

^{13}C sparse labeling for solid-state NMR investigations of biomolecular systems

Dissertation

Zur Erlangung des mathematisch-naturwissenschaftlichen Doktorgrades

”Doctor rerum naturalium”

der Georg-August-Universität Göttingen

vorgelegt von

Hannes Klaus Faßhuber

aus Judenburg, Österreich

Göttingen, 2014

Members of the thesis committee and examination board:

First Reviewer:

Prof. Dr. Adam Lange

Leibniz-Institut für Molekulare Pharmakologie und Institut für Biologie, Humboldt-Universität zu Berlin, Germany

Second Reviewer:

Prof. Dr. Marina Bennati

Electron Spin Resonance Spectroscopy Group, Max Planck Institute for Biophysical Chemistry, Göttingen, Germany

Dr. Lars Kuhn

NMR Spectroscopy, European Neuroscience Institute, Göttingen, Germany

Members of the extended thesis examination board:

Prof. Dr. Christian Griesinger

Department of NMR-based Structural Biology, Max Planck Institute for Biophysical Chemistry, Göttingen, Germany

Prof. Dr. Bert de Groot

Computational biomolecular dynamics group at the Max Planck Institute for Biophysical Chemistry

Prof. Dr. Kai Tittmann

Department of Bioanalytics, George-August-Universität Göttingen, Germany

Tag der mündlichen Prüfung: 4.12.2014

Affidavit

I hereby declare that this dissertation has been written independently and with no other sources and aids than quoted.

.....

Hannes Klaus Fäßhuber

List of contents

Members of the thesis committee and examination board:	III
Affidavit.....	IV
List of contents.....	V
List of Symbols and Abbreviations	IX
Abstract.....	XIII
Introduction	XV
Scope of the Thesis	XVII
PART I Ubiquitin.....	1
1 Introduction.....	2
2 Theoretical aspects of the applied methods	3
2.1 General principles of NMR.....	3
2.1.1 Quantum mechanical description of a statistical ensemble of spin $\frac{1}{2}$ nuclei	4
2.1.1.1 Density operator.....	4
2.1.1.2 Time evolution of the density operator.....	4
2.1.2 Magnetic nuclear spin Hamiltonian \hat{H}_s	5
2.1.2.1 External spin interaction.....	5
2.1.2.2 Internal spin interaction \hat{H}_{int}	6
2.1.2.2.1 Chemical shift \hat{H}_{CS}	6
2.1.2.2.2 Dipolar coupling \hat{H}_{DD}	7
2.1.2.2.3 Scalar or J coupling \hat{H}_J	8
2.1.2.3 Comparison of spin Hamiltonians.....	8
2.2 NMR Signal	9
2.3 Solid-state NMR.....	11
2.3.1 Definitions of various frames in ssNMR.....	11
2.3.2 Magic angle spinning (MAS)	12
2.3.3 Cross polarization (CP).....	12
2.3.4 2D Homonuclear correlation spectroscopy	13
2.3.5 2D Heteronuclear correlation spectroscopy.....	14
2.3.6 Band-selective homonuclear CO-C α CP transfer (BSH-CP)	15
2.4 Isotope labeling	16
2.4.1 General overview.....	16
2.4.2 Sparsely labeling strategies	16
2.5 Structure determination	17
2.6 Structure validation	19
3 Experimental aspects of the applied Methods.....	20
3.1 NMR experiments.....	20
3.1.1 Ubiquitin (MPD).....	20
3.1.1.1 1D Experiments of the isotopes ^{13}C and ^{15}N (MPD).....	20
3.1.1.2 2D Homonuclear experiments (MPD).....	20
3.1.1.3 2D Heteronuclear experiments (MPD)	20
3.1.1.4 BSH 2D Heteronuclear experiments (MPD).....	21
3.1.2 Ubiquitin (PEG)	21
3.2 Prediction of protein backbone torsion angles restraints	21
3.3 Structure calculation and refinement.....	21
3.4 Validation of the calculated structures.....	22
4 Materials.....	23
4.1 Isotope labeled reagents	23
4.2 Sample preparation of ubiquitin with MPD.....	23
4.3 Sample preparation of ubiquitin with PEG	24
5 Results	25
5.1 Results of Ubiquitin MPD.....	25
5.1.1 Verification of the sample quality.....	25
5.1.2 A comparison of spectral quality	25
5.1.3 Assignment of ubiquitin	26
5.1.4 Stereospecific assignment from sparsely labeling approach.....	28
5.1.5 Sequential assignment of ubiquitin	29

5.1.6	Secondary structure analysis of ubiquitin	33
5.1.7	Collecting of distance restraints information	35
5.1.8	Structure calculation.....	37
5.1.9	Comparison to different accessible PDB structures.....	40
5.1.9.1	Structural comparison to the X-ray structure 3ONS	40
5.1.10	Structure calculation of the second conformation	42
5.1.11	Structural comparison between the main and the second conformation.....	43
5.2	Results of ubiquitin PEG	44
5.2.1	Assignment of ubiquitin	44
5.2.2	Secondary structure analysis of conformation A.....	45
5.2.3	Collection of distance restraints information	46
5.2.4	Structure calculation.....	47
5.2.5	Comparison to different accessible PDB structures.....	48
5.2.5.1	Structural comparison to the X-ray structure 1UBQ.....	48
5.2.6	Structure calculation of conformation B.....	50
5.2.7	Structural comparison between conformation A and B	51
5.3	Comparison between MPD and PEG samples	52
5.3.1	Structural comparison between the main conformation of MPD and the conformation A of PEG	54
6	Conclusion	55
6.1	Conclusion MPD	55
6.2	Conclusion PEG	56
PART II	MxiH	57
7	Introduction	58
8	Material	60
8.1	Sample preparation	60
9	Methods	61
9.1	General aspects	61
9.2	Carbon-detected ssNMR.....	61
9.3	Proton-detected ssNMR	62
10	Results and Discussion.....	63
10.1	Isotopic labeling pattern.....	63
10.2	Dynamics in the biomolecular assembly	66
10.3	¹³ C- ¹³ C Distance restraints	67
10.4	¹⁵ N- ¹³ C Distance restraints	69
10.5	Proton-detected ssNMR	70
11	Conclusions.....	73
12	Appendix I.....	74
12.1	Ubiquitin MPD	74
12.1.1	SsNMR experiments (MPD)	74
12.1.2	Chemical shift lists (MPD).....	75
12.1.3	Secondary structure analysis of the second-conformation of Ubiquitin (MPD).....	78
12.1.4	Validation results of Ubiquitin (MPD).....	79
12.1.4.1	Validation result of the main conformation (MPD)	79
12.1.5	Validation result of the second conformation (MPD).....	81
12.1.6	Comparison to different structures (MPD)	83
12.1.6.1	Comparison between the main conformation (MPD) to the bundle of structures of 2JZZ	83
12.1.6.2	Comparison between the main conformation (MPD) to the bundle of structures of 1D3Z.....	83
12.2	Ubiquitin PEG.....	84
12.2.1	SsNMR experiments (PEG).....	84
12.2.2	Chemical shift lists (PEG)	85
12.2.3	Network distance restraints between I3Cy2 and V26Cα (PEG)	89
12.2.4	Secondary structure analysis of the conformation B (PEG)	90
12.2.5	Validation results of Ubiquitin (PEG)	91
12.2.5.1	Validation result of conformation A (PEG).....	91
12.2.5.2	Validation result of conformation B (PEG).....	93
12.2.6	Comparison between conformation A (PEG) to the bundle of structures of 1D3Z	95
12.3	Pulse sequence of 2D NCA and NCO experiments	98
12.4	Pulse sequence of 2D NCACX and NCOCX experiments	98
12.5	Structure calculation.....	99

12.5.1	Covalent bond energy term Ebond	99
12.5.2	Angle energy term Eangle.....	99
12.5.3	Dihedral energy term Edihedral	99
12.5.4	Van der Waal energy term Evdw.....	100
12.5.5	electrostatic energy term Eel	100
12.5.6	Distance restraint energy term Eexperimental	101
12.5.7	Inputfile of the x-plor-NIH calculation	102
13	Appendix II.....	108
13.1	Comparison between 1D ^{13}C spectra of [1- ^{13}C]-glc and [U- ^{15}N -[ILV- ^{13}C Methyl]]-labeled sample	108
13.2	Comparison between 1D ^{13}C spectra of [U- ^{13}C]-glc and [U- ^{15}N -[ILV- ^{13}C Methyl]]-labeled sample.....	108
13.3	Column projection of the ^{13}C chemical shift dimension	109
13.4	Pulse sequences for ^1H detection ssNMR experiments.....	110
References	113
Acknowledgements	117
Curriculum Vitae	118

List of Symbols and Abbreviations

α -ketoisovalerate	2-keto-3-(methyl- ¹³ C)-butyrate-4- ¹³ C
α -ketobutyrate	2-ketobutyrate -4- ¹³ C
CdCl ₂	Cadmium chloride
Cd-[acetate] ₂	Cadmium acetate
DSS	4,4-dimethyl-4-silapentane-1-sulfonic acid
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
MPD	2-methyl-2,4-pentandiol
¹⁵ NH ₄ Cl	¹⁵ N labeled ammonium chloride
PEG	polyethyleneglycol
[1- ¹³ C]-glc	[1- ¹³ C]-glucose
[2- ¹³ C]-glc	[2- ¹³ C]-glucose
[U- ¹³ C]-glc	[U- ¹³ C]-glucose
[1,3- ¹³ C]-glyc	[1,3- ¹³ C]-glycerol
[2- ¹³ C]-glyc	[2- ¹³ C]-glycerol
[1,2- ¹³ C]-pyr	[1,2- ¹³ C]-pyruvate
[1- ¹³ C]-pyr	[1- ¹³ C]-pyruvate
[2- ¹³ C]-pyr	[2- ¹³ C]-pyruvate
[3- ¹³ C]-pyr	[3- ¹³ C]-pyruvate
[LV- ¹³ C Methyl]	[Leu-Val- ¹³ C Methyl]
[U- ¹⁵ N-[ILV- ¹³ C Methyl]]	[U- ¹⁵ N-[Ile-Leu-Val- ¹³ C Methyl]

A	second-rank cartesian interaction tensor
\hat{A}	general expression of a macroscopic observable
a _j	amplitude of the signal
AQ	acquisition time
B₀	static external magnetic field
B₁	radio frequency field
B_{induced}	induced magnetic field
BSH-CP	band-selective homonuclear CP
c	slope of the asymptote of the soft square function
C	soft square function scaling factors
CP	cross-polarisation
C _{rep}	force constant
CSA	chemical shift anisotropy
CW	continuous wave
d	distance between two nuclei in the soft square function
DARR	dipolar recoupling enhancement through amplitude modulation
D_{jk}	second rank tensor for the dipole-dipole interaction between the nuclei _{j,k}
ΔE	energy differences
E _{angle}	angle energy
E _{bond}	covalent bond energy
E _{dihe}	normal dihedral energy
E _{dihedral}	total dihedral energy
E _{el}	electrostatic energy

$E_{\text{experimental}}$	energy of the experimental NMR data
$E_{\text{force field}}$	force field energy
E_{impr}	improper dihedral energy
E_{vdw}	van der Waals energy
FID	free-induction decay
$f_{el}(R)$	electrostatic energy function
$f_{\text{vdw}}(\mathbf{R})$	van der Waals function
\hbar	Planck constant (divided by 2π)
\hat{H}	time-independed Hamiltonian operator
\hat{H}_{CS}	chemical shift Hamiltonian
\hat{H}_{DD}	direct dipole-dipole coupling Hamiltonian
\hat{H}_{ext}	external spin Hamiltonian
\hat{H}_{int}	internal spin Hamiltonian
\hat{H}_{J}	scalar coupling (J coupling) Hamiltonian
\hat{H}_{rf}	radio frequency spin Hamiltonian
\hat{H}^{S}	spin Hamiltonian
\hat{H}_{static}	Zeeman spin Hamiltonian
$\hat{H}_{(t)}$	time-depended Hamiltonian operator
\mathbf{I}	nuclear spin operator
$\hat{\mathbf{I}}$	spin angular momentum operator
$\hat{I}_x \hat{I}_y \hat{I}_z$	spin angular momentum operator
\mathbf{J}_{jk}	second rank tensor for the scalar interaction between the nuclei _{j,k}
INEPT	Insensitive nuclei enhanced by polarization transfer
k	Boltzmann constant
k_{b}	bond force constants
k_{θ}	angle force constants
k_{ϕ_i}	dihedral angle force constants
kDa	kiloDalton
kHz	kiloHertz
K	Kelvin
\mathcal{L}	complex Lorentzian term
m_i	multiplicities
MAS	magic angle spinning
MD	molecular dynamic
MHz	megaHertz
\mathbf{M}_{-y}	net magnetization vector along the -y axes
\mathbf{M}_z	net equilibrium magnetization vector
n_i	periodicities
$N_{\alpha,\beta}$	population number of the eigenstates α,β
NMR	nuclear magnetic resonance
NS	number of scans
p	probability term
PAIN	proton-assisted insensitive nuclei
PAS	principal axis
PDB	protein data bank

PDSD	proton driven spin diffusion
PES	potential energy surface
PPP	pentose phosphate pathway
$Q_i Q_j$	electric charge of atoms
r	actual bond length
r_0	equilibrium bond length
rep, rexp, irexp	van der Waals parameters
rf	radio frequency
rmsd	root of mean squares deviation
R	rotation matrix
R^{-1}	inverse rotation matrix
\hat{R}	exponential rotation operator
\hat{R}^{-1}	inverse exponential rotation operator
R	actual van der Waals radius
REDOR	rotational echo double resonance
RFDR	radiofrequency-driven recoupling
RD	recycling delay
RDC	residual dipolar coupling
Ref.	reference
rmsd	root of mean squares deviation
$S_j(t)$	time domain signal
S	soft square function scaling factors
$S_j(\Omega)$	frequency domain signal
SPINAL	small phase incremental alteration
ssNMR	solid-state NMR
SW	spectrum width
t	time variable
T	temperature
TPP	thiamine pyrophosphate
TEDOR	transferred echo double resonance
T3SS	type three secretion system
WALTZ	wonderful alternating phase technique for zero residual splitting
X	magnetic field or nuclear spin or angular momentum operator
γ	gyromagnetic ratio
Γ	superoperator for describing relaxation mechanism
δ_i	phase shifts
ϵ_0	electric permittivity of free space
$\sigma_{zz}(\theta)$	principal value of the chemical shift tensor
θ_0	equilibrium constants
θ	angle of the chemical shield tensor in the PAS frame
θ_r	axis of rotation (magic angle)
Θ	actual value of the angle
λ_j	spin-lattice relaxation rate
$\hat{\mu}$	nuclear magnetic momentum operator
$\hat{\rho}(t)$	time-dependend density operator

σ	Lennard-Jones constant
$\hat{\sigma}_{(0)}$	time-depended reduced spin density operator in equilibrium
$\hat{\sigma}_{(t)}$	time-depended reduced spin density operator
ϕ_i	actual dihedral angle
ψ	time-independed wave function
$\Psi_{(t)}$	time-depended wave function
ω	Larmor frequency (in angular frequency units)
Ω	resonance frequency
\mathfrak{R}	distance between selected stets of atoms

Abstract

This thesis refers to ^{13}C sparsely labeled strategies, including protein structure characterization on two different biomolecular systems, to obtain structural information via ^{13}C - ^{13}C and ^{13}C - ^{15}N correlations by solid-state nuclear magnetic resonance (ssNMR).

By applying [1- ^{13}C]- and [2- ^{13}C]-glc labeling schemes to the folded globular protein ubiquitin, a strong reduction of spectral crowding and increase in resolution in ssNMR spectra can be achieved for two different precipitation conditions (MPD and PEG). This allowed spectral resonance assignment in a straightforward manner and the collection of an unprecedented wealth of long-range distance information. High precision solid-state NMR structures of microcrystalline ubiquitin with a backbone root mean squared deviation (rmsd) of 0.7 Å were calculated in both conditions. A backbone accuracy of 1.57 Å (MPD) and 1.88 Å (PEG) to the concerning X-ray structures could be calculated. In the comparison between the lowest energy structures of the two systems one can declare that the 3D fold of Ubiquitin is identical. A global backbone rmsd value of 1.63 Å is calculated (residue M1-70V). Small site specific conformational deviations can be identified in the regions (L8-T12, D21, E34-I36 and for E51-G53). Interestingly, one can resolve similar structural heterogeneity in both crystallization conditions. For the MPD system structural heterogeneity is present for β -strands β 1, β 2, β 3 and β 5 as well as for the loop regions β 1- β 2 and α 1- β 3. In the PEG condition one can distinguish structural heterogeneity for the first and second β -strand β 1, β 2, residue I23, at the tip of the α -helix α 1, residue D39, the third β -strand β 3, residue L50, residue I61, the fifth β -strand β 5 and residue L69. This structural polymorphism observed in the solid-state NMR spectra coincides with regions that were found to be involved in conformational dynamics of ubiquitin on the ns to μ s time scale, as reported in recent residual dipolar coupling (RDC)-based measurements and relaxometry experiments. We suggest that the conformational sampling of ubiquitin manifests itself as structural heterogeneity during the crystallization process.

The second isotope labeling strategy is based on the inclusion of two biosynthetic precursors in the bacterial growth medium, α -ketoisovalerate and α -ketobutyrate, leading to the production of leucine, valine and isoleucine residues that are only ^{13}C labeled on methyl groups. The resulting spectral simplification facilitates the collection of distance restraints, the verification of carbon chemical shift assignments and the measurement of methyl group dynamics. This approach is demonstrated on the Type-Three Secretion System needle of *Shigella flexneri*, where 33 unambiguous distance restraints could be collected. By combining this labeling scheme with ultra-fast MAS and proton detection, the assignment of methyl proton chemical shifts was achieved. This method can be applied for studying protein properties within large biological assemblies.

Introduction

Proteins play a central role in the structure and function of biological systems, and research on structure determination of proteins is still an important and crucial field in life sciences.

The first structural features of a protein have been realized on Myoglobin by X-ray diffraction¹, which was honored by a Nobel Prize for chemistry in 1962. The other important method for obtaining atomic resolution structure of proteins is nuclear magnetic resonance (NMR) spectroscopy, which was developed independently by Felix Bloch and Edward Mills Purcell in 1945 (Nobel Prize 1952). In general the research fields of NMR can be addressed to three different research categories (magnetic resonance imaging (MRI)², solution NMR³ and solid-state NMR (ssNMR)).⁴ In this thesis we focus only on ssNMR spectroscopy application for structure elucidation of proteins.

The brilliant implementation of Fourier-transformation NMR Spectroscopy by Richard Ernst (Nobel Prize 1991) and the development of magic angle spinning by Andrew et.al^{5,6} (1958) facilitate the recording of high resolution ssNMR spectra. Further approaches such as the use of high external magnetic fields (~20 Tesla), developing of new pulse sequences⁷⁻¹², proton detection at high magic angle spinning rates¹³, isotopic labeling strategies¹⁴ and highly sophisticated structure calculation programs^{15,16} permit the structural investigation of proteins using ssNMR. The major steps in structural studies of biomolecules by ssNMR are the assignment of chemical shifts, the collection of structural restraints such as long-range distance restraints, and the study of local dynamics.

Initial investigations to extract structural information can be performed by uniformly labeled samples. A major drawback of this labeling scheme is located in the large number of ¹³C labeled carbons, which contributes to low spectral resolution and high ambiguity of ¹³C-¹³C correlations (Figure 1).¹⁷ One further crucial disadvantage of uniformly labeled samples can be addressed to the presence of strong dipolar truncation which reduces the transfer of magnetization and hinders the detection of long-range distance correlations.

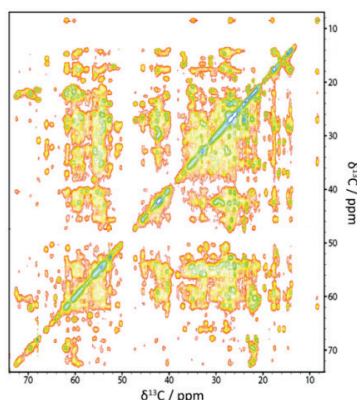


Figure 1 2D ¹³C-¹³C PDSD-spectrum of microcrystalline uniformly labeled ubiquitin, at a spinning frequency of 12kHz and a mixing time of 100ms. Ref.¹⁷

Successful improvements could be achieved by highly diluted uniformly labeled samples using minimal growth medium conditions with a ^{13}C carbon and an undetectable ^{12}C carbon labeled ratio of 10%:90%.¹⁸ Advantages of sparsely ^{13}C labeled protein samples result in an improvement of spectral resolution and in a strong reduction of cross-peak overlapping in multi-dimensional ssNMR spectra. Both improvements lead to an easier determination of unambiguous structural information. Recent examples of complementary labeling schemes used in ssNMR comprise [1,3- ^{13}C]-glycerol (glyc), [2- ^{13}C]-glyc¹⁹⁻²², [1- ^{13}C]-glucose (glc), and [2- ^{13}C]-glc.^{23,24}

In our group we could demonstrate significant improvements in the resolution of ^{13}C - ^{13}C ssNMR spectra by using [1- ^{13}C]-glc and [2- ^{13}C]-glc sparsely labeled proteins.^{24,25} With these complementary labeling schemes, a strategy was presented to determine intermolecular interactions of proteins in a self-assembled molecular system. Further benefits from the [2- ^{13}C]-glc labeling strategy allow the easy determination of stereospecific assignment for the amino acids valine and leucine²⁶ and the recording of long-range distance restraints in 3D spectra.²⁷

Other labeling strategies such as sparsely ^{13}C labeled methyl groups were employed to gain structural, dynamic and functional information about proteins.²⁸ In addition, unambiguous long-range correlations between methyl groups are crucial as they tend to be found in the hydrophobic core of proteins. The ^{13}C methyl labeling strategy leads to an improvement of the information content obtained from the methyl-methyl region of ^{13}C - ^{13}C 2D spectra by significant reduction in ambiguity of methyl-methyl cross-peaks. This approach allows us to obtain unambiguous distance restraints between carbons of methyl groups that are located closely in space.

Solid-state NMR spectroscopy is an ideal methodology for structural investigation due to the situation that no long-range order or fast tumbling is required as for X-ray or solution NMR studies, respectively. Biomaterials, which have already been studied by ssNMR, include microcrystalline proteins¹⁷, amyloid fibrils²⁹⁻³², membrane proteins³³⁻³⁵, virus capsids^{36,37}, bacterial filamentous proteins^{38,39} and fibrillar proteins such as collagen^{40,41}, keratin^{42,43} or silk.^{44,45}

Scope of the Thesis

Structural investigation of proteins by solid-state NMR (ssNMR) spectroscopy is primarily based on the long-range distance information of various nuclei such as proton, carbon and nitrogen. The resolution of the final protein structure is determined by the quantity, accuracy and ambiguity of this distance information. In the present thesis, these aspects are explored and investigated by various labeling strategies (with different amino acid precursors) and different crystallization procedures on two different biological systems. One approach will be demonstrated by the alternative [1-¹³C]- and [2-¹³C]-glc sparsely labeling strategy for determining the globular fold of the standard-model protein ubiquitin. Two different types of microcrystalline Ubiquitin samples are produced by the two common precipitation agents 2-methyl-2,4-pentandiol (MPD) and polyethylenglycol (PEG). The second labeling strategy is based on the inclusion of two biosynthetic precursors in the bacterial growth medium, α -ketoisovalerate and α -ketobutyrate, leading to the production of leucine, valine and isoleucine residues that are only ¹³C labeled on methyl groups. This approach was tested and analyzed on the Type-Three Secretion System needle of *Shigella flexneri*.

PART I Ubiquitin

1. Ubiquitin in MPD

2. Ubiquitin in PEG

Related Publications:

- Chaowei Shi, **Hannes K. Fasshuber**, Veniamin Chevelkov, Shengqi Xiang, Birgit Habenstein, Suresh Kumar Vasa, Stefan Becker, Adam Lange, BSH-CP based 3D solid-state NMR experiments for protein resonance assignment, *JBNMR*, 2014, 59, 1, 15-22.
- Veniamin Chevelkov, Chaowei Shi, **Hannes K. Fasshuber**, Stefan Becker, Adam Lange, Efficient band-selective homonuclear CO-CA cross-polarization in protonated proteins, *JBNMR*, 2013, 56,4,303-311.
- Guohua Lv, **Hannes K. Fasshuber**, Antoine Loquet, Jean-Philippe Demers, Vinesh Vijayan, Karin Giller, Stefan Becker, Adam Lange, A Straightforward Method for Stereospecific Assignment of Val and Leu Prochiral Methyl Groups by Solid-State NMR: Scrambling in the [2-¹³C]Glucose Labeling Scheme, *JMR*, 2013, 228, 45-49.

Note:

The BSH-CP based 3D ssNMR experiments of Ubiquitin precipitated in MPD were conducted by Dr. Chaowei Shi

1 Introduction

Ubiquitin, a small (8.5 kDa) regulatory protein, is a model protein well-studied by NMR spectroscopy.^{17,46-49} Main focus of both solution- and solid-state-NMR investigations is to address atomic resolution structural information as well as dynamic studies.⁵⁰⁻⁵² The accessibility to high-resolution crystal structures^{46,49} enables the direct comparison and cross validation of NMR based calculated structures (Figure 2).

In the context of this thesis microcrystalline ubiquitin crystals were obtained by the two prevailing precipitation agents polyethylenglycol (PEG) and 2-methyl-2,4-pentanediol (MPD), respectively. Both crystallization preparations end up in resulting high resolution ssNMR spectra.^{47,48} Different crystallization conditions results in small chemical shift variations for flexible regions as reported in recent publications.⁵³

Structure determination of proteins at atomic resolution remains a major challenge. One of the main issues in ssNMR is the detection and identification of distance restraints that reveal the 3D protein structure. The sequential assignment of proteins and the identification of distance restraints are often abrogated by the limits in the resolution of ssNMR spectra. New approaches in preparation of sparsely ¹³C labeled proteins using [1,3-¹³C]-glyc, [2-¹³C]-glyc¹⁹⁻²², [1-¹³C]-glc and [2-¹³C]-glc have made big advances in recording of high-resolution ssNMR spectra.^{23,24}

In our group, we demonstrated and explored the beneficial use of [1-¹³C]- and [2-¹³C]-glc as bacterial ¹³C source during heterologous expression of the α -synuclein protein⁵⁴ and the type three secretion needles of *Salmonella typhimurium*²⁴ and *Shigella flexneri*.²⁵ These labeling schemes result in a strong reduction of spectral crowding and enhancement in spectral resolution. In addition, a large number of long-range distance restraints is observable due to the concomitant reduction of dipolar truncation effects. In the present thesis, we developed a robust protocol for high-resolution protein structure determination using 1- and 2-glucose labeled samples. Our protocol is demonstrated and verified by using microcrystalline ubiquitin as model protein.

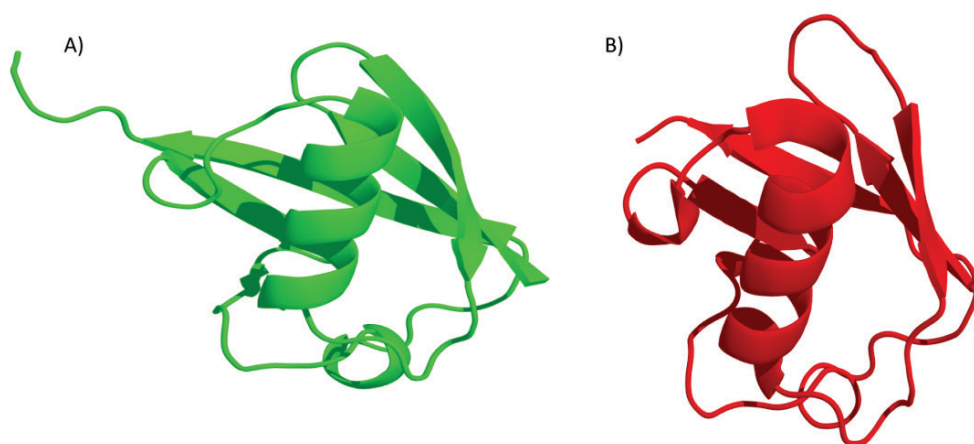


Figure 2 Illustration of the X-ray structures crystallized with two different precipitants A) PEG (green, PDF ID:1UBQ)⁴⁹ and B) MPD (red, PDB ID:3ONS).⁴⁶

2 Theoretical aspects of the applied methods

2.1 General principles of NMR

Nuclear magnetic resonance (NMR) spectroscopy was first observed independently by Bloch, Hansen and Packard and Purcell, Torrey and Pound in 1945. It is based on interactions between nuclear spins and a strong external magnetic field \mathbf{B}_0 (Figure 3). The external homogeneous magnetic field \mathbf{B}_0 , in the range of several Tesla, is generated by a superconducting magnet which has to be cooled down to around the boiling point of helium (4.18 K). The interaction between nuclei and an external magnetic field only takes place if the nuclei possess spin angular momenta. In this work only the nuclei ^1H , ^{13}C and ^{15}N with spin quantum number of $\frac{1}{2}$ were measured and will be further discussed. Due to the Zeeman splitting two eigenstates (α , β) will be generated by a nucleus with spin $\frac{1}{2}$. The transition from the low energy eigenstate α to the higher energy state β will be induced by an additional applied rectangular radio frequency pulse ($\sim 10^{18}$ photons).⁵⁵

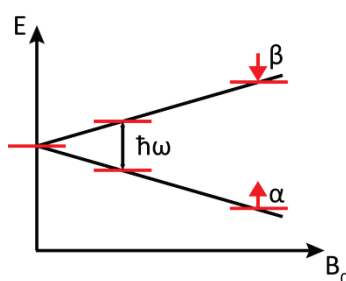


Figure 3 Energy level diagram of spin $\frac{1}{2}$ nuclei in presence of external magnetic field.

Equation 1 (Eq.1) describes the proportional relation between \mathbf{B}_0 to the resultant spin precession (Larmor frequency ω (in angular frequency units)) of the nucleus and the energy differences between the two eigenstates ΔE . The gyromagnetic ratio γ expresses the sensitivity response of a nucleus to the applied external magnetic field \mathbf{B}_0 .

$$\omega = \frac{\Delta E}{\hbar} = -\gamma \mathbf{B}_0 \quad (\text{Eq.1})$$

The population differences of the two eigenstates under equilibrium conditions can be explained by a Boltzmann distribution (Eq.2). For a 9.4 Tesla (400MHz Spectrometer) the population differences between the two eigenstates at room temperature (298 K) are in the order of 1 out of 10^4 spins for ^1H nuclei. According to the fact that the effective absorbed signal in the solenoid coil is depended on the population difference of the two eigenstates, NMR spectroscopy is counted as a low sensitive spectroscopy. An increase in population differences can be achieved by applying of a higher \mathbf{B}_0 field (Figure 3) or by measuring at lower temperatures.

$$\frac{N_\beta}{N_\alpha} = e^{-\Delta E/kT} \sim 1 - \frac{\Delta E}{kT} = 1 - \frac{\hbar \gamma \mathbf{B}_0}{kT} \quad (\text{Eq.2})$$

with k as the Boltzmann factor, T temperature in Kelvin, $N_{\alpha,\beta}$ population number of the eigenstates, \hbar Planck constant (divided by 2π).

2.1.1 Quantum mechanical description of a statistical ensemble of spin $\frac{1}{2}$ nuclei

2.1.1.1 Density operator

For describing the quantum mechanical behavior of an ensemble in thermal equilibrium, the implementation of density operator $\hat{\rho}_{(t)}$, is the method of choice (Eq.3). Each sub spin system(k) in an ensemble is described by a state function which can be expanded in terms of a complete orthonormal basis set ($|\psi^k\rangle = \sum_n c_n^k |n\rangle$) in Hilbert space. The population density of each eigenstate(k) in the ensemble can be obtained by the product of the diagonal elements of the corresponding density operator $\hat{\rho}_{(t)}$. For a nuclear spin ensemble, diagonal terms of $\hat{\rho}_{(t)}$ represent different spin states and off-diagonal terms represents different coherences between various spin states. The extension of $\hat{\rho}_{(t)}$ by the probability term p^k gives averaged population values of spin systems in an ensemble.

$$\hat{\rho}_{(t)} = \sum_k p^k |\psi_{(t)}^k\rangle \langle \psi_{(t)}^k| \quad \text{(Eq.3)}^{55}$$

2.1.1.2 Time evolution of the density operator

In order to have a macroscopic observation, the expected value of an observable \hat{A} has to be calculated by $\langle \hat{A} \rangle = \text{tr}[\hat{A} \hat{\rho}_{(t)}]$.⁵⁵ This expected value for the observable \hat{A} , is calculated based on the Schrödinger representation, where the time dependence of the system is associated with the density operator, while the Hamiltonian operator is time independent. For calculating the dynamics of quantum mechanical systems the Liouville von Neumann differential equation (Eq.4) can be employed as an appropriate method. The Liouville von Neumann equation can be solved by the right choice of rotating frame, which makes the Hamiltonian \hat{H}_t time-independent. The resultant propagator can be written in the form of $\exp(\frac{-i}{\hbar} \hat{H}t)$.⁵⁵

$$\frac{d}{dt} \hat{\rho}_{(t)} = -i[\hat{H}_{(t)}, \hat{\rho}_{(t)}] \quad \text{(Eq.4)}^{55}$$

The solution of the Liouville- von Neumann equation by consideration of a time-independent Hamiltonian \hat{H} can be derived in (Eq.5).

$$\hat{\rho}_{(t)} = \exp(\frac{-i}{\hbar} \hat{H}t) \hat{\rho}_{(0)} \exp(\frac{i}{\hbar} \hat{H}t) \quad \text{(Eq.5)}^{55}$$

However, for a separated description of nuclear or electronic spin observables of a system the employment of a reduced spin density operator $\hat{\sigma}_{(t)}$ is required. Remaining degrees of freedom contributing to spin relaxation effects are called 'lattice'. This simplification can be realized by separation of the total wave

function into a lattice and a spin part. The resultant reduced density operator, which is acting only on the spin observables, can be implemented into the Liouville- von Neumann equation. The general outcome, (Eq. 6) often called ‘quantum mechanical master equation’ (ME):

$$\frac{d}{dt} \hat{\sigma}_{(t)} = -i[\hat{H}^s, \hat{\sigma}_{(t)}] - \Gamma(\hat{\sigma}_{(t)} - \hat{\sigma}_{(0)}) \quad (\text{Eq. 6})^{55}$$

The spin Hamiltonian \hat{H}^s which operates only on the spin depending observables is obtained by averaging of the full Hamiltonian over the lattice coordinates (Eq. 6). The superoperator Γ contributes to relaxation behavior between the spin system and the lattice which forces the reduced density operator towards its equilibrium value $\hat{\sigma}_{(0)}$. Due to the situation that spin $\frac{1}{2}$ nuclei behave as a perfectly smooth magnetic ball all electronic spin observables vanish to zero and will not be considered further in this thesis. These simplifications of the resultant spin Hamiltonian \hat{H}^s contain only magnetic interactions of nuclei with a spin quantum number of a half.

2.1.2 Magnetic nuclear spin Hamiltonian \hat{H}^s

2.1.2.1 External spin interaction

The magnetic spin Hamiltonian \hat{H}^s can be separated into an external spin Hamiltonian \hat{H}_{ext} term and in an internal spin Hamiltonian \hat{H}_{int} term (Eq.7). The external spin Hamiltonian \hat{H}_{ext} (Eq.8) contains interactions of nuclei with the external static magnetic field \mathbf{B}_0 (\hat{H}_{static} ; Zeeman interaction) and with an applied linearly polarized radio-frequency pulse (\hat{H}_{rf}).

$$\hat{H}^s = \hat{H}_{\text{ext}} + \hat{H}_{\text{int}} \quad (\text{Eq.7})^{56}$$

$$\hat{H}_{\text{ext}} = \hat{H}_{\text{static}} + \hat{H}_{\text{rf}} \quad (\text{Eq.8})^{56}$$

The Zeeman interaction which provokes the splitting of degenerate energy eigenstates into the eigenstates α and β by applying a static magnetic field \mathbf{B}_0 , is expressed by the Zeeman Hamiltonian \hat{H}_{static} . Conventionally the static field is aligned along the z-axis of the laboratory frame which brings the Zeeman interaction in a form as shown in (Eq.9).

$$\hat{H}_{\text{static}} = -\hat{\boldsymbol{\mu}} \mathbf{B}_0 = -\gamma \mathbf{B}_0 \hat{I}_z = \hbar \omega \hat{I}_z \quad (\text{Eq.9})^{57}$$

with the nuclear momentum operator $\hat{\boldsymbol{\mu}} = \gamma \hbar \hat{\mathbf{I}}$, $\hat{\mathbf{I}}$ as the spin angular momentum operator and \hat{I}_z as the spin angular momentum operator in z-axis.

Before going into the details of the other interaction, consideration of the so-called secular- or high-field approximation is helpful. This approximation is based on perturbation theory, which reduces weak

interactions to those components that commute with the strong Zeeman interaction.⁵⁶ In NMR the general expression of weak magnetic spin Hamiltonians can be written as $\hat{H} = \hbar \mathbf{I} \mathbf{A} \mathbf{X}$ ⁵⁶, where \mathbf{I} is a nuclear spin operator, \mathbf{A} is a second-rank Cartesian interaction tensor and \mathbf{X} may be a magnetic field or another nuclear spin or angular momentum operator. This representation is used in defining the interaction as described below. For instance, the interaction of a nucleus with an oscillating radio frequency (rf) field \mathbf{B}_1 of an applied pulse becomes time-independent by using the correct rotating frame transformation that can be expressed as:

$$\hat{H}_{\text{rf}} = -\gamma \hat{\mathbf{I}} \mathbf{B}_1 = \omega_1 (\hat{I}_x \cos(\phi) + \hat{I}_y \sin(\phi)) \quad (\text{Eq.10})$$

with $\omega_1 = -\gamma |\mathbf{B}_1|$ and $\cos(\phi)$ $\sin(\phi)$ as the trigonometric function cosine(ϕ) and sine(ϕ), ϕ as the phase of the rf pulse, \hat{I}_x as the spin angular momentum operator in x-axis.

2.1.2.2 Internal spin interaction \hat{H}_{int}

For spin $\frac{1}{2}$ nuclei, the internal spin Hamiltonian \hat{H}_{int} can be separated into three main terms: the chemical shift (\hat{H}_{CS}), the direct dipole-dipole coupling (\hat{H}_{DD}) and the scalar coupling (J coupling) (\hat{H}_{J}) as shown in (Eq.11):

$$\hat{H}_{\text{int}} = \hat{H}_{\text{CS}} + \hat{H}_{\text{DD}} + \hat{H}_{\text{J}} \quad (\text{Eq.11})$$

According to the fundamental role of internal spin interactions in the field of NMR each interaction will be discussed in detail.

2.1.2.2.1 Chemical shift \hat{H}_{CS}

Each nucleus in an ensemble is surrounded by electrons that shield external magnetic field resulting in a local induced magnetic field $\mathbf{B}_{\text{induced}}$ around the nucleus. The shape and orientation of the electron cloud (chemical shielding) can be described by the chemical shift tensor σ as demonstrated in (Eq.12):

$$\mathbf{B}_{\text{induced}} = \sigma \mathbf{B}_0 \quad (\text{Eq.12})$$

Considering secular approximation and rotating frame transformations to the principal axes system (PAS), the chemical shift Hamiltonian, \hat{H}_{CS} , can be expressed in solids by:

$$\hat{H}_{\text{CS}}(\theta) = \gamma \hat{I}_z \sigma_{zz}(\theta) \mathbf{B}_0 \quad (\text{Eq.13})^{57}$$

with(θ) an angle to describe the molecular orientation with respect to the magnetic field.

The orientation-dependent part of the chemical shift is known as the chemical shift anisotropy(CSA).⁵⁷ Due to the weak induced magnetic field $\mathbf{B}_{\text{induced}}$ which is in the order of 10^{-6} to the external field \mathbf{B}_0 , chemical

shift values are presented in parts per million (ppm). Depending on the electronegativity differences of bonded nuclei, the nucleus of interest can be deshielded or even higher shielded reflected by observed chemical shift. As an example, the strong influence of the electronegativity of different types of nuclei can be demonstrated by the chemical shift differences between an aliphatic carbon (~30ppm) and a carboxylic carbon (~175ppm).

2.1.2.2.2 Dipolar coupling \hat{H}_{DD}

The direct interaction between two nuclei through space can be explained by the magnetic dipolar coupling strength. Each nucleus can be treated as a magnetic pole that generates a surrounding magnetic field. The dipole-dipole interaction will be treated as mutual, which indicates that each nuclear spin experiences the field generated by other spins. The dipole-dipole interaction Hamiltonian \hat{H}_{DD} can be described as shown in (Eq.14):

$$\hat{H}_{DD} = \hat{\mathbf{I}}_j \mathbf{D}_{jk} \hat{\mathbf{I}}_k \quad (\text{Eq.14})$$

By transforming into polar coordinates and consideration of the secular approximation (Eq.14) can be rewritten for homonuclear dipole-dipole interactions as:

$$\hat{H}_{jk}^{DD} = d_{jk} (3 \hat{I}_{jz} \hat{I}_{kz} - \hat{\mathbf{I}}_j \hat{\mathbf{I}}_k) \quad (\text{Eq.15})^{57}$$

with d_{jk} as the dipole-dipole coupling strength:

$$d_{jk} = b_{jk} \frac{1}{2} (3 \cos^2 \theta_{jk} - 1) \quad (\text{Eq.16})^{57}$$

The angle dependency term $(3 \cos^2 \theta_{jk} - 1)$ represents orientation dependency and its effect in ssNMR spectroscopy will be discussed in section 2.3. The term b_{jk} in (Eq.16) is named dipole-dipole coupling constant which describes the magnitude of the through space interaction between the two nuclei, j and k . The strength of the interaction is directly proportional to the multiplication of the gyromagnetic ratio of the two coupling spins and inversely cubic proportional to the internuclear distance.

$$b_{jk} = \frac{\mu_0}{4\pi} \frac{\gamma_j \gamma_k \hbar}{r_{jk}^3} \quad (\text{Eq.17})^{57}$$

In the case of heteronuclear nuclei the dipole-dipole interaction Hamiltonian \hat{H}_{DD} can be given by:

$$\hat{H}_{jk}^{DD} = d_{jk} 2 \hat{I}_{jz} \hat{I}_{kz} \quad (\text{Eq.18})^{57}$$

2.1.2.2.3 Scalar or J coupling \hat{H}_J

According to the direct dipole-dipole interaction which was discussed in the previous section another coupling interaction between two spins can be conducted in NMR. This scalar interaction is mediated by valence electrons, which contribute in chemical binding (particularly σ bonds), and is expressed by the J coupling Hamiltonian \hat{H}_J as shown in (Eq.19):

$$\hat{H}_{jk}^J = 2\pi \hat{I}_j J_{jk} \hat{I}_k \quad \text{(Eq.19)}^{57}$$

Following the same procedure as for the dipole-dipole Hamiltonian \hat{H}_{jk}^{DD} to describe the interaction between two spins, the expression of the homonuclear J coupling Hamiltonian \hat{H}_{jk}^J can be given as:

$$\hat{H}_{jk}^J = 2\pi J_{jk} \hat{I}_j \hat{I}_k \quad \text{(Eq.20)}^{57}$$

whereas heteronuclear J coupling (with secular approximation) is given as follows:

$$\hat{H}_{jk}^J = 2\pi J_{jk} \hat{I}_{jz} \hat{I}_{kz} \quad \text{(Eq.21)}^{57}$$

2.1.2.3 Comparison of spin Hamiltonians

The magnitudes (shown by the size of the circles) of the different spin interactions for spin $\frac{1}{2}$ nuclei solid state sample are represented in Figure 4. By comparison of the different spin Hamiltonians, the Zeeman Effect can be distinguished as the most significant spin interaction of a nucleus with a spin $\frac{1}{2}$ in a static magnetic field \mathbf{B}_0 . Because of the strong difference of magnitude between the Zeeman interaction to the weaker \hat{H}_{rf} and \hat{H}_{int} interactions strengths, it is reasonable to use the high field approximation.

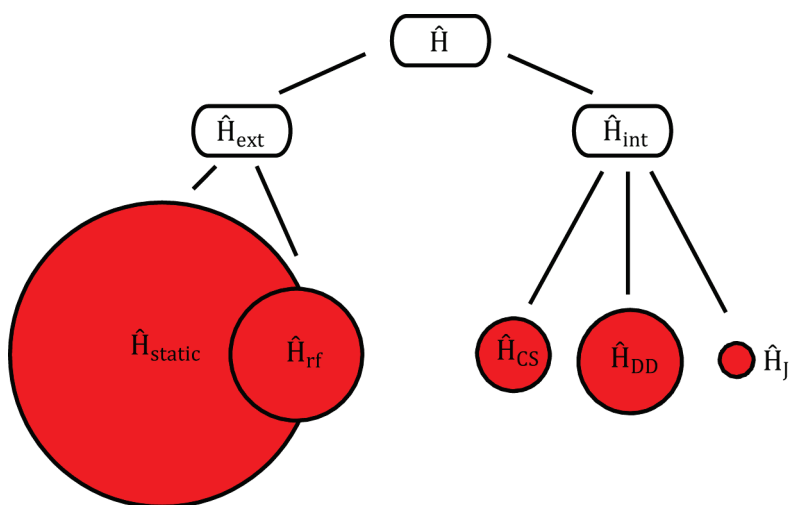


Figure 4 Relative comparisons between the different spin Hamiltonian strengths, for the sake of clarity the J coupling \hat{H}_J Hamiltonian is written outside the circle. Ref.⁵⁷

2.2 NMR Signal

According to linear response theory and the brilliant invention of Fourier-transformation NMR spectroscopy by Ernst (Nobel Prize 1991) and Anderson, the time evolution of a transverse nuclear magnetization (coherence) can be detected. The simplest transformation of a full spin ensemble from equilibrium to coherence can be generated by a simple 90 degree pulse (shown in Figure 5).

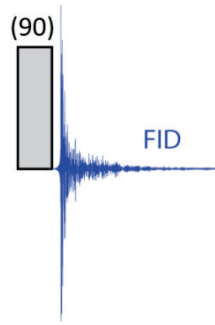


Figure 5 Illustration of a measured FID obtained by a 90 pulse on an arbitrary spin ensemble.

The evolution in transverse plane and relaxation of a spin system from coherence state to equilibrium induces an oscillating electric voltage in the receiver coil, which is called the free-induction decay (FID). Fourier transformation converts the FID time-domain signal $s_j(t)$ into a function of frequency $S_{j(\Omega)}$. A faithful representation of a detected signal in the frequency domain is determined by the Nyquist theorem which indicates that the sampling rate must be at least equal to twice the highest frequency contained in the signal.⁵⁵ The FID signal with a single frequency evolution is given as follows:

$$s_j(t) = a_j \exp [(i\Omega_j - \lambda_j) t] \quad (\text{Eq.22})^{57}$$

Here, Ω_j as the resonance frequency of the spin, λ_j as a spin-lattice relaxation rate and a_j as the amplitude of the signal.

For the Fourier transformation of a single spin signal one obtains (Eq.23):

$$S_{j(\Omega)} \propto \int_0^\infty s_j(t) e^{-i\Omega t} dt \quad (\text{Eq.23})^{57}$$

with Ω as the sum over all frequencies from a complete spin ensemble

The integration of (Eq.23) gives a so called complex Lorentzian term \mathcal{L} (Eq.24) which is a fundamental term of the interpretation of a NMR detected signals. The Lorentzian term \mathcal{L} can be separated into a real (absorption signal) and imaginary part (dispersion signal). Both peak shapes of a single spin are sketched in Figure 6.

$$\mathcal{L}_{(\Omega, \Omega_j, \lambda)} \sim \frac{1}{(\lambda + i(\Omega - \Omega_j))} \quad (\text{Eq.24})^{57}$$

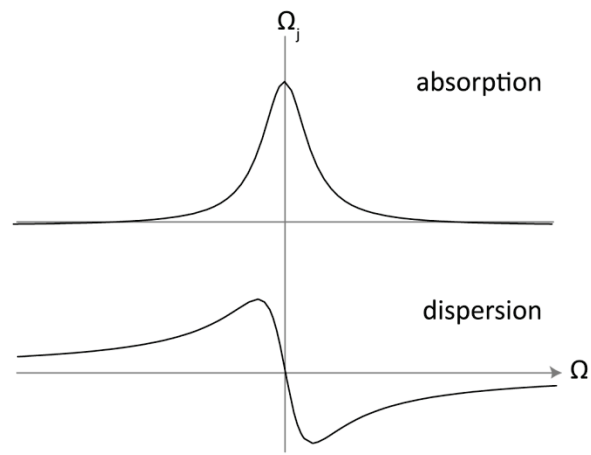


Figure 6 Representation of the absorption and dispersion part of the Lorentzian term. Ref.⁵⁷

Similar formalism can be extended to 2D or higher multidimensional NMR experiments (Figure 7). The Fourier-transformation expression for a two dimensional NMR spectrum is defined as:

$$S_{j(\Omega_1, \Omega_2)} \propto \iint_0^\infty s_{j(t_1, t_2)} e^{-i(\Omega_1 t_1 + \Omega_2 t_2)} \quad (\text{Eq.25})^{57}$$

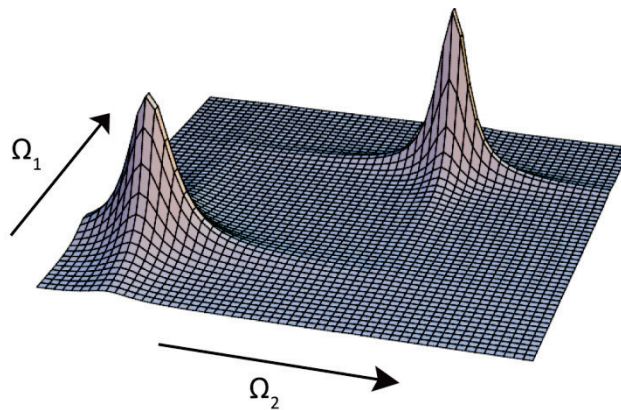


Figure 7 The real (absorption) part of two spins with different Larmor frequencies are shown in a surface plot. Ref.⁵⁷

2.3 Solid-state NMR

2.3.1 Definitions of various frames in ssNMR

The introduction of passive and active rotation transformation is required for a detailed description of ssNMR spectroscopy. The passive rotation is based on the linear transformation between different frames by keeping the operator unchanged. The prevailing used frames in ssNMR are sample-, laboratory-, molecular-, and principal axis frames as depicted in Figure 8. The transformation from one frame into another will be generated by rotation matrices, R , applied with the three Euler angles α , β , and γ . A general expression of the frame rotation is given by:

$$\hat{A}_{(XYZ)} = R \hat{A}_{(xyz)} R^{-1} \quad (\text{Eq.26})^{58}$$

where x , y , and z are the old coordinates and the X , Y , and Z are the new coordinates after the rotation.

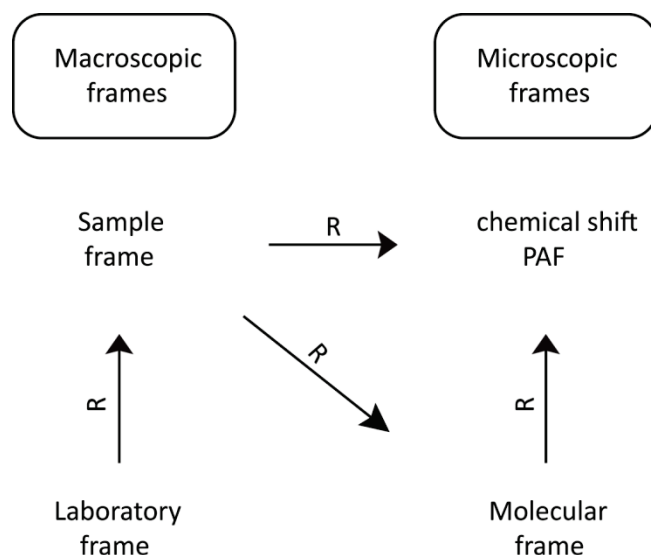


Figure 8 Sketched representation of coordinate frames transformation in ssNMR. Ref.⁵⁶

In NMR it is convenient to express the rotation of spin Hamiltonians (discussed in section 2.2) in form of exponential rotation operator \hat{R} . This active rotation transformation keeps the frame constant, while it transforms the operator in a specific direction. In the rotating frame the effect of a perfect 90° pulse to a spin system in equilibrium can be expressed in a simplified way by:

$$\hat{R}_x \hat{I}_z \hat{R}_x^{-1} = -\hat{I}_y \quad (\text{Eq.27})$$

with $\hat{R}_x = \exp(-i\frac{\pi}{2} \hat{I}_x)$; $\hat{I}_x, \hat{I}_y, \hat{I}_z =$ spin angular momentum operators.

In the established NMR vector model the rotation of \hat{I}_z to $-\hat{I}_y$ can be described by a transformation of the net equilibrium magnetization vector \mathbf{M}_z to \mathbf{M}_y .

2.3.2 Magic angle spinning (MAS)

In solution-state NMR, various anisotropic (orientation-dependent) interactions such as chemical shift anisotropy (CSA) and through-space dipolar couplings are averaged out by tumbling motion of the molecules. In the solid-state, this motion is absent which in turn leads to broader lines. Magic angle spinning (MAS) is a routinely used technique for obtaining narrow lines in ssNMR spectra. The basic principle behind the MAS approach is to mimic the tumbling motion by spinning the solid-state sample at an angle ($\theta_r = 54.74^\circ$) (magic angle) inclined to the static magnetic field B_0 which averages the $(3\cos^2\theta - 1)$ dependence of the anisotropic interactions to zero. For example, the relation between the positions of the sample due to chemical shield tensor in the PAS frame is depicted in Figure 9 .

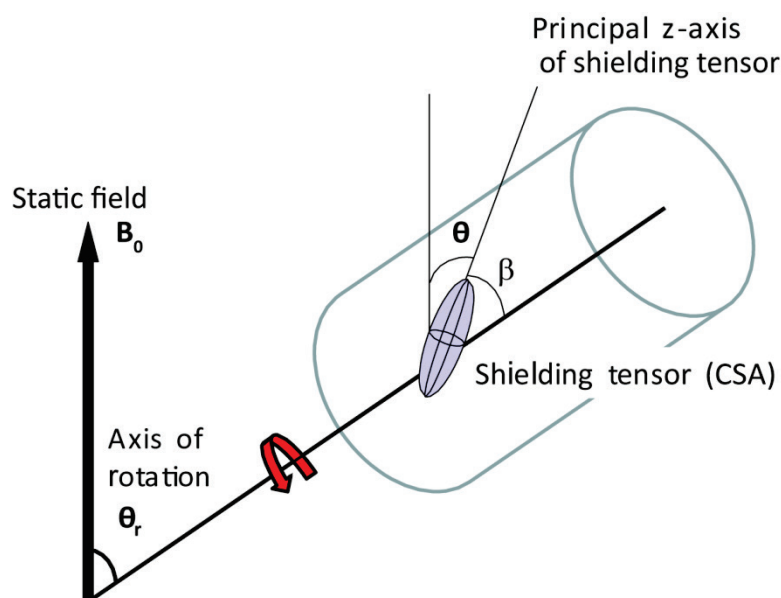


Figure 9 Schematic representation of magic-angle spinning and the orientation of principal axis frame of CSA tensor in the rotor. Ref.⁵⁸

The resulting equation for nuclear spin interaction in the PAS frame is given by

$$\langle 3\cos^2\theta - 1 \rangle = \frac{1}{2} (3\cos^2\theta_r - 1) (3\cos^2\beta - 1) \quad \text{(Eq.28)}^{58}$$

2.3.3 Cross polarization (CP)

The basic principle of cross polarization (CP) is to transfer polarization from a high- γ (e.g. ^1H proton) to a low- γ nucleus (e.g. ^{13}C carbon and /or ^{15}N nitrogen). An optimized CP transfer under MAS is provided by applying the Hartmann-Hahn condition⁵⁹ in the double rotating frame (Eq.29):

$$|\omega_{1H} \pm \omega_{13C/15N}| = n \omega_{MAS} \quad \text{(Eq.29)}$$

The concept of the CP process can be interpreted as an oscillation of magnetization between the heteronuclear spins. For large spin ensembles which are normally the case for solid-state samples the

polarization transfer can be defined in thermodynamic terms. A schematic representation of the pulse sequence to obtain a 1D ^{13}C spectrum based on a CP transfer is shown in Figure 10.

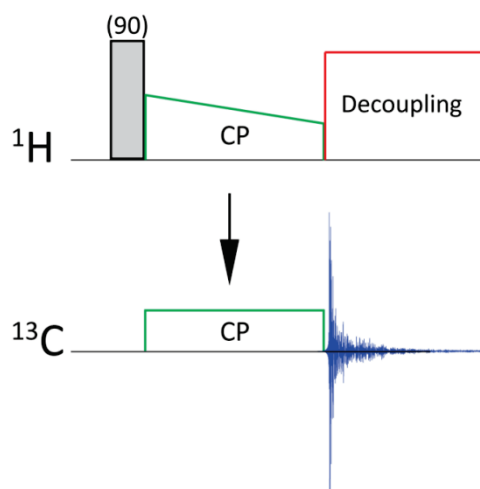


Figure 10 Illustration of the cross polarization experiment, the direction of the magnetization transfer is shown by an arrow.

For conducting 1D ssNMR spectra an efficient heteronuclear decoupling by a strong rf pulse ($\sim 83\text{kHz}$) on the proton channel has to be applied due to strong proton-proton coupling effecting heteronuclear nuclear dipolar coupling (Figure 10). Similarly, the decoupling of protons during signal acquisition periods is a fundamental requirement for the recording of high-resolution ^{13}C or ^{15}N ssNMR spectra.⁶⁰⁻⁶²

2.3.4 2D Homonuclear correlation spectroscopy

For the challenging and time-consuming task of sequential resonance assignment of proteins, many 2D ssNMR experiments have been developed in the last decades, in order to extract different homonuclear and heteronuclear correlation spectra. In the context of this thesis the common homonuclear pulse sequences, proton-driven spin diffusion (PDS) were recorded for ^{13}C - ^{13}C or ^{15}N - ^{15}N correlation ssNMR spectra. The corresponding pulse sequence is depicted in Figure 11. The variation of mixing time $t_{\text{Mixing time}}$ between 50ms up to 900ms in the PDS-block enables the collection of intra-residual and inter-residual correlations. Depending on the mixing time, various correlations can be recorded. As a rule of thumb, for uniformly labeled proteins, a mixing time of 20-50ms for intra-residue correlations, a mixing time of 100-250ms for sequential correlations and above 300ms for long-range correlations is required.

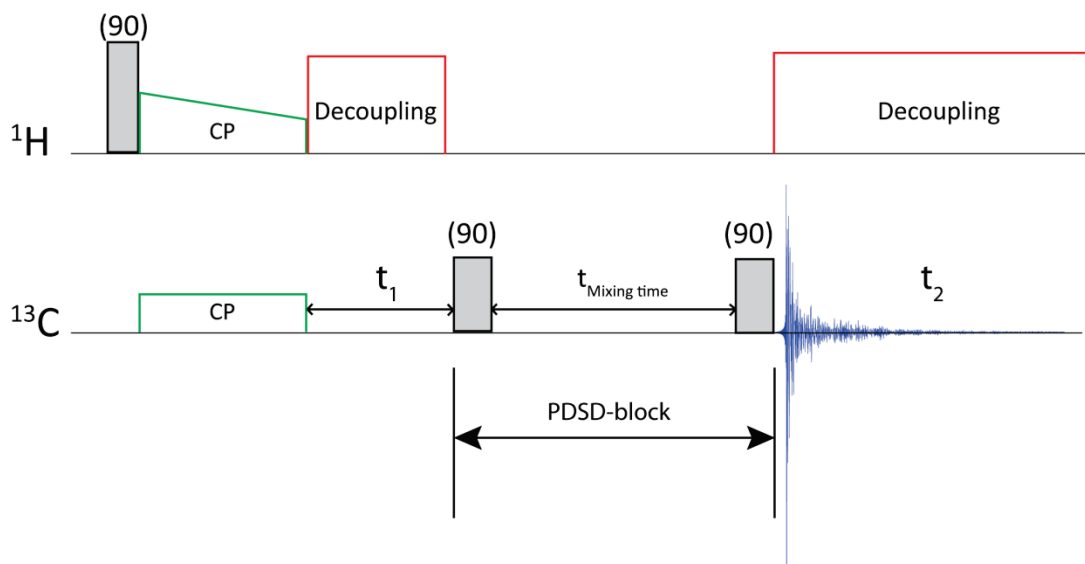


Figure 11 Representation of the 2D homonuclear PDSD pulse sequences, the PDSD-block is highlighted by the double-oriented arrow.

2.3.5 2D Heteronuclear correlation spectroscopy

In addition to homonuclear correlation spectra heteronuclear correlation (^{15}N - ^{13}C) spectra can be obtained by applying a second CP transfer (SPECIFIC-CP)⁸ in the pulse sequence (Figure 74, Appendix I). The large chemical shift differences between $^{13}\text{C}\alpha$ (~70-40ppm) and ^{13}CO (165-185ppm) and the irradiation of weak rf fields facilitate specific one-bond transfers from $^{15}\text{N}(i)$ - $^{13}\text{C}\alpha(i)$ (NCA-spectrum) or $^{15}\text{N}(i)$ - $^{13}\text{CO}(i-1)$ (NCO-spectrum) as shown in Figure 12. The implementation of further homonuclear correlation elements (PDSD or DARR (dipolar assisted rotational resonance)) in the pulse sequences permits additional intra-residual resonance information (NCACX-, NCOCX-spectrum). The pulse sequences are depicted in Figure 75 (Appendix I).

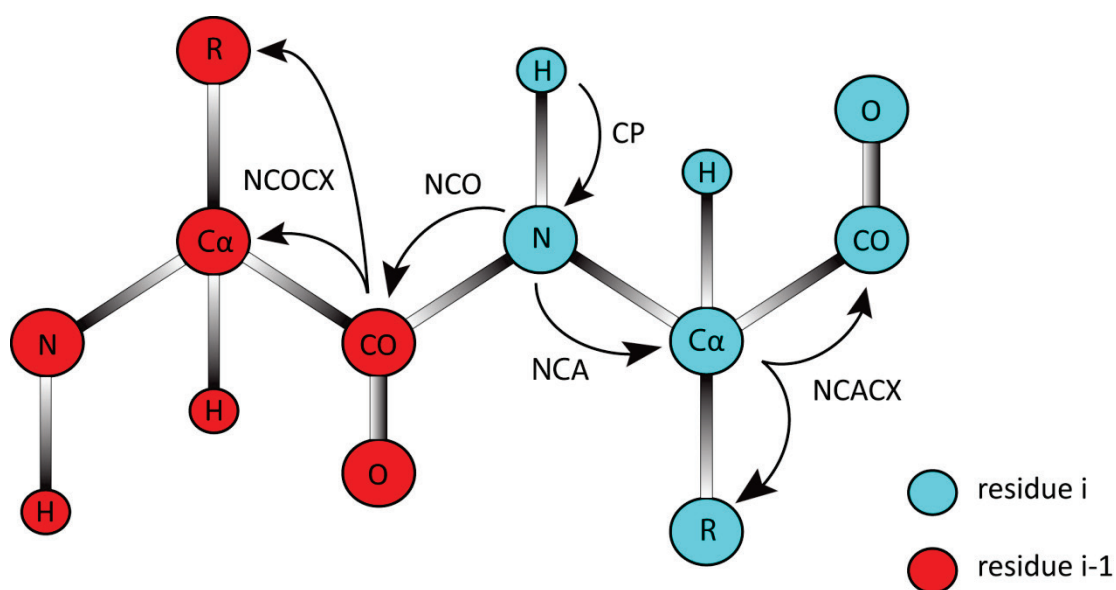


Figure 12 Demonstration of various possible heteronuclear magnetization transfers using different pulse sequences.

2.3.6 Band-selective homonuclear CO-C α CP transfer (BSH-CP)

The previous mentioned PDS and DARR transfers are not selective transfers that turn them to less sensitive experiments. Recently efficient band-selective magnetization transfer (BSH-CP) has been established in our group on highly deuterated and protonated samples.^{12,63} The essential component in this pulse sequence is the homonuclear magnetization transfer from CO to C α , which is the crucial step of the selective detection of sequential heteronuclear correlations from ^{15}N (i) to ^{13}C (i-1). Owing to the situation that the chemical shift variation in the CO band is smaller than in the C α band, it is optimal to apply the rf irradiation in the middle of the C α region in order to avoid a too narrow banded transfer. The theoretical rf amplitude ω_{rf} is given by:⁶⁴

$$\sqrt{\Omega^2 + \omega_{\text{rf}}^2} + \omega_{\text{rf}} = 2 \omega_{\text{MAS}} \quad (\text{Eq.30})^{64}$$

where Ω is the CO chemical shift offset in angular frequency and ω_{MAS} the spinning frequency. For the CO to C α magnetization transfer, the CO magnetization is flipped to the effective field by a hard trim pulse before BSH-CP. The flip angle θ is given by:

$$\theta = 90^\circ - \arctan(\omega_{\text{MAS}} / \Omega) \quad (\text{Eq.31})^{64}$$

After the BSH-CP transfer, a second hard trim pulse is applied on-resonance with C α to bring the CO magnetization to the transfer plane without affecting the C α magnetization. The pulse sequence for the BSH-CP magnetization transfer is depicted in Figure 13.

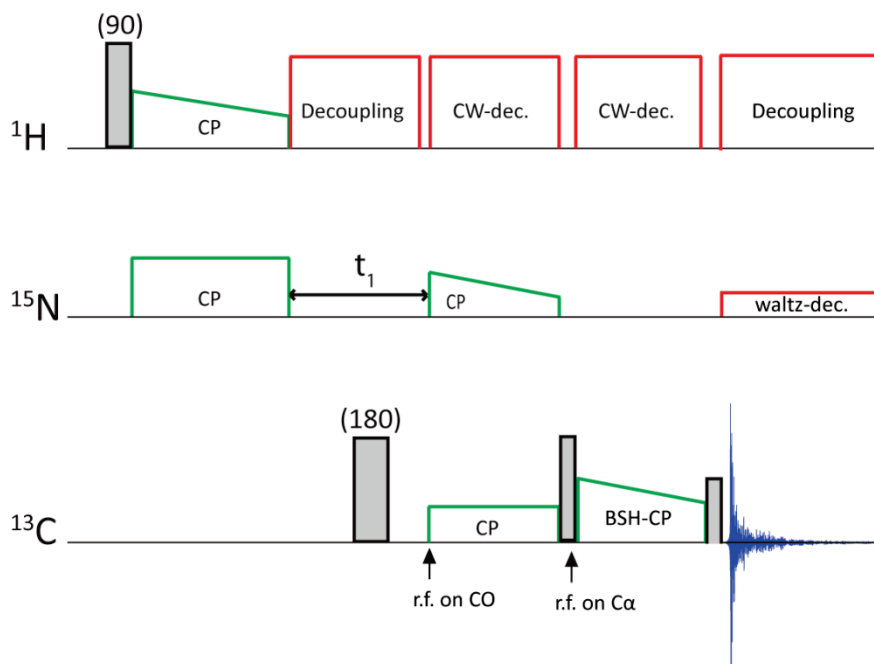


Figure 13 Illustration showing the implementation of the BSH-CP transfer in the NCOCA pulse sequence.

2.4 Isotope labeling

2.4.1 General overview

Since the first investigations on isotopic labeled proteins in the late 1960's^{65,66}, a lot of efforts have been put in to create different isotopic enriched labeling schemes for protein studies.^{67,68} In accordance with the inherent low sensitivity of NMR, a capable production of few milligrams of properly folded proteins is required to perform experiments with acceptable signal to noise ratio. The main approaches for obtaining reasonable amounts of labeled proteins with suitable yields can be ordered into three categories: the heterologous overexpression, the total synthesis and the cell-free-expression. In this thesis, the heterologous overexpression by bacteria will be discussed.

Along with the development of multidimensional NMR experiments in the 1980 and 1990's, ¹⁵N and ¹³C uniformly enriched proteins were required. The incorporation of ¹⁵N and ¹³C in proteins by cells growing in minimum media can be achieved by ¹⁵N labeled ammonium salts (¹⁵NH₄Cl) and ¹³C labeled glucose as the sole nitrogen and carbon sources, respectively. The uniform labeling strategy is commonly used for assignment purpose, which is necessary for recording of multidimensional homonuclear- (¹³C-¹³C, ¹⁵N-¹⁵N) and heteronuclear spectra, as discussed previously in sections 2.3.4 to 2.3.6. However, significant drawbacks in application of uniformly labeled proteins are found in the large number of ¹³C labeled carbons, which contributes to high orders of overlapping and corresponding low spectral resolution and high ambiguity of ¹³C-¹³C correlations. Major contributions to the line-broadening of ¹³C peaks in uniformly labeled samples are the one-bond ¹³C-¹³C dipolar and scalar couplings. Furthermore uniformly labeled samples show strong dipolar truncation, which slows down the magnetization transfer and suppresses the detection of long-range distance correlations.

2.4.2 Sparsely labeling strategies

Recently new approaches have been employed on alternative sparsely ¹³C labeling of proteins. The most common sparsely ¹³C labeling schemes comprise the use of [1,3-¹³C]-glyc, [2-¹³C]-glyc^{19,22,69}, [1-¹³C]-glc, [2-¹³C]-glc^{14,24,25,27}, [1-¹³C]-acetate⁷⁰ and [1,2-¹³C]-pyruvate (pyr) or [1-¹³C]-pyr with ¹³C labeled sodium bicarbonate.⁷¹ All these labeling strategies aim at dilution of ¹³C isotopes within the labeled protein, to enhance spectral resolution and reduce the overlap of cross-peaks in ssNMR spectra. Another significant advantage lies in the detection of long-range distance correlations by means of strong suppression of dipolar truncation effects.

In this thesis, the alternative [1-¹³C]- and [2-¹³C]-glc^{14,24,25,27} labeling scheme is utilized. In contrast to uniformly ¹³C labeled proteins, only one out of six carbons is ¹³C labeled in the labeled protein, which results in a high dilution of ¹³C labeled carbons within the protein. According to the biosynthetic pathway of amino acid synthesis the breaking of fructose-1,6-bisphosphate into dihydroxyacetone phosphate and glyceraldehyde-3-phosphate reduces the maximum labeling level to $\frac{1}{2}$, which leads to a theoretical

probability of $\frac{1}{4}$ for two directly bonded labeled sites. This strong reduction of ^{13}C to ^{13}C connections increases the spectral resolution by removal of one-bond dipolar and scalar ^{13}C - ^{13}C couplings. By consideration of the biosynthesis pathway of the different amino acids, labeling patterns for the two labeling approaches are predictable (Figure 14). The significant differences are related to the $^{13}\text{C}_\alpha$ presence in the $[2\text{-}^{13}\text{C}]$ -glc labeling scheme and the presence of ^{13}C -methyl in the $[1\text{-}^{13}\text{C}]$ -glc labeling pattern.¹⁴

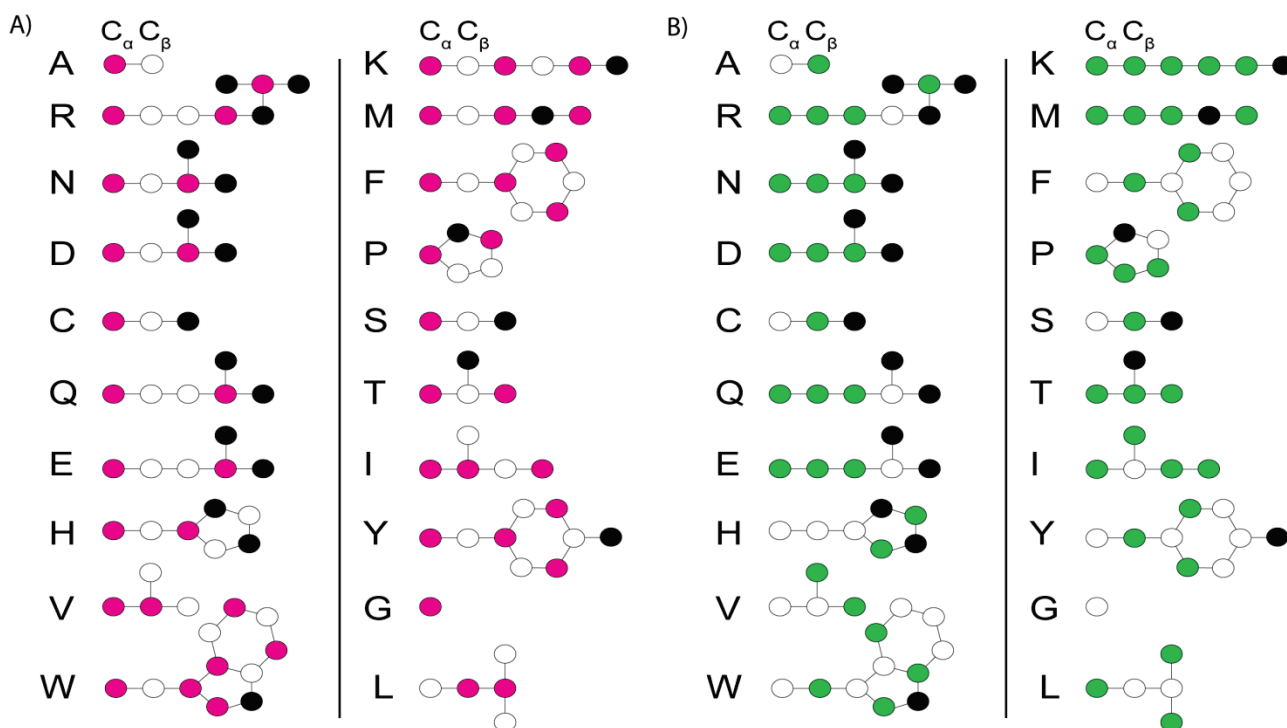


Figure 14 Demonstration of the predicted labeling pattern for $[2\text{-}^{13}\text{C}]$ -glc (magenta) and $[1\text{-}^{13}\text{C}]$ -glc (green), nuclei which are colored in black represent heteronuclei such as N, O, S.

2.5 Structure determination

In the last decades high efforts have been put in optimization of procedures to determine biomolecular structures. These include the use of highly sophisticated calculation algorithms and facilities of clustered computers to supercomputers. In the 1980's the introduction of molecular mechanics methodologies was a major and successful step in describing the potential energy surface (PES) of proteins.^{72,73} The general concept of force field methods is based on the fact that electrons are not considered as individual particles, in order to bypass the calculation of the electronic Schrödinger equation but rather to perform classical mechanics based calculations. This approximation leads to the description of the electronic energy in terms of experimentally founded functions based on nuclear coordinates. The different types of intramolecular interactions between the nuclei can be expressed by parametric functions (Figure 15). The individual parameters for the different functions can be taken from physicochemical analysis methods such as X-ray-, NMR-, Raman- or Infrared spectroscopy. For the calculation of protein structures based on NMR-data,

different common software packages are available (xplor-NIH¹⁵, CNS⁷⁴ and DYANA/CYANA¹⁶). In context of this thesis the software program xplor-NIH was used. The program provides the possibility to combine force field methods with experimentally determined structural information. The total energy E_{total} of the PES can be expressed as:

$$E_{\text{total}} = E_{\text{force field}} + E_{\text{experimental}}$$

(Eq.32)

With the consideration of the different intramolecular interactions, the utilized force field energy (in the xplor-NIH program named as topology energy) is given by:

$$E_{\text{force field}} = E_{\text{bond}} + E_{\text{angle}} + E_{\text{dihedral}} + E_{\text{vdw}} + E_{\text{el}} \quad \text{(Eq.33)}^{75}$$

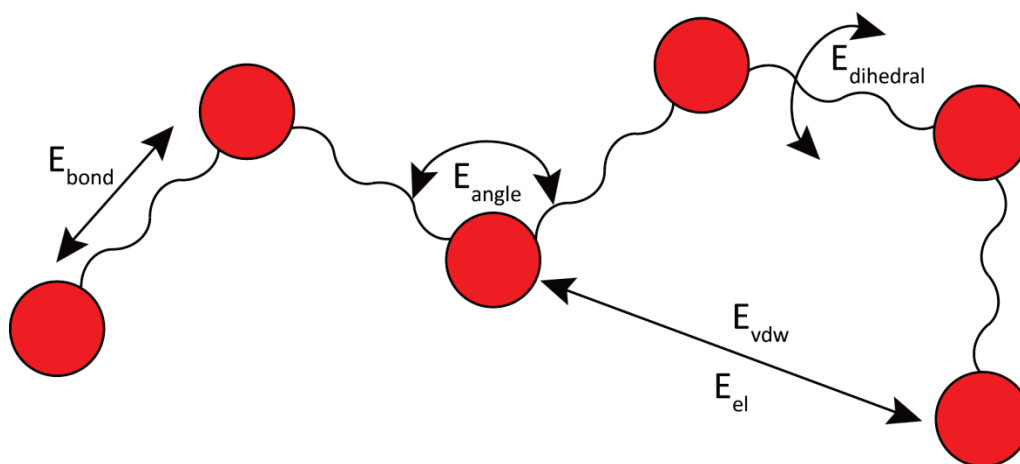


Figure 15 Representation of the fundamental force field energy terms (atoms are shown as red circles). Ref.⁷⁵

The high reduction of degrees of freedom due to the torsion angle dynamics approximation which is implemented in the xplor-NIH program¹⁵ provides reasonable calculation times (1000 structures in 24h).

The energy term $E_{\text{experimental}}$ is based on the distance restraint input file taken from the NMR experiments. The input file contains information about which nucleus has structural interaction to other nuclei in the protein. The distance correlations are ordered into three categories: unambiguous-, network- and ambiguous distance restraints. Each category is separated in two parts: medium-range [$1 < |i-j| \leq 5$] and long-range distance restraints [$|i-j| > 5$]. For the standard *de novo* structure calculation the first implementation of structural information will be given by the unambiguous distance correlations. After iterative applications of network and ambiguous distance correlations the final structure will be calculated. The right choice of the distance for the collected restraints will be evaluated by structure validation software programs.

The explicit mathematical expressions of the different energies in the xplor-NIH¹⁵ calculation are depicted in the section 12.5.(Appendix I)

2.6 Structure validation

The accessibility to online available protein structure validation programs enables the verification of the calculated structures in a straightforward way. In this work the validation program PSVS 1.5⁷⁶ was used for validation of the resulting ubiquitin structures. The program PSVS 1.5 employs several standard structure evolution tools such as RPF⁷⁷, Procheck⁷⁸, MolProbity⁷⁹, Verify3D⁸⁰, Prosa II⁸¹, the PDB validation software and various structure-validation tools developed by Bhattacharya et.al.⁷⁶ Main aspects for validation of a structure includes the Ramachandran plot, violation of distance restraints in the structure, the average of constraints per residue, dihedral angle violations, rmsd values between different structures and structure quality factors.

3 Experimental aspects of the applied Methods

3.1 NMR experiments

All samples were ultra-centrifuged and transferred either into a 4.0-mm or in a 3.2mm ZrO₂ MAS rotor. The MAS rotors were stored at 277K when not in use. The sample-temperature was calibrated from the ¹H chemical shift of water with respect to the DSS signal.⁸² Also, the chemical shift referencing of the spectra was done with respect to DSS signal. For probe safety we use for the 90° rf-pulses field strengths of 83kHz on ¹H, 50kHz on ¹³C and 35kHz on ¹⁵N. For proton decoupling, SPINAL-64 sequence⁶⁰ was used with field strength of 83kHz. All spectra were processed using Bruker Topspin 3.1 and analyzed using Sparky (version 3.100, T. D. Goddard & D.G. Kneller, University of California).⁸³

3.1.1 Ubiquitin (MPD)

The solid-state NMR experiments were recorded with a triple-resonance (¹H, ¹³C, ¹⁵N) MAS probe at either a 20, 18.8 or 14.1 Tesla spectrometer (Bruker Biospin, Germany) which corresponds to operating ¹H Larmor frequencies of 850, 800 and 600 MHz, respectively. The MAS spinning frequency was set to either 11 or 21kHz.

3.1.1.1 1D Experiments of the isotopes ¹³C and ¹⁵N (MPD)

All 1D experiments were conducted at a MAS frequency of 11kHz. The cross-polarization (CP) transfer from ¹H to ¹³C or ¹⁵N was acquired by the use of a ramp-shaped pulse (100%-80% pulse strength) on the proton channel. The contact time of the CP transfer varies between 200-1400 μs. The acquisition time was set to 20ms with a recycling delay of 2.3 seconds.

3.1.1.2 2D Homonuclear experiments (MPD)

2D homonuclear ¹³C-¹³C experiments were performed at a MAS frequency of 11kHz and correlation was achieved by using PDSM mixing. The parameters for the Hartmann Hahn conditions were similar to that of the 1D experiment explained above. The correlation time for ¹³C-¹³C transfer varies in the range from 50ms up to 900ms.

3.1.1.3 2D Heteronuclear experiments (MPD)

The parameters for the heteronuclear magnetization transfer from ¹H to ¹⁵N were taken from 1D experiments. The contact time for the second CP condition which transforms the magnetization from ¹⁵N to ¹³C varied between 3 and 5ms. The pulse strengths for ¹³C and ¹⁵N of the SPECIFIC CP^{8,9} transfer lay in the range of 4kHz up to 30kHz. During the double CP-transfer a high-power decoupling pulse of 83kHz on proton was applied. The ¹³C-¹³C transfer for the NCACX- and the NCOCX-spectra was obtained by the DARR pulse-sequence.

3.1.1.4 BSH 2D Heteronuclear experiments (MPD)

Heteronuclear correlation experiments like NCOCA are based on the efficient homonuclear band-selective (BSH) CO-CA magnetic transfer¹² at a spinning frequency of 21kHz at a 20 Tesla spectrometer. The achievement of an efficient recoupling can be realized by the sum (double-quantum condition) or difference (zero-quantum condition) of effective RF fields on CO and C α resonances equals one or two times the spinning rate. Experimentally, we implemented the double-quantum condition for low power conditions. The prior step of the BSH-CP is the flipping of the CO magnetization by a 63° hard trim pulse along the effective BSH-CP RF field. The BSH-CP was obtained by applying a RF irradiation during 4ms in the middle of the C α chemical shift region with an average RF strength of 13.2kHz and a linear ramp from 100-80%. Finally a second 50kHz hard trim pulse is applied during 4.5 μ s on-resonance with the C α transferring the CO magnetization to the transverse plane without affecting the C α magnetization. The parameters for the ¹H-¹⁵N and ¹H-¹³C CP and the ¹³C and ¹⁵N in the SPECIFIC CP were taken from 1D and 2D experiments as described before.

Various experiments recorded on all the samples are listed in Table 6 (Appendix I).

3.1.2 Ubiquitin (PEG)

The experimental parameters of the 1D, 2D homonuclear experiments and the 2D heteronuclear experiments are described in the section 3.2

Various experiments recorded on all the samples are listed in Table 11 (Appendix I).

The BSH 2D heteronuclear- and the BSH 3D heteronuclear experiments were taken from Ref.^{12,64}

3.2 Prediction of protein backbone torsion angles restraints

To confirm the secondary structure obtained from secondary chemical shift, dihedral angles are predicted for ubiquitin by TALOS+ program.⁸⁴ The prediction is conducted with the chemical shifts of N, C α , C β , CO. The prediction is based on the resemblance between observed chemical shift of proteins studied by NMR and whose structure is known.

3.3 Structure calculation and refinement

The calculation and refinement of the lowest energy structures of ubiquitin were realized with the program xplor-NIH¹⁵. A force field approach named “parallhdg.pro” integrates the torsion angle dynamics approximation. The force field is atomic based and includes all hydrogens and heavy atoms.

The final refinement file (section 12.5.7) includes the following required files:

- ubiq.psf
- gen_temp_ubt.pdb
- TALOS.tbl
- distance-restraints.tbl

The “ubiq.psf” file (topology file) contains all of the molecule-specific topology information needed to calculate the particular force field energy of ubiquitin. The main section of interest consists of atoms, bonds, dihedrals, improper terms (dihedral force terms used to maintain stereochemistry and planarity). The “gen_temp_ubt.pdb” file contains the initial extended structure (Figure 34.A).

The “TALOS.tbl” file integrates the dihedral angle restraints which were generated from the TALOS+ prediction. The “distance-restraints.tbl” file consists of distance restraints which have been collected from the sparsely labeled ^{13}C - ^{13}C PDS spectra. For each calculation step different distance restraints file were used: 59 distance restraints for the unambiguous-, 183 for the unambiguous and network- and 518 for the final structure calculation.

The protocol for the calculation of the structure consists of four stages: (i) 4000 steps of high-temperature torsion angle at 3000K. (ii) 4000 steps of slow-cooling torsion angle stage (iii) 2000 steps of slow-cooling Cartesian MD stage (iv) 200 steps for final minimization stage. The number of accepted structures is set to 1000. For academic purpose only the 10 lowest energy monomers from each structure calculation will be illustrated.

3.4 Validation of the calculated structures

The validation of the different calculated structures was evaluated with the program PSVS 1.5.⁷⁶ Listed files were used for the validation:

- pdb files of the ten lowest energy structures (resultant coordinates of the structures)
- TALOS.tbl (same as used for the calculation)
- distance-restraints.tbl (same as used for the calculation)

4 Materials

4.1 Isotope labeled reagents

Labeled $^{15}\text{NH}_4\text{Cl}$, $[\text{U-}^{13}\text{C}]$ -glc and selectively labeled $[1\text{-}^{13}\text{C}]$ -glc and the $[2\text{-}^{13}\text{C}]$ -glc were obtained from Sigma-Aldrich (Munich, Germany).

4.2 Sample preparation of ubiquitin with MPD

The recombinant expression, purification and the crystallization of ubiquitin were done by Karin Giller and Sebastian Wolff in the biochemical laboratory of our department under the management of Stefan Becker. The recombinant expression and purification of ubiquitin was carried out as described in Ref.⁴⁷ The crystallization condition (40 %v/v MPD and 0.2M CdCl_2) was identified by systematic screening using a commercial crystallization screen (Nextal MPD Suite, Qiagen).

Ubiquitin was recombinantly expressed in *E. coli*. Slight differences of the protocol⁴⁶ are implemented with respect to a change of crystallization conditions: the percentage of the precipitation agent MPD was diminished from 60% to 40% by concomitant addition of 0.2M CdCl_2 . These manipulations resulted in an increase of the pH to 6.5 instead of 4-4.5 for the buffer. The crystallization of ubiquitin could be realized by the hanging drop method. Two uniformly labeled samples (Ubiquitin and Ubiquitin-2) with ^{15}N and ^{13}C were generated by the use of $^{15}\text{NH}_4\text{Cl}$ and $[\text{U-}^{13}\text{C}]$ -glc as the only isotopic labeled source in the minimal growth medium. The expression of selectively labeled ubiquitin samples was obtained by growing bacteria exclusively on $[1\text{-}^{13}\text{C}]$ -glc or $[2\text{-}^{13}\text{C}]$ -glc as the ^{13}C carbon source.

Table 1 List of samples studied with MPD as precipitating agent

Type of protein	Labeling	Precipitation agent	Sample quantity / [mg]
Ubiquitin	$[\text{U-}^{13}\text{C}]$ -glucose $^{15}\text{NH}_4\text{Cl}$	MPD	6
Ubiquitin-2	$[\text{U-}^{13}\text{C}]$ -glucose $^{15}\text{NH}_4\text{Cl}$	MPD	20
Ubiquitin	$[2\text{-}^{13}\text{C}]$ -glucose $^{15}\text{NH}_4\text{Cl}$	MPD	30
Ubiquitin	$[1\text{-}^{13}\text{C}]$ -glucose $^{15}\text{NH}_4\text{Cl}$	MPD	40

4.3 Sample preparation of ubiquitin with PEG

The recombinant expression, purification is identical with the sample preparation described in section 4.2. The significant difference is the use of the second common precipitation agent PEG. The lyophilized ubiquitin sample was dissolved in water to a protein concentration of 28mg/ml. Then the solution was diluted by a 25% w/w PEG buffer (50mM Hepes, 0.2M Cd-[acetate]₂, PEG 2000, pH 7) to a protein concentration of 14mg/ml. The final solution was lyophilized to a protein concentration of 55mg/ml. For the crystallization process of ubiquitin no extra crystallization step was necessary. The labeling strategy employed for the preparation of the ubiquitin samples were following in the same procedure as described in section 4.2.

Table 2 List of samples studied with PEG as precipitating agent

Type of protein	Labeling	Precipitation agent	Sample quantity / [mg]
Ubiquitin	[U- ¹³ C]-glucose ¹⁵ NH ₄ Cl	PEG	24
Ubiquitin	[2- ¹³ C]-glucose ¹⁵ NH ₄ Cl	PEG	16.6
Ubiquitin	[1- ¹³ C]-glucose ¹⁵ NH ₄ Cl	PEG	25.3

5 Results

5.1 Results of Ubiquitin MPD

5.1.1 Verification of the sample quality

The sample quality is verified by recording simple 1D CP MAS spectra of ^{13}C and ^{15}N uniformly labeled ubiquitin. From the resolution and sensitivity of the 1D spectra (Figure 16) one can already recognize that the quality of the sample is very high in perspective of ssNMR spectroscopy sample preparation.

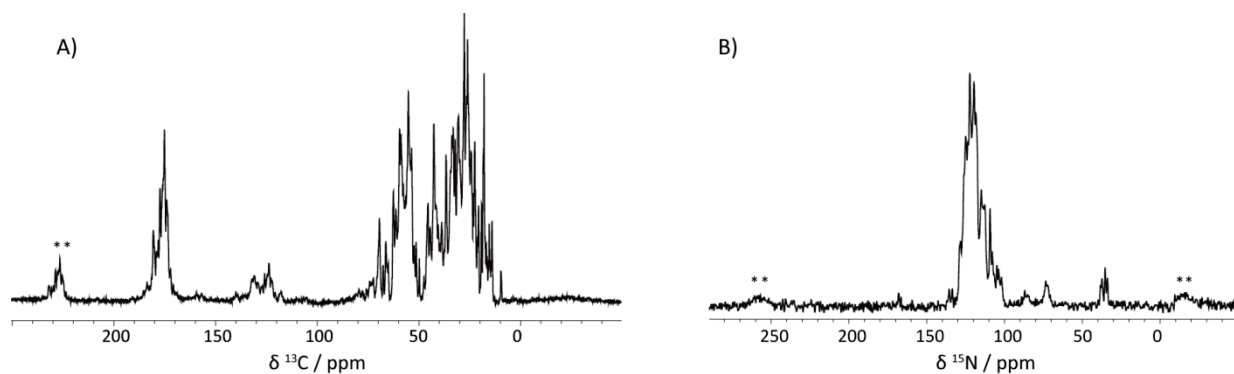


Figure 16 1D CP MAS spectra of Ubiquitin (MPD) A) ^{13}C spectrum, B) ^{15}N spectrum. Asterisks** represent sidebands due to the MAS spinning frequency of 11kHz.

The quality of the $[1-^{13}\text{C}]$ - and the $[2-^{13}\text{C}]$ -glc labeled samples is illustrated by the 1D ^{13}C CPMAS spectra (Figure 17). As expected, two alternative labeling schemes illustrated in Figure 14 show significant differences in the two spectra. For the $[2-^{13}\text{C}]$ -glc (magenta) labeled sample, a higher ^{13}C signal (labeling) observable for $\text{C}\alpha$ (around 60ppm) region and CO region (around 175ppm) carbon. Whereas, for the $[1-^{13}\text{C}]$ -glc sample (green), higher ^{13}C signal (labeling) observed in the aliphatic region (10-30ppm) in comparison with the $[2-^{13}\text{C}]$ -glc sample.

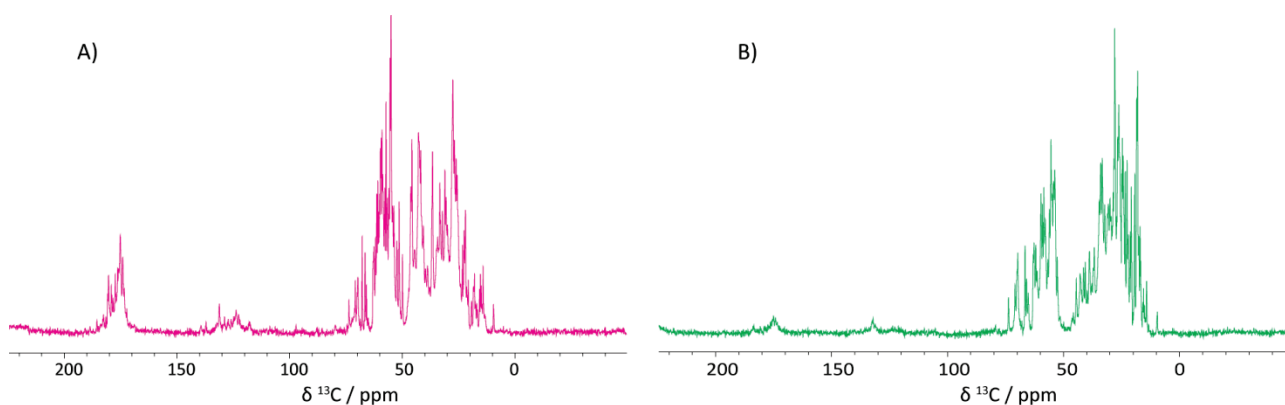


Figure 17 Excerpts of 1D ^{13}C CP-spectra of Ubiquitin (MPD) A) $[2-^{13}\text{C}]$ -glc (magenta) B) $[1-^{13}\text{C}]$ -glc (green).

5.1.2 A comparison of spectral quality

By comparison of the uniformly labeled sample to the two alternative labeling samples, a drastic increase of spectral resolution over the full spectral width can be observed. As an example, the improvement in

resolution for resonance peak of P19C δ 1 is demonstrated in Figure 18. For quantitative analysis, the improvement of spectral resolution can be estimated by the analyzing the full width at half maximum (FWHM) of the peak. For the uniformly labeled sample, a FWHM of 73Hz is obtained for the P19C δ 1 resonance. The [2- 13 C]-glc labeled sample has a FWHM of 57Hz corresponding to an improvement of spectral resolution of around 16Hz. However, P19C δ 1 is not labeled in the [1- 13 C]-glc approach. Here the reason for the improvement of resolution is due to the removal of one-bond 13 C- 13 C dipolar and J couplings by alternate labeling of the carbons in the amino acids.⁸⁵

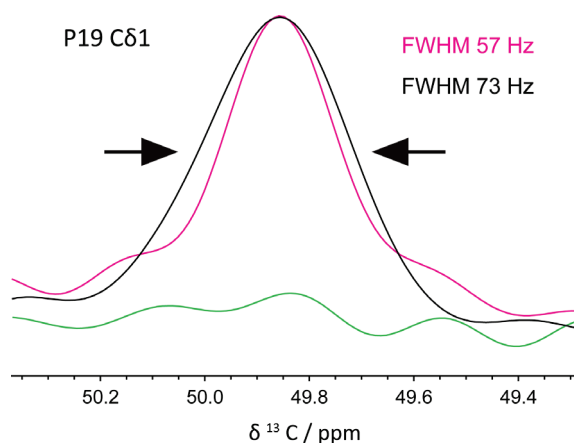


Figure 18 Excerpts of 1D 13 C CP-spectra of Ubiquitin for the resonance P19C δ 1; [2- 13 C]-glc (magenta), [1- 13 C]-glc (green) and [U- 13 C]-glc (black). Illustration of the improvement of resolution due to the reduction of one-bond J- and dipolar 13 C- 13 C couplings.

5.1.3 Assignment of ubiquitin

Higher-dimensional ssNMR-spectra are essentially needed for the assignments of resonances of biological systems. For instance, 2D homonuclear 13 C- 13 C correlation PDSB-spectra give already a good overview how many spin-systems of Ubiquitin are detectable (Figure 19.A). Overcoming overlapping effects in the 2D 13 C- 13 C-spectrum, the use of sparsely labeled samples such as [1- 13 C]-glc and the [2- 13 C]-glc proves to be as a successful labeling strategy. The strong effect of highly diluted 13 C carbon samples on resolution and assignment can be illustrated by recording of 2D 13 C- 13 C PDSB experiments as shown in Figure 19.A. A short mixing time for the 13 C- 13 C transfer in a PDSB-spectrum (50ms for [U- 13 C]-glc and 100ms for [2- 13 C]-glc [1- 13 C]-glc labeled samples) enable only the detection of intra-residual contacts. Longer mixing time for the sparsely labeled sample is mandatory due to the hindrance of the flip-flop spin diffusion processes inside the residues for achieving intra-residual information like in the uniformly labeled spectrum. The intra-residual cross-peaks for N60 and I61 residues are demonstrated as an example for the determination of amino acid spin systems (Figure 19). In a uniformly labeled sample, all carbons of a spin system/amino acid are 13 C labeled therefore one can record the full spin-system of the residue of interest. The simplification of the spectrum in the sparsely labeled sample allows the collection of spin-systems in a straightforward way as shown in Figure 19. In the [2- 13 C]-glc labeled sample C α -C β , C α -C γ 2, C β -C γ 1 and C β -C δ 1 correlations

were obtained for residue Ile61. For 2-glucose labeling scheme only $\text{C}\alpha\text{-C}\beta$, $\text{C}\alpha\text{-C}\delta 1$ and $\text{C}\beta\text{-C}\delta 1$ interactions should be detectable. The contacts between $\text{C}\alpha\text{-C}\gamma 2$ and $\text{C}\beta\text{-C}\gamma 1$ can be understood by scrambling effects due to the biosynthetic pathway of the amino acids. For the $[1\text{-}^{13}\text{C}]\text{-glc}$ labeled sample the collection of contacts between $\text{C}\alpha\text{-C}\beta$, $\text{C}\alpha\text{-C}\gamma 1$, $\text{C}\alpha\text{-C}\gamma 2$, $\text{C}\alpha\text{-C}\delta 1$, $\text{C}\beta\text{-C}\gamma 1$, $\text{C}\beta\text{-C}\gamma 2$, $\text{C}\gamma 1\text{-C}\gamma 2$, $\text{C}\gamma 1\text{-C}\delta 1$ and $\text{C}\gamma 2\text{-C}\delta 1$ can be observed. With respect to the prediction of the 1-glucose labeling scheme we distinguish scrambling effects for the contacts $\text{C}\beta\text{-C}\gamma 1$ and $\text{C}\beta\text{-C}\gamma 2$. For the residue N60 the interactions $\text{C}\alpha\text{-C}\beta$ and $\text{C}\beta\text{-C}\gamma$ reveal only in the $[1\text{-}^{13}\text{C}]\text{-glc}$ labeled sample which is consistent with the prediction of the 1-glucose labeling scheme.¹⁴

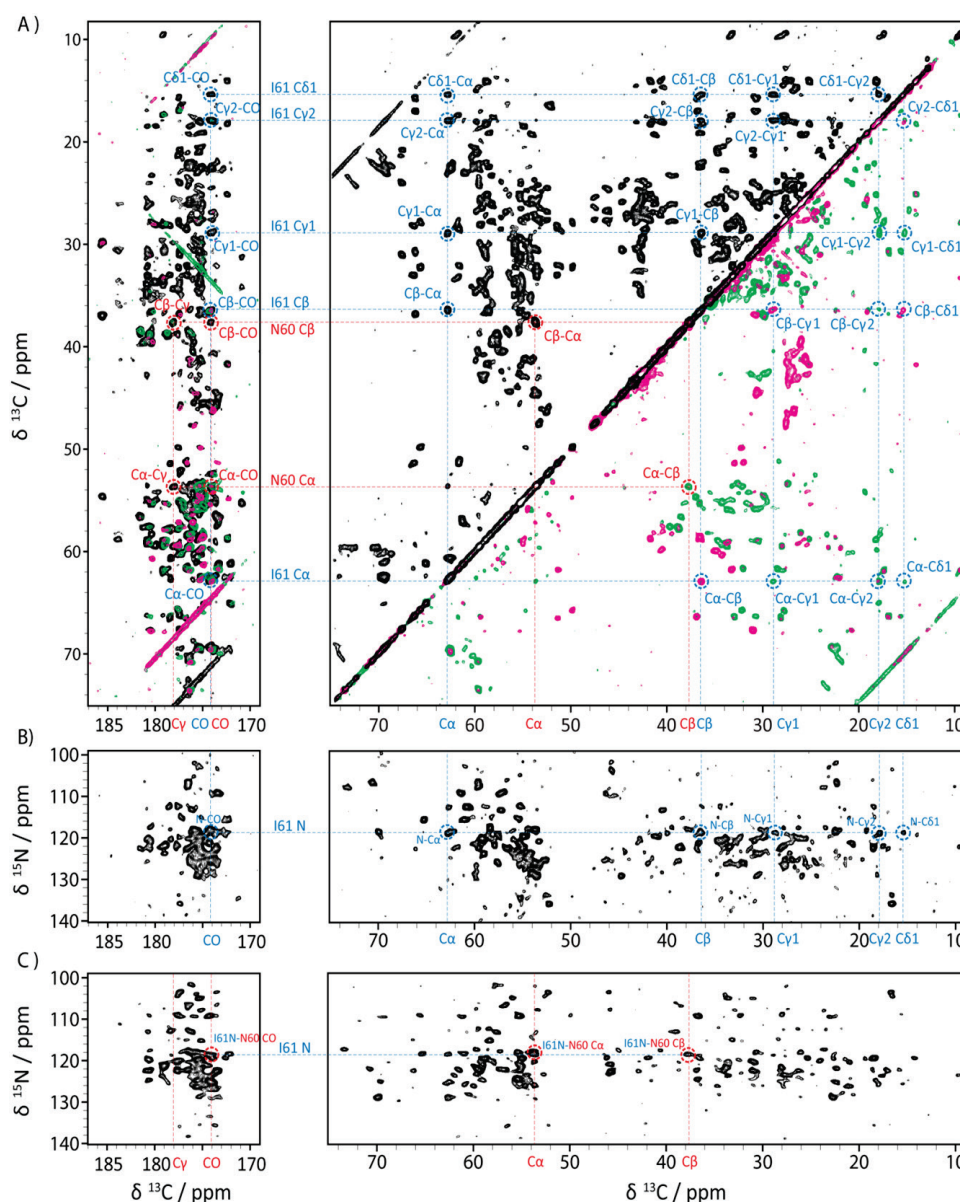


Figure 19 Illustration of the assignment strategy for microcrystalline ubiquitin A) 2D PDSD-spectra of $[U\text{-}^{13}\text{C}]\text{-glc}$ (black), $[2\text{-}^{13}\text{C}]\text{-glc}$ (magenta) and $[1\text{-}^{13}\text{C}]\text{-glc}$ (green) labeled ubiquitin recorded with mixing times of 50ms ($[U\text{-}^{13}\text{C}]\text{-glc}$) and 100ms ($[2\text{-}^{13}\text{C}]\text{-glc}$, $[1\text{-}^{13}\text{C}]\text{-glc}$). The full spin-system connections of N60 (red) and I61 (blue) are shown as an example. (B) 2D NACX-spectrum of $[U\text{-}^{13}\text{C}]\text{-glc}$ labeled ubiquitin, demonstration of the nitrogen assignment of I61 by means of the unambiguous resonance of I61C $\delta 1$ C) 2D NCOX-spectrum of $[U\text{-}^{13}\text{C}]\text{-glc}$ labeled ubiquitin, sequential assignment of the residue N60 using the unambiguous resonance of N60C β .

Nitrogen chemical shifts and sequential connection were obtained from NCA-, NCO-, NCACX- and NCOCX-spectra as shown in (Figure 19.B,C). Based on the knowledge of the nitrogen chemical shift of I61 and the unambiguous chemical shift of the N60 C β the sequential assignment of residue N60 can be facilitated in a 2D NCOCX-spectrum.

5.1.4 Stereospecific assignment from sparsely labeling approach

Based on the assignment of residue I61 by using [2-¹³C]-glc and [1-¹³C]-glc samples (Figure 19.A) we could distinguish scrambling products for the amino acids valine and leucine in the [2-¹³C]-glc labeled sample. Due to the knowledge of the stereospecific assignment of the prochiral methyl groups of valine and leucine from the 10% ¹³C labeled approach¹⁸ we were enabled to determine the assignment of the prochiral methyl groups of valine and leucine due to the scrambling effects in the [2-¹³C]-glc labeled sample.²⁶ As a representation the stereospecific assignments of the residue L56 and V17 are illustrated in Figure 20. For V17 we observed cross-peaks between C γ 1-C α and C γ 2-C β (Figure 20.A) and for L56 C δ 1-C β and C δ 2-C γ correlations were detectable (Figure 20.A). Such an observation of stereo-specific correlations between C γ 1-C α and C γ 2-C β of V17 and C δ 1-C β and C δ 2-C γ for L56 can be explained by the resultant biosynthetic precursors [2-¹³C]-pyr and [3-¹³C]-pyr.²⁶

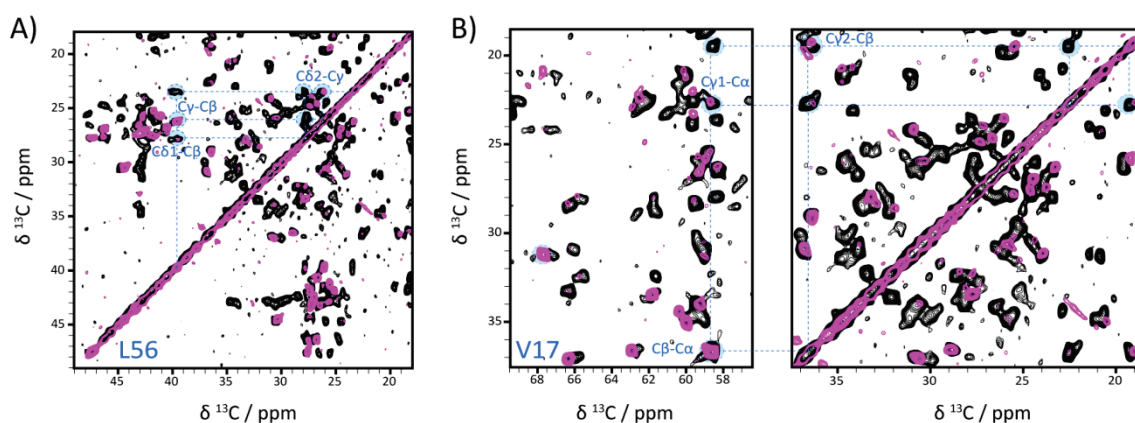


Figure 20 2D PDSD-spectra of [U-¹³C]-glc (black, mixing time is 50ms) and [2-¹³C]-glc (magenta, mixing time is 100ms) The stereospecific correlations are highlighted in light blue and dotted lines A) illustration of the stereospecific assignment of the prochiral methyl groups C δ 1/2 of L56 based on the correlations of C δ 1-C β and C δ 2-C γ B) the collection of the stereospecific assignment of V17 C γ 1/2 was obtained based on the correlations C γ 1-C α and C γ 2-C β .

As a consequence of the glycolysis of [2-¹³C]-glc only the production of [2-¹³C]-pyr can be predicted. The achievement of [3-¹³C]-pyr, generated from scrambling effects, requires a more detailed description of metabolic and catabolic processes of amino acids. Possible mechanisms which have to be taken into account are the followings: Entner-Doudroroff pathway⁸⁶, breakdown of oxaloacetate via gluconeogenesis, conversion of malate, amino acid degradation together with the C₁ metabolism.⁸⁷ The most reasonable biosynthetic pathway for explanation [3-¹³C]-pyr production is the pentose phosphate pathway (PPP) which leads to [3-¹³C]-pyr and [1,3-¹³C]-pyr.²⁶

However, the biosynthesis of Val and Leu generated from pyruvate is stereo-selective.^{88,89} The first step of the biosynthesis of valine is the reaction between pyruvate and the cofactor thiamine pyrophosphate (TPP). The resultant product (hydroxy-ethylidene-thiamine-pyrophosphate) reacts with pyruvate producing α -acetolacetate. After rearrangement of the tertiary ketol group by the stereo-selective migrating of the α -methyl group α -ketoisovalerate will be generated. The last step is the transformation of the keto group into an amino group receiving valine. Both possibilities of the different ^{13}C labeled pyruvates are illustrated in Figure 21.A.^{26,87}

The biosynthesis of leucine starts from α -ketoisovalerate, which reacts with acetyl-Co-A generating α -isopropylmalate. After decarboxylation and migration of the hydroxy group by the pinacol rearrangement α -ketoisocaproate will be produced. The last step of the synthesis of leucine is the transformation of the keto group into the amino groups at the α -position of the molecule (Figure 21.B).²⁶

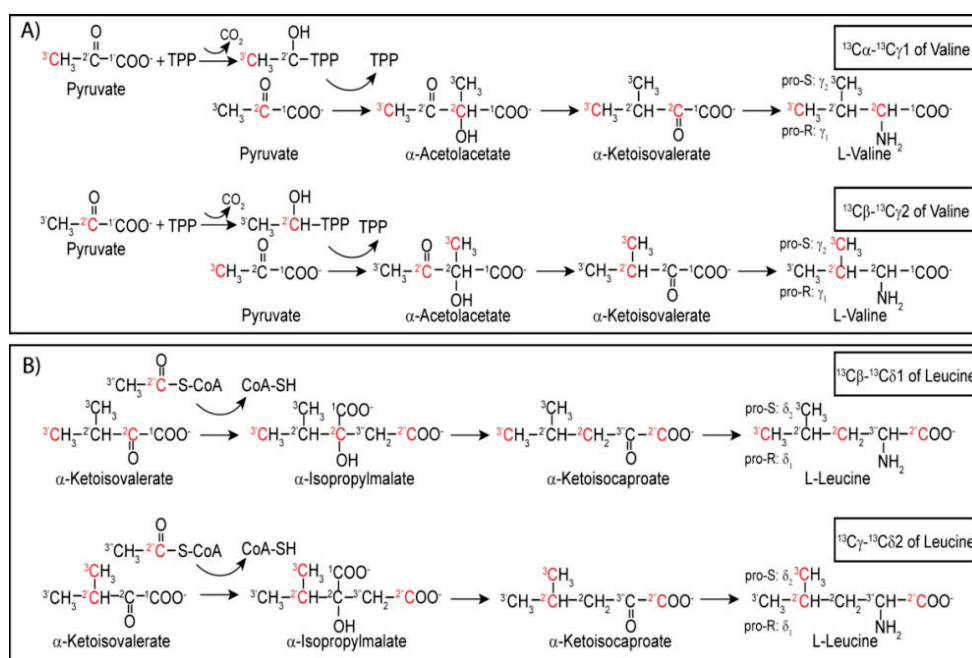


Figure 21 Illustration of the stereo-selective biosynthetic pathway of valine and leucine, A) formation of correlations between $^{13}\text{C}\alpha$ - $^{13}\text{C}\gamma1$ and $^{13}\text{C}\beta$ - $^{13}\text{C}\gamma2$ of valine and $^{13}\text{C}\beta$ - $^{13}\text{C}\delta1$ and $^{13}\text{C}\gamma$ - $^{13}\text{C}\gamma2$ of leucine originating from $[2\text{-}^{13}\text{C}]$ -pyr and $[3\text{-}^{13}\text{C}]$ -pyr. ^{13}C labeled carbons are highlighted in red. (Illustrations are taken from Ref.²⁶)

5.1.5 Sequential assignment of ubiquitin

The sequential assignment of I61 to N60 can be confirmed by higher selective ssNMR experiments such as the 2D $\text{CA}_i(\text{N}_i)\text{CO}_{i-1}\text{CA}_{i-1}$ experiment.¹² The magnetization transfer in this experiment starts at the proton of $\text{C}\alpha$ of i^{th} residue and goes through the backbone atoms N_i and CO_{i-1} and ends at the $\text{C}\alpha$ of $(i-1)$ th residue. In Figure 22, the sequential walk from I61 down to S57 is illustrated by contacts between the $\text{C}\alpha$ of residue (i) to the resonances $\text{C}\alpha$ and CO of residue $(i-1)$.

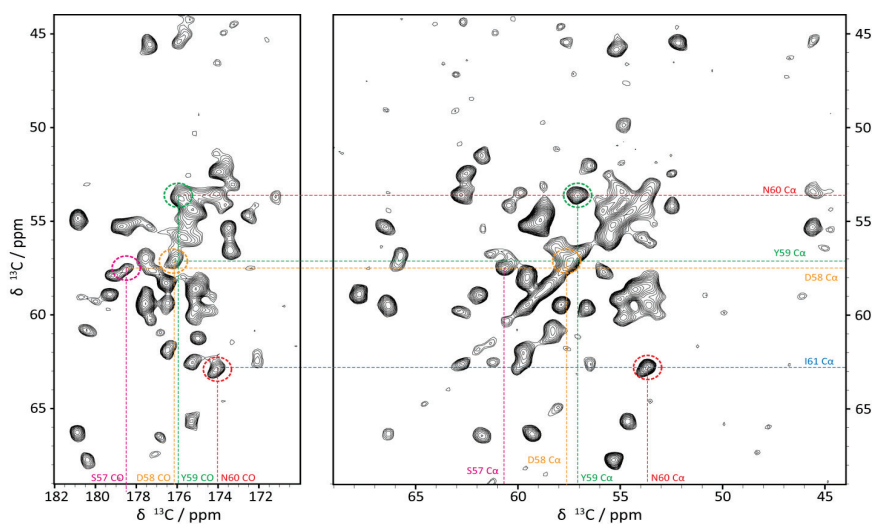


Figure 22 2D $CA_i(N_i)CO_{i-1}CA_{i-1}$ -spectrum of $[U-^{13}C]$ -glc, illustration of the sequential walk from residue I61 to residue S57, [I61 (blue),N60 (red), Y59 (green), D58 (orange) and S57 (magenta)].

Combining several multidimensional ssNMR experiments, including BSH-CP based 3D spectra (Figure 23), we were able to assign the full backbone from residue M1 to V70 for the first time. It is worth to mention that the C-terminal part of ubiquitin (residue R72 to G76) is not detectable in dipolar based ssNMR experiments probably due to its high dynamic behavior. We also observed multiple peaks for certain residues revealing the sample heterogeneity of certain parts of the protein. With the available high-resolution ssNMR spectra, we were able to determine three different conformations with almost equal populations (36% (main conformation), 31% (second conformation), and 33%) as judged from their spectral intensities of the NCA-spectrum (Figure 24.A).

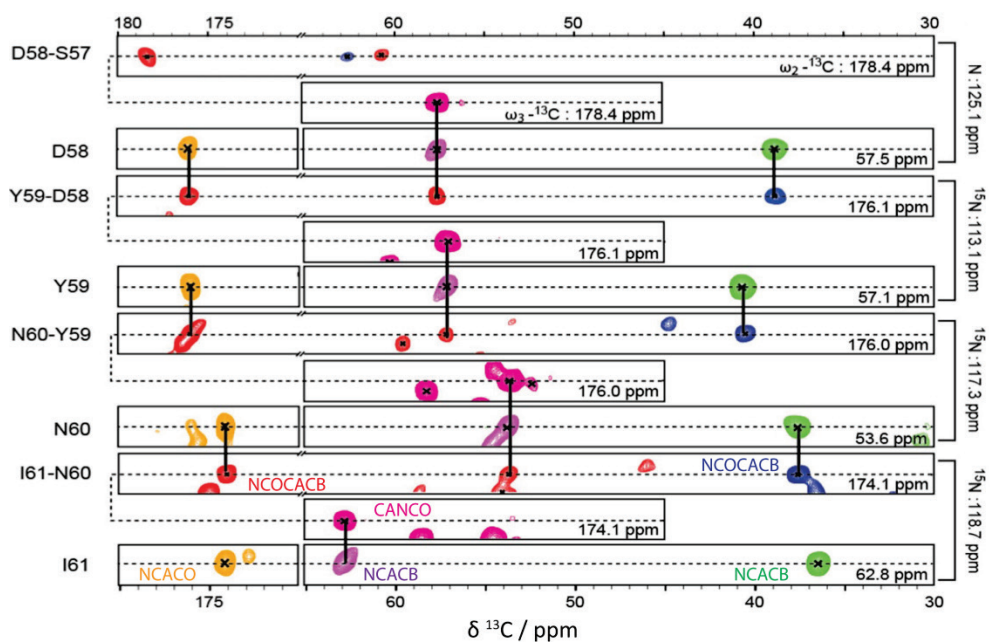


Figure 23 Strip plot of different 3D experiments showing the sequential walk for the amino acid stretch D58-I61. NCOCACB in red (positive signals) and blue (negative signals), CANCO in magenta, NCACO in orange and NCACB in purple (positive signals) and green (negative signals). Solid lines represent the assignments used in the sequential walk.

Furthermore, the observation of change in chemical shifts for different polymorphs indicates strong structural heterogeneity in the loop region β 1- β 3 (including the C-terminal tip of the β -helix β 1), the third β -strand β 3 (Q41-F45) and the adjacent residues up to Lys 48 as well as for the fifth β -strand β 5 (residues T66-V70) (Figure 24.B). Interestingly, in solution NMR, these regions showed increased dynamics on the supra- τ_c time scales as revealed by RDC-based studies.^{50,51,90} In particular loop β 1- β 2 and loop β 1- β 3, including the C-terminal tip of the α -helix, have been identified previously to be involved in large amplitude collective motion, resembling a “pincer like” motion, which was related to conformational sampling of ubiquitin during molecular recognition.⁵¹ It is conceivable that the conformational sampling of ubiquitin in solution will manifest itself as structural heterogeneity during the crystallization process of ubiquitin in MPD, leading to slightly different conformations, which can be identified, based on their chemical shift differences in the solid state. It should be noted that Ala46, close to Lys48, whose side chains is the major recognition site of ubiquitin, in particular during the poly-ubiquitination process⁹¹, shows the biggest difference in ^{15}N chemical shifts (Figure 24.B). More recent studies suggested that β -strands β 1, β 2, β 3 and β 5 are involved in a concerted motion across the β -sheet mediated by the hydrogen bond network.^{51,92}

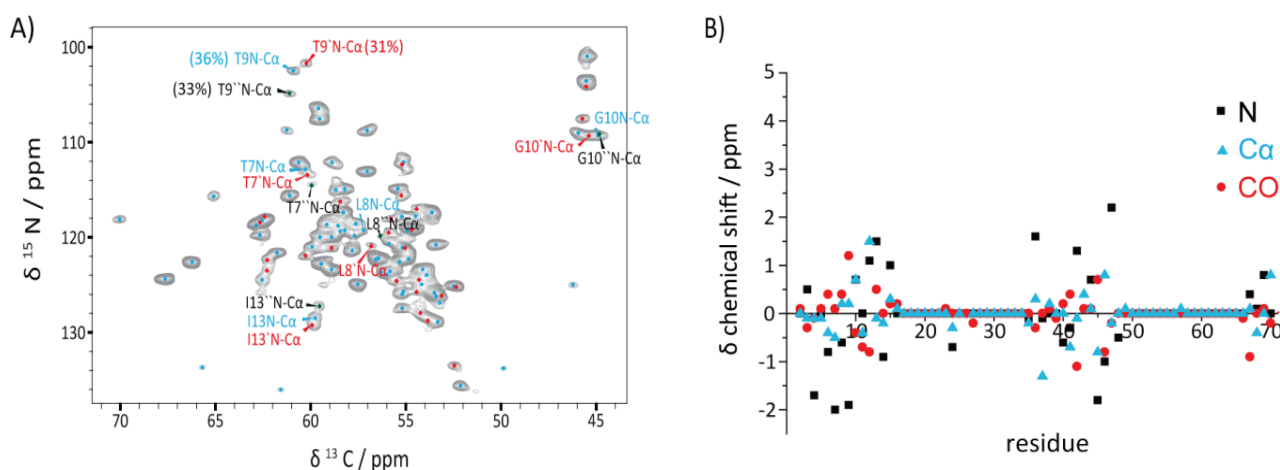


Figure 24 A) 2D NCA-spectrum of uniformly ^{13}C labeled sample, residues are colored as follows: main conformation (blue), second (red) and third conformation (black)). Population distribution is calculated by integration of the NCA cross-peaks of residue T9 B) Illustration of chemical differences of the backbone atoms between the main- and the second conformation.

Identification of highly flexible region of proteins can be realized by conducting INEPT-spectra which are based on J-coupling heteronuclear magnetization transfers. Only the precipitating agent MPD was detectable in the 2D INEPT-spectrum, which indicates that the undetectable C-terminal region of ubiquitin lies in the intermediate dynamic range (Figure 25).

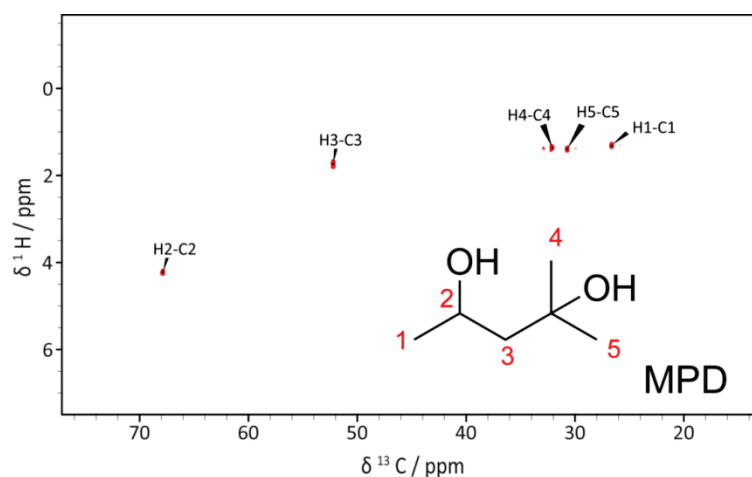


Figure 25 2D ^1H - ^{13}C INEPT-spectrum of microcrystalline ubiquitin, showing that only the precipitating agent (MPD) is detectable.

The structural heterogeneity observed in the solid-state localizes to the β -strand network and the described loop regions but is, for example, largely absent for the α -helix, further supporting the idea of conformational sampling of ubiquitin in solution being the origin of the polymorphism observed in the solid-state spectra. To investigate the origin of the revealed polymorphism, the recording of 2D experiments at different temperatures is required. As the freezing temperature of the buffer is around -11°C and the crystallization temperature is around 24°C the sample temperature window was set between -4°C to 20°C (Figure 26). Small chemical shift changes can be distinguished in NCA-spectra for the residues T9'' (third conformation), G47', G35' (second conformation) and E34 (main conformation) which belongs to flexible loop-regions.

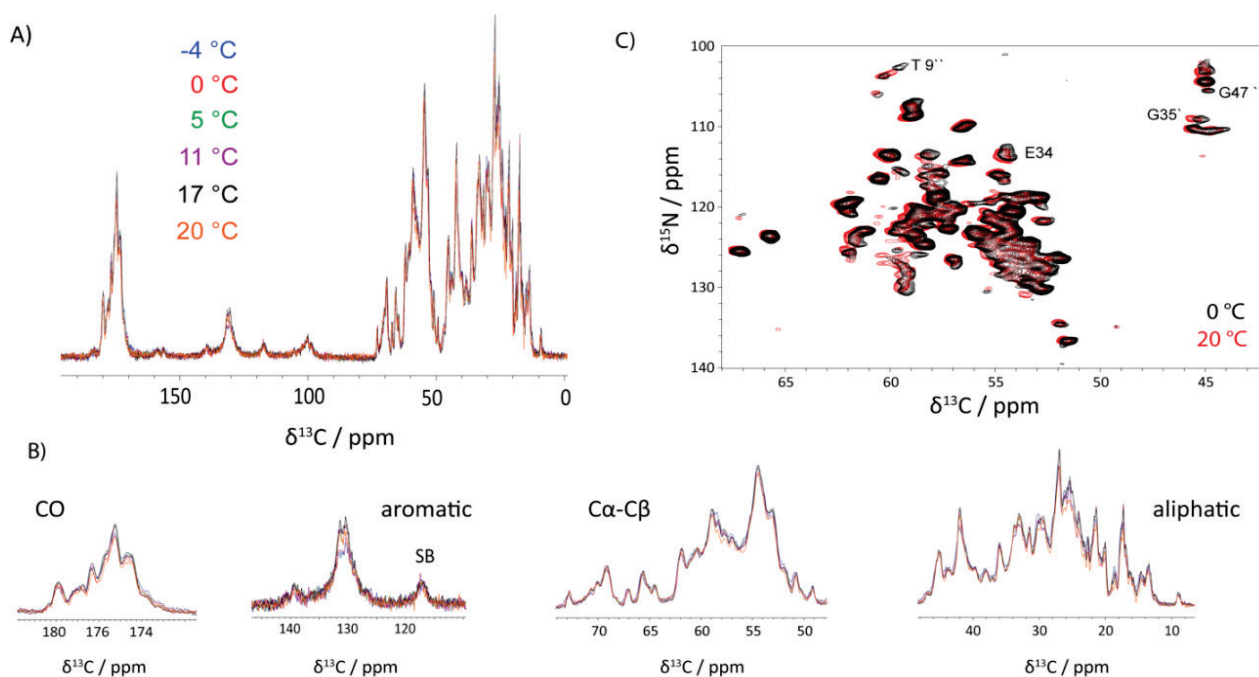


Figure 26 A) Presentation of 1D ^{13}C CP-spectra of uniformly labeled microcrystalline ubiquitin at different temperatures. (from -4°C to 20°C), B) Excerpts of different chemical shift regions of 1D ^{13}C CP-spectra (full spectra shown in panel A) C) Comparison of NCA-spectra at different temperatures (black (0°C), red (20°C)), residues which show observable chemical shift changes are highlighted.

In comparison, in the 2D PDSD-spectra the biggest chemical shift change of 0.32ppm can be identified for the resonance L50C δ 2 shown in Figure 27. The residue L50 is located at the beginning of the loop region between the fourth β -strand β_4 and the second α -helix α_2 .

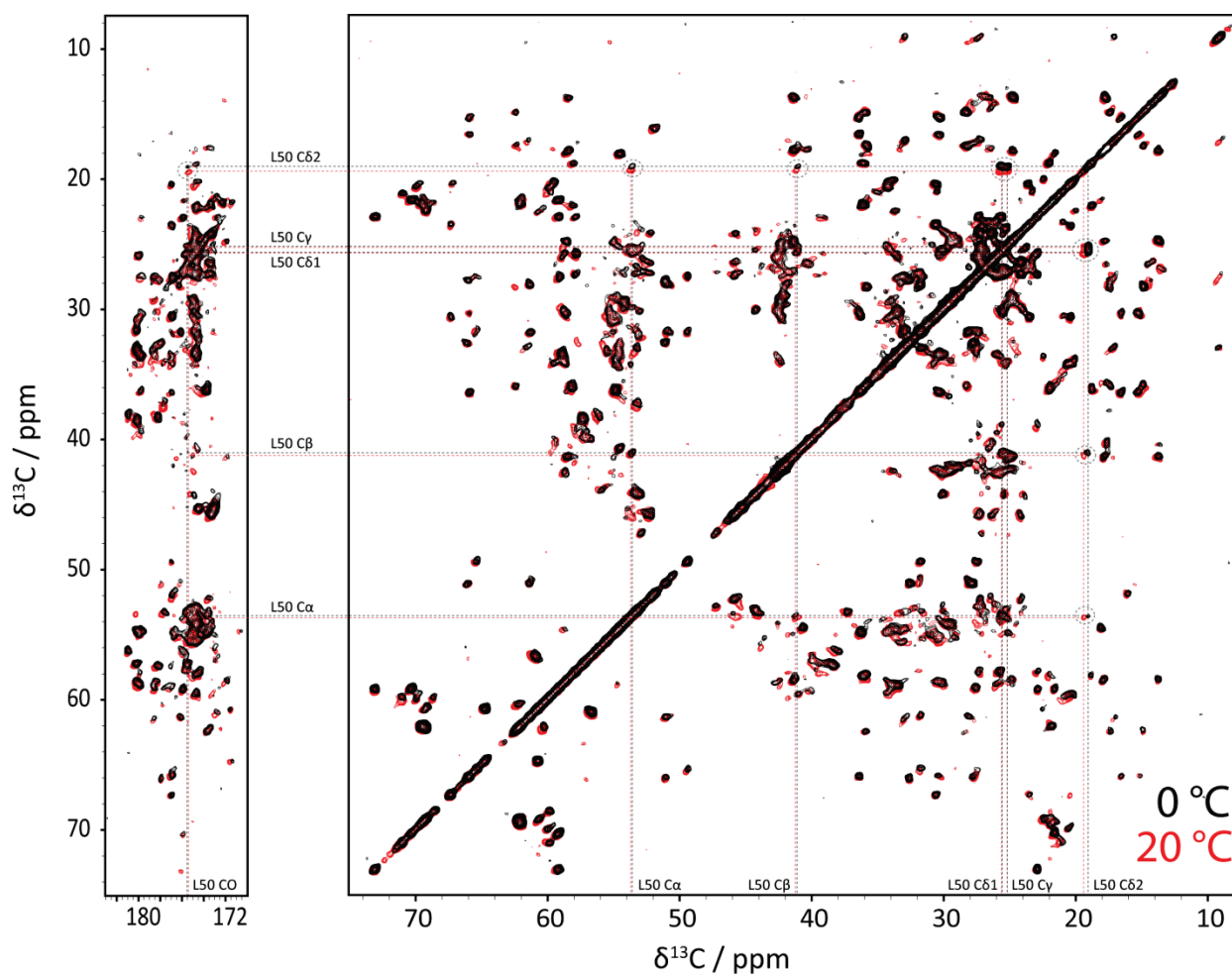


Figure 27 Comparison of 2D PDSD-spectra at two different temperatures (red (20°C), black (0°C)), the full spin-system of L50 is represented with dashed lines, and the intra residual correlation between C δ 2 to the other sites are indicated with a dashed grey circle.

5.1.6 Secondary structure analysis of ubiquitin

The identification of secondary structure of proteins can be determined in ssNMR with two prevailing methods; the secondary chemical shift analysis which compares experimentally observed carbon chemical shifts under MAS to standard isotropic random coil values⁹³, or with the backbone dihedral angle prediction obtained from the program TALOS+. For glycine, the second term $\Delta\delta C\beta$ of the equation for the secondary chemical shift analysis was set to zero. The result of the secondary chemical shift analysis for the residue range M1-V70 is illustrated in Figure 28. One can distinguish β -sheet like conformation for the residues M1-T7, L15-E18, Q40-F45, and between T66-V70, indicated by negative $\Delta\delta C\alpha - \Delta\delta C\beta$ values. Residue regions from T22-K33 and L56-D58 show α -helical like structure properties, appropriated by positive $\Delta\delta C\alpha - \Delta\delta C\beta$ values.

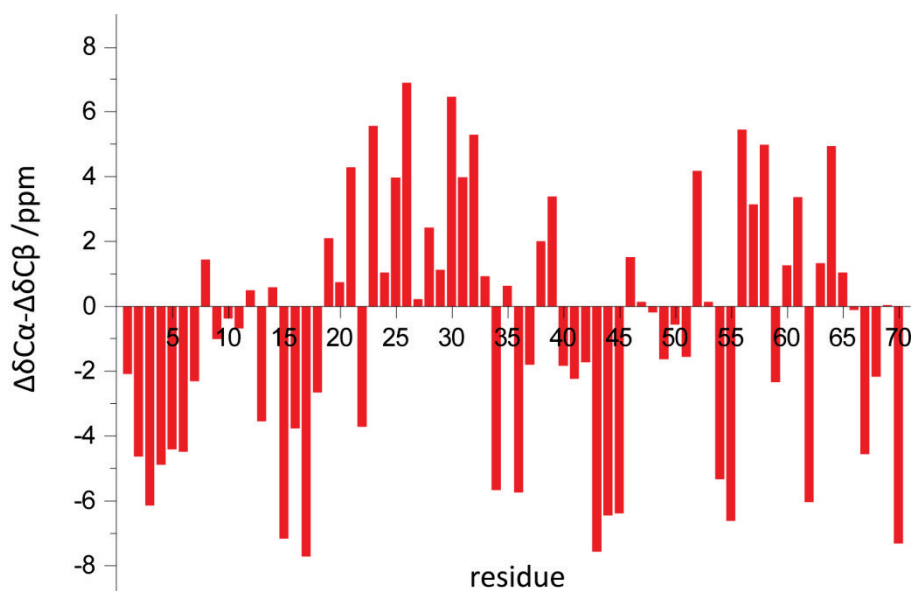


Figure 28 Secondary chemical shifts analysis of the main conformation of microcrystalline ubiquitin.

The result of the TALOS+ prediction is shown in Figure 29.A. For the sake of clarity, the TALOS+ prediction result is represented as a cartoon (Figure 29.B). From the cartoon, β -strands can be identified for the residue regions Q2-K6, K11-V17, R42-A46, Q49, and T66-V70, while α -helical structure revealed for residue regions I23-K33 and T55-Y59.

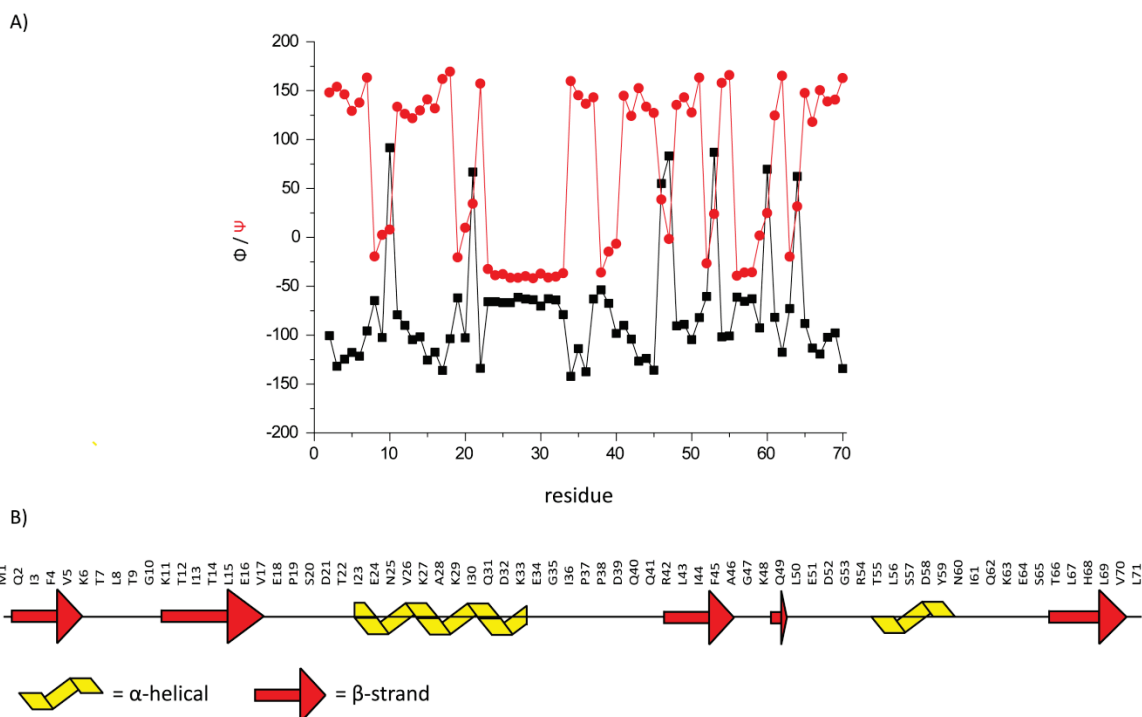


Figure 29 A) TALOS+ analysis, predicting Phi (black) and Psi (red) dihedral torsion angles of the main conformation of ubiquitin B) cartoon of the result of TALOS+ prediction, only residue E34 shows ambiguity for the secondary structure prediction.

5.1.7 Collecting of distance restraints information

The major aspect of calculation a protein structure in ssNMR is the collection of distance-restraints information. Due to the improvement of spectral resolution and the reduction of spectral crowding by the sparsely labeling schemes, one could obtain 518 distance-restraints (59 unambiguous-, 126 network- and 333 ambiguous distance restraints). As an example, the distance information by ^{13}C - ^{13}C correlations is given in Figure 30. The majority of restraints for the site V26 α lies in the medium-range [$1 < |i-j| < 5$], which is consistent due to its α -helical environment. The contact to D21 β gives useful information about the structural conformation of the loop between the β 2-strand and α 1-helix of ubiquitin (shown in Figure 30.C). The long-range [$|i-j| \geq 5$] contacts to I3 and L15 (demonstrated in Figure 30.C) are in a strong agreement with the X-ray structure 3ONS⁴⁶, which indicates already the right folding of the microcrystalline ubiquitin. For the resonance T55 β the medium-range contacts to D58 α , D58 β and S57 α depicted in Figure 30.D confirmed the secondary α -helical structure of ubiquitin, which was predicted by TALOS+ and estimated by the secondary chemical shift analysis. The long-range distance restraints to the residue S20, D21, T22 and I23 strengthened the right folding of the protein (Figure 30.D).

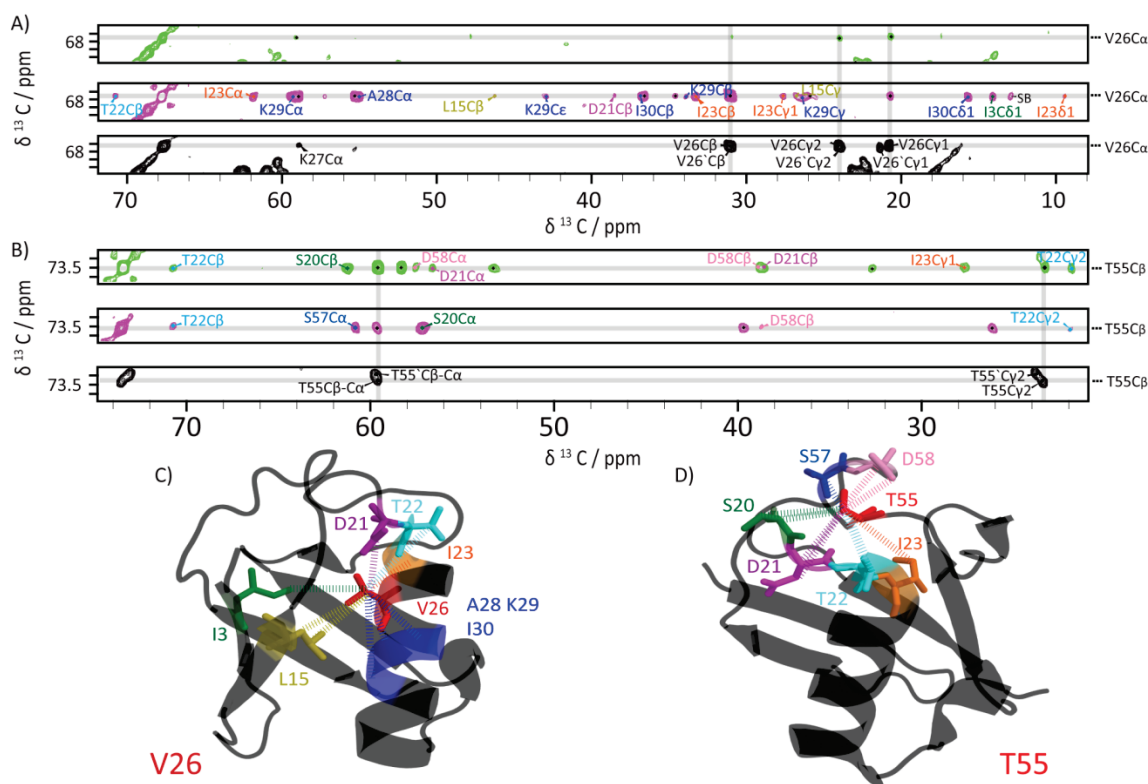


Figure 30 Distance restraints collected for V26 C α (A-C) and T55 C β (B-D). A-B Excerpts of 2D PDSD-spectra of [^{13}C]-glc labeled ubiquitin with a mixing time of 50ms (black) and of [$1\text{-}^{13}\text{C}$]- and [$2\text{-}^{13}\text{C}$]-glc labeled ubiquitin with mixing times of 900ms (green and magenta, respectively). Intra-residue and sequential correlations are labeled in black, medium- and long- range contacts are labeled in a residue-specific color: I3 (green), L15 (olive green), S20 (green), D21 (violet), T22 (cyan), I23 (orange), V26 (red), A28, K29 and I30 (all blue), T55 (red), S57 (blue) and D58 (pink). C-D) Illustration of the distance restraint collection for residues V26 C) and T55 D) on the X-ray structure 3ONS using the same color code as in (A-B). For the sake of clarity, no side-chains are shown for residues I23, A28, K29 and I30 in panel C.

Figure 31 illustrates the distance distribution of 59 unambiguous distance restraints as back-calculated in comparison to the X-ray structure 30NS.⁴⁶ The majority of the observed correlations were found to correspond to a distance between 4.5 to 7 Å.

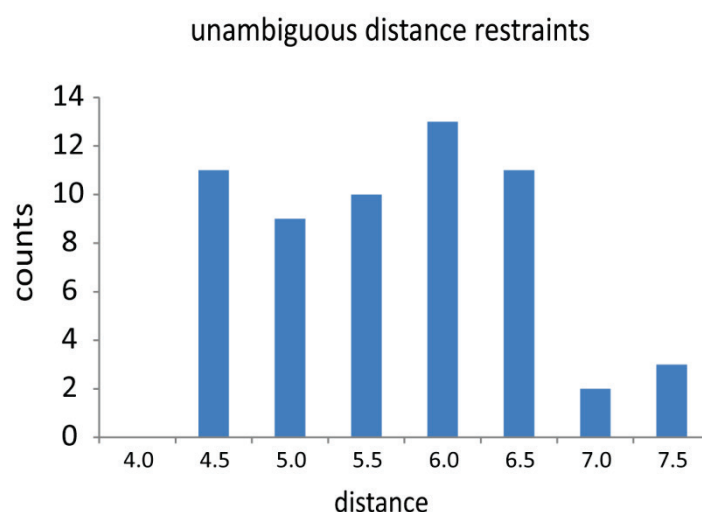


Figure 31 Distance distribution of 59 unambiguous distance restraints as collected from the spectra, the corresponding distance as extracted from the X-ray structure is shown.

Distance restraints which are originally named to be ambiguous can be specified as network distance restraints if the process of disambiguation of restraints due to unambiguous assignments involves the same residues. As an example for the definition of a network distance restraint the ¹³C-¹³C correlation between I61Cy2-F45Cβ is illustrated in Figure 33. The distance distribution of the 59 unambiguous- and the 126 network distance restraints as back-calculated in comparison to the X-ray structure are depicted in Figure 32. As illustrated in Figure 32 network correlations result in distance information increased to 9 Å.

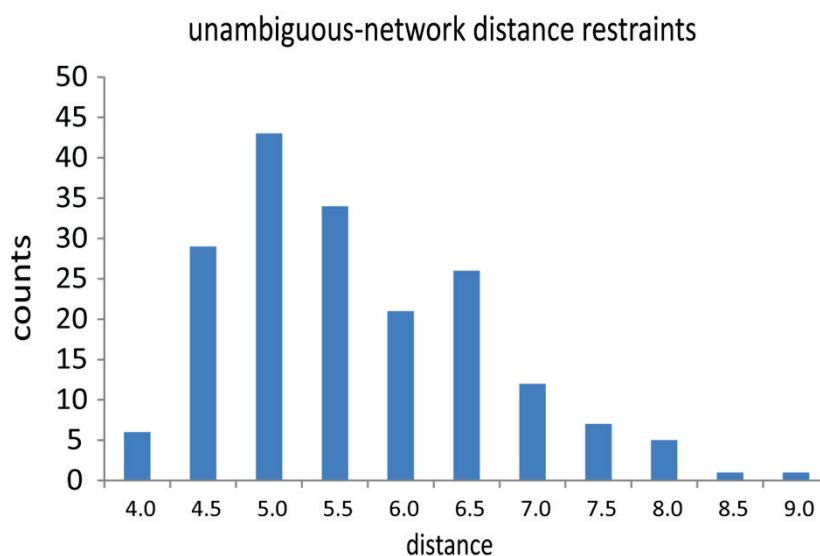


Figure 32 Distance distribution of 59 unambiguous and 126 network distance restraints as collected from the spectra, the corresponding distance as back calculated from the X-ray structure is shown.

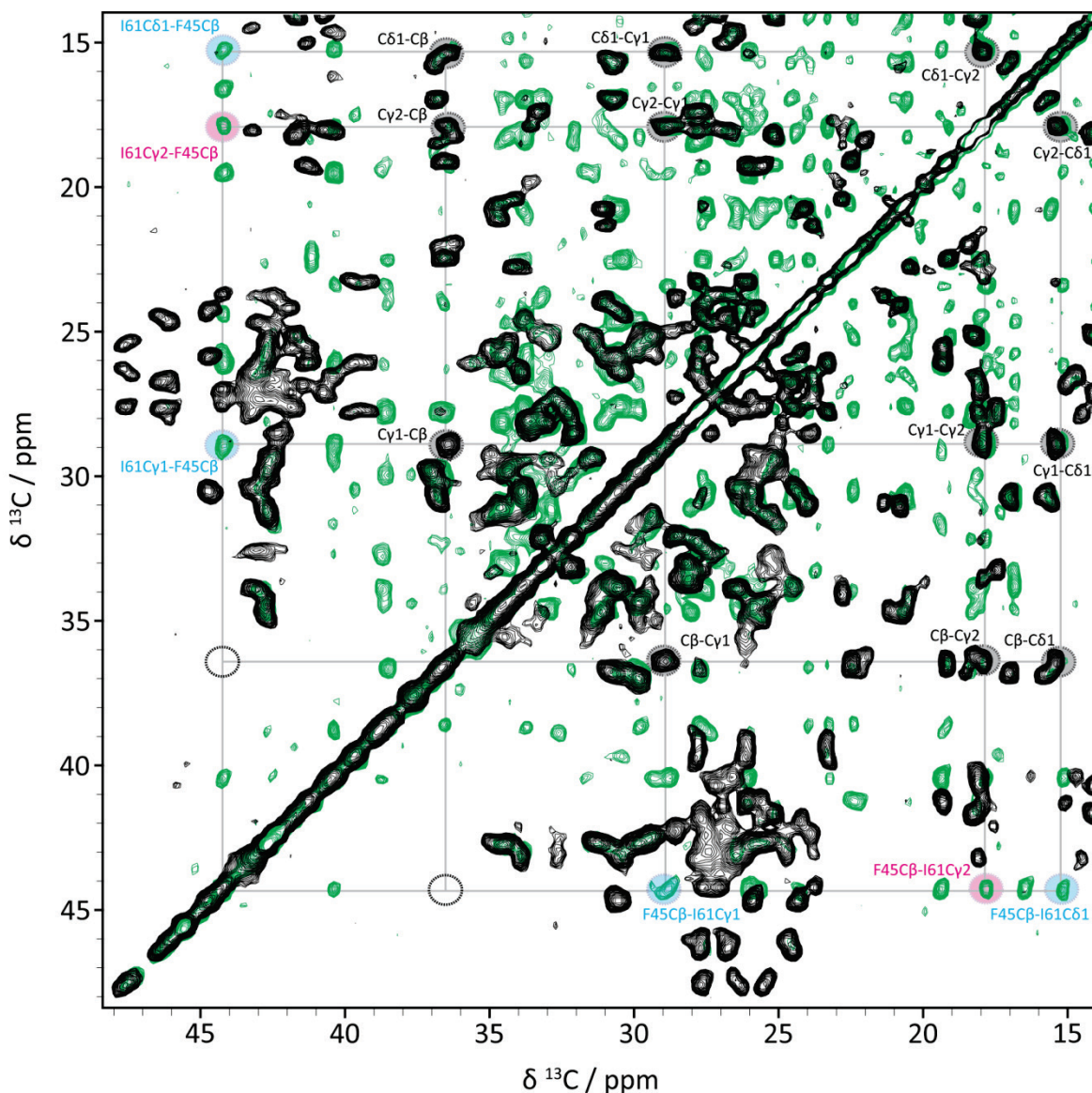


Figure 33 Comparison of 2D PDS-spectra of $[U-^{13}\text{C}]$ -glc (black: mixing-time of 50ms) and $[1-^{13}\text{C}]$ -glc (green: mixing time of 900ms) labeled ubiquitin. The correlations I61C δ 1-F45C β (light blue) and I61C γ 1-F45C β (light blue) are unambiguous due to the carbon sites labeled with 1-glucose²³, the cross-peak I61C γ 2-F45C β (magenta) is ambiguous with respect to the I61C γ 2 assignment. However, considering the unambiguous distance restraint assignments of I61C δ 1-F45C β and I61C γ 1-F45C β , the restraint I61C γ 2-F45C β is treated as “network unambiguous”. Networking describes in the context of distance restraint assignment the process of disambiguation of restraints due to unambiguous assignments involving the same residues. Note that the correlation I61C β -F45C β is not present in the $[1-^{13}\text{C}]$ -glc-spectrum, I61C β being unlabeled (marked with a black circle).

5.1.8 Structure calculation

Structures of microcrystalline ubiquitin were determined following the standard *de novo* structure calculation procedure with XPLOR-NIH which starts from a random coil and subsequently adds the collected distance restraints to fold the protein. In total, we used 518 inter-residue distance restraints, including 59 unambiguous, 126 network and 333 ambiguous distance restraints for the structure refinement. In order to calculate the structures, an allowed variation of the observed contacts was set to 1-7 Å (Figure 34). The correct 3D-fold of microcrystalline ubiquitin is already achievable by the use of the 59 unambiguous

distance restraints. Including the network correlation the precision of the resultant bundle of structures is sufficient for the collection of ambiguous distance restraints. (Figure 34)

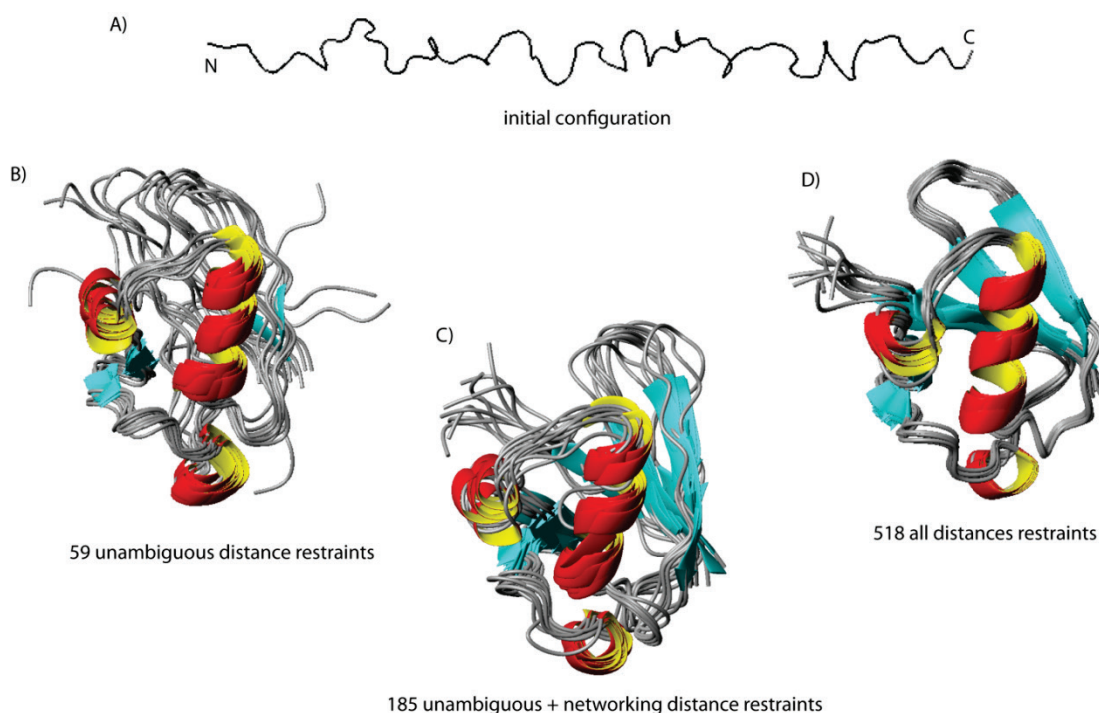


Figure 34 Representation of the various steps in de novo protein calculation. A) illustration of the extended configuration (random coil) used as starting point of the structure calculation of ubiquitin. Figures (B-D) show the ten lowest energy calculated structures for each distance restraints input. The 10 selected monomers were aligned on the backbone atoms using MOLMOL B) With the use of 59 unambiguous distance restraints, the bundle of monomers reveal already the correct 3D fold of Ubiquitin C) representation of 59 unambiguous and 126 networking distance restraints D) result of all 518 collected distance restraints (59 unambiguous, 126 network and 333 ambiguous distance restraints).

In Figure 35, the local rmsd values of the ten lowest calculated structures reveal same deviation in comparison to each other. The highest differences appear at the N-terminal and C-terminal ends of the protein as well as for the flexible loop regions (T7-G10), (K34-Q40).

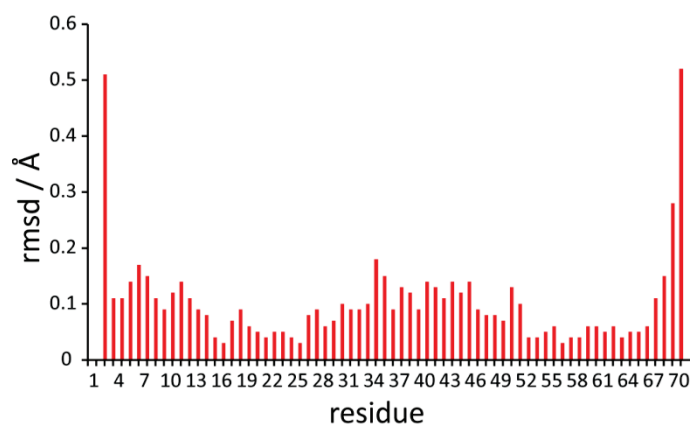


Figure 35 Plot of local backbone rmsd values of the 10 lowest energy structures (using distance restraints in the range of 1-7 Å).

For determination of the correct distance restraints range, for achieving the highest precision for the ten lowest energy monomers and the best accuracy to the X-ray structure 3ONS, structures were calculated with different distance ranges (shown in Figure 36). The upper limit range varied from 5 up to 20 Å. Interestingly the best result (see Table 3) could be generated by using the distance range 1-7 Å which is in a high agreement with the obtained unambiguous distances (see Figure 31) collected from the PSDS-spectra. The validation of the structure calculation by using the distance range 1-7 Å is listed in Table 9 (Appendix I).

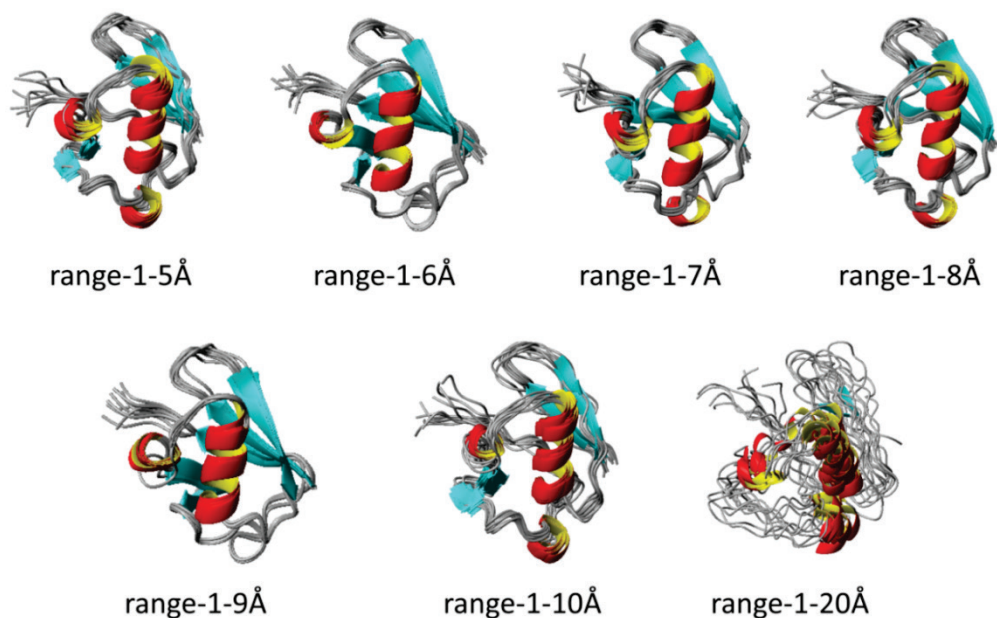


Figure 36 Comparison of the 10 lowest energy structures with varying upper distance bounds. Here the upper distance limit between two atoms that show a cross peak in the 2D PSDS-spectra was varied between 5 and 20 Å. For the final structure calculation a distance between 1 and 7 Å was allowed for each pair of atoms, as this was the lowest upper limit that did not result in any distance violations and gave good Ramachandran statistics while still exhibiting reasonably low rmsd of the bundle (compare Table 3 for details).

Table 3 Statistics of the calculated 10 lowest energy structures with different distance restraint limits

All rmsd values were calculated with MOLMOL⁹⁴ using all backbone atoms of residues 1-70. The rmsd values for validation of the precision were obtained using the 10 lowest energy monomers. Accuracy rmsd values were derived using the lowest energy monomer in comparison to the X-ray structure. The distance restraints violations and the Ramachandran analysis were taken from the PSVS⁷⁶ (Version 1.5, PDB Stat Version 5.9) results.

Range of Distance-Restraints (Å)	rmsd /precision (Å)	rmsd /accuracy (Å)	Violation of distance restraints > 0.5 Å	Disallowed regions in Ramchandran plot (%)
1-20	4.0	6.79	0	0
1-10	1.3	2.68	0	0
1-9	1.0	2.56	0	0
1-8	0.9	2.33	0	0
1-7	0.7	1.57	0	0.3
1-6	0.6	1.79	11	1.5
1-5	0.5	1.98	203	1.4

5.1.9 Comparison to different accessible PDB structures

For the comparison to the different structures only the structure calculated with the lowest energy value is taken (range 1-7 Å). Comparisons to structure PDB: ID 2JZZ¹⁷ and PDB ID: 1D3Z⁹⁵ are illustrated in the Figure 68 and Figure 69 (Appendix I).

5.1.9.1 Structural comparison to the X-ray structure 3ONS

A high accuracy of 1.57 Å rmsd for the backbone (residues 1-70) was calculated by comparison to the X-ray structure. Site specific differences larger than 0.3 Å are only observed for flexible loop regions and the N-terminus of the protein.

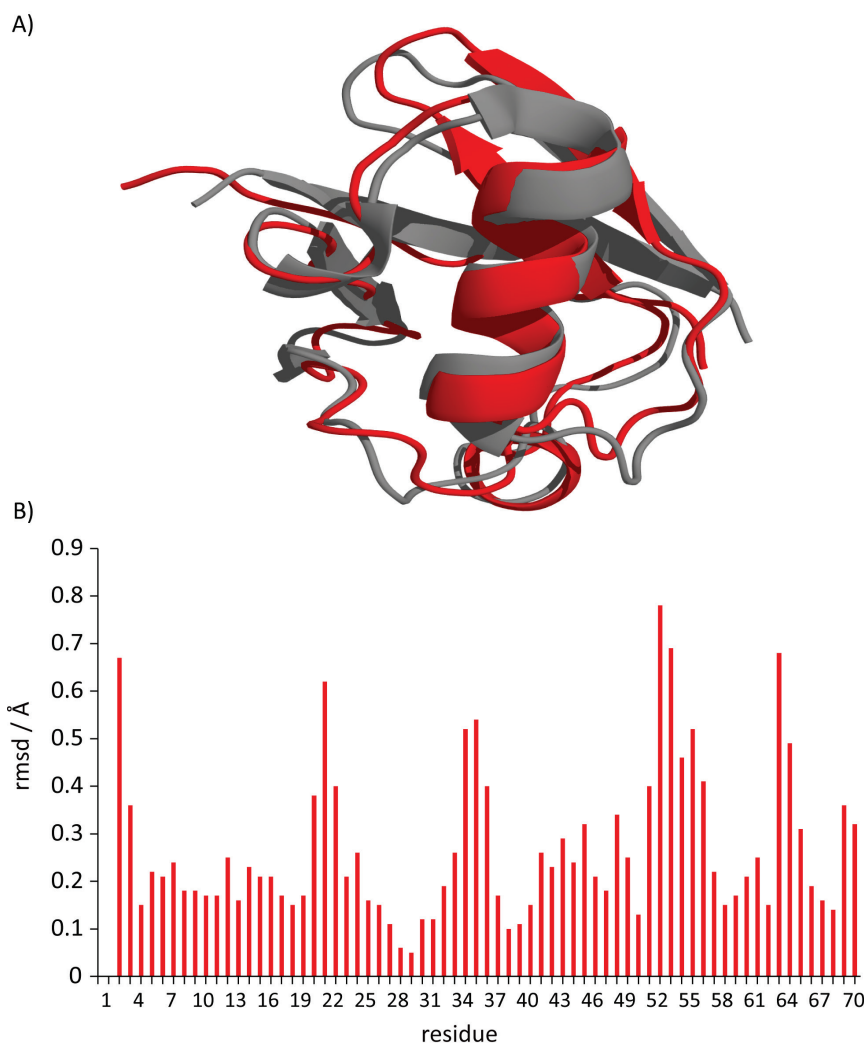


Figure 37 A) Comparison of the lowest energy calculated structure (red) to the MPD X-ray-structures 3ONS (global backbone rmsd value of 1.57 Å) B) Plot of local backbone rmsd values differences between the lowest energy calculated structure and the MPD X-ray-structures 3ONS.⁴⁶

Interestingly the structural heterogeneity observed in the solid-state involving the four β -strands and the first loop region (Figure 38.B), coincides with the regions in the X-ray structure where high B-factors are distinguishable (Figure 38.A).

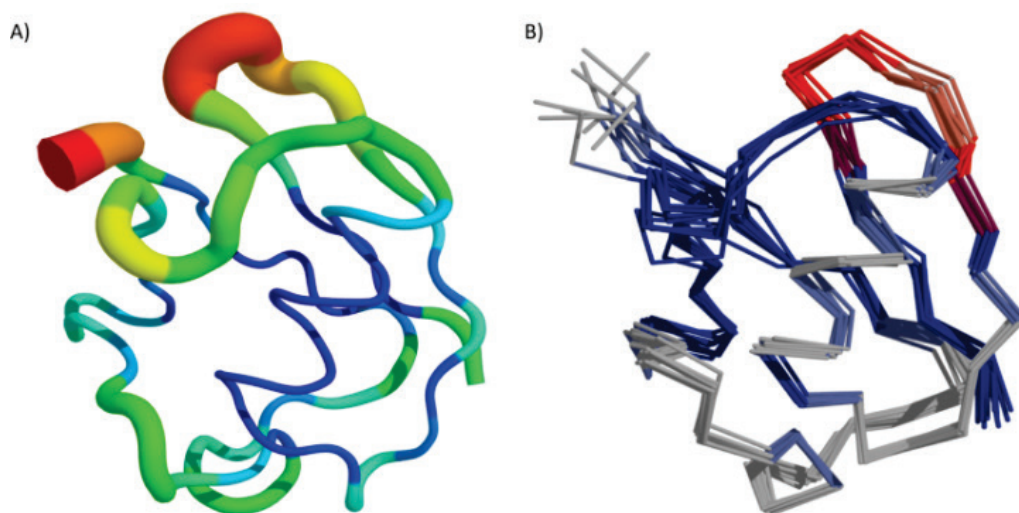


Figure 38 A) B-factors of the X-ray structure 3ONS⁴⁶ (value legend: blue (low-B-factor) - red (high-B-factor)), B) show the ten lowest energy calculated structure (in grey (residues with only one detectable conformation), in blue (residues with two conformations) and in red (residues with three observable conformations)).

5.1.10 Structure calculation of the second conformation

Due to the strong structural heterogeneity in the ubiquitin sample the detection of conformational selective distance restraints were accessible. In total 318 inter-residue distance restraints were used for the structure refinement of the second conformation. The 318 distance restraints can be separated into 55 unambiguous-, 143 network- and 120 ambiguous distance restraints. By removal of distance restraints which were assigned to both conformations one could collect 1 unambiguous-, 2 network- and 40 ambiguous distance restraints which belonged only to the second conformation. For the structure calculation an allowed variation of the observed correlations are set to 1-7 Å. For the ten lowest energy monomers a backbone rmsd of 1.4 Å was achieved (Figure 39).

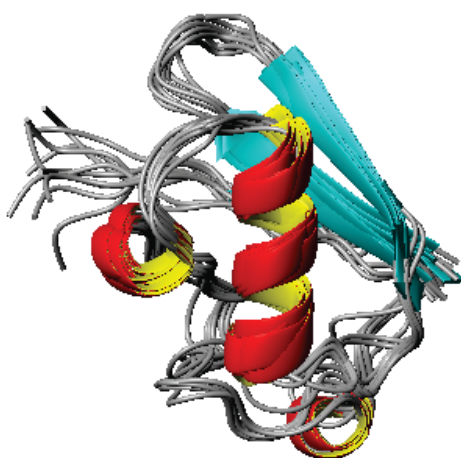


Figure 39 Illustration of the 10 lowest energy monomers of the second conformation.

Small variation of the local rmsd values could be distinguished for the second loop region (residue 17-21), the tip of the first alpha helix (residues (32-35)) and for the C-terminal part of ubiquitin (residue (61-70)) (Figure 40).

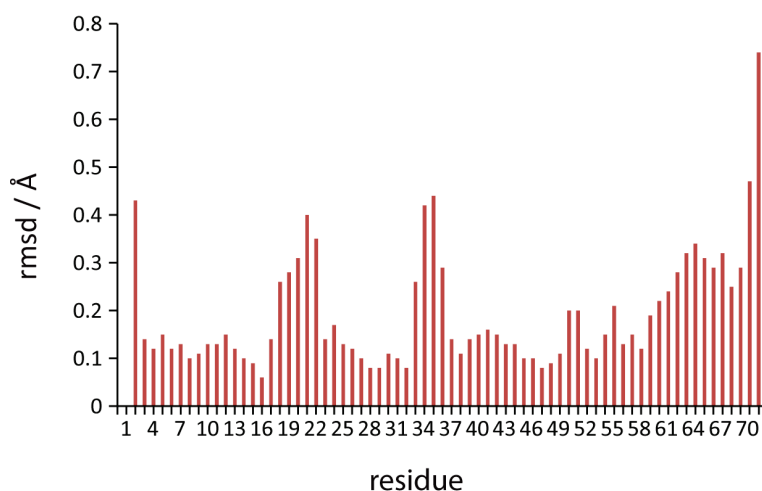


Figure 40 Plot of local backbone rmsd values of the 10 lowest energy structures of the second conformation (distance restraints range 1-7 Å).

5.1.11 Structural comparison between the main and the second conformation

The comparison between the two lowest energy structures from the main and the second conformation showed the same fold of ubiquitin. A global backbone rmsd value of 2.15 Å demonstrated no significant conformation change between the two structure models (Figure 41). Small local site specific differences are identified in the flexible loop region (residues K33-P37), in the region of the third and fourth β -strand (residues F45-L50) and at the C-terminal tail of ubiquitin (residues E64-V70) (Figure 42). Interestingly the regions with the strongest local differences showed large differences in chemical shifts as well. Only the area of the first loop was not affected, which could be explained by the fact that no long-range distance correlations were detectable for these residues.

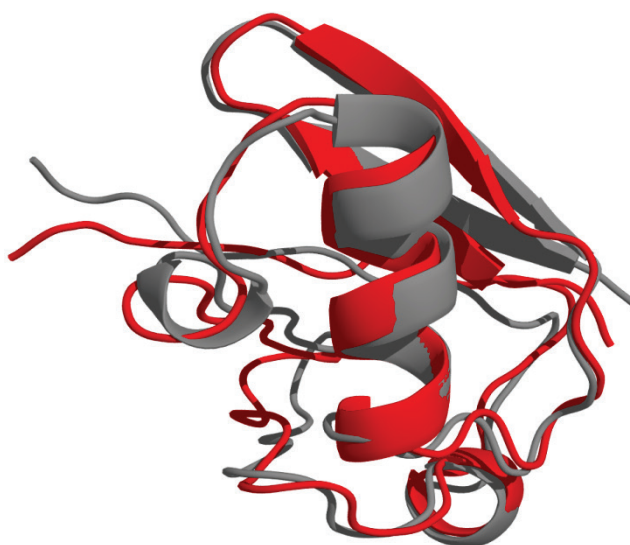


Figure 41 Comparison of the lowest energy calculated structures (main (red), second (grey)), the global backbone rmsd value is 2.15 Å).

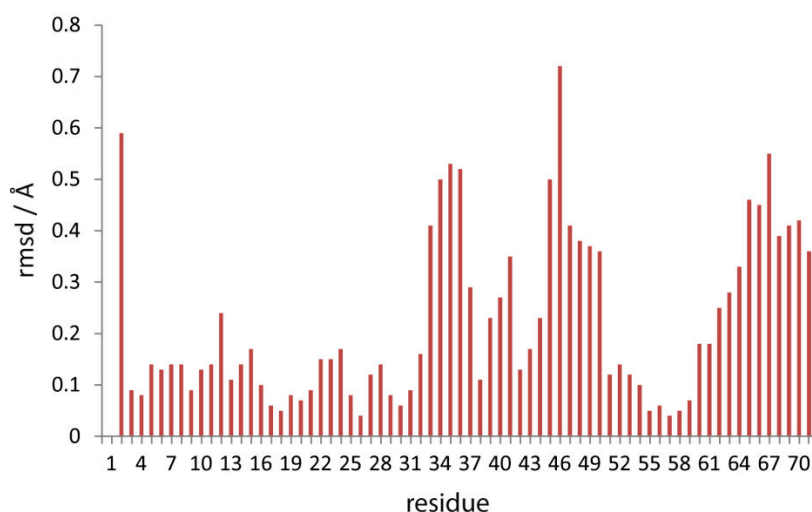


Figure 42 Plot of local backbone rmsd values differences between the two lowest energy calculated structures.

5.2 Results of ubiquitin PEG

5.2.1 Assignment of ubiquitin

Interestingly a structural heterogeneity of ubiquitin has been observed in the PEG sample as well. Polymorphism can be detected in the following regions (first and second β -strand β 1, β 2 (residues Q2-V5 and I13-L15), residue I23, at the tip of the α -helix α 1 (residues I30-G35), residue D39, the third β -strand β 3 (residues L43-G47), residue L50, residue I61, the fifth β -strand β 5 (residues S65-L67), and residue L69 (shown in Figure 43.B). The heteronuclear correlation of T9 is not visible in the NCA-spectrum (Figure 43.A), which points out that the loop region between the first and second β -strand β 1, β 2 seems to be even more flexible than after experiencing the MPD crystallization condition. By comparison of the (PEG) heterogeneity with the observation in the MPD sample one can easily see that the heterogeneity appears concentrated in the same highly dynamic regions of ubiquitin. This valuable result indicates that regions with the increased dynamics in the ns- μ s time scale cause structural heterogeneity during the crystallization process. A significant ^{15}N chemical shift change for A46 is shown in Figure 43.A,B. The intra-residual assignment of residue G53 (Figure 43.A) is ambiguous due to the fact that no sequential heteronuclear correlation is distinguished in the 2D NCO- or NCOCX-spectra.

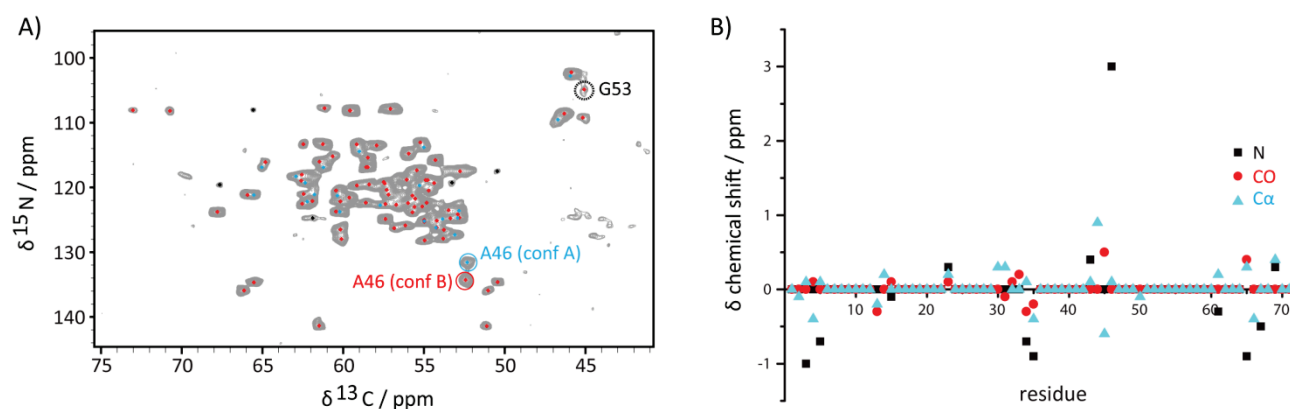


Figure 43 A) 2D NCA-spectrum of uniformly ^{13}C labeled sample, residues are colored as follows conformation A (blue) and conformation B (red), sequential correlations are colored in black, the ambiguous heteronuclear correlation of G53 is marked with a dashed circle B) illustration of the backbone atoms between conformation A and conformation B, due to the incomplete backbone assignment of residues M1, L8-K11, I44 and L71 the differences are set to zero.

In the 2D INEPT-spectrum only the cross-peak of the symmetric polyethylenglycol chain is detectable (Figure 44). This result is in agreement with the outcome of the MPD sample, in which the C-terminal region of ubiquitin is in the intermediate dynamic range.

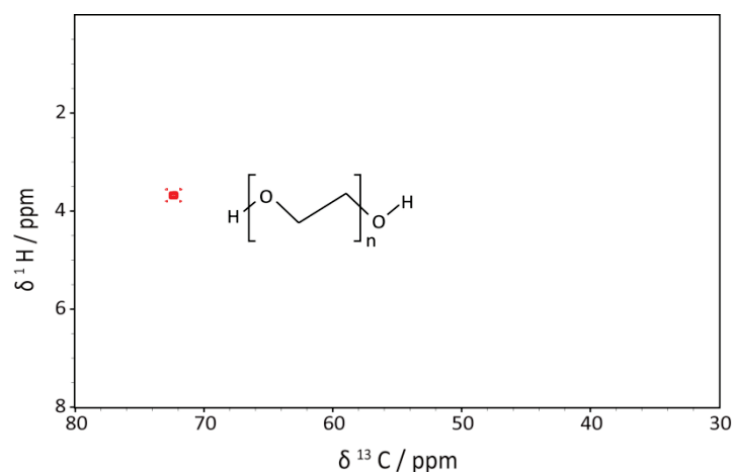


Figure 44 2D INEPT-spectrum of microcrystalline ubiquitin, showing that only the precipitating agent (PEG) is detectable.

5.2.2 Secondary structure analysis of conformation A

In the secondary chemical shift analysis the identification of β -strand conformation can be extracted for the following regions: M1-T7, I13-E18, Q40-F45 and L67-L71. Residue regions from I23-K33 and L56-D58 show α -helical conformation (Figure 45).

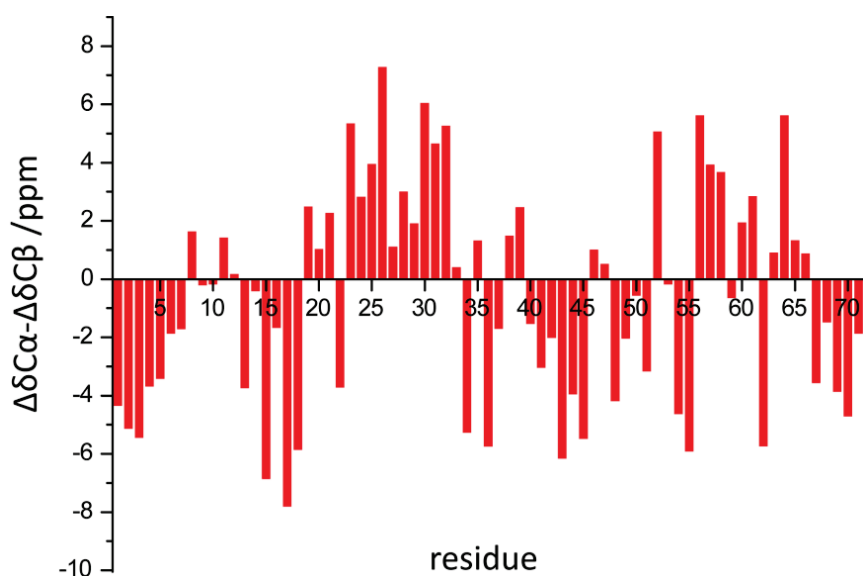


Figure 45 Secondary chemical shift analysis of conformation A of microcrystalline ubiquitin.

The TALOS+ prediction reveals β -strand conformation for the regions: Q2-K6, K11-V17, R42-A46, Q49 and T66-V70. For the regions I23-K33 and T55-Y59 α -helical conformation is predicted (Figure 46.A,B). Both results are consistent with the data given by the ubiquitin MPD sample.

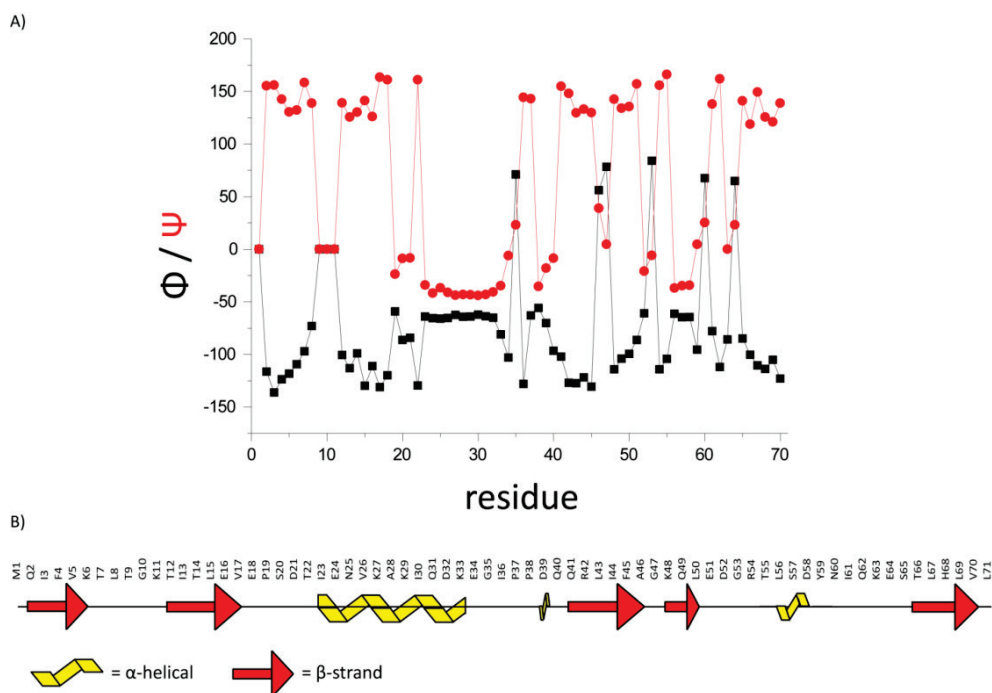


Figure 46 A) TALOS+ analysis, predicting Phi (black) and Psi (red) dihedral angles of conformation A of ubiquitin (PEG) B) cartoon of the result of TALOS+ prediction.

5.2.3 Collection of distance restraints information

As the resolution of the PEG sample is slightly decreased in comparison to the MPD preparation, only a set of 24 unambiguous-, 48 network- and 425 ambiguous distance restraints could be identified. The distance distribution of the 24 unambiguous distance restraints (Figure 47.A) back calculated by the X-ray structure 1UBQ⁴⁹ are identical with the distance distribution achieved using the MPD sample (Figure 31.A). For the network distance restraints a ¹³C-¹³C correlation between I3Cγ2-V26Cα with a back calculated distance of 9.69 Å can be detected (Figure 47.B). The contact is depicted on the X-ray structure 1UBQ in Figure 70 (Appendix I). The detection of this correlation in the PDS spectra can be explained by a possible rotation of the I3 side chain into another rotameric conformation by slight conformational conversion in the first β-strand in the ssNMR structure model from PEG sample.

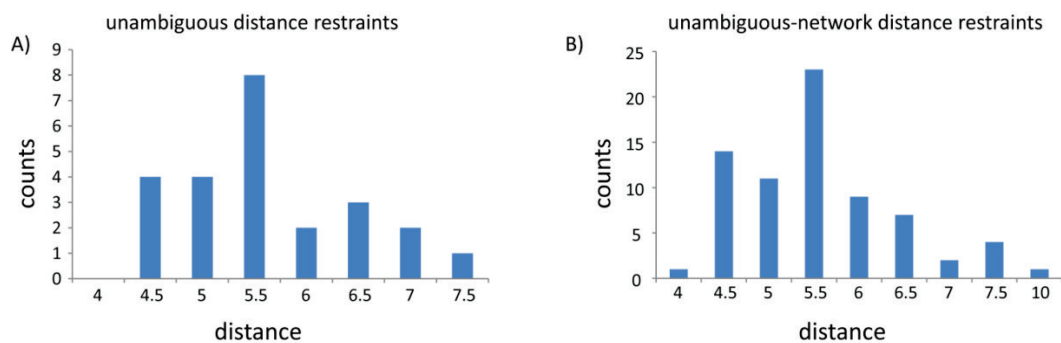


Figure 47 Distance distributions collected from PDS spectra, the corresponding distance as extracted from the X-ray structure is shown A) 24 unambiguous distance restraints, B) 73 unambiguous- and network distance restraints.

5.2.4 Structure calculation

The same structure refinement procedure, as already described in section 5.1.8, has been used for obtaining the different calculated structures for the PEG sample. The allowed variation of the observed contacts is set to 1-7 Å. Interestingly the general 3D-fold of Ubiquitin can already be achieved by the use of only 24 unambiguous distance restraints. The C-terminal part of Ubiquitin is aligned in a wrong direction although 3 unambiguous long-range distance correlations between the residues I3C δ 1-S65C β , I3C δ 1-L67C γ and F4C γ -T66C α were included in the distance restraint input file (Figure 48.B). The implementation of all collected distance restraints in the calculation gives the correct 3D fold of ubiquitin with a backbone rmsd value of 0.7 Å (Figure 48.D). The result of the structure validation is listed in Table 14 (Appendix I). The local rmsd values of the ten lowest energy structures of conformation A show significant differences in the first loop region (residues T9-T12), at the beginning of the α -helix α 1 (residues D21-T22), in the flexible loop region between β -strand β 4 and β -strand β 5, and at the C-terminal part of Ubiquitin (residues (68-71))(Figure 49).

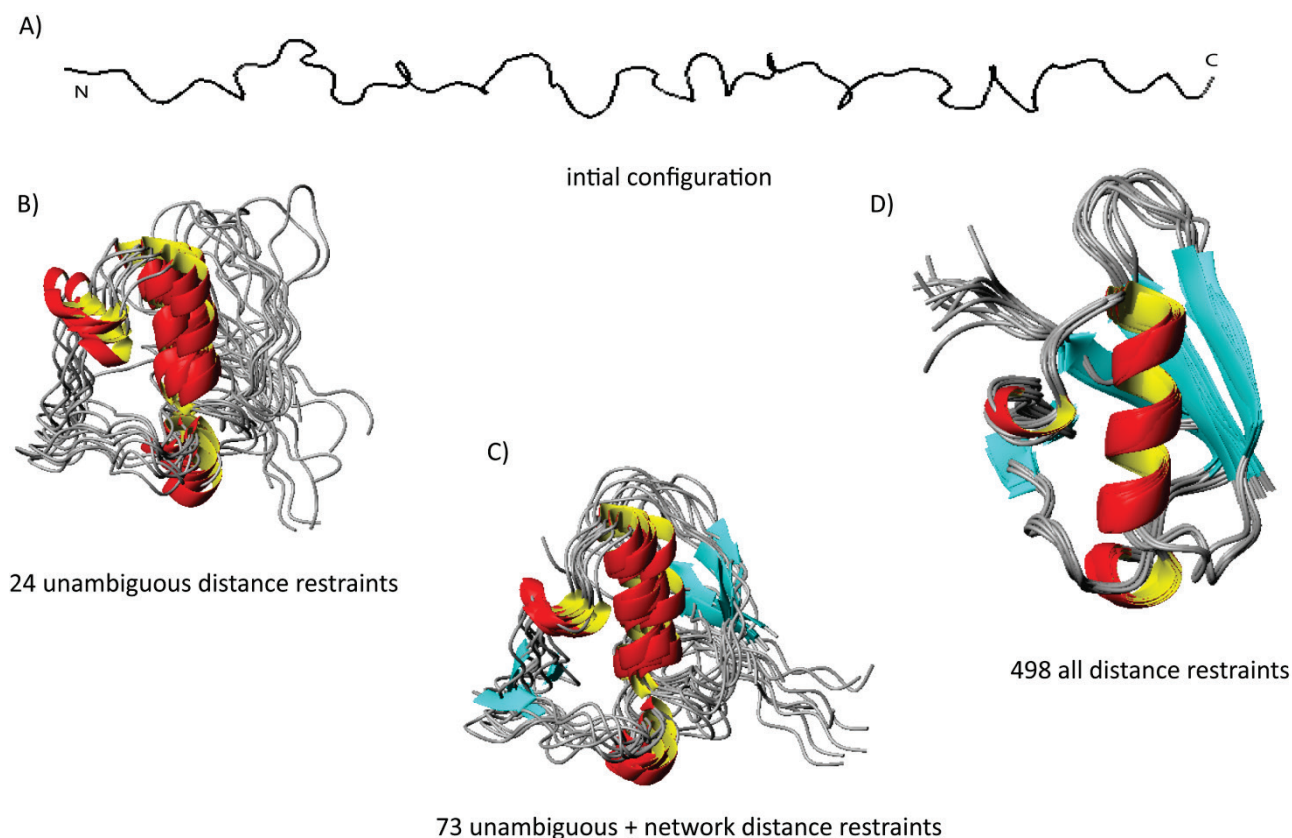


Figure 48 Representation of the various steps in *de novo* protein calculation. A) illustration of the extended configuration (random coil) used as starting point of the structure calculation of ubiquitin. Figures (B-D) show the ten lowest energy calculated structures for each distance restraints input. The 10 selected monomers were aligned on the backbone atoms using MOLMOL B) With the use of 24 unambiguous distance restraints, the bundle of monomers reveal already the correct 3D fold of Ubiquitin C) representation of 24 unambiguous and 49 networking distance restraints D) result of all 498 collected distance restraints (24 unambiguous, 49 network and 425 ambiguous distance restraints).

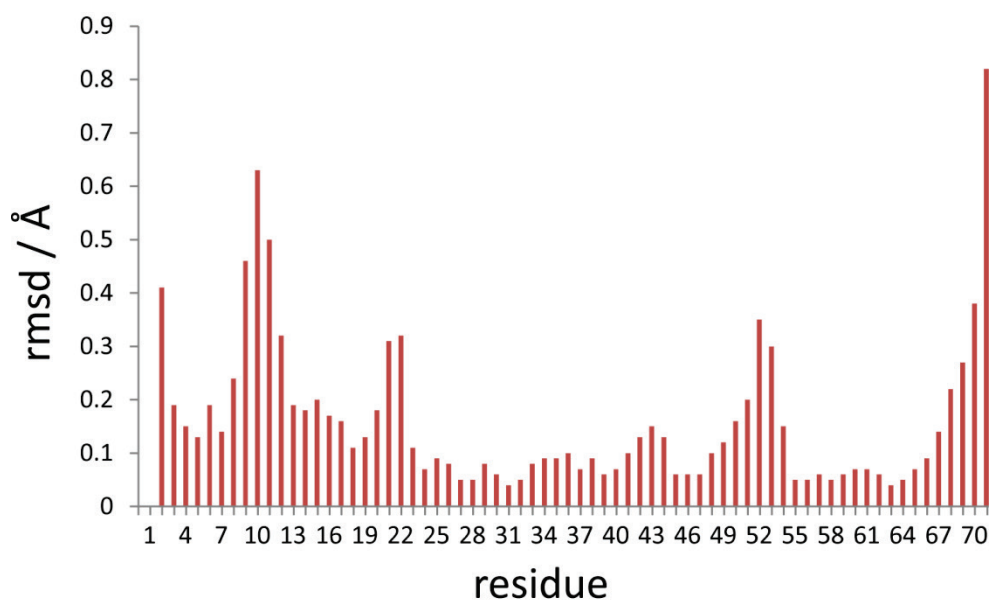


Figure 49 Plot of local backbone rmsd values of the 10 lowest energy structures of conformation A (distance restraints range 1-7 Å).

5.2.5 Comparison to different accessible PDB structures

For the comparison of different structures only the structure calculated with the lowest energy is taken (range 1-7 Å). The comparison to the solution NMR structures PDB ID: 1D3Z is illustrated in Figure 73 (Appendix I).

5.2.5.1 Structural comparison to the X-ray structure 1UBQ

An accuracy of 1.88 Å rmsd for the backbone (residues 1-70) was calculated by comparison to the X-ray structure (Figure 50.A). Major site specific differences can be distinguished for the flexible loop regions (residues (L8-K11), residues (21-22), residue (52-52) and residues (63-65)) (Figure 50.B).

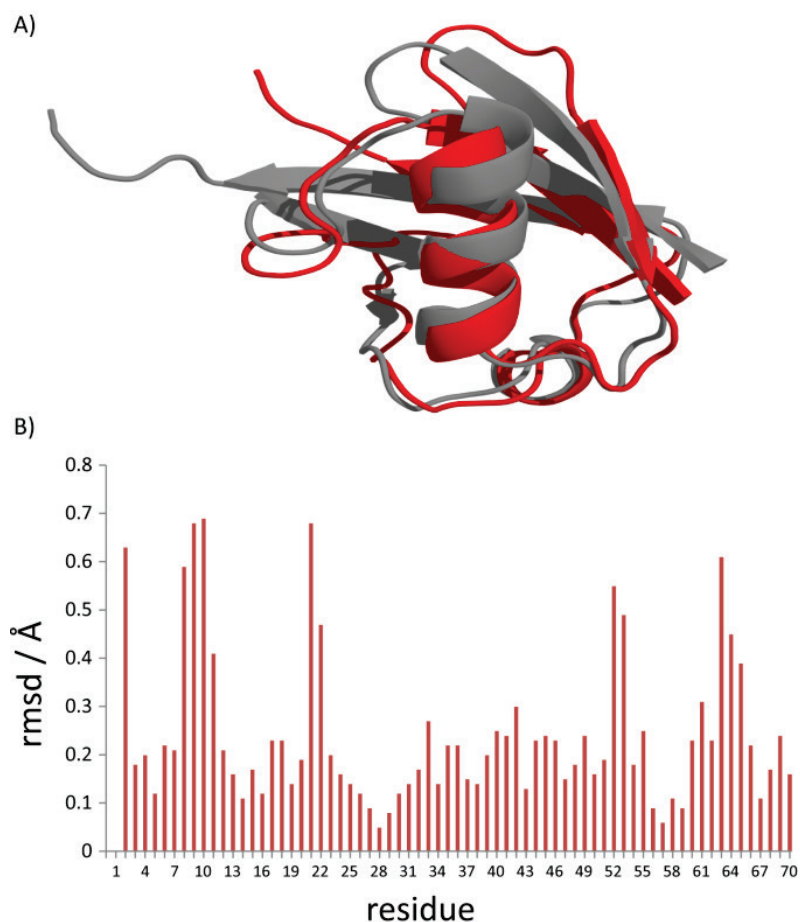


Figure 50 A) Comparison of the lowest energy calculated structure (red) to the X-ray structure (grey) (1UBQ) B) Plot of local backbone rmsd values differences between the lowest energy calculated structure and the X-ray-structures 1UBQ.

The structural heterogeneity observed in the solid-state coincides with the regions in the X-ray structure where high B-factors are localized (Figure 51.A). Interestingly the first loop region (residues (L8-K11)) in the X-ray structure 1UBQ shows a higher rigidity than the MPD structure 3ONS although in the ssNMR it is exactly contrary. As perviously mentioned the heteronuclear correlation of T9 is not visible in the NCA-spectrum (Figure 43.A), which integrates a higher flexibility in the PEG sample.

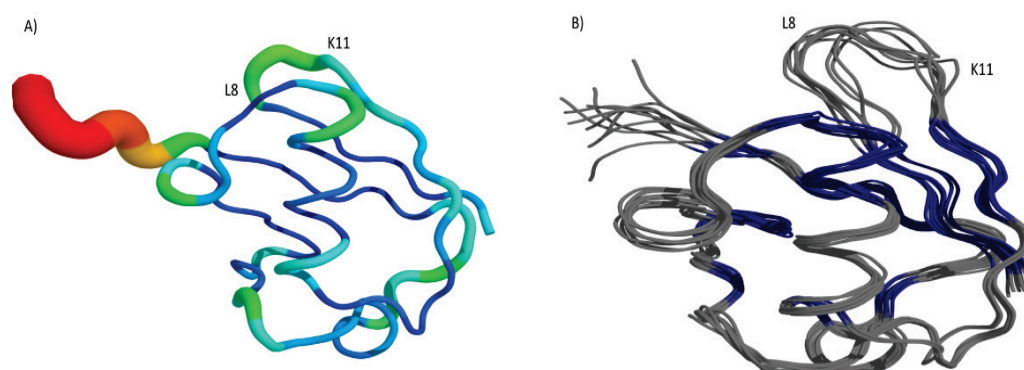


Figure 51 A) B-factors of the X-ray structure 1UBQ (value legend: blue (low-B-factor) - red (high-B-factor)), B) show the ten lowest energy calculated structure (in grey (residues with one detectable conformation) and in blue (residues with two observable conformations)).

5.2.6 Structure calculation of conformation B

The result of the secondary structure analysis and the TALOS+ prediction is depicted in Figure 71 and Figure 72 (Appendix I). The illustration of the 10 lowest energy structures is demonstrated in Figure 52. The backbone rmsd value for the ten lowest energy structures is 0.7 Å. Due to the lower resolution in the PEG sample the differentiation of ^{13}C - ^{13}C distance correlation between the two conformations is highly challenging. By removal of distance restraints which were assigned to both conformations one can only collect one unambiguous-, one network- and 11 ambiguous for conformation B. The local rmsd values result (Figure 53) is comparable to the resulting rmsd values for conformation A (Figure 49). The only difference concerns the region of residue D52, which delivers lower local rmsd values as for conformation A.

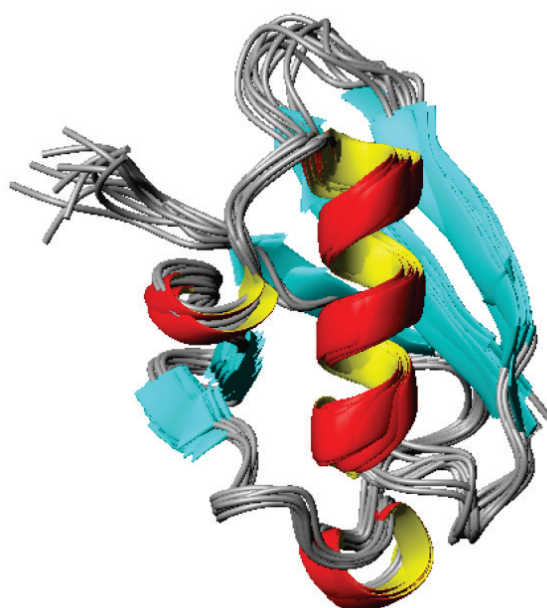


Figure 52 Illustration of the 10 lowest energy monomers of conformation B with a backbone rmsd value of 0.7 Å.

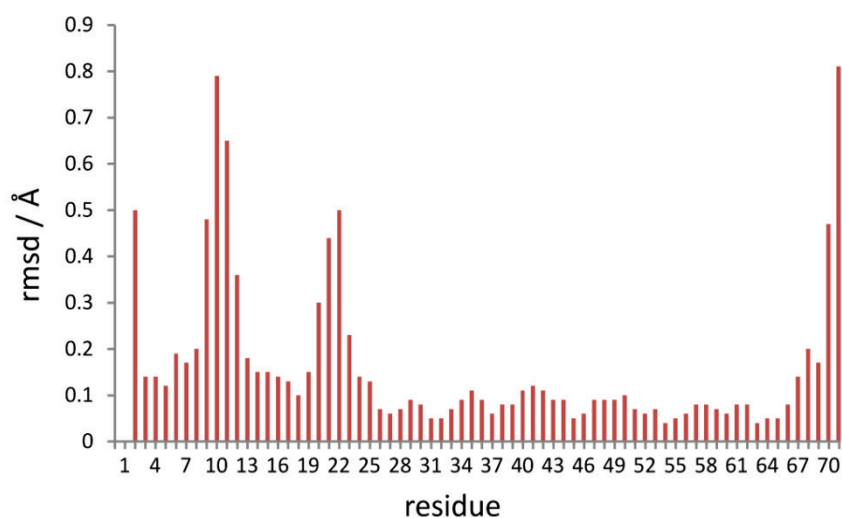


Figure 53 Plot of local backbone rmsd values of the 10 lowest energy structures of conformation B (distance restraints range 1-7 Å).

5.2.7 Structural comparison between conformation A and B

A backbone rmsd of 1.08 Å for the lowest energy structures of the two different conformations can be measured (Figure 54). The comparison between the two conformations shows slight differences in the flexible loop regions (residues (T7-K11), (P19-T22), (L50-D52)) and for the C-terminal part of ubiquitin although the number of distance restraints between the two conformations is comparable (conformation A (498) and conformation B (487)) (Figure 55).

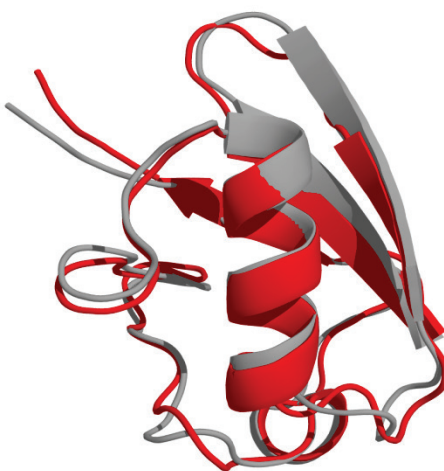


Figure 54 Comparison of the lowest energy calculated structures (conformation A (red) and conformation B (grey)), the global backbone rmsd value is 1.08 Å.

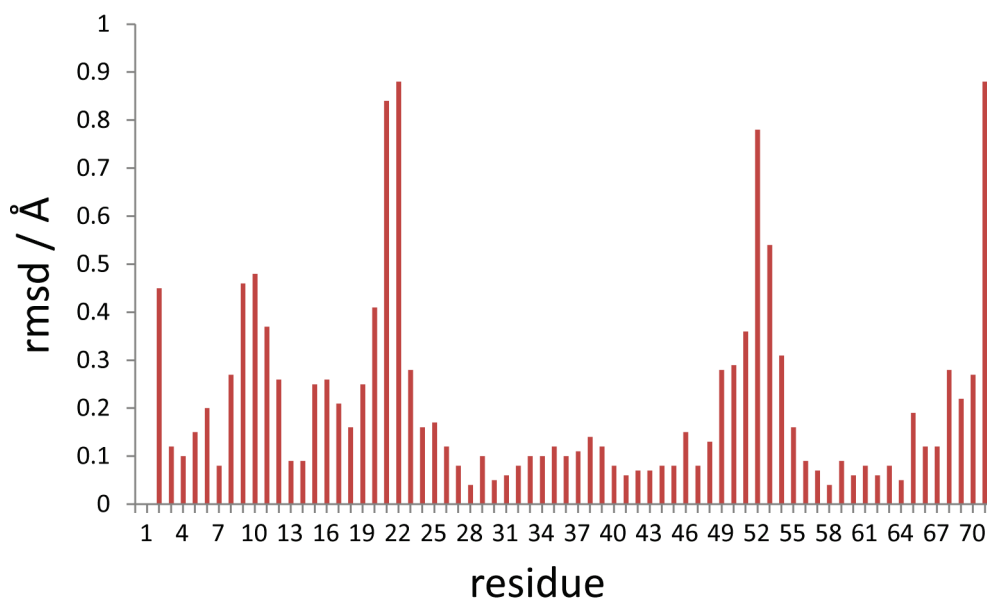


Figure 55 Plot of local backbone rmsd values differences between the two lowest energy calculated structures.

5.3 Comparison between MPD and PEG samples

The comparison between the two differently prepared samples is carried out on the main conformation of MPD and the conformation A of PEG.

By comparison of the backbone chemical shifts (^{15}N , ^{13}CO , $^{13}\text{C}\alpha$) of the two different samples significant chemical shift differences can be seen for ^{15}N of the residues V70 (14.4ppm) and L71 (6.9ppm) (Figure 56). These strong chemical shift changes indicate a conformational change of the C-terminal part of ubiquitin. Unfortunately both residues do not show long-range distance correlations to other residues in the protein, which would confirm the conformational change. The only reasonable explanation can be given, since both residues belong to the highly flexible C-terminus ubiquitin, which can be crystallized in different conformations. In Figure 56 one can notice that in general the ^{15}N nuclei show more chemical shift deviations than the ^{13}C atoms of the backbone, such as CO or $\text{C}\alpha$. Chemical shift differences bigger than 2ppm for ^{15}N can be determined for the residues (K6-L8), K29, K33, (R42-46), L50 and for the residues (D52-G53). This result is in strong agreement with the regions where structural heterogeneity can be observed in both samples. It seems that protein motions in the ns- μs time scale cause on the one hand structural heterogeneity during the crystallization process and on the other hand they are sensitive to the environment (crystallization condition).

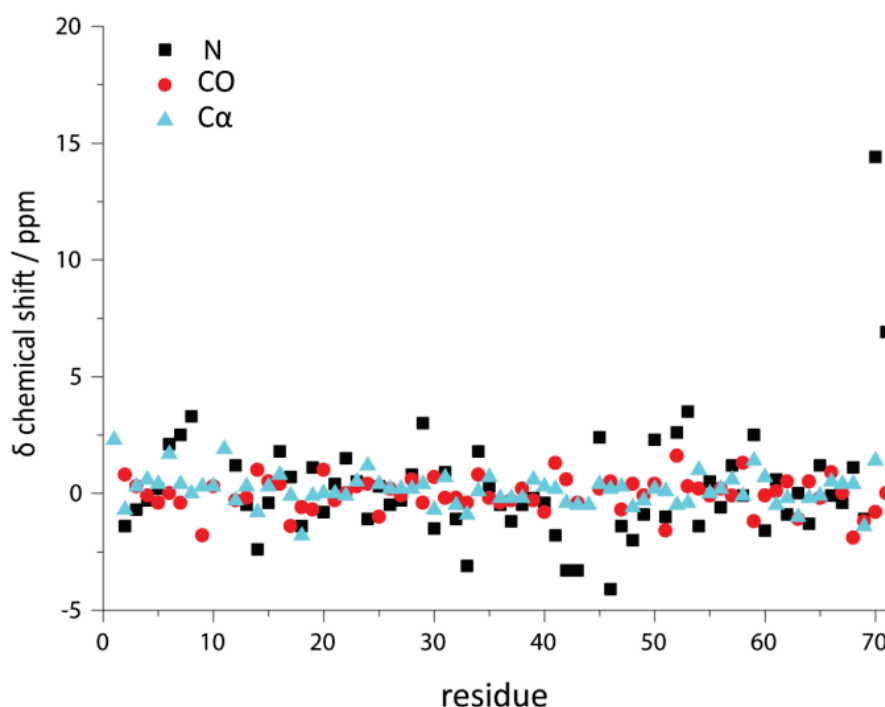


Figure 56 Illustration of chemical differences of the backbone atoms between the main conformation of the MPD sample and conformation A of the PEG sample.

Considering the differences in the ^{13}C - ^{13}C correlation 2D PDSD-spectra of the two mentioned samples, small chemical shift changes for the side chains can be identified all over the protein (Figure 57). For the residue L43 (highlighted in Figure 57) a strong 1.93ppm chemical shift change for $\text{C}\beta$ is detectable.

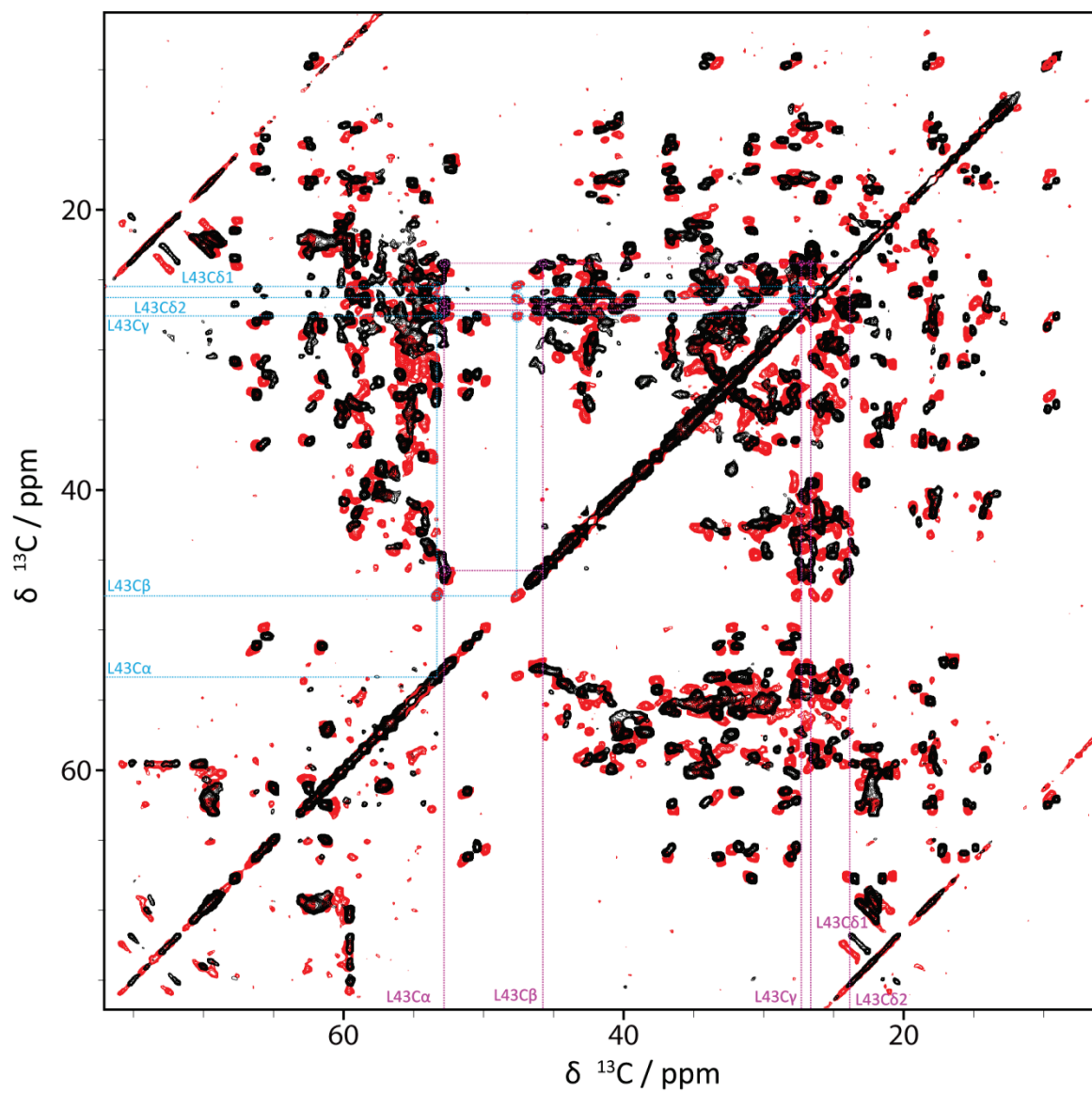


Figure 57 Comparison of 2D PDS-spectra of MPD sample (red) and PEG (black) with a mixing time of 50ms, the spin system of L43 is highlighted in the spectra (magenta (PEG) and light-blue (MPD)).

5.3.1 Structural comparison between the main conformation of MPD and the conformation A of PEG

By comparison between the lowest energy structures of the two systems, one can infer that the 3D fold of ubiquitin is identical (Figure 58.A). An accuracy of 1.63 Å for the backbone could be calculated (residue (M1-70V)). Small conformational deviations were observed in the loop regions (L8-T12, D21, 34-36 and for E51-G53)) (Figure 58.B).

A)



B)

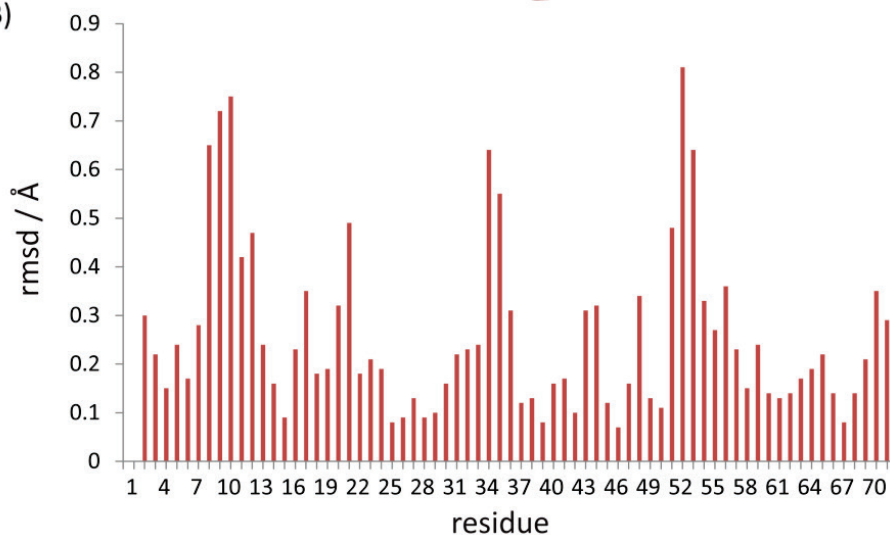


Figure 58 A) Comparison of the lowest energy calculated structures (main conformation of MPD (red) conformation A of PEG (grey), with a backbone rmsd of 1.63 Å, B) Plot of local backbone rmsd values differences between the two lowest energy calculated structures.

6 Conclusion

6.1 Conclusion MPD

The approach of the alternative [1-¹³C] - and [2-¹³C]-glc sparsely labeling strategy is demonstrated on the microcrystalline form of ubiquitin in MPD as a model protein. The improvement in spectral resolution and the reduction of spectral crowding allows us to perform a complete backbone assignment using solid-state-NMR in a straightforward manner. Most residues from M1 to V70 of microcrystalline ubiquitin were assigned accordingly. Another advantage of this labeling scheme is related to the decrease of dipolar truncation, which facilitates the collecting of inter-residual long-range distance information by running of 2D PDS experiments with long mixing times between 400-900ms. The identification of a structural polymorphism with up to 3 different conformations for the β -strands β 1, β 2, β 3 and β 5 as well as for the loop regions β 1- β 2 and α 1- β 3 coincide with regions showing increased dynamics on the ns- μ s time scale as previously revealed by solution NMR ¹⁵N relaxation and RDC-based experiments. This observation suggests that the conformational dynamics in solution manifests itself as a structural polymorphism during crystallization. Independently for each conformation, one highly precise solid-state NMR structure-model is calculated with a backbone rmsd of 0.7 Å for the main conformation and 1.4 Å for the second conformation (residue range M1-R72). The comparison of the two calculated structures ends up with the same 3D fold of ubiquitin (backbone rmsd of 2.15 Å for the residue range M1-V70). Small local site-specific differences were observed in highly dynamic loop regions and in the C-terminal part of ubiquitin. Notably the best accuracy (backbone rmsd of 1.57 Å) of the calculated structure to the X-ray structure 3ONS was obtained by allowing a distance restraint range of 1-7 Å for the extracted ¹³C-¹³C correlations. The upper limit distance range of 7 Å is in a high agreement with the distance distribution of unambiguous distance restraints as back-calculated by the X-ray structure 3ONS.

6.2 Conclusion PEG

A structural heterogeneity of microcrystalline ubiquitin precipitated in PEG is detected similar to MPD precipitation procedure. The locations of the polymorphism observed in similar regions as pointed out for the MPD ubiquitin sample. The assignment of the loop regions β 1- β 2 is incomplete due to undetectable intra-residual and inter-residual correlations. Independently, for each conformation, one high precision solid-state NMR structure is calculated with a backbone rmsd of 0.7 Å. The overlay of these two gained structures gives a calculation accuracy of 1.08 Å in the backbone (residue range M1-V70). Site specific differences between the two structures can be found for the loop regions and for the C-terminal part of ubiquitin. The comparison between conformation A and X-ray structure 1UBQ gives a globular backbone rmsd value of 1.88 Å.

By comparison of the backbone atoms between the MPD- and the PEG assignment, chemical shifts differences can be identified all over the protein atoms/residues. Strong ^{15}N chemical shift differences (>2.5ppm) are located in the same region where structural heterogeneity could be observed. These results indicate that protein motions in the ns- μ s time scale cause structural heterogeneity during the crystallization process.

PART II MxiH

Notes:

The assignment of MxiH was carried out by Dr. Jean-Philipp Demers

The proton detected ssNMR experiments were conducted by Dr. Veniamin Chevelkov

7 Introduction

Solid-state NMR (ssNMR) spectroscopy is a powerful technique, which provides site-specific information at the molecular level for complex biological systems. It is amenable to investigate challenging samples: high molecular weight, insoluble and non-crystalline systems. Biomaterials which have been studied by ssNMR include amyloid fibrils²⁹⁻³², membrane proteins³³⁻³⁵, virus capsids^{36,37}, bacterial filamentous proteins^{38,39}, fibrillar proteins such as collagen^{40,41}, keratin^{42,43} and silk^{44,45} and bone and teeth mineral materials.⁹⁶ SsNMR studies of large systems are technically challenging in part due to the overcrowding of spectra, in addition to the intrinsic low sensitivity and broad line-width of resonances in the solid state.

New techniques have been recently introduced to improve the resolution and sensitivity of Magic-Angle Spinning (MAS) solid-state NMR spectra, including the use of high external magnetic fields, pulse sequence implementation for observation of long-range distance correlations, increase in the magic-angle spinning rates, proton detection and the introduction of isotopic labeling strategies. Major steps in structural studies of biomolecules by solid-state NMR are the assignment of chemical shifts for various nuclei, the obtaining of structural restraints such as long-range distance restraints, and the study of local dynamics. Here we present a strategy for the isotopic ¹³C labeling of leucine, valine and isoleucine methyl groups and demonstrate how this labeling scheme facilitates the collection of distance restraints, the verification of carbon chemical shift assignments and the measurement of methyl group dynamics. We combined this labeling scheme with ultra-fast MAS and proton detection in order to obtain the chemical shift assignment of methyl protons.

In the past, different approaches have been presented to obtain structural information on uniformly labeled samples. One such approach is the development of new pulse sequences to collect long-range distance restraints: CHHC and NHHC for proton-proton distances^{10,97}, PDS^{7,98}, DARR^{99,100}, REDOR¹⁰¹ and PAR¹⁰² for carbon-carbon distances; TEDOR¹⁰³ and PAIN¹⁰⁴ for heteronuclear correlations such as ¹³C-¹⁵N. Solid-state-NMR spectra recorded on uniformly labeled samples contain a large number of carbon resonances which have broad line-widths mainly due to homonuclear couplings. Both factors contribute to the low resolution and high ambiguity of cross-peaks. Uniformly labeled samples present further disadvantages for the collection of long-range restraints: as the initial signal is shared by couplings among a large number of ¹³C labeled nuclei, the intensity of cross-peaks is reduced; as well, the presence of strong dipolar truncation in uniform labeled samples suppresses the transfer of magnetization, hindering the detection of long-range distances.

Former solutions to the dipolar truncation problem exploit sparsely ¹³C labeling schemes, in which different ¹³C labeled precursors act as the only ¹³C source in the growth medium during bacterial (over)expression of the protein. Sparsely ¹³C labeled protein samples have two important advantages: first, the spectra are simplified since a smaller number of ¹³C atoms are labeled in the protein; second, the resolution of ¹³C peaks is increased due to the removal of one-bond dipolar and scalar ¹³C-¹³C couplings.

Both result in a strong reduction of cross-peak overlapping in ssNMR spectra, leading to a simpler determination of unambiguous long-range distances restraints.

Examples of labeling schemes used in ssNMR comprise [1,3-¹³C]-glyc, [2-¹³C]-glyc¹⁹⁻²², [1-¹³C]-glc [2-¹³C]-glc^{23,24} and various ¹³C labeled pyruvates.⁷¹ Long-range distance restraints are obtained in combination with sparse labeling by recording ssNMR experiments utilized for the observation of long-range distances as previously mentioned, or also with ¹³C-¹³C first-order dipolar recoupling sequences such as RFDR^{62,105} as demonstrated using [2-¹³C]-glc labeling.¹⁰⁶ Significant improvements in the resolution of ¹³C-¹³C ssNMR spectra have been demonstrated by using [1-¹³C]-glc and [2-¹³C]-glc sparsely labeled proteins.⁸⁵ With these labeling schemes, a strategy was presented to determine intermolecular interactions of proteins in a self-assembled molecular system.¹⁰⁷ In the same way the [2-¹³C]-glc labeling allows the easy determination of stereospecific assignment for the amino acids valine and leucine and can be used for the acquisition of long-range distance restraints in 3D spectra.^{26,27}

Previous studies exhibit the prevalent application of sparsely labeled methyl groups to gain structural, functional and dynamics information about proteins.^{52,108-110} In addition unambiguous long-range correlations between methyl groups are important as methyl groups tend to be found deep in the hydrophobic core of proteins. However, a drawback of the glycerol and glucose labeling schemes lies in the high spectral crowding of the methyl ¹³C region. Many amino acids containing methyl groups — threonine, alanine, valine, isoleucine, leucine and methionine — often have theirs concurrently labeled in [1,3-¹³C]-glyc and [1-¹³C]-glc labeling schemes, maintaining a strong ambiguity of cross-peaks in ¹³C-¹³C ssNMR spectra. Our approach proposes to avoid this problem by utilizing the biosynthetic precursors α -keto-isovalerate¹¹¹ and α -keto-butyrate. Considering the biosynthetic metabolism of amino acids (Figure 59), one obtains only ¹³C labeled methyl groups on leucine and valine when using labeled α -keto-isovalerate. The C δ 1 methyl group of isoleucine can be ¹³C labeled by means of labeled α -keto-butyrate. This leads to an improvement in the information content obtained from the methyl-methyl region of ¹³C-¹³C 2D spectra as the ambiguity of methyl-methyl cross-peaks is largely reduced. This benefit facilitates collecting of unambiguous distance restraints between carbons of methyl groups that are close in space. Here, we demonstrate the recording of long-range restraints in the Type-Three Secretion System (T3SS) of *Shigella flexeneri*.²⁵ We also analyze the intensity of intra-residue cross-peaks and the relaxation properties of methyl carbons. Employing ultra-fast MAS rates and sparse ¹³C labeling of methyl groups facilitate high-resolution proton-detected ¹³C-¹H correlations yielding the assignment of methyl protons.

8 Material

8.1 Sample preparation

MxiH proteins were expressed recombinantly in *E. coli* strain BL21 (DE3) in minimal medium during 5 hours. For the production of the [Leu-Val ^{13}C Methyl] ([LV- ^{13}C Methyl])-labeled sample, α -ketoisovalerate Na [2-keto-3-(methyl- ^{13}C)-butyrate-4- ^{13}C] (125mg/L) was added as carbon source one hour prior to induction. For the [U- ^{15}N -[Ile-Leu-Val- ^{13}C Methyl] ([U- ^{15}N -[ILV- ^{13}C Methyl])-labeled sample, $^{15}\text{NH}_4\text{Cl}$ was used as nitrogen source. One hour before induction, α -ketoisovalerate Na [2-keto-3-(methyl- ^{13}C)-butyrate-4- ^{13}C] (125mg/L) and α -ketobutyrate Na [2-keto-butyrate-4- ^{13}C] (125mg/L) were added as carbon source. Both proteins were purified by following the protocol established for *Salmonella typhimurium* PrgI needles.²⁴ The N-terminal hepta-Histidine (His) tag was cleaved using tobacco etch virus protease, releasing MxiH proteins containing the non-native N-terminal residues glycine and histidine. The protein concentration was raised to 0.2mM during polymerization, which took place at 37°C for 16 days and approximately 20mg of needle proteins were produced.

9 Methods

9.1 General aspects

MxiH needles were ultra-centrifuged and transferred into a 4.0-mm and a 1.3-mm ZrO₂ ssNMR rotor. Solid-state NMR experiments for the [LV-¹³C Methyl]- and the [U-¹⁵N-[ILV-¹³C Methyl]]-labeled sample were recorded on an 18.8-Tesla spectrometer (Bruker Biospin, AVANCE III Germany) operating at 800 MHz ¹H Larmor frequency. Carbon-detected experiments were conducted at 11kHz MAS frequency (4.0-mm rotor) and a temperature of 7.5°C, while for proton-detected experiments sample was spun at 60kHz MAS frequency (1.3-mm rotor) and had an effective sample temperature of 35°C. Proton-detected experiments were conducted on only [U-¹⁵N-[ILV-¹³C Methyl]]-labeled sample. The temperature was calibrated from the ¹H chemical shift of water in reference to DSS.⁸²

All spectra were processed using Bruker Topspin 3.1 software. For 2D and 3D spectra, the free induction decay (FID) signals were apodized prior to Fourier transform, using sine squared window functions: the sine bell shift was 45° for all carbon-detected spectra, respectively 43° and 47° for the ¹H and ¹³C dimensions of the 2D (H)CH-spectrum (Figure 65.A) and 47° for all dimensions of the 3D (H)CCH-spectrum (Figure 65). The chemical shift assignment of spectra was carried out using the Sparky software 3.114.⁸³

9.2 Carbon-detected ssNMR

The 1D CP MAS spectra (Figure 60.A,B Figure 62.A) were acquired with the following parameters: maximal acquisition time (AQ)=50ms, recycling delay (RD) of 5s. No window function was applied in the processing. The build-up of ¹⁵N magnetization transferred during ¹⁵N-¹³C SPECIFIC-CP^{8,9} was determined in a series of 1D experiments (Figure 64.B) where the contact time of the ¹⁵N-¹³C CP was varied from 5ms up to 15ms in steps of 2.5ms. The experiments were recorded using the following parameters: AQ=10ms, RD of 2.5s and 5120 number of scans (NS). The nutation frequencies employed during the ¹⁵N-¹³C CP transfer were 8.8kHz rf on the ¹³C channel, 4.5kHz rf on the ¹⁵N channel and 58kHz rf on the ¹H channel.

For the [LV-¹³C Methyl]-labeled sample, two 2D PDS-spectra were acquired (Figure 60.B, Figure 63.A), with PDS mixing times of 50ms and 800ms, using the following parameters: AQ₁=12ms (indirect dimension), AQ₂=15ms (direct dimension), spectral window SW₁=22ppm (indirect dimension), SW₂=346ppm (direct dimension), RD of 3s, CP contact time of 0.8ms, NS=256 (for PDS 50ms) or NS=1856 (for PDS 800ms), for a total experimental times of 23h (for PDS 50ms) and 6days 19h (for PDS 800ms). For the [U-¹⁵N-[ILV-¹³C Methyl]]-labeled sample, a PDS-spectrum with mixing time of 800ms was recorded (Figure 63.A) using the following parameters: AQ₁=12ms (indirect dimension), AQ₂=15ms (direct dimension), spectral window SW₁=80ppm (indirect dimension), SW₂=345ppm (direct dimension), RD of 3s, CP contact time of 0.5ms, NS=320, for a total experimental time of 5days 11h. The spectrum of the [U-¹³C]-glc labeled sample (Figure 61) was recorded with a PDS mixing time of 50ms, using maximum acquisition times of

19ms (direct dimension) and 15ms (indirect dimension). The total experimental time was 1day 17h. The spectrum of the [2-¹³C]-glc labeled sample (Figure 61) was recorded with a PDSM mixing time of 50ms, using maximum acquisition times of 21ms (direct dimension) and 15ms (indirect dimension). The total experimental time was 1day 7h.

The NCX-spectrum (Figure 64.A) was recorded on the [U-¹⁵N-[ILV-¹³C Methyl]]-labeled sample with AQ₁=10ms (indirect dimension), AQ₂=10ms (direct dimension), spectral window SW₁=39.8ppm (indirect dimension), SW₂=282ppm (direct dimension), RD of 3.7s, initial ¹H-¹⁵N CP contact time of 1400μs. The ¹⁵N-¹³C SPECIFIC CP⁸ transfer was achieved using contact time of 12ms with a decoupling of 58kHz rf on ¹H, number of scan NS=1408, for a total experimental time of 3days 21h

The apparent ¹³C longitudinal relaxation rates T₁ were measured on the [U-¹⁵N-[ILV-¹³C Methyl]]-labeled sample in 1D ¹³C experiments (Figure 62.B,C). Following initial cross-polarization to ¹³C, carbon magnetization was placed along the longitudinal axis during a variable relaxation delay and then brought back to the transverse plane for detection. The relaxation delay was set in the range of 1μs to 4.5s. Experiments were recorded with AQ=30ms, NS=256, RD of 10s and a total experimental time for all experiments of 21h.

9.3 Proton-detected ssNMR

The proton-detected 2D ¹³C-¹H heteronuclear correlation (Figure 65.A) was recorded according conventional (H)CH pulse sequence¹¹² presented in Figure 79.A (Appendix II). Acquisition time for the direct and indirect dimensions was 8ms and 40.7ms, respectively. In 3D (H)CCH experiment (Figure 79.B, Appendix II), ¹³C-¹³C homonuclear dipolar interactions were recoupled by RFDR^{62,105}, thus giving information about carbon-carbon contacts. Maximal acquisition times were AQ₁=22.6ms, AQ₂=16.8ms (indirect dimensions) and AQ₃=7ms (direct dimension). RFDR recoupling was applied during 19.2ms and the applied carbon rf field was 45kHz. In all the experiments during ¹H-¹³C CP, the carbon rf field was ramped down from 100% to 70% with an average value of 15kHz, while the proton rf field was kept constant at 45kHz. The CP contact times were 520μs and 320μs for ¹H-¹³C and ¹³C-¹H transfer, respectively. During ¹³C evolution, ¹H-¹³C couplings were removed by application of low-power XiX¹¹³ on proton channel with an rf strength of 12kHz. During proton detection, the ¹H-¹³C scalar couplings were removed using WALTZ-16 decoupling sequence⁶¹ with an rf strength of 7kHz. The experiment times were 20.5h and 4days 12h for the 2D and 3D experiment, respectively.

10 Results and Discussion

10.1 Isotopic labeling pattern

The knowledge of the biosynthetic pathways of amino acids allows prediction of the labeling pattern which can be achieved in proteins over-expressed in *E. coli*.⁸⁹ To identify the labeling pattern obtained in our approach, we recorded solid-state NMR experiments on assembled MxiH needles. Two MxiH needle samples were produced: first, the [LV-¹³C Methyl]-labeled sample was prepared using the α -ketoisovalerate precursor ¹³C labeled at both methyl groups; second, the [U-¹⁵N-[ILV-¹³C Methyl]]-labeled sample was produced using three labeled precursors: sodium α -keto-isovalerate and sodium α -keto-butyrate^{114,115} labeled at the methyl groups and ¹⁵NH₄Cl (Figure 59.A,B).

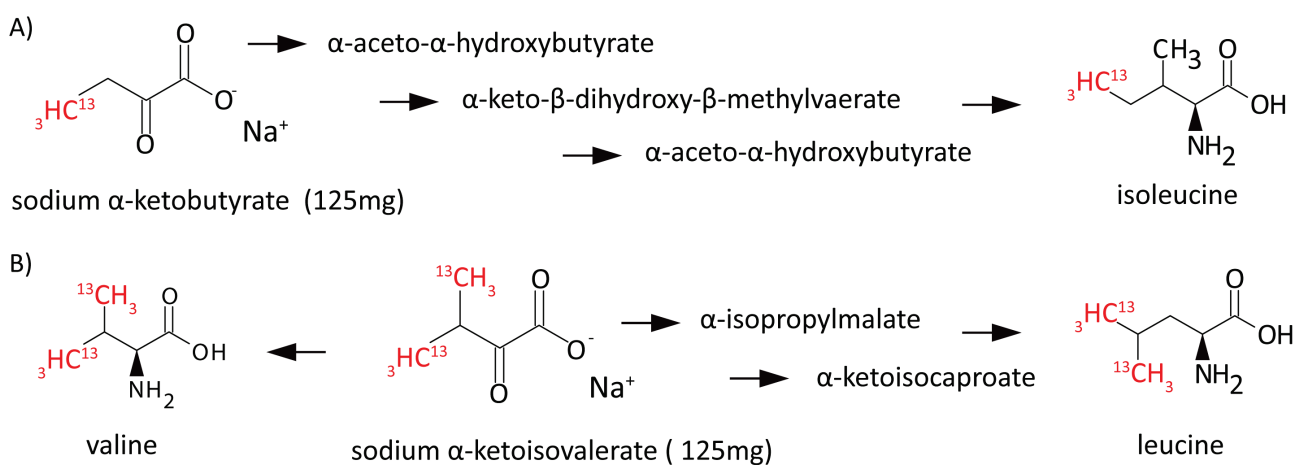


Figure 59 Representation of the biosynthetic pathways of the residues isoleucine, leucine and valine based on the induced precursors A) keto-butyrate and the intermediate steps achieving ¹³C-C δ 1 labeled isoleucine B) 2-ketoisovalerate and the resulting products ¹³C γ 1, ¹³C γ 2 labeled valine and ¹³C δ 1, ¹³C δ 2 labeled leucine.

Solid-state NMR 1D CP-spectra were recorded for these samples and compared with spectra obtained from two established ¹³C labeling strategies: uniform labeling [U-¹³C]-glc, and sparse labeling [1-¹³C]-glc. As previously demonstrated, the [1-¹³C]-glc- and [2-¹³C]-glc labeled spectra show a large improvement in ¹³C resolution and line-widths relative to [U-¹³C]-glc labeled spectra, due to the removal of a majority of one-bond ¹³C-¹³C dipolar and *J* couplings.⁸⁵ However, ¹³C spectra are highly crowded in the chemical shift range from 15ppm to 30ppm which corresponds to the methyl region of amino acids since methyl carbons are ¹³C labeled for all amino acid types in the [1-¹³C]-glc labeling. For some amino acids, carbon resonances from CH₂ groups can also be found in the same chemical shift range as methyl groups. In contrast, the [LV-¹³C Methyl]-labeled MxiH spectrum shows only signals from methyl carbons, in the chemical shift range from 15ppm to 26ppm (Table 4). Considering a tolerance window of ± 0.15 ppm, only four leucine or valine methyl carbon resonances are isolated in the [U-¹³C]-glc labeling, in contrast to six resonances for the [1-¹³C]-glc labeling and ten resonances for the [LV-¹³C Methyl] labeling.

The comparison between the [U-¹⁵N-[ILV-¹³C Methyl]]-labeled sample with the [LV-¹³C Methyl]-labeled

sample demonstrates the similarity of the valine and leucine methyl resonances both in resolution and signal intensity as well as the addition of the C δ 1 methyl carbons of isoleucines in the [U-¹⁵N-[ILV-¹³C Methyl]]-labeled sample (Figure 60.B), which are absent in the [LV-¹³C Methyl] spectra.

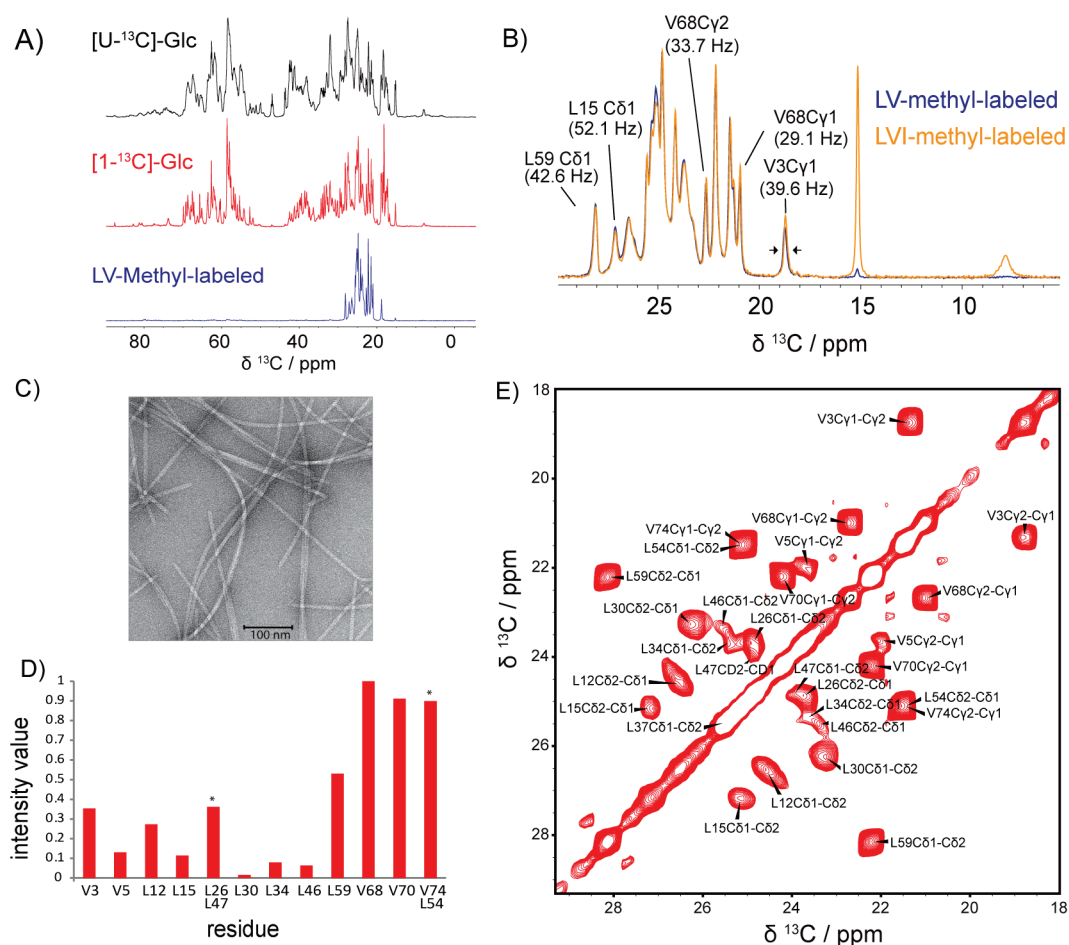


Figure 60 A) Excerpts of 1D CP-spectra of (black) [U-¹³C]-glc, (red) [1-¹³C]-glc and (violet) [LV-¹³C Methyl] B) Comparison of the [LV-¹³C Methyl]-labeled sample (blue) with the [U-¹⁵N-[ILV-¹³C Methyl]]-labeled sample (yellow) with full width at the half-height (FWHH) values are indicated for isolated resonances. C) Electron microscopy image of T3SS MxiH needles D) Measurements of the intensities of intra-residue cross-peaks calculated from the 2D PDSD-spectrum of [LV-¹³C Methyl] labeled sample, bars marked with a black star are ambiguous due to same chemical shift values E) 2D PDSD-spectrum of [LV-¹³C Methyl]-labeled MxiH needles recorded with a mixing time of 50ms. The cross-peak L37 C δ 1/C δ 2, located on the diagonal, was confirmed from a ¹³C-¹³C double-quantum single-quantum (DQ-SQ)-spectrum recorded on the uniformly labeled sample.²⁵

No significant signals could be detected in other regions of the 1D CP-spectra, demonstrating the absence of scrambling (i.e. undesired biosynthetic pathways) in the labeling patterns. This was further confirmed by recording 2D PDSD with short mixing time (50ms PDSD, Figure 60.B) in the LV labeled samples. Cross-peaks only appear where predicted according to the reported chemical shifts²⁵ (Table 4 and BMRB entry 18651).

Table 4 Isotropic chemical shifts of carbons and protons of the methyl groups of valine, leucine and isoleucine in the T3SS of *Shigella flexneri*.

Chemical shifts are indicated in ppm relative to DSS.

	C δ 1	C δ 2	C γ 1	C γ 2	H δ 1	H δ 2	H γ 1	H γ 2
V3	-	-	18.8	20.9	-	-	-0.4	0.5
V5	-	-	22.0	23.6	-	-	amb	amb
L12	26.5	24.7	-	-	1.0	1.2	-	-
L15	27.1	25.1	-	-	0.9	0.7	-	-
L26	23.8	24.9	-	-	1.1	0.9	-	-
L30	26.2	23.3	-	-	0.5	0.6	-	-
L34	23.7	25.4	-	-	0.6	0.1	-	-
L46	23.4	25.5	-	-	amb	0.7	-	-
L47	24.8	24.0	-	-	0.8	0.7	-	-
L54	21.5	25.1	-	-	amb	amb	-	-
L59	28.1	22.2	-	-	0.9	0.8	-	-
V68	-	-	21.0	22.7	-	-	0.9	1.1
V70	-	-	22.2	24.2	-	-	1.0	1.2
I71	8.0	-	-	-	0.7	-	-	-
V74	-	-	21.5	25.2	-	-	-	-
I78	amb	-	-	-	amb	-	-	-
I79	amb	-	-	-	amb	-	-	-

A remarkable advantage of the ^{13}C methyl labeling strategies is that the high resolution obtained in [1- ^{13}C]-glc- and [2- ^{13}C]-glc labeled samples is conserved (Figure 76), due to the absence of neighboring ^{13}C labeled nuclei. Indeed, the carbon resonances are very sharp, with line-widths measured on isolated resonances ranging from 29.1Hz (0.14ppm) to 52.1Hz (0.26ppm) (Figure 60.B).

The stereospecific assignment of the methyl groups was achieved by the approach of the [2- ^{13}C]-glc sparsely labeling strategy as previously introduced.¹¹⁶ The stereospecific assignment of valine residues was achieved based on the presence of strong correlation cross-peaks for C α -C γ 1 and C β -C γ 2 as highlighted for the assignment of V74C δ 1 and C δ 2 (Figure 61). For leucine residues, the unambiguous assignment of the prochiral methyl groups is extracted from the correlation cross-peaks for C γ -C δ 2, shown in Figure 61 for the L59 spin system. The stereospecific assignment of methyl groups is structurally relevant since side-chain rotameric conformations can be identified from the methyl ^{13}C chemical shifts.¹¹⁷

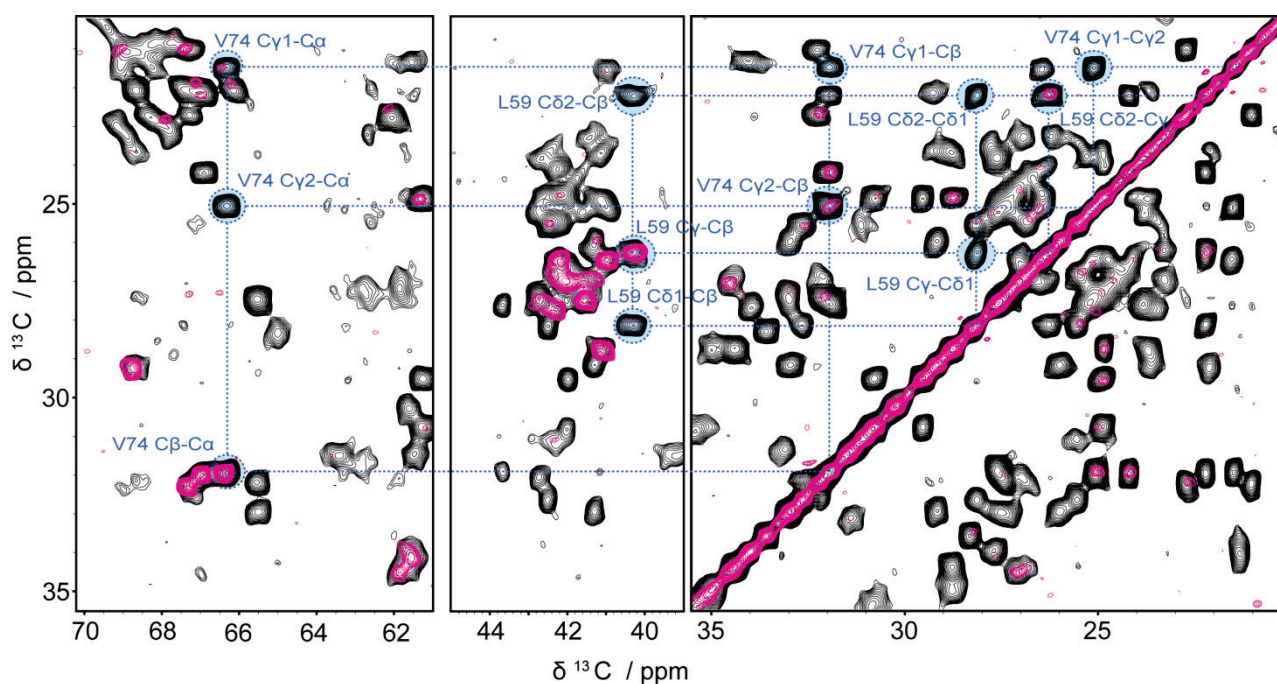


Figure 61 2D PDS spectra of [U-¹³C]-glc- (in black) and [2-¹³C]-glc labeled (in magenta) T3SS MxiH needles, recorded with a PDS mixing time of 50ms. Spin systems for V74, L59 are highlighted in blue. The stereospecific assignment for the methyl carbons of valine and leucine can be identified using the appearance of C γ -C δ 2 cross-peak in leucine, and C α -C γ 1 and C β -C γ 2 in valine, following the method introduced in Ref.²⁶

10.2 Dynamics in the biomolecular assembly

In addition to dynamic studies, we performed relaxation studies of these methyl groups. We observed strong differences in the signal intensities of intra-residue cross-peaks (Figure 60.D) measured from the 2D PDS spectrum short mixing time (Figure 60.E). Peaks corresponding to amino acids located in the C-terminal helix of the MxiH needle were more intense. The C-terminal helices, lining the inside lumen of the needle assembly^{24,25}, may be more tightly packed than the N-terminal helices present at the outer surface of the needle. Similar differences in cross-peak intensities were also observed in a DNP study of MxiH needles.¹¹⁸ It was also observed that the amino acid sequence of T3SS needle protein is more conserved in this C-terminal helix.²⁵

We measured the apparent ¹³C T₁ relaxation rates of Leu, Val and Ile methyl groups in MxiH, as knowledge about this parameter is highly recommended to choose appropriate mixing times in PDS experiments for recording distance restraints. Due to the highly resolved carbon resonances in the 1D CP-spectrum of [U-¹⁵N-[ILV-¹³C Methyl]]-labeled MxiH (Figure 62.A), we are able to select six isolated methyl peaks (Table 5). The ¹³C T₁ relaxation rates range from 318ms to 485ms for leucine and valine residues (Figure 62.B,C). The isoleucine resonances show a longer T₁: the I71C δ 1 resonance has a T₁ of 743ms; the isoleucine I78 and I79 resonances overlap; however, the fitted curve corresponds to a T₁ of 1.32s (Figure 62.C). Although the values for T₁ differ between residues, a structural interpretation of the T₁ values is difficult due to the strong influence of carbon-carbon spin diffusion during the measurement; higher MAS rates would be required to suppress this effect. For the observation of long-range distance cross-peaks, usual PDS mixing times of

longer than 400ms are employed to allow sufficient transfer of magnetization. However, the inspection of the ^{13}C T_1 decay curves for leucine and valine residues reveal that mixing times longer than 1s would result in a reduction of the magnetization amount to 10% of its initial level. We can suggest PDSM mixing times in the range of 400 to 800 milliseconds for the observation of long-range distance correlations.

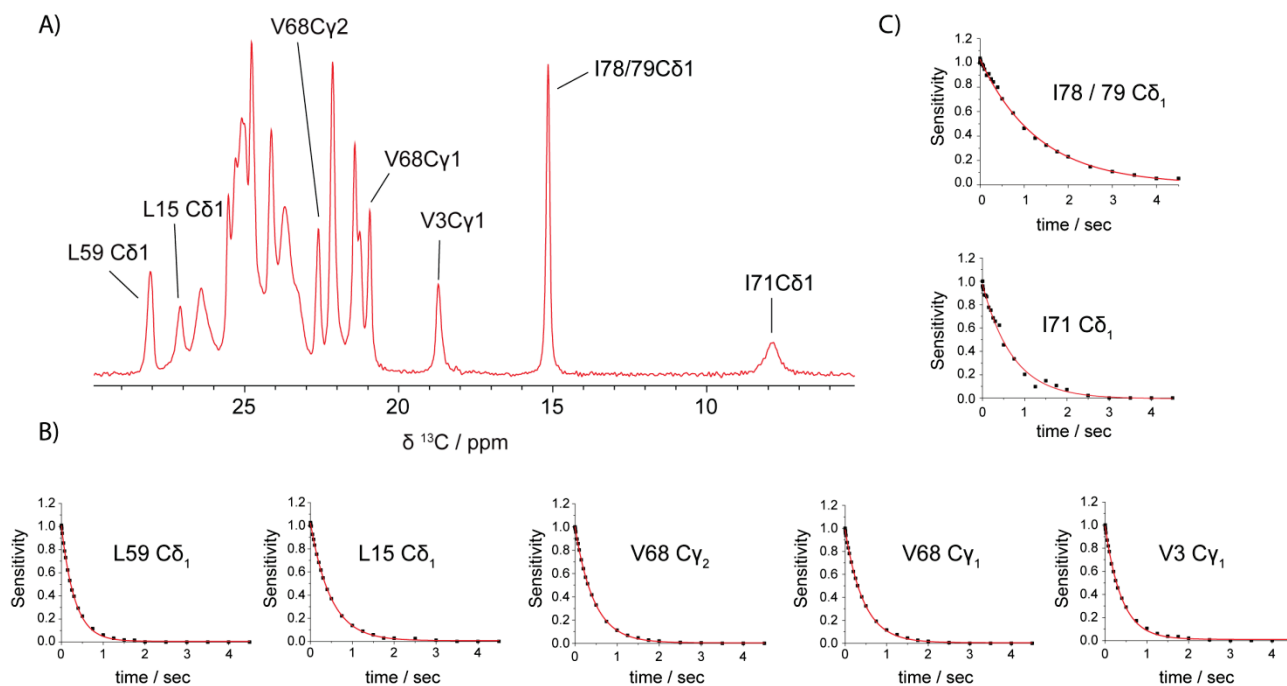


Figure 62 A) 1D CP-spectrum of $[\text{U-}^{15}\text{N-}[\text{ILV-}^{13}\text{C Methyl}]]$ B) Graphs of the ^{13}C T_1 -relaxation rates of the unambiguous leucine and valine methyl groups C) Graphs of the ^{13}C T_1 -relaxation rates of the isoleucines methyl groups.

Table 5 Longitudinal ^{13}C T_1 relaxation times of the methyl groups of valine, leucine and isoleucine in the T3SS of *Shigella flexneri*

Values are in milliseconds.

	Cδ1	Cδ2	Cγ1	Cγ2
V3	-	-	393 ± 6	-
L15	485 ± 6	-	-	-
L59	318 ± 4	-	-	-
V68	-	-	443 ± 4	447 ± 3
I71	743 ± 34	-	-	-
I78/79	1318 ± 42	-	-	-

10.3 ^{13}C - ^{13}C Distance restraints

In order to collect long-range distance restraints, we recorded 2D PDSM-spectra with a long mixing time of 800ms for both the $[\text{LV-}^{13}\text{C Methyl}]$ - and $[\text{U-}^{15}\text{N-}[\text{ILV-}^{13}\text{C Methyl}]]$ -labeled sample (Figure 63.A). By comparing those spectra with short mixing PDSM-spectra (Figure 60.E), multiple long-range cross-peaks can be observed. A long-range correlation is defined between residue i and j where $[|i-j| \geq 5]$. In our analysis, we picked all peaks in the spectrum and excluded all short-range $[1 < |i-j| < 2]$ correlations.

The long-range cross-peaks were classified as frequency unambiguous, structurally unambiguous, or ambiguous. For each cross-peak, all assignment possibilities were considered which had a chemical shift deviation of less than 0.15ppm between the resonance frequency and the frequency present in the BMRB (BMRB entry 18651). If only one assignment possibility is present in the tolerance window, the correlation is classified as frequency unambiguous (Table 17, Appendix II). The frequency unambiguous correlations indicate the right 3D fold of the subunit and the inter-molecular assemblies. As an example, all methyl-methyl contacts between residues L12 and V68 are frequency unambiguous and highlighted in magenta in Figure 63.A. Remaining long-range correlations can be assigned unambiguously by employing the 3D structure identified using the frequency unambiguous cross-peaks. Indeed, long-range correlations are usually assigned in an iterative procedure.¹¹⁹⁻¹²¹ To determine the number of structurally unambiguous correlations that can be obtained using our labeling approach, we used the atomic structure of the MxiH needle assembly determined on the basis of cryo-EM density map solid-state NMR restraints and Rosetta modelling.

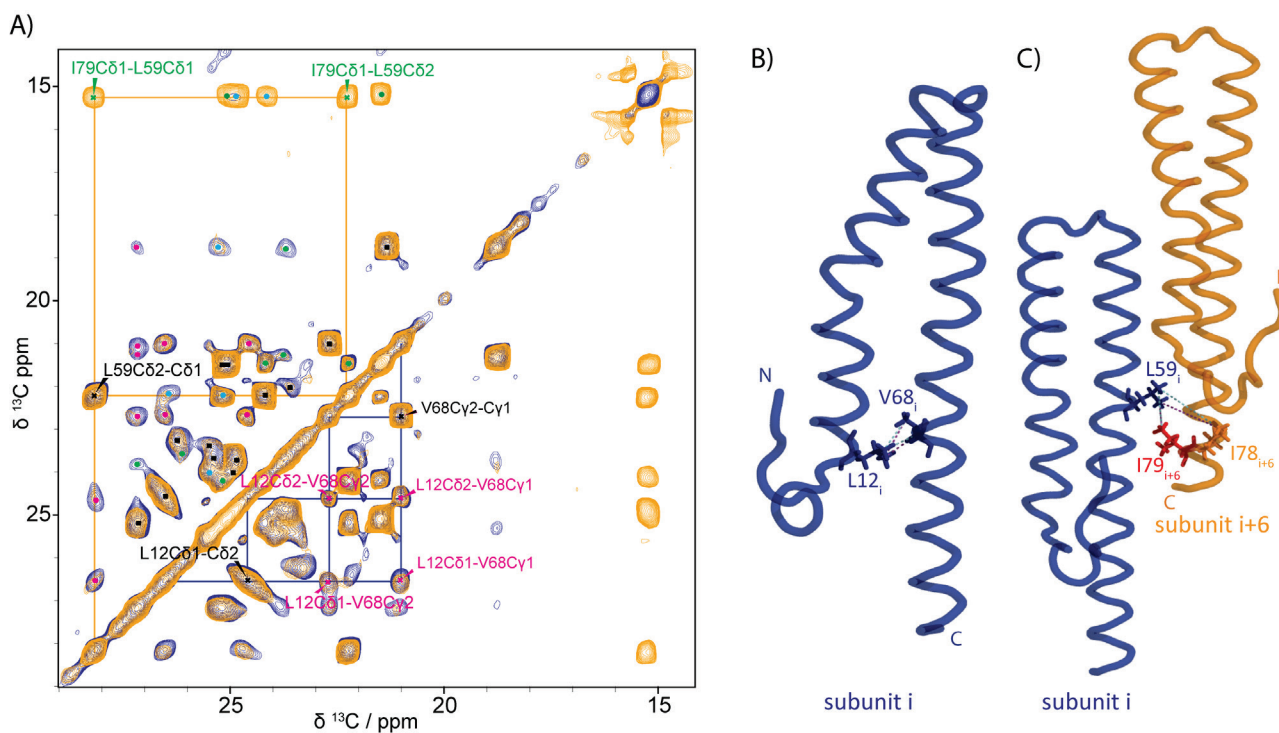


Figure 63 A) Comparison of the [LV-¹³C Methyl]-labeled sample (blue) with the [U-¹⁵N-[ILV-¹³C Methyl]]-labeled sample (yellow) PDSD-spectra with a mixing time of 800ms. The cross-peaks which are colored in black represent intra-residual interactions, in magenta unambiguous distance restraints and in green structurally unambiguous distances restraints. The shielded C δ 1 resonance of isoleucine I71 (8.0ppm) is not shown. B) The unambiguous distance restraints between L12 to V68, identified on the 2D PDSD-spectrum of the [LV-¹³C Methyl]-labeled sample, is illustrated on the atomic structure of MxiH needles from Ref.¹²²C) Representation of the structurally unambiguous distance restraints between L59 methyl carbons to I79C δ 1, as observed on the 2D PDSD-spectrum of the [U-¹⁵N-[ILV-¹³C Methyl]]-labeled sample.

For each long-range cross-peak, we calculated the distance for every assignment possibility, including intra-molecular and inter-molecular distances. A correlation was classified as structurally unambiguous when the

shortest distance was more than 2 Å shorter than any other assignment (Table 18, Appendix II). As an example, the L59 methyl carbons C δ 1 and C δ 2 have systematically shorter distances to I79C δ 1 than to I78C δ 1, as presented in Figure 63.C. The structurally unambiguous correlations I79C δ 1-L59C δ 1 and I79C δ 1-L59C δ 2 are highlighted in green in Figure 63.A. When more than a single possibility remained, correlations were classified as ambiguous (Table 19, Appendix II) and the corresponding cross-peaks are colored in blue. In this statistical analysis the 10 lowest energy conformations of the ensemble were considered: the distances listed in Table 17-19 (Appendix II) are reported as average \pm standard deviation format. The ambiguous correlations can be used further in a structure calculation procedure since many software packages allow the use of ambiguous restraints.^{119,123} Furthermore they can be used as an independent dataset for validation of the structure.

From the PDS 800ms spectrum of the [LV-¹³C Methyl]-labeled sample (Figure 63.A), we obtained ten frequency unambiguous correlations (Table 17, Appendix II), seven structurally unambiguous correlations (Table 18, Appendix II), and four ambiguous correlations (Table 19, Appendix II). The long-range interaction between the L12 and the V68 residues is highlighted in the protomer structure (Figure 63.B). This information confirms the intra-molecular packing of the second alpha-helix close to the beginning of the first alpha-helix of the MxiH-protein. By comparing the spectra of the [U-¹⁵N-[ILV-¹³C Methyl]] to those of the [LV-¹³C Methyl]-labeled sample, one can detect additional correlations between the L-V methyl carbons to the C δ 1 methyl group of I78 and I79: four structurally unambiguous and four ambiguous new correlations were found. Although the C δ 1 resonances of isoleucines I78 and I79 overlap, structural information on the intermolecular interfaces of the self-assembling system can be achieved. (Figure 63.A) One of these interactions corresponding to the lateral contact between the C-terminal helices of subunit *i* and *i*+6, is highlighted between the leucine L59 and the isoleucines I78 and I79 (Figure 63.C). In total, 21 unambiguous carbon-carbon distance restraints could be collected, which had distances varying from 3.9 Å to 10.3 Å, with an average of 6.7 Å.

10.4 ¹⁵N-¹³C Distance restraints

With ¹⁵NH₄Cl as the sole nitrogen source in the expression medium, the recombinant proteins become uniformly ¹⁵N labeled, such that ultra-sparse methyl ¹³C labeling schemes can be used to collect N-C distance restraints by recording a NCX-spectrum. In contrast to conventional ssNMR experiments such as NCA and NCO where a one-bond N-C transfer is required for the ¹⁵N-¹³C CP step, the backbone ¹⁵N and methyl ¹³C atoms are separated by three chemical bonds (valine) or four bonds (leucine, isoleucine). The intra-residual distances from the backbone ¹⁵N to the ¹³C labeled methyl carbons are larger compared to one-bond ¹⁵N-¹³C distances: on average 3.4 Å for valine, 4.3 Å for leucine and 4.4 Å for isoleucine. As a result, the contact time of the N-C transfer step has to be increased. The 1D build-up curves of the ¹⁵N-¹³C CP contact time starting from 5ms up to 15ms are shown in Figure 64.B.

A crucial element for this type of experiments is the control on the rf irradiation time. For probe safety, we

have chosen a CP contact time of 12ms with a ^1H decoupling strength of 58kHz and low-power irradiation on the carbon and ^{15}N channels (8.8kHz rf ^{13}C , 4.5kHz rf ^{15}N). The 2D NCX-spectrum contains intra-residue cross-peaks and sequential cross-peaks, which can be used to confirm the ^{15}N and ^{13}C chemical shift assignment. Indeed, we assigned six cross-peaks for N_i to C_{i-1} correlations and one cross-peak for N_i to C_{i+1} , with distances ranging from 3.5 Å to 6.4 Å. From the 2D NCX-spectrum, we obtained a total of twelve structurally unambiguous distance restraints (Table 17, Appendix II) and eight ambiguous restraints (Table 18, Appendix II). The information about the axial interface between needle subunits i and $i+11$ is illustrated in Figure 64.C by an unambiguous contact between E49N and I78C δ 1. Distance restraints were detected with distances ranging from 5.6 Å to 12.1 Å and an average of 7.3 Å.

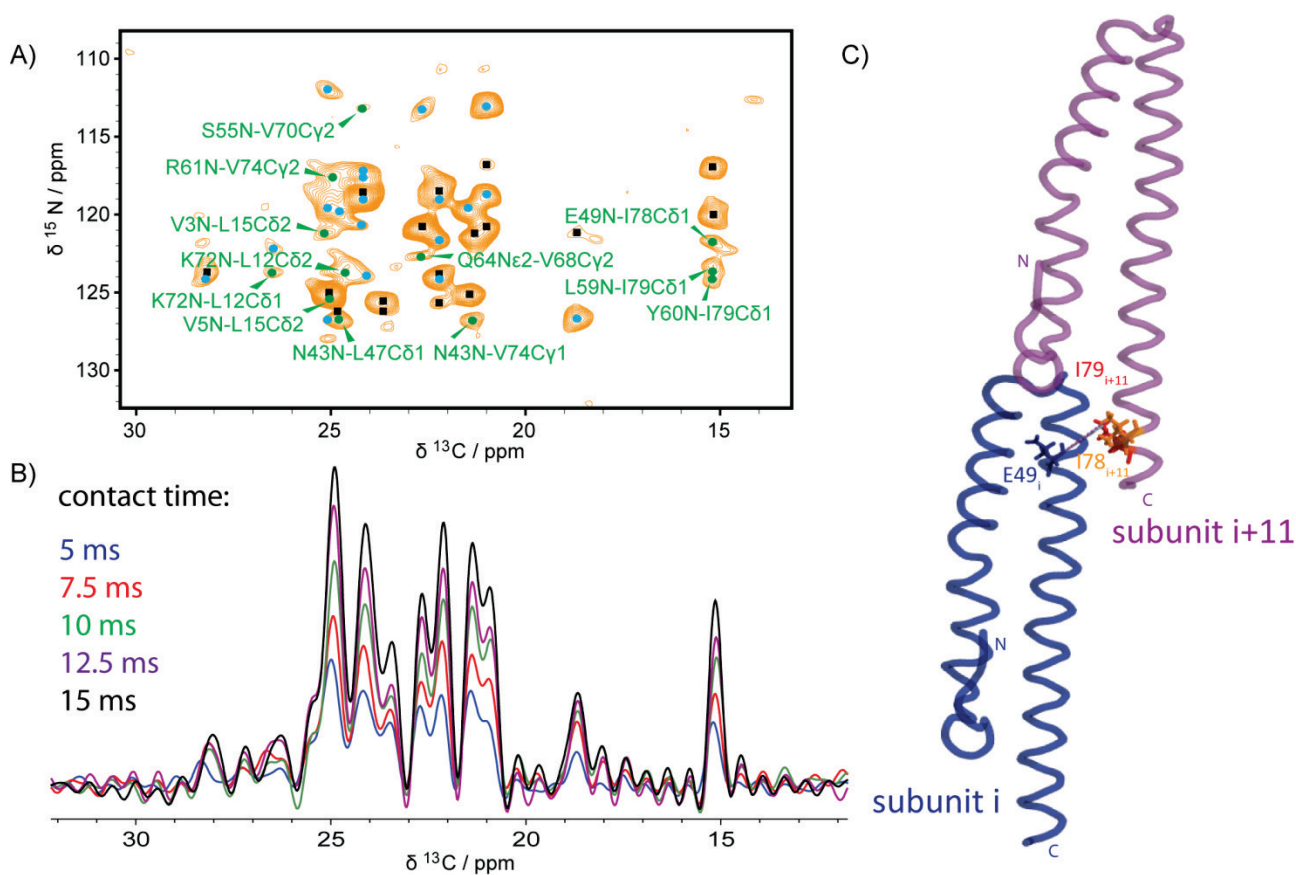


Figure 64 A) plot of the 2D NCX-spectrum of $[\text{U-}^{15}\text{N-}[\text{ILV-}^{13}\text{C Methyl}]]$ -labeled sample, cross-peaks which are colored in black represent intra-residual-, in blue ambiguous interactions and in green structurally unambiguous distance restraints. B) Build up curves of the double CP contact time in the range from 5ms up to 15ms. C) Illustration of the structurally unambiguous distance restraints between E49N to I78C δ 1, observed on the 2D NCX-spectrum, represented on the atomic structure of MxiH needles from Ref.¹²²

10.5 Proton-detected ssNMR

Ultra-fast MAS ($\sim 60\text{kHz}$ MAS) and high external magnetic fields provide attractive conditions to use proton detection in fully protonated samples.^{13,124} This approach provides better sensitivity compared to ^{13}C

detection and improved resolution via the additional proton dimension. In addition, low-power heteronuclear decoupling schemes can be efficiently employed at ultra-fast spinning rates¹¹³, and thus very long ¹³C evolution time can be used to gain high resolution originating from the very sharp carbon lines.

Figure 65.A shows a 2D proton-carbon heteronuclear correlation experiment which demonstrates the possibility to assign methyl proton chemical shifts, in part due to the reduced number of resonances. Proton line-widths vary from 160 to 300Hz, which indicates a high variation in the mobility of methyl groups as well as on the disorder and heterogeneity of the molecular assembly. The line-width of the ¹³C projection is presented in Figure 78 (Appendix II). We recorded a 3D (H)CCH experiment (Figure 65.B) using RFDR⁶² to recouple ¹³C-¹³C homonuclear dipolar interactions and obtain information about carbon-carbon contacts. Although the RFDR recoupling period lasted moderately long (19.2ms), we observed only intra-residual contacts as long-range magnetization transfers were suppressed by dipolar truncation.¹²⁵ The conducted experiments allowed the assignment of almost all methyl proton signals via the carbon resonance assignments, which were obtained previously²⁵ and confirmed in Figure 60.E and Figure 63.A.

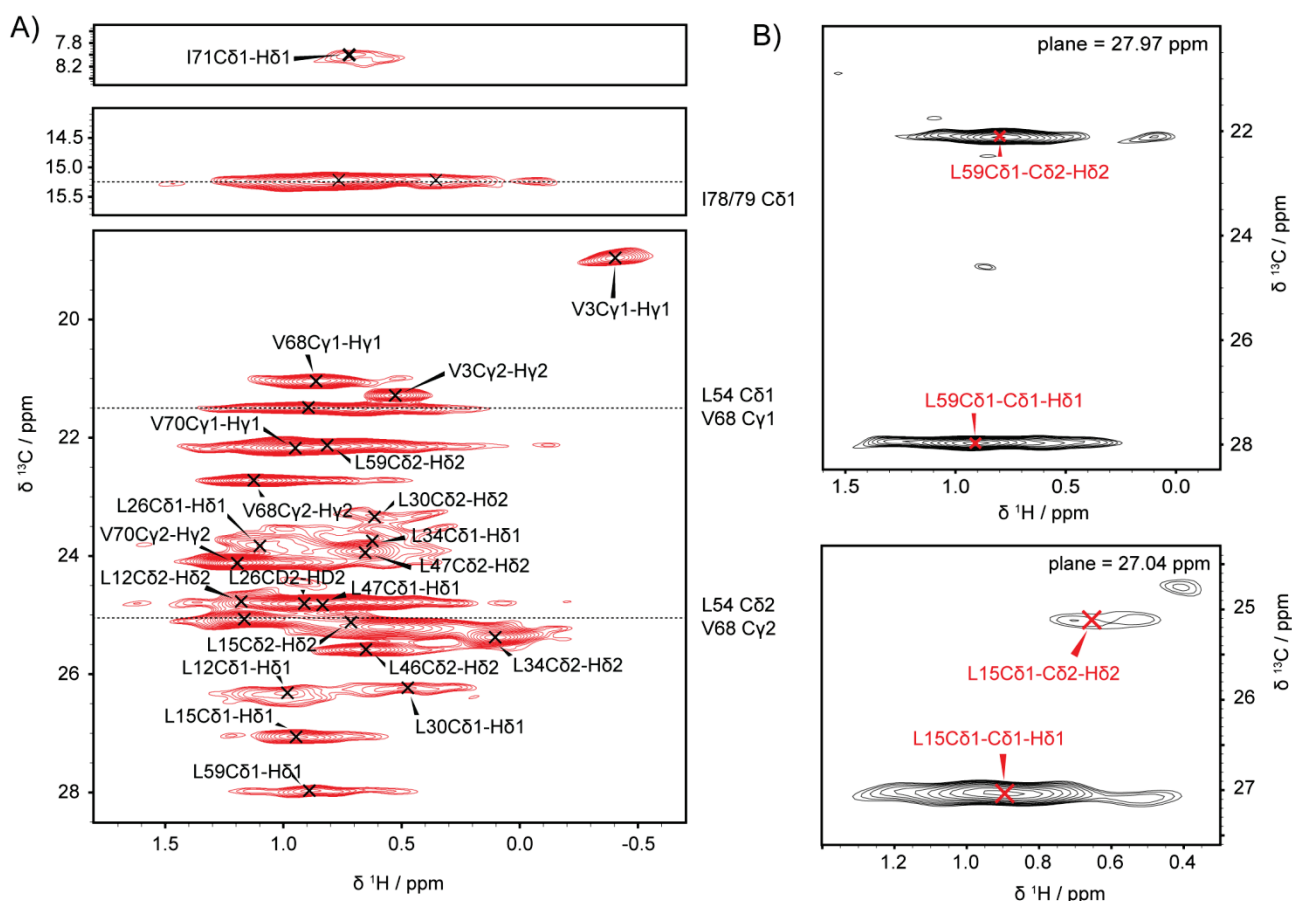


Figure 65 A) Two dimensional proton-detected (H)CH correlation spectrum. Three dashed lines show unresolved signals originating from known resonances, which are indicated on the right side of the panel. B) Two representative planes from 3D (H)CCH correlation experiment demonstrate the assignment of methyl protons from the same amino acid. Carbon isotropic chemical shift values in the indirect dimension are indicated at the top of the planes.

The chemical shift of methyl proton can reveal local structural features, such as the shielding and

deshielding effects due to aromatic rings currents. As proton chemical shifts are highly sensitive to the surrounding environment, they can also be used as a basis for the comparison of different sample preparations, for instance to detect the binding sites of ligands and to characterize the influence of binding partners. Considering the recent improvements in the performance of chemical shift prediction software packages¹²⁶⁻¹²⁸, the determination and validation of side-chain conformation could employ side-chain chemical shifts similarly to the present use of backbone chemical shifts.

11 Conclusions

We have reported a straightforward isotopic labeling strategy for the Type-Three Secretion System (T3SS) needle assembly of MxiH proteins and obtained 10 frequency unambiguous distance restraints, 23 structurally unambiguous distance restraints and 16 ambiguous distance restraints. We could show that the use of the two biosynthetic precursors, α -ketoisovalerate and α -ketobutyrate, is a valuable tool in achieving clear and ultra-sparse labeled ^{13}C methyl group spectra, with a high resolution for ^{13}C resonances (line-widths as low as 29.1Hz). We demonstrated the possibility of recording NCX-spectra where long-range distance correlations were detected. Our approach reveals a significant increase of detectable long range correlation distances up to 12 Å in comparison to primarily proton detected ssNMR experiments on partial protonated samples.¹¹¹ Additionally, the high resolution of peaks and the spectral simplification enabled many important aspects for the study of proteins: the assignment of methyl proton chemical shifts, the collection of long-range restraints, and the study of dynamics. The proposed labeling strategy can be employed to advance the study of large and complex protein systems.

12 Appendix I

12.1 Ubiquitin MPD

12.1.1 SsNMR experiments (MPD)

Table 6 List of solid-state NMR experiments performed on various samples

sample	Type of Experiment	Acquisition Times (ms)	Mixing Time (ms)	Recycling Delay (s)	Spectrum size (ppm)	Number of Scans	Total time	Spectrometer ¹ H larmor freq.
unif-Ubiquitin	PDS	14 / 14	50	3.0	354 / 240	32	1 d 18 h	850 MHz
unif-Ubiquitin	PDS	14 / 14	150	3.0	354 / 240	32	1d 18 h	850 MHz
unif-Ubiquitin	NCA	10 / 12		2.6	296 / 44	64	4h	800 MHz
unif-Ubiquitin	NCO	10 / 12		2.6	296 / 44	64	4h	800 MHz
unif-Ubiquitin	INEPT	10 / 10		1.1	334 / 14	32	2.5 h	850 MHz
unif-Ubiquitin	1D- ¹³ C	20		2.2	300	16	1m	850 MHz
unif-Ubiquitin	1D- ¹⁵ N	15		2.0	702	128	4m	800 MHz
unif-Ubiquitin	NCACX	14 / 11	30	3.3	283 / 60	1504	6d 5h	800 MHz
unif-Ubiquitin	NCOX	10 / 14	60	3.2	283 / 60	2080	10d 20h	800 MHz
unif-Ubiquitin	NCOCA	10 / 13.5		3.2	300 / 46	576	2d 8h	850 MHz
unif-Ubiquitin	NCACO	10 / 14		3.2	300 / 46	448	1d 21h	850 MHz
unif-Ubiquitin	CANCOCA	9 / 8		3.3	300 / 28	2304	8d 14h	850 MHz
1glu-Ubiquitin	1D- ¹³ C	20		2.5	334	32	1m	850 MHz
1glu-Ubiquitin	PDS	14 / 11	100	3.0	346 / 200	80	2d 19h	800 MHz
1glu-Ubiquitin	PDS	15 / 11	900	2.2	346 / 250	256	9d 1h	800 MHz
1glu-Ubiquitin	PDS	15 / 10	400	2.3	318 / 220	304	6d 19h	600 MHz
1glu-Ubiquitin	NCA	12 / 10		3.0	283 / 200	144	1d 15h	800 MHz
2glu-Ubiquitin	1D- ¹³ C	20		2.5	334	32	1m	850 MHz
2glu-Ubiquitin	PDS	14 / 11	100	2.8	249 / 100	96	2d 19 h	800 MHz
2glu-Ubiquitin	PDS	15 / 11	900	2.0	346 / 210	352	10d 7h	800 MHz
2glu-Ubiquitin	PDS	15 / 10	900	2.0	318 / 200	496	11d 1h	600 MHz
2glu-Ubiquitin	NCA	13 / 12		3.3	283 / 50	240	21h	800 MHz

12.1.2 Chemical shift lists (MPD)

Table 7 Observed chemical shifts for the main conformation of microcrystalline ubiquitin (MPD)

The chemical shifts values are indicated in ppm relative to DSS (deposited in the BMRB under accession ID: [25123](#)). Resonances marked with “*” symbol are ambiguous.

residue	N	CO	C α	C β	C γ / C γ 2	C δ / C δ 2	C ϵ / C ϵ 2	C ζ
Met 1	38.9	170.9	55.2	34.2	31.3		18.4	
Gln 2	125.0	175.1	54.2	31.5	35.3	180.1		
Ile 3	115.2	172.3	58.8	41.7	25.2 / 18.2	14.1		
Phe 4	119.2	175.3	54.7	41.4	139.2	132.1*	131.0*	127.1
Val 5	121.0	175.5	60.0	34.8	21.1 / 20.6			
Lys 6	126.2	177.2	53.4	33.8	24.5	29.5		
Thr 7	112.8	177.7	60.3	70.2	21.0			
Leu 8	119.3	177.4	57.3	42.4	27.4	28.5/ 23.9		
Thr 9	102.5	175.6	61.0	69.3	23.0			
Gly 10	109.4	173.4	44.9					
Lys 11	122.2	175.4	56.5	32.9	25.2	29.4	42.6	
Thr 12	119.8	174.4	62.8	69.6	22.7			
Ile 13	128.6	175.2	59.8	40.4	26.6 / 18.1	14.8		
Thr 14	124.4	174.0	62.6	69.5	22.4			
Leu 15	125.2	174.1	52.5	46.3	26.8	27.6 / 24.5		
Glu 16	123.4	175.1	54.2	30.4	36.7	178.7		
Val 17	118.8	175.1	58.6	36.5	22.5 / 19.2			
Glu 18	119.0	175.3	54.7	29.9	37.0	185.4		
Pro 19	133.7	177.5	65.8	32.2	27.8	49.9		
Ser 20	108.8	173.3	57.1	61.3				
Asp 21	122.3	176.6	56.7	38.5	181.5			
Thr 22	106.6	176.3	59.7	70.8	22.0			
Ile 23	121.7	179.2	61.8	33.3	27.7 / 17.4	9.5		
Glu 24	123.4	178.9	59.0	30.6	33.7	184.0		
Asn 25	121.0	180.4	55.3	36.6	174.2			
Val 26	124.4	177.4	67.7	31.0	20.7 / 24.0			
Lys 27	120.0	180.8	59.0	34.7	25.9	31.1	42.8	
Ala 28	122.2	179.9	55.0	18.5				
Lys 29	118.6	180.8	59.3	33.9	26.4	30.5	43.1	
Ile 30	122.7	177.5	66.4	36.9	30.7 / 16.9	15.7		
Gln 31	122.8	179.2	59.5	28.1	33.7	180.0		
Asp 32	121.4	177.6	57.9	38.7	178.8			
Lys 33	120.0	178.1	59.5	34.4	25.5	29.9	42.8	
Glu 34	112.0	177.4	55.0	33.3	34.9			
Gyl 35	109.2	173.9	46.0					
Ile 36	119.7	173.5	57.7	40.4	27.7 / 18.0	13.8		

residue	N	CO	C α	C β	C γ / C γ 2	C δ / C δ 2	C ϵ / C ϵ 2	C ζ
Pro 37	142.6	176.7	61.8	32.1	28.5	51.4		
Pro 38	136.6	178.4	66.4	33.0	28.1	51.4		
Asp 39	115.0	177.3	55.4	38.3	180.3			
Gln 40	117.8	175.9	55.2	29.5	34.0	179.2		
Gln 41	120.7	174.7	55.9	30.7	33.7			
Arg 42	125.7	173.5	55.2	30.8	27.8			
Leu 43	127.0	175.6	53.3	47.4	27.5	25.4 / 26.5		
Ile 44	119.3	176.1	58.2	41.5	27.7 / 17.6	12.8		
Phe 45	123.7	173.6	56.4	44.3	137.1	132.4*	132.3*	
Ala 46	135.7	177.3	52.2	16.5				
Gly 47	103.7	173.5	45.6					
Lys 48	122.5	175.0	55.3	31.4	25.3	29.4	42.7	
Gln 49	123.9	175.7	55.9	30.3	36.2	178.9		
Leu 50	123.9	176.4	54.1	41.3	25.6	26.1 / 19.4		
Glu 51	123.5	177.1	55.9	29.9	35.0			
Asp 52	118.6	175.9	57.7	39.7	181.2			
Gly 53	100.9	174.9	45.6					
Arg 54	120.8	175.0	53.4	32.6	27.4	43.0		
Thr 55	107.6	176.4	59.6	73.5	23.3			
Leu 56	117.4	180.3	58.3	39.7	26.1	27.8 / 23.3		
Ser 57	112.2	178.4	60.8	62.7				
Asp 58	125.1	176.1	57.6	38.8	180.6			
Tyr 59	113.0	176.0	57.1	40.5	130.8	131.1*	117.8*	156.9
Asn 60	117.4	174.1	53.7	37.7	178.1			
Ile 61	118.7	174.1	62.9	36.5	28.9 / 17.9	15.3		
Gln 62	125.8	175.2	53.6	32.2	33.2	180.8		
Lys 63	119.4	176.6	58.4	32.8		30.1		
Glu 64	114.9	174.9	58.2	26.1	35.6	183.9		
Ser 65	115.7	172.0	61.3	65.1				
Thr 66	118.2	172.9	62.5	70.0	22.3			
Leu 67	128.9	175.5	53.5	44.5	30.6	24.3 / 25.8		
His 68	117.8	175.6	54.5	30.4				
Leu 69	126.0	176.4	55.2	41.9	27.6	24.2 / 25.6		
Val 70	112.2	176.2	58.9	36.3	22.0 / 18.4			
Leu 71	119.7							

Table 8 Chemicals shift of the second and third conformation of microcrystalline Ubiquitin (MPD)

Chemical shifts are indicated in ppm relative to the DSS. The second and third conformation is highlighted by one or two primes respectively. Resonances marked with "*" symbol are ambiguous.

residue	N	CO	C α	C β	C γ / C γ 2	C δ / C δ 2	C ϵ / C ϵ 2
Met 1`	38.9	170.8	55.2	34.1	31.3		
Gln 2`	124.5	174.8	54.4	31.5	35.4	179.6	
Ile 3`	116.7	172.4	58.8	41.7	25.2 / 18.2	14.1	
Phe 4`	119.2	175.1	54.9	41.2	139.2	132.1*	
Val 5`	121.9	175.1	60.4	34.2	22.7 / 20.5		
Lys 6`	128.2	177.1	54.0	34.6	24.5	29.2	42.6
Thr 7`	113.5	177.3	60.2	71.3	21.4		
Thr 7``	114.7	177.1	60.0	70.1	20.7		
Leu 8`	121.1	176.2	57.0	43.1	26.9	26.6 / 23.8	
Leu 8``	120.1	177.2	56.3	42.2			
Thr 9`	101.8	176.1	60.3	69.1	22.3		
Thr 9``	104.8	175.7	61.2	69.8	22.9		
Gly 10`	109.3	174.1	45.4				
Gly 10``	109.3	173.5	44.9				
Lys 11`	121.1	176.0	55.0	33.3	25.1		
Thr 12`	118.4	174.0	62.8	70.0	22.6		
Thr 12``		174.3	62.4	69.5			
Ile 13`	129.3	175.2	60.0	41.4	26.9 / 18.1	15.0	
Ile 13``	127.3	175.1	59.6	41.4		18.7	
Thr 14`	123.3	173.8	62.3	69.6	22.2		
Leu 15`	125.2	173.8	52.4	45.9	26.8	27.6 / 24.5	
Thr 22`		176.3					
Ile 23`	122.5	179.2	62.2	33,6	28.3 / 17.9	9.7	
Val 26`	124.5	177.6	67.7	31.0	21.3 / 23.8		
Ile 30`		177.5	66.3	36.8	30.7 / 17.2	15.9	
Lys 33`		178.1					
Glu 34`	112.1	177.4	55.2	34.4	35.5	181.3	
Gyl 35`	107.6	174,2	45.7				
Ile 36`	119.8	173.5	59.0	40.8	27.2 / 18.0	14.1	
Pro 37`		176.6	61.6	32.2	28.7	51.0	
Pro 38`		178.5					
Asp 39`	115.6	177.1	55.4	38.3	180.2		
Gln 40`	118.2	175.5	55.8	29.9	34.0	179.4	
Gln 41`	119.5	175.8	55.9	32.2	33.7		
Arg 42`		173.5	54.8	30.4			
Leu 43`	126.2	175.5	53.1	46.1	28.1	24.8 / 26.5	
Ile 44`	121,1	175.4	58.9	43.2	27.9 / 18.0	14.5	
Phe 45`	124.7	174.4	55.5	44.1			
Ala 46`	133.6	177.5	52,4	16.4			
Gly 47`	104.2		45.5				
Lys 48`	122.5	175.0	55.3	31.7	25.1	29.4	42.5
Leu 56`			58.2	39.1	26.1	27.8 / 23.3	
Ser 65`		172.1	61.2	65.1			
Thr 66`	117.8	173.7	62.4	70.1	21.8		
Leu 67`	128.8	175.5	53.9	44.7	30.4	24.5 / 26.0	
His 68`	117.0	175.5	54.4	30.8			
Leu 69`	125.7	176.6	54.4	41.2			
Val 70`		176.4	59.1	36.8	22.1 / 18.5		

12.1.3 Secondary structure analysis of the second-conformation of Ubiquitin (MPD)

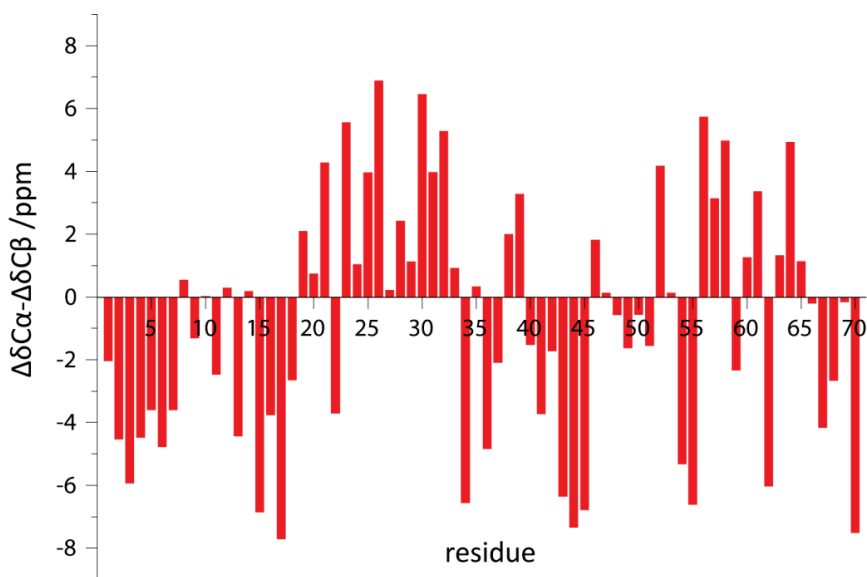


Figure 66 Secondary chemical shift analysis of the second conformation of Ubiquitin (MPD).

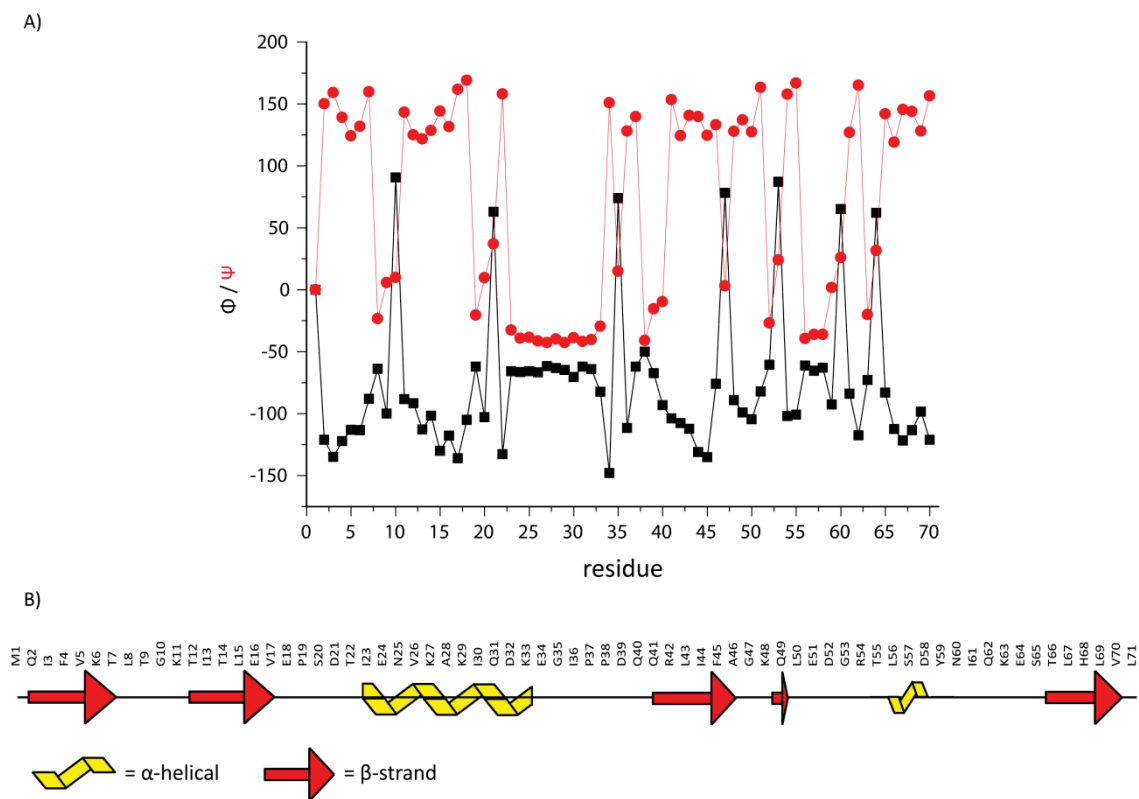


Figure 67 A) TALOS+ analysis, predicting Phi (black) and Psi (red) dihedral torsion angles of the second conformation B) cartoon of the result of TALOS+ prediction, only residue E34 shows ambiguity for the secondary structure prediction.

12.1.4 Validation results of Ubiquitin (MPD)

12.1.4.1 Validation result of the main conformation (MPD)

Table 9 Structural statistics for the 10 lowest-energy conformers of microcrystalline ubiquitin (MPD)

Conformers were calculated with XPLOR-NIH using 518 distance restraints with a defined distance range of 1-7 Å.

Summary of conformationally-restricting experimental constraints ^a

Distance constraints:

Total	518
intra-residue [i = j]	0
sequential [i - j = 1]	0
medium range [1 < i - j < 5]	241
long range [i - j ≥ 5]	277
Constraints per restrained residue ^b	7.3
Dihedral-angle constraints:	136
Total number of restricting constraints ^b	654
Total number of restricting constraints per restrained residue ^b	9.2
Restricting long-range constraints per restrained residue ^b	3.9
Total structures computed	1000
Number of structures used	10

Residual constraint violations ^{a,c}

Distance violations / structure	
0.1 - 0.2 Å	15.4
0.2 - 0.5 Å	10.3
> 0.5 Å	0
RMS of distance violation / constraint	0.05 Å
Maximum distance violation ^d	0.49 Å

Dihedral angle violations / structure

1 - 10 °	43
> 10 °	11.1
RMS of dihedral angle violation / constraint	6.26 °
Maximum dihedral angle violation ^d	30.50 °

Rmsd Values

	all	ordered ^e
All backbone atoms	0.7 Å	0.5 Å
All heavy atoms	1.3 Å	1.1 Å

Structure Quality Factors - overall statistics

	Mean score	SD	Z-score ^g
Procheck G-factor ^e (phi / psi only)	-0.54	N/A	-1.81
Procheck G-factor ^e (all dihedral angles)	-0.79	N/A	-4.67
Verify3D	0.31	0.02	-2.41
Prosall (-ve)	0.75	0.06	0.41
MolProbity clashscore	85.85	10.62	-13.21
Ramachandran Plot Summary from Procheck ^f			
Most favoured regions	86.0%		
Additionally allowed regions	8.7%		
Generously allowed regions	4.9%		
Disallowed regions	0.3%		
Ramachandran Plot Statistics from Richardson's lab			
Most favoured	91.6%		
Allowed regions	3.7%		
Disallowed regions	4.7%		

^a Analysed for residues 1 to 72

^b There are 71 residues with conformationally restricting constraints

^c Calculated for all constraints for the given residues, using average r^{-6}

^d Largest constraint violation among all the reported structures

^e Residues with sum of phi and psi order parameters > 1.8

Ordered residue ranges:

^f Residues selected based on: all residues

Selected residue ranges: all

^g With respect to mean and standard deviation for a set of 252 X-ray structures < 500 residues, of resolution $\leq 1.80 \text{ \AA}$, R-factor ≤ 0.25 and R-free ≤ 0.28 ; a positive value indicates a 'better' score

Generated using PSVS 1.5⁷⁶

12.1.5 Validation result of the second conformation (MPD)

Table 10 Structural statistics for the 10 lowest-energy conformers of microcrystalline ubiquitin (MPD)

Conformers were calculated with XPLOR-NIH using 318 distance restraints with a defined distance range of 1-7 Å.

Summary of conformationally-restricting experimental constraints ^a

Distance constraints:

Total	318
intra-residue [i = j]	0
sequential [i - j = 1]	0
medium range [1 < i - j <	148
long range [i - j ≥ 5]	170
Constraints per restrained residue ^b	4.5
Dihedral-angle constraints:	134
Total number of restricting constraints ^b	452
Total number of restricting constraints per restrained residue ^b	6.4
Restricting long-range constraints per restrained residue ^b	2.4
Total structures computed	1000
Number of structures used	10

Residual constraint violations ^{a,c}

Distance violations / structure	
0.1 - 0.2 Å	10.9
0.2 - 0.5 Å	10.2
> 0.5 Å	0.8
RMS of distance violation / constraint	0.06 Å
Maximum distance violation ^d	0.79 Å

Dihedral angle violations / structure

1 - 10 °	42.5
> 10 °	11.7
RMS of dihedral angle violation / constraint	6.30°
Maximum dihedral angle violation ^d	34.60 °

Rmsd Values

	all	ordered ^e
All backbone atoms	1.1 Å	0.9 Å
All heavy atoms	1.8 Å	1.6 Å

Structure Quality Factors - overall statistics

	Mean score	SD	Z-score ^g
Procheck G-factor ^e (phi / psi only)	-0.62	N/A	-2.12
Procheck G-factor ^e (all dihedral angles)	-0.89	N/A	-5.26
Verify3D	0.26	0.05	-3.21
Prosall (-ve)	0.57	0.15	-0.33
MolProbity clashscore	75.53	9.18	-11.44
Ramachandran Plot Summary from Procheck ^f			
Most favoured regions	83.5%		
Additionally allowed regions	12.2%		
Generously allowed regions	4.1%		
Disallowed regions	0.2%		
Ramachandran Plot Statistics from Richardson's lab			
Most favoured	86.4%		
Allowed regions	8.4%		
Disallowed regions	5.1%		

^a Analysed for residues 1 to 72

^b There are 71 residues with conformationally restricting constraints

^c Calculated for all constraints for the given residues, using average r^{-6}

^d Largest constraint violation among all the reported structures

^e Residues with sum of phi and psi order parameters > 1.8

Ordered residue ranges:

^f Residues selected based on: all residues

Selected residue ranges: all

^g With respect to mean and standard deviation for a set of 252 X-ray structures < 500 residues, of resolution $\leq 1.80 \text{ \AA}$, R-factor ≤ 0.25 and R-free ≤ 0.28 ; a positive value indicates a 'better' score

Generated using PSVS 1.5⁷⁶

12.1.6 Comparison to different structures (MPD)

12.1.6.1 Comparison between the main conformation (MPD) to the bundle of structures of 2JZZ

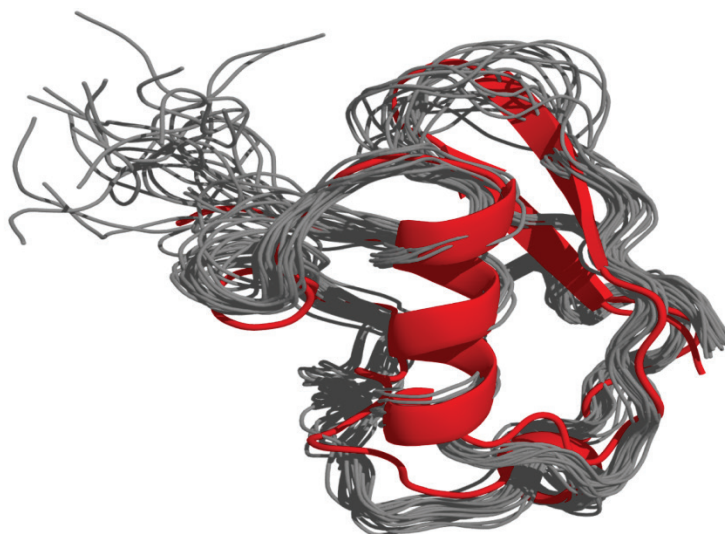


Figure 68 Comparison of the lowest energy calculated structure (red) to the 20 lowest energy structures (PDB ID: 2JZZ)¹⁷ (grey), (global backbone rmsd value of 1.98 Å).

12.1.6.2 Comparison between the main conformation (MPD) to the bundle of structures of 1D3Z

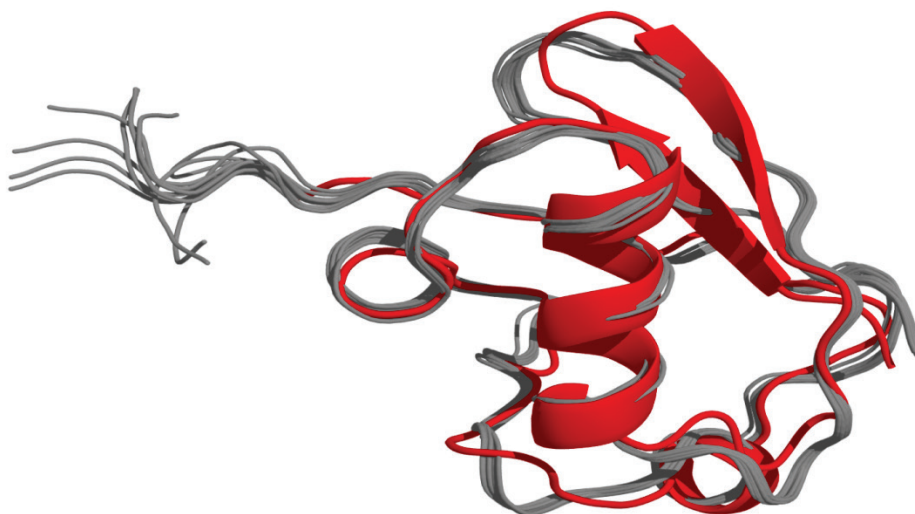


Figure 69 Comparison of the lowest energy calculated structure (red) to the 10 lowest energy solution NMR structures (PDB ID: 1D3Z)⁹⁵ (grey), (global backbone rmsd value of 1.32 Å).

12.2 Ubiquitin PEG

12.2.1 SsNMR experiments (PEG)

Table 11 List of solid-state NMR experiments performed on various samples (PEG)

sample	Type of Experiment	Acquisition Times (ms)	Mixing Time (ms)	Recycling Delay (s)	Spectrum size (ppm)	Number of Scans	Total time	Spectrometer ¹ H larmor freq.
unif-Ubiquitin	1D- ¹³ C	20		2.2	300	16	1m	850 MHz
unif-Ubiquitin	PDS	15 / 15	50	3.0	354 / 119	64	1d 23h	850 MHz
unif-Ubiquitin	NCA	13 / 12		3.5	283 / 60	256	1d 17h	800 MHz
unif-Ubiquitin	NCO	12 / 10		3.0	283 / 60	192	16h	800 MHz
unif-Ubiquitin	NCACX	15 / 14	50	4.2	283 / 60	608	4d 4h	800 MHz
unif-Ubiquitin	NCOCX	12 / 11	50	3.0	283 / 60	768	2d 22h	800 MHz
unif-Ubiquitin	INEPT	10 / 10		1.1	334 / 14	32	2.5h	850 MHz
1glu-Ubiquitin	PDS	14 / 10	100	2.5	356 / 250	96	2d 22h	800 MHz
1glu-Ubiquitin	PDS	12 / 10	400	2.5	331 / 155	96	2d 20h	600 MHz
1glu-Ubiquitin	PDS	15 / 12	850	2.0	354 / 119	320	6d 21h	850 MHz
1glu-Ubiquitin	NCA	12 / 10		2.5	283 / 60	288	19.5h	800 MHz
1glu-Ubiquitin	NCO	12 / 10		2.5	283 / 60	576	1d 15h	800 MHz
1glu-Ubiquitin	¹⁵ N- ¹⁵ N PDS	18 / 14	6	2.0	702 / 200	48	2d	800 MHz
2glu-Ubiquitin	PDS	12.4 / 8.4	100	2.5	249 / 195	128	2d 13h	800 MHz
2glu-Ubiquitin	PDS	15 / 12	400	2.5	354 / 220	176	6d 18h	850 MHz
2glu-Ubiquitin	PDS	12 / 10	800	2.5	331 / 255	160	5d 8h	600 MHz
2glu-Ubiquitin	NCA	12 / 8		2.5	283 / 100	544	2d 1h	800 MHz
2glu-Ubiquitin	NCO	12 / 8.3		2.5	283 / 100	576	1d 18h	800 MHz
2glu-Ubiquitin	¹⁵ N- ¹⁵ N PDS	18 / 16.5	6	2.0	702 / 200	32	1d 13h	800 MHz

12.2.2 Chemical shift lists (PEG)

Table 12 Observed chemical shift of the conformation A of microcrystalline Ubiquitin (PEG)

Chemical shifts are indicated in ppm relative to the DSS. The second and third conformation is highlighted by one or two primes respectively. Resonances marked with “*” symbol are ambiguous.

residue	N	CO	C α	C β	C γ / C γ 2	C δ / C δ 2	C ϵ / C ϵ 2	N δ 2	N ϵ 2
M1	-	-	57.5	38.7	32.3	-	17.7	-	-
Q2	123.6	175.9	53.5	31.2	33.8	179.3	-	-	111.2
I3	114.5	172.6	59.1	41.5	24.7 / 17.8	14.3	-	-	-
F4	118.9	175.2	55.3	40.8	138.7	132.0*	129.0*/131.0*	-	-
V5	121.2	175.1	60.4	34.0	22.1 / 21.1	-	-	-	-
K6	128.3	177.2	55.1	-	-	-	-	-	-
T7	115.3	177.3	60.7	69.9	22.0	-	-	-	-
L8	122.6	-	57.3	-	-	-	-	-	-
T9	-	173.8	61.3	69.0	22.3	-	-	-	-
G10	-	173.7	45.2	-	-	-	-	-	-
K11	-	-	58.4	-	-	-	-	-	-
T12	121.0	174.1	62.5	69.8	22.8	-	-	-	-
I13	128.1	175.0	60.1	40.8	27.0 / 17.5	14.1	-	-	-
T14	122.0	175.0	61.8	69.7	22.4	-	-	-	-
L15	124.8	174.6	52.8	46.3	26.5	27.2 / 24.4	-	-	-
E16	125.2	175.5	55.0	29.3	32.8	183.3	-	-	-
V17	119.5	173.7	58.5	36.5	21.8 / 18.6	-	-	-	-
E18	117.6	174.7	52.9	31.4	36.0	186.2	-	-	-
P19	134.8	176.8	65.7	31.7	27.6	50.5	-	-	-
S20	108.0	174.3	57.1	61.1	-	-	-	-	-
D21	122.7	176.3	56.7	40.6	-	-	-	-	-
T22	108.1	176.3	59.6	70.8	21.9	-	-	-	-
I23	122.2	179.5	62.3	33.9	27.7 / 18.0	9.1	-	-	-
E24	122.3	179.3	60.2	-	-	-	-	-	-
N25	123.9	179.4	55.7	36.9	179.6	-	-	-	105.8
V26	123.9	177.6	67.9	30.8	21.5 / 23.5	-	-	-	-
K27	119.7	180.7	59.2	33.9	26.4	30.6	42.4	-	-
A28	123.0	180.5	55.2	18.0	-	-	-	-	-
K29	121.6	180.4	59.7	33.6	26.0	31.2	42.7	-	-
I30	121.2	178.2	65.7	36.6	30.7 / 17.1	14.8	-	-	-
Q31	123.7	179.0	60.2	28.1	34.3	-	-	-	-
D32	120.3	177.4	57.4	38.3	179.5	-	-	-	-
K33	116.9	177.7	58.6	34.0	25.3	29.0	42.2	-	-
E34	113.8	178.2	55.1	33.0	35.8	181.1	-	-	-
G35	109.5	173.7	46.7	-	-	-	-	-	-
I36	119.2	173.1	57.5	40.2	27.1 / 17.9	13.3	-	-	-
P37	141.4	176.4	61.6	31.8	28.3	51.2	-	-	-
P38	136.1	178.6	66.2	33.2	28.0	51.1	-	-	-

residue	N	CO	C α	C β	C γ / C γ 2	C δ / C δ 2	C ϵ / C ϵ 2	N δ 2	N ϵ 2
D39	114.8	177.0	56.0	39.7	-	-	-	-	-
Q40	117.4	175.1	55.5	29.6	34.5	174.5	-	-	105.7
Q41	118.9	176.0	56.1	31.7	33.3	175.0	-	-	104.7
R42	122.4	174.1	54.8	-	-	-	-	-	-
L43	123.7	175.2	52.8	45.6	27.3	26.7 / 23.8	-	-	-
I44	-	-	57.7	-	-	-	-	-	-
F45	126.1	173.8	56.8	44.1	137.1	130.6* / 132.4*	132.3*	-	-
A46	131.6	177.8	52.4	17.2	-	-	-	-	-
G47	102.3	172.8	45.9	-	-	-	-	-	-
K48	120.5	175.4	54.7	34.7	24.7	29.4	42.1	-	-
Q49	123.0	175.6	55.6	30.2	-	-	-	-	-
L50	126.2	176.8	54.3	41.5	25.8	26.1 / 19.4	-	-	-
E51	122.5	175.5	56.0	31.8	36.5	183.6	-	-	-
D52	121.2	177.5	57.2	38.3	179.0	-	-	-	-
G53	104.4*	175.2	45.2	-	-	-	-	-	-
R54	119.4	175.2	54.4	32.9	27.6	42.9	-	-	-
T55	108.1	176.3	59.6	73.0	22.4	-	-	-	-
L56	116.8	180.5	58.5	39.5	26.4	26.7 / 22.7	-	-	-
S57	113.4	178.3	61.4	62.5	-	-	-	-	-
D58	125.0	177.4	57.5	40.0	180.2	-	-	-	-
Y59	115.5	174.8	58.5	40.3	132.6	-	-	-	-
N60	115.8	174.0	54.4	37.6	178.4	-	-	112.0	-
I61	119.3	174.2	62.4	36.5	27.7 / 17.8	15.4	-	-	-
Q62	124.9	175.7	53.4	31.7	33.3	180.5	-	-	111.1
K63	119.4	175.5	57.4	32.3	24.2	29.9	42.4	-	-
E64	113.6	175.4	58.0	25.0	34.3	-	-	-	-
S65	116.9	171.8	61.2	64.9	-	-	-	-	-
T66	118.1	173.8	63.0	69.6	21.9	-	-	-	-
L67	128.5	175.5	53.9	44.1	29.9	25.3 / 24.8	-	-	-
H68	118.9	173.7	54.9	30.1	-	-	-	-	-
L69	124.9	175.2	53.8	44.3	27.6	26.4 / 23.4	-	-	-
V70	126.6	175.4	60.3	35.2	21.9 / 21.1	-	-	-	-
L71	126.6		53.8						

Table 13 Observed chemicals shift of conformation B of microcrystalline Ubiquitin (PEG)

Chemical shifts are indicated in ppm relative to the DSS. The second and third conformation is highlighted by one or two primes respectively. Resonances marked with "*" symbol are ambiguous.

residue	N	CO	C α	C β	C γ / C γ 2	C δ / C δ 2	C ϵ / C ϵ 2	N δ 2	N ϵ 2
M1	-	-	57.5	38.7	32.3	-	17.7	-	-
Q2	123.6	175.9	53.4	31.2	33.8	179.3	-	-	111.2
I3	113.5	172.6	59.2	41.8	24.7 / 17.9	14.4	-	-	-
F4	118.9	175.3	54.9	41.0	139.5	131.9*	129.0*/131.0*	-	-
V5	120.5	175.1	60.5	34.0	22.1 / 21.1	-	-	-	-
K6	128.3	177.2	55.1	-	-	-	-	-	-
T7	115.3	177.3	60.7	69.9	22.0	-	-	-	-
L8	122.6	-	57.3	-	-	-	-	-	-
T9	-	173.8	61.3	69.0	22.3	-	-	-	-
G10	-	173.7	45.2	-	-	-	-	-	-
K11	-	-	58.4	-	-	-	-	-	-
T12	121.0	174.1	62.5	69.8	22.8	-	-	-	-
I13	128.1	174.7	59.9	40.6	26.3 / 17.7	14.0	-	-	-
T14	122.0	175.0	62.0	70.0	22.4	-	-	-	-
L15	124.7	174.7	52.8	45.9	26.8	27.2 / 24.0	-	-	-
E16	125.2	175.5	55.0	29.3	32.8	183.3	-	-	-
V17	119.5	173.7	58.5	36.5	21.8 / 18.6	-	-	-	-
E18	117.6	174.7	52.9	31.4	36.0	186.2	-	-	-
P19	134.8	176.8	65.7	31.7	27.6	50.5	-	-	-
S20	108.0	174.3	57.1	61.1	-	-	-	-	-
D21	122.7	176.3	56.7	40.6	-	-	-	-	-
T22	108.1	176.3	59.6	70.8	21.9	-	-	-	-
I23	122.5	179.6	62.5	34.2	28.3 / 18.3	9.6	-	-	-
E24	122.3	179.3	60.2	-	-	-	-	-	-
N25	123.9	179.4	55.7	36.9	179.6	-	-	-	105.8
V26	123.9	177.6	67.9	30.8	21.5 / 23.5	-	-	-	-
K27	119.7	180.7	59.2	33.9	26.4	30.6	42.4	-	-
A28	123.0	180.5	55.2	18.0	-	-	-	-	-
K29	121.6	180.4	59.7	33.6	26.0	31.2	42.7	-	-
I30	121.2	178.2	66.0	36.9	31.0 / 17.3	15.7	-	-	-
Q31	123.7	178.9	60.5	27.8	33.9	-	-	-	-
D32	120.3	177.5	57.4	38.7	179.5	-	-	-	-
K33	116.9	177.9	58.6	34.0	25.3	29.0	42.2	-	-
E34	113.1	177.9	55.2	33.3	35.7	181.1	-	-	-
G35	108.6	173.5	46.3	-	-	-	-	-	-
I36	119.2	173.1	57.5	40.2	27.1 / 17.9	13.3	-	-	-
P37	141.4	176.4	61.6	31.8	28.3	51.2	-	-	-
P38	136.1	178.6	66.2	33.2	28.0	51.1	-	-	-

residue	N	CO	C α	C β	C γ / C γ 2	C δ / C δ 2	C ϵ / C ϵ 2	N δ 2	N ϵ 2
D39	114.8	177.0	56.0	39.7	-	-	-	-	-
Q40	117.4	175.1	55.5	29.6	34.5	174.5	-	-	105.7
Q41	118.9	176.0	56.1	31.7	33.3	175.0	-	-	104.7
R42	122.4	174.1	54.8	-	-	-	-	-	-
L43	124.1	175.2	52.9	45.1	27.4	26.6/ 24.0	-	-	-
I44	122.5	175.8	58.6	41.2	28.0 / 17.9	12.8	-	-	-
F45	126.1	174.3	56.2	44.1	137.1	130.6*/ 132.4*	132.3*	-	-
A46	134.6	177.8	52.5	16.4	-	-	-	-	-
G47	102.3	172.8	45.9	-	-	-	-	-	-
K48	120.5	175.4	54.7	34.7	24.7	29.4	42.1	-	-
Q49	123.0	175.6	55.6	30.2	-	-	-	-	-
L50	126.2	176.8	54.2	42.3	25.8	26.0 / 19.0	-	-	-
E51	122.5	175.5	56.0	31.8	36.5	184.0	-	-	-
D52	121.2	177.5	57.2	38.3	179.0	-	-	-	-
G53	104.4*	175.2	45.2	-	-	-	-	-	-
R54	119.4	175.2	54.4	32.9	27.6	42.9	-	-	-
T55	108.1	176.3	59.6	73.0	22.4	-	-	-	-
L56	116.8	180.5	58.5	39.5	26.4	26.7 / 22.7	-	-	-
S57	113.4	178.3	61.4	62.5	-	-	-	-	-
D58	125.0	177.4	57.5	40.0	180.2	-	-	-	-
Y59	115.5	174.8	58.5	40.3	132.6	-	-	-	-
N60	115.8	174.0	54.4	37.6	178.4	-	-	112.0	-
I61	119.0	174.2	62.6	36.7	28.7 / 17.8	15.0	-	-	-
Q62	124.9	175.7	53.4	31.7	33.3	180.5	-	-	111.1
K63	119.4	175.5	57.4	32.3	24.2	29.9	42.4	-	-
E64	113.6	175.4	58.0	25.0	34.3	-	-	-	-
S65	116.0	172.2	61.5	64.9	-	-	-	-	-
T66	118.1	173.8	62.6	70.0	21.6	-	-	-	-
L67	128.0	175.5	53.9	44.1	29.9	25.3 / 24.8	-	-	-
H68	118.9	173.7	54.9	30.1	-	-	-	-	-
L69	125.2	175.2	54.2	43.8	27.6	26.5 / 23.8	-	-	-
V70	126.6	175.4	60.3	35.2	21.9 / 21.1	-	-	-	-
L71	126.6		53.8						

12.2.3 Network distance restraints between I3C γ 2 and V26C α (PEG)

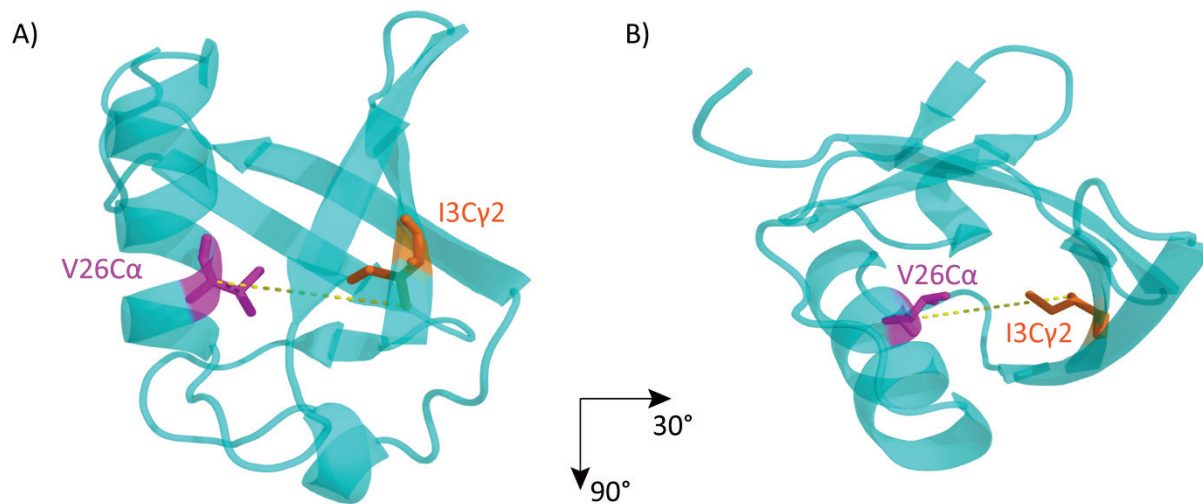


Figure 70 Illustration of the ^{13}C - ^{13}C correlation between I3C γ 2 (orange) and V26C α (magenta). The contact is highlighted by a yellow dashed line in the X-ray structure 1UBQ.⁴⁹

12.2.4 Secondary structure analysis of the conformation B (PEG)

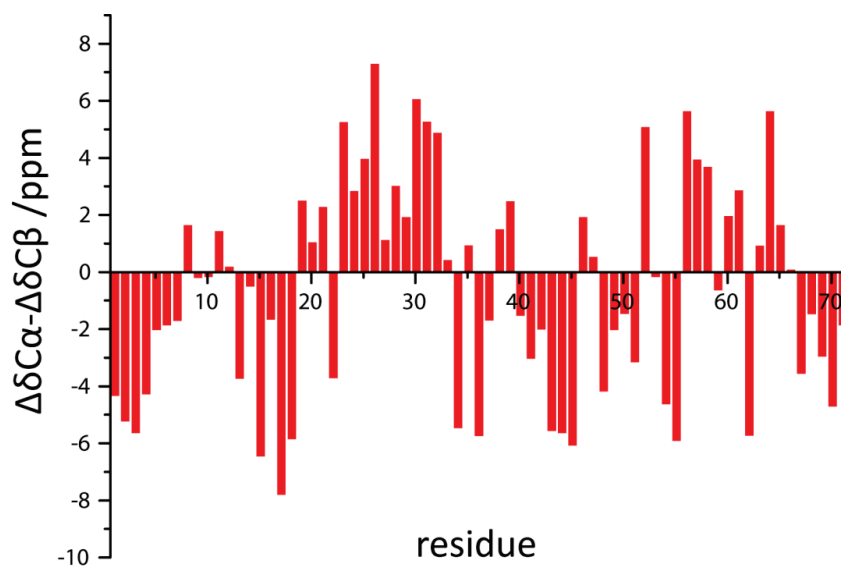


Figure 71 Secondary chemical shift analysis of conformation B of Ubiquitin (PEG).

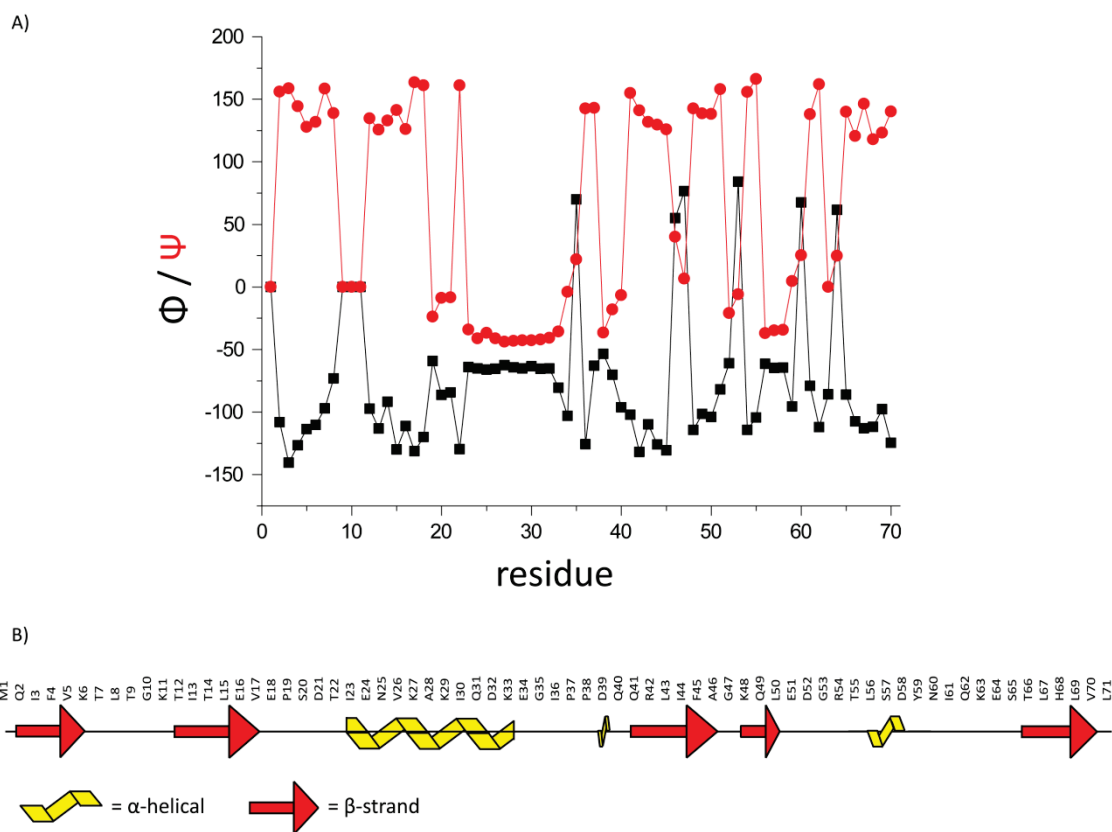


Figure 72 A) TALOS+ analysis, predicting Phi (black) and Psi (red) dihedral torsion angles of conformation B B) cartoon of the result of TALOS+ prediction.

12.2.5 Validation results of Ubiquitin (PEG)

12.2.5.1 Validation result of conformation A (PEG)

Table 14 Structural statistics for the 10 lowest-energy conformers of conformation A (PEG)

Conformers were calculated with XPLOR-NIH using 498 distance restraints with a defined distance range of 1-7 Å.

Summary of conformationally-restricting experimental constraints ^a

Distance constraints:

Total	498
intra-residue [i = j]	0
sequential [i - j = 1]	0
medium range [1 < i - j < 5]	264
long range [i - j ≥ 5]	234
Constraints per restrained residue ^b	7.0
Dihedral-angle constraints:	128
Total number of restricting constraints ^b	626
Total number of restricting constraints per restrained residue ^b	8.8
Restricting long-range constraints per restrained residue ^b	3.3
Total structures computed	1000
Number of structures used	10

Residual constraint violations ^{a,c}

Distance violations / structure	
0.1 - 0.2 Å	12.8
0.2 - 0.5 Å	5.2
> 0.5 Å	0.1
RMS of distance violation / constraint	0.03 Å
Maximum distance violation ^d	0.51 Å

Dihedral angle violations / structure

1 - 10 °	34
> 10 °	4.3
RMS of dihedral angle violation / constraint	4.98 °
Maximum dihedral angle violation ^d	35.80 °

Rmsd Values

	all	ordered ^e
All backbone atoms	0.7 Å	0.5 Å
All heavy atoms	1.4 Å	1.0 Å

Structure Quality Factors - overall statistics

	Mean score	SD	Z-score ^g
Procheck G-factor ^e (phi / psi only)	-0.47	N/A	-1.53
Procheck G-factor ^e (all dihedral angles)	-0.76	N/A	-4.49
Verify3D	0.32	0.04	-2.25
Prosall (-ve)	0.64	0.06	-0.04
MolProbity clashscore	56.18	9.47	-8.11
Ramachandran Plot Summary from Procheck ^f			
Most favoured regions	86.5%		
Additionally allowed regions	10.6%		
Generously allowed regions	2.4%		
Disallowed regions	0.5%		
Ramachandran Plot Statistics from Richardson's lab			
Most favoured	91.1%		
Allowed regions	3.4%		
Disallowed regions	5.4%		

^a Analysed for residues 1 to 72

^b There are 71 residues with conformationally restricting constraints

^c Calculated for all constraints for the given residues, using average r^{-6}

^d Largest constraint violation among all the reported structures

^e Residues with sum of phi and psi order parameters > 1.8

Ordered residue ranges:

^f Residues selected based on: all residues

Selected residue ranges: all

^g With respect to mean and standard deviation for a set of 252 X-ray structures < 500 residues, of resolution $\leq 1.80 \text{ \AA}$, R-factor ≤ 0.25 and R-free ≤ 0.28 ; a positive value indicates a 'better' score

Generated using PSVS 1.5⁷⁶

12.2.5.2 Validation result of conformation B (PEG)

Table 15 Structural statistics for the 10 lowest-energy conformers of conformation B (PEG)

Conformers were calculated with XPLOR-NIH using 487 distance restraints with a defined distance range of 1-7 Å.

Summary of conformationally-restricting experimental constraints ^a

Distance constraints:

Total	487
intra-residue [i = j]	0
sequential [i - j = 1]	0
medium range [1 < i - j < 5]	260
long range [i - j ≥ 5]	227
Constraints per restrained residue ^b	6.9
Dihedral-angle constraints:	132
Total number of restricting constraints ^b	619
Total number of restricting constraints per restrained residue ^b	8.7
Restricting long-range constraints per restrained residue ^b	3.2
Total structures computed	1000
Number of structures used	10

Residual constraint violations ^{a,c}

Distance violations / structure	
0.1 - 0.2 Å	13.9
0.2 - 0.5 Å	4.6
> 0.5 Å	0.2
RMS of distance violation / constraint	0.03 Å
Maximum distance violation ^d	0.58 Å
Dihedral angle violations / structure	
1 - 10 °	37.8
> 10 °	9.9
RMS of dihedral angle violation / constraint	5.68 °
Maximum dihedral angle violation ^d	36.10 °

Rmsd Values

	all	ordered ^e
All backbone atoms	0.7 Å	0.6 Å
All heavy atoms	1.4 Å	1.1 Å

Structure Quality Factors - overall statistics

	Mean score	SD	Z-score ^g
Procheck G-factor ^e (phi / psi only)	-0.49	N/A	-1.61
Procheck G-factor ^e (all dihedral angles)	-0.77	N/A	-4.55
Verify3D	0.37	0.03	-1.44
Prosall (-ve)	0.62	0.05	-0.12
MolProbity clashscore	62.74	8.10	-9.24
Ramachandran Plot Summary from Procheck ^f			
Most favoured regions	85.6%		
Additionally allowed regions	11.6%		
Generously allowed regions	2.4%		
Disallowed regions	0.2%		
Ramachandran Plot Statistics from Richardson's lab			
Most favoured	90.4%		
Allowed regions	4.6%		
Disallowed regions	5.0%		

^a Analysed for residues 1 to 72

^b There are 71 residues with conformationally restricting constraints

^c Calculated for all constraints for the given residues, using average r^{-6}

^d Largest constraint violation among all the reported structures

^e Residues with sum of phi and psi order parameters > 1.8

Ordered residue ranges:

^f Residues selected based on: all residues

Selected residue ranges: all

^g With respect to mean and standard deviation for a set of 252 X-ray structures < 500 residues, of resolution $\leq 1.80 \text{ \AA}$, R-factor ≤ 0.25 and R-free ≤ 0.28 ; a positive value indicates a 'better' score

Generated using PSVS 1.5⁷⁶

12.2.6 Comparison between conformation A (PEG) to the bundle of structures of 1D3Z

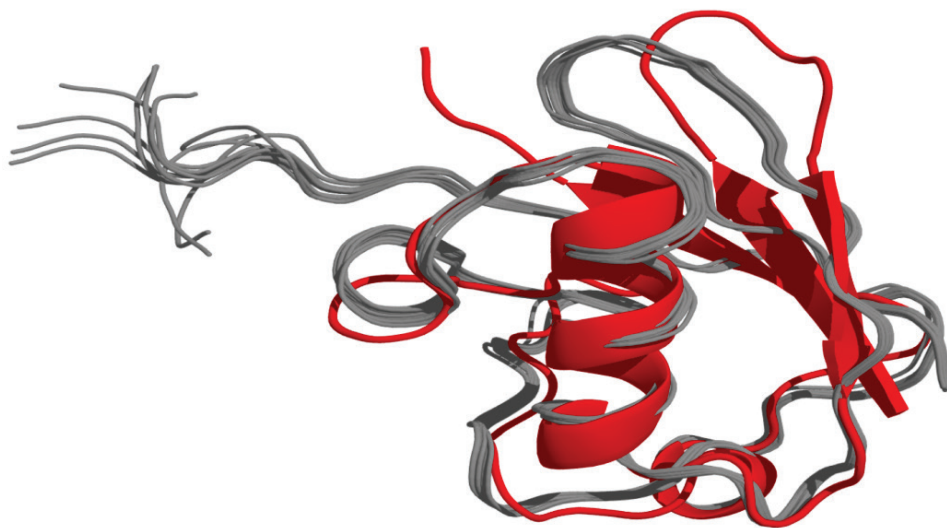


Figure 73 Comparison of the lowest energy calculated structure of conformation A (PEG) (red) to the 10 lowest energy solution NMR structures (PDB ID: 1D3Z)⁹⁵ (grey), (global backbone rmsd value of 1.81 Å).

Table 16 Backbone comparison between the main conformation of MPD and the conformation A of PEG

residue	ΔN	ΔCO	$\Delta C\alpha$
1	-	-	2.3
2	-1.4	0.8	-0.7
3	-0.7	0.3	0.3
4	-0.3	-0.1	0.6
5	0.2	-0.4	0.4
6	2.1	0	1.7
7	2.5	-0.4	0.4
8	3.3	-	0
9	-	-1.8	0.3
10	-	0.3	0.3
11	-	-	1.9
12	1.2	-0.3	-0.3
13	-0.5	-0.2	0.3
14	-2.4	1.0	-0.8
15	-0.4	0.5	0.3
16	1.8	0.4	0.8
17	0.7	-1.4	-0.1
18	-1.4	-0.6	-1.8
19	1.1	-0.7	-0.1
20	-0.8	1.0	0.0
21	0.4	-0.3	0.0
22	1.5	0.0	-0.1
23	0.5	0.3	0.5
24	-1.1	0.4	1.2
25	0.3	-1.0	0.4
26	-0.5	0.2	0.2
27	-0.3	-0.1	0.2
28	0.8	0.6	0.2
29	3.0	-0.4	0.4
30	-1.5	0.7	-0.7
31	0.9	-0.2	0.7
32	-1.1	-0.2	-0.5
33	-3.1	-0.4	-0.9
34	1.8	0.8	0.1
35	0.3	-0.2	0.7
36	-0.5	-0.4	-0.2
37	-1.2	-0.3	-0.2
38	-0.5	0.2	-0.2
39	-0.2	-0.3	0.6
40	-0.4	-0.8	0.3
41	-1.8	1.3	0.2
42	-3.3	0.6	-0.4
43	-3.3	-0.4	-0.5

residue	ΔN	ΔCO	$\Delta C\alpha$
44	-	-	-0.5
45	2.4	0.2	0.4
46	-4.1	0.5	0.2
47	-1.4	-0.7	0.3
48	-2.0	0.4	-0.6
49	-0.9	-0.1	-0.3
50	2.3	0.4	0.2
51	-1.0	-1.6	0.1
52	2.6	1.6	-0.5
53	3.5	0.3	-0.4
54	-1.4	0.2	1.0
55	0.5	-0.1	0.0
56	-0.6	0.2	0.2
57	1.2	-0.1	0.6
58	-0.1	1.3	-0.1
59	2.5	-1.2	1.4
60	-1.6	-0.1	0.7
61	0.6	0.1	-0.5
62	-0.9	0.5	-0.2
63	0.0	-1.1	-1.0
64	-1.3	0.5	-0.2
65	1.2	-0.2	-0.1
66	-0.1	0.9	0.5
67	-0.4	0.0	0.4
68	1.1	-1.9	0.4
69	-1.1	-1.2	-1.4
70	14.4	-0.8	1.4
71	6.9	-	-

12.3 Pulse sequence of 2D NCA and NCO experiments

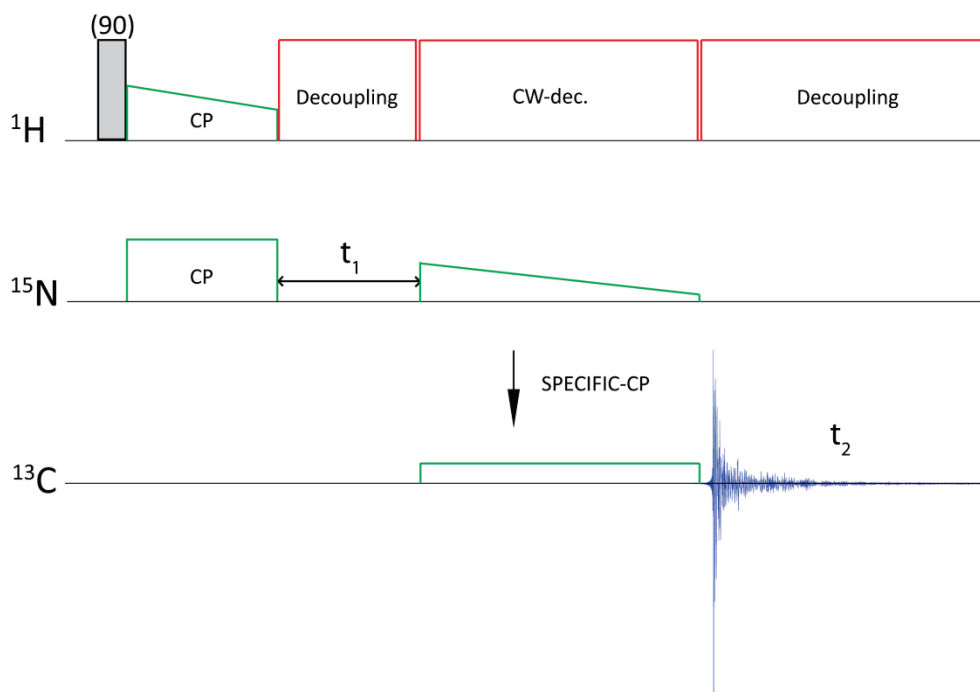


Figure 74 Illustration of the heteronuclear correlation pulse sequence to record 2D NCA and NCO experiments. The second magnetization transfer from ^{15}N to ^{13}C , highlighted by an arrow, is based on the SPECIFIC-CP magnetization transfer.

12.4 Pulse sequence of 2D NCACX and NCOCX experiments

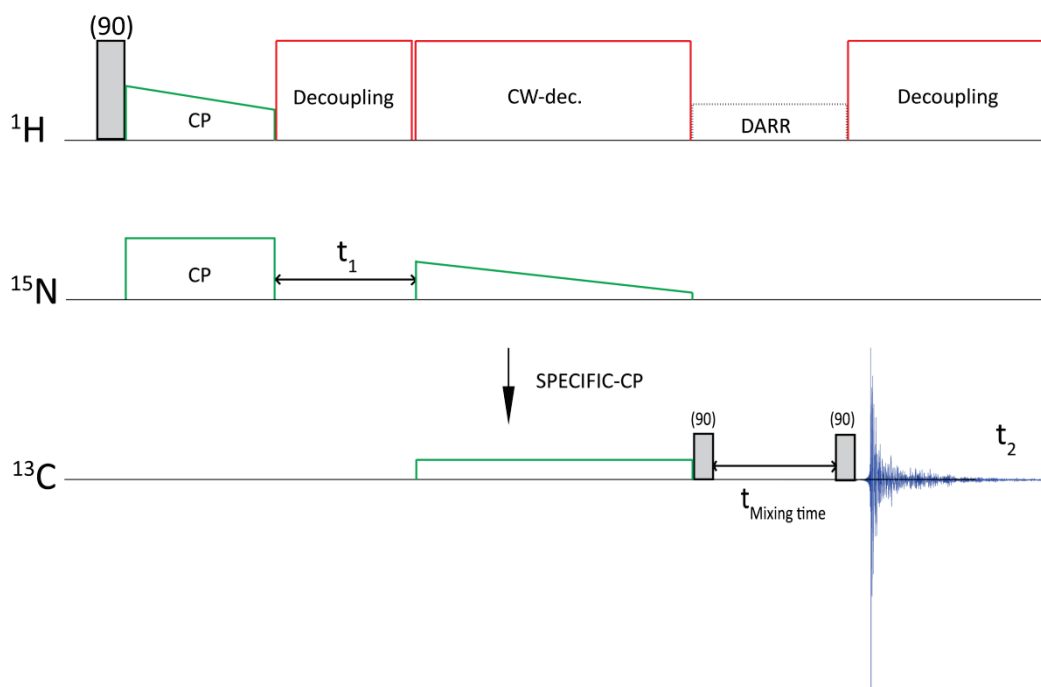


Figure 75 Illustration of the 2D heteronuclear correlation pulse sequence to recorded 2D NCACX and NCOCX experiments with DARR ^{13}C - ^{13}C correlation transfer.

12.5 Structure calculation

The detailed description of the different energy terms used for the total force field energy are listed as follows:

12.5.1 Covalent bond energy term E_{bond}

$$E_{\text{bond}} = \sum_{\text{bonds}} k_b (r - r_0)^2 \quad (\text{Eq.34})$$

with k_b = bond force constants, r_0 = equilibrium bond length (between specified nuclei),
 r = actual bond length

12.5.2 Angle energy term E_{angle}

$$E_{\text{angles}} = \sum_{\text{angles}} k_{\theta} (\theta - \theta_0)^2 \quad (\text{Eq.35})$$

with k_{θ} = angle force constants, θ_0 =equilibrium constants (between specified nuclei),
 θ = actual value of the angle

12.5.3 Dihedral energy term E_{dihedral}

In xplor-NIH program, the total dihedral energy term E_{dihedral} is separated into two terms.

$$E_{\text{dihedral}} = E_{\text{dihe}} + E_{\text{impr}} \quad (\text{Eq.36})$$

$$E_{\text{dihe}} = \sum_{\text{dihedrals}} \sum_{i=1, m} \begin{cases} k_{\phi_i} (1 + \cos (n\phi_i + \delta_i)) & \text{if } n_i > 0 \\ k_{\phi_i} (\phi_i + \delta_i)^2 & \text{if } n_i = 0 \end{cases} \quad (\text{Eq.37})$$

$$E_{\text{impr}} = \sum_{\text{impropers}} \sum_{i=1, m} \begin{cases} k_{\phi_i} (1 + \cos (n\phi_i + \delta_i)) & \text{if } n_i > 0 \\ k_{\phi_i} (\phi_i + \delta_i)^2 & \text{if } n_i = 0 \end{cases} \quad (\text{Eq.38})$$

with k_{ϕ_i} = dihedral force constants, ϕ_i = actual torsion angle, n_i = periodicities, m_i = multiplicities,
 δ_i = phase shifts

According to historical reasons the term E_{impr} is mostly used with $n = 0$ to maintain planarity or chirality whereas the term E_{dihe} is used for $n > 0$ to describe multi-minima for torsion potentials.

12.5.4 Van der Waal energy term E_{vdw}

For the xplor-NIH force field (parallhdg.pro) which is used in this work the van der Waals interaction term will be calculated with the so called repel function which exclude attraction- and electrostatic interactions terms.

$$E_{vdw} = \sum f_{vdw}(\mathbf{R}) \quad (\text{Eq.39})$$

$$f_{vdw}(\mathbf{R}) = C_{rep}[\max(0, (k^{rep} \mathbf{R}_{min})^{irexp} - \mathbf{R}^{irexp})]^{rexp} \quad (\text{Eq. 40})$$

with C_{rep} = force constant, $\mathbf{R}_{min} = \sigma \sqrt[6]{2}$, σ = Lennard-Jones constant (finite distance between two nuclei where the inter-particle potential is zero).

Parameter setting for E_{vdw} :

repel = 0.8

$C_{radius} = 1.5 \text{ \AA}$

$H_{radius} = 1.0 \text{ \AA}$

$N_{radius} = 1.35 \text{ \AA}$

rexp = 2

irexp = 2

rcon = 1

nbxmod = 3 (exclude nonbonded interactions between bonded atoms and atoms that are bonded to a common third atom)

wmin = 0.01 \AA (specific the threshold distance for close contact warnings, a warning is used when a pair of atoms gets closer than this distance unless the nonbonded interaction is excluded by nbxmod)

cutnb = 6.0 \AA (specifies the nonbonded interaction cutoff for the nonbonded list generation)

tolerance = 1.5 \AA (specifies how far atoms are allowed to move before the hydrogen-bond list gets updated)

12.5.5 electrostatic energy term E_{el}

$$f_{el}(R) = \begin{cases} Q_i Q_j \frac{C}{\epsilon_0 R} \text{ heavy } (R - R_{cut}), & \text{for pur truncation} \\ Q_i Q_j \frac{C}{\epsilon_0 R} \left(1 - \frac{R^2}{R_{off}^2}\right)^2, & \text{for shifted option} \\ Q_i Q_j \frac{C}{\epsilon_0 R^2} \text{ SW } (R, R_{on}, R_{off}), & \text{for } \frac{1}{R} \text{ option} \\ 0, & \text{for repel option} \end{cases}$$

Parameter setting for E_{el} :

$Q_i Q_j$ = electric charge of the atoms

C = scaling factor

ϵ_0 = electric permittivity of free space

R = distance

12.5.6 Distance restraint energy term $E_{\text{experimental}}$

For the distance restraints interaction term $E_{\text{experimental}}$ the soft square function is used

$$E_{\text{experimental}} = \min(\text{ceil}, SC) \begin{cases} a + \frac{b}{\Delta^{\text{softexp}}} + c\Delta & \mathfrak{R} > d + d_{\text{plus}} - d_{\text{off}} + r_{\text{sw}} \\ \Delta^{\text{exp}} & \mathfrak{R} < d + d_{\text{plus}} - d_{\text{off}} + r_{\text{sw}} \end{cases} \quad (\text{Eq.41})$$

S, C= scaling factors

\mathfrak{R} = distance between selected stets of atoms

c = slope of the asymptote

r_{sw} = specified by the switching function

definition of Δ :

$$\Delta = \begin{cases} \mathfrak{R} - (d + d_{\text{plus}} - d_{\text{off}}), & d + d_{\text{plus}} - d_{\text{off}} < \mathfrak{R} \\ 0, & d - d_{\text{minus}} < \mathfrak{R} < d + d_{\text{plus}} - d_{\text{off}} \\ (d - d_{\text{minus}} - \mathfrak{R}), & \mathfrak{R} < d - d_{\text{minus}} \end{cases} \quad (\text{Eq.42})$$

Parameter setting for $E_{\text{experimental}}$ (distance restraint range of (1-7 Å):

$d = 5 \text{ \AA}$

$d_{\text{plus}} = 2 \text{ \AA}$

$d_{\text{minus}} = 4 \text{ \AA}$

$d_{\text{off}} = 0 \text{ \AA}$

softexp = 1.0

$r_{\text{sw}} = 0.5$

c = 1.0

$\mathfrak{R} = \text{center} (R = (r_{\text{center}}^1 - r_{\text{center}}^2))$

exp = 2

12.5.7 Inputfile of the x-plor-NIH calculation

```
eval ($numStructs = 1000) ! total number of structures to calculate
cpyth "from os import environ as env"
cpyth "xplor.command('eval ($proc_num=%s)' % env['XPLOR_PROCESS'])"
cpyth "xplor.command('eval ($num_procs=%s)' % env['XPLOR_NUM_PROCESSES'])"
eval ($firstStruct = ($proc_num * $numStructs) / $num_procs)
eval ($lastStruct = (($proc_num+1) * $numStructs) / $num_procs)
!evaluate ($PARAMETERS = "@/progs/xplor-nih-2.30/toppar/parallhdg.pro")
evaluate ($STRUCTURE = "ubiq.psf" )
evaluate ($TEMPLATE = "gen_temp_ubi.pdb" )
evaluate ($NOE_TABLE1 = "distance-restraints.tbl" )
evaluate ($NOE_TABLE2 = "" )
evaluate ($DIHE_TABLE = "TALOS.tbl" )
evaluate ($PDBNAME = "ubi" )
evaluate ($tad_temp = 30000 )
evaluate ($tad_step = 4000 )
evaluate ($tad_vdw = 0.1 )
evaluate ($tad_noe = 150 )
evaluate ($tad_dih = 100 )
evaluate ($cool_t = 30000 )
evaluate ($cool_steps = 4000 )
evaluate ($cool_noe = 150 )
evaluate ($cool_dih = 100 )
evaluate ($vercool_temp = 1000 )
evaluate ($vercool_steps = 2000 )
evaluate ($vercool_noe = 150 )
evaluate ($vercool_dih = 100 )
evaluate ($pow_noe = 50 )
evaluate ($pow_dih = 200 )
evaluate ($pow_step = 300 )
evaluate ($end_count = 1000)
!parameter
! @$PARAMETERS
!parameter
! @/progs/xplor-nih-2.30/toppar/protein.par
!!!cluster
evaluate ($end_count = $lastStruct)
parameter
@/progs/xplor-nih-2.30/toppar/protein.par
end
structure
  @$STRUCTURE
end
coor @$TEMPLATE
coor copy end
```



```

set echo off message off end
noe
  nres = 10000
  class all
  @$NOE_TABLE1
end
set echo off message off end
restraints dihedral
  nass = 1000
  @$DIHE_TABLE
end
set echo on message on end
flags exclude * include bonds angle impr vdw noe cdih end
vector do (fbeta=10) (all)
vector do (mass=100) (all)

```

noe

```

ceiling=1000
averaging * cent
potential * soft
sqoffset * 0.0
sqconstant * 1.0
sqexponent * 2
soexponent * 1
asymptote * 1.0
rswitch * 0.5
end
parameter

```

nbonds

```

repel=0.80
rexp=2 irexp=2 rcon=1.
nbxmod=3
wmin=0.01
cutnb=6.0 ctonnb=2.99 ctofnb=3.
tolerance=1.5
end
end
set seed = 3 end
evaluate ($structure_number = 0)
evaluate ($accept_count = 0)
while ($end_count > $accept_count) loop main
  evaluate ($structure_number = $structure_number + 1)
  coor swap end
  coor copy end
  vector do ( vx = maxwell(1.5) ) ( all )

```

```

vector do ( vy = maxwell(1.5) ) ( all )
vector do ( vz = maxwell(1.5) ) ( all )
noe
  scale * $stad_noe
end
restraints dihedral
  scale = $stad_dih
end
constraints interaction (all) (all)
  weights * 1 vdw $stad_vdw end
end
parameter
  bonds ( name SG ) ( name SG ) 100. TOKEN
  angle ( name CB ) ( name SG ) ( name SG ) 50. TOKEN
end
dynamics torsion
  topology
    kdihmax = 300
    maxchn = 1000
    fix group ( resn PRO and not ( name c or name o ) )
    for $ss_rm_id_1 in id ( name SG ) loop SSRM
      for $ss_rm_id_2 in id ( name SG ) loop SSR2
        if ( $ss_rm_id_2 > $ss_rm_id_1 ) then
          free bond ( id $ss_rm_id_1 ) ( id $ss_rm_id_2 )
        end if
      end loop SSR2
    end loop SSRM
!   evaluate ( $nucl = 1 )
!   while ( $nucl le 24 ) loop cut
!     free bond ( resid $nucl and name C3' )
!       ( resid $nucl and name C4' )
!     evaluate ( $nucl = $nucl + 1 )
!   end loop cut
  end
  ntrfrq = 1
  nstep = $stad_step timestep = 0.015
  tcoupling = true tbath = $stad_temp nprint = 50
end
noe
  scale * $cool_noe
end
restraints dihedral
  scale = $cool_dih
end
parameter
  bonds ( name SG ) ( name SG ) 1000. TOKEN

```

```

    angle ( name CB ) ( name SG ) ( name SG ) 500. TOKEN
end
flag exclude * include noe cdih bond angl impr vdw end
evaluate ($final_t = 1000)
evaluate ($tempstep = 100)
evaluate ($ncycle = int(($cool_t-$final_t)/$tempstep))
evaluate ($nstep = int($cool_steps/$ncycle))
evaluate ($ini_vdw = $stad_vdw)
evaluate ($fin_vdw = 1.0)
evaluate ($vdw_step = ($fin_vdw-$ini_vdw)/$ncycle)
evaluate ($bath = $cool_t)
evaluate ($k_vdw = $ini_vdw)
evaluate ($i_cool = 0)
while ($i_cool < $ncycle) loop cool
    evaluate ($i_cool = $i_cool + 1)
    constraints interaction (all) (all)
    weights * 1. vdw $k_vdw end
end
dynamics torsion
    ntrfrq = 1
    nstep = $nstep timestep = 0.015
    tcoup = true tbath = $bath nprint = $nstep
end
evaluate ($bath = $bath - $tempstep)
evaluate ($k_vdw = $k_vdw + $vdw_step)
end loop cool
dynamics torsion
    topology
    reset
end
end
vector do ( vx = maxwell($vercool_temp) ) ( all )
vector do ( vy = maxwell($vercool_temp) ) ( all )
vector do ( vz = maxwell($vercool_temp) ) ( all )
noe
    scale * $vercool_noe
end
restraints dihedral
    scale = $vercool_dih
end
evaluate ($final_t = 300)
evaluate ($tempstep = 50)
evaluate ($ncycle = ($vercool_temp - $final_t)/$tempstep)
evaluate ($nstep = int($vercool_steps/$ncycle))
evaluate ($bath = $vercool_temp)
evaluate ($i_cool = 0)

```

```

while ($i_cool < $ncycle) loop cool
  evaluate ($i_cool=$i_cool+1)
  dynamics verlet
  nstep = $nstep time = 0.003
  iasvel = current firstt =$bath
  tcoup = true tbath = $bath
  nprint = $nstep iprfreq = 0
end
  evaluate ($bath = $bath - $tempstep)
end loop cool
noe
  scale * $pow_noe
end
restraints dihedral
  scale = $pow_dih
end

minimize powell nstep=$pow_step drop=10.0 nprint=25 end
evaluate ( $accept = 0 )
print threshold = 200 noe
! print threshold = 20 noe
evaluate ($noe5 = $violations)
print threshold = 0.5 noe
evaluate ($noe0 = $violations)
evaluate ($rms_noe = $result)
if ($noe5 > 0) then evaluate ( $accept = $accept + 1 ) end if
print threshold = 100.0 cdih
! print threshold = 50.0 cdih
evaluate ($cdih5 = $violations)
print threshold = 0.0 cdih
evaluate ($cdih0 = $violations)
evaluate ($rms_cdih = $result)
if ($cdih5 > 0) then evaluate ( $accept = $accept + 1 ) end if
print thres=2.5 bond
! print thres=0.5 bond
evaluate ($bond5=$violations)
evaluate ($rms_bond = $result)
if ($rms_bond > 0.02) then evaluate ( $accept = $accept + 1 ) end if
print thres=50.0 angle
evaluate ($angle5=$violations)
evaluate ($rms_angle = $result)
if ($rms_angle > 4.0) then evaluate ( $accept = $accept + 1 ) end if
print thres=50.0 improper
! print thres=50.0 improper
evaluate ($improper5=$violations)
evaluate ($rms_improper = $result)

```

```

if ($rms_improper > 4.0) then evaluate ( $accept = $accept + 1 ) end if
energy end
if ($accept = 0 ) then
  evaluate ($accept_count=$accept_count+1)
  remarks Accepted structure $accept_count of $structure_number structures
  remarks =====
  remarks      noe, cdih, bonds, angles, impropr
  remarks violations.: $noe5,$cdih5,$bond5,$angle5,$improper5
  remarks rmsd    .: $rms_noe,$rms_cdih,$rms_angle,$rms_bond,$rms_improper
  remarks 0-viol  .: $noe0,$cdih0
  remarks =====
  remarks overall = $ener
  remarks noe = $NOE
  remarks dih = $CDIH
  remarks vdw = $VDW
  remarks bon = $BOND
  remarks ang = $ANGL
  remarks imp = $IMPR
  remarks =====
  evaluate ($filename= $PDBNAME + "_" + encode($accept_count) + ".pdb")
  write coordinates output = $filename end
end if
end loop main
stop

```

13 Appendix II

13.1 Comparison between 1D ^{13}C spectra of $[1\text{-}^{13}\text{C}]$ -glc and $[\text{U-}^{15}\text{N-}[\text{ILV-}^{13}\text{C Methyl}]]$ -labeled sample

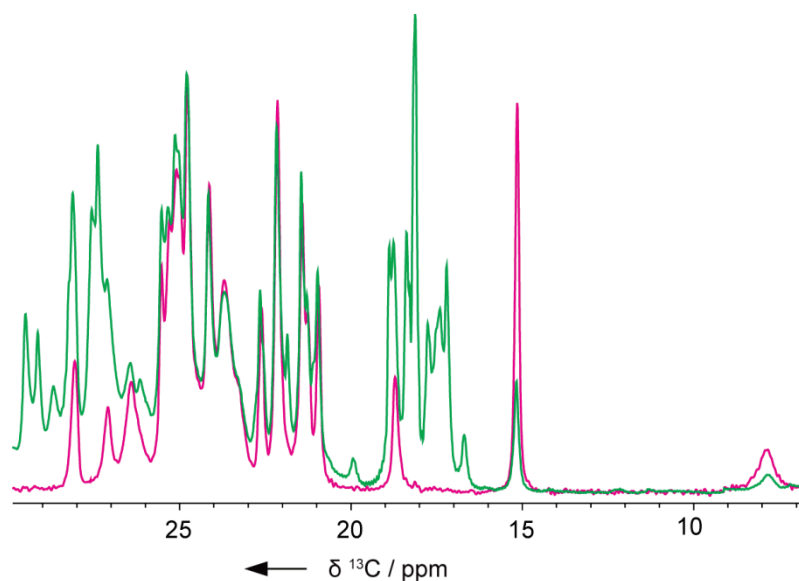


Figure 76 Excerpts of 1D CP ^{13}C spectra of $[1\text{-}^{13}\text{C}]$ -glc (green) and $[\text{U-}^{15}\text{N-}[\text{ILV-}^{13}\text{C Methyl}]]$ -labeled sample (magenta), For qualitative comparison, the spectra are scaled by equalizing the signal intensity of resonance at 25ppm.

13.2 Comparison between 1D ^{13}C spectra of $[\text{U-}^{13}\text{C}]$ -glc and $[\text{U-}^{15}\text{N-}[\text{ILV-}^{13}\text{C Methyl}]]$ -labeled sample

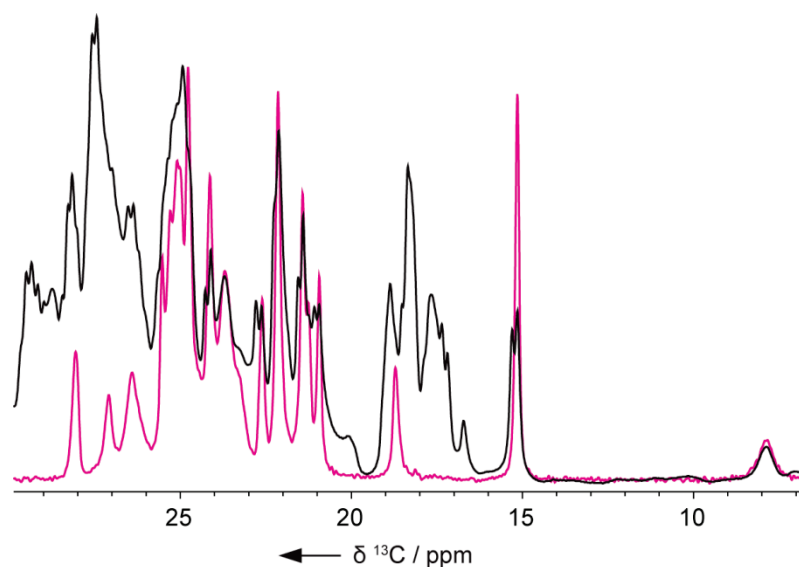


Figure 77 Excerpts of 1D CP ^{13}C spectra of $[\text{U-}^{13}\text{C}]$ -glc (black) and $[\text{U-}^{15}\text{N-}[\text{ILV-}^{13}\text{C Methyl}]]$ -labeled samples (magenta). For qualitative comparison, the spectra are scaled by equalizing the signal intensity of resonance at 25ppm.

13.3 Column projection of the ^{13}C chemical shift dimension

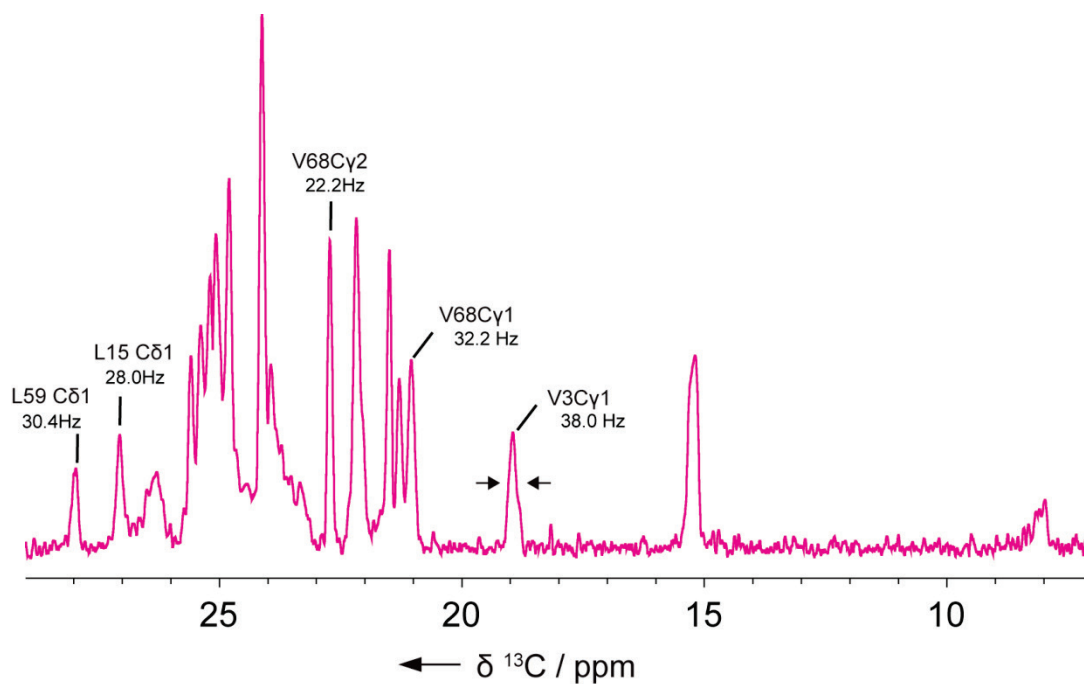
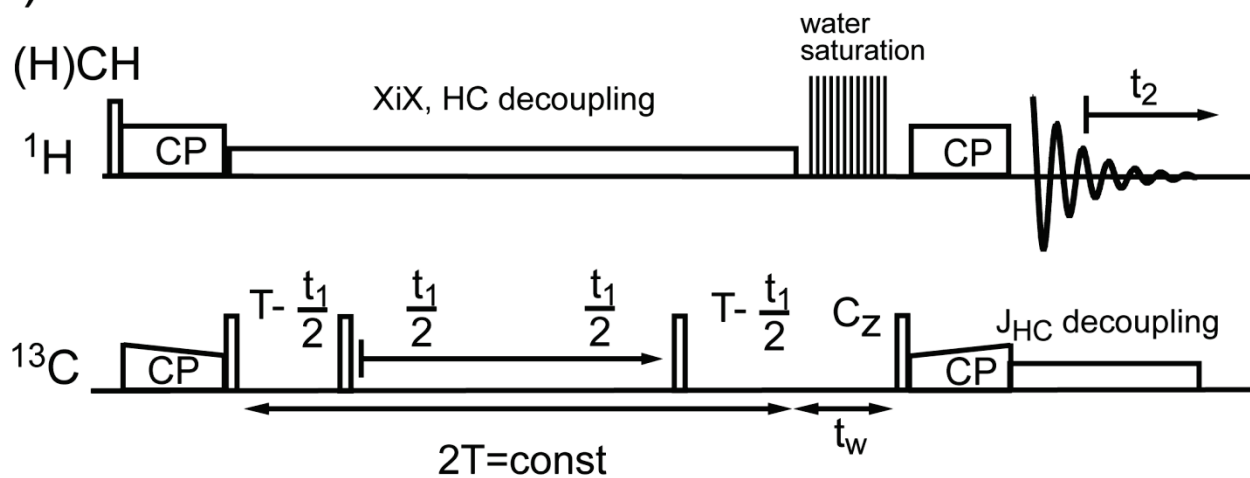


Figure 78 Column projection of the ^{13}C chemical shift dimension from the 2D (H)CH-spectrum (Figure 65.A), The full width at half maximum (FWHM) is indicated for isolated leucine and valine methyl groups. A sine squared window functions with sine bell shift of 47° was used to process the ^{13}C dimension.

13.4 Pulse sequences for ^1H detection ssNMR experiments

A)



B)

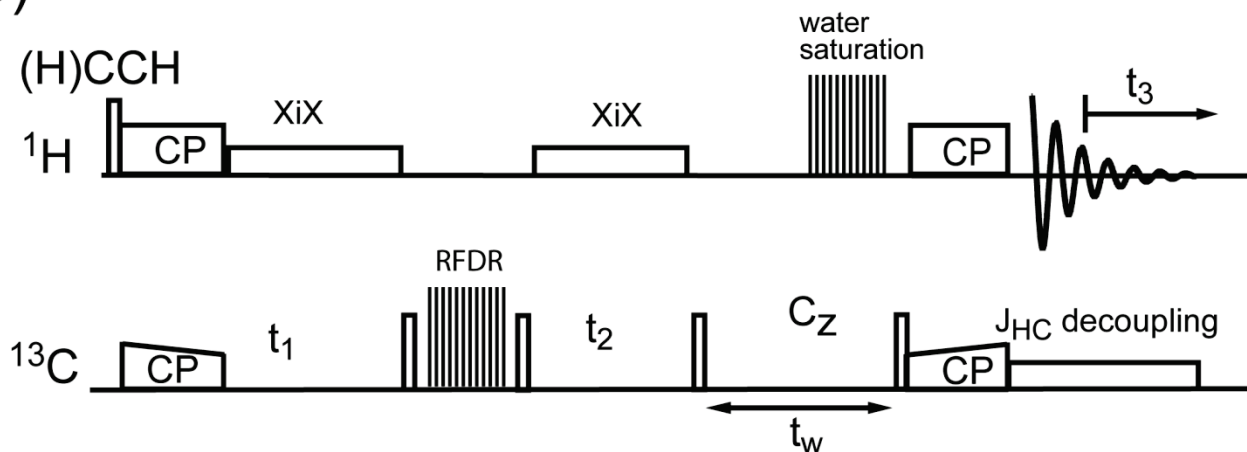


Figure 79 Pulse sequences to obtain A) 2D (H)CH and B) 3D (H)CCH correlations. Open bars represent 90° pulses. CP and RFDR are used for HC and CC magnetization transfer. During carbon and proton evolution periods, XiX and WALTZ-16 were applied for heteronuclear decoupling.

Table 17 Collection of intramolecular and intermolecular unambiguous long-range distance restraints of the self-assembling system of MxiH needles

Intermolecular distance restraints are indicated with a prime. Distances are shown in angstroms (Å).

res1	atom1	res2	atom2	Distance (Å)	Chemical-shift deviation (ppm)
V3	Cγ1	L15	Cδ1	9.67 ± 0.08	0.05
V3	Cγ2	L15	Cδ1	10.31 ± 0.08	0.07
L12	Cδ1	V68	Cγ1	3.94 ± 0.02	0.03
L12	Cδ1	V68	Cγ2	3.98 ± 0.03	0.04
L12	Cδ2	V68	Cγ1	4.07 ± 0.03	0.03
L12	Cδ2	V68	Cγ2	5.20 ± 0.03	0.02
L15	Cδ1	V68	Cγ1	7.59 ± 0.04	0.08
L15	Cδ1	V68	Cγ2	6.45 ± 0.05	0.04
L12'	Cδ1	L59'	Cδ1	10.24 ± 1.05	0.05
L12'	Cδ2	L59'	Cδ1	8.51 ± 1.10	0.05

Table 18 List of structurally unambiguous long-range distance restraints of the self-assembling system of MxiH needles

Intermolecular distance restraints are indicated with a prime symbol. Distances are shown in angstroms (Å).

res1	atom1	res2	atom2	Distance (Å)	Chemical-shift deviation (ppm)
L26	Cδ1	L30	Cδ2	9.56 ± 0.06	0.11
V74	Cγ1	I78	Cδ1	4.52 ± 0.12	0.03
V74	Cγ2	I78	Cδ1	6.92 ± 0.12	0.07
V3	N	L15	Cδ2	10.68 ± 0.09	0.08
V5	N	L15	Cδ2	12.05 ± 0.10	0.10
N43	N	L47	Cδ1	5.65 ± 1.20	0.03
K72	N	L12	Cδ1	5.59 ± 0.05	0.13
K72	N	L12	Cδ2	5.67 ± 0.05	0.03
Q64	Ne2	V68	Cγ2	6.07 ± 0.04	0.04
V3'	Cγ1	L34'	Cδ1	10.08 ± 0.07	0.06
V3'	Cγ2	L34'	Cδ1	8.88 ± 0.07	0.05
L15'	Cδ1	L34'	Cδ1	7.23 ± 0.06	0.02
L59'	Cδ1	I79'	Cδ1	4.23 ± 0.68	0.08
L59'	Cδ2	I79'	Cδ1	5.78 ± 0.97	0.05
V70'	Cγ1	L54'	Cδ1	3.90 ± 0.05	0.05
V70'	Cγ2	L54'	Cδ1	5.16 ± 0.08	0.02
V70'	Cγ2	L54'	Cδ2	4.02 ± 0.03	0.04
N43'	N	V74'	Cγ1	8.51 ± 0.20	0.01
E49'	N	I78'	Cδ1	7.16 ± 0.09	0.15
S55'	N	V70'	Cγ2	6.39 ± 0.08	0.02
L59'	N	I79'	Cδ1	6.35 ± 0.35	0.04
Y60'	N	I79'	Cδ1	6.80 ± 0.30	0.02
R61'	N	V74'	Cγ2	6.86 ± 0.10	0.06

Table 19 List of ambiguous long-range distance restraints for MxiH needles

Intermolecular distance restraints are indicated with a prime symbol. Distances are shown in angstroms (Å). Restraints that originate from the same cross-peaks are grouped together with the same index and can be used as a single ambiguous restraint.

Index	res1	atom1	res2	atom2	Distance (Å)	Chemical-shift deviation (ppm)
1	V3	Cγ1	L15	Cδ2	7.84 ± 0.09	0.14
1	V3'	Cγ1	L34'	Cδ2	7.79 ± 0.07	0.02
2	L12	Cδ1	V70	Cγ1	8.87 ± 0.05	0.04
2	L12'	Cδ1	L59'	Cδ2	9.11 ± 0.69	0.05
3	L37	Cδ2	L47	Cδ2	4.28 ± 1.06	0.05
3	L37	Cδ1	L47	Cδ2	4.57 ± 1.11	0.08
4	L47'	Cδ1	I78'	Cδ1	5.23 ± 0.83	0.09
4	L47'	Cδ1	I79'	Cδ1	5.71 ± 1.03	0.07
5	L47'	Cδ2	I78'	Cδ1	5.84 ± 0.87	0.06
5	L47'	Cδ2	I79'	Cδ1	5.88 ± 0.99	0.02
6	V70	Cγ1	V74	Cγ2	4.47 ± 0.08	0.03
6	V70'	Cγ1	L54'	Cδ2	4.14 ± 0.05	0.07
7	I71	Cδ1	V74	Cγ1	5.67 ± 0.10	0.14
7	I71'	Cδ1	L54'	Cδ1	4.65 ± 0.11	0.11
8	I71	Cδ1	L15	Cδ2	6.12 ± 0.04	0.16
8	I71	Cδ1	V74	Cγ2	4.89 ± 0.13	0.12
8	I71'	Cδ1	L54'	Cδ2	6.28 ± 0.11	0.15
9	S16	N	V68	Cγ1	7.32 ± 0.04	0.06
9	S55'	N	V68'	Cγ1	8.01 ± 0.04	0.09
10	S16	N	V68	Cγ2	5.26 ± 0.03	0.06
10	S55'	N	V68'	Cγ2	10.24 ± 0.04	0.07
11	A38'	N	L59'	Cδ2	10.43 ± 1.27	0.15
11	F82'	N	L59'	Cδ2	12.74 ± 0.61	0.10
11	F82'	N	V70'	Cγ1	10.49 ± 0.06	0.10
12	N43'	N	L15'	Cδ2	10.59 ± 0.05	0.10
12	N43'	N	L74'	Cγ2	10.24 ± 0.17	0.02
13	E49'	N	V3'	Cγ1	13.70 ± 0.15	0.13
13	E49'	N	L59'	Cδ2	12.95 ± 0.63	0.05
13	E49'	N	V70'	Cγ2	13.53 ± 0.05	0.05
14	Y50	N	L47	Cδ1	7.89 ± 0.94	0.02
14	I78'	N	L47'	Cδ1	6.89 ± 0.31	0.02
14	Q80'	N	L47'	Cδ1	8.21 ± 0.48	0.13
15	Q64	N	L12	Cδ1	10.20 ± 0.08	0.06
15	E56'	N	L12'	Cδ1	8.69 ± 0.07	0.08
15	K53'	N	L12'	Cδ1	8.99 ± 0.05	0.10
16	N66	N	V70	Cγ2	6.66 ± 0.02	0.02
16	R83'	N	V70'	Cγ2	6.89 ± 0.06	0.01

References

- (1) Perutz, M. F. *Nature* **1942**, *149*, 491.
- (2) Bullmore, E. T.; Sporns, O. *Nature Reviews Neuroscience* **2009**, *10*, 186.
- (3) Jaremko, L.; Jaremko, M.; Giller, K.; Becker, S.; Zweckstetter, M. *Science* **2014**, *343*, 1363.
- (4) Lange, A.; Giller, K.; Hornig, S.; Martin-Eauclaire, M. F.; Pongs, O.; Becker, S.; Baldus, M. *Nature* **2006**, *440*, 959.
- (5) Andrew, E. R.; Bradbury, A.; Eades, R. G. *Nature* **1958**, *182*, 1659.
- (6) Andrew, E. R.; Bradbury, A.; Eades, R. G. *Nature* **1959**, *183*, 1802.
- (7) Suter, D.; Ernst, R. *Physical Review B* **1985**, *32*, 5608.
- (8) Baldus, M.; Petkova, A. T.; Herzfeld, J.; Griffin, R. G. *Molecular Physics* **1998**, *95*, 1197.
- (9) Baldus, M.; Geurts, D. G.; Hediger, S.; Meier, B. H. *J. Magn. Reson. Ser. A* **1996**, *118*, 140.
- (10) Lange, A.; Seidel, K.; Verdier, L.; Luca, S.; Baldus, M. *Journal of the American Chemical Society* **2003**, *125*, 12640.
- (11) Lange, A.; Luca, S.; Baldus, M. *Journal of the American Chemical Society* **2002**, *124*, 9704.
- (12) Chevelkov, V.; Giller, K.; Becker, S.; Lange, A. *Journal of Magnetic Resonance* **2013**, *230*, 205.
- (13) Demers, J.-P.; Chevelkov, V.; Lange, A. *Solid state nuclear magnetic resonance* **2011**, *40*, 101.
- (14) Lundstrom, P.; Teilum, K.; Carstensen, T.; Bezsonova, I.; Wiesner, S.; Hansen, D. F.; Religa, T. L.; Akke, M.; Kay, L. E. *Journal of Biomolecular Nmr* **2007**, *38*, 199.
- (15) Schwieters, C. D.; Kuszewski, J. J.; Tjandra, N.; Clore, G. M. *Journal of Magnetic Resonance* **2003**, *160*, 65.
- (16) Guntert, P.; Mumenthaler, C.; Wuthrich, K. *Journal of Molecular Biology* **1997**, *273*, 283.
- (17) Manolikas, T.; Herrmann, T.; Meier, B. H. *Journal of the American Chemical Society* **2008**, *130*, 3959.
- (18) Schubert, M.; Manolikas, T.; Rogowski, M.; Meier, B. H. *Journal of Biomolecular Nmr* **2006**, *35*, 167.
- (19) Castellani, F.; van Rossum, B.; Diehl, A.; Schubert, M.; Rehbein, K.; Oschkinat, H. *Nature* **2002**, *420*, 98.
- (20) Hong, M.; Jakes, K. *Journal of Biomolecular Nmr* **1999**, *14*, 71.
- (21) Loquet, A.; Habenstein, B.; Chevelkov, V.; Vasa, S. K.; Giller, K.; Becker, S.; Lange, A. *Journal of the American Chemical Society* **2013**, *135*, 19135.
- (22) LeMaster, D. M.; Kushlan, D. M. *Journal of the American Chemical Society* **1996**, *118*, 9255.
- (23) Lundstroem, P.; Vallurupalli, P.; Hansen, D. F.; Kay, L. E. *Nature Protocols* **2009**, *4*, 1641.
- (24) Loquet, A.; Sgourakis, N. G.; Gupta, R.; Giller, K.; Riedel, D.; Goosmann, C.; Griesinger, C.; Kolbe, M.; Baker, D.; Becker, S.; Lange, A. *Nature* **2012**, *488*.
- (25) Demers, J.-P.; Sgourakis, N. G.; Gupta, R.; Loquet, A.; Giller, K.; Riedel, D.; Laube, B.; Kolbe, M.; Baker, D.; Becker, S.; Lange, A. *Plos Pathogens* **2013**, *9*, e1003245.
- (26) Lv, G.; Fasshuber, H. K.; Loquet, A.; Demers, J.-P.; Vijayan, V.; Giller, K.; Becker, S.; Lange, A. *Journal of Magnetic Resonance* **2013**, *228*, 45.
- (27) Habenstein, B.; Loquet, A.; Giller, K.; Becker, S.; Lange, A. *Journal of Biomolecular Nmr* **2013**, *55*, 1.
- (28) Tugarinov, V.; Kanelis, V.; Kay, L. E. *Nature Protocols* **2006**, *1*, 749.
- (29) Tycko, R. In *Annual Review of Physical Chemistry*; Leone, S. R., Cremer, P. S., Groves, J. T., Johnson, M. A., Eds. 2011; Vol. 62, p 279.
- (30) Chan, J. C. C. In *Solid state nuclear magnetic resonance*; Chan, J. C. C., Ed. 2012; Vol. 306, p 47.
- (31) Comellas, G.; Rienstra, C. M. In *Ann. Rev. Biophys.*; Dill, K. A., Ed. 2013; Vol. 42, p 515.
- (32) Tycko, R.; Wickner, R. B. *Accounts of Chemical Research* **2013**, *46*, 1487.
- (33) McDermott, A. In *Ann. Rev. Biophys.* 2009; Vol. 38, p 385.
- (34) Hong, M.; Zhang, Y.; Hu, F. H. In *Annual Review of Physical Chemistry*; Annual Reviews: Palo Alto, 2012; Vol. 63, p 1.
- (35) Franks, W. T.; Linden, A. H.; Kunert, B.; van Rossum, B. J.; Oschkinat, H. *European Journal of Cell Biology* **2012**, *91*, 340.
- (36) Sun, S.; Han, Y.; Paramasivam, S.; Yan, S.; Siglin, A. E.; Williams, J. C.; Byeon, I.-J. L.; Ahn, J.; Gronenborn, A. M.; Polenova, T. In *Protein NMR Techniques, Third Edition*; Shekhtman, A., Burz, D. S., Eds. 2012; Vol. 831, p 303.
- (37) Goldbourn, A. *Current Opinion in Biotechnology* **2013**, *24*, 705.
- (38) Loquet, A.; Habenstein, B.; Demers, J.-P.; Becker, S.; Lange, A. *Medecine sciences : M/S* **2012**, *28*, 926.
- (39) Loquet, A.; Habenstein, B.; Lange, A. *Accounts of Chemical Research* **2013**, *46*, 2070.
- (40) Batchelder, L. S.; Sullivan, C. E.; Jelinski, L. W.; Torchia, D. A. *Proceedings of the National Academy of Sciences of the United States of America-Biological Sciences* **1982**, *79*, 386.
- (41) Saito, H.; Tabeta, R.; Shoji, A.; Ozaki, T.; Ando, I.; Miyata, T. *Biopolymers* **1984**, *23*, 2279.
- (42) Mack, J. W.; Torchia, D. A.; Steinert, P. M. *Biochemistry* **1988**, *27*, 5418.

- (43) Yoshimizu, H.; Ando, I. *Macromolecules* **1990**, *23*, 2908.
- (44) Huster, D. In *Annual Reports on NMR Spectroscopy*; Webb, G. A., Ed. 2008; Vol. 64, p 127.
- (45) Arnold, A. A.; Marcotte, I. *Concepts in Magnetic Resonance Part A* **2009**, *34A*, 24.
- (46) Huang, K. Y.; Amodeo, G. A.; Tong, L.; McDermott, A. *Protein Science* **2011**, *20*, 630.
- (47) Seidel, K.; Etzkorn, M.; Heise, H.; Becker, S.; Baldus, M. *Chembiochem* **2005**, *6*, 1638.
- (48) Igumenova, T. I.; Wand, A. J.; McDermott, A. E. *Journal of the American Chemical Society* **2004**, *126*, 5323.
- (49) Vijaykumar, S.; Bugg, C. E.; Cook, W. J. *Journal of Molecular Biology* **1987**, *194*, 531.
- (50) Ban, D.; Funk, M.; Gulich, R.; Egger, D.; Sabo, T. M.; Walter, K. F. A.; Fenwick, R. B.; Giller, K.; Pichierri, F.; de Groot, B. L.; Lange, O. F.; Grubmueller, H.; Salvatella, X.; Wolf, M.; Loidl, A.; Kree, R.; Becker, S.; Lakomek, N.-A.; Lee, D.; Lunkenheimer, P.; Griesinger, C. *Angewandte Chemie-International Edition* **2011**, *50*, 11437.
- (51) Lakomek, N.-A.; Walter, K. F. A.; Fares, C.; Lange, O. F.; de Groot, B. L.; Grubmueller, H.; Brueschweiler, R.; Munk, A.; Becker, S.; Meiler, J.; Griesinger, C. *Journal of Biomolecular Nmr* **2008**, *41*, 139.
- (52) Sabo, T. M.; Bakhtiari, D.; Walter, K. F. A.; McFeeters, R. L.; Giller, K.; Becker, S.; Griesinger, C.; Lee, D. *Protein Science* **2012**, *21*, 562.
- (53) Schneider, R.; Seidel, K.; Etzkorn, M.; Lange, A.; Becker, S.; Baldus, M. *Journal of the American Chemical Society* **2010**, *132*, 223.
- (54) Lv, G.; Kumar, A.; Giller, K.; Orcelet, M. L.; Riedel, D.; Fernandez, C. O.; Becker, S.; Lange, A. *Journal of Molecular Biology* **2012**, *420*, 99.
- (55) Ernst, R. R.; Bodenhauser, G.; Wokaun, A. *Principles of Nuclear Magnetic Resonance in One and Two Dimensions*; Oxford Science Publications, 1987.
- (56) McDermott, A.; Polenova, T. *Solid-State NMR Studies of Biopolymers*; Wiley: West Sussex, United Kingdom, 2010.
- (57) Levitt, M. H. *Spin Dynamics*; second Edition ed.; Wiley: West Sussex, United Kingdom, 2008.
- (58) Duer, M. J. *Solid State NMR Spectroscopy: Principles and Applications*; Wiley, 2008.
- (59) Hartmann, S. R.; Hahn, E. L. *Physical Review* **1962**, *128*, 2042.
- (60) Fung, B. M.; Khitrin, A. K.; Ermolaev, K. *Journal of Magnetic Resonance* **2000**, *142*, 97.
- (61) Shaka, A. J.; Keeler, J.; Frenkiel, T.; Freeman, R. *Journal of Magnetic Resonance* **1983**, *52*, 335.
- (62) Bennett, A. E.; Griffin, R. G.; Ok, J. H.; Vega, S. *J Chem Phys* **1992**, *96*, 8624.
- (63) Chevelkov, V.; Shi, C.; Fasshuber, H. K.; Becker, S.; Lange, A. *Journal of Biomolecular Nmr* **2013**, *56*, 303.
- (64) Shi, C.; Fasshuber, H. K.; Chevelkov, V.; Xiang, S.; Habenstein, B.; Vasa, S. K.; Becker, S.; Lange, A. *Journal of Biomolecular Nmr* **2014**, *59*, 15.
- (65) Markley, J. L.; Putter, I.; Jardetzki, O. *Science* **1968**, *161*, 1249.
- (66) Crespi, H. L.; Katz, J. J. *Nature* **1969**, *224*, 560.
- (67) Cai, M. L.; Huang, Y.; Sakaguchi, K.; Clore, G. M.; Gronenborn, A. M.; Craigie, R. *Journal of Biomolecular Nmr* **1998**, *11*, 97.
- (68) Grzesiek, S.; Bax, A. *Journal of the American Chemical Society* **1993**, *115*, 12593.
- (69) Hong, M.; Jakes, K. *Journal of Biomolecular Nmr* **1999**, *14*, 71.
- (70) Eddy, M. T.; Belenky, M.; Sivertsen, A. C.; Griffin, R. G.; Herzfeld, J. *Journal of Biomolecular Nmr* **2013**, *57*, 129.
- (71) Guo, C.; Geng, C.; Tugarinov, V. *J Biomol NMR* **2009**, *44*, 167.
- (72) Weiner, P. K.; Kollman, P. A. *Journal of Computational Chemistry* **1981**, *2*, 287.
- (73) Bernard R. Brooks, R. E. B., * Barry D. Olafson, § David J. States; S. Swaminathan, ; Karplus, M. *Journal of Computational Chemistry* **1983**, *4*, 187.
- (74) Brunger, A. T.; Adams, P. D.; Clore, G. M.; DeLano, W. L.; Gros, P.; Grosse-Kunstleve, R. W.; Jiang, J. S.; Kuszewski, J.; Nilges, M.; Pannu, N. S.; Read, R. J.; Rice, L. M.; Simonson, T.; Warren, G. L. *Acta Crystallographica Section D-Biological Crystallography* **1998**, *54*, 905.
- (75) Jensen, F. *Introduction to Computational Chemistry*; second Edition ed.; Wiley: West Sussex, England, 2007.
- (76) Bhattacharya, A.; Tejero, R.; Montelione, G. T. *Proteins-Structure Function and Bioinformatics* **2007**, *66*, 778.
- (77) Huang, Y. J.; Powers, R.; Montelione, G. T. *Journal of the American Chemical Society* **2005**, *127*, 1665.
- (78) Laskowski, R. A.; MacArthur, M. W.; Moss, D. S.; Thornton, J. M. *Journal of Applied Crystallography* **1993**, *26*, 283.
- (79) Lovell, S. C.; Davis, I. W.; Adrendall, W. B.; de Bakker, P. I. W.; Word, J. M.; Prisant, M. G.; Richardson, J. S.; Richardson, D. C. *Proteins-Structure Function and Genetics* **2003**, *50*, 437.
- (80) Luthy, R.; Bowie, J. U.; Eisenberg, D. *Nature* **1992**, *356*, 83.
- (81) Sippl, M. J. *Proteins-Structure Function and Genetics* **1993**, *17*, 355.

- (82) Morcombe, C. R.; Zilm, K. W. *Journal of Magnetic Resonance* **2003**, *162*, 479.
- (83) Goddard, T. D.; Kneller, D. G.; 3.113 ed.; University of California, S. F., Ed. 2006.
- (84) Shen, Y.; Delaglio, F.; Cornilescu, G.; Bax, A. *Journal of Biomolecular Nmr* **2009**, *44*, 213.
- (85) Loquet, A.; Lv, G.; Giller, K.; Becker, S.; Lange, A. *J Am Chem Soc* **2011**, *133*, 4722.
- (86) Goldbourt, A.; Day, L. A.; McDermott, A. E. *Journal of Magnetic Resonance* **2007**, *189*, 157.
- (87) Voet, D.; Voet, J. G. *Biochemistry*; Wiley, 2011.
- (88) Crout, D. H. G.; Hedgecock, C. J. R.; Lipscomb, E. L.; Armstrong, F. B. *European Journal of Biochemistry* **1980**, *110*, 439.
- (89) Neri, D.; Szyperki, T.; Otting, G.; Senn, H.; Wuthrich, K. *Biochemistry* **1989**, *28*, 7510.
- (90) Sabo, T. M.; Smith, C. A.; Ban, D.; Mazur, A.; Lee, D.; Griesinger, C. *Journal of Biomolecular Nmr* **2014**, *58*, 287.
- (91) Lange, O. F.; Lakomek, N.-A.; Fares, C.; Schroeder, G. F.; Walter, K. F. A.; Becker, S.; Meiler, J.; Grubmueller, H.; Griesinger, C.; de Groot, B. L. *Science* **2008**, *320*, 1471.
- (92) Bryn Fenwick, R.; Esteban-Martin, S.; Richter, B.; Lee, D.; Walter, K. F. A.; Milovanovic, D.; Becker, S.; Lakomek, N. A.; Griesinger, C.; Salvatella, X. *Journal of the American Chemical Society* **2011**, *133*, 10336.
- (93) Luca, S.; Filippov, D. V.; van Boom, J. H.; Oschkinat, H.; de Groot, H. J. M.; Baldus, M. *Journal of Biomolecular Nmr* **2001**, *20*, 325.
- (94) Koradi, R.; Billeter, M.; Wuthrich, K. *Journal of Molecular Graphics* **1996**, *14*, 51.
- (95) Cornilescu, G.; Marquardt, J. L.; Ottiger, M.; Bax, A. *Journal of the American Chemical Society* **1998**, *120*, 6836.
- (96) Laurencin, D.; Wong, A.; Chrzanowski, W.; Knowles, J. C.; Qiu, D.; Pickup, D. M.; Newport, R. J.; Gan, Z. H.; Duer, M. J.; Smith, M. E. *Physical Chemistry Chemical Physics* **2010**, *12*, 1081.
- (97) Lange, A.; Becker, S.; Seidel, K.; Giller, K.; Pongs, O.; Baldus, M. *Angewandte Chemie* **2005**, *44*, 2089.
- (98) Szeverenyi, N. M.; Sullivan, M. J.; Maciel, G. E. *Journal of Magnetic Resonance* **1982**, *47*, 462.
- (99) Takegoshi, K.; Nakamura, S.; Terao, T. *Chemical Physics Letters* **2001**, *344*, 631.
- (100) Morcombe, C. R.; Gaponenko, V.; Byrd, R. A.; Zilm, K. W. *Journal of the American Chemical Society* **2004**, *126*, 7196.
- (101) Gullion, T.; Schaefer, J. *Journal of Magnetic Resonance* **1989**, *81*, 196.
- (102) De Paepe, G.; Lewandowski, J. R.; Loquet, A.; Bockmann, A.; Griffin, R. G. *The Journal of chemical physics* **2008**, *129*, 245101.
- (103) Hing, A. W.; Vega, S.; Schaefer, J. *Journal of Magnetic Resonance* **1992**, *96*, 205.
- (104) Lewandowski, J. R.; De Paepe, G.; Griffin, R. G. *Journal of the American Chemical Society* **2007**, *129*, 728.
- (105) Sodickson, D. K.; Levitt, M. H.; Vega, S.; Griffin, R. G. *J Chem Phys* **1993**, *98*, 6742.
- (106) Bayro, M. J.; Maly, T.; Birkett, N. R.; Dobson, C. M.; Griffin, R. G. *Angewandte Chemie-International Edition* **2009**, *48*, 5708.
- (107) Loquet, A.; Giller, K.; Becker, S.; Lange, A. *Journal of the American Chemical Society* **2010**, *132*, 15164.
- (108) Tugarinov, V.; Hwang, P. M.; Kay, L. E. *Annual Review of Biochemistry* **2004**, *73*, 107.
- (109) Gelis, I.; Bonvin, A. M. J. J.; Keramisanou, D.; Koukaki, M.; Gouridis, G.; Karamanou, S.; Economou, A.; Kalodimos, C. G. *Cell* **2007**, *131*, 756.
- (110) Bista, M.; Freund, S. M.; Fersht, A. R. *Proceedings of the National Academy of Sciences of the United States of America* **2012**, *109*, 15752.
- (111) Huber, M.; Hiller, S.; Schanda, P.; Ernst, M.; Bockmann, A.; Verel, R.; Meier, B. H. *Chemphyschem : a European journal of chemical physics and physical chemistry* **2011**, *12*, 915.
- (112) Zhou, D. H.; Rienstra, C. M. *Journal of Magnetic Resonance* **2008**, *192*, 167.
- (113) Detken, A.; Hardy, E. H.; Ernst, M.; Meier, B. H. *Chemical Physics Letters* **2002**, *356*, 298.
- (114) Lichtecker, R. J.; Weinhaeupl, K.; Reuther, L.; Schoerghuber, J.; Schmid, W.; Konrat, R. *Journal of Biomolecular Nmr* **2013**, *57*, 205.
- (115) Lichtecker, R. J.; Weinhaeupl, K.; Schmid, W.; Konrat, R. *Journal of Biomolecular Nmr* **2013**, *57*, 327.
- (116) Lv, G.; Fasshuber, H. K.; Loquet, A.; Demers, J. P.; Vijayan, V.; Giller, K.; Becker, S.; Lange, A. *J Magn Reson* **2013**, *228*, 45.
- (117) Hansen, D. F.; Kay, L. E. *Journal of the American Chemical Society* **2011**, *133*, 8272.
- (118) Fricke, P.; Demers, J.-P.; Becker, S.; Lange, A. *Chemphyschem : a European journal of chemical physics and physical chemistry* **2014**, *15*, 57.
- (119) Nilges, M.; Macias, M. J.; Odonoghue, S. I.; Oschkinat, H. *Journal of Molecular Biology* **1997**, *269*, 408.
- (120) Fossi, M.; Castellani, T.; Nilges, M.; Oschkinat, H.; van Rossum, B. J. *Angewandte Chemie-International Edition* **2005**, *44*, 6151.
- (121) Shahid, S. A.; Bardiaux, B.; Franks, W. T.; Krabben, L.; Habeck, M.; van Rossum, B.-J.; Linke, D. *Nature Methods* **2012**, *9*, 1212.

- (122) Jean-Philippe Demers, B. H., Antoine Loquet, Suresh Kumar Vasa, Stefan Becker, David Baker, Adam Lange, Nikolaos Sgourakis *nature communication* **2014**.
- (123) Guntert, P. *European Biophysics Journal with Biophysics Letters* **2009**, *38*, 129.
- (124) Zhou, D. H.; Shah, G.; Cormos, M.; Mullen, C.; Sandoz, D.; Rienstra, C. M. *Journal of the American Chemical Society* **2007**, *129*, 11791.
- (125) Hodgkinson, P.; Emsley, L. *Journal of Magnetic Resonance* **1999**, *139*, 46.
- (126) Beomsoo, H.; Yifeng, L.; Ginzinger, S. W.; Wishart, D. S. *Journal of Biomolecular NMR* **2011**, *50*, 43.
- (127) Meiler, J. *Journal of Biomolecular Nmr* **2003**, *26*, 25.
- (128) Rablen, P. R.; Pearlman, S. A.; Finkbiner, J. *Journal of Physical Chemistry A* **1999**, *103*, 7357.

Acknowledgements

I would like to thank my supervisor Prof. Dr. Adam Lange for the interesting project and his support throughout my Ph.D. Especially I want to thank him for his patience and the inspiring discussions about solid-state NMR.

I want to thank my thesis committee members Marina Bennati and Lars Kuhn for fruitful discussions during the thesis committee meetings.

I 'm grateful that I had the opportunity to work in such a nice environment of scientific richness at the Max-Planck Institute for Biophysical Chemistry in the NMR-based Structural Biology department and I want to thank Prof. Dr. Christian Griesinger for providing this excellent facility and friendly environment.

I would like to thank all members of the solid-state NMR group for their helping and their friendship. My special gratitude goes to my colleagues Dr. Jean-Philippe Demailly and Dr. Antoine Loquet who taught me the secrets of solid-state NMR. I want to thank Dr. Antoine Loquet for the excellent xplor-NIH script which is a fundamental part of the realization of this thesis. I want to thank Dr. Birgit Habenstein for her help on the Ubiquitin project and for great times of outdoor activities. I want to thank Dr. Nils-Alexander Lakomek for the interesting discussions about Ubiquitin and his helpful contribution of interpreting the ssNMR Ubiquitin data. I want to thank Dr. Veniamin Chevelkov and Dr. Chaowei Shi for their cooperation on the Ubiquitin project. I want to thank Brigitta Angerstein for her professional support all over the time. Special thanks go to Dr. Suresh Kumar Vasa and Dr. Davood Bakhtiari for their friendly support throughout my Ph.D. I also want to thank the members of the cubic for their productive discussions. I appreciate the nice outdoor adventures with Marta, Pablo and Dave and want to thank them for the nice time.

I want to thank Dr. Stefan Becker and his team for the production of the excellent samples and the abundant discussions about all and sundry.

Last but not least I want to thank my friends Elisabeth Bernhart, Oliver Bixner, Christian Rauch and my family for their support and understandings all over the years.

Curriculum Vitae

Personal data

Name Hannes Klaus Fasshuber
Gender male
Date and place of birth 25.09.1979 Judenburg, Austria
E-mail: hafa@nmr.mpibpc.mpg.de
Current Address: Fassberg 11, 37077, Göttingen, Germany
Tel: +49-(0)551-2012262

Academic education

2011- **Max Planck Institute for Biophysical Chemistry**, Goettingen Germany
PhD student in the department of NMR-based Structural Biology
Supervisor: Dr. Adam Lange
Dissertation: ^{13}C sparse labeling for solid-state NMR investigations of biomolecular systems

2004-2011 **University Vienna**, Vienna, Austria
Chemistry
Supervisor: Dr. Lothar Brecker
Thesis: NMR Based Structure Determination of Alkaloids
Isolated from Plants of the Genus *Psychotria* L. (Rubiaceae)

Publications:

- Chaowei Shi, **Hannes K. Fasshuber**, Veniamin Chevelkov, Shengqi Xiang, Birgit Habenstein, Suresh Kumar Vasa, Stefan Becker, Adam Lange, BSH-CP based 3D solid-state NMR experiments for protein resonance assignment, *JBNMR*, 2014, 59, 1, 15-22.
- Veniamin Chevelkov, Chaowei Shi, **Hannes K. Fasshuber**, Stefan Becker, Adam Lange, Efficient band-selective homonuclear CO-CA cross-polarization in protonated proteins, *JBNMR*, 2013, 56,4,303-311.
- Guohua Lv, **Hannes K. Fasshuber**, Antoine Loquet, Jean-Philippe Demers, Vinesh Vijayan, Karin Giller, Stefan Becker, Adam Lange, A Straightforward Method for Stereospecific Assignment of Val and Leu Prochiral Methyl Groups by Solid-State NMR: Scrambling in the [2- ^{13}C]Glucose Labeling Scheme, *JNMR*, 2013, 228, 45-49.
- Andreas Berger, **Hannes Fasshuber**, Johann Schinnerl, Lothar Brecker, Harald Greger, Various types of tryptamine-iridoid alkaloids from *Palicourea acuminata*, *Phytochemistry Letters*, 2012, 5, 3, 558-562.
- Ann-Christin Pöpler, Margret M. Meinholz, **Hannes Fasshuber**, Adam Lange, Michael John, Dietmar Stalke, Mixed Crystalline Lithium Organics and Interconversion in Solution, *Organometallics*, 2012, 31, 1, 42-45.
- Andreas Berger, **Hannes Fasshuber**, Johann Schinnerl, Wolfgang Robien, Lothar Brecker, Karin Valant-Vetschera, Iridoids as chemical markers of false ipecac (*Ronabea emetic*) a previously confused medicinal plant, *J.of Ethnopharmacology*, 2011, 138, 3,756-761.
- Martin Bernhard, **Hannes Fasshuber**, Wolfgang Robien, Lothar Brecker, Harald Greger, Dopamine-iridoid alkaloids in *Carapichea affinis* (*Psychotria borucana*) confirm close relationship to the vomiting root Ipecac, *Biochemical Systematics and Ecology*, 2001, 39, 3, 232-235.