Solid State NMR Characterization of Biosilicification Peptides

Ariel C. Zane

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington

2013

Reading Committee:

Gary Drobny, Chair

Munira Khalil

Bo Zhang

Program Authorized to Offer Degree:
Chemistry

University of Washington

Abstract

Solid State NMR Characterization of Biosilicification Peptides

Ariel C. Zane

Chair of the Supervisory Committee:
Professor Gary Drobny
Department of Chemistry

Biomimetic syntheses of useful inorganic and mineral materials are the subject of much interest in the biomineralization community, as organisms can often produce materials at ambient conditions and temperatures, as opposed to the extreme conditions of commercial methods. The focus here is on biosilicification peptides, which produce silica nanoparticles out of solutions of silicic acid. Solid-state NMR (SSNMR) investigations of the R5 peptide (part of a naturally occurring sequence in diatoms) and the synthetic, model LKα14 peptide are detailed here, utilizing 1D and 2D ^{13}C, ^{15}N and ^{29}Si CPMAS and DARR experiments and culminating in the elucidation of secondary structures for both peptides in their neat (lyophilized) form, and their form when complexed to biosilica. Additional relaxation experiments (^{13}C CP T1IR) probe dynamics of side chains. It is shown that secondary structure alone is not responsible for the morphology of the resulting silica nanoparticles, and the importance of larger scale aggregation is emphasized. Several self-assembly and silica templating models for the R5 and LKα14 peptide are proposed, with supporting SSNMR results and incorporating a review of relevant current literature in the field.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

GLOSSARY

ANN: Artificial Neural Network

CP: Cross Polarization

CPP: Cell Penetrating Peptides

CSA: Chemical Shift Anisotropy

DARR: Dipolar-Assisted Rotational Resonance

DPMAS: Direct Polarization Magic Angle Spinning

MAS: Magic Angle Spinning

PLL: Poly-L-lysine

R5-SIO_2: The R5 peptide, co-precipitated with silica, to form a composite

RR: Rotary Resonance

SPPS: Solid Phase Peptide Synthesis

SSNMR: Solid-State NMR

T_1IR: T_1 Inversion Recovery

TALOS-N: Torsion Angle Likeliness Obtained from Shift and Sequence Similarity

TIBALDH: Titanium(IV)-bis-ammonium-lactato-dihydroxide

TMOS: tetramethylorthosilicate

TMS: tetramethylsilane

ACKNOWLEDGMENTS

I would like to sincerely thank Professor Gary Drobny for his guidance and for taking an interest in my education as well as in my research projects. I would also like to thank all the members of the Drobny group for their encouragement and camaraderie, and for making me feel like part of a wonderful team.

Special thanks to: Jason Ash and Nicholas Breen, for supervising and helping me with my peptide synthesis, for teaching me how to set up and run NMR experiments, for patiently walking me through theory, and for allowing me to follow them around and ask lots of questions; Adrienne Roehrich, for being an excellent mentor and teacher; Prashant Emani, for humorously fielding all manner of theory and mathematical questions, for providing helpful perspective and for lending his powers of analysis; Matt Power, for his sage "real" chemist consulting prowess, and Christian Michelet, for his cheerful attitude and willing helpfulness.

And of course, I would like to thank my parents, for their support, my brother, who always makes me laugh, and my husband, for whom I am always thankful.

DEDICATION

For S.

Chapter 1

INTRODUCTION

Many biological organisms produce inorganic minerals in a controlled, purposeful method; bone, teeth, and shells are examples of biomineralization products. We strive to understand the methodology and mechanisms involved in these biomineralization processes in an attempt to mimic the creation of specific mineral materials - biomineral products like silica, titania, and calcium phosphate have much value and relevance in the modern world. Prosthetics and grafts are more easily accepted into the body when they are generated from non-alien materials, and biomimetic processes are often more moderate than current industrial methods. Various other examples of biomineralization that are important to human progress do exist, but this thesis focuses specifically on biosilicification - systems of proteins and peptides that co-precipitate with silica matrices, producing nanoparticles of uniform size and shape.

Synthetic silica materials are used today in a variety of applications, from biomedical technology (e.g., drug delivery systems) to common electronic devices like fiber optics. Current industrial methods of producing these silica-based materials require extreme temperatures, pressures and pH.[61] In contrast, certain biological organisms–namely, diatoms, a eukaryotic, unicellular microalgae–produce silica under much gentler conditions (ambient temperature and pressure, mildly acidic or neutral pH), and are capable of creating a diverse array of silica structures.[37] These diatoms are typically encased in cell walls composed of a combination of self-produced silica and other organic compounds, and are scientifically classified by the pattern of these self-constructed inorganic-organic matrices.

Studies in the field have focused on proteins extracted from various species of diatoms;[?] the most widely studied variant is termed "silaffin," and has been shown to be directly involved in silica formation.[53] More recent studies have focused on segments of silaffin proteins, such as the R5 peptide, one of several repeated segments in the *sil1-p* protein. Like its parent peptide, R5 precipitates silica nanospheres (with diameters of 400-700 nm) out of a buffered monosilicic acid solution, at neutral pH (pH = 5.5). The native *sil1p* protein precipitates silica nanospheres of similar diameters out of mildly acidic solutions.[14] Most of the existing literature on the topic centers on microbiology, marine biology, and materials science applications of the resulting silica (and silica engineering theories); this thesis attempts to deal with the R5-silica system on an atomic level, isolating protein sidechain and silica interactions, electrostatic and thermodynamic contributions, and other molecular level principles. Protein secondary structure and side chain dynamics will be analyzed, both in neat form and in silica, and compared. This is the subject matter of Chapter 3.

Extrapolating from the interaction picture drawn from the R5-silica system, we attempt to generalize protein-silica interactions by looking at synthetic, amphiphilic LK (Leucine, Lysine) peptides, which were previously used by the group in surface adsorption studies.[59, 12, 13, 65, 109] These peptides were developed by Degrado and Lear[20] in order to investigate the role of hydrophobic periodicity in the primary sequence of proteins. By varying the sequence of L (hydrophobic) and K (hydrophilic) residues, they created model peptides with different secondary structures at an apolar or water interface. Because they are relatively short (the most common LK peptides are less than 20 residues long) and have a diverse range of secondary structures that are easily controlled, they are a prime candidate for surface binding experiments. They also bind to silica surfaces and retain their original secondary structure in their bound states.[65] Because they have a large number of basic residues, we hypothesized and later demonstrated that these peptides do indeed co-precipitate with silica. Their easily controlled sequence length, conformation, and amphiphilic nature make

them worthy workhorses to demonstrate relationships that target specific protein properties, like amino acid sequence and presence, secondary structure (e.g., α-helix, β-sheet, etc.,). Exploration of these principles with respect to their relation with silica precipitation makes up the bulk of Chapter 4.

To study all of these proposed systems, we rely here mostly on solid-state NMR (SSNMR). It is a useful tool with which to probe the molecular level interactions between organic proteins and inorganic surfaces, and there are a wide range of experiments available to us. There are the routine 1D and 2D correlational spectra (e.g., DARR[103]) which enable us to assign chemical shifts to individual carbon atoms in amino acids. These chemical shifts can be fed into servers like TALOS-N [94], which draw upon a database of amino acid conformations (as they exist in proteins) to predict backbone torsion angles, and thus secondary structures for these primary sequences. Comparing chemical shifts of atoms in both the neat peptide and peptide-silica composite will show if the precipitation process induces any changes in peptide conformation, and can shed light on specific interactions occurring at the molecular level. In addition, relaxation experiments that measure T_1 for backbone and sidechain carbons and deuterons provide insight on the dynamics and relative time scales of motion of the side chain groups, which can be extrapolated into comments on the local mobility and environment of the particular side chain. The basic theory of SSNMR is explained in Chapter 2, and discussion of the specific experiments and results are part of Chapters 3 and 4.

Chapter 2

NUCLEAR MAGNETIC RESONANCE SPECTROSCOPY

NMR is the study of the interactions of atomic nuclei spin systems. Nuclear spin is a quantum mechanical property of some nuclei with spin quantum number $I > 0$, and is simply the angular momentum intrinsic to the particle. NMR active nuclei that will be discussed in this dissertation include ^{13}C, ^{1}H, ^{15}N, ^{29}Si. Each molecule is considered an individual spin system, and in NMR, an ensemble of spin systems is observed (thus, an average Hamiltonian is ultimately responsible for generating the response of the system). Nuclear spins of interest are placed in a large magnetic field, B_0 and subjected to radio frequency (RF) pulses of varying length and power levels; additionally, in SSNMR, physical manipulation (Magic Angle Spinning - MAS) is added. The response of the system to these treatments is observed and interpreted.

SSNMR is a widely used technique for characterizing peptides and proteins. It can be used to determine secondary and tertiary structures of proteins through the assignment of chemical shift values and the measurement of both heteronuclear and homonuclear dipolar interactions. The following chapter outlines general concepts of relevant SSNMR equations and theories, as applied to the biosilica system of interest. A more rigorous discussion of NMR and SSNMR theory are handled excellently in focused texts [24, 96] elsewhere.

2.1 Spin Interactions

Nuclear spin properties are defined by their spin quantum number, I, and their gyromagnetic ratio, γ. In general, the larger value γ, the larger magnetic moment of the nucleus. Nuclear spins interact with each other, with their local environments,

with the external radio frequency pulses, and with the external magnetic field $\mathcal{B}_{\prime}$. In the absence of a laboratory field, all spins of the same isotope (e.g., all ^{13}C spins) have the same energy. When subjected to a uniform, external magnetic field $\mathcal{B}_0$, the energy levels split into $2I + 1$ levels with respect to the direction of the field (in this case, along the z-axis); this is called the Zeeman splitting (Figure 2.1). Spin $\frac{1}{2}$ nuclei align either with or against the field ($2\left(\frac{1}{2}\right) + 1 = 2$ energy levels). As $\mathcal{B}_0$ increases, so does the energy difference between the levels, provoking the need for higher and higher magnetic fields $\mathcal{B}_0$.

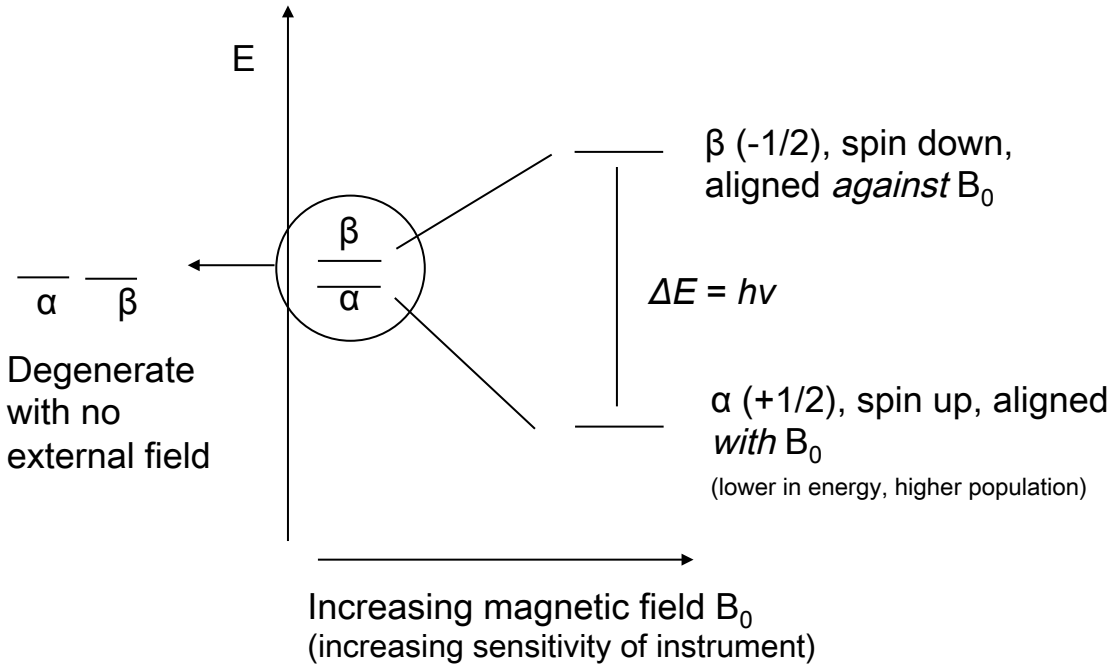

Figure 2.1: Zeeman splitting of degenerate energy levels for spin $\frac{1}{2}$ nuclei. Increasing magnetic fields $\mathcal{B}_0$ increases energy difference between the levels, reflecting the improved resolution at higher fields.

When a sample is placed in the magnetic field, the magnetization vector precesses about the magnetic field axis, which is conventionally the z-axis; the frequency of this precession around z is called the Larmor frequency, ω_0,

$$\omega_0 = -\gamma \mathcal{B}_0 \tag{2.1}$$

where γ is the gyromagnetic ratio of the nuclei.

The net magnetization along the external magnetic field is called the longitudinal magnetization, and is not observable through NMR. To flip the longitudinal magnetization down into the transverse x-y plane, a RF field designated $\mathcal{B}_1$, is applied orthogonally to $\mathcal{B}_\prime$. The frequency of the resulting nutation is

$$\omega_1 = -\gamma \mathcal{B}_1 \tag{2.2}$$

The transverse magnetization in the x-y plane is the only observable phase of the magnetization, corresponding to a single quantum coherence. The precessing magnetization in this transverse plane induces a current in the detection coil (which is the same as the excitation coil which generates $\mathcal{B}_1$). This creates a free induction decay, or FID, which is the NMR signal; this signal is Fourier transformed from time domain data into the frequency domain, to give the familiar frequency spectrum.

Nuclear spins interact with the external field, $\mathcal{B}_\prime$, applied RF pulses, and also with each other and their molecular surroundings. Local electronic information about a molecule is obtained through chemical shift analysis. The electronic environment changes at different sites in a molecule, and this changes the shielding effect that the nuclear spin experiences at that site; the chemical environment can shield or deshield the effects of $\mathcal{B}_0$. This difference is called the chemical shift, and manifests as a slight upfield or downfield shift of the Larmor frequency ω_0.

Spin-spin interactions make up the remaining interactions which can manifest as alterations to lineshape and width in a NMR spectrum, and there are several different ways the spins can interact with each other. Direct dipole-dipole coupling is a through-space interaction of nuclear spins. It is mandated by a dipolar coupling constant, which allows for measurement of internuclear distances through measurement of these

direct dipolar couplings. Through bond interactions are called J-couplings, and are generally ignored in SSNMR; the magnitude of J-couplings (1-2 Hz) is much smaller and considered negligible when compared to dipolar coupling (on the order of kHz).

Internal spin interactions that are dependent on the orientation of the molecule relative to the external field, they are anisotropic. Orientation dependent interactions are isotropic. In SSNMR, because there slower molecular motions and a variety of individual crystallite orientations, dipolar couplings and chemical shifts are anisotropic, though chemical shifts do have an isotropic component.

2.2 NMR Hamiltonians

Nuclear spin interactions are represented by Hamiltonian operators. The total NMR Hamiltonian is the sum of all of its component Hamiltonians - the sum of each interaction's Hamiltonian.

The interaction between the external magnetic field, $\mathcal{B}_0$, and a spin I is called the Zeeman interaction, and for a single spin, is given by

$$\hat{\mathcal{H}}_z = -\hbar\mathcal{B}_0(\gamma_I\hat{\mathcal{I}}_z) \tag{2.3}$$

where $\mathcal{I}_z$ is the spin operator corresponding to the z component of I. This can be extrapolated to a two-spin system:

$$\hat{\mathcal{H}}_Z = -\hbar\mathcal{B}_0(\gamma_I\hat{\mathcal{I}}_z + \gamma_S\hat{\mathcal{S}}_z) \tag{2.4}$$

where $\mathcal{S}_z$ is likewise the spin operator for the z component of the second spin S. This is the first external interaction to consider. The second, is the effect of the RF pulse. This is given as

$$\hat{\mathcal{H}}_{RF} = -\hbar\gamma_I\mathcal{B}_1\left(i\cos\omega t + j\cos\omega t\right)\cdot\hat{I} \tag{2.5}$$

where again, $\mathcal{B}_1$ is the applied RF field in a direction orthogonal to $\mathcal{B}_0$. To show the effect of a pulse θ (where θ usually some frequency ω_1 for some time t) along x on a spin I_z, we write

$$\hat{\mathcal{I}}_z \xrightarrow{\theta_x} \hat{\mathcal{I}}_z \cos\theta + \hat{\mathcal{I}}_y \sin\theta \tag{2.6}$$

The next interactions are internal to the sample; the first that we will look at is the chemical shift interaction, which is a measure of the spin's interaction with the local magnetic fields induced in the surrounding molecular orbitals by the external field. This is demonstrated by

$$\mathcal{B}_{induced}(t) = \sigma_I(t) \cdot \mathcal{B}_0 \tag{2.7}$$

where $\sigma_I(t)$ is the chemical shift tensor acting on spin I. For the two spin case,

$$\hat{\mathcal{H}}_{CS}(t) = \gamma_I \hbar I \cdot \sigma_I \cdot \mathcal{B}_0 + \gamma_S \hbar S \cdot \sigma_S(t) \cdot \mathcal{B}_0 \tag{2.8}$$

We have already said that there are two components of the chemical shift: an isotropic component and an anisotropic component. In general,

$$\sigma_{CS} = \sigma_{CS}^{iso} + \sigma_{CS}^{aniso} \tag{2.9}$$

where σ_{CS}^{iso} is a time-independent scalar, and σ_{CS}^{aniso} is actually the time dependent $\sigma_{aniso}(t)$.

The next term in the overall NMR Hamiltonian is the dipolar coupling Hamiltonian. The dipolar coupling quantifies the interaction of spins through space; spins that are in close proximity will induce changes in the local magnetic field affecting the other spin. The Hamiltonian is given by

$$\hat{\mathcal{H}}_D = -2\hat{I} \cdot D \cdot \hat{S} \tag{2.10}$$

where D is the dipole-coupling tensor, which encompasses the dipolar coupling constant d

$$d = \hbar \left(\frac{\mu_0}{4\pi r_{IS}^3} \right) \gamma_I \gamma_S \tag{2.11}$$

written in units of Hz, where μ_0 is the magnetic permeability of free space, and given as $4\pi x 10^{-7} N \cdot A^2$, and the internuclear distance between spins I and S is designated by r^{IS}. This shows that the dipolar coupling d (usually on the order of kHz) is proportional to $1/r^{IS}$; measurements of dipolar couplings between two spins I and S can then be extrapolated into approximations of the internuclear distance between these spins.

The dipolar coupling terms differ between the homonuclear and heteronuclear cases; the homonuclear dipolar coupling Hamiltonian is

$$\hat{\mathcal{H}}_D^{homo} = -d\frac{1}{2}\left(3\cos^2\theta - 1\right)\left[3\hat{I}_z\hat{S}_z - \hat{I}\hat{S}\right] \tag{2.12}$$

and for the heteronuclear case

$$\hat{\mathcal{H}}_D^{hetero} = -d\left(3\cos^2\theta - 1\right)\hat{I}_z\hat{S}_z \tag{2.13}$$

both in angular frequency units.

Next is the quadrupolar interaction, which is the interaction between a nuclear electric quadrupole and an electric field gradient, rather than between a nuclear magnetic dipole and a magnetic field. The quadrupolar interaction only arises for spins greater than $I = \frac{1}{2}$, such as ^{2}H. The Hamiltonian is

$$\hat{\mathcal{H}}_Q = \frac{eQ}{2I\left(2I-1\right)\hbar}\hat{I} \cdot V\hat{I} \tag{2.14}$$

where V is the electric field gradient, e is the charge of a proton, and Q is the nuclear electric quadrupole moment.

The last internal interaction term to consider is J-coupling, which is a weaker coupling than dipolar coupling, and in SSNMR is usually considered negligible. For completion, however: J-coupling, the indirect coupling of spins through bonds is written as

$$\hat{\mathcal{H}}_J = \hat{\mathcal{H}}_J^{iso} + \hat{\mathcal{H}}_J^{aniso}(t) \tag{2.15}$$

and

$$\hat{\mathcal{H}}_J^{iso} = hJ_{IS}^{iso}\hat{I} \cdot \hat{S} \tag{2.16}$$

$$\hat{\mathcal{H}}_J^{aniso}(t) = h\hat{I} \cdot J_{IS}^{aniso}(t) \cdot \hat{S} \tag{2.17}$$

for the time-independent isotropic component $\hat{\mathcal{H}}_J^{iso}$ and time-dependent anisotropic component $\hat{\mathcal{H}}_J^{aniso}(t)$, respectively.

The total NMR Hamiltonian is the sum of the external and internal components

$$\hat{\mathcal{H}}_{NMR} = \hat{\mathcal{H}}_{ext} + \hat{\mathcal{H}}_{int} \tag{2.18}$$

expanded to

$$\hat{\mathcal{H}}_{NMR} = \hat{\mathcal{H}}_Z + \hat{\mathcal{H}}_{RF} + \hat{\mathcal{H}}_{CS} + \hat{\mathcal{H}}_D + \hat{\mathcal{H}}_Q + \hat{\mathcal{H}}_J \tag{2.19}$$

2.3 MAS and High Resolution Spectra

SSNMR is a useful method in biomineralization systems for a number of reasons, such as the inherent ability to probe interactions as a function of hydration (e.g., performing the same experiment on a series of gradually more hydrated samples) and to tackle systems that may not have perfect crystal structures, or that have a different conformation in the solid state than in solution. Proteins can also be investigated in their bound state (i.e., adsorbed to a mineral surface or within a mineral matrix).

As in solution state, though, isotropic chemical shift information can be obtained, and still gives insight into the local electronic environment of a nuclear spin. Besides basic 1D and 2D experiments, other examples of useful techniques include dipolar recoupling experiments (to measure inter- and intra-molecular distances in proteins and between proteins and surfaces), and relaxation experiments.

In the solid state, molecular motion is limited or inhibited; a solid powder sample contains many individual crystallites with a random distribution of orientations. The NMR spectrum of a static (not spinning) powder sample is characterized by broad lines; each of the individual spins experiences a different local magnetic field, which leads to many different peaks across a range of frequencies. The broadness of the static lineshape is proportional to the chemical shift anisotropy.

For targeted, focused probing of protein structure and interactions, it is necessary to incorporate site-specific isotopic labeling in order to use SSNMR methods. The natural abundance of NMR-active nuclei (e.g., ^{13}C, ^{15}N) in biological compounds is very low (1.1% and 0.37%, respectively). Isotopically enriched amino acids are incorporated into the protein sequence through SPPS, as detailed in Appendix A. For practical reasons, usually labeling schemes are constructed such that several isotopically-enriched amino acids can be incorporated into the same sample, to minimize synthesis time and also, some experiments require the presence of specific pairs of atoms (e.g., dipolar coupling distance measurement experiments like REDOR[32]).

2.3.1 Cross Polarization

The problems inflicting SSNMR are low sensitivity and broadened lines, compared to solution state NMR. These are addressed by several methods, including CP and MAS, which are the subjects of the next several subsections.

Cross polarization, or CP, is a technique used in SSNMR to improve the signal-to-noise ratio of the spectrum. "Dilute" nuclei with low sensitivity and/or abundance such as ^{13}C and ^{15}N can have their signals enhanced by transferring polarization from

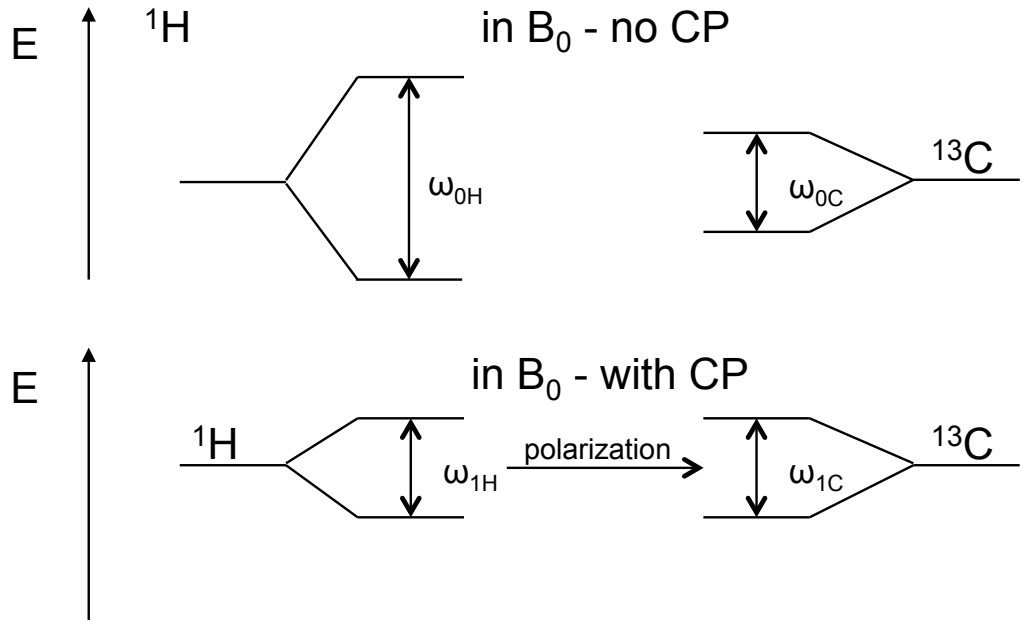

Figure 2.2: The top figure depicts the normal difference between rotational frequencies of ^{1}H and^{13}C; the bottom figure shows the Hartmann-Hahn matching of energy levels and rotational frequencies of ^{1}H and ^{13}C, enabling cross polarization of ^{13}C spins with proton spins.

more abundant nuclei (usually ^{1}H or ^{19}F).[35] In general, the abundant nucleus is excited by a $\frac{\pi}{2}$ (or 90°) RF pulse, and then during a "contact time," during which both the ^{1}H and ^{13}C spins are subjected to RF pulses, the polarization transfer from the abundant nucleus to the dilute spin. This is essentially because the flip-flop energy levels between ^{1}H and ^{13}C are now matched (Figure 2.2).

During the contact time, it is necessary to generate a ratio of RF power levels for ^{1}H and ^{13}C that is equal to the ratio of gyromagnetic ratios between ^{1}H and ^{13}C; that is, for ^{1}H and ^{13}C, a 4:1 power level ratio must be maintained in order to maintain the same excitation frequency for both spins.[35] The CP block is referred to as the contact time, and is usually performed as a ramped-amplitude of ^{1}H spin lock pulse for more efficient cross polarization transfer.[66] The ramped amplitude block effectively creates a series of Hartmann-Hahn conditions, eliminating the problem of amplitude

mismatches of the two nuclei spin-lock fields.[66] The basic CP pulse sequence is given in Figure 2.3 below.

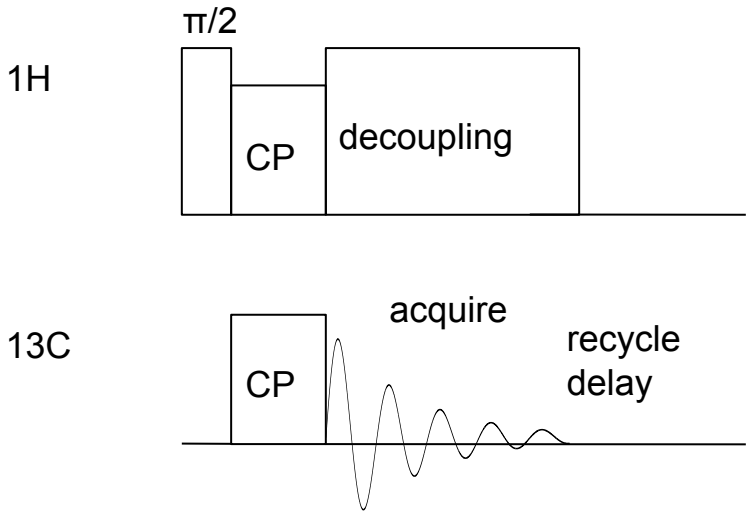

Figure 2.3: Pulse sequence for a 1D CPMAS experiment with acquisition on ^{13}C. Each block indicates an RF pulse; the width of the block is proportional to the width of the pulse, but typically schematics are not drawn to scale. The blocks are labeled with the type of pulse (e.g., $\frac{\pi}{2}$ indicates a 90° pulse along some phase), and the FID cartoon indicates the acquisition period. Delays are signified by a lack of pulse blocks, and are usually marked with τ or t. Generally, the recycle delay at the end (time between successive experiments) should be longer than the T_1 of the X nucleus (in this case, ^{13}C), but with cross polarization, it is sufficient to wait simply the T_1 of ^{1}H, which is another benefit of the CP technique.

Most SSNMR experiments utilize CP in some way, to enhance the signal strength of the dilute spin, which is often the detection nucleus. The downside is that the intensity of the peak is no longer transparently related to the number of spins at that resonance, as it is in most solution NMR. Also, in samples without abundant populations of protons (e.g., an SiO_2 matrix without water), efficiency is drastically reduced.

2.3.2 MAS

Magic-angle spinning (MAS) is the most commonly used technique in SSNMR, because it reduces the line broadening that comes from the chemical shift anisotropy and dipolar coupling. The CSA and dipolar coupling terms in the overall NMR Hamiltonian (equations 2.8, 2.12 and 2.13) contribute to line broadening in the solid-state; in the solution state, the constant molecular tumbling averages these interactions out, creating the familiar narrow peaks characteristic of solution state spectroscopy. In order to time-average out these interactions and remove the broadening from the CSA and dipolar coupling, the sample is spun at at an angle of $\theta = 54.74°$.[24] When $\theta = 54.74°$, the $3cos^2\theta$-1 term in the time-dependent $\hat{\mathcal{H}}_D$ (equations 2.12 and 2.13) averages out to zero, eliminating the dipolar contribution to the observable peak. The chemical shift anisotropy averages out to a non-zero term, so there still exists some line broadening due to the CSA. Figure 2.4 gives an example of how MAS enables better peak resolution.

2.3.3 Chemical Shifts

Just as in solution NMR, the isotropic chemical shift in a basic 1D ^{1}H, ^{13}C, or ^{15}N spectrum can give information about the local electronic environment of the respective nuclear spins of interest. In practical terms, the isotropic chemical shift is designated

$$\delta_{iso} = \frac{\nu - \nu_{ref}}{\nu_{ref}} \tag{2.20}$$

where ν_{ref} is the Larmor frequency in Hz units and ν is the spectral frequency of the spin of interest. The chemical shift scale, in ppm (parts per million), is referenced to TMS (tetramethylsilane) at 0 ppm. All other compounds are referenced to this external standard. In SSNMR, secondary standards are often used, such as adamantane for ^{13}C or isotopically labeled ammonium sulfate for ^{15}N.

In practical terms, the chemical shift is used to identify spins with particular local

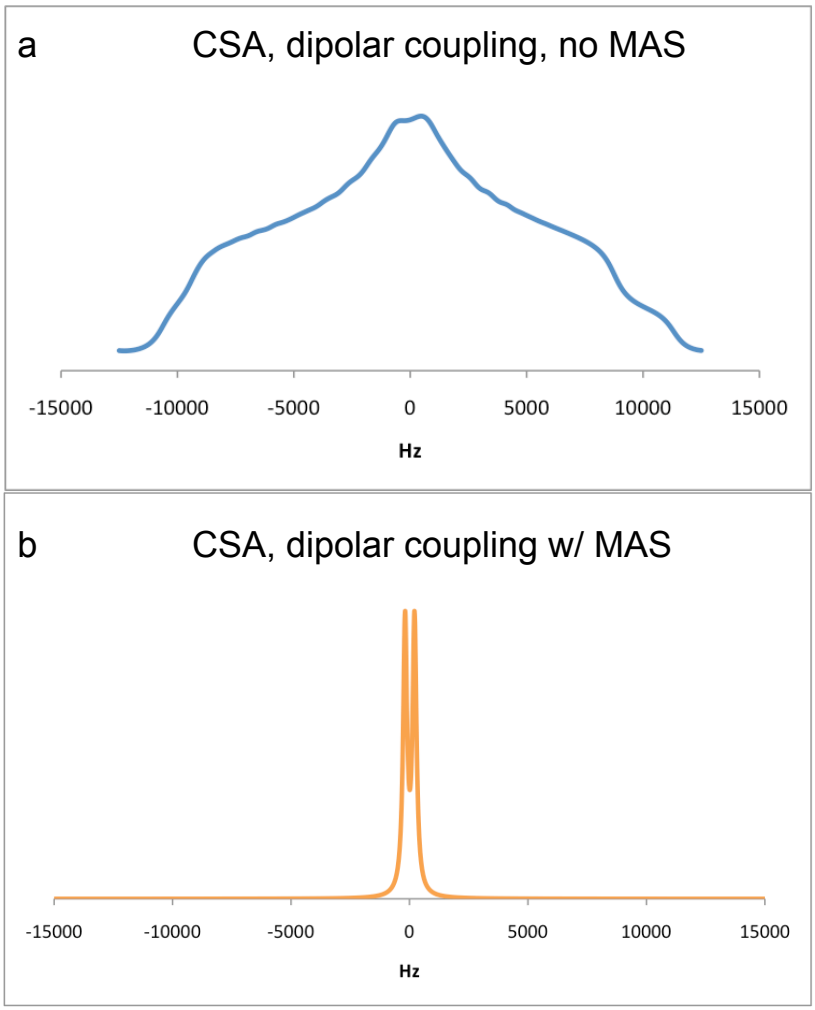

Figure 2.4: This example shows a simplified SSNMR ^{13}C spectrum with two ^{13}C spins. In a., the dipolar coupling between two ^{13}C spins and the chemical shift anisotropy, or CSA, broaden the peak. In b., MAS is added, which gives two separate ^{13}C peaks that are fairly well-resolved.

electronic environments; in solution NMR, ^{1}H and ^{13}C spectra can give information about what functional R groups are present in a hydrocarbon or other organic molecule. Likewise, in the solid state, especially for proteins, ^{13}C chemical shifts fall in identified regions of the spectra. It has been widely accepted that the chemical shifts of atoms in amino acids along a peptide chain, change with the secondary struc-

ture of the peptide[41]. Specifically, the carbonyl, alpha carbon, and beta carbon are closely correlated to secondary structure. Tables of amino acid chemical shifts have been compiled from protein databases, and the range of chemical shifts varies appropriately, representing a range of conformations and regular secondary structures.[106]

2.4 2D SSNMR

It is often the case that the 1D ^{13}C spectrum does not have the resolution to pick out individual peaks with very similar chemical shifts, as shown in Figure 2.5.

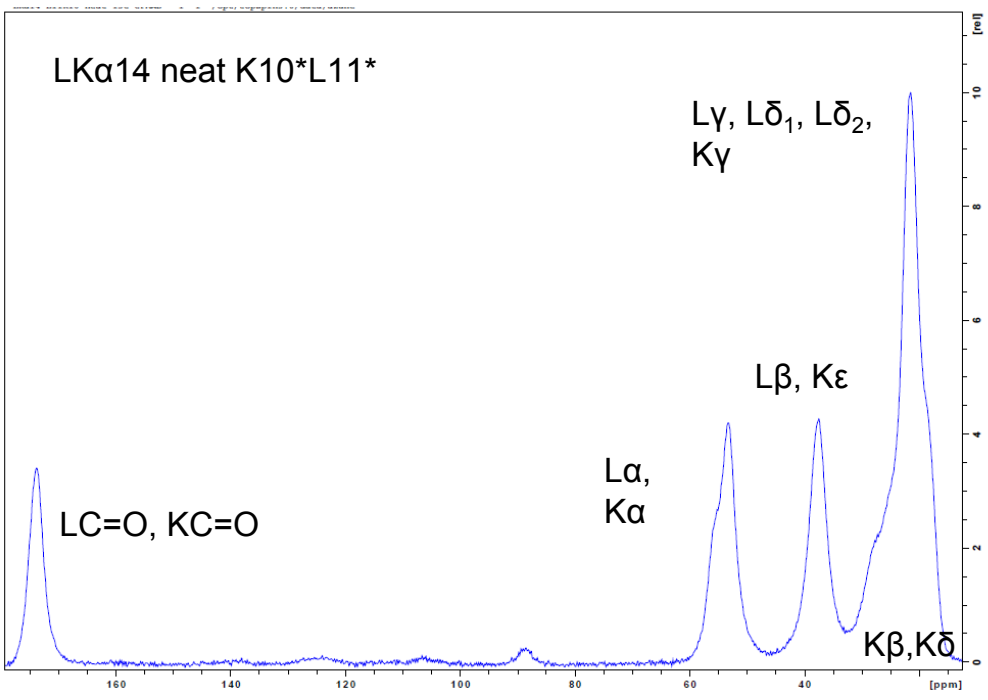

Figure 2.5: 1D ^{13}C CPMAS spectrum of the LKα14 peptide, with uniform ^{13}C and ^{15}N labels at amino acids K10 and L11. This is after 512 scans, with a 2 second recycle delay, spinning at 15±5 kHz, taken on a 700.18 MHz Bruker Avance III spectrometer, fitted with a ^{1}H{^{13}C, ^{15}N} 3.2 mm probehead. It is apparent from this spectrum that the chemical shifts for certain carbons in the backbone and sidechains of leucine and lysine are very similar; they overlap and cannot be individually resolved with a simple 1D CPMAS spectrum. Labels indicate accepted ranges of chemical shift for individual carbons in each L and K residue.

In this case, it is necessary to use 2D correlational NMR to separately resolve each residue's contribution to broadened peaks. This is done by correlating spin pair interactions with the cross peaks they give rise to. This can be done with homonuclear 2D experiments, or heteronuclear ones. Many different pulse sequences are used to this end; the one that is used in this thesis is DARR, or Dipolar-Assisted Rotational Resonance.[102]

2.4.1 The DARR pulse sequence

DARR[102], or Dipolar-Assisted Rotational Resonance, is a 2D homonuclear dipolar coupling experiment. It reintroduces the ^{13}C -^{13}C dipolar coupling that is traditionally removed by MAS, and enables us to pick out spin-spin interactions. ^{13}C spins that are close to each other in space will experience the effects of dipolar coupling, and this manifests itself as cross peaks in the 2D spectra. The pulse sequence for DARR is shown in Figure 2.6 below.

The DARR pulse sequence relies on ^{1}H driven broadband recoupling to reintroduce the ^{13}C -^{13}C dipolar coupling that is normally averaged out by MAS. Magnetization is generated on ^{1}H spins through a basic CP pulse and contact time on the ^{13}C channel. During the mixing time, ^{1}H spins are irradiated at the rotary resonance[76], which reintroduces the heteronuclear ^{1}H -^{13}C interaction, resulting in line broadening of the ^{13}C spectrum. This additional broadening increases the frequency range over which ^{13}C resonances overlap, effectively reintroducing the ^{13}C -^{13}C dipolar coupling, and giving a 2D spectrum of ^{13}C spin correlations.

The initial $\frac{\pi}{2}$ pulse on the ^{1}H channel is an initial excitation pulse for protons, and the subsequent CP pulse train transfers the polarization from abundant protons to ^{13}C spins. The t_1 period is the evolution period, during which the ^{13}C spins are allowed to evolve under the Zeeman interaction and MAS with proton decoupling (allowing it to evolve under its characteristic Larmor frequency, ω_1). Then, the $\frac{\pi}{2}$ pulse on the ^{13}C channel flips this magnetization onto the z-axis, and then in the following mixing

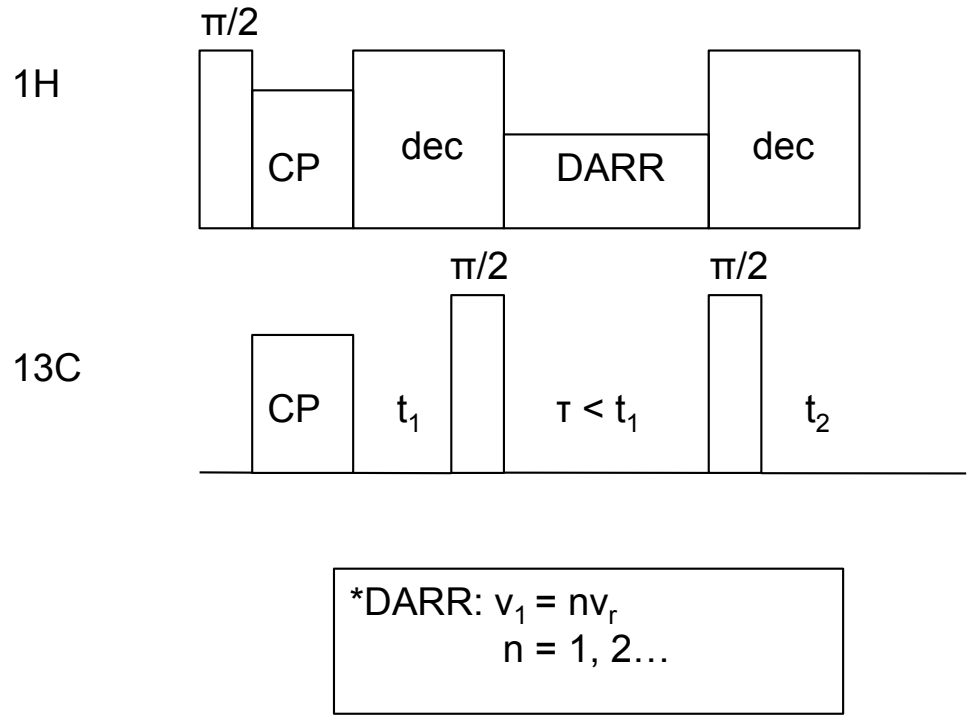

Figure 2.6: The DARR pulse sequence. The DARR block is a rotary resonance condition on the ^{1}H channel; ν_1 is the ^{1}H frequency, equal to integer (n multiples of the MAS speed,ν_r. The τ period is the mixing time, when ^{13}C spins mix and transfer polarization from one ^{13}C spin to another. Acquisition time on the ^{13}C channel is designated t_1. Adapted from [103].

period, τ, spins are allowed to interact, promoting spin exchange (which manifests as cross peaks in the resulting 2D spectrum). ^{1}H decoupling is also turned off to allow dephasing of unwanted coherences. Magnetization is then returned to the transverse plane with the last ^{13}C $\frac{\pi}{2}$, where it is observed (t_2).

At long mixing times τ, the magnetization transfer between spins becomes more efficient; typical DARR mixing times are on the order of 10-1000 ms, with inter-residue peaks becoming more prominent toward the end of the time range.[103]

2.5 TALOS-N

The chemical shifts that were assigned through 1D or 2D experiments, outlined above, can be used in several different analyses. The ΔCS is a measurement of a difference in chemical shift between, for example, two states of the same peptide; a relevant example is between a peptide in its native form and when it is bound to a mineral surface. The changes in chemical shifts of the same backbone and sidechain carbons can indicate a change in conformation, environment, or mobility. To generate empirical predictions for the backbone torsion angles of peptides and proteins under analysis, we submit experimentally obtained chemical shifts to a artificial neural network (ANN) based hybrid system, called TALOS-N.[94]

TALOS-N, like the original version of the system, TALOS (Torsion Angle Likeliness Obtained From Shift and Sequence Similarity),[95] is based on the principle that homologous protein sequences share similar chemical shifts. It searches a database (PDB - protein database, or BMRB - Biological Magnetic Resonance Bank)[106, 10] of high-resolution structures for the 10 best matches to the secondary chemical shifts of a residue, along with its immediate neighbors. If there is a match, the database triplet is used to make a prediction for the backbone torsion angles of the submitted residue. TALOS-N improves upon this original method by including the likelihood for a given residue to be in a α, β, or positive-ϕ conformation.

TALOS-N ranks backbone torsion angle predictions as "strong" if it is consistent and falls within a single clustered area of the Ramachandran plot[83]. "Bad" predictions deviate significantly from observed $\phi\psi$ angles.[94]

2.5.1 Protein Secondary Structure

It is now necessary to go through a brief description of protein structure. The primary sequence is taken to be the specific, unique order of amino acid units in a polypeptide unit. Each amino acid is connected to the next through a series of covalent peptide

bonds, and the sequence, by convention, runs from the N-terminus to the C-terminus. Secondary structure refers to the local, regular structure of these peptide sequences, such as an α-helix or a β-sheet. These conformations are characterized by regular hydrogen bonding patterns throughout the local sequence, as determined by Pauling et al.[80] In this way, the secondary structures are also characterized by a set of torsion angles, ϕ, ψ, ω.

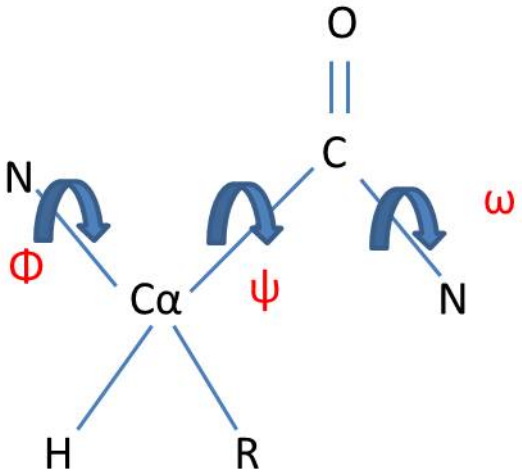

Figure 2.7: Torsion angles ϕ, ψ, and ω. The angles ϕ and ψ have a limited range of motion due to steric hindrances; ω is assumed to be planar (either 0 or 180°) and is usually not included in discussions.

It is the combination of ϕ and ψ angles that dictate regular secondary structure; the ranges of valid torsion angles for a specific conformation is indicated by a Ramachandran plot.[83] Repeating values of ϕ and ψ angles result in a regular structure. The range of valid ϕ and ψ angles for an α-helix are clustered around -57, -47, respectively; for a β-sheet, they are around -80, +150[41]. There are methods for experimentally inferring these torsion angles with NMR, through direct dipolar coupling measurements of intramolecular atom pairs, or through analysis of peptide chemical shifts with a program like TALOS-N.[94]

Tertiary structure (three-dimensional assembly of a single protein unit) and quaternary structure (assembly and/or aggregation of multiple protein units) are also necessary factors to consider when studying the interaction of proteins with any species, but are not dealt with in this thesis. Most of the peptide fragments (e.g., R5, LKα14) that are the subject of study here are considered too small for strict tertiary characterization, though aggregation and self-assembly are important topics of analysis and the resulting final discussion of the systems at hand.

2.6 Relaxation

In addition to characterizing static structures and assigning chemical shifts, it is also useful to assess the time scales of molecular motions and movements. Relaxation experiments are used to characterize these dynamic properties. Nuclear spin relaxation processes are mediated by fluctations resulting from molecular motion, which dictate the strength and rate of spin interactions like the dipolar coupling and chemical shift interactions (Equations 2.12, 2.13, 2.8), as well as an individual spin's interaction with the spin-lattice. Exchange rates of side chains flipping between rotamers, for example, can be calculated through measurement of the longitudinal (T_1) or transverse (T_2) time constants; these govern the relaxation rates of a perturbed spin returning to its equilibrium state, and are usually determined by experiments designed to measure a decay or buildup of intensities over variable delay periods τ.

2.6.1 CP ^{13}C T_1 Inversion Recovery

To measure the T_1 times for ^{13}C spins in biological molecules, the CPMAS T_1 Inversion Recovery (T1IR) pulse sequence that was developed by Torchia[105] is often used; the pulse sequence is given in Figure 2.8.

The CP T_1 Inversion Recovery sequence utilizes the CP efficiency and sensitivity boost to measure the T_1 of the ^{13}C (or other X) spins. In addition, it is only necessary to wait the ^{1}H relaxation time rather than the full ^{13}C T_1 between experiments, which

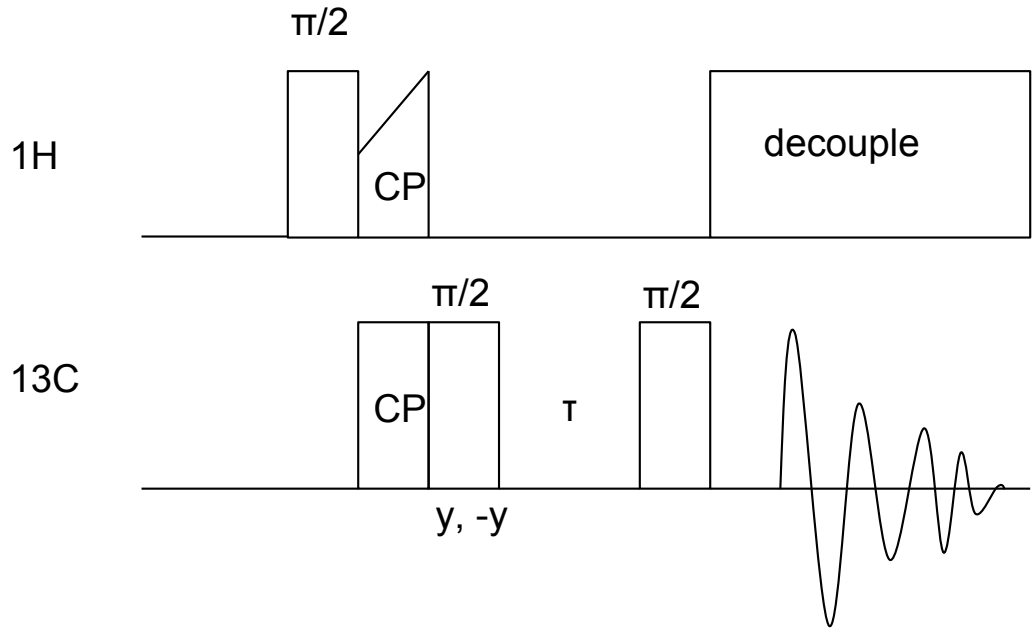

Figure 2.8: ^{13}C CPMAS T1 Inversion Recovery pulse sequence, developed by Torchia.[105] The "y,-y" under the ^{13}C $\frac{\pi}{2}$ indicates a phase change; for the first iteration, the $\frac{\pi}{2}$ pulse is along the y direction; in the second, it is along the -y direction.

greatly speeds up the acquisition process.

The sequence starts with a basic CP block, with the initial $\frac{\pi}{2}$ (90° pulse on the ^{1}H channel flipping the ^{1}H magnetization into the transverse plane; the successive spin lock keeps ^{1}H magnetization in the xy-plane, while a simultaneous RF field is applied to the ^{13}C spins, establishing a Hartmann-Hahn match and transferring polarization from the protons to the ^{13}C nuclei.[35] Immediately, another 90° pulse along the y-axis is applied to the ^{13}C spins, flipping the magnetization back onto the z-axis. The period immediately following this manipulation is then defined as $\tau = 0$ (where τ is the variable delay parameter n the experiment) and the corresponding CP-enhanced longitudinal magnetization (along z) at this initial state is denoted $M_{cp}(0)$. The delay period is incrementally set to longer and longer times in an arrayed set of experiments, to observe the loss of CP-longitudinal magnetization due to the ^{13}C magnetization

returning to its equilibrium value (that is, $M_{cp}(\tau)$ goes toward its equilibrium value, M_0). After the delay period, another ^{13}C 90° pulse brings the magnetization back into the xy-plane, where it can be observed. This is observed as a loss of intensity of the peak of interest with longer times τ. The time constant for this exponential decay of magnetization along z is denoted T_1.

Assuming that there is a little bit of ^{13}C z component magnetization that can arise from ^{13}C spin lattice interactions, denoted $M_A(\tau)$, we can write the full form of the total ^{13}C z magnetization as

$$M_{z1}(\tau) = [M_{cp}(0) - M_0] \exp\left(\frac{-\tau}{T_1}\right) + M_0 + [M_A(\tau)]_z \tag{2.21}$$

That is for the first iteration. The second iteration of the experiment for the same τ value is exactly the same, except that the ^{13}C $\frac{\pi}{2}$ pulse is along the -y direction (180° phase change) rather than the +y direction, as previously. Instead of putting the transverse magnetization onto the + z-axis, it is now flipped to the - z-axis, but otherwise, the rest of the sequence proceeds as normal. The other interactions that result in $M_A(\tau)$ are considered to be identical to the first sequence. Thus,

$$M_{z2}(\tau) = [-M_{cp}(0) - M_0]\, exp\left(\frac{-\tau}{T_1}\right) + M_0 + [M_A(\tau)]_z \tag{2.22}$$

Subtracting Equation 2.22 from 2.21, we get a net signal of

$$N M_{cp}(0) \exp\left(\frac{-\tau}{T_1}\right) \tag{2.23}$$

where N is the number of scans taken, and must be an even number. From Equation 2.23, we can see that the unwanted elements $M_A(\tau)$, M_0 are eliminated from the final signal.

The data is reviewed as an array of spectra - the intensity should decay with increasing time τ. The exponential decay curve is fit to

$$I(t) = I(0) \exp\left(\frac{-\tau}{T_1}\right) \tag{2.24}$$

where $I(t)$ indicates peak intensity, to extract the T_1 constant for the sample. If the individual peaks in a spectrum are well-resolved, it is possible to extract separate relaxation constants for each individual peak.

Chapter 3

BIOSILICIFICATION AND THE R5 PEPTIDE

The R5 peptide is a fragment of the larger *sil1p* protein with the sequence:

n-SSKKSGSYSGSKGSKRRIL-c

Figure 3.1: Primary sequence of the R5 peptide, a repeat fragment of the *sil1p* protein.

It is one of seven repeating segments of the *sil1p* protein, extracted from the diatom *Cylindrotheca fusiformis*, and precipitates silica at neutral pH.[14] In the first studies on this system, Kröger et al. monitored peptide-induced silica precipitation and found that the Silaffin-1A protein (a segment of *sil1p*)[100] is incorporated into the growing silica network.[53]

3.1 The R5 Peptide Forms Biosilica Composites

The protein that was first extracted from *Cylindrotheca fusiformis* (deemed Silaffin-1A) was treated with an HF solution. However, silaffins contain high numbers of lysines and hydroxyamines, typically candidates for post-translational modification of the side groups. However, the HF cocktail used to extract the protein from diatom cell walls also cleaves any modified functionalized groups (O-glycosidic and phosphate ester bonds)[69] a protein might have. In order to preserve any HF-sensitive regions of the silaffin protein, a gentler ammonium fluoride solution was used instead. This extraction preserved a segment of the silaffin protein's modified side chains, prompting the name natSil-1A for this native form protein.[53] In natSil-1A, a smaller, 6.5 kD peptide, lysine residues have modified side chains consisting of long-chain polyamines;

serines are phosphorylated. The original silaffins (Silaffins-1A and -1B, with 4 kD and 8 kD molecular weights, respectively, originating from endoproteolytic processing of *sil1p*) lack either of these modifications.

The natSil-1A protein is believed to provide a template for silicic acid polycondensation; in early stages, a mildly acidic solution containing both the peptide and silicic acid produced a branched network of intertwined silica and peptide; the presence of both the silica and peptide in the resulting composite was confirmed by elemental analysis. The branches pinch off into spheres of silica, which grow to diameters between 400-700 nm by the end of the ten-minute procedure. It is speculated that the zwitterionic nature of the natSil-1A peptide promotes self-aggregation, establishing a rigid framework that induces silica precipitation from solution.[53] Silica precipitation induced by the R5 fragment proceeds in similar fashion, although at neutral pH rather than mildly acidic pH (Appendix A),[14] and as the R5 peptide also lacks post-translational modification of its lysine and serine side chains, the presence of phosphate is necessary for silica activity.[84]

Compared to the natSil-1A protein, the R5 peptide is simpler; it has seven different residues and contains no post-translationally modified sidechains or functional groups. It still contains basic residues (K, R) and seven serines (S), which are shown to be common in diatom and sponge proteins.[54] Further biosilica studies have focused on the R5 peptide; Naik et al. tested the influence of applied mechanical stresses on R5-precipitated silica structures[72], and their findings, as well as those by Sumper et al.,[100] showed that conditions can be altered both chemically (e.g., peptide concentration, presence of multivalent anions) and mechanically (e.g., applied shears, streams of N_2 gas) to induce changes in peptide-controlled silica formation.

In order to try to tie specific residues to the silica precipitation process, Knecht et. al. tested various truncates and mutants of the R5 peptide for silica precipitation activity.[50] It was shown that the RRIL motif at the end of the R5 peptide appears to be integral to inducing silica precipitation; without it, truncates demonstrated

Table 3.1: Silica Precipitating Activity of R5 Truncates and Mutants, adapted from Knecht et. al. [50]

Peptide	Sequence	Specific Activity (nmol silica / min*nmol peptide)	Particle Size (nm)
1	SSKKSGSYSGSKGSKRRIL	3.59 ± 0.16	250-450
2	SSKKSGSY	0.08 ± 0.05	n/a
3	SGSKGSKRRIL	3.35 ± 0.25	180-400
4	KSGSYSGSKGSKRRIL	3.29 ± 0.21	125-200
5	SGSKGSKRR	2.70 ± 0.19	150-300
6	SSKKSGSYSGSKGSK	1.09 ± 0.23	85-130
7	LIRRSSKKSGSY	3.17 ± 0.22	60-300
8	SSKKSGSYRRIL	2.88 ± 0.32	60-300
9	SGSKGSKAAIL	0.24 ± 0.16	n/a
10	SGSKGSKEEIL	0.17 ± 0.08	n/a
11	SGSKGSKNNIL	0.16 ± 0.07	n/a

significantly reduced activity. When the RRIL motif was appended to other fragments of the sequence, silica activity returned to higher rates. Similarly, a truncate of the first three amino acids (SSK motif) showed silica activity levels nearly equal to that of the full R5 peptide. Table 3.1 gives the full list of truncates and mutants investigated by Knecht et al.

Other studies have highlighted the silica precipitating activity of polyamines and simple primary amines [51], polylysine [56], polyserine [58] and polyhistidine [57], suggesting that specific amino acid side chains do play a pivotal role in inducing silica precipitation and shaping the resulting silica particles. However, most work up to now has focused on examining protein and/or organic functional group interactions with silicic acid and silica at the macroscopic level;[53, 52, 50, 49] atomic level insight on the structures and behavior of peptides like R5 within the biosilica composite, could help further refine understanding of the complex relationship between biosilicification proteins and biosilica.

We focus the atomic level study around these questions:

- What is the secondary structure of the R5 peptide within the biosilica composite, and how does this differ from that of the free, neat peptide?
- What amino acid side chains and functional groups are integral to interaction with and control of silica?
- How are these amino acid side chains arrayed?
- What secondary or tertiary motifs might be utilized for the purpose of templating silica?

It is the subject of this chapter to address these questions as best as possible, through analysis via SSNMR and other methods.

3.2 Secondary Structure Characterization by SSNMR

The first step in understanding the protein's interactions with silicic acid and with silica is to assess the secondary structure of the R5 peptide. The biosilica-peptide composites are roughly 25-30% peptide by weight; this is a sufficient amount for detection with SSNMR. With SSNMR techniques, the secondary structure can be solved for the neat peptide, and then compared to the structure of the peptide within the biosilica composite, to see if there are any changes in peptide conformation when R5 co-precipitates with silica. There have been several models proposed for the tempting of silica structures onto polymers like poly-L-lysine,[107, 63, 36, 104, 78] which assume that the structure of the peptide or polymer in question remains the same throughout interactions with silica. However, other examples of proteins that bind to surfaces, like statherin,[73, 74, 62, 93] suggest that the more typical case involves a conformation change of the protein upon adsorption to mineral moieties. To validate any theories about aggregation schemes and methods of protein self-assembly, it is first necessary to determine whether the secondary structures of the neat and complexed forms of R5 (i.e., if any ordered structure prevails) are comparable.

In order to investigate the secondary structure of R5 with SSNMR, isotopically labeled amino acids had to be incorporated into the peptide sequence, so that examination of ^{13}C and ^{15}N spins are possible. To ensure that backbone ^{13}C shifts for particular amino acids can be unambigulously assigned, it's necessary to limit uniform labeling to 2-3 amino acids per sample, at most.

The basic strategy for determining the secondary structure and conformation of the neat and -SiO_2 R5 peptides was as follows:

1. Obtain chemical shift information with SSNMR 1D ^{13}C and^{15}N CPMAS, and 2D ^{13}C experiments (outlined in Chapter 2). It is well established that backbone chemical shifts (^{13}CO, ^{13}Cα, ^{13}Cβ, ^{15}N) are sensitive to local secondary structure.[111, 110, 86] It is also known that the side chain chemical shifts (i.e. ^{13}Cγ, ^{13}Cδ, etc.), while not directly correlated to backbone structure, can be perturbed by proximity to negatively charged silica surfaces.[45, 7, 8] Thus, chemical shift perturbations in the side chain ^{13}C spins can also tell us something about how the peptide and its side chains are situated relative to silica moieties. Also, since chemical shifts across several residues can sometimes overlap in a single 1D spectrum, it is necessary to use 2D correlational experiments like DARR[102, 103] to assign backbone ^{13}C shifts with any measure of certainty.

2. Submit experimentally assigned chemical shits into TALOS-N[94], which outputs torsion angle predictions for submitted sequences by drawing upon a database of high resolution protein structures.

3. Use the predicted torsion angles to generated peptide secondary structures using a molecular modeling program like Chimera.[81]

This results in a secondary structure for both the neat (lyophilized out of a controlled concentration solution) and the -SiO_2 R5 peptide; the two can then be compared, to identify any changes in conformation that might signify a targeted protein

response to silica, or any other molecular interaction mechanisms that could be involved in the silica precipitation catalyzation and shaping process.

3.2.1 Peptide Synthesis and Materials

Uniformly ^{13}C and ^{15}N labeled R5 peptides were synthesized de novo on a Rainin PS3 instrument using standard Fmoc chemistry (Appendix A).[3] Uniformly labeled, protected amino acids and Fmoc-Leu-Wang resin were purchased from Cambridge Isotope Laboratories (Andover, MA). Unlabeled amino acids were purchased from Sigma-Aldrich in protected form. Protection schemes for labeled isoleucine and serine are given in Appendix A. Other labeled amino acids were purchased in protected form, as necessary.

The complete set of synthesized, isotopically labeled R5 samples are given in Table 3.2.

Table 3.2: Uniformly labeled ^{13}C and^{15}N R5 peptide samples

Labels	Sample Name
SSKKSGSYSGSKG*S*K*RRIL	GSK-c
SSKKSGSYSG*S*K*GSKRRIL	GSK-m
SSKK*S*G*SYSGSKGSKRRIL	KSG
SSKKSGSYS*GSKGSKRR*I*L	SRI
S*SKKSGSYSGSKGSKRRIL	S
SS*K*KSGSYSGSKGSKRRIL	SK
SSKKSGS*YSGSKGSKR*RIL	SR
SSKKSGSYSGSKGSKR*RIL	R

Crude peptides were purified by reverse-phase HPLC and sequence confirmation was LC-MS. Neat R5 peptides were lyophilized out of 3 mM solutions. The R5-SiO_2 peptide composites were precipitated out of buffered (phosphate-citrate, 100 mM),

3 mM solutions of R5 with the addition of 1 M $Si(OH)_4$, which was prepared from TMOS as in Appendix A. The TMOS was purchased from Sigma-Aldrich.

R5-SiO_2 precipitates (the combination of the R5 peptide embedded in a silica matrix, resulting in an inorganic-organic composite compound) were imaged by SEM (FEI Sirion SEM, UW NTUF), as shown in Figure 3.2. Resulting nanospheres were approximately 500-750 nm in diameter, and roughly uniform in shape and size. These are extremely similar to the nanospheres that are precipitated with the parent protein, *sil1p*.[53]

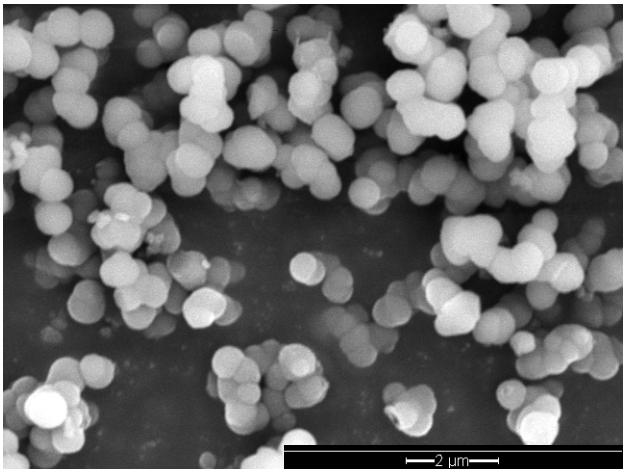

Figure 3.2: SEM image of R5-precipitated silica nanospheres. The spheres are mostly uniform in size and shape, with diameters ranging from 500-750 nm as found by Knecht et al. [50] The inset is the scale bar. R5-SiO_2 refers to the R5 peptide and silica composite through the remainder of this work.

3.2.2 SEM Imaging

SEM experiments were performed on a FEI Sirion scanning electron microscope; dry precipitates were dispersed onto double-sided sticky carbon tap and mounted on aluminum studs,and sputter-coated for 60 s with Au/Pd. Images were acquired with accelerating voltages of either 5 or 10 kV.

3.2.3 Experimental Methods

All SSNMR experiments were performed on a Bruker Avance III spectrometer, at a field of 700.18 MHz (^{1}H), fitted with a ^{1}H{^{13}C, ^{15}N} 3.2 mm MAS probe. The Larmor frequency for ^{13}C at this field is 176.07 MHz, and 70.95 MHz for ^{15}N. The spinning speed for all experiments was 15±2 kHz, regulated by a Bruker automated MAS controller. All samples were run at room temperature (∼22 °C).

The basic CP pulse sequence was given as Figure 2.3 in Chapter 2. The ^{13}C channel can be interchanged with ^{15}N as necessary. Parameters and times for ^{13}C experiments were optimized on adamantane (Sigma Aldrich), which also served as a secondary standard. The secondary standard used for ^{15}N was isotopically ^{15}N labeled ammonium sulfate, purchased from Cambridge Isotope Laboratories (Andover, MA); ammonium sulfate was also used for parameter optimization on ^{15}N.

For 1D ^{13}C CPMAS experiments, 1024 scans were run on both the neat and -SiO_2 versions of labeled R5 peptides. ^{1}H 90 pulse length was 2.75 μs, and the contact time was 1100 μs. The recycle delay was 2 s for all experiments. For 1D ^{15}N CPMAS experiments, 4-10K scans were taken; ^{1}H 90 times were again 2.75 μs, with a contact time of 500 μs. All ^{13}C experiments were referenced to adamantane as a secondary standard (relative to TMS). All ^{15}N experiments were referenced to labeled ^{15}N ammonium sulfate.

For 2D ^{13}C DARR, experiments were again carried out at room temperature, with a Bruker MAS unit regulating the spin speed to 15±2 kHz. The basic DARR experiment is given in Figure 2.6 in Chapter 2. In general, the ^{13}C-^{13}C DARR experiments were run with a 1100 μs contact time, a 2.75 μs ^{1}H 90 and a 4 μs ^{13}C 90 pulse length, and DARR continuous wave (CW) irradiation during the mixing period τ, with a ^{1}H RF field strength ν_1 equal to the MAS rotation frequency, ν_r ($n = 1$). Mixing times were typically 20 ms.

3.2.4 Chemical Shift Assignments and Results

1D ^{13}C and ^{15}N CPMAS experiments were performed in conjunction with 2D ^{13}C correlational DARR experiments in order to make unambiguous chemical shift assignments for residue triplets and pairs in the R5 sequence. Amino acid structures are shown in Figure 3.3; all amino acids exist in their L conformer.

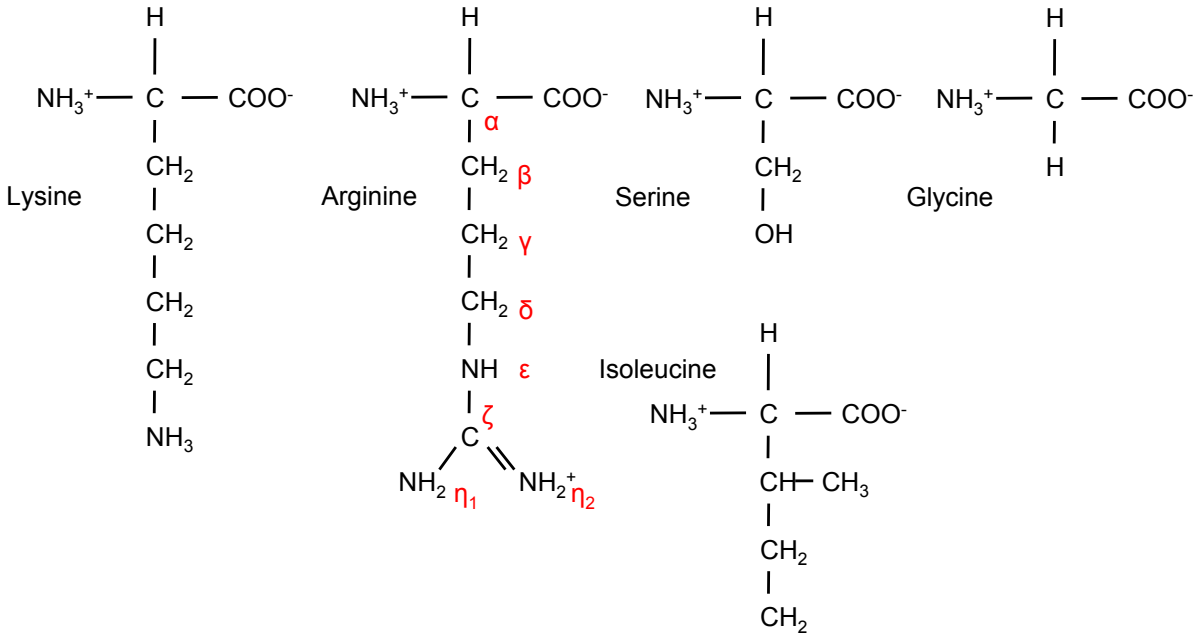

Figure 3.3: Structures of Amino Acids that are present in the R5 peptide. K = Lysine, R = Arginine, S = Serine, I = Isoleucine, G = Glycine. Structures are given at neutral pH = 7, so basic residues (K, R) have protonated amine groups. Also shown on arginine are conventional Greek letter side chain atom labels.[75]

1D ^{13}C spectra of the R5 samples are shown as a stacked plot in Figures 3.4 and 3.5. Viewing it this way, direct comparisons can be made between the ^{13}C 1D spectrum of the neat R5 peptide versus the R5-SiO_2 peptide. Chemical shift variances become more obvious, though for the most part, the general line shapes are similar between the two variations of peptide. Losses in resolution and decreases in peak intensity in the -SiO_2 peptide versus the neat version are due to the smaller amount of peptide

available in the silica composites for the same total sample volume. The downfield region (left end of the spectrum) contains the carbonyls, which typically fall around ∼170 ppm, and the upfield region contains the rest of the ^{13}C spins (^{13}Cα up through ^{13}Cζ). In Figure 3.4 it can be seen that the S and G carbonyls are mostly overlapping; the G ^{13}Cα can be distinguished at ∼40 ppm, and the S and K ^{13}Cα resonances are separated at around 52 and 50 ppm, respectively. S ^{13}Cβ is also identified at roughly 60 ppm, but the K ^{13}Cβ is buried in a large section of overlapping peaks in the region before 50 pppm. This illustrates the necessity of 2D spectroscopy.

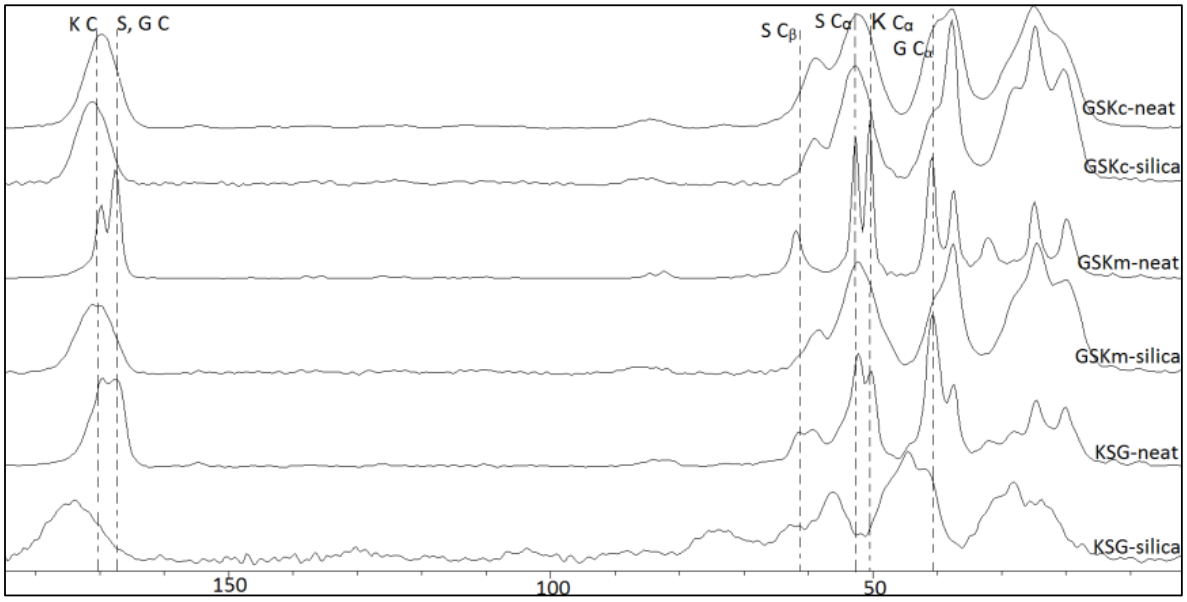

Figure 3.4: 1D CPMAS ^{13}C spectra of the GSK-c, GSK-m and KSG labeled samples of the R5 sequence, both neat and in SiO_2. Dashed lines guide the eye down peak centers and highlight chemical shift deviations between the two peptides (neat (lyophilized) vs. -SiO_2). The chemical shift scale (x-axis) is in ppm.

The ^{15}N spectra for R5 R, SK, and SRI samples are given in Figure 3.6. Arginines and lysines (R, K, respectively) have side chain amines, which give rise to an additional ^{15}N peak besides the characteristic backbone amide at ∼120 ppm. Arginine has three side chain nitrogens (a guanidinium group, with the ^{15}Nϵ showing up at ∼ 87 and the

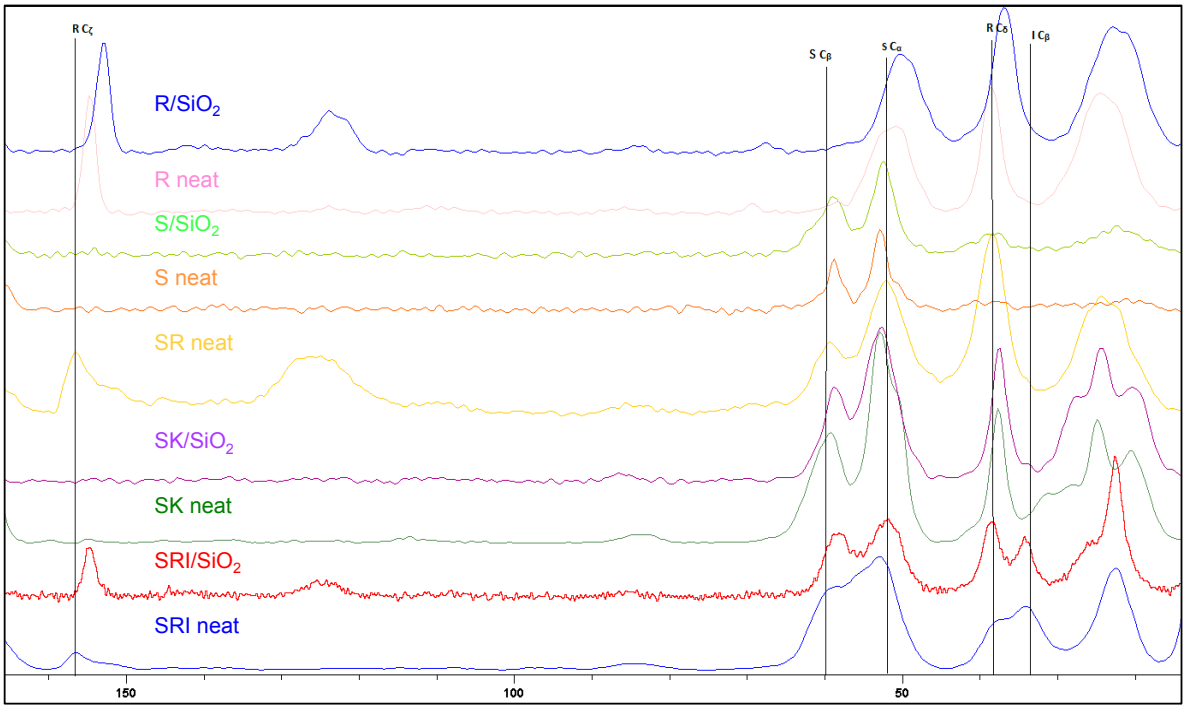

Figure 3.5: Additional 1D CPMAS ^{13}C spectra of the R5 sequence, in neat form and in SiO_2, shown as a stacked plot of the remaining four out of eight samples listed in Table 3.2. Lines guide the eye down chemical shift assignments. The chemical shift scale (x-axis) is in ppm.

equivalent $^{15}N\eta_1$, η_2 at 74 ppm), and lysine has one (ammonia group at $\sim$ 35 ppm). As seen in Figure 3.3, $^{15}N\eta_1$ and $^{15}N\eta_2$ are considered equivalent, so they both show up at the same resonance frequency. The backbone amides all occur around the $\sim$ 120 ppm region; using peaks from the singly labeled samples (i.e., individually R- and S- labeled samples in Table 3.2) as references, it is possible to identify individual shoulders of peaks in the ^{15}N backbone amide region in the remaining samples.

The ambiguities discussed previously and that can be seen in the 1D ^{13}C CPMAS spectra (Figures 3.4 and 3.5) demonstrated the necessity of a 2D ^{13}C-^{13}Ccorrelational DARR experiment, to show cross peak interactions of ^{13}C spins that are close in space. This introduces a logical sequence of ^{13}C spins to follow in making sequential

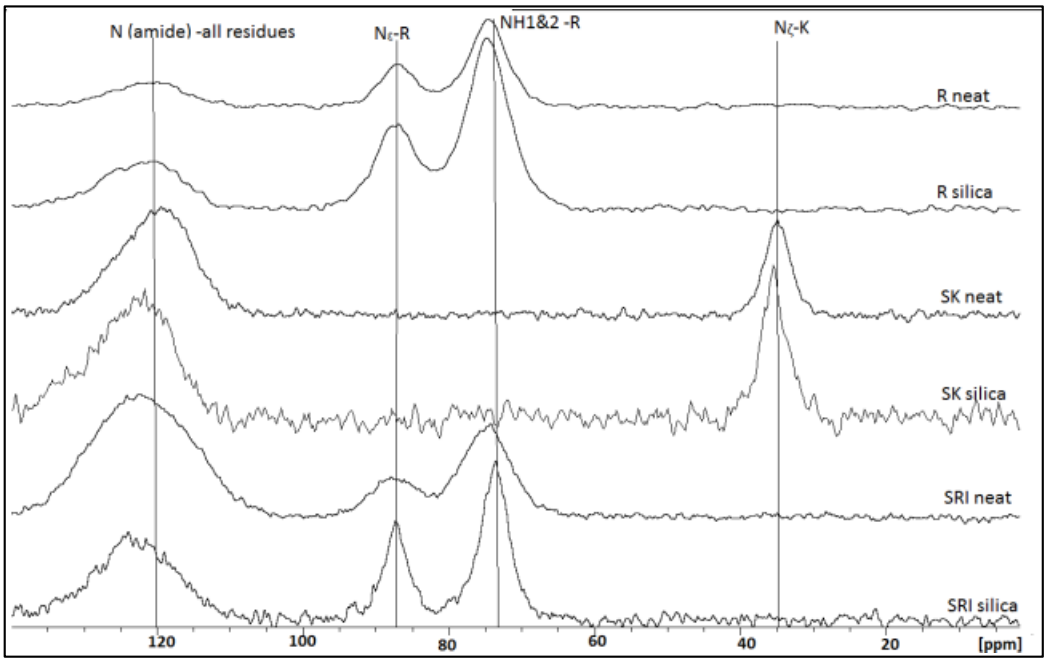

Figure 3.6: 1D CPMAS ^{15}N spectra of R5 R, SK, and SRI samples for neat and -SiO_2 peptides. Amino acids R and K contain side chain nitrogens, which are visible here at 87 and 74 ppm (R) and 35 ppm (K); these are easily distinguished from the amides in the backbone. For reference, the amino acid structures are given in Figure 3.3.

chemical shift assignments for individual residues, even in triply-labeled samples like GSK-m, where there is a lot of spectral overlap (Figure 3.7). The diagonal peaks have the same frequency coordinate on each axis, and run along the diagonal center of the spectrum. The cross peaks show interactions of spins, correlating a spin of frequency x with a spin of frequency y, and appearing as a contoured intersection of the two frequencies.

In Figure 3.7, the neat and -SiO_2 R5 peptide GSK-m spectra are superimposed in a single plot, to highlight the differences in chemical shift between the same residues in the neat versus the -SiO_2 versions of the sequence. The colored, labeled lines in the figure guide the eye along the sequence of ^{13}C spins along amino acids G10, S11, and K12, separately. Glycine typically has a $^{13}C\alpha$ shift that is $\sim$ 10 ppm upfield

of corresponding α carbons in either S or K (here, G10 has a $^{13}C\alpha$ chemical shift of 40.840 ppm, versus S11's 52.602 ppm and K12's 50.447 ppm), which makes it easier to pick out. Once the G10 $^{13}C\alpha$ is determined, it can be followed to its cross peak interaction with the corresponding G10 ^{13}CO peak. That eliminates one of the carbonyl resonances from the assortment of ^{13}CO region diagonal and cross peaks, and we have assigned both carbons in G10 (glycine has a side chain of a single proton, as shown in Figure 3.3). The next step is to pick out a salient feature of either of the remaining residues. The lysine $^{13}C\beta$ is distinguishable at 32.025 ppm. The same scheme used with G10 can now be used again to assign all of the K12 ^{13}C spins, and finally, repeated once more to identify and assign the S11 ^{13}C spins.

The complete set of ^{13}C and^{15}N backbone chemical shifts derived from 1D and 2D SSNMR experiments with these samples are in Tables 3.3 (neat R5) and 3.4 (R5 in SiO_2).

3.2.5 ^{29}Si CPMAS Spectra of R5-SiO_2 Particles

In addition to SSNMR investigation of dynamics and structure of the peptide, I also did ^{29}Si CPMAS experiments, performed on a Bruker Avance III spectrometer, at a field of 700.18 MHz (^{1}H). This spectrometer was fitted with a 1H-X-Y dual X channel MAS probehead, accommodating 2.5 mm rotors. The Larmor frequency of ^{29}Si at 700 MHz is 139.10 MHz. The ^{1}H 90° time was 6.0 μs, the contact time was 9.0 ms, and the recycle delay was 2.5 s. The number of scans was 131072, and the chemical shifts were referenced to Q8M8.

Even with 131072 scans, signal was paltry, suggesting that there is not very much ^{29}Si material in the rotor. A quick, rough calculation, operating on the assumption of roughly 15% by weight silica in the peptide-silica sample, boils down to approximately 0.15 mg ^{29}Si spins in the sample (with the ^{29}Si isotope being $\sim$4% natural abundance). The poor signal at such a high field suggests that this may be a reasonable estimate. The other problem could be loss of CP efficiency, due to a lack of ^{1}H spins. If

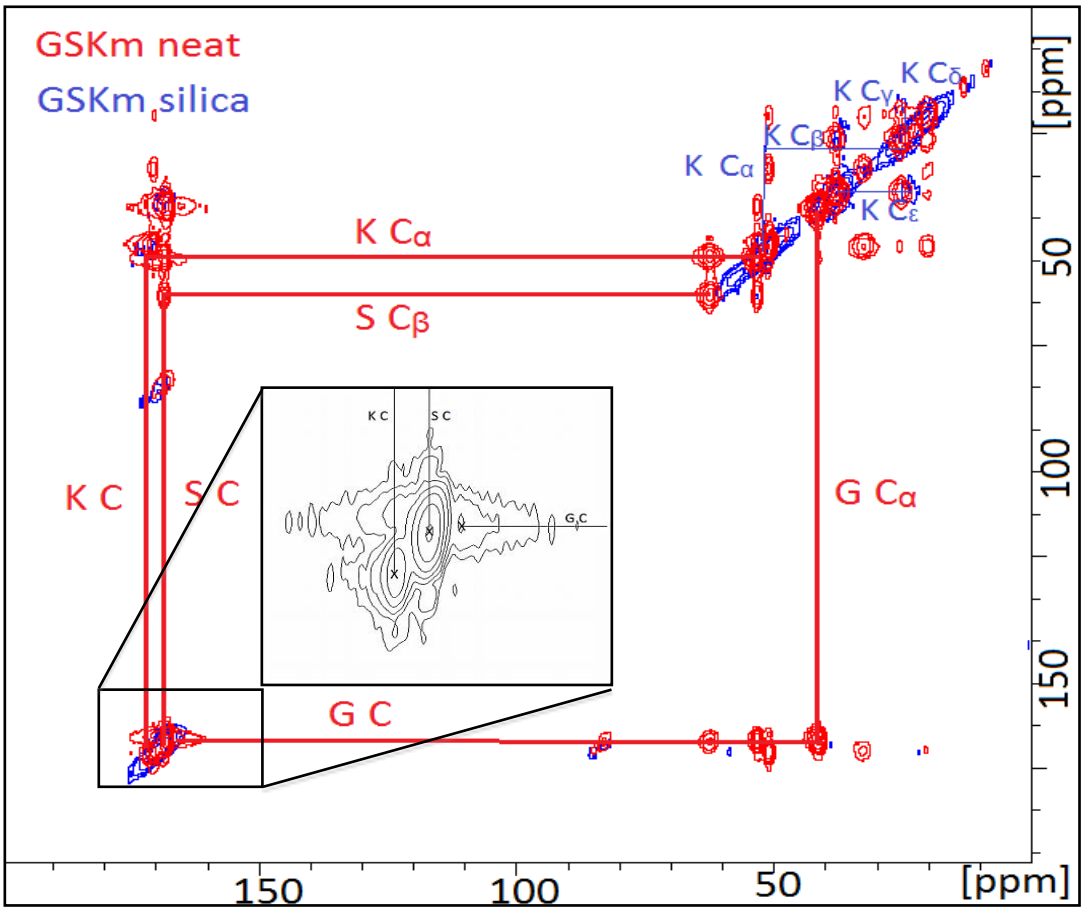

Figure 3.7: Superimposed ^{13}C DARR spectra of R5 GSK-m neat and -SiO_2 peptides. Superposition of the two spectra highlights slight changes in structure and also demonstrates how cross peaks in a 2D spectrum can help elucidate a single peptide structure by clarifying chemical shifts. Reprinted with permission from Roehrich, A.; Ash, J.; Zane, A.; Masica, D. L.; Gray, J. J.; Goobes, G.; Drobny, G., Solid-State NMR Studies of Biomineralization Peptides and Proteins. In Proteins at Interfaces III State of the Art, American Chemical Society: 2012; Vol. 1120, pp 77-96. Copyright 2012 American Chemical Society.[85]

Table 3.3: ^{13}C and^{15}N Chemical shift assignments for residues in the neat R5 peptide, made with 1D and 2D ^{13}C and^{15}N CPMAS and DARR experiments. Some carbons were unassigned due to prevailing ambiguities in the spectra.

Residue	CO (ppm)	Cα (ppm)	Cβ (ppm)	N (ppm)
S1	165.560	52.576	58.726	—
S2	167.690	52.772	59.160	119.086
K3	169.094	37.563	27.896	123.177
K4	170.279	59.257	27.944	122.843
S5	168.529	61.477	64.729	117.593
G6	167.269	40.699	n/a	112.705
S7	169.708	56.448	59.364	118.424
S9	170.200	53.314	58.855	117.382
G10	167.000	40.840	n/a	113.038
S11	167.400	52.602	61.842	116.330
K12	169.778	50.447	32.025	123.581
G13	107.814	39.746	n/a	107.814
S14	167.400	58.877	73.081	116.386
K15	171.286	52.364	28.686	—
R16	170.746	50.723	24.510	119.860
R17	169.600	—	19.820	112.001
I18	164.320	55.453	34.449	123.752

Table 3.4: ^{13}C and^{15}N Chemical shift assignments for residues in the R5 peptide in the SiO_2 matrix, made with 1D and 2D ^{13}C and^{15}N CPMAS and DARR experiments. Some carbons were unassigned due to ambiguities in the spectra.

Residue	CO (ppm)	Cα (ppm)	Cβ (ppm)	N (ppm)
S1	169.386	52.522	59.016	—
S2	173.409	52.784	58.791	121.630
K3	172.400	37.589	27.586	127.466
K4	171.299	52.243	24.109	121.209
S5	169.853	58.790	63.857	114.656
G6	168.069	40.486	n/a	111.451
S9	169.900	51.335	58.087	121.346
G10	169.800	40.151	n/a	112.817
S11	170.600	58.583	62.472	121.789
K12	171.411	55.300	28.141	126.957
G13	169.510	40.285	n/a	112.504
S14	171.042	58.860	—	118.034
K15	172.780	52.563	27.741	125.602
R16	169.455	50.286	22.879	120.868
R17	170.720	—	25.490	123.801
I18	170.346	52.334	34.222	128.196

Table 3.5: R5 Neat ^{13}C and ^{15}N shifts for the side chains in K and R

residue	Cγ (ppm)	Cδ (ppm)	Cϵ (ppm)	Cζ (ppm)	Nϵ (ppm)	Nζ (ppm)	N$\eta_{1,2}$ (ppm)
K3	20.4494	24.7701	31.2877	n/a	n/a	34.5244	n/a
K4	20.1246	24.6429	37.4902	n/a	n/a	34.4525	n/a
K12	19.8296	24.818	37.3928	n/a	n/a	34.5143	n/a
K15	21.4863	24.9863	37.7370	n/a	n/a	35.5915	n/a
R16	22.7507	38.5552	n/a	154.7349	86.9013	n/a	74.5974
R17	12.6515	37.6829	n/a	156.8686	87.4212	n/a	75.0916

Table 3.6: R5-SiO_2 ^{13}C and ^{15}N shifts for the side chains in K and R

residue	Cγ (ppm)	Cδ (ppm)	Cϵ (ppm)	Cζ (ppm)	Nϵ (ppm)	Nζ (ppm)	N$\eta_{1,2}$ (ppm)
K3	20.4036	24.4360	29.5000	n/a	n/a	35.3429	n/a
K4	19.8888	21.5094	38.1290	n/a	n/a	34.5968	n/a
K12	20.2148	24.7795	37.8007	n/a	n/a	34.9738	n/a
K15	20.2043	24.6480	37.5994	n/a	n/a	35.6205	n/a
R16	21.2387	36.9475	n/a	152.8955	86.8126	n/a	74.7103
R17	22.1730	38.4901	n/a	154.5870	87.0335	n/a	74.6358

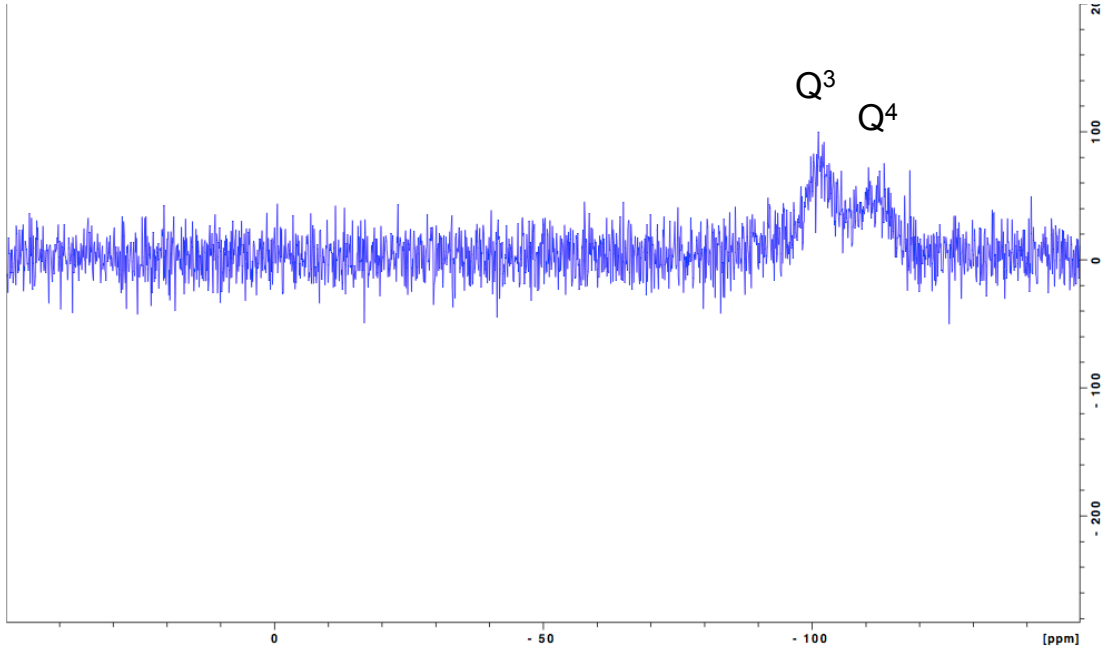

Figure 3.8: ^{29}Si CPMAS spectrum of the R5-SiO^2 composite, showing Q^n groups.

the peptides are buried in the center of the spherical aggregate and the silica forms only an external coating, it is reasonable to believe that the ^{29}Si spins would not be in close proximity to ^{1}H spins in the peptides themselves. However, biosilica is in general $SiO_2{\cdot}xH_2O$,[44, 9] meaning that there should be plenty of water (and therefore ^{1}H spins) bound to the silica particles themselves. This again brings us back to the conclusion that there simply is not a lot of silica in these biosilica-peptide composite particles.

3.3 Discussion

It is hypothesized that within the diatom, silaffin proteins self-assemble into a templating matrix that aids in the catalyzation of silica deposition into the diatom cell walls.[52] Upon satisfying certain conditions (i.e., circumneutral pH, adequate peptide concentration, ambient temperature), addition of silicic acid to buffered R5 peptide

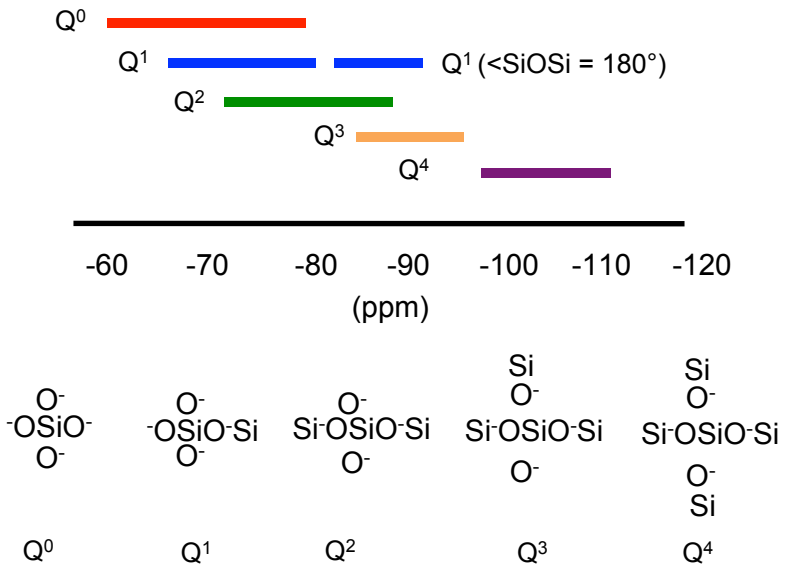

Figure 3.9: Q^n groups and characteristic ^{29}Si NMR chemical shifts. Adapted from reference [25].

solutions results in the formation of highly regular silica nanospheres.[50, 49] These nanospheres are similar in size and shape to the silica nanospheres that are generated through silicic acid interactions with the full parent natSil-1A protein, even though R5 is much smaller, lacks the post-translationally modified lysines and serines, and the presence of phosphate (via buffers, e.g., phosphate citrate or PBS) anions in solution are necessary in the case of R5.[50, 49, 53, 84]

The secondary and tertiary structures of the full natSil-1A are still unknown, but the R5 peptide is easier to synthesize, and the smaller sequence offers improved resolution over the larger protein. Elucidation of the R5 peptide's secondary structure serves as a benchmark point; this native form of the peptide can then be compared to the form it takes on when it forms complexes with silica. Proteins often undergo some kind of conformation change upon adsorption to a mineral surface[73, 74, 62, 93], so it is necessary to determine whether a similar conformational change occurs with R5.

The secondary structure of a peptide also affects the way that the peptide will ag-

gregate or self-assemble in solution; α-helices pack in a different way than β-strands, for example. If aggregation of the R5 peptide is indeed preemptive to silica precipitation, knowledge of the secondary structure will constrain further speculation on self-assembly models, and therefore on additional silica scaffolding and templating schemes, as well.

3.3.1 Comparison of the Secondary Structure of the neat and -SiO_2 R5 Peptide

Following the scheme outlined in Section 3.1, the experimentally obtained chemical shifts are input into TALOS-N, and the resulting predicted torsion angles are submitted into a molecular graphics program (here we use Chimera[81]), to generate graphical predictions of the R5 peptide in its neat (lyophilized) form, and in its biosilica form.

It must be said that this is operating on the assumption that the primary contributors to chemical shift perturbations in the backbone ^{13}C spins (ΔCS, Figure 3.11) are changes in conformation, which may not necessarily be the case; thus, while this is a fair prediction of secondary structure of R5 in the silica state, it does not account for aggregation effects or the proximity of silica or phosphate groups in all cases.

Still, a comparison of the two generated structures (Figure 3.10) does give a basic idea of the changes in the local environment of the peptide, as well as the general conformation of the R5 peptide, which seems to be mostly extended for maximum contact either with other R5 peptides and side chain functional groups, phosphate anions, or with silica moieties. It is also helpful to look at the ΔCS plots (Figures 3.11, 3.12) to see absolute differences between the two forms of R5 that are studied here.

Because R5 does not precipitate silica in the absence of phosphate anions in solution,[84], it appears that phosphate anions are necessary for self-assembly and aggregation of the R5 peptide. Elsewhere in biological systems, the RR motif (an arginine pair) occurs where binding to phosphate groups is necessary, as in the RR

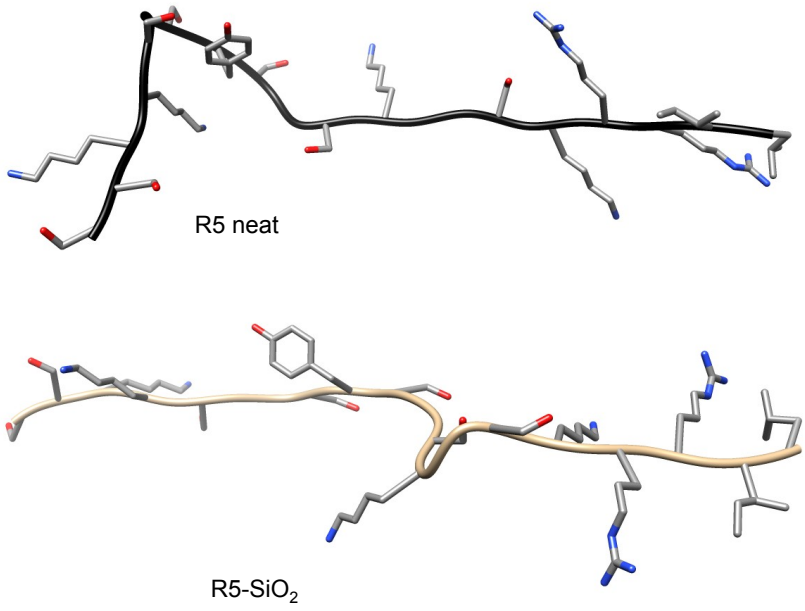

Figure 3.10: The torsion angles predicted by TALOS-N[94] were input into a molecular graphics program called Chimera[81] to generate predicted molecular structures for the R5 peptide, both in its neat form and when it is complexed with silica.

pair of the Tat protein binding to TAR RNA[113, 87, 112, 91] or with cell-penetrating peptides and lipid phosphate head groups[27, 90]. The charged guanidinium group of arginine shows an affinity for negatively charged phosphate interactions through both hydrogen bonding and electrostatic interactions.[91, 97] Single arginine interactions with phosphate groups are seen as well, as with the arginine in penetratin, a CPP that interacts with lipid phosphates.[97] This last work also demonstrates the preference of phosphate groups for arginine rather than lysine at ambient temperatures and pH, though both are shown to be in close proximity to phosphate groups via $^{13}C^{31}P$ REDOR.[97] It is also worth noting the work of Mitchell et al.[67], which shows that when lysines are substituted for wild-type arginines in CPPs, cellular uptake of the mutants become progressively weaker with increasing substitutions of lysines for arginines.

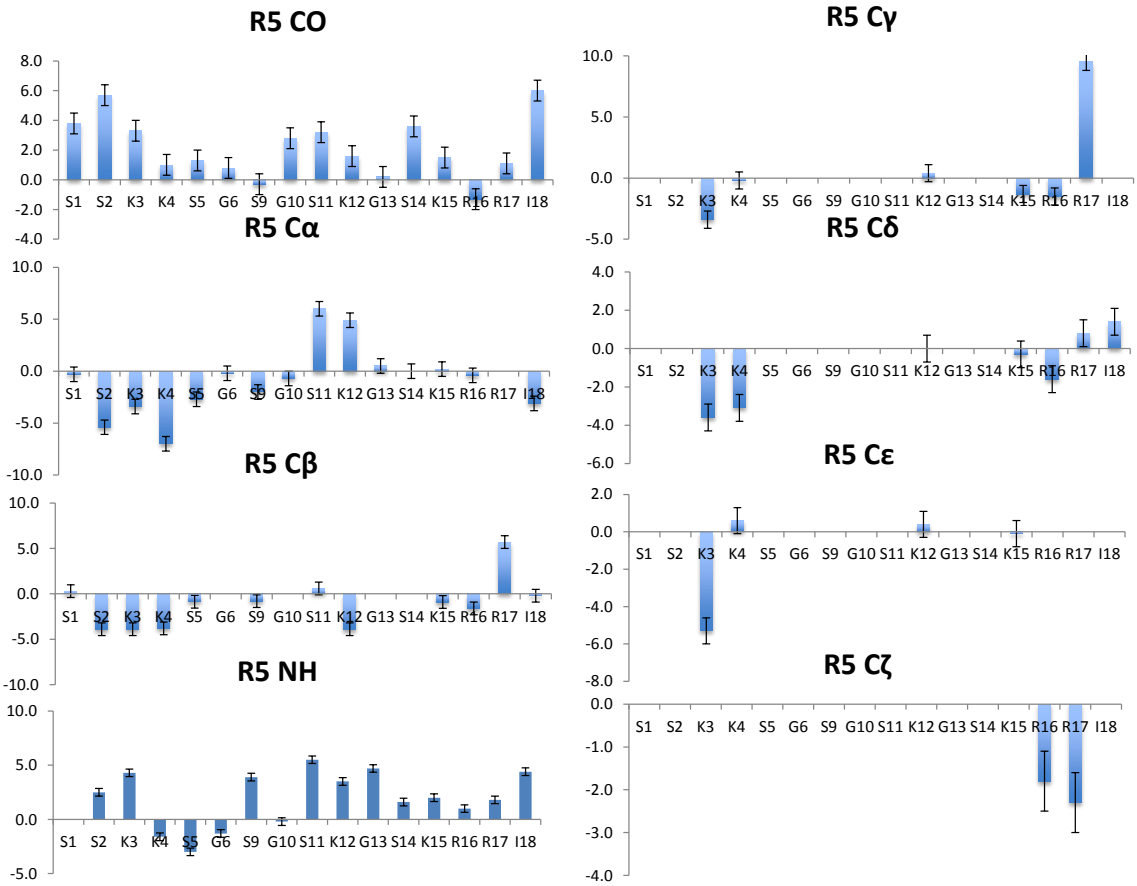

Figure 3.11: ΔCS plots for R5 neat versus -SiO_2 peptides. These show the perturbations of chemical shifts of backbone (left) and side chain (right) ^{13}C shifts in the R5 peptide that is co-precipitated with SiO_2, referenced to the neat peptide (technically, $^{13}C\beta$ is considered part of the side chain, but as the chemical shift contributes to torsion angle predictions, it may be more relevant to include it in discussions of the backbone). The side chains are primarily lysine and arginine (glycine has only hydrogens and serine has an OH attached to its β carbon) and a single isoleucine. Vertical scale is in parts per million. Positive changes indicate that the R5-SiO_2 had downfield shifts relative to the neat peptide; negative changes indicate upfield shifts relative to the neat peptide.

It is possible that either one or both of the arginines in the C-terminal RR pair of the R5 peptide could act in the same manner, binding phosphate ions in the buffered

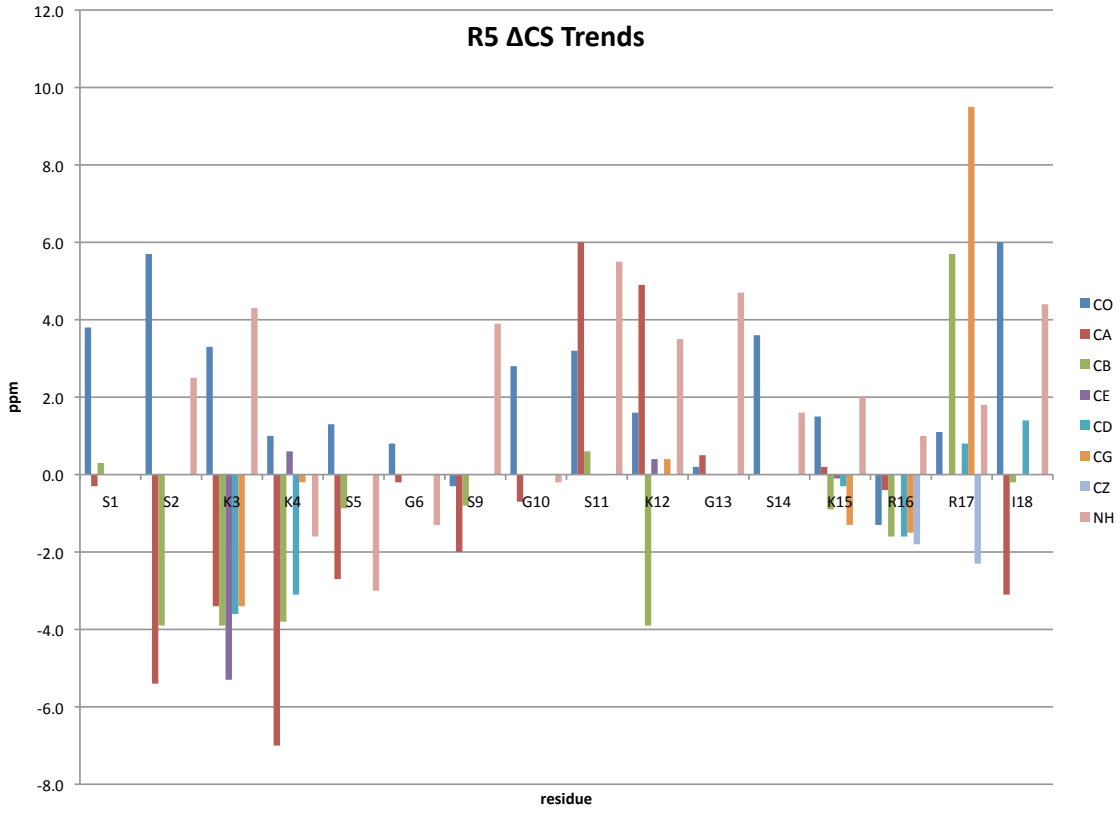

Figure 3.12: Alternate view of ΔCS histogram for the R5 neat vs. R5-SiO_2 peptide

solution and creating an effective salt bridge between adjacent R5 peptides - that is, an intermolecular bond. This could potentially be the basis of R5 self-assembly in solution and the basis of silica templating.

The neat R5 peptide here is a lyophilized peptide; R5 is dissolved in deionized water and then vacuum freeze-dried. Since this initial peptide solution lacks phosphate anions, it is assumed that no ordered self-assembly occurs. Conversely, the R5-SiO_2 consists of R5 peptide that has been dissolved in a phosphate citrate buffer (Appendix A), mixed with 1 M silicic acid, spun down, rinsed (with water, to remove excess peptide and to effectively quench the silicification process at that time point), and dried. The differences between the two forms, then, include the presence of both

phosphate and silica in one (R5-SiO_2), but not the other (R5 neat).

Thus is it is possible that the neat peptide, lyophilized out of an aqueous solution, does not exhibit effects of ordered aggregation. Looking at the proposed structure of the neat R5 peptide (Figure 3.10), the turn at the backbone occuring around residues K4 and S5, appears to straighten out in the -SiO_2 peptide, presumably flattening out to maximize interactions with other R5 peptides, silica, and phosphate groups, as previously hypothesized (though not necessarily in that order). Most notably, S5's torsion angles changes from being in the α-helical regime (-30° > ϕ > −90°, -10° > ψ > −80°) to the β-strand regime $\phi < 0°$, $\psi > 90°$); likewise, the general trend for conformational changes along the R5 peptide seems to be an extension (i.e., flattening) of the peptide. This may be a form of self-assembly, resulting in the scaffolding motif seen elsewhere in biomineralization.[57, 78, 114, 118, 34].

However, it is more probable that direct interactions with inorganic groups would happen at the side chains rather than the backbone itself; if the backbone of the R5 peptide has minimal contact with any surface or inorganic phases the bulk of the changes in chemical shift along the backbone can be attributed to structural rearrangements. Conversely, chemical shift perturbations in the side chains can be attributed to interactions with other R5 molecules, phosphate ions, or silica/silicic acid molecules.

The largest chemical shift perturbations along the backbone tend to occur toward both terminii: - 7.0 ppm for the $^{13}C\alpha$ at K4, and 6 ppm for the ^{13}CO at I18, suggesting that the ends are both more mobile, or have to make bigger movements in the accommodation of phosphate ions and induced aggregation in moving from the neat form to the form observed in the biosilica composite. However, significant chemical shift changes do occur throughout the entire length of the backbone (even if they are more exaggerated at the ends of the sequence), suggesting that aggregation affects the entirely chain almost uniformly, rather than being more biased toward one end versus the other.

The idea that the secondary structure of the R5-SiO_2 peptide must be slightly different than that of the lyophilized neat peptide is sound; these changes in conformation can be induced by self-assembly or simply be the effect of hydrogen bonding to phosphate groups, which in turn induces aggregation of the R5 units. In general, the peptide seems to adopt a mostly extended structure in both cases (Figure 3.10); there does seem to be an exaggerated extension in the case of the R5-SiO_2 peptide, which, as discussed earlier, can be optimal for maximizing interactions with inorganic molecules or other R5 peptides.

3.3.2 Role of R5 Aggregation in Silica Catalyzation and in Shaping Silica Nanospheres

Mizutani et al. showed that polylysine and polyarginine formed silica gels when introduced to solutions of silicic acid, but single amino acids or monoamines did not,[68], which emphasizes the importance of aggregation in silica condensation and precipitation. In addition, the work by Knecht et al. with R5 truncates demonstrates that without peptide aggregation (facilitated by the presence of arginines in the sequence), no silica precipitation occurs.[50] Further work by Rodríguez et al.[84] showed that no R5 aggregation or precipitation occurs in the absence of phosphate (which was mirrored by some of our own preliminary tests with the R5 peptide and various buffers). This highlights the role of phosphate anions in the self-assembly mechanism of the R5 peptide.

Continuing along this train of thought, Kröger et al. showed through observed broadening of ^{31}P NMR spectra of biosilica composites[53] that aggregates of about 700 nat-Sil1A protein molecules form the basis of a template for silica condensation. They proposed that this self-aggregation of proteins is driven by the zwitterionic nature of the nat-Sil1A protein: it has negatively charged phosphate groups and positively charged amines along the amino acid side chains at pH 5, which is the pH at which it induces silica precipitation. These charged groups participate in electrostatic interactions which facilitate protein self-assembly, creating aggregates along which

silica can then condense and template out. The silica proceeds by branching out into small spheres which are gradually absorbed by neighboring spheres until they reach a maximum, uniform size of roughly 500 nm in diameter.[53]

The R5 peptide only precipitates silica when it is dissolved in a buffered solution containing phosphate ions;[84] it doesn't have post-translationally modified phosphoserines or additional polyamine groups along its lysine side chains, so if it is to self-assemble in a nature similar to that of the parent protein nat-Sil1A, the phosphate anions in the buffer would be integral. It is suggested that these phosphate anions serve as cross-linking agents which could promote peptide self-assembly, substituting for the native protein's inherent phosphate groups. This could happen by way of arginine guanidinium groups, as hypothesized earlier in our discussion of arginine motifs occurring in biological systems[113, 87, 112, 91] (whether the arginine-phosphate bridges link intermolecularly across individual R5 peptides or first creates an intramolecular hydrogen bond within a single R5 unit prior to self-assembly is a topic of debate, which can hopefully be answer through analysis of side chain chemical shift perturbations in the following sections).

Several general schemes for peptide or polymer aggregation have been proposed,[63] especially as it relates to templating for silica or other biomineral growth.[104, 60, 78, 79, 115, 116, 114, 23, 118, 18, 77, 36, 34] McKenna et al. demonstrated that even without the addition of silicic acid or some other silica precursor, poly-L-lysine (PLL) in salt form (HCl, with molecular weights of 30 or 70 kDa) forms spherical aggregates in phosphate-citrate buffer, of diameters ca. 500-1000 nm.[63] They suggest that naturally occurring charged amino acid polymers can self-assemble into microspheres when paired with an appropriate counter-ion. The resulting surfaces of these spheres are functionalized and chemically active (because of the cationic nature of the side chains).

The PLL spheres were reacted with colloidal silica in one instance and then with prehydrolyzed TMOS in another (which is what is used in R5 precipitation studies in

this work). With the addition of colloidal silica, PLL spheres were redistributed and the colloidal silica formed a silica coating, with the polymers plastering themselves against the negatively charged silica coating, leaving a cavity in the center. In the case of prehydrolyzed TMOS addition, however, silica condensation proceeded at the surface of the PLL spheres while leaving the structural integrity of the original polymer assembly intact; that is, these spheres were not hollow.[63] This results in smooth surfaces, as evidenced by SEM and TEM imaging.

Other groups have found that PLL reacts with prehydrolyzed TMOS to form nano or microspheres (the size of the resulting silica particle seems to be related to polymer chain length).[78, 104, 107, 56] The surface of silica nanospheres generated in this fashion are smooth and seamless, suggesting that silica condensation is initiated by contact with the surface, rather than the adhesion of pre-formed silica oligomers (formed in solution, pre-contact with peptide moieties) to a functionalized surface. An example of the latter case is shown by Bamnolker et al.; they observed the adsorption of pre-formed silica molecules to polystyrene to give bumpy, textured silica surfaces that followed the outline of the polystyrene beads.[5] This is in stark contrast to the smooth surfaces (observed by SEM) seen by peptide-catalyzed silica condensation.

This appears to be what we observe with R5; it forms hydrogen bonds with phosphate ions, which introduces a cross-linking network of R5 peptides, resulting in a higher ordered self-assembly. This macrostructure appears to take the form of a sphere, perhaps because spherical forms minimize the surfacial interaction between neighboring charged particles.[44] Once silicic acid is introduced to the peptide-buffer system, negatively charged silicic acid molecules interact with the postively charged lysine or arginine side chains of R5, resulting in a nucleation of silica at the surface of the R5-peptide sphere. Silica condensation grows along the surface of the R5 sphere, either until the available exposed amine side chains have been used up, or until some energetically stable state of maximum silica growth has been reached.

The catalyzing force that drives silica condensation on the surface of the peptide

aggregates appears to be a combination of hydrogen bonding and electrostatic interactions between positively charged moieties on peptide side chains (namely, the ammonia group on lysine and the guanidinium group on arginine (Figure 3.3); both are protonated at physiological pH, which is relevant here).[75] This will be discussed in the next section.

3.3.3 Side Chain Chemical Shift Analysis

Chemical shift changes between the neat form of R5 and the form that is complexed with silica can be visualized with a ΔCS plot, as shown in Figure 3.11 and 3.12. A chemical shift change (i.e., ΔCS) between the two forms of the R5 peptide is obtained by subtracting the chemical shift for the ^{13}C or ^{15}N spin in the neat peptide, from the chemical shift of the same spin in the R5-SiO_2 complex. A positive ΔCS indicates that the chemical shift of the spin has increased upon complexation with silica (downfield shift, relative to TMS), and a negative ΔCS indicates that the chemical shift has decreased (upfield shift).

At pH 7, the pH at which co-precipitation with silica occurs, the side chains of lysine and arginine are positively charged. There is a lysine pair at the front end of the R5 sequence (K3, K4), two lysines throughout the middle region (K12, K15), and an arginine pair at the C-terminal end (R16, R17). SSNMR[28] and SFG[65] studies have shown that peptides interact with silica surfaces via positively charged functional groups along the peptide side chains, including lysine's NH_3^+ group and the guanidinium of arginine. Studies of polyamines inducing silica precipitation[68, 51, 99, 60] showed that molecules that were even simpler than PLL chains, but still contained the important amine group, achieved the desired interactions with silica and catalyzed silica condensation from a precursor like prehydrolyzed TMOS. Compared the the silica structures that are obtained when PLL chains are used, however, the biosilica created by polyamines lacks the regular size and patterned structure; the particles are described as colloidal silica ranging across a wider size range (100 nm -

1μm); this is because polyamines are simpler molecules than PLL and perhaps lack other defining features of PLL polymers.[60]

Still, the importance of the amine group - in whatever polymer or peptide form - in catalyzing silica condensation seems uncontested; Sumper et al., in proposing a phase separation model for the control of silicification in diatoms, suggested that silica precipitation is terminated when all of the polyamines have been consumed, indicating that the amines are the primary catalyst in silica precipitation from a silicic acid solution.[98] Similarly, ^{15}N{^{29}Si} REDOR studies and glycine-surface studies showed that monomeric alanine and glycine interact with silica through their backbone amide groups, relevant because these smaller amino acids lack extensive side chains - specifically, they lack amine side chains.[7, 8, 64].

Something to consider when discussing the side chain interactions observed by NMR; based upon our model of a R5 self-assembled sphere, the bulk of the peptide side chains would be buried in the peptide aggregate, with only some small percentage of side chains exposed at the surface and available to interact with silicic acid molecules. However, assuming that some fraction of the silicic acid molecules are actually able to burrow into the peptide aggregate, or infiltrate interstices, as suggested by Tomczak[104], then interactions of silicic acid and organic moieties within the bulk aggregate, as well as perturbations on carbon chemical shifts by silicic acid/silica, cannot be ignored.

However, through a silico-molybdate assay[44], Kröger et al. determined that silica precipitates made witih silaffin-1A[52], the molar ratio of silica/silaffin-1A is 12. Silaffin-1A has a reported molecular weight of 4 kD and silica has a molecular weight of 60.06 g/mol; this turns out to be particles that are, by weight, $\sim$ 15.2 % silica (ignoring for now the integration of phosphate groups into the biosilica peptide composite). Thus, it seems like most of the peptide interacts only with itself or with phosphate groups, with a very minimal amount interacting with silica (occurring at the surface of the R5 aggregate). We will assume here that the NMR signals (which are generated

by an ensemble average of spins) that are observed from the R5 peptide in both cases represent the bulk peptide, and chemical shift perturbations in the silica composite are largely due to (all or any, in some combination) proximity to and interactions with phosphate groups, phosphate-induced self-assembly and conformational changes, silica-induced conformation changes. Silicic acid and silica groups are believed to have a minimal impact, based solely on the low amount present in the organo-silica biocomposites.

Analysis of the side chains, then, must consider these possible phosphate interactions as well as structural rearrangements. Looking at Figures 3.11 and 3.12, it seems like side chain perturbations at lysines K3, K4 and R17 are the most pronounced; these are protonated side chains (Figure 3.11, which have been observed to interact with negatively charged phosphates.[91, 97, 67] These large perturbations also occur at the more terminal carbons of the side chains (i.e., $^{13}\mathrm{C}\delta, \epsilon$ for K and $^{13}\mathrm{C}\gamma, \zeta$ for R residues).

Looking more closely at the RR pair (Figure 3.12), it seems that the magnitude of chemical shift differences experienced by R17 are overall greater than those experienced by R16 (particularly at $^{13}\mathrm{C}\beta, \gamma$); it is possible that with natural torsions of the peptide backbone, the two arginine side chains would not be pointing in the same direction, thus allowing one arginine (e.g., R17) to interact with a phosphate group while the other does not. This might also explain the corresponding large backbone changes in I18, right next to it; if R17 hydrogen bonds to phosphate and through that, possibly another R5 molecule, it's possible that the peptide could torque around that locked interaction. Since I18 and L19 are at the ends (L19 wasn't labeled because it is attached to the resin), they would probably restructure themselves in a much more exaggerated fashion, as the ends of the peptide are usually more mobile.

Similarly large perturbations in chemical shift occur at the lysine residues K3 and K4; these are also charged side chains. PLL chains of various lengths have been shown to self-assemble into spheres in phosphate buffers[63] and other macrostruc-

tures (larger chains form hexagonal silica structures, implying that they template into some hexagonal macrostructure)[104], suggesting that lysines do provide an adequate positively charged group for interacting with phosphates and aggregating in solution. K3 and K4 in R5 could be demonstrating this kind of behavior - participation in self-assembly.

3.4 Future Goals

Here, we mostly focused on ascertaining the secondary structure of the R5 peptide in neat form, and comparing it to the form it takes when complexed with biosilica. SSNMR methods can also be used to determine the structure of protein assemblies and to investigate the local environments around these assemblies. In this work, I put forth some discussion as to the nature of the self-assemblies and the possible interaction of R5 side chains with phosphate groups. In order to test this hypothesis, it is possible to do $^{13}C\{^{31}P\}$ REDOR[32] to determine if there are any phosphate groups in close proximity to the arginine side chains, for example. If dephasing of the ^{13}C spins in the labeled arginine side chain occurs, then it is presumed that phosphates are near (^{31}P is helpfully both spin half and 100% naturally abundant). These internuclear distances can be measured by fitting of the resulting REDOR curve, as demonstrated by previous work in this group with statherin/SN15 and HAP[73, 74] and Tat peptides with TAR RNA.[43] Another, more labor-intensive approach to investigating the role of phosphates in the R5 peptide is to phosphorylate the serine residues in the sequence and then perform the same characterization (SEM) and SSNMR experiments. It would also be interesting to see whether the phosphate requirement disappears with the incorporation of phosphoserines in the sequence, as it does with natSil-1A.[53].

Another interesting mutant study, following the lead of Knecht et al.[50] is to eliminate all of the lysines in the sequence, perhaps with non-basic amino acids, to start, or with arginines, and in both the sequence with phosphoserines, and the wild type. It would be interesting to see whether the presence of lysines is necessary to

form aggregates on a large scale, if arginines throughout the sequence would suffice, or if the lysines aren't necessary in the presence of phosphoserines/phosphate groups.

To determine whether there are substantial interactions between R5 amino acid side chain groups and silica moieties (even within the bulk aggregate), dipolar recoupling experiments like ^{15}N{^{29}Si} REDOR[32] can also be used; these experiments will provide a measurement of the distances between targeted side chain spins (e.g., the ^{15}N on the end of a lysine, for example) and the surrounding silica phase; this was done by Ben Shir et al. with monomeric alanine and glycine, to show that the backbone amides interact with silica.[7, 8] In this group, the same ^{15}N{^{29}Si} REDOR experiments are being performed on existing samples (to start, samples R, S, and SR) to examine the distances between the guanidinium groups on the arginine side chains and the backbone amides in the serines and silica. REDOR experiments can also be used with selectively ^{13}C-enriched R5 samples to determine the nature of peptide assemblies (as shown by Balbach et al.with β-amyloid fibril assemblies[4]) in silica.

If there is not a lot of silica present in the particles, as hypothesized, and the silica-peptide interactions are limited to surfacial regions of the peptide aggregate, then it might be prudent to simply test dephasing of silica - that is, detect ^{29}Si dephasing by ^{13}C or ^{15}N instead. This would at least confirm silica proximity to specific labeled locations along a peptide, and would work best to to isolate possible side chains that serve as nucleation points of silica condensation (i.e., lysine or arginine side chains).

To further probe side chain dynamics and interactions, it is also possible to perform relaxation experiments like those outlined in Section 2.6, specifically CP ^{13}C or ^{15}N T_1IR experiments to measure the T_1 times of ^{13}C or ^{15}N spins along the backbone and side chains. These values would provide some insight as to the time scale of movement along the side chains, as well as hints as to the local environment and what steric constraints may be prevalent.

To address the issue of bulk peptide side chains versus surfacial side chains (that is, observing the interactions of peptide side chains that may be buried within the

aggregate versus those that actively interact with silica, at the very outer surface of the peptide aggregate sphere), it could be useful to repeat the 1D and 2D ^{13}C and ^{15}N CPMAS and DARR experiments with R5 adsorbed to a silica surface, like silica gel, rather than inducing co-precipitation of the peptide and biosilica matrix. Any changes in chemical shift and predicted structure between the two (adsorbed and co-precipitated R5) could provide more insight as to the nature of the organic/inorganic matrix nestled at the heart of the silica-peptide composites.

Finally, other R5 work in this group focuses on extrapolating the concept of biomimetic synthesis to another useful - though non-biological - oxide: TiO_2. Sewell et al. showed that R5 precipitates titania nanoparticles from TiBALDH (titanium(IV)bis (ammonium lactato) dihydroxide)[92], much in the same way that it co-precipitates with silica, and other groups have studied further protein-TiO_2 interactions.[23, 47, 55, 48] In this group, the secondary structure investigation will be repeated (^{13}C and ^{15}N 1D and 2D experiments for chemical shift assignments) to show whether the R5 peptide takes a different structure in titania, and then follow-up with analyses of chemical shift perturbations ΔCS values) to assess local areas along the peptide where titania-R5 interactions may occur.

Chapter 4

THE LKα14 PEPTIDE

In order to generalize interactions between proteins and silicic acid and silica, it seems logical to look at model peptides. Short, well-designed sequences offer a lot of control and the opportunity to focus in on certain protein qualities and how they affect silica precipitation (e.g., sequence length, amino acid composition, secondary structure, presence of hydrophobic/hydrophilic side chains and charged functional groups).

With R5, our work determined that the structure of the neat, lyophilized peptide and the R5 in a silica composite were mostly similar, based on torsion angles predicted by TALOS-N, motivated by chemical shifts that were experimentally obtained through ^{13}C and ^{15}N experiments. However, it was still unclear whether the changes in chemical shift could actually be attributed to changes in structure and conformation, since there is an assortment of possible contributions to ΔCS values, including the presence of phosphate and/or silicic acid and silica within the biocomposite. Other proposed models for peptide templating of silica condensation[104, 60, 78, 79, 115, 116, 114, 23, 118, 18, 77, 36, 34] assume that a peptide retains its specific secondary structure throughout its interaction with silicic acid, but no molecular level assessment and quantification of that was previously made.

In addition, much of the biosilicification and biomimetic approaches in recent literature used PLL chains,[78, 107, 36] polyamines,[51, 68, 15, 101], amino-containing peptides[18] or amphiphilic peptides containing lysines[118, 114] to produce regular silica nanoparticles, and SFG studies by Somorjai and coworkers[65, 82] that demonstrated adsorption of lysine side chains onto a silica surface, all pointed to the primary

role of lysines (and more importantly, its amine group) in the biosilicification process.

Since the LKα14 peptide was worked on previously in this group,[59, 13, 109, 12], I was familiar with the design of the amphiphilic peptide (alternating leucines and lysines in some fixed pattern to produce regular secondary structure), and their relatively short length made them appealing to work with - they are fairly easy to synthesize and require little to no extra purification because of their simplicity and lack of very complicated amino acid side chain protection schemes.

Ac-LKKLLKLLKKLLKL-c

Figure 4.1: Primary sequence of the LKα14 peptide; the N-terminus is capped with an acetyl group, designated "Ac."

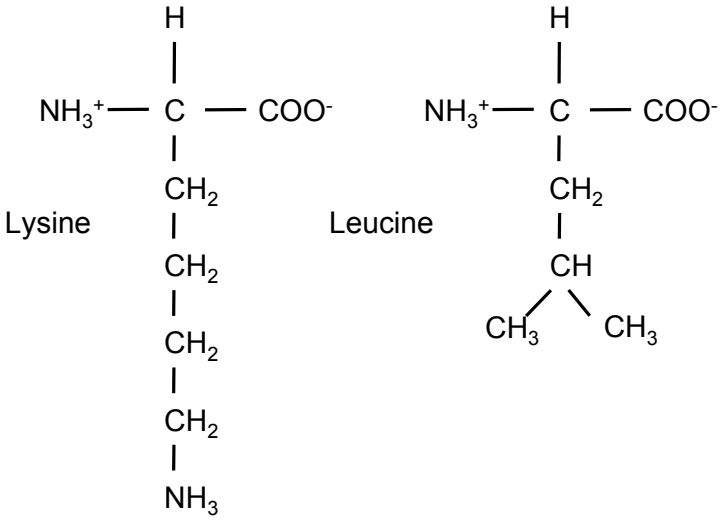

Figure 4.2: Structures for Leucine (L) and Lysine (K) at neutral pH (= 7); all amino acids take the L conformer. The lysine side chain is protonated at pH = 7, which is the pH at which the silica precipitations in this thesis (Appendix A) take place.[75]

As shown in Figure 4.1, the LKα14 peptide contains six lysines (R5 contains six total basic residues, as well, though in the case of R5 two of the six basic residues

are arginines), a substantial fraction of the total 14 residues. This led me to wonder if the LKα14 peptide would also induce silica precipitation. If so, it would provide a useful model system to probe.

The α-helix is designed such that the leucine side chains are aligned along one half of the peptide, and the lysine side chains are aligned along the other, as in Figure 4.3. It has been shown that this peptide preserves its α-helical conformation when adsorbed onto silica, polystyrene, and functionalized polystyrene surfaces[65, 59, 13, 109]. Long et al. found that the leucine side chains of the LKα14 peptide interacts with hydrophobic surfaces[59]; Bower et al. confirmed that lysine side chains are involved in adsorptions to hydrophilic surfaces, and that α-helical formation precedes adsorption[12].

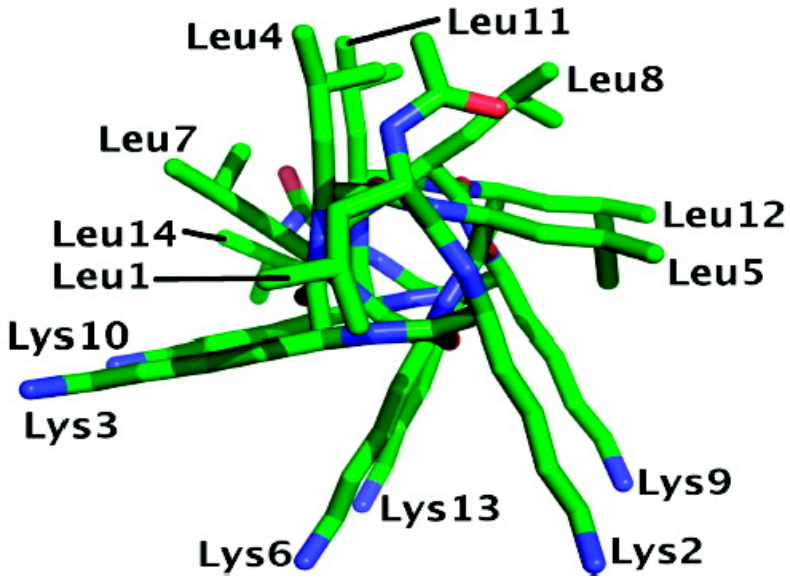

Figure 4.3: End-on view of LKα14 with residues labeled; this shows that leucine side chains align on one half of the helix, and lysine side chains align on the other, creating a hydrophobic side and a hydrophilic side of the helix, respectively. Reprinted with permission from Nicholas F. Breen; Tobias Weidner; Kun Li; David G. Castner; Gary P. Drobny; J. Am. Chem. Soc. 2009, 131, 14148-14149. Copyright 2009 American Chemical Society.

If the LKα14 peptide did induce silica formation, knowing whether it kept its - already solved and documented - secondary structure, would provide insight on how lysines and lysine-containing peptides are affected by incorporation of silica and the assembly of biosilica particles. Since the LKα14 peptide retains its structure when it is adsorbed to other surfaces, if it was also found to retain its structure when involved in a silica composite, that would say something about the nature of the interaction between the side chains and silicic acid, and the nature of possible peptide self-assembly mechanisms (since secondary structure affects tertiary and quaternary structure). This was the basis of launching a study of biosilicification and the LK peptides.

4.1 The LK Peptides and Silica Precipitation

The LK peptides are a family of amphiphilic peptides designed by Degrado et al.; [20, 22] hydrophobic leucines (L) and hydrophilic lysines (K) were arranged in selected motifs that generated specific secondary structures when placed at an interface (e.g., water and air). By alternating the pattern of amino acids and the length, it is possible to achieve several different secondary structures: α- and 3_{10}-helices, and β-strands.[20]

Because the LK peptides, like R5, are lysine rich, and based on studies which featured polylysine-silica precipitates[56], and demonstrated that the LKα14 binds to silica wafers,[65], I hypothesized that the LKα14 peptide would also co-precipitate with silica in a fashion similar to R5. To test this, I took some unlabeled LKα14 peptide from previous work in the group[13, 109], dissolved it in 1X PBS buffer, and added silicic acid (following the silica precipitation procedure outlined for R5 in Appendix A). Within a similar time frame, and over a peptide concentration range of 1 mM - 10 mM, I found that the LKα14 peptide does induce silica precipitation when combined with orthosilicic acid ($Si(OH_4)$) at neutral pH. Like with the R5 peptide, the resulting silica pellets from the LKα14 precipitation were centrifuged, washed, dried, and imaged with SEM. The SEM image of LKα14 (Figure 4.4)shows that the

resulting silica precipitates look nearly identical to the R5-SiO_2 nanospheres, with similar diameters of 500-700 nm.

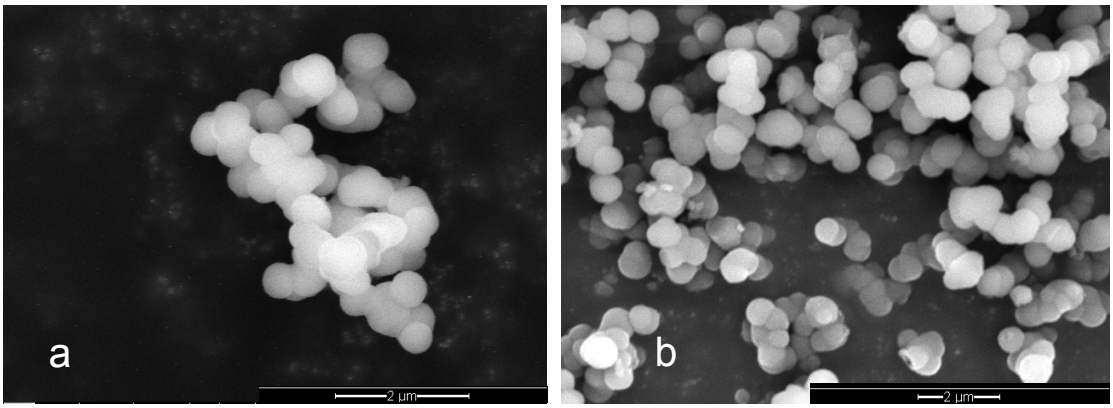

Figure 4.4: SEM images of (a) LKα14 and (b) R5 silica co-precipitates. Both peptides develop silica precipitate composite nanospheres, with similar diameters of 500-700 nm. The inset indicates the scale bar for size.

4.1.1 LK Peptide Sequence Length and Composition Effects on Silica Precipitation

This discovery led to further questions about the relationship between sequence length and sequence composition, and the readily customizable design of the LK peptides made them effective test subjects to use in investigation of these topics.

To start, I synthesized a set of unlabeled LK peptides of various lengths and secondary structures (Table 4.1), and followed the silica precipitation process again with each individual peptide.

Shorter peptides (seven residues in length) did not induce silica precipitation under conditions detailed in Appendix A; longer peptides (LKβ11, LK3_{10}13, LKα14 and LKβ15) did. All silica precipitates were spun down, washed, dried, and again imaged by SEM. Figure 4.5 shows that the different peptides (3_{10}-helix, α-helix, β-strand) co-precipitated with silica in different morphologies; the sizes and shapes of the resulting silica nanoparticles are different for each type of peptide. Despite the subtle difference

Table 4.1: List of Synthesized, Unlabeled LK Peptides

LK Peptide	sequence	Silica? (Y/N)	Silica Shape	diam/length (nm)
LKα7	LKKLLKL	N	N/A	N/A
LKβ7	LKLKLKL	N	N/A	N/A
LK3_{10}7	LLKLLKL	N	N/A	N/A
LKβ11	LKLKLKLKLKL	Y	amorphous	20,000
LK3_{10}13	LLKLLKLLKLLKL	Y	fibrils/tubes	500-850
LKα14	LKKLLKLLKKLLKL	Y	spheres	500-700
LKβ15	LKLKLKLKLKLKLKL	Y	amorphous	20,000

in helical character between the 3_{10} and α helices (and only one amino acid difference in sequence length), the LK3_{10}13 and LKα14 peptide made nanoparticles of very different structures (i.e., nanotubes/fibrils - it is currently unknown whether these particles are hollow - versus nanospheres, respectively). It is also worth noting that the β-strands of lengths 11 and 15 residues created amorphous particles of roughly the same size range, despite the four-amino acid difference in sequence length.

The SEM images of the silica composites that were precipitated from different LK peptides brought up an interesting question: since each secondary structure afforded a unique silica structure, was there a direct relationship between peptide secondary structure and the resulting silica morphology? In order to further explore the concept, I focused my attention on the LKα14 peptide, since the group had previously done work with it,[59, 13, 109] and because it made silica nanospheres that were nearly identical to those created with the R5 peptide.

4.2 Secondary Structure Analysis of the LKα14 Peptide in SiO_2

As mentioned previously, the LKα14 peptide was designed to adopt a helical conformation at interfaces[20], and I discovered that it precipitated silica nanospheres

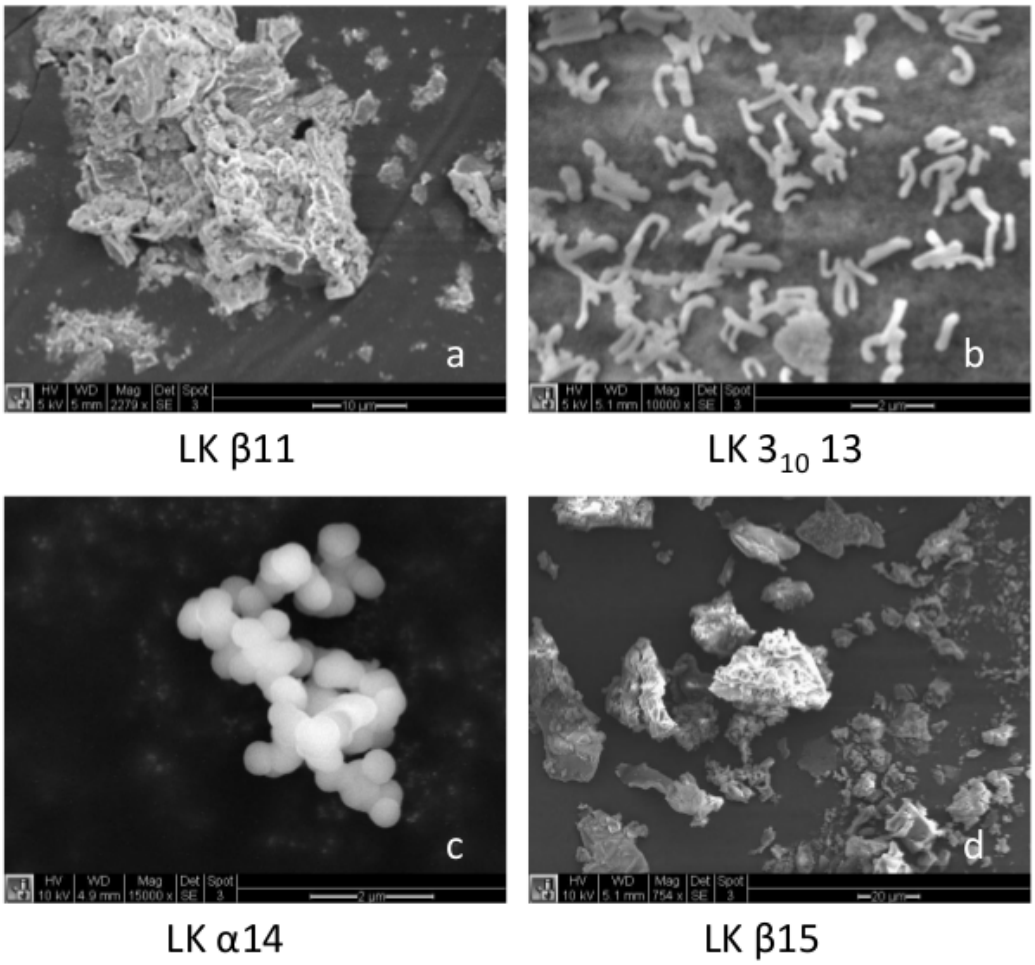

Figure 4.5: SEM images of the silica structures resulting from co-precipitation with several different LK peptides; each different secondary structures resulted in different silica nanoparticle morphologies.

similar to those that co-precipitated with the R5 peptide. To further investigate the relationship of secondary structure with a peptide, I followed the same strategy that was initially used with R5: to compare the secondary structure of the neat peptide with that of the peptide once it was co-precipitated with silica.

4.2.1 Peptide Synthesis and Materials

Unlabeled LK peptides were synthesized de novo on a Rainin PS3 instrument using standard Fmoc chemistry (Appendix A), followed by acetylation of the N-terminii; unlabeled, protected amino acids were purchased from Sigma-Aldrich. The list of synthesized,unlabeled LK peptides appears in Uniformly ^{13}C and ^{15}N labeled LK peptides were also synthesized de novo using standard Fmoc chemistry (Appendix A).[3] Uniformly labeled, protected amino acids and Fmoc-Leu-Wang resin were purchased from Cambridge Isotope Laboratories (Andover, MA). Unlabeled amino acids were purchased from Sigma-Aldrich in protected form. Protection schemes for labeled leucine are given in Appendix A. L-lysine was purchased in its protected form, also from Cambridge Isotope Laboratories (Andover, MA).

The complete set of synthesized, isotopically labeled LKα14 samples are given in Table 4.2.

In addition to peptides that were synthesized in house, two peptides were purchased from a commercial source: L5K6 (U^{13}C, ^{15}N labels) and L4L8 (^{13}CO on L4, ^{15}N on L8) were purchased custom-made from Anaspec (Fremont, CA).

The neat peptide is lyophilized from a solution of 3 mM LKα14 in deionized water; the LKα14-SiO_2peptides are co-precipitated with silica as outlined in Appendix A.

4.2.2 SEM Imaging

SEM experiments were performed on a FEI Sirion scanning electron microscope; dry precipitates were dispersed onto double-sided sticky carbon tap and mounted on aluminum studs,and sputter-coated for 60 s with Au/Pd. Images were acquired with accelerating voltages of either 5 or 10 kV.

Table 4.2: List of Synthesized, Isotopically Labeled LKα14 Peptides.

Peptide name	Labels
L1K2	U^{13}C, ^{15}N
L4	U^{13}C, ^{15}N
L5K6	U^{13}C, ^{15}N
L7	U^{13}C, ^{15}N
L8K9	U^{13}C, ^{15}N
K10L11	U^{13}C, ^{15}N
L11	U^{13}C, ^{15}N
L12K13	U^{13}C, ^{15}N

4.2.3 Experimental Methods

SSNMR experiments on LKα14 were taken at two fields. ^{15}N CPMAS experiments were taken on a 500 narrow bore Bruker spectrometer, with a proton field of 499 MHz, ^{13}C Larmor frequency of 125.58 MHz, and ^{15}N Larmor frequency of 50.61 MHz, fitted with a 3.2 mm ^{1}H{^{13}C, ^{15}N} MAS probe. 1D ^{13}C CPMAS and 2D SSNMR experiments were performed on a Bruker Avance III spectrometer, at a field of 700.18 MHz (^{1}H), fitted with a ^{1}H{^{13}C, ^{15}N} 3.2 mm MAS probe. The Larmor frequency for ^{13}C at this field is 176.07 MHz, and 70.95 MHz for ^{15}N. The spinning speed for all experiments was 15 $\pm$ 2 kHz, regulated by a Bruker automated MAS controller. All samples were run at room temperature ($\sim$22 °C).

CP ^{13}C experiments were run with ^{1}H 90 times of 3.2 μs, contact times of 1100 μs, and a recycle delay of 2 s; for the neat peptide, 512 scans were taken; for the -SiO_2 variant, 2048 scans were taken. For ^{15}N CPMAS experiments, 4096 scans were taken, with a 2.5 μs ^{1}H 90 pulse, 2000 μs contact time, and 2 s recycle delay. All ^{13}C experiments were referenced to adamantane as a secondary standard (relative to

TMS). All ^{15}N experiments were referenced to labeled ^{15}N ammonium sulfate.

DARR experiments were performed with mixing times of 30, 60, and 100 ms. The ^{1}H 90 pulse length was 3.2 μs, the contact time was 1100 μs, and the recycle delay was 1.5 s.

4.2.4 Results

As with the R5 peptide, both 1D and 2D ^{13}C and ^{15}N experiments were necessary to make assignments of all residues. Frustratingly, lysine and leucine chemical shifts show a lot of overlap (as shown in Figure 2.5).

The chemical shift assignments for backbone and side chain ^{13}C spins in the neat and -SiO_2 LKα14 peptide are summarized in Tables 4.3 and 4.4, respectively. An example of the 2D DARR data is shown in Figure 4.6; a difference in line shape of the terminal methyl side chain groups (Figure 4.2) is obvious here. This suggests some difference in mobility and motion or interactions of the side chain groups between the neat form and that in the biosilica. These will be discussed further in Section 4.4.3.

Another tool at our disposal is the variability of mixing times and the corresponding information incrementally long mixing times affords about the nature of spin correlations and polarization transfer. Sometimes, in order to distinguish between ^{13}C spins between a leucine and lysine residue, for example, it is helpful to compare the spectra taken at different mixing times. Typically, longer mixing times introduce longer-range couplings[102, 103]; thus, it becomes easier to distinguish between, for example, the lysine ^{13}Cβ - ^{13}Cϵ interaction and that of leucine's ^{13}Cβ - ^{13}Cγ interaction (the former only shows up in the 100 ms mixing time spectrum, as it is a three-bond length cross peak, versus the single bond distance interaction of the leucine ^{13}C spins).

The chemical shift assignments (externally referenced to adamantane and TMS) were used to generate TALOS-N input files, which in turn predicted torsion angles. This led to prediction of secondary structure for the LKα14 peptide in neat form

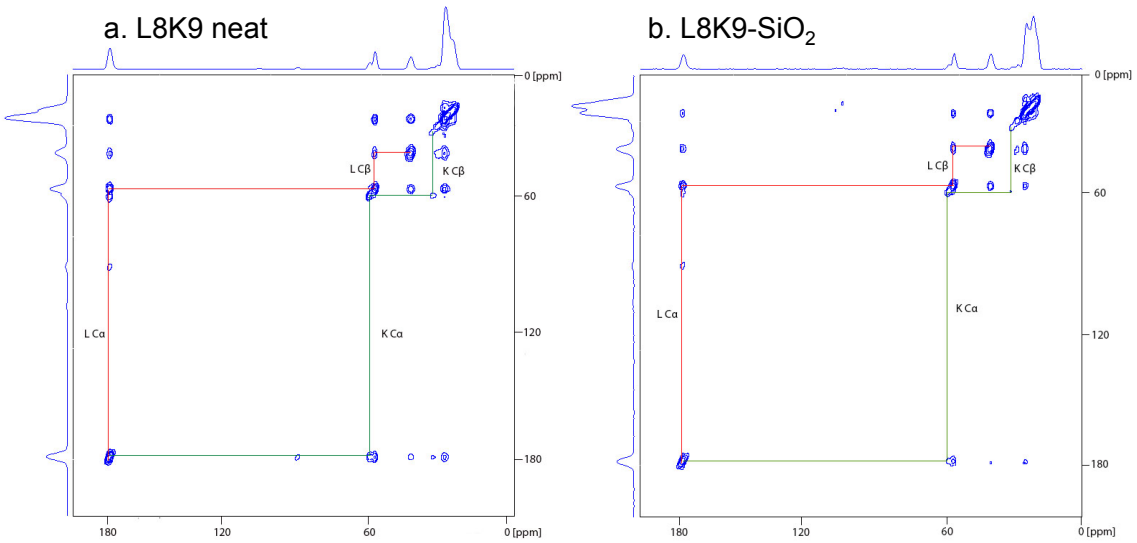

Figure 4.6: 2D DARR spectra of the L8K9 labeled LKα14 peptide. This compares the neat version of the peptide (a.) with that of it complexed with silica (b.), with 100 ms DARR mixing time. The green and red lines guide the eye down the backbone chemical shifts (and $^{13}C\beta$) of residues L8 and K9.

and when complexed to silica. It is also possible to compare chemical shift values to tables of reported values associated with each secondary structure type (i.e., α-helix, β-sheet), and TALOS-N also reports a residue's propensity for a certain conformation.[94, 106] Another useful tool of analysis is again examining ΔCS values - quantifying a change in chemical shift between the two forms of the peptide (numerical values are obtained by subtracting the neat peptide chemical shift from that of the -SiO_2 form; positive values indicate that -SiO_2 peaks fall downfield of the original neat peptide, and negative values indicate that the -SiO_2 peaks have moved upfield relative to the neat). ΔCS results for the neat vs. -SiO_2 forms of LKα14 are shown in Figures 4.10 and 4.11.

Table 4.3: Chemical Shift Assignments for the neat LKα14

residue	CO (ppm)	Cα (ppm)	Cβ (ppm)	Cγ(ppm)	C$\delta_{(1/2)}$(ppm)	Cϵ (ppm)
L1	177.603	57.272	41.622	25.861	22.827	n/a
K2	177.134	58.987	32.258	25.376	29.502	41.395
L4	177.272	57.979	42.027	26.132	23.316	n/a
L5	178.220	58.358	41.574	27.324	25.414	n/a
K6	178.640	60.316	32.483	25.730	29.401	42.371
L7	177.624	57.360	41.272	25.014	21.755	n/a
L8	178.233	58.097	41.738	26.518	22.992	n/a
K9	178.082	60.455	32.158	26.392	29.415	42.008
K10	177.539	59.599	32.348	26.086	28.978	41.942
L11	178.224	57.305	41.305	25.834	22.565	n/a
L12	177.927	57.509	41.305	25.991	22.623	n/a
K13	176.775	55.781	33.119	25.643	29.475	42.132

Table 4.4: Chemical Shift Assignments for LKα14-SiO_2

residue	CO (ppm)	Cα (ppm)	Cβ (ppm)	Cγ (ppm)	C$\delta_{(1/2)}$(ppm)	Cϵ (ppm)
L1	178.019	57.796	41.742	25.734	22.553	n/a
K2	178.494	59.437	29.231	24.511	29.160	42.091
L4	178.562	56.816	40.993	26.251	23.337	n/a
L5	178.546	57.952	41.209	26.144	22.870	n/a
K6	178.998	59.998	31.933	25.137	29.293	41.761
L7	177.469	57.507	41.450	26.478	22.826	n/a
L8	178.050	58.146	41.762	26.140	22.830	n/a
K9	178.414	60.452	32.175	25.621	29.990	42.126
K10	179.081	59.486	31.714	26.127	29.181	42.741
L11	177.899	57.713	41.523	26.371	23.195	n/a
L12	176.009	55.618	42.562	26.310	23.671	n/a
K13	178.394	58.129	32.896	26.670	30.029	42.865

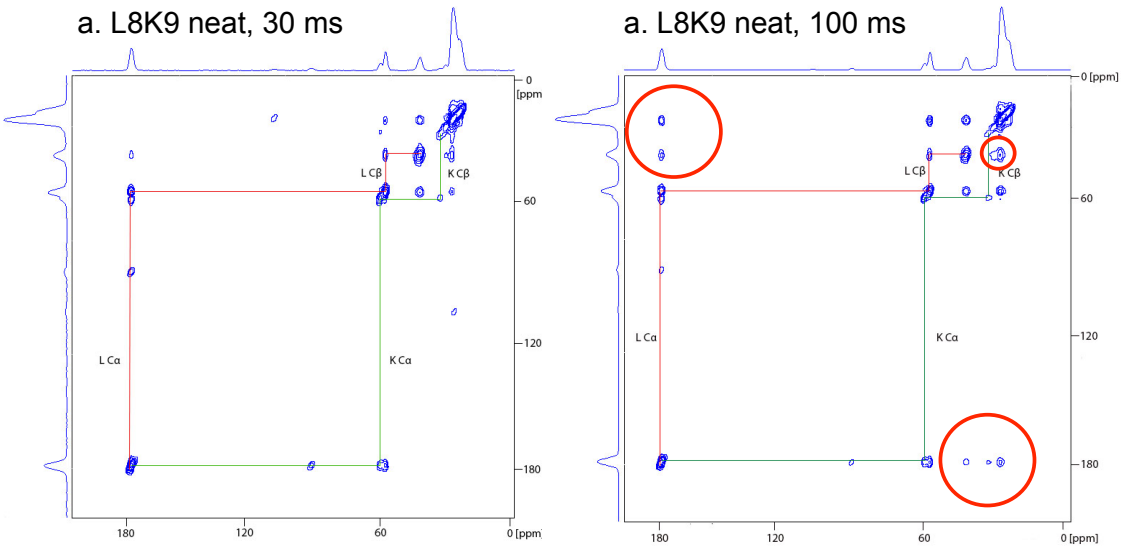

Figure 4.7: 2D DARR spectra of the neat L8K9 LKα14 peptide again, this time showing the spectrum at 30 ms DARR mixing time, and then at 100 ms mixing time. The red circles point out the long-time correlations that do not occur in the 30 ms spectrum. This allows us to distinguish between close-range interactions and those that are farther out (e.g., a ^{13}CO - ^{13}Cα interaction versus a ^{13}CO - ^{13}Cγ one).

4.3 Side Chain Chemical Shifts and Dynamics

To get a sense of the time scale for molecular motions in the LKα14 peptide, I performed ^{13}C CP T_1IR studies on the neat and -SiO_2 peptides, incorporating a uniformly labeled leucine (synthesized de novo according to the procedure outlined in Appendix A) at the seventh amino acid place in the sequence; this sample is referred to as UL7. An additional sample at L11 was made (identical labels) to test another region of the peptide; this sample is UL11.

The calculated T1 values for each of the carbons in the UL7 and UL11 samples (neat and in silica) are shown in Table 4.5.

Table 4.5: ^{13}C T_1 values for LKα14 UL7, UL11 neat and -SiO_2; values were calculated with results from ^{13}C T_1IR experiments.

^{13}C spin	UL7 ^{13}C T_1 (neat)/s	UL7 ^{13}C T_1 (-SiO_2)/s	UL11 ^{13}C T_1 (neat)/s	UL11 ^{13}C T_1 (-SiO_2)/s
CO	1.529 ± 0.029	1.574 ± 0.020	1.390 ± 0.027	2.187 ± 0.034
Cα	1.358 ± 0.007	1.397 ± 0.008	1.181 ± 0.046	1.460 ± 0.158
Cβ	1.391 ± 0.020	1.668 ± 0.013	1.119 ± 0.038	2.077 ± 0.075
Cγ	1.185 ± 0.013	1.238 ± 0.003	0.977 ± 0.019	1.131 ± 0.026
Cδ	1.291 ± 0.058	1.184 ± 0.061	0.955 ± 0.043	1.177 ± 0.013

4.3.1 Experimental Methods

^{13}C CP T_1IR experiments were performed on a 500 narrow bore Bruker spectrometer, with a proton field of 499 MHz and a ^{13}C Larmor frequency of 125.58 MHz, using a ^{1}H{^{13}C, ^{15}N} 3.2 mm MAS probe. All samples were run at room temperature (∼22 °C), and the MAS speed was 10 ± 2 kHz, regulated by a Bruker automated MAS controller.

Initial CP experiments were run with an ^{1}H 90 time of 2.75 μs and a contact time of 1600 μs. The ^{13}C 90 time was optimized at 4 μs using a basic direct detect pulse program on the ^{13}C channel (with ^{1}H decoupling). The ^{13}C CP T_1IR pulse sequence is outlined in Section 2.6.

For the neat structure, the delay increments were: 0.0001, 0.0005, 0.001, 0.010, 0.050, 0.100, 0.180, 0.500, 0.750, 1, 1.5, 2, 2.5, 3, 4, 5 seconds. For the -SiO_2 sample, the delay increments were 0.0001,0.001, 0.010, 0.050, 0.100, 0.500, 1, 5, 10, 20, 50, 80 s (12 total time points). Spectra were processed and T_1 values were calculated with Bruker's spectrometer software package, Topspin 3.0.

4.4 Discussion

Discovery of the different morphologies of silica precipitates generated by LK peptides of various secondary structures (i.e., the α-helix giving spheres, the β-strands giving irregular, unstructured precipitates, the 3_{10}-helix giving fibrils/tubes), it was thought that perhaps secondary structure played a direct, obvious role in controlling the size and shape of resulting silica particles. Thus, a SSNMR analysis of the secondary structure of the LKα14 peptide in neat form and in its silica composite was started. This also served as a method to test and validate the methodology used in R5 results; the LKα14 peptide has been well-studied and is well-documented in taking the α-helical form at interfaces and on various types of material surfaces through rigorous, well-accepted forms of NMR and SFG.[65, 12, 59, 13, 109] If TALOS-N generated an α-helical form for at least the neat peptide, then that lends more conviction to R5 results, as well.

Ultimately, though, the interest was in determining if LKα14 retained its helical form in the biosilica composite, and the possibilities for elucidating silica-interaction mechanisms that that discovery might uncover.

4.4.1 Comparison of the Secondary Structure of the neat and -SiO_2 LKα14 Peptide

Using the 1D and 2D ^{13}C CPMAS and DARR spectra, chemical shifts were assigned to leucines and lysines in the sequence, according to the labeled samples in Table 4.2. These experimentally obtained chemical shift values were then submitted to the TALOS-N server[94], which predicted torsion angles for the LKα14 peptide both in its neat, lyophilized (from a solution of deionized water and controlled peptide concentration) form, and the form it takes when it is involved with silica co-precipitation. All torsion angles for both forms of LKα14 were generated with a "strong," unambiguous prediction from TALOS-N, meaning that there were no outliers amongst the 25 best database matches.[94] The predicted structures are shown in Figures 4.8 and

4.9.

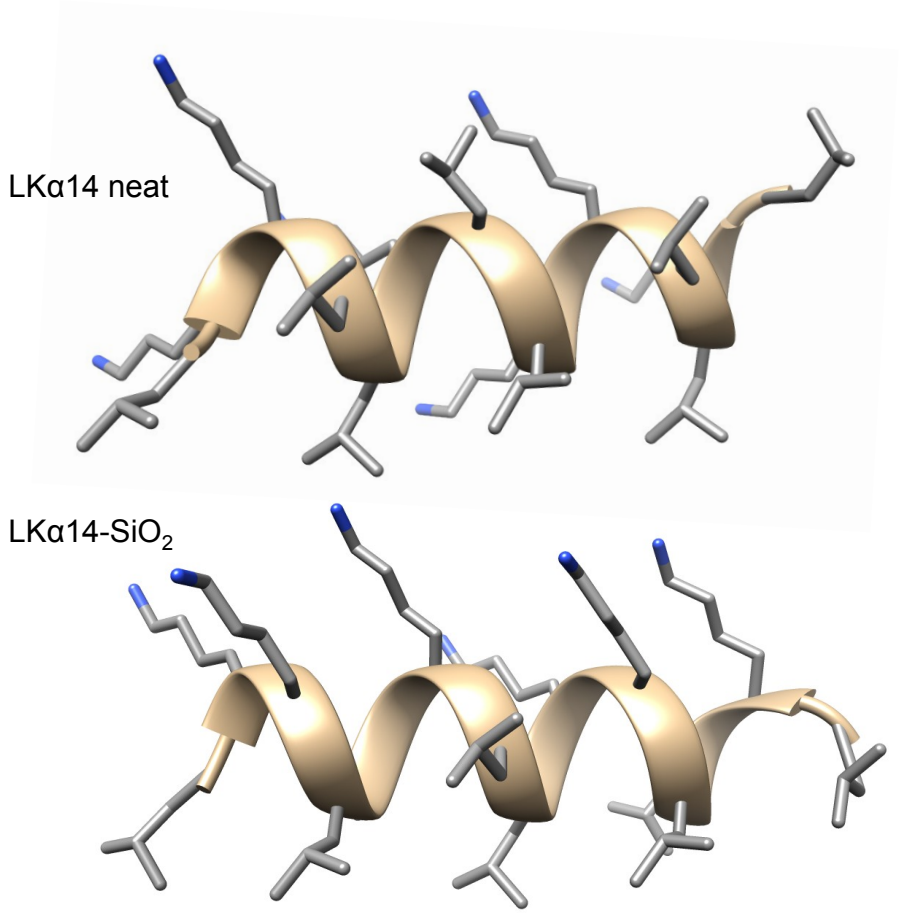

Figure 4.8: As with R5, the torsion angles predicted by TALOS-N were input into Chimera[81] to generate predicted molecular structures for the LKα14 peptide, both in its neat form and when it is co-precipitated with silica. This is a side view of both peptides. In practice, the N-terminus is capped with an acetyl group, but that is missing from this model.

Clearly (and this is also obvious by looking at the list of predicted torsion angles from TALOS-N; they all fall within the α-helical regime), the peptide is an α-helix in both cases - neat, and in silica. For the neat case, this falls in happy agreement with previous work done in the group[59, 12, 13, 109], which also serves to validate the use of TALOS-N on SSNMR experimentally obtained chemical shifts, as well.

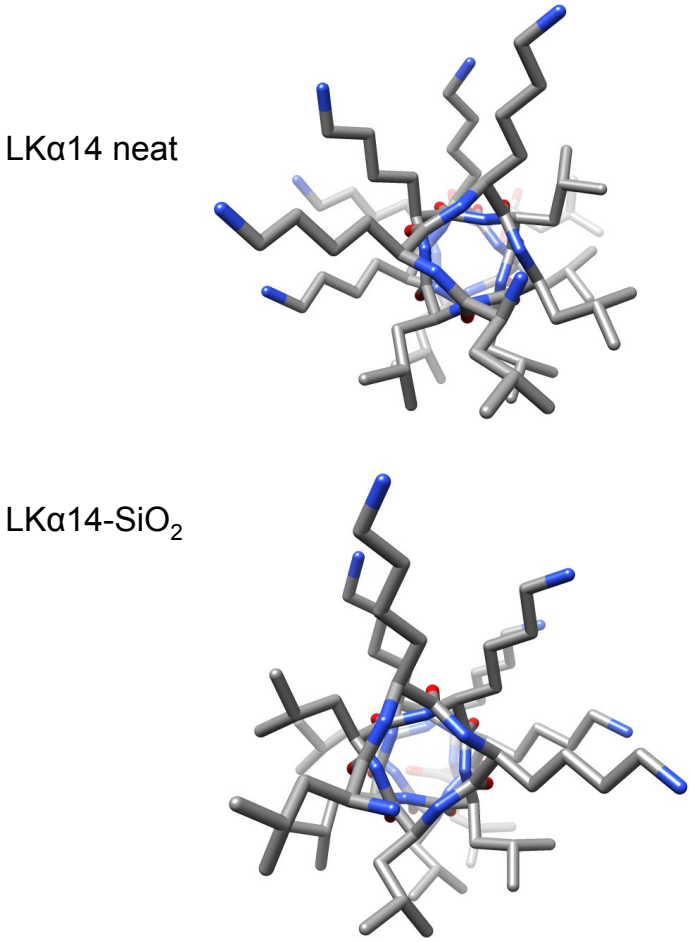

Figure 4.9: Top-down view for the same predicted structures for LKα14 (neat and in silica) shown in Figure 4.8; the top-down view emphasizes the alpha-helical structure of the peptide in both forms. Again, the N-terminal acetyl cap is not present in this model, to present a clearer view of side chains.

The result that LKα14 also retains its α-helical form in the silica biocomposite is also interesting. It helps to shape models for macrostructure self-assembly, and, because the α-helical structure here is so different from the predicted structures of R5, while producing very similar silica nanoparticles, this also suggests that perhaps the effect of secondary structure is not as straightforward as previously hypothesized.

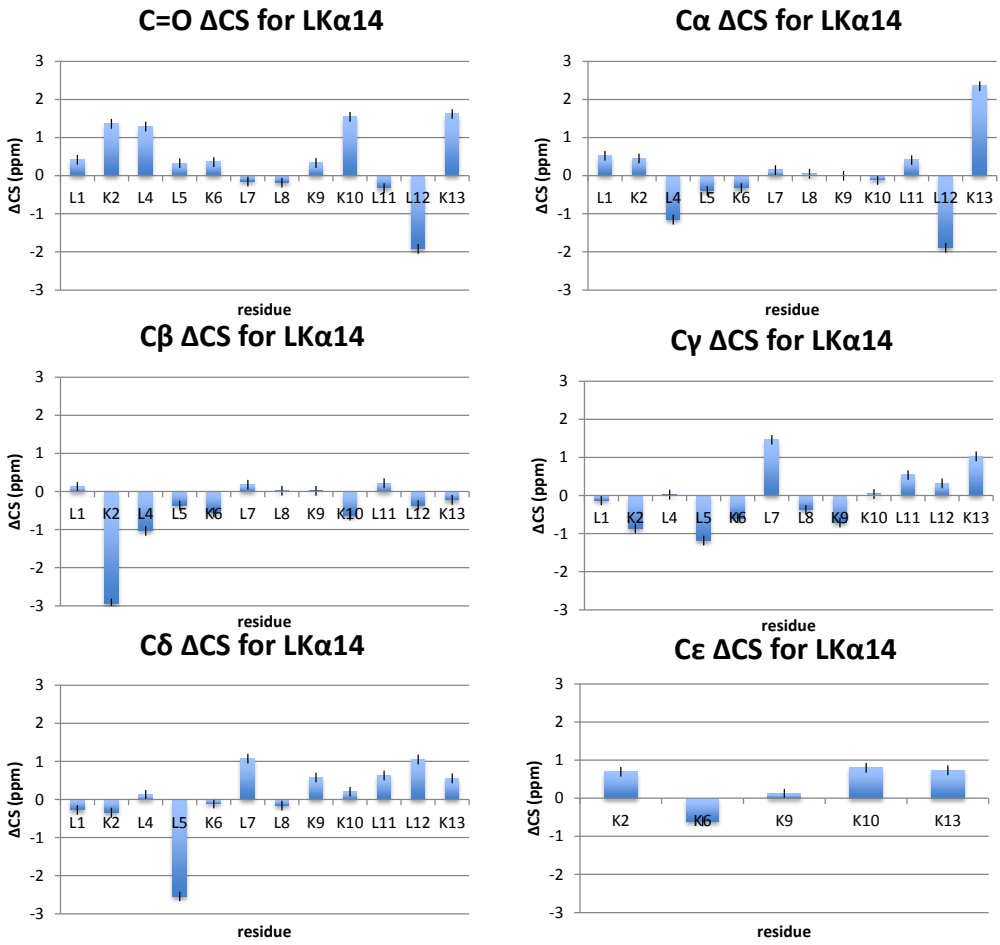

Figure 4.10: A histogram showing the ΔCS trends for ^{13}CO, ^{13}Cα, ^{13}Cβ, ^{13}Cγ, ^{13}Cδ, ^{13}Cϵ spins in LKα14.

But there are certain similarities in trends between the ΔCS values seen in R5 and in LKα14; like with R5, the backbone chemical shifts change more significantly

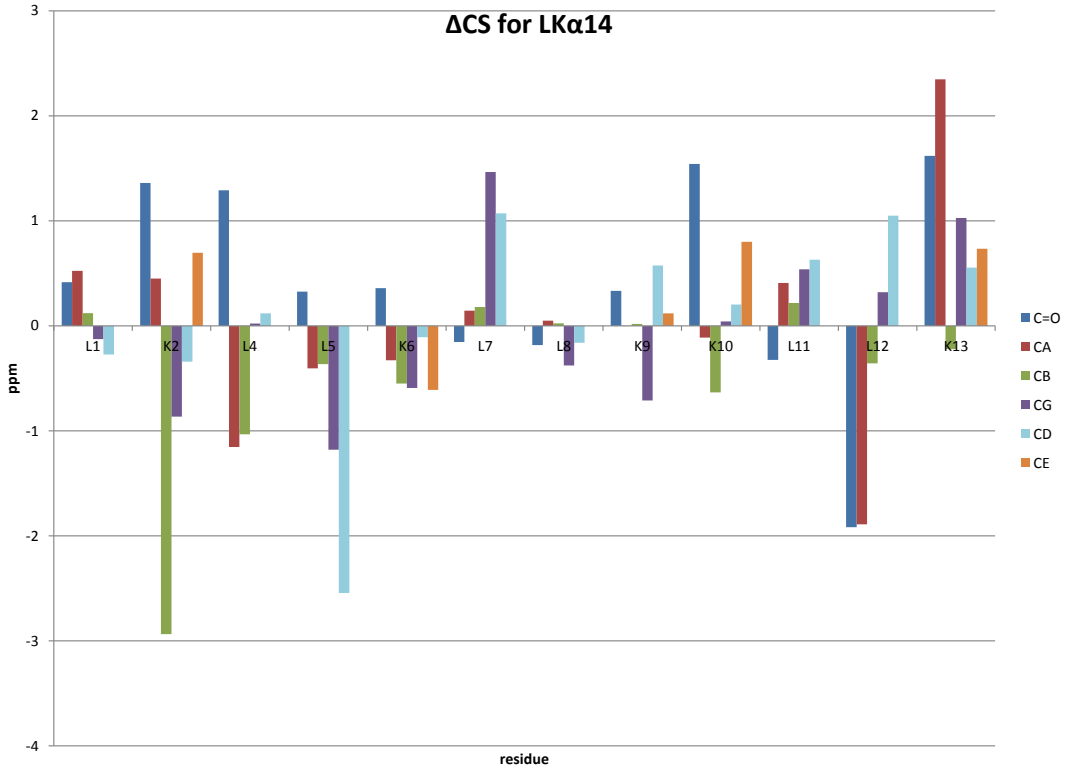

Figure 4.11: Another histogram of the ΔCS trends for ^{13}CO, ^{13}Cα, ^{13}Cβ, ^{13}Cγ, ^{13}Cδ, ^{13}Cϵ spins in LKα14. Showing each of the ΔCS for each residue clustered together helps highlight trends by residue.

toward the ends of the sequence (L1, K2, L12 and K13 show larger ΔCS values, as shown in Figures 4.10 and 4.11). This implies that like R5, the LKα14 peptides are more mobile toward the ends rather than the center, or the ends are more affected by aggregation effects or other conformation changes induced by self-assembly.

Overall, though, the magnitude of ΔCS values in R5 are much higher than those observed in LKα14, suggesting that each individual R5 peptide moves much more to get into a self-assembled aggregate macrostructure than the LKα14 peptide. This makes sense, since R5 doesn't seem to have a rigid structure in solution, whereas

LKα14 is α-helical. It is further postulated that the LKα14 peptides form tetrameric bundles in solution[65, 20, 22]. If that's the case, each individual LKα14 peptide is additionally held into place alongside three other α-helices, in a fixed position relative to the other three. It is presumed that bundle formation is driven by the hydrophobic effect; it is energetically favorable to bury the leucine side chains in the center of these bundles, rather than expose them to aqueous environements. With the leucines buried in the center, the lysines are all aligned on the outer halves of each individual helix (Figure 4.3). This would allow them to interact freely with silica, or, drawing more analogues from the R5 model, perhaps phosphate groups. That may lead to aggregation on an even larger scale - that is, a collection of tetrameric bundles.

The original work by Degrado et al.[20] found that at low concentrations of LKα14 peptide, the CD results predict random coils for the peptide; at higher concentrations, the α-helical character prevails. Thus, the LKα14 peptides appear to be α-helical when they are assembled into aggregates, and those aggregates are believed to be tetrameric bundles with hydrophobic cores. Thus, since NMR data previously documented[59, 12] and measured again here with a different method[102, 103] show that the LKα14 in its lyophilized form is α-helical, we can assume that it is already part of an aggregate. In contrast to R5, the amphiphilic nature and hydrophobic periodicity of the peptide ensure that it is energetically more favorable for the peptide to arrange itself such that there two faces of the helix - a lysine face, and a leucine face.[20] Thus, there is an inherent ability of the LKα14 peptide to form α helices in solution (aggregated α-helices, at that, which R5 lacks, as R5 requires phosphates to exhibit any sort of self-aggregating behavior.[84]

That the LKα14 peptide displays a very similar structure in the -SiO_2 form implies that it is already an aggregate upon interaction with silica, as well. This also limits interpretation of ΔCS values (Figures 4.10 and 4.11) to perturbations from interactions with possible phosphate or silicic acid moieties, as well as any structural rearrangements of the original aggregate that may have formed to accommodate interactions

with these newly introduced negatively charged species.

Based on studies of PLL spherical aggregates (with[78, 104, 107, 56] and without[63] the presence of silica) as referenced in Chapter 3 when hypothesizing the nature of R5 interactions, it is possible that even though the local secondary structure of the LKα14 and R5 peptides are different (the former adopting an α-helical conformation when assembled in aggregates, and the latter adopting a maximumly extended conformation), they self-assemble into larger spherical assemblies, on the order of about 700-1000 peptide molecules per silica nanosphere[53].

The LKα14 peptide already displays evidence of ordered aggregation on a small scale (i.e., tetrameric bundles of single monomers), but in order to produce a spherical aggregate of the size observed with SEM (Figure 4.4), assuming that silica condensation occurs only on the outer edges of a large peptide aggregate, as postulated by other groups[78, 104, 107, 56] and by evidence for the R5 peptide (Chapter 3 of this work), then each individual tetrameric bundle would have to participate in aggregation on another, larger scale - i.e., each tetrameric bundle is again a single unit of a greater self-assembled structure that produces a spherical macrostructure similar to those seen by McKenna et al.[63] with PLL chains.

4.4.2 Tetrameric Bundles of LKα14 and Aggregate-Initiated Silica Condensation

The prevailing model for tetrameric bundles of monomer amphiphilic helices seems to be four helices (either anti-parallel or parallel), with the helical axes arranged at a 20° tilt angle from one another.[20, 39, 108, 42, 21, 38, 88, 16, 89]. The driving force of bundle assembly appears to be that it is much more energetically favorable to bury the hydrophobic leucine side chains, exposing only the hydrophilic lysine chains to aqueous environments.[71, 70]

There has also been some suggestion of knobs-to-holes packing of these tetrameric 4-α-bundles, as first proposed by Crick[19]. Schemas for this kind of regular interdigitizing and packing of apolar side chains (i.e., hydrophobic side chains that would

be buried in the core of these tetramer bundles) have been developed by several groups.[39, 33, 17]. These suggest that the $^{13}C\alpha$ - $^{13}C\beta$ vector of one apolar side chain (i.e., leucine, here) fits into a corresponding $^{13}C\alpha$ - $^{13}C\alpha$ spatial gap in the adjacent helix.[33], as shown in Figure 4.13.

4.4.3 Chemical Shift Analysis of Side Chains

The ΔCS values for side chain ^{13}C spins in LKα14 that are shown in Figures 4.10 and 4.11 have maximum values of 3.03 ppm (including $^{13}C\beta$ - this is K2) and 2.54 ppm excluding β ^{13}C spins (this is at L5). L8 seems to be the least affected, but upon looking at the residue locations along the helix (Figure 4.3), and considering the tetramer aggregate unit arrangement (Figure 4.12), this seems reasonable, as L8 is in the middle of the helix (so it would not experience any possible exposure to solvents that may occur at either terminus of the sequence) and it appears to be buried in the approximate center of the hydrophobic core created by the four monomer helices. Even if phosphate or silicic acid molecules could burrow their way into certain crevices or interstices in a network of tetrameric bundles, it is not likely that they would make it into the very center where the L8 side chain appears to be. Since it is also in the center of the peptide, it is reasonable that it also would not experience a lot of torque on its backbone in either direction, if the peptide experiences more mobility toward the ends.

On the other end, residues L5 and L14 would experience the most possible exposure to solvents and any inorganic species (phosphate ions, silicic acid or silica moieties) that are in solution. L14 was not labeled because it was attached to the resin, but L5 does show significant perturbations in the chemical shifts of the $^{13}C\gamma$ (-2.5 ppm) and $^{13}C\delta$ (-1.2 ppm) spins (Figures 4.10, 4.11).

The next closest to the hydrophobic core boundary's edges are L12 and L7; again, L7 shows significant $^{13}C\gamma$ and $^{13}C\delta$ ΔCS values (1.5 ppm and 1.1 ppm, respectively), which could be a result of its position near the exterior of the hydrophobic core,

right at the transition between hydrophobic and hydrophilic areas of the bundle. The interpretation for L12 is not quite as straightforward; this residue has a position nearer to the terminal end of the sequence, so it is possible that there is more of a backbone effect here. Bearing in mind that each helix is believed to be set at an approximately 20° angle relative to each other, the ends of the helix are slightly offset. L12 shows more backbone movement (more significant ΔCS values at the carbonyl and Cα positions) and then another relatively large change at the Cδ position.

Looking at the lysines, most of them exhibit at least some large perturbation along some aspect of their side chains. K2 shows the largest side chain changes, and K13 shows the largest backbone changes. Both of these are toward the ends, where additional mobility may account for some of the change in movement and possibly conformation.

If the peptide aggregate-silica model on a large scale (greater than local secondary structure, which we have observed is different in R5 and in LKα14, though both preserve these inherent conformations throughout formation of biosilica complexed particles) of LKα14 is similar to that of R5, where the majority of each biosilica nanosphere is peptide (roughly 85%)[52], and silicic acid condenses only on the peptide aggregate sphere's surface to form a coating, then the interactions of peptide functional groups with silicic acid is limited. Since the NMR observed signal must be an ensemble average of all spins in a sample, this very small percentage of silicic acid/silica-affected peptide side chain ^{13}C and ^{15}N spins would not skew the observed peaks in a large manner.

4.4.4 Relaxation Studies of Side Chains

The ^{13}C CP T_1IR experiments on UL7 gave T_1 values in Table 4.5. The values are typical of ^{13}C spins in peptides and proteins in the solid state. It doesn't appear that the ^{13}CO and ^{13}Cα T_1 times are much affected by biosilica complexation; the times for the -SiO_2 spins are in general, slower, but only by about 45 ms (^{13}CO, with

roughly 20-30 ms error) and 40 ms ($^{13}C\alpha$, with roughly 10 ms error). In contrast, the $^{13}C\beta$ spin T_1 slows down by 277 ms (error = 20 ms) between the neat state and the -SiO_2 states. This suggests that while the backbone remains fairly rigid (possibly due to the tetramer bundles fixing helix axes in place), the β of L7 becomes more constrained in motion somehow in the complexed silica form. The ΔCS value for $^{13}C\beta$ at L7 demonstrates nearly negligible change (when accounting for error), which implies that it becomes more fixed in its original position.

Crick's original knobs-in-holes theory discusses the packing of the $^{13}C\alpha$ - $^{13}C\beta$ vector of one apolar side chain (like leucine) fitting into a corresponding hole in the adjacent helix;[19] Harbury et al. further expand on this theory for 4-α-bundles[33], and demonstrate that helical peptides with leucines in the apolar knobs-in-holes positions coordinate perpendicular relative to their neighboring helices to form these 4-α-bundles (Figure 4.13). If this is the case, it is reasonable to observe restricted movement for the $^{13}C\beta$ spin.

If the $^{13}C\gamma$ and $^{13}C\delta$ elements of the leucine side chains are buried in the center of the core, there is a large possibility that they might not be as restricted as the backbone, which is demonstrated by the T_1 values of both the UL7 neat and -SiO_2 peptides - the backbone carbons experience longer T_1 times than $^{13}C\gamma$ and $^{13}C\delta$. In general, these side chain carbons experience more movement, as demonstrated by previous work[13, 109, 6]. The $^{13}C\gamma$ in this case slows down by 53 ms and the $^{13}C\delta$ speeds up by 107 ms. This total picture seems to suggest that the rigid backbone carbons (^{13}CO and $^{13}C\alpha$) do not experience much change in mobility in the transition from neat peptide to silica-complex. Both the $^{13}C\beta$ and $^{13}C\gamma$ spins seem to experience more restriction of motion in the -SiO_2 LKα14 peptide, and perhaps this extra rigidity in the $^{13}C\beta$ - $^{13}C\gamma$ bond allows the $^{13}C\delta$ spin even more freedom of movement.

Comparing the two (UL7 and UL11), it seems that overall, UL11 neat T_1 times are faster than UL7 T_1 times, but become much slower than UL7 in the silica composites. The same phenomenon is also observed here, with the $^{13}C\beta$ slowing down much more

than the $^{13}C\alpha$ spin in the silica composite. Much of the same reasoning can be used to explain this slow $^{13}C\beta$ time. Overall, the slowing down of the times in the silica composite seems to suggest a constricting of the tetrameric bundles when packed into a larger aggregate. Previous work has shown similar results because of peptide aggregation.[109].

4.5 ^{29}Si NMR Analysis of LKα14-SiO$_2$

In addition to SSNMR investigation of dynamics and structure of the peptide, I also did ^{29}Si CPMAS experiments, performed on a Bruker Avance III spectrometer, at a field of 700.18 MHz (^{1}H). This spectrometer was fitted with a 1H-X-Y dual X channel MAS probehead, accommodating 2.5 mm rotors. The Larmor frequency of ^{29}Si at 700 MHz is 139.10 MHz. The ^{1}H 90° time was 6.0 μs, the contact time was 9.0 ms, and the recycle delay was 2.5 s. Multiple experiments were added together for a total of 223248 scans, and the chemical shifts were referenced to Q8M8.

Even with ∼200k scans, signal was low (as with R5, Figure 3.8), again suggesting that there is not very much ^{29}Si material in the rotor. The CP conditions should be similar in both the LKα14 and R5 peptide biosilica composites. We do see the same Q^3 and Q^4 groups here as we did with R5, which implies that the nature of the silica - coordination of the silicon atoms, number of oxygens - is nearly identical in both.

From the ^{29}Si CPMAS spectrum (Figure 4.15), we can identify Q^3 and Q^4 resonances;[25, 26, 9, 11] this corresponds with the findings of Gröger et al.[31] who found that the ^{29}Si DPMAS spectra of diatom cells showed populations of primarily Q^3 and Q^4 groups. This makes sense; the majority of silica moieties in the silica layer would probably be associated with four other SiO- groups, and only the terminal groups at the surface of the silica coating (exposed to the solvent/aqueous environment) would probably be terminated with hydroxy groups.

4.6 Comparison of LKα14 and the R5 peptide

As mentioned in the beginning of this chapter, the LK peptide investigation was launched because of the need to compare the results of R5 to a model peptide system. The use of TALOS-N with R5 was validated by a proof of principle type study with the LKα14 peptide, as the secondary structure of the latter is already well documented on various functionalized hydrophobic and hydrophilic surfaces.[20, 65, 59, 12, 13, 109]. This chemical shift + TALOS-N approach was performed with LKα14 and the results (α-helical conformation) matches the results of previous studies, adding weight to the secondary structure derived with this method for R5 in the previous chapter of this work.

An aspect of the R5 work that is unclear is the nature of aggregation. Since R5 requires the presence of phosphate ions to form aggregates and/or precipitate silica,[53, 84] it is possible that the structure for the neat form of R5 represents monomers that have not formed an ordered network as it would in the silica composite (R5-SiO_2). With the LKα14 peptide, again, it is determined to be α-helical primarily in aggregate form (with the hydrophobic interaction being the driving force of helical conformation for this amphiphilic peptide), so the predicted structure for the neat LKα14 peptide determined in this work confirm that it exists as part of some probably ordered self-assembled structure, even in the neat form. This is another possible difference between the R5 and LKα14 peptides. With the LKα14 peptide determined to be both α-helical and in an aggregate in both the neat and -SiO_2 forms, it is possible to attribute all ΔCS values between the two to interactions with external moieties i.e., phosphates or silicic acid/silica) or conformational changes induced by interactions with the same. This makes interpretation of the ΔCS values for the LKα14 peptide more straightforward and eliminates one possible factor. Of course, it must be said that it is possible that even though the LKα14 peptide forms tetrameric bundles in both cases, the larger scale aggregation is different for the measured neat

form and -SiO_2 forms, as the presence of phosphate ions in the latter (but not the former) could alter the higher order assembly in some way.

Another benefit of the simplified LKα14 peptide in contrast to the R5 peptide is that there are only two types of residues: hydrophobic, apolar leucines, and basic, charged lysine residues. It has been shown that the leucines orient themselves toward a hydrophobic surface like polystyrene,[13, 109, 65], while the lysines orient themselves toward hydrophilic surfaces like silica[65] and carboxylated polystyrene[13]. Now that it is has been determined that the LKα14 peptide retains its helical structure in these co-precipitated biosilica particles, it is possible to incorporate previous work on the peptide[13, 109, 65, 59] and suggest that the lysines are primarily the the major players in the LKα14 peptide's interactions with silicic acid to form biosilica composites. Lysines are part of the R5 sequence, as well, but since there are other residues (R5 contains serines, a single tyrosine, lysines, arginines, glycines, one isoleucine and one leucine), it is not as easy to assign cause and effect of observations to single amino acids.

Another conclusion to be drawn from comparison of the two peptides and their corresponding induced silica nanoparticles is that while the R5 peptide and the LKα14 peptide were shown to have different secondary structures, they both produce silica particles of similar shape and size (i.e., spheres of 500-800 nm diameters, as shown in Figure 4.4). This implies that local secondary structure of individual peptides is not the main factor controlling silica morphology, although it has an indirect effect, as secondary structure controls mechanisms of self-assembly and aggregation on a larger scale. In this sense, chain length and amino acid composition remain important, especially since the driving force of self-assemblies (intermolecular hydrogen bonding and electostatic interactions) appears to also contribute to interactions with negatively charged phosphate or silica groups. This multi-tasking of amino acids is quite efficient.

4.7 Future Goals

Here, as with R5, we focused mostly on determining the secondary structure of the LKα14 peptide in its biosilica form. In order to put merit in aggregation models, it's necessary to first ascertain whether the characteristic tendency of LKα14 to take a α-helical conformation in solution, at interfaces, and in aggregates, also held true throughout interactions with silicic acid and integration into a biosilica composite. Through analysis of SSNMR 1D and 2D ^{13}C and ^{15}N spectra, it appears that this is indeed the case. In this chapter, I discussed possibilities as to the nature of the self-assemblies of LKα14, drawing upon models first proposed by Degrado et al.[20] in the original work describing the design and characterization of these amphiphilic peptides. These self-aggregation models propose that the LKα14 peptide forms tetrameric bundles of helices in aqueous environments. It is unknown whether these tetrameric bundles also prevail as individual units of a greater aggregate structure in biosilica composites.

In order to investigate this, as with the R5 peptide, it is possible to perform ^{13}C$\{^{31}$P$\}$ REDOR[32, 73, 74] to determine whether the phosphate groups from the PBS buffer used with LK peptides contributes to the greater macrostructure aggregation observed by Mckenna et al. for PLL chains[63]; we believe that these spheres are also the prevailing form of large-scale aggregation with LKα14, which shares many similarities with PLL chains. Since the LKα14 peptide, unlike R5, already has a stable secondary structure configuration (and even perhaps a small-scale aggregation in tetramer bundles), phosphate groups do not appear to be necessary for the initial small scale aggregation. Degrado et al. observed tetrameric bundles in aqueous, phosphate-free environments[20]. It is possible, however, that the presence of phosphates coordinate the larger scale spherical aggregate that is used as a template for silica condensation. ^{13}C$\{^{31}$P$\}$ REDOR of side chains (particularly lysyl side chains) would demonstrate the plausibility of this theory; if there are contacts between ly-

sine side chains and phosphate groups, then it would seem likely that phosphates are buried in the peptide aggregate network, and are forming possible cross-links between individual tetrameric bundles or individual peptide strands.

Fujimoto et al.[29] discovered that lysines at certain intervals in short α-helical peptides (i.e., *i, i + 4*) serve as points of contact for small cross-linking molecules; these links between intramolecular lysines serve as additional stabilization for these small α-helices. It is possible that the lysines in LKα14 could act similarly as points of contact with intermolecular cross-linking, thus creating large networks of peptides or tetrameric bundles that could form aggregates on the order of the final silica nano particle (diameters ca. 500-800 nm, as in Figure 4.4).

As with R5, to address the issue of surfacial silica-peptide interactions versus more well-distributed interactions throughout the bulk peptide, ^{15}N{^{29}Si} REDOR[32] can be performed to elucidate any interactions between ^{15}N spins on lysyl side chains, and ^{29}Si spins in silica. It might also be interesting to perform T_1 measurements with singly-labeled K samples, to probe the dynamics and relative motions and timescales of the lysyl side chains, as these are the ones that are believe to be associated with both aggregate behavior (possible phosphate or intermolecular peptide interactions) as well as serve as nucleation points for silica condensation. The ^{13}C CPMAS T1IR experiments performed with UL7 could be repeated for various K residues in the sequence; it could be interesting to start with K2 or K13, which show the largest ΔCS values (Figure 4.11) and perhaps K6, which is toward the middle of the sequence and shows decidedly less perturbations in chemical shift between the neat and -SiO_2 forms of the peptide.

It might also be interesting to create mutants of the LKα14 peptide, as done by Somorjai and coworkers for SFG experiments with peptides on hydrophobic and hydrophilic surfaces.[40, 82, 117] Alanines share the apolar nature of leucines, but have much smaller side chains (a single methyl group), and arginines share the basic nature and charged side group of lysines. LR or AK peptides could be synthesized

and co-precipitated with silica in the same manner outlined in this work (Appendix A), imaged with SEM, and then subjected to the same SSNMR experiments covered here. Mutant work is useful for singling out the purpose and interactions of a single residue in a peptide; with the substitution of another similar - or completely different (e.g., acidic for basic) residues, it is possible to ascertain the role of the wild type amino acid. This is a common method for elucidating the function of specific residues in proteins and peptides.

In addition to this, returning to Figure 4.5 and result of different LK peptides precipitating different silica morphologies, it would be interesting to duplicate the experiments with the $LK3_{10}$ peptide to do an analogous study. Since the ultimate silica nanoparticles do not share an overall shape with either the LKα14 peptide or R5, presumably it aggregates in a different fashion, and in order to investigate possible mechanisms of self-assembly, it is again first necessary to establish what the secondary structure of the $LK3_{10}$ peptide is when it is part of a biosilica composite (and similarly, down the line, possibly LKβ11, as well). SSNMR methods of measuring tertiary or quaternary large-scale structures largely depend on first establishing a model to test,[4, 30] so this preliminary step is crucial. It is not possible to create a plausible large-scale aggregation model if the secondary structure is unknown.

The object of further work in this area is to be able to put forth a relatively well-defined model of the peptide-silica interactions on the molecular level, to introduce reasons for the resulting specificity of size and shape of the silica particles, and decipher the driving forces behind peptide self-assembly on the scale that we see with SEM - that is, the larger aggregates on the nanoscale, not just on the order of tetrameric bundles of individual LKα14 peptides. Evidence from the work of others in this field suggests a model in which peptides or polymers self-assemble to form a large sphere on the order of several hundreds of nanometers (perhaps with the aid of interlocking phosphate groups), and then once this organic macrostructure is assembled, silicic acid condenses on the surface to form a thin coating of silica.

Further investigations through NMR and other experimental methods by this group endeavor to flesh out aspects of this proposed model, starting with a series of dipolar recoupling measurements which serve to map out proximities of specific nuclei (e.g., whether the ^{31}P in a phosphate group is anywhere near the ^{13}Cϵ in a lysyl side chain), which could either support or invalidate certain assumptions made in the development of this model. Closer examination of dynamics and molecular motion timescales by T_1 measurement of side chain ^{13}C and even ^{15}N spins can help elucidate physical arrangements of both peptides and surrounding entities, casting light on which side chains or local areas of the peptide are subject to more motional restrictions; this can add to arguments that describe local environments and induced constraints on a side chain. In tandem, as demonstrated previously with the statherin-HAP protein-biomineral system[73, 74, 85], the results of numerous and varied SSNMR experiments can bring vast insight to the molecular level interactions of peptides and biomineral systems.

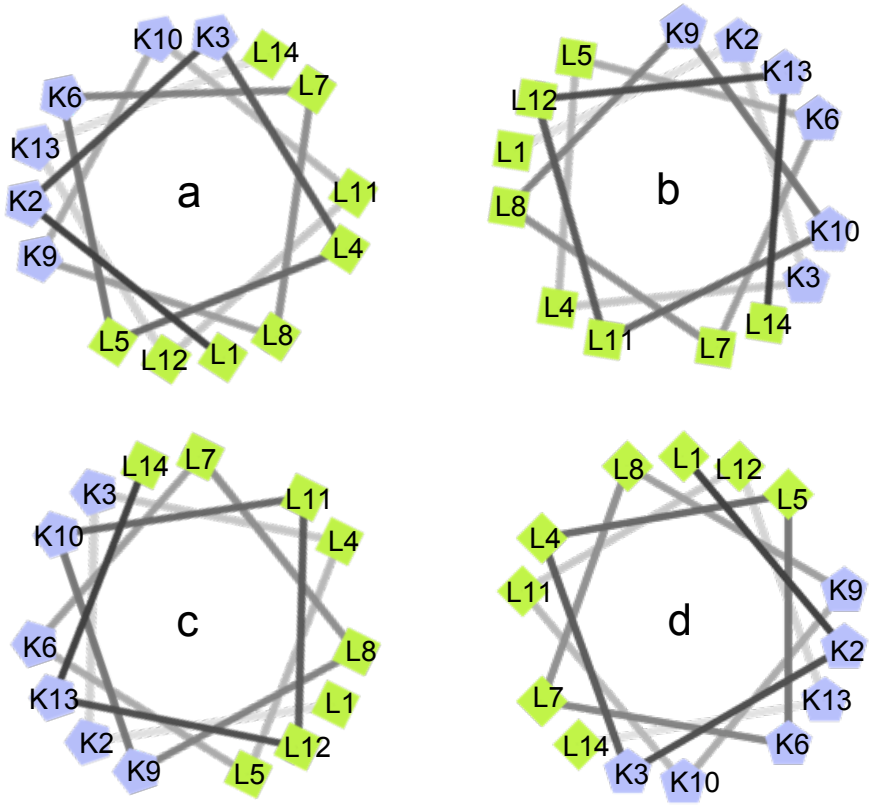

Figure 4.12: This figure shows one possible arrangement of individual LKα14 monomers to form a tetramer bundle; helices *a* and *d* have the N-terminus pointing out of the page, *b* and *c* have the C-terminus pointing out of the page (helices are in an anti-parallel alignment). Leucine residues are pointed inward to form a hydrophobic core, protected from interactions with water. Lysyl residues are thus free to interact with the aqueous environment. There is some freedom in the exact rotation of each helix, and alternate arrangements could have all four helices in parallel. This figure was made with the assistance of an online helical projection tool[2].

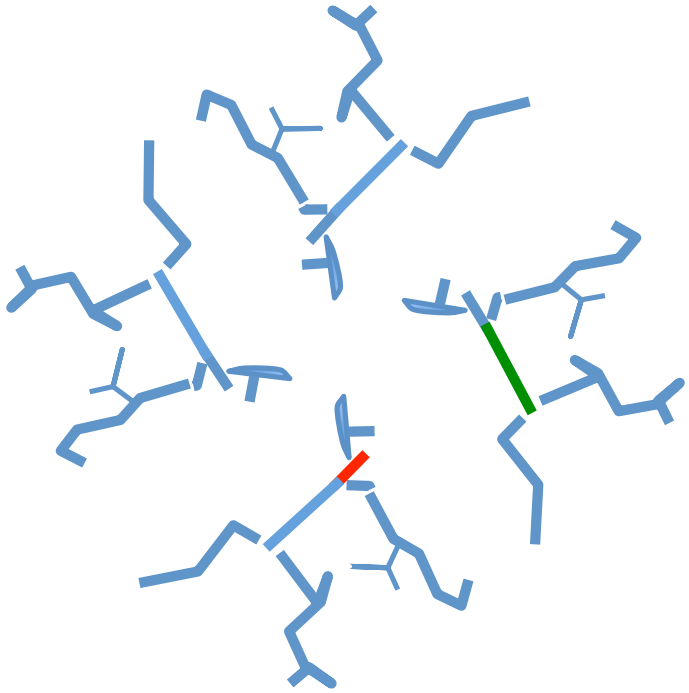

Figure 4.13: Model proposed by Harbury et al. for perpendicular packing of leucine side chains into a 4-α-bundle. This is a view down the barrel of an alpha helix; "knobs" of side chains (red, $^{13}C\alpha$ - $^{13}C\beta$ vector) fit into corresponding spatial "holes" (green). Figure adapted from reference[33].

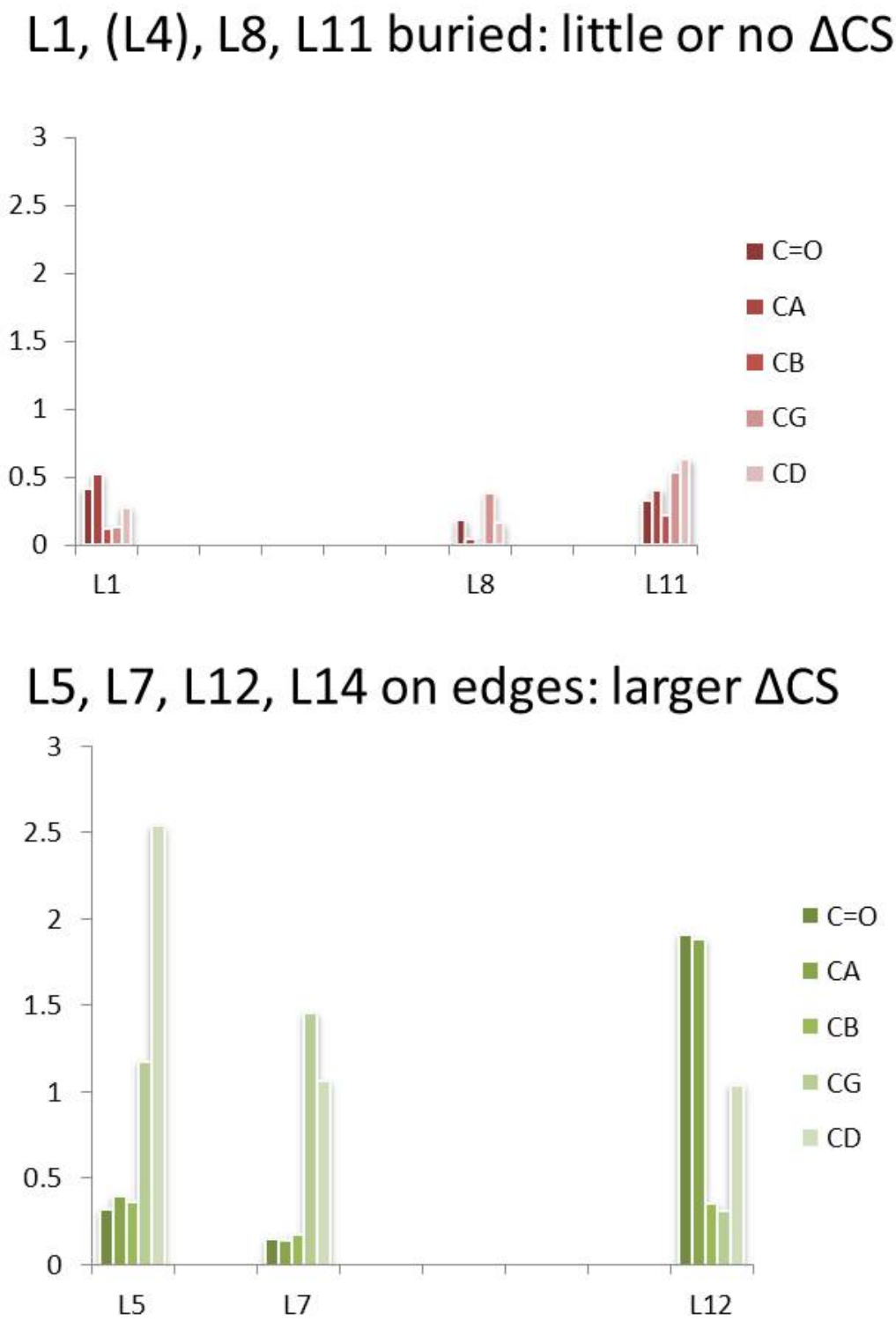

Figure 4.14: Comparison of ΔCS plots for leucines show that there is minimal perturbation of chemical shift between the neat and the -SiO_2 states of LKα14 for L1, L8 and L11, which should all be buried in the center of the tetrameric bundle (Figure 4.12). Conversely, perturbations at L5, L7, and L12 are considerably larger, as these leucine side chains could possibly be exposed to solvent or other molecules, being at the edges of the hydrophobic core.

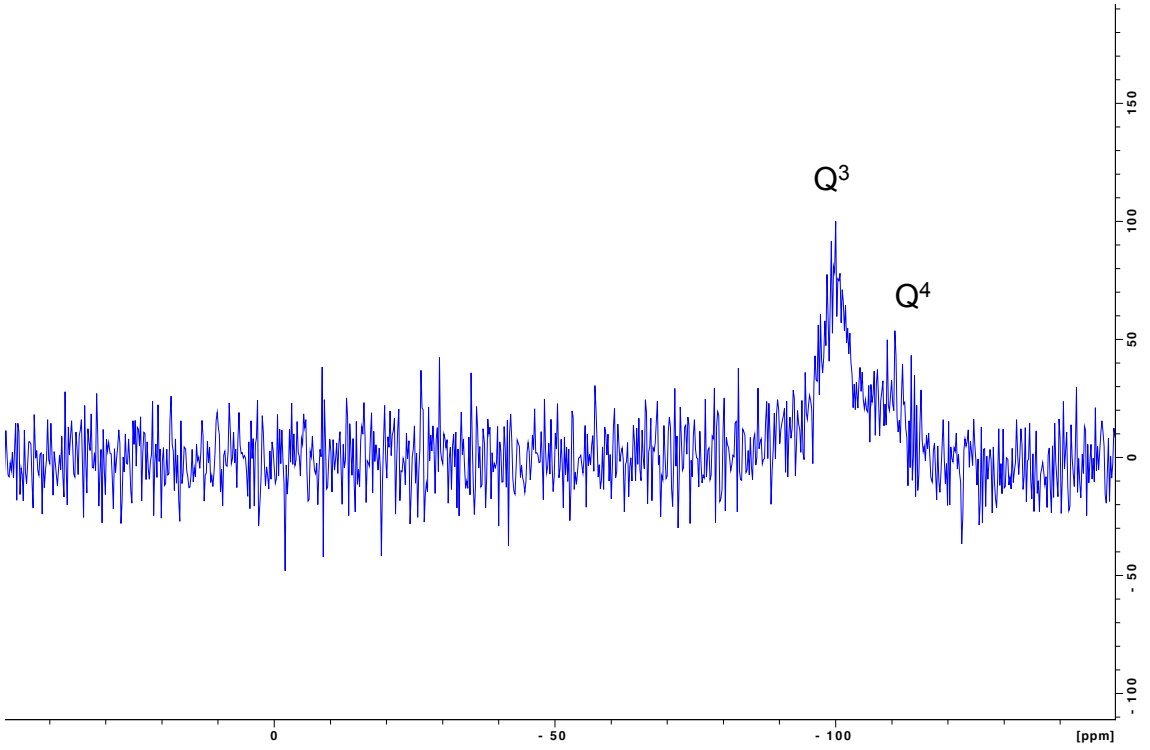

Figure 4.15: ^{29}Si CPMAS spectrum of LKα14 SiO_2 particles. Poor resolution even with 223248 scans at 700 MHz suggests that the amount of ^{29}Si present in the sample is very minimal. The Q^n groups signify Si bondings with n number of shared tetrahedra; n can range from 0 (a single Si bonded to four O atoms - essentially, monosilicic acid) to 4.[9] A pictorial representation of Q^n forms and their characteristic ^{29}Si NMR chemical shifts is given in Figure 3.9.[25, 26, 31, 9]

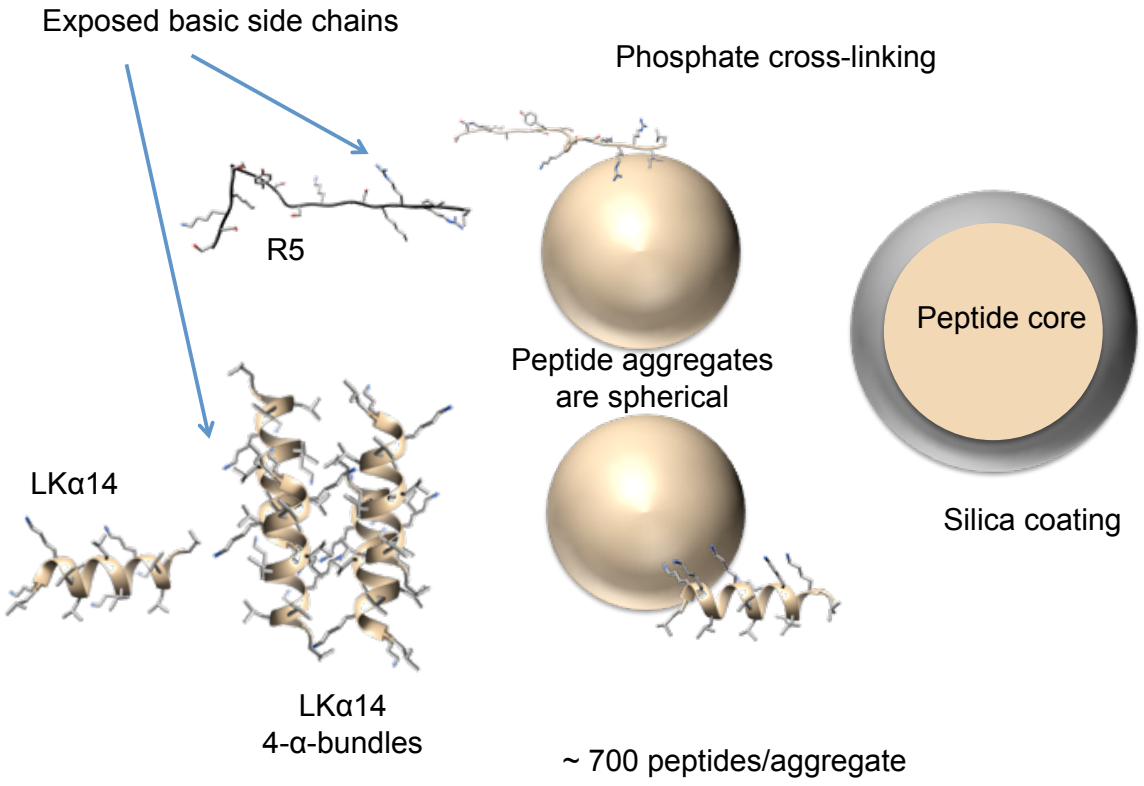

Figure 4.16: The proposed model for the R5 peptide and the LKα14 peptide. It is supposed that both peptides form large spherical aggregates, through a network of cross-linking phosphates and arginines and lysines in both peptides. Exposed side chains on the surface of the spherical aggregates serve as nucleation points for silica condensation, forming a silica coating on the surfaces of the spheres.

BIBLIOGRAPHY

[1] *Novabiochem Catalog*. 2002-2003.

[2] D. Armstrong and R. Zidovetzki. Helical wheel projections, July 2013.

[3] E. Atherton and R. Sheppard. *Solid Phase Peptide Synthesis, A Practical Approach*. Oxford University Press, Oxford, 1989.

[4] J. Balbach, Y. Ishii, O. Antzutkin, R. Leapman, N. Rizzo, F. Dyda, J. Reed, and R. Tycko. Amyloid fibril formation by aβ16-22, a seven-residue fragment of the Alzheimer's β-amyloid peptide, and structural characterization by solid state NMR. *Biochemistry*, 39:13748–59, 2000.

[5] H. Bamnolker, B. Nitzan, S. Gura, and S. Margel. New solid and hollow, magnetic and non-magnetic, organic-inorganic monodispersed hybrid microspheres: synthesis and characterization. *J. Mater. Sci. Lett.*, 16(16):1412–1415, 1997.

[6] L. S. Batchelder, C. E. Sullivan, L. W. Jelinski, and D. A. Torchia. Characterization of leucine side-chain reorientation in collagen fibrils by solid-state 2H NMR. *Proc. Natl. Acad. USA Biol. Struct.*, 79(2):386–389, 1982.

[7] I. Ben Shir, S. Kababya, T. Amitay-Rosen, Y. S. Balazs, and A. Schmidt. Molecular level characterization of the inorganic-bioorganic interface by solid state NMR: Alanine on a silica surface, a case study. *J. Phys. Chem. B.*, 114(18):5989–5996, 2010.

[8] I. Ben Shir, S. Kababya, and A. Schmidt. Binding specificity of amino acids to amorphous silica surfaces: Solid-state NMR of glycine on SBA-15. *J. Phys. Chem. C.*, 116(17):9691–9702, 2012.

[9] H. E. Bergna and W. O. Roberts. *Colloidal Silica Fundamentals and Applications*. Boca Raton, FL CRC Taylor & Francis, 2006.

[10] F. Bernstein, T. Koetzle, G. Williams, E. Meyer Jr., M. Brice, J. Rodgers, O. Kennard, T. Shimanouchi, and M. Tasumi. The protein data bank: A computer-based archival file for macromolecular structures. *J. Mol Biol.*, 112, 1977.

[11] R. Bertermann, N. Kröger, and R. Tacke. Solid-state 29Si MAS studies of diatoms: structural characterization of biosilica deposits. *Anal. Bioanal. Chem.*, 375(5):630–634, 2003.

[12] P. V. Bower, E. A. Louie, J. R. Long, P. S. Stayton, and G. P. Drobny. Solid-state NMR structural studies of peptides immobilized on gold nanoparticles. *Langmuir*, 21(7):3002–3007, 2005.

[13] N. F. Breen, T. Weidner, K. Li, D. G. Castner, and G. P. Drobny. A solid-state deuterium NMR and sum-frequency generation study of the side-chain dynamics of peptides adsorbed onto surfaces. *J. Am. Chem. Soc.*, 131(40):14148–14149, 2009.

[14] L. Brott, R. Naik, D. Pikas, S. Kirkpatrick, D. Tomlin, P. Whitlock, S. Clarson, and M. Stone. Ultrafast holographic nanopatterning of biocatalytically formed silica. *Nature*, 413:291–293, 2003.

[15] E. Brunner, K. Lutz, and M. Sumper. Biomimetic synthesis of silica nanospheres depends on the aggregation and phase separation of polyamines in aqueous solution. *Phys. Chem. Chem. Phys.*, 6:854–857, 2004.

[16] J. W. Bryson, S. F. Betz, H. S. Lu, D. J. Suich, H. X. X. Zhou, K. T. Oniel, and W. F. Degrado. Protein design - a hierarchal approach. *Science*, 270(5238):935–941, 1995.

[17] C. Chothia, M. Levitt, and D. Richardson. Helix to helix packing in proteins. *J. Mol. Biol.*, 145(1):215–250, 1981.

[18] T. Coradin, O. Durupthy, and J. Livage. Interactions of amino-containing peptides with sodium silicate and colloidal silica: A biomimetic approach of silicification. *Langmuir*, 18(6):2331–2336, 2002.

[19] F. H. Crick. The packing of alpha-helices - simple coiled-coils. *Acta Crystallogr.*, 6:689–97, 1953.

[20] W. F. DeGrado and J. D. Lear. Induction of peptide conformation at apolar water interfaces. 1. a study with model peptides of defined hydrophobic periodicity. *J. Am. Chem. Soc.*, 107(25):7684–7689, 1985.

[21] W. F. DeGrado, C. M. Summa, V. Pavone, F. Nastri, and A. Lombardi. De novo design and structural characterization of proteins and metalloproteins. *Ann. Rev. Biochem.*, 68:779–819, 1999.

[22] W. F. DeGrado, Z. R. Wasserman, and J. D. Lear. Protein design, a minimalist approach. *Science*, 243(4891):pp. 622–628, 1989.

[23] M. Dickerson, S. Jones, Y. Cai, G. Ahmad, R. Naik, N. Kröger, and K. Sandhage. Identification and design of peptides for the rapid formation of nanoparticulate tio_2 from aqueous solutions at room temperature. *Chem. Mater.*, 20:1578–84, 2008.

[24] M. Duer. *Introduction to Solid-State NMR Spectroscopy*. Wiley, 2005.

[25] G. Engelhardt. Multinuclear solid-state NMR in silicate and zeolite chemistry. *Trends Anal. Chem.*, 8(9):343–367, 1989.

[26] G. Engelhardt and D. Michel. *High-resolution Solid-state NMR of silicates and zeolites*. Wiley VCH, Chichester, 1987.

[27] E. K. Esbjörner, P. Lincoln, and B. Nordén. Counterion-mediated membrane penetration: Cationic cell-penetrating peptides overcome born energy barrier by ion-pairing with phospholipids. *Biochimica et Biophysica Acta (BBA) - Biomembranes*, 1768(6):1550 – 1558, 2007.

[28] V. Fernandez, J. Reimer, and M. Denn. Magnetic resonance studies of polypeptides adsorbed on silica and hydroxyapatite surfaces. *J. Am. Chem. Soc.*, 114:9634–42, 1992.

[29] K. Fujimoto, M. Kajino, and M. Inouye. Development of a series of cross-linking agents that effectively stabilize α-helical structures in various short peptides. *Chemistry*, 14:857–63, 2008.

[30] D. M. Gregory, T. L. Benzinger, T. S. Burkoth, H. Miller-Auer, D. G. Lynn, S. C. Meredith, and R. E. Botto. Dipolar recoupling NMR of biomolecular self-assemblies: determining inter- and intrastrand distances in fibrilized Alzheimer's -amyloid peptide. *Solid State Nucl. Magn. Reson.*, 13(3):149 – 166, 1998.

[31] C. Gröger, M. Sumper, and E. Brunner. Silicon uptake and metabolism of the marine diatom thalassiosira pseudonana: Solid-state 29si nmr and fluorescence microscopic studies. *J. Struct. Biol.*, 161:55–63, 2008.

[32] T. Gullion and J. Schaefer. Development of redor rotational-echo double-resonance NMR. *J. Magn. Reson.*, 81:196–200, 1989.

[33] P. B. Harbury, P. S. Kim, and T. Alber. Crystal-structure of an isoleucine-zipper trimer. *Nature*, 371(6492):80–83, SEP 1 1994.

[34] J. D. Hartgerink, E. Beniash, and S. I. Stupp. Self-assembly and mineralization of peptide-amphiphile nanofibers. *Science*, 294(5547):1684–1688, 2001.

[35] S. R. Hartmann and E. L. Hahn. Nuclear double resonance in the rotating frame. *Phys. Rev.*, 128:2042–2053, Dec 1962.

[36] K. M. Hawkins, S. S.-S. Wang, D. M. Ford, and D. F. Shantz. Poly-l-lysine templated silicas: Using polypeptide secondary structure to control oxide pore architectures. *J. Am. Chem. Soc.*, 126(29):9112–9119, 2004. PMID: 15264846.

[37] M. Hildebrand. Diatoms, biomineralization processes, and genomics. *Chem. Rev.*, 108(11):4855–4874, 2008.

[38] R. B. Hill and W. F. DeGrado. Solutions structure of alpha D-2, a nativelike de novo designed protein. *J. Am. Chem. Soc.*, 120(6):1138–1145, FEB 18 1998.

[39] S. P. Ho and W. F. Degrado. Design of a 4-helix bundle protein - synthesis of peptides which self-associate into a helical protein. *J. Am. Chem. Soc.*, 109(22):6751–6758, 1987.

[40] G. J. Holinga, R. L. York, R. M. Onorato, C. M. Thompson, N. E. Webb, A. P. Yoon, and G. A. Somorjai. An SFG Study of Interfacial Amino Acids at the Hydrophilic SiO2 and Hydrophobic Deuterated Polystyrene Surfaces. *J. Am. Chem. Soc.*, 133(16):6243–6253, 2011.

[41] S. Hovmöller, T. Zhou, and T. Ohlson. Conformations of amino acids in proteins. *Acta Crystallographica Section D*, 58:768–776, 2002.

[42] K. P. Howard, J. D. Lear, and W. F. DeGrado. Sequence determinants of the energetics of folding of a transmembrane four-helix-bundle protein. *Proc. Natl. Acad. Sci. USA*, 99(13):8568–8572, 2002.

[43] W. Huang, G. Varani, and G. P. Drobny. Interactions of protein side chains with RNA defined with REDOR solid state NMR. *J. Biomol. NMR*, 51(3):347–356, 2011.

[44] R. Iler. *The Chemistry of Silica: Solubility, Polymerization, Colloid and Surface Properties, and Biochemistry.* John Wiley and Sons, 1979.

[45] S. Kaplan, H. Resing, and J. Waugh. 13c NMR chemical shift anisotropy for benzene adsorbed on charcoal and silica gel. *J. Chem. Phys.*, 59:5681–87, 1973.

[46] S. Kates. *Solid phase synthesis: a practical guide*. Marcel Dekker, New York, 2000.

[47] E. Kharlampieva, C. Jun, V. Kozlovskaya, and V. Tsukruk. Secondary structure of silaffin at interfaces and titania formation. *J. Mater. Chem.*, 20:5242–50, 2010.

[48] E. Kharlampieva, T. Tsukruk, J. Slocik, H. Ko, N. Poulsen, R. Naik, N. Kröger, and V. Tsukruk. Bioenabled surface-mediated growth of titania nanoparticles. *Adv. Mater.*, 20:3274–79, 2008.

[49] M. Knecht and D. Wright. Amine-terminated dendrimers as biomimetic templates for silica nanosphere formation. *Langmuir*, 20:4728–32, 2004.

[50] M. R. Knecht and D. W. Wright. Functional analysis of the biomimetic silica precipitating activity of the R5 peptide from *Cylindrotheca Fusiformis*. *Chem. Commun.*, pages 3038–39, 2003.

[51] N. Kröger, R. Deutzmann, C. Bergdorf, and M. Sumper. Species-specific polyamines from diatoms control silica morphology. *Proc. Natl. Sci. USA*, 97:14133–14138, 2000.

[52] N. Kröger, R. Deutzmann, and M. Sumper. Polycationic peptides from diatom biosilica that direct silica nanosphere formation. *Science*, 286:1129–32, 1999.

[53] N. Kröger, S. Lorenz, E. Brunner, and M. Sumper. Self-assembly of highly phosphorylated silaffins and their function in biosilica morphogenesis. *Science*, 298(5593):584–586, 2002.

[54] N. Kröger and N. Poulsen. Diatoms from cell wall biogenesis to nanotechnology. *Annu. Rev. Genen.*, 42(1):83–107, 2008. PMID: 18983255.

[55] N. Kröger and K. Sandhage. From diatom biomolecules to bioinspired syntheses of silica- and titania-based materials. *MRS Bulletin*, 35:122–26, 2010.

[56] N. Li, X. Zhang, Q. Wang, F. Wang, and P. Shen. Biomimetic synthesis of silica hollow spheres using poly (l-lysine) and mechanism research. *RSC Adv.*, 2:3288–3297, 2012.

[57] M. Liang, S. Patwardhan, E. Danilovtseva, V. Annenkov, and C. P. Perry. Imidazole catalyzed silica synthesis: Progress toward understanding the role of histidine in biosilicification. *J. Mater. Res.*, 24:1700–08, 2009.

[58] K. D. Lobel, J. K. West, and L. Hench. Computational model for protein mediated biomineralization of the diatom frustule. *Mar. Biol.*, 126:353–360, 1996.

[59] J. R. Long, N. Oyler, G. P. Drobny, and P. S. Stayton. Assembly of α-helical peptide coatings on hydrophobic surfaces. *J. Am. Chem. Soc.*, 124(22):6297–6303, 2002.

[60] P. Lopez, C. Gautier, J. Livage, and T. Coradin. Mimicking biogenic silica nanostructures formation. *Curr. Nanosci.*, 1:73–83, 2005.

[61] S. Mann. Molecular tectonics in biomineralization and biomimetic materials chemistry. *Nature*, 365:499–505, 1993.

[62] D. L. Masica and J. J. Gray. Solution- and adsorbed-state structural ensembles predicted for the statherin-hydroxyapatite system. *Biophys. J.*, 96(8):3082 – 3091, 2009.

[63] B. J. McKenna, H. Birkedal, M. H. Bartl, T. J. Deming, and G. D. Stucky. Micrometer-sized spherical assemblies of polypeptides and small molecules by acid-base chemistry. *Angew. Chem. Int. Ed.*, 43(42):5652–5655, 2004.

[64] M. Meng, L. Stievano, and J.-F. Lambert. Adsorption and thermal condensation mechanisms of amino acids on oxide supports. 1. glycine on silica. *Langmuir*, 20:914–23, 2004.

[65] O. Mermut, D. C. Phillips, R. L. York, K. R. McCrea, R. S. Ward, and G. A. Somorjai. In situ adsorption studies of a 14-amino acid leucine-lysine peptide onto hydrophobic polystyrene and hydrophilic silica surfaces using quartz crystal microbalance, atomic force microscopy, and sum frequency generation vibrational spectroscopy. *J. Am. Chem. Soc.*, 128(11):3598–3607, 2006.

[66] G. Metz, X. Wu, and S. Smith. Ramped-amplitude cross polarization in magic-angle spinning nmr. *J. Magn. Reson.*, 110:219–27, 1994.

[67] D. Mitchell, D. T. Kim, L. Steinman, C. G. Fathman, and J. B. Rothbard. Polyarginine enters cells more efficiently than other polycationic homopolymers. *J. Pept. Res.*, 56:318–25, 2000.

[68] T. Mizutani, H. Nagase, and H. Ogoshi. Silicic acid polymerization catalyzed by amines and polyamines. *Bull. Chem. Soc. Jpn.*, 27:133–34, 1998.

[69] A. Mort and T. Lamport. Anhydrous hydrogen-fluoride deglycosylates glycoproteins. *Anal. Biochem.*, 82:289, 1977.

[70] M. Mutter, R. Hersperger, K. Gubernator, and K. Muller. The construction of new proteins 5. A template-assembled synthetic protein (TASP) containing both a 4-helix bundle and beta-barrel-like structure. *Proteins Struct. Funct. Bioinf.*, 5(1):13–21, 1989.

[71] M. Mutter and S. Vuilleumier. A chemical approach to protein design - template-assembled synthetic proteins (TASP). *Angew. Chem. Int. Ed.*, 28(5):535–554, 1989.

[72] R. R. Naik, P. W. Whitlock, F. Rodriguez, L. L. Brott, D. D. Glawe, S. J. Clarson, and M. O. Stone. Controlled formation of biosilica structures in vitro. *Chem. Commun.*, 0:238–239, 2003.

[73] M. Ndao, J. T. Ash, N. F. Breen, G. Goobes, P. S. Stayton, and G. P. Drobny. A 13c31P redor nmr investigation of the role of glutamic acid residues in statherin-hydroxyapatite recognition. *Langmuir*, 25(20):12136–12143, 2009.

[74] M. Ndao, J. T. Ash, P. S. Stayton, and G. P. Drobny. The role of basic amino acids in the molecular recognition of hydroxyapatite by statherin using solid state nmr. *Surf. Sci.*, 604(1516):L39 – L42, 2010.

[75] D. L. Nelson and M. M. Cox. *Lehninger Principles of Biochemistry*. W.H. Freeman and Company, 2005.

[76] T. Oas, R. Griffin, and M. Levitt. Rotary resonance decoupling of dipolar interactions in solid-state nuclear magnetic resonance spectroscopy. *J. Chem. Phys.*, 89:692, 1988.

[77] S. Paramonov, H. Jun, and J. Hartgerink. Self-assembly of peptide-amphiphile nanofibers: The roles of hydrogen bonding and amphiphilic packing. *J. Am. Chem. Soc.*, 128(22):7291–7298, JUN 7 2006.

[78] S. Patwardhan, R. Maheshwari, N. Mukherjee, K. Kiick, and S. Clarson. Conformation and assembly of polypeptide scaffolds in templating the synthesis of silica: An example of a polylysine macromolecular switch. *Biomacromolecules*, 7:491–7, 2006.

[79] S. V. Patwardhan, N. Mukherjee, M. Steinitz-Kannan, and S. J. Clarson. Bioinspired synthesis of new silica structures. *Chem. Commun.*, 0:1122–1123, 2003.

[80] L. Pauling, R. B. Corey, and H. Branson. The structure of proteins. *Proc. Natl. Acad. Sci.*, 37:205–11, 1951.

[81] E. Pettersen, T. Goddard, C. Huang, G. Couch, D. Greenblatt, E. Meng, and T. Ferrin. UCSF chimera - a visualization system for exploratory research and analysis. *J. Comput. Chem.*, 25:1605–12, 2004.

[82] D. C. Phillips, R. L. York, O. Mermut, K. R. McCrea, R. S. Ward, and G. A. Somorjai. Side chain, chain length, and sequence effects on amphiphilic peptide adsorption at hydrophobic and hydrophilic surfaces studied by sum-frequency generation vibrational spectroscopy and quartz crystal microbalance. *J. Phys. Chem. C.*, 111(1):255–261, 2007.

[83] G. N. Ramachandran, C. Ramakrishnan, and V. Sasisekharan. Stereochemistry of polypeptide chain configurations. *J. Mol. Biol.*, 7:95–99, 1963.

[84] F. Rodríguez, D. Glawe, R. Naik, K. Hallinan, and M. Stone. Study of the chemical and physical influences upon in vitro peptide-mediated silica formation. *Biomacromolecules*, 5:261–65, 2004.

[85] A. Roehrich, J. Ash, A. Zane, D. L. Masica, J. J. Gray, G. Goobes, and G. Drobny. *Solid-State NMR Studies of Biomineralization Peptides and Proteins*, chapter 5, pages 77–96. American Chemical Society, 2012.

[86] H. Saito. Conformation-dependent 13C chemical shifts: A new means of conformational characterization as obtained by high-resolution solid-state 13c NMR. *Magn. Reson. Chem.*, 24, 1986.

[87] K. A. Sannes-Lowery, P. Hu, D. P. Mack, H.-Y. Mei, and J. A. Loo. HIV-1 tat peptide binding to TAR RNA by electrospray ionization mass spectrometry. *Anal. Chem.*, 69(24):5130–5135, 1997.

[88] C. Schafmeister, S. LaPorte, L. Miercke, and R. Stroud. A designed four helix bundle protein with native-like structure. *Nature Struct. Biol.*, 4(12):1039–1046, DEC 1997.

[89] C. E. Schafmeister, L. J. W. Miercke, and R. M. Stroud. Structure at 2.5 angstrom of a designed peptide that maintains solubility of membrane-proteins. *Science*, 262(5134):734–738, OCT 29 1993.

[90] N. Schmidt, A. Mishra, G. H. Lai, and G. C. L. Wong. Arginine-rich cell-penetrating peptides. *FEBS lett.*, 584:1806–13, 2010.

[91] K. Schug and W. Lindner. Noncovalent binding between guanidinium and anionic groups: Focus on biological- and synthetic-based arginine/guanidinium interactions with phosph[on]ate and sulf[on]ate residues. *Chem. Rev.*, 105(1):67–113, 2005.

[92] S. Sewell and D. Wright. Biomimetic synthesis of titanium dioxide utilizing the R5 peptide derived from cylindrotheca fusiformis. *Chem. Mater.*, 18:3108–3113, 2006.

[93] W. J. Shaw, J. R. Long, A. A. Campbell, P. S. Stayton, and G. P. Drobny. A solid state NMR study of dynamics in a hydrated salivary peptide adsorbed to hydroxyapatite. *J. Am. Chem. Soc.*, 122(29):7118–7119, 2000.

[94] Y. Shen and A. Bax. Protein backbone and sidechain torsion angles predicted from NMR chemical shifts using artificial neural networks. *J. Biomol. NMR*, pages 1–15, 2013.

[95] Y. Shen, F. Delaglio, G. Cornilescu, and A. Bax. TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *J. Biomol. NMR*, 44:213–223, 2009.

[96] C. Slichter. *Principles of Magnetic Resonance. With Examples from Solid State Physics. [A Reduced Photographic Reprint of the Edition of 1963.].* New York; & John Weatherhill: Tokyo; printed in Japan, 1964.

[97] Y. Su, T. Doherty, A. J. Waring, P. Puchala, and M. Hong. Roles of Arginine and Lysine Residues in the Translocation of a Cell-Penetrating Peptide from C-13, P-31, and F-19 Solid-State NMR. *Biochemistry*, 48(21):4587–4595, JUN 2 2009.

[98] M. Sumper. A phase separation model for the nanopatterning of diatom biosilica. *Science*, 295(5564):pp. 2430–2433, 2002.

[99] M. Sumper and N. Kroger. Silica formation in diatoms: the function of long-chain polyamines and silaffins. *J. Mater. Chem.*, 14(14):2059–2065, 2004.

[100] M. Sumper and N. Kröger. Silica formation in diatoms: the function of long-chain polyamines and silafns. *J. Mater. Chem.*, 14:2059–2065, 2004.

[101] M. Sumper, S. Lorenz, and E. Brunner. Biomimetic control of size in the polyamine-directed formation of silica nanospheres. *Angew. Chem. Int. Ed.*, 42(42):5192–5195, 2003.

[102] K. Takegoshi, S. Nakamura, and T. Terao. 13C-1H dipolar-assisted rotational resonance in magic-angle spinning NMR. *Chem. Phys. Lett.*, 344:631–37, 2001.

[103] K. Takegoshi, S. Nakamura, and T. Terao. 13C-1H dipolar-driven 13C-13C recoupling without 13c RF irradiation in nuclear magnetic resonance of rotating solids. *J. Chem, Phys.*, 118:2325 – 42, 2003.

[104] M. M. Tomczak, D. D. Glawe, L. F. Drummy, C. G. Lawrence, M. O. Stone, C. C. Perry, D. J. Pochan, T. J. Deming, and R. R. Naik. Polypeptide-templated synthesis of hexagonal silica platelets. *J. Am. Chem. Soc.*, 127(36):12577–12582, 2005.

[105] D. Torchia. The measurement of proton-enhanced carbon-13 T_1 values by a method which supresses artifacts. *J. Magn. Reson.*, 30:613–16, 1978.

[106] E. L. Ulrich, H. Akutsu, J. F. Doreleijers, Y. Harano, Y. E. Ioannidis, J. Lin, M. Livny, S. Mading, D. Maziuk, Z. Miller, E. Nakatani, C. F. Schulte, D. E. Tolmie, R. Kent Wenger, H. Yao, and J. L. Markley. Biomagresbank. *Nucleic Acids Research*, 36(suppl 1):D402–D408, 2008.

[107] K. J. C. van Bommel, J. H. Jung, and S. Shinkai. Poly(l-lysine) aggregates as templates for the formation of hollow silica spheres. *Adv. Mater.*, 13(19):1472–1476, 2001.

[108] P. C. Weber and F. R. Salemme. Structural and functional diversity in 4-alpha-helical proteins. *Nature*, 287(5777):82–84, 1980.

[109] T. Weidner, N. F. Breen, K. Li, G. P. Drobny, and D. G. Castner. Sum frequency generation and solid-state nmr study of the structure, orientation, and dynamics of polystyrene-adsorbed peptides. *Proc. Natl. Acad. Sci. USA*, 107(30):13288–13293, 2010.

[110] D. Wishart and B. Sykes. The 13C chemical-shift index: A simple method for the identification of protein secondary structure using 13C chemical-shift data. *J. Biomol. NMR*, 4:171–80, 1994.

[111] D. Wishart, B. Sykes, and F. Richards. The chemical shift index: a fast and simple method for the assignment of protein secondary structure through nmr spectroscopy. *Biochemistry*, 31:1647–51, 1992.

[112] A. Woods and M. Huestis. A study of peptide-peptide interaction by matrix-assisted laser desorption/ionization. *J. Am. Chem. Soc. Mass Spectrom.*, 12(1):88–96, 2001.

[113] A. S. Woods and S. Ferr. Amazing stability of the arginine-phosphate electrostatic interaction. *J. Proteome Res.*, 4(4):1397–1402, 2005. PMID: 16083292.

[114] H. Xu, Y. Wang, X. Ge, S. Han, S. Wang, P. Zhou, H. Shan, X. Zhao, and J. Lu. Twisted nanotubes formed from ultrashort amphiphilic peptide I3K and their templating for the fabrication of silica nanotubes. *Chem. Mater.*, 22:5165–73, 2010.

[115] T. Yokoi, Y. Sakamoto, O. Terasaki, Y. Kubota, T. Okubo, and T. Tatsumi. Periodic arrangement of silica nanospheres assisted by amino acids. *J. Am. Chem. Soc.*, 128:13664–65, 2006.

[116] T. Yokoi, J. Wakabayashi, Y. Otsuka, W. Fan, M. Iwama, R. Watanabe, K. Aramaki, A. Shimojima, T. Tatsumi, and T. Okubo. Mechanism of formation of uniform-sized silica nanospheres catalyzed by basic amino acids. *Chem. Mater.*, 21:3719–29, 2009.

[117] R. L. York, W. K. Browne, P. L. Geissler, and G. A. Somorjai. Peptides adsorbed on hydrophobic surfaces - A sum frequency generation vibrational spectroscopy and modeling study. *Isr. J. Chem*, 47(1):51–58, 2007.

[118] V. M. Yuwono and J. D. Hartgerink. Peptide amphiphile nanofibers template and catalyze silica nanotube formation. *Langmuir*, 23(9):5033–5038, 2007.

Appendix A

PEPTIDE SYNTHESIS AND SILICA PRECIPITATION

A.1 Peptide Synthesis and Preparation

Native and isotopically labeled peptides were synthesized using a Rainin PS3 automated solid phase peptide synthesizer. For the labeled sequences, labeled amino acids were purchased from Cambridge Isotopes and then protected in house (with the exception of L-Lysine). When necessary, samples were purified by HPLC and then characterized with mass spectrometry. Peptides were studied neat and as part of silica composites; neat peptides were lyophilized, and silica composites were dried in a dessicator. The procedures for the syntheses, purifications, and characterizations of peptide samples are given here.

A.2 Fmoc-protection of Labeled Amino Acids

A.2.1 Fmoc-protection of Labeled Leucine/Isoleucine

In a 50 mL polyproprylene Falcon tube, 175 mg (1.267 mmol) of labeled L-Leucine/L-Isoleucine (purchased, Cambridge Isotope Laboratories, Andover, MA) is dissolved in 10 mL of saturated, aqueous $NaHCO_3$ (780 mg in 10 mL of DI, dissolved in a small beaker).

In a 50/100 mL round bottom flask, 447.76 mg (1.267 mmol) of Fmoc-Osu (Fmoc N-hydroxysuccinimide ester) is dissolved in 10 mL of acetonitrile.

The L-Leucine(or L-Isoleucine)/$NaHCO_3$ solution is poured into the acetonitrile solution and stirred for 12-24 hours. The round bottom flask is capped with a rubber septum or is covered with parafilm (with a few holes poked into the surface).

The acetonitrile is then evaporated on a rotovap,and the remaining oil is acidified

to pH ¡ 2 with 10% HCl; the white precipitate is extracted with ethyl acetate (3 x 10 mL ethyl acetate). The solvent is rotovapped off, and the final Fmoc-Leucine/L-Isoleucine precipitate is dried.[46]

A.2.2 Fmoc-Protection of Labeled Serine

In a 50 mL polyproprylene Falcon tube, 250 mg (2.38 mmol) of labeled L-serine (Cambridge Isotope Laboratories, Andover, MA) is dissolved in 12 mL of saturated, aqueous $NaHCO_3$.

In a 50/100 mL round bottom flask, 841 mg (2.38 mmol) of Fmoc-Osu is dissolved in 12 mL of acetonitrile.

The L-Serine/$NaHCO_3$ solution is poured into the acetonitrile solution and stirred for 12-24 hours. The round bottom flask is capped with a rubber septum or is covered with parafilm (with a few holes poked into the surface).

The acetonitrile is then evaporated on a rotovap,and the remaining oil is acidified to pH ¡ 2 with 10% HCl; the white precipitate is extracted with ethyl acetate (3 x 10 mL ethyl acetate). The solvent is rotovapped off, and the final Fmoc-Leucine precipitate is dried.[46]

A.3 Solid-Phase Peptide Synthesis

Labeled Leucines were Fmoc protected, as discussed previously; labeled α-N-Fmoc, ϵ-N-T-Boc L-Lysine was purchased from Cambridge Isotope Laboratories directly. Unlabeled amino acids were purchased from Sigma-Aldrich in protected form. All peptides were synthesized on a Rainin PS3 automated solid phase peptide synthesizer (Protein Technologies, Inc.). Both the R5 peptide and the LK variants started with a Fmoc-Leu-Wang resin. The solvent was N-methylpyrrolidone (NMP), the deprotector was 20% piperidine/NMP (by volume), and the activating agent was 0.4 M N-methyl morpholine (NMM) in N,N-dimethylformamide (DMF). The LK peptides were capped by reaction with acetic anhydride. The Rainin PS3 comes with a set of programs

which should be modified appropriately by the user; programs used to synthesize the R5 and LK peptides can be found in Table A.1. Single coupling programs were used for all labeled and unlabeled amino acids in the LKs, and double coupling was used for labeled residues in the R5 sequence.

The peptide was cleaved from the resin with a solution of (by volume) 9.5 TFA: 0.25 H_2O: 0.25 Triisopropylsilane (TIS).[1] Peptides were precipitated into cold tert-butyl-methyl ether, dried, suspended in DI, and lyophilized. The LK peptides did not require further purification. The R5 peptides were purified with an acetonitrile/water solvent system with 0.1% TFA on a Waters HPLC C18-reverse phase column. All peptides were characterized by mass spectrometry (Bruker Esquire LC-Ion Trap).

Peptides were made on the Rainin PS3 solid phase peptide synthesizer on the 100μmol scale; typically 400 μmol (4x excess) amounts were used for unlabeled amino acids, and 2x or 3x were used for labeled residues. Table A.2 gives frequently used amino acids and their respective masses.

A.4 Silica Precipitation

Orthosilicic acid was freshly prepared for each precipitation procedure. For most cases, 0.15 mL TMOS (tetramethylorthosilicate ester) was dissolved in 0.85 mL of 1 mM HCl to a final concentration of 1 M $Si(OH)_4$ (1 mL total volume).LK peptides were dissolved in either 1X PBS buffer solution (LK peptides) or 100 mM phosphate-citrate (R5 peptide) to a final concentration of 2-4 mM peptide. Typically, for NMR samples, 0.4 mL of 1 M $Si(OH)_4$ was added to a 5 mL peptide solution; the reaction was incubated for 5 minutes at ambient temperature before being centrifuged (3x with 2 mL washes of DI to remove excess silicic acid) and dried.[53]

Table A.1: Sample Rainin PS3 Programs Used for the LK and R5 Peptides

First and last amino acids in a sequence			
STEP	ACTION	TIME	REPS
1	SOLV	0:10:00	3
2	DEP	0:00:10	3
3	SOLV	0:00:30	6
4	ACT	0:00:30	1
5	AA	1:00:00	1
6	SOLV	0:00:30	3
Unlabeled AA (single coupling, LK)			
1	SOLV	0:00:30	3
2	DEP	0:05:00	2
3	SOLV	0:00:30	6
4	ACT	0:00:30	1
5	AA	1:00:00	1
6	SOLV	0:00:30	3
Unlabeled AA (double coupling, R5 peptide)			
1	SOLV	0:00:30	3
2	DEP	0:10:00	3
3	SOLV	0:00:30	6
4	ACT	0:00:30	1
5	AA	1:00:00	1
6	ACT	0:00:30	1
7	AA	1:00:00	1
8	SOLV	0:00:30	3
Labeled AA (single coupling, LK and R5 peptides)			
1	SOLV	0:00:30	3
2	DEP	0:10:00	2
3	SOLV	0:00:30	6
4	ACT	0:00:30	1
5	AA	4:00:00	1
6	SOLV	0:00:30	3

Table A.2: Amino Acid Masses for SPPS, 100μmolscale

Amino Acid	Mass, mg (unlabeled)	Mass, mg (labeled)
ARG	259.5	129.8
GLY	119.0	59.5
ILE	141.4	70.7
LEU	141.4	70.7
LYS	187.4	93.7
SER	153.4	65.5
TYR	183.8	—

Appendix B

EXAMPLES OF TALOS-N FILES

Example TALOS-N[94] input (.tab) and output (pred.tab) files for the R5 and LKα14 sequences are listed here.

B.1 LKα14 Input file (.tab)

```
#
# NMRPipe-format Chemical Shift table 2010.327
# This Chemical Shift table format is used by TALOS+, SPARTA+, MFR, etc.
#
# TALOS+ reads protein backbone chemical shifts from a text
# table and uses these shifts to predict backbone torsion
# angles phi and psi.
#
# The required columns for TALOS+ chemical shift input are
# RESID RESNAME ATOMNAME SHIFT, specified in any order. Other
# columns are ignored.
#
# The shifts used by TALOS+ are HN N CA CB C HA HA2 HA3,
# calibrated according to IUPAC recommendations.
# Other shifts are ignored.
#
# The GLY HA2 and HA3 shifts are treated identically by TALOS+, and so
# they do not have to be stereospecifically assigned.
```

```
#
# The complete amino acid sequence must be specified with
# DATA FIRST_RESID (the first residue ID for the sequence)
# and DATA SEQUENCE lines.
#
# The residue specification in the RESNAME column can use either
# one-letter or three-letter amino acid names.  The residue type
# given in the RESNAME column should match the corresponding
# residue type in the SEQUENCE specification.
#
# The residue specification is case sensitive:
#  Residue Name "CYS" or "C"  for reduced form  (CB ~28 ppm)
#  Residue Name "cys" or "c"  for oxydized form (CB ~42 ppm)
#  Residue Name "his" or "h" for protonated form (pH < ~6).
#  Residue Name "HIS" or "H" for unprotonated form.

# About the Table Format:
#  The table is in the NMRPipe Generic Data Base (GDB) format,
#  consisting of columns separated by one or more spaces.
#
#  The VARS line labels the column names (variables). The names
#  are case-sensitive.
#
#  The FORMAT line specifies the column data type and specifies
#  the output format when the table values are written. The
#  format specifications are adapted from those used by the
#  UNIX/C "printf" function: %d for integers, %f or %e for
#  floating-point values, and %s for text.
```

```
#
#  Blank lines are allowed, and comment lines like this one
#  can start with either the character # or the keyword REMARK.
#
#  Since columns are space-separated, no individual item in the
#  table may contain spaces, and no all-blank values are allowed.

REMARK Chemical Shift Table for LKa14

DATA FIRST_RESID 1

DATA SEQUENCE LKKLLKLLKK LLKL

VARS   RESID RESNAME ATOMNAME SHIFT
FORMAT %4d   %1s     %4s      %8.3f

   1 L   C  177.603
   1 L   CB  41.622
   1 L   CA  57.272
   2 K    C  177.134
   2 K   CB   32.258
   2 K   CA   58.987
   4 L   C  177.272
   4 L   CB  42.027
   4 L   CA  57.979
   4 L    N  119.504
   5 L   C  178.220
   5 L   CB  41.574
```

```
 5 L   CA  58.358
 6 K    C  178.640
 6 K   CB  32.483
 6 K   CA  60.316
 7 L   C  177.624
 7 L   CB   41.272
 7 L   CA   57.360
 7 L    N  119.256
 8 L   C  178.233
 8 L   CB  41.738
 8 L   CA  58.097
 9 K    C  178.082
 9 K   CB  32.158
 9 K   CA   60.455
10 K    C  177.539
10 K   CB  32.348
10 K   CA  59.599
11 L   C  178.224
11 L   CB  41.305
11 L   CA  57.305
11 L    N  119.512
12 L   C  177.927
12 L   CB  42.920
12 L   CA  57.509
13 K    C  176.775
13 K   CB  33.119
13 K   CA  55.781
```

B.2 LKα14 neat TALOS-N Output files

B.2.1 LKα14 neat TALOS-N Output: "pred.tab"

"pred.tab" file: Gives torsion angle predictions and ratings (strength of prediction).

```
REMARK TALOS-N Protein Backbone Torsion Angle Prediction Table
REMARK  Prediction Summary for Chemical Shift Input LKa14neatexp.tab
REMARK
REMARK  PHI is the predicted torsion angle C(i-1) N(i)  CA(i) C(i)   (degrees).
REMARK  PSI is the predicted torsion angle N(i)   CA(i) C(i)  N(i+1) (degrees).
REMARK
REMARK  DPHI and DPSI are the estimated standard deviations of the
REMARK  prediction errors in PHI and PSI (degrees).
REMARK
REMARK  DIST is the TALOS-N database matching score.
REMARK
REMARK  S2 is the Wishart RCI chemical shift order parameter,
REMARK  JACS, 127(43), 14970-14971.
REMARK
REMARK  COUNT is the number of database triplets used to form
REMARK  the torsion angle predictions.
REMARK
REMARK  CLASS is the classification of the prediction result:
REMARK    None: no torsion prediction was made.
REMARK
REMARK    Strong/Generous: majority consensus in database matches;
REMARK    prediction is likely to be good.
REMARK
REMARK    Warn: no consensus in database matches, do not use prediction.
```

```
REMARK
REMARK    Dyn:  RCI-S2 value indicates that residue has dynamic conformation.
REMARK
REMARK Reference:
REMARK  Y. Shen, and A. Bax:
REMARK  Protein backbone and sidechain torsion angles predicted from
REMARK  NMR chemical shifts using artificial neural networks
REMARK  J. Biomol. NMR (in press).
REMARK
REMARK  TALOS-N Version 4.01 Rev 2013.148.15.55 TALOSN_INFO

DATA FIRST_RESID 1
DATA SEQUENCE LKKLLKLLKK LLKL

VARS   RESID RESNAME PHI PSI DPHI DPSI DIST S2 COUNT CS_COUNT CLASS
FORMAT %4d %s %8.3f %8.3f %8.3f %8.3f %8.3f %5.3f %2d %2d %s

   1 L 9999.000 9999.000    0.000    0.000    0.000 0.000  0  6 None
   2 K  -64.743  -34.099    5.792    6.929    0.380 0.669 25  6 Strong
   3 K  -67.271  -37.700    5.577    4.165    0.286 0.737 25  7 Strong
   4 L  -63.023  -41.899    4.924    5.609    0.189 0.831 25  7 Strong
   5 L  -64.953  -39.543    5.263    7.969    0.137 0.843 25 10 Strong
   6 K  -65.780  -37.556    5.420    5.888    0.124 0.866 25 10 Strong
   7 L  -66.098  -39.123    5.188    5.049    0.100 0.863 25 10 Strong
   8 L  -65.187  -38.615    6.226    5.914    0.098 0.865 25 10 Strong
   9 K  -65.132  -37.964    6.318    5.250    0.108 0.868 25  9 Strong
  10 K  -66.597  -34.969    6.600    5.602    0.146 0.853 25 10 Strong
  11 L  -65.471  -32.995    6.646    6.607    0.172 0.807 25 10 Strong
```

```
  12 L  -66.001  -25.239    7.785    9.031    0.282 0.746 25 10 Strong
  13 K -100.130    2.214   13.813   14.470    0.539 0.714 25  6 Strong
  14 L 9999.000 9999.000    0.000    0.000    0.000 0.000  0  3 None
```

B.2.2 LKα14 neat TALOS-N Output: "predChi1.tab"

"predChi1.tab" file: χ_1 conformation predictions

```
REMARK TALOS-N Protein Chi1 Conformation Prediction Table
REMARK  Prediction Summary for Chemical Shift Input LKa14neatexp.tab
REMARK
REMARK  Q_Gm is the probability to be with a g- (gauche-) rotamer.
REMARK  Q_Gp is the probability to be with a g+ (gauche+)rotamer.
REMARK  Q_t  is the probability to be with a t (trans) rotamer.
REMARK
REMARK  CS_COUNT is the number of chemical shifts in the query triplets
REMARK  used to predict chi1 conformation.
REMARK
REMARK  CLASS is the classification of the prediction result:
REMARK    na: no chi1 prediction was made.
REMARK    g-/g+/t: 3-state chi1 prediction
REMARK
REMARK Reference:
REMARK  Y. Shen, and A. Bax:
REMARK  Protein backbone and sidechain torsion angles predicted from
REMARK  NMR chemical shifts using artificial neural networks
REMARK  J. Biomol. NMR (in press).
REMARK
```

```
REMARK  TALOS-N Version 4.01 Rev 2013.148.15.55 TALOSN_INFO


VARS   RESID RESNAME CS_COUNT Q_Gm Q_Gp Q_T CLASS
FORMAT %4d %s %2d %5.3f %5.3f %5.3f %s

   1 L  6 0.333 0.333 0.333 na
   2 K  6 0.538 0.089 0.373 na
   3 K  7 0.493 0.082 0.425 na
   4 L  7 0.562 0.040 0.399 na
   5 L 10 0.480 0.039 0.480 na
   6 K 10 0.436 0.128 0.436 na
   7 L 10 0.633 0.041 0.326 g-
   8 L 10 0.480 0.039 0.480 na
   9 K  9 0.456 0.088 0.456 na
  10 K 10 0.456 0.088 0.456 na
  11 L 10 0.692 0.045 0.263 g-
  12 L 10 0.633 0.041 0.326 g-
  13 K  6 0.584 0.080 0.336 na
  14 L  3 0.333 0.333 0.333 na
```