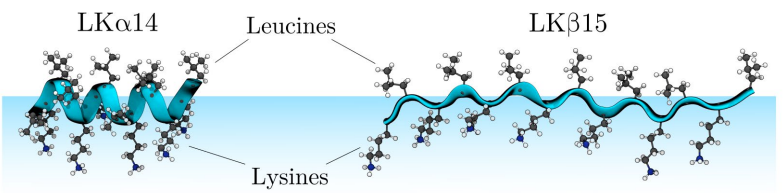


Figure 2.1: Secondary structure of LKα14 and LKβ15.

2.2 Results and Discussion

Figure 2.2 illustrates the sheet preparation procedure. First, an LK peptide film is assembled at the air-water interface. The LK concentration in the media below the LK film is reduced by repeated replacement of the subphase solution with phosphate buffered saline (PBS). Subsequently, a tetramethyl orthosilicate (TMOS) solution is injected into the trough (see Section 2.4 for details of the procedure). The LKα14 and LKβ15 templated silica sheets, obtained almost instantly after TMOS injection, are self-supported and can be removed from the air-solution interface and deposited onto a solid substrate.

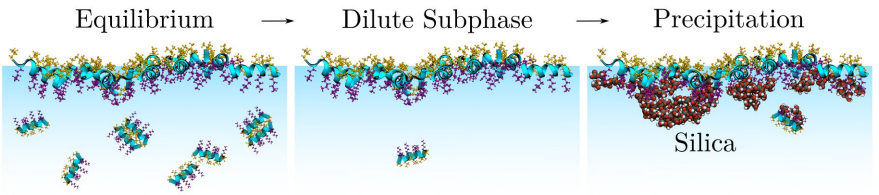


Figure 2.2: The peptide concentration in the bulk is reduced by subphase exchange, as it leads to undesired bulk biomineralization. Subsequently, TMOS is injected for the interfacial biomineralization of silica.

Scanning electron microscopy (SEM) was performed and interpreted by G. Glasser. SEM images for LKα14 and LKβ15 films deposited onto gold-coated silicon wafer pieces (Figure 2.3) provide more detailed information about the sheet morphologies. While LKα14 generated homogeneous, granular sheets, the LKβ15 sheet contains wire-like features similar to the fibrils observed for LKβ15 biosilicification reactions in solution.[44a]

Transmission electron microscopy (TEM) was conducted and interpreted by K. Kirchhoff. TEM images of self-supported biosilica sheets deposited onto a TEM grid are shown in Section 2.5 - Figure 1. The LKα14 sheet has a 'wrinkled' but homogeneous morphology, while the LKβ15 sheet appears folded with a string-like substructure. The wrinkles and folds are, in all likelihood, artifacts from the drying process.

The surface roughness of the newly formed peptide-silica sheets was studied with scanning force microscopy (SFM). SFM was performed and interpreted by H. Burg and R. Berger.

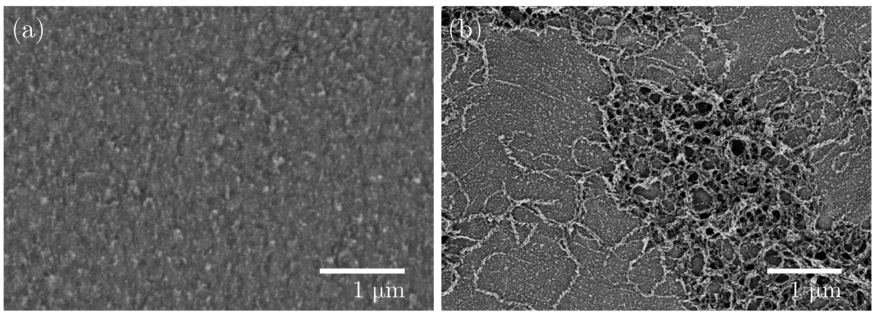


Figure 2.3: Biosilica films fabricated using LKα14 and LKβ15 peptides show different morphologies. SEM image of sheets prepared from LKα14 (a) and LKβ15 (b), deposited onto a gold-coated substrate.

Scanning across 110 nm of an LKα14-grown silica-sheet (Figure 2.4a) reveals that the surface consists of spherical structures roughly 2 nm in height. Consistent with this observation, Zane et al. reported solid-state nuclear magnetic resonance (SSNMR) data that suggests LKα14 forms a 4-helix bundle structure, approximately 2 nm in size in solution.[46] We speculate the spherical morphology is the result of LKα14-tetrameric interactions within the forming silica sheet. The LKβ15-grown silica-sheet (Figure 2.4b) appears to be composed of globular structures lined up like pearls on a string, indicating that protein aggregates are involved in the sheet assembly process.

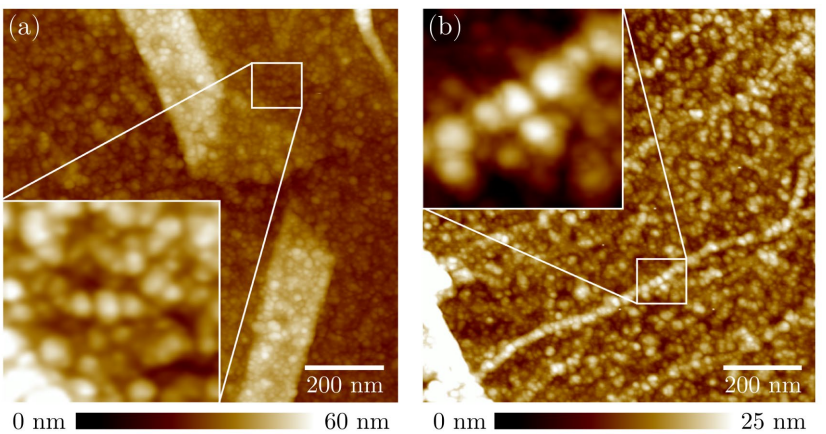


Figure 2.4: SFM images of biosilica films on gold coated substrates. Sheets prepared from LKα14 and LKβ15 are shown in (a) and (b), respectively.

SFM height profiles (Section 2.5 - Figure 2 and Figure 3) of the LKα14 and LKβ15 templated silica-sheets have an average film thickness of 2.8 nm and 2.1 nm, respectively. The silica-sheets produced by LKs are an order of magnitude thinner compared to previous attempts to grow silica films using surfactants and nanotubes, which are typically several hundred nanometers thick.[48]

X-ray photoelectron spectroscopy (XPS) was performed and interpreted by T. Weidner and H. Lu. XPS depth profiling of the sheets (Section 2.5 - Figure 4) shows that for LKα14 the silica sheet grows with silica and peptide contents distributed throughout. The LKβ15 sheet depth profile shows higher etch rates for the nitrogen component (related to the peptides) compared to the mineral-related silicon signal. This indicates that more LKβ15 is located near the water-silica interface during growth compared to the LKα14 sheets. The observation of distribution of peptide throughout the biosilica sheet – on a length scale of 3 nm – indicates the interfacial nucleation and growth process is different from silica precipitation occurring at grown peptide fibrils.[48b, 48c]

For a more detailed understanding of the molecular process involved in the film formation, we used VSFG spectroscopy. Details about the method are provided in Section 1.3. The obtained vibrational spectrum is inherently surface sensitive and will only contain features from preferentially oriented interfacial molecules. Here, we use VSFG to monitor side chain interactions and secondary structure of the LK sheets during mineralization.

Surface spectra of a solution of LKα14 and LKβ15 (0.025 mg/ml), before and after injecting the silica precursor TMOS, are shown in Figure 2.5. Exchanging the subphase of the solution three times to reduce the bulk solution concentration prior to the TMOS injection did not change the spectra. In agreement with previous studies of LK peptide films on hydrophobic surfaces, resonances near 2880, 2904, and 2921 cm^{-1} are observed and can be related to CH-stretching vibrations from ordered leucines pointing towards the air.[47c, 47d, 49] The broad spectral feature seen in Figure 2.5a for LKα14 in the range of 3000-3300 cm^{-1} is the result of stretching modes of differently hydrogen bonded water OH groups.[50] We verified that the resonances are related to water and not the LKs by changing the ionic strength of the PBS buffer, thereby depleting the water signal by altering the local electrical field at the interface (Section 2.5 - Figure 5).[51] Remarkably, no substantial water related resonances are observed in the case of LKβ15 (Figure 2.5b).

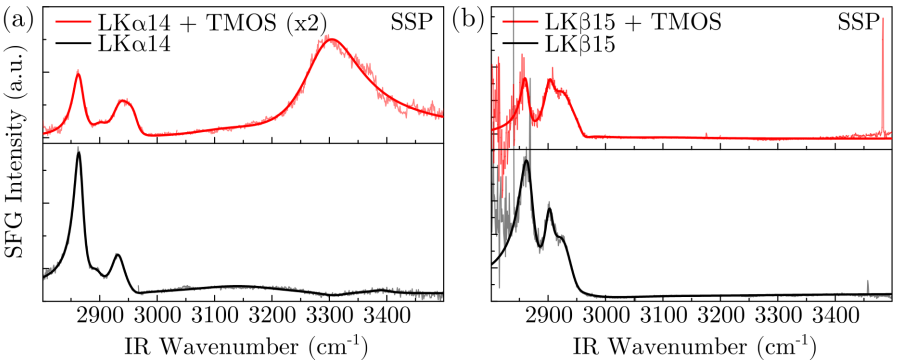


Figure 2.5: VSFG spectra before (black) and after TMOS addition (violet) in the spectral range of CH and water vibrations: LKα14 (a) and LKβ15 (b) at the air-water interface.

Upon TMOS addition, a peptide-silica film is formed at the surface. In the case of LKα14, a new feature appears at ~3300 cm^{-1} attributed to NH stretching vibrations of the lysine side chains (Figure 2.5a).[52] The emergence of the NH peak suggests a large increase in the collective molecular order of lysine amine groups, likely due to the peptide-mineral interaction. In contrast, the spectra of LKβ15 before and after TMOS addition (Figure 2.5b) reveal no significant changes in the NH stretch region. This observation indicates that, in contrast to the helical fold, lysine side chains in LKβ15 are not interacting in a collective manner with the silica during the sheet formation. The simulation results shown below indicate that backbone amide modes might also be more involved in the silica interactions.

The amide I VSFG spectral region provides further information about the secondary structures of the two peptides before and after biomineralization. The results for LKα14 (Figure 2.6a) show a resonance centered near 1642 cm^{-1}. Taking into account that all spectra were measured in D_2O, the band is assigned to an alpha helix, which is in agreement with previously published VSFG spectra of LKα14 on hydrophobic surfaces.[53] No significant spectral changes are apparent upon interaction with TMOS, thus we conclude that LKα14 retains its native folding in presence of the silica film.

Figure 2.6b shows the results for LKβ15 acquired in *ssp* and *sps* polarization combinations, respectively. For *ssp*, a resonance centered at ~1649 cm^{-1} can be assigned to the B1 β-strand mode, while in *sps* only the low frequency B2 mode can be observed near 1613 cm^{-1}.[54]

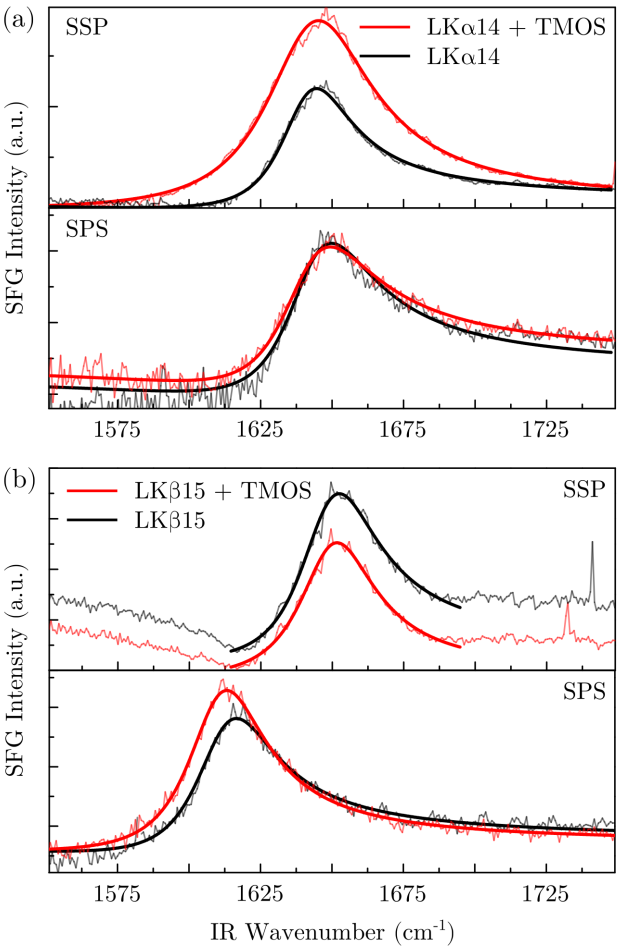


Figure 2.6: VSFG spectra before (black) and after TMOS addition (violet) in the spectral range of backbone vibrations: (a) *ssp* and *sps* spectrum of LKα14. (b) *ssp* and *sps* spectrum of LKβ15. Fit parameters are shown in Section 2.5 - Table 1.

Chirgadze and Nevskaya characterized the resonance splitting of amide vibrations by transition dipole coupling.[55] The mode splitting into B1 and B2 modes observed in the *ssp* and *sps* spectra indicates there is a certain degree of antiparallel-chain packing of the beta sheets at the interface. In addition, we observe the B1 mode in *ssp* and the B2 mode in *sps*, indicating the peptide is strongly oriented. The *ssp* polarization probes modes with transition dipole moments (TDMs) parallel and perpendicular to the surface, while *sps* probes TDMs oriented in the surface plane. Since the TDM of the

B2 mode points in the direction of the C=O bonds, we can conclude that the C=O bonds of LKβ15 are oriented predominantly parallel to the water surface.[55]

Molecular dynamics simulations were largely performed and analyzed by V. Jaeger to gain further insights into the mechanisms by which LKs induce the formation of silica films, especially the initial stages of mineralization. Simulation boxes contained 23 LKα14 or LKβ15 peptides at the vacuum–water interface with chlorine ions, phosphate ions and a silica oligomer ($H_{15}Si_7O_{22}^-$) present in the aqueous phase. Since the intermediate structures formed in the transition from silicic acid to silica are complex and varied, we selected one possible intermediate oligomer that (a) contained functional groups from silicic acid and (b) was small enough and flexible enough to reorient during a 250 ns simulation.[56] The contents of the simulation boxes, the structure of the silica oligomer, and a representative starting structure are presented in Section 2.5 - Table 2, Figure 6 and Figure 7. Simulations were initialized with LK peptides oriented as a monolayer with leucine residues pointing into the vacuum. The great majority of the peptides remain in a monolayer near the interface throughout the entire simulation, with only two LKβ15 peptides and no LKα14 traveling across the box to find the other vacuum-water interface. Moreover, the secondary structure of the peptides remains stable as measured by Ramachandran plots (Figure 2.7).

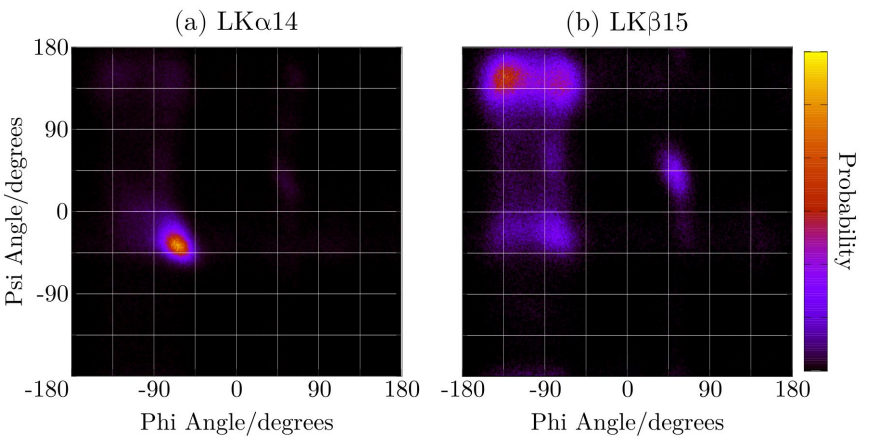


Figure 2.7: Ramachandran plots for backbone phi and psi angles over the last 100 ns of the simulation are shown in (a) and (b).

LKα14 is almost exclusively in the right-handed alpha helix conformation, while LKβ15 heavily favors the beta conformation with small populations in the right and left-handed helix conformations. Surface coverage in the simulations, as calculated by projecting van der Waals spheres of interfacial peptide atoms onto the xy plane, is determined to be 85.7 % and 78.6 % for LKα14 and LKβ15 respectively.

The high stability of the LK monolayer is in good agreement with the experimental observation that the peptide layer is not affected by removal of peptides from the solution subphase. Within 5 ns of the beginning of the simulations, phosphate and chlorine ions preferentially interact with the lysines near the interface driven by favorable electrostatic forces. Simultaneously, silica oligomers are likewise attracted to the lysine residues, and begin to aggregate into clusters of around 2.0 nm in diameter (Figure 2.8).

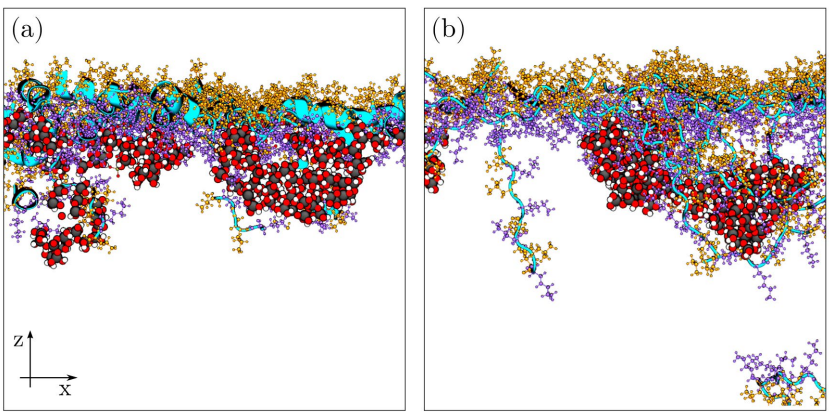


Figure 2.8: Side views of the simulation boxes containing LKα14 (a) and LKβ15 (b) interacting with silica.

The series of events is observed in both the LKα14 and the LKβ15 cases, and the structures that are formed (both the peptide monolayer and the silica clusters) remain stable for the entire 250 ns simulation. Snapshots of the simulations at 5 ns and 100 ns are presented in Section 2.5 - Figure 8.

Closer inspection of the simulation results reveals several important differences in the interactions of LKα14 and LKβ15 with the silica precursors. LKα14 shows a rather sharp angle distribution for the lysine side chains in contact with silica while LKβ15 shows a random orientation between ~30 and ~100 degrees (Figure 2.9a). This supports the view that the differences in NH

signal near 3300 cm^{-1} observed in the VSFG spectra are related to different degrees of order at the mineral interface. Analysis of the amide bond orientations within 1.5 nm of the interface shows that while the LKα14 bonds are broadly distributed the peptide bonds within LKβ15 are well aligned with an orientation parallel to the surface (Figure 2.9b).

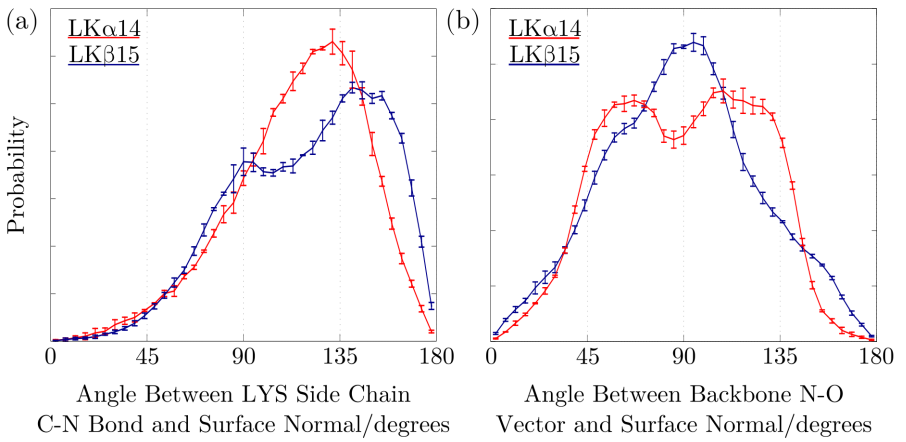


Figure 2.9: Simulated angle distributions for the angle between the terminal LYS side chain C-N bond and the surface normal are shown in (a). Simulated LK peptide angle distributions of the angle between the N-O vector (backbone nitrogen atoms - nearest carbonyl oxygen) and the surface normal are shown in (b).

The orientation distribution of leucine side chains within the LKβ15 film (Section 2.5 - Figure 9) shows the leucine side chains are oriented perpendicular to the surface – in agreement with the VSFG orientation analysis. Quantitative analysis of the silica-backbone distances (Section 2.5 - Figure 10) shows that LKβ15 interacts with silica more strongly via its backbone amide groups compared with LKα14. In addition, an analysis of the silica-backbone distance as a function of distance to the air-water interface shows that the LKα14 sheet is comparatively stable and draws the TMOS molecules into the interfacial layer. The silica-backbone distances in the LKβ15 sheet are less localized and peptides interacting with silica are spread over several nanometers into the bulk solution. Number densities of silica and side chain nitrogen atoms are shown in Section 2.5 - Figure 11. The selective ordering of the LKα14 side chain is also consistent with the observed XPS results. It is important to note that the experimentally observed sheets are assembled over minutes – the simulations capture only the first 250 ns of the

mineralization process. The fibrils and granules visible in the images for the grown films are not clearly visible in the simulation results. However, the marked differences in the distribution of peptide-silica clusters may trigger the onset of fibrillar (LKβ15) versus granular (LKα14) silica morphogenesis.

2.3 Conclusion

To conclude, amphiphilic LK peptides can precipitate nanometer thin self-supported silica sheets. The sheets are stable and can be transferred to solid substrates. TEM, SEM and SFM analysis demonstrates that the morphology of the sheets critically depends on the secondary structure of the peptide. The α-helical peptide LKα14 produced granular, homogeneous sheets, while the β-strand peptide LKβ15 produced a structure decorated with silica fibrils. VSFG analysis indicates that peptides, depending on their folding, control silica biogenesis differently – helices interact via the side chains, while the β-strand motif interacts via backbone groups. Together, this demonstrates that biomimetic peptides can be used to design novel silica and possibly other oxide thin sheets with controlled morphology and, if the organic content is removed by calcination, variable porosity.

2.4 Specific Experimental Details

Peptide synthesis (performed in collaboration with S. Pütz): LKα14 and LKβ15 were synthesized on a Liberty 12 (CEM) microwave peptide synthesizer using Fmoc-Leu-OH, Fmoc-Lys(Boc)-OH and Fmoc-Leu-Wang resin. Chemicals and resins were bought from Sigma Aldrich. The single amino acids are activated for the coupling reaction with an activator base mix containing N,N-diisopropylethylamine and 1-methyl-2-pyrrolidinone and subsequent addition of ethyl (hydroxyimino)cyanoacetate and N,N,N ,N -tetramethyl-O-(1H-benzotriazol-1-yl)uronium hexafluorophosphate. In advance of the coupling reaction, the synthesized chain was deprotected using piperazine, thus making it reactive towards the activated single amino acid. The last amino acid was capped using acetic anhydride. After washing the resin with dichloromethane, it was treated with a solution containing trifluoroacetic acid, triisopropylsilane and water (volume ratio 95:2.5:2.5) over night. Finally the peptide was precipitated with cold diethyl ether. The purity was analyzed with a Q Exactive Plus mass spectrometer (Thermo Scientific) coupled to an EASY-nLC 1000 uHPLC system (Thermo

Scientific). A self-packed C18 silica column (inner diameter 75 μm) with 1.9 μm bead size and a gradient of HPLC grade acetonitrile/0.1% formic acid in H_2O were used.

Biomineralization at the air-water interface: For silica biomineralization at the air-water interface, lyophilized peptides were dissolved in double ionized water (10 ml). The solution was mixed with 2x PBS (10 ml, 300 mM NaCl, 20 mM K_2HPO_4, 20 mM KH_2PO_4, pH = 7.4). With a final peptide concentration of 0.12 mM the solution was incubated in a Teflon trough for 50 min to allow for adsorption of the peptide at the air-water interface.[47c] To minimize the concentration of peptide in the bulk solution we exchanged the subphase (10 ml) with 1x PBS (10 ml) three times. The biomineralization process was initiated by injecting a precursor mix which was prepared by mixing HCl (850 μl, 1 mM) with TMOS (150 μl) and sonicating for 5 min.

X-ray photoelectron spectroscopy (performed and analyzed by T. Weidner and H. Lu): XPS data were collected on a Kratos AXIS Ultra DLD spectrometer. The instrument uses a monochromatic Al Kα X-ray source. The electron take off angle (between the sample surface plane and the axis of the analyzer lens) was 90° using small spot analysis detection mode with a spot size of 100 μm^2. The scans were acquired with an analyzer pass energy of 80 eV. The base pressure in the analysis chamber was better than $5 \cdot 10^{-9}$ mbar. The data analysis was performed using the Vision Processing software. A 5 keV Ar ion beam was used for sputtering with a crater size of 1 mm^2. The depth profiles of the samples were generated in cycles where each XPS spectrum was collected after exposing the sample to the ion beam.

Vibrational sum-frequency generation spectroscopy: The fundamentals of VSFG spectroscopy and a detailed description of the laser setup can be found in Section 1.3. The sum frequency signal was generated in reflection at an angle of 60° (IR) and 55° (VIS) respective to the surface normal. All data was acquired for 20 min at 22° C. In the range of 2800-3400 cm^{-1} we used H_2O as solvent whereas spectra in the range of 1600-1700 cm^{-1} were recorded using D_2O. After background correction, the energy was calibrated using the sum-frequency signal originating from the surface of a z-cut quartz crystal. The spectra were fitted according to Equation 1.13.

Scanning electron microscopy and scanning force microscopy (conducted and interpreted by G. Glasser, H. Burg and R. Berger): SiO_2-peptide nano-sheets were transferred from the air-water interface onto gold substrates via the Schaefer technique. Gold substrates were produced in-house by deposition of Chromium (1.2 nm) and Gold (50 nm) on silicon wafers (1 by 1 cm). These samples were mounted on aluminum stubs and analyzed with a Zeiss

Gemini 1530 instrument and an accelerating voltage of 750 V. Similar samples were prepared for the analysis with scanning force microscopy (Bruker, Dimension ICON) operated in PeakForce mode. For imaging we took Olympus OMCL 240TS probes having a nominal spring constant of 2 N/m.

Molecular dynamics simulations (performed and analyzed by V. Jaeger): A basic overview of molecular dynamics simulations can be found in Section 1.4. The experimental details are outlined in the following. Peptide, phosphate, and silica potentials came from the AMBER99SB-ildn,[31b] GAFF,[57] and modified version of an established force field,[58] respectively. Silica dihedral parameters were developed by fitting to B3LYP/6-311G(d) quantum mechanical calculations. Phosphate and silica electrostatic point charges were developed at the HF/6-31G(d) level of theory using the RESP method with Gaussian09 and the antechamber package in Ambertools (parameters are found in Section 2.5 - Figure 6, Table 3 and Table 4).[59] AMBER-type files were converted to GROMACS using acpype.[60] Molecular dynamics simulations were performed using GROMACS 4.6.[61] Freezing the hydrogen-heavy atom bonds and applying the LINCS algorithm allowed for a timestep of 1 fs. Systems of approximately 8.0 x 8.0 x 13.8 nm were built using Packmol.[62] Of the 13.8 nm in the z dimension, about 7.0 nm was vacuous. Periodic boundary conditions were applied in all three dimensions to use particle-mesh Ewald summations for long-range electrostatics.[63] This simulates x-y slabs separated by 7.0 nm of vacuum. Lennard-Jones potentials were shifted to zero at 1.2 nm. Temperature was maintained at 310 K using a stochastic velocity-rescaling thermostat.[64]

2.5 Additional Data

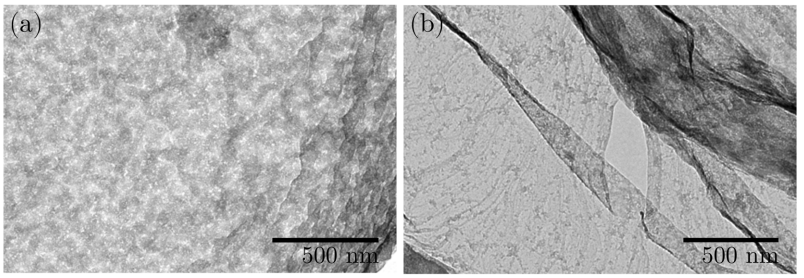


Figure 1: (a) TEM of free-standing biosilica films prepared from LKα14 and (b) LKβ15.

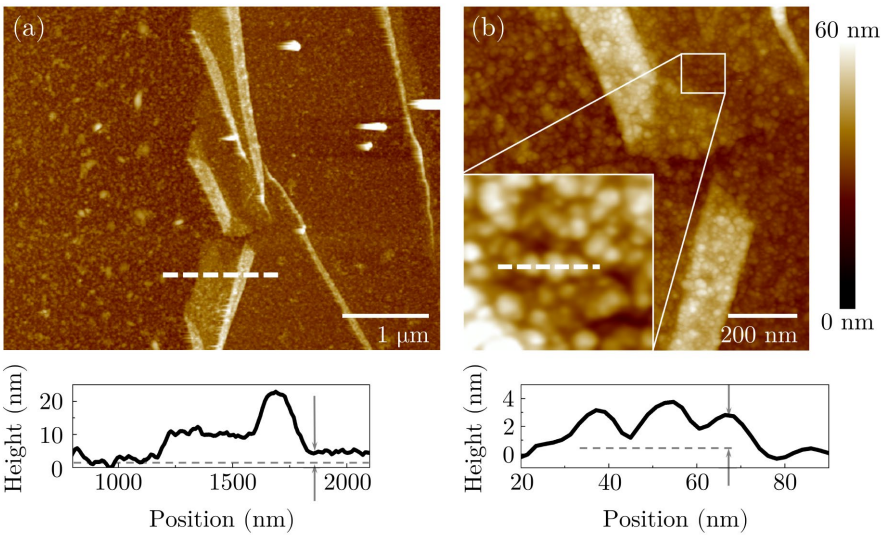


Figure 2: (a) SFM image of the edge of an LKα14-silica film lying on a gold substrate. (b) Zoom in on image (a). Corresponding height profiles taken from the dashed lines are shown below.

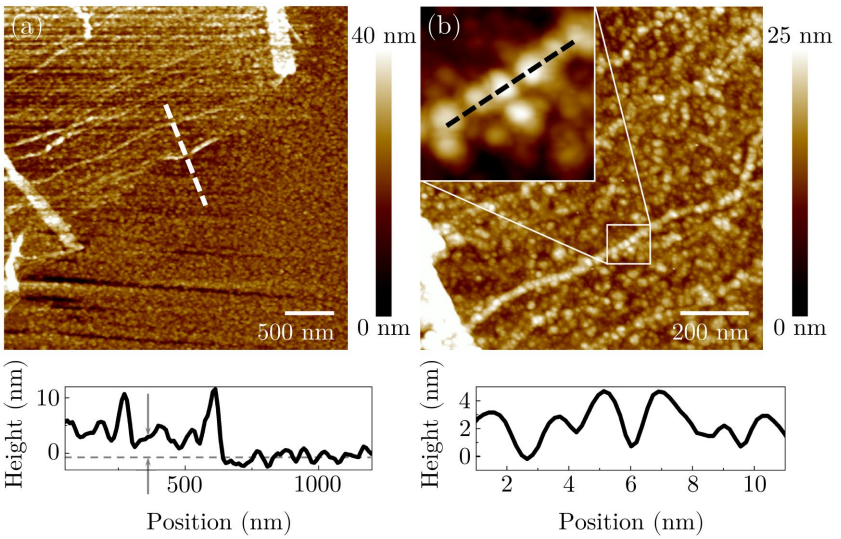


Figure 3: SFM images of an LKβ15-silica film. (a) The edge of the film lying on a gold substrate. (b) Zoom in on image (a). Height profiles taken from the dashed lines are shown below each SFM image, respectively.

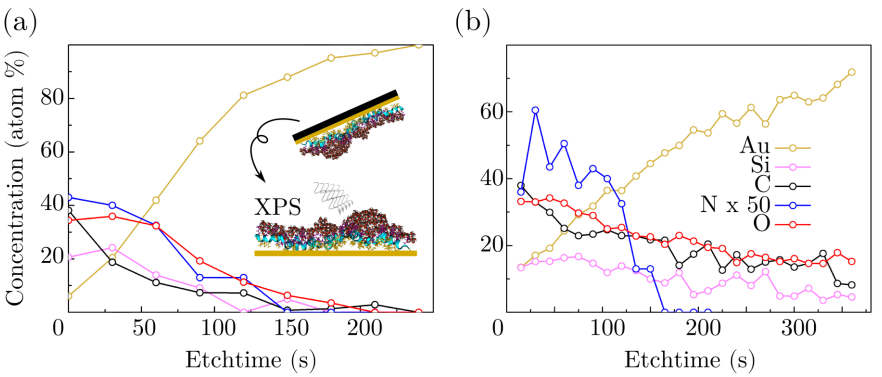


Figure 4: (a) XPS analysis of the LKα14-silica film. For depth profiling, the atom concentration in percent is plotted against the etching time in seconds. (b) Depth profiling of the LKβ15-silica film.

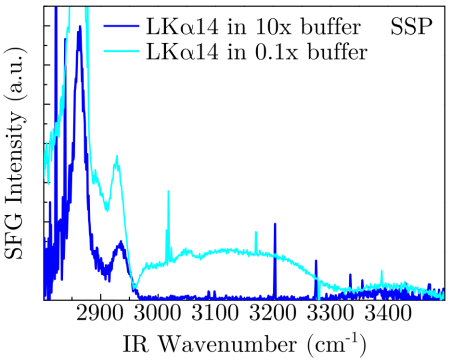


Figure 5: VSFG spectrum of LKα14 peptides at the air-water interface in 10x buffer (dark blue) and 0.1x buffer (light blue).

Table 1: Fit parameters for the spectra shown in Figure 2.6. Amplitude, frequency (in cm^{-1}) and half width at half max in (cm^{-1}) are denoted by A, w and Γ.

SSP				
	LKα14	**LKα14+TMOS**	**LKβ15**	**LKβ15+TMOS**
NR Amplitude	0.2	0.1	0.4	0.3
NR Phase ϕ	3.1	3.1	2.3	2.0
A	11.1	21.4	5.3	4.9
w	1640.6	1642.4	1648.3	1649.1
Γ	15.5	22.6	16.7	16.2
SPS				
	LKα14	**LKα14+TMOS**	**LKβ15**	**LKβ15+TMOS**
NR Amplitude	0.1	0.2	0.1	0.1
NR Phase ϕ	3.1	2.9	2.4	2.2
A	5.1	4.3	2.5	2.8
w	1642.8	1641.2	1612.6	1610.0
Γ	18.3	19.3	16.5	16.8

Table 2: The contents of the simulation boxes containing LKα14 or LKβ15

	LKα14		LKβ15	
	Molecules	*Atoms*	*Molecules*	*Atoms*
Peptide	23	6693	23	7199
Water	10700	32100	10667	32001
HPO_4^{2-}	5	30	5	30
$H_2PO_4^-$	5	35	5	35
Cl^-	77	77	100	100
Silica Model	23	1012	23	1012
Total	10808	39947	10823	40377

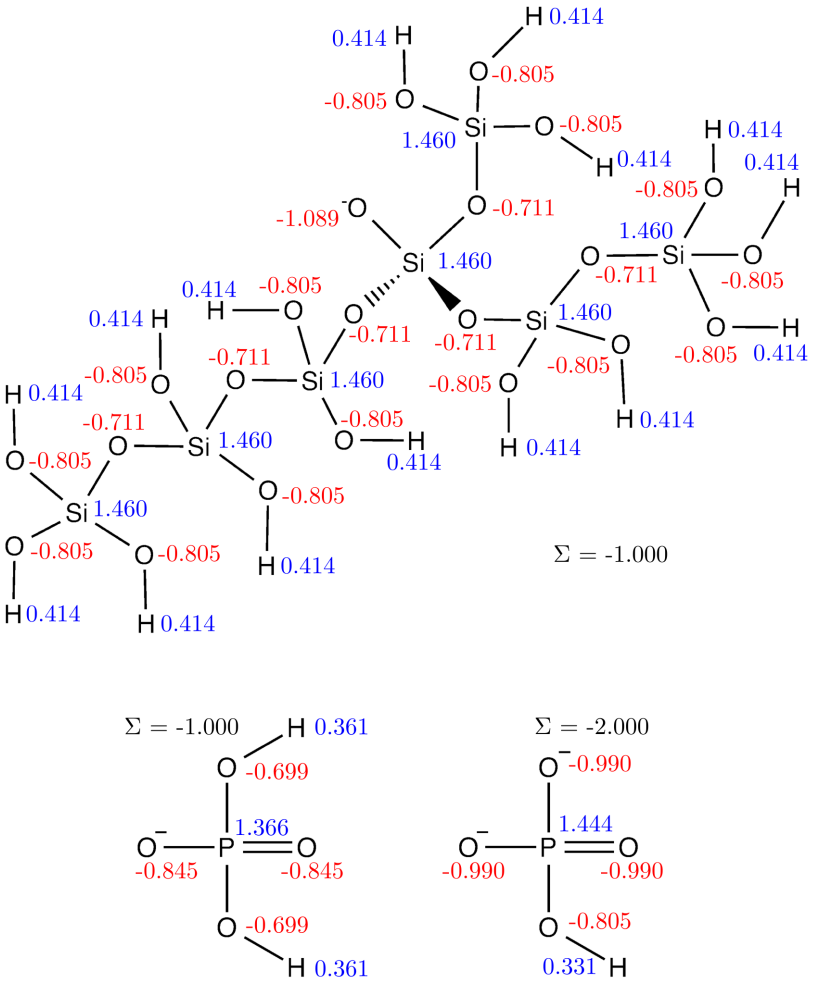


Figure 6: Chemical structure of the silica model precursor and the phosphate ions with their electrostatic point charges.

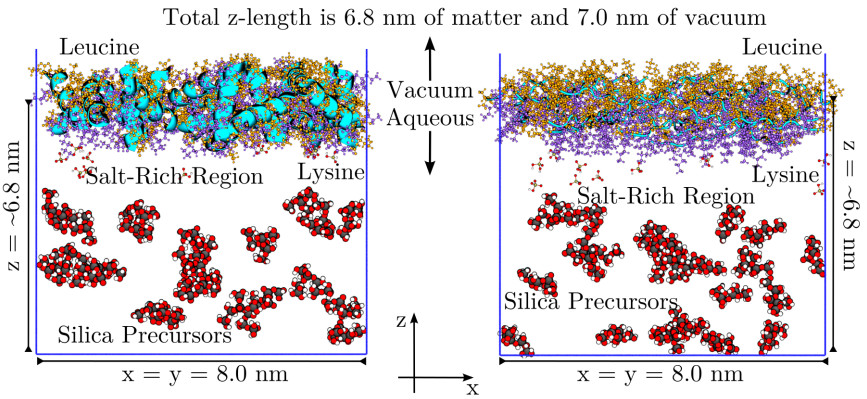


Figure 7: A representative box output from Packmol for LKα14 (left) and LKβ15 (right). Water is not shown for clarity. Phosphate ions are colored yellow, white, and red. Silica precursors are colored gray, white and red. Peptide backbones are cyan, leucines are orange and lysines are purple.

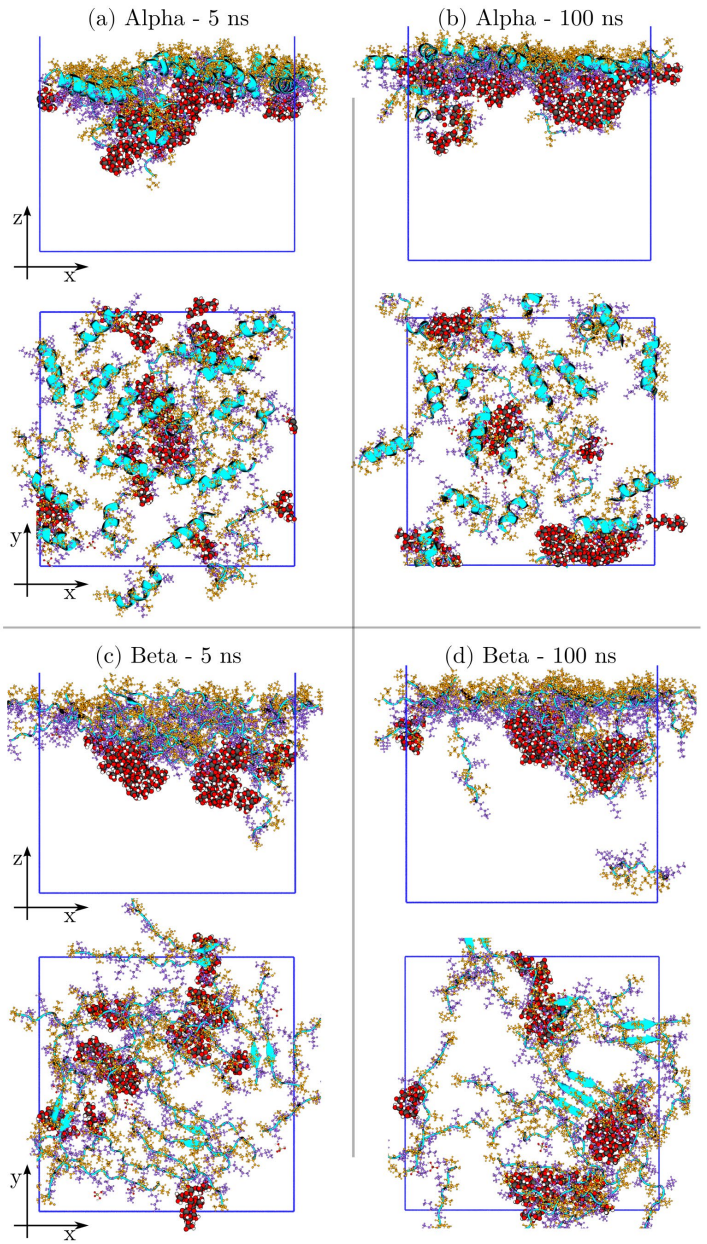


Figure 8: Snapshots from the MD simulations. (a) LKα14 after 5 ns of simulation. (b) LKα14 after 100 ns. (c) LKβ15 after 5 ns. (d) LKβ15 after 100 ns. Peptides are stable at the interface. Clusters of silica precursor molecules are stable. Coloring is the same as Figure 7 of this section. Some edges are clipped. Secondary structure for LKβ15 is shown only for β-sheets and not individual beta strands.

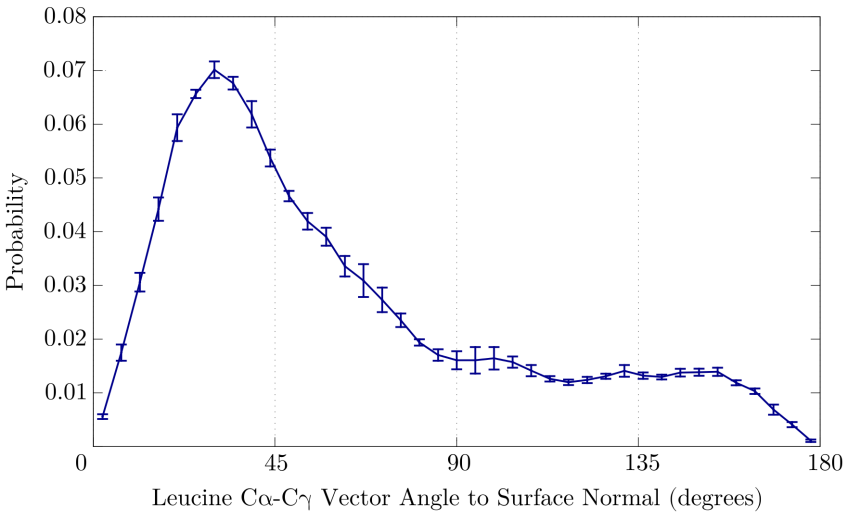


Figure 9: Simulated LKβ15 angle distributions for the angle between the LEU side chain Cα-Cγ vector and the surface normal. Alpha carbon atoms within 1.5 nm of the interface are included in the calculation. The last 100 ns of the simulation are used. The surface normal is the positive z-axis. Error bars are one standard deviation when the analysis is split into three 33.3 ns chunks.

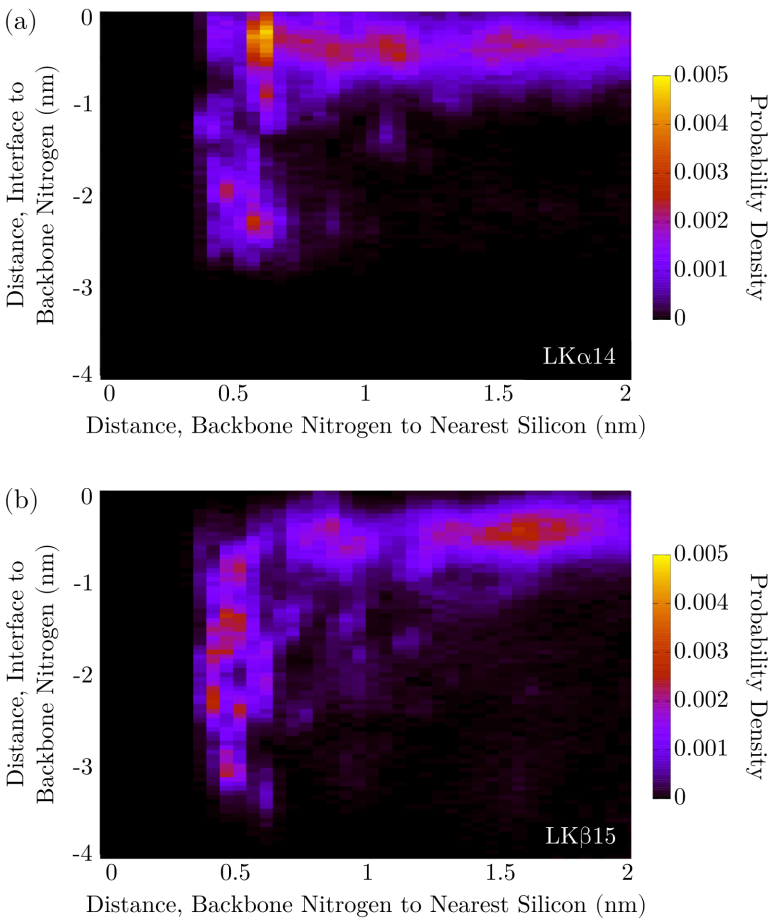


Figure 10: Simulated backbone nitrogen – nearest silicon distance distribution with respect to the distance from the interface for LKα14 (a) and for LKβ15 (b). Data taken from the last 100 ns of the simulation.

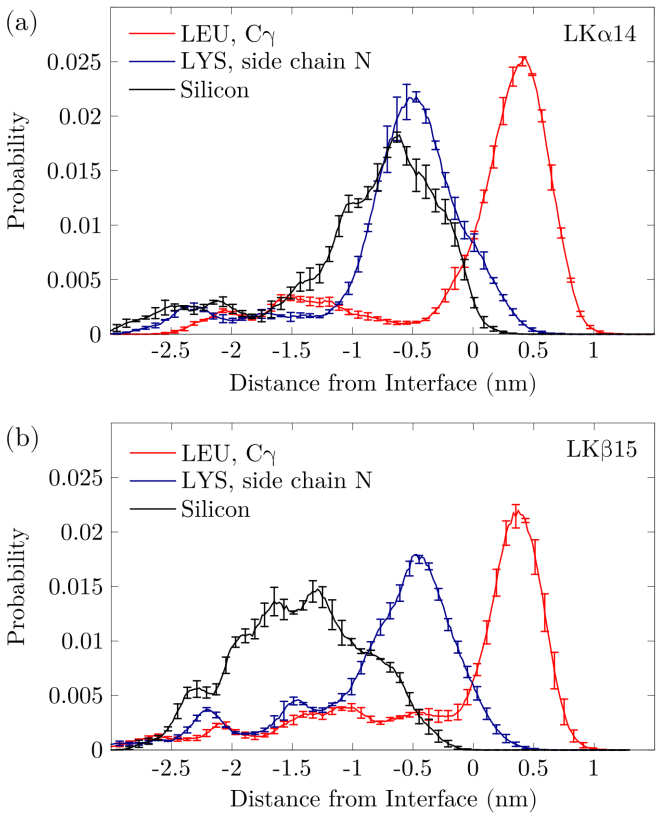


Figure 11: Simulated number density of certain important atoms with respect to distance from the interface for LKα14 (a) and for LKβ15 (b). The interface is defined as the intersection of the LEU and LYS curves.

Table 3: The following .frcmod file was used for silica model precursors.

MASS		
Si	28.085	0.000
os	16.000	0.465
oh	16.000	0.465
ho	1.008	0.135

BOND		
Si-os	885.10	1.610
Si-oh	428.00	1.420
oh-ho	545.00	0.960

ANGLE		
Si-os-Si	4.660	174.220
os-Si-os	159.570	110.930
os-Si-oh	153.260	111.090
Si-oh-ho	57.500	106.000
oh-Si-oh	89.620	116.260

DIHE				
Si-os-Si-os	1	0.000	0.000	1.000
Si-os-Si-oh	1	0.000	0.000	1.000
oh-Si-oh-ho	1	0.280	30.000	2.000
ho-oh-Si-os	1	0.630	0.000	-2.000
ho-oh-Si-os	1	0.280	180.000	1.000

IMPROPER			
Si-os-Si-os	1.1	180.0	2.0
Si-os-Si-oh	1.1	180.0	2.0

NONBON		
Si	2.1475	0.3000
os	1.7500	0.1700
oh	1.7700	0.2104
ho	0.2245	0.0460

Table 4: The following .frcmod file was used for phosphate ions.

MASS		
p5	30.970	1.538
o	16.000	0.434
oh	16.000	0.465
ho	1.008	0.135

BOND		
p5-o	487.70	1.481
p5-oh	321.20	1.625
oh-ho	369.60	0.974

ANGLE		
p5-oh-ho	44.150	110.140
o-p5-oh	69.980	115.260
o-p5-o	73.530	115.800
oh-p5-oh	71.250	102.450

DIHE				
o-p5-oh-ho	1	0.533	0.000	3.000
oh-p5-oh-ho	1	0.533	0.000	3.000

IMPROPER

NONBON		
p5	2.1000	0.2000
o	1.6612	0.2100
oh	1.7210	0.2104
ho	0.2245	0.0460

3 Influence of Peptide Acetylation on Silica Biomineralization

Acknowledgement: Dr. Vance Jaeger's contribution was in the design, conducting and analysis of the molecular dynamics simulations. Scanning electron microscopy was performed and interpreted by Gunnar Glasser. I conducted VSFG spectroscopy, FTIR spectroscopy and silica biomineralization experiments.

3.1 Introduction

In the last chapter and in reference [44a] it was shown that the morphology of silica films and particles can be controlled by simple binary peptides composed solely of hydrophobic leucine (L) and hydrophilic lysine (K), called LK peptides.[45] Short acetylated LK peptides of α-helical (Ac-LKα14, Ac-LKKLLKLLKKLLKL) and β-strand conformation (Ac-LKβ15, Ac-LKLKLKLKLKLKLKL) are capable of precipitating networks of silica spheres and silica fibers respectively, but also nanometer thin films and sheets in aqueous solution.[44a, 65] In conjunction with coarse grained (CG) simulations it was hypothesized that fibers could grow from pentameric or larger aggregates of the β-strand. In the case of the peptide with helical conformation the lowest energy aggregates were observed to be spherical tetramers.

Much to our surprise we found in later experiments that the non-acetylated LK variant of helical conformation, LKα14 precipitates a network of silica fibers with high resemblance to silica precipitates mineralized with Ac-LKβ15. The difference between LKα14 and Ac-LKα14 amounts to only 5 out of 285 atoms, and at a pH of 7.4 not more than one positive charge. This surprising result – showing the difference one charge can have on large scale biomineralization – led us to investigate how peptide-peptide and peptide-silica interactions may be altered by N-terminal acetylation of these LK peptides.

In the field of peptide research, N-terminal acetylation (Nt-acetylation) is a common modification of newly synthesized peptides. However, a recent study has shown that there is no general rule regarding to how Nt-acetylation impacts protein stability.[66] Depending on the set of investigated proteins there is indication for either destabilization, for stabilization or for neither.[67] Following a common approach in peptide-driven biosilicification,[68] we show that Nt-acetylation of the peptide has a tremendous impact on biosilicification. The morphological outcome of peptide-driven biosilicification could be determined in at least two ways: Either the peptide-silica interaction at the surface of aggregates is altered, or peptide-peptide interactions are altered to yield aggregates of a different size and shape.

To test for differences in surface interactions we monitored the peptide structure at silica surfaces by vibrational sum-frequency generation (VSFG) spectroscopy. We combined the VSFG experiments with coarse grained molecular simulations to detect significant changes in the free energy of aggregates of different morphologies to determine how acetylation might alter the mineralization pathway for LK peptides.

3.2 Results and Discussion

For Ac-LKα14 (Ac-LKKLLKLLKKLLKL) as well as the non-acetylated LKα14 (LKKLLKLLKKLLKL), biosilicification was initiated by the addition of a precursor solution containing silicic acid to a PBS buffered 2.5 mg/ml peptide solution. Scanning electron microscopy was performed and interpreted by G. Glasser. Representative SEM images of dried precipitates are shown in Figure 3.1. Surprisingly, depending on presence or absence of Nt-acetylation, two completely different morphologies are produced. Ac-LKα14 precipitates a network of nano-spheres with diameters of 500 nm while LKα14 forms a network of silica fibers, 20 nm in width, similar to

precipitates obtained with β-strand forming LKs.[44a] In a recent publication, reporting different silica morphologies for LK peptides with different hydrophobic periodicities, it was suggested that the amino acid sequence is the most important parameter for tuning the morphology of resulting precipitates.[44a] However, the results shown in Figure 3.1 indicate that a small perturbation in the form of Nt-acetylation can yield very different biosilica morphologies.

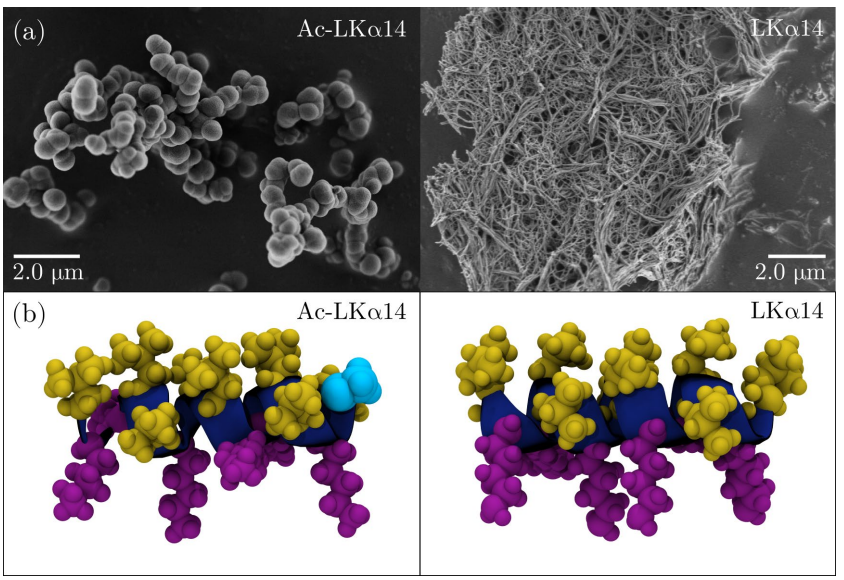


Figure 3.1: (a) Scanning electron microscope images of solution silica precipitates obtained from Ac-LKα14 (left) and LKα14 (right). (b) Idealized structures with leucine in yellow, lysine in violet, and acetylation in cyan.

The differences between the observed morphologies of silica can be explained either by a different molecular conformation of acetylated *vs.* non-acetylated peptides at the peptide-biosilica interface (be it at the nucleation site or on subsequent layers of peptides on the growing mineral), and/or by a difference in peptide aggregation, which would also expose the silica precursor to a different peptide surface.

To investigate possible differences in peptide arrangement between the two peptides, we mimic the peptide behavior at the interface of the hydrophobic aggregate core and the surrounding aqueous phase by studying LKα14 at the air-water interface – a model nonpolar-polar interface. We study this interface, before and after biosilicification, by VSFG spectroscopy,

which reports back on the peptide structure. First, we prepared monolayers of LKα14 and Ac-LKα14 at the air water interface and probed the peptide structure with VSFG. Then we injected a silica precursor solution (TMOS) beneath the peptide layer (Figure 3.2) and probed any changes of the peptide structure in a subsequent VSFG experiment.

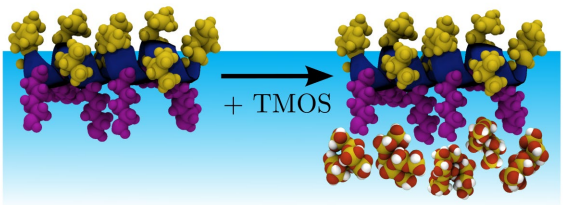


Figure 3.2: Scheme of interfacial biomineralization of silica at the air-water interface. Peptides were equilibrated as a surface adsorbed layer. Subsequent introduction of TMOS yielded a peptide-silica film of few nanometers thickness.[65] Colored with leucine (L) in yellow, lysine (K) in purple and oxygen, silicon, hydrogen of the silica in red, dark yellow and white, respectively.

Naturally, one would expect the hydrophobic leucine side chains to be excluded from the aqueous phase. In the α-helical conformation of Ac-LKα14/LKα14 all leucine side chains face the air, while the lysine side chains are solvated. Indeed, the VSFG spectrum in Figure 3.3a is consistent with substantial ordering of leucine side chains, given the typical methyl and methylene vibrational modes for a peptide layer adsorbed to the air-water interface.[49] Upon biosilicification with either LKα14 or Ac-LKα14 a new resonance near 3300 cm^{-1} evolves, as observed in Figure 3.3b, which can be assigned to amino groups of the hydrophilic lysine side chain becoming increasingly ordered.[52] This biosilicification-induced ordering appears to be more pronounced for Ac-LKα14, but clearly occurs for both peptides. It is not possible to quantify the changes to derive, for example, relative orientations of the amino groups, as surface pressure was not monitored in these measurements. However, it is evident that leucine side chains for both Ac-LKα14 and LKα14 are ordered at the air-water interface with and without *in-situ* mineralized silica, and that in both cases biosilicification leads to increased ordering in the K side chains.

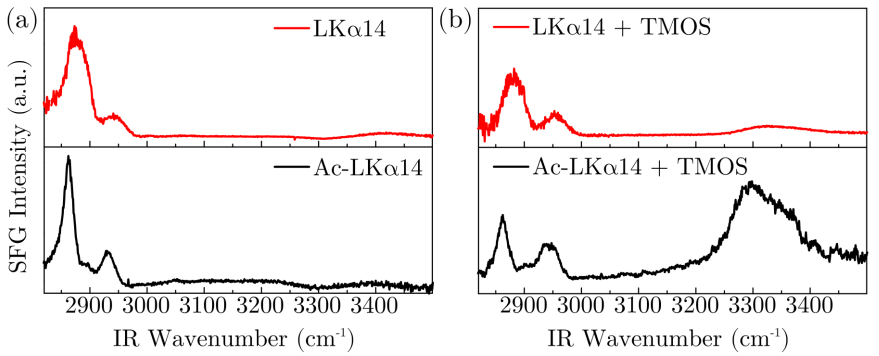


Figure 3.3: VSFG spectra of LKα14 (red) and Ac-LKα14 (black) at the air-water interface before (a) and after interfacial precipitation of silica (b). Spectra were acquired in the *ssp* polarization combination.

In the spectral range of the amide I vibrations shown in Figure 3.4a and b, a pronounced resonance near 1640 cm^{-1} is observed before and after biosilicification for both LKα14 and Ac-LKα14. Fit parameters are provided in Section 3.5 - Table 1 and Table 2.

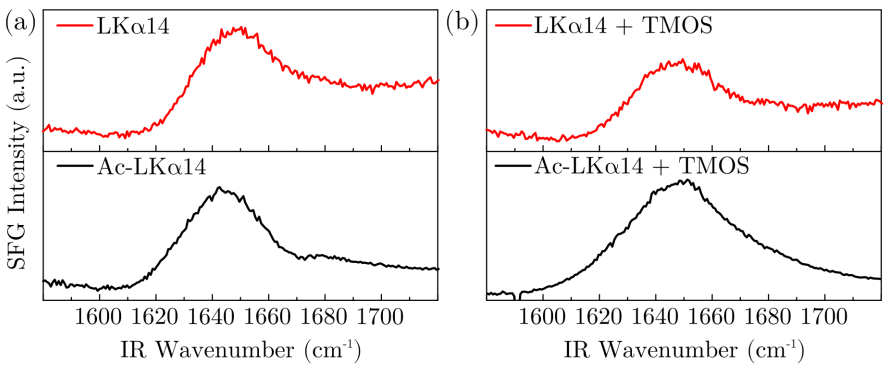


Figure 3.4: VSFG spectra of LKα14 (red) and Ac-LKα14 (black) in the amide I spectral range. Peptides were adsorbed to the air-water interface and spectra collected before (a) and after interfacial precipitation of silica (b). VSFG was detected in the *ssp* polarization combination.

This agrees well with a previous VSFG study of Ac-LKα14 where the amide I band was centered near 1642 cm^{-1}.[65] Considering that the spectra were acquired using D_2O solutions, the peak positions indicate predominantly α-helical conformation for both LKα14 and Ac-LKα14 before and after

biomineralization. From the relatively minor changes in both the side chain VSFG spectra as well as amide I VSFG spectra we conclude that the differences between silica morphologies can most likely not be attributed to different interfacial peptide structures. Another aspect to look at would certainly be the interfacial aggregation state of the peptides. Since we cannot deduce this information from VSFG spectra, we used SEM to image Langmuir-Schaefer deposited silica-peptide films that mineralized at the air-water interface. As can be seen in Section 3.5 - Figure 1 there are no major differences in the silica film morphology.

The fact that fibrillar precipitates are obtained with LKα14 (Figure 3.1, left panel) could hint that the conformation of LKα14 in solution might deviate from its interfacial conformation. However, Fourier transform infrared (FTIR) spectroscopy of peptide solutions with equal concentration as used in VSFG experiments show that most of the intensity is centered near 1645 cm^{-1} (Section 3.5 - Figure 2).

In order to explore whether LKα14 and Ac-LKα14 are likely to aggregate into different configurations in solution, we conducted molecular simulations. Ordered peptide-peptide association occurs on timescales beyond those explored in classical molecular dynamics simulations. To overcome this obstacle, we reduced number of degrees of freedom in the system and enhanced the sampling of peptide-peptide association by using coarse-grained metadynamics simulations.[36] The MARTINI force field is secondary-structure dependent, and therefore the secondary structure does not change during the coarse-grained simulations. We assume the peptide is helical for these simulations, which is supported by FTIR data (Section 3.5 - Figure 2). By biasing the association of the peptides with metadynamics, we are able to recover the free energy of the system as a function of the biased slow degree of freedom.

In Figure 3.5, we present the free energy of the system (replica at 311 K) as a function of the coordination number, which is a measure of the pairwise interactions among the five peptides in the simulation. The exact functional form of the coordination number is provided in Section 3.4. A coordination number near six corresponds to a peptide tetramer, and a coordination number near ten corresponds to a pentamer where all peptides in the box are part of the same aggregated structure. Convergence was assessed by computing the difference in free energy between the tetramer and pentamer wells of the free energy over the timespan of the simulation. See Section 3.4 and Section 3.5 - Figure 3 for more discussion of convergence. Figure 3.5

demonstrates that Ac-LKα14 prefers to form tetramers over pentamers or smaller oligomers.

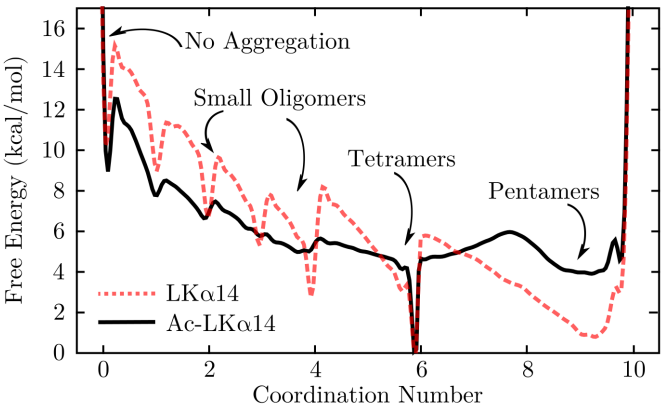


Figure 3.5. Free energy of the various oligomers. Coordination number approximates the sum of pairwise interactions.

We cannot exclude larger aggregation number, as our simulation box contained 5 peptides. However, the tetrameric state presented in Figure 3.6a corresponds to the previously described Ac-LKα14 aggregates found with solid state NMR experiments,[44a] lending additional credibility to our findings. In Figure 3.6a and c, it can be seen that the leucine core of the aggregate colored in white is tightly packed, and the aggregated tetrameric particles have diameters of approximately 3 nm. Snapshots were selected randomly from a selection of frames near a coordination numbers of 5.9 and 9.3, and the snapshots are generally representative of structures observed during the simulations. We hypothesize that the presented aggregate configurations of Ac-LKα14 nucleate the larger, roughly spherical biosilica nanoparticles. Figure 3.6b and d illustrate representative pentamers of both LKα14 and Ac-LKα14 with diameters of roughly 4 nm.

The low-energy well for a pentamer is shifted away from a perfectly packed structure with a coordination number of ten. This is because the pentameric structures tend to have one peptide slightly shifted away from a tighter tetrameric core. Figure 3.5 demonstrates that the pentameric and tetrameric states are nearly equally populated for LKα14 in contrast to the clear preference for the tetrameric state observed for Ac-LKα14. The distribution of aggregated states is much more heterogeneous for LKα14 compared to Ac-LKα14. This heterogeneity may lead to the more tangled

and random structures observed in Figure 3.1. It is also notable that the energy wells for smaller oligomers are sharper and deeper for LKα14 compared to Ac-LKα14.

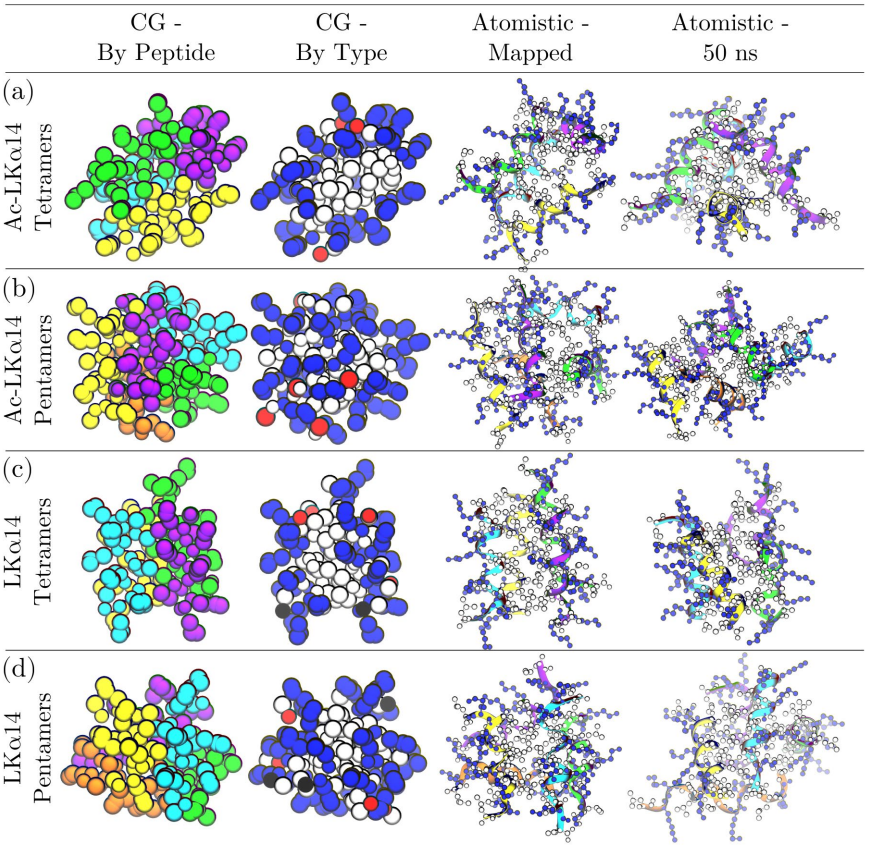


Figure 3.6: Predicted low energy structures of (a) Ac-LKα14 tetramer and (b) LKα14 tetramer (c) Ac-LKα14 pentamer and (d) LKα14 pentamer. Column 1: colored with each peptide having a unique color. Column 2: colored with leucine (L) in white, lysine (K) in blue, negatively charged ends of LKα14 in black, positive ends in red. Columns 3 and 4: Combination of coloring schemes from Columns 1 and 2 with atomistic detail.

We hypothesize that this is caused by a strong preference for anti-parallel helix orientations in the dimer state where charged termini align with a corresponding oppositely charged termini as observed in the simulation trajectories. This preference for anti-parallel alignment and the matching of

oppositely charged termini likely leads to a more energetically favorable structure. No such preference for parallel or anti-parallel arrangement is observed in the dimer state for Ac-LKα14. This is consistent with previously published the observation that capped LKα14 does not show a preference for either parallel or anti-parallel orientations.[69] Since the peptides in larger oligomers do not always align perfectly with each other, a comparison of parallel and anti-parallel arrangements is not available. Differences in peptide orientation and alignment could also lead to some of the differences in the structures observed in Figure 3.1.

The third and fourth columns of Figure 3.6 show atomistic structures of the system. Some structures are rotated for a better view. Column 3 contains a back-mapped structure of the structure illustrated in columns 1 and 2. Column 4 shows the structure after being allowed to evolve over nearly 50 ns of atomistic MD simulation at 300 K. The illustrations in column 4 demonstrate that the structures found using MARTINI CG and metadynamics are reasonably stable, and that the peptides remain roughly helical even in the aggregated solution state. This is consistent with the FTIR data presented in Section 3.5 - Figure 2. The hydrophobic cores of the aggregates tend to swell slightly when moving from CG to atomistic simulations, which indicates that the hydrophobic interactions in MARTINI might be too tight spatially.

3.3 Conclusion

Previously, Nt-acetylation of peptides has been widely employed but rarely considered to have a major impact on the functionality of a peptide. For the simple case of silica mineralization guided by LK peptides we show that the morphology of the precipitated biosilica is greatly affected by the presence of a terminal modification. Our study shows that it is unlikely that this behavior is guided by interfacial peptide-silica interactions but instead by peptide-peptide interactions, as demonstrated in coarse-grained molecular simulations. More specifically, it appears that larger aggregates are preferred in the case where termini are charged compared to the acetylated case. Furthermore, charged termini lead to a preference for anti-parallel oligomers, whereas there is no such preference in orientation observed in the acetylated case. In future studies it would be interesting to see how multiple peptide aggregates destabilized in such a way assemble and how these supramolecular assemblies differ from the case of acetylated peptides. The implications of our

findings extend beyond the field of biomineralization or peptide design to encompass all areas where chemical and biological phenomena that depend upon the precise ordering of peptides at interfaces.

3.4 Specific Experimental Details

Chemicals: Peptides were purchased from GenScript (Jiangning, Jiangsu Province, China) with a purity of > 90%. The amino acid sequence of Ac-LKα14 is Ac-LKKLLKLLKKLLKL while the sequence of LKα14 is lacking the Nt-acetylation. Buffer salts, HCl and tetramethoxysiliane (TMOS) were purchased from Sigma Aldrich (St. Louis, MO, USA).

Peptide studies at the air-water interface: Peptide solutions (20 ml, 0.12 mM peptide) were prepared in a Teflon trough with phosphate buffered saline (0.01 M phosphate, 0.138 M NaCl; 0.0027 M KCl, pH = 7.4). After an equilibration period of 1 h, the peptide concentration in the bulk solution was reduced by eight-fold dilution of the subphase with buffer. TMOS was mixed with 1 mM HCL in a v/v ratio of 3:16. The mix was sonicated for 5 min and injected into the trough yielding a TMOS concentration of 25 mM.[65]

Biomineralization in solution: Peptide solutions were prepared by dissolving lyophilized peptide in 500 µl MilliQ water (2.8 mM) and adding 500 µl of 2x concentrated PBS buffer. Biosilicification was started by adding TMOS/ 1 mM HCl (v/v 3:16) up to a final TMOS concentration of 25 mM.

Vibrational sum-frequency generation spectroscopy: A prototype *Spitfire Ace* (Spectra-Physics, Santa Clara, CA, USA, 10W, 1 kHz, ~120 fs pulses) was used as an amplifier. Part of the pulsed output laser beam was directed into an etalon to produce a narrow band (VIS) laser beam (800 nm, 20 cm^{-1} bandwidth) while another part was sent to a *TOPAS* and *NDFG stage* (Light Conversion, Vilnius, Lithuania) to produce a broad band (IR) laser beam. Sum-frequency generation was achieved by overlapping the VIS and IR beam spatially and temporally at angles of 55° and 60° respectively on the sample. Data was collected using a *Newton CCD* (Andor, Belfast, UK) with acquisition times of 20 min. Spectra recorded in the region of 2800-3600 cm^{-1} were recorded using H_2O, spectra recorded in the region of 1500-1800 cm^{-1} were recorded using D_2O solutions, to avoid interference with the H_2O bending mode. Spectra were background-corrected by subtracting a spectrum acquired with the IR beam off. The intensity was corrected by dividing raw

data by the VSFG spectrum from the surface of a z-cut quartz crystal. Fitting of VSFG spectra was accomplished using Equation 1.13.

Scanning electron microscopy (performed and interpreted by G. Glasser): SEM experiments were performed on a Zeiss Gemini 1530 SEM (Carl Zeiss Microscopy GmbH, Jena, Germany); dry precipitates were transferred onto sticky carbon tape and mounted on aluminum sample holders. The intricate morphology of the precipitates only became visible when no coating was used. The morphology of the precipitates was therefore analyzed without sputter-coating. The accelerating voltage was set to 750 V.

Molecular dynamics simulations (conducted and analyzed by V. Jaeger): We used coarse-grained metadynamics simulations to elucidate the free energy differences among various peptide oligomers and thus the relative probability of finding the peptides in a given configuration. The MARTINI 2.2 force field was used to describe peptides with a polarizable water force field,[70], wherein energy is conserved with a 20 fs time step for numerical integration.[71]. MARTINI 2.2 models the partitioning of hydrophobic residues into hydrophobic environments decently well, and it has been used previously to study the interactions of short peptides.[72]

In order to simulate the acetylation of the peptide, terminal backbone charges were neutralized. Simulations boxes contained five peptides and were cubic with a side length of 10 nm. CG chlorine and sodium ions were used to neutralize the box and to add 150 mM salt to the solution. The initial state of the simulation box was randomly generated using Packmol. Up to 10,000 steps of steepest descent minimization were used to ease bad contacts. Simulation boxes were further prepared with 100 ns of pressure equilibration in the NPT ensemble using the Parinello-Rahman barostat at 1 bar and 300 K.

Subsequent metadynamics simulations were 2.8-4.0 μs in length. This ensured convergence of the free energy calculations. Well-tempered metadynamics rather than standard metadynamics was used to ensure that the free energy of the system smoothly converged to a final value. The height of the hills added to the collective variables (CVs) was 2.0 kJ/mol with a width of 0.05 units in the CV space. Hills were deposited every picosecond. The bias factor for hill-height decay was 20. Two slow degrees of freedom (collective variables) were biased using metadynamics to sample protein oligomer conformations: (a) radius of gyration (nm) among the center of mass of each peptide in the simulation box and (b) the number of contacts (coordination number) among the LEU center of mass in each peptide. The pairwise interactions of the peptides in the system are assigned a value

between zero and one based upon a sigmoidal function. $S = (1-(r/r_0)^n)/ (1-(r/r_0)^m)$, S = pairwise coordination, r = pairwise LEU COM distance, r_0 = 1.6 nm, n = 8, m =16. The coordination number is the sum of these pairwise values.

Parallel tempering in the well-tempered ensemble was used to help the system overcome hidden free energy barriers.[39]Simulations were performed at temperatures ranging from 280 K to 420 K (280, 295, 311, 330, 349, 371, 394, 420) in the well-tempered ensemble with parallel tempering in GROMACS 5.1.2 patched with PLUMED 2.3.[73] The well-tempered ensemble bias on the potential energy of the system was built mostly over a short 10 ns simulation using a hill height of 2.0 kJ/mol deposited every picosecond decaying with a bias factor of 20. During the production simulation, bias was deposited more slowly at one hill every 50 ps.

From the metadynamics results, representative structures were selected from the low-energy tetrameric and pentameric states and back-mapped to atomistic structures.[74] In order to demonstrate that the low-energy states were plausibly stable, these atomistic structures were simulated using standard molecular dynamics for 50 ns with the CHARMM36 and TIP3P force fields for protein and water respectively.[75] The size of the simulation box was reduced to eliminate unnecessary water. Chlorine ions were added to neutralize the box. Simulations began with 10,000 steps of steepest descent minimization, followed by 1 ns of heating from 0 K to 300 K to ease bad contacts. The simulations continued an additional 49 ns using the Parinello-Rahman barostat at 1 bar. The time step of integration was set at 2 fs, with hydrogen-heavy atom bonds restrained using the LINCS algorithm. Temperature was maintained at 300 K with a stochastic velocity rescaling thermostat. PME was used for long range electrostatics. Short-range electrostatics and Lennard-Jones interactions were cut-off at 1.2 nm with a potential modifier to avoid artifacts.

Table 1: Important input parameters taken from MARTINI developer suggestions. Rows marked by a semicolon are ignored in GROMACS.

cutoff-scheme	= Verlet
nstlist	= 20
ns_type	= grid
pbc	= xyz
verlet-buffer-tolerance	= 0.005
coulombtype	= reaction-field
rcoulomb	= 1.1
epsilon_r	= 2.5
epsilon_rf	= 0
vdw_type	= cutoff
vdw-modifier	= Potential-shift-verlet
rvdw	= 1.1
tcoupl	= v-rescale
tc-grps	= system
tau_t	= 1.0
ref_t	= 300
Pcoupl	= no
;Pcoupltype	= isotropic (only for NPT equilibration)
;tau_p	= 12.0
;compressibility	= 3e-4
;ref_p	= 1.0
gen_vel	= no
gen_temp	= 300
gen_seed	= -1

3.5 Additional Data

Table 1: Fit results for amide I VSFG spectra of LKα14. Amplitude, frequency (in cm^{-1}) and half width at half max in (cm^{-1}) are denoted by A, w and Γ.

	without TMOS	**with TMOS**
NR Amplitude	0.1	0.1
NR Phase ϕ	2.8	2.6
A	2.3	1.9
w	1638.7	1638.5
Γ	17.1	17.4

Table 2: Fit results for amide I VSFG spectra of Ac-LKα14. Amplitude, frequency (in cm^{-1}) and half width at half max in (cm^{-1}) are denoted by A, w and Γ.

	without TMOS	**with TMOS**
NR Amplitude	0.2	0.2
NR Phase ϕ	2.5	3.1
A	6.9	22.9
w	1638.6	1641.9
Γ	14.8	23.2

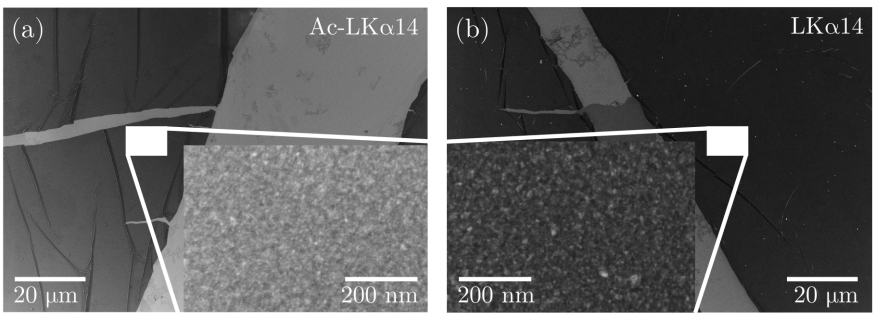


Figure 1: SEM images of Ac-LKα14 (a) and LKα14 (b) mineralized silica films at the air-water interface, deposited by the Langmuir-Schaefer method on silica substrates. The substrates were coated with 1.2 nm chrome and 50 nm gold by thermal evaporation in advance. Brightness and contrast in the above images vary based on sample and acquisition settings. Insets were taken from areas covered with peptide-silica film. Areas of lighter shading represent the gold surface of the substrate.

Fourier transform infrared spectroscopy: A drop of the respective peptide solutions used for bulk silica precipitation was placed on the crystal of a Tensor II FTIR spectrometer (Bruker, Billerica, MA, USA) and measured with 100 scans. The data was background corrected by subtracting a measurement of buffer only and compensated for atmospheric absorption by means of an internal algorithm of the measurement software OPUS 7.5 (Bruker).

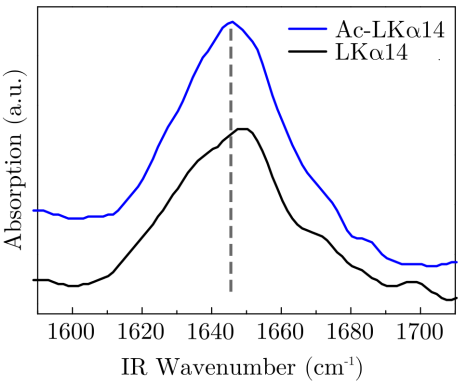


Figure 2: Background-corrected infrared spectra of LKα14 (black) and Ac-LKα14 (blue) at 0.2 mg/ml in PBS buffered D_2O solution. Both datasets were vertically displaced but scaled within the same dimensions to ensure comparability.

Convergence of the simulation: Free energy is taken as an integral over the free energy profile in the range of coordination numbers 5-7 for tetramers and coordination numbers 8-10 for pentamers.

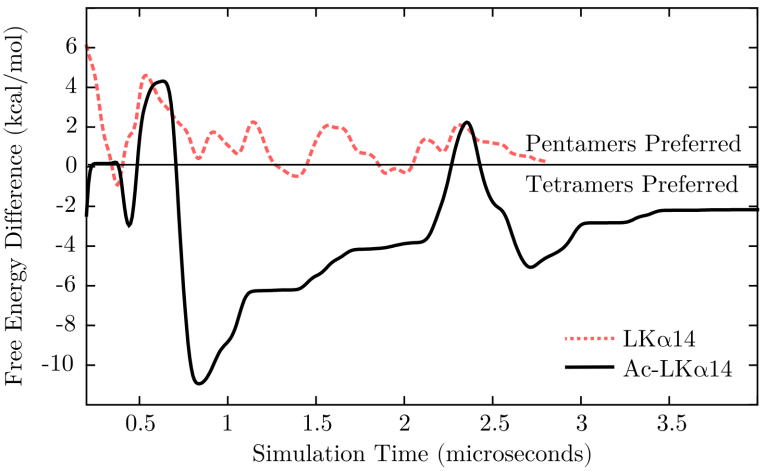


Figure 3: Convergence of the free energy difference between tetramers and pentamers.

4 Interfacial Silica Biomineralization by R5

Acknowledgement: Dr. Vance Jaeger conducted the molecular dynamics simulations. Dr. Lars Schmüser's contribution was in the calculation of VSFG spectra from molecular simulation frames. SEM and TEM were performed and interpreted by Gunnar Glasser and Katrin Kirchhoff, respectively. Helma Burg and Dr. Rüdiger Berger conducted and analyzed SFM measurements. The execution and analysis of VSFG spectroscopy measurements, biomineralization experiments and the analysis of simulation trajectories were performed by me.

4.1 Introduction

By using artificial model peptides we have learned that peptide agglomeration and structure influence the morphology of silica nanostructures. In this chapter we study interfacial silica formation using a more complex system. As mentioned in Section 1.2 of the introduction, the silica formation in diatoms is thought to be controlled by type-1 silaffin peptides and long chain polyamines (LCPAs).[17] Both groups of biomolecules are embedded in biogenic silica and have been observed to rapidly induce the precipitation of ~500 nm diameter spherical silica particles in vitro. While type-1 silaffin peptides are highly post-translationally modified phosphor-peptides, Kröger et al. have shown that an unmodified sequence based on silaffin-$1A_1$, called R5 (SSKKSGSYSGSKGSKRRIL), precipitates structures very similar to those of silaffin-$1A_1$ (see reference [19] and Section 4.5 - Figure 1).

Despite the fact that R5 is used in numerous nanotechnological studies,[76] structural information about R5-silica interactions is sparse and, in some regards, contradictory.[22, 43] A solution-state nuclear magnetic resonance (NMR) study by Senior et al.[43] has determined that R5 remains unstructured and monomeric when in contact with silica and that amino acid sequence rather than secondary structure dictates the silicification process. Solid-state nuclear magnetic resonance (ssNMR) studies by Roehrich and Drobny used backbone and side chain chemical shifts to follow the structure of R5 within silica precipitates and proposed that R5 maintains a strand-like structure with a micelle-like assembly in silica precipitates.[22] Within this model, the C termini are buried within the micelles with the RRIL motif mediating peptide-peptide interactions, while the N-terminal positively charged side chains are located at the micelle surface and drive the silica precipitation. Other groups have suggested physico-chemical processes like self-assembly of organic matrices and phase separation.[17]

While these previous studies reveal a wide range of structural properties of R5 peptides in aggregated or precipitated states in the presence or absence of silica, they tell us little about the interactions between R5 and two-dimensional interfaces. The biological silicification process occurs largely at existing, long-lived interfaces. Therefore, studies of the interfacial properties of R5 may provide additional details about the structural basis for biosilicification. A powerful technique capable of detecting interfacial protein structure in situ is vibrational sum-frequency generation (VSFG) spectroscopy. The selection rules dictate that the sum frequency spectrum represents modes exclusively from the interface. Amide VSFG spectra can thereby provide information about the folding of proteins at interfaces.[65]

4.2 Results and Discussion

Using VSFG, we have monitored the lateral peptide assembly and interfacial biomineralization within a peptide monolayer formed at the air-water interface. Here, R5 is assembled at the interface by injecting the peptide into the subphase of a 20 mL citrate buffer solution contained in a trough to reach a peptide concentration of 0.1 mM. Upon equilibration, monitored through the surface tension (Section 4.5 - Figure 2), the subphase is diluted by 8 fold to reduce the bulk peptide concentration. Subsequently, tetramethoxysilane (TMOS), the molecular precursor for silica, is injected into the subphase. The procedure leads to biosilica growth at the air-water

interface. As illustrated in Figure 4.1a, a thin layer of biosilica, consisting of silica and peptides, can be lifted off of the interface using a transmission electron microscopy (TEM) lacey support grid.

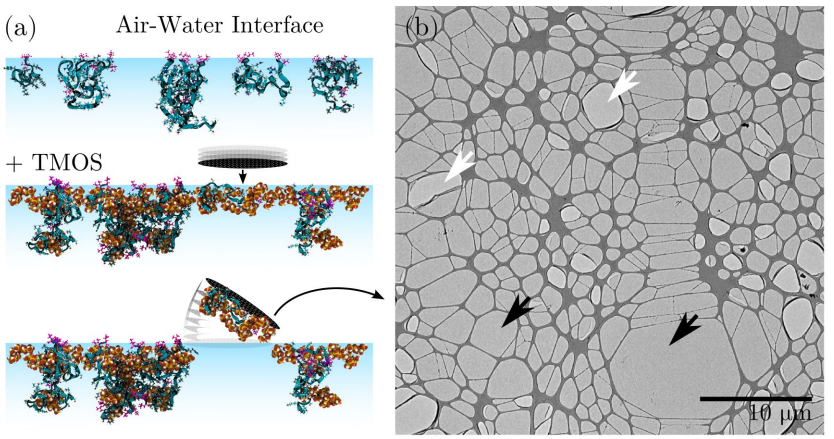


Figure 4.1: (a) Interfacial R5 mediated silica precipitation. (b) TEM image acquired by K. Kirchhoff of a R5 templated silica film collected from the air-water interface with a lacey grid. To highlight the stability of the film some of the combs where the silica film ruptured are indicated with white arrows, and some combs with an intact silica layer are indicated with black arrows.

The obtained thin silica sheets, shown in Figure 4.1b, provide an ideal platform to study interfacial silicification. In addition, with their stability over areas of up to 8 μm in the dry state they can potentially be harnessed in nanotechnological applications. Height profiles measured by H. Burg and R. Berger with scanning force microscopy (SFM) revealed that the film is 12 nm thin (Section 4.5 - Figure 3).

To determine the structure of R5 within the biomimetic silica film, we probed the conformation of R5 before and after biomineralization with VSFG. The secondary structure of the interfacial peptides can be deduced from the position of amide I resonances.[53a] All spectra were measured using a D_2O-based buffer to avoid interference with the water bending vibration.[53a] Spectra measured in ssp polarization (s polarized SFG, s polarized VIS and p polarized IR) before and after TMOS injection are depicted in Figure 4.2a.

While the amide I mode before silicification can be represented with two peaks, one near 1647 cm^{-1}, the other near 1664 cm^{-1}, the amplitude of the latter decreases about 30 % after biosilicification. We can assign the latter to

a β-sheet-like and the former to α-helical or random conformation.[53a, 77] VSFG spectra in the sps polarization (Figure 4.2b) exhibit a resonance near 1645 cm^{-1} both before and after biomineralization. The turn and sheet motifs observed in the ssp spectra near 1664 cm^{-1} are apparently not SFG-active in the sps polarization combination.

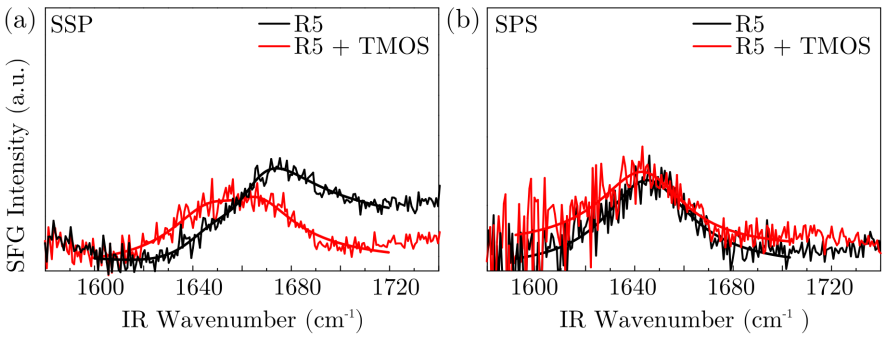


Figure 4.2: VSFG spectra of R5 at the air-water interface before (black) and after interfacial precipitation of silica. Spectra were acquired in *ssp* (a) and in *sps* polarization combination (b). Fits are overlaid with experimental data. Fit parameters are given in Section 4.5 - Table 1 and Table 2.

While the VSFG spectra indicate that R5 is rather ordered at the interface, undergoing subtle structural changes upon silicification, it is difficult to draw a detailed molecular picture from the VSFG spectra of this complex system by spectral inspection alone. We therefore obtained a more direct picture of the peptide conformation when in contact with silica using a recently developed method where the experimental data are connected to molecular dynamics (MD) simulations by theoretical VSFG spectra calculated from the MD snapshots.[78]

MD simulations were conducted by V. Jaeger. GROMACS 4.6[73a] with PLUMED 2.1[73b] was used to generate an ensemble of structures from which theoretical VSFG spectra could be calculated. An initial set of simulations was used to generate an R5 peptide structural ensemble. Boxes containing a vacuum-water interface were packed with water, ions, and a single peptide using Packmol.[62] AMBER99SB-ILDN,[31b] TIP3P,[75b] and a modified GAFF[57] force field were used for protein, water, and phosphate or silicon-containing additives[65] respectively. Parallel tempering in the Well-Tempered Ensemble[39] was used to ensure that a diverse peptide structural ensemble was sampled. Subsequently, a second set of 9 production simulations each

combined ten randomly chosen peptides from the earlier Well-Tempered Ensemble simulations into a single simulation box. These production simulations contained a vacuum-water interface, and the system was allowed to evolve without any bias in the Canonical Ensemble for 100 ns at a temperature of 300 K. A third set of simulations added a silica precursor to 9 similarly constructed silica-production simulation boxes in order to mimic the early stages of silica-peptide interactions, each at 300 K. Frames from the second and third sets of simulations used to calculate VSFG spectra are shown Figure 4.3. For VSFG calculations the bulk peptides have been removed from the structure file. Additional details on the MD simulation are provided in Section 4.4.

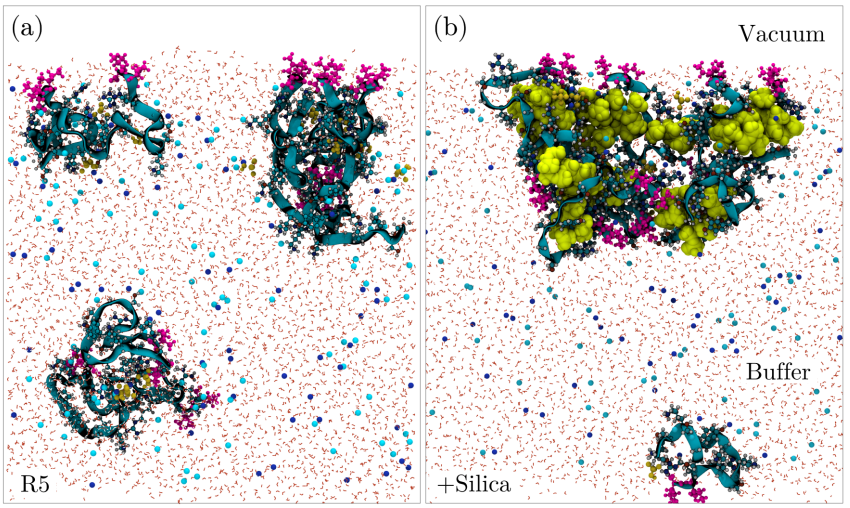


Figure 4.3: Snapshots of MD simulations after 80 ns without (a) and with silica (b). Color coding as follows: peptide backbone in cyan, leucine and isoleucine in magenta, silica in grey, phosphate ions in yellow, sodium and chlorine in dark and light blue. Periodic imaging was turned on.

VSFG spectra calculations were performed by L. Schmüser. The calculated spectra are shown in Figure 4.4. The calculated spectra capture the main resonance positions of the experimental data for both ssp and sps spectra. Importantly, the calculated ssp spectra also predict the significant red-shift of the spectral weight observed in the ssp spectra upon silica interaction. In the sps data, the experimental spectra show a shift of the resonance to the red side of the spectrum. This effect is also captured by the

calculations. Besides an overall amplitude scaling factor, no adjustable parameters were used when comparing the experimental and theoretical results. The general agreement of independent calculations and experimental results warrants a closer look into the predicted peptide conformation.

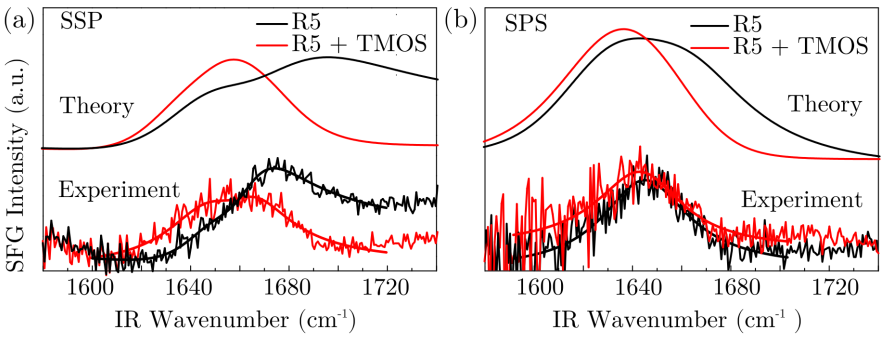


Figure 4.4: VSFG spectra of R5 as in Figure 4.2. Calculated spectra from MD simulation frames are shown with an offset above the experimental data.

A quantitative analysis of the production simulations was performed on 80 frames of the last 40 ns (one frame per 0.5 ns) of 9 replicate simulations. The sample window was selected to allow for side chains and hydrogen bonds to reorder between samples, a process which occurs largely on timescales in tens to hundreds of picoseconds.[79] Thus, the data presented in Figure 4.5 and Figure 4.6 represents an ensemble of 720 frames at a temperature of 300 K. The secondary structure of R5 was characterized by extracting the backbone dihedral angles ψ (C_β-C_α-C-O) and φ (H-N-C_α-C_β). Plotted as a Ramachandran plot, these angles provide a fingerprint of the backbone conformations (Figure 4.5a and b). Clearly the Ramachandran plot shows β-sheet (upper left corner) and helix content (middle, left) expected from the experimental VSFG data. For a better visualization of the changes within the peptide ensemble upon interaction with the silica, we subtracted the binned angle populations for peptide-only and peptide with silica to retrieve the difference Ramachandran plot shown in Figure 4.5c.

The upper left quarter, the region characteristic for β-strand conformation, loses up to 15 % per bin in angle population. The total fraction of beta sheet conformation changes from 31 % to 29 %.

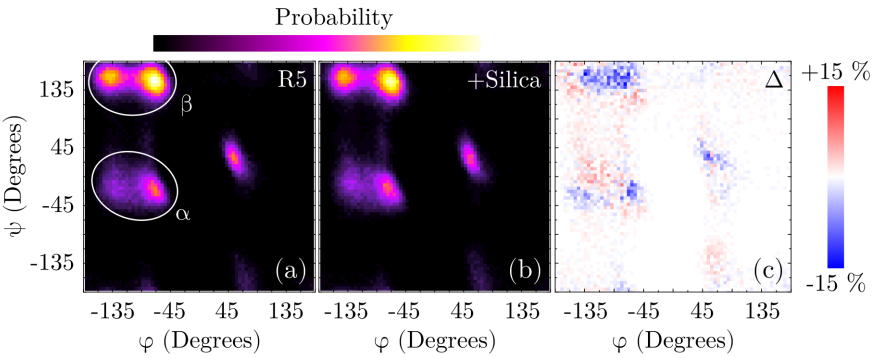


Figure 4.5: Ramachandran plot of the backbone dihedral angles ψ and φ in degrees extracted from 40 ns of simulation time without (a) and with silica clusters (b). Color scaling was performed by setting the largest angle population encountered to 100 %. The plots were subtracted from each other to generate the difference Ramachandran plot shown in (c).

This demonstrates that R5 is structured at interfaces and undergoes subtle but important reorganization when in contact with silica as shown by the changes presented in the theoretical curves in Figure 4.4a.

Reorientation of peptides in response to the silica interaction is unlikely in view of the relatively constant intensity ratio within the ssp and sps polarization spectra. Since this reorganization is also only slight on the secondary structure level, with apparently only small changes to patterns in β-strand population, it is likely that the changes in our spectra are due to hydrogen bonding rearrangement in the presence of silica (see Section 4.5 - Figure 4 for more detail). This leads to changes in amide I transition dipole coupling, which manifests itself in the theoretical and experimental spectra.

To address the question about changes in inter-peptide contacts that lead to shifting backbone hydrogen-bonding networks, we analyzed the trajectory using the GROMACS tool g_mdmat. By defining a cutoff of 0.4 nm we determined the mean number of residue-residue and backbone-backbone contacts of different chains. The standard error of mean was calculated as described in Section 4.4. Figure 4.6a and b show that inter-peptide contacts are generally reduced when silica is present, with C-terminal Ile and Leu residues as the exception. We attribute this loss of interactions to the intercalation of silica molecules into the R5 aggregates. In other words, some peptide-peptide contacts are replaced by peptide-silica contacts. This intercalation of silica, when R5 is at an interface, is in contrast to what has

been proposed by Roehrich and Drobny for R5 micellar aggregates in solution, where the interaction with silica was believed to be limited to the outer perimeter of the aggregates.[22]

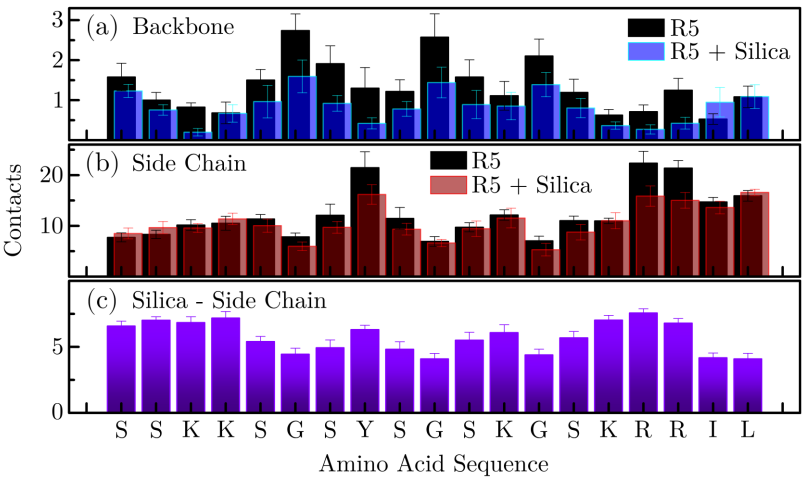


Figure 4.6: (a) Mean number of inter-peptide backbone contacts calculated over a total of 1000 frames, extracted from 10 independent production simulations. (b) Mean number of inter-peptide residue contacts. (c) Mean number of side chain-silica contacts. Error bars represent the standard error.

The Y8, R16 and R17 side chains have the highest number of peptide contacts before silica addition. With silica present, these numbers of peptide contacts are reduced here and, based on the analysis of the silica contacts in Figure 4.6c, replaced by silica molecules. The latter analysis also indicates that the silica is more or less affecting all residues with the most contacts towards the hydrophilic N-terminal residues. The Y8 residue in the middle of the sequence is an outlier here, with strong silica interactions. It is interesting to take a closer look at the C-terminal Ile and Leu units. These hydrophobic side chains anchor the peptides to the interface (Figure 4.3) and therefore interact rarely with silica because of their preference to be exposed to air-vacuum. The observation by other groups that a deletion of the terminal RRIL sequence renders the peptide incapable of precipitating silica supports our view that the role of the terminal IL group is to anchor the peptides during peptide assembly and mineralization.[21, 80]

4.3 Conclusion

In conclusion, we show that R5 can grow self-supported, nanometer thin biosilica sheets that are stable over several micrometers. In contrast to reports of unstructured R5 based on solution-state NMR data, we found that R5 is structured at interfaces and maintains a defined conformation within the silica sheets. During the mineralization process, R5 undergoes a conformational change where the peptide structure loses predominantly β-sheet content. The silica sheet grows by silica intercalation into the peptide aggregates, a process driven by silica contacts with all sites along the sequence. This intercalation was not observed for peptide aggregates in solution, where solid-state NMR data suggest interaction only at the C-terminus. Biosilicification is an interfacial process. This becomes clear from the differences of previous solution state data. Our results indicate that, in order to understand natural silica biogenesis by R5, it is important to take interfacial effects into account.

4.4 Specific Experimental Details

Chemicals: R5 was purchased from Genscript with a purity of $> 90\%$. Chemicals were purchased from Sigma Aldrich.

Biomineralization at the air-water interface: Peptide solutions (20 ml, 0.1 mM R5 peptide), were prepared by dissolving lyophilized peptides in citrate buffer (0.1 M Na_2HPO_4, 0.05 M citrate, pH 7). The solution was incubated for 1 h in a Teflon trough with 79 cm^2 of the solution exposed to air, to establish an equilibrium of surface adsorbed peptides. Additionally the peptide concentration in the bulk solution was reduced by exchange of the subphase (10 ml) with citrate buffer (10 ml) three times. TMOS was pre-hydrolyzed in 1 mM HCL while sonicating for 5 min and subsequently injected into the subphase of the peptide solution with a final TMOS concentration of 25 mM.

Sum frequency generation spectroscopy: VSFG experiments were conducted using the setup described in Section 1.3.2. The 800 nm (VIS) beam and IR beam were directed at the sample at an angle of 60° (IR) and 55° (VIS) respective to the surface normal to generate VSFG. All data was acquired for 20 min at 25° C. All spectra were recorded using D_2O as solvent. The data was corrected by subtracting background spectra (VIS only), and

energy was calibrated using the nonresonant VSFG response from z-cut quartz. The data was fit according to Equation 1.13.

Simulation of VSFG spectra (conducted by L. Schmüser): The method used for calculating the VSFG spectra has been described in detail before.[78] Briefly, the amide I excitonic Hamiltonian was constructed from snapshots of the calculated MD simulations. By solving the time-independent Schrödinger equation we calculated the delocalized eigenvectors and eigenvalues which were used to calculate the IR and Raman responses of the peptides. The VSFG hyperpolarizability was obtained by calculating the tensor product of the IR and Raman responses. The second order nonlinear susceptibility of the peptides $\chi^{(2)}$ was obtained by transforming the molecular frame to the lab frame by averaging the Euler transformation between both frames over the orientation distribution of the peptides. To account for laser polarization dependent refractive indexes, the $\chi^{(2)}$ response was further multiplied with the corresponding Fresnel factor. Additionally a non-resonant background can be added to account for differences between the resonant and the non-resonant phase. The shape of the calculated VSFG spectra was approximated to the experimental spectra by varying the non-resonant phase, the amplitude of the non-resonant background and the linewidth of the Lorentzian functions which were used to account for homogeneous broadening.

Scanning and transmission electron microscopy (performed and interpreted by G. Glasser and K. Kirchhoff): Solution silica precipitates were washed by centrifugation at 6000 rpm for 2 minutes, discarding the supernatant and adding water. The process was repeated three times. The solid pellet was freeze-dried overnight and a small sample was placed on sticky carbon tape for SEM. The samples were imaged on a Zeiss Gemini 1530 instrument using an accelerating voltage of 750 V. SiO_2-peptide nano-sheets were transferred from the air-water interface onto lacey copper TEM grids using the Schaefer technique. Samples for TEM imaging were prepared by placing a 300 mesh lacey copper grid (Plano) with coated side down on an interfacial peptide-silica film. After removal from the interface the grid was rinsed with 500 µl MilliQ water and dried in ambient conditions. Silica films immobilized in such a way were imaged with a FEI Tecnai F20 instrument operated at an accelerating voltage of 200 kV.

Scanning force microscopy (measured by H. Burg and interpreted by R. Berger): Sample thicknesses were measured in a Dimension ICON scanning force microscope (Bruker) in tapping mode using an OMCL AC160TS cantilever.

Molecular dynamics simulations (conducted by V. Jaeger): Parallel tempering simulations were conducted with 10 replicas at temperatures 273.0, 282.0, 291.4, 301.3, 311.7, 322.7, 334.3, 346.5, 359.4, and 370.0 K with exchanges attempted every 500 fs. The Well-Tempered Ensemble was used to increase fluctuations in potential energy and reduce the number of replicas needed to span the temperature range. A well-tempered bias was built over 10 ns using a bias factor of 50, a hill height of 2.0 kJ, and a hill width (σ) of 250 kJ at a rate of one hill per picosecond. The simulation box contained 20,000 water molecules and was 8.1 x 9.5 x 13.2 nm in dimension with the slab of water filling 7.6 nm of the 13.2 nm box height. A single R5 peptide was placed at the water-vacuum interface consisting of TIP3P waters and an appropriate amount of sodium and chlorine ions to neutralize the system and maintain a realistic ionic strength (150 mM). Subsequently, simulations were conducted for 100 ns in the Well-Tempered Ensemble using the bias applied during the bias-building simulation, but the rate of hill deposition was lowered to one per 50 ps with all other parameters remaining the same. Peptide structures were extracted at random points from the ensemble at 301.3 K and combined to create boxes for the production simulations. The volume of the water-slab in the box varied little with temperature, indicating that even at 370 K the Well-Tempered ensemble simulations are sampling a liquid water-vacuum interface. The charge of the peptide was defined according to the experimental pH of 7.4. Simulations were run with GROMACS 4.6 and the PLUMED 2.1 plugin. Classical MD was then used for production simulations. From the diverse set of structures obtained from PT-WTE simulations simulation boxes of 10 peptides were set up and run for 100 ns with 7 monovalent and 7 divalent phosphate ions in addition to the sodium and chloride ions from before as seen in Section 4.5 - Figure 5. These new simulations contained 20,000 water molecules with neutralizing sodium and chlorine ions (150 mM) in a box of 8.2 x 9.2 x 13.5 nm in dimension. The water slab occupied approximately 7.5 of the 13.5 nm box height. During all of the simulations, temperature was controlled using a stochastic velocity recalling thermostat with an integration timestep of 2 fs facilitated by the freezing of bonds using the LINCS algorithm.[64, 81]

The standard error $\sigma_{\bar{x}}$ of simulation analyses was calculated according to the following equation:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{\mathrm{n}}} \tag{4.1}$$

Here, σ represents the sample standard deviation and n is the sample size. In Figure 4.6, the mean contact numbers for each side chain and the standard error of the mean were obtained from 9 replicate simulations. For each of the 9 simulations, the contact numbers of each side chain were averaged for 80 frames (the last 40 ns). This procedure yields 9 mean contact numbers for each side chain. The mean of these 9 mean values is plotted in Figure 4.6. The standard error of the mean for each side chain was obtained by calculating the standard deviation of the mean of the 9 replicate simulations. Subsequently the standard deviation is divided by the square root of the sample size, in this case 9.

4.5 Additional Data

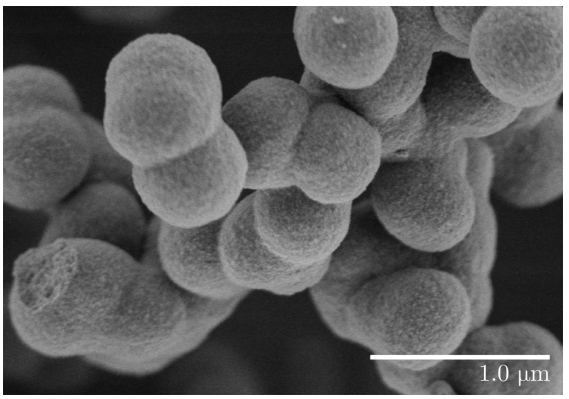


Figure 1: Silica precipitates from a 1 mg/ml solution of R5.

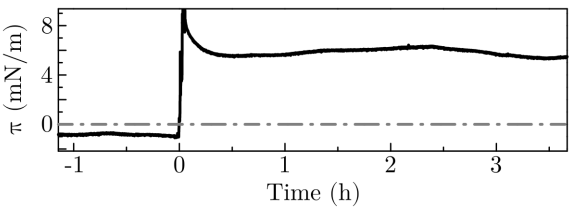


Figure 2: Surface Pressure measurement of R5 adsorption to the air-water interface. The surface pressure (π) was normalized with pure water to 0 mN/m. Upon exchange of water with citrate buffer, the surface pressure quickly stabilized to -1 mN/m. The time on the x-axis was set to 0 hours when the peptide was injected into the subphase. The spike after peptide injection was caused by detaching the needle from the surface and re-attaching it. This was followed by an equilibration of the surface pressure over the course of ~ 30 min.

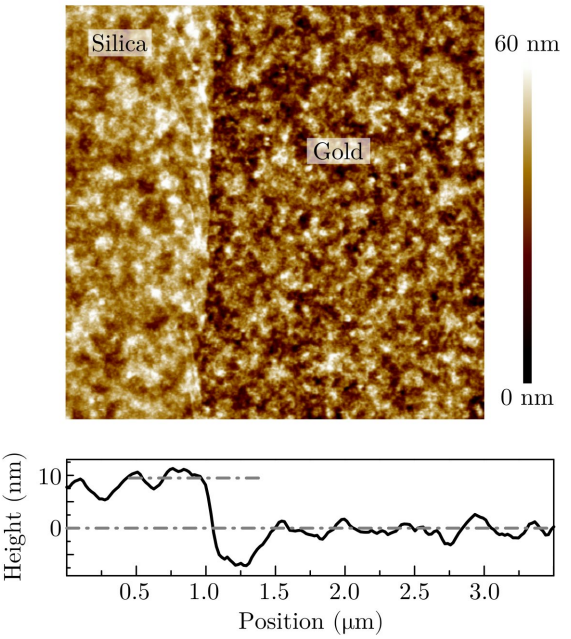


Figure 3: Illustration of a representative measurement of height profiles with SFM. The average film thickness is 12.2 nm with a standard deviation of 2.7 nm estimated from three measurements.

Table 1: Fit results for R5 without and with silica in the *ssp* polarization combination. Amplitude, frequency (in cm^{-1}) and half width at half max in (cm^{-1}) are denoted by A, w and Γ.

	R5	**R5 + Silica**
NR Amplitude	0.1	0.1
NR Phase ϕ	3.0	1.9
A_1	0.6	0.7
$\mathbf{w_1}$	1642.6	1644.4
Γ_1	20.8	20.1
A_2	0.6	0.4
$\mathbf{w_2}$	1665.2	1666.1
Γ_2	15.7	14.3

Table 2: Fit results for R5 without and with silica in the *sps* polarization combination. Amplitude, frequency (in cm^{-1}) and half width at half max in (cm^{-1}) are denoted by A, w and Γ.

	R5	**R5 + Silica**
NR Amplitude	0.1	0.1
NR Phase ϕ	1.6	1.5
A_1	0.8	0.6
$\mathbf{w_1}$	1644.6	1643.8
Γ_1	21.2	21.5

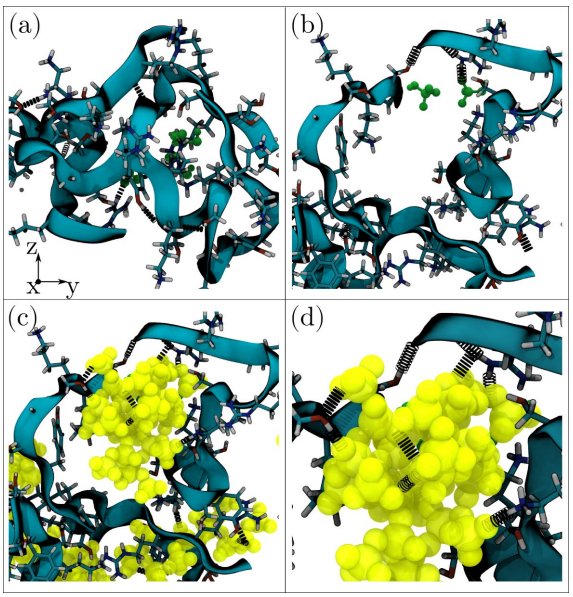


Figure 4: Hydrogen bonding network of peptide agglomerates in the simulation without (a) and with silica (b, c and d). Structures shown in (b) and (c) are identical except that silica and the H-bonds the silica is participating in are hidden. Without silica molecules (a), the peptides agglomerate with an extensive network of inter-peptide H-bonds. With silica molecules in the simulation (b and c), the peptide agglomerates are interspersed with silica molecules, which participate in the H-bond network and reduce the number of inter-peptide H-bonds. Panel (d) displays a two-fold magnification of (c). For clarity, only side chains that are able to form hydrogen bonds are shown. Color coding as follows: peptide backbone in cyan, phosphate ions in green, silica in yellow.

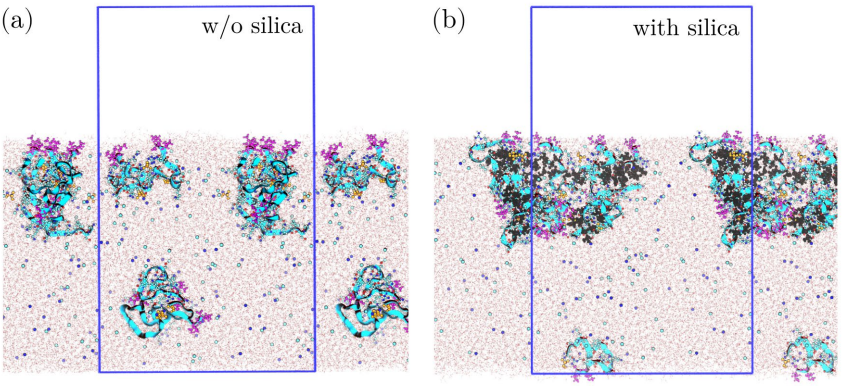


Figure 5: Simulation boxes of R5 without silica (a) and with silica precursors (b).

5 Peptide Structure Prediction at the Air-Water Interface

Description and acknowledgement of coauthors: This chapter represents preliminary findings and is not based on a published research article. The presented work is a close collaboration of Helmut Lutz and Dr. Vance Jaeger with equal contributions. Both conducted the molecular simulations and the analysis/interpretation of the simulation data. I conducted experimental VSFG measurements and Dr. Vance Jaeger calculated theoretical VSFG spectra.

5.1 Introduction

The experimental determination of protein structure at interfaces is a non-trivial problem. The most common methods of structural determination (i.e. x-ray crystallography, nuclear magnetic resonance (NMR) or cryo electron microscopy) probe protein conformation only in the solution or crystal phase.[82] However, none of these methods can inform us about the ways in which the protein conformation changes at any given interface. Of the tens of thousands of solved, high-resolution protein structures, none describes a protein at an interface. This is unfortunate, because the interfacial structure and dynamics of proteins is of utmost importance for understanding several relevant biological processes, e.g. the handling of protein or peptide drugs on industrial scales, food formulations, biofilm formation, and biomineralization.[65, 83]

A versatile method to determine the secondary structure and orientation of a protein at interfaces is vibrational sum-frequency generation (VSFG) spectroscopy, introduced in Section 1.3.[54a, 78, 84] Unfortunately, state of the art VSFG and other surface-sensitive techniques do not provide molecular-level detail in the same way as x-ray crystallography or NMR can. Therefore,

scientists often rely on molecular simulations to support VSFG experiments by producing an ensemble of possible protein conformational states based upon hypotheses about the interfacial coordinates.[85]

However, one consideration that has not been sufficiently explored in previous papers about proteins at interfaces is the effect of protein and water force fields on the protein's conformational ensemble. Since the major branches of both AMBER and CHARMM-based force fields are optimized to reproduce data of aqueous phase proteins and biomolecules,[31a, 86] there may be considerable effects when introducing an interface into the system, as has been observed for several solid-water interfaces. These effects include: (a) differences in solvation free energy of the backbone and side chains among protein force fields,[87] (b) differences in the properties of interfacial water compared to bulk water,[88] and (c) differences in peptide interactions among other considerations.[89]

A particularly simple and virtually omnipresent interface to study the behavior of protein and water force fields is the air-water interface. In order to determine whether some of the modern protein and water force fields can capture the physics of the air-water interface, we test several force field combinations in simulations of a peptide at this interface. The ability of these simulations to produce peptide ensembles of accurate structure is evaluated by calculating theoretical VSFG spectra from the simulations and comparing these to experimental VSFG data.

Several enhanced sampling techniques have been used to exhaustively sample ensembles of protein conformations at interfaces, including well-tempered metadynamics (WT-MetaD), parallel tempering,[35b, 90] solvent tempering,[91] and the well-tempered ensemble.[39, 92] We have chosen to enhance the exploration of the protein conformation using WT-MetaD,[36] introduced in Section 1.4.2. Because the air-water interface does not restrict the sampling of side chain and backbone conformations in the same way as a solid hydrophobic or charged interface, we hypothesize that WT-MetaD is sufficient to exhaustively sample the free energy of the system without any further biasing such as parallel tempering which might be used to overcome smaller hidden free-energy barriers associated with protein adsorption.

We combine the protein force fields AMBER99SB-ILDN (A99SB-ILDN), CHARMM27 (C27), AMBER99*SB-ILDN (A99SB*-ILDN) and CHARMM22* (C22*) along with three water force fields (TIP3P, TIP4P-D, and SPC/E).[31a, 31b, 32, 75b, 93] A list of force field combinations is provided in Table 5.1.

Table 5.1: Force field combinations used to simulate aurein 1.2. Protein force fields are abbreviated as follows: AMBER99SB-ILDN (A99SB-ILDN), CHARMM27 (C27), AMBER99*SB-ILDN (A99SB*-ILDN) and CHARMM22* (C22*). TIP3P, SPC/E and TIP4P-D denote the water force fields respectively.

A99SB-ILDN TIP3P	A99SB-ILDN SPC/E	A99SB-ILDN TIP4P-D
A99SB*-ILDN TIP3P	A99SB*-ILDN SPC/E	A99SB*-ILDN TIP4P-D
C27 TIP3P	C27 SPC/E	C27 TIP4P-D
C22* TIP3P	C22* SPC/E	C22* TIP4P-D

Six peptides of varying composition and secondary structure have been selected to test these force fields and our ability to reproduce VSFG spectra from molecular dynamics-derived ensembles. This chapter presents preliminary data for the antimicrobial peptide aurein 1.2 (helical, pdb: 1VM5) as a proof of concept and a starting point for future research.[94] The peptide aurein 1.2 was selected because of its amphiphilic nature, which makes it segregate to the air-water interface without any bias. Aurein 1.2 and the starting structure for simulations are shown in Figure 5.1.

From simulations and experiments on aurein 1.2, we determined which force field combinations best reproduced the experimental spectra collected herein, and thus these force fields likely reproduce the real ensemble of peptide conformations. Ultimately, we hope that the protocol developed in this chapter will encourage a more symbiotic relationship between MD and VSFG researchers, wherein both methods can be used in tandem to develop and test hypotheses about interfacial proteins.

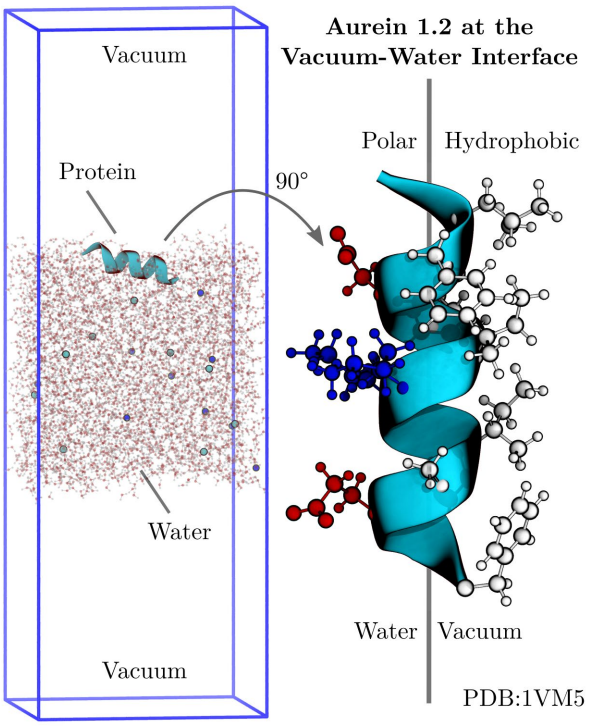


Figure 5.1: Left: Starting structure for simulations. The box boundaries are indicated in blue. The peptide backbone is colored cyan. Sodium and chloride ions are colored in dark and light blue. Water molecules are colored red and white. Right: Chemical structure of aurein 1.2 as determined from solution-state NMR, and its putative orientation at the water-vacuum interface. Hydrophobic side chains are colored white, acidic side chains are colored red and basic side chains are colored blue.

5.2 Results and Discussion

Enhanced sampling through WT-MetaD was used to assess the ability of different force fields to reproduce a natural ensemble of peptide secondary structures at the air-water interface. To this end, we biased the Cα radius of gyration and the number of structural hydrogen bonds of aurein 1.2 as collective variables. The free-energy landscapes auf aurein 1.2 with respect to these biased collective variables are shown in Figure 5.2. Here, we note that simulations with different combinations of water and protein force fields return very different results for free-energy minima.

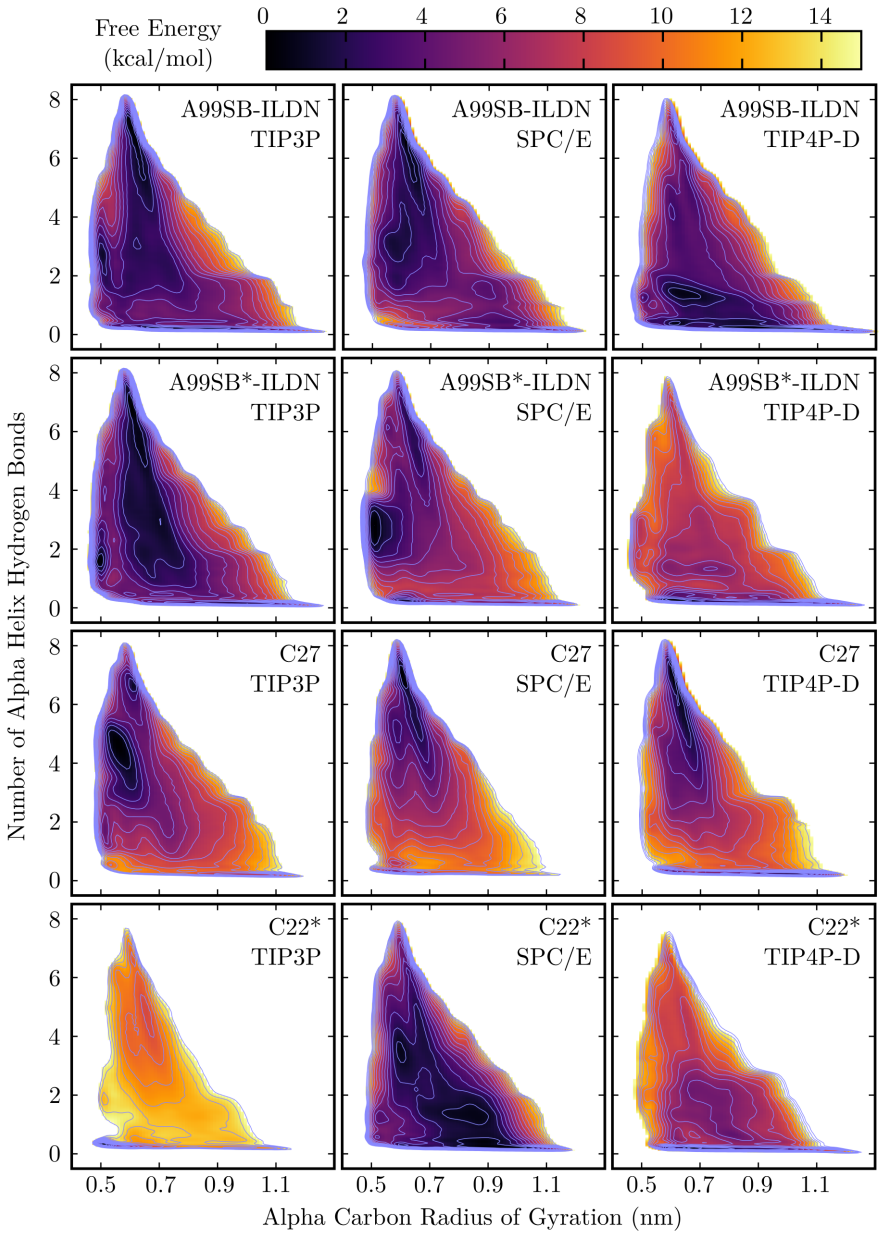


Figure 5.2: Free energy in CV space obtained from twelve 1 μs long WT-MetaD simulations of aurein 1.2. The CVs are denoted on the x- and y-axis as Cα radius of gyration and number of α helix hydrogen bonds. The protein/water force field combination used in each simulation is indicated in the top right corner of the free energy surface.

The native solution structure as determined from NMR sits roughly at a coordinate of 0.65, 6.0 in the plots shown in Figure 5.2. A common feature in all free-energy landscapes is the narrow horizontal region near zero α-helix hydrogen bonds. Peptides located in this area of the free-energy landscape form hydrogen bonds neither at the same positions as seen in the NMR structure nor in a perfect α-helical pattern. However this does not exclude the presence of other secondary structures. This horizontal band contains a range of diverse structures from highly compacted to completely extended, high-entropy states. Most of the 12 force field combinations retain a basin near the solution structure, except in the cases of C22* and some of the TIP4P-D simulations. Because the "*" variations of the force fields adjust the protein backbone parameters to change the population of α-helices versus coils, a change in the propensity to form a helix at the interface should be expected. TIP4P-D was parameterized in order to better sample intrinsically disordered proteins, and since proteins and peptides at interfaces often become disordered, we wanted to test whether this water model might give a population that differs greatly from the solution structure. This was in fact the case for three of the force field combinations containing TIP4P-D water. There are other states to be explored besides the extended state and the native folded state. For example, in several of the AMBER force field combinations a second well appears for a compacted, less helical structure near the coordinate 0.55, 3.0. If such a structure maintains some of its helical character, it could be expected to generate a VSFG signal that has α-helical character with some additional features.

One anomalous case exists for the free-energy profiles. That is for C22* with TIP3P water. In this case, we observe a high probability of being in a compact state with no α-helical hydrogen bonds. This is an effect of metadynamics failing to account for a hidden degree of freedom. This causes additional bias to be placed in a narrow region of phase space for an extended time leading to a downward shift in the minimum of the free-energy diagram for this force field combination. The relative populations of the other states in the system indicate that without this anomaly, states near the native solution structure would be highly populated. To demonstrate a measure of convergence for the plots in Figure 5.2, we plot the metadynamics hill height over the time of the simulation in Figure 5.3. While a low hill height does not prove the convergence of the metadynamics simulations, it does mean that for the last half of the simulations, no new regions of conformational space are being explored.

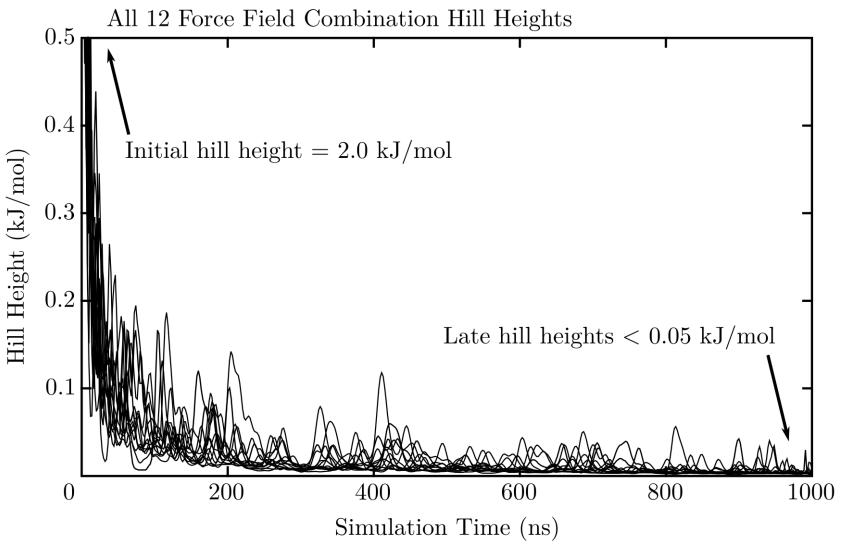


Figure 5.3: Metadynamics hill heights for all 12 simulations plotted over time. All hill height traces are uniformly plotted in black because the intention here is only to show that late hill heights are smaller than 0.05 kJ/mol. Lines are Bezier smoothed to make small peaks clearer. Hill height quickly decays from its initial height of 2.0 kJ/mol.

It is therefore very likely that all relevant energy barriers have been crossed, and all basins have likely been explored. Most hills added at the end of the trajectory are smaller than 0.02 kJ/mol in CV space, meaning that additional bias is not being added at any appreciable rate. Further convergence tests will be needed in the future to more completely assess the reproducibility of these results. Specifically, the relative weight of low free-energy basins should be compared over the simulation.

The structures represented in the low-energy basins presented in Figure 5.2 can be diverse even when their CV coordinates are very similar. Therefore, the trajectories were clustered according to the Cα-rmsd of a selection frames with respect to other frames in the trajectory. Populations of the clusters were assigned by the Torrie-Valleau reweighting algorithm. Representative structures for the four most populated clusters are shown in Figure 5.4. The relative weight of the clusters was determined by summing the weight of the respective frames and dividing by the weight of all frames.

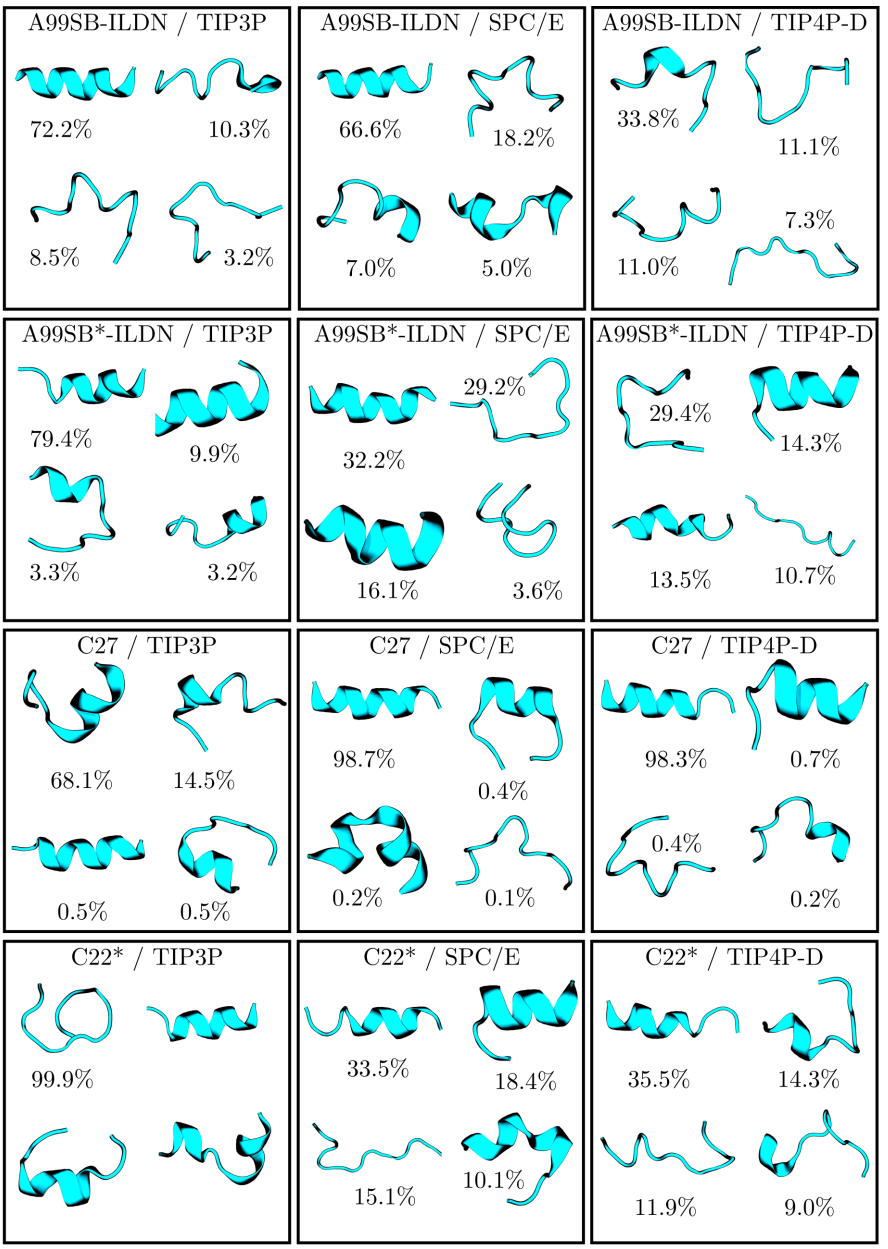


Figure 5.4: Each cell illustrates the central structure of each of the top four clusters ranked by probability. The force field combinations used in each simulation are depicted in the top right corner of each cell.

Inspection of the highly populated clusters reveals a large diversity in the structures sampled across the range of protein and water force fields used in this study. This supports our earlier claim that force fields selection does matter tremendously for problems involving interfaces. It can be seen that all of the force fields have varying degrees of α-helical character, the least of which are A99SB*-ILDN and A99SB-ILDN with TIP4P-D water. As mentioned before, the TIP4P-D water model is meant to reproduce intrinsically disordered structures, and low populations of helical peptides should be expected. Another interesting case is that of C22* with TIP3P water. Problems arising from improper metadynamics sampling caused the weight of a single cluster to dominate all other structures. From an initial visual inspection, it is unclear which degree of freedom is not sampled properly in this simulation nor which additional degree of freedom could be biased to prevent the same outcome. Other methods of general enhanced sampling such as parallel tempering may be beneficial in this case. Additionally, it appears that there is no protein or water force field that consistently produces similar secondary structures no matter which other force field is paired with them.

The experimental VSFG spectrum used for the comparison of simulations and experiment is shown in Figure 5.5. In the experimental VSFG spectrum we observed a non-resonant sum-frequency response on the high frequency side of the spectrum. This non-resonant response renders a comparison of experimental and calculated spectra difficult. To make experimental and calculated VSFG spectra comparable, we fit the experimental spectrum with Equation 1.13, which includes a non-resonant amplitude and phase. Subsequently, the amplitude and the phase of the non-resonant contribution were set to zero while keeping all of the other fit parameters constant.

In addition we have to account for the presence of vibrational modes originating from side chains of aurein 1.2, which are not taken into account by the VSFG calculations. The 2 resonant modes around 1587 and 1610 cm^{-1} can be assigned to the COO^- asymmetric stretching vibration (of glutamic and aspartic acid) and the phenyl ring in-plane vibration modes of phenylalanine.[95] Considering that the VSFG calculation cannot take these vibrational modes into account, the parameters of these modes were set to zero as well. Fit parameters before and after setting the parameters for side chain vibrations and the non-resonant contribution to zero are presented in Section 5.5 - Table 1.

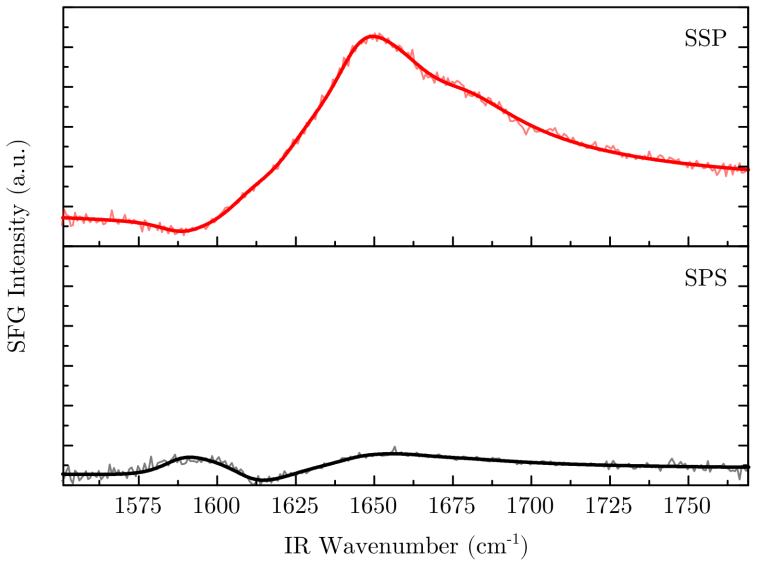


Figure 5.5: Experimental VSFG spectrum of aurein 1.2 (thin black/red line) with polarization combinations indicated in the panels. The result of the fit is superimposed with the experimental data (thick black/red line).

To test which of the force fields best represents reality, an ensemble of peptide structures obtained from the WT-MetaD simulation was used to calculate a theoretical VSFG spectrum. To this end, a sample of 100 frames was drawn from the trajectory. These 100 frames were accumulated by accepting or rejecting frames with a probability proportional to that obtained from reweighting. The weight of a respective frame was obtained by the Torrie-Valleau method. This set of 100 frames was used to generate a file which had each frame arranged in a roughly equally spaced manner within one plane. This file was subsequently used to calculate VSFG spectra using the method of Roeters et al.[78]

The calculated and experimental spectra are presented for each force field in Figure 5.6. The calculated VSFG spectra were compared to the experimental curves by first shifting both curves to zero at a wavenumber of 1550 cm^{-1} and then scaling the intensity of the calculated spectrum by a constant in order to minimize the root mean square difference (rmsd) between the calculated and experimental points. This rmsd value is reported in Table 5.2 for each force field combination.

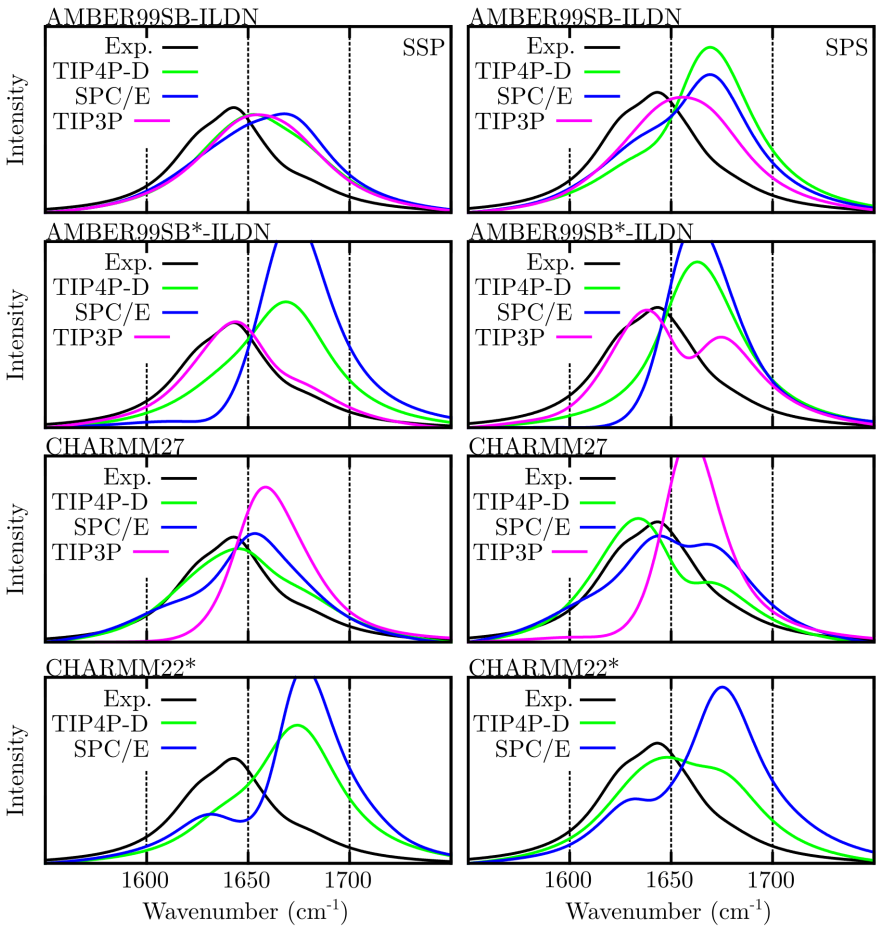


Figure 5.6: Comparison between experimental and calculated VSFG spectra for the ssp (left) and sps polarization (right). Prior to plotting, the experimental spectrum was fit and the non-resonant amplitude and phase were set to zero. The resulting spectrum is depicted as black curve in each graph. For comparison, each graph additionally depicts a scaled version of the calculated VSFG spectrum of one particular protein force field paired with the following water force fields: TIP4P-D (green), SPC/E (blue) and TIP3P (pink). CHARMM22* / TIP3P is excluded because of problems with metadynamics convergence

Table 5.2: Ranking the force fields according to minimum rmsd between calculated and experimental spectra. Scale of rmsd is arbitrary. Overall rank is the sum of rmsd of SSP and SPS data. A low rmsd indicates good agreement between theory and experiment. C22* / TIP3P is excluded because of problems with metadynamics convergence.

Force Fields	Rmsd SSP	Rank	Rmsd SPS	Rank	Overall
A99SB-ILDN / TIP3P	3.2	5	1.8	9	4
A99SB-ILDN / SPC/E	3.8	6	2.6	5	6
A99SB-ILDN / TIP4P-D	3.0	4	3.2	7	5
A99SB*-ILDN / TIP3P	1.0	1	1.6	3	2
A99SB*-ILDN / SPC/E	6.9	11	3.5	8	11
A99SB*-ILDN / TIP4P-D	5.0	7	3.0	10	8
C27 / TIP3P	5.3	8	3.2	11	9
C27 / SPC/E	2.9	3	1.6	4	3
C27 / TIP4P-D	1.5	2	1.0	2	1
C22* / TIP3P	-	-	-	-	-
C22* / SPC/E	6.6	10	3.4	1	10
C22* / TIP4P-D	5.6	9	1.8	6	7

Since, the relative amplitude and location of VSFG peaks are the important features in a spectrum, a simple rescaling of the amplitude by multiplication of a constant is valid. Calculated points were linearly interpolated in order to match the exact discrete wavenumbers at which data were collected. Table 5.2 indicates that the best performing force fields for this model peptide are C27 / TIP4P-D and A99SB*-ILDN / TIP3P with C27 / SPC/E performing slightly less optimally overall. On the other hand, the rest of the force fields perform quite poorly relative to the three top force fields listed. The outstanding performance of the C27 / TIP4P-D force field is surprising. It has been shown that the C27 force field disproportionately stabilizes helical structures. However, this behavior appears to be balanced by an increase in London dispersion interactions of water molecules by using the TIP4P-D water model. In contrast to C27, the A99SB-ILDN force field underestimates the stability of alpha helices. The mediocre performance of force field combinations including A99SB-ILDN is therefore not surprising. The A99SB*-ILDN and the C22* force field are specifically balanced to yield a good representation of the helix-coil transition.[93b] Based upon the performance of these force fields with respect to their solution-state alpha-helix propensity, we speculate that C22* / TIP3P performs well compared to

other force fields. Unfortunately, the problem of proper sampling could not be resolved at the time of writing this thesis, but it will be explored in the near future.

The reproducibility of VSFG calculations was tested by drawing five samples of 100 peptide structures from the ensemble of the A99SB*-ILDN/TIP3P simulation. The spectra calculated from these samples are depicted in Figure 5.7.

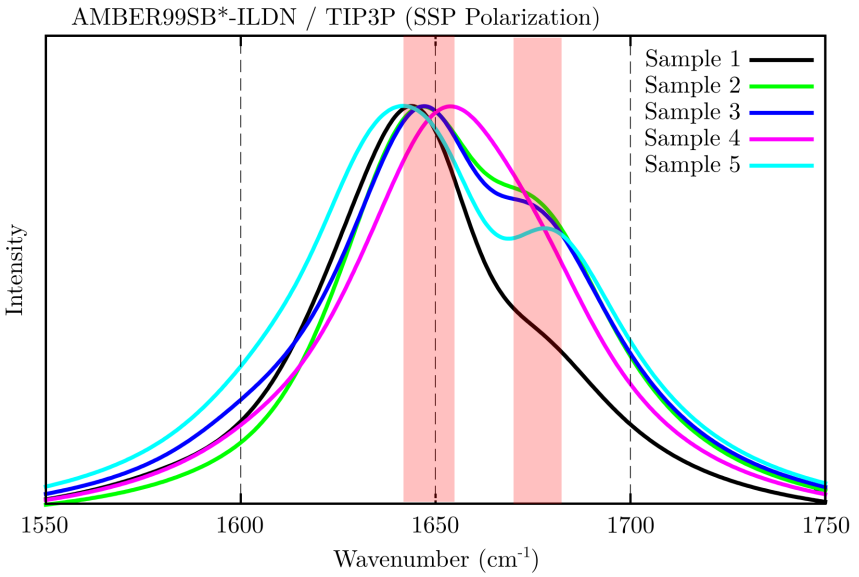


Figure 5.7: Sampling of the ensemble produced by A99SB*-ILDN / TIP3P for evaluation of the reproducibility of calculated spectra. The general range of the two peaks is highlighted in pink.

The superimposed calculated spectra shown in Figure 5.7 provide a good picture of the impact of sampling on the resulting VSFG spectra. Sample 4 does not appear to reproduce both the two major features near 1650 cm^{-1} and 1670 cm^{-1}, which are present in the other samples. The other four samples differ in their relative weights for these two features. Figure 5.7 indicates that either more samples are needed per spectrum or that there is a shortcoming in the methods by which we choose frames. The five rmsd values of these curves, as calculated by the same method as in Table 5.2, average to 2.4 with a standard deviation of 1.0. This average value would make the force field combination the second-best fit for the SSP polarization combination as

opposed to first-best fit that was observed for the single spectrum. Qualitatively, this observation changes little. A99SB*-ILDN/TIP3P is still a very good force field combination relative to the other combinations tested, but it appears that additional samples will be needed for all force field combinations to increase the certainty of our conclusions. The diversity of the spectra from these five samples may be useful for further analysis to determine which structural features give rise to which spectral features. Since small changes in relative free energy (<1 k_BT) can have large effects on the relative population of structures, and since force fields often have errors larger than k_BT, one might expect the relative height of spectral peaks to be skewed. Thus, by having multiple samples of the structural ensemble with different relative peak heights, one can theoretically rebalance the ensemble to reproduce spectra more exactly. This possibility is currently being explored.

5.3 CONCLUSION

These preliminary findings indicate that force field selection is of utmost importance for the simulation of peptides at interfaces. Moreover, enhanced sampling, in the form of WT-MetaD is necessary to overcome free energy barriers to produce a diverse ensemble of interfacial structures. The method of sampling structural ensembles using molecular simulation and comparing those ensembles to spectral data has been demonstrated to be a feasible and useful technique for generating hypothetical molecular-level structures for interfacial peptides. The combination of molecular simulation and VSFG can thus be symbiotic and helpful for researchers in determining specific, detailed peptide structures.

The generality of these findings must be verified with other peptides and eventually proteins as well. It is possible that our findings concerning one peptide having one secondary structure feature may not apply across all peptides containing that feature. The other five peptides are currently being examined to determine the reproducibility of these results for other helical and beta-like structures. However, based upon the preliminary findings for the antimicrobial peptide aurein 1.2, we have can provide some suggestions for researchers interested in performing MD simulations or VSFG experiments of helical peptides or proteins at an air-water interface.

(1) WT-MetaD simulations are a valid way to generate structural ensembles for adding atomistic detail to VSFG and other experimental data

at the air-water interface. This can be used to generate or test hypothetical structures to be taken together with experimental data. Other enhanced-sampling techniques may need to be explored to overcome hidden free energy barriers in systems that become stuck in certain conformations.

(2) In the case of aurein 1.2, the C27 / TIP4P-D and A99SB*-ILDN / TIP3P force field combinations produce structural ensembles whose simulated spectra best match experimental data. Thus we suggest using these force fields in future simulations of helical peptides at the air-water interface.

(3) Clustering analysis shows that a peptide's interfacial structural ensemble can be diverse. Therefore, we suggest using some enhanced sampling such as metadynamics or replica exchange in order to generate realistic ensembles.

(4) The secondary structure of some peptides is greatly affected by the air-water interface. This should be considered in industrial or medical applications in which air bubbles might be introduced into a solution such as in fermentation, cell cultures, protein purification, and protein storage.

These findings are based upon the assumption that the peptides are largely independent of each other within experiments as we know they are in the simulation box. The introduction of high concentrations of peptides at the air-water interface would likely have the effect of causing the peptides to aggregate and move into solution because of strong hydrophobic interactions, as is found in LKα14 experiments. In the future, this assumption might be tested by generating structural ensembles from other methods such as parallel bias metadynamics[96], solvent tempering[91b], or parallel tempering[90] with guidance from the protocols presented in this chapter.

5.4 Specific Experimental Details

Vibrational sum-frequency generation spectroscopy: The peptide structure at the air-water interface was probed with VSFG spectroscopy. A detailed description of the fundamentals of VSFG and the laser setup can be found in Section 1.3. The sample preparation is described in the following. A Teflon trough was filled with a solution of NaCl in D_2O (40 ml, 150 mM NaCl) and placed on a motorized rotation stage. The surface tension was recorded with a Kibron DeltaPi tensiometer (Helsinki, Finland), calibrated on the NaCl solution. While measuring the surface tension, a peptide solution (1 ml, 1.4 mM aurein 1.2 in D_2O, 150 mM NaCl) was injected through the water

surface of the NaCl solution. The peptide layer adsorbed to the air-water interface was assumed to be equilibrated when the surface tension did not change significantly for half an hour. After the equilibration, VSFG spectra were collected. All data was acquired for 60 min at 22°C in the range of 1550-1750 cm^{-1}. The raw data was processed according to the procedure outlined in Section 1.3.2. The spectra were fit according to Equation 1.13.

Molecular Dynamics Simulations: In order to elucidate the effect of different atomistic force fields on the conformation of peptides at the air-water interface we used molecular dynamics simulations. Simulation boxes were constructed by placing the peptide with a random spatial orientation at the edge of a water box. These water boxes were cubic with a side length of 2.4 nm longer than the longest axis of the peptide, and 150 mM sodium chloride was added along with any additional ions needed to neutralize the system. Then, the dimension of the box normal to the peptide-containing interface was expanded by a factor of three in order to introduce a vacuum, which is used as a proxy for air. Boxes are simulated with periodic boundary conditions in order to allow for the application of the particle mesh Ewald sum method for long-range electrostatic calculations.[40] Such a simulation setup produces infinite slabs of water separated by several nanometers in space, which is large enough for the electrostatic interactions among the slabs to largely decay. GROMACS 4.6 patched with PLUMED 2.1 was used for all MD simulations.[73] Bonds were constrained using the LINCS algorithm to allow for a numerical integration time step of 2 fs.[97] The temperature was held at 300 K using a stochastic velocity-rescaling thermostat.[64] Lennard-Jones and electrostatic interactions were cut-off at 1.0 nm with a Verlet integration scheme, and long-range electrostatics were handled with the particle mesh Ewald summation method. Steepest descent minimization of 1000 steps was used to relax the system before subsequent production simulations in the canonical ensemble. WT-MetaD simulations were 1 μs in length.

Metadynamics simulations were facilitated by the use of PLUMED 2.1. An introduction to the enhanced sampling methods of metadynamics and in particular WT-MetaD is given in Section 1.4.2.[98] In short, the bias potential V, given by Equation 1.18, is constructed during the simulation by accumulating small Gaussian hills of potential with a width of σ that are deposited periodically to slow degrees of freedom known as collective variables (CVs) denoted by the letter S. The height of the Gaussian kernel is initially W but in WT-MetaD subsequent hills decrease in height according to Equation 1.20. The decrease in hill height is a function of the magnitude of

the bias previously added and a temperature parameter ΔT (not to be confused with the temperature of the simulation T) which controls the exponential decay of the hill height. Thus, the system is lifted out of free-energy minima, and eventually the full conformational landscape of these CVs can be sampled. The parameters for the size of the hills are W = 2.0 kJ/mol and σ = 0.03 and 0.01 (for radius of gyration and hydrogen bonds respectively) in CV space to be applied every picosecond, and ΔT = 2700 K. For these simulations, we selected the alpha carbon (Cα) radius of gyration and the numbers of structural hydrogen bonds as our two CVs. The radius of gyration

$$g = \left(\frac{\sum_i^n m_i |r_i - r_{COM}|^2}{\sum_i^n m_i}\right)^{1/2}, \quad \text{with } r_{COM} = \frac{\sum_i^n m_i r_i}{\sum_i^n m_i}. \tag{5.1}$$

m_i is the mass and r_i is the position of atom i. r_{COM} is the position of the center of mass.

Structural hydrogen bonds are those found in the experimental solution NMR structure of aurein 1.2 plus any other possible hydrogen bonds that would arise in a perfect alpha helix. The presence of a hydrogen bond was determined by applying a sigmoidal function s_{ij} to the distance between the hydrogen and oxygen in question.

$$s_{ij} = \frac{1-\left(\frac{r_{ij}}{r_0}\right)^n}{1-\left(\frac{r_{ij}}{r_0}\right)^m}. \tag{5.2}$$

s_{ij} decays from one to zero as the distance between the two atoms i and j grows. We used the constants m = 8, n = 6, and r_0 = 0.25 nm. The CV biased during the simulations was the sum of s_{ij} over all defined hydrogen bonds. From the bias applied to these CVs during a WT-MetaD simulation, the free-energy landscape can be computed as in Equation 1.21.

Clustering: In order to identify the most probable structures from a WT-MetaD simulation a clustering algorithm was employed. The trajectory was clustered using a sample of 20,000 of the 1,000,000 frames from the simulation with the g_cluster tool within GROMACS 4.6 and the gromos method.[99] This smaller sample was used, because the calculation time and memory required to construct the necessary root mean square displacement matrix among the frames scales roughly with the number of frames squared. Thus, clustering of the whole trajectory would be around 2500 times more costly than the reduced set. The cut-off for cluster members in rmsd space

was set to 0.3 nm for the Cα-atoms with a goal of keeping the number of clusters near ten. The remaining 980,000 frames were then compared to the central member of the top ten existing clusters, and assigned to the first cluster within the cut-off. If no cluster was within the cut-off, the frame was assigned to a separate "junk" cluster that was not considered in the subsequent clustering analysis. Using the Torrie-Valleau method,[100] weights were assigned to frames in the trajectory based upon the metadynamics bias applied during a given frame. Frames were taken after the transient period where the majority of the metadynamics bias was applied. Each cluster was assigned a weight by the sum of the weights of these frames.

Simulating VSFG spectra: For the purpose of comparing simulations with experimental results we simulated VSFG spectra from representative conformational ensembles of peptides. Simulation frames were selected from the whole ensemble of conformations with a probability proportional to the frame weight, assigned by the Torrie-Valleau method, these random frames were included in a sample of 100 frames in total. This sample of 100 peptide structures was arranged in an array in a single pdb file. The pdb file was used to compute a VSFG spectrum using the method of Roeters et al..[78] This method calculates the second-order non-linear susceptibility $\vec{\vec{\chi}}^{(2)}$ which was introduced in Section 1.3.1. The calculation includes nearest- and non-nearest neighbor coupling and hydrogen bond effects. Nearest-neighbor coupling is included in the calculation by a map of the dihedral angle dependent coupling. The coupling was obtained by *ab initio* calculations at the 6-31G+(d) B3LYP level of theory.[101] Non-nearest neighbor couplings were estimated using transition-dipole interactions.[102]

5.5 Additional Data

Table 1: (a) Fit parameters before setting the non-resonant amplitude and phase and the low frequency vibrational modes (w1 and w2) to zero. (b) Fit parameters used for the comparison to theoretical VSFG spectra.

	(a) SSP	**(b) SSP**	**(a) SPS**	**(b) SPS**
NR	0.3	0	0.1	0
phi	2.7	0	2.6	0
A1	-0.5	0	0.8	0
w1	1586.7	1586.7	1586.7	1586.7
G1	11.1	0	11.1	0
A2	0.6	0	-1.0	0
w2	1609.5	1609.5	1609.5	1609.5
G2	11.9	0	11.9	0
A3	2.7	2.7	0.8	0.8
w3	1627.4	1627.4	1627.4	1627.4
G3	18.1	18.1	18.1	18.1
A4	2.9	2.9	0.7	0.7
w4	1643.7	1643.8	1643.8	1643.8
G4	14.9	14.9	14.9	14.9
A5	0.9	0.9	0.4	0.4
w5	1656.4	1656.4	1656.4	1656.4
G5	15.3	15.3	15.3	15.3
A6	1.0	1.0	0.2	0.2
w6	1675.9	1675.9	1675.9	1675.9
G6	18.1	18.1	18.1	18.1

Bibliography

[1] U. G. K. Wegst, H. Bai, E. Saiz, A. P. Tomsia, R. O. Ritchie, *Nat. Mater.* **2015**, *14*, 23-36.

[2] C. E. Hamm, R. Merkel, O. Springer, P. Jurkojc, C. Maier, K. Prechtel, V. Smetacek, *Nature* **2003**, *421*, 841-843.

[3] G. R. Halse, E. E. Syvertsen, in *Identifying Marine Diatoms and Dinoflagellates*, Academic Press, San Diego, **1996**, pp. 5-385.

[4] E. M. Valliant, J. R. Jones, *Soft Matter* **2011**, *7*, 5083-5095.

[5] B. Fadeel, A. E. Garcia-Bennett, *Adv. Drug Del. Rev.* **2010**, *62*, 362-374.

[6] H. C. Schroder, X. Wang, W. Tremel, H. Ushijima, W. E. G. Muller, *Natural Product Reports* **2008**, *25*, 455-474.

[7] E. V. Armbrust, *Nature* **2009**, *459*, 185-192.

[8] S. Hazelaar, H. J. Van Der Strate, W. W. C. Gieskes, E. G. Vrieling, *J. Phycol.* **2005**, *41*, 354-358.

[9] a) R. K. Iler, *The Chemistry of Silica: Solubility, Polymerization, Colloid and Surface Properties and Biochemistry of Silica*, Wiley-Interscience, **1979**; b) M. A. Borowitzka, B. E. Volcani, *J. Phycol.* **1978**, *14*, 10-21; c) A.-M. M. Schmid, D. Schulz, *Protoplasma* **1979**, *100*, 267-288.

[10] C. Lechner, C. Becker, *Mar. Drugs* **2015**, *13*, 5297.

[11] N. Kröger, R. Deutzmann, C. Bergsdorf, M. Sumper, *Proceedings of the National Academy of Sciences* **2000**, *97*, 14133-14138.

[12] a) M. Sumper, E. Brunner, G. Lehmann, *FEBS Lett.* **2005**, *579*, 3765-3769; b) M. Sumper, G. Lehmann, *ChemBioChem* **2006**, *7*, 1419-1427; c) M. Sumper, E. Brunner, *Adv. Funct. Mater.* **2006**, *16*, 17-26.

[13] a) K. M. Delak, N. Sahai, *The Journal of Physical Chemistry B* **2006**, *110*, 17819-17829; b) T. N. Mizutani, H.; Fujiwara, N.; Ogoshi, H., *Bull. Chem. Soc. Jpn.* **1998**, *71*, 2017-2022; c) T. Coradin, J. Livage, *Colloids Surf. B. Biointerfaces* **2001**, *21*, 329-336.

[14] M. Sumper, *Science* **2002**, *295*, 2430-2433.

[15] M. Sumper, *Angew. Chem. Int. Ed.* **2004**, *43*, 2251-2254.

[16] R.-H. Jin, D.-D. Yao, R. Levi, *Chemistry – A European Journal* **2014**, *20*, 7196-7214.
[17] N. Kroger, N. Poulsen, *Annu. Rev. Genet.* **2008**, *42*, 83-107.
[18] N. Kröger, S. Lorenz, E. Brunner, M. Sumper, *Science* **2002**, *298*, 584-586.
[19] N. Kröger, R. Deutzmann, M. Sumper, *Science* **1999**, *286*, 1129-1132.
[20] R. R. Naik, P. W. Whitlock, F. Rodriguez, L. L. Brott, D. D. Glawe, S. J. Clarson, M. O. Stone, *Chem. Commun.* **2003**, 238-239.
[21] C. C. Lechner, C. F. W. Becker, *J. Pept. Sci.* **2014**, *20*, 152-158.
[22] A. Roehrich, G. Drobny, *Acc. Chem. Res.* **2013**, *46*, 2136-2136.
[23] W. Demtröder, *Experimentalphysik 2, Elektrizität und Optik*, 4 ed., Springer Berlin Heidelberg New York, **2006**, 214.
[24] R. W. Boyd, in *Nonlinear Optics (Third Edition)*, Academic Press, Burlington, **2008**, pp. 1-67.
[25] A. G. Lambert, P. B. Davies, D. J. Neivandt, *Applied Spectroscopy Reviews* **2005**, *40*, 103-145.
[26] G. J. Simpson, *ChemPhysChem* **2004**, *5*, 1301-1310.
[27] R. W. Boyd, in *Nonlinear Optics (Third Edition)*, Academic Press, Burlington, **2008**, pp. 135-206.
[28] Y. R. Shen, Wiley-Interscience, **2003**.
[29] E. L. D. van der Spoel, B. Hess, and the GROMACS development team, *GROMACS User Manual version 4.6.5*, www.gromacs.org, **2013**.
[30] A. R. Leach, *Molecular Modelling Principles and Applications*, 2nd ed., Pearson Education Limited, **1996**, 353-406.
[31] a) A. D. MacKerell, D. Bashford, M. Bellott, R. L. Dunbrack, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiórkiewicz-Kuczera, D. Yin, M. Karplus, *J. Phys. Chem. B* **1998**, *102*, 3586-3616; b) K. Lindorff-Larsen, S. Piana, K. Palmo, P. Maragakis, J. L. Klepeis, R. O. Dror, D. E. Shaw, *Proteins: Structure, Function, and Bioinformatics* **2010**, *78*, 1950-1958; c) W. L. Jorgensen, J. Tirado-Rives, *J. Am. Chem. Soc.* **1988**, *110*, 1657-1666.
[32] A. D. Mackerell, *J. Comput. Chem.* **2004**, *25*, 1584-1604.
[33] K. A. Beauchamp, Y.-S. Lin, R. Das, V. S. Pande, *J Chem Theory Comput* **2012**, *8*, 1409-1414.
[34] C. Abrams, G. Bussi, *Entropy* **2014**, *16*.

[35] a) U. H. E. Hansmann, *Chem. Phys. Lett.* **1997**, *281*, 140-150; b) Y. Sugita, Y. Okamoto, *Chem. Phys. Lett.* **1999**, *314*, 141-151; c) A. Barducci, M. Bonomi, M. Parrinello, *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2011**, *1*, 826-843.

[36] A. Barducci, G. Bussi, M. Parrinello, *Phys. Rev. Lett.* **2008**, *100*, 020603.

[37] D. Sindhikara, Y. L. Meng, A. E. Roitberg, *J. Chem. Phys.* **2008**, *128*.

[38] N. M. Laurendeau, *Statistical Thermodynamics: Fundamentals and Applications*, Cambridge University Press, Cambridge, **2005**, 347.

[39] M. Bonomi, M. Parrinello, *Phys. Rev. Lett.* **2010**, *104*, 190601.

[40] U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, L. G. Pedersen, *The Journal of Chemical Physics* **1995**, *103*, 8577-8593.

[41] a) D. Y. Zhao, J. L. Feng, Q. S. Huo, N. Melosh, G. H. Fredrickson, B. F. Chmelka, G. D. Stucky, *Science* **1998**, *279*, 548-552; b) A. Corma, *Chem. Rev.* **1997**, *97*, 2373-2420.

[42] S. Mann, G. A. Ozin, *Nature* **1996**, *382*, 313-318.

[43] L. Senior, M. P. Crump, C. Williams, P. J. Booth, S. Mann, A. W. Perriman, P. Curnow, *J Mater Chem B* **2015**, *3*, 2607-2614.

[44] a) J. E. Baio, A. Zane, V. Jaeger, A. M. Roehrich, H. Lutz, J. Pfaendtner, G. P. Drobny, T. Weidner, *J. Am. Chem. Soc.* **2014**, *136*, 15134-15137; b) F. Rodriguez, D. D. Glawe, R. R. Naik, K. P. Hallinan, M. O. Stone, *Biomacromolecules* **2004**, *5*, 261-265; c) M. B. Dickerson, K. H. Sandhage, R. R. Naik, *Chem. Rev.* **2008**, *108*, 4935-4978.

[45] W. F. Degrado, J. D. Lear, *J. Am. Chem. Soc.* **1985**, *107*, 7684-7689.

[46] A. C. Zane, C. Michelet, A. Roehrich, P. S. Emani, G. P. Drobny, *Langmuir* **2014**, *30*, 7152-7161.

[47] a) T. Weidner, N. F. Breen, K. Li, G. P. Drobny, D. G. Castner, *P Natl Acad Sci USA* **2010**, *107*, 13288-13293; b) T. Weidner, J. S. Apte, L. J. Gamble, D. G. Castner, *Langmuir* **2010**, *26*, 3433-3440; c) T. Weidner, N. T. Samuel, K. McCrea, L. J. Gamble, R. S. Ward, D. G. Castner, *Biointerphases* **2010**, *5*, 9-16; d) O. Mermut, D. C. Phillips, R. L. York, K. R. McCrea, R. S. Ward, G. A. Somorjai, *J. Am. Chem. Soc.* **2006**, *128*, 3598-3607; e) L. Fu, J. Liu, E. C. Yan, *J. Am. Chem. Soc.* **2011**, *133*, 8094-8097; f) M. Deighan, J. Pfaendtner, *Langmuir* **2013**, *29*, 7999-8009.

[48] a) H. Yang, N. Coombs, I. Sokolov, G. A. Ozin, *Nature* **1996**, *381*, 589-592; b) S. J. Wang, X. Ge, J. Y. Xue, H. M. Fan, L.

J. Mu, Y. P. Li, H. Xu, J. R. Lu, *Chem. Mater.* **2011**, *23*, 2466-2474; c) J. E. Meegan, A. Aggeli, N. Boden, R. Brydson, A. P. Brown, L. Carrick, A. R. Brough, A. Hussain, R. J. Ansell, *Adv. Funct. Mater.* **2004**, *14*, 31-37; d) B. Liu, Y. Y. Cao, Z. H. Huang, Y. Y. Duan, S. N. Che, *Adv. Mater.* **2015**, *27*, 479-497.

[49] N. Ji, Y. R. Shen, *J. Chem. Phys.* **2004**, *120*, 7107-7112.

[50] Q. Du, E. Freysz, Y. R. Shen, *Science* **1994**, *264*, 826-828.

[51] R. L. York, O. Mermut, D. C. Phillips, K. R. McCrea, R. S. Ward, G. A. Somorjai, *Journal of Physical Chemistry C* **2007**, *111*, 8866-8871.

[52] T. Weidner, N. F. Breen, G. P. Drobny, D. G. Castner, *J. Phys. Chem. B* **2009**, *113*, 15423-15426.

[53] a) A. Barth, C. Zscherp, *Q. Rev. Biophys.* **2002**, *35*, 369-430; b) T. Weidner, D. G. Castner, *PCCP* **2013**, *15*, 12516-12524.

[54] a) K. T. Nguyen, J. T. King, Z. Chen, *J. Phys. Chem. B* **2010**, *114*, 8291-8300; b) B. R. Singh, *Infrared Analysis of Peptides and Proteins, Vol. 750*, **2000**, 54-95.

[55] a) Y. N. Chirgadze, N. A. Nevskaya, *Biopolymers* **1976**, *15*, 607-625; b) Y. N. Chirgadze, N. A. Nevskaya, *Biopolymers* **1976**, *15*, 627-636.

[56] A. Depla, E. Verheyen, A. Veyfeyken, M. Van Houteghem, K. Houthoofd, V. Van Speybroeck, M. Waroquier, C. E. A. Kirschhock, J. A. Martens, *Journal of Physical Chemistry C* **2011**, *115*, 11077-11088.

[57] J. M. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, D. A. Case, *J. Comput. Chem.* **2004**, *25*, 1157-1174.

[58] a) F. S. Emami, V. Puddu, R. J. Berry, V. Varshney, S. V. Patwardhan, C. C. Perry, H. Heinz, *Chem. Mater.* **2014**, *26*, 2647-2658; b) S. V. Patwardhan, F. S. Emami, R. J. Berry, S. E. Jones, R. R. Naik, O. Deschaume, H. Heinz, C. C. Perry, *J. Am. Chem. Soc.* **2012**, *134*, 6244-6256.

[59] a) M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R.

Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, Ö. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, D. J. Fox, *Gaussian 09, Revision B.01*, Gaussian, Inc., Wallingford, CT, **2009**; b) C. I. Bayly, P. Cieplak, W. D. Cornell, P. A. Kollman, *J. Phys. Chem.* **1993**, *97*, 10269-10280; c) J. M. Wang, W. Wang, P. A. Kollman, D. A. Case, *Journal of Molecular Graphics & Modelling* **2006**, *25*, 247-260; d) D. A. Case, V. Babin, J. T. Berryman, R. M. Betz, Q. Cai, D. S. Cerutti, T. E. Cheatham III, T. A. Darden, R. E. Duke, H. Gohlke, A. W. Goetz, S. Gusarov, N. Homeyer, P. Janowski, J. Kaus, I. Kolossváry, A. Kovalenko, T. S. Lee, S. LeGrand, T. Luchko, R. Luo, B. Madej, K. M. Merz, F. Paesani, D. R. Roe, A. Roitberg, C. Sagui, R. Salomon-Ferrer, G. Seabra, C. L. Simmerling, W. Smith, J. Swails, R. C. Walker, J. Wang, R. M. Wolf, X. Wu, P. A. Kollman, *AMBER 14*, University of California, San Francisco, **2014**.

[60] A. Sousa da Silva, W. Vranken, *BMC Research Notes* **2012**, *5*, 367.

[61] B. Hess, C. Kutzner, D. van der Spoel, E. Lindahl, *J Chem Theory Comput* **2008**, *4*, 435-447.

[62] L. Martinez, R. Andrade, E. G. Birgin, J. M. Martinez, *J. Comput. Chem.* **2009**, *30*, 2157-2164.

[63] T. Darden, D. York, L. Pedersen, *J. Chem. Phys.* **1993**, *98*, 10089-10092.

[64] G. Bussi, D. Donadio, M. Parrinello, *J. Chem. Phys.* **2007**, *126*, 014101

[65] H. Lutz, V. Jaeger, R. Berger, M. Bonn, J. Pfaendtner, T. Weidner, *Adv Mater Interfaces* **2015**, *2*, 1500282.

[66] T. Arnesen, *PLoS Biol.* **2011**, *9*, e1001074.

[67] a) A. O. Helbig, S. Rosati, P. W. W. M. Pijnappel, B. van Breukelen, M. H. T. H. Timmers, S. Mohammed, M. Slijper, A. J. R. Heck, *BMC Genomics* **2010**, *11*, 685; b) H. Jornvall, *J. Theor. Biol.* **1975**, *55*, 1-12; c) B. Persson, C. Flinta, G. Vonheijne, H. Jornvall, *Eur. J. Biochem.* **1985**, *152*, 523-527; d) C. S. Hwang, A. Shemorry, A. Varshavsky, *Science* **2010**, *327*, 973-977; e) A. Varshavsky, *J. Biol. Chem.* **2008**, *283*, 34469-34489.

[68] X. B. Zhao, F. Pan, H. Xu, M. Yaseen, H. H. Shan, C. A. E. Hauser, S. G. Zhang, J. R. Lu, *Chem. Soc. Rev.* **2010**, *39*, 3480-3498.

[69] C. Dalgicdir, M. Sayar, *J. Phys. Chem. B* **2015**, *119*, 15164-15175.

[70] a) D. H. de Jong, G. Singh, W. F. D. Bennett, C. Arnarez, T. A. Wassenaar, L. V. Schafer, X. Periole, D. P. Tieleman, S. J. Marrink, *J Chem Theory Comput* **2013**, *9*, 687-697; b) S. O. Yesylevskyy, L. V. Schafer, D. Sengupta, S. J. Marrink, *PLoS Comp. Biol.* **2010**, *6*, e1000810.

[71] D. H. de Jong, S. Baoukina, H. I. Ingolfsson, S. J. Marrink, *Comput. Phys. Commun.* **2016**, *199*, 1-7.

[72] W. J. M. FrederixPim, G. G. Scott, Y. M. Abul-Haija, D. Kalafatovic, C. G. Pappas, N. Javid, N. T. Hunt, R. V. Ulijn, T. Tuttle, *Nat Chem* **2015**, *7*, 30-37.

[73] a) S. Pronk, S. Pall, R. Schulz, P. Larsson, P. Bjelkmar, R. Apostolov, M. R. Shirts, J. C. Smith, P. M. Kasson, D. van der Spoel, B. Hess, E. Lindahl, *Bioinformatics* **2013**, *29*, 845-854; b) G. A. Tribello, M. Bonomi, D. Branduardi, C. Camilloni, G. Bussi, *Comput. Phys. Commun.* **2014**, *185*, 604-613.

[74] T. A. Wassenaar, K. Pluhackova, R. A. Bockmann, S. J. Marrink, D. P. Tieleman, *J Chem Theory Comput* **2014**, *10*, 676-690.

[75] a) R. B. Best, X. Zhu, J. Shim, P. E. M. Lopes, J. Mittal, M. Feig, A. D. MacKerell, *J Chem Theory Comput* **2012**, *8*, 3257-3273; b) W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, M. L. Klein, *The Journal of Chemical Physics* **1983**, *79*, 926-935.

[76] a) A. J. Mieszawska, L. D. Nadkarni, C. C. Perry, D. L. Kaplan, *Chem. Mater.* **2010**, *22*, 5780-5785; b) K. E. Cole, A. N. Ortiz, M. A. Schoonen, A. M. Valentine, *Chem. Mater.* **2006**, *18*, 4592-4599; c) C. W. P. Foo, S. V. Patwardhan, D. J. Belton, B. Kitchel, D. Anastasiades, J. Huang, R. R. Naik, C. C. Perry, D. L. Kaplan, *P Natl Acad Sci USA* **2006**, *103*, 9428-9433; d) S. L. Sewell, D. W. Wright, *Chem. Mater.* **2006**, *18*, 3108-3113; e) W. D. Marner, A. S. Shaikh, S. J. Muller, J. D. Keasling, *Biotechnol. Progr.* **2009**, *25*, 417-423; f) C. C. Lechner, C. F. W. Becker, *Biomater Sci-Uk* **2015**, *3*, 288-297; g) T. Martelli, E. Ravera, A. Louka, L. Cerofolini, M. Hafner, M. Fragai, C. F. W. Becker, C. Luchinat, *Chem-Eur J* **2016**, *22*, 425-432; h) R. Plowright, N. Dinjaski, S. Zhou, D. J. Belton, D. L. Kaplan, C. C. Perry, *Rsc Adv* **2016**, *6*, 21776-21788.

[77] E. Goormaghtigh, V. Cabiaux, J. M. Ruysschaert, *Sub-cellular biochemistry* **1994**, *23*, 405-450.

[78] S. J. Roeters, C. N. van Dijk, A. Torres-Knoop, E. H. G. Backus, R. K. Campen, M. Bonn, S. Woutersen, *The Journal of Physical Chemistry A* **2013**, *117*, 6311-6322.

[79] D. E. Shaw, P. Maragakis, K. Lindorff-Larsen, S. Piana, R. O. Dror, M. P. Eastwood, J. A. Bank, J. M. Jumper, J. K. Salmon, Y. Shan, W. Wriggers, *Science* **2010**, *330*, 341-346.

[80] M. R. Knecht, D. W. Wright, *Chem. Commun.* **2003**, 3038-3039.

[81] a) B. Hess, *J Chem Theory Comput* **2008**, *4*, 116-122; b) B. Hess, H. Bekker, H. J. C. Berendsen, J. G. E. M. Fraaije, *J. Comput. Chem.* **1997**, *18*, 1463-1472.

[82] *RCSB Protein Data Bank* **Retrieved 03.04.2017**, http://pdbbeta.rcsb.org/pdb/statistics/holdings.do.

[83] a) S. Mauri, T. Weidner, H. Arnolds, *PCCP* **2014**, *16*, 26722-26724; b) J. M. Rodríguez Patino, C. Carrera Sánchez, M. R. Rodríguez Niño, *Adv. Colloid Interface Sci.* **2008**, *140*, 95-113; c) K. Meister, A. Bäumer, G. R. Szilvay, A. Paananen, H. J. Bakker, *The Journal of Physical Chemistry Letters* **2016**, *7*, 4067-4071.

[84] R. Hennig, J. Heidrich, M. Saur, L. Schmüser, S. J. Roeters, N. Hellmann, S. Woutersen, M. Bonn, T. Weidner, J. Markl, D. Schneider, *Nature Communications* **2015**, *6*, 7018.

[85] a) D. Schach, C. Globisch, S. J. Roeters, S. Woutersen, A. Fuchs, C. K. Weiss, E. H. G. Backus, K. Landfester, M. Bonn, C. Peter, T. Weidner, *J. Chem. Phys.* **2014**, *141*; b) J. S. Apte, G. Collier, R. A. Latour, L. J. Gamble, D. G. Castner, *Langmuir* **2010**, *26*, 3423-3432.

[86] W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, P. A. Kollman, *J. Am. Chem. Soc.* **1996**, *118*, 2309-2309.

[87] M. R. Shirts, V. S. Pande, *J. Chem. Phys.* **2005**, *122*.

[88] Y. Nagata, T. Ohto, E. H. G. Backus, M. Bonn, *J. Phys. Chem. B* **2016**, *120*, 3785-3796.

[89] P. Bjelkmar, P. Larsson, M. A. Cuendet, B. Hess, E. Lindahl, *J Chem Theory Comput* **2010**, *6*, 459-466.

[90] R. M. Neal, *Stat Comput* **1996**, *6*, 353-366.

[91] a) L. B. Wright, T. R. Walsh, *PCCP* **2013**, *15*, 4715-4726; b) P. Liu, B. Kim, R. A. Friesner, B. J. Berne, *P Natl Acad Sci USA* **2005**, *102*, 13749-13754.

[92] a) K. G. Sprenger, Y. He, J. Pfaendtner, *Molec Model Simul* **2016**, 21-35; b) K. G. Sprenger, J. Pfaendtner, *Langmuir* **2016**, *32*, 5690-5701.

[93] a) R. B. Best, G. Hummer, *J. Phys. Chem. B* **2009**, *113*, 9004-9015; b) S. Piana, K. Lindorff-Larsen, D. E. Shaw, *Biophys. J.* **2011**, *100*, L47-L49; c) S. Piana, A. G. Donchev, P. Robustelli, D. E. Shaw, *J. Phys. Chem. B* **2015**, *119*, 5113-5123; d) H. J.

C. Berendsen, J. R. Grigera, T. P. Straatsma, *J. Phys. Chem.* **1987**, *91*, 6269-6271.
[94] G. S. Wang, Y. F. Li, X. Li, *J. Biol. Chem.* **2005**, *280*, 5803-5811.
[95] a) A. Barth, *Biochim. Biophys. Acta* **2007**, *1767*, 1073-1101; b) S. Y. Venyaminov, N. N. Kalnin, *Biopolymers* **1990**, *30*, 1243-1257; c) B. Sjöberg, S. Foley, B. Cardey, M. Enescu, *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy* **2014**, *128*, 300-311.
[96] J. Pfaendtner, M. Bonomi, *J Chem Theory Comput* **2015**, *11*, 5062-5067.
[97] B. Hess, *J Chem Theory Comput* **2008**, *4*, 116-122.
[98] A. Laio, M. Parrinello, *P Natl Acad Sci USA* **2002**, *99*, 12562-12566.
[99] X. Daura, K. Gademann, B. Jaun, D. Seebach, W. F. van Gunsteren, A. E. Mark, *Angew Chem Int Edit* **1999**, *38*, 236-240.
[100] G. M. Torrie, J. P. Valleau, *J Comput Phys* **1977**, *23*, 187-199.
[101] R. D. Gorbunov, D. S. Kosov, G. Stock, *The Journal of Chemical Physics* **2005**, *122*, 224904.
[102] a) S. Krimm, Y. Abe, *Proceedings of the National Academy of Sciences* **1972**, *69*, 2788-2792; b) H. Torii, M. Tasumi, *Journal of Raman Spectroscopy* **1998**, *29*, 81-86.

Summary

Biominerals are fascinating composites of organic and inorganic matter that have evolved over hundreds of millions of years. With their intricate nano- to microscale architecture, biominerals display extraordinary properties in terms of toughness, strength and weight. These properties are found for example in the silica skeletons of diatoms. The formation of silica in diatoms has been well-studied on the macroscopic scale. However, the implementation of biomimetic silica formation in technological applications requires a molecular-level understanding of how silica morphology is controlled by organic molecules. Peptides are one class of organic molecules that have been used extensively to mimic biomolecules involved in diatom biosilicification. The aim of this thesis – as defined in the introduction – was to expand knowledge of how peptide structure and ordering at interfaces influences the morphology of artificially generated silica. In these concluding remarks, I will summarize what we have discovered about peptide controlled silica mineralization.

First, the relationship between peptide ordering and silica morphology in solution was studied using simplified model peptides (LK peptides), before moving on to complex peptides and interfacial systems in later studies. LK peptides are composed of hydrophobic and hydrophilic building blocks, typically arranged in a characteristic sequence. It has been hypothesized that differences in silica morphologies generated by different model peptides in solution can be attributed to different agglomeration behavior resulting from the peptide's characteristic hydrophobic/hydrophilic sequence. In this thesis I show that the morphology of silica is not only influenced by the overall hydrophobic/hydrophilic pattern of the peptide sequence but also by small, almost inconspicuous changes of the system e.g. the presence or absence of peptide acetylation. Surprisingly, the acetylated and non-acetylated variants of an LK peptide generate completely different silica morphologies, even though their solution structure appears to be the same. Interface-specific vibrational spectroscopy in combination with coarse grained molecular simulations indicate that the hydrophobic core of peptide agglomerates is

stabilized by the absence of N-terminal charge in acetylated peptides. This result shows that peptide agglomeration plays a critical role in the formation of silica nanostructures but is highly sensitive to the smallest of changes in the system. Further insights into the agglomeration of biomineralization active peptides will be highly valuable for large-scale biotechnological silica mineralization in solution. Methods of solution and solid-state nuclear magnetic resonance could provide a molecular picture of peptide agglomerates in solution or within silica precipitates.

LK peptides were initially designed to adopt classical structural elements of proteins at polar/nonpolar interfaces. Therefore, the effect of different peptide structures on the interfacial mineralization of silica was studied with LK peptides adsorbed to the air-water interface. In Chapter 2, I show that helical and β-strand LK peptides mineralize different silica morphologies at the air-water interface. Molecular simulations provided evidence that helical LK peptides interact with silica predominantly via their side chains, whereas β-strand LK peptides interact via backbone groups. It is unclear whether these findings are generally applicable. However, the experimental approach can be transferred to investigate the interfacial biomineralization of other model peptides and systems such as calcium carbonate and calcium phosphate.

Rather than changing the mineral system, we increased the peptide complexity by investigating the interfacial silica formation mediated by the peptide R5. R5 is derived from a diatom peptide found in the silica cell wall and has relevance for numerous biotechnological applications, such as hybrid materials or mediating protein-silica interactions. During silica mineralization at the air-water interface, we found that R5 remains well-structured. Silica that is mineralized by the R5 peptide at the air-water interface appears to grow by silica intercalation into the interfacially adsorbed peptide agglomerates. Due to its biotechnological relevance, the results presented herein will likely contribute to a better understanding of R5-directed interfacial mineralization of silica.

Each of the studies contained in this thesis rely on molecular dynamics simulations – as do many complementary studies in the field of biomolecular surface science. Molecular simulations are used on a regular basis to aid the interpretation and verification of experimental data obtained from interfaces. In the particular case of peptides and proteins at the air-water interface, the reproduction of accurate structural ensembles has not been explored so far. In the last chapter of this thesis, we evaluate the performance of molecular simulation parameters for peptides at the air-water interface. By comparing

sum frequency generation spectra and calculated spectra from molecular simulations we show that only two parameter sets (out of 12 tested sets) accurately reproduce the conformational ensemble of a given peptide at the air-water interface. These parameter sets are the CHARMM27 protein force field in combination with the TIP4P-D water model and the AMBER99SB*-ILDN protein force field in combination with the TIP3P water model. We plan to expand this approach to other peptides with different structures and eventually to more complex proteins. With additional studies, we hope to gain understanding of how to adjust force field models to accurately represent molecular interactions at the air-water interface.

In conclusion, the molecular-level understanding of peptide-mediated biosilicification we sought at the onset of this project is gradually taking shape. Using advanced nonlinear vibrational spectroscopy in combination with molecular dynamics simulations, we have expanded our ability to determine the conformation of various peptides at interfaces, and we have shown that the silica morphology is indeed dependent on the structure of a given interfacially adsorbed peptide template. We have also considered the effects of peptide agglomeration on silicification. Several aspects of interfacial biosilicification could not be addressed in this thesis but could be interesting to explore further. It appears that the effect of peptide agglomeration on the morphology of mineralized silica is more pronounced than the effect of different peptide structures. An interesting route for research would be to investigate two-dimensional agglomeration of model peptides at interfaces. In addition to the insights about the role of structure and agglomeration behavior on artificial silica structures, we evaluated a model case for enhanced molecular simulations of peptides at the air-water interface. The results of this interdisciplinary approach will provide a guideline for future studies and hopefully inspire the development of better simulation protocols and force field parameterization to predict the structure of peptides and proteins at interfaces.

Samenvatting

Biomineralen zijn fascinerende composieten van organische en anorganische stoffen die over honderden miljoenen jaren zijn geëvolueerd. Met hun ingewikkelde architectuur op nano- tot microschaal tonen biomineralen buitengewone eigenschappen wanneer het gaat om taaiheid, kracht en gewicht. Deze eigenschappen worden bijvoorbeeld gevonden in de silicaskeletten van diatomen. De vorming van silica in diatomen is goed onderzocht op macroscopische schaal. Echter, de implementatie van biomimetische silicavorming in technologische toepassingen vereist inzicht op moleculair niveau in hoe de silicamorfologie wordt gecontroleerd door organische moleculen. Peptiden zijn een klasse organische moleculen die extensief gebruikt zijn om biomoleculen die betrokken zijn bij biosilicificatie van diatomen na te bootsen. Het doel van dit proefschrift, zoals gedefinieerd in de inleiding, was het vergroten van de kennis over hoe peptidestructuur en ordening aan grensvlakken de morfologie van kunstmatig gegenereerd silica beïnvloeden. In deze slotopmerkingen zal ik samenvatten wat we ontdekt hebben over peptide-gecontroleerde silicamineralisatie.

Ten eerste werd de relatie tussen peptideordening en silicamorfologie in oplossing onderzocht met behulp van vereenvoudigde modelpeptiden (LK-peptiden). In vervolgonderzoeken zijn complexe peptiden en grensvlaksystemen bestudeerd. LK peptiden zijn samengesteld uit hydrofobe en hydrofiele bouwstenen, die typisch in een karakteristieke volgorde worden georganiseerd. Er is verondersteld dat verschillen in silicamorfologieën, die worden gegenereerd door verschillende modelpeptiden in oplossing, kunnen worden toegeschreven aan verschillend agglomeratiegedrag dat voortvloeit uit de kenmerkende hydrofobe / hydrofiele sequentie van het peptide. In dit proefschrift laat ik echter zien dat de morfologie van silica niet alleen wordt beïnvloed door het algehele hydrofobe / hydrofiele patroon van de peptidesequentie, maar ook door kleine, bijna onopvallende veranderingen van het systeem, zoals de aan- of afwezigheid van peptideacetylering. Verrassend genoeg genereren de geacetyleerde en niet-geacetyleerde varianten

van een LK-peptide volledig verschillende silicamorfologieën, hoewel hun structuur in oplossing hetzelfde lijkt te zijn. Grensvlakspecifieke vibratiespektroscopie in combinatie met coarse grained moleculaire simulaties wijzen erop dat de hydrofobe kern van peptideagglomeraten wordt gestabiliseerd door de afwezigheid van N-terminale lading in geacetyleerde peptiden. Dit resultaat laat zien dat peptideagglomeratie een cruciale rol speelt in de vorming van silicananostructuren, maar zeer gevoelig is voor de kleinste veranderingen van het systeem. Verdere inzichten in de agglomeratie van biomineralisatie-actieve peptiden zullen zeer waardevol zijn voor grootschalige biotechnologische silicamineralisatie in oplossing. Kernspinresonantie (NMR) in oplossing en vaste stof kunnen een moleculair beeld van peptideagglomeraten in oplossing of binnen silicaprecipitaten verschaffen.

LK peptiden werden oorspronkelijk ontworpen om aan polaire / apolaire grensvlakken klassieke structurele elementen van eiwitten aan te nemen . Daarom werd het effect van verschillende peptidestructuren op de grensvlakmineralisatie van silica onderzocht met LK-peptiden geadsorbeerd aan het lucht-water grensvlak. In hoofdstuk 2 laat ik zien dat helix en β-streng LK-peptiden verschillende silicamorfologieën mineraliseren aan het lucht-water grensvlak. Moleculaire simulaties leverden bewijs dat helische LK-peptiden voornamelijk met hun zijketens met silica interacteren, terwijl β-streng LK-peptiden met de ruggengraatgroepen interacteren. Het is onduidelijk of deze bevindingen algemeen toepasbaar zijn. De experimentele aanpak kan echter worden toegepast om de grensvlakbiomineralisatie van andere modelpeptiden en -systemen zoals calciumcarbonaat en calciumfosfaat te onderzoeken.

In plaats van het veranderen van het mineraalsysteem hebben we de peptidecomplexiteit vergroot door het onderzoeken van de grensvlaksilicavorming bemiddeld door het peptide R5. R5 is afgeleid van een diatoompeptide dat aanwezig is in de silicacelwand en heeft relevantie voor talrijke biotechnologische toepassingen, zoals hybride materialen of het bemiddelen van eiwit-silica interacties. Tijdens silicamineralisatie aan het lucht-water grensvlak hebben we gevonden dat R5 goed gestructureerd blijft. Silica dat gemineraliseerd is door het R5-peptide aan het lucht-water grensvlak lijkt te groeien door middel van silica-intercalatie in aan het grensvlak geadsorbeerde peptide-agglomeraten. Vanwege zijn biotechnologische relevantie zullen de hierin gepresenteerde resultaten waarschijnlijk bijdragen aan een beter begrip van de R5-gestuurde grensvlakmineralisatie van silica.

Elk van de studies in dit proefschrift is afhankelijk van moleculaire dynamica simulaties - net zoals vele complementaire studies op het gebied van biomoleculaire oppervlaktewetenschappen. Moleculaire simulaties worden op regelmatige basis gebruikt om te helpen bij de interpretatie en verificatie van experimentele data verkregen aan grensvlakken. In het specifieke geval van peptiden en eiwitten aan het lucht-water grensvlak is de reproductie van nauwkeurige structurele ensembles tot op heden nog niet onderzocht. In het laatste hoofdstuk van dit proefschrift evalueren we het effect van moleculaire simulatieparameters voor peptiden aan het lucht-water grensvlak. Door het vergelijken van somfrequentie-generatiespectra en berekende spectra uit moleculaire dynamica simulaties tonen we aan dat slechts twee parametersets (van de 12 geteste sets) het conformationele ensemble van een bepaald peptide aan het lucht-water grensvlak nauwkeurig reproduceren. Deze zijn het CHARMM27 eiwitkrachtveld in combinatie met het TIP4P-D watermodel en het AMBER99SB*-ILDN eiwitkrachtveld in combinatie met het TIP3P watermodel. We zijn van plan om deze aanpak uit te breiden naar andere peptiden met andere structuren en uiteindelijk tot complexere eiwitten. Met aanvullende studies hopen we te begrijpen hoe de krachtveldmodellen aangepast dienen te worden om moleculaire interacties aan het lucht-water grensvlak nauwkeurig weer te geven.

Het begrip van de peptidebemiddelde biosilicificatie op moleculair niveau, dat we aan het begin van dit project zochten, begint dus geleidelijk vorm aan te nemen. Met behulp van geavanceerde niet-lineaire vibratiespectroscopie in combinatie met moleculaire dynamica simulaties hebben we onze mogelijkheden om de conformatie van verschillende peptiden aan grensvlakken vast te stellen uitgebreid en hebben we aangetoond dat de silicamorfologie inderdaad afhankelijk is van de structuur van een gegeven grensvlakgeadsorbeerd peptidesjabloon. We hebben ook rekening gehouden met de effecten van peptideagglomeratie op silicificatie. Enkele aspecten van grensvlakbiosilicificatie konden niet in dit proefschrift worden behandeld, maar zouden wel interessant kunnen zijn om verder te onderzoeken. Het blijkt dat het effect van peptideagglomeratie op de morfologie van het gemineraliseerde silica meer uitgesproken is dan het effect van verschillende peptidestructuren. Een interessante route voor onderzoek zou het onderzoeken van de tweedimensionale agglomeratie van modelpeptiden aan grensvlakken zijn. Naast de inzichten in de rol van structuur en agglomeratiegedrag in kunstmatige silicastructuren, hebben we een modelgeval voor verbeterde moleculaire simulaties van peptiden aan het lucht-water grensvlak geëvalueerd. De resultaten van deze interdisciplinaire

aanpak zullen een richtlijn bieden voor toekomstige studies en hopelijk inspireren tot de ontwikkeling van betere simulatieprotocollen en veldsterkteparametrisatie om de structuur van peptiden en eiwitten aan grensvlakken te voorspellen.

Acknowledgements

This work would not have been possible without the support of many people. In this part of the thesis I wish to express my gratitude to the mentors, colleagues, family and friends who have helped me to complete my Ph.D.

First, I wish to thank my supervisors. For the last four years, Tobias Weidner has been a great mentor and motivator. Tobias's excitement for science has kept me going and has always inspired me to explore uncharted territory. Thank you for believing in me and helping me to become a better scientist. I would also like to thank Mischa Bonn for his guidance, support and many helpful discussions. For the time as a guest researcher at the University of Washington, I wish to thank Jim Pfaendtner for his interest in my research and for being a great mentor. I would also like to thank my doctoral committee for taking the time to evaluate my thesis and for their participation in my Ph.D. defense ceremony.

I would like to express my very great appreciation to all collaborators who devoted their time and effort towards my projects. I am particularly grateful to Vance Jaeger for being my unwavering ally in the realm of computational chemistry throughout my Ph.D. Using "computer magic" he managed to overcome all problems, that we were faced with on the path to publications. I am greatly indebted to you for all your efforts in helping me to graduate – including critically reading most of this thesis and teaching me the art of molecular simulation. Last but not least, thank you for your friendship.

Thank you to all members of the Weidner group, the Bonn group and the Pfaendtner group for critical discussions, valuable advice, and for inspiring my scientific work. I wish you the best of luck on all of your future endeavors. In particular, I would like to thank: Maksim, Yuki and Sapun for carefully reading sections of this thesis and providing valuable feedback and corrections. Leonie for a wonderful Dutch translation of the thesis summary. Yuki for providing the fundamental idea and valuable input to the project described in Chapter 5. Sergio for being a patient teacher and introducing me

to the operation of SFG setups. Mike, Lars and Johannes who have been great companions in the lab. I also wish to acknowledge the help provided by the students whom I have had the pleasure to supervise: Nina and Konrad. A special thanks to the scientific technicians who worked with me on most of the projects: Sabine, Gunnar, Gabi, Katrin, Helma, Marc-Jan and Florian. I would also like to thank Laurie for making my Ph.D. life so much easier and less stressful. You are wonderful people and I enjoyed all of the great conversations and this wonderful community.

Thank you to all of my family, in particular, my parents Johann and Christina, for their faith in me and for allowing me to follow my ambitions throughout my childhood. Thank you to all my friends for their emotional support and the good times we had.

Finally, and most importantly, I would like to thank my wife Sarah for standing beside me throughout the ups and downs of the Ph.D. She helped me with all the little tasks of life so I could focus on finishing the Ph.D. Without her love and sacrifice this thesis would have never become reality. More importantly, Sarah always gave me strength and encouraged me to handle the difficulties that emerged over the last four years. Thank you Sarah, this thesis is dedicated to you.