New Methods in NMR Spectroscopy for the Study of Protein Dynamics

Dissertation
zur Erlangung des Doktorgrades
der Mathematisch-Naturwissenschaftlichen Fakultäten
der Georg-August-Universität zu Göttingen

vorgelegt von

Nils-Alexander Lakomek
aus Göttingen

Göttingen 2008

D7
Referent: Prof. Dr. Reiner Kree
Korreferent: Prof. Dr. Christian Griesinger
Tag der mündlichen Prüfung: 28. Oktober 2008

Abstract

The focus of this work has been the investigation of internal protein dynamics in a time window between the overall tumbling correlation time τ_c (a few nanoseconds) and about 50 μs (supra-τ_c time window). This time window is inaccessible by NMR relaxation methods, but contains functionally important dynamics which have remained undetected so far. With the use of Residual Dipolar Couplings (RDCs) in NMR spectroscopy this previously hidden time window is made accessible. We used ubiquitin as a model-system, a 8.5 kDa protein involved in many regulatory processes in the cell, such as protein degradation. During this PhD work the experimental basis of backbone amide group NH RDCs has been significantly enhanced. The experimental basis, consisting of NH RDC data sets from altogether 36 different alignment conditions is the broadest available and is of highest accuracy and homogeneity [Lakomek, Walter et al. 2008]. From the RDC data, dynamic information can be extracted using the RDC-based model-free approach, originally invented by Griesinger and coworkers [Meiler et al., 2001, Peti et al., 2002]. The RDC-based model-free approach relies on the measurement of NH RDCs for five linearly independent alignment tensor orientations in at least five different alignment media. Using a high-resolution structure to determine the alignment tensors, structural as well as dynamic information can be deduced. During this PhD work the RDC-based model-free approach was rigorously re-evaluated mathematically and the algorithm further improved, for example, with respect to filtering out of experimental noise [Lakomek et al., 2006]. Using this new approach, RDC-based order parameters $S_{rdc}^2(NH)$ were derived with unprecedented accuracy revealing new modes of motion in the supra-τ_c time window. Charged and polar residues show more mobile backbone amide groups than hydrophobic residues. To our surprise, according to the RDC-based model-free analysis, a correlation between side-chain orientation and backbone mobility could be observed for ubiquitin [Lakomek et al., 2005]. Residues with solvent-exposed side chains showed a tendency to be more mobile in the protein backbone, whereas those with side chains pointing towards the hydrophobic core appeared more rigid in the protein backbone. This finding was manifested as an alternating pattern of $S_{rdc}^2(NH)$ order

parameters in the β-strands and challenges the current picture of a rigid protein backbone uncoupled to side-chain dynamics.

In order to become independent from the accuracy of the structural model used for alignment tensor calculation, a Self-Consistent RDC-based Model-free analysis (SCRM) was developed. It delivers RDC-based order parameters independently of the details of the structure used for alignment tensor calculation, as well as the dynamic average orientation of the inter-nuclear vectors in the protein structure in a self-consistent manner. The SCRM approach corroborated the presence of motions slower than the correlation time τ_c. Indeed, the inclusion of the supra-τ_c window increases the averaged amplitude of mobility observed in the sub-τ_c window by about 34%. Furthermore, a RDC-refined structural ensemble of ubiquitin was derived (Lange, Lakomek et al., 2008). In contrast to relaxation-based ensembles of solution structures considering only motions faster than τ_c, the RDC-derived ensemble includes solution dynamics up to microseconds. The ensemble covers the complete structural heterogeneity observed in 46 ubiquitin crystal structures, most of which are in complexes with other proteins. From that we conclude that conformational selection, rather than induced fit motion suffices to explain the molecular recognition dynamics of ubiquitin.

Kurzzusammenfassung

Wesentlicher Inhalt dieser Doktorarbeit ist die Untersuchung der inneren Dynamik von Proteinen in einem Zeitfenster, langsamer als die rotatorische Korrelationszeit τ_c (im Bereich von wenigen Nanosekunden) und schneller als ca. 50 μs (supra-τ_c Zeitfenster). Dieser Zeitbereich ist für Relaxationsmethoden der kernmagnetischen Resonanz (NMR) nicht zugänglich. Für die Funktion wichtige Proteindynamik in diesem Zeitbereich blieb daher bislang unerforscht. Mit Hilfe von Residualen Dipolaren Kopplungen (RDCs) in der NMR-Spektroskopie konnte dieser bislang verborgene Zeitbereich zugänglich gemacht werden. Als Modellsystem wurde Ubiquitin verwendet, ein 8.5 kDa grosses Protein, das in viele regulatorische Prozesse in der Zelle involviert ist, z.B. im Abbauprozess von Proteinen. Während dieser Doktorarbeit wurde die experimentelle Basis der RDCs von NH Amidgruppen im Proteinrückgrat deutlich ausgebaut. Die experimentellen Grundlagen basieren nun auf NH RDC Datensätzen von Präzisionsmessungen in insgesamt 36 verschiedenen Orientierungsmedien. Das ist die breiteste und homogenste zur Zeit verfügbare Datensammlung [Lakomek, Walter et al. 2008]. Aus den RDCs kann mit Hilfe der sogenannten RDC-basierten Modell-freien Analyse Information über Proteindynamik extrahiert werden. Dieses Analyseverfahren wurde ursprünglich von Griesinger und Mitarbeitern entwickelt [Meiler et al., 2001, Peti et al., 2002] und basiert auf der Messung von fünf linear unabhängigen Orientierungstensoren in mindestens fünf verschiedenen Orientierungsmedien. Unter Zuhilfenahme einer hochaufgelösten Struktur zur Berechnung der Orientierungstensoren zu den dazugehörigen RDC Datensätzen, kann damit strukturelle und dynamische Information abgeleitet werden. Im Rahmen dieser Doktorarbeit wurde die RDC-basierte Modell-freie Analyse mathematisch rigoros re-evaluiert und der Algorithmus erheblich optimiert, z.B. im Hinblick auf die Ausfilterung experimentellem Rauschens [Lakomek et al., 2006]. Mit Hilfe des neuen Analyse -Verfahrens konnten RDC-basierte Ordnungsparameter $S^2_{rdc}(NH)$ mit bislang unerreichter Genauigkeit bestimmt werden. Diese $S^2_{rdc}(NH)$ Ordnungsparameter machen neue, bislang unentdeckte Bewegungsmodi von Ubiquitin im Supra-τ_c-Zeitbereich sichtbar. Geladene und polare Aminosäuren zeigen eine höhere Beweglichkeit der Amidgruppen als Aminosäuren mit hydrophoben Seit-

enketten. Den Ergebnissen unserer Analyse zufolge liegt eine Korrelation zwischen Seitenketten-Orientierung und der Beweglichkeit des Proteinrückgrates vor. Aminosäuren mit Seitenketten, die der Lösung zugewandt sind, erscheinen im Proteinrückgrat beweglicher als Aminosäuren mit hydrophoben Seitenketten [Lakomek et al., 2005]. Diese Beobachtung zeigte sich als alternierendes Muster von $S^2_{rdc}(NH)$ Ordnungsparametern im β-Faltblatt und stellte das bisherige Bild eines verhältnismässig starren Proteinrückgrates in Frage, das nicht an die Dynamik der Seitenketten gekoppelt ist.

Um von der Genauigkeit des Strukturmodells, das bei der Berechnung des Orientierungstensors benutzt wird, unabhängig zu werden, wurde die Selbst-Consistente RDC-basierte Model-freie Analyse (SCRM) entwickelt. Dieses Verfahren liefert RDC-basierte Ordnungsparameter und dynamisch gemittelte NH Vektor-Orientierungen, unabhängig von den Details der Struktur, die zur Berechnung des Orientierungstensors herangezogen wird. Das SCRM-Verfahren untermauert das Vorhandensein von Bewegung langsamer als die Korrelationszeit τ_c. Der Einschluss des Supra-τ_c-Zeitbereichs vergrössert die mittlere Amplitude der Bewegung gegenüber dem Sub-τ_c-Zeitbereich um 34%.

Im weiteren konnte ein RDC-basiertes Struktur-Ensemble von Ubiquitin entwickelt (Lange, Lakomek et al., 2008) werden. Im Gegensatz zu bisherigen Ensembles, die auf Relaxationsdaten basieren und damit nur Bewegung schneller als die Korrelationszeit τ_c berücksichtigen, schliesst das RDC-basierte Struktur-Ensemble auch Dynamik im Mikrosekunden Bereich mit ein. Dieses neu entwickelte Ensemble deckt die gesamte strukturelle Heterogenität von Ubiquitin ab, die in verschiedenen Kristallstrukturen von Ubiquitin beobachtet wird, wenn dieses im Komplex an andere Proteine gebunden ist. Daraus schliessen wir, dass die molekulare Erkennung durch Ubiquitin über Konformations-Selektion erfolgt. Demzufolge liegen alle Konformationen von Ubiquitin schon in freier Lösung vor und werden nicht erst durch den Bindungspartner induziert.

Related Publications

The work presented in this PhD thesis led to the following publications. I wish to express my kindest regards to all co-authors for fruitful cooperation.

1. Lakomek, N. A., Fares, C., Becker, S., Carlomagno, T., Meiler, J., Griesinger, C., Side-chain orientation and hydrogen-bonding imprint supra-τ_c motion on the protein backbone of ubiquitin, *Angew. Chem. Int. Ed.*, 44, 7776-7778, (2005). [Lakomek et al., 2005]

2. Lakomek, N. A., Carlomagno, T., Becker, S., Griesinger, C., Meiler, J., A thorough dynamic interpretation of residual dipolar couplings in ubiquitin. *J. Biomol. NMR*, 34, 101-115, (2006). [Lakomek et al., 2006]

3. Lakomek*, N. A., Walter*, K. F. A., Farés, C., Lange, O. F., de Groot, B. L., Grubmuller, H., Bruschweiler, R., Munk, A., Becker, S., Meiler, J., Griesinger, C., Self-consistent residual dipolar coupling based model-free analysis for the robust determination of nanosecond to microsecond protein dynamics, *J. Biomol. NMR*, 41,139-155 (2008). [Lakomek, Walter et al., 2008]

4. Lange*, O. F., Lakomek*, N. A., Farés, C., Schroder, G. F., Walter, K. F. A., Becker, S., Meiler, J., Grubmuller, H., Griesinger, C., de Groot, Bert L., Recognition dynamics up to microseconds revealed from an RDC-derived ubiquitin ensemble in solution, *Science*, 320, 1471-1475, (2008). [Lange, Lakomek et al., 2008]

*= both authors contributed equally

Acknowledgements

Furthermore, I would like to thank Michael Funk for carefully reading this dissertation.

Finally, I would like to thank all my colleagues in the department NMR-based structural biology for a nice and friendly atmosphere making it a good place to work.

Contents

Chapter 1

Introduction

The study of protein dynamics with different biophysical methods

Even after the human genome was sequenced in 2003, which can be considered one of the major milestones in biology, many secrets towards an understanding of life at a molecular level remain. In most cases the primary sequence of a protein cannot be directly related to its function; knowledge about the protein structure and dynamics is usually needed. X-ray crystallography, Nuclear Magnetic Resonance (NMR) spectroscopy and electron microscopy have been very powerful in determining the three-dimensional structures of proteins and protein complexes. Since Perutz solved the first X-ray structure of Haemoglobin in 1960 [Perutz et al., 1960] and Kendrew the first of Myoglobin in the same year [Kendrew et al., 1960], X-ray crystallography has elucidated more than 44, 000 protein structures (PDB statistics, see www.rcsb.org). Currently, protein complexes up to several Mega Daltons (10^6 u) can be investigated by X-ray crystallography [Ban et al., 2000]. A more recent technique, cryo electron microscopy provides models of large protein systems with a resolution up to 10 Å and better [Halic et al., 2006a, Halic et al., 2006b].

NMR spectroscopy has been developed in parallel as an alternative approach to investigate the structure of proteins. After the first observation of the nuclear magnetic resonance phenomenon in 1946 by Purcell and Bloch [Purcell et al., 1946, Bloch, 1946, Bloch et al., 1946], the invention of high-power, pulse NMR spectroscopy and subsequent Fourier transformation by Ernst and co-workers [Ernst and Anderson, 1966] and its first application to biomolecular systems by Wuethrich and co-workers have led NMR spectroscopy to become an important method in structural biology and biophysics [Wagner and Wuthrich, 1978, Wuthrich, 1986]. The invention of three-dimensional NMR spectroscopy [Griesinger et al., 1987, Oschkinat et al., 1988] and use of labeling strategies

[Oh et al., 1988, Marion et al., 1989, Kay et al., 1989b] have enabled the study of higher molecular weight complexes. Today, monomeric proteins up to 50 kDa are routinely studied with NMR spectroscopy, and sophisticated labeling strategies enable the study of proteins up to 670 kDa [Sprangers and Kay, 2007] [Sprangers et al., 2007]. With respect to the size of protein structures that can be solved, NMR spectroscopy is inferior to X-ray crystallography, but in contrast to X-ray crystallography, NMR can study the dynamics of proteins in solution [Kay, 1998, Palmer, 2004, Mittermaier and Kay, 2006].

Understanding protein dynamics is crucial to understanding protein function. Many molecular processes, like enzymatic reactions or ligand recognition, are governed by protein dynamics [Wang et al., 2001, Kern and Zuiderweg, 2003] [Henzler-Wildman et al., 2007, Zhang et al., 2007]. Nevertheless, the understanding of protein dynamics is still limited due to severe experimental difficulties. Besides NMR spectroscopy, protein dynamics can be studied by a number of different techniques, such as optically time-resolved methods after temperature jumps [Ervin et al., 2000, Fierz et al., 2007], electric methods [Eigen and de Maeyer, 1955, Porschke and Eigen, 1971], infrared-spectroscopy [Hamm et al., 1995, Snow et al., 2002] and fluorescence correlation spectroscopy [Oehlenschlager et al., 1996, Kettling et al., 1998]. Promising techniques such as time-resolved X-ray crystallography [Srajer et al., 1996, Ihee et al., 2005a], two-dimensional infrared spectroscopy [Bredenbeck et al., 2007a] [Bredenbeck et al., 2007b], or temperature jump fluorescence based methods [Ma and Gruebele, 2005] [Ervin et al., 2000] and temperature jump Fourier transformation infrared spectroscopy (FT-IR) [Chung et al., 2007] are emerging to access protein dynamics with atomic resolution. However, as with every technique, these methods also suffer various bottlenecks: Time-resolved X-ray crystallography has the prerequisite of simultaneous excitation of the molecular ensemble in order to study protein dynamics with atomic resolution and is limited to very short time-scales on the order of femto- to picoseconds. Moreover, since the crystal packing cannot change, unfolding reactions are hard to study – as for example seen by a comparison with solution state NMR results for the photoactive yellow protein (PYP) [Ihee et al., 2005b, Rubinstenn et al., 1998]. Fluorescence-based methods require a labeling of the protein with a fluorescent dye that can change the structural and dynamic properties of the protein or are limited to photo-active amino acids like tryptophan, tyrosine and phenylalanine. Two-dimensional infrared spectroscopy is limited to vibrational interactions in the femto- to picosecond time-scale. While all these methods have provided insight, they are unable to study the equilibrium dynamics of a protein with atomic resolution over a wide time range. From the theoretical side, molecular dynamics simulations point a possible solution. However, state-of-the-art molecular dynamics trajectories of

protein motion are limited to less than a few microseconds because of limited computer resources [Maragakis et al., 2008]. Furthermore, the force-fields used need experimental cross-validation [Karplus, 2003].

Protein dynamics studied by NMR spectroscopy

When it comes to the experimental study of internal dynamics in a protein with atomic resolution and time resolution from femtoseconds to hours, NMR spectroscopy is unique in the sense that it covers a wide time range and requires only minimal interference with the native protein. With multidimensional NMR spectroscopy most of the atoms in a molecule are accessible for NMR-detected observation and — except for replacing the naturally occurring carbon ^{12}C by ^{13}C and the naturally occurring nitrogen ^{14}N by ^{15}N — modification of the protein is not necessary and it can be observed in a near-physiological environment. Also, in contrast to time resolved measurements, kinetics can be measured in equilibrium: the equilibrium does not need to be disturbed.

Up to now, most NMR methods to study protein dynamics have been based on relaxation methods which reveal motion in the pico- to nanosecond time scale. Rates and amplitude of mobility can be determined by measuring T1 and T2 relaxation times as well as HetNOEs (Heteronuclear Nuclear Overhauser Effect) [Kay, 1998, Palmer, 2004]. Usually backbone amide NH groups are the focus of studies to investigate protein backbone motion and methyl groups likewise to measure side chain dynamics [Mittermaier and Kay, 2006]. As explained later in detail (cf. chapter 3), only motion faster than the overall tumbling correlation τ_c of a protein (the time a protein needs for one rotation, on the order of a few nanoseconds) can be studied. In the following we will call these kind of motions sub-τ_c motions. Such motion has been proposed to contribute mostly to the entropy of proteins [Schneider et al., 1992, Li et al., 1996] [Prompers and Bruschweiler, 2000, Lee and Wand, 2001]. Motion in a time-range between $50\mu s$ and 10ms is accessible by relaxation dispersion measurements, based on the stochastic fluctuations of isotropic chemical shifts, which are independent of rotational tumbling [Akke and Palmer, 1996, Mittermaier and Kay, 2006]. Kinetic rates and populations of different conformations can be determined. These measurements have been used to characterize major structural changes and enzymatic reactions [Wang et al., 2001] [Henzler-Wildman et al., 2007]. Conformational heterogeneity slower than 50 ms can be directly observed as peak splitting in NMR spectra. For backbone amides, motions faster than 50 μs do not result in sufficient line broadening to be detectable for relaxation dispersion measurements. These measurements therefore probe motions slower than about 50 μs up to approximately 50 ms.

Based on data from NMR relaxation studies, X-ray crystallography and molecular
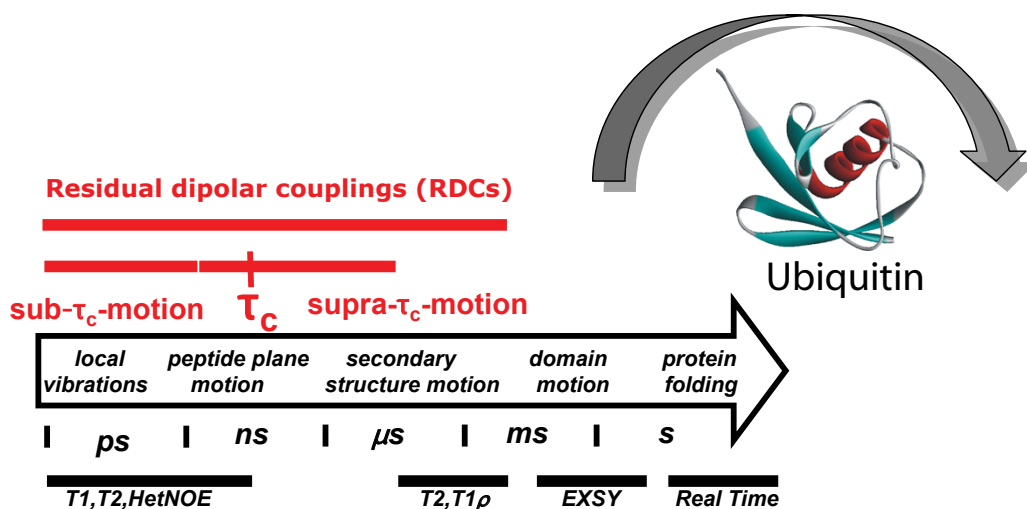
Figure 1.1: Time scales in NMR spectroscopy. Residual dipolar couplings are sensitive to motion from ps to ms. They provide information complementary to relaxation methods, especially about the before inaccessible supra-τ_c time-window.

dynamics simulation protein were, briefly, divided into three main categories:(i) small random atomic fluctuations on the sub-picosecond timescale (e.g. bond librations) that are, in general, uniform throughout the protein; (ii) correlated motions involving the concerted movement of small or large groups of atoms that can span the picosecond to second timescale (e.g. crank-shaft motions along the polypeptide backbone, which are both frequent and fast; aromatic ring flips, which are fast but infrequent; local unfolding, which is slow and infrequent; and interdomain motions); and (iii) triggered conformational changes. The first category always represents random, essentially axially symmetric, excursions about an equilibrium conformation and is driven by the inherent kinetic energy of a protein. The second category is also driven by inherent kinetic energy and can involve either random excursions about an equilibrium conformation or transitions from one equilibrium state (or sub-state) to another. The third category always involves a transition from one equilibrium state to another and, while triggered by an external event such as ligand binding, still requires inherent kinetic energy and, hence, the existence of the other two motional classes, to take place [Clore and Schwieters, 2006].

That means relaxation methods are only sensitive to protein motion in the sub-τ_c range and in the range between 50 μs and 50 ms. Thus, there is a window be-

4

tween the correlation time τ_c and 50 μs that has been unexplored (see Figure 1). We will call this the *supra-τ_c* time window in the following. Functionally relevant dynamic processes in proteins might have remained undetected so far because of the inaccessibility of this time window. To our knowledge, the only report in which such motions are described with atomic detail is the photodissociation of carbon monoxide from myoglobin that has been studied by Moffat and colleagues in a crystal with Laue diffraction [Srajer et al., 1996]. Here, conformational re-arrangements on the small μs range were reported. However, such investigations have not been possible in solution so far.

RDCs as a tool to study protein dynamics

A few years ago the group of Joel Tolman (Johns Hopkins University, Baltimore) and the group of Christian Griesinger independently developed methods to access this hidden supra-τ_c time window by using so-called Residual Dipolar Couplings (RDCs) [Meiler et al., 2001, Tolman, 2002]. Dipolar couplings can be observed as a splitting of the resonance of a given spin that is dipolar-coupled to another spin. Since this dipolar coupling has an orientational angular dependence with respect to the external magnetic field, it is generally averaged out in solution state NMR spectroscopy because of the isotropic tumbling of the protein in solution and the isotropic distribution of orientations of the molecule. However, proteins can be aligned in solution by weak liquid crystalline media which still allow an almost isotropic tumbling of the molecule but induce an anisotropic distribution of orientations that reintroduces a small amount of the dipolar coupling interaction, the *residual* dipolar coupling (RDC) [Tolman et al., 1995, Tjandra and Bax, 1997]. The alignment happens in very non-invasive manner so that the resulting residual dipolar coupling is about 1000 times smaller than the original dipolar coupling interaction, thus providing orientation information without affecting the shape and dynamics of the protein. RDCs report on the direction of inter-nuclear vector orientations; for example, the direction of the backbone amide group in a protein:

$$D = D_a \left((3\cos^2\theta - 1) + \frac{3}{2} R \sin^2\theta \cos 2\phi \right) \qquad . \qquad (1.1)$$

Here, D_a is the axial component of the alignment tensor and R the rhombicity. The angles θ and ϕ are the polar coordinates in the alignment frame. The alignment frame is fixed to the molecular frame of the molecule.

That RDCs are sensitive to protein dynamics was recognized early on [Tolman et al., 1997]. If there are internal dynamics of an inter-nuclear vector within the molecular frame, this will have an effect on the orientation of the inter-nuclear vector in the molecular frame as well as in the alignment frame.

5

Because of their orientation dependence, the experimentally measured RDC values are dynamically averaged and can function as a probe for protein dynamics. Since RDCs are fixed to the molecular frame independently of the tumbling of the molecule, they pick up motions faster and slower than rotational tumbling correlation time τ_c of the molecule up to the same limit that was active for chemical shift modulations discussed above. Thus, RDCs are also sensitive to the supra-τ_c window between τ_c and 50 μs and enable the exploration of a previously hidden time window (Figure 1).

In the Griesinger group, the so-called *RDC-based model free approach* has been developed [Meiler et al., 2001] and, likewise, the DIDC (Direct Interprepation of Dipolar Couplings) approach was introduced by [Tolman, 2002]. Both methods similarly rely on the measurement of NH RDC data sets in at least five linear-independent alignment media and have been applied to the protein ubiquitin [Briggman and Tolman, 2003, Peti et al., 2002]. Ubiquitin is a key protein involved in many regulatory processes in the cell. Proteins are marked by polyubiquitination for degradation in the proteasome.

The first experimental studies by the groups of Tolman and Griesinger, both on the protein ubiquitin, had observed a significant amount of supra-τ_c motion, i.e. motion slower than the overall tumbling correlation time τ_c of the protein. The main outcome of both methods are RDC-based order parameters $S^2_{rdc}(NH)$ that are sensitive to the amplitudes of backbone NH groups up to the millisecond time-scale (see before). These are compared to Lipari-Szabo S^2_{LS} order parameters derived from relaxation measurements that are only sensitive for motion faster than the correlation time τ_c of a protein which is around 4 ns for ubiquitin at room temperature [Lipari and Szabo, 1982a, Chang and Tjandra, 2005]. Conclusions about additional dynamics slower than τ_c are mainly derived from a comparison of these order parameters.

The discovery of motion slower than τ_c by Tolman, Griesinger, and co-workers has challenged the current perception of protein mobility that considers the protein backbone as rigid in comparison to the more mobile side-chains. The widely accepted picture is that most of the dynamics occur in the sub-τ_c time window and only rare events occur in the μs to ms time range, for example, the breaking of hydrogen bonds [Clore and Schwieters, 2004b, Clore and Schwieters, 2006]. This point of view has been challenged by the works of Tolman, Griesinger and co-workers. As a result, their work has led to an intense debate during the last few years. As has been pointed out by Marius Clore: "... If this were true, it would represent a paradigm shift in our understanding of protein motions" [Clore and Schwieters, 2006] . Possible problems of the model-free approach include noise in the experimental RDC data and the influence of *structural noise*, that is the influence of an inaccuracy of the structure used for alignmnent tensor

calculation, These problems were noted in [Clore and Schwieters, 2004b]. The influence of structural noise on the dynamic interpretation of RDC data has been well investigated by [Zweckstetter and Bax, 2002]. Thus, in 2004, the year when this PhD work started, a rigorous examination and re-evaluation of the "RDC-based" model-free approach was needed. Experimentally it turned out to be very difficult to match the five linearly independent alignment tensor orientations by measuring in only five different alignment media. The explanation for this is that most alignment media show some linear dependence and it is hard to cover all five dimensions equally. That is why, experimentally, a much larger number of RDC data sets was needed. That was the starting point of this PhD thesis.

PhD thesis:

- The initial goal of this PhD work was a rigorous mathematical re-evaluation of the RDC-based model-free approach and systematic enlargement of the experimental basis of NH RDC data for ubiquitin to reduce the effects of experimental uncertainties. The refined model-free approach and the obtained results will be presented in chapter 4.

- With the results of the refined rdc-based model-free approach we were able to identify a correlation between side-chain orientation and backbone mobility for ubiquitin. These findings challenge the current picture of a rigid protein backbone uncoupled to side-chain dynamics and will be presented in chapter 5.

- A driving force through this PhD work was the continuous improvement and further development both of the mathematical analysis of RDC data as well the experimental basis towards higher accuracy. These efforts culminated in the newly developed Self-Consistent RDC-based Model-free (SCRM) approach which alleviates the influence of structural noise. Several parameters were introduced that assess the quality of fit of the obtained results to the experimental data. An intrinsic scaling problem of the model-free approach with respect to relaxation data has been accessed in statistically robust manner. Through these measures, the S_{rdc}^2 have become more reliable. The developed SCRM approach and the obtained results are presented in chapter 6.

- In parallel to the SCRM approach, a molecular dynamics based approach has been developed in order to describe the observed dynamics with a *structural ensemble* that provides an atomic picture of the amplitudes of motion. This work was a cooperative effort of the research group Computational Biomolecular Dynamics (head: Dr. Bert de Groot) and our group and is described in chapter 7.

Chapter 2

Basics of NMR-Spectroscopy

2.1 The phenomenon of nuclear magnetic resonance

All atom nuclei with a nuclear spin quantum number I unequal to zero possess a magnetic moment μ_I:

$$\boldsymbol{\mu}_I = \gamma \mathbf{I} \tag{2.1}$$

with the nuclear spin $\mathbf{I}$ and the gyromagnetic ratio γ. The nuclear spin shows the characteristics of quantum angular momentum (compare e.g.
[Haken and Wolf, 2001] [Cohen-Tannoudji, 2001a, Cohen-Tannoudji, 2001b].

In an external static magnetic field $\mathbf{B}_0$ the previously undistinguishable energy levels of $\mathbf{I}$ split into distinct levels [Haken and Wolf, 2001]. It is convention to put $\mathbf{B}_0$ along the z-axis. Then the magnetic moment of the nucleus μ_I has the energy in the external magnetic field:

$$E = -\boldsymbol{\mu}_I \cdot \mathbf{B}_0 = -\mu_{I,z} B_{0,z} = -\gamma \hbar I_z B_0 \tag{2.2}$$

Note that in order to be consistent with the NMR literature [Sorensen, 1989], in the following we will use I_z as a unit-less number without $\hbar$ (in contrast to most introductory textbooks for quantum mechanics such as for example [Cohen-Tannoudji, 2001a].)

For all fermionic nuclei with nuclear spin number $I = 1/2$, such as ^{1}H, ^{13}C, ^{15}N, ^{31}P, which are often detected in biomolecular NMR spectroscopy, I_z has the two eigen-

states $I_z = -1/2$ and $I_z = +1/2$ and we get the two distinct energy levels:

$$E(\alpha) = -\frac{1}{2}\gamma\hbar B_0 \qquad \text{for the nuclear spin in the } \alpha\text{-state } I_z = \tfrac{1}{2} \text{ and}$$

$$E(\beta) = \frac{1}{2}\gamma\hbar B_0 \qquad \text{for the nuclear spin in the } \beta\text{-state } I_z = -\tfrac{1}{2} \qquad .$$

Irradiation with an electromagnetic wave of an energy equal to the difference between those two energy levels will lead to a transition from the lower energy state $E(\alpha)$ to the higher state $E(\beta)$:

$$-h\nu_L = -\hbar\omega_L = \Delta E = \gamma\hbar B_0 \tag{2.3}$$

The corresponding resonance frequency ν_L is called *Larmor frequency* (see e.g. [Cavanagh et al., 1996]):

$$\nu_L = \frac{\omega_L}{2\pi} = -\frac{\gamma B_0}{2\pi} \qquad . \tag{2.4}$$
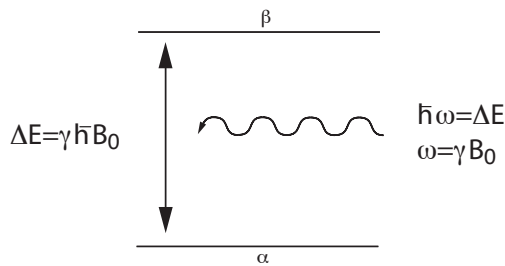
Figure 2.1: Resonance transition of the nuclear spin from state α to state β upon irradiation with an electromagnetic wave of Larmor frequency

This stimulated absorptive process is the basic principle of nuclear magnetic resonance. After the invention of NMR spectroscopy by Purcell and Bloch in 1947, for a long time resonance signals were detected using *continuous wave* techniques. Only the *high power pulse* technique and subsequent fourier transformation of the spectral data invented by Richard Ernst, made an application to biomolecular systems feasible. For the high power pulse technique, a pulsed electromagnetic wave of Larmor frequency ω_L is applied perpendicular to the static magnetic field B_0.

To understand the basic interactions in NMR spectroscopy we will take a closer look at the Hamiltonian describing a two spin system in an external magnetic

field. We will distinguish between internal interactions between the two nuclear spins themselves, $\mathscr{H}_{int}$, and external interactions with the external magnetic field, $\mathscr{H}_{ext}$.

$$H = \mathscr{H}_{ext} + \mathscr{H}_{int} \qquad . \tag{2.5}$$

The external interaction can be separated in an interaction $\mathscr{H}_{ext,0}$ with the static magnetic field B_0 and an interaction $\mathscr{H}_{ext,1}$ with the field B_1 induced by the oscillating electromagnetic wave:

$$\mathscr{H}_{ext} = \mathscr{H}_{ext,0} + \mathscr{H}_{ext,1} \tag{2.6}$$

The effect of pulsed electromagnetic waves is described by $\mathscr{H}_{ext,1}$ and has to be treated with perturbation theory methods. A thorough description of those methods and a convenient simplification called the *product operator formalism* can be found, for example, in textbooks like [Cavanagh et al., 1996, Ernst et al., 1987].

The term $\mathscr{H}_{ext,0}$ is the *Zeeman* Hamiltonian and has the following energy eigen values:

$$\frac{E}{\hbar} = \omega_1 I_{1z} + \omega_2 I_{2z} \tag{2.7}$$

with the resonance frequencies ω_1 and ω_2 of the different nuclear spins I_1 and I_2; Because I_z can take two discrete eigenvalues values $I_{1z} = -1/2$ and $I_{2z} = +1/2$, we get four different discrete energy levels:

$$\frac{E}{\hbar}(\alpha\alpha) = \frac{1}{2}(\omega_1 + \omega_2) \tag{2.8}$$

$$\frac{E}{\hbar}(\beta\alpha) = \frac{1}{2}(-\omega_1 + \omega_2) \tag{2.9}$$

$$\frac{E}{\hbar}(\alpha\beta) = \frac{1}{2}(\omega_1 - \omega_2) \tag{2.10}$$

$$\frac{E}{\hbar}(\beta\beta) = \frac{1}{2}(-\omega_1 - \omega_2). \tag{2.11}$$

$$\tag{2.12}$$

with $\omega_1 = -\gamma B_{loc}$ and ω_2 respectively. B_{loc} differs slightly from the static external field B_0 and depends on the local electronic environment of nucleus 1. Both field are related by the so-called nuclear shielding σ:

$$B_{loc} = (1 - \sigma)B_0 \qquad . \tag{2.13}$$

The nuclear shielding σ has the negative value of the more frequently used chemical shift δ: $\delta = -\sigma$. The effect of chemical shift arises because of Lenz's rule: Motions of electrons induced by the external magnetic field B_0 generate a secondary magnetic field that is opposed to the inducing field B_0. This is the most

prominent effect, so called diamagnetic shielding or also diamagnetic ring currents. There are also paramagnetic effects to be observed for example from Carbonyl groups which rely on the small energy difference between n/π^* transitions and lead to a magnetic moment that increases the external magnetic field.

Following the *selection rules*

$$\Delta I_{1z} = \pm 1 \qquad \Delta I_{2z} = \pm 1 \qquad \Delta I_{z,sum} = \pm 1 \tag{2.14}$$

with $I_{z,sum} = I_{1z} + I_{2z}$ we get two possible spectral transitions that lead to the resonance lines $\nu_1 = \omega_1/2\pi$ and $\nu_2 = \omega_2/2\pi$ in the spectrum.

2.2 Interactions between nuclear spins

Interaction between the nuclear spin are described by $\mathscr{H}_{int}$. We distinguish between indirect *scalar* interactions that are mediated via binding electrons and *dipolar* interactions that are direct interaction trough the space:

$$\mathscr{H}_{int} = \mathscr{H}_{sc} + \mathscr{H}_{dip} \qquad . \tag{2.15}$$

The *scalar* or *J-coupling* interaction between two nuclear spins is mediated via binding electrons by the *Fermi contact interaction* between the magnetic moment of the nucleus and the magnetic moments of the s-electrons. In the limit of weak coupling($2\pi J_{12} \ll |\omega_1 - \omega_2|$) that is fulfilled for most homonuclear and heteronuclear J-couplings the internal scalar Hamiltonian can be reduced to

$$\mathscr{H}_{sc} = 2\pi J_{12} I_{1z} I_{2z} \qquad . \tag{2.16}$$

The scalar interaction term leads to a shift of energy levels. Neglecting the direct dipolar interaction the expression for the energy levels of the system becomes

$$\frac{E}{\hbar} = \omega_1 I_{1z} + \omega_2 I_{2z} + 2\pi J_{12} I_{1z} I_{2z} \tag{2.17}$$

In addition to equation 2.7 we get the additional term $2\pi J_{12} I_{1z} I_{2z}$ that leads to a shift in energy levels:

$$\frac{E}{\hbar}(\alpha\alpha) = \frac{1}{2}(\omega_1 + \omega_2 + \pi J_{12}) \tag{2.18}$$

$$\frac{E}{\hbar}(\beta\alpha) = \frac{1}{2}(-\omega_1 + \omega_2 - \pi J_{12}) \tag{2.19}$$

$$\frac{E}{\hbar}(\alpha\beta) = \frac{1}{2}(\omega_1 - \omega_2 - \pi J_{12}) \tag{2.20}$$

$$\frac{E}{\hbar}(\beta\beta) = \frac{11}{2}(-\omega_1 - \omega_2 + \pi J_{12}). \tag{2.21}$$

$$\tag{2.22}$$
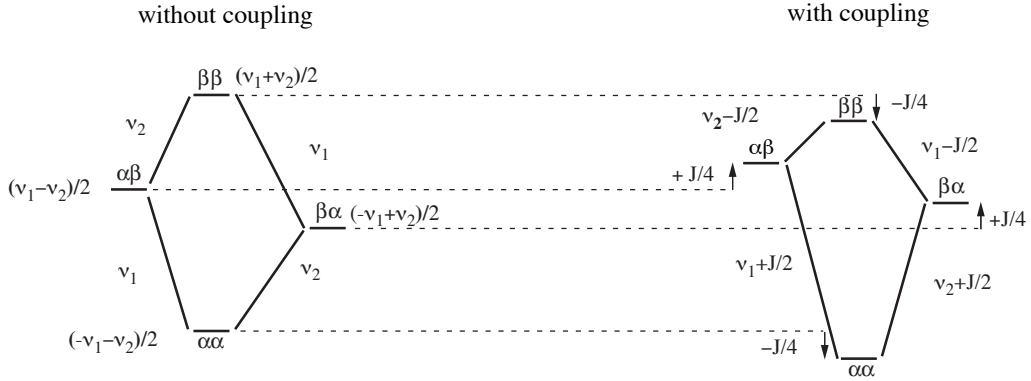
Figure 2.2: Shift of energy level diagrams due to the scalar interaction

This leads to a splitting of the original resonance line two equal resonance lines separated by J.

Due to the similarity of dipolar and J coupling Hamiltonian as is obvious from comparisons of equations (2.16) and (3.5) the same spectroscopic *phenotype* is induced by the dipolar couplings as by the J-coupling. The scalar interaction is between two and four orders of magnitude smaller than the direct dipolar interaction which will be treated in the next chapter.

Chapter 3

Residual dipolar couplings as a tool to study protein dynamics

Since the work and results of this thesis rely on the dipolar coupling interaction in NMR spectroscopy, a detailed but comprehensive derivation of the dipolar coupling interaction is provided in the following. It will be shown in detail how the most frequently used dipolar coupling equation in the alignment frame (see below) can be derived. Furthermore it will be explained why residual dipolar couplings can provide information complementary to relaxation methods.

3.1 Dipolar coupling interaction between nuclear spins

We start from the interaction energy between two magnetic dipoles, which is well known from classical electrodynamics [Jackson, 1983, Nolting, 2000]:

$$E = \frac{\mu_0}{4\pi} \frac{1}{r_{12}^3} \cdot \left(\boldsymbol{\mu}_1 \cdot \boldsymbol{\mu}_2 - \frac{3}{r_{12}^2} \left(\boldsymbol{\mu}_1 \cdot \mathbf{r}_{12} \right) \left(\boldsymbol{\mu} \cdot \mathbf{r}_{12} \right) \right), \tag{3.1}$$

Using the quantum mechanic correspondence principal [Cohen-Tannoudji, 2001a] and $\boldsymbol{\mu} = \gamma \mathbf{I}$ we get the following description for the dipolar Hamiltonian:

$$\mathscr{H}_{dip} = \frac{\mu_0}{4\pi} \hbar^2 \frac{\gamma_1 \gamma_2}{r_{12}^3} \left(\mathbf{I_1} \cdot \mathbf{I_2} - \frac{3(\mathbf{I}_1 \cdot \mathbf{r}_{12})(\mathbf{I}_2 \cdot \mathbf{r}_{12})}{r_{12}^2} \right) \tag{3.2}$$

with r_{12} being the distance between nucleus 1 and nucleus 2.

If we consider the heteronuclear case, for example ^{1}H and ^{15}N, in the high field limit only the secular terms remain active: I_{1z} and I_{2z}. Using the definition of the scalar product

$$\frac{\mathbf{I}_z \cdot \mathbf{r}_{12}}{r_{12}} = I_{1z} \cos \theta \tag{3.3}$$

we get:

$$\mathscr{H}_{dip} = \frac{\mu_0}{4\pi} \hbar^2 \frac{\gamma_{I_1} \gamma_{I_2}}{r_{12}^3} \Big(I_{1z} I_{2z} - 3(I_{1z} \cos \theta)(I_{2z} \cos \theta) \Big) \tag{3.4}$$

The Hamiltonian is expressed in units of energy, we want to express it in units of frequency. Because of $E = \hbar\omega = \hbar 2\pi\nu$ we have to divide $\mathscr{H}_E$ by $2\pi\hbar$ to obtain $\mathscr{H}_\nu$. We derive:

$$\mathscr{H}_{dip,\nu} = -\frac{\mu_0 h \gamma_1 \gamma_2}{8\pi^3 r_{12}^3} I_{1z} I_{2z} P_2(\cos \theta) \tag{3.5}$$

with the second Legendre polynomial $P_2(\cos \theta) = \frac{1}{2}(3\cos^2 \theta - 1)$.

The dipolar splitting, the measured dipolar coupling between two nuclei i and j, is the difference between two transitions:

$$D_{ij} = \Big(<\mathscr{H}_{ij}>_{\beta\beta} - <\mathscr{H}_{ij}>_{\beta\alpha} \Big) - \Big(<\mathscr{H}_{ij}>_{\alpha\beta} - <\mathscr{H}_{ij}>_{\alpha\alpha} \Big)$$

$$\Rightarrow \quad D_{ij} = -\frac{\mu_0 h \gamma_1 \gamma_2}{8\pi^3 r_{12}^3} P_2(\cos \theta) \Big(\big((-\frac{1}{2})(-\frac{1}{2}) - (-\frac{1}{2})\frac{1}{2}\big) - \big((\frac{1}{2})(-\frac{1}{2}) - \frac{1}{2}\frac{1}{2}\big) \Big)$$

$$\Rightarrow \quad D_{ij} = -\frac{\mu_0 h \gamma_1 \gamma_2}{8\pi^3 r_{12}^3} P_2(\cos \theta) \tag{3.6}$$

Equation 3.6 describes the dipolar coupling in the laboratory frame. Because the internal Hamiltonian is the sum of the scalar and the dipolar Hamiltonian (compare equation 2.5) the dipolar coupling D adds to the J-coupling and we measure $J + D$ instead of J.

3.2 Residual Dipolar Couplings (RDCs) in solution state NMR

For the NH group, for example, the dipolar coupling adopts its maximum value for $\theta = 0°$ of 11 kHz. This coupling is much larger than the through-bond J-coupling of approximately 90 Hz between the two nuclei that can be observed as a doublet splitting on the nitrogen and proton resonances. The dipolar coupling interaction contains a wealth of information because of its orientation dependence. On the other hand dipolar coupling as a relaxation mechanism leads to broad

line width and deteriorated spectral quality in the first solid-state NMR spectra (before Magic angle spinning and sophisticated re-coupling techniques had been applied).

In solution state NMR, however, the dipolar coupling is generally not observed as a splitting of resonances for two reasons: a) in isotropic solution the average over all orientations yields 0 for $(3\cos^2\theta - 1)$ and b) the rotational tumbling with a correlation time τ_c of approximately 4 ns for ubiquitin at room temperature [Lienin et al., 1998] occurs approximately 10,000 times faster than the magnitude of the dipolar coupling leading to efficient averaging of the powder spectrum.

Nevertheless, information about the orientation of inter-nuclear vectors obtained from dipolar couplings would be desirable. Indeed, we can regain the dipolar coupling information in solution state NMR at good spectral quality by applying the following technique: Proteins can be aligned in solution by the use of highly diluted anisotropic media, like for example weak liquid crystalline media of polyacrylamide gels. These weak anisotropic media allow an almost isotropic tumbling of the molecule but induce an anisotropic distribution of orientations that reintroduces a small amount of the dipolar coupling interaction, the *residual* dipolar coupling (RDC) [Tolman et al., 1995, Tjandra and Bax, 1997]. The alignment happens in very non-invasive manner so that the resulting residual dipolar coupling is about 1000 times smaller than the original dipolar coupling interaction, thus providing orientation information without affecting the shape and dynamics of the protein. The alignment of the protein can be described by an *alignment tensor* that describes the average orientation of the molecule with respect to the external magnetic field and will be introduced in the next paragraph.

3.3   Dipolar coupling equation in alignment frame

In the following we transform the dipolar coupling equation from the laboratory frame to the alignment frame (see below). Let our reference frame be an arbitrarily chosen *molecular frame* rigidly attached to the molecule. The orientation of the static magnetic field $\mathbf{B_0}$ with respect to the molecular frame can be described by the three angles β_x, β_y and β_z between the axis of the molecular frame and $\mathbf{B_0}$. The orientation of an internuclear vector $\mathbf{r_{ij}}$ with respect to the molecular frame can be described by three angles α_x, α_y and α_z, too. In the following let $\mathbf{b_0}$ and $\mathbf{r_{ij}}$ be the normalized vectors. Now $\cos\theta_{ij}$ with the the angle θ_{ij} between

$\mathbf{b_0}$ and $\mathbf{r_{ij}}$ can be expressed by the scalar product:

$$\cos\theta_{ij} = \mathbf{b_0} \cdot \mathbf{r_{ij}} = \begin{pmatrix} \cos\beta_x \\ \cos\beta_y \\ \cos\beta_z \end{pmatrix}^T \cdot \begin{pmatrix} \cos\alpha_x \\ \cos\alpha_y \\ \cos\alpha_z \end{pmatrix} = \sum_k \cos\beta_k cos\alpha_k \qquad . \qquad (3.7)$$

Expressing the cosine in the second legendre polynomial using this relationship gives:

$$P_2(\cos\theta) = \frac{1}{2}(3\cos^2\theta - 1) = \frac{1}{2}\left(3\left(\sum_k \cos\beta_k \cos\alpha_k\right)^2 - 1\right) = \qquad (3.8)$$

$$= \frac{3}{2}(\sum_k \cos\beta_k \cos\alpha_k \sum_l \cos\beta_l \cos\alpha_l) - \frac{1}{2} \qquad . \qquad (3.9)$$

Because $\mathbf{r_{ij}}$ is normalized, we can use the relationship $\cos^2\alpha_x + \cos^2\alpha_y + \cos^2\alpha_z = 1$. That gives:

$$P_2(\cos\theta) = \frac{3}{2}\sum_{k,l}(\cos\beta_k\cos\beta_l)(\cos\alpha_k\cos\alpha_l) - \frac{1}{2}\underbrace{\sum_{k,l}\cos\alpha_k\cos\alpha_l\delta_{kl}}_{=1} \qquad .(3.10)$$

For each single internuclear vector $\mathbf{r_{ij}}$ the dipolar coupling equation in the laboratory frame can be transformed then to:

$$D_{ij} = -\frac{\mu_0 h \gamma_1 \gamma_2}{8\pi^3 r_{12}^3}\sum_{kl} S_{kl}\cos\alpha_k\cos\alpha_l \qquad (3.11)$$

with the so-called Saupe order matrix elements [Saupe, 1968, Losonczi et al., 1999]

$$S_{kl} = \frac{3}{2}(\cos\beta_k\cos\beta_l) - \frac{1}{2}\delta_{kl} \qquad (3.12)$$

Since the Saupe order matrix is real and symmetric, it is always possible to define a new molecular axis system where $\mathbf{S}$ becomes diagonal. More common than the Saupe order matrix is to use the so-called *alignment tensor* which is related to the previous by:

$$\mathbf{S} = \frac{3}{2}\mathbf{A} \qquad (3.13)$$

The principal axis frame of the alignment tensor $\mathbf{A}$ where $\mathbf{A}$ becomes diagonal, is called *alignment frame*. The order of the principal component is defined as: $|A_{zz}| \geq |A_{yy}| \geq |A_{xx}|$. In the following we will continue this notation of the alignment tensor.

Since the dipolar coupling value is a scalar it will not be affected by the change of reference frame. Equation (3.11) transformed to the alignment frame reads:

$$D_{ij} = \frac{3}{2}d\left(A_{xx}\cos^2\xi_x + A_{yy}\cos^2\xi_y + A_{zz}\cos^2\xi_z\right) \tag{3.14}$$

with A_{xx}, A_{yy} and A_{zz} being the eigen values of the alignment tensor

$$d := -\frac{\mu_0 h\gamma_1\gamma_2}{8\pi^3 r_{12}^3} \tag{3.15}$$

To determine the unknown angles $\cos\xi_x, \cos\xi_y$ and $\cos\xi_z$ we consider the inter-nuclear vector as the local basis vector $\mathbf{e_r}$ in spherical coordinates. Let $(\mathbf{e_x}, \mathbf{e_y}, \mathbf{e_z})$ be the principal axis frame in Cartesian form. The local basis vector $\mathbf{e_r}$ can be expressed in spherical coordinates:

$$\mathbf{e_r} = \frac{\partial\mathbf{r}}{\partial r} = \begin{pmatrix} \sin\theta\cos\phi \\ \sin\theta\sin\phi \\ \cos\theta \end{pmatrix} \qquad . \tag{3.16}$$

It holds:

$$\begin{aligned} \cos\xi_x &= \mathbf{e_r}\cdot\mathbf{e_x} = \sin\theta\cos\phi \\ \cos\xi_y &= \mathbf{e_r}\cdot\mathbf{e_y} = \sin\theta\sin\phi \\ \cos\xi_z &= \mathbf{e_r}\cdot\mathbf{e_z} = \cos\theta \end{aligned} \tag{3.17}$$

Using the relationship (3.17), equation (3.14) can be transformed to:

$$D_{ij} = \frac{3}{2}d\left(A_{xx}\sin^2\theta\cos^2\phi + A_{yy}\sin^2\theta\sin^2\phi + A_{zz}\cos^2\theta\right) \tag{3.18}$$

Usually the alignment tensor is described using an axial component A_a and a rhombic component A_r. Following the notation by Bax [Bax and Grishaev, 2005] they are defined as:

$$\begin{aligned} A_a &:= \frac{3}{2}A_{zz} \\ A_r &:= (A_{xx} - A_{yy}) \end{aligned} \tag{3.19}$$

We get:

$$D_{ij} = d\Big(\frac{3}{2}A_{xx}sin^2\theta\cos^2\phi + \frac{3}{2}A_{yy}\sin^2\theta\sin^2\phi + A_a\cos^2\theta +$$

$$+\underbrace{\frac{A_a}{2}\cos^2\theta - \frac{A_a}{2}(\cos^2\theta + \sin^2\theta) + \frac{A_a}{2}\sin^2\theta(\cos^2\phi + \sin^2\phi)}_{=0}\Big). \tag{3.20}$$

The alignment tensor is traceless. Together with the definition of A_a in equation 3.19 we obtain:

$$A_{xx} + A_{yy} + A_{zz} = 0 \Rightarrow A_a = -\frac{3}{2}A_{xx} - \frac{3}{2}A_{yy} \tag{3.21}$$

and further:

$$D_{ij} = d\left(\frac{3A_{xx}}{4}\sin^2\theta(\cos^2\phi - \sin^2\phi) + \frac{3A_{yy}}{4}\sin^2\theta(\sin^2\phi - \cos^2\phi) + \frac{A_a}{2}(3\cos^2\theta - 1)\right)$$

$$\Rightarrow D_{ij} = \frac{d}{2}\left(A_a(3\cos^2\theta - 1) + \frac{3}{2}(A_{xx} - A_{yy})\sin^2\theta(\cos^2\phi - \sin^2\phi)\right)$$

$$\Rightarrow D_{ij} = \frac{d}{2}\left(A_a(3cos^2\theta - 1) + \frac{3}{2}A_r\sin^2\theta\cos 2\phi\right)$$

That gives the dipolar coupling equation in the alignment frame:

$$D_{ij} = -\frac{\mu_0 h \gamma_i \gamma_j}{16\pi^3 r_{ij}^3}\left(A_a(3\cos^2\theta - 1) + \frac{3}{2}A_r\sin^2\theta\cos 2\phi\right) \tag{3.22}$$

Instead of referring to the magnitude of alignment A_a which is a unitless number (on the order of 10^{-3}), in liquid state NMR often the dipolar coupling principal component D_a is used which is directly measurable. D_a is related to A_a by: $D_a = \frac{d}{2}A_a$. Using this relation and $R = \frac{A_r}{A_a}$ we arrive at the well-known dipolar coupling equation in the alignment frame:

$$D_{ij} = D_a\left((3\cos^2\theta - 1) + \frac{3}{2}R\sin^2\theta\cos 2\phi\right) \tag{3.23}$$

3.4   Calculation of the alignment tensor

To determine the alignment tensor of a protein with known structure, we start from equation 3.11 that can be written down for each residue or inter-nuclear vector respectively (for that a RDC can be measured). By that we obtain a system of linear equations with the five independent Saupe order matrix elements $(S_{yy}, S_{zz}, S_{xy}, S_{xz}, S_{yz})$ as unknowns, considering that the Saupe matrix is a second rank tensor, symmetric and traceless. The system of linear equations can be expressed as a matrix equation of the form:

$$\mathbf{C}S = D \tag{3.24}$$

where C is the Nx5 structure matrix containing the cosine terms in equation 3.11, S is a 5-dimensional vector of the Saupe order matrix elements and D is

a N-dimensional vector of the RDCs for each of the N residues. The vector D is obtained from the measured RDCs. If the protein structure is known, the structure matrix $\mathbf{C}$ can be determined easily from the coordinates of the pdb structure file. Now the unknown variables in S can be determined by doing a singular value decomposition (SVD, cf. [Werner, 1992a, Werner, 1992b]) of $\mathbf{C}$ and applying standard numerical methods [Losonczi et al., 1999]. Because the Saupe order matrix is real and symmetric, it always possible to define a new molecular axis system where the Saupe matrix or the alignment tensor respectively (cf. 3.13) becomes diagonal. This principal axis frame is called alignment frame. The procedure to calculate an alignment tensor and its alignment frame is automated by programs such as PALES [Zweckstetter and Bax, 2000] or DIPOCOUP [Meiler et al., 2000].

3.5 Homogeneity of RDCs measured in different alignment conditions

In order to quantify the similarity of structure and dynamics in the different alignment media (homogeneity of RDC data), a SECONDA analysis can be applied on the RDC data [Hus and Bruschweiler, 2002, Hus et al., 2003] . The SECONDA method analyzes the covariance matrix constructed of all RDC data obtained under different alignment conditions. It performs a principal component analysis (PCA) of the RDC covariance matrix, which is equivalent to a singular value decomposition (SVD) of the RDC matrix. The singular values are sorted according to decreasing size. Structural and dynamic information is contained in the first five singular values, since dipolar couplings are a second rank symmetric tensor interaction and hence reside in a linear 5-dimensional space. Accordingly, only noise, systematic errors, and structural and dynamic heterogeneity may cause the 6th and higher singular values to differ from zero. The ratio of the 5th and 6th singular values (called SECONDA gap in the following) is a measure of the homogeneity of RDC data and the magnitude of noise. The larger the SECONDA gap, the more self-consistent are the RDC data in the different alignment media.

Since there is no straightforward way to treat missing entries in the RDC covariance matrix, a specific residue can only be considered if experimental data for this residue are available in all selected alignment media (compare [Hus et al., 2003], Scenario I). All experimental RDCs are normalized by division through the largest absolute RDC value in each medium. A quantitative breakdown of inconsistencies into heterogeneous behavior and noise is not straightforward, too. The SECONDA analysis, however, allows one to obtain an upper noise limit, assuming that no heterogeneities are present. By adding various amounts of Gaussian

noise to RDCs back-calculated from the N-H vector orientations of a given average protein structure, it can be estimated to how much noise the SECONDA gap corresponds to. The difference between the thus determined noise level and the measurement error can, in principal, be attributed to structural and dynamic inconsistencies between the different alignment conditions.

3.6   Limitations of NMR relaxation methods

As already mentioned in the introduction, NMR is a unique tool to experimentally study protein dynamics with atomic resolution. NMR relaxation methods are the state-of the art methods to study fast time-scale motion from picoseconds up to a few nanoseconds as well as slow time-scale motions from approximately 50 μs up to about 50 ms [Kay et al., 1989a, Kay, 1998, Palmer, 2004, Mittermaier and Kay, 2006] . Relaxation of nuclear magnetization is caused by magnetic interactions of the nuclei that fluctuate about a mean value induced by the rotational tumbling of the molecule. One of the main interactions leading to relaxation is the through space dipolar coupling between two nuclei as introduced before. It depends on the orientation of the internuclear vector with respect to the magnetic field and the distance between the two nuclei. To analyze motion, we focus on nuclei that have a fixed distance, for example like the nitrogen and proton backbone amide group NH. Then the angular fluctuation is the only degree of freedom. As explained in section (3.2) the dipolar coupling interaction is averaged out in isotropic solution. Nevertheless, the fluctuations of the dipolar coupling due to the tumbling lead to relaxation which is governed by the overall tumbling. If we concentrate on a single NH vector then its fluctuations will be determined by the tumbling time of the molecule unless there is faster motion of this vector with respect to the molecular frame. If there is motion of the vector slower than the correlation time τ_c it will not influence the fluctuations of the dipolar coupling. Therefore, relaxation reports on motion in a molecule only up to the rotational tumbling correlation time correlation time τ_c [Palmer, 2004]. All these fluctuations detected by NMR relaxation that are faster than the correlation time are considered to contribute mainly to the entropy of proteins [Akke et al., 1993, Lee and Wand, 2001, Frederick et al., 2007] but are typically much smaller than the structural changes involved in molecular recognition.

There is another mechanism that reports on motion of proteins in the equilibrium: the modulation of the isotropic chemical shift δ by conformational fluctuations. Each nucleus has a given resonance frequency, the Larmor frequency ω_L that is determined not only by the gyromagnetic ratio γ of the spin and the external

magnetic field B_0 but also by the chemical shift that considers the magnetic field at the position of the nucleus, induced by the electronic environment of the nucleus: $\omega_L = \gamma B_0 (1 + \delta)$ (cf. chapter 2). The chemical shift leads to the dispersion of the resonances in the spectrum. Otherwise, nuclei, e.g. hydrogen atoms, in a molecule would have the same Larmor frequency ω_L. Since the isotropic chemical shift is independent of the molecular tumbling, the previously mentioned limit at the correlation time does not exist here. The chemical shift not only depends on the composition of the molecule but also on its conformation. Thus, conformational transitions of a molecule will lead to a modulation of the chemical shift of those resonances whose immediate chemical environment changes. The observable in this case is a contribution to the line width Γ of the resonances that depends on the change of Larmor frequencies as a function of the change of the chemical shifts $\Delta\delta$ and the time-scale of the chemical shift change τ_e:

$$\Gamma = \tau_e (\gamma B_0 \Delta\delta)^2 \tag{3.25}$$

NMR relaxation methods that are based on the detection of chemical shift dispersion are called NMR relaxation dispersion measurements [Akke and Palmer, 1996, Palmer et al., 2001]. As is apparent from equation 3.25, the line width contribution Γ will become undetectably small when τ_e becomes small. Chemical shifts changes $\Delta\delta$ due to conformational changes are in the order of a few ppm. At current resonance frequencies of NMR magnets up to $\gamma B_0/2\pi$ =900 MHz processes faster than approximately τ_e= 50μs cannot be studied due to limitations in the pulse repetition rate in CPMG (Carr-Purcell-Meiboom-Gill) experiments and sample heating problems [Carr and Purcell, 1954, Meiboom and Gill, 1958, Palmer et al., 2001]. Thus, only motion slower than about 50μs can be detected by NMR relaxation dispersion methods. On the slow end, chemical exchange can be detected as several resonances when the exchange is slower than approximately 50 ms. To summarize, motion in a time range between 50μs and 50ms can be detected with NMR relaxation dispersion methods. The modulation of chemical shifts has been used to characterize major structural changes and enzymatic reactions and one of the major breakthroughs in the past years has been the detection of functionally relevant dynamics in enzymes based on these measurements.

To summarize, NMR relaxation methods are sensitive both to motions faster than the overall tumbling correlation time τ_c and motions in a time window between 50μs and 50 ms. However, there is a gap between τ_c and 50μs that has been inaccessible by NMR relaxation methods so far. In the second next paragraph it will be explained how RDCs can help to access this previously hidden time window which we call *supra-τ_c time window*. Before, however, the concept of order parameters should be introduced. Order parameters are frequently used in NMR spectroscopy to characterize the amplitudes of motion.

23

3.7 S^2 order parameters in NMR spectroscopy

Amplitudes of dynamics in NMR spectroscopy are usually described by so-called generalized S^2 order parameters introduced by Lipari and Szabo in 1982 [Lipari and Szabo, 1982a, Lipari and Szabo, 1982b]:

$$S^2 = \frac{4\pi}{5} \sum_{M=-2}^{2} < Y_{2M}^*(\theta, \phi) >< Y_{2M}(\theta, \phi) > \qquad (3.26)$$

This expression can be derived as following: For the correlation function of motions of inter-nuclear vectors, like, for example, the NH vector of the backbone amide group, we assume that the overall and internal motions are independent. Then the the total correlation function can be factored as:

$$C(t) = C_0(t)C_I(t) \qquad (3.27)$$

where $C_0(t)$ is the correlation function for overall motion and $C_I(t)$ is the correlation function for internal motions, on which we will focus in the following. The correlation function for internal motions is

$$C_I(t) = < P_2(\mu(\mathbf{0}) \cdot \mu(\mathbf{t})) > \qquad , \qquad (3.28)$$

where the unit vector μ describes the orientation of the interaction vector in a reference frame that is rigidly attached to the macromolecule, and $P_2(x) = \frac{1}{2}(3x^2 - 1)$ is the second Legendre-Polynomial. The spherical harmonics addition theorem relates the second Legendre-Polynomial to second order spherical harmonics. For $L = 2$ it states:

$$P_2(\cos(\theta_{ij})) = \frac{4\pi}{5} \sum_{M=-2}^{2} Y_{2M}^*(\theta_i, \phi_i) Y_{2M}(\theta_j, \phi_j) \qquad . \qquad (3.29)$$

where θ_{ij} is the inter-nuclear angle between two different interaction vectors $\mu_{\mathbf{i}}$ and $\mu_{\mathbf{j}}$. Instead of comparing two different interaction vectors we can also compare the orientation of the same interaction vector at different points in time t_i and t_j with $t_i < t_j$. Without loss in generality we can set $t_i = 0$. Using the spherical harmonic addition theorem we derive for the internal correlation function $C_I(t)$:

$$C_I(t) = < P_2(\hat{\mu}(0) \cdot \hat{\mu}(t)) > = P_2(\cos(\theta_{ij})) = \frac{4\pi}{5} \sum_{M=-2}^{2} Y_{2M}^*(\theta_i, \phi_i) Y_{2M}(\theta_j, \phi_j) =$$

$$= \frac{4\pi}{5} \sum_{M=-2}^{2} Y_{2M}^*(\theta(t_i = 0), \phi(t_i = 0)) Y_{2M}(\theta(t_j = t), \phi(t_j = t))$$

Using the property of correlation functions that

$$\lim_{t \to \infty} < A(0)B(t) >=< A >< B >$$ (3.30)

we get:

$$C_I(t \to \infty) = \frac{4\pi}{5} \sum_{M=-2}^{2} < Y_{2M}^*(\theta, \phi) >< Y_{2M}(\theta, \phi) >$$ (3.31)

where the brackets denote the time average [Bruschweiler and Wright, 1994]. The generalized order parameter is defined as

$$S^2 := \lim_{t \to \infty} C_I(t) \qquad .$$ (3.32)

Since the internal relaxation is governed by the overall tumbling of the molecule with rotational correlation time correlation time τ_c as explained in section 3.6, S^2 will, however, only report on motion faster than τ_c. It can take values between 0 and 1. For a value of 1 the respective inter-nuclear vector is completely rigid. The smaller the order parameter becomes the more mobile the inter-nuclear vector is. This can be directly understood by having the definition of the correlation function in mind. Nowadays the generalized order parameter is usually called Lipari-Szabo order parameter according the inventors, Lipari and Szabo.

Thus, the amplitudes of motion faster than the overall tumbling correlation time τ_c are described by the so called Lipari- Szabo order parameter S_{LS}^2 that reflects sub-τ_c motion. The Lipari-Szabo order parameters used in this thesis include libration motion which is a ps angular fluctuation of each NH vector about its averaging position which reduces the order parameter homogeneously for each NH vector by 12%. Further reductions of the Lipari -Szabo order parameter affect the NH vectors individually and occur on slower but still sub-τ_c time-scales.

3.8 Why residual dipolar couplings are sensitive to protein dynamics

In the following it will be explained why RDCs are sensitive to motion both faster as well as slower than the overall tumbling correlation time τ_c of a protein.

In the alignment frame, rdcs can be described by the following equation (compare 3.22):

$$\langle D_{ij} \rangle = D_{a,i} \left\{ \langle 3 \cos^2 \theta_{al} - 1 \rangle + \frac{3}{2} R_i \langle \sin^2 \theta_{al} \cos 2\phi_{al} \rangle \right\}$$ (3.33)

where i is the residue number and j the alignment medium. Because of their orientational dependence, described by the angles θ_{al} and ϕ_{al} with respect to the alignment frame, rdcs D_{ij} carry a wealth of orientation information, for example, for the backbone amide NH group, on which we will focus in the following.

If there is internal dynamics in a protein, this will have an effect on the orientation of the backbone NH group and, due to their orientational dependence, also on the dipolar couplings D_{ij}. Thus, the measured rdcs are dynamically averaged $< D_{ij} >$ [Blackledge, 2005, Tolman and Ruan, 2006]. For the measured rdcs, two different ensemble averaging processes occur: First, the averaging process due to the weak alignment process which is considered to happen in the picosecond time-scale much faster than the overall tumbling correlation time τ_c of a protein. Second the rdcs are averaged because of internal dynamics. According to the *ergodic hypothesis*, well known from statistical mechanics, the time evolution of a system can be described by an ensemble of systems. All different dynamic conformations of a protein are present at one single point in time. Since there are so many proteins in solution, they can be considered as an ensemble sampling the time trajectory of the protein. All different dynamic conformations will align and give rise to the rdc signal. In the presence of local internal motion the measured dipolar coupling can be represented as :

$$< D_{ij} >= \sum_n p_n D_{ij,n} \tag{3.34}$$

where p_n represents the respective weighting of the n conformational sub-states, whose individual dipolar couplings $D_{ij,n}$ are defined by equation 3.33. Thus, RDCs sample the dynamics of the considered inter nuclear vector, e.g. the NH vector, with respect to the alignment frame. Since the alignment frame is fixed to the molecular frame, the lower rate limit for the RDC conformational averaging process is the difference of the RDCs in two or more exchanging conformations. Because RDCs are not larger than on the order of 10 Hz, averaging up to approximately 10 to 100 ms would result. However, since RDCs are observed on proton or nitrogen resonances whose chemical shifts tend to be more affected by chemical exchange than the RDCs, they average up to approximately 1 ms. Our goal is now to derive S^2 order parameters from RDCs that are sensitive for all motion faster than the millisecond time-scale. This will be discussed in the next chapter. RDCs are especially valuable since they can provide information complementary to relaxation methods, in particular in the supra-τ_c time window between τ_c and 50μs.

Chapter 4

The RDC-based model-free approach

4.1 Introduction

After the renaissance of RDCs in liquid-state protein NMR spectroscopy [Tolman et al., 1995, Tjandra and Bax, 1997], the potential of RDCs to study protein dynamics was recognized quickly [Tolman et al., 1997] and several methods have been developed to extract dynamic information from RDCs. In 2001, a mathematical method, the so-called *RDC-based model-free approach*, was derived that extracts structural as well as dynamic information from RDCs [Meiler et al., 2001]. Using a high-resolution structure to determine the alignment tensors, RDC-based order parameters, directions and amplitude of the anisotropy of motion, as well as average inter-nuclear vector orientations were extracted from the measurement of RDCs in at least five linear independent alignment media. Thus, motions in a previously hidden time window between the correlation time τ_c (about 4 ns for ubiquitin at room temperature) and 50 μs (supra-τ_c time window) became accessible by the use of RDCs, as explained in the previous chapter (3). The first experimental studies on NH RDCs for the protein ubiquitin using this novel technique detected a significant amount of motion slower than τ_c [Peti et al., 2002]. These results were supported in an independent work by Joel Tolman and colleagues [Tolman, 2002, Briggman and Tolman, 2003] on the same protein using the *Direct Interpretation of Dipolar Couplings (DIDC) approach* [Tolman, 2002]. Also the DIDC approach relies on the measurement of RDCs in five linear alignment conditions but does not require a structural model of the protein. Both the structural and dynamical models as well as the underlying alignment tensors are obtained simultaneously. However, in order to become in-

dependent from a structural model, the DIDC approach has to rely on an initial assumption about the protein dynamics by minimizing the variation of RDC-based order parameters (see, e.g., [Tolman, 2002] Equation 15). This is also true for a recent extension of this approach [Yao et al., 2008].

Both the RDC-based model-free approach and DIDC approach had a qualitative agreement in the finding of additional motion slower than τ_c (supra-τ_c motion) that has stayed undetected by NMR relaxation methods previously. The finding of supra-τ_c motion questioned the general picture of globular proteins that emerged from NMR relaxation studies, X-ray crystallography and molecular dynamics simulations. According to those methods, backbone motions occurred on a time-scale faster than τ_c and – with the exception of flexible loop regions – were of small overall amplitudes, described to full extent by Lipari-Szabo S_{LS}^2 order parameters. Larger motions could occur on slower time scales, but generally represented rare events, such as local unfolding and solvent access for backbone amide exchange and aromatic ring flips [Clore and Schwieters, 2004b]. Thus, the finding of supra-τ_c motion led to an intense debate. Indeed, the existence of supra-τ_c motion has been challenged by Clore and Schwieters [Clore and Schwieters, 2004b] who applied a two-site jump model on the same NH RDC data set as used for the first application of the RDC-based model-free approach [Peti et al., 2002] . The two-site jump model approach is basically a structural ensemble approach with ensemble size $N_e = 2$ [Clore and Schwieters, 2004b]. From a heuristic perspective it represents the simplest description of anisotropic motion. Clore and Schwieters observed a comparable rigid backbone of ubiquitin with $S^2(jump)$ order parameters ≥ 0.85, but some exceptions with substantial anisotropic motion with $S^2(jump)$ order parameters in the 0.3-0.8 range. According to [Clore and Schwieters, 2004b] the experimental RDC data by [Peti et al., 2002] provided no basis for invoking large-scale concerted motions as observed by [Peti et al., 2002, Meiler et al., 2003]. Additionally, the experimental quality of 3 NH RDC data sets used in the first RDC-based model-free analysis [Peti et al., 2002] was questioned.

Thus, by the time this PhD work started two very opposed interpretations of the same experimental RDC data by [Peti et al., 2002] were discussed. The debate boiled down to the question whether the results obtained by the RDC-based model-free approach were artifacts due to cumulative errors and the effect of structural noise (inaccuracy in the coordinates of the structure used for alignment tensor calculation) or whether they, indeed, showed substantial dynamics slower than τ_c both in loop regions and secondary structure elements. A rigorous mathematical re-evaluation of the RDC-based model-free approach and a systematic enlargement and re-measuring of the experimental basis of NH RDC data for ubiquitin was needed to reduce the effects of experimental uncertainties.

This was the starting point of this PhD work. In the following the theoretical background of the RDC-based model-free approach will be rigorously mathematically proven. Afterwards, it will be explained how the experimental basis was significantly enlarged and how the effect of cumulative errors could be reduced.

4.2   Theory

4.2.1   RDC-based order parameters S^2_{rdc}

The size of rdcs observed in an alignment medium i can be described by the dipolar coupling equation in the alignment frame (cf. equation 3.33):

$$\langle D_{ij} \rangle = D_{a,i} \left\{ \langle 3\cos^2\theta_{al} - 1 \rangle + \frac{3}{2} R_i \langle \sin^2\theta_{al} \cos 2\phi_{al} \rangle \right\} \tag{4.1}$$

The angular brackets $< \cdot >$ express the dynamic averaging of the angular terms. $D_{a,i}$ is the principal component and R_i the rhombicity of the alignment tensor induced by the alignment medium i. Note, that the following equations assume that the average orientation as well as the dynamics of each NH vector are independent of the alignment medium and are not correlated with the alignment tensor of the whole molecule. That assumption has recently been supported by molecular dynamics simulations [Salvatella et al., 2008]. That means we can work with a single dynamical averaged alignment tensor which is considered to be rigid. Molecular dynamics simulations [Meiler et al., 2001] have shown that isotropic dynamics will only lead to a down-scaling of the principal component of the alignment tensor. By using the definition of the second order spherical harmonics (see for example [Zare, 1988])

$$Y_{20}(\theta, \phi) = \sqrt{\frac{5}{16\pi}} (3cos^2\theta - 1)$$

$$Y_{2\pm2}(\theta, \phi) = \sqrt{\frac{5}{32\pi}} e^{\pm 2i\phi} \sin^2\theta \tag{4.2}$$

we replace the angular terms and express the dipolar coupling equation with time-averaged second order spherical harmonics:

$$\frac{< D_{ij} >}{D_{zz}} = \sqrt{\frac{4\pi}{5}} \left(< (Y_{20}(\theta_{al}, \phi_{al}) > + \sqrt{\frac{3}{8}} R(< Y_{2-2}(\theta_{al}, \phi_{al}) > + < Y_{22}(\theta_{al}, \phi_{al}) >) \right) \qquad .$$

$$\tag{4.3}$$

From the RDCs we want to derive a RDC-based order that is constructed very similar to the Lipari-Szabo order parameter from relaxation [Lipari and Szabo, 1982a], however with the important difference that the rdc-based order parameters are sensitive up to the millisecond time scale — in contrast to Lipari-Szabo order parameters which are only sensitive for motions faster than the overall tumbling correlation time τ_c (about 4ns for ubiquitin at room temperature) [Blackledge, 2005, Tolman and Ruan, 2006]:

$$S^2_{LS} = \frac{4\pi}{5} \sum_{M=-2}^{2} \langle Y_{2M}(\theta, \phi) \rangle \langle Y^*_{2M}(\theta, \phi) \rangle \Big|_0^{\tau_c} \qquad (4.4)$$

$$S^2_{rdc} = \frac{4\pi}{5} \sum_{M=-2}^{2} \langle Y_{2M}(\theta, \phi) \rangle \langle Y^*_{2M}(\theta, \phi) \rangle \Big|_0^{ms} \qquad (4.5)$$

4.2.2   Transformation to the molecular frame

To calculate order parameters, first the time averaged second order spherical harmonics $Y_{2M}(\theta, \phi)$ have to be determined. Therefore, the dipolar coupling equations are transformed to a common single reference frame that can be arbitrarily chosen, for example, the pdb-frame of the structure used for alignment tensor calculation. The new reference frame will be called molecular frame in the following. The transformation of the second order spherical harmonics from the alignment frame $Y_{2M}(\theta_{al}, \phi_{al})$ towards the molecular frame $Y_{2M}(\theta_{mol}, \phi_{mol})$ is done by a Wigner rotation $\mathbf{P}_R(\alpha, \beta, \gamma)$ that is defined as:

$$Y_{2M}(\theta_{al}, \phi_{al}) = \mathbf{P}_R(\alpha, \beta, \gamma) Y_{2M}(\theta_{mol}, \phi_{mol}) = \sum_{M'=-2}^{2} D_{M'M}(\alpha, \beta, \gamma) Y_{2M'}(\theta_{mol}, \phi_{mol}) =$$

$$\sum_{M'=-2}^{2} e^{-i\alpha M'} d^{(2)}_{M'M}(\beta) \cdot e^{-i\gamma M} Y_{2M'}(\theta_{mol}, \phi_{mol})$$

$$(4.6)$$

The angles α, β and γ are the Eulerian angles describing the orientation of the alignment frame with respect to the molecular frame. The Eulerian angles are defined in the standard way of connected rotations with respect to a newly rotated reference frame:

$$\tilde{R}(\alpha, \beta, \gamma) = \tilde{R}_{z''}(\gamma) \tilde{R}_{y'}(\beta) \tilde{R}_z(\alpha) \qquad (4.7)$$

However, very often it is convenient to express the rotations within a single reference frame, for example with respect to the molecular frame. As shown in for example in [Zare, 1988] it holds:

$$\tilde{R}(\alpha, \beta, \gamma) = \tilde{R}_{z''}(\gamma)\tilde{R}_{y'}(\beta)\tilde{R}_z(\alpha) = \tilde{R}_z(\alpha)\tilde{R}_y(\beta)\tilde{R}_z(\gamma) = \tilde{R}_{fixed}(\gamma, \beta, \alpha) \quad (4.8)$$

In order to choose the correct Eulerian angles when using information from other programs, for example PALES [Zweckstetter and Bax, 2000] or DipoCoup [Meiler et al., 2000], that determine the alignment tensor by fitting the experimental RDCs to a pdb structure, it is extremely important to be aware of different conventions used in different programs and textbooks. We refer to the convention used by [Zare, 1988] who uses a right-handed coordinate system with counterclockwise rotation. (Reduced Wigner matrix elements $d^{(2)}_{M'M}(\beta)$ have been calculated according to formula (3.57) in [Zare, 1988]. Note that the tabulated values in [Zare, 1988] show some typos. Values tabulated in [Buckmaster, 1964, Buckmaster, 1966] are consistent with the formula (3.57) derived in [Zare, 1988]. We use the Wigner-Rotation to do a passive transformation from the alignment-frame coordinates to the molecular-frame coordinates $R(\alpha, \beta, \gamma)$. This passive transformation is inverse to the active transformation $\tilde{R}(\alpha, \beta, \gamma)$ that rotates the alignment frame axes to the molecular frame axes frame axes: $R(\alpha, \beta, \gamma) = (\tilde{R}(\alpha, \beta, \gamma))^{-1}$

PALES uses the following rotation matrix in a clockwise(!)sense:

$$\begin{pmatrix} x_{al} \\ y_{al} \\ z_{al} \end{pmatrix}_{mol} = R'(\alpha_{Pales}, \beta_{Pales}, \gamma_{Pales}) \begin{pmatrix} x_{mol} \\ y_{mol} \\ z_{mol} \end{pmatrix}_{mol}$$

$$= \begin{pmatrix} \cos\alpha\cos\beta\cos\gamma - \sin\alpha\sin\alpha & \sin\alpha\cos\beta\cos\gamma + \cos\alpha\sin\gamma & -\sin\beta\cos\gamma \\ \sin\alpha\cos\beta\cos\gamma - \sin\alpha\cos\gamma & -\sin\alpha\cos\beta\sin\gamma + \cos\alpha\cos\gamma & \sin\beta\sin\gamma \\ \cos\alpha\sin\beta & \sin\alpha\sin\beta & \cos\beta \end{pmatrix} \begin{pmatrix} x_{mol} \\ y_{mol} \\ z_{mol} \end{pmatrix}_{mol}$$

Thus, the PALES rotation acts as an active transformation and rotates the molecular frame axes to the alignment frame axes while both axes systems are still defined in the molecular frame. This rotation is inverse to $\tilde{R}(\alpha, \beta, \gamma)$ and uses a clockwise rather than a counterclockwise rotation. A second important difference is that it is defined with respect to a fixed reference frame. With 4.8 we get:

$$R(\alpha_{mf}, \beta_{mf}, \gamma_{mf}) = (\tilde{R}(\alpha_{mf}, \beta_{mf}, \gamma_{mf}))^{-1} = R'(-\gamma_{Pales}, -\beta_{Pales}, -\alpha_{Pales}) \qquad . \tag{4.9}$$

where the subscript mf stands for the described implementation of the model-free approach and $Pales$ for the PALES output after alignment tensor calculation.

By comparison we identify:

$$\begin{aligned}
\alpha_{mf} &= -\gamma_{Pales} \\
\beta_{mf} &= -\beta_{Pales} \\
\gamma_{mf} &= -\alpha_{Pales}
\end{aligned} \tag{4.10}$$

Note that the PALES output does not stick to the conventional nomenclature for Eulerian angles but specifies them in a fixed reference frame. That is why α_{mf} is identified with $-\gamma_{Pales}$, etc. as derived. For DIPOCOUP we can analogously identify:

$$\begin{aligned}
\alpha_{mf} &= \gamma_{Dipocoup} \\
\beta_{mf} &= \beta_{Dipocoup} \\
\gamma_{mf} &= \alpha_{Dipocoup} \quad .
\end{aligned} \tag{4.11}$$

4.2.3 The F-matrix

With the transformation from the alignment frame coordinates to the molecular frame coordinates at hand we can now express equation (4.3) in the molecular frame:

$$\frac{D_{ij}^{exp}}{D_{i,zz}} = \sum_{M=-2}^{2} F_{i,M} < Y_{2M}(\theta_j^{mol}, \phi_j^{mol}) > \tag{4.12}$$

with

$$F_{i,M} = \sqrt{\frac{4\pi}{5}} \left(D_{M0}^2(\alpha^i, \beta^i, \gamma^i) + \sqrt{\frac{3}{8}} R\Big(D_{M2}^2(\alpha^i, \beta^i, \gamma^i) + D_{M-2}^2(\alpha^i, \beta^i, \gamma^i) \Big) \right)$$

$$= \sqrt{\frac{4\pi}{5}} \left(e^{-iM\alpha^i} d_{M0}^{(2)}(\beta^i) + \sqrt{\frac{3}{8}} R\Big(e^{-iM\alpha^i} d_{M2}^{(2)}(\beta^i) e^{-2i\gamma^i} + e^{-iM\alpha^i} d_{M-2}^{(2)}(\beta^i) e^{2i\gamma^i} \Big) \right) \tag{4.13}$$

where i is the alignment medium and j is the residue number.

For each residue j we obtain a system of K linear equations for the K different alignment conditions and the 5 time-averaged second order spherical harmonics as unknown variables $< Y_{2M}(\theta_j^{mol}, \phi_j^{mol}) >$.

We can express this system of linear equations by a matrix equation which we call F-Matrix equation:

$$\begin{pmatrix} \frac{D_{1j}^{exp}}{D_{1,zz}} \\ \vdots \\ \frac{D_{kj}^{exp}}{D_{k,zz}} \end{pmatrix} = \begin{pmatrix} F_{1-2} & \dots & F_{12} \\ & \ddots & \\ F_{k-2} & \dots & F_{k2} \end{pmatrix} \begin{pmatrix} < Y_{2-2}(\theta_j^{mol}, \phi_j^{mol}) > \\ \vdots \\ < Y_{22}(\theta_j^{mol}, \phi_j^{mol}) > \end{pmatrix} \tag{4.14}$$

The Pseudo-Inverse of the F-Matrix can be calculated with Singular Value Decompostion (SVD). (For an introduction to singular value decomposition see for example [Werner, 1992a, Werner, 1992b]). Left-Multiplication with the Pseudo-inverse yields the unknown time-averaged second order spherical harmonics. Considering the relationship between second order spherical harmonics $Y_{22}(\theta, \phi)^* = Y_{2-2}(\theta, \phi)$ and $Y_{21}(\theta, \phi)^* = -Y_{2-1}(\theta, \phi)$ the complex F-Matrix equation remains a five dimensional problem. Therefore, at least five different alignments have to be measured in order to determine the five spherical harmonics $< Y_{2M}(\theta_j^{mol}, \phi_j^{mol}) >$. From these, the rdc-based order parameter S_{rdc}^2 is calculated.

4.2.4   The overall scaling problem

Since all measured rdcs are intrinsically scaled by motional averaging, their overall fitting to a single rigid NMR structure necessarily yields a motionally averaged alignment tensor characterized by $\tilde{D}_{i,zz}$ and $\tilde{R}_i$ as well as the angles $\tilde{\alpha}_i$, $\tilde{\beta}_i$ and $\tilde{\gamma}_i$. As has been pointed out in [Meiler et al., 2001], the orientation $(\tilde{\alpha}_i, \tilde{\beta}_i, \tilde{\gamma}_i)$ and the rhombicity $\tilde{R}_i$ of the scaled tensor are virtually indistinguishable from those of the true tensor and the motion is reflected only in a scaling of the principal value according to:

$$\tilde{D}_{i,zz} = S_{overall} \cdot D_{i,zz} \qquad . \tag{4.15}$$

However, for the RDC-based model-free analysis of dynamics a static alignment tensor has to be invoked because all dynamics are represented by the time-averaged spherical harmonics by definition. That is why the static alignment tensor (no tilde) has to be estimated from the experimentally determined dynamically averaged one (indicated by tilde). It turned out that the $\tilde{D}_{i,zz}$ is determined such that the average over the $S_{rdc,unscaled}^2$ equals 1 (when taking all residues for alignment tensor calculation), because the average dynamics is absorbed in the alignment tensor [Meiler et al., 2001]. Therefore, the $S_{rdc,unscaled}^2$ provide only relative values for the dynamic amplitudes, but have to be scaled against the Lipari-Szabo S_{LS}^2 order parameters which contain absolute mobility information given a fixed distance between the amide proton and nitrogen and assuming a constant chemical shift anisotropy.

Using definition 4.15, equation 4.12 can be rewritten:

$$\frac{D_{ij}^{exp}}{\tilde{D}_{i,zz}} \cdot S_{overall} = \sum_{M=-2}^{2} F_{i,M} < Y_{2M}(\theta_j^{mol}, \phi_j^{mol}) > \tag{4.16}$$

This scaling leads to scaled spherical harmonics
$< Y_{2M}(\theta, \phi) >_{(scaled)} = S_{overall} < Y_{2M}(\theta, \phi) >_{(unscaled)}$ since the spherical harmonics are determined by inverting equation (4.16). It should be noted that it is impossible to obtain the value of $S_{overall}$ directly from the RDCs. However, a reasonable estimate can be obtained as will be discussed in the paragraph *Estimation of $S_{overall}$* . For the following, the superscript mol will be omitted: θ and ϕ will be used when referring to the molecular frame of the protein. Unlike the Lipari-Szabo order parameter S^2_{LS} from relaxation, S^2_{rdc} reflects motion up to the time scale that is defined by the inverse of the differences of chemical shifts of RDCs of exchanging conformations (normally in the ms range). Therefore, it must hold $S^2_{LS} \geq S^2_{rdc}$ within the experimental error, which takes into account all motions contained in the Lipari-Szabo order parameter, such as vibrations. The RDC-derived order parameter must always be smaller than the Lipari-Szabo derived one as the former picks up motion on all time scales up to the ms range while the latter does not. An upper limit for the scaling factor $S_{overall}$ in equation (4.16) has to be estimated by requiring $S^2_{LS} \geq S^2_{rdc}$ for every individual amino acid in ubiquitin. If the $S_{overall}$ scaling was neglected in the above set of equations the derived $< Y_{2M}(\theta, \phi) >_{(unscaled)} = < Y_{2M}(\theta, \phi) >_{(scaled)} / S_{overall}$ and $S_{rdc(unscaled)} = S_{rdc(scaled)} / S_{overall}$ would be increased by a factor $1/S_{overall}$. We will refer to these values as *unscaled*. The scaling factor $S_{overall}$ can be estimated by analyzing the distribution of $S^2_{LS}/S^2_{rdc(unscaled)}$ values (see below). It should be noted that it is intrinsically impossible to derive the value of $S_{overall}$ directly from the RDCs.

4.2.5 Average inter-nuclear vector orientation (θ, ϕ), amplitude η and direction of anisotropy ϕ'_{rdc}

Besides allowing the extraction of S^2_{rdc} in a model-free way, the $< Y_{2M}(\theta, \phi) >$ values provide a detailed picture of the on-going motional averaging. To determine the dynamical averaged orientation of each individual NH vector, the coordinate system is transformed by a Wigner-Rotation into a new (primed) frame such that $< Y_{20}(\theta', \phi') >$ is maximized:

$$Max[< Y_{20}(\theta', \phi') >] = \sum_{M=-2}^{2} D^2_{M0}(\phi_{av}, \theta_{av}, 0) < Y_{2M}(\theta, \phi) >= \qquad (4.17)$$

$$= \sqrt{\frac{4\pi}{5}} \sum_{M=-2}^{2} Y^*_{2M}(\theta_{av}, \phi_{av}) < Y_{2M}(\theta, \phi) > \qquad (4.18)$$

Maximizing $< Y_{20}(\theta', \phi') >$ places the new z axis into the center of the distribution for the given NH vector and thus defines the polar angles (θ_{av}, ϕ_{av}) , and results

in $< Y_{21}(\theta', \phi') >$ and $< Y_{2-1}(\theta', \phi') >$ becoming equal to zero. The polar angles (θ_{av}, ϕ_{av}) describe the average NH vector orientation.

$< Y_{22}(\theta', \phi') >$ and $< Y_{2-2}(\theta', \phi') >$ represent the asymmetry of the motion. To obtain a better understanding of the asymmetric part of motion, the amplitude η of anisotropic motion is introduced:

$$\eta_{rdc} = \sqrt{\frac{\sum_{M=-2,2} < Y_{2M}(\theta', \phi') >< Y_{2M}^*(\theta', \phi') >}{\sum_{M=-2,0,2} < Y_{2M}(\theta', \phi') >< Y_{2M}^*(\theta', \phi') >}} \qquad (4.19)$$

The direction of the anisotropic motion is described by the angle ϕ'_{rdc} with respect to the x' axis in the $x'y'$ plane of the primed frame. This angle ϕ'_{rdc} can be determined by application of an Euler rotation $R(\phi_{av}, \theta_{av}, \phi'_{rdc})$ that transforms the molecular frame axes into the axes of a (doubly primed) frame, so that the real part of the second order spherical harmonic $< Y_{22}(\theta'', \phi'') >$ becomes maximized in that double primed frame:

$$Max[Re[< Y_{22}(\theta'', \phi'') >]] = R(\phi_{av}, \theta_{av}, \phi'_{rdc}) < Y_{22}(\theta, \phi) > \qquad . \qquad (4.20)$$

Furthermore, the ϕ'_{rdc} dependence of $< Y_{22}(\theta', \phi') >$ and $< Y_{2-2}(\theta', \phi') >$ results in a π-periodicity of ϕ'_{rdc}. The translation of the five averaged spherical harmonics into the primed coordinate system amounts to the definition of five new parameters, namely: $< Y_{20}(\theta', \phi') >$ which reflects the axial order, θ_{av} and ϕ_{av} which represent the average orientation of the vector, and η_{rdc} and ϕ'_{rdc} which reflect the amount of anisotropic disorder and the direction of this anisotropic motion in the plane. Note that θ_{av}, ϕ_{av}, η_{rdc} and ϕ'_{rdc} are not affected by the scaling procedure and therefore also not sensitive to errors in estimating $S_{overall}$. In order to visualize the direction of anisotropy more easily, the angle ξ'_{rdc} is defined as the direction of the NCα vector in the x'', y'' plane (compare Fig. 4.1):

$$\xi'_{rdc} = \arctan \frac{y''_{NC\alpha}}{x''_{NC\alpha}} \qquad (4.21)$$

where x'' and y'' are the coordinates of the NCα vector in the doubly primed frame. The angle ξ'_{rdc} can be regarded as the direction of the anisotropy with respect to the peptide plane. In contrast to ϕ'_{rdc}, ξ'_{rdc} is no longer model-free, since ξ'_{rdc} is calculated using the NMR ubiquitin structure (1d3z) for the definition of the peptide plane.
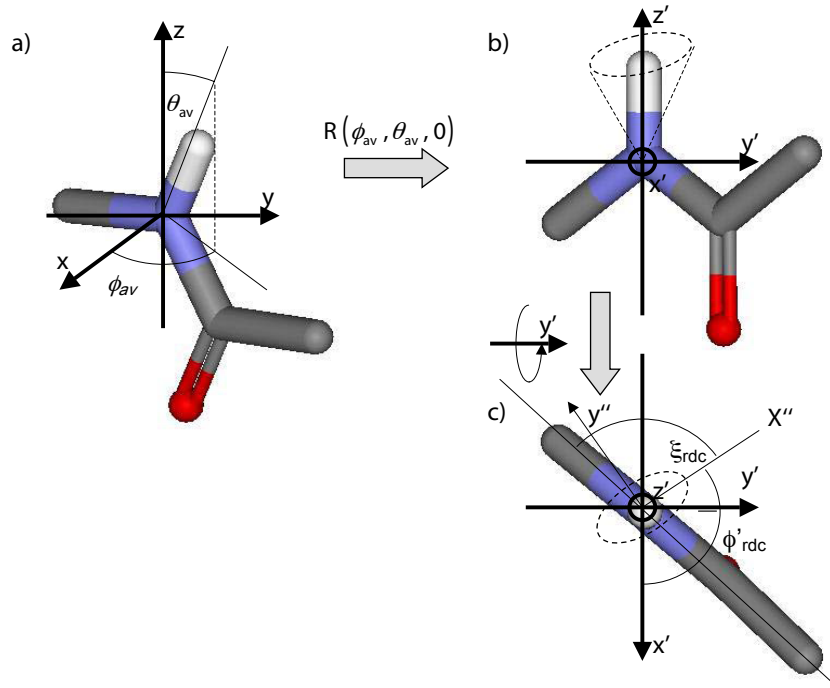
Figure 4.1: The angles (θ_{av}, ϕ_{av}) in panel (a) describe the effective orientation of the NH vector in the molecular frame x, y, z. θ_{av} and ϕ_{av} are determined by maximizing $< Y_{20} >$ and describe therefore also the transformation of the molecular frame x, y, z into the primed coordinate system x′, y′, z′. It is visualized in panel b) along the x′-axis and in panel c) along the y′ axis. The NH vector points along the z′-axis. The dotted cone illustrates the dynamics of the NH vector in terms of the model of diffusion in an elliptical cone. The direction of asymmetry is either measured model-free with respect to the coordinate system x′, y′, z′ as ϕ'_{rdc} , or utilizing a structure of ubiquitin with respect to the peptide plane as ξ'_{rdc}.

4.3   Materials and methods

4.3.1   Sample preparation

Compared to previous work [Peti et al., 2002] in which 11 different conditions for alignment were used, the data basis was substantially broadened by performing the following experiments: A first data set was recorded using five alignment media: DMPC/DHPC bicelles (D1), CHAPSO/DLPC/SDS (D2), n-dodecyl-

penta(ethylene glycol)/n-hexanol (D3), purple membrane fragments (D4) and Pf-1 phages (D5). The second set of alignment media included polyacrylamide gel (E1), cetylpyridinium bromide/n-hexanol (E2), n-dodecyl-penta(ethylene glycol)/n-hexanol (E3), purple membrane fragments (E4) and Pf-1 phages (E5).

^{15}N, ^{13}C-labeled human ubiquitin (wt) was purchased from VLI Research, Inc. (Malvern, PA) and used without further purification for the data sets D1, D2, D3, D4, D5, E1 and E4. For the data sets E2, E3 and E5 ^{15}N, ^{13}C-labeled human ubiquitin was used, expressed according to Handel et al. [Johnson et al., 1999]. Both ubiquitin samples gave identical Heteronuclear Single Quantum Correlation (HSQC)-spectra (vide infra). To estimate the strength of alignment, the water ^{2}H quadrupolar splitting (QS) has been measured for each alignment medium.

Preparation of the first set:

For the first set of alignment media, all samples contained 0.4 mM to 0.6 mM of ^{15}N, ^{13}C-labeled human ubiquitin in 10mM Na phosphate buffer, pH = 6.5, including 10% of D$_2$O. Two bicelle media, DMPC:DHPC = 3:1 (5 % w/v, Avanti Polar Lipids, sample D1) [Cornilescu et al., 1998] and CHAPSO:DLPC:SDS = 125:25:1 (5 % w/v, Avanti Polar Lipids, sample D2) [Losonczi and Prestegard, 1998] were prepared. Sample D3 contained dodecyl-penta(ethylene glycol) (C12E5) and n-hexanol in 10mM Na phosphate buffer, pH = 6.5. The C12E5 surfactant-to-water ratio was 5% (w/w) and the molar ratio of surfactant/alcohol was 0.985 [Ruckert and Otting, 2000]. Sample D4 was prepared by adding purple membrane fragments at 4 mg/ml to the protein solution. Additionally, 100 mM NaCl (final concentration) was added to reduce electrostatic interactions between purple membrane and ubiquitin [Koenig et al., 1999]. For the fifth alignment condition (sample D5), 17 mg/ml Pf-1 Phages (ASLA Ltd., Riga, Latvia) were dissolved in 10 mM Na phosphate buffer, pH =6.5. Additionally, NaCl was added up to a final concentration of 350 mM in order to reduce electrostatic interactions between phages and ubiquitin [Zweckstetter and Bax, 2001]. A 0.4 mM isotropic solution of 15N, 13C-labeled human ubiquitin was used for reference (D6).

Preparation of the second set:

For all samples of the second set of alignment media the concentration of ^{15}N, ^{13}C-labeled human ubiquitin was increased to 0.6 - 1.1 mM. The Na phosphate buffer, pH = 6.5, was increased from 10 mM to 50mM and 10% to 15% D$_2$O was added. Six samples, termed E1-E6, were prepared. In sample E1 the protein solution was soaked into a 7% uncharged polyacrylamide gel. In these gels alignment is achieved by radially compressing and stretching the gel in the direction of the magnetic field [Tycko et al., 2000, Sass et al., 2000]. Alignment in sample E2 was achieved by preparing a 3.5% (w/v) solution of cetylpyridinium bromide and n-hexanol in 50 mM Na phosphate buffer, pH = 6.5, and 25 mM NaBr. The molar

ratio of CPBr and n-hexanol was 1:1 [Barrientos et al., 2000]. Sample E3 was very similar to sample D3, except for the Na phosphate buffer concentration, which was 50mM instead of 10mM. Sample E4 was prepared similar to sample D4 except increasing the concentration of the Na phosphate buffer from 10mM to 50mM and decreasing the concentration of NaCl from 100mM to 50mM. For sample E5 the Pf-1 Phage concentration was reduced to 15mg/ml (compared to 17mg/ml in sample D5), whereas the NaCl concentration was increased from 350mM to 400mM. For reference a 50mM Na phosphate buffer (pH= 6.5) with 1.8mM 15N,13C-labeled ubiquitin was used as isotropic solution (Sample E6).

4.3.2   NMR spectroscopy

NMR experiments for the first dataset were performed on a Bruker-avance 700 MHz spectrometer (Bruker AG, Karlsruhe, Germany) at 308 K. A two dimensional ^{15}N, ^{1}H-HSQC experiment was run to measure NH rdcs. The time domain was TD1 x TD2= 1k x 2k. For the second data set all NMR experiments were performed on Bruker-DRX-600 MHz and Bruker-avance-800 MHz spectrometers (Bruker AG, Karlsruhe, Germany). To measure the NH rdcs of ubiquitin in the different alignment media 2D- inphase/antiphase (IPAP)-^{15}N, ^{1}H- HSQC experiments [Ottiger et al., 1998] were recorded for each aligned sample (D1 to D5 and E1 to E5) as well as for the isotropic sample (D6 and E6). All IPAP-HSQC spectra were recorded in the interleaved mode at 308 K. Temperatures were calibrated by using 1,2-ethanediol. The temperature-dependent chemical shift difference between the OH protons and those of the methylene groups was measured. For the Bruker-DRX-600 MHz spectra the time domain was TD1 x TD2 = 512 x 2048 complex points, the spectral width F1 x F2=1818 Hz x 8389 Hz, NS = 48 and the total experimental time 25 h, for the Bruker-avance-800 MHz spectra the time domain was TD1 x TD2 = 768 x 2048, the spectral width F1 x F2= 2432 Hz x 11160 Hz, NS = 32 and the total experimental time 20h. After zero-filling to TD1x TD2 = 32k x 4k and processing the spectra the couplings were extracted using the NMRPipe software package [Delaglio et al., 1995]. Alignment tensors were calculated using the DipoCoup software [Meiler et al., 2000]. One bond ^{15}N, ^{1}H rdcs were derived from the difference in splitting between the aligned samples (1-5) and the isotropic state (6). A conservative estimate for the experimental error is 0.3 Hz for the measured NH rdc, since the digitalization is 0.055 Hz/point after zero-filling. As shown in Figure 2, the correlation of measured rdcs versus back-calculated values is excellent. Using the NMR-structure 1d3z [Cornilescu et al., 1998] for back-calculation the correlation coefficient is ρ =0.99, the Q-value varies between 0.10 for polyacrylamide gel and 0.14 for cetylpyridinium bromide/ n-hexanol. When the X-ray structures (1ubi [Ramage et al., 1994], 1ubq [Vijay-Kumar et al., 1987]

are used for back-calculation, the correlation decreases only slightly, while the maximum Q-value increases to 0.19.
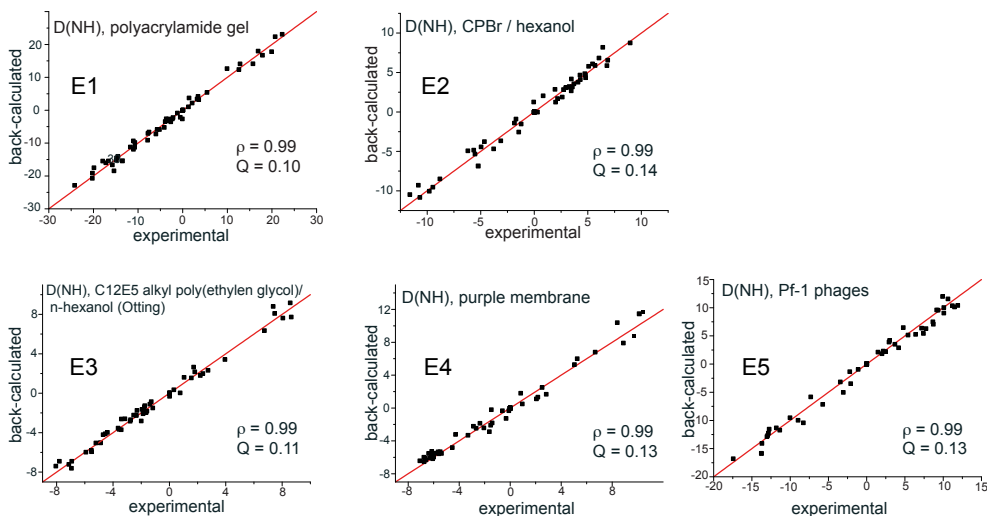
Figure 4.2: Correlation of experimental residual dipolar couplings with values that were back-calculated from a single static structure (1d3z) for a set of five different alignment media. An excellent agreement between a single structural model and the data is already observed (Q-values between 0.10 and 0.14). However, the small deviations are not only experimental uncertainty but also reflect dynamics. The alignment media shown are polyacrylamide gel (E1), cetylpyridinium bromide/n-hexanol (E2), n-dodecyl-penta(ethylene glycol)/n-hexanol (E3), purple membrane fragments (E4), and Pf-1 phages (E5).

4.3.3   Input for analysis

For the first rdc-based model free analysis by Peti et al. [Peti et al., 2002] (A), 11 sets of NH rdcs were used. 9 more data sets were published by Tolman and coworkers [Briggman and Tolman, 2003](B) and one dataset, obtained in polyacrylamide gel, was taken from the literature [Sass et al., 2000, Tycko et al., 2000] (C). For the presented work the experimental basis was significantly extended by the measurements described in the experimental section yielding two sets of five different alignment conditions (D and E). Altogether, a total of 31 experimental datasets were available for the analysis. By using a large number of experimental data sets as input for the analysis, the experimental noise could be reduced and the accuracy of the inversion of the matrix improved. This procedure resulted in more accurate RDC-derived order parameters. Even if some alignment conditions

are very similar we find it desirable to use as many RDC-datasets as available to ensure the convergence of the order parameters.

Quality measures and error assessment

Out of the original 11 datasets used in the previous work (Peti et al., 2002), four (CHAPSO/DLPC, CHAPSO/DLPC/SDS, purple membrane, and polyacrylamide gel) were not used for the new analysis (leaving a total of 27 data sets), as more accurate measurements have since then become available in these cases, three out of these four data sets had been marked as less accurate by [Clore and Schwieters, 2004b]. In addition, the rather large Q-values for these datasets suggested relatively high experimental uncertainties as already identified by [Clore and Schwieters, 2004b]. To identify RDC data affected by high experimental noise, the complete matrix of all RDCs for all 76 amino acids in all 27 alignment media was analyzed with singular value decomposition, as proposed by Tolman [Tolman, 2002]. In such an analysis the first five eigenvalues represent the five structural as well as dynamical degrees of freedom (compare above), while all other eigenvalues reflect noise. A matrix of RDCs was back-calculated using only the largest five eigenvalues and respective eigenvectors. Those RDC values that deviated by more than 12% of D_{zz} (three times the standard deviation) from the back-calculated ones were not considered further. About 3% of all data points were removed by this procedure. More than 1400 measured NH couplings have been used for the analysis. To evaluate the influence of each single data set on our analysis each of the 27 remaining data sets was consecutively removed and the analysis was performed without that dataset. This has been done using tensors derived from three different structures, two X-ray (PDB codes 1ubi, 1ubq) as well as one NMR structure (PDB code 1d3z) to evaluate the influence of structural noise on the analysis. Finally the average and standard deviations of all cycles were taken as results for the model free parameters. All protocols used for this analysis were implemented with the MATHEMATICA5 software package. The RDC-based model-free results were accepted as long as the 5th singular value of the F-matrix was larger than 0.6.

4.3.4   Motion of the α-helix

For the α-helix an aniostropic concerted motion has been suggested [Meiler et al., 2003] and observed by high-pressure NMR [Kitahara et al., 2005]. In order to compare model-free derived directions of anisostropy ϕ'_{rdc} to the high-pressure results, the high and low pressure ubiquitin structure were interpreted

as two different states of a two-side jump model and directions of anisotropy ϕ'_{rdc} were derived: Both the low pressure and high pressure structures 1V80 (30 bar) and 1V81, (3 kbar) [Kitahara et al., 2005] were overlayed with the nmr structure of ubiquitin (1d3z) [Cornilescu et al., 1998] in order to work in the same reference frame. Always the first structure of all ensembles was used. From both structures 1V80 and 1V81, spherical harmonics were calculated from the structures' polar angles and the average was taken as input spherical harmonics for the model-free algorithm to derive the direction of anisotropies, $\phi'_{rdc,HP}$. The data calculated for the α-helix are listed in Table 2 together with the rdc-derived ϕ'_{rdc} values for comparison. These have to be considered in terms of modulo 180°.

4.3.5   Structural noise

The propagation of inaccuracies in the coordinates of the protein structure used for alignment tensor calculation on the alignment tensor calculation itself is called the effect of *structural noise*. To evaluate the influence of structural noise [Zweckstetter and Bax, 2002] on the model-free analysis of RDCs, the rigid structure used to derive the individual alignment tensors was systematically varied. We used three different starting structures including two X-ray structures [Vijay-Kumar et al., 1987, Ramage et al., 1994] and the first model of the NMR 1d3z ensemble [Cornilescu et al., 1998]. As the ten models derived from NMR data (1d3z) are all very similar, only the first of the ten models was used. The X-ray structures (PDBcode: 1ubq and 1ubq) reflect about 5% of structural noise with respect to theNMT structure [Zweckstetter and Bax, 2002].

4.4   Results and discussion

4.4.1   Higher number of datasets reduces noise

The results of a preliminary analysis [Peti et al., 2002] suggested, that the number of available alignment media critically influences the quality of the structural and dynamical parameters derived. Although in theory a set of five well-chosen alignment media is sufficient to determine the two structural parameters as well as the three dynamical parameters, the accuracy of the latter in particular improves significantly using additional experimental datasets. This is due to the small size of the changes caused by motional averaging, which approaches the range of experimental uncertainties. Besides the reduction of experimental noise, an efficient way to separate this noise from real dynamical information resides in

the improvement of the statistics obtained by the measurement of additional experimental data.

4.4.2 Estimation of $S_{overall}$
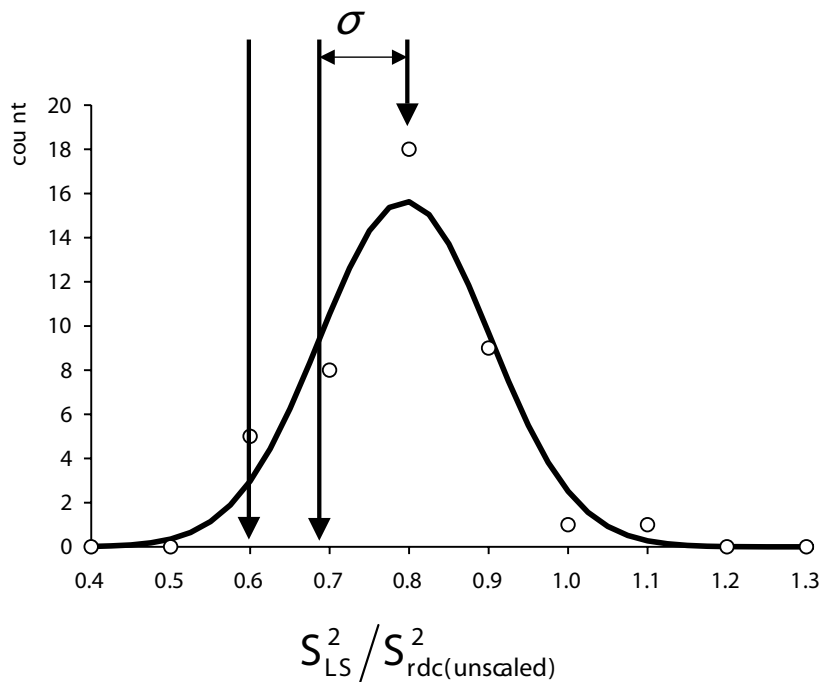
Figure 4.3: $S_{overall}$ is estimated utilizing the inequality $S_{LS}^2 \geq S_{overall}^2 S_{rdc,unscaled}^2$. A histogram analysis of the ratios $S_{LS}^2/S_{rdc,unscaled}^2$ in ubiquitin (circles) was performed. To satisfy this inequality for all data points, $S_{overall}^2$ would need to be as low as 0.59. In order to interpret the data conservatively and consider experimental uncertainties, we added 1 standard deviation ($\sigma = 0.1$) of the distribution of $S_{LS}^2/S_{rdc,unscaled}^2$ values to this value and arrived at 0.69 as $S_{overall}^2$ which corresponds to $S_{overall}$=0.83.

In order to calibrate $S_{overall}$, we assume that the most rigid backbone amide groups with the highest $S_{rdc(unscaled)}^2$ experience no additional motion beyond the correlation time τ_c. For those residues it holds: $S_{LS}^2 = S_{rdc(scaled)}^2 = S_{overall}^2 S_{rdc(unscaled)}^2$. We require $S_{LS}^2 \geq S_{rdc(scaled)}^2 = S_{overall}^2 S_{rdc(unscaled)}^2$ or $S_{LS}^2/S_{rdc(unscaled)}^2 \geq S_{overall}^2$

42

for all amino acids. Taking the minimum of all $S^2_{LS}/S^2_{rdc(unscaled)}$-ratios as a value which is $S^2_{overall} = 0.59$ would ensure that this inequality is always fulfilled. In order to have a statistically more significant basis and to take experimental error into account, we made a histogram analysis of the distribution of $S^2_{LS}/S^2_{rdc(unscaled)}$-ratios (cf. Figure 3) and added the standard deviation $\sigma = 0.1$ of this distribution to the value of $S^2_{overall} = 0.59$ and arrived at $S^2_{overall} = 0.69$ ($S_{overall} = 0.83$) as a very conservative upper limit for $S^2_{overall}$. Note, that the exact shape of the distribution is not relevant for this conclusion. In the current analysis all rdc-based order parameters are (within the error bars) smaller than the Lipari-Szabo counter part, with a few exceptions for Leu8, Asp32, Gln49 and Ser57 that can be explained by the very conservative estimate for $S_{overall}$.

4.4.3 Model free parameters

Residue number	$S^2_{rdc}(NH)$	ϕ_{av}	θ_{av}	ϕ'_{rdc}	η_{rdc}	ξ'_{rdc}
1	n	n	n	n	n	n
2	0.63	-112.2	93.7	-84.8	0.08b	123.8b
3	0.73	30. 4	61.9	-44.9	0.04	76.6
4	0.72	-171.5	118.5	84.2	0.10	89.1
5	0.74	-19.1	64.7	-32.5	0.06	103.8
6	0.71	169.2	97.5	-80.0	0.04	74.4
7	0.59	-43.9	81.7	-75.0	0.14	148.5
8	0.74	101.6	60.5	36.9	0.38	134.2
9	n	n	n	n	n	n
10	n	n	n	n	n	n
11	0.54	96.4	95.6	69.3	0.15	122.0
12	0.61	-36.9	84.9	-50.8	0.15	99.8
13	0.70	152.1	114.2	-76.2	0.02	86.9
14	0.62	1. 6	73.8	-65.4	0.08	98.0
15	0.72	-150.9	122.4	0.7	0.12	39.3
16	0.72	42.3	72.5	39.5	0.06	136.5
17	0.73	-106.2	111.7	-82.6	0.02	169.0
18	0.64	84.0	93.9	77.6	0.08	31. 7
19	n	n	n	n	n	n
20	0.59	-56.2	49.8	85.7	0.10	82.2
21	0.77	-79.3	110.9	-68.6	0.10	134.0
22	n	n	n	n	n	n
23	0.91	162.8	97.3	-26.9	0.11	07. 2

24	n	n	n	n	n	n
25	0.90	-166.6	85.8	26. 4	0.03	103.5
26	n	n	n	n	n	n
27	0.78	157.6	84.2	-4.9	0.03	121.2
28	0.69	-177.0	73.0	71.6	0.11	117.8
29	0.76	-174.5	97.3	-26.7	0.02	115.5
30	0.73	163.9	90.2	-47.3	0.03	36.6
31	n	n	n	n	n	n
32	0.92	-168.8	76.5	46.3	0.05	100.2
33	0.65	177.6	98.7	-75.2	0.03	129.1
34	0.69	169.1	78.9	68.3	0.05	87.8
35	0.71	-140.3	41.3	0.1	0.14	43.0
36	0.67	-111.3	80.5	61.6	0.06	64.9
37	n	n	n	n	n	n
38	n	n	n	n	n	n
39	0.63	28. 7	47.3	-58.8	0.02	119.4
40	0.79	-33.3	41.9	60.5	0.03	115.1
41	0.74	110.7	26. 1	-78.1	0.15	28. 3
42	0.67	-28.0	123.0	-89.1	0.06	48.5
43	0.82	138.9	62.1	29. 7	0.04	80.2
44	0.71	-22.9	96.5	-81.5	0.08	38.5
45	0.87	135.8	95.0	-9.4	0.04	105.7
46	n	n	n	n	n	n
47	0.66	-18.7	84.7	-11.5	0.04	64.9
48	0.72	-70.4	62.3	52.0	0.06	126.0
49	0.84	139.2	120.0	09. 6	0.10	57.5
50	0.62	-16.8	104.0	-90.2	0.06	99.0
51	0.74	-141.9	85.6	-52.1	0.05	105.3
52	0.61	67.0	124.5	46.7	0.09	160.0
53	n	n	n	n	n	n
54	0.51	17. 8	132.2	-26.6	0.06	150.5
55	0.72	-174.4	93.9	-4.4	0.03	113.8
56	0.72	10. 7	38.3	8. 4	0.12	108.8
57	0.93	41.3	35.7	0.0	0.17	169.3
58	0.74	39.2	57.2	-80.7	0.09	147.0
59	0.61	-1.9	49.9	12. 8	0.12	148.7
60	0.76	117.1	29. 4	-74.8	0.13	43.0

61	0.75	95.4	50.4	-75.8	0.07	158.0
62	0.55	-94.7	172.2	47.8	0.29	118.5
63	0.75	-149.7	58.3	-4.0	0.09	124.6
64	0.79	44.6	43.0	-61.1	0.03	27. 6
65	0.63	140.8	21.0	73.6	0.37	86.8
66	0.68	-158.7	130.2	73.1	0.11	89.4
67	0.73	-17.2	55.5	-21.3	0.06	119.5
68	0.72	159.5	89.5	65.2	0.07	70.1
69	n	n	n	n	n	n
70	0.63	152.3	48.2	83.4	0.07	46.2
71	0.59	-48.9	147.0	67.5	0.06	180.3
72	0.59	-49.4	8.5	63.6	0.16	90.7
73	n	n	n	n	n	n
74	0.37	16.5	95.7	-1.4	0.44	80.1
75	n	n	n	n	n	n
76	0.01	23. 6	63.3	-7.1	0.26	90.7

Table1: Model-free parameters $(S^2_{rdc}(NH), \phi_{av}, \theta_{av}, \phi'_{rdc}, \eta_{rdc})$ determined from the analysis and the ξ'_{rdc} values are shown. Presented are the average values using three different structures (1d3z, 1ubi, 1ubq), n means no value available.

In Figure 4 the effective orientation of each NH vector (θ_{av}, ϕ_{av}) is compared with the NH vector orientation found in the ubiquitin structure determined by NMR spectroscopy (PDBcode: 1d3z). All deviations are smaller than 10° except for some ϕ angles when θ is close to 0° or 180°. In these cases, the deviation becomes undefined since the vector is almost parallel with the $\pm z$ axis of the coordinate frame. Therefore, for the deviations of ϕ_{av} from the ϕ value of the NMR structure, $\Delta\phi_{av} \sin\theta_{av}$ was calculated. Since the effective polar coordinates of the NH vectors (θ_{av}, ϕ_{av}) are very similar to the coordinates (θ, ϕ) of the NMR structure (1d3z), this structure can be considered in a first approximation as a good dynamically averaged single-structure representation of ubiquitin.

The residue specific S^2_{rdc} values are shown in Figure 5a as derived from the individually averaged spherical harmonics according to equation (1.4). All secondary structure elements show reduced averaged values compared to the S^2_{LS} Lipari-Szabo order parameters, revealing large amounts of motion slower than the inverse overall tumbling correlation time of the protein. This observation is most prominent for the β-strands and loop regions: The average RDC-based order parameters S^2_{rdc} are 0.77 for the α-helix, 0.71 for the β-sheet and 0.68 for the loop regions, compared to the average S^2_{LS} values 0.83, 0.80 and 0.72. For some loop regions, a higher mobility is already detected on the relaxation time scale. The decrease of order parameters is more pronounced on the S^2_{rdc} time scale
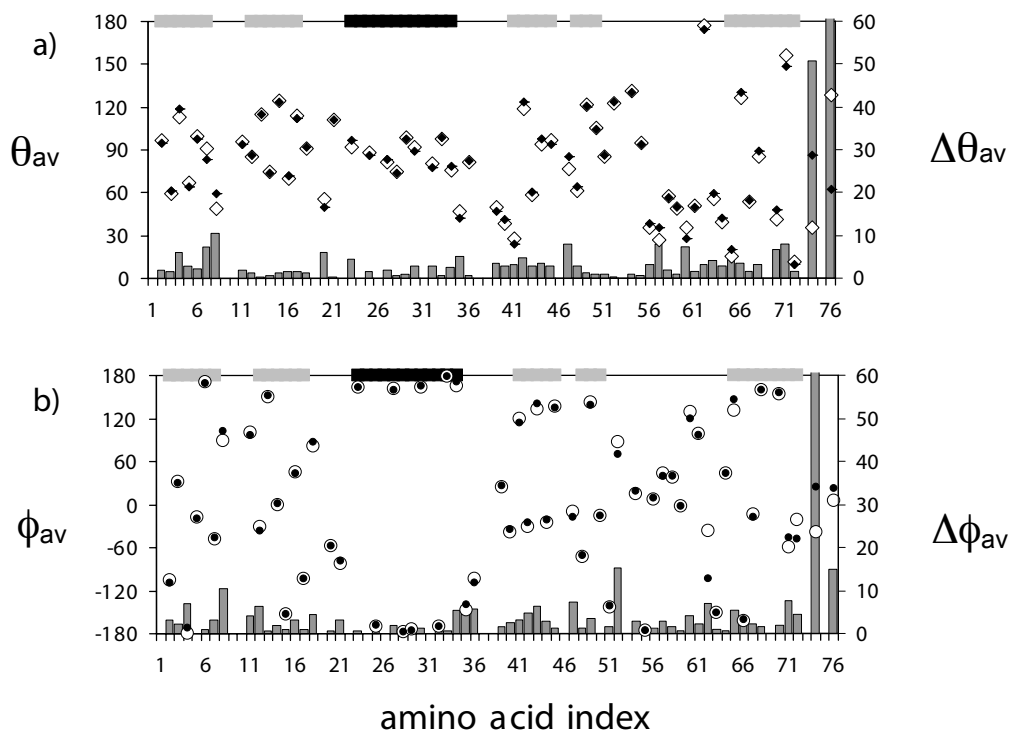
Figure 4.4: The effective orientation of each NH vector described by the polar angles θ_{av} (a) and ϕ_{av} (b) (degrees on the left y-axis) as derived from residual dipolar couplings (filled diamonds/circles) are compared with the ones obtained from the NMR structure 1d3z (open diamonds/circles) [Cornilescu et al., 1998]. The deviations $\Delta\theta_{av}$ (a) and $\Delta\phi_{av}$ (b) between the model-free derived values and the latter ones are plotted in degrees as gray bars (right y-axis). The deviations of ϕ_{av} are scaled with $\sin(\theta_{av})$ to account for the spherical distortion.

suggesting that for most loop regions the amplitudes of fast motions scale with the amplitudes of slow motion. For example, the V-shape of values (Figure 5a) in the region spanning residues 4-16 is deepened and becomes more pronounced in the S^2_{rdc} values.

However, for the α-helix and β-sheet additional new modes of motion beyond the inverse overall tumbling correlation time can be detected by the RDC-based order parameters S^2_{rdc}. While the α-helix is rather rigid on the relaxation time scale (all S^2_{LS} are large and about the same size), RDcs detect differences between individual residues in the helix: Amino acids 23, 25, and 32 have an increased S^2_{rdc} order parameter compared with the remaining amino acids in the α-helix. The N-terminal part of the α-helix appears very rigid, while the mobility increases to-

wards the C-terminal part. Together with the observed directions of anisotropic motion (vide infra) this supports the idea of a slow anisotropic excursion of the helix that has also been described in an earlier publication [Meiler et al., 2003] and has recently been observed by Kithara & al. in high pressure NMR experiments [Kitahara et al., 2005]. They attribute a large population of 20% at normal pressure and 80% at high pressure of the alternative conformation. If the interconversion would be on the microsecond/millisecond time scale, very broad lines would have to be expected. The fact that this is not the case indicates a fast interconversion at least on a time scale of less than 200 μs assuming a change of chemical shifts of at least 1 ppm.

Interestingly, the nature of the side chains seems to influence the backbone amide group mobility. Charged and polar residues reveal decreased S^2_{rdc} order parameters compared to hydrophobic residues, indicating that backbone amide groups are more mobile for charged and polar residues. In particular, the arginines Arg42, 54, 72 and 74 show very small order parameters like 0.67, 0.51, 0.59 and 0.37. These arginines are considered to play important roles in the interaction with the ubiquitin-activating enzyme E1 [Burch and Haas, 1994] and with the ubiquitin-conjugating enzyme E2 [Miura et al., 1999] as well as deubiquinating enzymes [Wilkinson et al., 1999]. A more detailed analysis is given in the next chapter.

Those amino acids that show an extremely large deviation between the S^2_{rdc} and the S^2_{LS} value are of special interest. The difference between S^2_{LS} and S^2_{rdc} of Thr7, Asp20, Arg54, Ser65 and Val70 is greater than 0.20, indicating that the NH vectors of these amino acids have large amplitude motions on a time scale slower than the correlation time. For some of these residues, one can expect motion on a time scale close to where relaxation dispersion measurements become sensitive. Indeed, for Val70 a large dispersion effect was observed at -13°C [Mills and Szyperski, 2002], which corresponds to a conformational exchange rate of 133 μs. Val70 forms a hydrophobic patch together with Leu8 and Ile44 that confers specifity for the binding to different enzymes, [Beal et al., 1996, Haas and Siepmann, 1997].

Whereas large motions are detected for Ile23 and Asn25 in relaxation dispersion measurements [Fushman and Cowburn, 1998, de Alba et al., 1999]
[Mills and Szyperski, 2002, Dittmer and Bodenhausen, 2004, Wist et al., 2004]
[Massi et al., 2005], both residues appear only moderately mobile in our analysis. To explain why the S^2_{rdc} order parameter cannot detect the motion observed in the relaxation dispersion experiments, we hypothesize that the amide chemical shift of Ile23 and Asn25 is modulated either by the breaking of a hydrogen bridge or by side chain reorientation. Both kinds of motion would not change the orientation of the NH vector and thus would not be detected by RDC measurements. Disruption

of the hydrogen bond between Ile23 and Asp54 as a potential mechanism of exchange broadening for Ile23 would lead to a minor state population of $p_B \approx 0.02$ [Massi et al., 2005] and would be undetectable by order parameters.

In Figure 5b) and c) the rdc-derived amplitude η and direction of the asymmetry ϕ_{rdc} (with respect to the peptide plane) are presented and compared with the respective values determined from molecular dynamics [Meiler et al., 2001]. The average amplitude of asymmetry derived from rdc data is higher than the one obtained from the MD trajectory calculated by [Meiler et al., 2001]. Whereas the MD-derived directions of anisotropy have values close to 90° in agreement with the 3D-Gaf model for motions of peptide planes on the Lipari-Szabo time scale [Bremi and Brueschweiler, 1997, Bremi et al., 1997, Brutscher et al., 1997, Lienin et al., 1998], larger deviations occur for the ξ_{rdc} values indicating additional modes of motion in the ns to μs/ms time scale.

4.4.4 Anisotropic motion

residue	$\phi'_{rdc,HP}[°]$	$\phi'_{rdc}[°]$	$S^2_{rdc}(NH)$	η_{rdc}
22	-33.9	n	n	n
23*	68.8	-26.9	0.91	0.11
24	-56.9	n	n	n
25	42.2	26. 4	0.90	0.03
26	-42.1	n	n	n
27	5. 7	-4.9	0.78	0.03
28	-77.6	71.6	0.69	0.11
29	-19.4	-26.7	0.76	0.02
30*	77.5	-47.3	0.73	0.03
31	-2.1	n	n	n
32	40.0	46.3	0.92	0.05
33*	35.8	-75.2	0.65	0.03
34	46.1	68.3	0.69	0.05
35	4. 9	0.1	0.71	0.14

Table 2: Comparison of model-free directions of anisotropy ϕ'_{rdc} and those obtained from high-pressure measurements $\phi'_{rdc,HP}$. The angles have to be considered in terms of modulo 180°, n means no value available. Residues 23, 30, 33 with high deviations between ϕ'_{rdc} and $\phi'_{rdc,HP}$ are indicated by *. Additionally, the $S^2_{rdc}(NH)$ and amplitude of anisotropy η_{rdc} are shown.

Except for the assumption of the independence of structure and dynamics from the alignment medium, the analysis presented here is *model-free*, since no models

about the potential modes of motion of the NH vectors are made. Generally, asymmetries in loop regions are larger on average than in the secondary structure elements. Indeed, the average amplitude of anisotropy is $< \eta_{rdc} > = 0.12$ for loop regions (without C-terminus), whereas the average values are $< \eta_{rdc} >= 0.06$ and $< \eta_{rdc} >= 0.07$ for the α-helix and β-strands. Exceptionally high anisotropy was observed for the residues L8, Q62 and S65 and the residues of the C-terminus, all of them located in loop regions. Interestingly, the anisotropic motion of the α-helix detected by rdcs correlates quite well with the excursion of the α-helix observed by high pressure studies [Kitahara et al., 2005]. We compared model-free derived η_{rdc} with the high pressure NMR results by back-calculating η_{rdc} values from a two structure-ensemble of the 1V80 (30 bar) and 1V81 (3 kbar) structures (see Table 2). For most of the helix residues both values deviate by less than 20°. Outliers are the residues 23, 30 and 33. An explanation for the deviation might be that the high and low pressure structures have been determined without the use of RDCs. Thus, the non-refined low and high pressure structures might contain errors on the directions of the NH vectors. Moreover, 23 is very rigid and 30 and 33 show only a very low amplitude of anisotropy. Thus, the direction of anisotropy has only little impact for these residues.

4.4.5 Structural noise

For the three different structures 1ubi, 1ubq and 1d3z alignment tensors differ by less than 2% in size, less than 5% in rhombicity and less than 5° in the angles defining the Euler rotation of the alignment frame with respect to the molecular frame. The obtained S^2_{rdc} values depend only weakly on the structure that is used for determination of alignment tensors if the amount of structural noise is less than 5%. The correlation coefficients between the three sets of S^2_{rdc}-values are larger than $\rho = 0.99$. Structural noise has therefore only a small influence on the results of the calculation. Thus, the observed deviations of S^2_{rdc} from the S^2_{LS} order parameters cannot be accounted for by structural noise but indeed reveals additional motions present on a time-scale slower than the correlation time τ_c. The order parameters presented in Figure 5a) are the average values of S^2_{rdc} that were obtained for the three different structures (1 nmr and 2 X-ray) used for the determination of alignment tensors (see above). The error margins reflect the standard deviation. An even more robust treatment of the structural noise problem will be presented in chapter 6.

4.5 Conclusions

Based on the enlarged RDC data set of 27 NH rdc data sets, the RDC-based model-free approach was rigorously re-evaluated mathematically and the algorithm further improved, for example, with respect to filtering out of experimental noise (cf. chapter 4,[Lakomek et al., 2006]). The new results corroborated the initial finding of supra-τ_c motion by [Peti et al., 2002]. Using the improved RDC-based model-free approach, RDC-based order parameters $S^2_{rdc}(NH)$ were derived with unprecedented accuracy, revealing new modes of motion in the supra-τ_c time window. However, the amount of observed supra-τ_c motion with an average RDC-based order parameter $< S^2_{rdc}(NH) >= 0.72$ appeared less than in the initial study with $< S^2_{rdc}(NH) >= 0.61$ [Peti et al., 2002], compared to $< S^2_{LS}(NH) >= 0.78$ [Chang and Tjandra, 2005]. Charged and polar residues show more mobile backbone amide groups than hydrophobic residues. In particular the biologically relevant arginines reveal strongly decreased $S^2_{rdc}(NH)$ order parameters. Besides the $S^2_{rdc}(NH)$, effective orientations of the NH internuclear vectors described by the polar angles(θ_{av}, ϕ_{av}) and the amplitudes η and directions of anisotropic motions ϕ'_{rdc} were determined. For the α-helix, the determined ϕ'_{rdc} support an excursion of the α-helix as suggested earlier [Meiler et al., 2003] and recently observed by high pressure NMR [Kitahara et al., 2005].

The obtained results differ from those of Clore and Schwieters [Clore and Schwieters, 2004b]. In the following a possible explanation is given. For this, first we explain the approach by [Clore and Schwieters, 2004b] in more detail: Fitting the NH RDC data first with a single structure representation, Clore and Schwieters obtained a residue-specific dipolar coupling R-factor $R^{(NH)}_{dip}(work)$ between 5-10% that is defined as $R_{dip} = \frac{1}{\sqrt{2}}Q_{dip} = RMS(D^{obs}_j - D^{pred}_i)/RMS(D^{obs}_j)$ where RMS refers to the root mean square function and D^{obs}_j and D^{pred}_j are the observed and predicted RDCs for residue j [Bax and Grishaev, 2005]. In the following, the authors claim this $R^{(NH)}_{dip}(work)$ factor to lie within the range of accuracy of measured NH RDCs by comparing and correlating RDC data sets obtained in very similar alignment conditions. In our experience, however, it is very difficult to exactly reproduce the alignment conditions. In fact, the alignment tensor will slightly change, making the suggested procedure for assessing the accuracy of RDC data very questionable. With a main principal axis value of $D_{zz} \approx 20$ Hz on average, an R_{dip} between 5 and 10% would correspond to an experimental NH RDC error between 1 and 2 Hz which is between two and four times the value of the experimental error of ± 0.5 Hz estimated in [Peti et al., 2002]. With a larger error tolerance without energy penalty for the harmonic RDC-potential term in a molecular dynamics simulated annealing protocol, all NH RDC data can be fulfilled by a single structure. We had similar findings during the development

of the EROS approach which is described in chapter 7. However, in our opinion the obtained single structure has to be considered as a dynamic average structure. Note that the existence of a moderately fitting average structure does not contradict the presence of dynamics.

In order to take anisotropic motion into account, Clore and Schwieters have extended the single structure refinement to a two-site jump model approach that is basically an ensemble approach with ensemble size $N_e = 2$ [Clore and Schwieters, 2004b]. From a heuristic perspective it represents the simplest description of anisotropic motion. Each ensemble member was allowed to take its own value of the axial alignment tensor component D_a and the rhombicity R, allowing for alignment tensors consistent with ensemble members having different shapes. The authors introduced a pseudo-energy to restrain the spread of the axial component of the alignment tensor D_a and the rhombicity R. Furthermore, a molecular shape term and a relative atomic position term have been introduced by the authors: The molecular shape term derived from a massless inertia tensor prevented excessive rotation and deformation of one ensemble member relative to another. The relative atomic position (RAP) term restrains the C_α atomic positions in each ensemble such that they do not stray too far from their respective ensemble-averaged positions. According to the authors this has only minimal impact in N-H bond vector orientations between members of an ensemble. In my opinion, however, a bias towards a too tight ensemble introduced by these additional restraints cannot be excluded. Especially the RAP term applied on C_α atom coordinates appears questionable considering work by Zuiderweg an colleagues who detected increased dynamics of the $C(O)C_\alpha$ vectors in ubiquitin with cross-correlated relaxation experiments [Wang et al., 2003] and recent work by Voegli and Bax who detected increased $H_N H_\alpha$ mobility for protein G [Vogeli et al., 2008]. Allowing a higher error tolerance without energy penalty for the harmonic RDC-potential term in a molecular dynamics simulated annealing protocol leads to a tighter and less flexible ensemble. That appears counter-intuitive at first glance but becomes clear when keeping in mind, that restraints by the RDC potential can be more easily fulfilled when the allowed error tolerance is larger. Thus, the impact of RDCs becomes less and that of structural geometry restraints becomes higher, making the ensemble more tight. These effects have been observed by us during the development of the EROS ensemble approach, too (cf. chapter 7). Indeed, high accuracy measurements of RDCs, as presented in this work, are crucial to a correct assessment of protein dynamics.

The 2-site jump ensemble approach by Clore and Schwieters is a model-based approach, since it assumes the protein dynamics to be consistent with force-fields implemented in the Xplor-NIH structure determination package

[Schwieters et al., 2003]. In contrast, the rdc-based model-free approach is *model-free* in that sense that it does not make any assumptions about a dynamic model. Another approach, model-free in the same sense, is the Direct Interpretation of Dipolar Couplings (DIDC) approach by Tolman [Tolman, 2002], mentioned already in the introduction. Also the DIDC approach relies on the measurement of RDCs in five linear alignment conditions. An RDC matrix $\mathbf{D}$ is formed directly from the measured RDCs with residue number as rows and alignment media as columns. The matrix $\mathbf{D}$ is decomposed into two matrices $\mathbf{A}$ and $\mathbf{B}$ and a constant pre-factor K:

$$\mathbf{D} = K\mathbf{BA}. \tag{4.22}$$

The columns of $\mathbf{A}$ represent the alignment tensors for each of the alignment media and the rows of $\mathbf{B}$ contain the structural and dynamic information that should be determined, that is the mean orientation and accompanying description of dynamics for each inter-nuclear vector RDCs are available, e.g. the NH vector. If the alignment tensor matrix $\mathbf{A}$ is given, for example by using programs such as DipoCoup [Meiler et al., 2000] or Pales [Zweckstetter and Bax, 2000] and a structural model, the DIDC approach becomes, in fact, conceptually very similar to the RDC-based model-free approach. However, the DIDC approach is constructed to work without a structural model. Having no ab initio alignment tensor information, the system of linear equations, represented by equation 4.22 becomes underdetermined. To overcome this problem, Tolman derived after some matrix algebra a boundary condition which basically requires that the variation in generalized S^2 order parameters becomes minimized (cf. [Tolman, 2002], equation 15). In my opinion this procedure has the drawback that it makes assumptions about this distributions of order parameters. Furthermore the procedure is highly sensitive to experimental errors. Only one residue with incorrect RDC data leading to a highly dynamic (low) S^2 order parameter can result in a wrong smoothening of the remaining (discrete) order parameter curve and to an underestimation of the dynamics. A very recent extension of the DIDC approach by Bax and coworkers [Yao et al., 2008] that applies the DIDC in an iterative fashion suffers the same shortcomings. For protein G, a comparably smooth S^2_{rdc} order parameter is observed, similar to the Lipari-Szabo S^2_{LS} order parameter curve, with a few highly dynamic exception indicated by very low S^2 order parameters. According to Joel Tolman, the DIDC approach was originally designed to determine correct dynamically averaged inter-nuclear vector orientations with less consideration of the S^2 order parameter (personal communication with Joel Tolman). A sophisticated grid-search based approach towards the determination of dynamically averaged inter-nuclear vector orientations had been developed very recently by Joel Tolman and coworkers [Ruan et al., 2008].

While the improved RDC-based model-free analysis corroborated the existence of

supra-τ_c motion two major short comings remain: First, the RDC-based model-free approach depends on a highly accurate protein structure used for alignment tensor calculation. While the results were stable for less than $5°$ of structural noise, higher deviations were observed for larger amounts of structural noise (cf. chapter 6 , Fig. 10). Second, since the RDC-derived order parameters S^2_{rdc} contain only relative dynamic information and have to be scaled against the Lipari-Szabo S^2_{LS} order parameters, the amount of supra-τ_c motion observed critically depends on the $S_{overall}$ scaling factor. While first attempts for a correct assessment of $S_{overall}$ have been undertaken, a sophisticated statistically appropriate scaling method, taking the experimental error into account, is desired. These points have lead to the development of the Self-Consistent Rdc-based Model-free Approach (SCRM) that is presented in chapter 6.
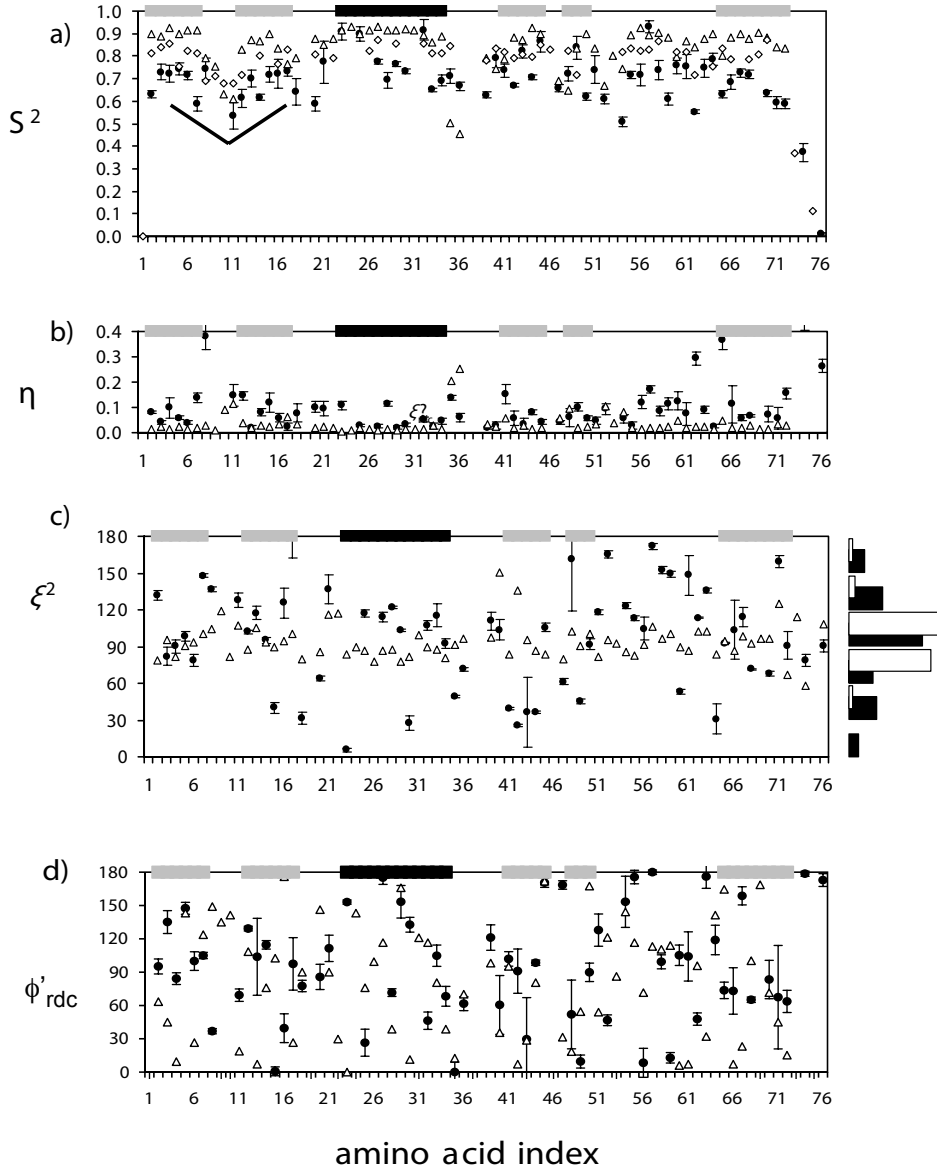
Figure 4.5: a) Shows a comparison of ubiquitin order parameters $S^2_{rdc}(NH)$ obtained from 27 sets of NH rdcs (filled circles) with the values obtained from relaxation measurements S^2_{LS} (diamonds) as well as values derived from a 10ns molecular dynamics trajectory S^2_{traj} [Meiler et al., 2001] (triangles). Helical regions in ubiquitin are marked with a black bar on top of the diagram, strand regions with a gray bar. b) Compares the asymmetry parameter derived from rdcs η_{rdc} (filled circles) with values derived from a 10ns molecular dynamics trajectory η_{traj} (triangles) [Meiler et al., 2001]. In (c) the ξ'_{rdc} angles as determined from experimental data (filled circles) are compared with the ones from the trajectory (triangles). The ξ'_{rdc} angle measures the direction of the asymmetry with respect to the peptide plane. In the right panel the distribution plot illustrates the strong preference for this angle to be 90° in the trajectory and how this preference is weakened when looking at longer time scales utilizing experimental rdcs. Diagram d) compares the ϕ'_{rdc} angles as determined from experimental data (circles) with the ones from the trajectory (triangles). The ϕ'_{rdc} describe the direction of the anisotropic motion in a model-free manner.

Chapter 5

Side chain orientation and hydrogen bonding imprint supra-τ_C-motion on the protein backbone of ubiquitin

To differentiate between the sub-τ_c and supra-τ_c time-scale, the RDC-based order parameters $S^2_{rdc}(NH)$ (cf. chapter 4, [Lakomek et al., 2006]) were compared to Lipari-Szabo order parameters $S^2_{LS}(NH)$ that are derived from conventional relaxation time measurements and only sensitive to the sub-τ_c time scale [Lipari and Szabo, 1982b, Lipari and Szabo, 1982a].

Interestingly, a periodic variation of the S^2_{rdc} can be observed with a periodicity of 2 residues in the β-strands of ubiquitin (aa 2-6, 12-16, 41-45, 66-71), while it is largely absent from the S^2_{LS} and exchange rate data. Most prominently, the S^2_{rdc} values are larger for residues Gln41, Leu43 and Phe45, and are smaller for residues Arg42 and Ile44 in β-strand 41-45 (cf. Figure 1a). Correspondingly, the side chains of residue Gln41, Leu43, Phe45 point towards the hydrophobic core (core residues), whereas the side chains of Arg42 and Ile44 are solvent-exposed (exposed residues) (for classification of solvent exposure, see Appendix). Thus, the alternating pattern of S^2_{rdc} values seems to correlate with the orientation of the side chains towards the solvent or away from it (cf. Figure 1b).

The observation that solvent-exposed residues exhibit reduced S^2_{rdc} values as compared to core residues holds not only for the β-strands, but also for the rest of the protein. In Figure 2, the amino acids of ubiquitin (1d3z) are colour-coded according to the S^2_{rdc} value of the backbone amide groups. Residues with less mobile NH vectors (blue and green) predominantly have side chains pointing to

the hydrophobic core (a), while for those with more mobile NH vectors (yellow, orange and red), the side chains are solvent-exposed (b).

In order to investigate this effect quantitatively, an order parameter for only the supra-τ_c time-scale is derived: $S^2_{rdc}/S^2_{LS} = \frac{4\pi}{5} \sum_{M=-2}^{2} \langle Y_{2M}(\theta, \phi) \rangle \langle Y^*_{2M}(\theta, \phi) \rangle \big|_{\tau_c}^{ms}$. For residues with supra-τ_c motions, we expect a value for S^2_{rdc}/S^2_{LS} smaller than one while it should be one in the absence of such motions (see also Appendix).

Core residues show an average S^2_{rdc}/S^2_{LS} value of 0.91 ± 0.02 as compared to 0.86 ± 0.02 for solvent-exposed residues (Table 1), while the exposure of the previous side chain does not have a significant influence (0.86 ± 0.02 vs. 0.87 ± 0.02). 90% of the core residues have a S^2_{rdc}/S^2_{LS} greater than 0.85. In contrast, this is true for only 55% of the solvent-exposed residues (cf. Figure 3a). Thus, the supra-τ_c order parameter S^2_{rdc}/S^2_{LS} of the NH vector is on average 0.05 ± 0.03 larger if the side chain is buried in the protein core.

We also investigated the influence of backbone hydrogen bonds on the supra-τ_c-order parameter S^2_{rdc}/S^2_{LS}. For this, each NH group was classified according to the number of backbone hydrogen bonds on the corresponding peptide plane includ-
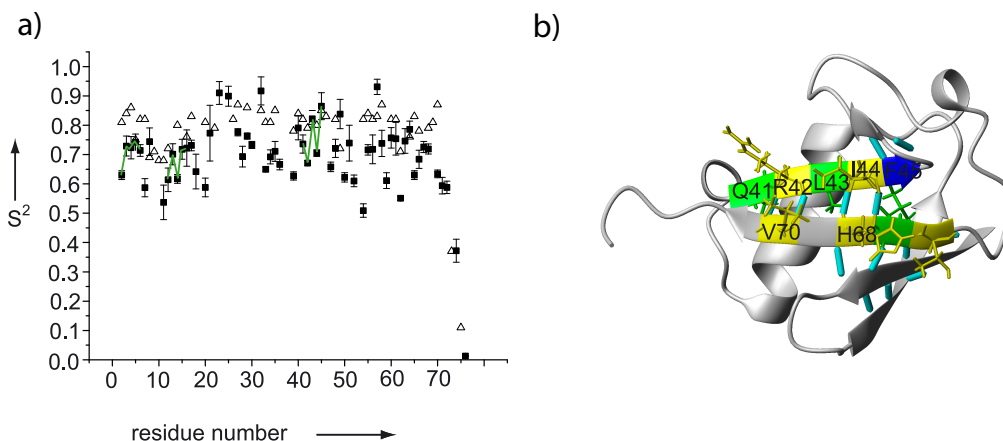
Figure 5.1: a) Comparison of the Lipari-Szabo order parameter S^2_{LS} (empty triangles) and the rdc- derived order parameter S^2_{rdc} (filled squares) as a function of residue number. The marked periodic variations of the S^2_{rdc} are indicated by lines connecting sequential residues.
b) The residues of β-strands 41-45 and 66-71 of ubiquitin (1d3z) are colour-coded according to the RDC-based order parameters S^2_{rdc} (yellow $0.63 \leq S^2_{rdc} < 0.73$, green $0.73 \leq S^2_{rdc} < 0.83$, blue $S^2_{rdc} \geq 0.83$, gray unknown). Hydrogen bridges are displayed as cyan rods. The alternating pattern of S^2_{rdc} order parameters in β-strand 41-45 is clearly correlated with the side-chain orientation.

56

	Exposure				Hydrogen Bridges		
	solvent(i)	core(i)	solvent(i-1)	core(i-1)	0	1	2
S^2_{LS}	0.80	0.82	0.80	0.81	0.74	0.81	0.83
	±0.01	±0.01	± 0.01	±0.01	±0.02	±0.01	±0.01
S^2_{rdc}	0.69	0.75	0.71	0.70	0.65	0.70	0.77
	±0.02	±0.02	±0.01	± 0.02	±0.03	±0.02	± 0.02
S^2_{rdc}/S^2_{LS}	0.86	0.91	0.87	0.86	0.82	0.86	0.90
	± 0.02	±0.02	±0.02	±0.02	±0.05	±0.02	±0.02

Table 5.1: Average order parameters for core and exposed residues and for residues involved in 0, 1, or 2 hydrogen bonds. The NH_i order parameter was correlated with the side-chain orientation of the same residue (i) and the orientation of the previous residue (i-1). Whereas for the first case, a strong dependence on the side-chain orientation is obtained, no correlation is observed for the second case, thus indicating that mobility is transferred only along the ϕ and not the ψ angle.
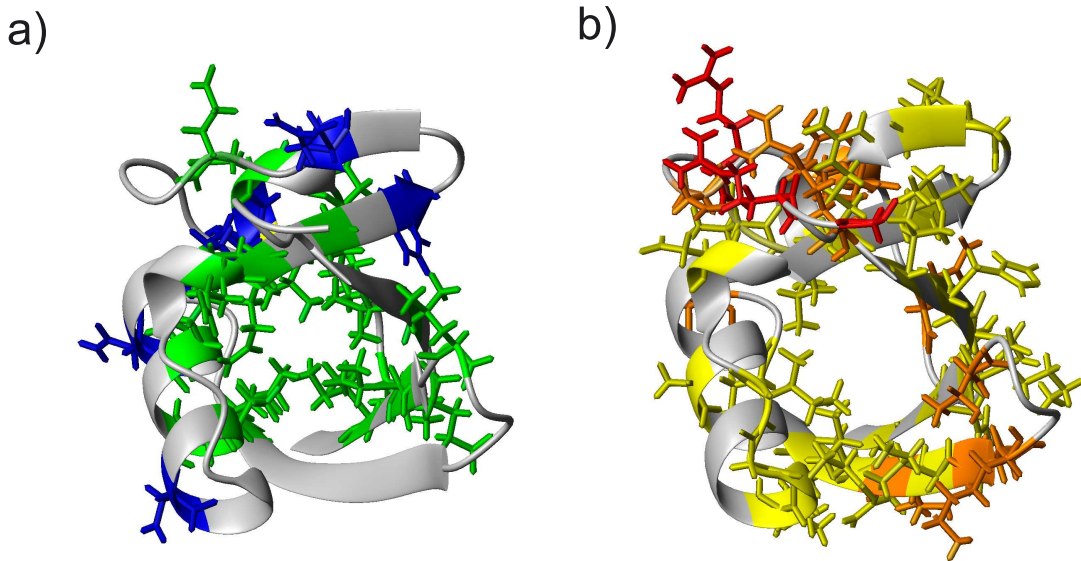
Figure 5.2: Residues of ubiquitin (1d3z) are colour-coded with respect to the RDC-based order parameters (red $S^2_{rdc} < 0.53$, orange $S^2_{rdc} < 0.63$, yellow $S^2_{rdc} < 0.73$, green $S^2_{rdc} < 0.83$, blue $S^2_{rdc} \geq 0.83$, gray unknown) and are distinguished between large a) and small b) values. Interestingly, most of the residues with larger backbone amide S^2_{rdc} values have side chains pointing to the core (a), whereas residues with solvent-exposed side chains show smaller order parameters (b).

ing the amino acids NH_i and the preceding carbonyl group CO_{i-1} (cf. Appendix, [Cordier and Grzesiek, 1999]).

The average S^2_{rdc}/S^2_{LS} values (Table 1) reveal a dependence on the number of hydrogen bonds as well, increasing from 0.82 ± 0.05 to 0.86 ± 0.02 to 0.90 ± 0.02 with the number of hydrogen bonds being zero, one or two. 83% of residues with two backbone hydrogen bonds have values greater than 0.85. This is true for 61% of the residues that are involved in only one hydrogen bond (Figure 3b).

In conclusion, both side-chain orientation and hydrogen bonds influence supra-τ_c motions of the backbone amide groups. The hydrogen bridge dependency correlates with the secondary structure of the protein, whereas solvent exposure is a property of the tertiary structure. It is interesting that the side-chain mobility is apparently transmitted only along the ϕ and not the ψ angle (cf. Appendix). The investigation of this effect is beyond the scope of this work. One possibility would be that the side chain motion is coupled with pyramidalization of the neighbouring NH in a dissipative manner. This NH is closer than the NH of the following amino acid. However, the reason could also be the different rotation barriers around the and the angle due to the constitutional difference of these two moieties. Very recent results of a RDC-based model-free analysis on methyl group RDCs for
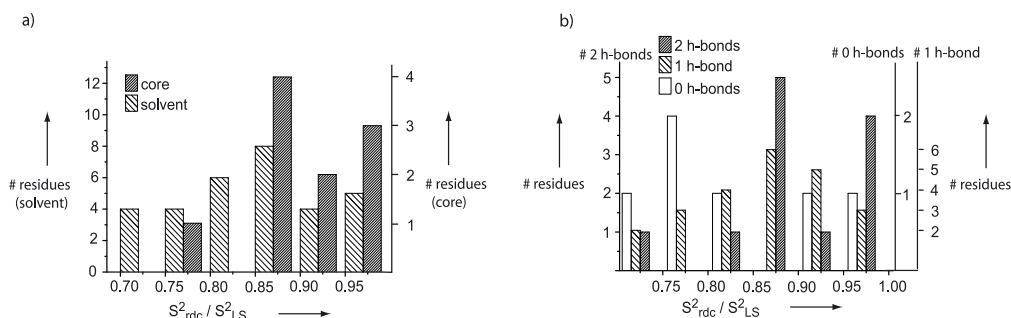
Figure 5.3: Histogram plot of the distribution of S^2_{rdc}/S^2_{LS} values. The residues are classified as core or solvent-exposed residues (a) or with respect to the number of backbone hydrogen bonds (b). The average S^2_{rdc}/S^2_{LS} order parameter is 0.91 ± 0.02 for core and 0.86 ± 0.02 for solvent-exposed residues. It is 0.90 ± 0.02 for residues involved in two backbone hydrogen bonds, 0.86 ± 0.02 for one hydrogen bond and 0.82 ± 0.05 for residues that are not involved in backbone hydrogen bonds. a) Core residues have strong tendency for higher S^2_{rdc}/S^2_{LS} order parameters: 9 out of 10 core residues have a S^2_{rdc}/S^2_{LS} value greater than 0.85. b) Residues involved in two hydrogen bonds have higher S^2_{rdc}/S^2_{LS} values than those without any.

ubiquitin, show a correlation of $\rho = 0.34$ between $S_{rdc}^2(NH)$ and $S_{rdc}^2(CH_3, axial)$ (data not shown). These first results for side-chain dynamics (without a residue-specific analysis) support the idea of a coupling between backbone and side-chain dynamics and corroborate the presented results.

Our findings have been supported by independent results for protein G by Martin Blackledge and coworkers [Bouvignies et al., 2005b]. Using a broad data set collection of different backbone RDCs, a three-dimensional Gaussian axial fluctuation model has been derived to analyze the nature and amplitude of backbone motion in the third IgG binding domain of protein G that consists of 61 residues [Derrick and Wigley, 1994]. The three-dimensional Gaussian axial fluctuation model, originally invented to analyze NMR relaxation data has been derived by comparison to a Langevin molecular dynamics simulation [Bremi and Brueschweiler, 1997, Lienin et al., 1998]. The motions of the peptide planes are assumed to be harmonic and are described by independent Gaussian axial fluctuations about three orthogonal axes α, β, γ. They can be visualized as ellipsoids with principal axes parallel to α, β, γ and lengths that are proportional to the fluctuation amplitudes $\sigma_\alpha, \sigma_\beta, \sigma_\gamma$. In the original Langevin dynamics simulation the largest fluctuations occur about the γ axis which is nearly collinear to the $C_{(i-1)}^\alpha$–$C_{(i)}^\alpha$ vector. The same parametrization of the motional model has now been transferred to RDCs. In the same way as for the RDC-based model-free approach, the dipolar coupling equation in the alignment frame is expressed by dynamically averaged second order spherical harmonics (cf. 4.3). However, these dynamically averaged second order spherical harmonics are not interpreted in a model-free way but are parameterized according to the Gaussian axial fluctuation model. Thus, for each residue the RDCs are fitted to the motional amplitudes $\sigma_\alpha, \sigma_\beta, \sigma_\gamma$ (cf. [Bouvignies et al., 2005b], equations 1-3). Motions slower than τ_c could be observed for loop regions, the β-sheet and the α-helix. The extracted σ_γ amplitudes of mobility and S^2 order parameters show an alternating pattern along the sequence in the β-strands. The alternating nodes along the sequence match the alternation of strongly hydrophobic side chains buried in the protein core, in concordance with our findings for ubiquitin. By doing a hydrogen-bond scalar coupling analysis Blackledge and coworkers could show the motion in the β-sheet to be correlated.

On the relaxation time scale, meaning motion faster than the correlation time τ_c, a very similar effect has been observed earlier by Palmer and coworkers for Ribonuclease H [Mandel et al., 1995, Mandel et al., 1996]. In the β-sheet region backbone dynamics are correlated with solvent-exposure as well. Thus, similar patterns of motions seem to be observed on different time-scales. This finding fits well into a hypothesis recently proposed by Karplus, Kern and coworkers who postulate a linkage of time-scales for enzyme catalysis. Based on relaxation

measurements and molecular dynamics simulation they observed that pico- to nano-second timescale atomic fluctuations in hinge regions of adenylate kinase facilitate the large-scale, slower lid motions that produce a catalytically competent state [Henzler-Wildman et al., 2007]. Here the slower motions occur on the μs to ms time-scale. Results derived with the RDC-based model-free approach could in principle provide the missing link in the previously hidden time window between τ_c and 50 μs.

Recently, dynamics of ubiquitin in the microcrystalline state have been investigated by solid state NMR [Lorieau and McDermott, 2006]. These measurements were made using the dipolar tensor. Since the proteins do not reorient isotropically like in solution, the molecular frame remains fixed in solid-state, and thus a direct measurement of the tensor is possible. Motions can be measured that are faster than the frequency of the solid state dipolar coupling Hamiltonian, which is typically the inverse of tens of microseconds. In practice order parameters are measured by comparing the experimental tensor breath to that expected for an analogous tensor probed for a rigid molecule. Order parameters were determined for $^{13}C^1H(\alpha)$, $^{13}C^1H_2(\alpha)$, $^{13}C^1H(\beta)$, $^{13}C^1H_2(\beta)$ and several leucine, lysine and proline side chains. For the backbone, a weak positive correlation between solid state oder parameters and our solution state $S^2_{rdc}(NH)$ order parameter can be observed. In both cases the C-terminal part of the α-helix appears relatively mobile, Glu-64 appears most rigid in the solid state NMR measurements and compared to the surrounding residues also for RDC-based model-free $S^2_{rdc}(NH)$ order parameters. Slow time-scale motions have also been detected in several cross-correlated relaxation experiments in solution state NMR [Pelupessy et al., 2003, Dittmer and Bodenhausen, 2004, Ferrage et al., 2006]. .

Chapter 6

The self-consistent RDC-based model-free analysis (SCRM)

6.1 The problem of structural noise - Motivation for the SCRM analysis

The RDC-based model-free approach relies on the measurement of NH RDCs for five linearly independent alignment tensor orientations in at least five different media [Meiler et al., 2000, Peti et al., 2002, Lakomek et al., 2005] [Lakomek et al., 2006]. Using a high-resolution structure to determine the alignment tensors, structural as well as dynamic information can be deduced. A possible problematic point is the dependence on the accuracy of the protein structure used for alignment tensor calculation. For the original RDC model-free approach a 1.8 Å high-resolution X-ray structure(1ubi, [Ramage et al., 1994]) has been used. Inaccuracies in this structure which are called *structural noise* will propagate to the model-free results. A systematic error in the alignment tensor-calculation introduced by structural noise has been pointed out by [Zweckstetter and Bax, 2002]. Consequences for the dynamic interpretation of RDCs have been analyzed by [Bouvignies et al., 2005a, Clore and Schwieters, 2004b] [Clore and Schwieters, 2004a, Clore and Schwieters, 2006] . Second, since the RDC-derived order parameters S^2_{rdc} contain only relative dynamic information and have to be scaled against the Lipari-Szabo S^2_{LS} order parameters, the observed amount of supra-τ_c motion critically depends on the $S_{overall}$ scaling factor. Thus, a sophisticated statistically appropriate scaling method, taking the experimental error into account, is desired. Moreover, Annila and coworkers tried to explain the variation in S^2_{rdc} by possible alignment tensor fluctuation challenging

the presence of supra-τ_c motion [Louhivuori et al., 2007]. Indeed, the model-free approach assumes that possible alignment tensor fluctuation are not correlated with alignment tensor fluctuations.

These concerns have motivated us to develop a Self-Consistent RDC-based Model-free (SCRM) analysis that addresses all these points. Different quality criteria have been applied and introduced to assess the accuracy of experimental data and SCRM results. The analysis has been applied to the protein ubiquitin. RDC data sets used in the first work by [Peti et al., 2002] have been remeasured on newly prepared samples with higher ubiquitin concentration in order to achieve better resolution and higher intensity of the peaks for highest experimental accuracy and most consistent RDC data.

6.2 Experimental part

6.2.1 Alignment media preparation

In addition to the previously measured data sets E1 to E5, 13 new NH rdc data sets (A1-A13) were measured and used for the new analysis. This has been done in order to replace previous data sets that had been prepared with a smaller concentration. Alltogether, 36 NH RDC data sets from the backbone of the wild-type human ubiquitin were available for the new SCRM analysis. Previous measurements [Peti et al., 2002], data sets D1-D5 in [Lakomek et al., 2006] were replaced by measurements with increased concentration of ubiquitin. ^{15}N, ^{13}C-labeled human ubiquitin (wt) was expressed according to [Johnson et al., 1999]. Thirteen new alignment conditions, A1 - A13, were prepared as described in the following. In every case, 2.5 mg of ubiquitin were dissolved in a 50mM Na phosphate buffer at pH 6.5. The final ubiquitin concentration varied between 0.75 and 0.9 mM, and 10 % - 20 % (v/v) D_2O were added for field locking. The following briefly describes the individual new alignment conditions:

- A1: A 7% positively charged gel sample was prepared according to [Cierpicki and Bushweller, 2004]. The positive charge was introduced by addition of (3-acrylamidopropyl)-trimethylammonium chloride (APTMAC, Sigma-Aldrich, Inc.) in a ratio of APTMAC:acrylamide=1:3.

- A2: A 7% positively charged gel was prepared as for A1 but with a ratio APTMAC:acrylamide=1:1.

- A3: A 5% negatively charged gel was prepared according to [Cierpicki and Bushweller, 2004]. The negative charge was introduced by addition of acrylic acid (Sigma-Aldrich, Inc.) in a ratio of acrylic acid: acryl amide = 1:1.

- A4: The ubiquitin solution was added to dodecyl-penta(ethylene glycol) (C12E5) stock solution (15% w/v) in a ratio of 2:1 and vortexed. The solution became opalescent after addition of 1.5% (v/v) hexanol [Ruckert and Otting, 2000].

- A5: Ubiquitin was dissolved in a suspension of 25 mg/ml Pf-1 phage (ASLA Ltd., Riga, Lativa) in 50 mM Na phosphate buffer with 100 mM NaCl [Zweckstetter and Bax, 2001].

- A6: Same as A5 but with a Pf-1 phage concentration of 20 mg/ml.

- A7: A 1,2-dimyristoyl-sn-glycero-3-phosphatidylcholine (DMPC) / 1,2-dihexanoyl-sn-glycero-3-phosphatidylcholine (DHPC) = 3:1 mixture (Avanti Polar Lipids, Alabama) of 15% w/v was dissolved in buffer containing 50mM NaCl and 50 mM Na phosphate buffer (pH = 6.5) with 0.02% sodium azide, and 10% D_2O. The final ubiquitin concentration was 0.9mM [Triba et al., 2005].

- A8: DMPC, DHPC and SDS (sodium dodecyl sulfate, Serva, Heidelberg, Germany) were mixed in a ratio of 30:10:2 and dissolved in a 50mM Na phosphate buffer with pH=6.5, containing 15-20 % D_2O until a total lipid concentration of 5%(w/v) was reached. The composition was vortexed at 0°C until the solution became clear. Ubiquitin was dispersed in this solution with a final concentration of 0.75 mM.

- A9-A13: Bicelle media were prepared similarly to A7 and A8. Ingredients and total lipid concentration can be found in Table 1. The ubiquitin concentration was 0.75 mM.

To complement the data sets obtained from these alignment conditions (in order to better span the 5-dimensional RDC space) the following data sets have been used:

Alignment condition	Liquid crystalline media	Mixing ratio	Total lipid concentration / (w/v)
A9	DLPC:DHPC:SDS	30:12:02	10%
A10	DMPC:DHPC: C14PC	30:10:01	5%
A11	DMPC: DHPC :C14PC	30:10:02	10%
A12	DMPC: CHAPSO: CTAB	50:10:01	10%
A13	DMPC: DHPC: CTAB	30:10:01	8%

Table 6.1: Bicelle media preparation. (See chapter Materials and Methods, alignment media preparation). The following abbreviation are used, additionally to those already explained before: C14PC= tetradecylphosphatidylcholine, CHAPSO = 3-(cholamidopropyl)dimethylammonio-2-hydroxyl-1-propane sulfonate (Sigma-Aldrich, Inc.), CTAB= N-cetyl-N,N,N-trimethylammonium bromide (Roth, Karlsruhe, Germany).

- A14-A18: NH data set E1 to E5 measured for the previous analysis [Lakomek et al., 2006].

- A19-A36: NH RDC data sets published by Bax and coworkers [Ottiger and Bax, 1998] and Tolman and coworkers [Briggman and Tolman, 2003, Ruan and Tolman, 2005]. The collection of 36 NH RDC data sets will be referred to as D36M. A table containing all newly measured NH RDC data sets (A1-A18) used for the analysis is provided in the Appendix. All relevant alignment conditions used for the different data sets are listed there as well.

6.2.2 NMR spectroscopy

NH RDC data for all new aligned media A1-A13 as well as the isotropic reference experiment were measured using 2D-^{15}N,^{1}H-IPAP HSQC experiments [Ottiger et al., 1998]. All data were recorded at a sample temperature of 308 K. Measurements were performed on either a Bruker-Avance 700 MHz (Bruker AG, Karlsruhe, Germany), a Bruker-DRX 600 MHz or a Bruker-Avance 600 MHz spectrometer equipped with a TXI cryogenic probe head or a Bruker 800 MHz spectrometer equipped with a TCI cryogenic probe head. The time domain was either TD1xTD2 = 1k x 2k or TD1xTD2=2k x 2k with a spectral width of = 25ppm x 15ppm (besides = 30ppm x 15ppm for data set A5 and A6). The number of scans was 32 and higher. For processing, data were zero-filled to TD1xTD2 = 32kx4k and processed with the NMRPipe software package

[Delaglio et al., 1995]. (J+D)- and J-coupling constants were extracted using NMRPipe. One-bond ^{15}N, ^{1}H-RDCs were derived from the difference in coupling between aligned samples and the isotropic sample. The experimental error was conservatively estimated to be less than 0.3 Hz. Alignment tensors were calculated using the software PALES [Zweckstetter and Bax, 2000].

6.2.3 SECONDA analysis

In order to quantify the similarity of structure and dynamics in the different alignment media (homogeneity of RDC data), a SECONDA analysis was applied [Hus and Bruschweiler, 2002, Hus et al., 2003] . The SECONDA method analyzes the covariance matrix constructed of all RDC data obtained under different alignment conditions. It performs a principal component analysis (PCA) of the RDC covariance matrix, which is equivalent to a singular value decomposition (SVD) of the RDC matrix. The singular values are sorted according to decreasing size. Structural and dynamic information is contained in the first five singular values, since dipolar couplings are a second rank symmetric tensor interaction and hence reside in a linear 5-dimensional space. Accordingly, only noise, systematic errors, and structural and dynamic heterogeneity may cause the 6th and higher singular values to differ from zero. The ratio of the 5th and 6th singular values (called SECONDA gap in the following) is a measure of the homogeneity of RDC data and the magnitude of noise. The larger the SECONDA gap, the more self-consistent are the RDC data in the different alignment media.

Since there is no straightforward way to treat missing entries in the RDC covariance matrix, a specific residue can only be considered if experimental data for this residue are available in all selected alignment media (compare [Hus et al., 2003], Scenario I). All experimental RDCs are normalized by division through the largest absolute RDC value in each medium. For D36M, the SECONDA gap was 5.66 with experimental data available for 21 residues in all media. Starting from these 36 NH RDC data sets, a subset was selected that maximizes the SECONDA gap by reducing the number of data sets and number of amino acids, with the restraint of having a 5th EV of the F-matrix with a value larger than 1 (see below). In a Monte Carlo search that randomly removed amino acids and alignment media in order to maximize the SECONDA gap, a maximum SECONDA gap of 15.18 could be obtained with the alignment media A 2,4,9-10,13-16,19-20,22-29,32-36 out of D36M and the amino acids 2, 5, 7, 11, 14, 25, 27, 32, 33, 36, 40, 44, 49, 51, 55, 58, 63, 66 and 68. This subset of 23 data sets will be called D23M in the following. Considering all residues for which experimental data were available (36 residues) the SECONDA gap was 6.81 after normalization. For comparison, selecting out of the 36 residues the 21 residues that were used for D36M gives a SCONDA

gap of 8.69. These rather large SECONDA gaps indicate a highly homogenous behaviour of structure and dynamics across all alignment media contributing to D36M and even more so for D23M.

6.3 Self-consistent RDC-based model-free (SCRM) approach

6.3.1 Method

The Self-Consistent RDC-based Model-free (SCRM) method was developed from the theory of the original RDC-based model-free approach. The goal of the Self-Consistent RDC-based Model-free (SCRM) approach was to remove a possible bias due to the protein structure used for the initial determination of the alignment tensors. To this aim, we proceeded as follows: The first step of the SCRM method was the application of the standard RDC-based model-free analysis as described in (cf. chapter 4, [Lakomek et al., 2006]): First, alignment tensors were calculated from the measured RDCs using the PALES
[Zweckstetter and Bax, 2000] software and from the X-ray structure 1ubi
[Ramage et al., 1994] to which protons were added with MOLMOL
[Koradi et al., 1996]. Naturally, the dynamically averaged orientations (θ_{av}, ϕ_{av}) calculated from the RDC-based model-free approach will exhibit deviations from the orientations in the X-ray structure. In an iterative fashion, the NH vector orientations used for the tensor calculation were replaced by the resulting (θ_{av}, ϕ_{av}) after each model-free analysis cycle (Figure 1). An N-H bond length of 1.04 [Ottiger and Bax, 1998] Å has been used to calculate the new proton coordinates. (Note that the SCRM results are independent of the used bond-length.) Each cycle i of the model-free analysis an alignment tensor recalculation was conducted until convergence of the order parameters $|S^2_{rdc,i} - S^2_{rdc,i-1}| = \frac{1}{n}\sum_{j=1}^{n} |S^2_{rdc,i}(NH_j) - S^2_{rdc,i-1}(NH_j)| \leq 0.01$ was achieved. Note that in every step of the SCRM approach the five time-averaged second order spherical harmonics are calculated in the same way as in the original rdc-based model-free approach for each residue. From those the five rdc-based model-free parameters $S^2_{rdc}, \phi_{av}, \theta_{av}, \eta, \phi'_{rdc}$ are derived. In the following discussion we will focus on the first three parameters S^2_{rdc} and (ϕ_{av}, θ_{av}) .

The SCRM procedure was implemented in a semi-automated manner using a Mathematica 5.2 protocol, PALES, and several Python scripts. To assess the fit of the alignment tensor to the experimental RDC data, static Q-factors $Q_{static} = \sqrt{2}R_{dip}$ as defined in [Bax and Grishaev, 2005] and the Pearson correlation coef-

ficients ρ were extracted from PALES after each cycle. Additional criteria and error measures were implemented for the SCRM analysis as explained below.
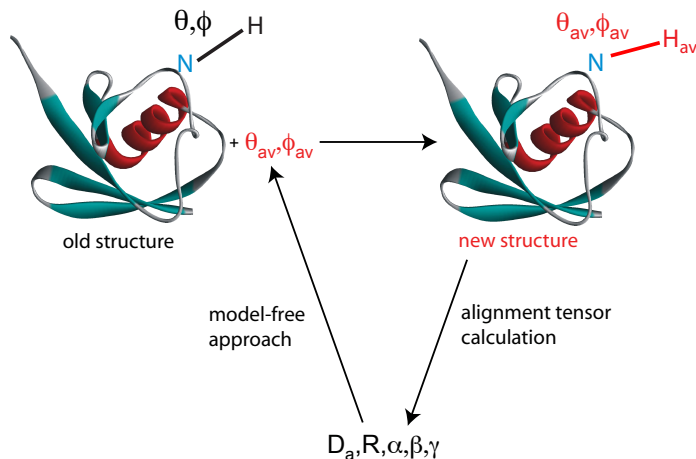
Figure 6.1: Sketch of the SCRM cycles: RDC-based model-free results are calculated using NH RDC data as well as inter-nuclear NH vectors for tensor calculation. Starting from a first structural model (e.g. the X-ray structure 1ubi with protons added with standard geometry), the inter-nuclear NH vector orientations are adjusted in each SCRM cycle towards the true dynamic average orientation. Consequently, the fit of the alignment tensor to the RDC data is improved in each step and as result the fit of the model-free results to the RDCs as well.

6.3.2 Back-calculated RDCs and dynamic Q-values assess the fitting quality of SCRM

RDCs were back-calculated from the time-averaged second order spherical harmonics $< Y_{2M}(\theta, \phi) >$ using the F-Matrix equation. Consistency of the model-free approach on a per-residue-basis (running index j) was assessed by computing the root mean square deviation (rmsd) between back-calculated RDCs D_{MF} and experimental RDCs D_{exp} (RDC-rmsd):

$$rmsd(rdc, j) = \sqrt{\frac{1}{K} \sum_{i=1}^{K} (D_{ij,MF} - D_{ij,exp})^2} \tag{6.1}$$

where K is the number of alignment media. We also introduced *dynamic* Q-values Q_{dyn} utilizing the D_{MF} RDCs:

$$Q_{dyn,i} = \sqrt{\frac{\sum_{j=1}^{N}(D_{ij,MF} - D_{ij,exp})^2}{\sum_{j=1}^{N} D_{ij,exp}^2}} \tag{6.2}$$

where N is the number of RDCs available in medium i. The dynamic Q-value Q_{dyn} measures the quality of fit of the model-free solution for the different alignment media. It is a straightforward extension of the well-known Q-value that measures the quality of fit of experimental RDCs over a single static structure.

6.3.3 Error calculation for the SCRM analysis

To estimate the error of the RDC-based order parameters $S_{rdc}^2(j)$ for each residue j, the experimental error was modelled by adding Gaussian noise to the measured RDCs. The input RDCs with added noise $D_{ij,noise}$ were generated by drawing N=1000 random samples $D_{ij,noise} = random[p(D_{ij})]$ from a Gaussian distribution $p(D)$ with standard deviation σ_j. Two different standard deviations σ_j were considered to study the propagation of different sources of errors to the SCRM derived order parameters: first, in order to assess the impact of the experimental error from the RDC data alone $\sigma_j^{exp} = 0.3$ Hz was used. Second, the residue-specific $\sigma_j^{rmsd} = rmsd(rdc, j)$ was used in order to assess the combined effect of experimental error and additional systematic errors introduced by the model-free analysis. Examples for the latter are a possible correlation between alignment tensor fluctuations and internal dynamics, or the single tensor approximation. This analysis was repeated N=1000 times. The error is evaluated as the standard deviation of the resulting N=1000 $S_{rdc}^2(j)$:

$$\Delta S_{rdc}^2(j) = \sqrt{\frac{1}{N} \sum_{k=1}^{N} (S_{rdc,k}^2(j) - <S_{rdc}^2(j)>)^2} \tag{6.3}$$

6.3.4 Selection of the set of RDCs for alignment tensor calculation

The RDC-based model-free analysis assumes that internal protein dynamics of the backbone NH vectors are not correlated with the alignment tensor modulations. This assumption allows working with a single average alignment tensor

for each medium. Simulations indicated that this assumption is correct for secondary structure elements, at least for steric alignment [Louhivuori et al., 2006, Salvatella et al., 2008]. Correlations between alignment tensor fluctuations and backbone NH vector dynamics have been observed only for more mobile loop regions in ubiquitin [Salvatella et al., 2008]. Consequently, we excluded the most mobile residues from the alignment tensor calculation. However, reducing the number of residues from which the alignment tensor is determined may lead to an inhomogeneous sampling of the three principal axes and may amplify structural noise. Thus, a consensus set of RDCs had to be found, which provides a nearly complete sampling of orientations while still avoiding correlations between tensor modulations and internal protein dynamics. To this end, we followed an approach similar to the one introduced by [Bouvignies et al., 2005b]: In a first step, alignment tensors are calculated using the experimental data for all residues 2-72. (The highly flexible C-terminus of ubiquitin (residues 73-76) was always excluded.) Four iterations of the SCRM protocol were performed. Then, the 20% most mobile residues ($S^2_{rdc,unscaled} \leq 0.95$) were excluded from the alignment tensor calculation and the SCRM analysis was re-started using the remaining set of residues for alignment tensor calculation. To ensure an adequate sampling of the three principal axes of the alignment tensor, an eigenvalue analysis of the matrix

$$C = \mathbf{B^T B} \tag{6.4}$$

where $\mathbf{B} = (\mathbf{e_1}, \mathbf{e_2}, \ldots, \mathbf{e_N})$ is the 3 x N matrix containing the normalized NH vectors of the average structure, is performed and the diagonalized matrix $\mathbf{D} = (\mathbf{d(1)}, \mathbf{d(2)}, \mathbf{d(3)})$ is obtained that contains the three eigenvalues of $\mathbf{C}$ sorted according to magnitude (see [Prompers and Bruschweiler, 2001] [Prompers and Bruschweiler, 2002]). The ratio r of the first and third eigenvalue of $\mathbf{D}$ as well as the asymmetry a defined below are a measure for the goodness of directional sampling:

$$r = \sqrt{\frac{d(1)}{d(3)}} \tag{6.5}$$

$$a = \frac{d(2) - d(3)}{d(1)} \tag{6.6}$$

In the ideal case of complete and homogeneous sampling, all eigenvalues of $\mathbf{D}$ are equal to 1, and the ratio $r = 1$ and the asymmetry $a = 0$. The reduced NH vector set (number of residues N= 49) after four runs of SCRM gives a ratio $r = 2.17$ and an asymmetry of $a = 0.32$ for the procedure applied on D36M and $r = 2.01$ and $a = 0.28$ for D23M (N=48). For comparison, the best possible benchmarks are $r = 1.61$ and $a = 0.22$ when all possible NH vectors (N=68) are used for the analysis. Using NH vectors of secondary structure elements only (N=40) gives

$r = 2.55$ and $a = 0.22$ which is slightly better with respect to asymmetry but significantly worse for r compared to the dynamic selections.

With r and a, we measure the completeness and homogeneity of the sampling of the orientations. For a successful RDC-based model free analysis the alignment tensors also need to homogeneously and completely sample the five dimensional space. A measure for completeness is the 5th eigenvalue and the condition number of the F-Matrix (see above). This is an empirical value without physical meaning, but compared to other experimental data sets it gives a first impression of the completeness of RDC-data. The 5th EV of the F-matrix is 1.02 for D23M and 1.31 for D36M, the condition numbers are 6.57 for D23M and 6.28 for D36M. This compares favourably with earlier analyses in which the condition numbers were also between 6 and 7 [Peti et al., 2002, Lakomek et al., 2005, Lakomek et al., 2006].

6.3.5   Determination of $S_{overall}$

Since RDC-based order parameters are sensitive up to the millisecond time scale while the Lipari-Szabo ones only up to the overall tumbling correlation time τ_c, the condition $S^2_{rdc} \leq S^2_{LS}$ or $S^2_{overall} \cdot S^2_{rdc,unscaled} \leq S^2_{LS}$ must hold within experimental error as explained already for the orignial RDC-based model-free approach in chapter 4. This relationship is used to estimate the overall scaling factor $S^2_{overall}$ by requiring $S^2_{overall} \leq S^2_{LS}/S^2_{rdc,unscaled}$ within the experimental error of S^2_{rdc} and S^2_{LS}. It is further assumed that several residues do not show supra-τ_c motion resulting in identical S^2_{rdc} and S^2_{LS} for those residues. Lipari-Szabo order parameters S^2_{LS} measured at 308 K by Tjandra and co-workers were used [Chang and Tjandra, 2005]. The scaling of the unscaled RDC-based S^2_{rdc} order parameters based on the Lipari-Szabo order parameter S^2_{LS} becomes non-trivial when taking the experimental error σ^2 into account.

Discussions with colleagues and cooperation partners encouraged us to develop a more sophisticated and robust statistical scaling method that will be explained in the following. Our main goal is to derive a conservative estimate for the overall scaling factor $S^2_{overall}$ such that a confidence level $1 - \alpha$ can be provided for the estimated $S^2_{overall}$.

Let us assume we knew the true RDC-based order parameters $S^2_{rdc,unscaled,true,i}$ for the different residues i. We will call them f_i. They have to be scaled with respect to the known $S^2_{LS,i}$ that will be called g_i. The constant overall scaling factor $S^2_{overall}$ will be called λ in the following. First, we define a true underlying scaling factor λ^* that is the maximal factor λ, such that $\lambda f_i \leq g_i$ is fulfilled for

all residues i:

$$\lambda^* = Max_\lambda\left\{\lambda f_i \leq g_i\right\} = Min_{i=1,...,n}\left\{\frac{g_i}{f_i}\right\} = \frac{1}{Max_{i=1,...,n}\left\{\frac{f_i}{g_i}\right\}} = \frac{1}{Max_{i=1,...,n}\left\{f_i^*\right\}}$$

(6.7)

with $f_i^* = \frac{f_i}{g_i}$ and n as the number of residues for which both f_i and g_i are available and non-zero.

As explained in detail in the paragraph *Error calculation for the SCRM analysis*, the effect of the experimental error on the RDCs leads to an error σ_i^2 on the derived order parameters $S^2_{rdc,unscaled,i}$ that will be called Y_i in the following. Because the true overall scaling factor λ^* is unknown, it will be estimated by

$$\hat{\lambda} = \frac{1}{Max_{i=1,...,n}\{Y_i^*\}}$$

(6.8)

with $Y_i^* = \frac{Y_i}{g_i}$.

Now we want to give an upper bound for the true λ^* with a certain classified probability. For that, confidence-levels $1 - \alpha$ are used. We ask: What is the probability that our upper bound $\hat{\lambda}_\alpha^u$ exceeds the true unknown scaling factor λ^* : $P(\lambda^* < \hat{\lambda}_\alpha^u) = 1 - \alpha$? Now we show how to determine $\hat{\lambda}_\alpha^u$ such that the previous equation is fulfilled for a certain confidence level, like e.g. $P(\lambda^* < \hat{\lambda}_\alpha^u) = 1 - \alpha = 0.95$. Unfortunately, this turned out not be distribution-independent. Therefore, we considered the distribution function

$$F_{max}(b) = P\left(Max_i Y_i^* - Max_i f^* \leq b\right)$$

(6.9)

that describes the probability, that the difference between the maximum value of the measurement $Y^* = \frac{Y}{g}$ and the maximum value of the true curve f^* is smaller than b. Indeed, $F_{max}(b)$ is dependent on the true curve f^* which means, that the problem is not distribution-independent, as mentioned before. In order to overcome this difficulty, let us assume now that the true curve f^* is close to the curve of measured data Y^*. We assume that the measured order parameter can be found within a distribution of order parameters with variance σ_i^2 around the unknown true order parameter $Y_i = f_i + \epsilon(\sigma_i)$ (where the error $\epsilon(\sigma_i)$ is one element of this distribution).

The propagation of this experimental error is estimated by adding Gaussian noise on the RDCs and repeating the SCRM analysis N= 1000 times. That yields $j = 1 \ldots 1000$ Y_{ij}^* order parameter values for each residue i forming a distribution with variance σ_i^2. Accordingly, we have generated new data with expectation of our experimental vector $Y^* = EY_j$. The distribution function is constructed out

of the N different Y_j^* curves. We calculated N times $Max_i Y_{ij}^* - Max_i f_{ij}$ for the different Y_j^* and by this generated the distribution function $F_{max}(b)$ (compare Fig. 2).

Starting from $1 - \alpha = F_{max}(b) = P\left(Max_i Y_i^* - Max_i f^* \leq b\right)$ we can invert for the upper bound $\hat{\lambda}_\alpha^u$. Using the relationship $\lambda^* = \frac{1}{Max_{i=1,\ldots,n}\left\{f_i^*\right\}}$ gives

$$1 - \alpha = F_{max}(b) = P\left(\frac{1}{\hat{\lambda}} - \frac{1}{\lambda^*} \leq b\right) \tag{6.10}$$

which can be transformed into

$$1 - \alpha = P\left(\lambda^* \leq \frac{1}{1/\hat{\lambda} - b}\right) = P\left(\lambda^* \leq \hat{\lambda}_\alpha^u\right) \qquad . \tag{6.11}$$

Now we obtain the required confidence bound $\hat{\lambda}_\alpha^u$ for the estimated scaling factor

$$\hat{\lambda}_\alpha^u = \frac{1}{1/\hat{\lambda} - b} \qquad . \tag{6.12}$$

This means: The probability that the true (unknown) scaling factor λ^* is smaller than the upper bound $\hat{\lambda}_\alpha^u$ is $P(\lambda^* \leq \hat{\lambda}_\alpha^u) = 1 - \alpha$. Back to our experimental data we calculate the distribution function $F_{max}(b)$ (compare Fig. 2). For the experimental error $\sigma = 0.3$ Hz, we calculate $\hat{\lambda}_\alpha^u = 0.80$ which is equivalent to $P(S_{overall} \leq 0.89) = 0.95$. That means that the probability for the true overall scaling factor to be lower than 0.89 is 95% - which is a conservative estimate.

Since our derivation has been based on the assumption that the true curve is close to the curve of measured data points, we tested the stability of our derivation by exchanging the true curve f^* by one of the simulated experimental curves Y_j^* and leaving b unaltered. This has been done M= 200 times. A slight tendency for an increase in confidence level has been observed with an average confidence level of $< 1 - \alpha > = 0.97$.

This scaling method has been applied to scale RDC-based S_{rdc}^2 order parameters derived from D36M with respect to the Lipari-Szabo order parameters S_{LS}^2. The confidence level was estimated taking the error propagation of the experimental error on the RDCs of $\sigma_j^{exp} = 0.3$ Hz into account. If we consider the experimental plus model error estimated from the RDC-rmsd $\sigma_j^{rmsd} = rmsd(rdc, j)$, the confidence level is $P(S_{overall} \leq 0.89) = 0.67$. The derived overall scaling factor is transferred to D23M as well.
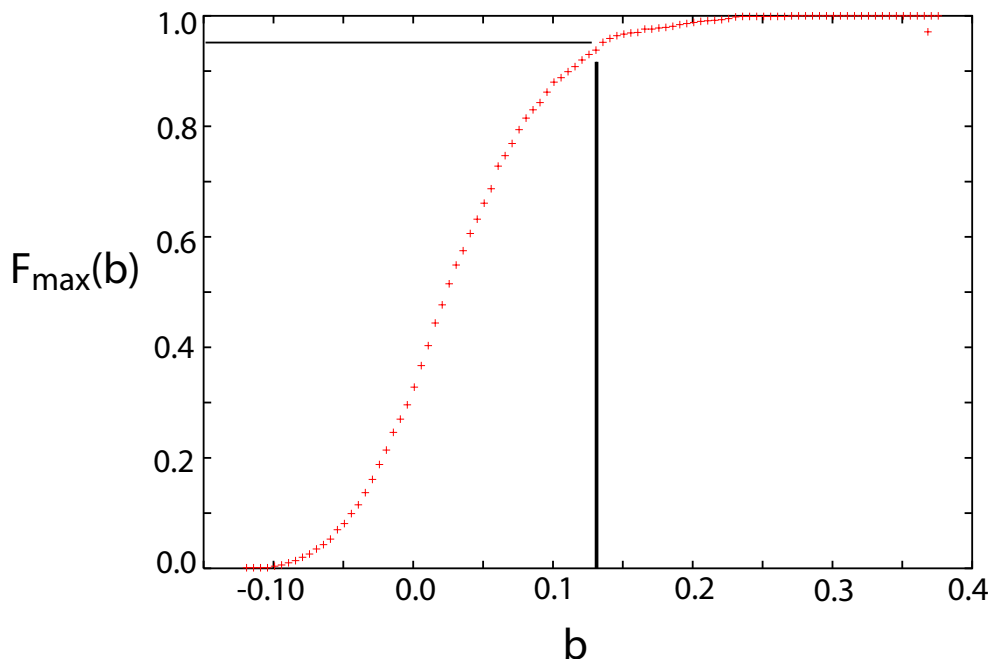
Figure 6.2: Distribution function $F_{max}(b)$

6.3.6 Application of SCRM approach to experimental data

The SCRM method has been applied to both experimental NH RDC data sets D36M and D23M using the X-ray structure 1ubi (Ramage et al., 1994) as starting structure (with protons added in standard positions with MOLMOL (Koradi et al., 1996) using a bond length of 1.04 Å). The influence of structural noise on the SCRM analysis was tested as described in the following paragraph.

6.3.7 Structural noise analysis

The influence of structural noise on the SCRM approach was tested for two different scenarios, A and B: In scenario A, synthetic Gaussian noise was added on the NH vector orientation of the X-ray 1ubi structure (with hydrogen atoms added according to standard geometry). Using PALES [Zweckstetter and Bax, 2000], the NH vector is tilted Gaussian distributed with opening angle θ and an equally distributed polar angle ϕ as described in (Zweckstetter and Bax, 2002). For the stan-

dard deviation of the Gaussian distribution values of a) $\sigma = 10°$, b)$\sigma = 20°$ or c) $\sigma = 30°$ were chosen, subsequently referred to as structural noise of $10°$, $20°$ or $30°$ respectively. For each case a)-c) three different random noise structures have been generated. These random noise structures were used as starting structures for the alignment tensor calculation in the SCRM analysis.

For scenario B nine crystal structures of ubiquitin bound to its recognition proteins were used as input. These structures deviate from the free ubiquitin crystal structure 1ubi by backbone RMSD values between 0.3 and 0.6 Å. These structures are 1cmx [Johnston et al., 1999], 1uzx [Teo et al., 2004], 1xd3 [Misaghi et al., 2005], 1yiw [Bang et al., 2005] , 2c7n [Penengo et al., 2006], 2d3g
(two structures, [Hirano et al., 2006]), 2fif (two structures, [Lee et al., 2006]). For both scenarios, the RDC-based order parameters S^2_{rdc} are compared to those derived from the noise-free 1ubi X-ray structure to analyze the influence of structural noise.

6.3.8   Statistical analysis of S^2_{rdc} and S^2_{LS} distributions

We describe the spread of the S^2_{rdc} and S^2_{LS} distributions over all residues of ubiquitin in terms of P-percentiles. The 25th percentile P25 is the value compared to which 25% of the distribution is lower. For the 75th percentile P75, 75% of the distribution have lower values. The interquantile range (IQR) is defined as the difference between P75 and P25. The IQR covers 50% of the distribution and is a direct measure for the spread of a distribution.

6.4   Results and Discussion

6.4.1   SCRM on experimental NH RDC data (D23M and D36M)

The SCRM method was applied to both NH RDC experimental data sets D23M and D36M using the X-ray structure 1ubi as starting input structure for the first cycle of the SCRM method.

The static X-ray structure 1ubi (with hydrogen atoms added according to standard geometry) yields static Q-values of $< Q_{static} >= 0.178$ for D23M and $< Q_{static} >= 0.193$ for D36M averaged over all alignment conditions. The Pearson correlation coefficients between experimental RDCs and those back-calculated from the static X-ray structure are $< \rho >= 0.977$ for D23M and $< \rho >= 0.972$

for D36M respectively, on average over all conditions. As described in the *Material and Methods* section, the SCRM method was designed to iteratively improve the accuracy of the alignment tensor determination and to adjust the average inter-nuclear vector orientations, and as a result, to further reduce the static Q-values $< Q_{static} >$ and increase the Pearson correlation coefficient ρ. Indeed, after already 4 SCRM-cycles, the static Q-values decreased to less than half of the original value with $< Q_{static} >$=0.062 for both D23M and D36M (Fig. 3a, b). Simultaneously, the correlation coefficients ρ increased to $< \rho >$=0.997 on average (Fig.3 c ,d). Convergence was attained in already 4 cycles of SCRM after which the inter-nuclear vector orientations were found to deviated by less than 0.5° between consecutive SCRM cycles. Thus the iterative procedure rapidly improves the fit of the static structure to the RDCs as compared to the input X-ray structure.

In Fig. 4(a,b) the inter-nuclear angles κ_j enclosed between the dynamic average NH vector orientations and the NH vectors of the starting X-ray structure 1ubi are shown. For better comparison, the dynamic average NH vector orientations have been rotated to a best-fit superposition with the NH vectors of the 1ubi structure (with protons added according to standard geometry). Most of the dynamic average inter-nuclear vector orientations obtained after 4 SCRM cycles differ from those of the 1ubi X-ray structure by less than 10° for κ_j (Fig. 4 a,b). The average angular deviation is 6.97° for D23M and 6.87° for D36M. Deviations larger than 10° are observed for Lys6, Lys11, Ile13, Ser20, Lys33, Glu34, Arg42, Lys48, Leu50, Asp52, Arg54, Leu67, Val70 and Leu71 for D23M. The largest deviation is 20.3° for Arg54 (compare Fig. 4 a,b). Most of these residues are highly mobile with $S_{rdc}^2(NH) < 0.7$. Exceptions are Lys6 in the first β-strand, Lys33 and Glu34 in the α-helix, Leu43 in the third β-strand and Leu67 and Val70 in the fifth β-strand. Despite the fact that $S_{rdc}^2(NH) > 0.7$ for Lys33 and Glu34, they appear relatively mobile compared to the surrounding residues in the α-helix. Values for all residues have been listed in table S3 for D23M and S4 for D36M. The derived dynamic average NH orientations have been compared to the NMR 1d3z structure [Cornilescu et al., 1998] (first structure of the ensemble) as well (Fig. 4c,d). The average deviation to the NMR structure is 4.84° for D23M and 4.52° for D36M. Thus, the agreement between the derived dynamic average NH vector orientations and the NMR structure is significantly better than for the 1ubi structure. Interestingly, most of those dynamic average NH vectors that showed the largest deviations to the 1ubi X-ray structure, for example K6, L67 and V70, did not show large deviations compared to the NMR structure. Only Lys11 and Asp52 show large discrepancies both for the X-ray and NMR structure. Both are highly dynamic. In the 1ubi structure Lys6, Lys48 and Arg54 appear to be affected by crystal packing. Indeed, the largest deviation between 1ubi and 1d3z is
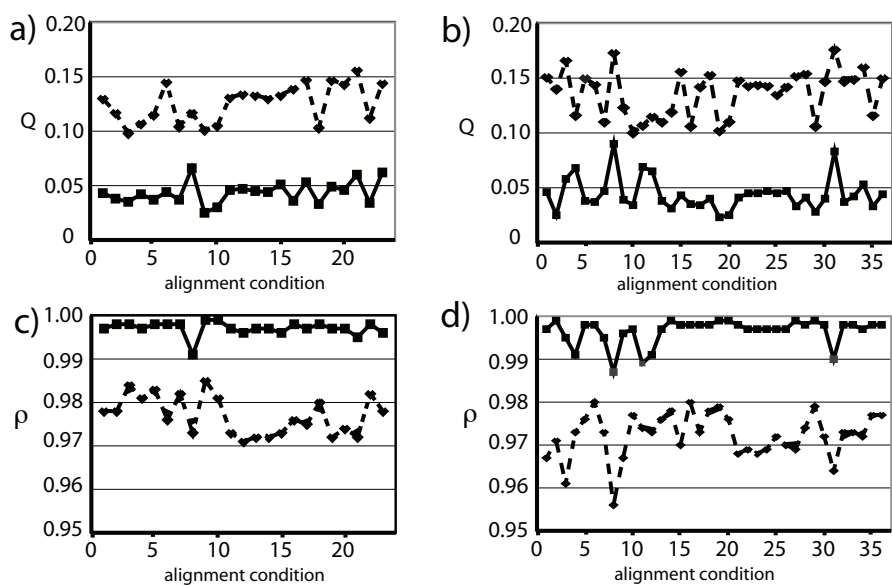
Figure 6.3: a) and b) show Q-values for back-calculated RDCs using the X-ray structure 1ubi for alignment tensor determination (dashed line) and after 4 SCRM cycles using the fitted dynamic average NH vector orientations (black line) both for a) D23M and b) D36M. The fit of inter-nuclear vector orientations and determined alignment tensor to the experimental data is improved significantly: starting from $< Q_{static} >=0.178$ for D23M and $< Q_{static} >=0.193$ for D36M on average, the Q-values decrease to $< Q_{static} >=0.062$ for both D23M and D36M after 4 SCRM-cycles. c) and d) same as a) and b) but for $< \rho >$ instead of Q-values. Starting from $< \rho >=0.977$ for D23M and $< \rho >=0.972$ for D36M on average, improves to $< \rho >=0.997$ after 4 SCRM cycles.

observed for Arg54 with $\kappa=22.4°$, also for Lys48 and Lys6 the deviations are high with $\kappa=9.3°$ and $\kappa=12.2°$. A tendency was observed that NH vectors involved in hydrogen bonds became more collinear to the electron donating carbonyl groups upon application of SCRM. Considering only changes greater than 3° of the NH vector orientation, 15 out 23 backbone amide groups became more parallel to the carbonyl group. For comparison, the SCRM analysis was repeated using the 1d3z NMR structure as starting structure. As expected, the results are almost identical and corroborate the robustness of the SCRM approach. The resulting $S_{rdc}^2(NH)$ and κ_j are listed in Table 3 for D23M and Table 4 for D36M, compare also Fig. 5 (a, b) .

In parallel with the improvement of the static Q-values $< Q_{static} >$ and correlation
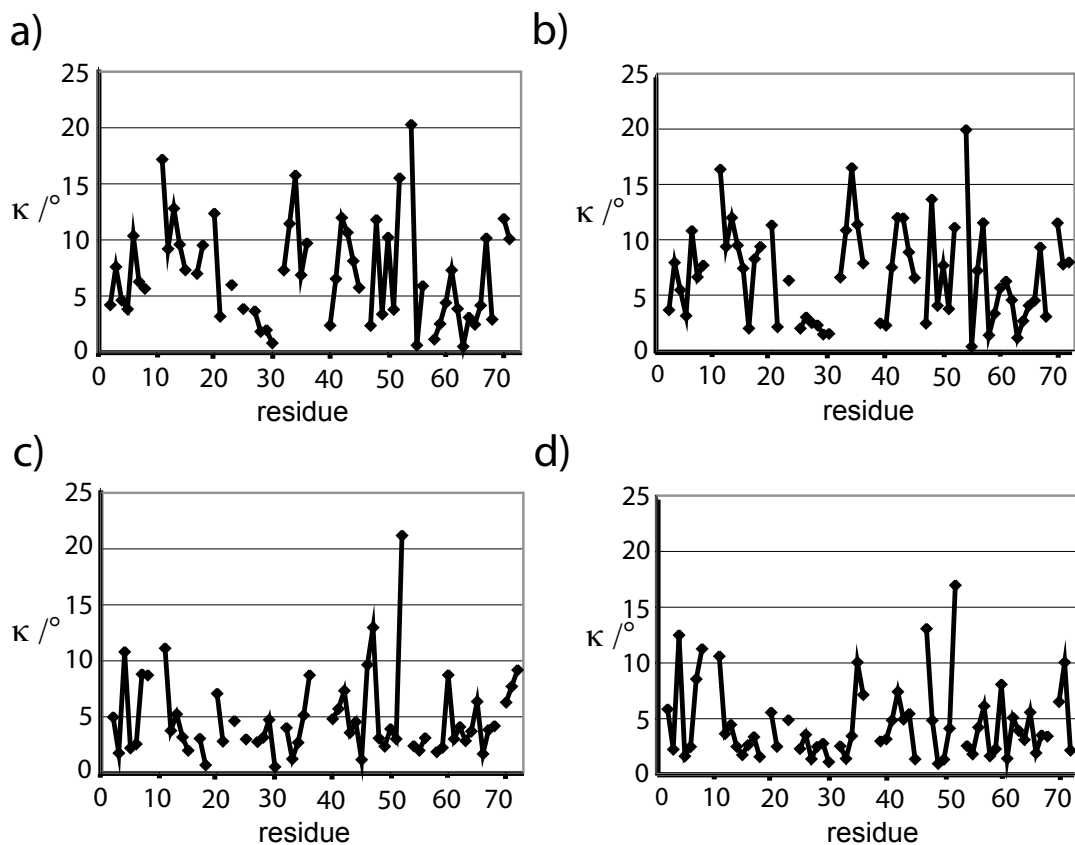
Figure 6.4: a) and b): The inter-nuclear angles κ_j enclosed between the dynamic average NH vector orientations and the NH vectors of the starting X-ray structure 1ubi are shown. The average angular deviation is 6.97° for D23M and 6.87∘ for D36M. c) and d): Same as for c) and d), but compared to the 1d3z NMR structures. The average deviation to the NMR structure is 4.84° for D23M and 4.52° for D36M. Thus, the agreement between the derived dynamic average NH vector orientations and the NMR structure is significantly better than for the 1ubi structure.

coefficients ρ, the RDC-based order parameters also converged after 4 cycles of SCRM (Fig. 6 a,b).

A more specific measure of the fit of the SCRM results to the experimental RDCs are residue-specific RDC-rmsd values which can be back-calculated from the model-free derived dynamic averaged second order spherical harmonics (compare Fig. 6 c,d). For D23M, the average RDC-rmsd was strongly reduced to <
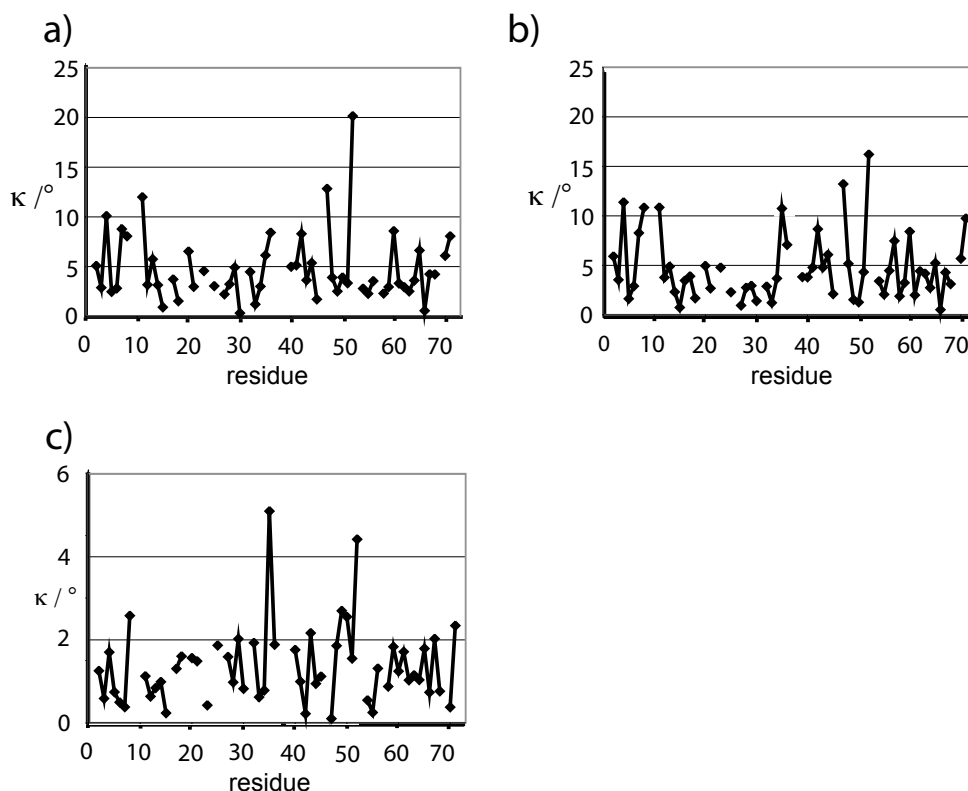
Figure 6.5: a) and b): The inter-nuclear angles κ_j enclosed between the dynamic average NH vector orientations derived from a) D23M and b) D36M using the 1d3z NMR structure as starting structure and the NH vectors of the starting 1d3z NMR structure itself are shown. The average angular deviation is 4.76° for D23M and 4.71° for D36M. c) The inter-nuclear angle κ_j enclosed between the dynamic average vectors derived from D23M and derived from D36M agree very well with and average value of 1.4° . A higher deviation is observed for Gly35 and Asp52 which also show a higher discrepancy of the S^2_{rdc}.

$rmsd(rdc, j) > = 0.28$ Hz after four SCRM-cycles compared to $< rmsd(rdc, j) >= 0.52$ Hz for D36M. To estimate the remaining inhomogeneity in the data, we added Gaussian noise to the noise-free back-calculated RDCs until the SECONDA gap reached 6.8, the value found for D23M. That analysis yields an estimated inhomogeneity for the D23M dataset of 0.22 Hz. Since it is not expected that removal of only 13 data sets reduces the RDC rmsd by almost a factor of 2, this result indicates that the resulting set D23M is more homogeneous, consistent with the SECONDA analysis. As mentioned above, SECONDA homogeneity is neither compatible with significant structural changes induced by the alignment
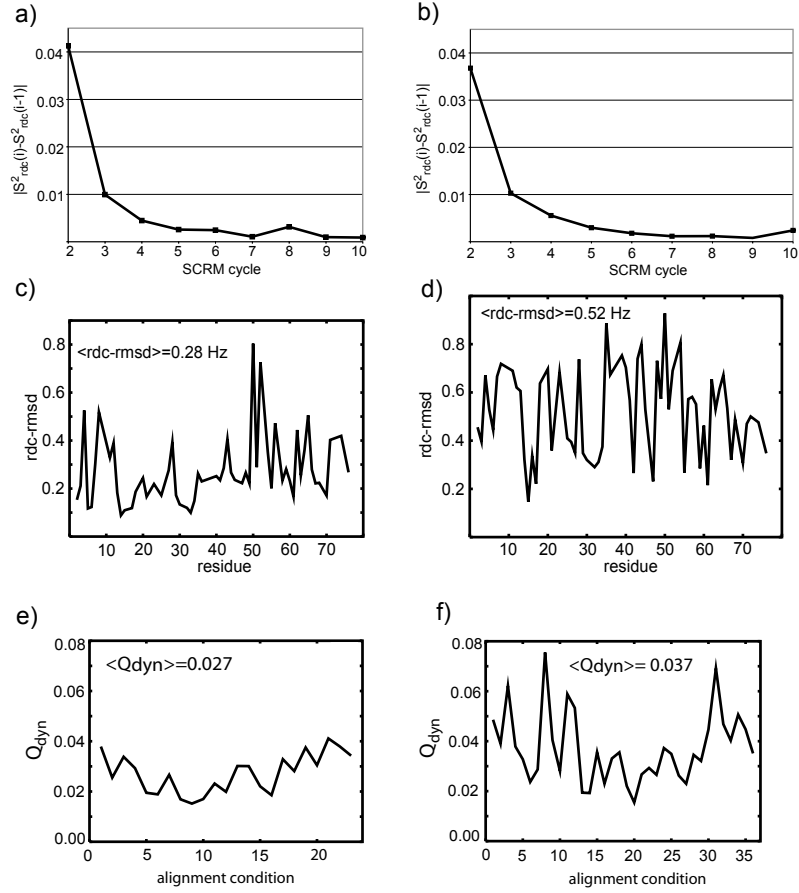
Figure 6.6: a) and b): The average difference of RDC-based $S^2_{rdc}(NH)$ order parameter between subsequent SCRM cycles for a) D23M and b) D36M is shown: $|S^2_{rdc,i} - S^2_{rdc,i-1}| = \frac{1}{n} \sum_{j=1}^{n} |S^2_{rdc,i}(NH_j) - S^2_{rdc,i-1}(NH_j)|$. RDC-based order parameters have converged after 4 cycles of SCRM with less than 0.01 difference $|S^2_{rdc,i} - S^2_{rdc,i-1}|$ between subsequent SCRM cycles. c) and d): Residue-specific RDC-rmsd values $rmsd(rdc, j)$ are shown for c) D23M and d) D36M after 4 SCRM cycles. For D23M, the average RDC-rmsd is $< rmsd(rdc, j) >= 0.28$ Hz and 0.52 Hz for D36M. e) and f): Dynamic Q-values Q_{dyn} for the different alignment conditions are back-calculated, for D23M the average dynamic Q-value is $< Q_{dyn} > = 0.027$ and for D36M $< Q_{dyn} >= 0.037$.

media nor with significant correlation of the vector fluctuations and the alignment tensor. Thus, for D23M the error introduced by ignoring a possible correlation between internal protein dynamics and alignment tensor fluctuation is small. Further, also use of a single dynamically average alignment tensor does not seem to

introduce a considerable error. Significant deviations of $< rmsd(rdc, j) >$ values from the average can mainly be observed for loop regions indicating a possible correlation between internal dynamics and alignment tensor fluctuations for these regions, in agreement with [Salvatella et al., 2008]. Possible complications as addressed in [Louhivuori et al., 2007] are thus unlikely for the alignment conditions in the D23M subset.

For both D23M and D36M the resulting S^2_{rdc} RDC-based order parameters are identical within the error, with a very few exceptions for Gly35, Lys63 and Leu71 (Fig. 7c). The correlation coefficient between the S^2_{rdc} derived from both data sets D23M and D36M is ρ=0.945. The inter-nuclear angle κ_j enclosed between the dynamic average vectors derived from D23M and derived from D36M agree very well with an average $< \kappa >$ value of 1.4° (compare Figure 5c). A higher deviation is observed for Gly35 and Asp52 which also shows a higher discrepancy of the S^2_{rdc}.

A second measure of the fit of the SCRM results to the experimental data are the dynamic Q-values $< Q_{dyn} >$. Those were obtained from the correlation of the experimental data to the RDCs back-calculated from the model-free derived dynamically averaged second order spherical harmonics $< Y_{2M}(\theta_j^{mol}, \phi_j^{mol}) >$ in the different alignment conditions (Figure 6 e, f). For D23M the average dynamic Q-value is $< Q_{dyn} >= 0.027$ and for D36M $< Q_{dyn} >= 0.037$ which indicates that the SCRM results agree very well with the experimental RDC results. Conceptually, the RDC-based model-free method resembles a residue-wise least-square fit to the experimental RDCs. Thus the dynamic Q-values indicate the best-fit solution to a restraint-free minimization of the second order spherical harmonics $< Y_{2M}(\theta_j^{mol}, \phi_j^{mol}) >$ to the experimental RDCs. In terms of a possible RDC-based molecular dynamics ensemble refinement, a minimization of the NH vector orientation without additional force-field restraints should give a distribution of NH vector orientations that resembles the SCRM-derived results and order parameters.

6.4.2 Determination of $S_{overall}$

As shown before $S_{overall}$ is smaller than 0.89 with a confidence level of 95% using only the experimental error σ_j^{exp} of 0.3 Hz (6.3.5). The confidence level drops to 67% using the total error σ_j^{rmsd}. The found $S_{overall}$ in this work deviates from the first analysis ($S_{overall} = 0.83$, see chapter 4) since now the most mobile NH amide groups have been excluded from the alignment tensor calculation resulting in a smaller down-scaling of caused by isotropic internal dynamics. Consequently, $S_{overall}$ is expected to be larger than the value of 0.83 obtained in the analysis

in chapter (4). Note that the average $< S^2_{rdc} >= 0.72$ are equal in this and the previous analysis (vide infra).

6.4.3   Analysis of $S^2_{rdc}(NH)$ order parameter distribution shows supra-τ_c motion

In Fig. 7 (a, b) the derived NH RDC-based order parameters $S^2_{rdc}(NH)$ for D23M and D36M are compared to the Lipari-Szabo $S^2_{LS}(NH)$ order parameters. The derived $S^2_{rdc}(NH)$ order parameters are listed in Table 2 and Table 3. While for some residues S^2_{rdc} and S^2_{LS} are very similar, for others, mainly in loop regions but also in secondary structure elements, significantly lower S^2_{rdc} values are observed. The average RDC-based order parameter is $< S^2_{rdc} >= 0.72 \pm 0.02$ for D23M and for D36M compared to $< S^2_{LS} >= 0.778 \pm 0.003$ for the Lipari-Szabo order parameter. The order parameter $< S^2_{LS} >$ is a measure for the remaining rigidity in the sub-τ_c window. For the mobility in that window, $1- < S^2_{LS} >$ is the appropriate measure. Similarly, the supra-τ_c mobility is measured by

$$1- < S^2_{supra-\tau_c} >= 1 - \frac{< S^2_{rdc} >}{< S^2_{LS} >} \qquad . \qquad (6.13)$$

Accordingly, inclusion of the supra-τ_c window increases the averaged amplitude of mobility observed in the sub-τ_c window by:

$$\frac{supra - \tau_c mobility}{sub - \tau_c mobility} = \frac{1- < S^2_{supra-\tau_c} >}{1- < S^2_{sub-\tau_c} >} = \frac{1 - \frac{<S^2_{rdc}>}{<S^2_{LS}>}}{1- < S^2_{LS} >} = 34\%. \qquad (6.14)$$

For D23M, N=57 S^2_{rdc} order parameters was derived and N=62 for D36M. For the S^2_{LS} N=49 S^2_{LS} were available. The correlation coefficient between S^2_{rdc} and S^2_{LS} is ρ=0.45 for D23M (and ρ=0.41 for D36M).

For both data sets,S^2_{rdc} order parameters show a significantly broader spread than those derived from relaxation, S^2_{LS}. With an interquantile range of IQR= 0.12 for D23M (and 0.12 for D36M) compared to IQR= 0.05 for the Lipari-Szabo order parameters, the distribution of S^2_{rdc} order parameters is 2.4 times wider than for S^2_{LS}(Fig. 8 a, b). In conclusion the RDC-based order parameters sample additional motion beyond τ_c. The fact that all S^2_{rdc} must be smaller than the S^2_{LS}, together with the much wider spread of the S^2_{rdc} distribution, leads on average to lower RDC-based order parameters S^2_{rdc} .

AA #	S^2_{rdc}	$\Delta_{exp}S^2_{rdc}$	$\Delta_{exp+mod}S^2_{rdc}$	AA #	S^2_{rdc}	$\Delta_{exp}S^2_{rdc}$	$\Delta_{exp+mod}S^2_{rdc}$
1	n	n	n	39	n	n	n
2	0.85	0.06	0.06	40	0.67	0.03	0.03
3	0.68	0.07	0.07	41	0.80	0.03	0.03
4	0.79	0.06	0.10	42	0.69	0.06	0.06
5	0.75	0.04	0.04	43	0.77	0.06	0.08
6	0.78	0.04	0.04	44	0.75	0.03	0.03
7	0.64	0.04	0.04	45	0.80	0.06	0.06
8	0.67	0.04	0.06	46	n	n	n
9	n	n	n	47	0.77	0.05	0.05
10	n	n	n	48	0.59	0.03	0.03
11	0.45	0.02	0.02	49	0.72	0.04	0.04
12	0.64	0.05	0.07	50	0.59	0.03	0.08
13	0.68	0.03	0.03	51	0.74	0.07	0.07
14	0.78	0.05	0.05	52	0.69	0.03	0.06
15	0.72	0.11	0.11	53	n	n	n
16	n	n	n	54	0.69	0.03	0.03
17	0.86	0.06	0.06	55	0.77	0.05	0.05
18	0.77	0.04	0.04	56	0.72	0.04	0.06
19	n	n	n	57	n	n	n
20	0.55	0.05	0.05	58	0.87	0.08	0.08
21	0.86	0.05	0.05	59	0.81	0.04	0.04
22	n	n	n	60	0.80	0.03	0.03
23	0.80	0.04	0.04	61	0.90	0.04	0.04
24	n	n	n	62	0.55	0.01	0.02
25	0.83	0.05	0.05	63	0.79	0.03	0.03
26	n	n	n	64	0.88	0.05	0.07
27	0.82	0.04	0.04	65	0.66	0.04	0.08
28	0.85	0.04	0.06	66	0.87	0.06	0.06
29	0.73	0.04	0.04	67	0.82	0.04	0.04
30	0.77	0.04	0.04	68	0.80	0.04	0.04
31	n	n	n	69	n	n	n
32	0.84	0.04	0.04	70	0.70	0.04	0.04
33	0.76	0.04	0.04	71	0.57	0.03	0.04
34	0.73	0.04	0.04	72	n	n	n
35	0.82	0.04	0.04	73	n	n	n
36	0.76	0.04	0.04	74	0.17	0.02	0.04
37	n	n	n	75	n	n	n
38	n	n	n	76	0.02	0.01	0.01

Table 6.2: The order parameters S^2_{rdc} are obtained using the experimental NH RDC data collection D23M. The first line is the amino acid number, second line the order parameter S^2_{rdc} , third the error $\Delta_{exp}S^2_{rdc}$ on the S^2_{rdc} values propagated from the experimental error $\sigma^{exp}_j = 0.3$ Hz and fourth line the error $\Delta_{exp+mod}S^2_{rdc}$ on the S^2_{rdc} propagated from the experimental plus model error $\sigma^{rmsd}_j = rmsd(rdc, j)$ (derived using the RDC-rmsd as the input error for the SCRM analysis.)

AA #	S^2_{rdc}	$\Delta_{exp}S^2_{rdc}$	$\Delta_{exp+mod}S^2_{rdc}$		AA #	S^2_{rdc}	$\Delta_{exp}S^2_{rdc}$	$\Delta_{exp+mod}S^2_{rdc}$
1	n	N	n		39	0.68	0.03	0.07
2	0.89	0.05	0.08		40	0.71	0.03	0.08
3	0.75	0.04	0.06		41	0.77	0.02	0.04
4	0.73	0.03	0.07		42	0.74	0.03	0.03
5	0.76	0.04	0.06		43	0.75	0.03	0.08
6	0.79	0.03	0.05		44	0.75	0.03	0.09
7	0.65	0.03	0.07		45	0.78	0.04	0.07
8	0.67	0.04	0.10		46	n	n	n
9	n	n	n		47	0.79	0.05	0.05
10	n	n	n		48	0.58	0.02	0.06
11	0.45	0.02	0.05		49	0.77	0.04	0.08
12	0.70	0.05	0.10		50	0.63	0.03	0.10
13	0.68	0.04	0.07		51	0.84	0.05	0.08
14	0.78	0.04	0.04		52	0.62	0.01	0.03
15	0.74	0.06	0.06		53	n	n	n
16	0.77	0.06	0.06		54	0.64	0.03	0.08
17	0.83	0.05	0.05		55	0.77	0.04	0.04
18	0.80	0.04	0.08		56	0.76	0.02	0.04
19	n	n	n		57	0.86	0.02	0.04
20	0.59	0.05	0.11		58	0.84	0.05	0.09
21	0.78	0.04	0.05		59	0.83	0.03	0.03
22	n	n	n		60	0.80	0.02	0.04
23	0.83	0.04	0.08		61	0.83	0.03	0.03
24	n	n	n		62	0.57	0.01	0.03
25	0.86	0.04	0.06		63	0.68	0.03	0.05
26	0.78	0.06	0.08		64	0.79	0.04	0.07
27	0.82	0.04	0.04		65	0.63	0.02	0.05
28	0.82	0.04	0.09		66	0.81	0.03	0.06
29	0.78	0.04	0.05		67	0.83	0.03	0.04
30	0.79	0.03	0.04		68	0.83	0.03	0.05
31	n	n	n		69	n	n	n
32	0.85	0.04	0.04		70	0.72	0.03	0.03
33	0.75	0.03	0.04		71	0.67	0.02	0.04
34	0.75	0.03	0.04		72	0.56	0.03	0.05
35	0.64	0.03	0.08		73	n	n	n
36	0.77	0.04	0.08		74	0.20	0.02	0.03
37	n	n	n		75	n	n	n
38	n	n	n		76	0.02	0.01	0.01

Table 6.3: Same as before but with D36M
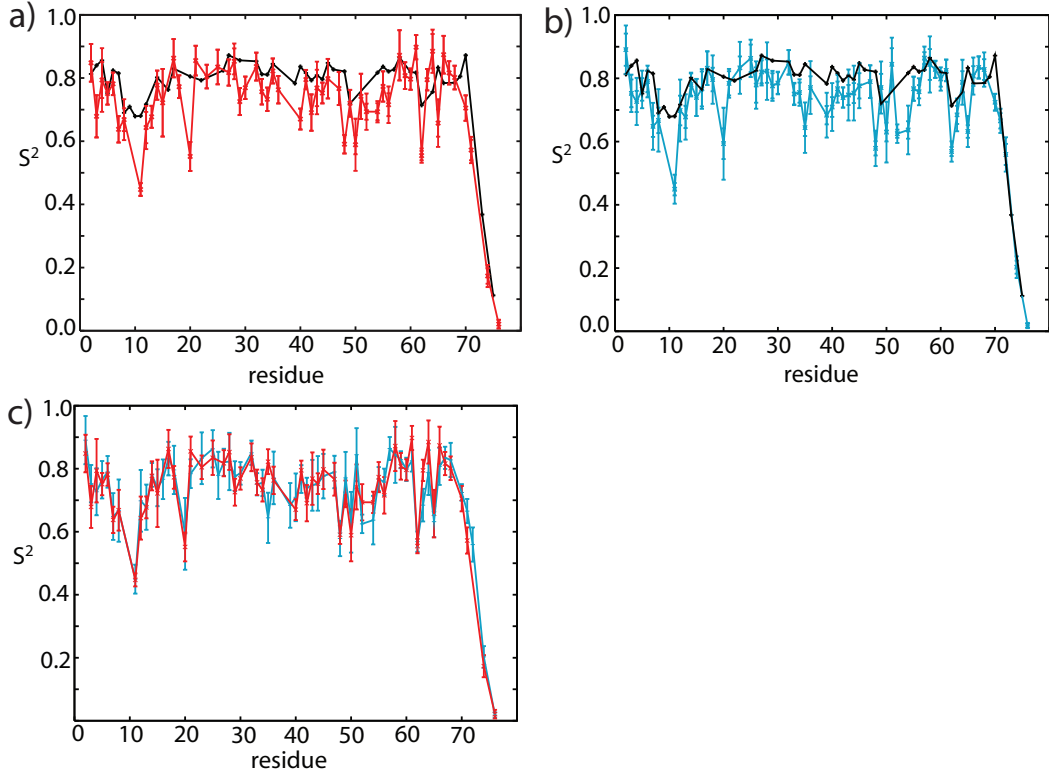
Figure 6.7: RDC-based $S^2_{rdc}(NH)$ order parameters (red and blue) scaled according to the method described in chapter 6.3.5 are compared to the Lipari-Szabo (black) for a) D23M and b) D36M. Both error bars for $\sigma^{exp}_j = 0.3$ Hz and $\sigma^{rmsd}_j = rmsd(rdc, j)$ are indicated as horizontal lines. While for some residues S^2_{rdc} and S^2_{LS} have almost equal values, for others, mainly in loop regions but also in secondary structure elements, S^2_{rdc} values are significantly lower. The average RDC-based order parameter is $S^2_{rdc} = 0.72 \pm 0.02$ for D23M and $S^2_{rdc} = 0.72 \pm 0.02$ for D36M compared to $S^2_{LS} = 0.778 \pm 0.003$ for the Lipari-Szabo order parameter. (c) RDC-based order parameters S^2_{rdc} derived from D23M (red) and D36M (blue) are compared. Both data sets D36M and D23M give S^2_{rdc} that are identical within the error, with a very few exceptions for Gly35, Lys63 and Leu71. The correlation coefficient is $\rho = 0.945$.

Supra-τ_c motion is observed mainly in loop regions like (7-11, 20, 36-40, 46-47, 50-56, 60-65, 72-76), but also for several residues in secondary structure elements (2-6, 12-17, 22-35, 41-45, 48-49, 57-59, 66-71). The average RDC-based order parameter is $< S^2_{rdc,loop} > = 0.66 \pm 0.04$ from D23M and $< S^2_{rdc,loop} > = 0.65 \pm 0.04$ from D36M for loop regions. These values are about 10% smaller than the Lipari-Szabo
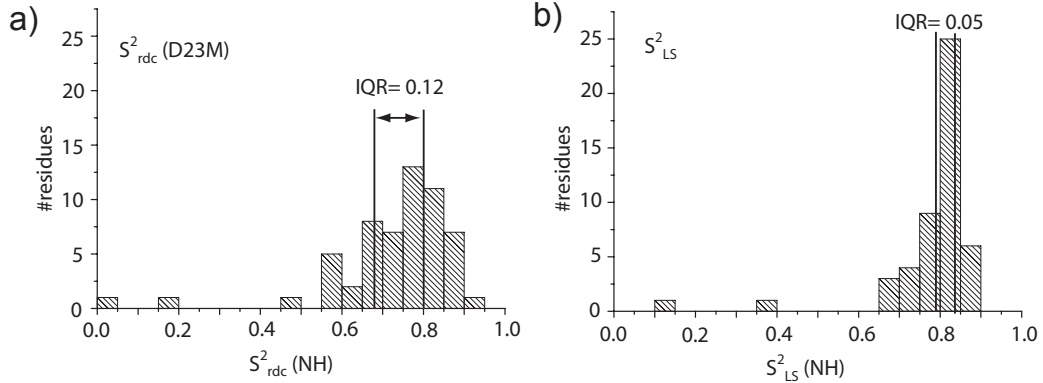
Figure 6.8: Comparison of a) $S^2_{rdc}(NH)$ and b) $S^2_{LS}(NH)$ order parameter distributions, The 25th percentile of the S^2_{rdc} distribution is P25=0.68, the 75th percentile is P75=0.80 for D23M, giving and interquantile range (P25 to P75) of IQR= 0.12. (Identical values are obtained for D36M.) In contrast, the distribution of Lipari-Szabo order parameters S^2_{LS} is 2.4 times narrower with P25= 0.78, P75=0.83 and an interquantile range of IQR= 0.05. For the RDC-based order parameter S^2_{rdc} the IQR is more than double than that for the Lipari-Szabo S^2_{LS} showing that the RDC-based order parameters detect a much wider range of mobility.

value of $< S^2_{LS,loop} >= 0.72\pm0.04$ for the loop regions. For secondary structure elements the average RDC-based order-parameter is $< S^2_{rdc,sec} >= 0.77\pm0.01$ (both for D23M and D36M) and still about 5% smaller than $< S^2_{LS,sec} >= 0.81 \pm 0.01$. The presence of supra-$\tau_c$ motion in secondary structure elements is emphasized by comparing the 25th and 75th percentile P25=0.72 and P75=0.82 of the RDC-based order parameter in secondary structure elements $< S^2_{rdc,sec} >$ derived from D23M (P25=0.75 and P75=0.83 in the case of D36M) to P25=0.79 and P75=0.85 for the Lipari-Szabo ones $< S^2_{LS,sec} >$. Table 4 lists all parameters describing the distribution and the distribution.

Interestingly, an alternating pattern of $S^2_{rdc}(NH)$ order parameters was extracted for residues Lys48 to Leu50 in the 4th β-strand. The backbone of Lys48, whose side chain is known to be involved in the poly-ubiquitination process that leads to protein trafficking and degradation, appears very mobile with an order parameter of $S^2_{rdc}(NH) = 0.59 \pm 0.03$ for D23M ($S^2_{rdc}(NH) = 0.58 \pm 0.07$ for D36M) compared to $S^2_{LS}(NH)$=0.82 for the Lipari-Szabo value. Other alternating patterns

of $S^2_{rdc}(NH)$ in β-sheets like Gln41 to Phe45 that have been described before in chapter(5, [Lakomek et al., 2005]) are reproduced in this analysis for D23M, however with reduced amplitude. The same alternating pattern is observed weakly also for Lipari-Szabo order parameters $S^2_{LS}(NH)$ [Chang and Tjandra, 2005]. These findings are consistent with our earlier analyses , independent findings for protein G using the 3D-GAF analysis [Bouvignies et al., 2005b] and even earlier results by Palmer and co-workers for Ribonuclease H [Mandel et al., 1995] [Mandel et al., 1996] and Fibronectin type III using relaxation methods. A correlation between backbone mobility and side-chain orientation has recently also been extracted from ultra high-resolution X-ray structures [Davis et al., 2006].

	all $S^2_{rdc}(NH)$	all S^2_{LS}	Sec. Struct. S^2_{rdc}	Sec. Struct. S^2_{LS}	Loops S^2_{rdc}	Loops S^2_{LS}
	(a) D23M					
$< \cdot >$	0.72	0.778	0.77	0.81	0.66	0.72
	± 0.02	± 0.003	± 0.01	± 0.01	± 0.04	± 0.04
P25	0.68	0.78	0.72	0.79	0.59	0.69
P75	0.80	0.83	0.82	0.85	0.77	0.82
IQR	0.12	0.05	0.10	0.05	0.18	0.13
N	57	49	34	31	23	18
	(b) D36M					
$< \cdot >$	0.72	0.778	0.77	0.81	0.65	0.72
	± 0.02	± 0.003	± 0.01	± 0.01	± 0.04	± 0.04
P25	0.68	0.78	0.75	0.79	0.61	0.69
P75	0.80	0.83	0.83	0.85	0.78	0.82
IQR	0.12	0.05	0.08	0.05	0.17	0.13
N	62	49	37	31	25	18

Table 6.4: a) Statistics on the RDC-based order parameters $S^2_{rdc}(NH)$ (derived from D23M). The average values, the 25th and 75th percentile as well as the interquantile range IQR and the number of analysed residues N are shown and compared to the data from relaxation $S^2_{LS}(NH)$. The 1st and 2nd column show the values for the complete distribution, in the 3rd and 4th columns and the 5th and 6th columns we distinguish between secondary structure elements and loop regions. The percentile values reveal a much wider distribution of $S^2_{rdc}(NH)$ order parameter than for the Lipari-Szabo $S^2_{LS}(NH)$. This effect is most visible for loop regions but also for secondary structure elements. b): The same statistics as in a) but for $S^2_{rdc}(NH)$ derived from D36M.

6.4.4 Focus on supra-τ_c motion

To distinguish supra-τ_c motion from sub-τ_c motion, the distribution of $S^2_{rdc}/S^2_{LS}(NH)$ was analyzed along the amino acid sequence of ubiquitin (Table 5). For residues with solvent-exposed side chains, the backbone amide groups appear more mobile, while residues with side chains pointing to the hydrophobic core of the protein appear more rigid in the protein backbone, in agreement with the previous analysis in chapter(5), see also [Lakomek et al., 2005]. The analysis has been applied in the same way as described before. A very simple two-state model has been applied. All residues with a solvent accessibility less than 11.5% were considered as core residues, all others as solvent-exposed. Solvent accessibility has been calculated with the help of MOLMOL [Koradi et al., 1996]. The average $S^2_{rdc}/S^2_{LS}(NH)$ value is 0.90 ± 0.02 for solvent-exposed residues and 0.93 ± 0.03 for core residues in the case of D23M (and 0.90 ± 0.02 and 0.93 ± 0.02 for D36M) which reveals a tendency of core residues to be more rigid. The 25th percentile is P25=0.81 for the class of solvent-exposed residues and P25= 0.87 for the core residues for D23M (and P25= 0.81 and P25= 0.92 for D36M). This indicates a tendency for residues with solvent exposed side-chains to be more mobile in the protein backbone than those with side chains pointing towards the hydrophobic core [Lakomek et al., 2005].

The dependence of $S^2_{rdc}/S^2_{LS}(NH)$ values on the number of hydrogen bonds on the corresponding peptide plane (including the amino group NH(i) and the preceding carbonyl group CO(i-1)) is analyzed in the same way as before in chapter (5, see also Lakomek2005). Peptide planes that are not involved in a hydrogen bond appear more mobile than those that are hydrogen-bonded: The average RDC-based order parameter is $S^2_{rdc}/S^2_{LS}(NH) = 0.81 \pm 0.05$ for D23M (0.82 ± 0.05 for D36M) when the peptide plane is not involved in hydrogen bonds, compared to 0.92 ± 0.02 (D23M and D36M) when the peptide plane is involved in at least one hydrogen bond. For details, see Table 5.

6.4.5 Comparison to previous analyses

The correlation coefficient between the S^2_{rdc} derived in this analysis and the previous one (cf. chapter 4, [Lakomek et al., 2006]) is $\rho =0.80$ for D23M and $\rho=0.82$ for D36M. Both analyses yield an average $< S^2_{rdc} >$ of 0.72, which underlines the presence of motion beyond the overall tumbling correlation time τ_c . These results also highlight that it is important to remove the possible bias introduced by the structure used for the tensor calculation. In the previous analysis [Lakomek et al., 2006] some outliers were present, for which $S^2_{rdc}(NH)$ order parameters were larger than the corresponding Lipari-Szabo $S^2_{LS}(NH)$. These

$S^2_{rdc}/S^2_{LS}(NH)$	Core(i)	Solvent(i)	0 hb	1 hb	2 hb
(a) Derived from D23M					
$< S^2_{rdc}/S^2_{LS}(NH) >$	0.93	0.90	0.81	0.92	0.92
	±0.03	±0.02	±0.05	±0.02	±0.02
P25	0.87	0.81	0.78	0.87	0.85
P75	1.00	0.97	0.85	1.00	0.98
IQR	0.13	0.16	0.07	0.13	0.13
N	10	28	5	22	11
(b) Derived from D36M					
$< S^2_{rdc}/S^2_{LS}(NH) >$	0.93	0.90	0.82	0.92	0.93
	± 0.02	± 0.02	±0.04	±0.02	±0.02
P25	0.92	0.81	0.78	0.92	0.91
P75	1.00	1.00	0.87	1.00	1.00
IQR	0.08	0.19	0.09	0.08	0.09
N	11	31	6	23	13

Table 6.5: a): Statistics on the RDC-based order parameters $S^2_{rdc}/S^2_{LS}(NH)$ (derived from D23M) which describe the supra-τ_c contribution to mobility. The average values, the 25th and 75th percentile as well as the interquantile range IQR and the number of analyzed residues N are shown. The 1st and 2nd column distinguish between residues with solvent-exposed side chains and those pointing towards the hydrophobic core. The average $S^2_{rdc}/S^2_{LS}(NH)$ reveals a slight tendency for core residues to be more rigid. 25% of the $S^2_{rdc}/S^2_{LS}(NH)$ distribution are lower than the 25th percentile which is P25=0.81 for the class of solvent-exposed residues and P25= 0.87 for the core residues in the case of D23M. This reveals a tendency for residues with solvent exposed side-chains to be more mobile in the protein backbone [Lakomek et al., 2005]. The dependence of $S^2_{rdc}/S^2_{LS}(NH)$ values on the number of hydrogen bonds the residue is involved in is analyzed columns 3 to 5. (Each NH group is classified according to the number of backbone hydrogen bonds on the corresponding peptide plane, including the amino acids NH(i) and the preceding carbonyl group CO(i-1).) Peptide planes that are not involved in a hydrogen bond (hb0) appear more mobile than those that are hydrogen-bonded. b): The same statistics as in a) but for $S^2_{rdc}/S^2_{LS}(NH)$ derived from D36M.

were Leu8, Asp32, Gln49 and Ser57. In the SCRM analysis these residues show $S^2_{rdc}(NH)$ values lower than the corresponding $S^2_{LS}(NH)$ and are less conspicuous. These previous outliers are attributed to the influence of structural noise. As described in the next paragraph, the new SCRM method can efficiently avoid such outliers.

6.4.6 SCRM analysis is robust against the influence of structural noise

For synthetic structural noise added to the starting structure for the SCRM analysis (see Materials and Methods case A), the resulting $S^2_{rdc}(NH)$ order parameters (using D36M) after 4 SCRM cycles are in excellent agreement with those obtained using the noise-free structure 1ubi), both for 10° and 20° Gaussian noise, as seen in Figure 9 (a, b). Even for 30° structural noise the agreement is reasonably good (see Fig. 9 c).
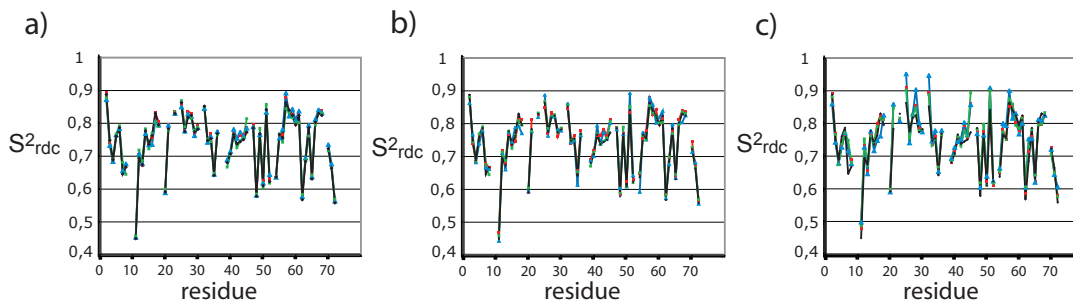
Figure 6.9: SCRM-derived S^2_{rdc} after addition of a) 10° b) 20° and c) 30° synthetic structural noise on the 1ubi X-ray structure used as starting structure for alignment tensor calculation. The agreement between the calculated S^2_{rdc} and those derived using the noise-free structure (black) shows that SCRM is robust against the influence of structural noise.

The SCRM approach has been tested on nine different input structures that deviate considerably from the free ubiquitin crystal (see Materials and Methods case B). While the original model-free approach [Lakomek et al., 2006] is affected by structural differences of the input structures used for tensor calculation (Fig. 10 a, b), the new SCRM method alleviates the effect of structural noise (Fig. 10 c, d). After only 4 SCRM cycles the $S^2_{rdc}(NH)$ order parameters of the nine different test cases have converged and agree very well with those for the free form 1ubi (Fig. 10 a, b). The standard deviation of $S^2_{rdc}(NH)$ order parameters

is σ =0.033 for the original RDC-based model-free approach applied on D23M (σ=0.039 for D36M) and σ=0.010 after 4 SCRM cycles applied on D23M (σ= 0.006 for D36M). This illustrates nicely that the SCRM method is able to accurately determine alignment tensors and inter-nuclear vector orientations almost independently from the quality of the starting structure (within a certain range). This is an important prerequisite for reliable quantification of macromolecular dynamics.

6.5 Conclusions

To become independent from the accuracy of the structural model used for alignment tensor calculation, a <u>S</u>elf-<u>C</u>onsistent <u>R</u>DC-based <u>M</u>odel-free analysis (SCRM) has been developed that delivers RDC-based order parameters independent of the details of the structure used for alignment tensor calculation, as well as the dynamic average orientation of the inter-nuclear vectors in the protein structure in a self-consistent manner . SCRM was applied on two NH RDC data set collections, D36M and D23M. For both NH RDC data collections, the new SCRM approach gives almost identical order parameters (correlation factor of ρ= 0.945). For D23M it was concluded that there are neither significant structural changes induced by the alignment media nor significant correlations of the vector fluctuations and the alignment tensors. The correlation between alignment tensor fluctuations and internal dynamics, which has been proposed as a possible source of error by [Louhivuori et al., 2006, Louhivuori et al., 2007], is therefore likely to be small. This finding agrees very well with theoretical predictions [Salvatella et al., 2008].

It has been shown in this work that the influence of structural noise on the alignment tensor determination and the resulting order parameter can be alleviated by the SCRM method. Synthetic structural noise has been added to the 1ubi X-ray structure and different structures of ubiquitin from several complexes have been used as input for the SCRM analysis. The resulting order parameters were found to agree within 0.01 irrespective of the starting structure. Thus, RDC-derived $S^2_{rdc}(NH)$ order parameters were determined with unprecedented accuracy independent of structural noise [Lakomek, Walter et. al 2008].

For ubiquitin, the SCRM analysis yields an average RDC-derived order parameter of the NH vectors $< S^2_{rdc}(NH) >= 0.72 \pm 0.02$, compared to $< S^2_{LS}(NH) >= 0.778 \pm 0.003$ for the Lipari-Szabo order parameters. Consistent with the first analysis described in chapter 4, the inclusion of the supra-τ_c window increases the averaged amplitude of mobility observed in the sub-τ_c window by about 34%. The

backbone of Lys48, whose side chain is known to be involved in the polyubiquina-tion process that leads to protein degradation, is very mobile on the supra-τ_c time scale ($S^2_{rdc}(NH) = 0.59 \pm 0.03$), while it is inconspicuous ($S^2_{LS}(NH) = 0.82$) on the sub-τ_c as well as on μs - ms relaxation dispersion time scales. This find-ing motivated the investigation of the role of supra-τ_c motion for protein-protein recognition .

Promising techniques are emerging to reduce the experimental effort of collecting enough linearly independent RDC data sets [Ruan and Tolman, 2005] [Yao and Bax, 2007]. Thus, RDCs are expected to become a routine tool to complement the analysis of biomolecular dynamics.
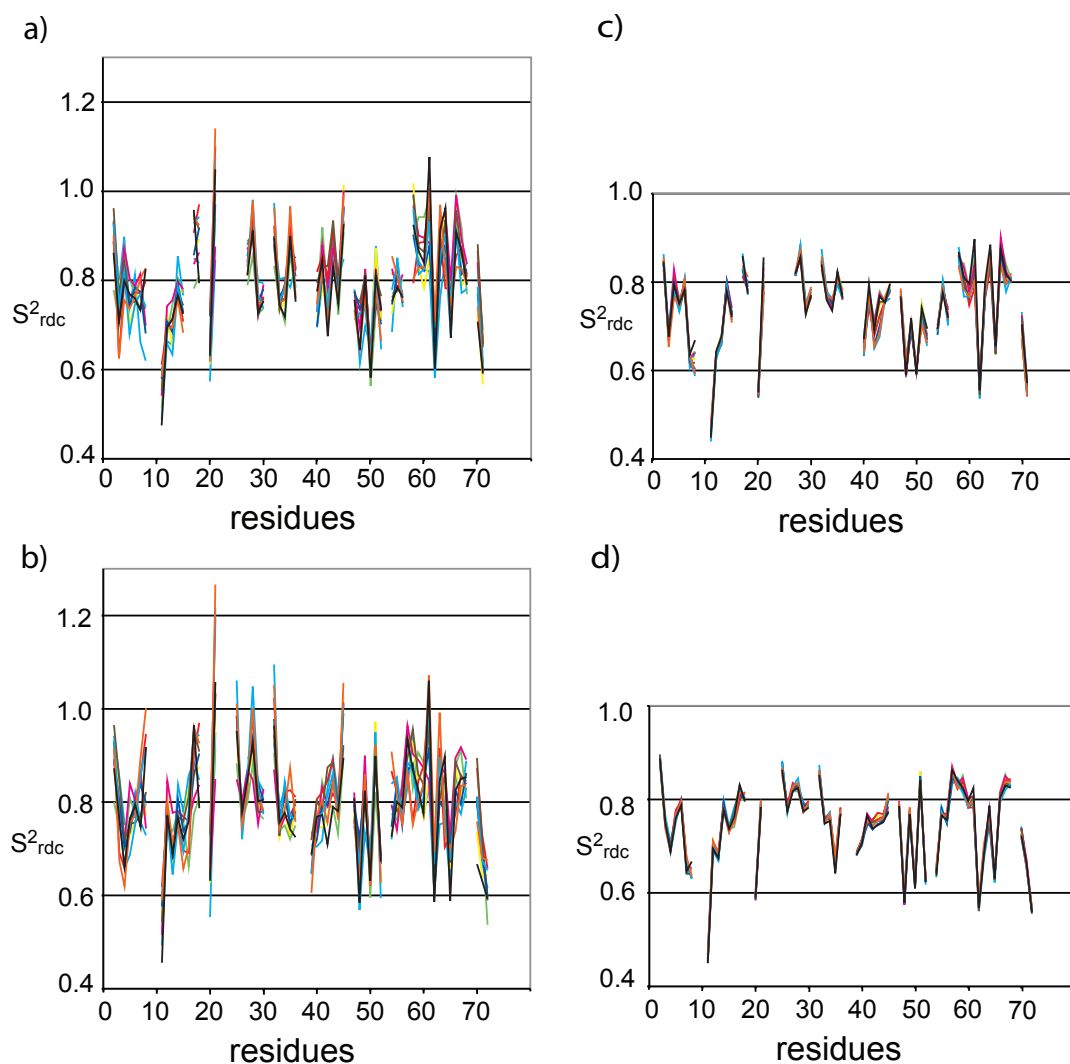
Figure 6.10: a) and b): $S^2_{rdc}(NH)$ order parameter are derived for a) D23M and b) D36M using the original RDC-based model-free analysis and nine different structures of ubiquitin bound in several complexes as starting structure for the analysis. The different structures have a backbone rmsd between 0.3 and 0.6 Å to the free Xray structure 1ubi [Ramage et al., 1994] (black line) and serve as a test case for severe structural noise. c) and d): $S^2_{rdc}(NH)$ order parameter are derived for c) D23M and d) D36M using the new SCRM method and the same nine different structures of ubiquitin (free X-ray structure 1ubi black line). While the original model-free approach [Lakomek et al., 2006] is affected by structural difference of the input structure used for tensor calculation, the new SCRM method alleviates the effect of structural noise.

92

Chapter 7

Molecular recognition in ubiquitin studied with an RDC-derived NMR solution ensemble

Protein function relies on structural protein dynamics with time-scales ranging from picoseconds to beyond seconds. For molecular recognition, for example, proteins adapt their structure to different binding partners, often exhibiting large structural heterogeneity. In the past 30 years, atomic information on many dynamical processes has been accumulated from a broad variety of techniques [Kolano et al., 2006, Frauenfelder and Wolynes, 1985] . As already described before in chapter 3, Nuclear Magnetic Resonance (NMR) relaxation has been used to quantitatively probe protein dynamics at the fast end (picoseconds to nanoseconds) as well as in a much slower range (microseconds to milliseconds) of this broad spectrum of time scales [Mittermaier and Kay, 2006, Kay et al., 1989a] [Akke and Palmer, 1996, Wang et al., 2001]. Relaxation of nuclear magnetization is caused by fluctuations of magnetic interactions between nuclei due to the nanosecond rotational tumbling of the molecule and internal dynamics. The amplitudes of these motions are expressed by Lipari-Szabo order parameters S_{LS}^2 [Lipari and Szabo, 1982a]. Internal dynamics slower than the rotational tumbling time τ_c have no impact on the overall fluctuation of the magnetic interactions. Therefore, S_{LS}^2 order parameters reflect only sub-τ_c motions, at the fast end of timescales.

The slow range of timescales is accessible by relaxation dispersion measurements, based on the stochastic fluctuations of isotropic chemical shifts, which are inde-

pendent of rotational tumbling [Akke and Palmer, 1996]
[Mittermaier and Kay, 2006]. Conformational heterogeneity slower than 10 ms can be directly observed as peak splitting in NMR spectra. For backbone amides, motions faster than 50 μs do not result in sufficient line broadening to be detectable for relaxation dispersion measurements. These measurements therefore probe motions slower than about 50 μs up to approximately 10 ms, and have been used to characterize major structural changes and enzymatic reactions [Wang et al., 2001, Henzler-Wildman et al., 2007]. Except for certain favorable cases [Boehr et al., 2006] it is, however, difficult to translate these fluctuations into ensembles of structures. Therefore, relaxation-based ensembles of solution structures take only motions faster than τ_c into account: they are limited to sub-τ_c dynamics [Lindorff-Larsen et al., 2005, Richter et al., 2007]. These sub-τ_c motions are typically much smaller than the structural changes involved in molecular recognition and likely to contribute mainly to the entropy of proteins [Akke et al., 1993, Yang et al., 1997, Frederick et al., 2007]. As a consequence, the structural heterogeneity observed in protein complexes has frequently been assumed to be inaccessible to equilibrium fluctuations in solution, thus favoring induced-fit models [Goh et al., 2004, Grunberg et al., 2004] .

7.1 Residual dipolar couplings probe supra-τ_c dynamics

Residual Dipolar Couplings (RDCs) are sensitive to motion from picoseconds to milliseconds, which includes the previously invisible time window between τ_c and 50 μs which we call supra-τ_c. Indeed, RDCs recorded for ubiquitin as well as for the B1 domain of protein G hint at significant dynamics between nano- and microseconds [Meiler et al., 2001, Peti et al., 2002, Tolman, 2002]
[Lakomek et al., 2005, Lakomek et al., 2006]
[Bouvignies et al., 2005b, Bouvignies et al., 2006, Markwick et al., 2007]. Here, we present a structural ensemble of ubiquitin based on an extensive RDC data set (Fig. 1). Ubiquitin is a key to many cellular signaling networks [Hicke et al., 2005, Harper and Schulman, 2006], as for example in protein degradation, and is recognized by a broad variety of proteins with high specificity [Brzovic and Klevit, 2006]. Accordingly, ubiquitin crystal structures of 46 different complexes show a particularly pronounced structural heterogeneity (cf. Fig 2), which cannot be explained from the available sub-τ_c ensembles refined against NMR relaxation data [Lindorff-Larsen et al., 2005, Richter et al., 2007] (Fig. 2c, 2e).

To assess the supra-τ_c time scale for ubiquitin, we measured RDCs for the backbone amide NH couplings in 18 different alignment conditions as well as back-
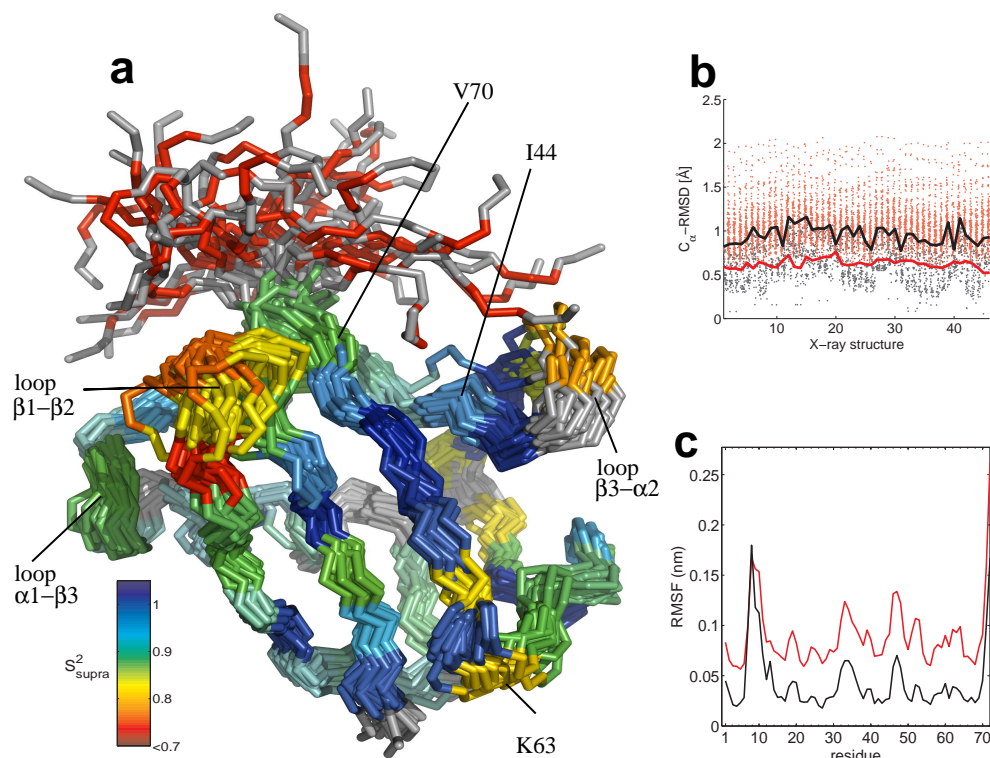
Figure 7.1: Structure ensemble of ubiquitin. **(a)** Backbone trace of 40 randomly chosen structures from the EROS ensemble. Residues are colored by the amount of additional (supra-τ_c) mobility as compared to the Lipari-Szabo order parameters (Fig. 3c) $S^2_{\mathrm{supra}} = S^2_{\mathrm{EROS}}/S^2_{\mathrm{LS}}$. **(b)** For each X-ray structure (x-axis; numbering suppl. Tab. S3), the backbone RMSDs of residues 1-70 are shown for superpositions with each EROS structure (red) and each (other) X-ray structure (black). The minimal RMSD for EROS structures (red) and the maximal RMSD for X-ray structures (black) are highlighted as a line in the respective color to guide the eye. **(c)** C_α root mean square fluctuations (RMSF) of EROS structures (red) and of 46 known ubiquitin X-ray structures (black).

bone H_NC(O) and NC(O) RDCs from 4 different alignment media. Together with data from the literature [Briggman and Tolman, 2003, Ruan and Tolman, 2005, Ottiger and Bax, 1998] , 36 NH RDC data sets and 6 H_NC(O) and NC(O) RDC data sets were available. To probe side-chain dynamics as well, side-chain methyl group RDCs measured for 11 alignement media have been included in the analysis.
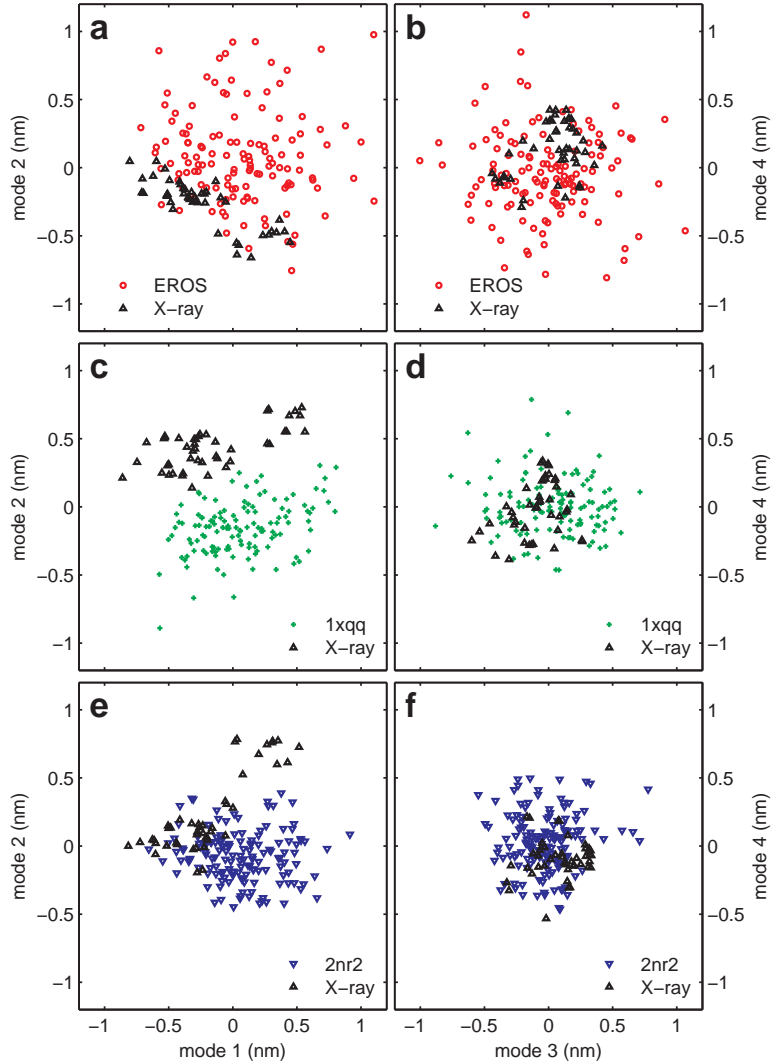
Figure 7.2: Comparison of supra-τ_c and sub-τ_c solution ensembles (colors) to the collection of 46 x-ray structures (black) of ubiquitin by principal component analysis (PCA): EROS **(a,b)**, 1xqq **(c,d)**, and 2nr2 **(e,f)**. The PCA was carried out over the merged 2 ensembles that are displayed (in each case the x-ray ensemble and one NMR ensemble: EROS, 1xqq and 2nr2). Panels a, c, and e show projections onto the principal modes 1 and 2, panels b, d and f show projections onto eigenvectors 3 and 4. Systematic deviations are observed along the principal modes for both sub-τ_c ensembles, but not for the supra-τ_c EROS ensemble.

7.2 Supra-τ_c ubiquitin ensemble reveals conformational selection

To extract a structual ensemble from these data, we carried out cross-validated ensemble refinement from unfolded structures in explicit solvent subjected simultaneously to restraints from NMR Nuclear Overhauser Enhancement (NOE) and RDC data (henceforth referred to as EROS for Ensemble Refinement with Orientational restraintS). Interestingly, the unperturbed protein exhibits considerable flexibility, with a substantial fraction (color coded, Fig. 1a) attributed to supra-τ_c. Slower motions, at the micro- to millisecond time scale, have previously only been observed for a very limited number of residues [Massi et al., 2005] thus confining the additional motion to the time range between the correlation time and approximately $50\,\mu$s. As a cross validation, the ensemble was also calculated without NOEs. The resulting ensemble was found to be virtually unchanged, indicating that the ensemble is predominantly defined by the RDC data.

Surprisingly, this supra-τ_c ensemble comprises the complete range of crystallographically observed structural changes during interface engagement (Fig. 1b, 2a), in contrast to the known fast dynamics (Fig. 2c, 2e) [Lindorff-Larsen et al., 2005, Richter et al., 2007] . Indeed, each of the X-ray structures is similar to members of the solution ensemble within less than 0.8 Å root mean square (RMS) backbone deviation (Fig. 1b) despite the fact that no crystallographic data has been used during refinement. Conformational selection, rather than induced fit, thus suffices to explain all known structural adaptations that the ubiquitin backbone undergoes upon complex formation with different binding partners. Remaining induced fit motions are restricted to rotameric side chain rearrangements and minor backbone changes.

As an independent validation of our ensemble, we have compared the results to those of the self-consistent RDC-based model-free (SCRM) analysis (cf. chapter 6, [Lakomek et al., 2008]). For comparison, generalized order parameters were also computed from the EROS ensemble. A correlation coefficient $r = 0.74$ between S^2_{SCRM} and S^2_{EROS} is found (Fig. 3a). This agreement between two independent approaches shows that the dynamics observed in the EROS ensemble is indeed strongly determined by the experimental RDC data. This conclusion is supported by rigorous cross-validation implemented in EROS by systematically leaving out all RDCs between backbone amide N and carbonyl C, as well as all scalar couplings, from refinement. The ensemble averaged free RDC R-factor of 18.5% is considerably lower than for other solution ensembles ($> 24\%$, cf. supplement, Tab. S2). Combining all X-ray structures into an "ensemble" [Best et al., 2006] we obtained a similarly low R-factor of 18.3%. Compared to
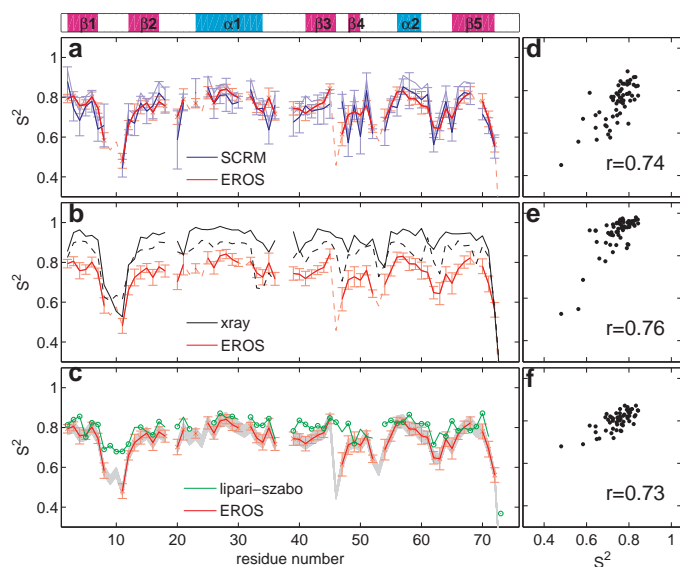
Figure 7.3: Comparison of NH-order parameters of ubiquitin. The order parameters of the presented EROS ensemble (red) are compared to **(a,d)** SCRM order parameters (blue) derived from the NH part of the RDC data used for EROS. **(b,e)** order parameters intrinsic to the ensemble of 46 crystallographic structures (black). The dashed curve is obtained when the 46 structures are relaxed at 300K by short molecular dynamics simulations of 10 ps. **(c,f)** generalized order parameters obtained from NMR relaxation data (green) for the sub-τ_c dynamics of ubiquitin via Lipari-Szabo model-free analysis [Chang and Tjandra, 2005]. Green circles mark the data-points taken from the most recent and accurate measurement [Chang and Tjandra, 2005], remaining data points are taken from previously published data [Tjandra et al., 1995]. The latter [Tjandra et al., 1995] were rescaled such that they align with the newer results [Chang and Tjandra, 2005]. The EROS order parameters were scaled by 0.93 to account for limited ensemble size and underestimation of the librational contribution. Error bars (1 sigma) for the EROS ensemble (light-red) comprise intrinsic sampling and force-field errors as well as propagated experimental errors. The uncertainty in the libration correction was estimated as $\pm 4\%$ and is represented grey-shaded. A solid line is shown for residues where sufficient RDC data was available to determine a robust value with SCRM; For the other positions, EROS order parameters are shown as dashed line. **(d-f)** Scatter plots for a direct comparison of the two sets of order parameters that are shown to the left of the respective plot.

the R-factor of $25 \pm 4\%$ for individual X-ray conformers, this result confirms that the conformational heterogeneity as found in the EROS ensemble and in

the X-ray data considerably improves the description of the experimental solution NMR data. In addition, the correlation between order parameters derived from the X-ray "ensemble", particularly when relaxed in short (10 ps) molecular dynamics simulations at 300K (Fig. 3b and supplement), and the RDC-derived order parameters S^2_{EROS} and S^2_{SCRM} suggests that the interconversion between the different ubiquitin conformations in the X-ray ensemble strongly contributes to the solution dynamics.

To assess how much of the solution dynamics is slower than τ_c, we compare S^2_{EROS} and S^2_{SCRM} to order parameters derived from NMR relaxation measurements. The pico- to nanosecond time scale dynamics of the ubiquitin backbone was probed previously by NMR relaxation techniques, yielding a set of S^2_{LS} order parameters as derived from a Lipari-Szabo analysis [Lipari and Szabo, 1982a, Chang and Tjandra, 2005]. Fig. 3c compares order parameters S^2_{EROS} from the ensemble presented in Fig. 1a with S^2_{LS} order parameters. For most residues additional mobility is seen, thus quantifying the supra-τ_c motion in the EROS ensemble, shown as color code in Fig. 1a. For EROS, absolute order parameters were derived from the RDC-refined ensemble and corrected for limited ensemble size and libration effects. For SCRM, the absolute scale was determined relative to S^2_{LS} order parameters, with S^2_{LS} an upper bound for S^2_{SCRM}, within the error bars. It is interesting to note that, despite the fact that the RDCs do not provide the absolute amplitude of the dynamics, the overall scale of the independently determined S^2_{EROS} and S^2_{SCRM} is nearly identical.

7.3 Solution fluctuations allow for interface contact formation

As noted above, the supra-τ_c motion accesses all the conformations that are observed in complex structures. To rationalize this unexpected result, Fig. 4a overlays all interface-contacts (gray spheres) of the different binding partners found in the X-ray structures with a single structure of ubiquitin whose coloring represents the solution dynamics as given by S^2_{EROS} (NH). It is striking that helix $\alpha 1$, for which no contacts are observed, shows only little motion in solution (blue), whereas high flexibility (orange-red) is observed in regions that form many different protein-protein interfaces. A quantitative analysis of the number of interface contacts per residue (Fig. 4c) shows an unexpectedly high similarity to the conceptually unrelated order parameters S^2_{EROS} (NH), which corroborates this initial observation.

Two prominent exceptions from the observed high-flexibility in the binding re-
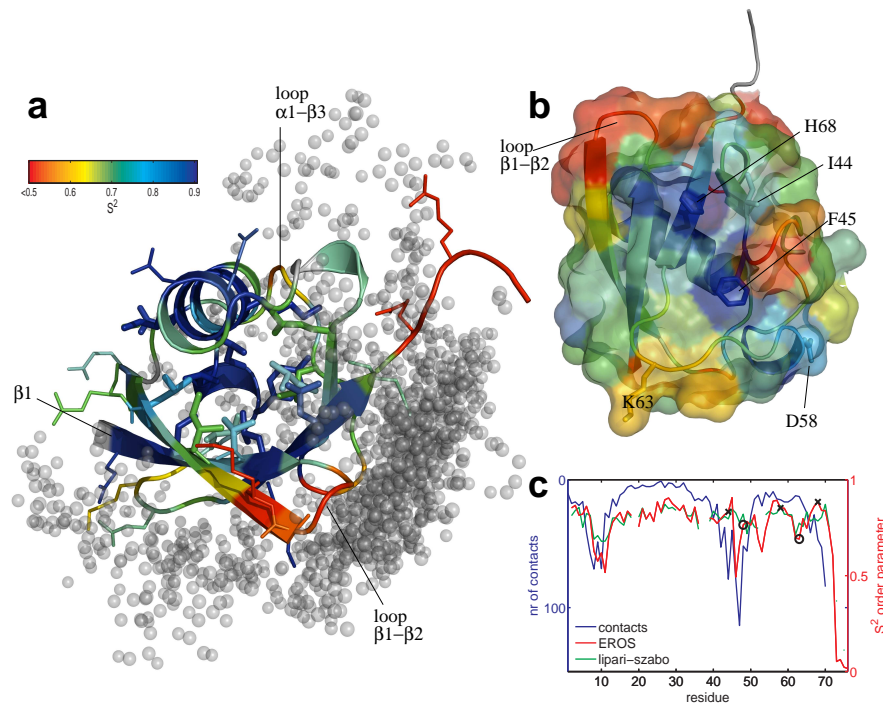
Figure 7.4: Solution dynamics correlates with molecular recognition sites. **(a-b)** The apo-structure of ubiquitin (1UBI) is colored by backbone flexibility in solution as given by S^2_{EROS}. **(a)** Positions of contacting atoms of complexing proteins ($< 5\,\text{Å}$ distance) are shown as gray spheres. **(b)** View towards the surface at the most prominent recognition site around residues ILE44/HIS68. HIS68 (sticks) lies within a rigid crevice that connects via PHE45 to the other known recognition site centered at ASP58. The walls of this crevice are formed by regions with high flexibility. Around HIS68 rigidity is provided by packing of core residues LEU67 and LEU69 (not shown) against the central helix; at ASP58 packing of LEU55 and a long-range hydrogen bond from TYR59 to GLU51 provides stability. **(c)** Ubiquitin-binding protein contacts per residue (blue) and the flexibility in solution for the sub-τ_c time regime (geen) and the supra-τ_c time range, as extracted from the EROS ensemble (red). A remarkable correlation between contacts and solution fluctuations is observed, particularly for the EROS ensemble. Exceptions from the observed correlation are found for known molecular recognition hot spots (marked with crosses: ILE44/HIS68, ASP58), which may act as rigid anchors, allowing flexibility for neighboring residues. Lysines responsible for polyubiquitination are marked with circles (LYS48, LYS63).

gions are residue I44 and H68 (crosses in Fig. 4c). Both are known from mutation studies to be central hot spot [Clackson and Wells, 1995] residues of a binding motif (cf. Fig. 4b) that is involved in recognition of many different binding partners [Hicke et al., 2005, Kiel and Serrano, 2006]. Recently, the first crystal structure with a new recognition motif centered at hot spot D58 (cross in Fig. 4c) has been found [Penengo et al., 2006]. Our results show that in solution this residue is as rigid as I44/F45 and H68.

At first sight, the observed fluctuations appear incompatible with the proposed conformational selection scenario. In particular, it seems combinatorially highly unlikely to find all involved residues simultaneously in the proper configuration required for binding, thus imposing a high entropic barrier. Only concerted fluctuations, implying reduced entropic cost, would explain the observed high physiological on-rates and affinities [Kiel and Serrano, 2006].

7.4 Collective molecular recognition dynamics

To check whether such concerted fluctuations are actually observed in the ubiquitin ensemble, we have carried out a principal component analysis. The conformational changes observed in X-ray structures are well described within the first five principal components. Although the number of degrees of freedom is reduced from 1839 to only five, all X-ray structures can be described up to a backbone RMS deviation (RMSD) of $0.45 \pm 0.04\,\text{Å}$. From linear combinations of these five principal components we found a single collective mode that corresponds to a pincer-like motion of predominantly those residues that are frequently involved in interfaces, and accounts for 25% (RMSD) of all backbone fluctuations in the solution ensemble (cf. Fig. 5b).

A stringent test whether this mode indeed describes the molecular recognition dynamics is to predict the bound ubiquitin conformations using only information from the binding partner. To this end we systematically varied the ubiquitin structure along this mode for each of altogether 41 interfaces until the highest number of contacting interface atoms, i.e., atoms within $3\,\text{Å}$ to $8\,\text{Å}$ of the binding partner, was reached. A correlation of 0.94 between the projection of the thus predicted and the actual X-ray structure was found for the pincer-like mode (cf. Fig. 5a). Analogously, correlations of 0.90 and 0.84 were obtained for the linearly combined first three principal components and the third principal component, respectively. These consistently high correlations for collective modes indicate that the interface adaptation dynamics of ubiquitin is indeed well described within a few collective degrees of freedom that dominate the solution ensemble. Moreover,
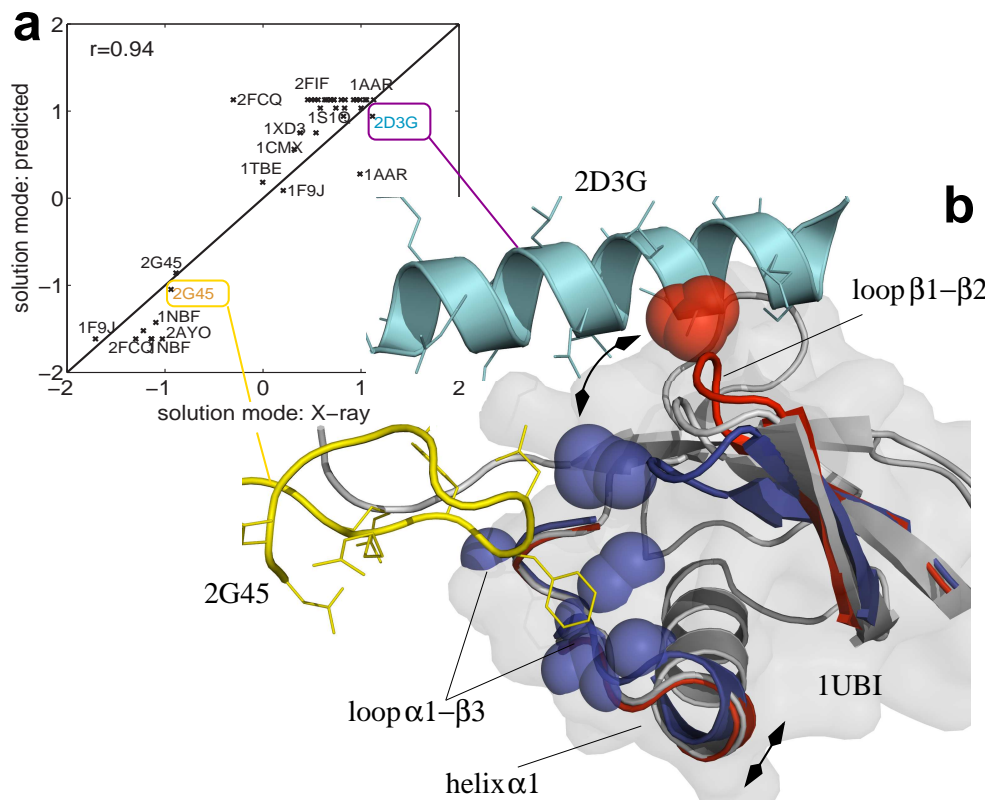
Figure 7.5: Equilibrium supra-τ_c dynamics is dominated by conformational selection dynamics. A large amplitude collective solution mode entails a pincer-like motion of loop $\beta 1$-$\beta 2$ and loop $\alpha 1$-$\beta 3$ including the C-terminal tip of helix $\alpha 1$. For each of altogether 41 binding partners this collective solution mode was systematically varied to find a predicted position that maximized contacts. **(a)** The position on the mode of the thus predicted selected structures are plotted on the y-axis, whereas the projected position onto this mode for the actual crystal structures are plotted on the x-axis. **(b)** To illustrate the conformational selection along the collective solution mode, two of the selected snapshots (dark blue and red) are shown together with relevant parts of their respective binding partners, the Zinc finger ubiquitin binding domain of isopeptidase T (2G45; yellow) and hepatocyte growth factor-regulated tyrosine kinase substrate (HRS) (2D3G; cyan). Contacts affected by the motion along the collective mode are shown as spheres. The crystal structure of apo-ubiquitin (1UBI) is shown at relevant regions as gray cartoon. The full protein is shown as half transparent surface.

this analysis indicates that the ability to optimize contacts with binding partners via backbone interface adaptation is important for ubiquitin to reach sufficient

affinity with many different binding partners. As illustrated in Fig. 5b, for the ubiquitin interfaces with hepatocyte growth factor-regulated tyrosine kinase substrate (HRS) and the Zinc finger ubiquitin binding domain of isopeptidase T (Protein Data Bank (PDB) codes 2D3G and 2G45), the collective solution mode allows molecular recognition by enabling ubiquitin to adapt to different protein interfaces.

The slow supra-τ_c time scale of ubiquitin's interface adaptation dynamics is corroborated by the observation that collective solution modes obtained from the first five principal components of nano-second ensembles 1xqq and 2nr2 [Lindorff-Larsen et al., 2005, Richter et al., 2007] were less adept in describing the interface adaptation. For these modes the correlation between predicted and crystallized position dropped from 0.94 to 0.68 and 0.55, respectively. The supra-τ_c time scale has previously been speculated to be important in the context of signal propagation of the Ig-binding domain of protein G [Bouvignies et al., 2005b] as well as for aggregation dynamics [Bertoncini et al., 2005].

7.5 Summary

Taken together, we have determined a solution ensemble of a globular protein from experimental data that comprises all solution dynamics up to the microsecond time scale at atomic resolution. A large part of this solution dynamics is concentrated in a collective pincer-like motional mode that strongly contributes to the interface adaptation dynamics during molecular recognition events. All available crystallographic structures of ubiquitin complexed to different binding proteins were shown to be accessible in solution. Conformational selection rather than induced fit is thus the main contributor to the observed interface adaptations. The observed conformational selection dynamics lowers entropic barriers, thereby explaining physiologically observed high affinity and fast on-rates which otherwise would need to be explained by induced fit motions.

These findings suggest how ubiquitin recognizes many different partner proteins with a high degree of specificity and sufficient affinity. To reach sufficient affinity a certain degree of structural plasticity is required that is thermally accessible in solution. To nevertheless maintain high specificity, the binding interfaces are centered around the rigid hot spot [Clackson and Wells, 1995] residues H68/I44 and D58. The rigidity of these mutational hot spots [Hicke et al., 2005, Penengo et al., 2006, Kiel and Serrano, 2006] might prevent promiscuous binding because only precisely aligned partner interfaces benefit from the high hot spot energy contribution. Structurally, the observed rigidity is maintained for H68

by packing with its neighbors L67 and L69 tightly into the protein core, whose rigidity is reinforced by helix 1. Similarly, I44 is anchored via F45 and decoupled from the adjacent flexible loop via an alanine-glycine linker (A46/G47). At D58 packing of L55 and a long-range hydrogen bond from Y59 to E51 provides stability. Since the solution dynamics are dominated by the collective pincer-like interface adaptation, it seems that only functionally essential flexibility is present. Apparently, ubiquitin has evolved to be as rigid as possible while remaining as flexible as necessary to engage in different interfaces.

Our finding that conformational selection is responsible for protein/protein binding of ubiquitin is in line with recent findings of conformational selection occurring for antibodies and enzymes [Eisenmesser et al., 2005, Tang et al., 2007, James et al., 2003]. For the latter, relaxation dispersion experiments that are sensitive to micro- to millisecond time scales, i.e., a 1000 times slower than the processes we described here, show conformational selection for all steps in enzymatic reactions of Dihydrofolate reductase [Boehr et al., 2006]. It should be noted that our findings differ from the stepwise model proposed for the binding of unfolded proteins to folded ones [Sugase et al., 2007], and thus open up a whole range of possible molecular recognition mechanisms.

Conformational sampling has very recently been observed for the transactivation response element (TAR) RNA from the human immunodeficiency virus type-1 (HIV-1) when binding to different ligands [Zhang et al., 2007]. The investigated TAR RNA fragment consisted of two RNA helices that were linked by a tri-nucleotide buldge. Dynamics up to milliseconds could be measured as motion of one helix relative to the other by using multiple sets of RDCs and a helix-elongation strategy [Zhang et al., 2006]. By separately elongating one of the two helices, the elongated helix dominated the overall alignment with the elongated axis being on average oriented parallel to the magnetic field. Thus, the frame of reference was anchored on the elongated helix and information on motions of the non-elongated helix relative to the elongated one could be extracted from the dynamically averaged RDCs. The RDCs, together with a idealized A-form helix geometry [Musselman et al., 2006], were used to determine five order-tensor elements describing ordering of the non-elongated helix relative to the elongated helix (see also [Tolman, 2001]). Using this approach, large motional amplitudes could be observed for TAR RNA which reported a high degree of motional asymmetry and a preference for inter-helical bending through rotations about a principal direction that is nearly orthogonal to both helices. To visualize the spatially non-random inter-helical motions, the RDCs were fitted with an ensemble with up to three (N=3) equally populated inter-helical conformers twisting in a clockwise manner. Remarkably, it was found that seven conformations of TAR when bound to different ligands fall along various positions of the dynamical tra-

jectory connecting the three different structures of the (N=3) ensemble of the free TAR RNA [Zhang et al., 2007]. Thus, also for TAR RNA the free form samples different conformations of TAR when bound to various ligands, supporting the idea of conformational sampling.

Recently, Blackledge and coworkers have used Accelerated Molecular Dynamics (AMD) to explore multiple time scale motions up to milliseconds in protein G by a more complete sampling of the conformational subspace of the protein energy landscape [Markwick et al., 2007]. A key aspect of the AMD simulation is to enhance the escape rate from one low-energy conformational state to another. This is achieved by lowering the free energy barriers separating the low-energy states on the rugged potential energy landscape, thus accelerating the occurrence of slower dynamic events by application of a biasing potential. The average backbone coordinates of the AMD simulation compare very closely with those of the high resolution 1.1 Å X-ray structure 1igd [Derrick and Wigley, 1994], indicating that although the accelerated MD trajectories sample broader conformational space, they appear to be distributed about a mean conformation that resembles the time- and ensemble-averaged experimentally determined structures. Markwick et al. could show that the RDC-derived order parameters are well reproduced by the AMD simulation [Markwick et al., 2007]. Interestingly, Lipari-Szabo S_{LS}^2 order parameter, calculated from 60 separate 2-ns MD simulations, using starting coordinates extracted from different conformational states from the AMD simulation and subsequently averaged, compared better to the experimental values than those derived from 60 classical MD simulations. Thus the AMD simulation seems to sample the phase space more accurately. The collective nature of the slower motions present in GB3 has been analyzed using quasi-harmonic principal component analysis of the AMD ensemble. Interestingly, the principal mode concerns a correlated motion relating the C-terminal region of the α-helix continuing into the loop III 41-45 and loop I between strands $\beta1$ and $\beta2$. This mode also describes a collective motion in the β-sheet, the outer strand $\beta2$ of which constitutes the active site of the molecule. Considering the structural similarity between protein G and ubiquitin, the concordance of results of the PCA analyses of the EROS ensemble of ubiquitin on the one side and the AMD ensemble of protein G on the other side is striking.

Chapter 8

Summary

Development of RDC-based NMR methods to access a hidden time window in protein dynamics

The focus of this work has been the investigation of protein dynamics in a time window between the correlation time τ_c and about 50 μs (supra-τ_c time window). This time window is inaccessible by relaxation methods (see chapter 3), however contains functionally important dynamics which, until now, remained undetected. With the use of Residual Dipolar Couplings (RDCs) this previously hidden time window has been made accessible. The enhancement of the experimental NH RDC data basis, together with the improvement and new development of mathematical methods to extract dynamic information from RDCs, has led to an improved accuracy of RDC-derived order parameters which make a quantitative interpretation feasible [Lakomek et al., 2005, Lakomek et al., 2006] [Lakomek, Walter 2008]. New dynamic modes of the protein ubiquitin in the supra-τ_c time window were observed that help elucidate the process of molecular recognition in ubiquitin. According to our results, conformational sampling rather than induced fit motion seems to be the dominant mechanism for molecular recognition in ubiquitin [Lange, Lakomek et al., 2008].

Enhancement of the experimental NH RDC basis

This PhD work started with the RDC-based model-free approach, originally invented by Griesinger and co-workers (Meiler et al., 2001). The RDC-based model-free approach relies on the measurement of NH RDCs for five linearly independent alignment tensor orientations in at least five different media [Meiler et al., 2001, Peti et al., 2002]. Using a high-resolution structure to determine the alignment tensors, structural as well as dynamic information can be deduced. Experimentally, it was very difficult to match the five linearly independent alignment tensor

orientations by the measurement of only five different alignment media. The reason for this is that most alignment media show some linear dependence, and it is hard to cover all five dimensions equally. That is why, experimentally, a much larger number of RDC data sets was needed.

Thus, an important part of this work was the measurement of RDCs in new additional alignment conditions with highest accuracy to provide a significantly enhanced robust experimental basis for the model-free analysis and further work. Compared to the first analysis [Peti et al., 2002] with 11 alignment media, the number of data sets, together with data from the literature could be significantly enhanced to 27 NH RDC data sets [Lakomek et al., 2006] and later on to 36 NH RDC data sets (cf. chapter 6). In the latest data set collection, the original 11 data sets have been fully replaced by more accurate measurements. A Seconda analysis (cf. chapter 6) of the latest 36 NH RDC data set revealed a ratio of the 5th to 6th singular value of 5.66; by a selection of 23 datasets (D23M) out of the 36, the Seconda ratio could be further improved to 6.81. D36M and D23M are the broadest experimental collection of RDC data sets now available with highest accuracy. They are the experimental basis for the work published in [Lakomek, Walter et al. 2008] and [Lange, Lakomek et al. 2008] and provide the experimental basis for the future development of RDC-based methods to study protein dynamics by our and other groups.

Rigorous mathematical re-evaluation of the RDC-based model-free approach

Based on the enlarged RDC data set of 27 NH rdc data sets, the RDC-based model-free approach was rigorously re-evaluated mathematically and the algorithm further improved, for example, with respect to filtering out of experimental noise (cf. chapter 4,[Lakomek et al., 2006]). Using this new approach, RDC-based order parameters $S^2_{rdc}(NH)$ were derived with unprecedented accuracy, revealing new modes of motion in the supra-τ_c time window. Charged and polar residues show more mobile backbone amide groups than hydrophobic residues. In particular the biologically relevant arginines reveal strongly decreased $S^2_{rdc}(NH)$ order parameters. Besides the $S^2_{rdc}(NH)$, effective orientations of the NH internuclear vectors described by the polar angles (θ_{av}, ϕ_{av}) and the amplitudes η_{rdc} and directions of anisotropic motions ϕ'_{rdc} were determined. For the α-helix, the determined ϕ'_{rdc} support an excursion of the α-helix as suggested earlier [Meiler et al., 2003] and recently observed by high pressure NMR [Kitahara et al., 2005].

Correlation of side-chain orientation and backbone mobility

According to the RDC-based model-free analysis, and to our surprise, a correlation between side-chain orientation and backbone mobility could be observed for

ubiquitin (cf. chapter 5 , [Lakomek et al., 2005]). Residues with solvent-exposed side chains showed a tendency to be more mobile in the protein backbone, whereas those with side chains pointing towards the hydrophobic core appeared more rigid in the protein backbone. This finding was manifested as an alternating pattern of $S^2_{rdc}(NH)$ order parameters in the β-strands and challenges the current picture of a rigid protein backbone uncoupled to side-chain dynamics. However, a very similar finding supporting our results has been independently observed for protein G by Martin Blackledge and co-workers [Bouvignies et al., 2005b] and earlier for the relaxation time-scale by Art Palmer and coworkers in the case of Ribonuclease H [Mandel et al., 1995, Mandel et al., 1996].

A self-consistent RDC-based model-free approach has been developed that alleviates the influence of structural noise

To become independent from the accuracy of the structural model used for alignment tensor calculation, a Self-Consistent RDC-based Model-free analysis (SCRM) has been developed that delivers RDC-based order parameters independent of the details of the structure used for alignment tensor calculation, as well as the dynamic average orientation of the inter-nuclear vectors in the protein structure in a self-consistent manner (cf. chapter 6) . SCRM was applied on two NH RDC data set collections, D36M and D23M. For both NH RDC data collections, the new SCRM approach gives almost identical order parameters (correlation factor of ρ= 0.945). For D23M it was concluded that there are neither significant structural changes induced by the alignment media nor significant correlations of the vector fluctuations and the alignment tensors. The correlation between alignment tensor fluctuations and internal dynamics, which has been proposed as a possible source of error by [Louhivuori et al., 2006, Louhivuori et al., 2007], is therefore likely to be small. This finding agrees very well with theoretical predictions [Salvatella et al., 2008].

It has been shown in this work that the influence of structural noise on the alignment tensor determination and the resulting order parameter can be alleviated by the SCRM method. Synthetic structural noise has been added to the 1ubi X-ray structure and different structures of ubiquitin from several complexes have been used as input for the SCRM analysis. The resulting order parameters were found to agree within 0.01 irrespective of the starting structure. Thus, RDC-derived $S^2_{rdc}(NH)$ order parameters were determined with unprecedented accuracy independent of structural noise [Lakomek, Walter et. al 2008].

For ubiquitin, the SCRM analysis yields an average RDC-derived order parameter of the NH vectors $< S^2_{rdc}(NH) >= 0.72 \pm 0.02$, compared to $< S^2_{LS}(NH) >= 0.778 \pm 0.003$ for the Lipari-Szabo order parameters. Consistent with the first analysis described in chapter 4, the inclusion of the supra-τ_c window increases the

averaged amplitude of mobility observed in the sub-τ_c window by about 34%. The backbone of Lys48, whose side chain is known to be involved in the polyubiquination process that leads to protein degradation, is very mobile on the supra-τ_c time scale ($S^2_{rdc}(NH) = 0.59 \pm 0.03$), while it is inconspicuous ($S^2_{LS}(NH) = 0.82$) on the sub-τ_c as well as on μs - ms relaxation dispersion time scales. This finding motivated the investigation of the role of supra-τ_c motion for protein-protein recognition .

An RDC-refined structural ensemble of ubiquitin reveals that conformational sampling is the dominant mechanism for molecular recognition in ubiquitin

In a cooperative effort with Oliver Lange, Bert de Groot and Helmut Grubmueller (Department for Theoretical and Computational Biophysics, MPI for Biophysical Chemistry), a RDC-refined structural ensemble of ubiquitin has been derived (cf. chapter 7). In contrast to relaxation-based ensembles of solution structures that take only motions faster than τ_c into account [Lindorff-Larsen et al., 2005, Richter et al., 2007], the RDC-derived ensemble includes solution dynamics up to microseconds (Lange, Lakomek et al., 2008). The ensemble covers the complete structural heterogeneity observed in 46 ubiquitin crystal structures, most of which are in complexes with other proteins. It is therefore concluded that conformational selection, rather than induced fit motion suffices to explain the molecular recognition dynamics of ubiquitin. Remarkable correlations are seen between the flexibility of the ensemble and contacts formed in ubiquitin complexes. A large part of the solution dynamics is concentrated in one concerted mode which accounts for most of ubiquitin's molecular recognition heterogeneity, and ensures a low entropic cost for complex formation. The finding that conformational selection is responsible for protein/protein binding of ubiquitin agrees with recent findings of conformational selection occurring for antibodies and enzymes [Eisenmesser et al., 2005, Tang et al., 2007, James et al., 2003]. For the latter, relaxation dispersion experiments that are sensitive to micro- to millisecond time scales, i.e. 1000 times slower than the processes seen in our studies, show conformational selection for all steps in enzymatic reactions of dihydrofolate reductase [Boehr et al., 2006]. According to our results, conformational selection seems to play a dominant role not only for enzymatic reactions but for molecular recognition as well, on the much faster supra-τ_c time scale which has been previously inaccessible by other methods.

To summarize, RDCs can provide additional information about protein dynamics (supra-τ_c motion), complementary to relaxation methods. Since promising techniques are emerging to reduce the experimental effort of collecting enough linearly independent RDC data sets [Ruan and Tolman, 2005, Yao and Bax, 2007], RDCs are expected to become a routine tool to complement the analysis of biomolecular

dynamics. These dynamic measurements can contribute to a better understanding of enzymatic reactions and recognition events in proteins.

Chapter 9

Outlook

Residual dipolar couplings are measured both for further backbone inter-nuclear vectors, such as $D(H_N C')$ and $D(NC')$, as well as for side-chain methyl groups. Applying the SCRM method on these RDC data, will provide a more precise picture of the backbone peptide plane dynamics and the nature of side-chain dynamics, especially with respect to the proposed coupling of side-chain and backbone dynamics. While the amplitudes of motion of the backbone amide NH groups have been determined precisely in this work and the dynamics have been attributed to the supra-τ_c time window (between 4 ns and 50 μs for ubiquitin at room temperature), the information about time scale cannot be narrowed down further with NMR measurements. Therefore, dielectric spectroscopy measurements are being conducted for ubiquitin in cooperation with PD Dr. Lunkenheimer (Institute for Experimental Physics V, University of Augsburg). Initial results indicate a slow relaxation process at 110 ns, attributed to internal dynamics. In order to test whether the novel results observed for ubiquitin might be of general nature, we are going to measure RDCs and perform the SCRM analysis on another protein. Moreover, we are going to explore the recognition dynamics of ubiquitin further by performing NMR kinetic binding studies on ubiquitin with its recognition binding partners. According to the literature, the affinities lie in the range of 10 μM to several 100 μM and on- and off-rates have not been published yet. Deviations from a diffusion controlled assembly are expected if molecular recognition is indeed governed by conformational sampling.

The research performed in this PhD thesis was dedicated towards more insight into internal protein dynamics, studied on the protein ubiquitin as a model system. We consider a better knowledge and understanding of protein dynamics essential to understand their function and interactions of these macromolecules. Besides making a new time window, the supra-τ_c time window, accessible for the

study of protein dynamics, the developed methods, based on RDCs, have the potential to explore protein dynamics of systems, e.g. membrane proteins, where relaxation measurements are hampered by the size of the system but RDCs can be measured.

Appendix

Appendix A

Appendix SCRM

Table 1: The inter-nuclear angles κ_j enclosed between the dynamic average NH vector orientations and the NH vectors of the starting X-ray structure 1ubi are shown in column 4 (ubi/ubi) for D23M. In column 5 (ubi/1d3z) the same dynamic average NH vector orientations are compared to the 1d3z NMR structure. For comparison, the SCRM analysis was repeated using the 1d3z NMR structure as starting structure. The resulting $S_{rdc}^2(NH)$ and κ_j are listed in column 3 and 6. In column 2 the derived $S_{rdc}^2(NH)$ using the 1ubi X-ray structure as starting structure are shown again. Since the C-terminal part of ubiquitin (residues 72-76) are highly mobile, these residues have not been considered for the determination of inter-nuclear angles κ_j.

AA #	S_{rdc}^2 (1ubi)	S_{rdc}^2 (1d3z)	κ ubi/ubi	κ ubi/1d3z	κ 1d3z/1d3z
1	n	n	n	n	n
2	0.85	0.86	4.2	5	5.1
3	0.68	0.68	7.6	1.8	2.9
4	0.79	0.81	4.6	10.8	10.1
5	0.75	0.76	3.8	2.2	2.5
6	0.78	0.8	10.4	2.6	2.8
7	0.64	0.62	6.3	8.8	8.8
8	0.67	0.64	5.7	8.7	8.1
9	n	n	n	n	n
10	n	n	n	n	n
11	0.45	0.46	17.2	11.1	12
12	0.64	0.63	9.2	3.8	3.2

13	0.68	0.68	12.8	5.2	5.7
14	0.78	0.8	9.6	3.2	3.1
15	0.72	0.73	7.3	2	0.9
16	n	n	n	n	n
17	0.86	0.85	7	3.1	3.7
18	0.77	0.79	9.5	0.7	1.5
19	n	n	n	n	n
20	0.55	0.55	12.4	7.1	6.5
21	0.86	0.83	3.2	2.8	3
22	n	n	n	n	n
23	0.8	0.81	6	4.6	4.6
24	n	n	n	n	n
25	0.83	0.84	3.9	3	3
26	n	n	n	n	n
27	0.82	0.82	3.7	2.8	2.2
28	0.85	0.85	1.8	3.2	3.2
29	0.73	0.74	2	4.7	4.9
30	0.77	0.78	0.8	0.5	0.3
31	n	n	n	n	n
32	0.84	0.84	7.3	4	4.5
33	0.76	0.77	11.4	1.2	1.2
34	0.73	0.75	15.8	2.7	3
35	0.82	0.81	6.9	5.1	6.1
36	0.76	0.77	9.7	8.7	8.4
37	n	n	n	n	n
38	n	n	n	n	n
39	n	n	n	n	n
40	0.67	0.67	2.4	4.8	5
41	0.8	0.78	6.5	5.7	5.1
42	0.69	0.7	12	7.3	8.3
43	0.77	0.76	10.7	3.6	3.6
44	0.75	0.76	8.1	4.6	5.4
45	0.8	0.78	5.8	1.2	1.7
46	n	n	n	n	n

47	0.77	0.77	2.4	13	12.8
48	0.59	0.59	11.8	3.1	3.9
49	0.72	0.71	3.4	2.4	2.5
50	0.59	0.6	10.2	3.9	3.9
51	0.74	0.73	3.8	3	3.3
52	0.69	0.68	15.5	21.2	20.1
53	n	n	n	n	n
54	0.69	0.7	20.3	2.4	2.8
55	0.77	0.79	0.6	2	2.3
56	0.72	0.72	5.9	3.1	3.5
57	n	n	n	n	n
58	0.87	0.86	1.1	1.9	2.3
59	0.81	0.82	2.5	2.3	3
60	0.8	0.78	4.4	8.7	8.6
61	0.9	0.85	7.3	3	3.3
62	0.55	0.55	3.9	4.1	3
63	0.79	0.77	0.5	2.9	2.5
64	0.88	0.86	3.1	3.7	3.6
65	0.66	0.65	2.5	6.4	6.6
66	0.87	0.88	4.2	1.7	0.6
67	0.82	0.83	10.2	3.8	4.2
68	0.8	0.81	2.9	4.2	4.2
69	n	n	n	n	n
70	0.7	0.72	11.9	6.3	6.1
71	0.57	0.57	10.1	7.7	8.1
72	n	n	n	n	n

Table 2: Same as Table 1 but for D36M.

AA #	S^2_{rdc} (1ubi)	S^2_{rdc} (1d3z)	κ ubi/ubi	κ ubi/1d3z	κ 1d3z/1d3z
1	n	n	n	n	n
2	0.89	0.89	3.7	5.9	5.9
3	0.75	0.76	7.9	2.2	3.5
4	0.73	0.71	5.5	12.5	11.4
5	0.76	0.78	3.1	1.6	1.6
6	0.79	0.8	10.8	2.5	2.9
7	0.65	0.64	6.6	8.6	8.3
8	0.67	0.64	7.7	11.3	10.8
9	n	n	n	n	n
10	n	n	n	n	n
11	0.45	0.45	16.4	10.6	10.9
12	0.7	0.69	9.4	3.7	3.8
13	0.68	0.69	12	4.5	4.9
14	0.78	0.8	9.5	2.5	2.3
15	0.74	0.74	7.4	1.8	0.7
16	0.77	0.77	2	2.6	3.5
17	0.83	0.81	8.3	3.4	3.9
18	0.8	0.8	9.4	1.6	1.7
19	n	n	n	n	n
20	0.59	0.59	11.3	5.6	5
21	0.78	0.77	2.1	2.5	2.7
22	n	n	n	n	n
23	0.83	0.84	6.3	4.9	4.8
24	n	n	n	n	n
25	0.86	0.87	2	2.3	2.3
26	0.78	0.79	3	3.6	n
27	0.82	0.82	2.5	1.4	1
28	0.82	0.83	2.3	2.5	2.8
29	0.78	0.79	1.5	2.8	3
30	0.79	0.79	1.5	1.1	1.4
31	n	n	n	n	n
32	0.85	0.85	6.6	2.5	2.9
33	0.75	0.77	10.9	1.4	1.2

34	0.75	0.76	16.5	3.5	3.7
35	0.64	0.65	11.4	10.1	10.8
36	0.77	0.78	7.9	7.1	7.1
37	n	n	n	n	n
38	n	n	n	n	n
39	0.68	0.69	2.5	3	3.8
40	0.71	0.71	2.3	3.2	3.8
41	0.77	0.76	7.5	4.9	4.8
42	0.74	0.75	12	7.4	8.7
43	0.75	0.75	12	4.9	4.8
44	0.75	0.75	8.9	5.4	6.1
45	0.78	0.77	6.6	1.4	2.1
46	n	n	n	n	n
47	0.79	0.79	2.4	13.1	13.2
48	0.58	0.58	13.6	4.9	5.2
49	0.77	0.77	4.1	1	1.5
50	0.63	0.61	7.7	1.4	1.3
51	0.84	0.84	3.7	4.1	4.3
52	0.62	0.62	11.1	17	16.2
53	n	n	n	n	n
54	0.64	0.64	19.9	2.6	3.4
55	0.77	0.78	0.4	1.8	2.1
56	0.76	0.76	7.2	4.2	4.5
57	0.86	0.85	11.5	6.1	7.5
58	0.84	0.84	1.4	1.7	1.9
59	0.83	0.84	3.3	2.3	3.3
60	0.8	0.8	5.6	8.1	8.4
61	0.83	0.8	6.3	1.4	2
62	0.57	0.57	4.6	5.1	4.4
63	0.68	0.68	1.1	3.9	4.2
64	0.79	0.77	2.7	3.1	2.7
65	0.63	0.64	4.1	5.6	5.2
66	0.81	0.82	4.5	1.9	0.5
67	0.83	0.85	9.3	3.5	4.3

68	0.83	0.83	3.1	3.4	3.1
69	n	n	n	n	n
70	0.72	0.74	11.5	6.5	5.7
71	0.67	0.67	7.8	10	9.7
72	n	n	n	n	n

Table 3: NH RDCs in different alignment media

The following RDCs (in Hz) have been newly measured for the different alignment conditions A1-A18. Alignment conditions A19-A36 have been taken from the literature as indicated (cf. Table 4). Please note: Data sets A8-A13 have been measured by Korvin Walter during his diploma work [Walter, 2007] For convenience, all NH RDC data sets used for the SCRM analysis are provided in Table 3, including data sets taken from the Supporting Information by Tolman and coworkers, with permission by J. Tolman. Please note the references to the original literature.

The NH rdc data sets A19 to A29 are those provided in the supplement to (Brigmann and Tolman, 2003). Data sets A30 to A36 have been taken from the Supporting Information to (Ruan and Tolman, 2005). Data sets taken from the literature are referenced in the same way as in the original literature. We refer to the original literature for further information.

NH RDCs [Hz] A1-A8:

	A1	A2	A3	A4	A5	A6	A7	A8
source	SCRM	SCRM	SCRM	SCRM	SCRM	SCRM	SCRM	SCRM
Identifier	A1	A2	A3	A4	A5	A6	A7	A8
D23M		x		x				
1	n	n	n	n	n	n	n	n
2	-1.81	-3.22	-20.99	14.72	10.46	8.41	13.66	4.53
3	2.25	5.07	1.11	-5.68	-18.17	-14.13	-10.97	-3.32
4	2.13	3.67	3.16	-4.61	-12.9	-10.2	-9.2	-2.59
5	-0.89	-0.94	2.62	-2.28	-3.99	-4.13	-6.9	-3.46
6	-1.91	-2.77	-12.16	-3.23	12.32	9.76	-6.72	-0.36
7	-2.01	-3.18	-12.17	7.1	22.04	17.48	1.33	-0.46
8	n	n	-3.75	-4.11	6.69	6.84	-11.58	-6.22
9	n	n	n	n	n	n	n	n
10	n	n	n	n	n	n	n	n
11	-3.47	-6.76	-17.91	16.49	16.71	14.21	14.45	4.12
12	-1.5	n	-12.22	1.48	21.37	17.65	-5.69	-1.79
13	-1.93	-2.78	n	2.18	0.12	-0.58	-1.72	-2.21
14	-0.81	-0.76	-5.42	-4.71	-1.46	-1.54	-8.56	-1.64
15	3.19	6.98	9.9	-7.41	n	n	-12.71	-4.98
16	0.57	1.75	-10.35	1.75	-6.23	-4.51	-1.42	0.58

17	2.63	5.74	-10.4	4.71	n	3.17	-1.3	-1.31
18	-2.68	-5.65	-20.85	19.34	16.41	13.47	18.9	5.65
19	n	n	n	n	n	n	n	n
20	-3.22	-6.87	9.42	11.04	n	n	11.52	n
21	2.55	5.98	n	3.52	18.16	16.26	-3.6	-2.74
22	n	n	n	n	n	n	n	-7.86
23	-1.78	-2.78	-11.93	-3.28	n	n	-6.98	-0.41
24	n	n	n	n	n	n	n	n
25	-2.47	-4.9	-18.62	-3.12	10.56	8.41	-1.11	4.3
26	-2.03	-2.99	-10.98	-2.57	10.65	7.89	-5.92	-0.75
27	-1.18	-1.75	-16.58	-4.8	n	20.12	-7.35	0.36
28	-2.72	-5.33	n	-4.39	16.18	14.77	-3.2	4.34
29	-1.64	-2.67	-11.78	-4.37	4.4	3.25	-5.53	1.14
30	-1.92	-2.79	-14.93	-2.79	19.07	15.99	-6.54	-0.16
31	-1.13	-2.32	-12.75	-7.25	n	18.55	-6.57	1.92
32	-2.61	-5.5	-19.19	-3.66	n	12.98	-0.64	5.48
33	-1.52	-2.79	-9.76	-3.65	6.74	5.09	-6.15	0.08
34	-1.25	-2.08	-16.13	-6.31	23.99	18.84	-6.97	1.57
35	-2.51	-7.13	8.08	1.35	-3.94	-4.21	13.73	5.09
36	-4.01	-8.63	-22.13	18.96	14.24	10.53	22.26	6.59
37	n	n	n	n	n	n	n	n
38	n	n	n	n	n	n	n	n
39	3.78	7.51	16.55	-10.06	-27.45	-22.09	-15.74	-6.54
40	-0.54	-1.56	24.71	1.59	-23.04	-19.67	2.07	-3.38
41	7.26	14.3	23.3	-18.55	-16.55	-11.88	-20.24	-7
42	2	3.77	-8.39	-10.76	n	18.88	-11.49	-0.44
43	2.43	5.23	-12.07	-8.19	22.94	n	-13.17	-2.47
44	-1.47	-1.98	-17.74	-3.43	n	n	-6.79	0.21
45	-2.4	-3.53	-14.55	6.71	24.82	20.3	1.21	-0.51
46	n	n	n	n	n	n	n	n
47	-1.83	-2.96	-11.26	-3.52	n	n	-7.15	n
48	-5.15	-10.39	-5.53	19.21	8.91	5.59	22.38	5.54
49	-2.36	-4.1	6.65	5.36	-2.37	-3.38	3.08	-2.99
50	0.24	-0.58	n	-4.82	n	n	-5.45	3.35

51	-3.5	-7.08	-21.66	7.6	5.29	3.86	10.97	7.13
52	-5.56	-12.26	-4.39	15.47	n	-0.2	24.86	7.78
53	n	n	n	n	n	n	n	n
54	-1.22	-4.31	0.9	-4.29	6.52	5.37	4.1	5.92
55	-1.72	-3.08	-13.2	-4.59	8.05	5.68	-5.64	1.51
56	4.72	8.68	n	-10.68	n	-30.37	-14.7	-6.92
57	6.61	13.21	33.68	-15.27	-39.89	-32.08	-20.14	-8.8
58	3.23	6.65	3.99	-5.97	-19.49	-15.67	-11.02	-4.39
59	1.32	3.11	18.26	-6.63	-24.78	-21.47	-10.57	-6.45
60	6.79	13.7	20.63	-18.09	-11.07	-7.34	-20.16	-6.09
61	6.24	12.98	2.02	-8.01	2.35	4.16	-15.78	-6.12
62	n	7.78	29.6	-14.32	-28.53	n	-11.07	-5.51
63	-3.53	-8.73	-9.76	2.42	6.92	4.91	13.06	7.97
64	6.39	13.42	21.55	-12.09	-30.39	-22.99	-18.96	-7.8
65	3.84	10.02	23.22	-17.5	-15.2	-11.38	-15.67	-5.31
66	3.12	6.91	14.14	-8.69	-26.04	-20.96	-14.92	-5.95
67	-0.28	0.36	4.63	-3.31	-14.65	-12.85	-7.64	-5.29
68	-1.58	-2.37	-17.91	-3.29	23.62	18.08	-7.53	-0.4
69	1.69	n	-15.96	-8.04	n	n	-11.95	-5.83
70	2.69	4.87	-1.26	-13.37	14.68	12.02	-12.89	-1.1
71	5.97	10.93	13.76	-16.17	-2.38	-0.87	-17.62	-5.35
72	3.14	5.91	n	-9.56	-30.26	-24.97	-7.62	-4.92
73	0.72	2.24	15.22	-1.78	n	n	-7.77	-1.48
74	n	n	3.37	-5.71	-9.35	-7.36	-8.14	-3.68
75	n	n	n	n	n	n	n	n
76	0.26	0.51	1.6	-1.59	-1.86	-1.39	-2.21	0.22

NH RDCs A9-A16:

	A9	A10	A11	A12	A13	A14	A15	A16
Source	SCRM	SCRM	SCRM	SCRM	SCRM	L2006	L2006	L2006
Identifier	A9	A10	A11	A12	A13	E1	E2	E3
D23M	x	x			x	x	x	x
1	n	n	n	n	n	n	n	n
2	11.16	7.87	14.76	17.41	17.6	15.74	2.61	6.74
3	-13.39	-7.56	-13.68	-12.15	-12.55	-11.03	-3.79	-2.78
4	-14.09	-8.34	-17.43	-11.25	-10.35	-10.74	-0.05	-2.76
5	-18.23	-8.1	-16.44	-11	-7.39	-7.8	3.46	-1.75
6	-12.74	-7.57	-14.68	-9.48	-5.02	n	4.3	-1.66
7	-13.9	-2.76	-4.87	-1.75	2.66	3.44	6.78	2.74
8	-26.33	-6.24	-13.84	-12.4	-17.16	n	n	n
9	n	n	n	n	n	n	n	n
10	n	n	n	n	n	n	n	n
11	7.69	7.58	16.24	14.35	19.19	16.9	6.04	7.47
12	-17.41	-6.26	n	-7.45	-4.25	n	n	n
13	-14.33	-5.37	-10.95	-6.08	-1.49	-3.59	5.07	0.31
14	-12.51	-8.2	-15.99	-11.49	-7.75	-7.55	1.94	-2.54
15	-17.59	-8.75	-17.91	-15.28	-16.32	-14.52	-4.97	-3.64
16	-1.09	-1.4	-2.42	-0.61	0.48	-0.09	-1.45	0.76
17	-8.94	-0.16	-0.38	-0.23	-2.64	1.74	-3.13	2.19
18	12.84	10.08	16.51	20.94	23.35	19.92	5.4	8.64
19	n	n	n	n	n	n	n	n
20	n	n	n	n	n	n	n	n
21	n	-2.32	-3.9	-3.71	-6.88	1.41	-1.71	1.72
22	-25.5	-10.87	-20.94	-21.77	-27.32	n	n	n
23	n	-7.7	-18.3	-9.07	-5.64	-3.92	4.72	-1.94
24	n	n	n	n	n	n	n	n
25	6.56	-2.24	-5.57	-2.9	2.72	1.3	3.29	-1.39
26	-13.73	-7.54	-13.32	-9.16	-4.93	n	n	n
27	-11.08	-7.02	-14.14	-9.18	-6.56	-2.33	4.08	-2.35
28	12.12	-1.47	-3.02	0.09	1.76	2.12	2.73	-1.86
29	-5.1	-6.6	-12.58	-6.78	-3.49	n	n	n
30	-13.58	-7.43	-14.05	-8.67	-5.38	-2.67	4.78	-1.6

31	n	-4.43	-7.66	-5.64	n	n	n	n
32	11.16	-1.09	-1.95	0.03	3.71	3.23	3.27	-1.29
33	-9.02	-7.18	-14.48	-7.98	-4.26	-4.13	3.7	-1.87
34	-6.08	-5.88	-11.01	-7.39	-6.12	n	n	n
35	n	11.13	23.8	18.51	16.77	9.93	0.83	1.04
36	24.65	13.24	22.17	21.07	28.89	22.3	5.44	8.57
37	n	n	n	n	n	n	n	n
38	n	n	n	n	n	n	n	n
39	-20.05	-10.23	-19.32	-20.36	n	-17.22	-6.2	-4.55
40	-5.88	-0.88	-1.43	-1.95	0.83	-5.21	2.17	0.01
41	-19.92	-8.77	-18.92	-20.6	-30.17	-19.89	-11.62	-8.06
42	n	-5.35	-9.81	-9.94	-13.59	-5.72	-1.85	-4.7
43	-20.76	-7.89	-11.09	-13.24	-16.21	-6	-1.74	-3.42
44	-14.02	-7.02	-11.95	-8.72	-6.03	-2.16	4.27	-1.78
45	-16.37	-3.11	-4.72	-2.66	2.3	3.58	6.86	2.37
46	n	n	n	n	n	n	n	n
47	n	n	n	n	n	n	n	n
48	15.7	10.99	21.97	21.56	26.82	17.87	8.95	8.05
49	-11.67	-2.22	-4.88	-1.77	3.23	-1.18	5.68	1.55
50	-3.9	-4.26	-8.56	-4.8	-3.49	-0.54	3.45	-2
51	21.33	6.49	14.17	14.86	18.47	12.59	3.56	3.94
52	34.79	15.97	n	25.9	32.72	20.72	6.39	7.36
53	n	n	n	n	n	n	n	n
54	25.73	4.46	9.43	7.2	7.47	5.45	0.3	-1.18
55	-3.46	-5.72	-11	-7.27	-2.89	-3.7	2.99	-2.3
56	-21.59	-10.67	-19.82	-18.88	-21.02	-20.23	-5.54	-5.22
57	-25.52	-10.72	-22.87	-24.15	-29.51	-24.21	-10.68	-6.93
58	-12.86	-6.85	-15.35	-9.78	-13.84	-11.74	-5.65	-3.16
59	-20.54	-9.29	-17.61	-15.29	-14.63	-14.87	-1.22	-3.56
60	-16.6	-7.82	-12.97	-15.86	-28.26	-17.93	-10.82	-7.81
61	-24.88	-7.24	-13.06	-14.47	-22.72	-11.06	-8.16	-3.41
62	-3.96	-3.3	-4.37	-8.87	-18.15	-14.82	-8.81	-5.93
63	34.75	9.83	18.13	16.54	18.87	12.84	2.93	1.8
64	-24.42	-10.08	-20.04	-20.26	-26.67	-20.25	-9.47	-5.54

65	-6.47	-5.24	-8.33	-12.32	-22.35	-15.77	-9.8	-7.14
66	-20.72	-10.22	-18.45	-18.75	-18.75	-16.64	-4.65	-4.4
67	-18.86	-7.76	-16.14	-12.9	-9.13	-10.78	1.99	-2.37
68	-14.75	-7.85	-15.51	-10.19	-5.69	-3.19	4.79	-1.67
69	-16.2	-7.22	-13.69	-11.67	-13.66	n	n	n
70	-5.96	-5.25	-7.92	-9.66	-16.05	-7.87	-4.05	-5.53
71	-15.48	-6.69	-12.51	-16.39	-24.68	-15.4	-8.36	-6.95
72	-3.75	-3.12	-5.44	-8.81	-13.59	-13.51	-5.23	-4.89
73	-19.87	-5.97	-11.41	-7.74	-8.97	-1.19	1.9	n
74	-7.27	-4.52	-10.12	-8.12	-7.95	n	n	n
75	n	n	n	n	n	n	n	n
76	0.88	-1.16	-2.06	-2.41	-2.33	-1.64	-0.96	-0.65

NH rdcs [Hz] A17-A22:

Pub	A17	A18	A19	A20	A21	A22
Source	L2006	L2006	O1998	O1998	B2003	B2003
Identify	E4	E5	Bic1	Bic2	PM1	PM2
D23M			x	x		X
1	n	n	n	n	n	n
2	-6.16	5.4	8.17	15.52	-4.73	-4.04
3	-1.48	-8.97	-8.27	-10.52	-0.95	-0.65
4	-1.63	-7.32	-10.49	-9.65	-1.61	-1.27
5	-2.07	-3.45	-9.87	-6.08	-1.85	-1.54
6	-6.16	n	-9.15	-3.85	-5.21	-4.28
7	-5.59	9.92	-3.7	3.4	-4.97	-3.94
8	n	n	-6.46	-15.05	0.24	0.46
9	n	n	n	n	n	n
10	n	n	n	n	n	n
11	-5.49	9.14	7.53	16.94	-4.51	-3.77
12	n	n	-7.38	-2.5	-4.56	-3.54
13	-2.67	-1.11	-6.95	-0.7	-2.42	-1.77
14	-4.54	-2.22	-9.71	-6.43	-3.96	-3.05
15	0.95	-11.82	-9.85	-14.2	1.08	0.81
16	-4.29	-3.05	n	n	n	n
17	-2.88	2.47	-0.04	-1.96	-2.16	-1.94
18	-6.12	8.71	10.51	20.11	-5.06	-4.11
19	n	n	n	n	n	n
20	n	n	4.07	12.09	0.69	0.74
21	-3.31	10.08	-2.12	-4.58	-2.48	-2.39
22	n	n	n	n	n	n
23	-6.05	4.78	-9.1	-3.83	-4.68	-3.95
24	n	n	n	n	n	n
25	-6.75	3.69	-2.95	2.55	-5.68	-4.53
26	n	n	n	n	n	n
27	-6	10.08	-8.22	-5.07	-4.91	-3.95
28	-6.18	8.65	-1.5	1.76	-5.18	-4.04
29	n	n	-7.18	-2.72	-4.91	-3.61
30	-6.69	7.75	-8.71	-3.62	-5.19	-4.3

31	n	n	n	n	n	n
32	-6.61	6.4	-1.26	3.47	-5.43	-4.43
33	-5.39	1.56	-8.26	-3.36	-4.77	-3.82
34	n	n	-6.53	-4.36	-5.13	-4.18
35	3.79	-1.83	12.5	13.49	2.65	2.19
36	-5.91	7.18	13.88	24.37	-4.55	-3.66
37	n	n	n	n	n	n
38	n	n	n	n	n	n
39	2.83	-12.94	n	n	n	n
40	5.04	-11.36	-1.94	-0.16	3.6	2.94
41	8.88	-6.15	-7.87	-26.4	7.08	5.75
42	-1.42	10.62	-5.72	-11.61	-1.28	-1.05
43	-2.21	12.31	-7.68	-13.33	-1.97	-1.67
44	-6.77	11.48	-7.98	-4.37	-5.55	-4.51
45	-6.4	11.19	-4.24	3	-5.08	-3.97
46	n	n	n	n	n	n
47	n	n	-9.05	-4.39	-4.45	-3.53
48	-2.37	4.17	10.83	22.97	-2.24	-1.55
49	-0.11	-2.09	-3.5	2.83	-0.68	-0.37
50	-5.65	11.9	-5.86	-3.92	-4.88	-3.95
51	-7.09	2.16	6.79	15.69	-5.09	-4.06
52	-0.32	1.47	16.98	27.26	-0.84	-0.76
53	n	n	n	n	n	n
54	1.11	2.95	6.3	5.53	0.53	0.6
55	-6.08	2	-6.76	-2.25	-4.81	-3.82
56	6.66	-17.44	-11.02	-19.04	5.25	3.95
57	9.67	-18.22	n	n	n	n
58	-0.61	-10.03	-7.86	-12.08	0.07	-0.28
59	2.5	-13.04	-10.4	-12.03	2.52	1.61
60	8.17	-3.77	-7.36	-25.03	6.38	5.13
61	2.17	2.92	-6.91	-19.07	1.77	1.18
62	10.44	-12.75	-2.51	-16.25	8.05	6.51
63	-1.53	2.99	11.05	15.43	-1.44	-1.09
64	5.26	-13.7	-10.72	-22.6	4.48	3.27

65	8.4	-5.75	-4.03	-19.75	6.08	5.07
66	2.06	-12.83	-11.32	-16.22	1.37	1.09
67	0.82	-8.28	-9.85	-8.13	0.24	0.31
68	-6.32	9.33	-8.74	-4.57	-5.39	-4.6
69	n	n	n	n	n	n
70	1.01	7.41	-4.55	-13.5	0.78	0.43
71	6.1	0.33	-6.23	-21.3	4.82	3.74
72	10.12	-13.75	n	n	n	n
73	-5.07	14.45	n	n	n	n
74	n	n	-5.46	-6.6	0.19	-0.3
75	n	n	n	n	n	n
76	0.16	-0.88	-1.12	-1.7	0.02	-0.02

NH RDCs [Hz] A23-A28

Pub	A23	A24	A25	A26	A27	A28
Source	B2003	B2003	B2003	B2003	B2003	B2003
Identify	PM3	PM4	PM5	PM6	Helfrich	Phage
D23M	x	x	x	x	x	x
1	n	n	n	n	n	n
2	-3.35	-2.85	-3.81	-3.47	4.05	1.58
3	-0.75	-0.44	-0.41	-0.54	-5.65	-1.52
4	-1.21	-0.95	-0.37	-1.21	-0.8	-0.33
5	-1.16	-0.91	-0.07	-1.44	4.68	1.17
6	-3.39	-2.87	-2.81	-3.93	6.78	3.42
7	-3.18	-2.71	-2.69	-3.71	9.7	4.38
8	-0.11	-0.02	-0.67	0.22	-7.58	0.27
9	n	n	n	n	n	n
10	n	n	n	n	n	n
11	-3.13	-2.67	-3.41	-3.43	9.15	2.7
12	-2.82	-2.61	-2.38	-3.19	7.34	4.06
13	-1.5	-1.3	-0.42	-1.98	7.15	1.76
14	-2.51	-2.16	-1.75	-3.07	3.09	1.66
15	0.62	0.47	0.94	0.75	-7.42	-2.5
16	n	n	n	n	n	n
17	-1.68	-1.49	-2.11	-1.62	-4.7	0.18
18	-3.27	-2.8	-3.65	-3.47	7.93	2.38
19	n	n	n	n	n	n
20	0.58	0.59	1.22	0.64	8.28	-0.55
21	-2.27	-1.88	-2.35	-2.18	-2.36	2.42
22	n	n	n	n	n	n
23	-3.13	-2.75	-2.32	-3.45	7.2	3.78
24	n	n	n	n	n	n
25	-3.8	-3.29	-3.61	-4.24	5.87	2.93
26	n	n	n	n	n	n
27	-3.39	-2.96	-2.95	-3.69	6.73	4.83
28	-3.48	-2.79	-3.32	-3.84	5	3.46
29	-3.24	-2.76	-2.52	-3.51	4.69	2.44
30	-3.58	-3.11	-2.88	-3.95	7.53	4.3

31	n	n	n	n	n	n
32	-3.59	-3.12	-3.67	-4.25	6.14	3.31
33	-3.21	-2.7	-2.31	-3.44	5.59	2.79
34	-3.33	-2.85	-3.04	-3.76	5.15	4.14
35	1.84	1.61	1.21	2.05	1.83	-2.16
36	-3.21	-2.69	-3.77	-3.43	8.52	1.76
37	n	n	n	n	n	n
38	n	n	n	n	n	n
39	n	n	n	n	n	n
40	2.56	2.21	3.48	2.75	1.76	-2.82
41	4.49	3.69	4.06	5.52	-17.3	-4.02
42	-1.03	-0.9	-1.64	-1.17	-2.24	2.67
43	-1.52	-1.32	-1.81	-1.57	-2.24	3.15
44	-3.76	-3.11	-3.38	-4.5	6.94	4.91
45	-3.13	-2.6	-2.6	-3.72	10.03	4.77
46	n	n	n	n	n	n
47	-2.85	-2.28	-2.24	-3.3	6.52	3.68
48	-1.21	-0.99	-1.06	-1.56	12.71	1.43
49	-0.26	-0.11	0.63	-0.33	7.64	0.65
50	-3.15	-2.73	-2.84	-3.73	5.27	4.35
51	-3.49	-3.09	-3.77	-3.92	6.05	1.44
52	-0.6	-0.76	-0.91	-0.69	9.71	-0.6
53	n	n	n	n	n	n
54	0.3	0.22	-0.07	0.41	0.98	-0.03
55	-3.21	-2.75	-2.63	-3.48	4.86	2.4
56	3.43	3.47	4.02	3.85	-9.93	-4.73
57	n	n	n	n	n	n
58	0.03	-0.22	-0.03	-0.11	-7.84	-2.12
59	1.43	1.26	2.31	1.69	-2.72	-2.33
60	4.15	3.42	3.51	4.69	-16.23	-3.26
61	0.85	0.72	0.2	1.25	-12.29	-0.7
62	5.32	4.51	5.18	5.87	-13.58	-5.74
63	-0.81	-0.93	-1.45	-1.15	5.06	0.33
64	2.68	2.34	2.58	3.18	-14.99	-4.22

65	4.24	3.46	3.93	4.73	-13.81	-3.89
66	0.98	0.88	1.57	1.18	-7.39	-2.55
67	0.31	0.33	1.39	0.27	1.96	-0.69
68	-3.7	-3.01	-3	-4.32	7.3	4.65
69	n	n	n	n	n	n
70	0.23	0.17	-0.07	0.39	-5.26	1.11
71	2.93	2.41	2.41	3.73	-12.31	-2.18
72	n	n	n	n	n	n
73	n	n	n	n	n	n
74	0.08	-0.1	0.11	-0.15	-2.28	-0.17
75	n	n	n	n	n	n
76	-0.02	-0.03	-0.09	n	-0.83	-0.23

NH RDCs [Hz] A29-A34:

pub	A29	A30	A31	A32	A33	A34
source	B2003	R2005	R2005	R2005	R2005	R2005
identify	C12E5	A	B	C	D	E
D23M	x			x	x	x
1	n	n	n	n	n	n
2	11.4	10.13	13.01	10.31	9.09	7.41
3	-3.66	-12.11	-15.41	-10.51	-7.79	-7.7
4	-4.23	n	n	-10.11	-8.11	-6.75
5	-1.7	-2.43	-1.34	-3.41	-3.63	-1.68
6	-2.05	5.43	14.09	4.7	3.08	4.58
7	5.36	15	23.41	12.83	8.65	9.34
8	-2.48	0.53	2.74	-0.81	0.41	-2.18
9	n	n	n	n	n	n
10	n	n	n	n	n	n
11	12.31	15.19	19.05	14.3	10.92	10.63
12	1.02	11.51	22.09	10.63	6.38	8.4
13	1.55	1.49	3.75	0.19	-0.83	0.56
14	-3.15	-1.02	1.27	-2.49	-2.34	-0.57
15	-5.42	n	n	n	n	n
16	n	n	n	n	-2.13	n
17	3.94	1.82	3.81	2.83	2.28	n
18	14.25	14.52	n	14.79	11.9	10.83
19	n	n	n	n	n	n
20	7.25	-1.82	-4.7	-1.17	-1.37	-0.51
21	3.53	12.16	18.04	10.79	7.88	5.78
22	n	n	n	n	n	n
23	-2.14	8.76	n	n	n	4.24
24	n	n	n	n	n	n
25	-1.81	8.35	13.3	6.28	4.65	6.6
26	n	n	n	n	n	n
27	-2.88	n	n	11.33	8.08	8.85
28	-3.36	14.2	n	n	n	n
29	-2.89	3.78	6.72	1.92	1.56	3.11
30	-1.84	11	19.97	7.96	5.45	6.48

31	n	n	n	n	n	n
32	-2.43	11.16	18.63	9.12	7.32	8.77
33	-2.47	4.05	8.69	2.35	1.14	3.09
34	-4.52	13.23	21.78	10.37	7.37	8.63
35	0.1	-1.52	-5.27	-0.19	1.17	1.63
36	13.82	13.4	16.11	13.84	11.01	11.19
37	n	n	n	n	n	n
38	n	n	n	n	n	n
39	n	n	n	n	n	n
40	0.31	n	n	-11.85	-9.56	-8.41
41	-13.87	-15.57	-21.38	-14.7	-10.77	-13.36
42	-8.05	n	n	n	5.41	n
43	-5.91	n	n	n	7.41	6.72
44	-2.3	13.79	19.53	10.39	7.72	9.12
45	4.97	n	n	n	n	10.45
46	n	n	n	n	n	n
47	-2.26	8.99	15.34	6.59	4.08	5.93
48	13.61	10.12	n	9.74	7.67	8.05
49	3.7	-0.94	-0.52	-1.65	-1.88	-0.06
50	-4.37	n	n	7.33	6.64	7.45
51	5.8	6.66	n	7.04	6.52	7.08
52	10.64	5.46	4.21	6.27	6.64	5.7
53	n	n	n	n	n	n
54	-3.68	4.43	5.4	4.51	3.97	4.8
55	-3.17	n	n	2.8	1.71	3.19
56	-8.47	n	n	n	-16.96	-15.48
57	n	n	n	n	n	n
58	-3.99	-14.21	n	-13.55	-9.93	-9.69
59	-5.08	n	-34.84	-16.14	n	-11.9
60	-13.71	-12.46	-15.93	-11.08	-8.89	-10.98
61	-5.55	-1.72	-0.39	-1.85	-1.7	-4.47
62	-10.95	-20.6	-29.04	-19.04	-13.98	-14.71
63	1.34	6.71	7.48	6.91	6.47	6.89
64	-8.78	-20.86	-28.68	-20.99	-14.98	-18.25

65	-13.13	n	n	n	n	n
66	-6.13	-20.07	-25.83	-18.45	-14.16	-14.28
67	-2.62	-10.54	-14.04	-9.61	-8.82	-6.72
68	-2.12	13.56	20.54	9.44	6.99	7.46
69	n	16.26	n	13.51	9.85	n
70	-9.89	6.33	10.67	3.91	3.32	2.14
71	-12.16	n	n	-7.05	-6.48	-7.43
72	n	n	n	n	n	n
73	n	-12.26	-16.54	-11.37	-9.09	-9.05
74	-4.31	-6.96	-9.02	-6.73	-4.85	-4.46
75	n	n	n	n	n	n
76	-0.94	-1.56	-1.85	-1.5	-1.16	-1.14

NH RDCs [Hz] A35-A36:

pub	A35	A36		pub	A35	A36
source	R2005	R2005		source	R2005	R2005
identify	F	G		identify	F	G
D23M	x	x		D23M	x	x
1	n	n		39	n	-2.88
2	8.44	1.58		40	-3.15	-2.82
3	-5.87	-1.52		41	-9.26	-4.02
4	-6.77	-0.33		42	-2.4	2.67
5	-4.79	1.17		43	-2.5	3.15
6	-3.01	3.42		44	-1.28	4.91
7	1.86	4.38		45	1.94	4.76
8	-3.46	0.27		46	n	n
9	n	n		47	-3.06	3.68
10	n	n		48	9.35	1.43
11	8.88	2.7		49	-0.99	0.65
12	-1.41	4.06		50	-0.38	4.35
13	-2.42	1.76		51	6.33	1.44
14	-4.77	1.66		52	11.11	-0.6
15	-7.76	-2.5		53	n	n
16	n	-0.38		54	3.22	-0.03
17	1.62	0.18		55	-2.85	2.4
18	11.01	2.38		56	-11.01	-4.73
19	n	n		57	n	-6.35
20	2.93	-0.55		58	-6.1	-2.12
21	1.95	2.42		59	-8.51	-2.33
22	n	n		60	-8.55	-3.26
23	-2.71	3.78		61	-4.68	-0.7
24	n	4.22		62	-7.22	-5.74
25	0.08	2.93		63	6.97	0.33
26	n	3.22		64	-10.2	-4.22
27	-1.41	4.83		65	-5.57	-3.88
28	1.13	3.46		66	-9.3	-2.55
29	-2.6	2.44		67	-6.51	-0.69
30	-1.96	4.3		68	-1.76	4.65

31	n	n	69	-0.2	4.63
32	1.24	3.31	70	-3.44	1.11
33	-3.12	2.79	71	-7.02	-2.18
34	-1.52	4.14	72	n	n
35	5.33	-2.16	73	-6.33	-2.37
36	11.76	1.76	74	-3.27	-0.17
37	n	n	75	n	n
38	n	n	76	-0.95	-0.23

Table 4: For NH rdc data sets taken from the literature the following nomenclature has been used:

Name	Alignment conditions	Reference	Ref. Abbreviation
A14	E1	Lakomek et al., 2006	L2006
A15	E2	Lakomek et al., 2006	L2006
A16	E3	Lakomek et al., 2006	L2006
A17	E4	Lakomek et al., 2006	L2006
A18	E5	Lakomek et al., 2006	L2006
A19	Bic1	Ottiger and Bax, 1998	O1998
A20	Bic2	Ottiger and Bax, 1998	O1998
A21	PM1	Briggman and Tolman, 2003	B2003
A22	PM2	Briggman and Tolman, 2003	B2003
A23	PM3	Briggman and Tolman, 2003	B2003
A24	PM4	Briggman and Tolman, 2003	B2003
A25	PM5	Briggman and Tolman, 2003	B2003
A26	PM6	Briggman and Tolman, 2003	B2003
A27	Helfrich	Briggman and Tolman, 2003	B2003
A28	Phage	Briggman and Tolman, 2003	B2003
A29	C12E5	Briggman and Tolman, 2003	B2003
A30	A	Ruan and Tolman, 2005	R2005
A31	B	Ruan and Tolman, 2005	R2005
A32	C	Ruan and Tolman, 2005	R2005
A33	D	Ruan and Tolman, 2005	R2005
A34	E	Ruan and Tolman, 2005	R2005
A35	F	Ruan and Tolman, 2005	R2005
A36	G	Ruan and Tolman, 2005	R2005

Bibliography

[Akke et al., 1993] Akke, M., Bruschweiler, R., and Palmer, A. G. (1993). Nmr order parameters and free-energy - an analytical approach and its application to cooperative ca2+ binding by calbindin-d(9k). *J. Am. Chem. Soc.*, 115(21):9832–9833. 22, 94

[Akke and Palmer, 1996] Akke, M. and Palmer, A. G. (1996). Monitoring macromolecular motions on microsecond to millisecond time scales by r1rho-r1 constant relaxation time nmr spectroscopy. *J. Am. Chem. Soc.*, 118:911–912. 3, 23, 93, 94

[Ban et al., 2000] Ban, N., Nissen, P., Hansen, J., Moore, P. B., and Steitz, T. A. (2000). The complete atomic structure of the large ribosomal subunit at 2.4 a resolution. *Science*, 289(5481):905–20. 1

[Bang et al., 2005] Bang, D., Makhatadze, G. I., Tereshko, V., Kossiakoff, A. A., and Kent, S. B. (2005). Total chemical synthesis and x-ray crystal structure of a protein diastereomer: [d-gln 35]ubiquitin. *Angew Chem Int Ed Engl*, 44(25):3852–6. 74

[Barrientos et al., 2000] Barrientos, L. G., Dolan, C., and Gronenborn, A. M. (2000). Characterization of surfactant liquid crystal phases suitable for molecular alignment and measurement of dipolar couplings. *J Biomol NMR*, 16(4):329–37. 38

[Bax and Grishaev, 2005] Bax, A. and Grishaev, A. (2005). Weak alignment nmr: a hawk-eyed view of biomolecular structure. *Curr. Opin. Struct. Biol.*, 15:563 – 570. 50, 66

[Beal et al., 1996] Beal, R., Deveraux, Q., Xia, G., Rechsteiner, M., and Pickart, C. (1996). Surface hydrophobic residues of multiubiquitin chains essential for proteolytic targeting. *Proc Natl Acad Sci U S A*, 93(2):861–6. 47

[Bertoncini et al., 2005] Bertoncini, C. W., Jung, Y. S., Fernandez, C. O., Hoyer, W., Griesinger, C., Jovin, T. M., and Zweckstetter, M. (2005). Release of long-range ter-

tiary interactions potentiates aggregation of natively unstructured alpha-synuclein. *Proc. Natl. Acad. Sci.*, 102(5):1430–5. 103

[Best et al., 2006] Best, R. B., Lindorff-Larsen, K., DePristo, M. A., and Vendruscolo, M. (2006). Relation between native ensembles and experimental structures of proteins. *Proc Natl Acad Sci U S A*, 103(29):10901–6. 97

[Blackledge, 2005] Blackledge, M. (2005). Recent progress in the study of biomolecular structure and dynamics in solution from residual dipolar couplings. *Prog. Nucl. Magn. Reson. Spectrosc.*, 46:23 – 61. 26, 30

[Bloch, 1946] Bloch, F. (1946). Nuclear induction. *Phys. Rev.*, 70:460. 1

[Bloch et al., 1946] Bloch, F., Hansen, W. W., and Packard, M. (1946). Nuclear induction. *Phys. Rev.*, 69:127. 1

[Boehr et al., 2006] Boehr, D. D., McElheny, D., Dyson, H. J., and Wright, P. E. (2006). The dynamic energy landscape of dihydrofolate reductase catalysis. *Science*, 313(5793):1638–1642. 94, 104, 110

[Bouvignies et al., 2005a] Bouvignies, G., Bernado, P., and Blackledge, M. (2005a). Protein backbone dynamics from n-h-n dipolar couplings in partially aligned systems: a comparison of motional models in the presence of structural noise. *J. Magn. Res.*, 173(2):328–338. 61

[Bouvignies et al., 2005b] Bouvignies, G., Bernado, P., Meier, S., Cho, K., Grzesiek, S., Bruschweiler, R., and Blackledge, M. (2005b). Identification of slow correlated motions in proteins using residual dipolar and hydrogen-bond scalar couplings. *Proc. Natl. Acad. Sci. U. S. A.*, 102(39):13885–13890. 59, 69, 86, 94, 103, 109

[Bouvignies et al., 2006] Bouvignies, G., Markwick, P., Bruscheweiler, R., and Blackledge, M. (2006). Simultaneous determination of protein backbone structure and dynamics from residual dipolar couplings. *J. Am. Chem. Soc.*, 128(47):15100–15101. 94

[Bredenbeck et al., 2007a] Bredenbeck, J., Helbing, J., Kolano, C., and Hamm, P. (2007a). Ultrafast 2d-ir spectroscopy of transient species. *Chemphyschem*, 8(12):1747–56. 2

[Bredenbeck et al., 2007b] Bredenbeck, J., Helbing, J., Nienhaus, K., Nienhaus, G. U., and Hamm, P. (2007b). Protein ligand migration mapped by nonequilibrium 2d-ir exchange spectroscopy. *Proc Natl Acad Sci U S A*, 104(36):14243–8. 2

[Bremi and Brueschweiler, 1997] Bremi, T. and Brueschweiler, R. (1997). Locally anisotropic internal polypeptide backbone dynamics by nmr relaxation. *J. Am. Chem. Soc.*, 119:6672–6673. 48, 59

[Bremi et al., 1997] Bremi, T., Brueschweiler, R., and Ernst, R. R. (1997). A protocol for the interpretation of side-chain dynamics based on nmr relaxation - application to phenylalanines in antamanide. *J. Am. Chem. Soc.*, 119(18)):4272–4284. 48

[Briggman and Tolman, 2003] Briggman, K. B. and Tolman, J. R. (2003). De novo determination of bond orientations and order parameters from residual dipolar couplings with high accuracy. *J. Am. Chem. Soc.*, 125(34):10164–10165. 6, 27, 39, 64, 95

[Bruschweiler and Wright, 1994] Bruschweiler, R. and Wright, P. E. (1994). Nmr order parameters of biomolecules - a new analytical representation and application to the gaussian axial fluctuation model. *J. Am. Chem. Soc.*, 116(18):8426–8427. 25

[Brutscher et al., 1997] Brutscher, B., Bruschweiler, R., and Ernst, R. R. (1997). Backbone dynamics and structural characterization of the partially folded a state of ubiquitin by 1h, 13c, and 15n nuclear magnetic resonance spectroscopy. *Biochemistry*, 36(42):13043–53. 48

[Brzovic and Klevit, 2006] Brzovic, P. S. and Klevit, R. E. (2006). Ubiquitin transfer from the e2 perspective: why is ubch5 so promiscuous? *Cell Cycle*, 5(24):2867–73. 94

[Buckmaster, 1964] Buckmaster, H. A. (1964). Tables of angular momentum transformation matrix elements. *Canadian Journal of Physics*, 42:386 – 391. 31

[Buckmaster, 1966] Buckmaster, H. A. (1966). Further tables of angular momentum transformation matrix elements. *Canadian Journal of Physics*, 44:2525 – 2530. 31

[Burch and Haas, 1994] Burch, T. J. and Haas, A. L. (1994). Site-directed mutagenesis of ubiquitin. differential roles for arginine in the interaction with ubiquitin-activating enzyme. *Biochemistry*, 33(23):7300–8. 47

[Carr and Purcell, 1954] Carr, H. Y. and Purcell, E. M. (1954). Effects of diffusion on free precession in nuclear magnetic resonance experiments. *Phys. Rev.*, 94(3):630–638. 23

[Cavanagh et al., 1996] Cavanagh, J., Fairbrother, W. J., Palmer, A. G. I., and Skelton, N. J. (1996). Protein nmr spectroscopy. *Academic Press.* 10, 11

143

[Chang and Tjandra, 2005] Chang, S.-L. and Tjandra, N. (2005). Temperature dependence of protein backbone motion from carbonyl 13c and amide 15n nmr relaxation. *J. Magn. Res.*, 174:43–53. 6, 50, 70, 86, 98, 99

[Chung et al., 2007] Chung, H. S., Ganim, Z., Jones, K. C., and Tokmakoff, A. (2007). Transient 2d ir spectroscopy of ubiquitin unfolding dynamics. *Proc Natl Acad Sci U S A*, 104(36):14237–42. 2

[Cierpicki and Bushweller, 2004] Cierpicki, T. and Bushweller, J. (2004). Charged gels as orienting media for measurement of residual dipolar couplings in soluble and integral membrane proteins. *J. Am. Chem. Soc.*, 126:16259–16266. 62, 63

[Clackson and Wells, 1995] Clackson, T. and Wells, J. A. (1995). A hot spot of binding energy in a hormone-receptor interface. *Science*, 267(5196):383–6. 101, 103

[Clore and Schwieters, 2004a] Clore, G. M. and Schwieters, C. D. (2004a). Amplitudes of protein backbone dynamics and correlated motions in a small alpha/beta protein: correspondence of dipolar coupling and heteronuclear relaxation measurements. *Biochemistry*, 43(33):10678–91. 61

[Clore and Schwieters, 2004b] Clore, G. M. and Schwieters, C. D. (2004b). How much backbone motion in ubiquitin is required to account for dipolar coupling data measured in multiple alignment media as assessed by independent cross-validation? *J Am Chem Soc*, 126(9):2923–38. 6, 7, 28, 40, 50, 51, 61

[Clore and Schwieters, 2006] Clore, G. M. and Schwieters, C. D. (2006). Concordance of residual dipolar couplings, backbone order parameters and crystallographic b-factors for a small alpha/beta protein: A unified picture of high probability, fast atomic motions in proteins. *J. Mol. Biol.*, 355(5):879–886. 4, 6, 61

[Cohen-Tannoudji, 2001a] Cohen-Tannoudji, C. (2001a). Quantenmechanik 1. 9, 15

[Cohen-Tannoudji, 2001b] Cohen-Tannoudji, C. (2001b). Quantenmechanik 2. 9

[Cordier and Grzesiek, 1999] Cordier, F. and Grzesiek, S. (1999). Direct observation of hydrogen bonds in proteins by interresidue (3h)j(nc ') scalar couplings. *J. of the Am. Chem. Soc.*, 121(7):1601–1602. 58

[Cornilescu et al., 1998] Cornilescu, G., Marquardt, J. L., Ottiger, M., and Bax, A. (1998). Validation of protein structure from anisotropic carbonyl chemical shifts in a dilute liquid crystalline phase. *J. Am. Chem. Soc.*, 120:6836–6837. 37, 38, 41, 46, 75

[Davis et al., 2006] Davis, I. W., Arendall, W. B., r., Richardson, D. C., and Richardson, J. S. (2006). The backrub motion: how protein backbone shrugs when a sidechain dances. *Structure*, 14(2):265–74. 86

[de Alba et al., 1999] de Alba, E., Baber, J. L., and Tjandra, N. (1999). The use of residual dipolar coupling in concert with backbone relaxation rates to identify conformational exchange by nmr. *J. Am. Chem. Soc.*, 121(17):4282–4283. 47

[Delaglio et al., 1995] Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., and Bax, A. (1995). Nmrpipe: a multidimensional spectral processing system based on unix pipes. *J. Biomol. NMR*, 6(3):277–93. 38, 65

[Derrick and Wigley, 1994] Derrick, J. P. and Wigley, D. B. (1994). The third iggbinding domain from streptococcal protein g. an analysis by x-ray crystallography of the structure alone and in a complex with fab. *J Mol Biol*, 243(5):906–18. 59, 105

[Dittmer and Bodenhausen, 2004] Dittmer, J. and Bodenhausen, G. (2004). Evidence for slow motion in proteins by multiple refocusing of heteronuclear nitrogen/proton multiple quantum coherences in nmr. *J. Am. Chem. Soc.*, 126(5):1314–5. 47, 60

[Eigen and de Maeyer, 1955] Eigen, M. and de Maeyer, L. (1955). Die geschwindigkeit der neutralisierungsreaktion. *Naturwissenschaften*, 42:413. 2

[Eisenmesser et al., 2005] Eisenmesser, E. Z., Millet, O., Labeikovsky, W., Korzhnev, D. M., Wolf-Watz, M., Bosco, D. A., Skalicky, J. J., Kay, L. E., and Kern, D. (2005). Intrinsic dynamics of an enzyme underlies catalysis. *Nature*, 438(7064):117–121. 104, 110

[Ernst and Anderson, 1966] Ernst, R. R. and Anderson, W. A. (1966). Application of fourier transform spectroscopy to magnetic resonance. *Rev. Sci. Instr.*, 37:93–102. 1

[Ernst et al., 1987] Ernst, R. R., Bodenhausen, G., and Wokaun, A. (1987). Principles of nuclear magnetic resonance in one and two dimensions. *Oxford University Press*. 11

[Ervin et al., 2000] Ervin, J., Sabelko, J., and Gruebele, M. (2000). Submicrosecond real-time fluorescence sampling: application to protein folding. *J Photochem Photobiol B*, 54(1):1–15. 2

[Ferrage et al., 2006] Ferrage, F., Pelupessy, P., Cowburn, D., and Bodenhausen, G. (2006). Protein backbone dynamics trough 13c'-13calpha cross-relaxation in nmr spectroscopy. *J. Am. Chem. Soc.*, 128:11072–11078. 60

145

[Fierz et al., 2007] Fierz, B., Joder, K., Krieger, F., and Kiefhaber, T. (2007). Using triplet-triplet energy transfer to measure conformational dynamics in polypeptide chains. *Methods Mol Biol*, 350:169–87. 2

[Frauenfelder and Wolynes, 1985] Frauenfelder, H. and Wolynes, P. G. (1985). Rate theories and puzzles of hemeprotein kinetics. *Science*, 229(4711):337–45. 93

[Frederick et al., 2007] Frederick, K. K., Marlow, M. S., Valentine, K. G., and Wand, A. J. (2007). Conformational entropy in molecular recognition by proteins. *Nature*, 448(7151):325–U3. 22, 94

[Fushman and Cowburn, 1998] Fushman, D. and Cowburn, D. (1998). Model-independent analysis of 15n chemical shift anisotropy from nmr relaxation data. ubiquitin as a test example. *J. Am. Chem. Soc.*, 120:7109–7110. 47

[Goh et al., 2004] Goh, C. S., Milburn, D., and Gerstein, M. (2004). Conformational changes associated with protein-protein interactions. *Current Opinion in Structural Biology*, 14(1):104–109. 94

[Griesinger et al., 1987] Griesinger, C., Sorensen, O. W., and Ernst, R. R. (1987). Novel 3-dimensional nmr techniques for studies of peptides and biological macromolecules. *J. Am. Chem. Soc.*, 109(23):7227–7228. 1

[Grunberg et al., 2004] Grunberg, R., Leckner, J., and Nilges, M. (2004). Complementarity of structure ensembles in protein-protein binding. *Structure*, 12(12):2125–36. 94

[Haas and Siepmann, 1997] Haas, A. L. and Siepmann, T. J. (1997). Pathways of ubiquitin conjugation. *Faseb J*, 11(14):1257–68. 47

[Haken and Wolf, 2001] Haken, H. and Wolf, H. (2001). Atom- und quantenphysik. 7. 9

[Halic et al., 2006a] Halic, M., Blau, M., Becker, T., Mielke, T., Pool, M. R., Wild, K., Sinning, I., and Beckmann, R. (2006a). Following the signal sequence from ribosomal tunnel exit to signal recognition particle. *Nature*, 444(7118):507–11. 1

[Halic et al., 2006b] Halic, M., Gartmann, M., Schlenker, O., Mielke, T., Pool, M. R., Sinning, I., and Beckmann, R. (2006b). Signal recognition particle receptor exposes the ribosomal translocon binding site. *Science*, 312(5774):745–7. 1

[Hamm et al., 1995] Hamm, P., Zurek, M., Mantele, W., Meyer, M., Scheer, H., and Zinth, W. (1995). Femtosecond infrared spectroscopy of reaction centers from

rhodobacter sphaeroides between 1000 and 1800 cm-1. *Proc Natl Acad Sci U S A*, 92(6):1826–30. 2

[Harper and Schulman, 2006] Harper, J. W. and Schulman, B. A. (2006). Structural complexity in ubiquitin recognition. *Cell*, 124(6):1133–6. 94

[Henzler-Wildman et al., 2007] Henzler-Wildman, K. A., Lei, M., Thai, V., Kerns, S. J., Karplus, M., and Kern, D. (2007). A hierarchy of timescales in protein dynamics is linked to enzyme catalysis. *Nature*, 450(7171):913–U27. 2, 3, 60, 94

[Hicke et al., 2005] Hicke, L., Schubert, H. L., and Hill, C. P. (2005). Ubiquitin-binding domains. *Nat Rev Mol Cell Biol*, 6(8):610–21. 94, 101, 103

[Hirano et al., 2006] Hirano, S., Kawasaki, M., Ura, H., Kato, R., Raiborg, C., Stenmark, H., and Wakatsuki, S. (2006). Double-sided ubiquitin binding of hrs-uim in endosomal protein sorting. *Nat Struct Mol Biol*, 13(3):272–7. 74

[Hus and Bruschweiler, 2002] Hus, J.-C. and Bruschweiler, R. (2002). Principal component method for assessing structural heterogeneity across multiple alignment media. *J. Biomol. NMR*, 24:123–132. 21, 65

[Hus et al., 2003] Hus, J.-C., Peti, W., Griesinger, C., and Bruschweiler, R. (2003). Self-consistency analysis of dipolar couplings in multiple alignment of ubiquitin. *J. Am. Chem. Soc.*, 125:5596–5597. 21, 65

[Ihee et al., 2005a] Ihee, H., Lorenc, M., Kim, T. K., Kong, Q. Y., Cammarata, M., Lee, J. H., Bratos, S., and Wulff, M. (2005a). Ultrafast x-ray diffraction of transient molecular structures in solution. *Science*, 309(5738):1223–7. 2

[Ihee et al., 2005b] Ihee, H., Rajagopal, S., Srajer, V., Pahl, R., Anderson, S., Schmidt, M., Schotte, F., Anfinrud, P. A., Wulff, M., and Moffat, K. (2005b). Visualizing reaction pathways in photoactive yellow protein from nanoseconds to seconds. *Proc Natl Acad Sci U S A*, 102(20):7145–50. 2

[Jackson, 1983] Jackson (1983). Klassische elektrodynamik. *de Gruyter*, 2. 15

[James et al., 2003] James, L. C., Roversi, P., and Tawfik, D. S. (2003). Antibody multispecificity mediated by conformational diversity. *Science*, 299(5611):1362–7. 104, 110

[Johnson et al., 1999] Johnson, E. C., Lazar, G. A., Desjarlais, J. R., and Handel, T. M. (1999). Solution structure and dynamics of a designed hydrophobic core variant of ubiquitin. *Structure Fold Des*, 7(8):967–76. 37, 62

[Johnston et al., 1999] Johnston, S. C., Riddle, S. M., Cohen, R. E., and Hill, C. P. (1999). Structural basis for the specificity of ubiquitin c-terminal hydrolases. *Embo J*, 18(14):3877–87. 74

[Karplus, 2003] Karplus, M. (2003). Molecular dynamics of biological macromolecules: a brief history and perspective. *Biopolymers*, 68(3):350–8. 3

[Kay, 1998] Kay, L. (1998). Protein dynamics from nmr. *Nat. Struct. Biol.*, NMR supplement:513–516. 2, 3, 22

[Kay et al., 1989a] Kay, L. E., Torchia, D. A., and Bax, A. (1989a). Backbone dynamics of proteins as studied by 15n inverse detected heteronuclear nmr spectroscopy: application to staphylococcal nuclease. *Biochemistry*, 28(23):8972–9. 22, 93

[Kay et al., 1989b] Kay, L. E., Torchia, D. A., and Bax, A. (1989b). A novel approach for sequential assignment of 1h, 13c, and 15n spectra of proteins: heteronuclear triple-resonance three-dimensional nmr spectroscopy. application to calmodulin. *Biochemistry*, 29(19):4659–67. 2

[Kendrew et al., 1960] Kendrew, J. C., Dickerson, R. E., Strandberg, B. E., Hart, R. G., Davies, D. R., Phillips, D. C., and Shore, V. C. (1960). Structure of myoglobin - 3-dimensional fourier synthesis at 2 a resolution. *Nature*, 185(4711):422–427. 1

[Kern and Zuiderweg, 2003] Kern, D. and Zuiderweg, E. R. P. (2003). The role of dynamics in allosteric regulation. *Curr. Opin. Struct. Biol.*, 13(6):748–757. 2

[Kettling et al., 1998] Kettling, U., Koltermann, A., Schwille, P., and Eigen, M. (1998). Real-time enzyme kinetics monitored by dual-color fluorescence cross-correlation spectroscopy. *Proc Natl Acad Sci U S A*, 95(4):1416–20. 2

[Kiel and Serrano, 2006] Kiel, C. and Serrano, L. (2006). The ubiquitin domain superfold: structure-based sequence alignments and characterization of binding epitopes. *J Mol Biol*, 355(4):821–44. 0022-2836 (Print) Journal Article Research Support, Non-U.S. Gov't. 101, 103

[Kitahara et al., 2005] Kitahara, R., Yokoyama, S., and Akasaka, K. (2005). Nmr snapshots of a fluctuating protein structure: ubiquitin at 30 bar-3 kbar. *J. Mol. Biol.*, 347(2):277–85. 40, 41, 47, 49, 50, 108

[Koenig et al., 1999] Koenig, B. W., Hu, J.-S., Ottiger, M., Bose, S., Hendler, R. W., and Bax, A. (1999). Nmr measurement of dipolar couplings in proteins aligned by transient binding to purple membrane fragments. *J. Am. Chem. Soc.*, 121:1385–1386. 37

148

[Kolano et al., 2006] Kolano, C., Helbing, J., Kozinski, M., Sander, W., and Hamm, P. (2006). Watching hydrogen-bond dynamics in a beta-turn by transient two-dimensional infrared spectroscopy. *Nature*, 444(7118):469–72. 93

[Koradi et al., 1996] Koradi, R., Billeter, M., and Wuthrich, K. (1996). Molmol: A program for display and analysis of macromolecular structures. *Journal of Molecular Graphics*, 14(1):51–&. 66, 87

[Lakomek et al., 2006] Lakomek, N. A., Carlomagno, T., Becker, S., Griesinger, C., and Meiler, J. (2006). A thorough dynamic interpretation of residual dipolar couplings in ubiquitin. *J. Biomol. NMR*, 34:101–115. iii, v, vii, 50, 55, 61, 62, 64, 66, 70, 87, 89, 92, 94, 107, 108

[Lakomek et al., 2005] Lakomek, N. A., Carlomagno, T., Becker, S., Meiler, J., and Griesinger, C. (2005). Side-chain orientation and hydrogen-bonding imprint supra-τ_c motion on the protein backbone of ubiquitin. *Angew. Chem. Int. Ed.*, 44:7776–7778. iii, vi, vii, 61, 70, 86, 87, 88, 94, 107, 109

[Lakomek et al., 2008] Lakomek, N. A., Walter, K. F., Fares, C., Lange, O. F., de Groot, B. L., Grubmuller, H., Bruschweiler, R., Munk, A., Becker, S., Meiler, J., and Griesinger, C. (2008). Self-consistent residual dipolar coupling based model-free analysis for the robust determination of nanosecond to microsecond protein dynamics. *J Biomol NMR*, 41(3):139–55. 97

[Lee and Wand, 2001] Lee, A. L. and Wand, A. J. (2001). Microscopic origins of entropy, heat capacity and the glass transition in proteins. *Nature*, 411(6836):501–504. 3, 22

[Lee et al., 2006] Lee, S., Tsai, Y. C., Mattera, R., Smith, W. J., Kostelansky, M. S., Weissman, A. M., Bonifacino, J. S., and Hurley, J. H. (2006). Structural basis for ubiquitin recognition and autoubiquitination by rabex-5. *Nat Struct Mol Biol*, 13(3):264–71. 74

[Li et al., 1996] Li, Z. G., Raychaudhuri, S., and Wand, A. J. (1996). Insights into the local residual entropy of proteins provided by nmr relaxation. *Protein Science*, 5(12):2647–2650. 3

[Lienin et al., 1998] Lienin, S. F., Bremi, T., Brutscher, B., Brueschweiler, R., and Ernst, R. R. (1998). Anisotropic intramolecular backbone dynamics of ubiquitin characterized by nmr relaxation and md computer simulation. *J. Am. Chem. Soc.*, 120:9870–9879. 17, 48, 59

[Lindorff-Larsen et al., 2005] Lindorff-Larsen, K., Best, R. B., Depristo, M. A., Dobson, C. M., and Vendruscolo, M. (2005). Simultaneous determination of protein structure and dynamics. *Nature*, 433(7022):128–32. 94, 97, 103, 110

[Lipari and Szabo, 1982a] Lipari, G. and Szabo, A. (1982a). Model-free approach to the interpretation of nuclear magnetic-resonance relaxation in macromolecules .1. theory and range of validity. *J. Am. Chem. Soc.*, 104(17):4546–4559. 6, 24, 30, 55, 93, 99

[Lipari and Szabo, 1982b] Lipari, G. and Szabo, A. (1982b). Model-free approach to the interpretation of nuclear resonance relaxation in macromolecules. 2. analysis of experimental results. *J. Am. Chem. Soc.*, 104:4559–4570. 24, 55

[Lorieau and McDermott, 2006] Lorieau, J. L. and McDermott, A. E. (2006). Conformational flexibility of a microcrystalline globular protein: Order parameters by solid-state nmr spectroscopy. *J. Am. Chem. Soc.*, 128(35):11505 – 11512. 60

[Losonczi and Prestegard, 1998] Losonczi, J. and Prestegard, J. (1998). Improved dilute bicelle solutions for high-resolution nmr of biological macromolecules. *J. Biomol. NMR*, 12(3):447–451. 37

[Losonczi et al., 1999] Losonczi, J. A., Andrec, M., Fischer, M. W. F., and Prestegard, J. H. (1999). Order matrix analysis of residual dipolar couplings using singular value decomposition. *J. Magn. Res.*, 138:334–342. 18, 21

[Louhivuori et al., 2006] Louhivuori, M., Otten, R., Lindorff-Larsen, K., and Annila, A. (2006). Conformational fluctuations affect protein alignment in dilute liquid crystal media. *J. Am. Chem. Soc.*, 128(13):4371–4376. 69, 90, 109

[Louhivuori et al., 2007] Louhivuori, M., Otten, R., Salminen, T., and Annila, A. (2007). Evidence of molecular alignment fluctuations in aqueous dilute liquid crystalline media. *J. Biomol. Nmr*, 39(2):141–152. 62, 80, 90, 109

[Ma and Gruebele, 2005] Ma, H. and Gruebele, M. (2005). Kinetics are probe-dependent during downhill folding of an engineered lambda6-85 protein. *Proc Natl Acad Sci U S A*, 102(7):2283–7. 2

[Mandel et al., 1995] Mandel, A. M., Akke, M., and Palmer III, A. G. (1995). Backbone dynamics of escherichia coli ribonuclease hi: correlations with structure and function in an active enzyme. *J. Mol. Biol.*, 246:144 – 163. 59, 86, 109

[Mandel et al., 1996] Mandel, A. M., Akke, M., and Palmer III, A. G. (1996). Dynamics of ribonuclease h: temperature dependance of motions on multiple time scales. *Biochemistry*, 35:16009 – 16023. 59, 86, 109

[Maragakis et al., 2008] Maragakis, P., Lindorff-Larsen, K., Eastwood, M. P., Dror, R. O., Klepeis, J. L., Arkin, I. T., Jensen, M. O., Xu, H., Trbovic, N., Friesner, R. A., Iii, A. G., and Shaw, D. E. (2008). Microsecond molecular dynamics simulation shows effect of slow loop dynamics on backbone amide order parameters of proteins. *J Phys Chem B*, 112(19):6155–8. 3

[Marion et al., 1989] Marion, D., Kay, L. E., Sparks, S. W., Torchia, D. A., and Bax, A. (1989). 3-dimensional heteronuclear nmr of n-15-labeled proteins. *J. Am. Chem. Soc.*, 111(4):1515–1517. 2

[Markwick et al., 2007] Markwick, P. R. L., Bouvignies, G., and Blackledge, M. (2007). Exploring multiple timescale motions in protein gb3 using accelerated molecular dynamics and nmr spectroscopy. *J. Am. Chem. Soc.*, 129(15):4724–4730. 94, 105

[Massi et al., 2005] Massi, F., Grey, M. J., and Palmer, A. G. (2005). Microsecond timescale backbone conformational dynamics in ubiquitin studied with nmr r1r. *Protein Science*, 14:735–742. 47, 48, 97

[Meiboom and Gill, 1958] Meiboom, S. and Gill, D. (1958). Modified spin-echo method for measuring nuclear relaxation times. *Review of Scientific Instruments*, 29(8):688–691. 23

[Meiler et al., 2000] Meiler, J., Peti, W., and Griesinger, C. (2000). Dipocoup: A versatile program for 3d-structure homology comparison based on residual dipolar couplings and pseudocontact shifts. *J Biomol NMR*, 17(4):283–94. 21, 31, 38, 52, 61

[Meiler et al., 2003] Meiler, J., Peti, W., and Griesinger, C. (2003). Dipolar couplings in multiple alignments suggest a helical motion in ubiquitin. *J. Am. Chem. Soc.*, 125:8072 – 8073. 28, 40, 47, 50, 108

[Meiler et al., 2001] Meiler, J., Prompers, J. J., Peti, W., Griesinger, C., and BrÃ¼schweiler, R. (2001). Model-free approach to the dynamic interpretation of residual dipolar couplings in globular proteins. *J. Am. Chem. Soc.*, 123:6098 – 6107. iii, v, 5, 6, 27, 29, 33, 48, 54, 94, 107

[Mills and Szyperski, 2002] Mills, J. L. and Szyperski, T. (2002). Protein dynamics in supercooled water: the search for slow motional modes. *J Biomol NMR*, 23(1):63–7. 47

[Misaghi et al., 2005] Misaghi, S., Galardy, P. J., Meester, W. J., Ovaa, H., Ploegh, H. L., and Gaudet, R. (2005). Structure of the ubiquitin hydrolase uch-l3 complexed with a suicide substrate. *J Biol Chem*, 280(2):1512–20. 74

[Mittermaier and Kay, 2006] Mittermaier, A. and Kay, L. E. (2006). New tools provide new insights in nmr studies of protein dynamics. *Science*, 312(5771):224–8. 2, 3, 22, 93, 94

[Miura et al., 1999] Miura, T., Klaus, W., Gsell, B., Miyamoto, C., and Senn, H. (1999). Characterization of the binding interface between ubiquitin and class i human ubiquitin-conjugating enzyme 2b by multidimensional heteronuclear nmr spectroscopy in solution. *J Mol Biol*, 290(1):213–28. 47

[Musselman et al., 2006] Musselman, C., Pitt, S. W., Gulati, K., Foster, L. L., Andricioaei, I., and Al-Hashimi, H. M. (2006). Impact of static and dynamic a-form heterogeneity on the determination of rna global structural dynamics using nmr residual dipolar couplings. *J Biomol NMR*, 36(4):235–49. 104

[Nolting, 2000] Nolting, W. (2000). Grundkurs theoretische physik. *vieweg*, 3. 15

[Oehlenschlager et al., 1996] Oehlenschlager, F., Schwille, P., and Eigen, M. (1996). Detection of hiv-1 rna by nucleic acid sequence-based amplification combined with fluorescence correlation spectroscopy. *Proc Natl Acad Sci U S A*, 93(23):12811–6. 2

[Oh et al., 1988] Oh, B. H., Westler, W. M., Darba, P., and Markley, J. L. (1988). Protein c-13 spin systems by a single two-dimensional nuclear magnetic-resonance experiment. *Science*, 240(4854):908–911. 2

[Oschkinat et al., 1988] Oschkinat, H., Griesinger, C., Kraulis, P. J., Sorensen, O. W., Ernst, R. R., Gronenborn, A. M., and Clore, G. M. (1988). Three-dimensional nmr spectroscopy of a protein in solution. *Nature*, 332(6162):374–6. 1

[Ottiger and Bax, 1998] Ottiger, M. and Bax, A. (1998). Determination of relative n-hn, n-c', ca-c', and ca-ha effective bond lengths in a protein by nmr in a dilute liquid crystalline phase. *J. Am. Chem. Soc.*, 120:12334–12341. 64, 66, 95

[Ottiger et al., 1998] Ottiger, M., Delaglio, F., and Bax, A. (1998). Measurement of j and dipolar couplings from simplified two-dimensional nmr spectra. *J. Magn. Res.*, 131:373–378. 38, 64

[Palmer et al., 2001] Palmer, A. G., Kroenke, C. D., and Loria, J. P. (2001). Nuclear magnetic resonance methods for quantifying microsecond-to-millisecond mo-

tions in biological macromolecules. *Nuclear Magnetic Resonance of Biological Macromolecules, Pt B*, 339:204–238. 23

[Palmer, 2004] Palmer, A. G., r. (2004). Nmr characterization of the dynamics of biomacromolecules. *Chem. Rev.*, 104(8):3623–40. 2, 3, 22

[Pelupessy et al., 2003] Pelupessy, P., Ravindranathan, S., and Bodenhausen, G. (2003). Correlated motions of successive amide n-h bonds in proteins. *J. Biomol. NMR*, 25:265–280. 60

[Penengo et al., 2006] Penengo, L., Mapelli, M., Murachelli, A. G., Confalonieri, S., Magri, L., Musacchio, A., Di Fiore, P. P., Polo, S., and Schneider, T. R. (2006). Crystal structure of the ubiquitin binding domains of rabex-5 reveals two modes of interaction with ubiquitin. *Cell*, 124(6):1183–95. 74, 101, 103

[Perutz et al., 1960] Perutz, M. F., Rossmann, M. G., Cullis, A. F., Muirhead, H., Will, G., and North, A. C. T. (1960). Structure of haemoglobin - 3-dimensional fourier synthesis at 5.5-a resolution, obtained by x-ray analysis. *Nature*, 185(4711):416–422. 1

[Peti et al., 2002] Peti, W., Meiler, J., Brueschweiler, R., and Griesinger, C. (2002). Model-free analysis of protein backbone motion from residual dipolar couplings. *J. Am. Chem. Soc.*, 124:5822–5833. iii, v, 6, 27, 28, 36, 39, 41, 50, 61, 62, 70, 94, 107, 108

[Porschke and Eigen, 1971] Porschke, D. and Eigen, M. (1971). Co-operative non-enzymic base recognition. 3. kinetics of the helix-coil transition of the oligoribouridylic–oligoriboadenylic acid system and of oligoriboadenylic acid alone at acidic ph. *J Mol Biol*, 62(2):361–81. 2

[Prompers and Bruschweiler, 2000] Prompers, J. J. and Bruschweiler, R. (2000). Thermodynamic interpretation of nmr relaxation parameters in proteins in the presence of motional correlations. *J. Phys. Chem. B*, 104(47):11416–11424. 3

[Prompers and Bruschweiler, 2001] Prompers, J. J. and Bruschweiler, R. (2001). Reorientational eigenmode dynamics: A combined md/nmr relaxation analysis method for flexible parts in globular proteins. *J. Am. Chem. Soc.*, 123(30):7305–7313. 69

[Prompers and Bruschweiler, 2002] Prompers, J. J. and Bruschweiler, R. (2002). General framework for studying the dynamics of folded and nonfolded proteins by nmr relaxation spectroscopy and md simulation. *Journal of the American Chemical Society*, 124(16):4522–4534. 69

[Purcell et al., 1946] Purcell, E. M., TORREY, H. C., and POUND, R. V. (1946). Resonance absorption by nuclear magnetic moments in a solid. *Phys. Rev.*, 69:37–38. 1

[Ramage et al., 1994] Ramage, R., Green, J., Muir, T. W., Ogunjobi, O. M., Love, S., and Shaw, K. (1994). Synthetic, structural and biological studies of the ubiquitin system: the total chemical synthesis of ubiquitin. *Biochem. J.*, 299:151–8. 38, 41, 61, 66, 92

[Richter et al., 2007] Richter, B., Gsponer, J., Varnai, P., Salvatella, X., and Vendruscolo, M. (2007). The mumo (minimal under-restraining minimal over-restraining) method for the determination of native state ensembles of proteins. *J. Biomol. NMR*, 37(2):117 – 135. 94, 97, 103, 110

[Ruan et al., 2008] Ruan, K., Briggman, K. B., and Tolman, J. R. (2008). De novo determination of internuclear vector orientations from residual dipolar couplings measured in three independent alignment media. *J Biomol NMR*, 41(2):61–76. 52

[Ruan and Tolman, 2005] Ruan, K. and Tolman, J. R. (2005). Composite alignment media for the measurement of independent sets of nmr residual dipolar couplings. *J. Am. Chem. Soc.*, 127(43):15032–15033. 64, 91, 95, 110

[Rubinstenn et al., 1998] Rubinstenn, G., Vuister, G. W., Mulder, F. A., Dux, P. E., Boelens, R., Hellingwerf, K. J., and Kaptein, R. (1998). Structural and dynamic changes of photoactive yellow protein during its photocycle in solution. *Nat Struct Biol*, 5(7):568–70. 2

[Ruckert and Otting, 2000] Ruckert, M. and Otting, G. (2000). Alignment of biological macromolecules in novel nonionic liquid crystalline media for nmr experiments. *J. Am. Chem. Soc.*, 122(32):7793–7797. 37, 63

[Salvatella et al., 2008] Salvatella, X., Richter, B., and Vendruscolo, M. (2008). Influence of the fluctuations of the alignment tensor on the analysis of the structure and dynamics of proteins using residual dipolar couplings. *J Biomol NMR*, 40(1):71–81. 29, 69, 80, 90, 109

[Sass et al., 2000] Sass, H.-J., Musco, G., Stahl, S. J., Wingfield, P. T., and Grzesiek, S. (2000). Solution nmr of proteins within polyacrylamide gels: Diffusional properties and residual alignment by mechanical stress or embedding of oriented purple membranes. *J. Biomol. NMR*, 18:303 – 309. 37, 39

[Saupe, 1968] Saupe, A. (1968). *Angew Chem Int Ed Engl*, 7:97. 18

[Schneider et al., 1992] Schneider, D. M., Dellwo, M. J., and Wand, A. J. (1992). Fast internal main-chain dynamics of human ubiquitin. *Biochemistry*, 31(14):3645–52. 3

[Schwieters et al., 2003] Schwieters, C. D., Kuszewski, J. J., Tjandra, N., and Clore, G. M. (2003). The xplor-nih nmr molecular structure determination package. *J Magn Reson*, 160(1):65–73. 52

[Snow et al., 2002] Snow, C. D., Nguyen, H., Pande, V. S., and Gruebele, M. (2002). Absolute comparison of simulated and experimental protein-folding dynamics. *Nature*, 420(6911):102–6. 2

[Sorensen, 1989] Sorensen, O. W. (1989). Polarization transfer experiments in high-resolution nmr spectroscopy. *Prog. NMR Spectrosc.*, 21:503–569. 9

[Sprangers and Kay, 2007] Sprangers, R. and Kay, L. E. (2007). Quantitative dynamics and binding studies of the 20s proteasome by nmr. *Nature*, 445(7128):618–22. 2

[Sprangers et al., 2007] Sprangers, R., Velyvis, A., and Kay, L. E. (2007). Solution nmr of supramolecular complexes: providing new insights into function. *Nat Methods*, 4(9):697–703. 2

[Srajer et al., 1996] Srajer, V., Teng, T., Ursby, T., Pradervand, C., Ren, Z., Adachi, S., Schildkamp, W., Bourgeois, D., Wulff, M., and Moffat, K. (1996). Photolysis of the carbon monoxide complex of myoglobin: nanosecond time-resolved crystallography. *Science*, 274(5293):1726–9. 2, 5

[Sugase et al., 2007] Sugase, K., Dyson, H. J., and Wright, P. E. (2007). Mechanism of coupled folding and binding of an intrinsically disordered protein. *Nature*, 447(7147):1021–5. 104

[Tang et al., 2007] Tang, C., Schwieters, C. D., and Clore, G. M. (2007). Open-to-closed transition in apo maltose-binding protein observed by paramagnetic nmr. *Nature*, 449(7165):1078–82. 104, 110

[Teo et al., 2004] Teo, H., Veprintsev, D. B., and Williams, R. L. (2004). Structural insights into endosomal sorting complex required for transport (escrt-i) recognition of ubiquitinated proteins. *J Biol Chem*, 279(27):28689–96. 74

[Tjandra and Bax, 1997] Tjandra, N. and Bax, A. (1997). Direct measurment of distances and angles in biomolecules by nmr in a dilute liquid crystalline medium. *Science*, 278:1111–1113. 5, 17, 27

[Tjandra et al., 1995] Tjandra, N., Feller, S. E., Pastor, R. W., and Bax, A. (1995). Rotational diffusion anisotropy of human ubiquitin from 15n nmr relaxation. *J. Am. Chem. Soc.*, 117(50):12562–12566. 98

[Tolman, 2001] Tolman, J. R. (2001). Dipolar couplings as a probe of molecular dynamics and structure in solution. *Curr Opin Struct Biol*, 11(5):532–9. 104

[Tolman, 2002] Tolman, J. R. (2002). A novel approach to the retrieval of structural and dynamic information from residual dipolar couplings using several oriented media in biomolecular nmr spectroscopy. *J. Am. Chem. Soc.*, 124(40):12020–12030. 5, 6, 27, 28, 40, 52, 94

[Tolman et al., 1995] Tolman, J. R., Flanagan, J. M., Kennedy, M. A., and Prestegard, J. H. (1995). Nuclear magnetic dipole interactions in field-oriented proteins: Information for structure determination in solution. *Proc. Natl. Acad. Sci. USA*, 92:9279–9283. 5, 17, 27

[Tolman et al., 1997] Tolman, J. R., Flanagan, J. M., Kennedy, M. A., and Prestegard, J. H. (1997). Nmr evidence for slow collective motions in cyanometmyoglobin. *Nat. Struct. Biol.*, 4(4)):292–297. 5, 27

[Tolman and Ruan, 2006] Tolman, J. R. and Ruan, K. (2006). Nmr residual dipolar couplings as probes of biomolecular dynamics. *Chemical Reviews*, 106(5):1720–1736. 26, 30

[Triba et al., 2005] Triba, M. N., Warschawski, D. E., and Devaux, P. F. (2005). Reinvestigation by phosphorus nmr of lipid distribution in bicelles. *Biophys J*, 88(3):1887–901. 63

[Tycko et al., 2000] Tycko, R., Blanco, F., and Ishii, Y. (2000). Alignment of biopolymers in strained gels: A new way to create detectable dipole-dipole couplings in high-resolution biomolecular nmr. *J. Am. Chem. Soc.*, 122(38):9340 –9341. 37, 39

[Vijay-Kumar et al., 1987] Vijay-Kumar, S., Bugg, C. E., and Cook, W. J. (1987). Structure of ubiquitin refined at 1.8 a resolution. *J. Mol. Biol.*, 194(3):531–44. 38, 41

[Vogeli et al., 2008] Vogeli, B., Yao, L., and Bax, A. (2008). Protein backbone motions viewed by intraresidue and sequential hn-halpha residual dipolar couplings. *J Biomol NMR*, 41(1):17–28. 51

[Wagner and Wuthrich, 1978] Wagner, G. and Wuthrich, K. (1978). Dynamic model of globular protein conformations based on nmr studies in solution. *Nature*, 275(5677):247–8. 1

[Walter, 2007] Walter, K. F. A. (2007). Bestimmung der dynamik von ubiquitin in loesung mittels orientierungsabhaengiger nmr-parameter. *Diplomarbeit*. 123

[Wang et al., 2001] Wang, L., Pang, Y., Holder, T., Brender, J. R., Kurochkin, A. V., and Zuiderweg, E. R. (2001). Functional dynamics in the active site of the ribonuclease binase. *Proc Natl Acad Sci U S A*, 98(14):7684–9. 2, 3, 93, 94

[Wang et al., 2003] Wang, T., Cai, S., and Zuiderweg, E. R. (2003). Temperature dependence of anisotropic protein backbone dynamics. *J Am Chem Soc*, 125(28):8639–43. 0002-7863 (Print) Journal Article Research Support, U.S. Gov't, Non-P.H.S. 51

[Werner, 1992a] Werner, J. (1992a). Numerische mathematik 1. *vieweg*. 21, 33

[Werner, 1992b] Werner, J. (1992b). Numerische mathematik 2. *vieweg*. 21, 33

[Wilkinson et al., 1999] Wilkinson, K. D., Laleli-Sahin, E., Urbauer, J., Larsen, C. N., Shih, G. H., Haas, A. L., Walsh, S. T., and Wand, A. J. (1999). The binding site for uch-l3 on ubiquitin: mutagenesis and nmr studies on the complex between ubiquitin and uch-l3. *J Mol Biol*, 291(5):1067–77. 47

[Wist et al., 2004] Wist, J., Frueh, D., Tolman, J. R., and Bodenhausen, G. (2004). Triple quantum decoherence under multiple refocusing: Slow correlated chemical shift modulations of c ' and n nuclei in proteins. *J. Biomol. Nmr*, 28(3):263–272. 47

[Wuthrich, 1986] Wuthrich, K. (1986). Nmr of proteins and nucleic acids. *Wiley Interscience*. 1

[Yang et al., 1997] Yang, D. W., Mok, Y. K., FormanKay, J. D., Farrow, N. A., and Kay, L. E. (1997). Contributions to protein entropy and heat capacity from bond vector motions measured by nmr spin relaxation. *Journal of Molecular Biology*, 272(5):790–804. 94

[Yao and Bax, 2007] Yao, L. and Bax, A. (2007). Modulating protein alignment in a liquid-crystalline medium through conservative mutagenesis. *J Am Chem Soc*, 129(37):11326–7. 91, 110

[Yao et al., 2008] Yao, L., Vogeli, B., Torchia, D. A., and Bax, A. (2008). Simultaneous nmr study of protein structure and dynamics using conservative mutagenesis. *J Phys Chem B*, 112(19):6045–56. 28, 52

157

[Zare, 1988] Zare, R. (1988). Angular momentum. *John Wiley and Sons.* 29, 31

[Zhang et al., 2007] Zhang, Q., Stelzer, A. C., Fisher, C. K., and Al-Hashimi, H. M. (2007). Visualizing spatially correlated dynamics that directs rna conformational transitions. *Nature*, 450(7173):1263–7. 2, 104, 105

[Zhang et al., 2006] Zhang, Q., Sun, X., Watt, E. D., and Al-Hashimi, H. M. (2006). Resolving the motional modes that code for rna adaptation. *Science*, 311(5761):653–6. 104

[Zweckstetter and Bax, 2000] Zweckstetter, M. and Bax, A. (2000). Prediction of sterically induced alignment in a dilute crystalline phase: Aid to protein structure determination by nmr. *J. Am. Chem. Soc.* 21, 31, 52, 65, 66, 73

[Zweckstetter and Bax, 2001] Zweckstetter, M. and Bax, A. (2001). Characterization of molecular alignment in aqueous suspensions of pf1 bacteriophage. *J. Biomol. NMR*, 20:365–377. 37, 63

[Zweckstetter and Bax, 2002] Zweckstetter, M. and Bax, A. (2002). Evaluation of uncertainty in alignment tensors obtained from dipolar couplings. *J. Biomol. NMR*, 23(2):127–37. 7, 41, 61

Lebenslauf

Name	**Nils-Alexander Lakomek**
Geburtsdatum	23. April 1979
Geburtsort	Göttingen
Staatsangehörigkeit	Deutsch

Seit September 2004	Doktorand in der Abteilung NMR-basierte Strukturbiologie am Max-Planck-Institut für biophysikalische Chemie in Göttingen; Titel der Arbeit unter Anleitung von Prof. Dr. Christian Griesinger: *Neue Methoden in der Kernspinresonanz-Spektroskopie zur Untersuchung der Dynamik von Proteinen*
Juli 2005	Teilnahme an der NMR Sommerschule: *Protein NMR: Recording, structure calculation and evaluation*, organisiert von Prof. Flemming Poulsen, Universität Kopenhagen, Dänemark
Juli 2004	Diplom in Physik (Universität Göttingen)
Mai 2003 – Mai 2004	Diplomarbeit in der Abteilung NMR-basierte Strukturbiologie am Max-Planck-Institut für biophysikalische Chemie in Göttingen; Titel der Arbeit unter Anleitung von Prof. Dr. Christian Griesinger: *Entwicklung und Anwendung von NMR-Methoden zur Untersuchung der Struktur und Dynamik von Proteinen*
Oktober 1999 – Juli 2004	Studium der Physik an der Universität Göttingen, Vordiplom im Juli 2001
Oktober 1998 – Oktober 1999	Zivildienst im Universitätsklinikum Göttingen
Juli 1998	Abitur
1991 – 1998	Besuch des Max-Planck-Gymnasiums in Göttingen