

Protein Chemical Shift Tensor Calculation with Bond Polarization Theory: A New Approach for the Study of Orientation and Dynamics in Biological Systems

Zur Erlangung des akademischen Grades eines
DOKTORS DER NATURWISSENSCHAFTEN

(Dr. rer. nat.)

Fakultät für Chemie und Biowissenschaften

Karlsruher Institut für Technologie (KIT) – Universitätsbereich

genehmigte

DISSERTATION

von

Dipl.-Biochem. Igor Jakovkin

aus St.-Petersburg

Dekan: Prof. Dr. Stefan Bräse

Referent: Prof. Dr. Burkhard Luy

Korreferent: Prof. Dr. Anne S. Ulrich

Tag der mündlichen Prüfung: 14. 07. 2011

Table of Contents

Table of contents	I
Overview and scope of this study	1
1. Introduction	4
1.1 Nuclear magnetic resonance spectroscopy	4
1.2 Chemical shift tensor	5
1.3 Spin-lattice and spin-spin relaxation times.....	10
1.4 Dipolar coupling.....	12
1.5 Residual dipolar coupling.....	13
2. Protein chemical shift calculation	15
2.1 <i>Ab initio</i> route to chemical shift tensors and its limits.....	15
2.2 Bond polarization theory	23
2.3 ¹⁵ N and ¹³ C parameterization	28
2.4 Crystalline tripeptides: the test case for ¹⁵ N parameterization	30
2.5 Unbiased ¹³ C parameterization for proteins.....	38
2.6 Application to ubiquitin.....	41
2.7 Chemical shift driven protein structure refinement.....	44
2.8 Computational efficiency: BPT vs. DFT	49
2.9 Chemical shift and protein structure validation	51
3. Dynamics and orientation of membrane peptides	56
3.1 Biological background.....	56
3.2 PISEMA spectroscopy	56
3.3 Macroscopically aligned samples	57
3.4 Application to gramicidin A – the role of dynamics	59
3.5 Molecular dynamics with orientational constraints.....	69
3.6 Molecular dynamics simulation of gramicidin A	70

3.7 Local order tensors	75
3.8 Local order parameters in gramicidin A	78
3.9 PISEMA simulation for nontrivial cases	85
4. Dynamics of solid-state proteins	90
4.1 CSA order parameters	90
4.2 Application to thioredoxin	92
4.3 Application to immunoglobulin-binding protein GB1	102
4.4 Protein collective motions in solid state	109
5. Methods	112
5.1 Molecular Modeling	112
5.2 BPT chemical shift calculation	113
5.3 DFT chemical shift calculation	114
5.4 Chemical shift driven geometry optimization	115
5.5 Molecular dynamics simulation	115
6. Summary	118
6.1 Protein chemical shift calculation	118
6.2 Dynamics and orientation of membrane peptides	120
6.3 Dynamics of solid-state proteins	121
List of abbreviations	123
References	125
Appendix	A
Deutsche Zusammenfassung	A
Ehrenwörtliche Erklärung	B
Lebenslauf	C
Danksagung	D

Overview and scope of this study

Up to 30% of known proteins are embedded into biological membranes¹. Membrane proteins play an important role in cell membrane adhesion, signal transduction and membrane transport. The elucidation of membrane protein structures remains challenging due to difficulties with their solubilization and crystallization. Solid-state nuclear magnetic resonance spectroscopy (NMR) is a powerful tool for analysis of protein structure in membrane proteins and peptides². Unfortunately, in many cases the connection between protein structure and function is unknown. The knowledge of protein membrane orientation and dynamics can provide the missing link between structure and function.

NMR chemical shift (CS) tensors yield a variety of information on protein structure, dynamics and membrane orientation. Empirical methods for protein chemical shift calculation facilitate the prediction of torsion angles in protein backbone³. A number of recent reports demonstrate their potential for chemical shift driven structure elucidation^{4,5,6}. However, empirical methods can provide only isotropic chemical shift values and not the full chemical shift tensors. *Ab initio* methods can compute the full chemical shift tensors but they are computationally much too demanding for biopolymers⁷. Semi-empirical methods offer a solution to this problem combining the capacity of the full tensor calculation with low computational cost. The scope of this study is therefore to adapt the semi-empirical bond polarization theory⁸ (BPT) to protein chemical shift tensor calculation and to apply such calculations to interpretation of the solid-state NMR data.

The introductory chapter of this thesis provides a brief outline of the few NMR basic concepts. The second chapter is dedicated to BPT protein chemical shift calculation and protein chemical shift driven structure refinement. It explains the basics of the bond polarization theory and deals with the evaluation and fine-tuning of the BPT parameterization for ¹⁵N and ¹³C chemical shift calculation. Test calculations and comparison of the computational efforts with density functional theory (DFT) calculations are presented for several biological systems including peptide crystals and small globular protein ubiquitin. In addition BPT enables chemical shift gradient

calculation and consequently chemical shift driven geometry optimization. The development of the protocol for chemical shift driven protein structure refinement with BPT chemical shift gradients is also described in the second chapter.

After having presented isotropic chemical shift calculations in the second chapter, the thesis proceeds to applications which require chemical shift tensor calculations. PISEMA spectroscopy correlating ^{15}N chemical shift with ^1H - ^{15}N dipolar is a widespread method for the study of peptide and protein orientation in biological membranes⁹. In routine PISEMA applications the assignment of signals relies on a priori assumptions about the chemical shift tensors and their orientation (i.e. all tensors can be regarded as identical) and for data interpretation the molecule is treated as a rigid body. This strategy is well-suited for rigid α -helical peptides but the application to non-helical structures, molecules with DL-amino acids substitutions and molecules with different dynamic behavior on different amino acid sites is not possible in many cases. The third chapter deals with PISEMA spectra prediction using explicitly calculated ^{15}N chemical shift tensors and molecular dynamics simulations. Simultaneous determination of membrane orientation and dynamics is presented for gramicidin A – a membrane-active antibiotic peptide containing D-amino acids. Local order parameters for gramicidin A are derived directly from the calculated ^{15}N chemical shift tensors. These local order parameters are used in order to predict PISEMA spectra of gramicidin A which do not follow the standard signal pattern.

Orientation of proteins in macroscopically aligned lipid bilayers is not always possible. The fourth chapter of this thesis is therefore dedicated to study of dynamics of non-oriented solid-state samples. Novel order parameters based on comparison of the calculated and measured ^{15}N chemical shift tensors are introduced. They are evaluated on the small redox protein thioredoxin since dynamics of thioredoxin in the solid state has been thoroughly studied. This approach is applied for investigation of collective motions in microcrystalline immunoglobulin-binding protein GB1.

The fifth chapter is the methodology chapter. It contains the detailed settings for the calculations presented in other chapters. The sixth chapter summarizes the results of this study.

On the whole, this thesis demonstrates that semi-empirical bond polarization theory can be successfully applied to protein chemical shift tensor calculation at a very low computational cost. This results in an advancement of solid-state NMR data interpretation. On the fly chemical shift tensor computations during molecular dynamics simulation simplify the interpretation of PISEMA spectra of non-helical and highly dynamic biological systems in macroscopically aligned lipid bilayers. Furthermore they allow the derivation of local order parameters based on chemical shift tensors for single amino acids sites. For non-oriented biological systems like microcrystalline proteins novel order parameters can be computed using bond polarization theory calculations and chemical shift tensor measurements. They allow an insight into protein dynamics on microsecond to millisecond time scales. The new methodology presented in this thesis should enhance the ability to study biological systems using solid-state NMR.

1. Introduction

1.1 Nuclear magnetic resonance spectroscopy

Atomic nuclei have a quantum-mechanical property named spin¹⁰. The spin of a nucleus can be defined as an operator generating rotations of the nuclear wavefunction. In semi-classical treatment spin is regarded as a quantized angular momentum. It is a vectorial quantity and its z-component takes the form:

$$I_z = \hbar m \quad (1.1)$$

with m being the magnetic quantum number with the constraint $m = -I, -I+1, \dots, +I$. In this approach a nucleus is regarded as a spinning charged particle, thus producing a circular current and behaving like an electromagnet. The magnetic dipole moment $\vec{\mu}$ of a nucleus is collinear with its spin angular momentum and the proportionality constant between these values is the gyromagnetic ratio γ :

$$\vec{\mu} = \gamma \vec{I}. \quad (1.2)$$

In an external magnetic field the behavior of the nuclear spin is similar to the behavior of the spinning magnetic moment. The spin is precessing around the direction of the external magnetic field (by definition this direction is set along the z axis of the laboratory coordinate system). The angular velocity ω_0 of the precession depends on the magnetic flux density \vec{B} (B_0 for the static magnetic field):

$$\omega_0 = \gamma B_0 \quad (1.3)$$

and the characteristic frequency of the precession ν_0 is called the Larmor frequency:

$$\nu_0 = \omega_0 / 2\pi. \quad (1.4)$$

The potential energy of the magnetic moment in an external magnetic field is

$$E = -\vec{\mu} \times \vec{B}. \quad (1.5)$$

The energy of the nuclear spin in an external magnetic field is quantized and the equation (1.5) can be expressed considering the spin orientation:

$$E = -\gamma I_z B_0 = -\gamma \hbar m B_0. \quad (1.6)$$

This corresponds to the splitting into $2I+1$ energy levels called the Zeeman effect. The population of the energy levels follows the Boltzmann distribution. Nuclei with spin $I=1/2$ like ^1H , ^{13}C or ^{15}N possess two Zeeman energy levels in an external magnetic field. Their spins assume either parallel (α -state) or anti-parallel (β -state) orientation to the magnetic field direction. For a macroscopic sample the Boltzmann distribution predicts higher population for the lower energy α -state resulting in net magnetization in the direction of the external magnetic field if the gyromagnetic ratio is positive or vice versa if it is negative.

The irradiation of an oscillating magnetic field \vec{B}_1 (the so-called radio-frequency pulse) at the Larmor frequency can influence the orientation of the net magnetization. The flip angle α of the net magnetization between the z axis (i.e. the direction of the static external field) and the xy (transverse) plane of the laboratory coordinate system depends on the pulse duration t_p :

$$\alpha = \gamma B_1 t_p. \quad (1.7)$$

In the simplest NMR experiment, a sample is positioned inside a coil of conducting wire in the static external magnetic field. Zeeman splitting of the nuclear spin state energy levels in the magnetic field is associated with the formation of the macroscopic magnetization parallel to the magnetic field direction. A radio-frequency pulse at the Larmor frequency is applied in order to flip the net magnetization to the transverse plane of the laboratory coordinate system. In the transverse plane the vector of the macroscopic magnetization is precessing with the Larmor frequency around the direction of the external magnetic field and acts as an alternating current generator. Precessing magnetic dipole then induces an oscillating current in the coil. The frequency of the induced current is again the Larmor frequency. This current is the signal detected in NMR spectroscopy.

1.2 Chemical shift tensor

The electrons around a nucleus induce secondary magnetic fields in the sample. Secondary magnetic fields can shield the external magnetic field, thus altering the magnetic field experienced by the nucleus. This alteration is reflected in a change of

the nuclear spin energy levels and correspondingly observed Larmor frequencies (cf. equations 1.3 and 1.6):

$$\omega_{observed} = (1 - \sigma)\omega_0. \quad (1.8)$$

Chemical shielding (also called nuclear shielding) σ depends on the chemical surrounding of the nucleus and molecular geometry. This fact explains the utility of NMR spectroscopy for chemistry. Chemical shielding values are usually given as a factor in parts per million (ppm). Commonly, chemical shifts are reported instead of chemical shieldings. The chemical shift δ of a nucleus is the shielding difference between a nucleus in the substance under study and the same nuclear species in a reference compound:

$$\delta_{observed} = \frac{\sigma_{reference} - \sigma_{observed}}{1 - \sigma_{reference}} \approx \sigma_{reference} - \sigma_{observed}. \quad (1.9)$$

Addition of the reference compound to the sample simplifies the comparison of the NMR spectra. By definition, the chemical shift scale is oriented in the opposite direction to the chemical shielding scale. It is customary to plot the chemical shift scale from right to left. The zero point of the chemical shift scale is set to the chemical shielding of the reference substance – liquid NH₃ at 25 °C for ¹⁵N and TMS solvated in MeOD at 25 °C for ¹³C and ¹H. In biomolecular NMR TSP and DSS (both at 25 °C in aqueous solution) are also used.

The induced field $\vec{B}_{observed}$ is not necessary parallel to the external magnetic field \vec{B} . The shielding generated by the electrons around the nucleus depends on the molecular orientation in the external magnetic field. Therefore mathematical description of the chemical shift (or of the nuclear shielding) in laboratory coordinate system requires a tensor of second rank (3×3 matrix) δ . The observed chemical shift value $\delta_{observed}$ is the chemical shift tensor component aligned with the external magnetic field (by definition this is the zz component δ_{zz}). For crystals the chemical shift tensor is symmetric, so that $\delta_{xy} = \delta_{yx}$. This is a physical property of real crystals and is not necessary valid for all tensors. Nevertheless this assumption is often applied to NMR data analysis for liquids and amorphous solids. The chemical shift tensor can be decomposed into a symmetric and antisymmetric part:

$$\delta = \delta^{symmetric} + \delta^{antisymmetric}, \quad (1.10)$$

where

$$\delta_{\alpha\beta}^{symmetric} = \frac{1}{2}(\delta_{\alpha\beta} + \delta_{\beta\alpha}) \text{ for } \alpha \neq \beta \text{ and } \delta_{\alpha\beta}^{symmetric} = \delta_{\alpha\beta} \text{ for } \alpha = \beta \quad (1.11)$$

and

$$\delta_{\alpha\beta}^{antisymmetric} = \frac{1}{2}(\delta_{\alpha\beta} - \delta_{\beta\alpha}) \text{ for } \alpha \neq \beta \text{ and } \delta_{\alpha\beta}^{antisymmetric} = 0 \text{ for } \alpha = \beta. \quad (1.12)$$

Only the symmetric part of the tensor enters in the first-order term of the average Hamiltonian (the operator corresponding to the total energy of the system) of the spin system. The antisymmetric part provides non-zero contributions to the higher-order terms in the average Hamiltonian of the spin system. Therefore if the Zeeman interaction is significantly larger than any other interaction in the spin system (including chemical shielding) the antisymmetric part of the chemical shift tensor can be neglected:

$$\delta \approx \delta^{symmetric}. \quad (1.13)$$

The advantage of this approach is that any symmetric tensor of rank two can be transformed from a laboratory coordinate system into the principal axis coordinate system. The principal axis system is defined as the coordinate system in which only the diagonal components of the chemical shift tensor are non-zero and the characterization of the tensor is simpler. Two conventions can be applied in order to describe the chemical shift tensor in principal axis systems¹¹. According to Herzfeld-Berger convention¹¹ the principal (diagonal) components of the tensor are sorted by their magnitude:

$$\delta_{11} \geq \delta_{22} \geq \delta_{33} \quad (1.14)$$

and the tensor is characterized by the isotropic chemical shift δ_{iso} , span Ω and its skew κ :

$$\delta_{iso} = (\delta_{11} + \delta_{22} + \delta_{33})/3, \quad (1.15)$$

$$\Omega = \delta_{11} - \delta_{33}, \quad (1.16)$$

$$\kappa = 3(\delta_{22} - \delta_{iso})/\Omega. \quad (1.17)$$

In Haeberlen-Mehring convention¹¹ the principal components are ordered by their separation from isotropic value:

$$|\delta_{33} - \delta_{iso}| \geq |\delta_{11} - \delta_{iso}| \geq |\delta_{22} - \delta_{iso}| \quad (1.18)$$

and the chemical shift tensor is characterized by its isotropic chemical shift δ_{iso} (cf. equation 1.15), reduced anisotropy δ (not to be confused with chemical shift δ), anisotropy $\Delta\sigma$ and its asymmetry η :

$$\delta = \delta_{33} - \delta_{iso}, \quad (1.19)$$

$$\Delta\sigma = \delta_{33} - (\delta_{11} + \delta_{22})/2 = 3\delta/2, \quad (1.20)$$

$$\eta = (\delta_{22} - \delta_{11})/\delta. \quad (1.21)$$

Isotropic chemical shift and chemical shift anisotropy and asymmetry are equivalent to the principal components of the chemical shift tensor. They provide a second description for the observed chemical shift value $\delta_{observed}$ of a nucleus in a molecule oriented in external magnetic field:

$$\delta_{observed}(\theta, \varphi) = \delta_{iso} + \frac{1}{2}\delta(3\cos^2\theta - 1 + \eta\sin^2\theta \cdot \cos 2\varphi), \quad (1.22)$$

where θ and φ are the polar angles defining the orientation of the external magnetic field in the principal axis coordinate system of the chemical shift tensor. Therefore it is possible to derive the molecular orientation from observed chemical shift values of a nucleus if the chemical shift tensor is known a priori and vice versa.

In polycrystalline samples and lipid dispersions molecules assume all possible spatial orientations with respect to the external magnetic field. All frequencies described by equation 1.22 can be observed simultaneously resulting in a broad NMR signal called “powder pattern” (cf. figure 1.1). If the environment of the nucleus under study has at least one threefold symmetry axis or molecular motions around such a symmetry axis are present (e.g. methyl group rotation), axially symmetric tensor can be observed. In this case all non diagonal elements of the tensor are zero and in Herzfeld-Berger convention δ_{22} equals either δ_{11} or δ_{33} and the skew is $\kappa = \pm 1$. In Haeberlen-Mehring convention the asymmetry is $\eta = 0$ and principal tensor components are denoted as $\delta_{11} = \delta_{22} = \delta_{\perp}$ and $\delta_{33} = \delta_{\parallel}$. Powder pattern of the nucleus with axially symmetric tensor has a much larger intensity at the frequency corresponding to δ_{\perp} (for this orientation

the z axis of the principal axis coordinate system is orthogonal to the direction of the external magnetic field and the number of such orientations is infinite) than at the frequency corresponding to δ_{\parallel} (for this orientation the z axis of the principal axis coordinate system is parallel to the direction of the external magnetic field and there is only one such orientation). For tensors without axial symmetry it is also possible to determine principal tensor components directly from the powder pattern, but it is not possible to define the corresponding axis of the principal axis coordinate system.

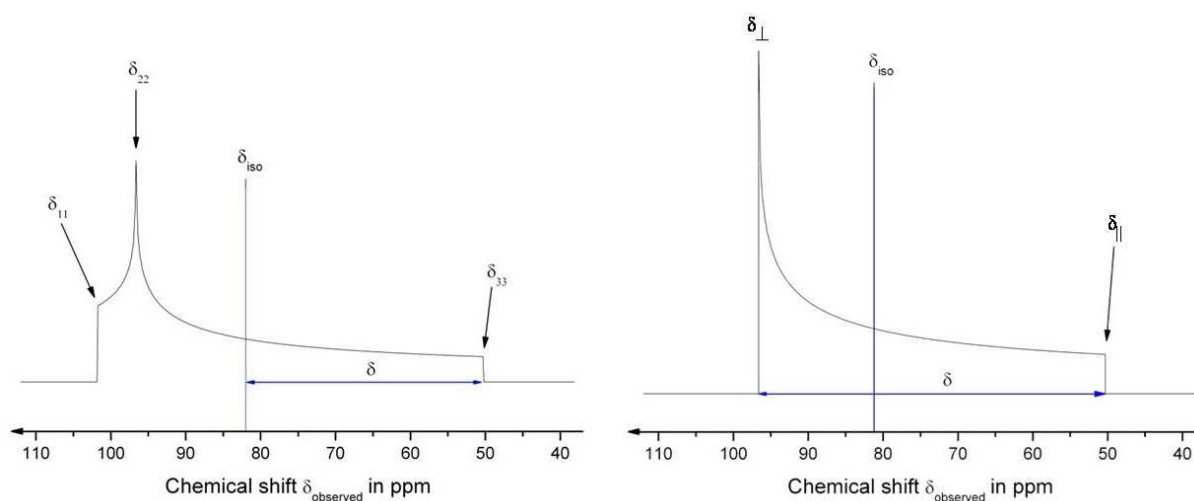


Figure 1.1: Example of an NMR powder spectrum (left – powder spectrum of a nucleus with an axially asymmetric chemical shift tensor, assignment of principal tensor components is valid for both conventions; right – powder spectrum of a nucleus with an axially symmetric chemical shift tensor).

Molecular mobility leads to the averaging of the chemical shift tensor and therefore to narrowing of observed NMR signals. In solid samples the correlation time of molecular motion is comparable to the duration of the NMR experiment. Thus anisotropic interactions remain preserved but the lineshape and linewidth of the spectra are influenced by molecular motion. Their analysis can provide information on molecular dynamics. For an axially symmetric chemical shift tensor of a nucleus rotating around an axis the influence of molecular motion can be expressed by the rotational order parameter S_{α} ¹²

$$S_{\alpha} = \left\langle \frac{3\cos^2 \alpha - 1}{2} \right\rangle, \quad (1.23)$$

where α denotes the angle between the z axis of the principal axis coordinate system and the axis of averaging and brackets refer to the time-average value. The range of S_α is $0 \leq S_\alpha \leq 1$. $S_\alpha = 0$ corresponds to the maximal possible averaging (averaging to the isotropic value) while $S_\alpha = 1$ corresponds to the absence of any motional averaging. In this case equation 1.22 can be written as

$$\delta_{observed}(\alpha, \beta) = \delta_{iso} + S_\alpha \cdot \frac{1}{2} \delta(3 \cos^2 \beta - 1), \quad (1.24)$$

where β is the angle between the axis of averaging and the direction of the external magnetic field. This order parameter formalism can also be applied to other anisotropic NMR parameters like the dipolar coupling tensor (cf. section 1.4).

In isotropic fluids molecules are in a rapid tumbling motion and the chemical shift tensor of a nucleus is averaged over all orientations of the molecule during this motion (the rotational correlation time of the tumbling motion is much shorter than the duration of NMR experiment: $\tau_c \ll \tau_{NMR}$). All orientations are equally probable and therefore only the isotropic chemical shift (the average value of the principal chemical shift tensor components) can be observed.

1.3 Spin-lattice and spin-spin relaxation times

After flipping the net magnetization into the transverse plane of the laboratory coordinate system by a radio-frequency pulse (cf. section 1.1) the original thermal equilibrium state must be restored (the precession of the net magnetization around the direction of the external magnetic field cannot go on forever). That implies that the magnetization in the transverse plane must disappear while net magnetization along the external magnetic field direction (longitudinal magnetization) must arise again. Two relaxation phenomena are involved in this process. Spin-lattice relaxation is responsible for the buildup of the longitudinal magnetization. It restores the Boltzmann distribution by exchanging energy with the molecular environment. Spin-spin relaxation is responsible for the decay of the transverse magnetization. After buildup of the net magnetization in the transverse plane the magnetic dipole vectors $\vec{\mu}$ of the individual spins are precessing with only slightly different phases. Mutual

exchange of spin energies leads to the loss of this phase coherence. Spins start precessing with arbitrary phases and cancel each other out, causing transverse magnetization to vanish. Relaxation times for spin-lattice and spin-spin relaxation are denoted T_1 and T_2 , respectively. In the simplest description of relaxation (the Bloch equations), relaxation of isolated spins is characterized by the phenomenological exponential decay rate constants $R_1 = T_1^{-1}$ and $R_2 = T_2^{-1}$.

A variety of physical mechanisms contributes to relaxation phenomena. For spin $I = 1/2$ nuclei these mechanisms include (but are not limited to) the relaxation due to dipolar interactions between the magnetic moments of the spins and the relaxation due to chemical shift anisotropy. Both mechanisms result in internal forces acting on individual spins. These forces can alter the spin orientation. They are a prerequisite for relaxation. The dipolar interaction between the two nuclei experiencing the magnetic dipolar field of each other depends on the distance between the nuclei and their mutual orientation. As already mentioned in section 1.2 electrons induce local secondary magnetic fields around the nuclei. Since chemical shielding tensors are anisotropic, the interaction of nuclear spin with local magnetic fields depends on molecular orientation. Distance and mutual orientation between the nuclei as well as molecular orientation and interaction with local magnetic fields depend on molecular motion. Thus molecular motion affects relaxation mechanisms and therefore relaxation times are sensitive probes of molecular motion. Higher relaxation rates of a nucleus usually correlate with higher molecular mobility. Furthermore relaxation time measurements allow to derive order parameters for molecular mobility without any a priori assumptions on the direction and distribution of the motions¹³. These parameters are called Lipari-Szabo model-free generalized order parameters and are denoted as S^2 . Lipari-Szabo order parameters are a measure of the spatial restriction of the motion. They are numbers between 0 and 1, where 0 describes a nucleus in a completely flexible and 1 in a completely rigid bond. Lipari-Szabo order parameters allow quantitative comparisons of molecular mobility.

1.4 Dipolar coupling

Dipolar coupling is the direct interaction between magnetic dipoles. In classical treatment each spin acts like a magnet creating its own magnetic field experienced by other spins. This interaction leads to splitting of the Zeeman energy levels and therefore to line splitting and broadening in solid-state NMR spectra. Dipolar coupling between two spins varies with their orientation in the external magnetic field and therefore the description of dipolar coupling requires a tensor of second rank \mathbf{D} . Dipolar coupling is an axially symmetric interaction. Thus \mathbf{D} is an axially symmetric traceless tensor with principal values $\mathbf{D}_{\alpha\alpha} = \{-d/2; -d/2; d\}$, where d is the dipole-coupling constant given by

$$d = \hbar \left(\frac{\mu_0}{4\pi} \right) r_{IS}^{-3} \gamma_I \gamma_S \quad (1.25)$$

with the reduced Planck's constant \hbar , magnetic field constant $\mu_0(4\pi)^{-1}$, the internuclear distance r_{IS} and the gyromagnetic ratios γ_I and γ_S of the coupling nuclei. The first-order Hamiltonian for heteronuclear dipolar coupling is:

$$\hat{H} = -d\hbar(3\cos^2\theta - 1)\hat{I}_z\hat{S}_z, \quad (1.26)$$

where θ is the angle defining the orientation of the internuclear connecting vector \vec{r}_{IS} to the direction of the external magnetic field and \hat{I}_z and \hat{S}_z are the spin operators of the coupling nuclei. The observed heteronuclear dipolar coupling $D_{observed}$ is the dipolar coupling tensor component aligned with the external magnetic field (by definition this is the zz component D_{zz}):

$$D_{zz} = -d(3\cos^2\theta - 1). \quad (1.27)$$

This is why it is possible to derive the molecular orientation from observed dipolar coupling values if the dipolar coupling tensor is known a priori. Maximal dipolar coupling with the absolute value of $2d$ is observed when the internuclear vector of the coupling nuclei is aligned with the external magnetic field ($\theta = 0^\circ$). Dipolar couplings are usually given in Hz.

Molecular mobility in solid samples leads to the averaging of the dipolar coupling tensor. In the case of rotation around an axis, the observed dipolar splitting is scaled

by the rotational order parameter S_α described in section 1.2. In the case of wobbling around an equilibrium position the scaling of observed dipolar splitting is expressed by a phenomenological order parameter $S_{dipolar}$ simply derived from the observed dipolar splitting and the absolute value of maximal dipolar splitting:

$$S_{dipolar} = \frac{D_{observed}}{D_{max}} = \frac{D_{observed}}{2d}. \quad (1.28)$$

The range of $S_{dipolar}$ is identical to the range of S_α (cf. section 1.2). Furthermore the $S_{dipolar}$ order parameter can be applied to the approximate description of the chemical shift anisotropy averaging by wobbling motion. In this case $S_{dipolar}$ must be inserted into equation 1.24 instead of S_α . It should be emphasized that any kind of order parameter for dipolar couplings as well as for the chemical shift anisotropy may be given either locally for a single nucleus or internuclear vector or globally for all anisotropic interactions in the whole molecule (rigid body approach). If several motional modes are present their influence is given by the product of order parameters for each single mode.

In isotropic fluids, molecules are in a rapid tumbling motion, all possible values of θ exist and dipolar coupling averages to zero. However, at any particular time point the dipolar interaction can be different from zero, thus contributing to relaxation (cf. section 1.3).

1.5 Residual dipolar coupling

In an isotropically tumbling molecule dipolar couplings are averaged to zero. Anisotropic solvents like liquid crystalline solvents (e.g. bicelles or filamentous phage Pf1) and stretched polyacrylamide gels can partially align dissolved molecules through steric and anisotropic interactions with the solvent¹⁴. In this case, motionally averaged dipolar coupling (residual dipolar coupling) can be observed in liquid-state NMR. The magnitude of the residual dipolar coupling (RDC) depends on the degree of alignment induced by the anisotropic solvent. Furthermore, the observed residual dipolar coupling depends on the motion of the molecule. It can provide information on molecular dynamics up to the time scale defined by the inverse of the measured

dipolar coupling. In proteins dipolar coupling can be used for description of slow motion up to the millisecond time scale. If a set of residual dipolar couplings is known for a protein residue it is possible to derive an order parameter and motional amplitudes in three spatial dimensions (three-dimensional Gaussian axial fluctuation (GAF) model¹⁵) for the peptide bond plane of this residue¹⁶. In this way RDCs allow quantitative evaluation of protein dynamics.

2. Protein chemical shift calculation

2.1 *Ab initio* route to chemical shift tensors and its limits

Biomolecular NMR spectroscopy permits the detailed study of the structure and dynamics of macromolecules both in the liquid and in the solid state. NMR chemical shifts are virtually always available even in the case where determination of NMR distance constraints is not possible. Moreover, NMR chemical shifts provide information on every single atom in the molecule under study. Empirical relationships between protein structure and NMR chemical shift can contribute to the protein structure elucidation³. However, the performance of empirical methods is limited by the data used for their development. Posttranslational modifications (e.g. glycosylation), L- to D-amino acids substitutions and such systems as pseudo-peptides pose large problems if studied with empirical methods. Furthermore, current empirical methods can predict only isotropic chemical shift values and not the full chemical shift tensors required for study of protein dynamics and orientation. This is why the development of *ab initio* quantum chemistry methods for nuclear shielding calculation gained a lot of attention during the last decades.

Application of an external magnetic field affects the electronic structure of the molecule under study. In the *ab initio* framework¹⁷, energy contributions arising from this interaction are responsible for nuclear shielding. Energy contributions from the interaction with an external magnetic field are significantly smaller than the ground-state molecular energy and therefore they can be treated in terms of a time-independent perturbation theory. This approach begins with an unperturbed Hamiltonian \hat{H}_0 with known energies and eigenstates. The perturbation of the Hamiltonian can be written as a power series dependent on a parameter λ (for simplicity energies are assumed to be discrete):

$$\hat{H} = \hat{H}_0 + \sum_{i=1}^n \lambda^i \hat{H}_i \quad (2.1)$$

and the energy levels and the eigenstates of the perturbed Hamiltonian are given by

$$E = E_0 + \sum_{i=1}^n \lambda^i E_i \quad (2.2)$$

and

$$\psi = \psi_0 + \sum_{i=1}^n \lambda^i \psi_i. \quad (2.3)$$

Perturbed Hamiltonian depending on two parameters λ and κ can be expressed as

$$\hat{H} = \hat{H}_0 + \lambda \hat{H}_{10} + \kappa \hat{H}_{01} + \lambda^2 \hat{H}_{20} + \kappa^2 \hat{H}_{02} + \kappa \lambda \hat{H}_{11}. \quad (2.4)$$

For simplicity only the terms up to the second order are given in equation 2.4. In this case the mixed second-order energy E_{11} is the derivative of the total energy:

$$E_{11} = \left(\frac{\partial^2 E}{\partial \lambda \partial \kappa} \right)_{\lambda=0, \kappa=0}. \quad (2.5)$$

The energy of a nucleus shielded by the electrons in an external magnetic field can be derived from equations 1.5 and 1.8 and written as:

$$E = \sum_{\alpha, \beta} \mu_\alpha \sigma_{\alpha\beta} B_\beta. \quad (2.6)$$

Thus chemical shielding tensor components can be expressed as the derivatives of the energy:

$$\sigma_{\alpha\beta} = \left(\frac{\partial^2 E}{\partial B_\alpha \partial \mu_\beta} \right)_{B=0, \mu=0}. \quad (2.7)$$

The comparison of equations 2.5 and 2.7 shows that if an external magnetic field and the magnetic dipole moment of the nucleus under study are considered as parameters perturbing the electronic state, the contribution to the energy responsible for the nuclear shielding is the mixed second-order energy:

$$E = \sum_{\alpha, \beta} \mu_\alpha \sigma_{\alpha\beta} B_\beta = \sum_{\alpha, \beta} \mu_\alpha \left(\frac{\partial^2 E}{\partial B_\alpha \partial \mu_\beta} \right)_{B=0, \mu=0} B_\beta = E_{11} \sum_{\alpha, \beta} \mu_\alpha B_\beta \quad (2.8)$$

Double perturbation theory provides an expression for description of the energy term responsible for the nuclear shielding:

$$E_{11} = \langle \psi_0 | \hat{H}_{11} | \psi_0 \rangle + 2 \langle \psi_0 | \hat{H}_{10} | \psi_{01} \rangle = \langle \psi_0 | \hat{H}_{11} | \psi_0 \rangle + 2 \langle \psi_0 | \hat{H}_{01} | \psi_{10} \rangle \quad (2.9)$$

The first order perturbation function required by equation 2.9 is given by:

$$(\hat{H}_0 - E_0)\psi_{01} + (\hat{H}_{01} - E_{01})\psi_0 = 0. \quad (2.10)$$

The development of ψ_0 into the eigenstates $\psi_{0,n}$ yields the expression:

$$E_{11} = \langle \psi_0 | \hat{H}_{11} | \psi_0 \rangle - \sum_{n \neq 0} (E_{11,n} - E_{11,0})^{-1} \left\{ \langle \psi_{0,0} | \hat{H}_{10} - E_{10} | \psi_{0,n} \rangle \langle \psi_{0,n} | \hat{H}_{01} - E_{01} | \psi_{0,0} \rangle \right\}. \quad (2.11)$$

The Hamiltonians required in equations 2.9 and 2.11 describe the energy of an electron in an external magnetic field. For non-relativistic closed-shell systems in an external magnetic field these Hamiltonians can be derived from the Dirac equation (relativistic quantum mechanical wave equation for description of spin- $1/2$ particles) at its non-relativistic limit (i.e. the small components of the four-component wave function in the Dirac equation are neglected). In this case, the Hamiltonian in the Dirac equation can be written as a series with its second-order component corresponding to the energies of interactions with an external magnetic field. For calculation of nuclear shieldings the sum of the spin-dependent contributions over the occupied spin orbitals is zero for a non-degenerate electron state and therefore spin-dependent contributions in the Hamiltonian can be neglected, so that one finally obtains*:

$$\hat{H}_1 = (\hat{H}_{10} + \hat{H}_{01}) + \hat{H}_{11} = \frac{e}{2m_e} \sqrt{\frac{\mu_0}{4\pi}} (\vec{A}_O \cdot \vec{p} + \vec{A}_N \cdot \vec{p}) + \frac{\mu_0}{4\pi} \cdot \frac{e^2}{2m_e} \vec{A}_O \cdot \vec{A}_N, \quad (2.12)$$

where $\mu_0(4\pi)^{-1}$ is the magnetic field constant, m_e is the electron mass, e is the elementary charge and \vec{p} is the momentum of the electron. \vec{A}_O and \vec{A}_N refer to the magnetic vector potentials of the external magnetic field and of the magnetic field of the nucleus N . The generalization from the one-electron energy contribution to the total energy of the n -electron system is achieved by the following substitution:

$$\vec{p} \rightarrow \sum_{i=1}^n \vec{p}_i. \quad (2.13)$$

In his pioneering work on the theory of nuclear shielding Norman F. Ramsey followed the ansatz implied by equation 2.11¹⁸. However, in Ramsey's approach the influence

* All formulae in this section are given in SI system of units. CGS or atomic units are occasionally used in the literature, sometimes causing confusion. The factor $\mu_0(4\pi)^{-1}$ in SI system corresponds to c^{-2} in electrostatic CGS and to c^{-1} in Gaussian CGS systems, respectively. In SI-based system of atomic units it corresponds to c^{-2} and in Hartree system of atomic units it is equal to 1 and $m_e = e = \hbar(2\pi)^{-1} = 1$ a.u. c refers to the velocity of light.

of the external magnetic field and the influence of the nuclear magnetic moment are treated as a single perturbation (cf. equation 2.1). Furthermore, since the spin-dependent contributions can be neglected in the calculation of the nuclear shielding the Hamiltonian for a spinless charged particle in the electromagnetic field is used a priori (V is the nuclear potential)

$$\hat{H} = \frac{1}{2m_e} (\vec{p} + \sqrt{\frac{\mu_0}{4\pi}} \cdot e\vec{A})^2 + V \quad (2.14)$$

and the derivation of the Hamiltonian is strongly simplified:

$$\hat{H} = \hat{H}_0 + \hat{H}_{10} + \hat{H}_{11} = \left(\frac{1}{2m_e} \vec{p}^2 + V \right) + \sqrt{\frac{\mu_0}{4\pi}} \cdot \frac{e}{m_e} (\vec{A} \cdot \vec{p}) + \frac{\mu_0}{4\pi} \cdot \frac{e^2}{2m_e} \vec{A}^2. \quad (2.15)$$

It is noteworthy that the equation 2.15 is only valid in a special case, where the total magnetic vector potential is

$$\vec{A} = \frac{1}{2} \vec{B} \times \vec{r} + \frac{\vec{\mu}_N \times \vec{r}}{r^3} = \vec{A}_0 + \vec{A}_K \quad (2.16)$$

with $\vec{R}_O = \vec{R}_N = \vec{0}$ and $\nabla \cdot \vec{A} = 0$ (cf. page 20 for a detailed explanation) and the second-order Hamiltonian from the equation 2.15 can be expressed as:

$$\hat{H}_1 = \hat{H}_{10} + \hat{H}_{01} + \hat{H}_{11} = \sqrt{\frac{\mu_0}{4\pi}} \cdot \frac{e}{m_e} (\vec{A}_O \cdot \vec{p}) + \sqrt{\frac{\mu_0}{4\pi}} \cdot \frac{e}{m_e} (\vec{A}_K \cdot \vec{p}) + \frac{\mu_0}{4\pi} \cdot \frac{e^2}{2m_e} \vec{A}_O \cdot \vec{A}_K. \quad (2.17)$$

Ramsey's approach yields the famous formula often found in NMR textbooks. In this formula chemical shielding tensor of a nucleus is considered as a sum of a diamagnetic term σ^d and a paramagnetic term σ^P :

$$\sigma = \sigma^d + \sigma^P. \quad (2.18)$$

From the physical point of view only the sum of both terms is meaningful. As already pointed out by Ramsey the separation into diamagnetic and paramagnetic terms is artificial (in the relativistic framework not even the corresponding terms do exist) and the exact magnitude of their individual contributions is strongly dependent on the choice of the magnetic vector potential. Therefore an attempt to interpret the diamagnetic term as the shielding by the core electrons and the paramagnetic term as the deshielding by the excited electronic states with paramagnetic properties is pointless.

The diamagnetic term depends only on the electronic ground state:

$$\sigma_{N,\alpha\beta}^d = \frac{\mu_0}{4\pi} \frac{e^2}{2m_e} \left\langle \Psi_0 \left| \sum_i \frac{(\vec{r}_{iA} \vec{r}_{iN}) \delta_{\alpha\beta} - r_{iA,\beta} r_{iN,\alpha}}{r_{iN}^3} \right| \Psi_0 \right\rangle, \quad (2.19)$$

where Ψ_0 denotes the ground-state wavefunction, \vec{r}_{iN} and \vec{r}_{iA} denote the distance of electron i from the nucleus N and from the origin of the external magnetic vector potential, respectively, and $\delta_{\alpha\beta}$ is the Kronecker delta function (not to be confused with chemical shift). The corresponding term in equation 2.9 is $\langle \psi_0 | \hat{H}_{11} | \psi_0 \rangle$ which is therefore referred to as the diamagnetic or the Lamb term.

The paramagnetic term is given by

$$\sigma_{N,\alpha\beta}^p = -\frac{\mu_0}{4\pi} \frac{e^2}{2m_e^2} \sum_{n \neq 0} (E_n - E_0)^{-1} \left\{ \left\langle \Psi_0 \left| \sum_i \hat{l}_{Ni,\alpha} r_{iN}^{-3} \right| \Psi_n \right\rangle \left\langle \Psi_n \left| \sum_i \hat{l}_{Ai,\beta} \right| \Psi_0 \right\rangle \right. \\ \left. + \left\langle \Psi_0 \left| \sum_i \hat{l}_{Ai,\beta} \right| \Psi_n \right\rangle \left\langle \Psi_n \left| \sum_i \hat{l}_{Ni,\alpha} r_{iN}^{-3} \right| \Psi_0 \right\rangle \right\}, \quad (2.20)$$

where Ψ_0 denotes the ground-state wavefunction and E_n and E_0 refer to the energies of the corresponding electronic states Ψ_n . The angular momentum operator with respect to the gauge origin (cf. equation 2.4) \hat{l}_A represents the interaction of electron i with the external magnetic field while the angular momentum operator with respect to nuclear position \hat{l}_{Ni} describes the interaction of electron i with the nuclear magnetic dipole field. The corresponding term in the equation 2.9 is $2 \langle \psi_0 | \hat{H}_{10} | \psi_{01} \rangle$ which is therefore referred to as the paramagnetic, the high-frequency or the Ramsey term.

In equation 2.12 as well as in the Ramsey's approach the external magnetic field \vec{B} at the spatial point \vec{r} is represented by its magnetic vector potential:

$$\vec{A}_O(\vec{r}) = \frac{1}{2} \vec{B} \times (\vec{r} - \vec{R}_O) \quad (2.21)$$

This definition of magnetic vector potential is not unique. Mathematical definition of the vector potential allows arbitrary addition of curl-free components to the magnetic vector potential without changing the observed magnetic field. This property of the magnetic vector potential is called gauge invariance. In equation 2.21 vector \vec{R}_O

points to a fixed point in space called the gauge origin and the choice of \vec{R}_O is arbitrary. For nuclear shielding calculations of single atoms the possible choice of the gauge origin is a natural gauge origin. In this case the vector potential \vec{A}_N created by the magnetic moment of a nucleus can be described by a magnetic point dipole $\vec{\mu}_N$ at nuclear position \vec{R}_N :

$$\vec{A}_N(\vec{r}) = \frac{\mu_0}{4\pi} \cdot \frac{\vec{\mu}_N \times (\vec{r} - \vec{R}_N)}{|\vec{r} - \vec{R}_N|^3}. \quad (2.22)$$

Unfortunately usually there is neither a natural gauge origin in molecules nor a generally applicable rule for the gauge origin choice of the external magnetic field vector potential. Ramsey's approach relies on a single gauge origin with $\vec{R}_O = \vec{R}_N$. It satisfies the condition $\nabla \cdot \vec{A} = 0$. The last setting is denoted as Coulomb gauge and is usually applied for semi-classical calculations where the magnetic vector potential is quantized but Coulomb interactions are not.

The energy of an atom or a molecule in an external magnetic field depends on the magnetic flux density (cf. equation 1.5) and therefore it is invariant under a change of the gauge origin of the vector potential (like the observed magnetic field is). The same holds for Zeeman energy levels and observed chemical shielding. For the Ramsey equations (2.19 and 2.20) this condition is fulfilled only for the exact solutions of the Schrödinger equation. For approximate solutions usually provided by computational methods the equations 2.19 and 2.20 are not individually gauge invariant and lead to large errors¹⁹. Furthermore Ramsey's approach requires the knowledge of all the excited states of the system under study. However, many *ab initio* methods do not calculate excited states of the system under study explicitly. Therefore, though Ramsey equations can facilitate the qualitative interpretation of the nuclear shieldings, they are not well-suited for quantitative calculations.

Current methods for quantum chemical nuclear shielding calculations use equation 2.9 as an ansatz. They apply the concept of distributed gauge origins in order to overcome the gauge origin problem. A fundamental property of the operator of finite translation in space \hat{T} is that it commutes with the momentum operator and atomic orbitals can be translated to each point in space without changing the kinetic energy.

In presence of the external magnetic field, the magnetic vector potential must be included into the operator of finite translation in space. The operator can then be applied to one-electron wave functions translating it to a chosen gauge origin without changing the energy responsible for nuclear shielding. Such orbitals are known as London orbitals or gauge-including atomic orbitals²⁰ (GIAO) and can be written as²¹:

$$\chi_{\mu}(\vec{r}, \vec{B}) = \hat{T}(\vec{R}_O) \chi_{\mu}^o(\vec{r}) = \exp \left\{ -\frac{i}{2} \cdot \frac{e^2}{4\pi\epsilon_0\hbar c} (\vec{B} \times (\vec{R}_{\mu} - \vec{R}_O) \cdot \vec{r}) \right\} \chi_{\mu}^o(\vec{r} - \vec{R}_O), \quad (2.23)$$

where $\chi_{\mu}^o(\vec{r} - \vec{R}_O)$ is the standard atomic orbital with the coordinates \vec{R}_{μ} as center and \vec{R}_O is the chosen gauge origin, ϵ_0 denotes the electric constant. The energy responsible for nuclear shielding can then be obtained from the solutions of the coupled Hartree-Fock equations (Hartree-Fock (HF) is the standard *ab initio* method for approximate determination of the ground-state wave function and ground-state energy in many-electron systems) with the external magnetic field as perturbation and nuclear shielding can be calculated using the equation 2.8. This approach does not lead to the true gauge invariance, however, it achieves gauge-origin independence since the terms containing the gauge origin cancel out during the calculation. An alternative approach for nuclear shielding calculation relies on the application of the localized orbitals where a gauge origin can be assigned to each orbital. This approach is known as the IGLO²² (individualized gauges for localized orbitals) method.

The conversion of *ab initio* calculated nuclear shieldings into chemical shifts requires the chemical shielding of the reference substance (cf. equation 1.9). The shielding of the reference substance has to be determined separately for each basis set and computational method. Difficulties can arise from the fact that computational methods (especially density functional theory where empirical data is used in the development of hybrid functionals) perform differently for different classes of compounds and that *ab initio* calculations do not include thermal effects and vibrational corrections to nuclear shielding.

In marked contrast to empirical methods, *ab initio* methods can calculate the full chemical shift tensor and they are not restricted to a specific class of compounds.

Therefore at a first glance the application of *ab initio* methods could circumvent the limitations of empirical methods. Accurate prediction of nuclear shieldings often requires sufficiently large and flexible basis sets and electron correlation in the molecule (not considered by the HF method) must be taken into account. The GIAO approach has been implemented for post Hartree-Fock methods (post HF methods include electron correlation) like second order Møller-Plesset perturbation theory²³ (MP2) and for density functional theory (DFT). DFT nuclear shielding calculations are available for oligopeptides²⁴ but unfortunately such computations are too demanding for full-size proteins and their application to real problems is in many cases far out of reach.

Chemical shielding is dominated by the local environment. A significant speed-up of *ab initio* computations can be achieved if this fact is taken into account. In standard *ab initio* methods all electrons of the system contribute to chemical shielding of every single atom. Beer et al. introduced a sublinear-scaling method where unphysical long-range contributions to chemical shielding are truncated²⁵. Alternatively, it is possible to reduce the size of the system under study (and therefore the computational burden) to the nucleus of interest and its environment. Vila et al. performed calculations on proteins by treating each amino acid X in the protein sequence as a terminally blocked tripeptide Ac-GXG-Me in a conformation obtained from experimental protein structure²⁴. If no changes in the topology of the molecule are desired *ab initio* methods can be combined with force field methods²⁶ (QM/MM). In this case the nucleus of interest and its environment form the region for quantum chemical calculations and the rest of the molecule is described by molecular mechanics (e.g. nuclei are represented as point charges). Unfortunately, the approaches described above are restricted to ¹³C chemical shift calculation as ¹⁵N chemical shifts are too sensitive to long-range effects²⁷. Moreover *ab initio* calculations of chemical shifts cannot be applied to chemical shift driven molecular structure refinement. Calculation of forces derived from the chemical shift target function would require the derivatives of the chemical shifts with respect to nuclear coordinates but no methods for *ab initio* calculation of such derivatives have been developed up till now.

Semi-empirical methods are able to provide the full chemical shift tensor at low computational cost. Therefore this thesis is concerned with the adaptation of a semi-empirical method – the bond polarization theory⁸ (BPT) to the calculation of protein NMR chemical shifts and chemical shift driven protein structure refinement. This chapter deals with the evaluation and fine-tuning of the BPT parameterization for ¹⁵N and ¹³C protein chemical shift calculation and of the BPT protocols for chemical shift driven geometry optimization. Applications for the study of biomolecular orientation in lipid membranes and protein dynamics are presented in the following chapters.

2.2 Bond polarization theory

NMR chemical shifts are localized properties and localized orbitals can be applied for their calculation. Moreover localized orbitals can be used in order to describe the chemical concept of bonds and electronic inner shells and lone pairs. Bond polarization theory is a semi-empirical method for the calculation of localized molecular properties. As the name implies, bond polarization theory is based on bond orbitals being approximations to localized orbitals. Hybrid wave functions $\chi(\vec{r})$ pointing into the bond direction are constructed as linear combinations of Slater atomic orbitals (method of Del Re²⁸). Bond and anti-bond orbitals $\psi(\vec{r})$ and $\psi^*(\vec{r})$ centered on the bond i between the atoms designated as a and b are set up as linear combinations of hybrid wave functions $\chi(\vec{r})$:

$$\psi_i(\vec{r}) = c_{ai}\chi_a^i(\vec{r}) + c_{bi}\chi_b^i(\vec{r}), \quad (2.24)$$

$$\psi_i^*(\vec{r}) = c_{bi}\chi_a^i(\vec{r}) - c_{ai}\chi_b^i(\vec{r}). \quad (2.25)$$

The ground state wave function Ψ_0 is a Slater determinant of n bond orbitals of the molecular system under study:

$$\Psi_0 = \frac{1}{\sqrt{2n!}} \text{Det} \left| \psi_1^+ \psi_1^- \dots \psi_i^+ \psi_i^- \dots \psi_n^+ \psi_n^- \right|, \quad (2.26)$$

where + and – indicate the spin states. Excited configurations Ψ_i are introduced by substituting one or two bond orbitals in the ground state Slater determinant by anti-bond orbitals:

$$\Psi_I = \Psi \begin{pmatrix} j^* \\ i \end{pmatrix} = \frac{1}{\sqrt{2n!}} \text{Det} |\dots \psi_j^{*+} \psi_i^- \dots|. \quad (2.27)$$

Addition of the excited configurations with $i = j$ and $i \neq j$ to the ground state wave function describes polarization and delocalization. In this way molecular wave function Ψ can be developed into configurations and calculated by perturbation theory:

$$\Psi = \Psi_0 + \sum_I c_I \Psi_I = \Psi_0 + \sum_{ij^*} \frac{\langle \Psi_0 | \hat{H} | \Psi \begin{pmatrix} j^* \\ i \end{pmatrix} \rangle}{E_0 - E \begin{pmatrix} j^* \\ i \end{pmatrix}} \Psi \begin{pmatrix} j^* \\ i \end{pmatrix} + \sum_{\text{doubly excited configurations}}. \quad (2.28)$$

Neglecting delocalization $\Psi \begin{pmatrix} j^* \\ i \end{pmatrix}_{i \neq j}$ and double-excitation $\Psi \begin{pmatrix} j^* l^* \\ i k \end{pmatrix}$ contributions to the molecular wave function, the molecular system can be divided into subunits (e.g. bonds belonging to the nucleus of interest), each one represented by its own localized wave function Ψ_C . It is noteworthy that orbital localization does not exclude long-range electrostatic effects since bond polarization in one subunit still depends on the charge distribution of the whole molecular system (cf. equation 2.30).

Large contributions to nuclear shielding arise from the local environment (i.e. the bonds) near the nucleus under study. Therefore in BPT framework nuclear shielding of an atom is a sum of individual bond contributions of the bonds connected to this atom. Sternberg and Prieß demonstrated that in this case nuclear shielding or chemical shift can be expressed as a sum of expectation values of one electron operators²⁹:

$$\langle \Psi_C | \hat{O}^{CS} | \Psi_C \rangle = 2 \sum_i^{i \in C} \langle i | \hat{O}_i | i \rangle, \quad (2.29)$$

where the index i denotes the bonds i_{a-b} of the nucleus under study a forming a molecular subunit C and i as a short notation for the bond orbital ψ_i .

Ab initio methods for nuclear shielding calculation require the knowledge of the molecular wave function Ψ from the very beginning. Moreover the solution to the gauge origin problem must be found. Bond polarization theory avoids time consuming

perturbation calculations by deriving the expression for the change of the nuclear shielding expectation value of an atom in a molecular subunit C when the surrounding X of C is changing. If the influence of the molecular surrounding X is described by the atomic point charges q_x (i.e. the Hamiltonian in equation 2.28 is replaced by the point charge electrostatic potential operator \hat{V}_x), the application of the formalism expressed in the equation 2.29 to the perturbed wave function of the bond (cf. equation 2.28) yields:

$$\langle \Psi_C | \hat{O}^{CS} | \Psi_C \rangle = 2 \sum_i^{i \in C} \left(\langle i | \hat{O}_i | i \rangle + \frac{2}{\Delta E_i} \langle i | \hat{O}_i | i^* \rangle \langle i^* | \hat{V}_x | i \rangle \right). \quad (2.30)$$

For simplicity equation 2.30 can be written in shorthand notation:

$$\langle \Psi_C | \hat{O}^{CS} | \Psi_C \rangle = \sum_i^{i \in C} \left(2P_i^0 + 2^2 A_i \langle i^* | \hat{V}_x | i \rangle \right). \quad (2.31)$$

The term P_i^0 depends only on the ground state wave function. It can be interpreted as the nuclear shielding contribution from an unpolarized bond (hypothetical bond without any surrounding charges). The second term in equation 2.31 describes the influence of the molecular charge distribution on the bond contribution to the nuclear shielding. The term $\langle i | \hat{V}_x | i^* \rangle$ is referred as polarization energy V_i of the bond. The terms P_i^0 and A_i can be treated as constants for each bond type. Therefore no explicit operator evaluation is required and computational burden can be significantly reduced compared to *ab initio* methods. Furthermore, the explicit form of the nuclear shielding operator \hat{O}^{CS} vanishes in the constant terms and no explicit reference to the magnetic vector potential is required. In this way gauge invariance is achieved.

In the practical implementation of equation 2.31 bond occupation numbers n_i are introduced to treat molecular systems with conjugated bonds:

$$\langle \Psi_C | \hat{O}^{CS} | \Psi_C \rangle = \sum_{i \in C} \left(n_i P_i^0 + n_i^2 A_i \sum_{x \in C} \left[\left\langle \chi_a^i \left| \frac{q_x}{\vec{R}_x - \vec{r}} \right| \chi_a^i \right\rangle - \left\langle \chi_b^i \left| \frac{q_x}{\vec{R}_x - \vec{r}} \right| \chi_b^i \right\rangle \right] \right). \quad (2.32)$$

Nuclear shielding of a nucleus a is the sum of the expectation values arising from single bonds connected to this nucleus. Nuclear shielding operator \hat{O}^{CS} singles out a configuration of bonds C forming the quantum-mechanical region of the calculation. The first term P_i^0 in equation 2.32 remains unchanged compared to equation 2.31.

The second term in equation 2.32 represents the bond polarization energies in a quantum-mechanical region calculated as a sum over all partial charges q_x outside the quantum mechanical region. For the calculation of the second term only the partial charges q_x and their positions \vec{R}_x are required since the hybrid functions χ can be constructed solely from the bond geometry. The sum in equation 2.32 runs over all bonds i between the nucleus a and its neighbours (each neighbouring atom is denoted as b). Bond occupation numbers are estimated with the help of the empirical relationship between the valence of the bond v_{a-b} , bond length d_{a-b} , and the empirical bond valence parameter R_{a-b} (which is in a formal sense an ideal single bond length) defined by O'Keeffe and Breese³⁰:

$$v_{a-b} = \exp\left[\frac{(R_{a-b} - d_{a-b})}{0.37}\right]. \quad (2.33)$$

The occupation numbers n_i of conjugated π -bonds are obtained as $n_i = 2(v_i - 1)$ and the occupation number of σ -bonds is set to $n = 2$.

Depending on the choice of the constant terms P_i^0 or A_i , bond polarization theory can provide either nuclear shieldings or directly chemical shifts. The isotropic chemical shift δ is the sum of the bond contributions δ_i consisting of the chemical shift of an unpolarized bond $P_i^0 = \delta_i^0$ and depending on the parameter A_i^δ describing the change of the CS expectation value with bond polarization by surrounding atomic point charges:

$$\delta = \sum_i (n_i \delta_i^0 + n_i^2 A_i^\delta V_i). \quad (2.34)$$

The parameters δ_i^0 and A_i^δ can be obtained by a linear parameterization procedure if a sufficiently large set of experimental or calculated CS values and corresponding molecular structures are known. Every element combination $a-b$ and bond type (σ or π) requires a separate set of two parameters. The same is valid for the isotropic nuclear shielding. For chemical shift (or nuclear shielding) tensor calculations the parameters δ_i^0 and A_i^δ become tensors. These tensors can be regarded as diagonal in the bond coordinate system (with the z-axis pointing into bond direction; for amide nitrogen in a peptide bond the z axis is pointing along the N-H or N-C bond vector and the x axis is perpendicular to the π -plane) and 6 parameters per bond type have

to be determined. For the parameterization procedure in the tensor form the equation 2.34 can be expressed as:

$$\begin{pmatrix} \sigma_{11}^{calc} \\ \sigma_{22}^{calc} \\ \sigma_{33}^{calc} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & V_{pol} & 0 & 0 \\ 0 & 1 & 0 & 0 & V_{pol} & 0 \\ 0 & 0 & 1 & 0 & 0 & V_{pol} \end{pmatrix} \begin{pmatrix} \sigma_{11}^0 \\ \sigma_{22}^0 \\ \sigma_{33}^0 \\ A_{11}^\sigma \\ A_{22}^\sigma \\ A_{33}^\sigma \end{pmatrix}, \quad (2.35)$$

or in the shorter notation:

$$\vec{\sigma}^{calc} = \mathbf{M}\vec{p}. \quad (2.36)$$

For simplicity equations 2.34 and 2.35 refer to the ideal case of a molecular system consisting only of one bond surrounded by a charge distribution. In this case equation 2.34 must be augmented by transformation matrices \mathbf{D} for the transformation from the bond coordinate system to a common molecular coordinate system:

$$\sigma_{\alpha\beta} = \sum_i^{i \in A} \sum_{\alpha'\beta'} D_{\alpha\alpha'}^i D_{\beta\beta'}^i (n_i \delta_{\alpha'\beta'}^i + n_i^2 A_{\alpha'\beta'}^i V^i). \quad (2.37)$$

Chemical shift prediction with bond polarization theory requires accurate parameterization for the empirical factors P_i^0 and A_i as well as the accurate description of atomic point charges. *Ab initio* charge calculation in proteins is computationally very demanding. Fortunately bond polarization theory can also be applied to the calculation of atomic point charges³¹. The formalism for the BPT charge calculation is analogous to the formalism for the BPT CS calculation. In the BPT framework, partial atomic charges are the sums of bond contributions consisting of the partial charge of an unpolarized bond and the change of the charge with bond polarization. For the calculation of the partial atomic charges q_A the knowledge of all other atomic charges q_x of the molecular system is required. Therefore, all charges in the molecule have to be determined simultaneously leading to a set of equations with the dimension equal to the number of atoms in the system under study. The solution of this set of equations is the time-determining step of the BPT CS calculation. The BPT charge calculation is implemented in the COSMOS-NMR force field³² and atomic charges can be calculated in every step of molecular dynamics

(MD) simulation or geometry optimization. A well established parameterization for the BPT charge calculation on the NPA/6-31G(d,p) level of theory³³ exists for the majority of elements found in biomolecules. Therefore no additional adjustments of the parameterization for the BPT charge calculation are required and the following sections of this thesis are focused on the evaluation and correction of the BPT parameterization for the protein CS calculation.

2.3 ¹⁵N and ¹³C parameterization

The performance of semi-empirical methods critically depends on their parameterization. The original parameterization for the BPT chemical shift calculation^{29,34} (cf. table 2.1) was focused on the ¹³C nucleus. It was obtained on experimental solid-state NMR chemical shift tensors and corresponding neutron diffraction crystal structures. The parameterization on the experimental data allows the direct prediction of the chemical shifts and no conversion of the nuclear shieldings into chemical shifts is needed. However, the set of molecules used for the original parameterization did not include peptides. Moreover the amount of molecules with known neutron diffraction structures and measured NMR chemical shift tensors available at once is very limited. These circumstances resulted in a parameterization bias towards carbohydrates³⁵.

BPT parameterization for the ¹⁵N chemical shift tensor calculation (cf. table 2.2) relied on *ab initio* calculations on the MP2/TZVPP level of theory^{36,37}. The set of compounds for the ¹⁵N parameterization included molecules containing nitrogen in typical amid surrounding mimicking the protein backbone. *Ab initio* calculations yield only nuclear shieldings and the comparison of the calculated with the experimental data requires their transformation into chemical shifts (cf. equation 1.9). BPT calculated ¹⁵N nuclear shieldings can be converted into chemical shifts using the reference value $\sigma_{reference}$ of 273.3±0.1 ppm (referencing to liquid NH₃) as proposed by Gauss et al. for the MP2 calculations³⁸ used for BPT parameterization.

First chemical shift calculations on a pseudo-peptide using the original ¹³C parameterization³⁹ appeared to be very promising³⁹. Even so a careful evaluation of

Table 2.1: BPT parameters for the ^{13}C nuclear shielding calculation. All parameters are given with respect to the bond coordinate system (with the z-axis pointing into bond direction; for amide nitrogen in a peptide bond the z axis is pointing along the N-H or N-C bond vector and the x axis is perpendicular to the π -plane).

Chemical shift parameters							
		Bond parameters in ppm			Polarization parameters in ppm/Hartree		
Parameter type	Bond type	δ_{xx}^0	δ_{yy}^0	δ_{zz}^0	A_{xx}^σ	A_{yy}^σ	A_{zz}^σ
C(sp ³)-H	σ	-7.53341	-7.53341	24.2796	170.520	170.520	-201.744
C(sp ²)-H	σ	26.0823	26.0823	66.7377	-104.062	-104.062	658.061
C(sp ³)-C*	σ	10.1435	10.1435	23.3047	-55.1875	-55.1875	-36.9642
C(sp ²)-C	σ	22.8709	22.8709	83.4502	378.360	378.360	115.146
C(sp ²)-C	π	106.098	-46.7495	-58.9496	91.5204	-106.387	-7.49204
C(sp ³)-N*	σ	16.4508	16.4508	4.91665	22.4814	22.4814	181.914
C(sp ²)-N	σ	13.3698	13.3698	52.0428	-248.964	-248.964	520.119
C(sp ²)-N	π	160.079	-6.64771	45.9487	-319.140	401.796	-660.0
C(sp ³)-O*	σ	51.4422	51.4422	25.9644	-260.012	-260.012	64.4191
C(sp ²)-O	σ	54.7308	54.7309	61.4657	-569.301	-569.301	428.003
C(sp ²)-O	π	38.3434	16.7368	-51.3646	475.929	-49.3557	119.727

the ^{13}C parameterization is necessary. The quality of the ^{15}N parameterization was assessed in a self-consistent way. *Ab initio* calculated nuclear shieldings used for the parameterization were compared to the nuclear shieldings back-calculated by the BPT using the determined parameters. BPT ^{15}N isotropic nuclear shielding calculations were able to reproduce *ab initio* calculations with a correlation coefficient

Table 2.2: BPT parameters for the ^{15}N nuclear shielding calculation. All parameters are given with respect to the bond coordinate system.

Nuclear shielding parameters							
		Bond parameters in ppm			Polarization parameters in ppm/Hartree		
Bond type		σ_{xx}^0	σ_{yy}^0	σ_{zz}^0	A_{xx}^σ	A_{yy}^σ	A_{zz}^σ
N-H(σ)		-147.202	-123.794	-141.027	-215.973	-327.562	-50.1749
N-C(σ)		-176.435	-78.8433	-113.169	-2987.10	-1317.63	-374.717
N-C(π)		657.830	63.6845	192.951	3935.61	1169.85	1841.42

of $R=0.988$ and standard deviation of 4.7 ppm^{36} . In marked contrast to the self-consistent evaluation first attempts to compute the peptide ^{15}N chemical shifts using the BPT ^{15}N parameterization were unsatisfactory⁴⁰. However, it should be emphasized that a reliable parameterization is not the only prerequisite for realistic ^{15}N chemical shift prediction. Cheeseman et al. evaluated the performance of the GIAO DFT and MP2 ^{15}N chemical shift computation⁴¹. Despite the fact that both methods have proven to be reliable, Cheeseman et al. observed mean absolute errors of 28.4 ppm and 11.8 ppm for DFT and MP2, respectively. The successful prediction of the ^{15}N chemical shifts requires a very accurate description of molecular structure and dynamics. Small changes in molecular surrounding of the ^{15}N nucleus can have a significant impact on its chemical shift. Reto Bader demonstrated that even a single charge of $+0.25$ proton units in 3.5 \AA distance from amide ^{15}N nucleus can induce chemical shift differences of up to 6 ppm^{27} . Woolf et al. investigated the influence of molecular dynamics on the chemical shift of an amide ^{15}N nucleus in peptide backbone and predicted large chemical shift fluctuations even on the short time scale of 0.05 ps^{42} . Therefore the evaluation of the ^{15}N parameterization requires a very careful choice of the test system.

2.4 Crystalline tripeptides: the test case for ^{15}N parameterization

As already mentioned in section 2.3 molecular surrounding and molecular motion must be taken into account for the prediction of the ^{15}N chemical shifts. Therefore crystalline tripeptides Ala-Gly-Gly, Ala-Pro-Gly and Gly-Gly-Val were selected as a test system for the evaluation of the ^{15}N BPT parameterization. Waddell et al. carried out single-crystal NMR investigations on these peptides ensuring that crystal modifications obtained for the chemical shift measurements were the same as originally defined by crystallography^{43,44}. Moreover Waddell et al. did not observe ^{15}N chemical shift tensor narrowing or changes in ^{15}N chemical shift tensor components in the temperature range between $+20$ and $-123 \text{ }^\circ\text{C}$. This fact indicates a relatively low impact of molecular dynamics on the ^{15}N chemical shifts. Waddell et al. also tried to estimate the description of the molecular surrounding sufficient for the ^{15}N chemical shift computation. Calculations on the DFT/B3LYP 6-311++G(d,p) level of theory revealed that tripeptide monomers were not suitable for the correct prediction

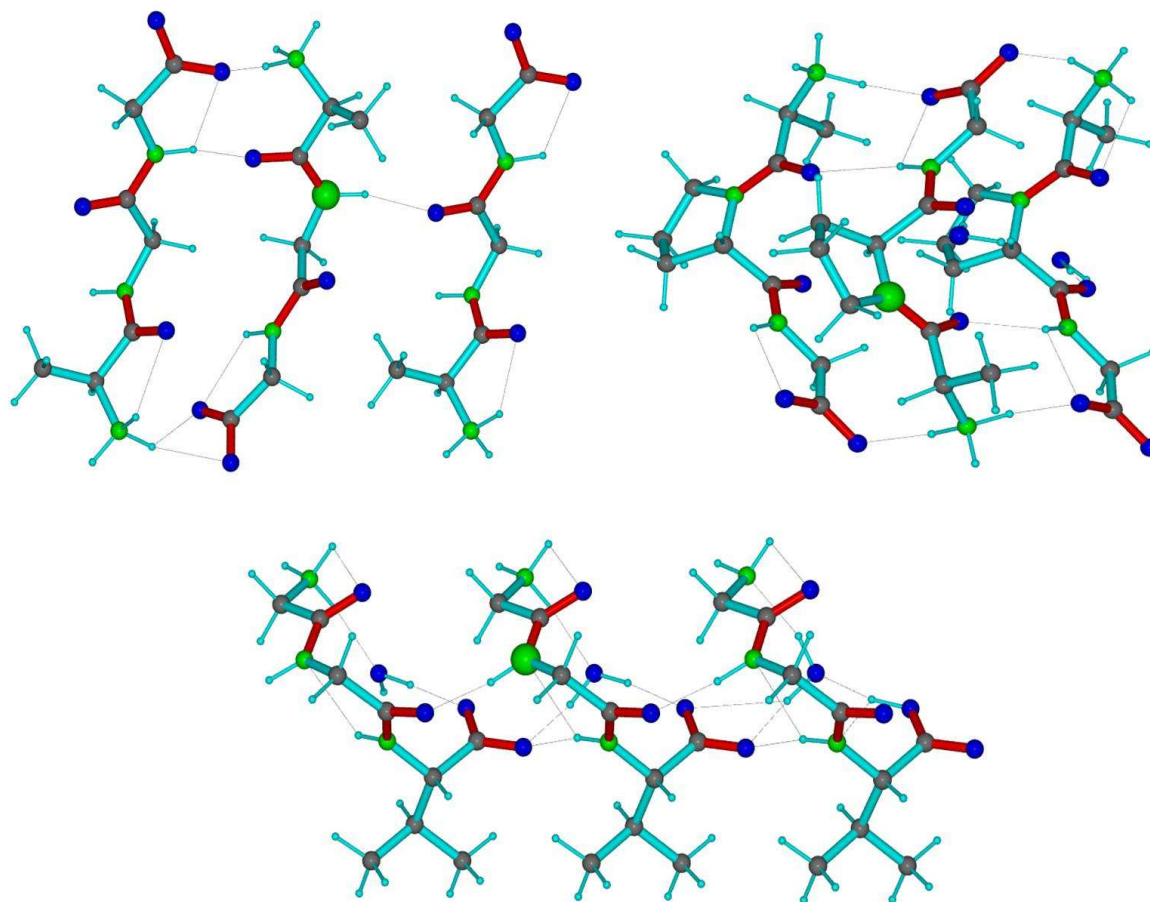


Figure 2.1: Structures of the AGG (at the top, left), APG (at the top, right) and GGV (at the bottom) trimers used in DFT and BPT calculations. Target nitrogen is emphasized and hydrogen bonds are shown in gray.

of the ^{15}N chemical shift in crystal structures. An attempt to consider the intermolecular interactions inside the crystal structure by construction of trimer peptide clusters (cf. figure 2.1) did not allow to reproduce experimental results. The mean absolute error (MAE) of DFT calculations was 8.7 ppm and the results did not improve with larger basis sets.

Initial BPT ^{15}N chemical shift calculations confirmed that neither tripeptide monomers nor trimer clusters provide a sufficient description of the corresponding crystal structures (cf. table 2.3 and figure 2.3). Significant chemical shift differences were observed between the tripeptide monomers and the trimers. The mean absolute error of BPT calculations was 21.2 ppm for monomers and 4.6 ppm for trimer clusters, respectively. A more accurate description of the tripeptide crystal structure was

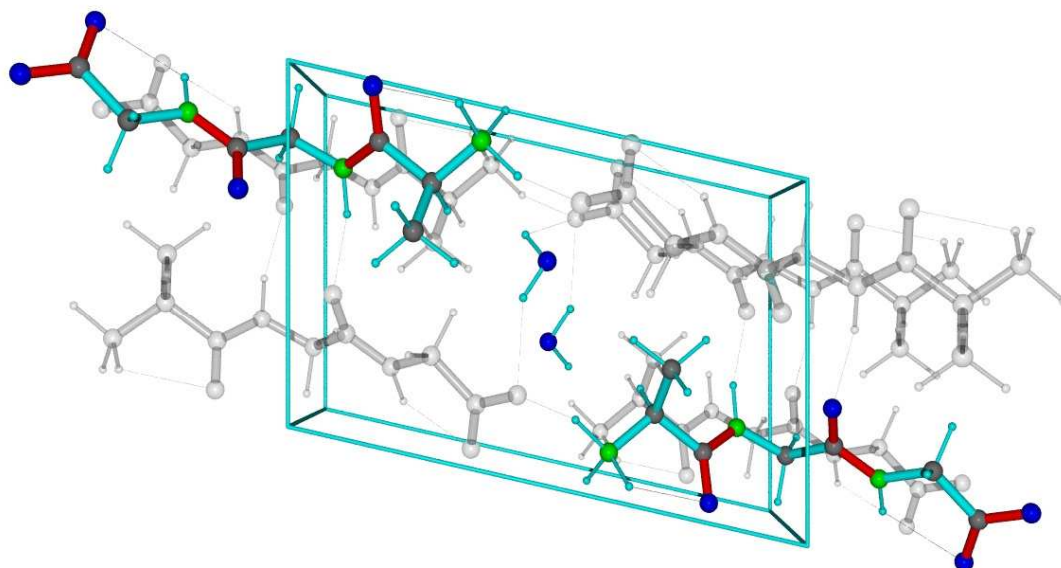


Figure 2.2: Unit cell of AGG tripeptide. Structures belonging to the asymmetric unit are shown in full colour.

obtained by modeling crystal unit cells (cf. figures 2.2 and 2.4). Atomic charges and chemical shifts in crystal unit cells were calculated under periodic boundary conditions. This approach significantly improved BPT calculated chemical shifts leading to a mean error of 2.1 ppm. Furthermore it became obvious that hydrogen atom positions seem to play an important role for ^{15}N chemical shift calculations in crystal structures. For modeling of the tripeptide unit cells, hydrogen atom positions were subjected to conjugate gradient geometry optimization (cf. the methodology chapter 5 for details). The resulting geometry was sufficient for reproducing the experimental ^{15}N chemical shifts of Ala-Gly-Gly and Ala-Pro-Gly tripeptides. First calculations of the ^{15}N chemical shift in the Gly-Gly-Val unit cell structure led to an absolute error of 22.6 ppm. An additional hydrogen position optimization using molecular dynamics under periodic boundary conditions was required to reproduce experimental results.

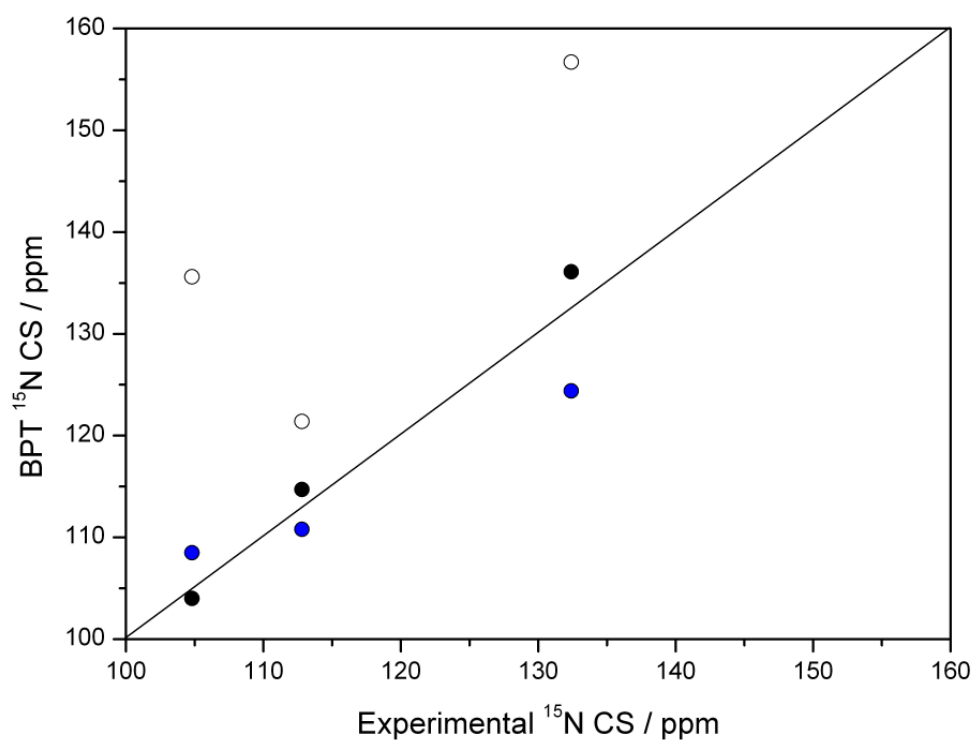


Figure 2.3: Comparison of the experimental and BPT calculated ^{15}N chemical shifts in central residues of crystalline tripeptides (solid black circles – calculation on unit cells, solid blue circles – calculation on trimer clusters, open black circles – calculation on monomers).

Table 2.3: Comparison of the experimental and BPT calculated ^{15}N chemical shifts in central residues of crystalline tripeptides (MAE denotes mean absolute error).

^{15}N Chemical shift in ppm					
Molecule	Experiment	BPT – Unit cell	BPT – Trimer	DFT – Trimer	BPT – Monomer
APG	132.4±0.5	136.1	124.4	141	156.7
AGG	104.8±0.5	104.0	108.5	113	135.6
GGV	112.8±0.5	114.7	110.8	122	121.4
MAE		2.1	4.6	8.7	21.2

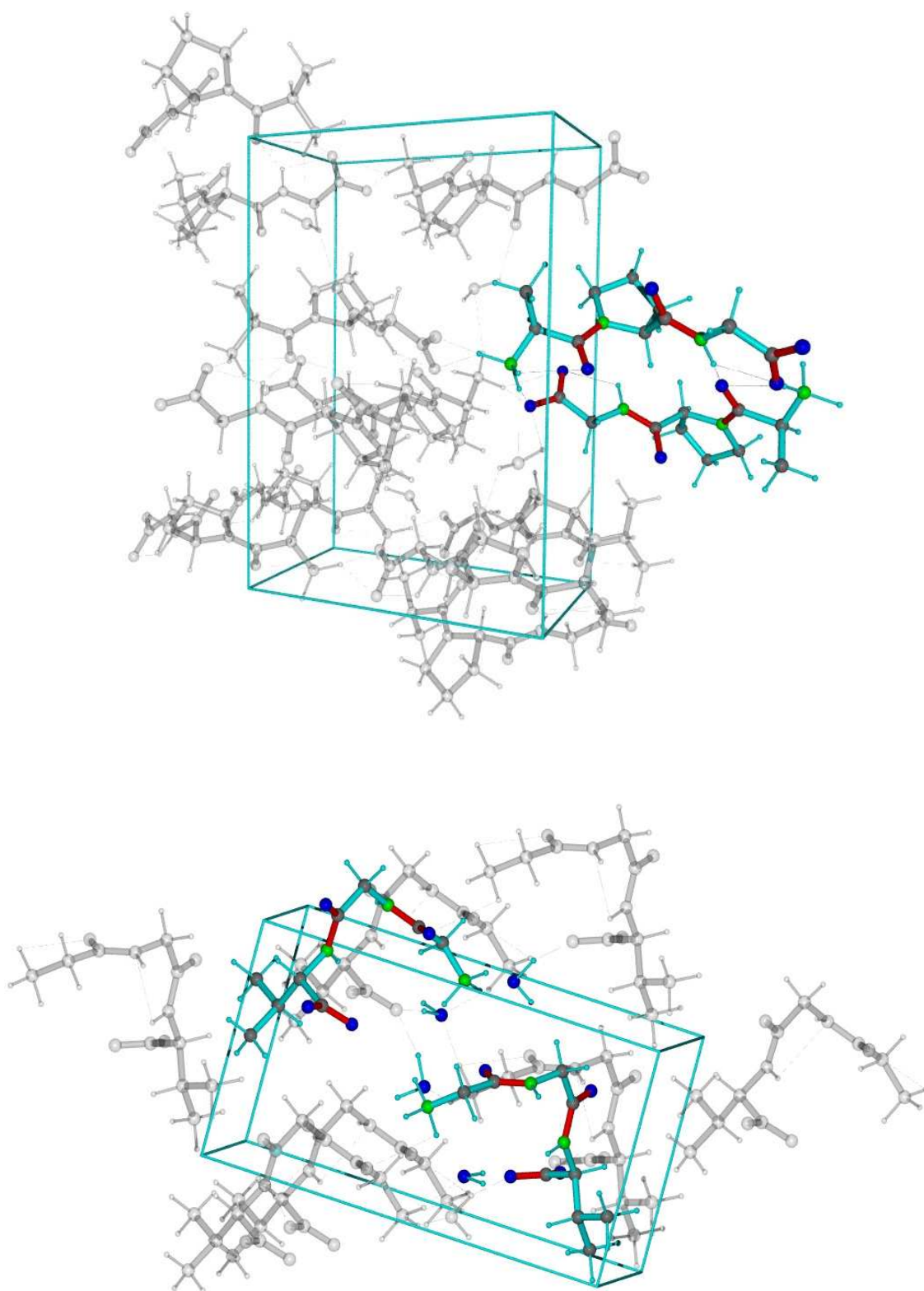


Figure 2.4: Unit cells of APG (at the top) and GGV (at the bottom) tripeptides used in BPT calculations with periodic boundary conditions. Structures belonging to the asymmetric unit are shown in full colour.

The next system selected for the evaluation of the BPT ^{15}N chemical shift calculation is the chemotactic peptide N-formyl-L-Met-L-Leu-L-Phe-OH (f-MLF-OH). Rienstra et al. determined the solid-state NMR structure of f-MLF-OH with high accuracy using precise ^{13}C - ^{13}C and ^{13}C - ^{15}N distance measurements and torsion angle measurements⁴⁵. The three-dimensional structure of the f-MLF-OH monomer was calculated by a sophisticated full structure search method (cf. figure 2.5). At first sight, ^{15}N chemical shift calculations on f-MLF-OH do not appear promising since the crystal structure of f-MLF-OH remains unknown. Nevertheless initial calculations using the experimental f-MLF-OH structure of Rienstra et al. led to a mean absolute error of 3.1ppm (cf. table 2.4). Optimization of the hydrogen positions in f-MLF-OH structure did not improve the calculated ^{15}N CS values. The next logical step was an attempt to include the influence of molecular dynamics on the ^{15}N chemical shift in f-MLF-OH in BPT calculations. BPT ^{15}N chemical shift calculations can be performed “on the fly” in every step of molecular dynamics. On the one hand, unconstrained molecular dynamics simulation can easily perturb molecular structure and therefore lead to significant errors in ^{15}N chemical shift computation. On the other hand, calculations on the Gly-Gly-Val tripeptide demonstrated that hydrogen atom positions (and therefore their dynamics) can have a strong influence on the ^{15}N chemical shift.

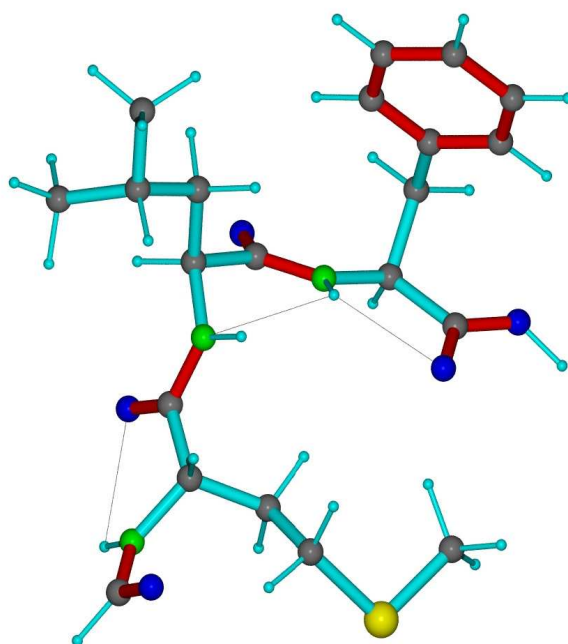


Figure 2.5: Three-dimensional structure of N-formyl-L-Met-L-Leu-L-Phe-OH peptide⁴⁵.

Protein chemical shift calculation

Table 2.4: Experimental and calculated chemical shifts in f-MLF-OH (all ^{13}C CS are referenced to DSS): MD – averaged CS values over 10 ps molecular dynamics trajectory with all heavy atoms fixed; FSS – CS values of structure obtained via full structure search of the conformational space; ^1H opt. – CS values of structure obtained via full structure search after hydrogen position optimization. MAE denotes mean absolute error.

Chemical shift in ppm					
Residue	Experiment	BPT – MD	BPT – FSS	BPT – ^1H opt.	DFT – ^1H opt.
^{15}N					
Met	125.5	124.3	125.2	127.1	146.6
Leu	116.2	117.5	119.2	124.6	127.2
Phe	107.6	109.3	113.6	101.4	109.6
MAE		1.4	3.1	5.4	11.4
$^{13}\text{C}_\alpha$					
Met	52.0	54.8	54.9	55.0	58.4
Leu	56.8	55.2	55.3	55.1	63.9
Phe	54.4	54.4	54.8	54.1	62.7
MAE		1.5	1.6	1.7	7.3
$^{13}\text{C}_\beta$					
Met	37.9	39.7	39.7	40.1	47.4
Leu	40.7	41.1	41.0	41.4	49.8
Phe	36.9	42.0	41.9	41.8	37.9
MAE		2.4	2.4	2.6	6.5

Therefore molecular dynamics (MD) was performed keeping the position of all non-hydrogen atoms fixed. Chemical shifts were calculated after each molecular dynamics step and an exponential memory decay function (cf. the methodology chapter 5 for details) was applied for the calculation of the time averaged chemical shifts in order to minimize the influence of the molecular dynamics equilibration phase (cf. figure 2.6). Small chemical shift fluctuations (<0.5 ppm) still could be observed after 10 ps of molecular dynamics simulation and calculated chemical shifts did not converge to constant equilibrium values. Nevertheless dynamically averaged ^{15}N chemical shifts after 10 ps were in a very good agreement with experimental results (mean absolute error of 1.4 ppm; cf. figure 2.7).

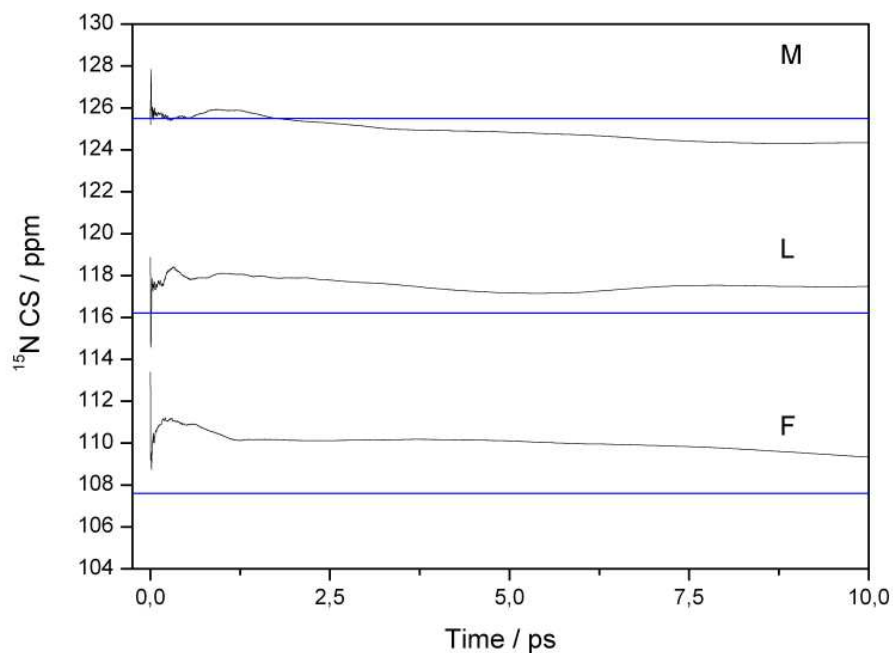


Figure 2.6: The trajectory of the calculated averaged ^{15}N chemical shifts in f-MLF-OH during molecular dynamics simulation.

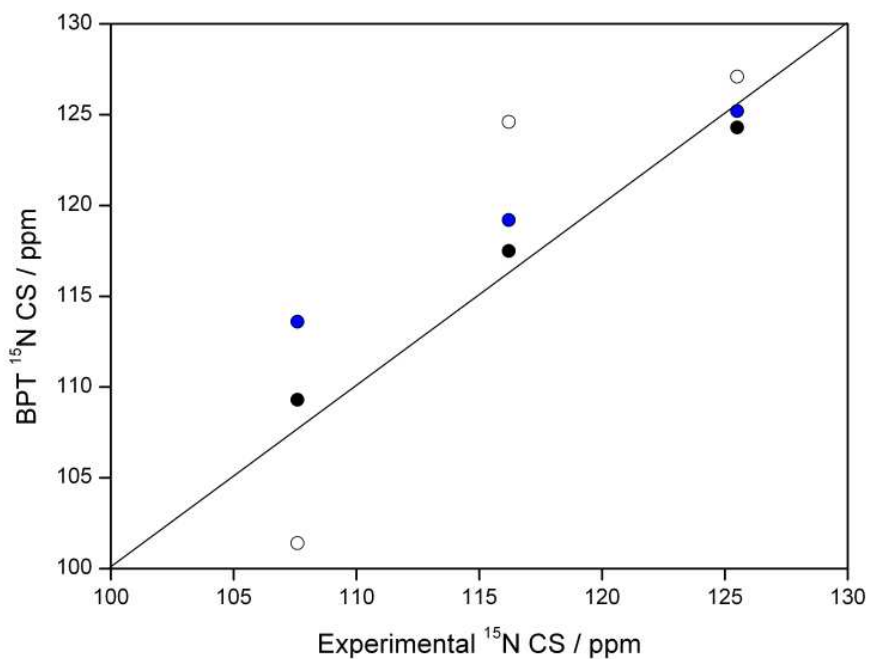


Figure 2.7: Comparison of the experimental and BPT calculated ^{15}N chemical shifts in f-MLF-OH (solid black circles – averaged CS values over 10 ps MD trajectory, solid blue circles – CS values of structure obtained via full structure search of the conformational space, open black circles – CS values of structure obtained via full structure search after hydrogen position optimization).

The calculations on crystalline tripeptides demonstrated the validity of the BPT parameterization for the ^{15}N chemical shift calculation. However, the realistic prediction of the ^{15}N chemical shifts must explicitly include long-range electrostatic interactions and molecular dynamics and it is not possible to estimate the impact of these effects a priori.

2.5 Unbiased ^{13}C parameterization for proteins

The original BPT parameterization for the ^{13}C chemical shift calculation introduced a bias towards carbohydrates. ^{13}C chemical shifts are less affected by electrostatic interactions and molecular motions than ^{15}N chemical shifts. Therefore the precise structure of f-MLF-OH is a good test system for the BPT $^{13}\text{C}_\alpha$ and $^{13}\text{C}_\beta$ chemical shift computation. Computed $^{13}\text{C}_\alpha$ and $^{13}\text{C}_\beta$ BPT chemical shifts in f-MLF-OH were in a very good agreement with the measured chemical shift values determined by

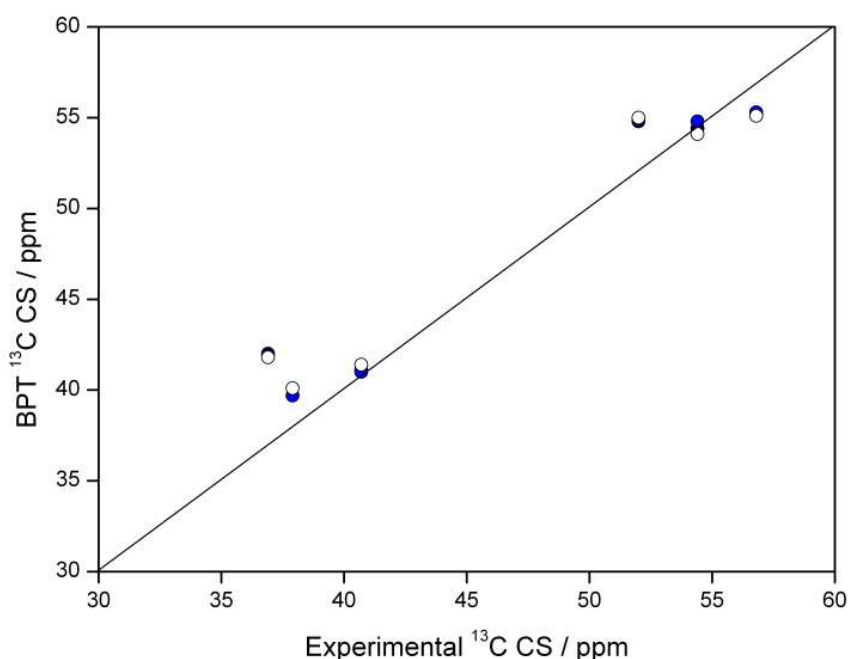


Figure 2.8: Comparison of the experimental and BPT calculated $^{13}\text{C}_\alpha$ and $^{13}\text{C}_\beta$ chemical shifts in f-MLF-OH (solid black circles – averaged CS values over 10 ps MD trajectory, solid blue circles – CS values of structure obtained via full structure search of the conformational space, open black circles – CS values of structure obtained via full structure search after hydrogen position optimization). Correlation coefficient R of 0.97 is achieved between the calculated and the experimental CS.

Rienstra et al.⁴⁵ (MAE of 1.5 ppm and 2.4 ppm respectively, cf. table 2.4 and figure 2.8). As expected no significant dependence of the ^{13}C chemical shifts on hydrogen atom positions could be observed.

Regardless of the successful ^{13}C chemical shift prediction in f-MLF-OH initial BPT ^{13}C calculations on the small globular protein ubiquitin (cf. section 2.6 for details) revealed systematic deviations of the calculated ^{13}C CS values from experimental values at the α -carbons of glycine, isoleucine, lysine, proline, serine, threonine and valine and the β -carbons of arginine, isoleucine, glutamine, glutamic acid and valine. To prove that this tendency is not ubiquitin-specific and to allow a system-independent correction of the deviations, $^{13}\text{C}_\alpha$ and $^{13}\text{C}_\beta$ CS were calculated in the models of the ideal α -helical polypeptides containing the corresponding amino acids ($\text{NH}_3^+(\text{Ala}_4\text{X})_4\text{Ala}_4\text{-COO}^-$ in the consensus α -helical geometry ($\varphi = -57.80^\circ$ and $\psi = -47.00^\circ$) with X being the corresponding amino acid; cf. figure 2.9). The computed chemical shifts were compared to the statistic mean ^{13}C chemical shift values of amino acids in α -helical conformations as defined by Zhang et al.⁴⁶ Mean deviations between the calculated chemical shifts in ideal α -helical polypeptides and the values of Zhang et al. exhibit no differences outside the error limits derived from calculated and experimental chemical shifts in ubiquitin (cf. table 2.5). The correlation coefficient between these two sets of deviations is 0.98. The deviations obviously originate from the bias in original BPT ^{13}C parameterization. For the ^{13}C chemical shift calculation of the corresponding amino acids the bond parameters $\delta_{\alpha\beta}^0$ given in table 2.1 have to be corrected by the mean observed deviations. Therefore in all future calculations the chemical shift tensor components of a non-polarized bond $\delta_{\alpha\beta}^0$ for the carbons of the corresponding amino acids were shifted by 4.8 ppm for $^{13}\text{C}_\alpha$ (16.2 ppm for proline $^{13}\text{C}_\alpha$) and -9.0 ppm for $^{13}\text{C}_\beta$.

If the corrected parameterization is used to compute the $^{13}\text{C}_\alpha$ and $^{13}\text{C}_\beta$ CS in α -helical polypeptides, a correlation coefficient of 0.98 and a standard deviation of 3.09 ppm between the calculated CS and the averaged values of Zhang et al. can be achieved (cf. figure 2.10).

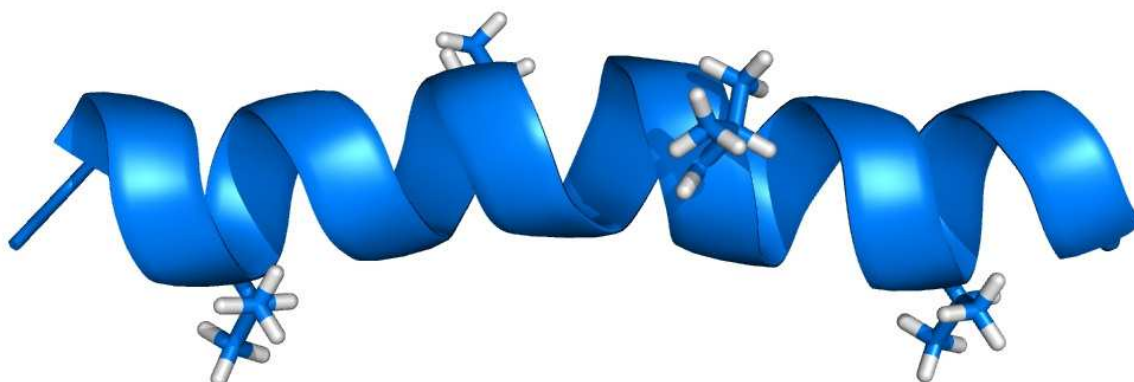


Figure 2.9: An example of the α -helical polypeptide used for correction of ^{13}C parameterization. $(\text{NH}_3^+(\text{Ala}_4\text{Val})_4\text{Ala}_4\text{-COO}^-)$ is shown in ribbon representation; the valine side chains are shown as “sticks”).

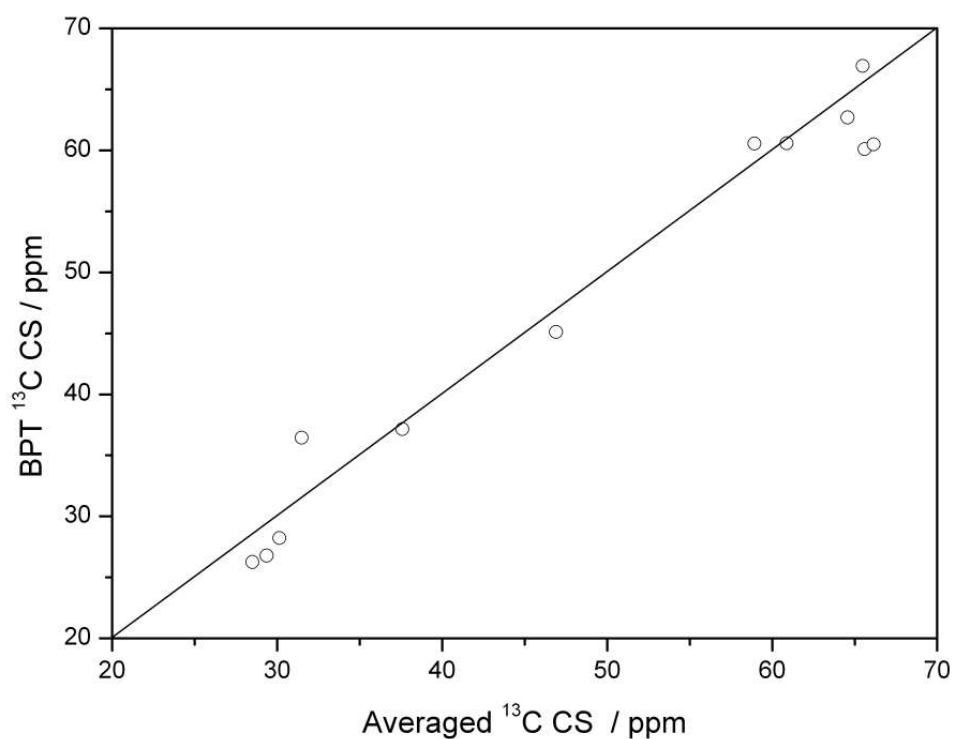


Figure 2.10: Correlation between the BPT calculated $^{13}\text{C}_\alpha$ and $^{13}\text{C}_\beta$ chemical shifts in peptides with consensus α -helical geometry and the averaged experimental $^{13}\text{C}_\alpha$ and $^{13}\text{C}_\beta$ chemical shifts in helical structures.

Table 2.5: Comparison of the BPT calculated $^{13}\text{C}_\alpha$ and $^{13}\text{C}_\beta$ chemical shifts in $\text{NH}_3^+(\text{Ala}_4\text{X})_4\text{Ala}_4\text{-COO}^-$ (X=corresponding amino acid) peptides with the consensus α -helical geometry with the averaged experimental $^{13}\text{C}_\alpha$ and $^{13}\text{C}_\beta$ chemical shifts in helical structures (all ^{13}C CS are referenced to DSS).

Amino acid	BPT CS (in ppm)	Averaged CS (in ppm)	ΔCS (in ppm)	ΔCS in Ubiquitin (in ppm)
$^{13}\text{C}_\alpha$ nucleus				
GLY	40.33±0.19	46.91±1.10	6.58±1.29	4.59±2.11
ILE	57.90±0.47	64.57±1.74	6.67±2.21	5.29±2.31
LYS	55.77±0.29	58.93±1.44	3.16±1.73	2.69±2.22
SER	55.78±0.11	60.88±1.61	5.10±1.72	4.58±2.86
THR	55.29±0.05	65.61±2.39	10.32±2.44	5.77±1.98
VAL	55.70±0.15	66.16±1.55	10.46±1.70	5.97±4.24
Reference $^{13}\text{C}_\alpha$				4.8 ppm
PRO	50.73±0.21	65.49±1.08	14.76±1.29	16.24±0.16
Reference $^{13}\text{C}_\alpha$ PRO				16.2 ppm
$^{13}\text{C}_\beta$ nucleus				
ARG	37.22±0.22	30.14±1.14	-7.08±1.36	-7.77±0.39
GLN	35.26±0.21	28.51±0.92	-6.75±1.13	-8.02±0.95
GLU	35.79±0.38	29.37±0.99	-6.42±1.37	-8.39±1.86
ILE	46.15±0.12	37.60±1.15	-8.55±1.27	-9.91±2.95
VAL	45.46±0.19	31.49±0.72	-13.97±0.91	-11.13±2.43
Reference $^{13}\text{C}_\beta$				-9.0 ppm

2.6 Application to ubiquitin

The small globular protein ubiquitin (cf. figure 2.11) contains 18 of 20 standard amino acid types. It is a popular system to evaluate the performance of computational methods. Direct ^{13}C chemical shift prediction on the DFT level of theory²⁴ and ^{13}C chemical shift prediction using the chemical shielding surfaces computed on the Hartree-Fock level of theory⁴⁷ were tested on ubiquitin. Therefore chemical shift prediction in ubiquitin allows to compare the results of BPT computations with the results obtained by other methods. $^{13}\text{C}_\alpha$ and $^{13}\text{C}_\beta$ chemical shifts for the liquid-state NMR⁴⁸ (PDB ID 1D3Z) and crystal⁴⁹ structures (PDB ID 1UBQ) of ubiquitin were calculated with bond polarization theory. Chemical shifts were computed for each

structure in the NMR structure ensemble of ubiquitin and ensemble averaged CS values were calculated for each residue. ^{13}C chemical shifts computed with bond polarization theory are in very good agreement with the experimental chemical shifts. The correlation coefficient between the calculated and the experimental $^{13}\text{C}_\alpha$ and $^{13}\text{C}_\beta$ chemical shifts is 0.87 and 0.95, respectively (cf. table 2.6 and figure 2.12). The calculated $^{13}\text{C}_\beta$ chemical shifts exhibit a larger mean absolute error than the calculated $^{13}\text{C}_\alpha$ chemical shifts (3.5 ppm vs. 1.9 ppm). This fact could be explained by amino acid side chain mobility in liquids so that $^{13}\text{C}_\beta$ chemical shifts cannot be sufficiently represented by a single side chain conformation (in less mobile f-MLF-OH crystal the $^{13}\text{C}_\beta$ CS MAE was only 2.4 ppm, cf. section 2.5).

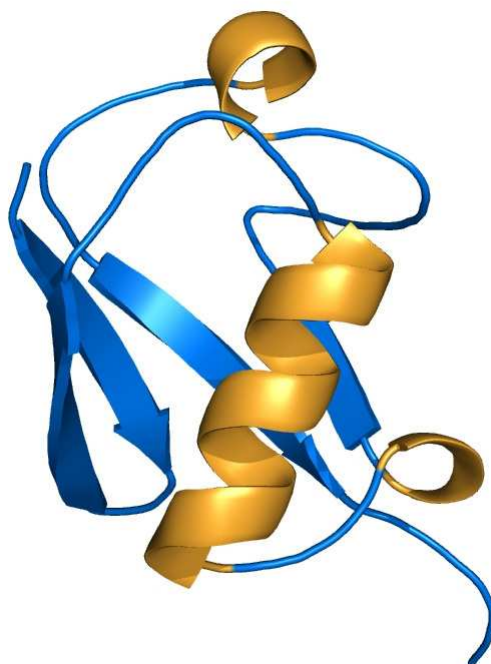


Figure 2.11: Structure of ubiquitin (PDB ID 1D3Z) in ribbon representation.

The performance of the computational methods for the ^{13}C chemical shift prediction can be compared in terms of the conformationally averaged $^{13}\text{C}_\alpha$ chemical shift root mean square deviation between calculated and experimental chemical shift values ($^{13}\text{C}_\alpha$ -RMSD) introduced by Harold A. Scheraga and coworkers²⁴:

$$^{13}\text{C}_\alpha - \text{RMSD} = \left[(1/N) \sum_{i=1}^N (\delta_{\text{exp},i} - (1/\Omega) \sum_{j=1}^{\Omega} \delta_{\text{calc},i,j})^2 \right]^{1/2} \quad (2.38)$$

with N – number of amino acids in protein sequence, Ω – number of protein structures in ensemble and δ - chemical shift of $^{13}\text{C}_\alpha$. The $^{13}\text{C}_\alpha$ -RMSD between the BPT calculated and the experimental chemical shifts in ubiquitin is 2.4 ppm for the NMR and 2.9 ppm for the X-ray structure. This is in excellent agreement with the $^{13}\text{C}_\alpha$ -RMSD of 2.5 ppm for the NMR and 2.7 ppm for the X-ray ubiquitin structure obtained by Vila et al. in DFT calculations²⁴. However, Vila et al. conducted calculations on the ubiquitin structure decomposed into Ac-GXG-Me tripeptides (for each amino acid X in the protein sequence in experimental conformation) in order to reduce the computational burden while BPT calculations were conducted on the full-size protein structure. Furthermore the results of the BPT chemical shift prediction are better than the results of the chemical shift prediction using the computed $^{13}\text{C}_\alpha$ shielding surfaces obtained by Sun et al.⁴⁷ (2.7 ppm for the NMR and 3.6 ppm for the X-ray structure; $^{13}\text{C}_\alpha$ -RMSD was calculated by Vila et al.²⁴ using the data of Sun et al.).

Table 2.6: Correlation between the BPT calculated and the experimental chemical shifts in ubiquitin before and after chemical shift driven structure optimization (SD and MAE are given in ppm).

Chemical shift calculation in ubiquitin													
Nucleus	1D3Z			optimized 1D3Z			1UBQ			optimized 1UBQ			
	R	SD	MAE	R	SD	MAE	R	SD	MAE	R	SD	MAE	
$^{13}\text{C}_\alpha$	0.87	2.4	1.9	0.98	1.2	0.9	0.80	2.9	2.3	0.98	1.2	0.9	
$^{13}\text{C}_\beta$	0.95	4.2	3.5	0.97	3.1	2.5	0.95	4.3	3.4	0.97	3.0	2.4	
^{15}N	0.47	13.7	11.7	0.93	3.3	2.8	0.05	25.9	20.9	0.90	3.5	2.8	

BPT calculated ^{15}N chemical shifts of the ubiquitin structure do not show a good correlation with the observed chemical shift values (cf. figure 2.13). Experimental protein chemical shifts are average values over a variety of structural conformations. Protein structure ensembles usually provided by NMR spectroscopy do not necessary represent all coexisting protein conformations. Protein X-ray crystallography provides only one molecular conformation and it does not provide exact hydrogen positions (hydrogen atom positions can have a significant influence on the ^{15}N CS; cf. section 2.4). Even so quantum chemical ^{13}C chemical shift prediction in proteins with experimental accuracy geometry is possible (^{13}C chemical

shifts are less affected by electrostatic interactions and molecular dynamics; cf. section 2.5). In marked contrast to the ^{13}C chemical shift prediction, ^{15}N chemical shift prediction remains a challenging task as ^{15}N chemical shifts are very sensitive to molecular environment. The molecular geometry in the proximity of the ^{15}N nucleus has to be optimized on a quantum chemical level similar to the quantum chemical level required for the ^{15}N chemical shift computation. Structural differences that appear insignificant for structural biology (e.g. tiny imperfections of bond length) can lead to significant chemical shift errors in *ab initio* and semi-empirical calculations.

The analysis of protein membrane orientation and dynamics requires the prediction of full ^{15}N chemical shift tensors while usually only the isotropic ^{15}N chemical shifts can be determined experimentally or are already known from previous studies. BPT investigations on the ^{13}C chemical shift tensors of cellulose revealed that chemical shift driven geometry optimization with isotropic chemical shifts as a target function yields the cellulose structure perfectly matching the experimentally determined chemical shift tensors in quantum chemical calculations⁵⁰. The same strategy can be applied to the protein ^{15}N chemical shifts. This approach, combining protein structure refinement with the isotropic chemical shifts with the consecutive ^{15}N chemical shift tensor calculation, enables investigations described in the following chapters.

2.7 Chemical shift driven protein structure refinement

Bond polarization theory is the only quantum chemical method with analytically derived chemical shift gradients⁵¹. That is why the influence of the experimental chemical shifts can be included into the COSMOS-NMR hybrid QM/MM force field³² as an additional energy term enabling chemical shift driven protein structure refinement with molecular mechanics. Chemical shift gradients provide pseudo-forces required for the chemical shift driven geometry optimization:

$$F_{\alpha}^{\delta_i} = s_i^{\Delta\delta} k_i^{\delta} (\delta_i^{\text{theo}} - \delta_i^{\text{exp}}) \frac{\partial \delta_i^{\text{theo}}}{\partial x_{\alpha}} \quad (2.39)$$

with $F_{\alpha}^{\delta_i}$ – CS pseudo-force acting on the nucleus i , δ_i^{exp} – experimental CS of the nucleus i , δ_i^{theo} – calculated CS of the nucleus i , x_{α} – atomic coordinates of the

nucleus i , α is the index of the force or coordinate vector component, k_i^δ – force constant adjusting the magnitude of the chemical shift pseudo-force and $s_i^{\Delta\delta}$ – scaling function controlling the width of the chemical shift potential.

Experimentally derived protein structures usually do not exactly match experimental chemical shifts leading to the very large chemical shift pseudo-forces in the beginning of the geometry optimization. If linearly rising chemical shift pseudo-forces are applied the geometry optimization either will not converge or it will deform the whole structure. Therefore the pseudo-forces are multiplied with scaling factors that depend on the difference between the theoretical and the experimental CS values:

$$s_i^{\Delta\delta} = k_0^\delta \frac{f - f^{-1}}{f + f^{-1}} \quad (2.40)$$

$$f = e^{\frac{(\delta_i^{\text{theo}} - \delta_i^{\text{exp}})}{\Delta\delta}}$$

If the deviations between the calculated and the experimental CS values are small the scaled pseudo-forces behave similar to the harmonic pseudo-forces. For large deviations between the calculated and the experimental chemical shifts the pseudo-forces approach constant values. While the function f used for the definition of the pseudo-force scaling function $s_i^{\Delta\delta}$ (cf. equation 2.40) is individual for each nucleus the force constant k_0^δ allows to step up or diminish all pseudo-forces at once. In addition, the nucleus-specific force constants k_i^δ are introduced in order to provide the correct range of magnitudes and units for the chemical shift pseudo-forces. These constants can be calculated from the BPT parameters for the charge and chemical shift of the atoms⁵¹:

$$k_i^\delta = \sum_{j=1}^N \frac{|A_j^q|}{en_j^2 (A_j^\delta)^2}, \quad (2.41)$$

where N is the number of bonds of the atom i with the bond occupation number n_j and e is the elementary charge.

In the beginning of the chemical shift driven geometry optimization the CS pseudo-energy should be on the order of the magnitude of the non-bonding energy terms. Therefore a very thoroughful choice of the potential width control parameter

$\Delta\delta$ (cf. equation 2.40) and force constant k_0^δ is required or the geometry optimization will not converge. To define the optimal settings for these parameters, a series of chemical shift driven geometry optimizations was performed on the first conformation of the ubiquitin NMR structure ensemble (PDB ID 1D3Z). Separate parameter settings were applied for the ^{15}N and ^{13}C nuclei to account for their different chemical shift sensitivity upon molecular geometry errors. The initial setting for the parameter $\Delta\delta$ reflected the mean absolute error in previous chemical shift calculations. The parameters were varied in each geometry optimization run in order to minimize the deviation between the calculated chemical shifts and the experimental chemical shifts. The results of these calculations are given in table 2.7. Optimal settings for the potential width control parameter $\Delta\delta$ and the force constant k_0^δ of the chemical shift pseudo-force are 30 and 15 ppm for ^{15}N and 2.5 and 2 ppm for ^{13}C , respectively. Deviations from these values tend to increase the differences between the calculated and the experimental chemical shifts after CS driven geometry optimization. To demonstrate the transferability of this protocol the first conformations of the seven randomly chosen protein structure ensembles from the

Table 2.7: Settings applied in a series of the CS driven geometry optimizations of ubiquitin and the corresponding correlation coefficients (R) and standard deviations (SD) of the BPT calculated and experimental ^{15}N and $^{13}\text{C}_\alpha$ CS achieved after each CS driven geometry optimization (SD and $\Delta\delta$ are given in ppm).

^{15}N		$^{13}\text{C}_\alpha$		^{15}N		$^{13}\text{C}_\alpha$		^{15}N		$^{13}\text{C}_\alpha$		^{15}N		$^{13}\text{C}_\alpha$	
k_0^δ	$\Delta\delta$	k_0^δ	$\Delta\delta$	R	SD	R	SD	k_0^δ	$\Delta\delta$	k_0^δ	$\Delta\delta$	R	SD	R	SD
50	20	0.15	2.50	0.94	2.38	0.63	4.04	30	20	0.15	2.50	0.91	2.64	0.82	2.63
20	15	5.00	2.50	0.29	53.7	0.66	3.76	30	15	2.25	2.25	0.91	2.56	0.86	2.35
20	15	1.50	2.50	0.85	3.43	0.81	2.78	30	15	2.50	2.25	0.91	2.50	0.87	2.24
20	15	2.00	2.50	0.83	3.51	0.85	2.48	30	15	2.75	2.25	0.90	2.56	0.89	2.18
20	15	1.25	2.50	0.85	3.38	0.80	2.81	30	15	3.00	2.25	0.90	2.64	0.90	2.03
25	15	2.00	2.50	0.89	2.83	0.86	2.44	30	15	2.00	2.00	0.91	2.57	0.86	2.37
30	15	2.00	2.50	0.90	2.69	0.87	2.29	30	15	2.25	2.00	0.91	2.52	0.86	2.37
25	15	1.25	2.50	0.90	2.74	0.79	2.90	30	15	2.50	2.00	0.89	2.78	0.93	1.76
30	15	1.25	2.50	0.91	2.50	0.78	2.90	30	15	2.75	2.00	0.90	2.61	0.87	2.23
25	20	0.15	2.50	0.90	2.70	0.81	2.78	30	15	3.00	2.00	0.89	2.83	0.91	2.01

protein data bank (PDB ID 1UEM, 2JUO, 2K14, 2K19, 2K53, 2K75 and 1Q02) were subjected to the chemical shift driven protein geometry optimization. Correlation coefficient and root mean square deviation for all seven structures after the CS driven geometry optimization is 0.86 and 2.72 ppm for the backbone ^{15}N nuclei and 0.85 and 2.57 ppm for the $^{13}\text{C}_\alpha$ nuclei*. These values are similar to those obtained on ubiquitin.

NMR and X-ray structures of ubiquitin were subjected to the chemical shift driven geometry optimization. Each structure in the NMR structure ensemble of ubiquitin was refined and ensemble averaged CS values after CS driven geometry optimization were calculated for each residue (for one conformer of the NMR ensemble no local CS minimum could be obtained and therefore it was excluded from the evaluation). The agreement between the calculated and the experimental $^{13}\text{C}_\alpha$ and $^{13}\text{C}_\beta$ chemical shifts was slightly improved after the CS driven geometry optimization (cf. table 2.6 and figure 2.12). The agreement between the calculated and the experimental protein backbone ^{15}N chemical shifts improved significantly (correlation coefficient of 0.93 and standard deviation of 3.3 ppm, cf. table 2.6 and figure 2.13). These results for the ^{15}N CS are comparable to the results obtained on peptide crystals. CS driven geometry optimization can find only the local minimum of the protein structure with respect to the chemical shift target function. No significant structure perturbations could be observed after the CS driven geometry optimization. Backbone atoms RMSD of the ubiquitin structures before and after optimization did not exceed 0.04 Å. This fact indicates that in the BPT framework chemical shift hypersurface possesses a variety of local minima and protein conformations corresponding to these local minima are very similar to the experimentally derived structures.

*First geometry optimization results were evaluated with OriginPro 7 program (OriginLab Corporation, Northampton, MA, 2002) using a linear regression procedure. SD values given in table 2.7 and on the page 44 refer to the root mean square errors obtained after linear regression of the experimental and the calculated CS.

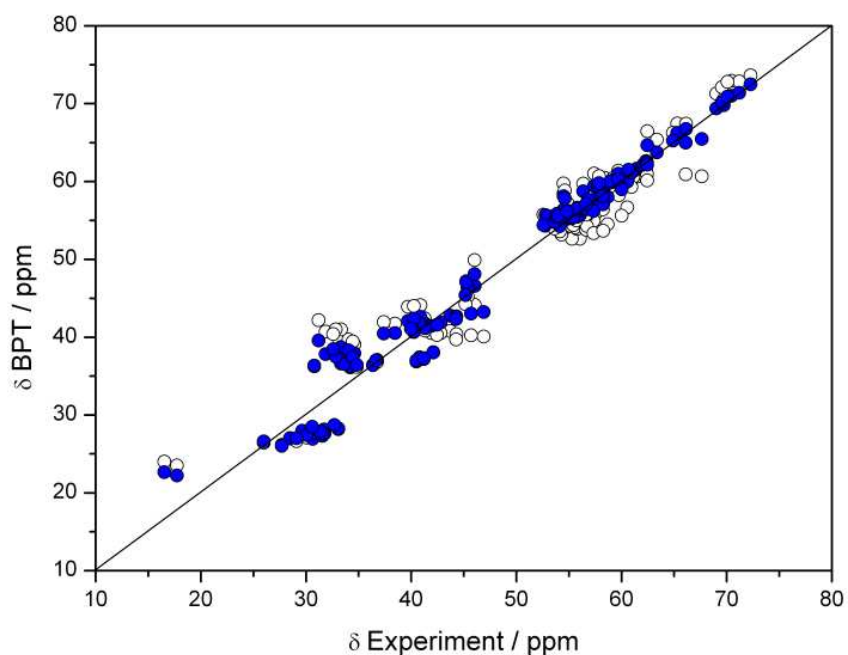


Figure 2.12: Correlation between experimental and BPT calculated $^{13}\text{C}_\alpha$ and $^{13}\text{C}_\beta$ chemical shifts of ubiquitin before (open black circles) and after (blue solid circles) chemical shift driven geometry optimization.

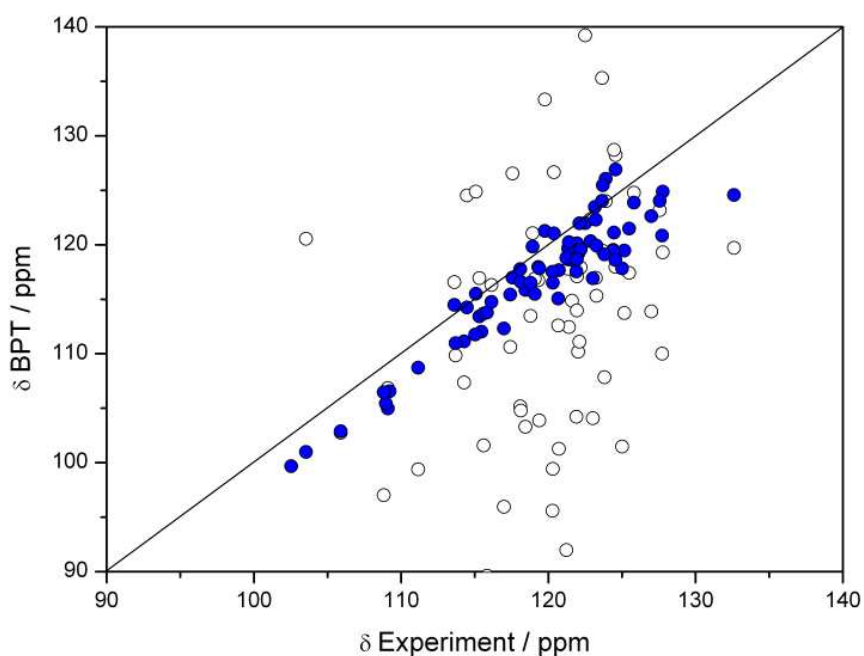


Figure 2.13: Correlation between experimental and BPT calculated ^{15}N chemical shifts of ubiquitin before (open black circles) and after (blue solid circles) chemical shift driven geometry optimization.

2.8 Computational efficiency: BPT vs. DFT

One of the major challenges of the *ab initio* chemical shift computation in biological systems is the high computational cost of such calculations. Bond polarization theory is capable to predict protein ^{13}C chemical shifts with the quality comparable to the performance of the density functional theory. Therefore it is interesting to compare the computational cost of both methods. DFT and BPT chemical shift calculations with the corresponding CPU time measurements were performed on several biological systems described in previous sections (cf. table 2.8). A desktop computer with 2 GHz CPU and 3 GB RAM was applied for all calculations. DFT calculations were carried out on the B3LYP/6-311++G(d,p) level of theory with the Gaussian 03 program⁵². The total CPU times were extracted from the Gaussian 03 output files. BPT calculations were carried out with the COSMOS PRO program⁵³. Times required for the BPT chemical shift and charge computation were determined with the help of the COSMOS molecular dynamics log file. Times of interest were calculated as the mean time differences over 10 MD steps in MD simulations with and without charge and chemical shift calculation. DFT CS calculations on the full-size ubiquitin and tripeptide unit cell under periodic boundary conditions are too demanding for the desktop computer and therefore the corresponding computation times are not given in table 2.8. Vila et al. reported DFT $^{13}\text{C}_\alpha$ CS computation time of 6686.93 s per residue for ubiquitin decomposed into tripeptides on the B3LYP/6-311+G(2d,p) level of theory⁵⁴. Vila et al. determined the computation time on a Beowulf-type cluster with 64 1.15 GHz CPUs and 256GB of shared RAM. The comparison of the computation times given in table 2.8 demonstrates that the computational cost of the BPT calculations is significantly smaller than the computational cost of the DFT calculations.

Waddell et al.⁴⁴ evaluated the basis sets and the reference value for the ^{15}N chemical shift calculation in peptides on the DFT B3LYP level of theory. According to Waddell et al. optimal results can be achieved with the 6-311++G(d,p) basis set and the reference value of 244.6 ppm (cf. equation 1.9). Vila et al.⁵⁴ evaluated the basis sets and the reference value for the $^{13}\text{C}_\alpha$ chemical shift calculation in peptides on the DFT B3LYP level of theory. According to Vila et al.⁵⁴ optimal results can be achieved with

Table 2.8: Comparison of the COSMOS BPT and DFT chemical shift computation times (n.a. – calculation is not available).

CPU Time		
Molecule	BPT	DFT B3LYP 6-311++G(d,p)
f-MLF-OH Monomer	0.02 s	22 h 8 min 25.0 s
Ala-Gly-Gly Monomer	0.01 s	2 h 36 min 38.0 s
Ala-Gly-Gly Trimer	0.04 s	2 d 3 h 17 min 5.0 s
Ala-Gly-Gly Unit cell	2.18 s	n. a.
Full-size ubiquitin	20.46 s	n. a.

the 6-311+G(2d,p) basis set which is similar to the basis set proposed by Waddell et al. Vila et al. noticed that the reference value for the $^{13}\text{C}_\alpha$ chemical shift calculation has to be defined separately for each basis set. For comparison of the computation time DFT CS calculations were conducted on f-MLF-OH. The reference value for conversion of the ^{13}C nuclear shieldings into chemical shifts was determined as recommended by Vila et al.⁵⁴ TMS model in T_d symmetry conformation was subjected to the geometry optimization with extremely tight convergence criteria^{55,56} followed by the nuclear shielding calculation (cf. table 2.9). In marked contrast to the BPT calculated chemical shifts, DFT calculated chemical shifts could not match the experimental values (cf. table 2.4). The results of the DFT calculation could be improved by adjusting the reference values. However, this is outside of the scope of this thesis. The fact that a priori defined reference values do not always lead to correct chemical shift prediction (even if the same reference values were successfully applied for compounds similar to the compound under study) is the problem often associated with DFT CS calculations. No such problems were encountered in BPT CS calculations.

Table 2.9: Experimental^{57,58} and calculated (DFT B3LYP with 6-311++G(d,p) basis set) geometry parameters and chemical shieldings of TMS. Calculated nuclear shielding of TMS is used for conversion of the calculated ^{13}C nuclear shieldings into chemical shifts.

Parameters of the calculated and experimental TMS geometry and chemical shielding				
Molecule	Si–C in Å	C–H in Å	Si–C–H in degrees	σ in ppm
Experimental	1.877±0.004	1.110±0.003	111.0±0.2	188.1
DFT B3LYP	1.891	1.095	111.3	184.2

2.9 Chemical shift and protein structure validation

Quantum chemical CS calculation allows to validate the quality of a NMR derived protein structure ensemble. Single point chemical shift calculation revealed that the NMR structure ensemble of ubiquitin ($^{13}\text{C}_\alpha$ -RMSD of 2.4 ppm; cf. section 2.6) is a better representation of the experimental NMR chemical shifts than the ubiquitin X-ray structure ($^{13}\text{C}_\alpha$ -RMSD of 2.9 ppm). The difference in the $^{13}\text{C}_\alpha$ -RMSD can be observed despite the structural similarity between the X-ray and the NMR structures (the RMSD of backbone C_α atoms between the first structure of the ubiquitin NMR ensemble and the X-ray structure is 0.35 Å and all heavy atoms RMSD is 0.38 Å).

^{15}N chemical shifts are more sensitive to molecular structure than $^{13}\text{C}_\alpha$ chemical shifts. The correlation between experimental and calculated ^{15}N CS of ubiquitin also confirms that the NMR structure is a much better representation of the experimental ^{15}N CS than the X-ray structure (correlation coefficient R of 0.47 for the NMR structure and R of 0.05 for the X-ray structure; cf. table 2.6). This trend vanishes after the CS driven structure optimization ($^{13}\text{C}_\alpha$ -RMSD of 1.2 ppm and ^{15}N CS correlation coefficient R of 0.93 for the NMR ensemble vs. 1.2 ppm and 0.90 for the X-ray structure) demonstrating that structural differences between the NMR and the X-ray structure of ubiquitin are insignificant.

Vila et al. investigated the quality of four protein structure ensembles using DFT $^{13}\text{C}_\alpha$ chemical shift computations²⁴. The set of Vila et al. (cf. table 2.10) included ubiquitin⁴⁸ (PDB ID 1D3Z), histidine-containing phosphocarrier protein⁵⁹ (PDB ID 1HDN), zinc-binding domain of tristetraprolin protein⁶⁰ (PDB ID 1M90) and calcium-binding protein parvalbumin⁶¹ (PDB ID 1TTX). Vila et al.²⁴ attempted to rank the protein structure ensembles according to their $^{13}\text{C}_\alpha$ -RMSD per residue (cf. table 2.11). This approach can be disputed since the $^{13}\text{C}_\alpha$ -RMSD already refers to the CS deviation per residue and calculation of the $^{13}\text{C}_\alpha$ -RMSD per residue will introduce a bias towards larger proteins. The ranking with respect to the $^{13}\text{C}_\alpha$ -RMSD per residue does not correlate with the ranking with respect to the $^{13}\text{C}_\alpha$ -RMSD. As expected the best results in terms of the $^{13}\text{C}_\alpha$ -RMSD per residue are achieved for the largest

protein parvalbumin (PDB ID 1TTX) while the best results in terms of the $^{13}\text{C}_\alpha$ -RMSD are achieved for ubiquitin (PDB ID 1D3Z; cf. table 2.11). It is noteworthy that Vila et al. did not compare the ranking obtained by DFT $^{13}\text{C}_\alpha$ CS computation to the resolution of the protein structure ensembles or to their stereochemical quality. The results of Vila et al. exhibit no correlation with these parameters (cf. table 2.11).

Table 2.10: Experimental conditions and BMRB data available for proteins analyzed by Vila et al.

PDB ID	No. of models in ensemble	pH	T in K	Reference	No. of observed $^{13}\text{C}_\alpha$ CS	Accession number	Missing data
1D3Z	10	6.6	298	TSP	76	15410	none
1HDN	30	6.5	303	TSP	85	2371	^{15}N CS
1M9O	23	5.8	293	TSP	38	5525	$^{13}\text{C}_\beta$ CS
1TTX	20	6.5	298	DSS	109	6705	none

Table 2.11: Structure validation of proteins analyzed by Vila et al. Stereochemical quality is estimated in % of the dihedral angles within the regions of the Ramachandran plot (MF – most favoured region; AA – additionally allowed region; GA – generously allowed region; DA – disallowed region). DFT and BPT $^{13}\text{C}_\alpha$ -RMSD are given in ppm, R denotes the $^{13}\text{C}_\alpha$ CS correlation coefficient.

PDB ID	Resolution in Å	PROCHECK stereochemical quality				DFT $^{13}\text{C}_\alpha$ -RMSD	DFT $^{13}\text{C}_\alpha$ -RMSD per residue	BPT $^{13}\text{C}_\alpha$ -RMSD	BPT R
		MF	AA	GA	DA				
1D3Z	0.12	96.7	3.3	0.0	0.0	2.5	0.032	2.4	0.87
1HDN	0.88	83.3	15.6	0.9	0.1	6.9	0.081	3.6	0.78
1M9O	0.51	29.2	53.3	13.7	3.8	3.5	0.100	3.2	0.76
1TTX	0.65	80.7	16.5	1.8	1.0	2.7	0.025	5.1	0.76

The NMR ensemble of ubiquitin (PDB ID 1D3Z) exhibits the best backbone resolution (ensemble backbone C_α RMSD of 0.12 Å) and the best stereochemical quality as estimated with PROCHECK program⁶² (ca. 97% of all dihedral angles lie within the most favoured regions of the Ramachandran plot and no dihedral angles lie within the disallowed regions). The resolution of the remaining three protein structure ensembles is similar (ensemble backbone C_α RMSD between 0.51 Å and 0.88 Å).

Histidine-containing phosphocarrier protein (PDB ID 1HDN) and parvalbumin (PDB ID 1TTX) also possess comparable stereochemical quality (ca. 80% of all dihedral angles lie within the most favoured regions of the Ramachandran plot). Stereochemical quality of the zinc-binding domain of tristetraprolin protein (PDB ID 1M90) is slightly poorer (ca. 29% of all dihedral angles lie within the most favored regions and ca. 4% lie within the disallowed regions of the Ramachandran plot). However, the tristetraprolin zinc-binding domain exhibits the best ensemble resolution of the remaining three proteins (0.51 Å) and 4% of disallowed dihedral angles correspond to one or two residues with bad geometry per structure in an ensemble and not to completely erroneous structures. Therefore a reliable estimator of the protein structure quality should rank the ubiquitin ensemble as the best structure ensemble and yield similar results for the remaining three protein ensembles.

Chemical shifts of proteins analyzed by Vila et al.²⁴ were calculated with bond polarization theory. The analysis focussed on $^{13}\text{C}_\alpha$ chemical shifts. The results of the $^{13}\text{C}_\alpha$ CS calculation provide the most reliable criteria for the local backbone structural quality since $^{13}\text{C}_\beta$ chemical shifts are too dependent on side chain dynamics and ^{15}N chemical shifts are even stronger influenced by electrostatic long-range effects. A BPT $^{13}\text{C}_\alpha$ -RMSD of 2.4 ppm was computed for ubiquitin and BPT $^{13}\text{C}_\alpha$ -RMSDs of 3.2 and of 3.6 ppm were computed for the tristetraprolin zinc-binding domain and the histidine-containing phosphocarrier protein, respectively (cf. table 2.11). The largest BPT $^{13}\text{C}_\alpha$ -RMSD of 5.1 ppm was observed in the parvalbumin ensemble. The results obtained for ubiquitin, tristetraprolin zinc-binding domain and histidine-containing phosphocarrier protein are in good agreement with the ensemble resolution and stereochemical quality. Large BPT $^{13}\text{C}_\alpha$ -RMSD for the parvalbumin ensemble cannot be explained by its stereochemical quality.

It is noteworthy that large $^{13}\text{C}_\alpha$ -RMSD in parvalbumin was not observed in DFT calculations of Vila et al. However, the DFT calculations of Vila et al.²⁴ provided a very high $^{13}\text{C}_\alpha$ -RMSD value of 6.9 ppm for the histidine-containing phosphocarrier protein NMR structure ensemble (PDB ID 1HDN). This result is in marked

contradiction with ensemble resolution and stereochemistry data. To address this issue, Vila et al. computed the DFT $^{13}\text{C}_\alpha$ -RMSD for the corresponding X-ray structure⁶³ (PDB ID 1CM2) and obtained the value of 2.7 ppm. The BPT $^{13}\text{C}_\alpha$ -RMSD of 3.4 ppm was computed for the same X-ray structure. The X-ray structure (PDB ID 1CM2) is comparable to the NMR structure (PDB ID 1HDN). The backbone C_α RMSD between the first structure of the NMR ensemble and the X-ray structure is 0.82 Å. Therefore these calculations and the BPT CS calculation on the NMR structure imply that the large DFT $^{13}\text{C}_\alpha$ -RMSD computed for the NMR structure of the histidine-containing phosphocarrier protein does not indicate poor structural quality but probably reflects the shortcomings of the DFT calculations.

The calculations on the protein set of Vila et al.²⁴ demonstrate that BPT $^{13}\text{C}_\alpha$ CS computation can be applied for protein structure validation but great care must be taken if the $^{13}\text{C}_\alpha$ -RMSD is used as the only estimator of protein structural quality. A better estimator of structural quality is provided by the correlation coefficient between experimental and BPT calculated $^{13}\text{C}_\alpha$ chemical shifts (cf. table 2.11). A correlation coefficient of 0.87 was obtained for ubiquitin and correlation coefficients of 0.76 and 0.78 were obtained for the remaining three proteins. This ranking is in agreement with ensemble resolution and stereochemical quality.

An alternative approach to protein structure validation using the BPT chemical shift computation relies on chemical shift driven structure refinement. This type of geometry optimization can only find the closest local minimum with respect to the target function. Therefore CS driven geometry optimization of a valid NMR structure will in most cases improve the agreement with the experimental chemical shifts without perturbing the whole structure. An optimization attempt (with a small initial step width) of an invalid structure is more likely to result in disagreement with experimental CS than in significant alterations of the structure. This general property of the geometry optimization procedure should allow to involve the ^{15}N chemical shifts (which usually cannot be reliably predicted using experimental protein geometry) into protein structure validation. The set of protein ensembles analyzed by Vila et al. therefore was subjected to the chemical shift driven geometry optimization.

As a result, all chemical shifts for the four structures could be matched with correlation factors $R \geq 0.93$ between calculated and experimental chemical shifts (cf. table 2.12). All four protein structures under study are of high-quality and therefore no significant differences in the correlation coefficients can be observed among different structure ensembles after geometry optimization.

It becomes obvious that for well-defined structures protein structure validation using CS driven structure refinement can only confirm the high-quality of the data and no further ranking of protein structures can be defined. However, this approach might be applied in order to single out a set of valid structures from a large set of poorly defined conformers (e.g. several thousand of structures calculated from sparse NMR data). The protein set of Vila et al. does not include lower-quality structures and therefore no statement can be made if the approach based on CS driven geometry optimization is really able to provide a meaningful structure ranking or at least to exclude the structures of very low quality. This subject deserves further investigation but it is outside of the scope of this thesis.

Table 2.12: Correlation between BPT-calculated and experimental chemical shifts in the NMR structure ensembles of ubiquitin (1D3Z), histidine-containing phosphocarrier protein (1HDN), tristetraprolin zinc-binding domain (1M9O) and parvalbumin (1TTX) before and after chemical shift driven structure optimization (SD and MAE are given in ppm).

PDB ID	1D3Z			1HDN			1M9O			1TTX		
Nucleus	R	SD	MAE	R	SD	MAE	R	SD	MAE	R	SD	MAE
Single point chemical shift calculation												
$^{13}\text{C}_\alpha$	0.87	2.4	1.9	0.78	3.6	2.9	0.76	3.2	2.4	0.76	5.1	4.5
$^{13}\text{C}_\beta$	0.95	4.2	3.5	n.a.	n.a.	n.a.	0.96	5.4	4.3	0.96	3.7	3.2
^{15}N	0.47	13.7	11.7	0.56	28.3	26.5	n.a.	n.a.	n.a.	0.38	12.3	10.5
Chemical shift driven geometry optimization												
$^{13}\text{C}_\alpha$	0.98	1.2	0.9	0.98	1.0	0.8	0.99	1.0	0.7	0.97	1.8	1.5
$^{13}\text{C}_\beta$	0.97	3.1	2.5	n.a.	n.a.	n.a.	0.97	4.3	3.4	0.98	2.8	2.1
^{15}N	0.93	3.3	2.8	0.94	2.1	1.8	n.a.	n.a.	n.a.	0.93	2.4	2.1

3. Dynamics and orientation of membrane peptides

3.1 Biological background

Biological membranes form the interface between cells and their environment. They consist of two layers of lipid molecules oriented according to their hydrophobicity (cf. figure 3.1). Polar head groups of membrane lipids face the aqueous surrounding and the non-polar lipid chains form the hydrophobic membrane core. Membrane-active peptides participate in a variety of biological processes associated with membrane remodelling or permeabilization. Antimicrobial peptides constitute a class of membrane-active peptides playing an important role in innate immunity⁶⁴. Antimicrobial peptides are capable of binding to the membrane bilayer surface of bacteria and fungi and killing them either by disrupting the cell membrane or by crossing the membrane and targeting the cell organelles or interfering with metabolic pathways. The disruption of lipid bilayers can be achieved in a non-structured manner by high local concentration of antimicrobial peptides on the membrane surface⁶⁵ (the “carpet” model). Another mechanism of membrane disruption by antimicrobial peptides is based on pore formation across the membrane⁶⁶. The pores can consist of the antimicrobial peptides alone (“barrel-stave” model) or combined with charged membrane lipid head groups (toroidal “wormhole” model). The action of the pore-forming antimicrobial peptides can often be explained in terms of the Shai-Matzusaki-Huang model⁶⁷. At low peptide-to-lipid ratio pore-forming antimicrobial peptides align on the membrane surface perpendicular to the membrane normal in the surface-bound “S-state”. With the increase of the peptide-to-lipid ratio antimicrobial peptides undergo oligomerization and assume a fully upright transmembrane alignment in the immersed “I-state”. “I-state” is associated with pore formation that is lethal to a cell. In addition to the states proposed by the Shai-Matzusaki-Huang model an obliquely immersed tilted “T-state” was also observed⁶⁸.

3.2 Macroscopically aligned samples

The example of antimicrobial peptides illustrates that the knowledge of membrane orientation of the membrane-active peptides is crucial for the understanding of their

mechanism of action. Solid-state NMR allows the study of peptide membrane orientation using macroscopically aligned samples⁶⁹. Macroscopically aligned samples consist of a stack of glass plates oriented in the external magnetic field (cf. figure 3.1). Lipids dispersed between the glass plates spontaneously form bilayers mimicking biological membranes. The bilayer surface is aligned parallel to the plate surface and up to several thousand bilayers can be accommodated between two glass plates. The peptide under study can be situated and aligned in the lipid environment. Peptide alignment in lipid bilayers fixes their orientation in the external magnetic field. Therefore no powder spectrum is observed and peptide orientation in the bilayer can be deduced from the observed peptide NMR signals using the orientational dependence of NMR interactions (cf. sections 1.2 and 1.4).

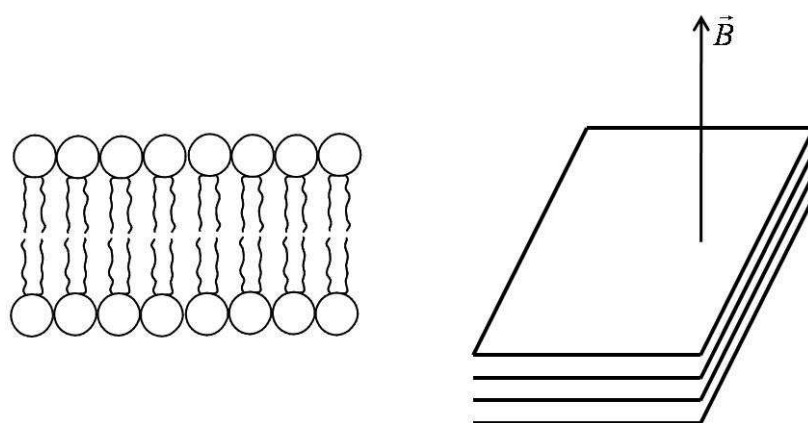


Figure 3.1: Schematic cross sectional profile of a lipid bilayer (left) and macroscopically aligned sample for solid-state NMR spectroscopy on lipid bilayers (right). Circles and lines denote lipid head groups and chains. Membrane normal is aligned parallel to the external magnetic field.

3.3 PISEMA spectroscopy

PISEMA (polarization inversion spin exchange at magic angle) spectroscopy is a solid-state NMR method for study of protein and peptide orientation in lipid bilayers⁹. Resonances observed in PISEMA spectra correlate the observed components of the backbone amide ^{15}N chemical shift tensor and the corresponding ^{15}N - ^1H dipolar coupling. NMR interactions observed in macroscopically oriented samples depend on the sample orientation in the external magnetic field, on the time-averaged molecular orientation in lipid bilayers and also on the orientation of the interaction tensor in the

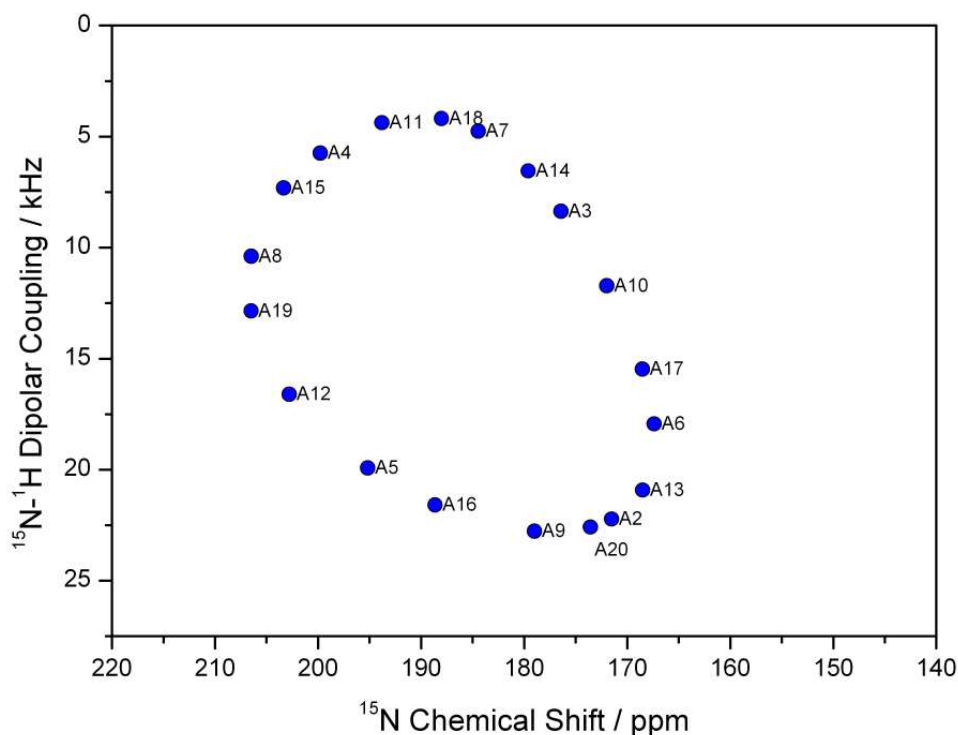


Figure 3.2: Calculated PISEMA spectrum of a 30° tilted polyalanine α -helix ($\text{NH}_3^+\text{-Ala}_{20}\text{-COO}^-$ with $\varphi = -57.80^\circ$ and $\psi = -47.00^\circ$) exhibits a PISA “wheel” signal pattern. The spectrum was calculated with the COSMOS PRO program using fixed ^{15}N chemical shift tensors.

molecular frame of reference. Peptides with α -helical secondary structure exhibit a repeating geometry pattern. The periodicity of this geometry pattern is given by the backbone winding of 3.6 residues in an α -helix turn. The repeating geometry pattern results in periodic orientation of the NMR interaction tensors in the molecular frame of reference. Therefore periodic signal patterns, the so-called PISA (polarity index slant angle) “wheels”, can be observed in PISEMA spectra of α -helical peptides (cf. figure 3.2). If a single or a few observed resonances belonging to the PISA “wheel” pattern are assigned with the help of selective ^{15}N labeling of the peptide under study, the assignment of the remaining signals can be deduced from the PISA “wheel” periodicity. Therefore no extensive selective labeling of each residue and no additional experiments for resonance assignment are required. Membrane orientation of the α -helix can be determined with the help of the data obtained in PISEMA spectra assuming that the ^{15}N chemical shift tensor orientation defined in such model systems as the Ala- ^{15}N -Leu dipeptide⁷⁰ is valid for all amino acids in α -helical

peptides⁷¹ (excluding glycine that is treated separately). This approach is referred to as the fixed tensor approach.

Transmembrane helix dynamics can have a significant impact on the spectral appearance of a PISA “wheel”⁷². The influence of the peptide whole-body motions can be accommodated into the analysis of PISEMA spectra and the amplitudes of the fluctuations can be extracted from the spectra⁷³. However, significant problems in the interpretation of PISEMA spectra may arise if the system under study is not α -helical or if the cylindrical rigid-body approach is not sufficient for the description of molecular orientation and motion. Moreover, the fixed tensor approach cannot always be applied to non- α -helical molecules. This chapter is focussed on the interpretation of PISEMA spectra of the β -helical antimicrobial peptide gramicidin A using molecular dynamics simulation combined with explicit ¹⁵N chemical shift tensor computation.

3.4 Application to Gramicidin A – the role of dynamics

Gramicidin A is a linear antimicrobial pentadecapeptide produced by the soil bacterium *Bacillus brevis*⁷⁴. Its primary sequence (N-formyl-L-Val-L-Gly-L-Ala-D-Leu-L-Ala-D-Val-L-Val-D-Val-L-Trp-D-Leu-L-Trp-D-Leu-L-Trp-D-Leu-L-Trp-ethanolamine) consists of alternating L- and D- amino acids. In lipid bilayers gramicidin A adopts a right-handed helical conformation. L-amino acids of gramicidin A exhibit β -sheet-type torsion angles (cf. figure 3.3). D-amino acids exhibit torsion angles that are disallowed for L-amino acids due to steric clashes. This geometry of the gramicidin A was originally proposed by Urry et al. and designated as the $\pi_{(L,D)}$ -helix⁷⁵ (also referred to in the literature as the β -helix; not to be confused with β -helix tertiary structure formed by association of parallel β -sheets into helical pattern). Gramicidin A molecules can undergo a head-to-head dimerization forming a transmembrane ion channel permeable for monovalent cations. Gramicidin A in DMPC bilayers was extensively studied by solid-state NMR. The structure of gramicidin A was determined from ¹H-¹⁵N and ¹⁵N-¹³C dipolar couplings using single site selective labeling⁷⁶. Vogt et al. recorded PISEMA spectra of fully ¹⁵N labeled gramicidin A⁷⁷ (cf. figure 3.4). At first glance, the signals observed in PISEMA spectra of gramicidin A do not follow any regular pattern. Therefore the systematic assignment of signals in the

PISEMA spectrum of gramicidin A appears impossible and selective labeling of every single site seems to be inevitable in order to determine the membrane orientation.

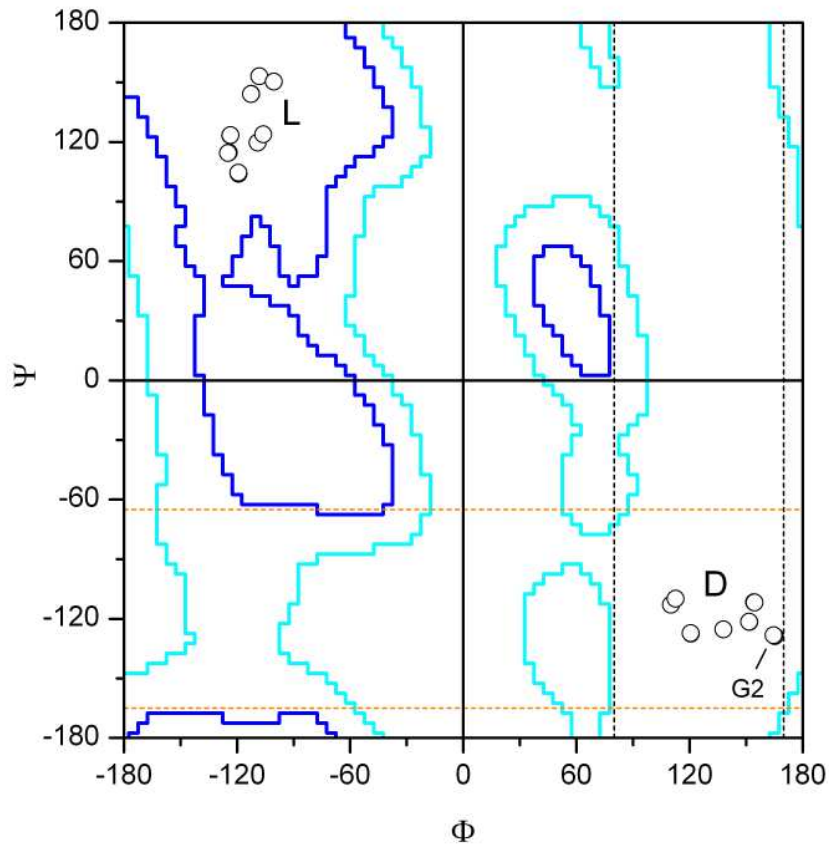


Figure 3.3: Ramachandran plot of gramicidin A. The borders of most favorable regions are shown in blue; the borders of generally allowed regions are shown in cyan⁷⁸. Dashed lines indicate steric restrictions for L-amino acids⁷⁹ (black dashed line for $O_{i-1}...C_{\beta}$ and orange dashed line for $C_{\beta}...N_{i+1}$ steric clashes). L-amino acids (L) lie within the β -sheet region while D-amino acids (D) adopt an extended geometry disallowed for L-amino acids.

The PISEMA spectrum of an α -helical peptide with a fully upright transmembrane alignment (as expected for gramicidin A channel) can collapse into a single spot. For a non-tilted (tilt angle of 0° between the helix long axis and the membrane normal) polyalanine α -helix (cf. the caption of figure 3.2) the PISA “wheel” predicted with fixed ^{15}N CS tensors has a chemical shift range of 0.049 ppm and a dipolar coupling range of 0.025 kHz. In this case, the signal dispersion is beyond the experimental spectral resolution. However, α -helical peptides usually do not possess the same torsion

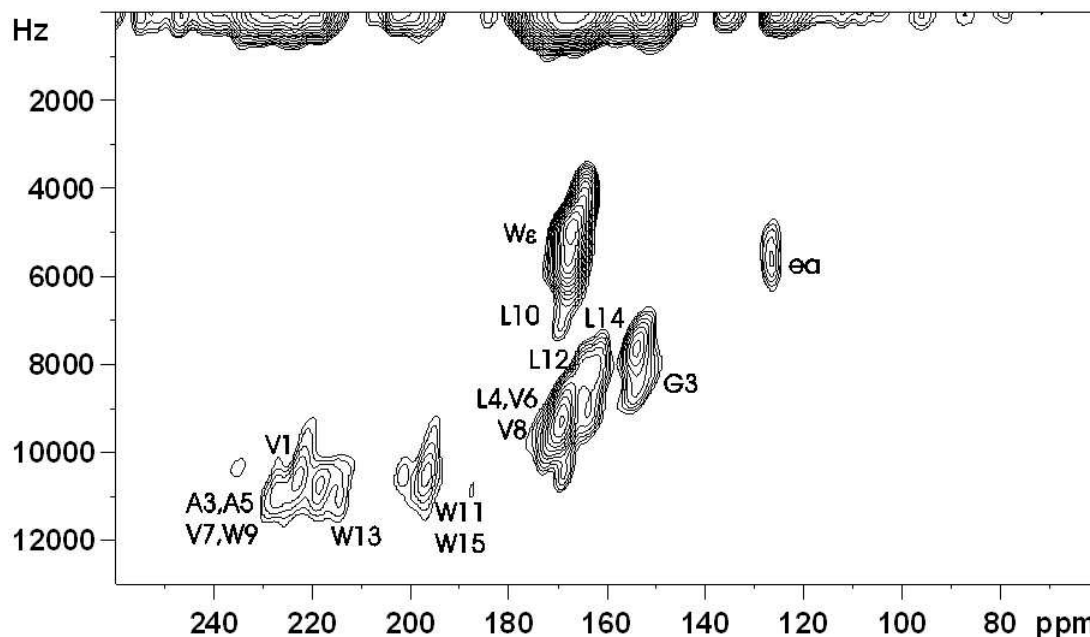


Figure 3.4: PISEMA spectrum of gramicidin A with the tentative assignment of resonances using the data published by Ketchem et al.⁷⁶ (data courtesy of Prof. Burkhard Bechinger). W_{ϵ} denotes the resonances of the tryptophane side chains and ea denotes the resonance of ethanolamine.

angles for all amino acid residues. Torsion angle variations can lead to perturbations of the PISA “wheel” signal pattern and resolve the signals of a non-tilted helical peptide⁷². The torsion angles of amino acids in gramicidin A are not identical. Moreover L- and D-amino acids alternate in the sequence of gramicidin A and their torsion angles occupy two distinct regions of the Ramachandran plot. As a result, two distinct signal regions with two strongly perturbed PISA “wheels” can be expected in the PISEMA spectrum of gramicidin A. It might even be assumed that a simple calculation of the gramicidin A PISEMA spectrum using the fixed tensor approach can facilitate its assignment, or, generalizing the problem, that calculations with the fixed tensor approach for different tilt angles with the following comparison of the calculated and experimental PISEMA spectra enable to deduce the membrane orientation of non- α -helical molecules.

The PISEMA spectrum of the gramicidin A solid-state NMR structure⁷⁶ (PDB ID 1MAG) in fully upright transmembrane orientation was calculated with the COSMOS PRO program using the fixed ^{15}N chemical shift tensors. The calculated PISEMA

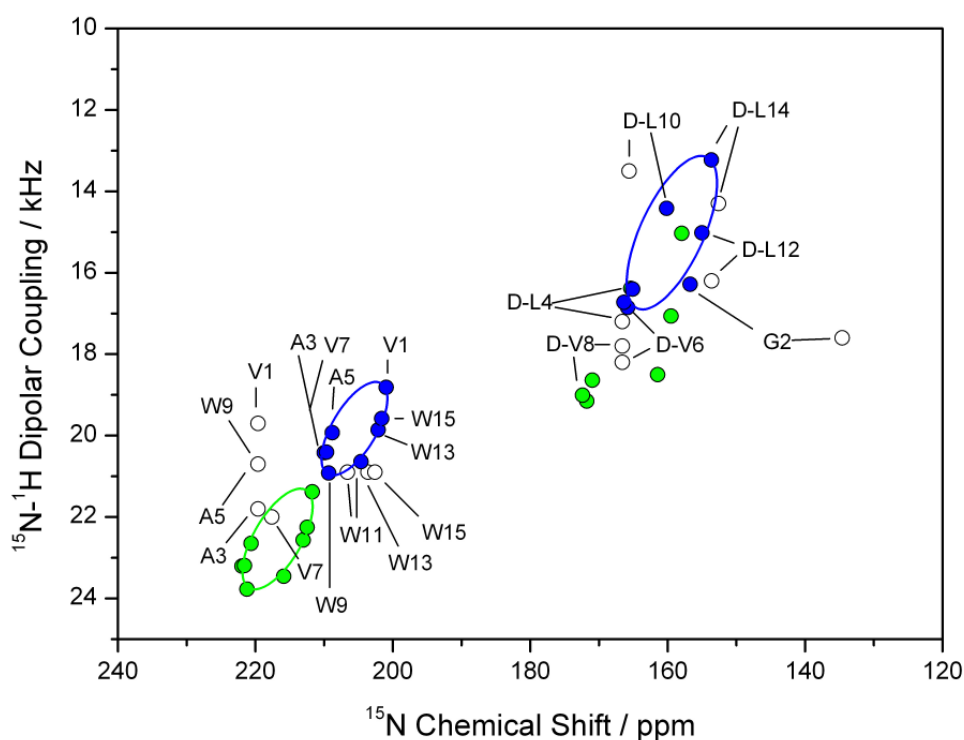


Figure 3.5: Calculated PISEMA spectra of gramicidin A (green circles – unscaled; blue circles – scaled with an order parameter of 0.88; green and blue lines are guides to the eye) compared to the data measured by Ketchem et al.⁷⁶ (open circles) using single site selective labeling.

spectrum exhibits two PISA “wheel” signal patterns, one for the L- and one for the D-amino acids. The comparison to experimental chemical shift and dipolar coupling values measured by Ketchem et al. with single site selective ^{15}N labeling reveals that the position of the calculated values appears shifted to the position of the experimental ones. Internal mobility of the peptide can shift the position of the PISA “wheel” in the spectrum. Therefore, as a first attempt to account for the molecular mobility, the calculated values were scaled with the order parameter S^i of 0.88, which was originally introduced for the initial description of dynamics in the samples containing deuterium-labeled alanine⁸⁰. The calculated PISEMA spectra and the corresponding experimental values are shown in figure 3.5. Calculated PISA “wheel” of the D-amino acids exhibits larger signal position perturbation than those of the L-amino acids, thus indicating the higher conformational variability of the D-amino acid residues. The order of the amino acid signals in both calculated PISA “wheels” corresponds to the experimentally observed position of the signals relative to each other (except Trp 9 and Gly 2). The experimental values of L-amino acids, however,

do not follow the wheel pattern. Furthermore, in the experimental PISEMA spectrum of the fully ^{15}N labeled sample recorded by Vogt et al. (cf. figure 3.4) very broad, superimposing signals are observed (in section 3.9 it will be demonstrated that the primary reason of the signal broadening is not relaxation or insufficient quality of alignment but the intrinsic mobility in the peptide backbone). At first glance, the experimental PISEMA spectrum seems to have nothing in common with the calculated one. Therefore, though PISEMA spectra calculated using the fixed tensor approach can assist the explanation of the chemical shift and dipolar coupling values measured by Ketchum et al.⁷⁶, they cannot facilitate the assignment and evaluation of the experimental PISEMA spectrum of the fully ^{15}N labeled gramicidin A.

It seems natural that the calculated ^{15}N chemical shift tensors might provide a more accurate prediction of PISEMA spectra than the fixed tensor values. As discussed in chapter one, experimental peptide and protein structures are usually not suitable for the quantum chemical ^{15}N chemical shift tensor prediction. BPT isotropic chemical shift driven structure refinement allows to obtain a structure that is not only a good representation of the experimental chemical shifts but is also suitable for the ^{15}N CS tensor computation. Therefore the solid-state NMR structure of gramicidin A (PDB ID 1MAG) was subjected to the isotropic chemical shift⁸¹ driven geometry optimization with subsequent ^{15}N CS tensor calculation. A correlation coefficient R of 0.94 and standard deviation of 1.9 ppm between the experimental and the BPT calculated isotropic ^{15}N CS was achieved after the CS driven geometry optimization (cf. figure 3.6). The backbone C_α RMSD between the experimental and the ^{15}N CS optimized structure is 0.08 Å. The principal components of the ^{15}N chemical shift tensor calculated before and after isotropic ^{15}N CS driven geometry optimization show significant differences (cf. figure 3.7). It is noteworthy that the δ_{11} component (in Haeberlen-Mehring convention) of the Gly 2 ^{15}N CS tensor has a non-physical value of -4.66 ppm before the CS driven geometry optimization. After optimization all BPT calculated principal components of the ^{15}N CS tensors lie within the expected range of values for these parameters.

BPT calculated ^{15}N chemical shift tensors of gramicidin A can be compared to the experimental ^{15}N chemical shift tensors measured by Mai et al.⁸² (cf. figure 3.8). The

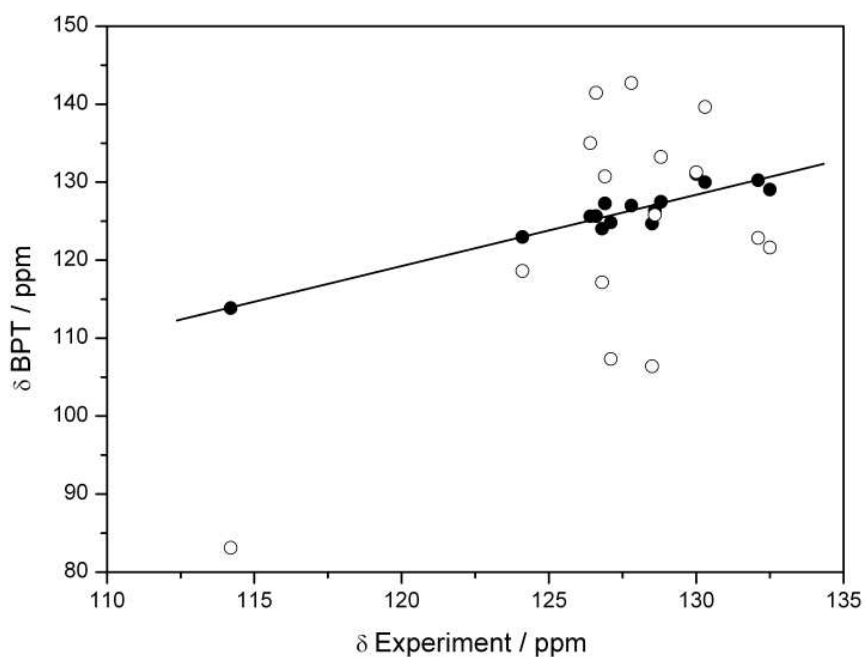


Figure 3.6: Calculated isotropic ^{15}N CS of gramicidin A before (open circles) and after (black circles) chemical shift driven geometry optimization (black line represents the linear fit between the calculated and the experimental CS values).

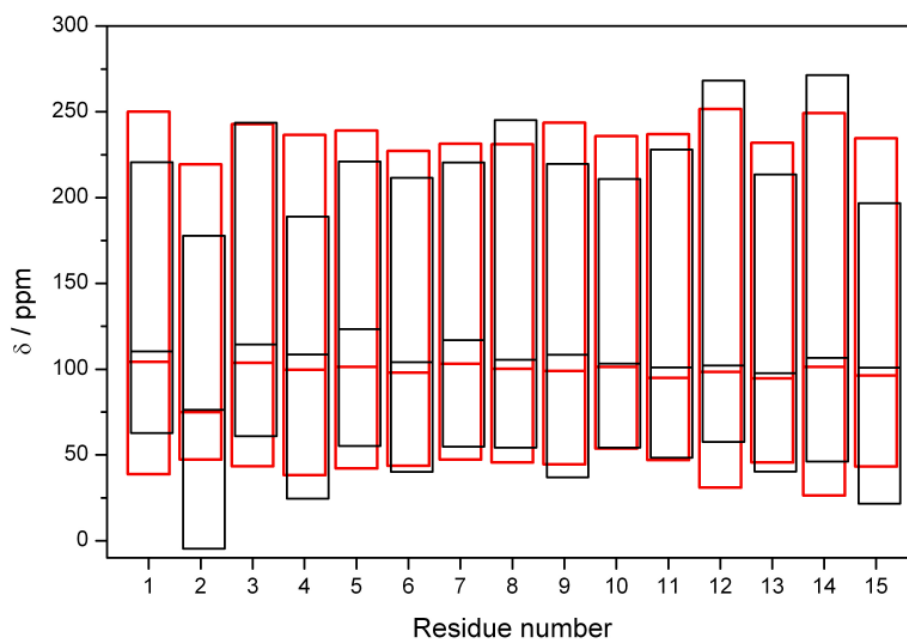


Figure 3.7: Calculated principal components of the ^{15}N CS tensors of the gramicidin A backbone before (black) and after (red) chemical shift driven geometry optimization.

comparison should be taken with care since Mai et al. measured these values on the dry mixture of gramicidin A and dimyristoylphosphatidylcholine (DMPC) lipid with the unknown gramicidin A conformation. The calculated δ_{11} components are systematically smaller than the experimental ones while the δ_{22} and δ_{33} components are systematically too large. This trend results in a larger span of the calculated ^{15}N CS. Quantum chemical shift calculations provide the tensor components of the purely static molecules. Therefore the source of the experimentally observed tensor span narrowing can be attributed to molecular motion. Furthermore it is possible to compare the experimental and the calculated ^{15}N CS tensor principal components with the tensor component values applied in the fixed tensor approach calculations (cf. figure 3.8). The fixed tensor approach leads to the errors of approximately up to 10 ppm if compared with experimental values and of approximately up to 20 ppm if compared with the calculated values.

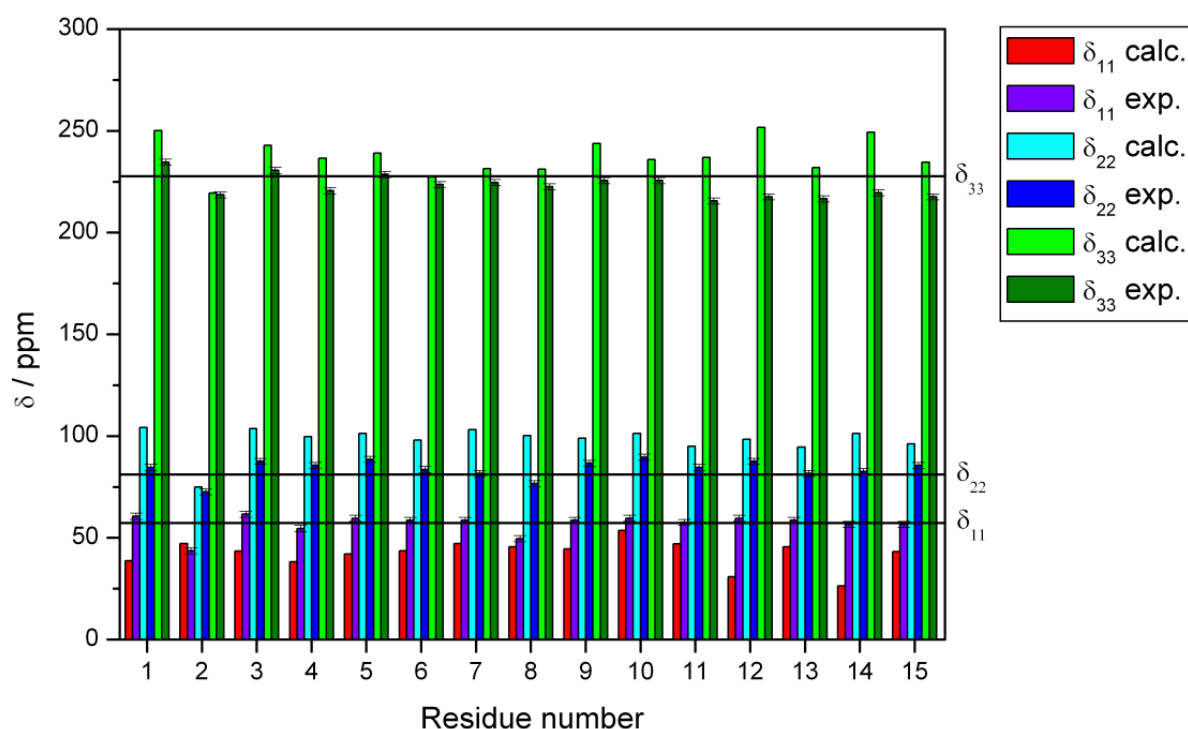


Figure 3.8: Comparison of the BPT calculated principal ^{15}N CS tensor components with values of Mai et al. obtained on amorphous DMPC/gramicidin A samples. Horizontal black lines indicate the principal components applied in fixed tensor approach⁷¹.

In contrast to the calculated principal components of the ^{15}N CS tensors, calculated ^{15}N CS tensor orientations within the molecular framework are in agreement with experimental results (cf. figure 3.9). The tensor orientation is almost the same for all residues of gramicidin A excluding Val 1 and Gly 2 and no significant differences can be observed between the BPT calculated and the experimental⁸² values (cf. table 3.1). The only considerable deviation between the calculated and the experimental tensor orientation is found for the α_D angle of Val 1. The experimental value of the α_D angle of Val 1 is unusually large. Calculated ^{15}N CS tensors are visualized in figure 3.10.

Table 3.1: Calculated and experimental ^{15}N CS tensor orientation in gramicidin A dimer.

Residue	β_D calc. (deg)	β_D exp. (deg)	α_D calc. (deg)	α_D exp. (deg)
Val 1	98.6 and 98.3	106 ± 2	0	28 ± 5
Gly 2	100.3 and 103.6	98 ± 2	0	0 ± 5
Ala 3 to Trp 15	104-106	$104-106 \pm 2$	0	0 ± 5

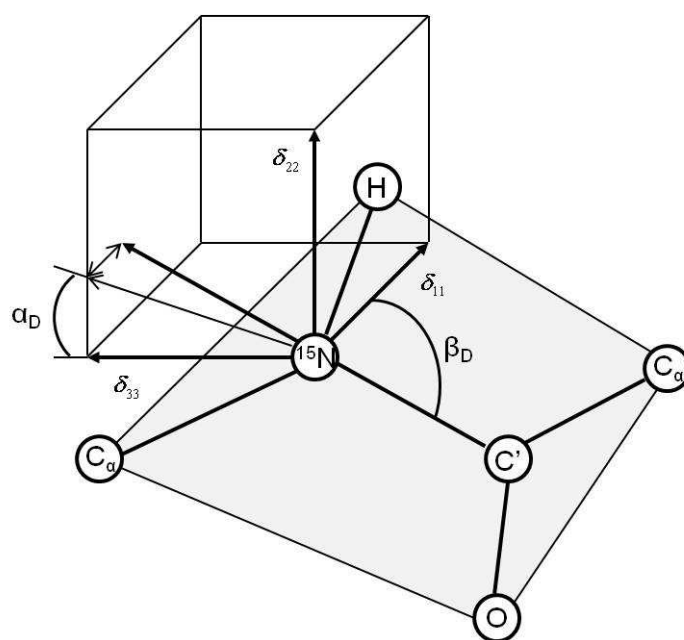


Figure 3.9: ^{15}N CS tensor orientation in the molecular frame as defined by Mai et al.⁸² The peptide plane is shown in light gray.

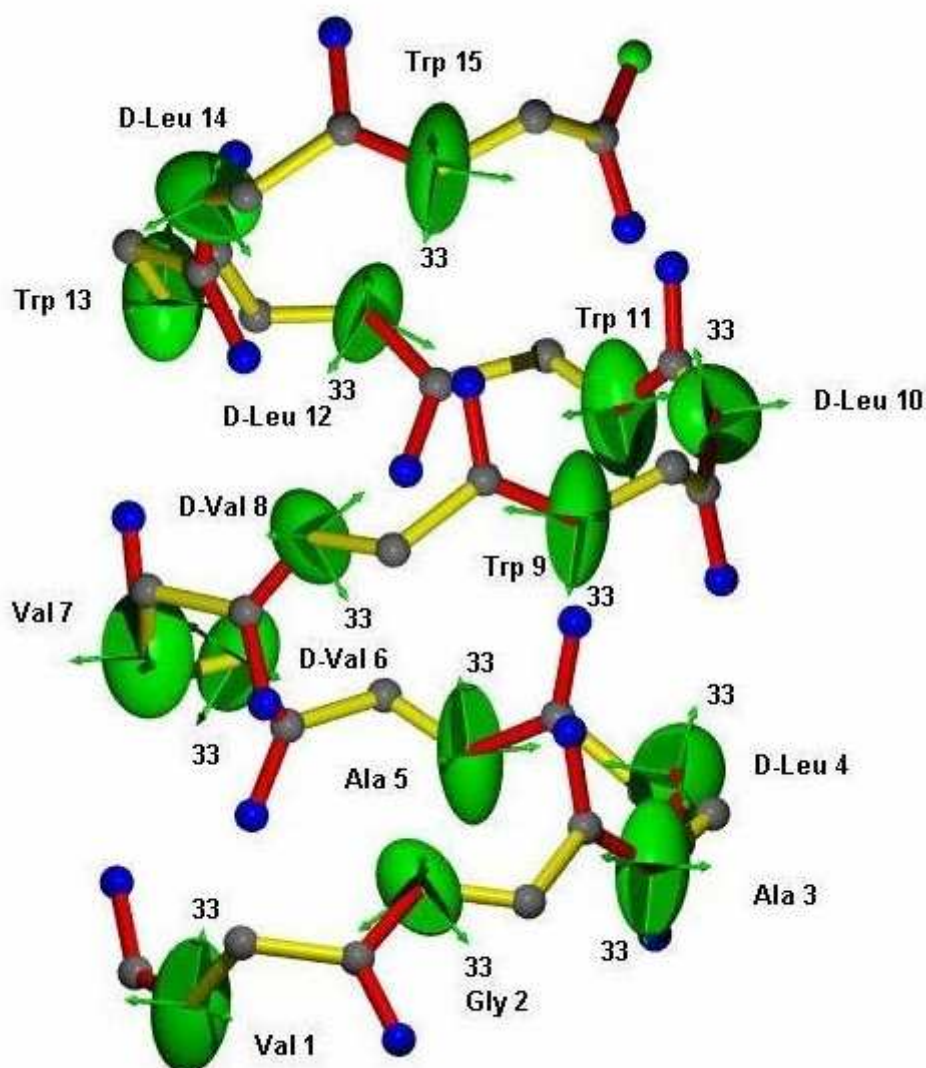


Figure 3.10: Backbone of the gramicidin A structure optimized with isotropic ^{15}N chemical shifts. CS tensor ellipsoids and their orientations are displayed for backbone amide ^{15}N atoms.

PISEMA spectra of the optimized structure of the gramicidin A in fully upright transmembrane orientation were calculated with the COSMOS PRO program using fixed and BPT calculated ^{15}N chemical shift tensors. Calculated PISEMA spectra do not exhibit the PISA “wheel” signal pattern (cf. figure 3.11). In contrast to the spectra calculated based on the solid-state NMR structure (cf. figure 3.5), the predicted signals of L-amino acids occupy the same spectral region as the experimental signals. However, the predicted signals of D-amino acids are shifted downfield from the spectral region of the experimental signals. Moreover, signals in calculated spectra cannot match the experimental values of Ketchem et al.⁷⁶ and the order of

the amino acid signals in both calculated spectra does not correspond to the experimentally observed position of the signals relative to each other. It becomes obvious that the source of the unusual signal pattern in the PISEMA spectrum of gramicidin A is neither the structural fidelity nor the incorrect assumptions for the ^{15}N CS tensor (these issues were eliminated by the BPT structure refinement and CS calculation) but rather molecular dynamics. Therefore molecular dynamics of the gramicidin A backbone is investigated in the next sections and its influence on the PISEMA spectra is considered explicitly using molecular dynamics simulation with orientational constraints.

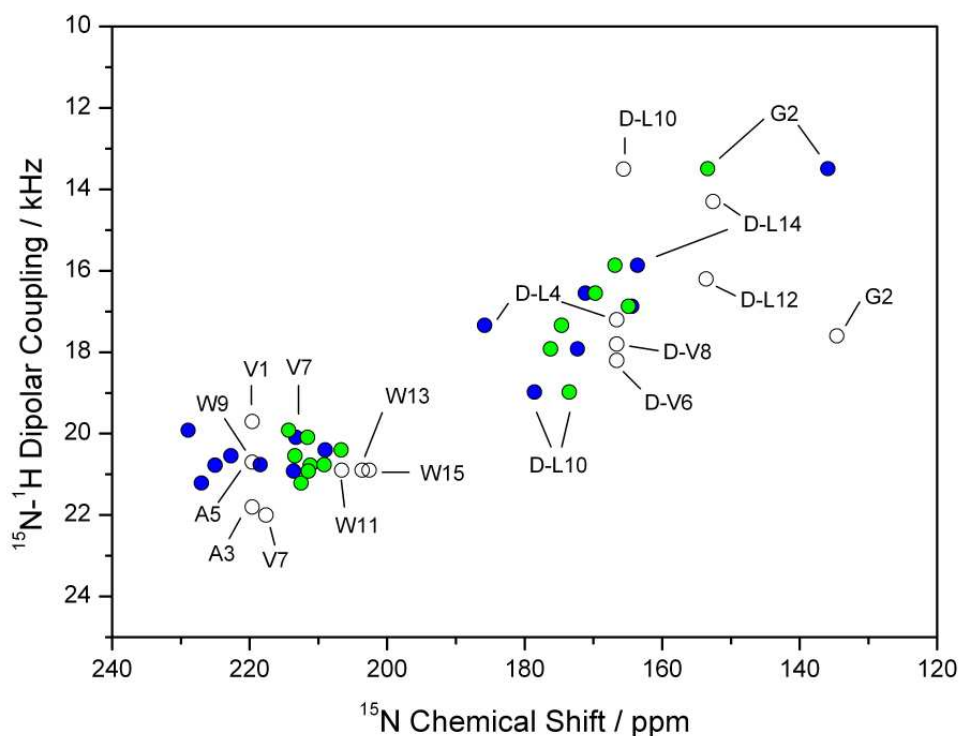


Figure 3.11: Calculated PISEMA spectra of the gramicidin A structure optimized with isotropic ^{15}N CS (green circles – fixed ^{15}N CS tensor; blue circles – BPT calculated ^{15}N CS tensor) compared to the data measured by Ketchem et al.⁷⁶ (open circles) using single site selective labeling. All values were scaled with an order parameter of 0.88. Even though CS driven geometry optimization did not significantly alter the structure (backbone C_α RMSD of 0.08 Å), significant perturbations of the PISEMA spectrum calculated with the fixed tensor approach can be observed compared to the spectrum shown in figure 3.5.

3.5 Molecular dynamics with orientational constraints

Molecular dynamics (MD) simulations can predict physical movements of molecules in a classical framework. Experimental restraints can be accommodated into the molecular dynamics simulation by adding pseudo-energy terms depending on experimental parameters. The derivatives of the pseudo-energy terms provide forces for molecular dynamics. Sternberg et al. introduced molecular dynamics simulations with NMR orientational constraints⁸³. In the approach of Sternberg et al., the pseudo-energy term representing the influence of a tensorial property (in this case the dipolar coupling) is given by

$$E_{pseudo} = \frac{1}{2} k_0 s_t s_i^{\Delta D_{\alpha\beta}} \sum_{\alpha\beta} \sum_i (D_{\alpha\beta}^{theo_i} - D_{\alpha\beta}^{exp_i})^2 \quad \text{with} \quad \alpha, \beta = \{x, y, z\}, \quad (3.1)$$

where E_{pseudo} denotes the dipolar coupling pseudo-energy, $D_{\alpha\beta}^{exp_i}$ – the components of the experimental dipolar coupling tensor assigned to the nucleus i (e.g. to the ^1H nucleus in the case of the ^1H - ^{15}N dipolar coupling), $D_{\alpha\beta}^{theo_i}$ – the components of the calculated dipolar coupling tensor assigned to the nucleus i , k_0 – a force constant adjusting the magnitude and the unit of the pseudo-energy, s_t and $s_i^{\Delta D_{\alpha\beta}}$ – scaling functions controlling the dipolar coupling potential.

In the first steps of the MD simulation the calculated dipolar couplings do not match the experimental motionally averaged dipolar couplings. Many MD steps are required in order to approach the experimental dipolar coupling values. That is why the largest mismatch between the experimental and the calculated dipolar couplings resulting in very high dipolar coupling pseudo-forces is usually observed during the equilibration phase of the MD simulation. Too high dipolar coupling pseudo-forces can easily deform the whole structure of the molecule. Therefore a time-dependent scaling factor s_t and the scaling function $s_i^{\Delta D_{\alpha\beta}}$ are introduced in order to control the magnitude of the dipolar-coupling pseudo-energy. The value of the scaling factor s_t increases exponentially in the course of the MD simulation and finally approaches the value of 1.0 at the MD time point $t = \rho$. It can be described by the equation

$$s_t = 1 - e^{-t/\rho}, \quad (3.2)$$

where ρ should be set to the value comparable with the MD equilibration time. The scaling function $s_i^{\Delta D_{\alpha\beta}}$ depends on the difference between the calculated and the experimental dipolar coupling values. The function $s_i^{\Delta D_{\alpha\beta}}$ is defined analogously to the function controlling the chemical shift pseudo-forces and its value depends on the potential width control parameter $\Delta D_{\alpha\beta}$ (cf. equation 2.40). If the difference between the calculated and the experimental dipolar coupling values is small the resulting dipolar coupling pseudo-forces behave similar to the harmonic pseudo-forces. If this difference is large the dipolar coupling pseudo-forces approach constant values.

In addition, to minimize the influence of the MD equilibration phase on the evaluation of the MD trajectory, time averaged dipolar coupling values can be calculated using the exponential memory decay function originally introduced by Torda and van Gunsteren⁸⁴:

$$\langle D_{\alpha\beta} \rangle_t = \frac{1}{N} \int_0^t e^{-t/\tau} D_{\alpha\beta} dt, \quad (3.3)$$

where τ is the decay constant for the exponential decay and N is the norm of the integral (number of MD steps in which the dipolar couplings were computed). The averaged dipolar couplings do not reflect the influence of short fluctuations in the system under study but rather the trends in system's behaviour.

3.6 Molecular dynamics simulation of gramicidin A

A 5 ns molecular dynamics simulation with ^1H - ^{15}N dipolar couplings measured by Ketchum et al.⁷⁶ as orientational constraints was performed on the solid-state NMR structure of the gramicidin A dimer by Dr. Ulrich Sternberg. The NMR experiment time scale is in the range of several milliseconds. Therefore, at first glance, a MD simulation on the nanosecond timescale will not be able to reproduce the NMR data. However, orientational constraints provide additional energy to the molecular system under study and accelerate the molecular dynamics. That is why molecular dynamics simulation with orientational constraints allows to sample the conformations observed on the NMR time scale on the nanosecond time scale. Furthermore, molecular dynamics with orientational constraints does not require explicit modeling of the

membrane environment since the orientational influence of the membrane is already represented by the orientational constraints. 2000 snapshots of molecular coordinates (one after every 2.5 ps; data courtesy of Dr. Ulrich Sternberg) were extracted from the molecular dynamics trajectory and subjected to the ^{15}N CS driven geometry optimization with the following ^{15}N CS tensor calculation (this approach is justified since structural changes introduced by the CS driven geometry optimization are negligibly small (cf. section 2.7) compared to the molecular structure and orientation fluctuations during the MD run). The behavior of the zz tensor components of the dipolar coupling and ^{15}N CS tensors during the MD simulation represented by the geometry optimized MD snapshots was analyzed and compared to the values measured by Ketchem et al.⁷⁶

Time averaged ^1H - ^{15}N dipolar couplings were computed as described in equation 3.3 with the decay constant τ of 200 ps. Experimental dipolar couplings provided the driving forces for the MD simulation. Therefore the averaged dipolar couplings calculated from 2000 MD snapshots are in good agreement with the experimental dipolar couplings of Ketchem et al. (cf. figure 3.12). The correlation coefficient R of 0.997 and the standard deviation SD of 0.1 kHz are observed between the calculated and the experimental dipolar couplings. A typical ^1H - ^{15}N dipolar coupling trajectory is shown in figure 3.13. The largest fluctuations of the ^1H - ^{15}N dipolar coupling occur in the first 200 ps of the MD simulation, and thus the setting of 200 ps for the decay constant τ appears well justified.

The trajectory of the calculated ^{15}N CS tensor zz components exhibits large fluctuations during the whole MD simulation (cf. figure 3.13). This is in good agreement with the results obtained by Woolf et al.⁴² Woolf et al. analyzed the molecular dynamics trajectory of gramicidin A and performed DFT ^{15}N nuclear shielding tensor calculations for several sites of gramicidin A in MD geometry by substituting the neighbouring residues of the sites under study by N-methyl-acetamid moieties. Woolf et al. reported large fluctuations of the CS tensor components but their approach was not able to reproduce the experimental ^{15}N CS with quantitative accuracy. In the case of the rapid conformational exchange (the time step between two MD snapshots is only 2.5 ps long) the observed NMR chemical shifts should be

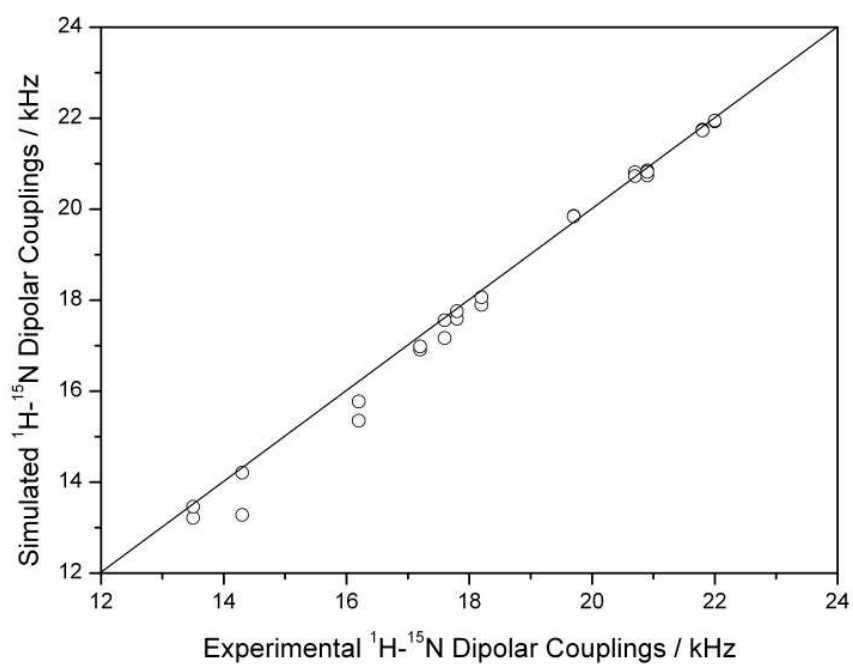


Figure 3.12: Correlation of the experimental and averaged calculated ^1H - ^{15}N dipolar couplings in gramicidin A.

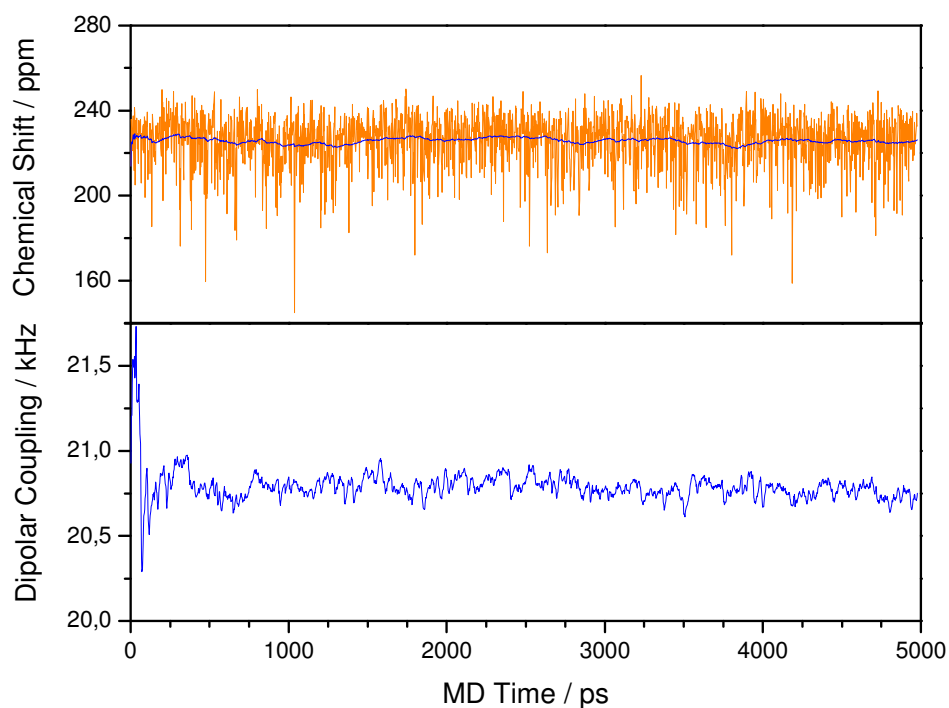


Figure 3.13: Typical trajectories of ^1H - ^{15}N dipolar couplings and zz components of the ^{15}N CS tensor during MD simulation with orientational constraints (depicted for residue Ala 5). The fluctuations of δ_{zz} are shown in orange and the running mean averaged values (cf. equation 3.3) are shown in blue.

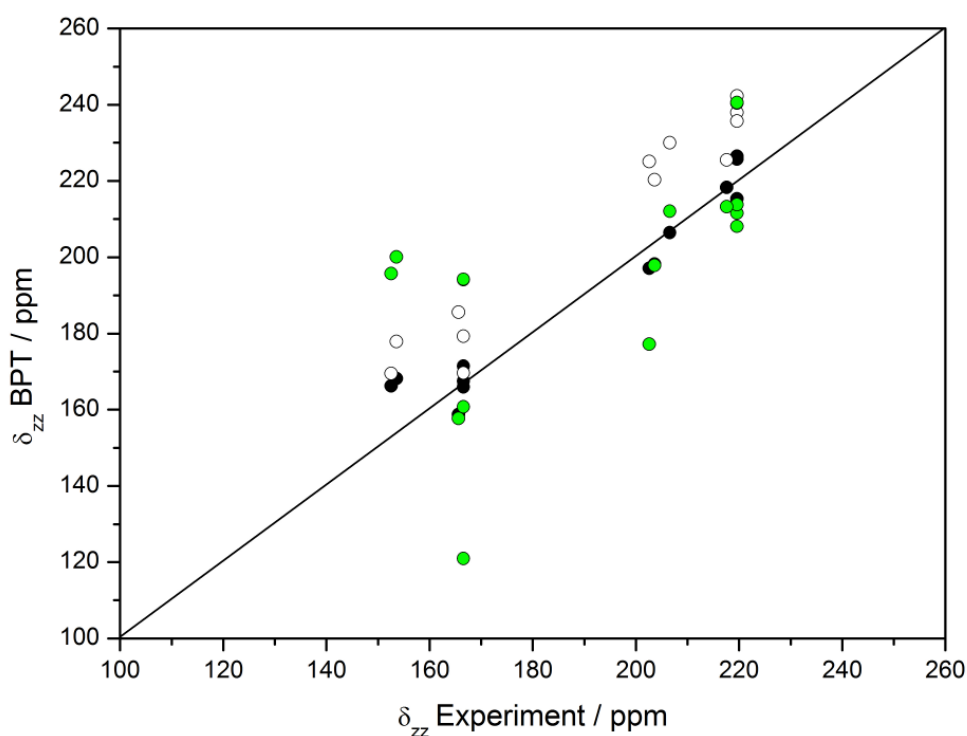


Figure 3.14: Comparison of the experimental and BPT calculated zz components of the ^{15}N CS tensor in gramicidin A (solid green circles – values obtained on the original solid-state NMR structure; open black circles – values after geometry optimization with isotropic ^{15}N chemical shifts; solid black circles – time averaged values after MD with dipolar coupling orientational constraints).

computed as the averaged values over the MD chemical shift trajectory. Time averaged ^{15}N CS values were computed analogously to the time averaged dipolar couplings using equation 3.3 with the decay constant $\tau = 200$ ps (cf. figure 3.13).

In marked contrast to the ^1H - ^{15}N dipolar couplings, ^{15}N chemical shift tensors were treated as free parameters during the MD simulation. ^{15}N chemical shifts were obtained on MD snapshots after the MD run and there were no CS pseudo-forces during the MD simulation. All snapshots were subjected to the isotropic ^{15}N chemical shift driven geometry optimization and no tensorial ^{15}N constraints were applied. Isotropic ^{15}N chemical shift driven geometry optimization of the solid-state NMR gramicidin A structure (PDB ID 1MAG) results in the standard deviation SD of 19.2 ppm between the calculated and the experimental ^{15}N CS tensor zz components (note that the SD for the isotropic ^{15}N CS is only 1.9 ppm). Time averaged ^{15}N CS tensor zz components computed on 2000 MD snapshots after the isotropic ^{15}N CS

driven geometry optimization exhibit the standard deviation SD of 6.9 ppm between the calculated and the experimentally observed values (cf. figure 3.14). This improvement of the calculated ^{15}N CS tensor zz component values indicates that molecular dynamics with orientational constraints is able to reproduce the realistic dynamic behavior of gramicidin A.

Orientation of the vector connecting the centers of gravity of the two β -helices forming a gramicidin A dimer (in the following referred to as the dimer alignment vector) can describe the membrane alignment of gramicidin A. First inspection of the gramicidin A alignment during the MD simulation (cf. figure 3.15) reveals that no significant whole-body reorientations of gramicidin A in the membrane occur and the gramicidin A dimer is wobbling around the upright transmembrane alignment. Therefore the primary reason of the signal pattern perturbation in the PISEMA spectrum is probably not the whole-body motion but the individual dynamics of the Gramicidin A residues which is analyzed in the next sections.

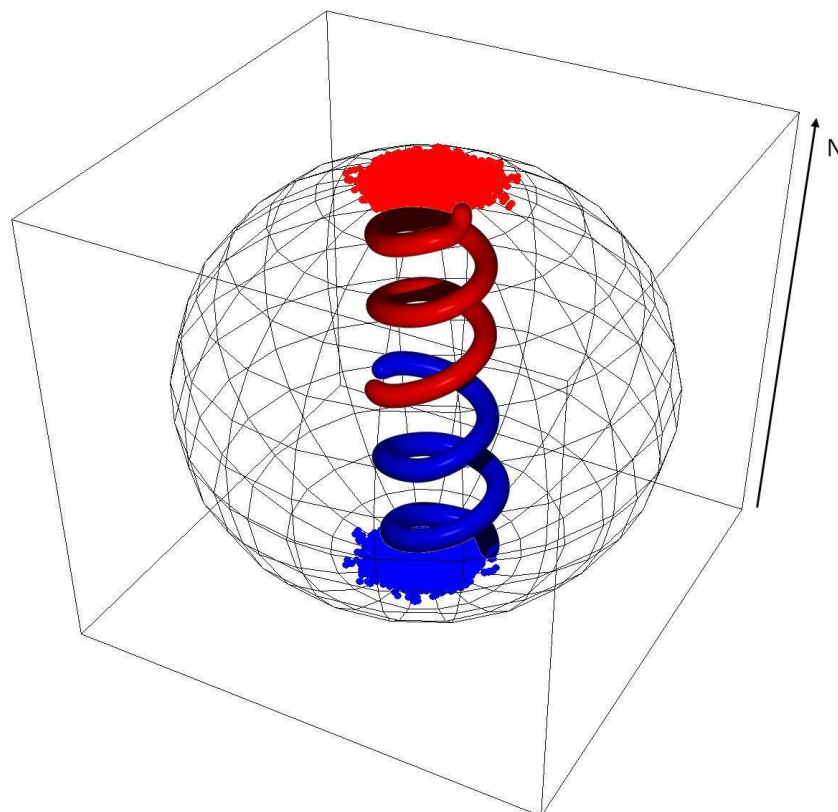


Figure 3.15: Scatter plot depicting the orientation fluctuations of the dimer alignment vector during MD simulation with orientational constraints. The backbone of gramicidin A remains oriented parallel to the membrane normal N and the external magnetic field.

3.7 Local order tensors

In the MD simulations with orientational constraints introduced by Sternberg et al.⁸³ dipolar coupling tensor \mathbf{D} is defined as a constant tensor in the bond frame of reference. The bond frame of reference is identical to the principal axis system of the dipolar coupling tensor. The z axis of the principle axis system is aligned with (and points through) the interconnecting vector between the coupled nuclei (for the ^1H - ^{15}N dipolar couplings observed in PISEMA experiments the interconnecting vector is the ^1H - ^{15}N bond). The principal values of \mathbf{D} are $\mathbf{D}_{\alpha\alpha} = \{-d^\dagger/2; -d^\dagger/2; d^\dagger\}$, where d^\dagger is the dipole-coupling constant given by

$$d^\dagger = 2\hbar \left(\frac{\mu_0}{4\pi} \right) \bar{r}^{-3} \gamma_I \gamma_S \quad (3.4)$$

with the reduced Planck's constant \hbar , magnetic field constant $\mu_0(4\pi)^{-1}$, the gyromagnetic ratios γ_I and γ_S of the coupling nuclei and the averaged internuclear distance \bar{r} . The internuclear distance is treated as a constant and the influence of the bond vibration is neglected.

The calculation of the pseudo-energies and pseudo-forces for the MD simulation is performed in the laboratory frame of reference. Therefore a transformation from the bond frame of reference into the laboratory frame of reference is required for each dipolar coupling constraint:

$$D_{\alpha\beta}^{lab} = R_{\alpha\alpha'} R_{\beta\beta'} D_{\alpha'\beta'}^{bond}. \quad (3.5)$$

The corresponding transformation matrices \mathbf{R} consist of the unit vectors along the coordinate system axes in the molecular frame of reference:

$$\mathbf{R} = (\vec{e}_x, \vec{e}_y, \vec{e}_z), \quad (3.6)$$

where the coordinates of the unit vectors are given in the laboratory frame. Then the time averaged dipolar coupling can be expressed as:

$$\langle D_{\alpha\beta}^{lab} \rangle_t = \langle R_{\alpha\alpha'} R_{\beta\beta'} \rangle_t D_{\alpha'\beta'}^{bond} \quad (3.7)$$

with the brackets indicating averaging over time.

It should be noted that, if a bond between the two coupled nuclei is statically aligned in the external magnetic field, it is common to express the orientational dependence of the observed dipolar coupling with the help of the second Legendre polynomial instead of the transformational matrices:

$$\begin{aligned}
 |D_{observed}| &= d^\dagger \cdot P_2 \\
 &= 2\hbar \left(\frac{\mu_0}{4\pi} \right) \bar{r}^{-3} \gamma_I \gamma_S \frac{(3\cos^2\theta - 1)}{2}. \quad (3.8) \\
 &= d(3\cos^2\theta - 1)
 \end{aligned}$$

In this case, the expression for the observed heteronuclear dipolar coupling resembles the expression given in equation 1.27 and yields the dipolar coupling constant d given in equation 1.25.

Transformation matrices connecting the dipolar coupling tensor in the principal axis system (i.e. in the bond frame of reference) with its averaged value reflect the molecular mobility and allow to obtain the local order tensor \mathbf{W}^i ⁸⁵:

$$W_{\alpha\beta}^i = \frac{1}{2} \langle 3R_{\alpha z}^i R_{\beta z}^i - \delta_{\alpha\beta} \rangle_t, \quad (3.9)$$

where $\delta_{\alpha\beta}$ is the Kronecker delta function. The local order tensors \mathbf{W}^i are calculated in every step of the molecular dynamics simulation with orientational constraints for each site i with assigned dipolar coupling (e.g. for the ^1H nuclei in the case of the ^1H - ^{15}N dipolar coupling).

A convenient interpretation of the local order tensors \mathbf{W}^i is possible if they are transformed (diagonalized) into the so-called principal frame of order S as described by Low⁸⁶:

$$\mathbf{S}^i = \begin{pmatrix} -\frac{1}{2}S - \xi & 0 & 0 \\ 0 & -\frac{1}{2}S + \xi & 0 \\ 0 & 0 & S \end{pmatrix}. \quad (3.10)$$

The order parameter S and biaxiality parameter ξ in the principal frame of order refer to the average direction of motion along which the system under study may be

regarded as aligned – the so-called director. The director of a molecule in a fully upright transmembrane alignment (e.g. gramicidin A) is parallel to the membrane normal. Fully rigid sites of a molecule permanently oriented parallel to the membrane normal will exhibit the parameter $S = 1$. Motional fluctuations of the site under study result in the fluctuations of the dipolar coupling principal axis system around the director (as seen in the laboratory frame of reference). The value of the order parameter S decreases if the site mobility and the amplitude of such fluctuations increase. For an isotropically moving site with no preferred direction of motion the value of S drops to zero. The biaxiality parameter ξ indicates if a second preferred direction of non-random motion exists. In this case $\xi \neq 0$ and molecular motions are characterized by two axes being orthogonal to each other.

This approach for derivation of local order parameters is not restricted to dipolar couplings but is valid for every tensorial NMR property. All MD snapshots applied for the MD trajectory evaluation of gramicidin A were subjected to the isotropic ^{15}N CS driven geometry optimization with the following ^{15}N CS tensor calculation. Therefore the analysis of the MD trajectory of gramicidin A yields not only the dipolar coupling order parameters S_{DC} and the corresponding biaxialities ξ_{DC} for each residue but also the chemical shift order parameters S_{CS} and the corresponding biaxialities ξ_{CS} . The ^1H - ^{15}N dipolar interaction is very sensitive to the motion of the hydrogen atom and therefore local order parameters obtained from dipolar coupling usually describe the dynamics of the corresponding ^1H - ^{15}N bonds and not necessary the dynamics of the backbone. ^{15}N chemical shift tensor is influenced by all the atoms in its surrounding and it is less influenced by the ^1H - ^{15}N bond motion⁸⁷. That is why local order parameters obtained from ^{15}N chemical shift tensors represent the dynamics of the whole corresponding residues and their surroundings. Therefore the methodology presented in this chapter does not only permit the simulation of the PISEMA spectra of non- α -helical molecules (cf. section 3.9) but additionally enables the computation of the ^{15}N CS tensor derived order parameters and biaxialities enhancing the dynamics analysis of membrane peptides.

3.8 Local order parameters in gramicidin A

Local order parameters in gramicidin A were calculated as described in section 3.7. The calculation was restricted to the last 1000 MD coordinate snapshots in order to assure that the calculated order parameters describe the equilibrium dynamics of gramicidin A.

Order parameters S_{DC} obtained from dipolar couplings (cf. figure 3.16) reveal substantial differences in the mobility of L- and D-amino acids. All D-amino acids display a remarkably higher mobility (i.e. lower order parameters) than the L-amino acids. Moreover the order parameters of L-amino acids do not differ significantly (the only exception is the very mobile residue Gly 2) while the mobility of D-amino acids increases towards the C-terminus of gramicidin A and therefore towards the surface of the lipid bilayer.

The biaxiality values ξ_{DC} obtained from dipolar couplings (cf. figure 3.17) are very low and therefore should be interpreted with great care since very low biaxiality values might be an artefact of the calculation. The biaxiality values describing the motion of D-amino acids are higher than those of L-amino acids and a significant increase of D-amino acid biaxiality values is observed towards the C-terminus of gramicidin A and consequently towards the surface of the lipid bilayer.

All in one the mobility of all L-amino acids in gramicidin A appears to be almost the same (excluding the very mobile residue Gly 2). It could be described as wobbling around the membrane normal (the director) and there are almost no non-random motions around other directions. The mobility of D-amino acids is higher than those of L-amino acids, it increases towards the bilayer surface and motional fluctuations around the second axis can be observed in D-amino acid residues in the proximity of the C-terminus of gramicidin A. However, the same order parameters for all L-amino acids do not comply with L-amino acid signal pattern perturbations observed in the PISEMA spectrum. In addition, it remains unclear why the differences in the D-amino acid mobility are not reflected in the mobility of the neighbouring L-amino acids.

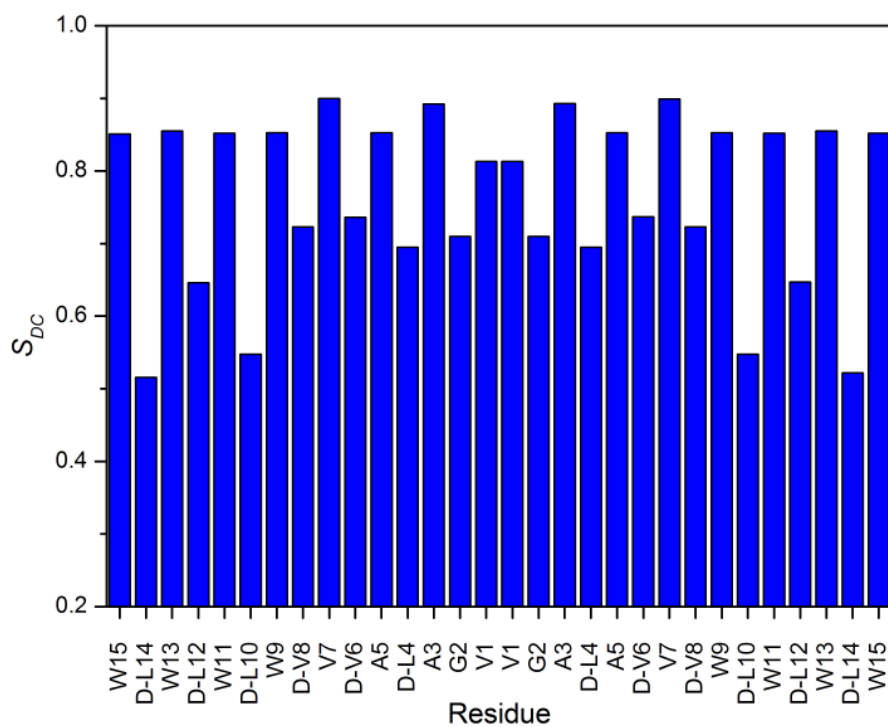


Figure 3.16: Dipolar coupling order parameters in gramicidin A dimer.

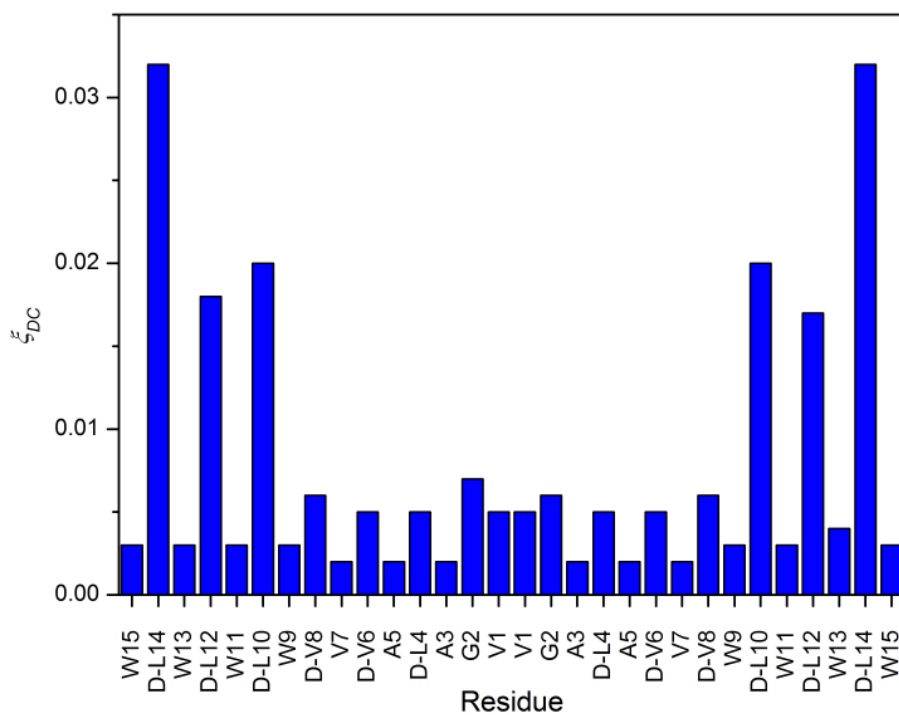


Figure 3.17: Dipolar coupling derived biaxiality values in gramicidin A dimer.

Order parameters S_{CS} obtained from the ^{15}N CS tensors (cf. figure 3.18) are almost unbiased by ^1H - ^{15}N bond motions and provide a better insight into the backbone mobility of gramicidin A. They confirm that D-amino acids display a remarkably higher mobility than the L-amino acids (excluding the very mobile residue Gly 2). The mobility of L-amino acids increases towards the termini of gramicidin A and the lowest mobility is observed for the residue Ala 5. D-amino acids exhibit a similar mobility distribution but the lowest D-amino acid mobility is observed in the middle of the gramicidin A sequence for the residue D-Val 8. However, the mobility of the residue D-Leu 10 does not follow this pattern and it is remarkably low.

All biaxiality values ξ_{CS} obtained from ^{15}N CS tensors are significantly larger than the biaxiality values ξ_{DC} obtained from dipolar couplings (cf. figure 3.19). The biaxiality values describing the mobility of D-amino acids and of the residue Gly 2 are higher than those of L-amino acids. They exhibit a repetitive pattern with a periodicity of six residues (cf. figure 3.20). The highest biaxialities in this pattern are observed for Gly 2 and D-Val 8, the middle biaxialities for D-Leu 4 and D-Leu 10 and the lowest biaxialities for D-Val 6 and D-Leu 12. The biaxiality values describing the motion of L-amino acids increase towards the termini of gramicidin A and the lowest biaxiality is observed for the residue Val 7.

Previous theoretical investigations on the functional mechanism of gramicidin A suggested that correlated librations are present among the peptide planes⁸⁸. Roux and Karplus proposed a hypothesis in which an ion passing the gramicidin A channel is not “hopping” between the local energy minimum positions but the local energy minimum position “flows” with the ion and correlated librations among the peptide planes are associated with this process⁸⁹. North and Cross stated that the influence of cations on the structure and dynamics of gramicidin A is very small and extensive motional correlations can be anticipated even in the absence of cations. ^{15}N T_1 relaxation measurements of North and Cross demonstrated that local backbone motions on the nanosecond time scale are present in gramicidin A⁹⁰. The ion translocation time through the gramicidin A channel is also on the nanosecond time scale^{91,92}. North and Cross hypothesized that the deformation induced in the structure of gramicidin A (e.g. by an interaction with a cation) could propagate via correlated

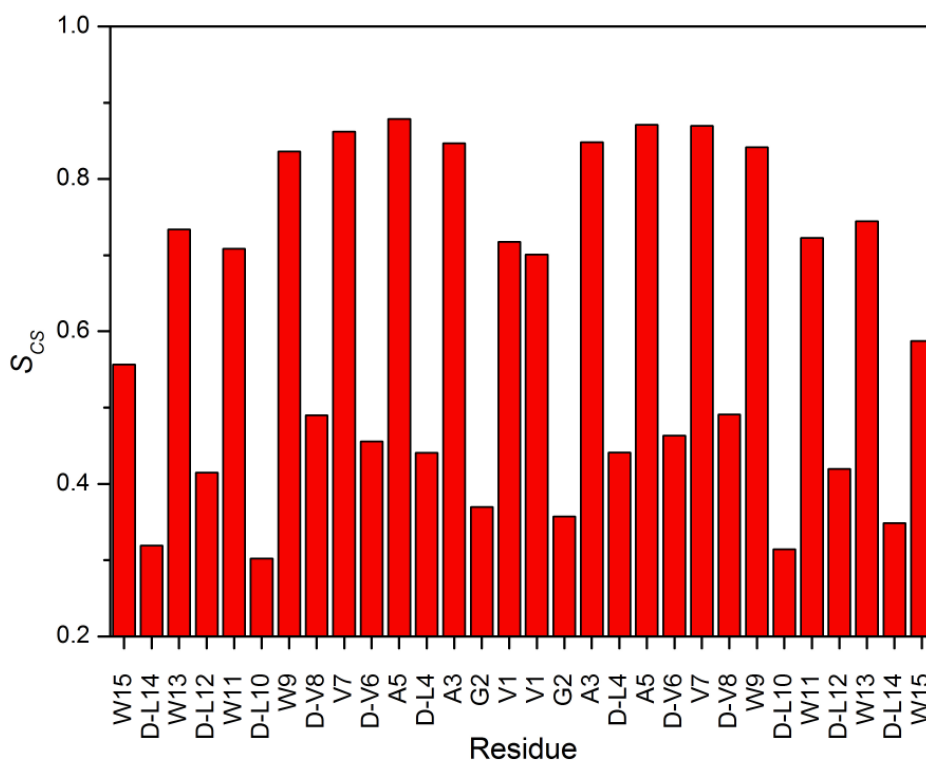


Figure 3.18: ^{15}N CS tensor derived order parameters in gramicidin A dimer.

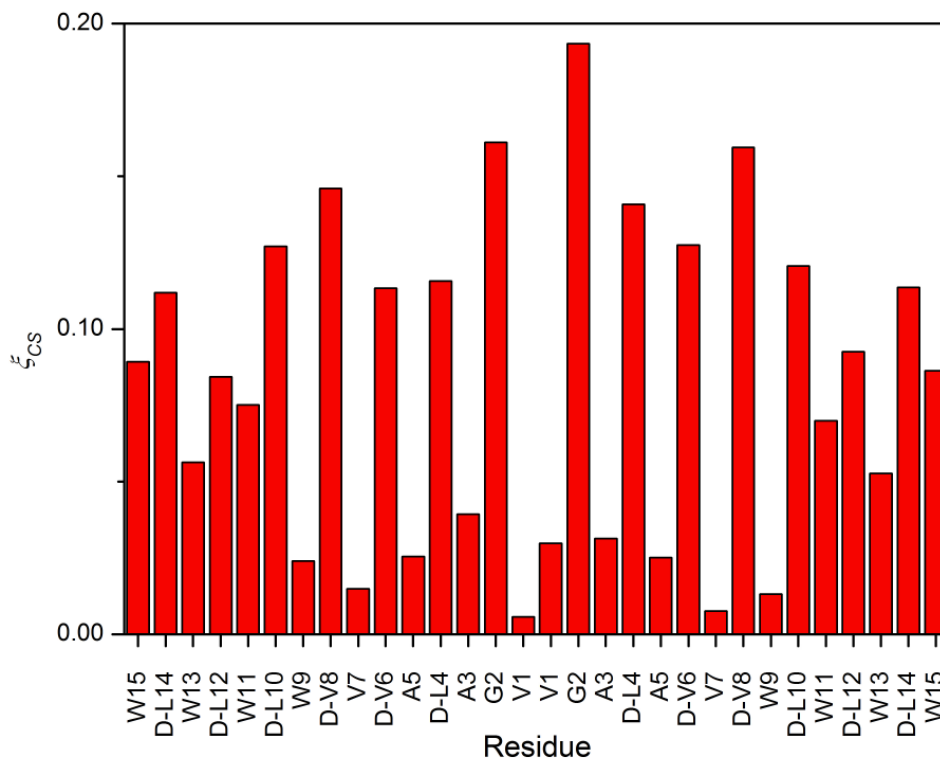


Figure 3.19: ^{15}N CS tensor derived biaxiality values in gramicidin A dimer.

motions and therefore backbone dynamics of gramicidin A might be directly connected to the mechanism of ion translation through the channel⁹⁰. North and Cross proposed three mechanisms for the propagation of the correlated motions⁹⁰: propagation via covalent bonds to the C_α carbon of the adjacent peptide planes, propagation via hydrogen bonds between the distant residues (hydrogen bonds connect every sixth peptide plane) and propagation via electrostatic interactions with the water molecules in the channel.

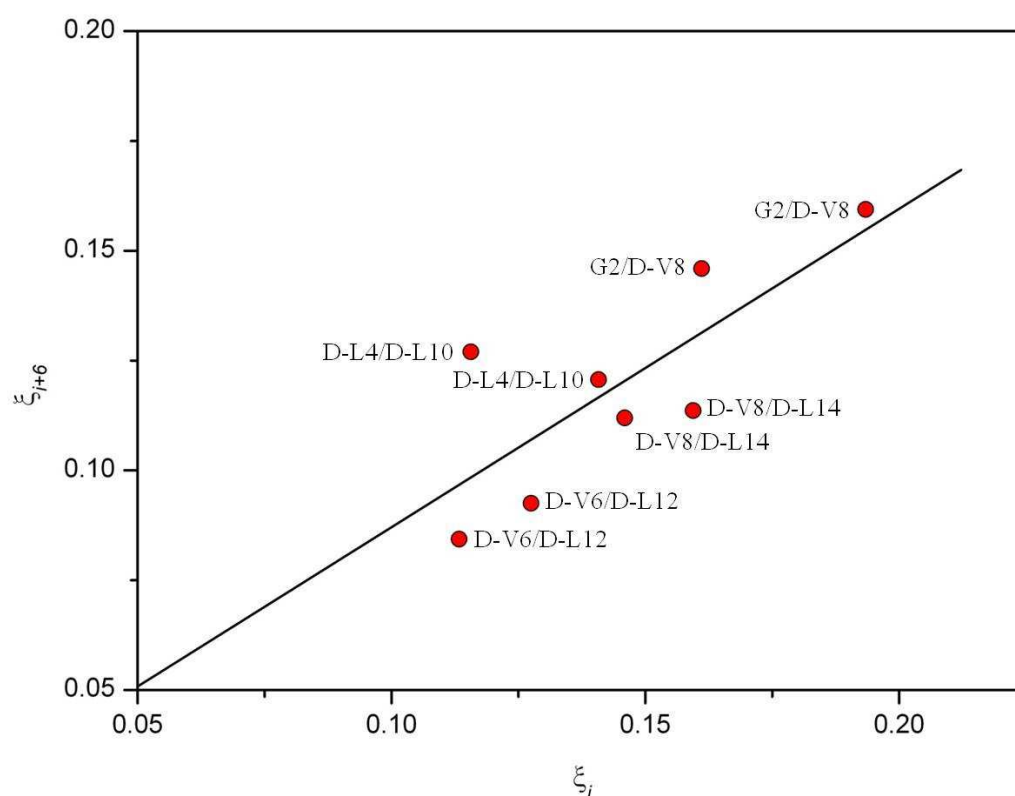


Figure 3.20: Biaxiality correlation (correlation coefficient R of 0.77) between the amino acid i and the amino acid $i+6$ observed for D-amino acids and glycine in gramicidin A dimer backbone (black line represents the linear fit between the biaxiality values). Correlation coefficient R of 0.93 can be achieved if one of the data points for D-L4 and D-L10 is omitted.

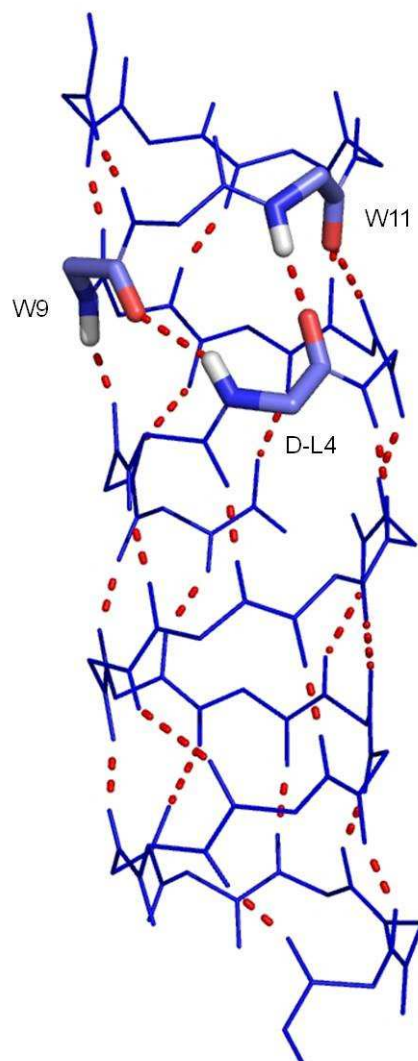


Figure 3.21: Hydrogen bonds in the backbone of gramicidin A dimer. A minimal unit consisting of three hydrogen bonded amino acids (an example is shown as “sticks”) is required for the propagation of correlated motions.

The biaxiality values ξ_{CS} obtained from the BPT ^{15}N CS tensor calculations on the snapshots of molecular coordinates from the trajectory of the MD simulation with NMR orientational constraints provide the experimental evidence that correlated motions are present in the backbone of gramicidin A (cf. figure 3.20). The repetitive pattern observed for the biaxiality values of D-amino acids and the residue Gly 2 does not merely indicate the presence of correlated motions but also sheds light on the mechanism of their propagation. The periodicity of the repetitive pattern (six residues) implies the propagation via hydrogen bonds as suggested by North and Cross⁹⁰. A thorough inspection of gramicidin A structure reveals that the residues of the gramicidin A are hydrogen bonded to their fifth ($\text{HN}_i \cdots \text{O}_{i+5}$) and their seventh

($O_i \cdots HN_{i+7}$) neighbours (cf. figure 3.21; note that North and Cross erroneously equate hydrogen bonding to every sixth peptide plane with hydrogen bonding to the sixth neighbouring residue in their discussion – this is not necessarily the same and is not the case for the gramicidin A structure). Thus the mechanism for the propagation of the correlated motions involves both the hydrogen and the covalent bonds. The motion of the D-amino acid (or Gly 2) with the residue number i is transmitted to the L-amino acids with the residue numbers $i+5$ and $i+7$. The conformational space of L-amino acids is smaller than those of D-amino acids or glycine and therefore they transmit the mechanical torque to their D-configured neighbours and the biaxiality values describing the motion of the residues i and $i+6$ appear correlated. Assuming that the hydrogen bonds to the terminal formyl and ethanolamine moieties of gramicidin A (no experimental data available and therefore not shown in figure 3.19) participate in this mechanism, the propagation of the correlated motions is possible throughout the whole gramicidin A dimer.

In addition to local fluctuations, order parameters S_{CS} obtained from the BPT ^{15}N CS tensors allow the interpretation of gramicidin A dynamics in lipid bilayers in terms of wobbling-in-a-cone model (cf. figure 3.22). The orientation of gramicidin A fluctuates around the membrane normal and the amplitude of motional fluctuations increases towards the bilayer surface. N-terminal Val 1 and Gly 2 residues responsible for head-to-head dimerization of gramicidin A also exhibit higher mobility. The mobility of D-amino acids and glycine is significantly higher than the mobility of L-amino acids, thus reflecting their higher conformational flexibility (cf. figure 3.3). In addition biaxiality values of D-amino acids and glycine indicate that these residues exhibit correlated motions orthogonal to the membrane normal.

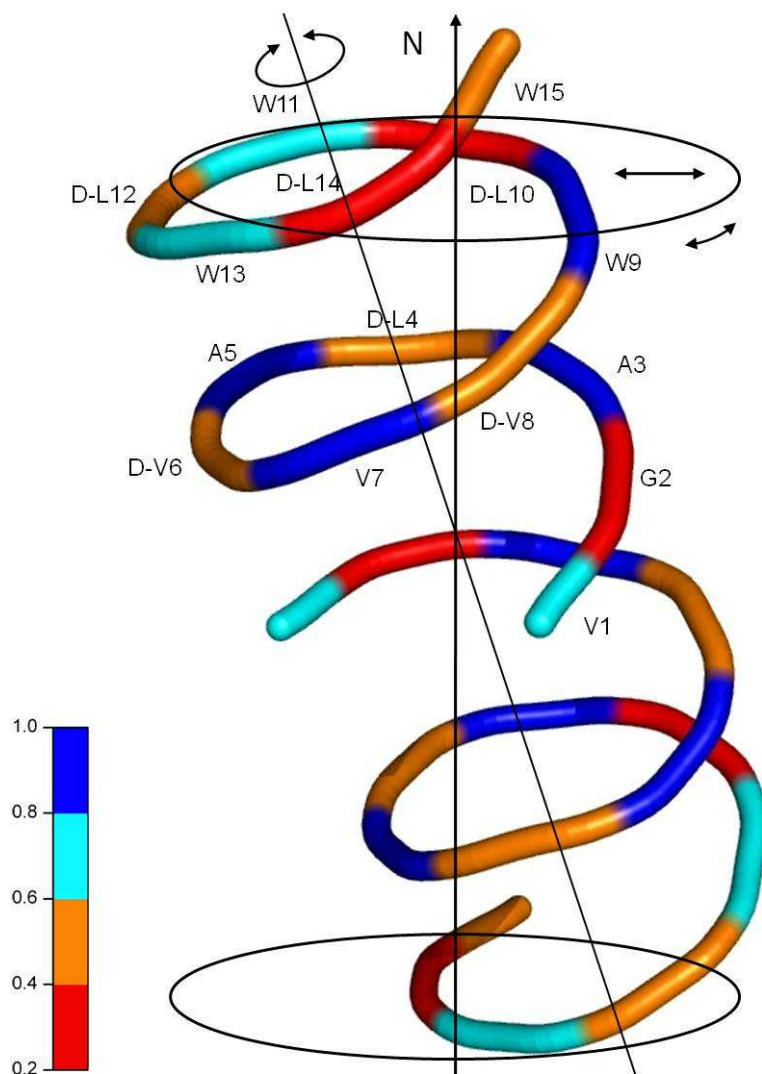


Figure 3.22: Distribution of local order parameters in the gramicidin A backbone (colors indicate the size of the ^{15}N CS tensor derived order parameters). Arrows indicate the directions of possible backbone fluctuations in gramicidin A aligned in lipid bilayers.

3.9 PISEMA simulation for nontrivial cases

The information content of the PISEMA spectrum can be described as the distribution of signal intensity in two dimensions: the dipolar coupling and the CS dimension. For a simulation the area of the PISEMA spectrum can be divided into pixels with each pixel specified by a triple consisting of dipolar coupling coordinate, CS coordinate and the corresponding intensity. To avoid an infinite number of pixels the coordinates of a single pixel are given as dipolar coupling and CS ranges and the pixels appear as quadratic bins. The intensity of the pixel is then given by the number of the

calculated signal resonances (pairs of dipolar coupling and CS values) lying within the bin.

Molecular coordinate snapshots obtained in MD simulation with NMR orientational constraints were used for the simulation of the gramicidin A PISEMA spectrum with quantitative accuracy. 2000 MD snapshots were subjected to the ^{15}N CS driven geometry optimization with the following ^{15}N CS tensor calculation. Averaged dipolar coupling and ^{15}N CS tensor zz components were computed using the exponential memory decay function given in equation 3.3 for each time step (i.e. for each snapshot) of the MD simulation resulting in a set of 2000 pairs of dipolar coupling and chemical shift values for each residue. The calculated set of values was subjected to a binary count. The range of chemical shift bins was set from 99 to 241 ppm and the CS bin width was set to 2 ppm yielding 30 CS bins. The range of dipolar coupling bins was set from 10 to 25 kHz and the dipolar coupling bin width was set to 0.5 kHz yielding 30 dipolar coupling bins. This procedure corresponds to the decomposition of the PISEMA spectrum into 900 square bins (pixels) within the spectral range specified above. The results of the binary count (number of “hits” lying within the bins or, with other words, signal intensities) were visualized as a contour plot. As a first attempt to consider the fact that very weak signals cannot be distinguished from the background noise all signals with intensity four times smaller than the maximum observed intensity were not visualized (signal-to-noise ratio of 4). For simplicity no attempts were made to include relaxation effects or any other phenomena in the simulation of line shape and the distance between the contour lines was arbitrary set to 40 (the number refers to absolute number of “hits” in the bin describing the intensity). Binary count and visualization were performed with the program Mathematica 5.0 (Wolfram Research, Inc., Champaign, IL, 2003).

The simulated PISEMA spectrum was compared with the values obtained by Ketchum et al. with single site selective ^{15}N labeling at room temperature in DMPC bilayers⁷⁶ (cf. figure 3.23). In marked contrast to PISEMA spectra computed on the static gramicidin A structures, the PISEMA spectrum computed on the MD trajectory resembles the experimental signal pattern. However, the resonances in the

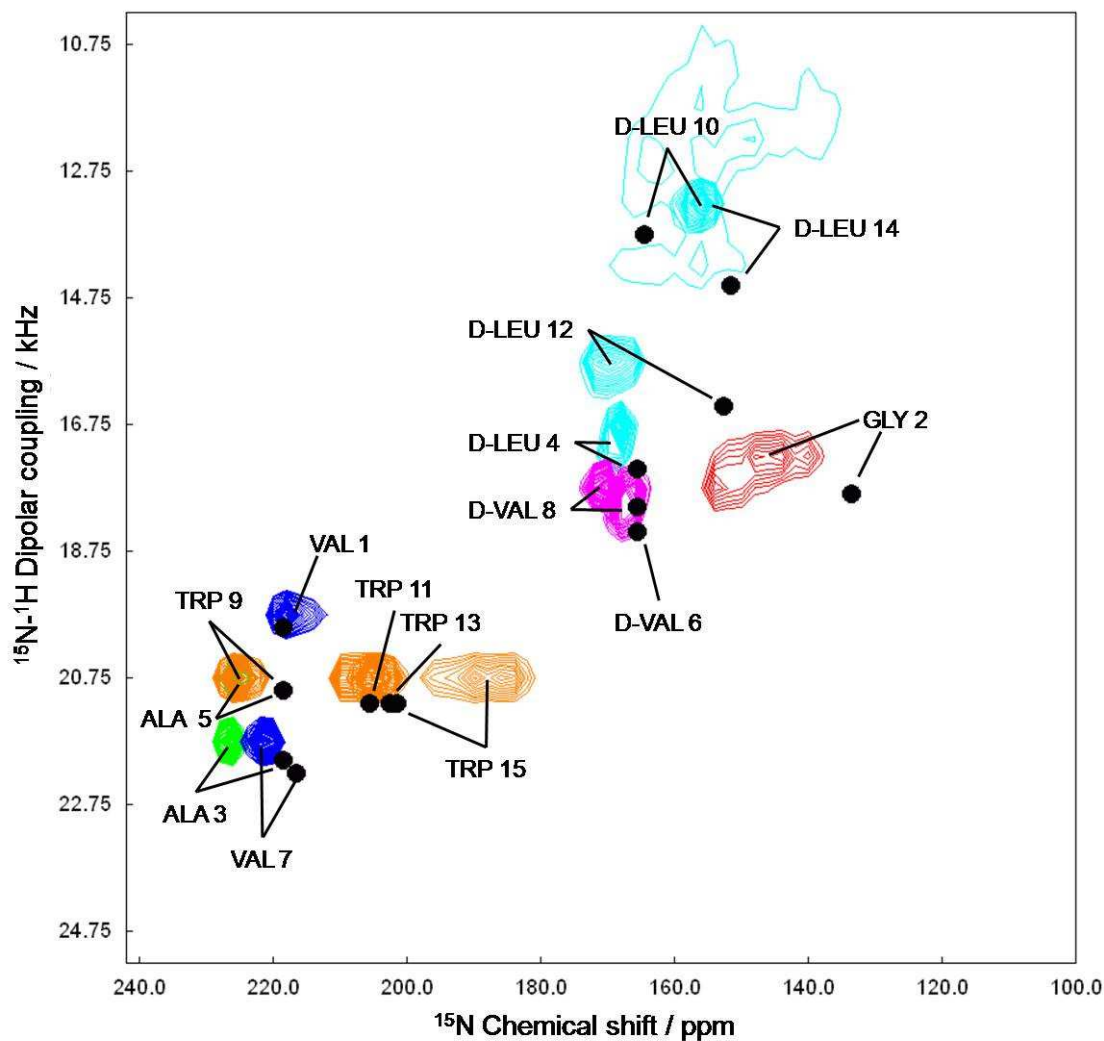


Figure 3.23: PISEMA spectrum of gramicidin A simulated using orientational MD and BPT ^{15}N CS tensor computation (contours of different amino acids are shown in different colours: ALA in green, GLY in red, TRP in orange, VAL in blue, D-VAL in magenta and D-LEU in cyan; black solid circles denote the experimental values of Ketchem et al.⁷⁶).

computed PISEMA spectrum are very broad. The signals of D-Leu 10 and D-Leu 14 cannot be resolved in the calculated spectrum. The calculated signals of D-Leu 12 and Gly 2 are shifted downfield from their experimental positions in the chemical shift dimension.

Homonuclear decoupling is applied in the indirect dimension of the PISEMA experiment leading to the scaling of the observed dipolar couplings. In the ideal case the scaling factor is 0.82 for the FSLG decoupling as applied by Vogt et al., however,

the scaling factor is strongly dependent on the experimental settings. The comparison of the PISEMA spectrum of Vogt et al. with the values measured by Ketchem et al. with single site selective ^{15}N labeling implies the scaling factor of approximately 0.25. Therefore for the comparison of the simulated PISEMA spectrum with the measured PISEMA spectrum of Vogt et al. all calculated dipolar couplings were scaled by the factor of 0.25 and the simulated spectrum was recalculated with the range of dipolar coupling bins set from 0.5 to 6.5 kHz.

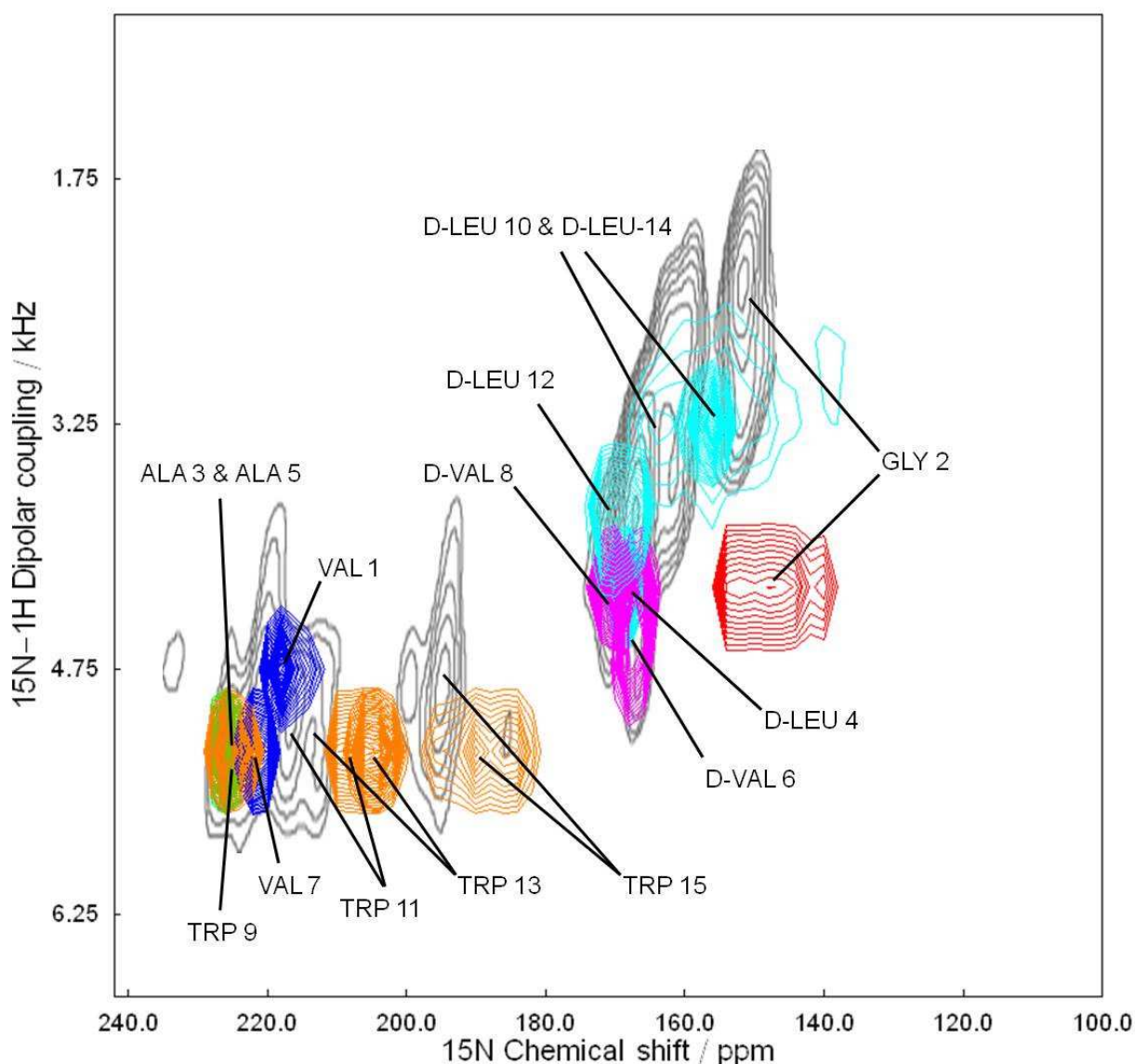


Figure 3.24: Simulated PISEMA spectrum of gramicidin A with dipolar couplings scaled by the factor of 0.25 (contours of different amino acids are shown in different colours: ALA in green, GLY in red, TRP in orange, VAL in blue, D-VAL in magenta and D-LEU in cyan) compared to the PISEMA spectrum measured by Vogt et al.⁷⁷ (contours are shown in gray; data courtesy of Prof. Burkhard Bechinger).

The simulated PISEMA spectrum is in good agreement with the measured spectrum (cf. figure 3.24). The broad lines predicted in the simulated spectrum can be observed in the measured spectrum confirming that the primary reason of the signal broadening is not relaxation or insufficient quality of alignment but the intrinsic mobility in the peptide backbone. It is noteworthy that compared to the spectrum obtained by Vogt et al.⁷⁷ (measured in DMPC bilayers at 310 K) the calculated signals of D-Leu 12 and Gly 2 have the right position in the chemical shift dimension but Gly 2 exhibits the deviation of ca. 2 kHz in the dipolar coupling dimension. The calculated signals of Trp 11, 13 and 15 are slightly shifted upfield from their experimental positions in the chemical shift dimension. Thus the comparison of the simulated spectrum with data obtained by Ketchem et al.⁷⁶ and Vogt et al.⁷⁷ demonstrates that the accuracy of the simulation resembles the accuracy of experimental measurements.

Good agreement between the simulated and the experimental PISEMA spectra confirms that molecular dynamics with orientational constraints (applied to obtain the simulated spectrum) yields a realistic description of gramicidin A mobility. It becomes obvious that the prediction of the PISEMA spectra of non- α -helical molecules with high intrinsic mobility requires explicit inclusion of molecular dynamics. If no experimental data describing the molecular dynamics are available, a priori no reliable prediction of PISEMA spectra is possible and expensive and labour-intensive selective labeling is inevitable for spectral assignment. However, the methodology presented in this section allows to validate the assignment of PISEMA spectra by their simulation. If the signals of several residues can be assigned using selective labeling, the position of other signals could be deduced using a trial and error assignment strategy based on the comparison of calculated and experimental spectra. In this way the required amount of selectively labeled peptides might be reduced in future investigations.

4. Dynamics of solid-state proteins

4.1 CSA order parameters

Protein dynamics plays a crucial role in various biological functions, such as enzyme catalysis, allosteric regulation and ion transport by ion channel proteins⁹³. These processes are often associated with large-amplitude internal motions of protein residues on the time scales of microseconds to milliseconds. ^{15}N chemical shift anisotropy is sensitive to protein dynamics up to the millisecond time scale². In solid-state NMR, motional averaging of the chemical shift tensors results in the narrowing of the observed chemical shift tensor span (cf. figure 4.1). Furthermore ^{15}N chemical shift tensors are very sensitive to electrostatic effects (cf. section 2.4) and depend on all the atoms of the corresponding residues and their surroundings. Therefore observed ^{15}N chemical shift tensors are less influenced by the ^1H - ^{15}N bond vibrational motion (pico- to nanosecond timescales) and predominantly reflect molecular motions of the backbone⁸⁷.

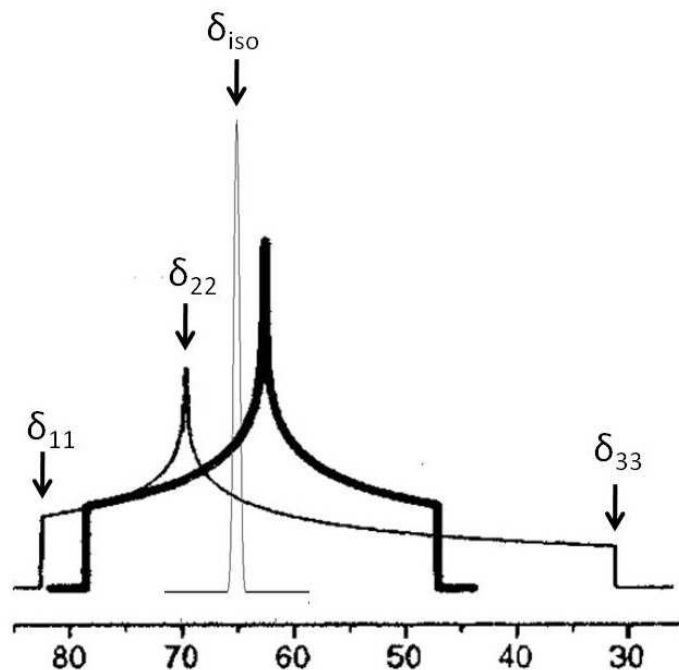


Figure 4.1: Example of an NMR powder spectrum (thin black line – powder spectrum of an immobile nucleus thick black line – motionally averaged powder spectrum).

Quantum chemical calculations yield ^{15}N chemical shift tensors of an immobilized “frozen” protein structure. Motionally averaged protein ^{15}N chemical shift tensors can be experimentally determined⁹⁴. The comparison of the calculated and the experimental ^{15}N chemical shift tensors allows to estimate the mobility of the corresponding residue. Two parameters describe the motional averaging of the chemical shift tensors: the chemical shift tensor span indicating the separation of the δ_{33} and δ_{11} CS tensor principal components (cf. equation 1.16), and the chemical shift (reduced) anisotropy (cf. equations 1.19 and 1.20) indicating the separation of the δ_{33} principal component and the isotropic chemical shift value. Both parameters refer to the linewidth of the corresponding powder spectrum and their evaluations would yield identical results if all CS tensor principal components are influenced in the same way by the molecular motion. However, motional averaging of the chemical shift tensors depends not only on the amount but also on the direction of motion (e.g. axially symmetric tensors can be observed in the presence of rotations around the threefold symmetry axis, cf. section 1.2). Therefore the same influence of the molecular mobility on all CS tensor principal components cannot be assumed a priori and the evaluation of the chemical shift tensor span, thus explicitly considering both δ_{33} and δ_{11} principal components, appears to be a more adequate approach.

Order parameters describing the molecular mobility can be derived from the comparison of the experimental and the calculated chemical shift tensor span analogously to the definition of the order parameters for dipolar couplings given in equation 1.28:

$$S_{\Omega} = \frac{\Omega_{\text{exp}}}{\Omega_{\text{calc}}} = \frac{\delta_{11}^{\text{exp}} - \delta_{33}^{\text{exp}}}{\delta_{11}^{\text{calc}} - \delta_{33}^{\text{calc}}}. \quad (4.1)$$

The approach expressed in equation 4.1 relies on the anisotropic nature of the nuclear shielding in the solid-state. That is why order parameters derived in equation 4.1 are referred to as the chemical shift anisotropy order parameters, despite the fact that they are obtained from the CS tensor span.

Until now, ^{15}N chemical shift tensors were measured only for two proteins, thioredoxin and immunoglobulin-binding domain of protein G (GB1). CSA order

parameters were derived for both proteins using BPT ^{15}N CS tensor calculations and their molecular mobility is analyzed in the following sections.

4.2 Application to thioredoxin

E.coli thioredoxin is a 108-amino acid residue disulfide reductase playing an important role in many metabolic pathways⁹⁵. Complementary fragments of *E.coli* thioredoxin obtained by proteolytic cleavage at a specific position can spontaneously reassemble in solution⁹⁶. This property of thioredoxin allows selective, uniform ^{13}C and ^{15}N labeling of thioredoxin fragments and simplifies the interpretation of thioredoxin NMR spectra. Microcrystalline oxidized thioredoxin reassemblies (fragments containing the residues 1 to 73 and 74 to 108, cf. figure 4.2) were extensively studied by solid-state NMR. ^{15}N and ^{13}C isotropic chemical shifts assignments are available for both thioredoxin fragments^{96,97} and ^{15}N chemical shift tensors were measured in the thioredoxin fragment containing the residues 1 to 73⁹⁸. In addition molecular dynamics of the thioredoxin fragment containing the residues 1 to 73 was subjected to a detailed investigation including spin-lattice relaxation time measurements and the determination of the ^1H - ^{15}N dipolar order parameters⁹⁸. Mobility distributions in thioredoxin backbone follow a similar pattern on different time scales and therefore thioredoxin is a good model system for the evaluation of the CSA order parameters.

Unfortunately, a solid-state NMR thioredoxin structure is not yet available. However, backbone dihedral angles predicted from the solid-state NMR chemical shifts do not exhibit significant differences to those of the liquid-state thioredoxin structure⁹⁷. All structures of the liquid-state thioredoxin structure ensemble⁹⁹ (oxidized form, PDB ID 1XOA) were subjected to the chemical shift driven geometry optimization with amide backbone ^{15}N and $^{13}\text{C}_\alpha$ and $^{13}\text{C}_\beta$ isotropic chemical shifts measured in solid-state thioredoxin reassemblies^{96,97} with the subsequent ^{15}N CS tensor calculation. BPT calculated ensemble averaged CS values exhibit adequate correlation with the experimental chemical shifts (cf. table 4.1 and figures 4.3 and 4.4). Next, ensemble averaged BPT ^{15}N CS tensor spans were computed and CSA order parameters were derived for all thioredoxin residues with experimentally determined ^{15}N CS tensors.

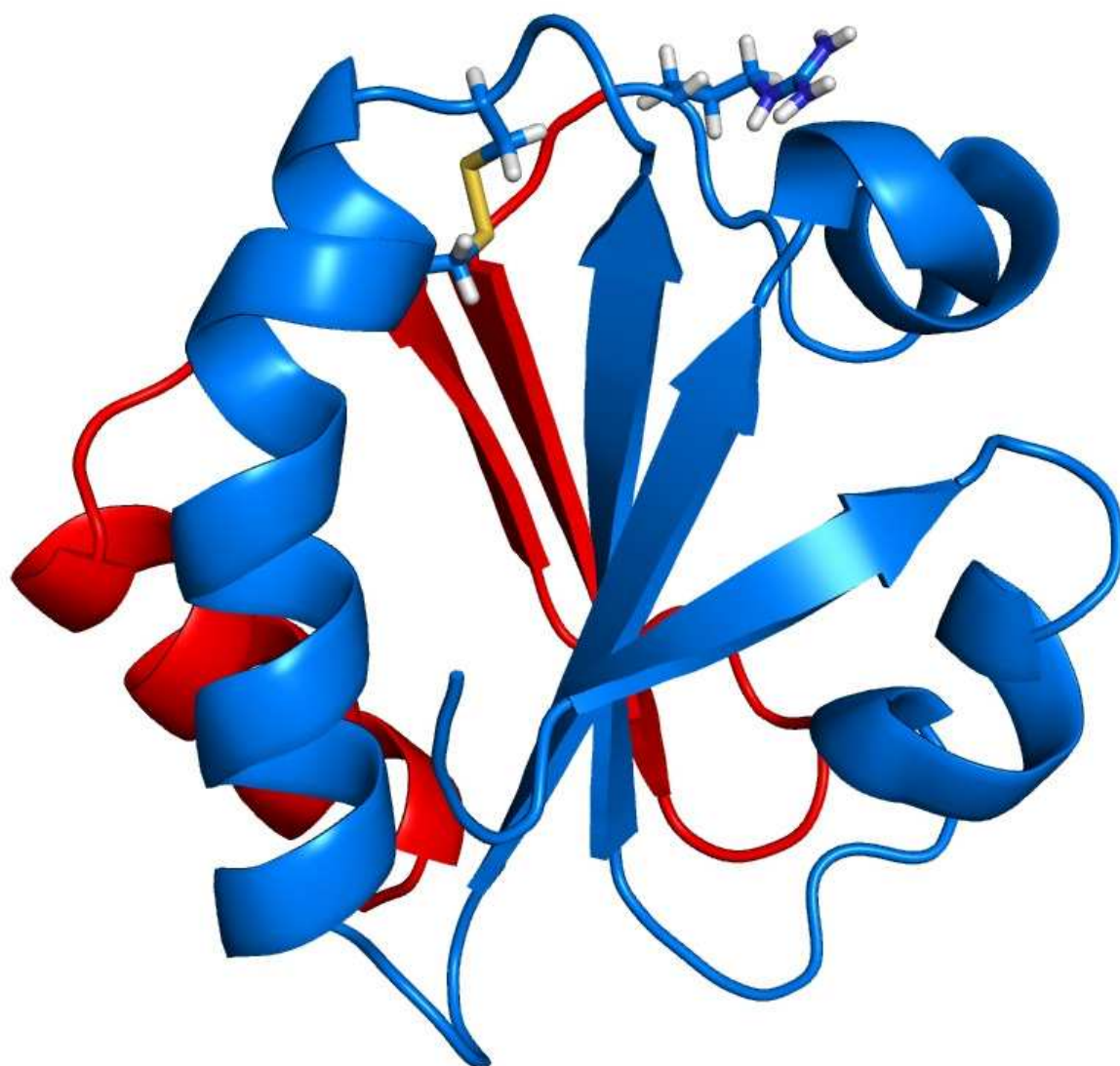


Figure 4.2: Liquid-state structure of thioredoxin (oxidized form, PDB ID 1XOA) in ribbon representation (fragment containing the residues 1 to 73 is shown in blue, fragment containing the residues 74 to 108 is shown in red; side chain of the cleavage site of the solid-state thioredoxin reassemblies (Arg 73) and disulfide bond between Cys 32 and Cys 35 are shown as “sticks”).

Table 4.1: Correlation between the BPT calculated and the experimental chemical shifts in thioredoxin before and after chemical shift driven structure optimization (SD and MAE are given in ppm).

Nucleus	1XOA			optimized 1XOA		
	R	SD	MAE	R	SD	MAE
$^{13}\text{C}_\alpha$	0.79	3.6	2.9	0.94	1.8	1.4
$^{13}\text{C}_\beta$	0.94	4.4	3.5	0.97	3.3	2.3
^{15}N	0.54	11.6	9.0	0.95	3.3	2.8

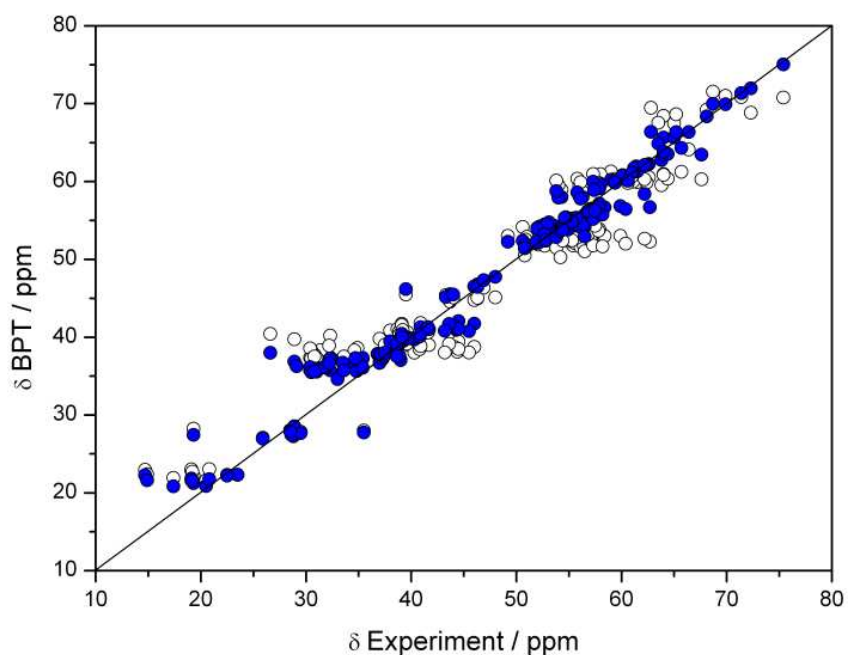


Figure 4.3: Correlation between experimental and BPT calculated $^{13}\text{C}_\alpha$ and $^{13}\text{C}_\beta$ chemical shifts of thioredoxin before (open black circles) and after (blue solid circles) chemical shift driven geometry optimization.

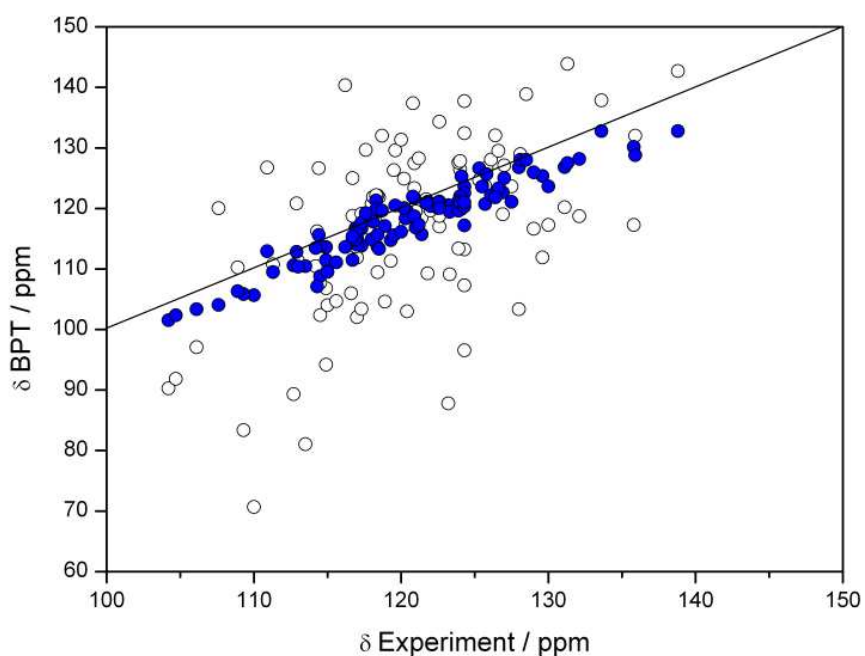


Figure 4.4: Correlation between experimental and BPT calculated ^{15}N chemical shifts of thioredoxin before (open black circles) and after (blue solid circles) chemical shift driven geometry optimization.

As expected, BPT computed ^{15}N chemical shift tensor spans, corresponding to the immobile thioredoxin structure, are systematically larger than experimentally measured⁹⁸, motionally averaged ^{15}N CS tensor spans (cf. figure 4.5; mean span values of 195.6 ± 13.1 ppm vs. $148.3.0 \pm 17.1$ ppm can be observed; note that the values of 13.1 and 17.1 ppm do not refer to the experimental errors but to the standard deviation of the CS tensor span values in thioredoxin). Calculated ^{15}N CS tensor spans of the residues Asp 2 (calculated span value of $\Omega=230.8$ ppm), Ala 29 ($\Omega=214.1$ ppm) and of all proline residues ($\Omega=234.8 \pm 23.9$ ppm; the value of 23.9 ppm does not refer to the experimental error but to the standard deviation of the calculated proline CS tensor spans) are significantly larger than those of the other residues. In proline residues this trend can be attributed to the molecular topology (tertiary amide backbone nitrogen atom). The conformation of the terminal residue

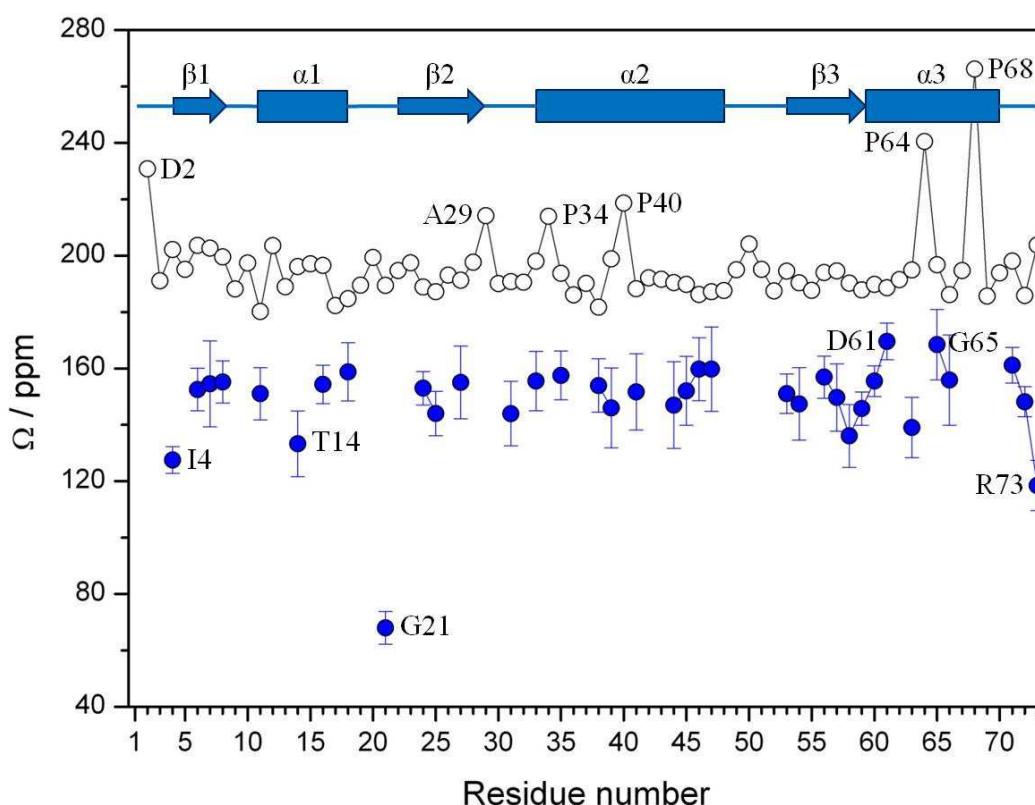


Figure 4.5: Experimental (solid blue circles) and BPT calculated (open black circles) ^{15}N CS tensor spans of the thioredoxin fragment containing the residues 1 to 73. The secondary structure (blue arrows denote the β -sheets, blue rectangles denote α -helices) is defined on the basis of the liquid-state NMR thioredoxin structure (PDB ID 1XOA).

Asp 2 is disordered throughout the structure ensemble and an unusually large calculated ^{15}N CS tensor span might result from the atypical backbone geometry (e.g. positive φ torsion angles). However, there is no evident reason for the large ^{15}N CS tensor span of the residue Ala 29.

Unfortunately, not all resonances could be resolved in experimental ^{15}N CS tensor measurements⁹⁸. However, the amount of observed resonances is sufficient for the description of molecular mobility in the thioredoxin fragment containing the residues 1 to 73. At first glance, the magnitude of the experimental ^{15}N CS tensors allows to recognize the most mobile (Ile 4 with $\Omega=202.2\pm 4.8$ ppm, Thr 14 with $\Omega=133.3\pm 11.7$ ppm, Gly 21 with $\Omega=68.0\pm 5.8$ ppm, Arg 73 with $\Omega=118.5\pm 8.9$ ppm) and the most immobile (Asp 61 with $\Omega=169.6\pm 6.5$ ppm, Gly 65 with $\Omega=168.5\pm 12.5$ ppm) residues using the assumption that smaller CS tensor spans always correspond to a higher degree of the motional tensor span narrowing (cf. figure 4.5). A more accurate description of the molecular mobility is obtained if the differences between the experimental and the calculated ^{15}N CS tensor spans are examined (cf. figure 4.6). These differences reveal that the mobility of the residue Gly 65 ($\Delta\Omega=28.2\pm 12.5$ ppm) is not significantly lower than those of the other residues (mean $\Delta\Omega=44.0\pm 18.8$ ppm can be observed; the value of 18.8 ppm does not refer to the experimental error but to the standard deviation of the $\Delta\Omega$ values in thioredoxin). The example of the residue Gly 65 demonstrates that the results obtained using only the experimental ^{15}N CS tensors can be elusive. For a reliable comparability the differences between the experimental and the calculated ^{15}N CS tensor spans should be related to the calculated tensor spans, thus CSA order parameters should be calculated. CSA order parameters describing the molecular mobility of thioredoxin (cf. figure 4.7) confirm the tendencies already observed from the differences between the experimental and the calculated ^{15}N CS tensor spans.

CSA order parameters can be compared to the spin-lattice relaxation rates and to the ^1H - ^{15}N dipolar order parameters measured in the microcrystalline thioredoxin fragment containing residues 1 to 73⁹⁸. Spin-lattice relaxation rates provide information on the molecular mobility on the pico- to nanosecond time scale (cf. figure 4.8). Relatively high spin-lattice relaxation rates (cf. figure 4.9) are observed for

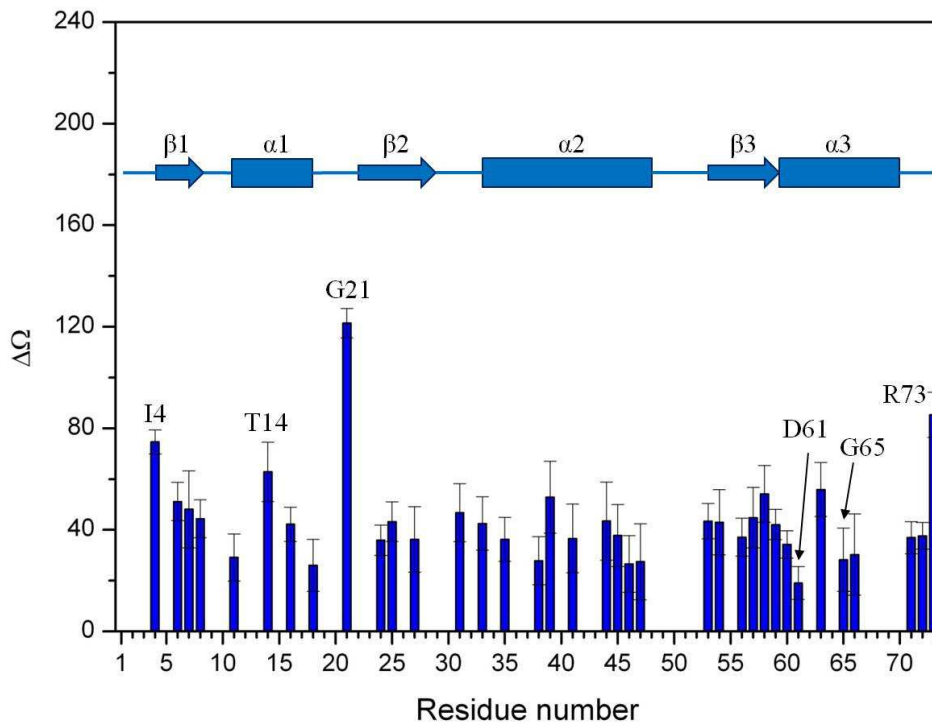


Figure 4.6: Differences between experimental and BPT calculated ^{15}N CS tensor spans in the thioredoxin fragment containing residues 1 to 73.

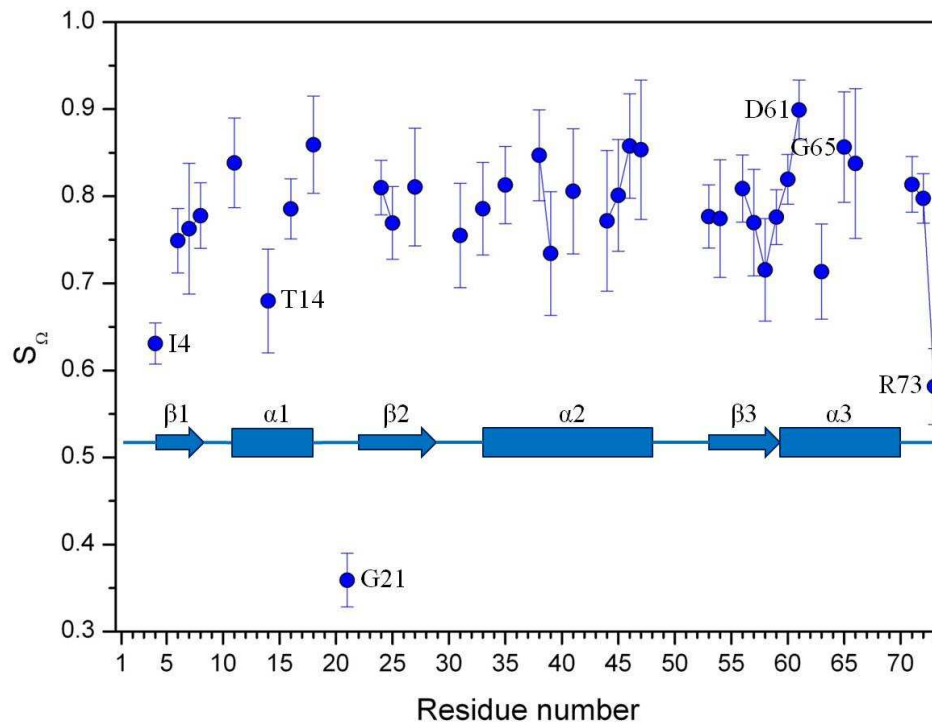


Figure 4.7: ^{15}N CSA order parameters in the thioredoxin fragment containing residues 1 to 73.

the residues Ile 4 ($R_1 = 0.051 \pm 0.012 \text{ s}^{-1}$; mean relaxation rate of $R_1 = 0.037 \pm 0.025 \text{ s}^{-1}$ can be observed in thioredoxin (calculated for all residues with available data excluding the residue Gly 21); note that the value of 0.025 s^{-1} does not refer to the experimental error but to the standard deviation of the spin-lattice relaxation rates in thioredoxin), Thr 14 ($R_1 = 0.053 \pm 0.004 \text{ s}^{-1}$), Gly 21 ($R_1 = 0.636 \pm 0.167 \text{ s}^{-1}$), Gly 33 ($R_1 = 0.070 \pm 0.019 \text{ s}^{-1}$), Thr 66 ($R_1 = 0.063 \pm 0.004 \text{ s}^{-1}$) and Arg 73 ($R_1 = 0.150 \pm 0.100 \text{ s}^{-1}$). Four of these residues (Ile 4, Thr 14, Gly 21 and Arg 73) exhibit relatively low CSA order parameters (0.63 ± 0.02 , 0.68 ± 0.06 , 0.36 ± 0.03 and 0.58 ± 0.04 , respectively; the mean order parameter in thioredoxin is 0.77 ± 0.10 ; the value of 0.10 does not refer to the experimental error but to the standard deviation of the order parameters). Thus high mobility is observed for the residues Ile 4, Thr 14, Gly 21 and Arg 73 on both the picoseconds to nanosecond and microsecond to millisecond time scales. This is in line with the observation of Yang et al.⁹⁸ that fast librations in thioredoxin backbone modulate the slower motions. An exception to this trend is observed for the residue Gly 65 ($R_1 = 0.043 \pm 0.008 \text{ s}^{-1}$ vs. order parameter of $S_\Omega = 0.86 \pm 0.06$) which exhibits a slightly above-average molecular mobility on the pico- to nanosecond time scale but is relatively immobile on the microsecond to millisecond time scale.

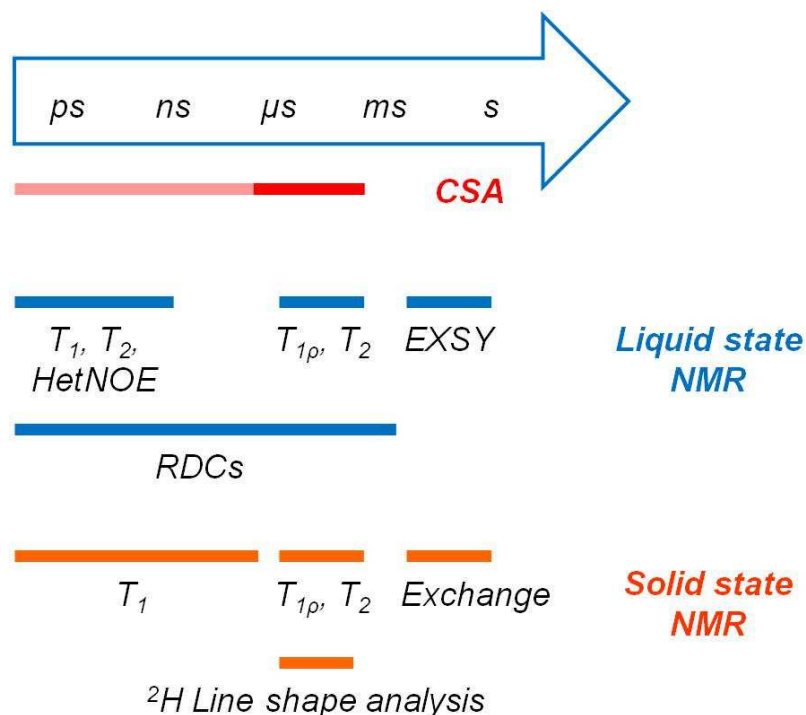


Figure 4.8: Time scale of motions detected by NMR.

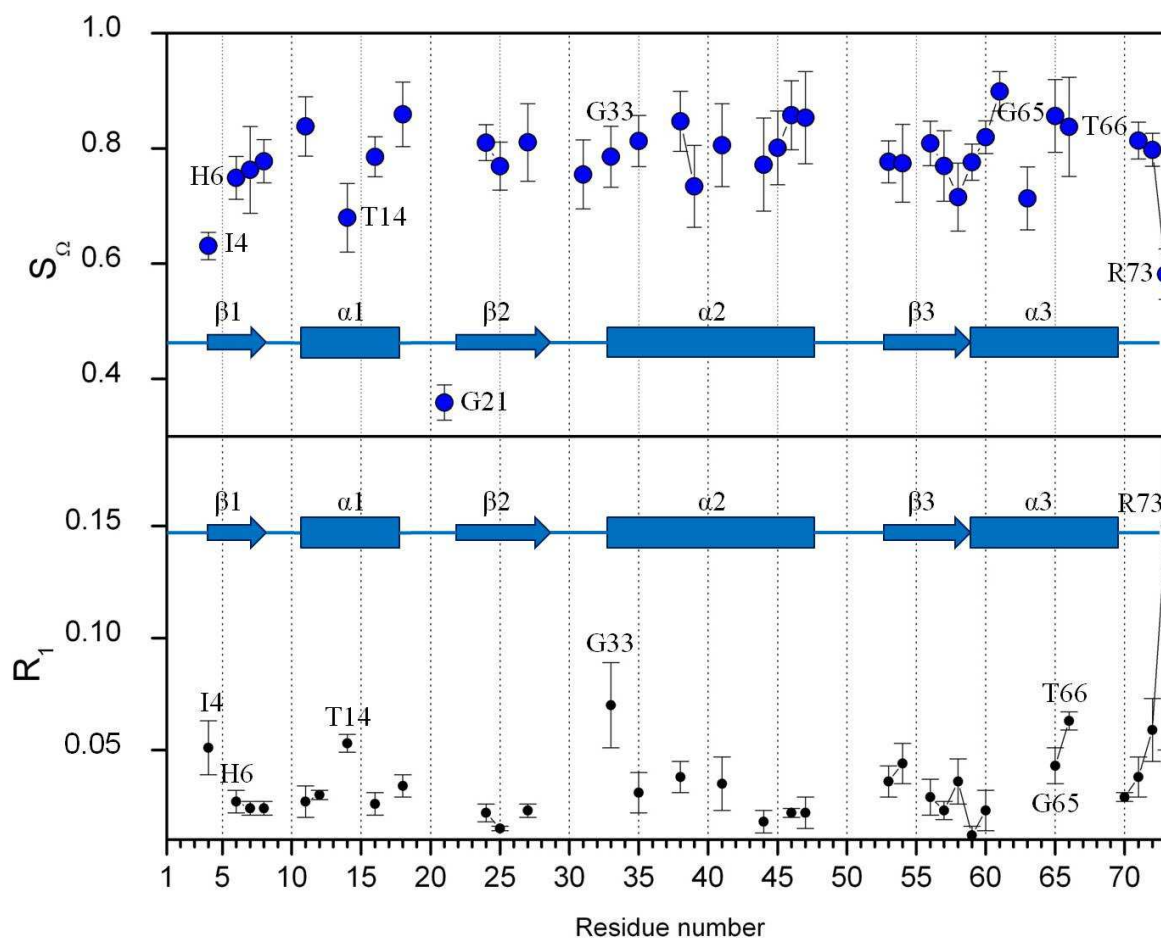


Figure 4.9: ^{15}N CSA order parameters (at the top) and spin-lattice relaxation rates⁹⁸ (given in s^{-1} ; at the bottom; relaxation rate of the residue Gly 21 ($0.636 \pm 0.167 \text{ s}^{-1}$) is out of the plot range) in the microcrystalline thioredoxin fragment containing residues 1 to 73.

^1H - ^{15}N dipolar order parameters provide information on the molecular mobility on the time scale from picoseconds to milliseconds¹⁰⁰. Therefore ^1H - ^{15}N dipolar order parameters are strongly influenced by the ^1H - ^{15}N bond motions on the pico- to nanosecond time scale and do not necessarily reflect the slower mobility of the whole residues. Relatively low ^1H - ^{15}N dipolar order parameters are observed for the residues Ile 4 (order parameter of $S_d = 0.85 \pm 0.05$; mean ^1H - ^{15}N dipolar order parameter of $S_d = 0.89 \pm 0.10$ can be observed in thioredoxin; note that the value of 0.10 does not refer to the experimental error but to the standard deviation of the spin-lattice relaxation rates in thioredoxin), His 6 ($S_d = 0.84 \pm 0.05$), Thr 14 ($S_d = 0.84 \pm 0.04$), Gly 21 ($S_d = 0.35 \pm 0.11$), Gly 33 ($S_d = 0.83 \pm 0.05$), Thr 66 ($S_d = 0.82 \pm 0.05$) and Arg 73 ($S_d = 0.85 \pm 0.05$; cf. figure 4.10). These residues (excluding His 6) also exhibit relatively high spin-lattice relaxation rates. For the

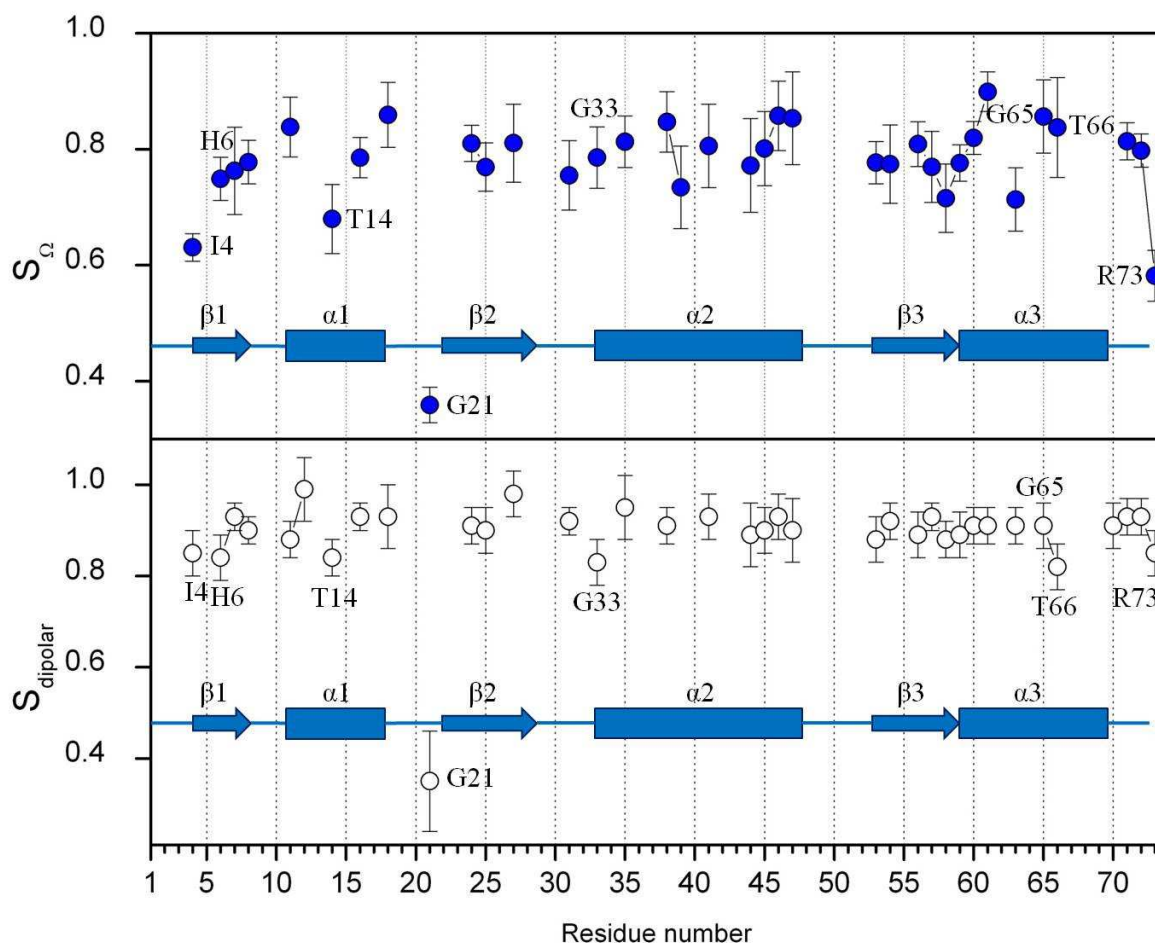


Figure 4.10: ^{15}N CSA order parameters (at the top) and ^1H - ^{15}N dipolar order parameters⁹⁸ (at the bottom) in the microcrystalline thioredoxin fragment containing the residues 1 to 73.

residues Ile 4, Thr 14, Gly 21 and Arg 73 high mobility on the microsecond to millisecond time scale can be confirmed by the CSA order parameters. Thus, as already stated above, the mobility of the residues Ile 4, Thr 14, Gly 21 and Arg 73 is above average on the pico- to nanosecond time scale as well as on the micro- to millisecond time scale. Residues Gly 33 and Thr 66 exhibit above-average mobility on the pico- to nanosecond time scale and their mobility obviously has a strong impact on their ^1H - ^{15}N dipolar order parameters. However, CSA order parameters reveal that the residues Gly 33 ($S_{\Omega} = 0.79 \pm 0.05$) and Thr 66 ($S_{\Omega} = 0.84 \pm 0.09$) do not exhibit high mobility on the micro- to millisecond time scale. Residue His 6 exhibits neither a high spin-lattice relaxation rate nor a low CSA order parameter. Therefore the motions of this residue probably take place on the time scale slower than those described by the CSA order parameters.

It can be summarized that information provided by the CSA order parameters (cf. figure 4.11) is in good agreement with previous investigations on the thioredoxin dynamics⁹⁸. ^{15}N CSA order parameters provide a more accurate description of molecular mobility than experimental ^{15}N CS tensor spans and allow to distinguish between the residues with really low mobility (and therefore large observed CS tensor spans) and the residues with slightly below-average mobility and large observed CS tensor span due to molecular geometry (e.g. the residue Gly 65).

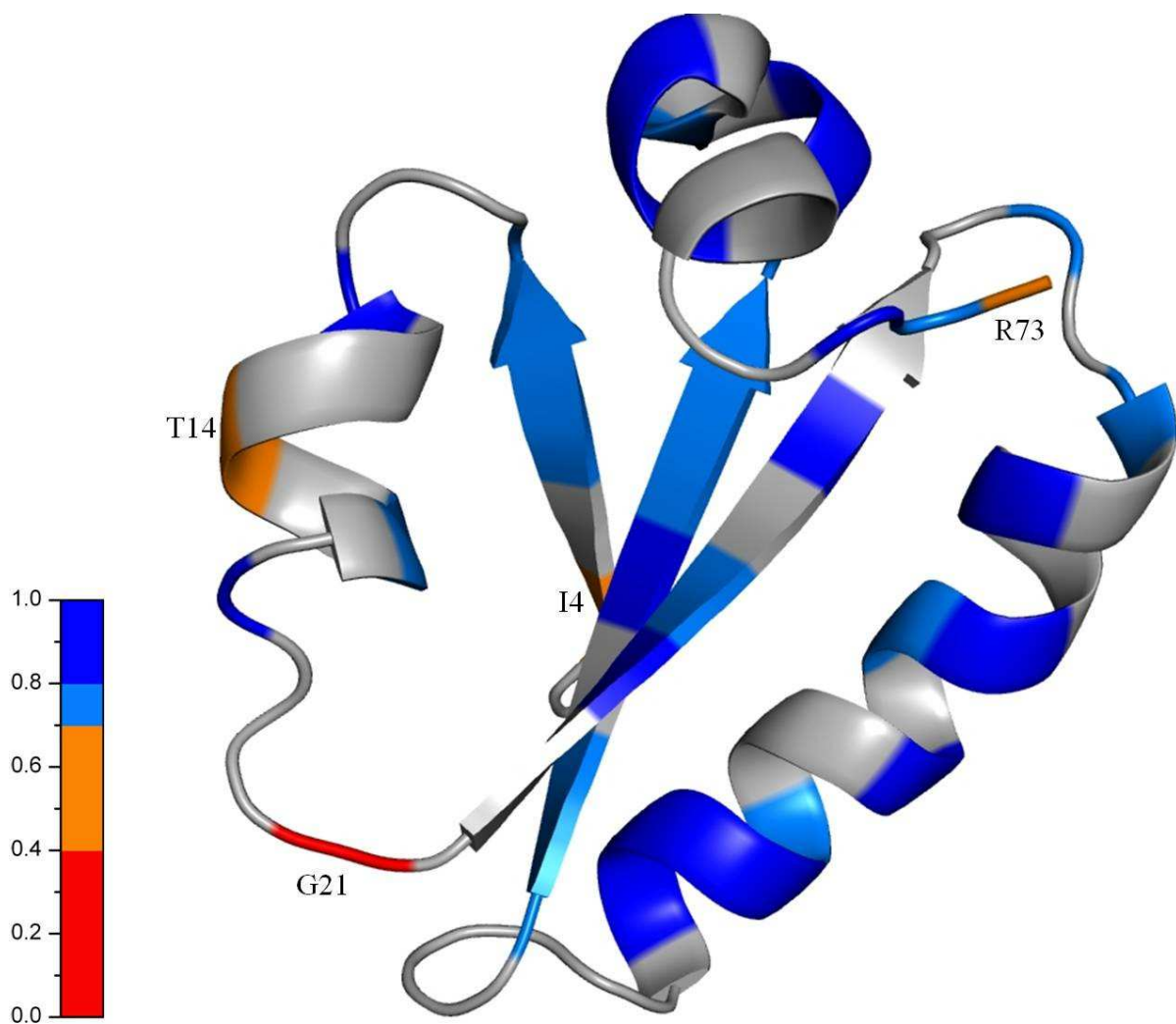


Figure 4.11: Distribution of the ^{15}N CSA order parameters in the backbone of the microcrystalline thioredoxin fragment containing the residues 1 to 73. Labels indicate the residues with the highest mobility (Ile 4, Thr 14, Gly 21 and Arg 73); residues with missing experimental data are shown in gray.

4.3 Application to immunoglobulin-binding protein GB1

Protein G produced by group G streptococci bacteria can bind and neutralize the antibodies of the host organism and defend the bacteria against the immune response¹⁰¹. Protein G consists of repetitively arranged homologous domains. The β 1 immunoglobulin-binding domain of protein G (GB1) is a 56-amino acid residue protein often used as a model system for development of solid-state NMR methodology. Full chemical shift assignment¹⁰², ^{15}N chemical shift tensors¹⁰³ and a solid-state NMR structure¹⁰⁴ of microcrystalline GB1 were determined by Rienstra and coworkers. ^{13}C spin-lattice relaxation rates in GB1 were measured by Lewandowski et al.¹⁰⁵ Furthermore, Franks et al. tried to determine the ^1H - ^{15}N dipolar order parameters in GB1 backbone; however, the estimated experimental errors were too large for their detailed interpretation¹⁰². Reliable results could only be obtained for the residues Asp 40, Gly 41 Trp 43 and Val 54, which seem to exhibit higher mobility than the other residues and therefore the mobility of GB1 in the solid state still remains an open question.

To facilitate the CSA order parameters based investigations on the GB1 dynamics, all structures of the solid-state GB1 structure ensemble¹⁰⁴ (PDB ID 2JSV) were subjected to the chemical shift driven geometry optimization with amide backbone ^{15}N and $^{13}\text{C}_\alpha$ and $^{13}\text{C}_\beta$ isotropic chemical shifts¹⁰² (BMRB accession number 15156) with the subsequent ensemble averaged ^{15}N CS tensor span calculation. BPT calculated ensemble averaged isotropic CS values are in agreement with the experimental chemical shifts (cf. table 4.2 and figures 4.12 and 4.13).

Wylie et al. performed two experiments in order to measure ^{15}N CS tensor spans in GB1¹⁰³: an experiment with $^{13}\text{C}_\alpha$ signal detection and an experiment with backbone carbonyl carbon atom signal detection. None of these experiments allowed to resolve all resonances of GB1, however, the amount of the observed resonances is sufficient for describing the molecular mobility in the GB1 backbone. It should be noted that Wylie et al. were not able to provide an accurate estimation of the experimental error limits for each experiment. However, both experiments applied by Wylie et al. were performed and evaluated independently. Mean ^{15}N CS tensor span values were

calculated for all residues with data available from both experiments (in total 37 mean experimental values) and deviations from mean and experimental values were adopted as error limits. 6 additional ^{15}N CS tensor span values from the $^{13}\text{C}_\alpha$ -detected experiment and 8 additional ^{15}N CS tensor span values from the backbone carbonyl carbon atom detected experiment were included into the evaluation without estimation of the error limits. This allowed the determination of the CSA order parameters for 51 out of the 56 GB1 residues.

Table 4.2: Correlation between the BPT calculated and the experimental chemical shifts in GB1 before and after chemical shift driven structure optimization (SD and MAE are given in ppm).

Nucleus	2JSV			optimized 2JSV		
	R	SD	MAE	R	SD	MAE
$^{13}\text{C}_\alpha$	0.81	3.0	2.3	0.94	1.7	1.3
$^{13}\text{C}_\beta$	0.97	3.9	3.2	0.98	3.0	2.2
^{15}N	0.55	14.3	12.3	0.89	4.0	3.3

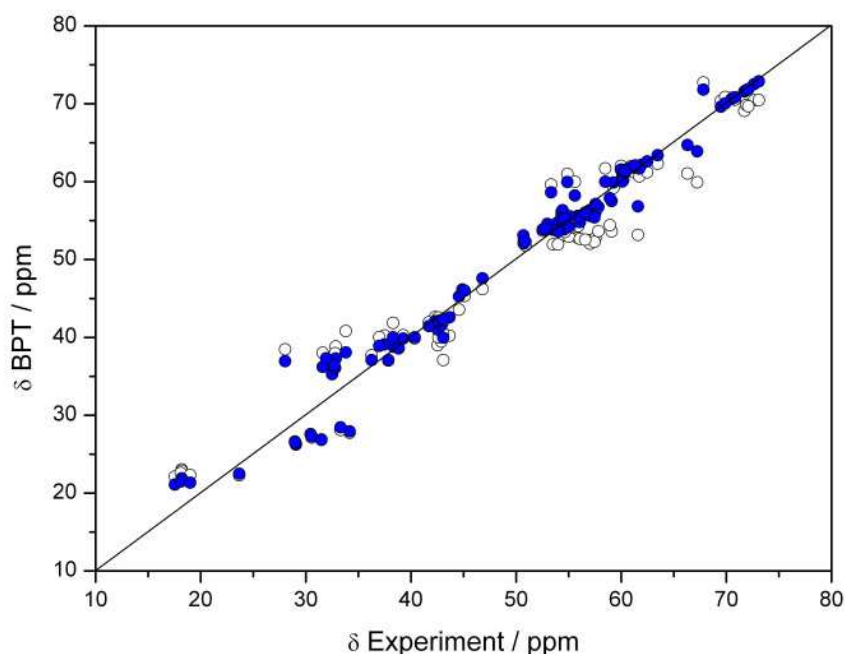


Figure 4.12: Correlation between experimental and BPT calculated $^{13}\text{C}_\alpha$ and $^{13}\text{C}_\beta$ chemical shifts of GB1 before (open black circles) and after (blue solid circles) chemical shift driven geometry optimization.

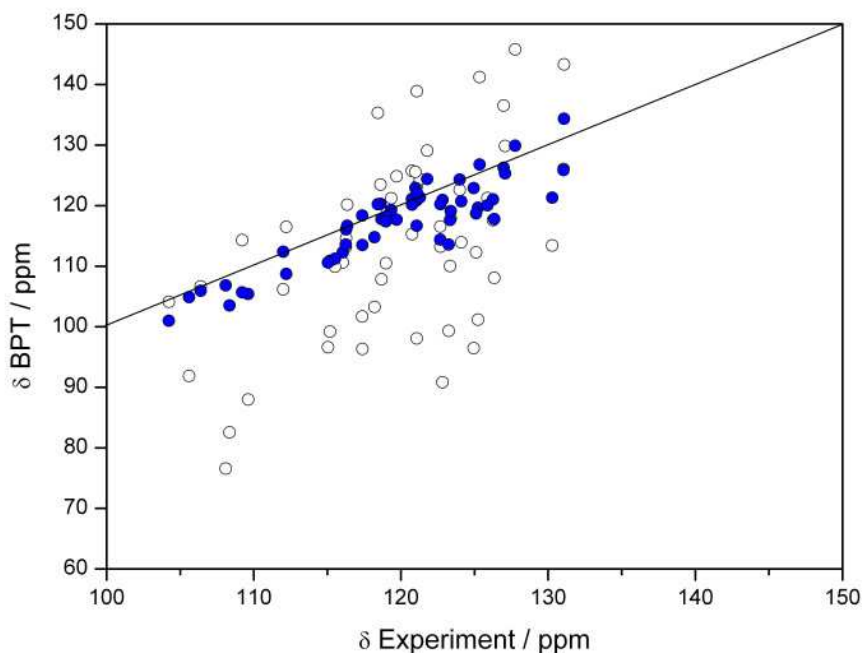


Figure 4.13: Correlation between experimental and BPT calculated ^{15}N chemical shifts of GB1 before (open black circles) and after (blue solid circles) chemical shift driven geometry optimization.

^{15}N chemical shift tensor spans of GB1 were computed and compared to the mean experimental ^{15}N CS tensor span values (cf. figure 4.14). As already demonstrated for thioredoxin, BPT computed ^{15}N chemical shift tensor spans, corresponding to the immobile structure, are systematically larger than experimentally measured¹⁰³, motionally averaged ^{15}N CS tensor spans (mean span values of 189.0 ± 8.0 ppm vs. 173.0 ± 9.5 ppm can be observed; note that the values of 8.0 and 9.5 ppm do not refer to experimental errors but to the standard deviation of the CS tensor span values in GB1). The only exception to this trend is the residue Lys 31. The experimental ^{15}N CS tensor span of the residue Lys 31 is larger than the calculated tensor span ($185.5 \pm \text{n.a.}$ ppm vs. 172.9 ppm; n.a. (not available) denotes missing data for error estimation, cf. above). This can result either from the shortcomings of the BPT computation or from the experimental error. Therefore the residue Lys 31 was excluded from further evaluation.

It is tempting to evaluate the mobility in GB1 backbone by comparing the magnitudes of the experimental ^{15}N CS tensors. Using the assumption that a smaller CS tensor

span corresponds to a higher degree of motional tensor span narrowing, it is possible to identify the most mobile (Gly 41 with the span of $\Omega=132.6\pm 10.9$ ppm, Asp 47 with $\Omega=153.5\pm n.a.$ ppm) and the most immobile (Lys 50 with $\Omega=192.5\pm 2.8$ ppm) residues (cf. figure 4.14). In contrast to thioredoxin, where a few residues with smaller experimental ^{15}N CS tensor spans were surrounded by the residues with larger spans, clear tendencies in the distribution of experimental ^{15}N CS tensor spans can be observed in the GB1 backbone. Experimental ^{15}N CS tensor spans in the β -sheets (mean value of 171.1 ± 4.0 ppm) are systematically smaller than in the α -helix (mean value of 180.0 ± 4.2 ppm). At first glance, it is not obvious whether this tendency can be attributed to the motional tensor span narrowing or to the influence of the molecular geometry on the chemical shift tensors. Wylie et al.¹⁰³ discussed both possibilities and Saito et al.¹⁰⁶ mentioned only the interpretation based on the influence of the molecular geometry in their recent review on NMR CS tensors in proteins.

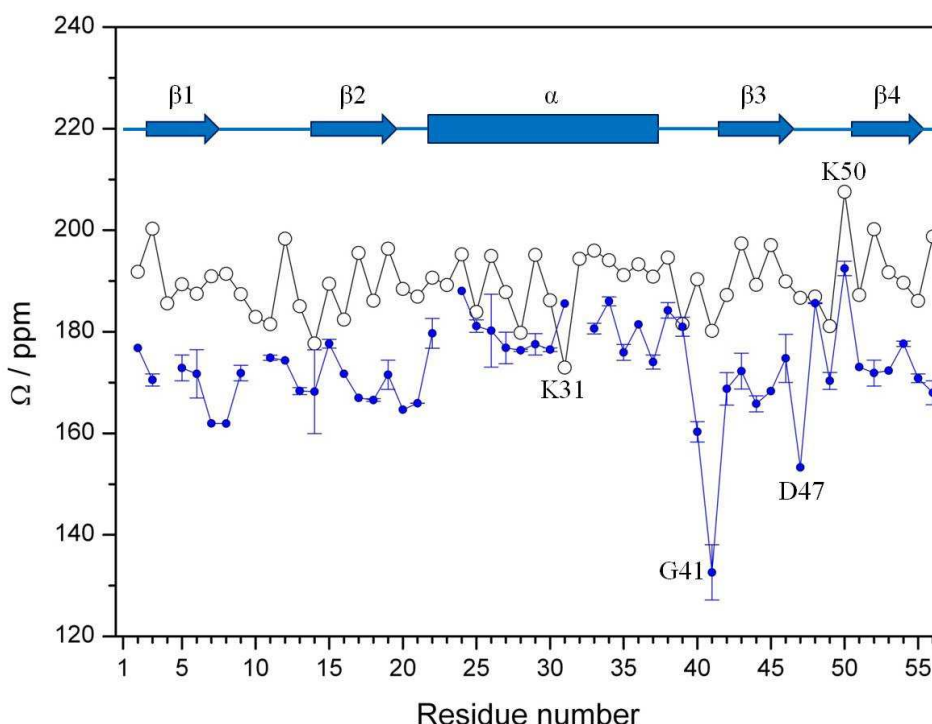


Figure 4.14: Experimental (solid blue circles) and BPT calculated (open black circles) backbone ^{15}N CS tensor spans of GB1. The secondary structure (blue arrows denote the β -sheets, the blue rectangle denotes the α -helix) is defined on the basis of the solid-state NMR GB1 structure (PDB ID 2JSV).

An unambiguous answer to this question is obtained if the differences between the experimental and the calculated ^{15}N CS tensor spans (cf. figure 4.15) and the corresponding CSA order parameters S_Ω are computed (cf. figure 4.16). CSA order parameters confirm the high mobility of the residues Gly 41 (order parameter of $S_\Omega = 0.74 \pm 0.06$) and Asp 47 ($S_\Omega = 0.82 \pm n.a.$; mean CSA order parameter in GB1 is $S_\Omega = 0.91 \pm 0.05$, note that the value of 0.05 does not refer to the experimental error but to the standard deviation of the CSA order parameter values in GB1). On the other hand, CSA order parameters reveal that the mobility of the residue Lys 50 ($S_\Omega = 0.93 \pm 0.01$) is not significantly lower than the mobility of the other residues. The unusually large ^{15}N CS tensor span of Lys 50 can probably be explained by its atypical backbone geometry (positive backbone torsion angle φ , e.g. φ of 26° in the first structure of the solid-state NMR structure ensemble with PDB ID 2JSV; this explanation was already considered by Wylie et al.¹⁰³). The lowest mobility can be observed for the residues Val 39 ($S_\Omega = 0.99 \pm 0.02$) and Ala 48 ($S_\Omega = 0.99 \pm 0.00$). This tendency could not be clearly detected in the evaluation of the experimental ^{15}N CS tensor spans ($\Omega = 181.0 \pm 3.8$ ppm and $\Omega = 186.0 \pm 0.2$ ppm for the residues Val 39 and Ala 48, respectively).

It is noteworthy that the residues Val 39 and Ala 48 are located in loops. In this case high mobility is usually expected. CSA order parameters predominantly describe the molecular mobility on the microsecond to millisecond time scale and do not necessarily reflect the nanosecond time scale fluctuations in protein loops. However, spin-lattice relaxation rates provide information on the molecular mobility on the pico- to nanosecond time scale. A comparison to the spin-lattice relaxation rates measured by Lewandowski et al.¹⁰⁵ revealed that on the pico- to nanosecond time scale the residues Val 39 ($R_1 = 0.244 \pm 0.047$ s⁻¹; mean relaxation rate of $R_1 = 0.461 \pm 0.187$ s⁻¹ can be observed in GB1; note that the value of 0.187 s⁻¹ does not refer to the experimental error but to the standard deviation of the spin-lattice relaxation rates in GB1) and Ala 48 ($R_1 = 0.431 \pm 0.055$ s⁻¹) exhibit average mobility (cf. figure 4.17). It should be noted, that, in contrast to the tendency observed in thioredoxin, in GB1 the mobility on the pico- to nanosecond time scale does not modulate the mobility on the micro- to millisecond time scale. The most mobile residues on the pico- to nanosecond time scale are the residues Thr 11 ($R_1 = 0.973 \pm 0.186$ s⁻¹) and Asp 40

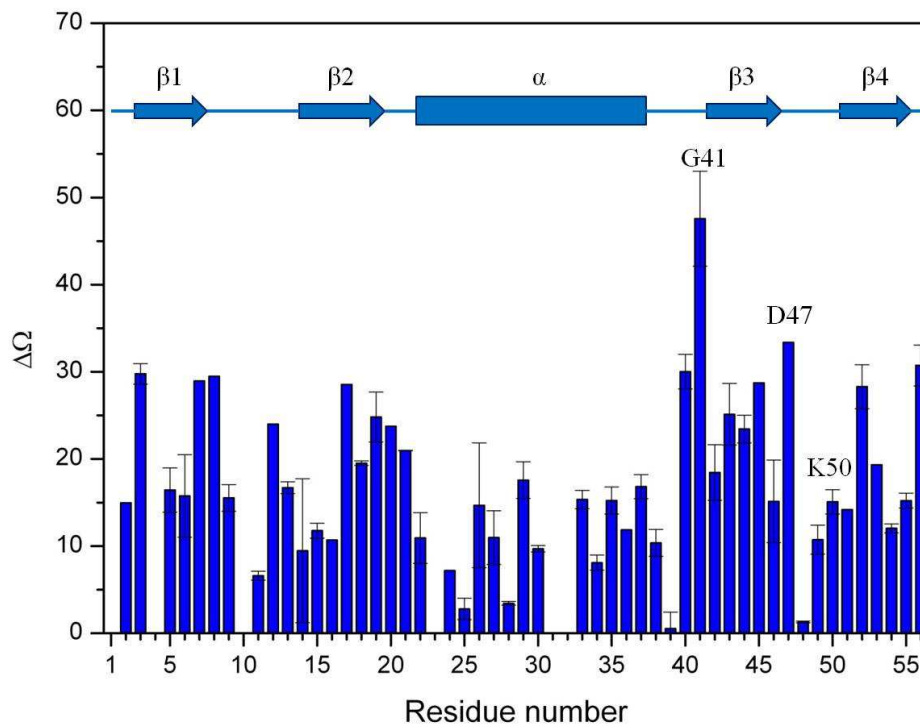


Figure 4.15: Differences between the experimental and the BPT calculated backbone ^{15}N CS tensor spans in GB1.

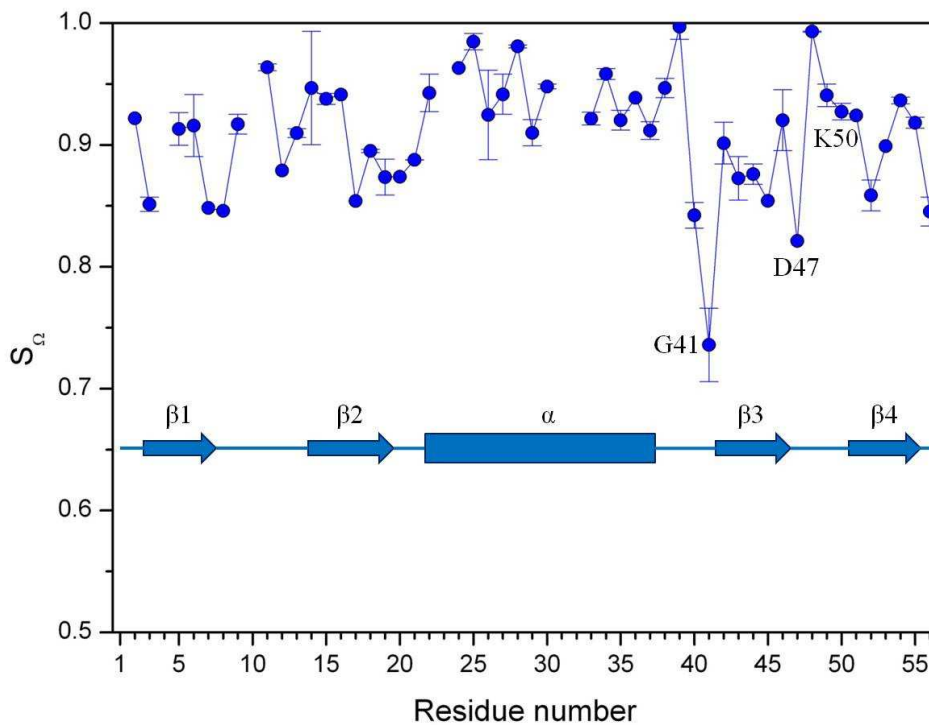


Figure 4.16: ^{15}N CSA order parameters in GB1.

($R_1 = 0.816 \pm 0.187 \text{ s}^{-1}$). On the micro- to millisecond time scale the residue Thr 11 ($S_\Omega = 0.96 \pm 0.00$) is rather immobile and the residue Asp 40 ($S_\Omega = 0.84 \pm 0.01$) exhibits lower mobility than the residues Gly 41 and Asp 47. Curiously, though the residue Gly 41 is located in the loop, its high mobility predominantly takes place on the micro- to millisecond and not on the pico- to nanosecond time scale.

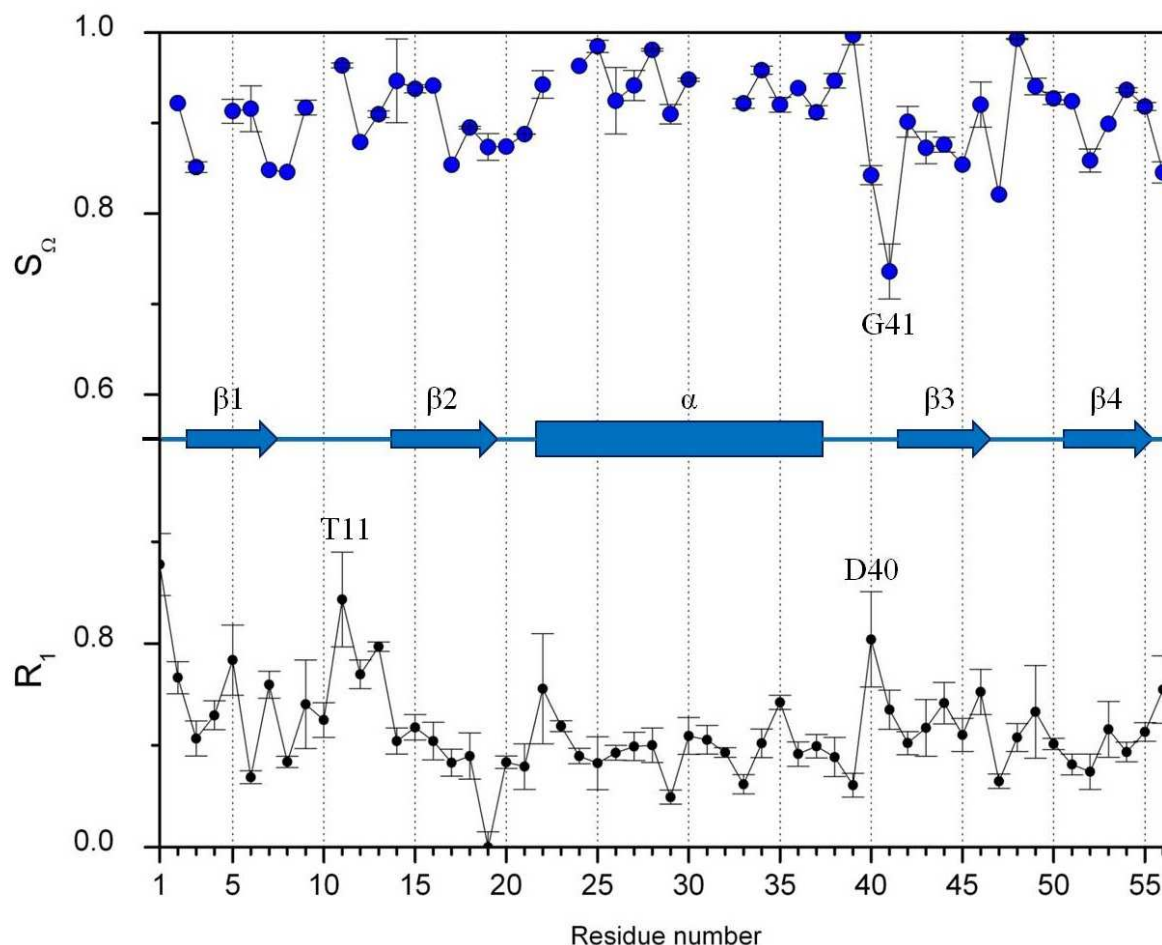


Figure 4.17: ^{15}N CSA order parameters (at the top) and $^{13}\text{C}_\alpha$ spin-lattice relaxation rates¹⁰⁵ (given in s^{-1} ; at the bottom) of microcrystalline GB1.

Last but not least it becomes obvious that the difference of the experimental ^{15}N CS tensor spans between the β -sheets and the α -helix can be explained by the impact of molecular dynamics and not by the influence of molecular geometry (cf. figure 4.18). On the micro- to millisecond time scale β -sheets (mean order parameter of $S_\Omega = 0.90 \pm 0.03$) are more mobile than the α -helix (mean order parameter of

$S_{\Omega} = 0.95 \pm 0.04$; note that the values of 0.03 and 0.04 do not refer to the experimental errors but to the standard deviation of the order parameters).

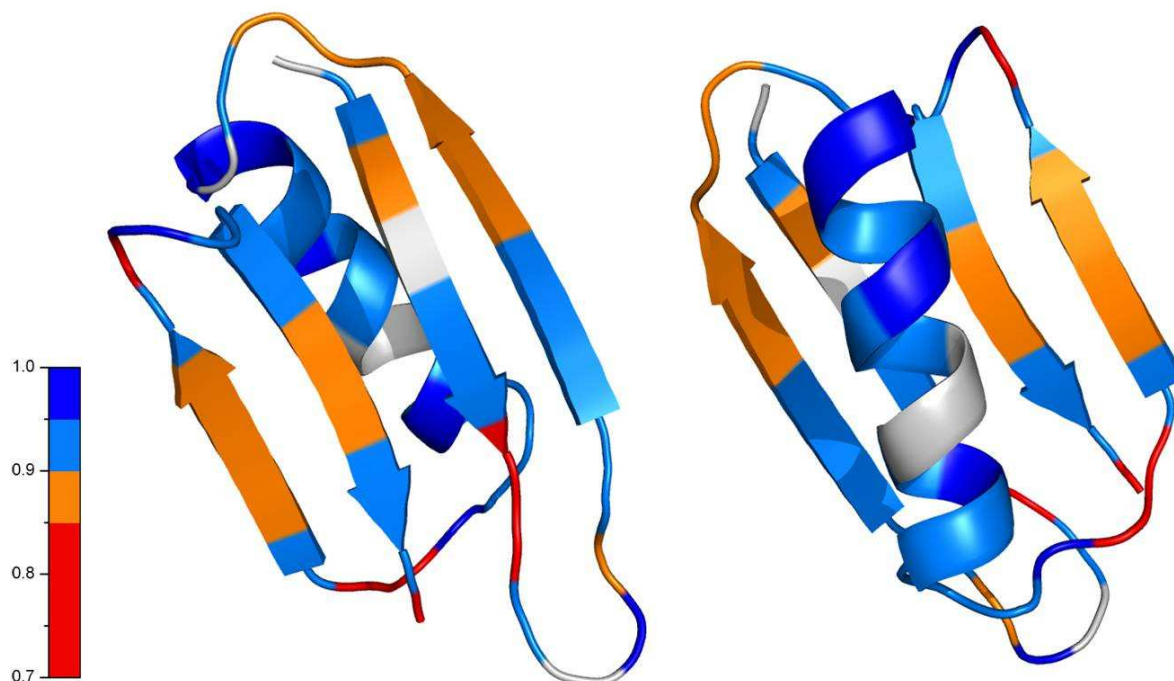


Figure 4.18: Distribution of the ^{15}N CSA order parameters in the backbone of microcrystalline GB1. Residues with missing experimental data are shown in gray.

The investigations on GB1 as well as the investigations on thioredoxin demonstrate that the evaluation of the experimental backbone ^{15}N CS tensors can shed light on the molecular mobility. However, until now it was not possible to extract reliable information on the molecular dynamics from the experimental ^{15}N CS tensors if other experimental data were missing. The approach described in this chapter presents an available solution to this problem.

4.4 Protein collective motions in solid state

CSA order parameters revealed relatively high β -sheet mobility and relatively low α -helix mobility on the micro- to millisecond time scale in the microcrystalline protein GB1. The $\beta 1$ immunoglobulin-binding domain of protein G (GB1) is very similar to the $\beta 3$ immunoglobulin-binding domain of protein G (GB3). Both proteins share

sequence identity for 50 out of 56 amino acids. The C_{α} backbone RMSD between the X-ray structure of GB1 crystal modification similar to the microcrystals used in solid-state NMR studies¹⁰⁷ (PDB ID 2QMT) and the GB3 X-ray structure refined with residual dipolar couplings¹⁰⁸ (PDB ID 2OED) is merely 0.41 Å. Liquid-state NMR investigations using residual dipolar couplings were performed on the micro- to millisecond time scale dynamics of GB3¹⁶. These investigations indicated relatively high β -sheet mobility and relatively low α -helix mobility and demonstrated that the motions across the β -sheets are mutually correlated and transmitted through the interstrand hydrogen bonds. High similarity between GB3 and GB1 implies that the same results are also valid for GB1. Interstrand hydrogen bonds remain preserved in the microcrystalline GB1. Therefore β -sheet motion transmitted via hydrogen bonds can explain the mobility distribution observed in microcrystalline GB1.

The comparison of the GB1 CSA order parameters to the largest liquid-state motional amplitudes of the GB3 peptide planes calculated by Bouvignies et al.¹⁶ (cf. figure 4.19) shows that the patterns of the backbone dynamics distribution in GB3 and GB1 are similar. However, residual dipolar couplings appear more sensitive to dynamics on the pico- to nanosecond time scale and detect dynamics in the loop between the β -sheets denoted as β_1 and β_2 , which is not evident from the CSA order parameters. In addition, the amplitudes of the peptide plane motions described by the RDCs appear to alternate across the β -sheets, thus reflecting the correlation of motions of the hydrogen-bonded peptide planes. As already mentioned above, CSA order parameters do not merely indicate the local motions of the peptide planes, but they also depend on the surrounding of the ^{15}N nuclei. Therefore, CSA order parameters do not clearly reflect the alternations of motional amplitudes detected by RDCs. Nevertheless, the approach presented in this chapter is obviously able to detect protein collective motions on the microsecond to millisecond time scale in microcrystalline proteins, where RDC-based liquid-state methodology cannot be applied.

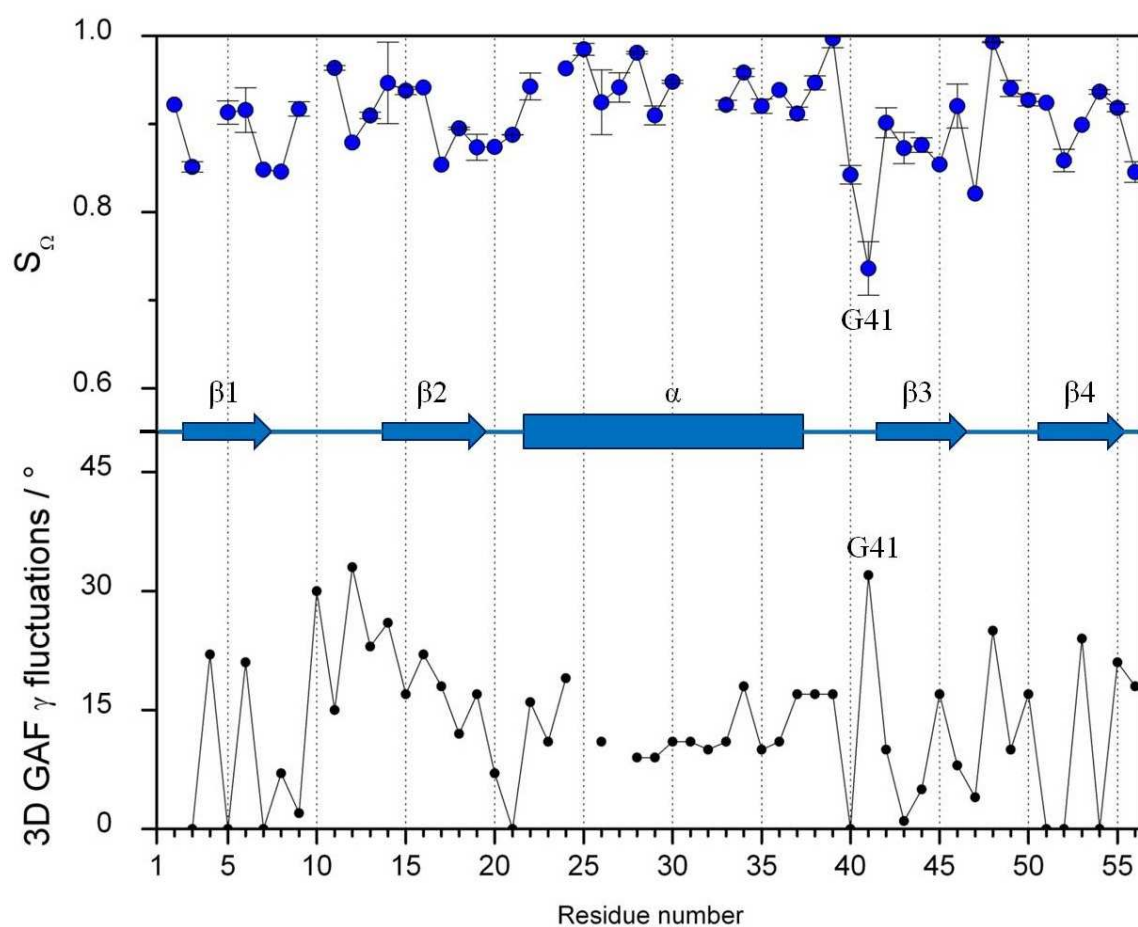


Figure 4.19: ^{15}N CSA order parameters in microcrystalline GB1 (at the top) and liquid-state 3D GAF (cf. section 1.5) motional γ -amplitudes¹⁶ in GB3 (at the bottom).

5. Methods

This chapter summarizes the settings required to reproduce the calculations described in this thesis. All calculations, except stated otherwise, were performed with the programs COSMOS PRO or COSMOS-Backend (COSMOS PRO version without graphic user interface)⁵³.

5.1 Molecular Modeling

5.1.1 Tripeptide crystals and α -helical polypeptide models.

Initial tripeptide crystal structures for CS calculation (APG·H₂O (CSD code FUDGIU), GGV·2H₂O (CSD code CUWRUH) and AGG·H₂O (CSD code CALXES), for details see Waddell et al.⁴⁴ and the papers cited therein) were obtained from the Cambridge structural database (<http://www.ccdc.cam.ac.uk>). Position of the hydrogen atoms in the tripeptide structures was optimized (cf. section 5.1.3) and crystal unit cells were created using the COSMOS PRO program. The coordinates of the tripeptide trimer clusters were adopted from Waddell et al.

The experimental structure of f-MLF-OH was obtained from Rienstra et al.⁴⁵ The positions of the hydrogen atoms in f-MLF-OH were optimized (cf. section 5.1.3) before calculation.

NH₃⁺-(Ala₄X)₄Ala₄-COO⁻ (X=amino acid cf. section 2.5) peptides with the consensus α -helical geometry ($\varphi=-57.80^\circ$ and $\psi=-47.00^\circ$) were constructed using the COSMOS Pro peptide builder tool and were subjected to geometry optimization.

5.1.2 Protein structures

All structures were retrieved from the protein data bank (<http://www.pdb.org>) using the PDB IDs as specified in the corresponding sections. To all protein structures hydrogen atoms were added if necessary with the program COSMOS PRO.

5.1.3 Geometry optimization

Positions of the hydrogen atoms in the tripeptide structures were subjected to geometry optimization with the QM/MM hybrid COSMOS-NMR force field³². BPT charge computation³¹ on the quantum chemical level is included in the COSMOS-NMR force field and is available in every step of geometry optimizations or molecular dynamics simulations. A conjugate gradient geometry optimization with charge calculation in each step and an initial slope of 1.0 (this corresponds to the step width of 1.0×10^{-3} Å) was applied in all cases. Positions of non-hydrogen atoms were fixed during geometry optimizations because they were regarded to be reliable (the structures were derived from the X-ray experiments and the NMR experiments with accurate ^{13}C - ^{13}C and ^{13}C - ^{15}N distance measurements).

5.2 BPT chemical shift calculation

5.2.1 BPT charge calculation

The BPT charge calculation³¹ is a prerequisite for BPT chemical shift calculations²⁹. All BPT charge calculations were conducted using a distance cutoff radius of 100 Å for polarization integrals (i.e. nuclei more than 100 Å away from the nucleus of interest were not considered for charge calculation). In all peptide and protein structures charged groups were defined (i.e. the charges of the charged amino acid side chains were prespecified) before the BPT partial charge calculation (this step corresponds to the setting of total molecular charges in *ab initio* methods). BPT partial charges and chemical shifts in crystal unit cells were calculated under periodic boundary conditions.

5.2.2 BPT chemical shift calculation

All BPT CS calculations²⁹ were conducted using distance cutoff radii of 100 Å for polarization integrals (i.e. charges more than 100 Å away from the nucleus of interest were not considered) and nuclear shielding and chemical shift parameters as given in tables 2.1, 2.2 and 2.5. Calculated BPT ^{15}N nuclear shieldings were converted into

chemical shifts using the reference of 273.3 ppm (referencing to liquid NH₃)³⁸. Calculated BPT ¹³C chemical shifts (parameterized on data referenced to TMS) were converted from TMS to TSP or DSS chemical shift scale by adding 1.82 ppm or 1.7 ppm, respectively¹⁰⁹.

5.2.3 Averaged chemical shifts and chemical shift order parameters

In systems with strong chemical shift dependence on molecular dynamics averaged BPT chemical shifts were calculated on-the-fly during molecular dynamics simulation after each MD step. To minimize the influence of the MD equilibration phase on the averaged CSs, time averaging was performed using the exponential memory decay function⁸⁴

$$\langle \delta \rangle_t = \frac{1}{N_{MD\ Steps}} \int_0^t e^{t/\tau} \delta(t) dt, \quad (5.1)$$

where τ is the decay constant for the exponential decay and $N_{MD\ Steps}$ is the norm of the integral (number of MD steps in which the chemical shifts were computed). For f-MLF-OH (cf. section 2.4), the decay constant τ was set to 10 ps (the duration of the whole MD simulation, cf. section 5.5) and averaged CSs were calculated after each MD time step. For gramicidin A the decay constant τ was set to 200 ps (approximately the duration of the MD equilibration phase) and averaged CS were calculated after every 2.5 ps (note that gramicidin A MD coordinate snapshots were subjected to isotropic CS driven geometry optimization before CS tensor computation). Chemical shift order parameters were calculated using equations specified in sections 3.7 and 4.1.

5.3 DFT chemical shift calculation

DFT chemical shift calculations were carried out with the Gaussian 03 program⁵². All DFT CS calculations were conducted with B3LYP functional and 6-311++G(d,p) basis set. The reference of 244.6 ppm was used in order to convert ¹⁵N nuclear shieldings into chemical shifts (cf. Waddell et al.⁴⁴ for basis set and reference evaluation). Vila et al. demonstrated that observed TMS nuclear shielding of 188.1

ppm in the gas phase does not provide acceptable results for DFT ^{13}C computation and that calculated TMS nuclear shielding (which must be specified for each basis set) provides a better reference value⁵⁴. TMS ^{13}C nuclear shielding was computed after carrying out geometry optimization of TMS conformation in T_d symmetry with extremely tight convergence criteria^{55,56} (the Gaussian 03 program applies Berny geometry optimization algorithm¹¹⁰; tight convergence criteria in Gaussian 03 refer to the maximum force component of 2.0×10^{-6} , RMS force of 1.0×10^{-6} , maximum step component of 6.0×10^{-6} and an RMS step of 4.0×10^{-6} ; cf. table 2.9 for optimized geometry parameters). The computed shielding value of 184.2 ppm was used as a reference in order to convert ^{13}C nuclear shieldings into chemical shifts. Calculated DFT ^{13}C chemical shifts were converted from TMS to DSS chemical shift scale by adding 1.7 ppm¹⁰⁹.

5.4 Chemical shift driven geometry optimization

Conjugate gradient geometry optimization with charge calculation in each step was applied for the isotropic chemical shift driven structure refinement⁵¹. The minimization target function of the CS driven geometry optimization in the COSMOS program consists of the force field energy and the CS pseudo-energy. The convergence criterion is given by the change of the energy gradients of the two consecutive optimization steps and the optimization is performed until the convergence criterion drops under 1.0×10^{-6} . The initial slope for the CS driven geometry optimization was set to 0.2 (0.4 for the X-ray ubiquitin structure; initial slope of 1.0 corresponds to the initial step width of 1.0×10^{-3} Å). The CS potential width control parameter $\Delta\delta$ and the force constant k_0^δ of the chemical shift pseudo-forces were set to 30 and 15 ppm for ^{15}N and 2.5 and 2 ppm for ^{13}C , respectively.

5.5 Molecular dynamics simulation

5.5.1 Gly-Gly-Val unit cell and f-MLF-OH

The Gly-Gly-Val unit cell required additional hydrogen position optimization to reproduce experimental ^{15}N chemical shifts. Molecular dynamics simulation with the

COSMOS-NMR force field under periodic boundary conditions was applied in order to escape the local energy minimum found by the initial geometry optimization procedure. In addition, a molecular dynamics simulation with the COSMOS-NMR force field was applied for the time averaged CS calculation in f-MLF-OH. In both cases positions of non-hydrogen atoms were fixed during the MD simulation because they were regarded as reliable (Gly-Gly-Val structure was derived from the X-ray experiments and f-MLF-OH structure was derived from the NMR experiments with accurate ^{13}C - ^{13}C and ^{13}C - ^{15}N distance measurements) and unconstrained MD can easily perturb the whole structure. BPT partial charges were calculated after each MD step as described in section 5.2.1. Cutoff radii for van der Waals and electrostatic interactions during the MD simulation were set to 6 Å and 30 Å, respectively. The temperature was set to 293 K and the coupling constant for the thermostat was 0.01 ps. Both MD runs consisted of 20000 steps of 0.5 fs duration (total MD time of 10 ps).

5.5.3 Gramicidin A dimer

A 5 ns molecular dynamics simulation with COSMOS-NMR force field using ^1H - ^{15}N dipolar couplings measured by Ketchum et al.⁷⁶ as orientational constraints was performed based on the solid-state NMR structure of the gramicidin A dimer⁷⁶ (PDB ID 1MAG) by Dr. Ulrich Sternberg. The BPT charge calculation was conducted using a distance cutoff radius of 30 Å for polarization integrals before the MD simulation and calculated charges were retained during the whole MD run. Cutoff radii for van der Waals and electrostatic interactions during the MD simulation were set to 6 Å and 30 Å, respectively. The temperature was set to 293 K and the coupling constant for the thermostat was 0.01 ps. The MD run consisted of 10^6 steps of 0.5 fs duration (total MD time of 5 ns).

Distance constraints of 2.0 ± 1.0 Å were introduced for all hydrogen bonds observed in the solid-state NMR gramicidin A structure and the force constant of the distance constraints potential was set to 0.5. These very soft settings for distance constraints potential do not prevent hydrogen bond length fluctuations but they are sufficient to ensure that no major distortions of gramicidin A structure occur during the MD simulation.

The dipolar coupling potential width control parameter $\Delta D_{\alpha\beta}$ and the force constant k_0 of the dipolar coupling pseudo-forces were set to 1 kHz and 0.02, respectively. The averaged internuclear distance \bar{r} for dipolar coupling calculation was set to 1.0 \AA^{111} (note that different settings for the averaged internuclear distance will result in scaling of the calculated order parameters by a constant value but will not affect their distribution). The time constant ρ adjusting the time-dependent scaling factor s_i controlling the rise of the pseudo forces in the beginning of the MD simulation (cf. equation 3.2) was set to 200 ps. The averaged dipolar couplings obtained in the course of the MD simulation were calculated using an exponential memory decay function (cf. equation 3.3) and a decay constant τ of 200 ps. Dipolar coupling order parameters were calculated using equations specified in section 3.7.

6. Summary

Membrane proteins and membrane active-peptides remain a significant challenge to structural biology. Recent advances in solid-state NMR spectroscopy enabled structure determination of microcrystalline proteins and membrane alignment determination of α -helical membrane-active peptides. Investigations on solid-state protein dynamics and membrane alignment of non- α -helical peptides still remain a non trivial task. NMR chemical shift tensors can provide a wealth of information on protein structure, dynamics, and membrane alignment. However, *ab initio* methods for chemical shift tensor calculation are usually computationally too demanding for proteins. The work presented here is concerned with the adaption of the computationally inexpensive semi-empirical bond polarization theory (BPT) to protein chemical shift tensor calculations and its application to the study of membrane alignment and solid-state molecular dynamics in proteins and peptides.

6.1 BPT chemical shift calculation in proteins

Bond polarization theory was successfully adapted to the calculation of protein $^{13}\text{C}_\alpha$, $^{13}\text{C}_\beta$ and ^{15}N chemical shifts. Test calculations on tripeptide crystals demonstrated the accuracy of BPT chemical shift predictions. Mean absolute errors of 1.5, 2.4 and 1.4 ppm were observed between experimental and computed $^{13}\text{C}_\alpha$, $^{13}\text{C}_\beta$ and ^{15}N chemical shifts in the chemotactic peptide N-formyl-L-Met-L-Leu-L-Phe-OH⁴⁵. A mean absolute error of 2.1 ppm was observed between experimental and computed ^{15}N chemical shifts for the three peptide crystals of Ala-Gly-Gly, Ala-Pro-Gly, and Gly-Gly-Val⁴⁴. Test calculations confirmed that long range electrostatic effects may have a significant impact on ^{15}N chemical shifts and therefore the application of crystalline periodic boundary conditions is required in order to reproduce the experimental chemical shift values of peptide crystals. Moreover, the impact of the hydrogen atom positions and dynamics on ^{15}N chemical shifts can be comparable with the impact of long range electrostatic interactions; however, the influence of the both factors cannot be reliably estimated a priori.

Ensemble averaged $^{13}\text{C}_\alpha$, $^{13}\text{C}_\beta$ and ^{15}N chemical shifts were computed for the small globular protein ubiquitin. BPT-computed $^{13}\text{C}_\alpha$ and $^{13}\text{C}_\beta$ chemical shifts were in good agreement with experimentally observed chemical shifts. The obtained results are comparable to results obtained by computationally highly demanding *ab initio* calculations on the DFT / B3LYP level of theory with 6-311++G(d,p) basis set²⁴. Standard deviations of 2.4 ppm and 2.5 ppm between experimental and computed $^{13}\text{C}_\alpha$ chemical shifts were observed for BPT and DFT calculations, respectively. It is noteworthy that the BPT computation time for atomic charges and chemical shifts of the full-size ubiquitin structure was only 20.46 s on a desktop computer with 2 GHz CPU and 3 GB RAM.

The liquid-state NMR ubiquitin structure ensemble does not possess sufficiently accurate geometries for BPT-based ^{15}N chemical shift calculations. Bond polarization theory, however, makes it possible to calculate a gradient for chemical shifts and, consequently, chemical shift driven geometry optimization can be applied to overcome this problem. Within this thesis, the protocol for protein chemical shift driven geometry optimization was successfully developed and applied to a number of molecules. Its application for ubiquitin, for example, led to a good agreement between calculated and experimental ^{15}N chemical shifts with a low standard deviation of 3.3 ppm. In addition, isotropic chemical shift driven geometry optimization yields not only the structure matching the experimentally determined isotropic chemical shifts, but also leads to the derivation of full chemical shift tensors.

Another topic, the analysis of protein orientation and dynamics in a membrane, requires the prediction of full ^{15}N chemical shift tensors. Isotropic NMR chemical shifts are virtually always available even in the case where NMR distance constraints cannot be determined. Furthermore, chemical shift driven geometry optimization produces a structure that is merely a local minimum with respect to experimental chemical shifts and is extremely similar to the initial structure (the backbone RMSD between initial and optimized structures of the liquid-state NMR ubiquitin structure ensemble did not exceed 0.04 Å). Therefore, isotropic chemical shift based protein structure refinement with consecutive ^{15}N chemical shift tensor calculation provides

the foundation for the analysis of molecular orientation and dynamics using either static molecular structure models or molecular dynamics simulation trajectories.

6.2 Dynamics and orientation of membrane peptides

Gramicidin A is an antimicrobial peptide with ion channel activity that adopts β -helix geometry with a very flexible backbone. As the molecule does not represent a conventional rigid α -helical membrane-spanning peptide, the corresponding prediction of PISEMA spectra is a challenging task and membrane orientation of such peptides cannot be always easily determined. In this case, a molecular dynamics simulation combined with chemical shift tensor computation allows to simultaneously determine the mobility of individual residues and the overall alignment of gramicidin A in membrane.

A 5 ns molecular dynamics simulation with ^1H - ^{15}N dipolar couplings as orientational constraints was performed on the structure of the gramicidin A dimer. 2000 snapshots of molecular coordinates were extracted from the molecular dynamics trajectory and subjected to the ^{15}N chemical shift driven geometry optimization with the following ^{15}N chemical shift tensor calculation. Order parameters S_{CS} and biaxiality parameters ξ for all residues were calculated from the ^{15}N chemical shift tensor trajectory. The order parameter S_{CS} in this case describes the fluctuations around the average direction of motion along which the system under study may be regarded as aligned, while the biaxiality parameter indicates if a second preferred direction of non-random motion exists. In contrast to ^1H - ^{15}N dipolar coupling order parameters, ^{15}N chemical shift tensor derived order parameters are almost unbiased by ^1H - ^{15}N bond motions and provide a more accurate estimate of the backbone mobility.

As a result, the order parameters revealed that D-amino acids display a remarkably higher mobility than the L-amino acids and that the mobility of the backbone increases towards the termini of gramicidin A and therefore towards the bilayer surface. Thus it can be concluded that the gramicidin A dimer shows upright transmembrane alignment and is wobbling around the membrane normal. More

specifically, the dynamics of gramicidin A in lipid bilayers can be described in terms of a wobbling-in-a-cone model.

In addition to the overall motion of the molecule, the correlation of biaxiality parameters in the gramicidin A backbone indicates the presence of correlated motions in the mutually hydrogen-bonded peptide planes. Correlated motions of the mutually hydrogen-bonded peptide planes have been suggested in previous theoretical investigations on gramicidin A and have been proposed to play an important role in the mechanism of ion translation through the channel⁹⁰. Application of molecular dynamics simulations with ^1H - ^{15}N dipolar couplings as orientational constraints allowed now to obtain an evidence of these motions based on experimental data. The subsequent calculation of the order and biaxiality parameters from the ^{15}N chemical shift tensor trajectories so far appears to be the only method allowing the direct detection of correlated motions in the gramicidin A backbone. Moreover, molecular dynamics simulation with ^1H - ^{15}N dipolar couplings as orientational constraints and BPT ^{15}N chemical shift tensor computation enabled the prediction of the PISEMA spectrum of gramicidin A with the accuracy resembling the accuracy of experimental measurements.

As a general result, the approach developed in this work allows to validate the assignment of PISEMA spectra by their simulation. If the signals of several residues can be assigned using selective labeling, the position of other signals could be deduced using a trial and error assignment strategy based on the comparison of calculated and experimental spectra. In this way, the required amount of selectively labeled peptides might be reduced in future investigations.

6.3 Dynamics of solid-state proteins

Protein dynamics on the micro- to millisecond time scale play a crucial role in protein function. ^{15}N chemical shift tensors are very sensitive to their environment. They are less influenced by ^1H - ^{15}N bond pico- to nanosecond timescale vibrational motions and predominantly reflect molecular motions of the backbone on the millisecond time

scale. BPT computations yield ^{15}N chemical shift tensors of an immobilized “frozen” protein structure, while motionally averaged chemical shift tensors can be measured experimentally. In this thesis, a new chemical shift anisotropy order parameter S_{Ω} , based on the comparison of calculated and experimental tensor spans, is introduced.

Until now, ^{15}N chemical shift tensors in the solid state were measured only for two proteins, thioredoxin for which also extensive investigations of molecular dynamics have been performed and the immunoglobulin-binding domain of protein G (GB1). Within this thesis, CSA order parameters were computed for both proteins and compared to previously published data.

The results obtained for thioredoxin were in good agreement with previous investigations including specific variations of mobility for residues Ile 4, Thr 14, Gly 21, and Arg 73.

The results obtained for GB1 indicate relatively high β -sheet mobility and relatively low α -helix mobility. This mobility distribution resembles the liquid-state mobility distribution of the very similar protein GB3 determined previously by RDCs and H-bond mediated scalar couplings. β -sheet collective motion transmitted via interstrand hydrogen bonds, previously observed in GB3, can explain the mobility distribution observed in GB1.

In summary, several methods for the study of molecular membrane orientation and dynamics from solid-state NMR data were developed and evaluated in this thesis. They will provide an extended toolbox for the examination and characterization of soluble and membrane-bound proteins and peptides within the program COSMOS and will hopefully lead to a better understanding of the complex and highly dynamic interplay of biomolecules in the future.

List of Abbreviations

a.u.	Expressed in atomic units
B3LYP	Becke 3-parameter, Lee-Yang-Parr exchange-correlation functional
BMRB	Biological magnetic resonance data bank
BPT	Bond polarization theory
CGS	Centimetre-gram-second system of units
CPHF	Coupled Perturbed Hartree-Fock method
CPU	Central processing unit
CS	Chemical shift
CSA	Chemical shift anisotropy
CSD	Cambridge structural database
DFT	Density functional theory
DMPC	Dimyristoylphosphatidylcholine
DSS	2,2-dimethyl-2-silapentane-5-sulfonate, sodium salt
GAF	Gaussian axial fluctuation
GIAO	Gauge-including atomic orbitals
Hz	Hertz
IGLO	Individualized gauges for localized orbitals
MAE	Mean absolute error
MD	Molecular dynamics simulation
MP2	Second-order Møller-Plesset perturbation theory
NMR	Nuclear magnetic resonance
NPA	Natural population analysis
PDB	Protein data bank
PDB ID	Protein data bank identification code
PISA	Polarity index slant angle
PISEMA	Polarization inversion spin exchange at magic angle
ppm	Parts per million
RAM	Random-access memory
RDC	Residual dipolar coupling
RMSD	Root mean square deviation

List of Abbreviations

SD	Standard deviation
SI	International system of units
TMS	Tetramethylsilane
TSP	3-(trimethylsilyl)-propionate, sodium salt
TZVPP	Triple-zeta valence plus polarization basis set

References

- ¹ Tan, S.; Tan, H. T.; Chung, M. C. M. Membrane Proteins and membrane proteomics; *Proteomics* **2008**, *8*, 3924
- ² Saito, H.; Ando, I.; Naito, A. Solid State NMR Spectroscopy for Biopolymers: Principles and Applications; Springer: Dordrecht, **2006**
- ³ Sternberg, U.; Witter R.; Ulrich, A. S. 3D Structure Investigation Using NMR Chemical Shifts; *Ann. Rep. NMR Spectrosc.* **2004**, *52*, 53
- ⁴ Cavalli, A.; Salvatella, X.; Dobson, C. M.; Vendruscolo, M. Protein structure determination from NMR chemical shifts; *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 9615
- ⁵ Shen, Y.; Lange, O.; Delaglio, F.; Rossi, P.; Aramini, J. M.; Liu, G. H.; Eletsky, A.; Wu, Y. B.; Singarapu, K. K.; Lemak, A.; Ignatchenko, A.; Arrowsmith, C. H.; Szyperski, T.; Montelione, G. T.; Baker, D.; Bax, A. Consistent blind protein structure generation from NMR chemical shift data; *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 4685
- ⁶ Wishart D. S.; Arndt, D.; Berjanskii, M.; Tang, P.; Zhou, J.; Lin, G. CS23D: a web server for rapid protein structure generation using NMR chemical shifts and sequence data; *Nucleic Acids Res.* **2008**, *36*, 496
- ⁷ Helgaker, T.; Jaszunski, M.; Ruud, K. Ab initio methods for the calculation of NMR shielding and indirect spin–spin coupling constants; *Chem. Rev.* **1999**, *99*, 293
- ⁸ Sternberg, U. Theory of the influence of the second co-ordination sphere on the chemical shift; *J. Mol. Phys.* **1988**, *63*, 249
- ⁹ Ramamoorthy, A.; Wie, Y.; Lee, D. PISEMA Solid-State NMR Spectroscopy; *Ann. Rep. NMR Spectrosc.* **2004**, *52*, 1
- ¹⁰ Sloan, T. The spin structure of the nucleon; *Phil. Trans. R. Soc. Lond. Ser. A* **2001**, *359*, 379
- ¹¹ Mason, J. Conventions for the reporting of nuclear magnetic shielding (or shift) tensors suggested by participants in the NATO ARW on NMR Shielding Constants at the University of Maryland, College Park, July 1992; *Solid State Nucl. Mag.* **1993**, *2*, 285

References

- ¹² Resing, H. A.; Slotfeld-Ellingsen, D. Order parameters from chemical shift anisotropy pattern in two-dimensional systems; *J. Magn. Reson.* **1980**, *38*, 401
- ¹³ Lipari, G.; Szabo, A. Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 1. Theory and range of validity; *J. Am. Chem. Soc.* **1982**, *104*, 4546
- ¹⁴ Bax, A.; Grishaev, A. Weak alignment NMR: a hawk-eyed view of biomolecular structure; *Curr. Opin. Struct. Biol.* **2005**, *15*, 563
- ¹⁵ Bremi, T.; Brüschweiler, R. Locally Anisotropic Internal Polypeptide Backbone Dynamics by NMR Relaxation; *J. Am. Chem. Soc.* **1997**, *119*, 6672
- ¹⁶ Bouvignies, G.; Bernado, P.; Meier, S.; Cho, K.; Grzesiek, S.; Brüschweiler, R.; Blackledge, M. Identification of slow correlated motions in proteins using residual dipolar and hydrogen-bond scalar couplings; *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 13885
- ¹⁷ Helgaker, T.; Jaszunski, M.; Ruud, K. Ab initio methods for the calculation of NMR shielding and indirect spin–spin coupling constants; *Chem. Rev.* **1999**, *99*, 293
- ¹⁸ Ramsey, N. Magnetic Shielding of Nuclei in Molecules; *Phys. Rev.* **1950**, *78*, 699
- ¹⁹ Juselius, J.; Sundholm, D.; Gauss, J. Calculation of current densities using gauge-including atomic orbitals; *J. Chem. Phys.* **2004**, *121*, 3952
- ²⁰ Wolinski, K.; Hinton, J. F.; Pulay, P. Efficient Implementation of the Gauge-Independent Atomic Orbital Method for NMR Chemical Shift Calculations; *J. Am. Chem. Soc.* **1990**, *112*, 8251
- ²¹ Brunner, E.; Sternberg, U. Solid-state NMR investigations on the nature of hydrogen bonds; *Prog. Nucl. Mag. Res. Sp.* **1998**, *32*, 21
- ²² Kutzelnigg, W. Theories of magnetic susceptibilities and NMR chemical shifts in terms of localized quantities; *Isr. J. Chem.* **1980**, *19*, 193
- ²³ Gauss, J. Calculation of NMR chemical shifts at second-order many-body perturbation theory using gauge-including atomic orbitals; *Chem. Phys. Lett.* **1992**, *191*, 614

References

- ²⁴ Vila, J. A.; Villegas, M. E.; Baldoni, H. A.; Scheraga, H. A. Predicting $^{13}\text{C}^\alpha$ chemical shifts for validation of protein structures; *J. Biomol. NMR* **2007**, *38*, 221
- ²⁵ Beer, M.; Kussmann, J.; Ochsenfeld, C. Nuclei-selected NMR shielding calculations: A sublinear-scaling quantum chemical method; *J. Chem. Phys.* **2011**, *134*, 74102
- ²⁶ Cui, Q.; Karplus, M. Molecular Properties from Combined QM/MM Methods. 2. Chemical Shifts in Large Molecules; *J. Phys. Chem. B* **2000**, *104*, 3721
- ²⁷ Bader, R. Utilizing the Charge Field Effect on Amide ^{15}N Chemical Shifts for Protein Structure Validation; *J. Phys. Chem. B* **2009**, *113*, 347
- ²⁸ Del Re, G. A Simple MO-LCAO Method for the Calculation of Charge Distributions in Saturated Organic Molecules; *J. Chem. Soc.* **1958**, *812*, 4031
- ²⁹ Sternberg, U.; Prieß, W. New Semi-empirical Approach for the calculation of ^{13}C Chemical Shift Tensors; *J. Magn. Reson.* **1997**, *125*, 8
- ³⁰ O'Keefe, M.; Breese, N. E. Atom Sizes and Bond Lengths in Molecules and Crystals; *J. Am. Chem. Soc.* **1991**, *113*, 3226
- ³¹ Sternberg, U.; Koch, F.-Th.; Möllhoff, M. New Approach to the Semiempirical Calculation of Atomic Charges for Polypeptides and Large Molecular Systems; *J. Comput. Chem.* **1994**, *15*(5), 524
- ³² Möllhoff, M.; Sternberg, U. Molecular mechanics with fluctuating atomic charges – a new force field with a semi-empirical charge calculation; *J. Mol. Model.* **2001**, *7*, 90
- ³³ Reed, A. E.; Weinstock R. B. and Weinhold, F. Natural population analysis; *J. Chem. Phys.* **1985**, *83*, 735
- ³⁴ Prieß, W.; Sternberg, U. Fast calculation of ^{13}C NMR chemical shift tensors using bond polarization model; *J. Mol. Struct. (Theochem)* **2001**, *544*, 181
- ³⁵ Sternberg, U.; Koch, F.-Th.; Prieß, W.; Witter, R. Crystal structure refinements of cellulose polymorphs using solid state ^{13}C chemical shifts; *Cellulose* **2003**, *10*, 189

References

- ³⁶ Jakovkin, I.; Klipfel, M.; Muhle-Goll, C.; Ulrich, A.S.; Luy, B.; Sternberg, U. Rapid calculation of protein chemical shifts using bond polarization theory and its application to protein structure refinement; in preparation
- ³⁷ Klipfel, M. Semiempirische Berechnung von Tensoren der chemischen Verschiebung zur Interpretation der NMR-Spektren von Biomolekülen; Dissertationsschrift, Universität Karlsruhe (TH), **2007**
- ³⁸ Gauss, J. Effects of Electron Correlation in the Calculation of Nuclear Magnetic Resonance Chemical Shifts; *J. Chem. Phys.* **1993**, *99*, 3629
- ³⁹ Witter, R.; Seyfart, L.; Greiner, G.; Reissmann, S.; Weston, J.; Anders, E.; Sternberg, U. NMR Structure determination of a pseudotriptide zinc complex with the COSY-NMR force field and DFT methods; *J. Biomol. NMR* **2002**, *24*, 277
- ⁴⁰ Dr. rer. nat. habil. Ulrich Sternberg, private communication
- ⁴¹ Cheeseman, J. R.; Trucks, G. W.; Keith, T. A.; Frisch, M. J. A comparison of model for calculating nuclear magnetic resonance shielding tensors; *J. Chem. Phys.* **1996**, *104*, 5497
- ⁴² Woolf, T. B.; Malkin, V. G.; Malkina, O. L.; Salahub, D. R.; Roux, B. The backbone ¹⁵N chemical shift tensor of the gramicidin channel. A molecular dynamics and density functional study; *Chem. Phys. Lett.* **1995**, *239*, 186
- ⁴³ Chekmenev E. Y.; Zhang, Q.; Waddell, K. W.; Mashuta, M. S.; Wittebort R. J. ¹⁵N Chemical Shielding in Glycyl Tripeptides: Measurement by Solid-State NMR and Correlation with X-ray Structure; *J. Am. Chem. Soc.* **2004**, *126*, 379
- ⁴⁴ Waddell, K. W.; Chekmenev E. Y.; Wittebort R. J. Single-Crystal Studies on Peptide Prolyl and Glycyl ¹⁵N Shielding Tensors; *J. Am. Chem. Soc.* **2005**, *127*, 9035
- ⁴⁵ Rienstra, C. M.; Tucker-Kellogg, L.; Jaroniec, C. P.; Hohwy, M.; Reif, B.; McMahon, M. T.; Tidor, B.; Lozano-Perez, T.; Griffin, R. G. *De novo* determination of peptide structure with solid-state magic-angle spinning NMR spectroscopy; *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 10260
- ⁴⁶ Zhang, H.; Neal, S.; Wishart, D. S. RefDB: A database of uniformly referenced protein chemical shifts; *J. Biomol. NMR* **2003**, *25*, 173

References

- 47 Sun, H.; Sanders, L.K.; Oldfield, E. Carbon-13 NMR Shielding in the Twenty Common Amino Acids: Comparisons with Experimental Results in Proteins; *J. Am. Chem. Soc.* **2001**, *124*, 5486-5495
- 48 Cornilescu, G.; Marquardt, J.L.; Ottiger, M.; Bax, A. Validation of Protein Structure from Anisotropic Carbonyl Chemical Shifts in a Dilute Liquid Crystalline Phase; *J. Am. Chem. Soc.* **1998**, *120*, 6836
- 49 Vijay-Kumar, S.; Bugg, C. E.; Cook, W. J. Structure of Ubiquitin refined at 1.8 Å resolution; *J. Mol. Biol.* **1987**, *194*, 513
- 50 Witter, R.; Sternberg, U.; Hesse, S.; Kondo, T.; Koch, F.-Th.; Ulrich, A. S. ¹³C Chemical Shift Constrained Crystal Structure Refinement of Cellulose I_α and Its Verification by NMR Anisotropy Experiments; *Macromolecules* **2006**, *39*, 6125
- 51 Witter, R.; Prieß, W.; Sternberg, U. Chemical Shift Driven Geometry Optimization; *J. Comput. Chem.*, **2002**, *23*, 298
- 52 Gaussian 03, Revision E.01, Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, Jr., J. A.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; and Pople, J. A.; Gaussian, Inc., Wallingford CT, **2004**.
- 53 COSMOS Pro, Sternberg, U.; Koch, F.-Th. and Losso, P. COSMOS-Software, Jena, Germany, **2011**, www.cosmos-software.de
- 54 Vila, J. A.; Baldoni, H. A.; Scheraga, H. A. Performance of Density Functional Models to Reproduce Observed ¹³C^α Chemical Shifts of Proteins in Solution; *J. Comput. Chem.* **2009**, *30*, 884

References

- 55 Csaszar, P. and Pulay, P. Geometry optimization by direct inversion in the iterative subspace; *J. Mol. Struct. (Theochem)* **1984**, *114*, 31
- 56 Farkas, Ö. and Schlegel, H. B. Methods for optimizing large molecules. II. Quadratic search; *J. Chem. Phys.* **1999**, *111*, 10806
- 57 Campanelli, A. R.; Ramondo, F.; Domenicano, A.; Hargittai, I. Toward a more accurate silicon stereochemistry: an electron diffraction study of the molecular structure of tetramethylsilane; *Struct. Chem.* **2000**, *11*, 155
- 58 Jameson, A. K.; Jameson, C. J. Gas-phase ^{13}C chemical shifts in the zero-pressure limit: refinements to the absolute shielding scale for ^{13}C ; *J. Chem. Phys. Lett.* **1997**, *134*, 461
- 59 van Nuland, N. A. J.; Hangyi, I. W.; van Schaik, R. C.; Berendsen, H. J. C.; van Gunsteren, W. F.; Sheek, R. M.; Robillard, G. T. The high-resolution structure of the histidine-containing phosphocarrier protein HPr from *Escherichia coli* determined by restrained molecular dynamics from nuclear magnetic resonance nuclear Overhauser effect data; *J. Mol. Biol.* **1994**, *237*, 544
- 60 Amann, B. T.; Worthington, M. T.; Berg, J. M. A. A Cys₃His zinc-binding domain from Nup475/Tristetraprolin: a novel fold with a disklike structure; *Biochem.* **2003**, *42*, 217
- 61 Babini, E.; Bertini, I.; Capozzi, F.; Del Bianco, C.; Hollender, D.; Kiss, T.; Luchinat, C.; Quattrone, A. Solution structure of human β -parvalbumin and structural comparison with its paralog α -parvalbumin and with their rat orthologs; *Biochem.* **2004**, *43*, 16076
- 62 Laskowski, R. A.; MacArthur M. W.; Moss, D. S.; Thornton, J. M. PROCHECK - a program to check the stereochemical quality of protein structures; *J. App. Cryst.* **1993**, *26*, 283
- 63 Napper, S.; Delbaere, L. T. J.; Waygood, B. E. J. Histidine-containing protein, HPr, of the *Escherichia coli* phosphoenolpyruvate: sugar phosphotransferase system can accept and donate a phosphoryl group; *J. Biol. Chem.* **1999**, *274*, 21776
- 64 Zasloff, M. Antimicrobial peptides of multicellular organisms; *Nature* **2002**, *415*, 389
- 65 Shai, Y. Mechanism of the binding, insertion and destabilization of phospholipid bilayer membranes by α -helical antimicrobial and cell non-selective membrane-lytic peptides; *Biochim. Biophys. Acta.* **1999**, *1462*, 55

References

- ⁶⁶ Strandberg, E.; Ulrich, A. S. NMR Methods for Studying Membrane-Active Antimicrobial Peptides; *Concepts Magn. Res.* **2004**, *23A*, 89
- ⁶⁷ Huang, H. W. Action of Antimicrobial Peptides: Two-State Model; *Biochem.* **2000**, *39*, 8347
- ⁶⁸ Strandberg, S.; Wadhvani, P.; Termouilhac, P.; Dürr, U. H. N., Ulrich, A. S. Solid-State NMR Analysis of the PGLa Peptide Orientation in DMPC Bilayers: Structural Fidelity of ²H-Labels versus High Sensitivity of ¹⁹F-NMR; *Biophys. J.* **2006**, *90*, 1676
- ⁶⁹ Moll, F. 3rd; Cross, T. A. Optimizing and characterizing alignment of oriented lipid bilayers containing gramicidin D. *Biophys. J.* **1990**, *57*, 351
- ⁷⁰ Wu, C. H.; Ramamoorthy, A.; Gierasch, L. M.; Opella, S. J. Simultaneous characterization of the amide ¹H chemical shift, ¹H-¹⁵N dipolar, and ¹⁵N chemical shift interaction tensors in a peptide bond by three-dimensional solid-state NMR spectroscopy; *J. Am. Chem. Soc.* **1995**, *117*, 6148
- ⁷¹ Shi, L.; Traaseth, N. J.; Verardi, R.; Cembran, A.; Gao, J.; Veglia, G. A refinement protocol to determine structure, topology, and depth of insertion of membrane proteins using hybrid solution and solid-state NMR restraints; *J. Biomol. NMR* **2009**, *44*, 195
- ⁷² Page, R. C.; Kim, S.; Cross, T. A. Transmembrane helix uniformity examined by spectral mapping of torsion angles; *Structure* **2008**, *16*, 787
- ⁷³ Santi, E.-M.; Strandberg, E.; Fuertes, G.; Ulrich, A. S.; Salgado, J. Influence of Whole-Body Dynamics on ¹⁵N PISEMA NMR Spectra of Membrane Proteins: A Theoretical Analysis; *Biophys. J.* **2009**, *96*, 3233
- ⁷⁴ Kelkar, D. A.; Chattopadhyay, A. The gramicidin ion channel: A model membrane protein; *Biochim. Biophys. Acta.* **2007**, *1768*, 2011
- ⁷⁵ Urry, D. W.; Goodall, M. C.; Glickson, J. D.; Mayers, D. F. The Gramicidin A Transmembrane Channel: Characteristics of Head-to-Head Dimerized $\pi_{(L,D)}$ Helices; *Proc. Natl. Acad. Sci. U.S.A.* **1971**, *68*, 1907
- ⁷⁶ Ketchum, R. R.; Hu, W.; Cross, T. A. High-Resolution Conformation of Gramicidin A in a Lipid Bilayer by Solid-State NMR; *Science*, **1993**, *261*, 1457

References

- 77 Vogt, B. T. C.; Schinzel, S.; Bechinger, B. Biosynthesis of isotopically labeled gramicidins and tyrocidins by *Bacillus brevis*; *J. Biomol. NMR* **2003**, *26*, 1
- 78 Lovell, S. C.; Davis, I. W.; Arendall, B. W.; de Bakker, P. I. W.; Word, J. M.; Prisant, M. G.; Richardson, J. S.; Richardson, D. C. Structure Validation by C α Geometry: ϕ, ψ and C β Deviation; *Protein Struct. Funct. Genet.* **2003**, *50*, 437
- 79 Ho, B. K.; Thomas, A.; Brasseur, R. Revisiting the Ramachandran plot: Hard-sphere repulsion, electrostatics, and H-bonding in the α -helix; *Protein Science* **2003**, *12*, 2508
- 80 Strandberg, E.; Santi, E.-M.; Salgado, J.; Ulrich, A. S. Orientation and Dynamics of Peptides in Membranes Calculated from ^2H -NMR; *Biophys. J.* **2009**, *96*, 3223
- 81 Fares, C.; Sharom, F. J.; Davis, J. H. $^{15}\text{N}, ^1\text{H}$ Heteronuclear Correlation NMR of Gramicidin A in DMPC- d_{67} ; *J. Am. Chem. Soc.* **2002**, *124*, 11232
- 82 Mai, W.; Hu, W.; Wang, C.; Cross, T. A. Orientational constraints as three-dimensional structural constraints from chemical shift anisotropy: The polypeptide backbone of gramicidin A in a lipid bilayer; *Protein Sci.* **1993**, *2*, 532
- 83 Sternberg, U.; Witter, R.; Anne, S. U. All-atom molecular dynamics simulations using orientational constraints from anisotropic NMR samples; *J. Biomol. NMR* **2007**, *38*, 23
- 84 Torda, A. E.; van Gunsteren, W. F. The refinement of NMR structures by molecular dynamics simulation; *Comput. Phys. Commun.* **1991**, *62*, 289
- 85 Limmer, S. Physical principles underlying the experimental methods for studying the orientational order of liquid crystals; *Fortschritte der Physik* **1991**, *37*, 879
- 86 Low, R. J. Measuring order and biaxiality; *Eur. J. Phys.* **2002**, *23*, 111
- 87 Sishi, T.; Case, D. A. Vibrational averaging of chemical shift anisotropies in model peptides; *J. Biomol. NMR* **2007**, *38*, 255
- 88 Roux, B.; Karplus, M. Ion Transport In a Gramicidin-like Channel: Dynamics and Mobility; *J. Phys. Chem.* **1991**, *95*, 4856
- 89 Roux, B.; Karplus, M. Ion Transport in a Gramicidin-like Channel: Structure and Thermodynamics; *Biophys. J.* **1991**, *59*, 961

References

- ⁹⁰ North, C. L.; Cross, T. A. Correlations between Function and Dynamics: Time Scale Coincidence for Ion Translocation and Molecular Dynamics in the Gramicidin Channel Backbone; *Biochem.* **1995**, *34*, 5883
- ⁹¹ Andersen, O.S. Ion movement through gramicidin A channels. Single-channel measurements at very high potentials; *Biophys. J.* **1983**, *41*, 119
- ⁹² Becker, M. D.; Koeppe, R. E.; Andersen, O.S. Amino acid substitutions and ion channel function. Model-dependent conclusions; *Biophys. J.* **1992**, *62*, 25
- ⁹³ Henzler-Wildman, K.; Kern, D. Dynamic personalities of proteins; *Nature*, **2007**, *450*, 964
- ⁹⁴ Wylie, B. J.; Rienstra, C. M. Multidimensional solid state NMR of anisotropic interactions in peptides and proteins; *J. Chem. Phys.* **2008**, *128*, 52207
- ⁹⁵ Arnér, E.S.J.; Holmgren, A. Physiological functions of thioredoxin and thioredoxin reductase; *Eur. J. Biochem.*, **2000**, *267*, 6102
- ⁹⁶ Marulanda, D.; Tasayco, M. L.; McDermott, A.; Cataldi, M.; Arriaran, V.; Polenova, T. Magic Angle Spinning Solid-State NMR Spectroscopy for Structural Studies of Protein Interfaces. Resonance Assignments of Differentially Enriched *E.coli* Thioredoxin Reassembled by Fragment Complementation; *J. Am. Chem. Soc.* **2004**, *126*, 16608
- ⁹⁷ Yang, J.; Paramasivam, S.; Marulanda, D.; Cataldi, M.; Tasayco, M. L.; Polenova, T. Magic Angle Spinning NMR Spectroscopy of Thioredoxin Reassemblies; *Magn. Reson. Chem.* **2007**, *45*, S73
- ⁹⁸ Yang, J.; Tasayco, M. L.; Polenova, T. Dynamics of Reassembled Thioredoxin Studied by Magic Angle Spinning NMR: Snapshots from Different Time Scales; *J. Am. Chem. Soc.* **2009**, *131*, 13690
- ⁹⁹ Jeng, M. F.; Campbell, A. P.; Begley, T.; Holmgren, A.; Case, D. A.; Wright, P. E.; Dyson, H. J. High-resolution solution structures of oxidized and reduced *Escherichia coli* thioredoxin; *Structure* **1994**, *2*, 853
- ¹⁰⁰ Peti, W.; Meiler, J.; Brüschweiler, R.; Griesinger, C. Model-Free Analysis of Protein Backbone Motion from Residual Dipolar Couplings; *J. Am. Chem. Soc.* **2002**, *124*, 5822

References

- ¹⁰¹ Sjöbring, U.; Björck, L.; Kastern, W. Streptococcal Protein G. Gene structure and protein binding properties; *J. Biol. Chem.* **1991**, *266*, 399
- ¹⁰² Franks, W. T.; Donghua, H. Z.; Wylie, B. J.; Money, B. G.; Graesser, D. T.; Frericks, H. L.; Sahota, G.; Rienstra, M. C. Magic-Angle Spinning Solid-State NMR Spectroscopy of the $\beta 1$ Immunoglobulin Binding Domain of Protein G (GB1): ^{15}N and ^{13}C Chemical Shift Assignments and Conformational Analysis; *J. Am. Chem. Soc.* **2005**, *127*, 12291
- ¹⁰³ Wylie, B. J.; Franks, W. T.; Rienstra, M. C. Determinations of ^{15}N Chemical Shift Anisotropy Magnitudes in a Uniformly $^{15}\text{N},^{13}\text{C}$ -Labeled Microcrystalline Protein by Three-Dimensional Magic-Angle Spinning Nuclear Magnetic Resonance Spectroscopy; *J. Phys. Chem. B* **2006**, *110*, 10926
- ¹⁰⁴ Franks, W. T.; Wylie, B. J.; Frericks, H. L.; Nieuwkoop, A. J.; Mayrhofer, R. M.; Shah, G. J.; Graesser, D. T.; Rienstra, M. C. Dipole tensor-based atomic-resolution structure determination of a nanocrystalline protein by solid-state NMR; *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 4621
- ¹⁰⁵ Lewandowski, J. R.; Sein, J.; Sass, H.-J.; Grzesiek, S.; Blackledge, M.; Emsley, L. Measurement of Site-Specific ^{13}C Spin-Lattice Relaxation in a Crystalline Protein; *J. Am. Chem. Soc.* **2010**, *132*, 8252
- ¹⁰⁶ Saito, H.; Ando, I.; Ramamoorthy, A. Chemical shift tensor – the heart of NMR: Insights into biological aspects of proteins; *Prog. Nucl. Mag. Res. Sp.* **2010**, *57*, 181
- ¹⁰⁷ Frericks Schmidt, H. L.; Sperling, L. J.; Gao, Y. G.; Wylie, B. J.; Boettcher, J. M.; Wilson, S. R.; Rienstra, M. C. Crystal polymorphism of protein GB1 examined by solid-state NMR spectroscopy and X-ray diffraction; *J. Phys. Chem. B* **2007**, *111*, 14362
- ¹⁰⁸ Ulmer, T. S.; Ramirez, B. E.; Delaglio, F.; Bax, A. GB3 solution structure obtained by refinement of X-ray structure with dipolar couplings; *J. Am. Chem. Soc.* **2003**, *125*, 9179
- ¹⁰⁹ Wishart, D. S.; Bigam, C. G.; Yao, J.; Abildgaard, F.; Dyson, H. J.; Oldfield, E.; Markley, J. L.; Sykes, B. D. ^1H , ^{13}C and ^{15}N chemical shift referencing in biomolecular NMR; *J. Biomol. NMR* **1995**, *6*, 135
- ¹¹⁰ Schlegel, H. B. Optimization of Equilibrium Geometries and Transition Structures; *J. Comp. Chem.* **1982**, *3*, 214

References

- ¹¹¹ Allen, F. H.; Kennard, O.; Watson, D. G.; Brammer, L.; Orpen A. G.; Tylor, R. Tables of Bond Lengths determined by X-Ray and Neutron Diffraction. Part 1. Bond Lengths in Organic Compounds; *J. Chem. Soc. Perkin. Trans. II* **1987**, 12, S1

Deutsche Zusammenfassung

Festkörper-Kernresonanzspektroskopie ist eine leistungsfähige Methode mit welcher die Struktur und Dynamik der Peptide und Proteine in Membranen und Mikrokristallen, sowie die Orientierung der Peptide in der Membran untersucht werden können. Chemische Verschiebungstensoren erlauben Rückschlüsse auf die Struktur, Dynamik und Orientierung in der Membran auch dann, wenn weitere experimentelle Daten nicht verfügbar sind. *Ab initio* Methoden zur Berechnung der chemischen Verschiebung sind zu rechenintensiv, um damit Hilfe Proteine untersuchen zu können. Die Zielsetzung der vorliegenden Arbeit besteht deshalb darin, die Bindungspolarisationstheorie – ein semiempirisches Verfahren zur Berechnung der chemischen Verschiebungstensoren – für die Berechnung der chemischen Verschiebung in Proteinen und Peptiden derart zu adaptieren, dass mit ihrer Hilfe Rückschlüsse auf die Dynamik (und Orientierung) von Proteinen und Peptiden, wie z.B. dem membranständigen Gramicidin A oder den mikrokristallinen Thioredoxin und GB1, zu gewinnen.

Bindungspolarisationstheorie konnte, nach einer Korrektur der ^{13}C -Parameterisierung, erfolgreich auf die Berechnung der chemischen Verschiebung in Proteinen angepasst werden. So ließen sich die $^{13}\text{C}_\alpha$ chemischen Verschiebungen im Protein Ubiquitin auf dem DFT / B3LYP 6-311++G(d,p) Niveau in nur 20.46 s berechnen. Es wurde deutlich, dass experimentell gewonnene Proteinstrukturen i.d.R. nicht hinreichende Qualität für die Berechnung der ^{15}N chemischen Verschiebung haben. Daher wurde ein Protokoll für die Geometrieoptimierung an Proteinen mit chemischen Verschiebungen als Zielfunktionen ausgearbeitet.

Die Dynamik und die Orientierung des antimikrobiellen Peptids Gramicidin A in der Membran wurde mit Hilfe der Moleküldynamiksimulation mit orientierenden Randbedingungen (experimentellen ^1H - ^{15}N dipolaren Kopplungen) untersucht. Für jeden Schritt der Dynamiksimulation wurden die ^{15}N chemischen Verschiebungstensoren berechnet. Daraus wurden die Ordnungstensoren bestimmt, mit deren Hilfe zum ersten Mal die theoretisch vorhergesagten korrelierten Bewegungen im Gramicidin A-Rückgrat in einer Untersuchung mit experimentellen Daten detektiert werden konnten. Zusätzlich konnten mit Hilfe der Moleküldynamiksimulation und der Berechnungen der ^{15}N chemischen Verschiebungstensoren das unregelmäßig aussehende PISEMA-Spektrum von Gramicidin A erfolgreich simuliert und erklärt werden.

Für die Dynamikuntersuchungen an den mikrokristallinen Proteinen Thioredoxin und GB1 ist ein neuer Ordnungsparameter, der auf dem Vergleich des berechneten ^{15}N chemischen Verschiebungstensors mit dem experimentellen, durch die Bewegung beeinflussten, ^{15}N chemischen Verschiebungstensor basiert, eingeführt worden. Die Ordnungsparameter für Thioredoxin waren im Einklang mit den bisherigen Ergebnissen zur Dynamik von Thioredoxin. Für GB1 zeigt sich schließlich, dass seine β -Faltblätter eine höhere Dynamik als die entsprechende α -Helix aufweisen. Eine ähnliche Verteilung der Dynamik im Proteinrückgrat ist im fast identischen Protein GB3 in Flüssigkeit beobachtet worden.

Ehrenwörtliche Erklärung

Ich erkläre hiermit, dass ich die vorliegende Arbeit selbständig verfasst und keine anderen als die angegebenen Hilfsmittel verwendet habe. Darüber hinaus erkläre ich, dass die Arbeit nicht anderweitig als Prüfungsarbeit oder als Dissertation bei einer anderen Fakultät verwendet wird oder wurde.

Karlsruhe, den 18. 05. 2011

Lebenslauf

Persönliche Daten

Name	Igor Jakovkin
Geburtsdatum	06.10.1982
Geburtsort	Sankt-Petersburg
Staatsangehörigkeit	deutsch
Familienstand	ledig

Schulischer Werdegang

1989-1996	Grundschule und Gymnasium in St.-Petersburg
1997-1998	Wolffskeel-Realschule in Würzburg
1998-2003	Hans-Geiger-Gymnasium in Kiel

Studium

2003-2007	Studium der Biochemie an der Christian-Albrechts-Universität in Kiel
Abschluss	Diplom (Note 1.0)

Promotion

seit 2008	Doktorarbeit am Karlsruhe Institute of Technology
Mai 2008	Abgabe der Dissertation

Danksagung

Meinem Betreuer Dr. rer. nat. habil. Ulrich Sternberg danke ich ganz herzlich für ein spannendes Arbeitsthema, sein stetiges Interesse an meiner Arbeit, die wunderbare Arbeitsatmosphäre und seine uneingeschränkte Unterstützung.

Herrn Professor Dr. Burkhard Luy danke ich herzlich für die Übernahme des Referats dieser Dissertation und die sehr freundliche Aufnahme in seiner Arbeitsgruppe.

Frau Professor Dr. Anne S. Ulrich danke ich herzlich für die Übernahme des Korreferats und dafür, dass sie es mir ermöglichte, die ersten zwei Jahre der Dissertation in ihrem Arbeitskreis zu arbeiten.

Professor Dr. Burkhard Bechinger danke ich für die Abbildung des PISEMA-Spektrums von Gramicidin A und für die Erlaubnis diese Abbildung zu benutzen.

Professor Dr. Tatyana Polenova danke ich für die Überlassung von Daten zur Zuordnung der chemischen Verschiebungen von Thioredoxin.

Herrn Dr. Raiker Witter danke ich für viele anregende Diskussionen.

Außerdem gilt mein ausdrücklicher Dank Frau Professor Dr. Anne S. Ulrich, Frau Dr. Marina Berditsch und Herrn Dr. Sergiy Afonin für die angenehme und fruchtbare Zusammenarbeit an den Projekten, die nicht in die vorliegende Arbeit eingegangen sind.

Mein besonderer Dank gilt meinen Eltern, Boris und Bronislava Jakovkin, die mir das Studium ermöglichten und mich in jeder Lebenslage unterstützen.