Habilitation à Diriger des Recherches (HDR) Characterising disordered proteins using nuclear magnetic resonance spectroscopy: Recent developments and future perspectives Caractérisation des protéines désordonnées par résonance magnétique nucléaire: Développements récents et perspectives futures

Malene Ringkjøbing Jensen

▶ To cite this version:

Malene Ringkjøbing Jensen. Habilitation à Diriger des Recherches (HDR) Characterising disordered proteins using nuclear magnetic resonance spectroscopy: Recent developments and future perspectives Caractérisation des protéines désordonnées par résonance magnétique nucléaire: Développements récents et perspectives futures. Structural Biology [q-bio.BM]. Université Joseph Fourier, 2012. tel-01297937

HAL Id: tel-01297937
https://hal.univ-grenoble-alpes.fr/tel-01297937

Submitted on 5 Apr 2016

Habilitation à Diriger des Recherches (HDR)
Université Joseph Fourier, Grenoble
UFR de Chimie

Characterising disordered proteins using nuclear magnetic resonance spectroscopy: Recent developments and future perspectives

Caractérisation des protéines désordonnées par résonance magnétique nucléaire: Développements récents et perspectives futures

DR. MALENE RINGKJØBING JENSEN
INSTITUT DE BIOLOGIE STRUCTURALE JEAN-PIERRE EBEL
GRENOBLE, FRANCE

Members of the habilitation jury:

Prof. Angela Gronenborn	(rapporteur)
Prof. H. Jane Dyson	(rapporteur)
Prof. Julian Garcia	(rapporteur)
Dr. Teresa Carlomagno	(examinateur)
Dr. Anja Böckmann	(examinateur)
Dr. Anne Imberty	(examinateur)
Prof. Eva Pebay-Peyroula	(examinateur)

Table of contents

1. Preface

This research habilitation focuses on the use of Nuclear Magnetic Resonance (NMR) spectroscopy for studying intrinsically disordered proteins and their interactions at atomic resolution. In particular, special emphasis is put on the development of ensemble descriptions of the proteins on the basis of experimental NMR data and the perspective of combining several NMR parameters with diverse averaging properties into a unified representation of the disordered state.

The habilitation is divided into two main parts, where the first part (chapter 3-7) describes some of the work carried out after my Ph.D. thesis in 2006 focusing mainly on the development of ensemble descriptions, while the last part concerns future research projects aiming at the application of ensemble descriptions to novel biological systems involving disordered domains (chapter 8-9).

The work described in this habilitation is based on the following key publications:

Chapter 3:	Nodet et al. J. Am. Chem. Soc. (2009), 131, 17908-17918. *"Quantitative description of backbone conformational sampling of unfolded proteins at amino acid specific resolution from NMR residual dipolar couplings"*
	Gabel et al. J. Am. Chem. Soc. (2009), 131, 8769-8771. *"Quantitative model-free analysis of urea binding to unfolded ubiquitin using a combination of small angle X-ray and neutron scattering"*
Chapter 4:	Salmon et al. J. Am. Chem. Soc. (2010), 132, 8407-8418. *"NMR characterization of long-range order in intrinsically disordered proteins"*
Chapter 5:	Jensen et al. J. Am. Chem. Soc. (2010), 132, 1270-1272. *"Defining conformational ensembles of intrinsically disordered and partially folded proteins directly from chemical shifts"*
Chapter 6:	Jensen and Blackledge. J. Am. Chem. Soc. (2008), 130, 11266-11267. *"On the origin of NMR dipolar waves in transient helical elements of partially folded proteins"*
	Jensen et al. J. Am. Chem. Soc. (2008) 130, 8055-8061. *"Quantitative conformational analysis of partially folded proteins from residual dipolar couplings: Applications to the molecular recognition element of Sendai virus nucleoprotein"*
Chapter 7:	Jensen et al. Proc. Natl. Acad. Sci. U.S.A (2011), 108, 9839-9844. *"Intrinsic disorder in measles virus nucleocapsids"*

The work described in this habilitation has been the collaborative effort of many people to whom I should not forget to express my gratitude.

First of all, I would like to thank Dr. Martin Blackledge for accepting me as a postdoctoral fellow in his group, for the stimulating scientific discussions, for introducing me to the world of intrinsically disordered proteins, for the numerous conversations about the latest football results (even the Danish ones), for so many things that it will be impossible to list them all. Thanks Martin, you provide a true scientific and pleasant working environment!

I would like to thank all the members of the FDP group – past and present – who have all contributed to this work: Guillaume Communie, Alexander Grimm, Dr. Jie-rong Huang, Jaka Kragelj, Antoine Licinio, Damien Maurin, Dr. Luca Mollica, Dr. Gabrielle Nodet, Valéry Ozenne, Dr. Loïc Salmon, Dr. Robert Schneider and Mingxi Yao. In particular, my appreciation goes to Loïc for being brave enough to undertake my proposed project: "relaxation measurements (week 1), grand fit (week 2), manuscript writing (week 3)". Although, my timeline was a bit off (I apologize for this), I very much enjoyed working with you, and I wish you the best of luck for your postdoctoral stay in the US. A special thanks also goes to the members of the ASTEROIDS team: Gabrielle, Loïc and Valéry. Thank you for the many versions of ASTEROIDS with different functionalities and the discussions on whether Python or C is more efficient for coding a genetic algorithm.

To this list of acknowledgments, I should not forget to add my collaborators outside FDP who have contributed to this work. First of all, I would like to thank Frank Gabel (IBS-Grenoble) for his expertise on small angle scattering and Dr. Eric Condamine and Dr. Adrien Favier (IBS-Grenoble) for assistance at the spectrometers. I would like to thank Prof. Rob W. H. Ruigrok and Prof. Marc Jamin (UVHCI-Grenoble) and the members of their groups, Dr. Euripedes Ribeiro, Dr. Cédric Leyrat, Filip Yabukarski, Nicolas Martinez and Ivan Ivanov, as well as Dr. Sonia Longhi (AFMB-Marseille) for the fruitful collaboration on intrinsic disorder in viral proteins. I would also like to thank Dr. Nico van Nuland and Dr. Jose-Luis Ortega Roldan (VIB-Brussels/University of Granada) for the collaboration on the determination of the structure and dynamics of weak protein-protein complexes as well as Prof. Lyndon Emsley and Dr. Guido Pintacuda (CRMN-Lyon) for the collaboration on solid state NMR on viral nucleocapsids.

Finally, I would like to thank all the jury members for taking the time to read and evaluate my work.

DR. MALENE RINGKJØBING JENSEN

OCTOBER 2011

2. Describing intrinsically disordered proteins at atomic resolution

2.1. Intrinsically disordered proteins

The central dogma that has motivated massive worldwide investment in structural genomic projects has been founded upon the assumption that the resolution of the three-dimensional structure of a finite number of proteins will provide the key to understanding biological activity (classical structure-function paradigm). However, over the last decade it has become clear that a large fraction (up to 40%) of the proteins encoded by the human genome are intrinsically disordered or contain disordered regions of significant length (> 50 aa) (*1–5*). Intrinsically disordered proteins (IDPs) remain functional despite a lack of a well-defined structure. The classical structure-function paradigm therefore breaks down for this class of proteins, and new insight into the relationship between primary sequence and molecular function is necessary. The importance of such investigations is underlined by the fact that numerous IDPs are strongly associated with human diseases, including cancer, cardiovascular disease, amyloidosis, neurodegenerative disease and diabetes (*6–8*). A better understanding of the residual structure and dynamics of IDPs in solution and, in particular, insight into how IDPs are recognised by their partner proteins, will not only improve our understanding of fundamental aspects of molecular biology, but will inevitably lead to the discovery of more efficient drugs targeting protein-protein interactions involving IDPs (*9*, *10*).

The conformational plasticity of IDPs and their intrinsic lack of rigid structure provide them with unique capabilities to act in functional modes not achievable by folded, globular proteins. A number of different scenarios have been identified for the binding of IDPs to their partner proteins (*11*). This includes the folding-upon-binding mechanism, where the IDPs fold into specific conformations upon binding to the partner proteins (*12–14*). Other IDPs are capable of folding into different conformations depending on the partner, illustrating the complexity of the different binding modes (*15*). Finally, some IDPs engage in complexes that are much more dynamic, where the IDPs do not necessarily adopt a specific conformation in the complex but rather sample various states on the surface of the partner (fuzziness) (*16*). It is not entirely clear how the intrinsic structural propensity of the primary sequence of an IDP is related to the kinetics of the binding reaction and the final conformation adopted in the complex. Therefore, in order to fully understand how IDPs carry out their function, atomic resolution models are necessary of the proteins both in their free, pre-recognition state and in complex with their partner proteins. Both of these states are very probably highly dynamic, and ensemble descriptions have emerged as the preferred tool for representing the structural and dynamic properties of IDPs and their complexes (*17*). Within such descriptions, it is assumed that the protein adopts a continuum of rapidly inter-converting structures, and the determination of these representative ensembles is one of the major challenges in the studies of IDPs.

2.2. NMR as a tool for studying disordered proteins

NMR spectroscopy has recently evolved into one of the most powerful techniques for studying IDPs (*18*–*22*). Even though the protein adopts many different conformers in solution, NMR allows site-specific characterization of average properties over all conformers of the disordered chain (Figure 1). NMR is a rich source of structural and dynamic information, and a number of experimental parameters can be used to characterize IDPs and their complexes in solution as described below.

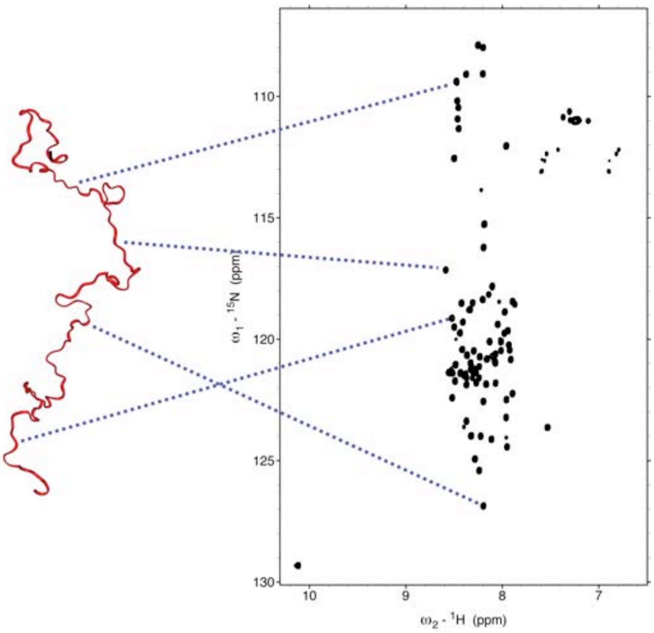

Figure 1: Illustration of the site-specific information available from NMR spectroscopy. Each peak in the ^{1}H-^{15}N HSQC spectrum corresponds to a specific amide group in the protein.

Chemical shifts are the most readily accessible NMR parameters and are sensitive to local backbone conformations in proteins (*23*–*25*). They can be used to locate transiently populated secondary structures in disordered proteins as well as for estimating the populations of these elements (*26*). Chemical shift changes upon partner protein binding (titration experiments) can also be used to provide information about the regions of interaction, dissociation constants, exchange rates and possible structural and dynamic changes upon binding. A straightforward interpretation of the chemical shifts of course requires that the signals of the complex are visible in the NMR spectra and that they can be assigned. In many cases, however, the coupled folding-and-binding reaction and, more generally, the dynamic complexes of the IDPs result in exchange broadening of the NMR signals. This is one of the major obstacles to overcome in future studies of complexes involving IDPs, and relaxation dispersion is potentially very powerful in this context (*12*). Relaxation dispersion measurements are sensitive to exchange processes occurring on the μs-ms time scale. It can be applied to exchanging systems involving a dominant NMR observable state inter-converting with a weakly populated (> 1%) state that cannot be probed directly (*27*, *28*). The advantage of relaxation dispersion for studies of IDPs and their complexes is that chemical shifts and exchange kinetics of the complex state can in principle be derived through measurements on signals of the free state. IDPs rarely exhibit exchange on the μs-ms time scale in their free states at ambient temperatures allowing a clear identification of exchange processes related to complex formation.

Scalar coupling constants, for example H^N-Hα couplings, are also sensitive reporters of local backbone conformations in proteins. The dependence of the ^3J coupling on the backbone dihedral angle has been parameterized according to a Karplus relationship (*29–31*):

$$J(\varphi) = A\cos^2(\varphi - 60^\circ) + B\cos(\varphi - 60^\circ) + C \qquad \text{(Eq. 1)}$$

A, *B*, and *C* have been optimized using coupling constants measured in several proteins of known structure and Eq. 1, therefore, provides a constraint on the distribution of φ angles in conformational ensembles of IDPs (*32*, *33*).

Besides chemical shifts and scalar couplings, residual dipolar couplings (RDCs) are one of the most powerful reporters of residual structure in the disordered state. In solution NMR, the dipolar coupling between two spins is effectively averaged to zero because all orientations of the protein molecule are equally probable (isotropic solution). A small part of the dipolar coupling can be re-introduced by weakly aligning the protein in the magnetic field for example using lipid bicelles (*34*), filamentous phages (*35–37*), polyacrylamide gels (*38*, *39*), liquid crystals (*40*) or by exploiting the magnetic anisotropy of paramagnetic metal ions (*41*). RDCs report on orientations of inter-nuclear bond vectors (e.g ^{15}N-$^1H^N$ and ^{13}Cα-^{1}Hα) with respect to the direction of the static magnetic field:

$$D_{ij} = -\frac{\gamma_i \gamma_j \hbar \mu_0}{8\pi^2 r^3}\left\langle \frac{3\cos^2\theta - 1}{2} \right\rangle \qquad \text{(Eq. 2)}$$

Here, γ_i and γ_j are the gyromagnetic ratios of the two nuclei *i* and *j*, θ is the angle between the bond vector and the static magnetic field, and the brackets indicate an average of all conformations sampled on time scales faster than the millisecond. The dipolar coupling can conveniently be expressed through a tensor that describes the overall alignment of the molecule in the magnetic field. In the absence of internal dynamics, Eq. 2 can be recast to (*42*, *43*):

$$D_{ij} = -\frac{\gamma_i \gamma_j \hbar \mu_0}{8\pi^2 r^3}\left[A_a(3\cos^2\upsilon - 1) + \frac{3}{2}A_r \sin^2\upsilon \cos(2\eta) \right] \qquad \text{(Eq. 3)}$$

Here, A_a and A_r are the axial and rhombic components of the alignment tensor and (υ,η) describes the orientation of the inter-nuclear vector with respect to this tensor. RDCs are powerful structural probes in IDPs and can be used to report on local backbone conformational behaviour and transiently populated secondary structures (*21*, *44–47*). Recently, it has been shown that RDCs are also sensitive to long-range contacts between different parts of the unfolded chain (*48*, *49*).

In general, long-range information (< 25Å) in disordered states is most easily extracted from the nuclear relaxation enhancements induced by an unpaired electron artificially introduced into the protein normally through an MTSL spin label attached to the side chain of a cysteine residue (*50*, *51*). The dipolar interaction induces a paramagnetic relaxation enhancement (PRE) manifest as a line broadening of the signal of the observed spin that depends on the inverse sixth power of the distance to the unpaired electron (*52*). Spin labels have been used to report on long-range contacts between different parts of the chain in IDPs (*53–56*) and to map protein folding by identifying important long-range contacts along the folding pathway (*57*).

Nuclear relaxation rates, for example ^{15}N, report on the local dynamics of proteins on the ps-ns time scale. For proteins with a folded, stable structure a characterization of the dynamics from these parameters is relatively straightforward, relying on the assumption that internal motion can be decoupled from overall tumbling of the protein (*58*, *59*). Analysis of NMR relaxation data to extract timescales and amplitudes of the underlying motions is not straightforward in the case of disordered proteins, since the common decoupling of global and internal motions no longer has any relevance. For this reason most studies so far of spin relaxation in IDPs provide only a qualitative data interpretation for example identifying less mobile protein segments or hydrophobic clusters from local increases in ^{15}N R_2 relaxation rates and steady-state nuclear Overhauser enhancements (nOes) (*60*).

In conclusion, a number of parameters are available from NMR that in principle allows a mapping of both the residual structure and the dynamics on various time scales of IDPs in solution. The major challenge, however, is how to interpret the experimental NMR data in terms of explicit conformational ensembles that are representative of a given protein in solution.

2.3. Ensemble descriptions from NMR data

The description of the highly heterogeneous conformational space available to IDPs is inherently difficult due to the vast number of degrees of conformational freedom available to the protein. Even so, progress has been made over the last fifteen years in the description of intrinsically disordered states using experimental NMR data. One of the first studies employed PREs converted into distance restraints to calculate a structural ensemble of the denatured state of staphylococcal nuclease using a combined distance geometry and molecular dynamics approach (*51*). Similar approaches were later used to map long-range order in denatured acyl Co-A binding protein (ACBP) (*53*) and α-synuclein (*61*) using PRE data from extensive spin labelling measurements. The combined use of RDCs and PREs in ensemble descriptions to map simultaneously local and long-range order was introduced in a study of the conformational landscape of urea-denatured ubiquitin using X-PLOR ensemble structure calculations showing that ubiquitin, although denatured, retains some degree of native structure (*56*). Similar conclusions were obtained in another study employing only the RDC data of ubiquitin as input in restrained molecular dynamics simulations (*62*).

One common feature of the methods mentioned above is that they rely on molecular dynamics simulations or on direct restraining of conformational ensembles using experimental data. Such approaches run the risk of under-sampling the available conformational space, and the effect of using direct constraints depends strongly on the relative weighting of experimental data and classical force field terms. It is clear that the accuracy of ensemble descriptions of disordered proteins will depend strongly on the efficiency of the "structure generator". For this reason, the Blackledge group has developed the Flexible-Meccano algorithm for creating structural ensembles of disordered states (*63*). The approach is conceptually simple, creating a large number of conformers using an amino acid specific random coil φ/ψ database derived from loop regions of high-resolution crystal structures. Flexible-Meccano allows for a very efficient, restraint-free sampling to flood the conformational space available to disordered proteins. This statistical coil description of the unfolded state thus provides a straightforward method for calculating RDC, PRE, chemical shift and *J*-coupling profiles that would be expected, if the protein behaved as a random coil, devoid of any specific or persistent local or long-range structure. In its initial application, Flexible-Meccano was shown to provide structural ensembles in agreement with experimentally measured RDCs in urea-denatured apo-myoglobin and staphylococcal nuclease $\Delta131\Delta$ as well as in a two-domain viral protein containing a three-helix bundle and a long disordered chain (*63*).

Often, IDPs are not well described by an ensemble of random-coil conformers, but contain some degree of residual structure. In these cases, Flexible-Meccano can be used to detect deviations from random coil behaviour and by modifying the ϕ/ψ-database of Flexible-Meccano, different conformational sampling regimes can be tested in the regions of the protein where the experimental data deviate significantly from those predicted from the random coil state. This approach has successfully identified highly populated turn motifs in the IDP Tau (*33*) and α-helical propensity in the molecular recognition element of the N-terminal transactivation domain of the tumour suppressor p53 (*64*).

Testing different conformational sampling regimes is a hypothesis-driven method and therefore severely limits the practical applications as well as the potential for discovery. To overcome this problem and select conformational ensembles directly from experimental data without the need for prior hypotheses, we have developed **A S**election **T**ool for **E**nsemble **R**epresentations **O**f **I**ntrinsically **D**isordered **S**tates (ASTEROIDS) (*65*). ASTEROIDS is a sample-and-select approach, where sub-ensembles in agreement with experimental data are selected from a large pool of conformers that is assumed to efficiently sample the conformational space (local and long-range order) available to the protein under investigation. The selection of conformers follows the evolution scheme of a genetic algorithm, where generations are obtained by random selection, mutations and crossings of conformers.

The ASTEROIDS approach is conceptually similar to other sample-and-select approaches such as ENSEMBLE, developed by Forman-Kay and co-workers (*66–70*). While the development of this kind of ensemble sampling is rapidly gaining in popularity, so far they have not been shown to be predictive. Indeed, the representation of IDPs by ensembles containing a restricted number of conformers is a classical ill-posed problem that has no unique solution. In the face of such potential for under-determination, and hence over-fitting, it would appear to be essential that the validity of the ensembles is independently verified. This can be achieved, for example, by using cross-validation procedures (consistency checks), where a fraction of the experimental data is left out of the selection and subsequently compared to the same data back-calculated from the selected ensemble. It is also necessary to consider how to construct the initial pool from which the sub-ensembles are selected. Under-representation in the pool of certain structural propensities will most likely lead to erroneous ensembles and poor cross-validation of independent experimental data. Another point to take into account is the number of conformers that are needed to describe a given system and the averaging and convergence characteristics over an ensemble of the different NMR parameters. For example, RDCs require many thousands of structures to converge, while parameters such as ^3J scalar couplings converge much faster (hundreds of structures). This inherently poses a problem when selecting ensembles comprising only a small number of structures and if special attention is not drawn to this fact, the derived conformational sampling could simply be wrong. These aspects, together with the development, validation and application of the ASTEROIDS approach, have been the focus of my research over the last few years as described in the following chapters.

3. Conformational sampling of disordered proteins from RDCs

3.1. Introduction

IDPs populate a vast conformational space, and the mapping of this landscape represents a classical ill-posed problem, in which the number and complexity of the available degrees of conformational freedom far outweigh the accessible experimental data that can be measured for a particular system. This implies that conformational ensembles can be selected that agree with experimental data, however, that do not necessarily represent the true conformational sampling of the protein. The development of robust procedures that address this issue is of paramount importance.

This chapter describes the application of ASTEROIDS to determine the site-specific conformational sampling of disordered proteins from RDCs. RDCs are sensitive to bond vector orientations averaged over the ensemble of disordered states and in principle allow a mapping of conformational space, provided that the RDCs are measured for different types of bond vectors. In order to characterise the behaviour of sample-and-select approaches for the interpretation of RDCs, we test the ASTEROIDS approach using extensive simulation. We demonstrate the importance of taking appropriate account of the convergence characteristics of the RDCs, as well as of determining the number of structures necessary to accurately characterize the conformational sampling of the protein. Importantly, we define parametric ranges over which ASTEROIDS can be used to accurately define conformational space directly from RDCs, and we demonstrate that agreement with experimental data, achieved outside these parametric ranges, does not necessarily capture the true conformational sampling of the protein.

Having established the ASTEROIDS protocol for targeting experimental RDCs, we apply it to the determination of site-specific conformational sampling in urea-denatured ubiquitin for which a large data set of experimental RDCs was obtained previously in 8 M urea at pH 2.5 (*71*). Analysis of the selected ASTEROIDS ensemble shows that ubiquitin on average samples more extended regions of Ramachandran space (poly-proline II and β-strand) compared to the standard random coil library. From the derived site-specific conformational sampling, we identify specific residue types that are more affected by urea binding than others.

3.2. Convergence characteristics of RDCs over structural ensembles

When applying ASTEROIDS to select small sub-ensembles in agreement with experimental RDCs, it is necessary to take into account the particular convergence characteristics of RDCs when averaged over structural ensembles. We say that convergence of a parameter has been reached when the addition of one more conformer to the ensemble does not perturb the calculated average parameter within a predefined limit. Figure 2 illustrates the convergence of the ^{15}N-$^{1}H^{N}$ RDC of residue 41 of ubiquitin over a structural ensemble with an increasing number of conformers. The alignment tensor was estimated on the basis of the shape of each conformer using PALES assuming steric alignment of the molecules (*72*, *73*). Only above 10K structures, convergence is reached, and it is clear that selection of a sub-ensemble consisting of for example 200 conformers will severely hamper the reliability of the determined conformational sampling. One way to overcome this problem is to use the so-called Local Alignment Windows (LAWs) (*74*), where the alignment tensor is calculated for a smaller segment (window) of the protein and only the RDCs of the central amino acid are kept. The LAW is then moved along the protein sequence to calculate the RDCs for all residues. As shown in Figure 2 the use of LAWs significantly improves the convergence characteristics of the RDCs going from thousands of conformers for the global tensor calculation to only a few hundreds for shorter LAWs.

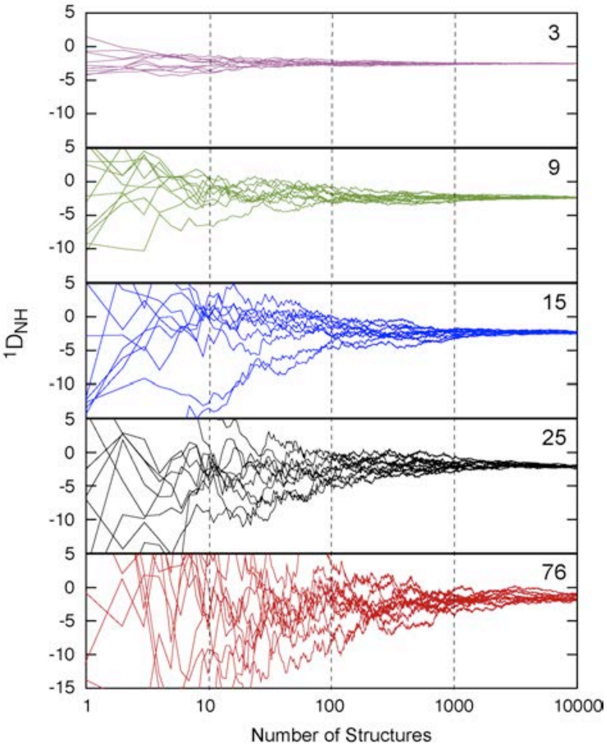

Figure 2: Convergence of the ^{15}N-^{1}H^N RDC of residue 41 of ubiquitin over a structural ensemble with an increasing number of conformers. Results are shown for the calculation using a global alignment tensor for the 76 amino acids of ubiquitin (red) and different sizes of LAWs: 25 (black), 15 (blue), 9 (green) and 3 (pink) amino acids.

In general, ensemble-averaged RDCs follow a bell-shaped distribution (baseline) on which local fluctuations are superimposed according to amino acid type and residual structure. This baseline can be analytically described using simulations of for example a random coil poly-valine chain where fluctuations due to local structure variations are essentially absent (Figure 3):

$$B(i) = 2b\cosh(a(i - d)) - c \qquad \text{(Eq. 4)}$$

Here, $d=(L+1)/2$ where L is the length of the protein, and a, b and c are parameters that depend on the length of the protein and the type of coupling. The calculation of RDCs using the LAW approach effectively removes the contribution from the baseline. The parameterization of the baseline is therefore useful, as we can re-introduce the baseline on the RDCs calculated using the LAWs by simple multiplication. In this way, RDCs obtained using the LAW approach closely resemble those predicted using a global alignment tensor (Figure 4).

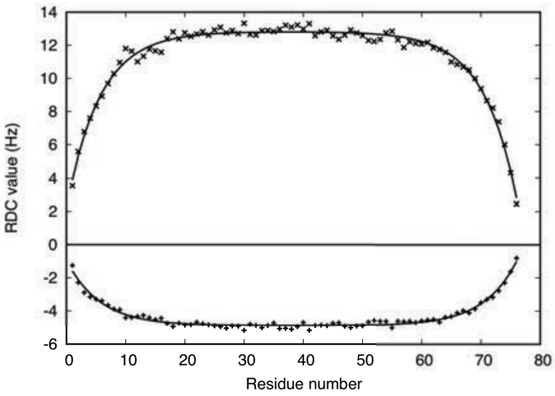

Figure 3: Ensemble-averaged ^{15}N-^{1}H^{N} (below zero) and ^{13}Cα-^{1}Hα (above zero) RDCs (100K conformers) of a random coil poly-valine chain of 76 amino acids. The solid line corresponds to the best fit of Eq. 4 to the RDCs.

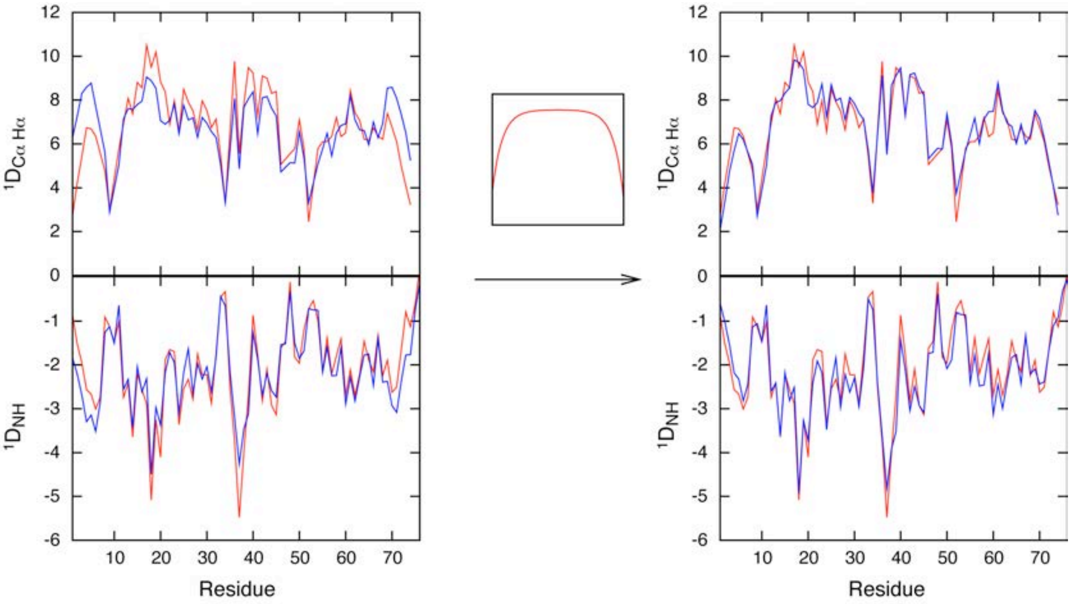

Figure 4: Comparison of ^{15}N-^{1}H^{N} and ^{13}Cα-^{1}Hα RDCs in ubiquitin using 15 amino acids LAWs (blue, left) with RDCs calculated using a global tensor (red, left and right). The multiplication of the LAW predicted RDCs with the parameterized baseline (blue, right) closely resembles the RDCs predicted using a global tensor. In all cases an average over 50K conformers was used.

It is also important to consider the size of the LAW used to calculate the RDCs from a structural ensemble. If a window is used that is significantly smaller than the persistence length of the chain, erroneous RDCs will be predicted. For example, a LAW of three amino acids fails to reproduce the RDCs calculated using a global alignment tensor, independent of how many structures are used in the average (Figure 5). Nine and 15 amino acid windows produce satisfactory results for a standard conformational sampling regime (Figure 5). The persistence length of an unfolded chain may, however, vary depending on the conformational sampling. Using a more rigid statistical coil sampling, a window of nine amino acids also fail to reproduce the RDCs predicted using a global tensor. Therefore, a 15 amino acid window seems to be the most suitable compromise, allowing relatively fast convergence of the RDCs over a structural ensemble and providing a good reproduction of RDCs calculated using a global tensor.

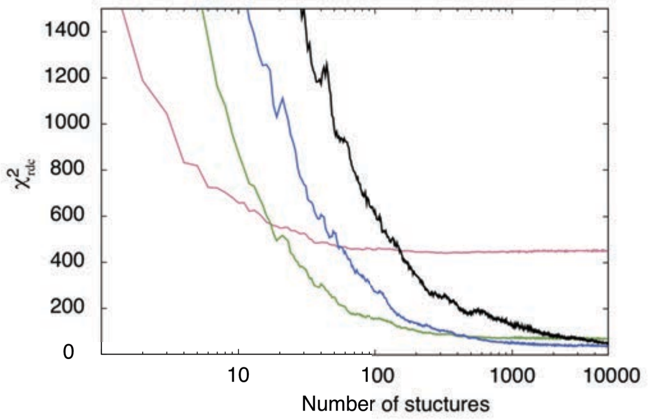

Figure 5: Reproduction of RDCs calculated using a global alignment tensor by RDCs calculated using LAWs of different sizes: three (pink), nine (green), 15 (blue) and 25 amino acids (black).

3.3. Testing ASTEROIDS on simulated RDC data

Initially, the performance of ASTEROIDS for determining the amino acid specific conformational sampling of ubiquitin was tested using simulated data. Thus, a synthetic data set was generated consisting of different types of ensemble-averaged RDCs over 50K Flexible-Meccano random coil conformers calculated using a global alignment tensor and a standard statistical coil sampling. Figure 6 shows the reproduction of the synthetic data set for ASTEROIDS-selected ensembles comprising 20 and 200 structures, respectively, employing LAWs of 15 amino acids in length. Both selected ensembles reproduce well the simulated data, however, a significant difference is observed in how well the site-specific conformational sampling of ubiquitin is captured by the selected ensembles (Figure 7). Thus, 200 structures perform significantly better than 20 structures showing that a minimum number of conformers are necessary in the selected ensemble to guarantee accurate determination of the site-specific conformational sampling.

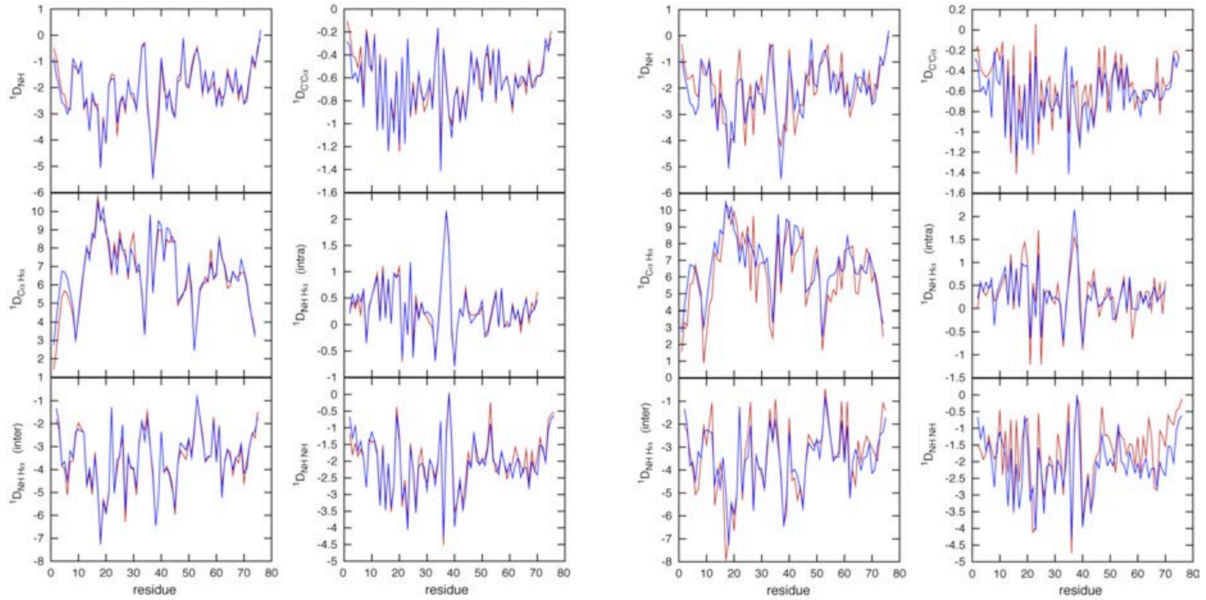

Figure 6: Comparison of synthetic RDC data of ubiquitin (blue) and back-calculated RDCs using 15 amino acid LAWs (red) from an ASTEROIDS-selected ensemble comprising 200 (left) and 20 (right) conformers.

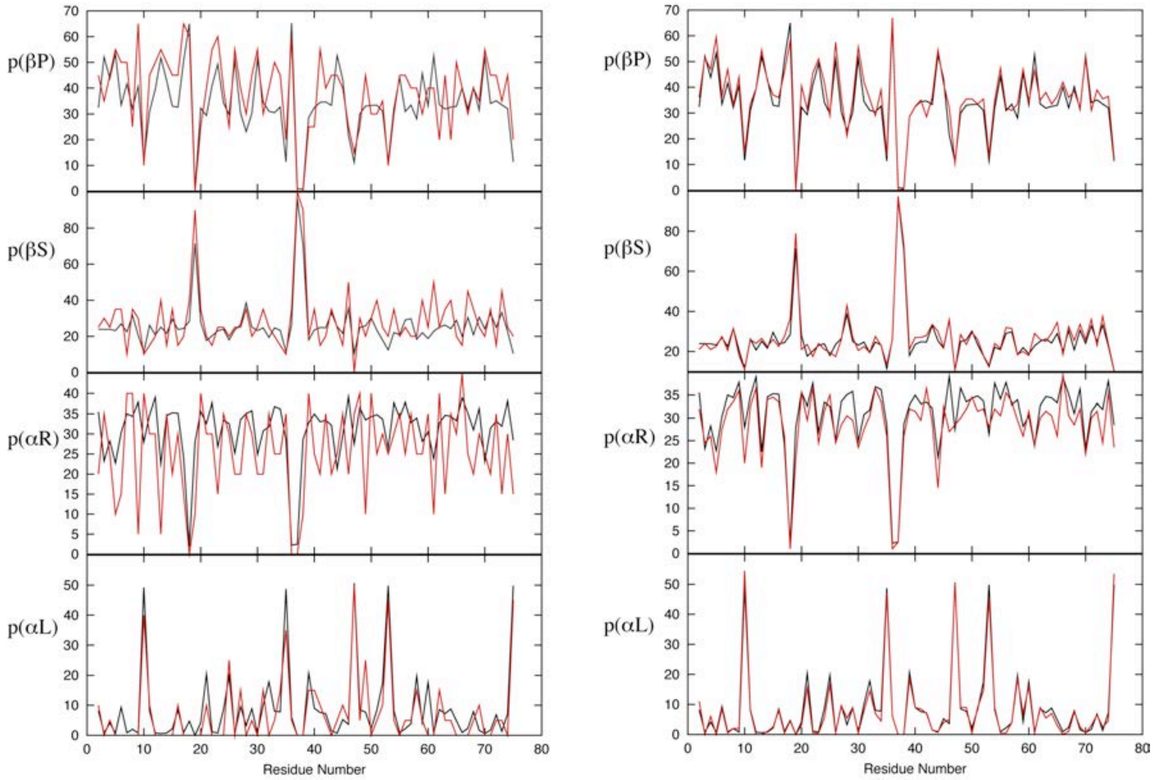

Figure 7: Reproduction of site-specific conformational sampling by ensembles selected using ASTEROIDS targeting synthetic RDC data of ubiquitin. The population in different regions of Ramachandran space is used as a metric of the accuracy of the determined conformational sampling. The regions are defined as follows: αL {$\phi > 0°$}; αR {$\phi < 0$, $-120° < \psi < 50°$}; βP {$-90° < \phi < 0°$, $\psi > 50°$ or $\psi < -120°$}; βS {$-180° < \phi < -90°$, $\psi > 50°$ or $\psi < -120°$}. The population of these quadrants is denoted p(αL), p(αR), p(βP) and p(βS), respectively. Reproduction of the conformational sampling of the synthetic data (black) is shown for two different ASTEROIDS ensembles comprising 20 (red, left) and 200 (red, right) conformers.

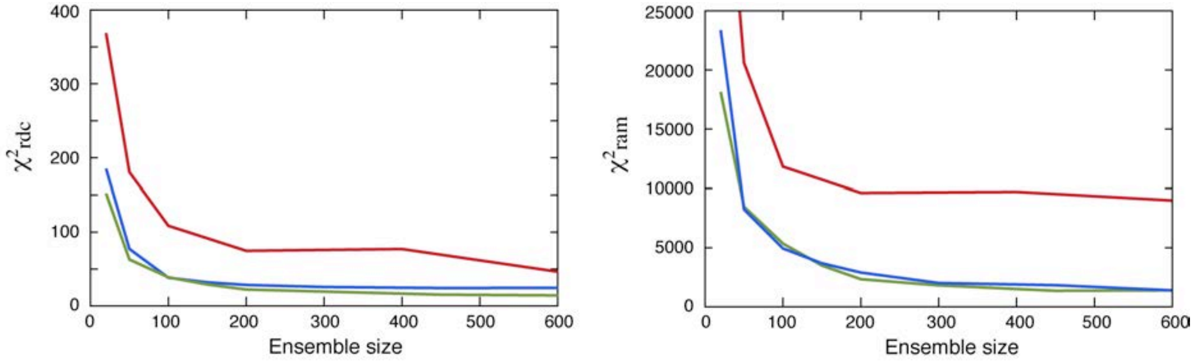

Figure 8: Accuracy of the reproduction of the synthetic RDC data of ubiquitin (left) and the site-specific conformational sampling (right) by ASTERIODS ensembles comprising an increasing number of conformers. RDCs were calculated using a global alignment tensor (red), nine amino acid LAWs (green) or 15 amino acid LAWs (blue).

To determine the appropriate number of conformers in the ASTEROIDS-selected ensembles of ubiquitin, we carried out selections for different ensemble sizes and monitored the reproduction of the RDCs and the site-specific conformational sampling (Figure 8). For a protein of the size of ubiquitin, it appears that around 200 structures are necessary to accurately reproduce the site-specific conformational sampling. In addition, it is noteworthy that selected sub-ensembles, where RDCs are obtained using a global alignment tensor, completely fails to reproduce the conformational sampling demonstrating the importance of employing LAWs for the calculation of RDCs (Figure 8).

3.4. Applying ASTEROIDS to experimental RDCs in urea-denatured ubiquitin

The ASTEROIDS protocol was then applied to data from ubiquitin, providing good agreement between the experimental RDCs and those back-calculated from the selected ASTEROIDS ensemble comprising 200 conformers (Figure 9). In order to check the validity of the selected ensemble, cross-validation was carried out i.e. 10% of the experimental RDC data were left out of the ASTEROIDS selection and subsequently back-calculated from the resulting ensemble. Reasonable agreement between back-calculated and experimental RDCs is found, and the cross-validation procedure shows that the 200-fold ensemble size is within the range where the cross-validation target function is essentially flat (Figure 10).

The site-specific conformational sampling of urea-denatured ubiquitin derived from the selected ASTEROIDS ensemble shows, in agreement with a hypothesis-driven analysis of the same RDCs (*71*), that ubiquitin samples on average more extended conformations of Ramachandran space (Figure 11). Close inspection of the sampling reveals that the residues most affected by the presence of urea are threonines, glutamic acids and arginines which all contain potential hydrogen-bond donor moieties in their side chains, while only very few hydrophobic residues display a significant difference in sampling compared to the standard coil library. A recent study using vibrational spectroscopy demonstrated that at low pH urea orients with the carboxyl group pointing towards the protein surface (*75*), an observation that supports the suggestion that hydrogen-bond donor groups may interact preferentially with urea as a step in the protein denaturation process.

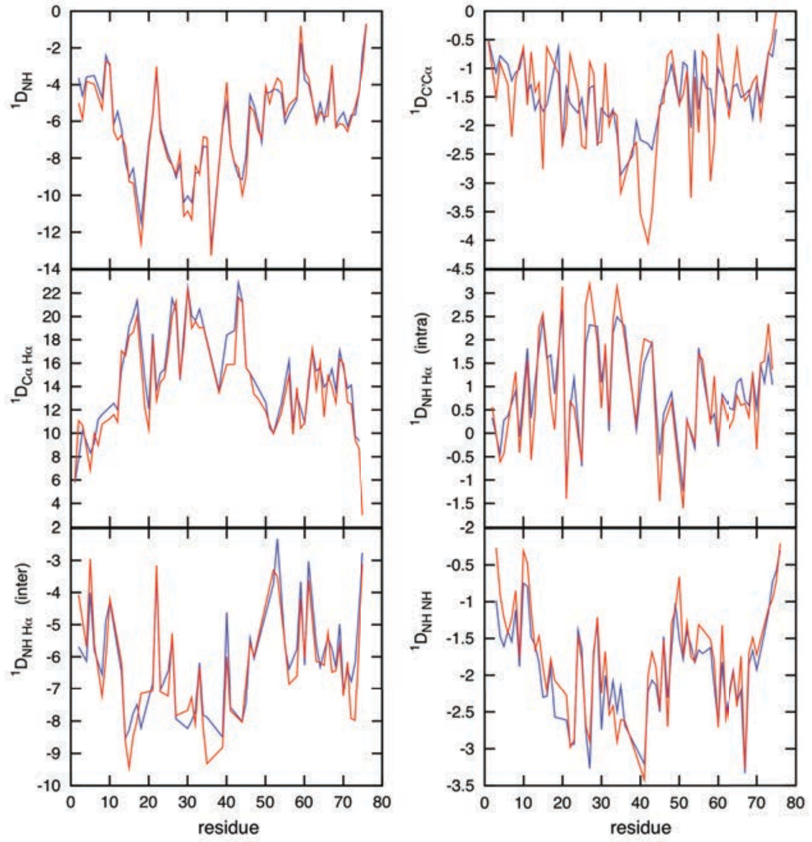

Figure 9: Comparison of experimental RDCs in urea-denatured ubiquitin (blue) and the RDCs back-calculated from the selected ASTEROIDS ensemble comprising 200 conformers (red). Two different scaling factors were applied to covalently bound and proton-proton RDCs, respectively.

Our results suggest that at low pH, urea binds to ubiquitin and thereby forces a more extended sampling of the backbone for certain residues. In this context, it is interesting to determine the effective binding constant of urea to the protein. ^{1}H-^{15}N HSQC experiments carried out on ubiquitin at pH 2.5 at different urea concentrations show that ubiquitin unfolds at urea concentrations above 2 M. In addition, systematic chemical shift changes are observed between 2 M and 8 M urea according to a fast-exchange regime (J.-R. Huang and S. Grzesiek, unpublished results). The chemical shift changes in principle allow a site-specific determination of the urea binding constant to unfolded ubiquitin, however, an analysis of the data shows that ubiquitin is far from saturated at 8M urea (the chemical shifts depend linearly on the urea concentration between 2 and 8 M). An accurate determination of the binding constant is therefore practically impossible using these titrations.

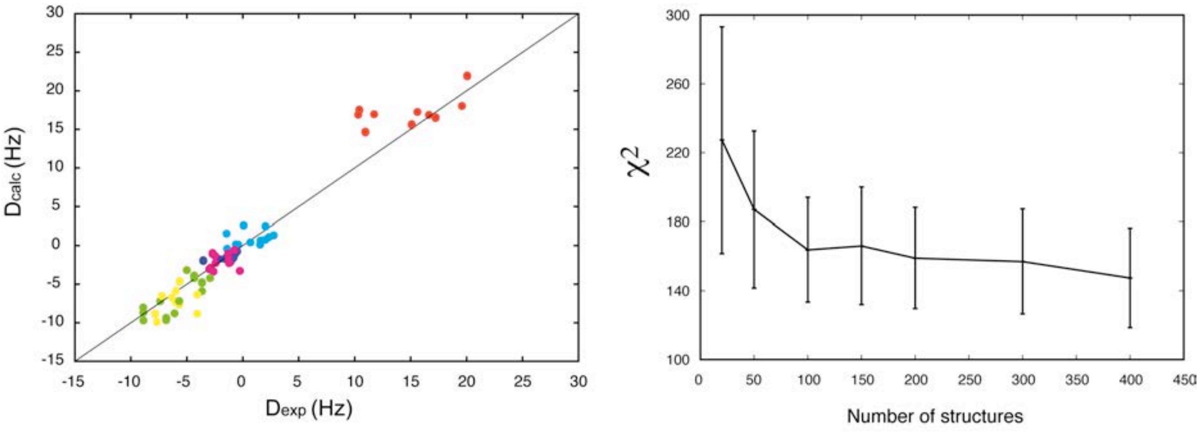

Figure 10: Reproduction (cross-validation) of experimental RDCs not used in the ASTEROIDS selection. (Left) The ASTEROIDS ensemble selection procedure was repeated, taking 10% of the RDCs out of the analysis and comparing the predicted values from the resulting ensemble with the experimental RDCs. Color code: ^{15}N-^{1}H^{N} (green), ^{13}Cα-^{1}Hα (red), ^{13}Cα-^{13}C' (dark blue), ^{1}H^{N}-^{1}Hα (cyan), ^{1}H^{N}-^{1}Hα(i-1) (yellow), ^{1}H^{N}-^{1}H^{N}(i+1) (magenta). (Right) Average $\chi2$ over ten cross-validations as function of the number of structures in the selected ensemble.

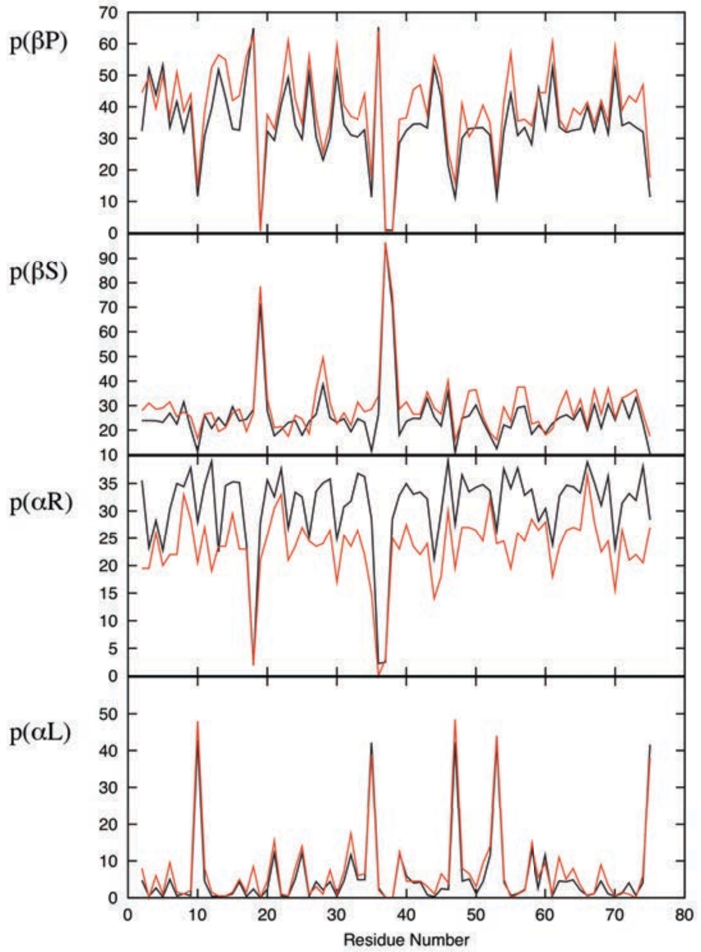

Figure 11: Comparison of the site-specific conformational sampling in the standard coil library (black) and in the ASTEROIDS selected ensemble (200 conformers) of urea-denatured ubiquitin on the basis of experimental RDCs (red).

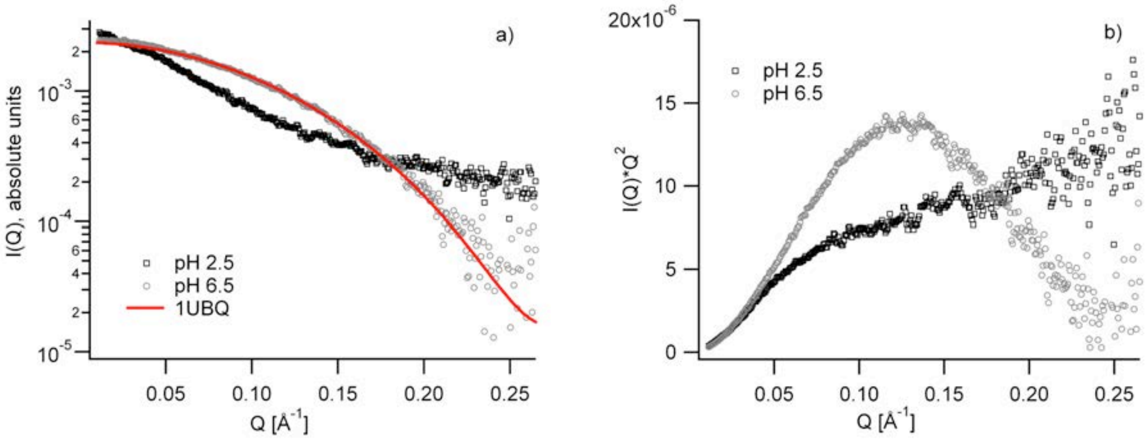

Figure 12: (a) SAXS scattering curves of ubiquitin in 8 M urea/H_2O at pH 6.5 and 2.5. The red line represents the calculated scattering curve of ubiquitin (PDB: 1UBQ) using CRYSOL. (b) Kratky plot of the same data showing that ubiquitin is folded at pH 6.5 and undergoes unfolding, when the pH is lowered to pH 2.5.

To obtain further insight into the denaturation process of ubiquitin we have employed a powerful combination of small angle X-ray (SAXS) scattering and small angle neutron scattering (SANS) (*76*). SAXS and SANS data obtained in 8 M urea at pH 6.5 and 2.5 in H_2O and D_2O (8 data sets) show that ubiquitin is folded at pH 6.5 and unfolds when the pH is lowered to 2.5 (Figure 12). By comparing the coherent intensities scattered at zero angle from SAXS and SANS and exploiting the different scattering densities of H_2O, D_2O, ubiquitin and urea for X-rays and neutrons, the number of urea molecules, N_{urea}, that are preferentially recruited during the unfolding transition from neutral to acidic pH was determined to be around 20 molecules (*76*). From these results, the apparent binding constant, K_b, of urea to ubiquitin can be calculated assuming that all binding sites are independent and have the same binding constant:

$$K_b = \frac{N_{\text{urea}}}{c_{\text{urea}}(k - N_{\text{urea}})} \qquad \text{(Eq. 5)}$$

Here, c_{urea} is the urea concentration and k is the total number of binding sites in ubiquitin (assuming binding to backbone amide groups, k=72 for ubiquitin). From Eq. 5 an apparent binding constant of 0.05 M^{-1} is obtained – an estimate that is in reasonable agreement with the determination of the same binding constant from NMR (0.1 - 0.3 M^{-1}) (*77*, *78*), fluorescence (0.08 - 0.3M^{-1}) (*79*), calorimetry (0.04 - 0.08 M^{-1}) (*80*, *81*) and end-to-end diffusion (0.3 M^{-1}) (*82*).

3.5. Conclusions

An accurate description of IDPs implies an atomic resolution determination of local conformational propensities. Here, we developed and tested a sample-and-select approach, ASTEROIDS, designed to map site-specific conformational sampling directly from different types of RDCs. In this context, we demonstrated the importance of determining the appropriate number of conformers in the selected sub-ensembles and of carefully considering the convergence characteristics of the RDCs. Importantly, we showed that if these aspects are not taken into account, an erroneous conformational sampling could be derived.

We applied ASTEROIDS to select a sub-ensemble of conformers of the protein ubiquitin under urea-denaturing conditions, for which a large number of experimental RDCs have been measured. An analysis of the selected ensemble reveals that urea-denatured ubiquitin samples on average more extended regions of Ramachandran space compared to the standard random coil database. This supports the model whereby urea binds to the backbone of ubiquitin as a step in the denaturation process. This hypothesis was further substantiated by small angle scattering measurements of ubiquitin showing that approximately 20 urea molecules are recruited to the protein during the unfolding transition from neutral to acidic pH.

4. Probing long-range order in IDPs using paramagnetic relaxation

4.1. Introduction

A coherent picture of the conformational behaviour of IDPs and partially folded proteins requires not only a mapping of local structure but also long-range order. Long-range interactions in IDPs are often transient in nature and their detection, therefore, requires a strong probe that is active over a few nanometers such as that provided by an unpaired electron. One of the most efficient ways of introducing an unpaired electron is by attaching an MTSL spin label to the protein through a cysteine residue. The dipolar interaction between the unpaired electron and the protein nuclei induces paramagnetic relaxation enhancements (PREs) that depend strongly on the electron-nucleus distances. By introducing spin labels at several different positions in the protein, a mapping becomes possible of long-range interactions in the disordered state.

This chapter describes the use of ASTEROIDS for interpreting experimental PREs in terms of molecular ensembles in order to probe long-range order present in disordered states. In this context, we demonstrate the importance of taking into account the motion of the MTSL side chain in calculations of the relaxation enhancements. We use the resulting ASTEROIDS ensembles for predicting the effect that the long-range interactions has on RDC baselines and, thereby, provide a method for combining RDCs and PREs in a single ensemble description to obtain both local and long-range information in disordered proteins.

4.2. Calculating relaxation enhancements in disordered proteins

The transverse relaxation enhancement, Γ_2, arising from the dipolar interaction between an unpaired electron and a proton is given by (*52*):

$$\Gamma_2 = \frac{1}{15}\left(\frac{\mu_0}{4\pi}\right)^2 \gamma_H^2 g_e^2 \mu_B^2 S(S+1)\left[4J(0)+3J(\omega_H)\right] \qquad \text{(Eq. 6)}$$

Here, μ_0 is the permittivity of free space, γ_H is the gyromagnetic ratio of the proton, g_e is the electron *g*-factor, μ_B is the Bohr magneton and *S* is the electron spin quantum number. Normally, PREs are calculated over structural ensembles by considering a fixed position of the MTSL side chain (for example on the Cβ atom of the cysteine) and by invoking the spectral density function:

$$J(\omega) = r_{\text{H-e}}^{-6}\left[\frac{\tau_c}{1+\omega^2\tau_c^2}\right] \qquad \text{(Eq. 7)}$$

Here, $r_{\text{H-e}}$ is the distance between the electron and the proton, and the correlation time is given by:

$$\frac{1}{\tau_c} = \frac{1}{\tau_r} + \frac{1}{\tau_s} \qquad \text{(Eq. 8)}$$

where τ_r and τ_s are the rotational and electron spin correlation times, respectively. By assuming that the inter-conversion rate between members of the ensemble is much slower than the correlation time of the dipolar interaction, the PREs can be calculated for each conformer independently and subsequently averaged over an ensemble composed of *N* structures:

$$\Gamma_2^{\text{calc}} = \frac{1}{N}\sum_{k=1}^{N}\Gamma_2^{k} \qquad \text{(Eq. 9)}$$

Here, we extent the calculation of PREs in inherently dynamic systems to include motion of the MTSL spin label for each member of the ensemble. We assume that the motion of the MTSL side chain occurs on a time scale that is much faster than the exchange between individual conformers of the ensemble. We invoke the motion of the MTSL spin label for each Flexible-Meccano conformer by sampling different states of the side chain using known rotamer libraries (*83*) (Figure 13). It is seen that the MTSL side chain covers a rather large volume space, and it is expected that this motion will contribute significantly to measured PREs in disordered states.

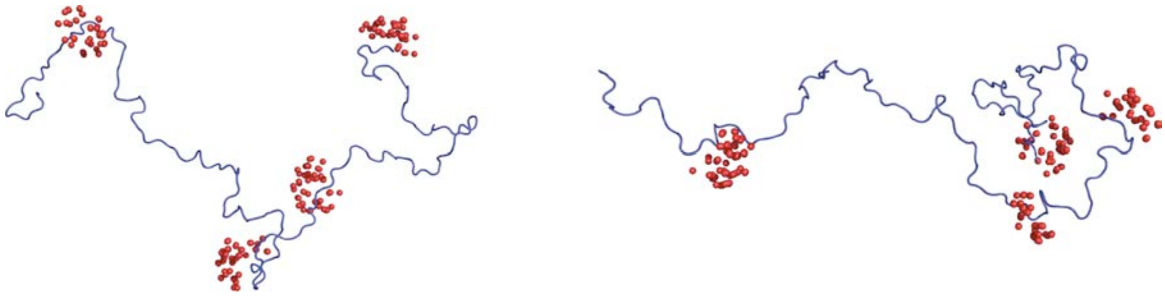

Figure 13: Two Flexible-Meccano conformers (blue) of α-synuclein with distributions of the side chain MTSL spin labels (red) for four different spin label positions (residues 18, 76, 90 and 140). Previously proposed MTSL rotameric libraries were randomly sampled for a total of 600 conformers for each site. Each position was retained and included in the averaging procedure, if no steric clashes were found with the given backbone conformation.

To take into account the motion of the spin label in the calculation of PREs, we apply the spectral density function originally proposed for calculating ^{1}H-^{1}H cross-relaxation rates in the presence of rapid motion (*84*) and later adapted to calculate PREs in proteins with flexible paramagnetic tags (*85*):

$$J(\omega) = \left\langle r_{\text{H-e}}^{-6} \right\rangle \left[\frac{S_{\text{H-e}}^2 \tau_\text{c}}{1+\omega^2\tau_\text{c}^2} + \frac{(1 - S_{\text{H-e}}^2)\tau_\text{e}}{1+\omega^2\tau_\text{e}^2} \right] \qquad \text{(Eq. 10)}$$

where

$$\frac{1}{\tau_\text{e}} = \frac{1}{\tau_\text{r}} + \frac{1}{\tau_\text{s}} + \frac{1}{\tau_\text{i}} \qquad \text{(Eq. 11)}$$

and τ_i represents the effective correlation time of the spin label. The order parameter, $S_{\text{H-e}}^2$, that depends on both the orientation and length of the electron-nucleus vector, can be calculated for each conformer using (*84*):

$$S_{\text{H-e}}^2 = \frac{4\pi}{5} \left\langle r_{\text{H-e}}^{-6} \right\rangle^{-1} \left\langle r_{\text{H-e}}^{-3} \right\rangle^{2} \sum_{m=-2}^{2} \left| \left\langle Y_2^m(\Omega^{\text{mol}}) \right\rangle \right|^2 \qquad \text{(Eq. 12)}$$

Here, $Y_2^m(\Omega^{\text{mol}})$ are the second-order spherical harmonics and Ω^{mol} are the Euler angles in the molecular frame describing the orientation of the interaction vector. In agreement with previous studies, we assume correlation times of 5 ns for τ_c (*54*, *56*), 500 ps for τ_i (*86*) and 10^{-7} s for τ_s (*87*). Experimentally, PREs are often measured as the ratio of the intensities in the ^{1}H-^{15}N HSQC spectrum of the protein carrying the MTSL spin label in its oxidized and reduced state, respectively. In this case, the intensity ratio can be estimated from the calculated relaxation enhancements for each residue using (*88*):

$$\frac{I}{I^0} = \frac{\Gamma_2^{\text{red}} \exp(-\Gamma_2^{\text{calc}} \tau_{\text{mix}})}{\Gamma_2^{\text{red}} + \Gamma_2^{\text{calc}}} \qquad \text{(Eq. 13)}$$

Here, Γ_2^{red} is the intrinsic relaxation rate of the amide proton and τ_{mix} is the mixing time of 10 ms during which relaxation occurs in the HSQC pulse sequence.

4.3. Testing ASTEROIDS on simulated PRE data

In order to test the performance of ASTEROIDS for mapping long-range interactions in disordered states on the basis of PREs, we initially generated two sets of synthetic PRE data from ensembles of a model protein of 100 amino acids with known long-range contacts. We define a contact between two contiguous strands of the protein, if any Cβ atom in the first strand is less than 15 Å from any Cβ atom in the second strand. We generated synthetic PRE data sets from two ensembles with persistent long-range contacts between the strands 41-50 and 81-90 as well as between 11-20 and

61-70. The motion of the MTSL spin label was taken into account as described above and PRE data were calculated from four different positions of the spin label (residues 20, 40, 60 and 80).

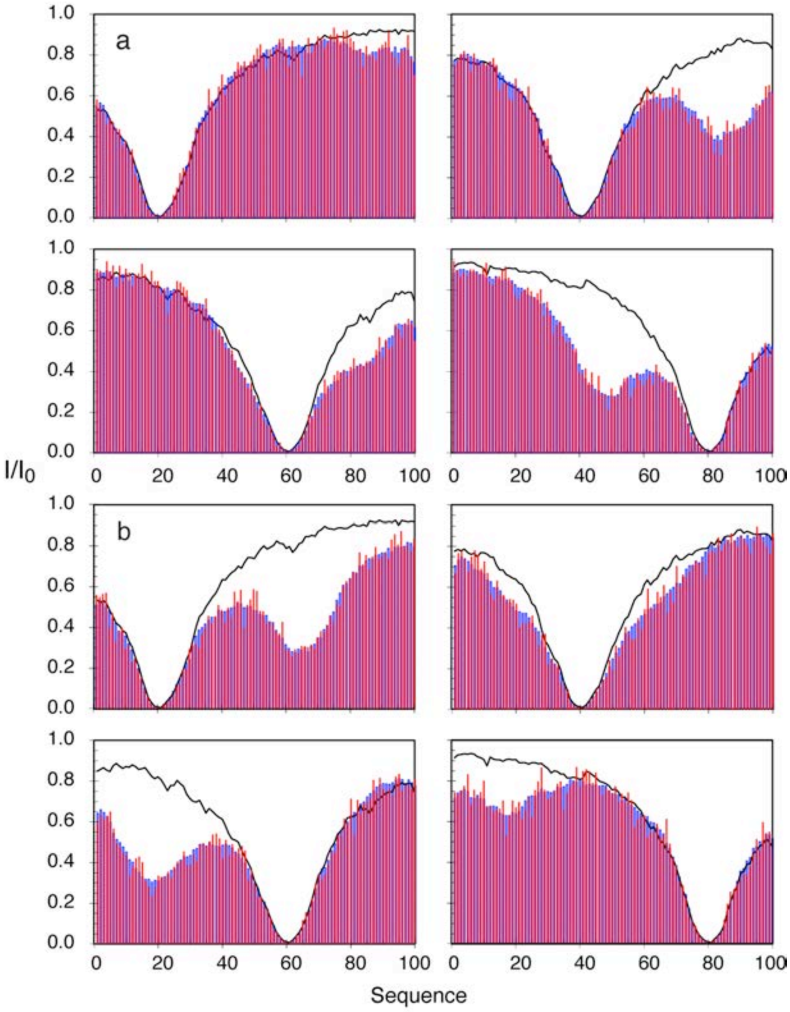

Figure 14: Reproduction of synthetic PRE data sets (blue) with ASTEROIDS-selected ensembles comprising 80 conformers (red). (a) The synthetic data set was obtained from an ensemble carrying a persistent long-range contact between the regions 41-50 and 81-90. (b) The synthetic data set was obtained from an ensemble carrying a persistent long-range contact between the regions 11-20 and 61-70. The black lines correspond to the calculated PRE profile from an ensemble carrying no specific contacts. In both (a) and (b) the spin label was located at residues 20, 40, 60 and 80.

From a large pool (10K) of Flexible-Meccano conformers carrying no specific contacts, ASTEROIDS was used to select sub-ensembles comprising 80 structures in agreement with the synthetic PRE data sets. The ASTEROIDS-selected ensembles nicely reproduce the simulated PRE data (Figure 14). Long-range contacts within the selected ensembles were identified using the metric Δ_{ij} that compares the distance distribution of the selected ensemble with that of the reference ensemble carrying no specific long-range contacts:

$$\Delta_{ij} = -\log\left(\left\langle d_{ij}\right\rangle / \left\langle d_{ij}^{0}\right\rangle\right) \qquad \text{(Eq. 14)}$$

Here, d_{ij} is the distance between residues i and j in the ASTEROIDS ensemble, while d_{ij}^{0} is the corresponding distance in the reference ensemble. The results show that ASTEROIDS successfully locates the long-range contacts present in the ensembles used to create the synthetic PRE data sets (Figure 15).

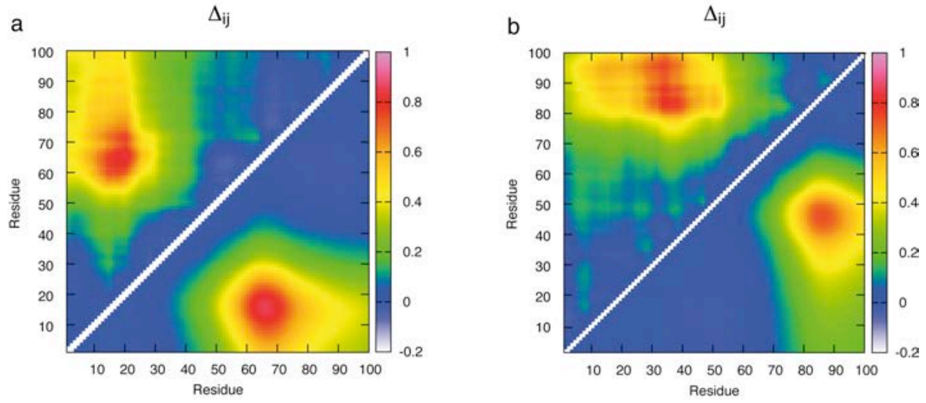

Figure 15: Contact maps showing chain proximity (Eq. 14) in the ASTEROIDS-selected ensembles on the basis of the synthetic PRE data shown in Figure 14 (above the diagonal) in comparison with target ensembles (below the diagonal). The long-range contact is between 11-20 and 61-70 (a) and between 41-50 and 81-90 (b). The scale for the data above the diagonal in each panel has been multiplied by a factor of 0.50 for ease of identification of the contact.

4.4. Combining PREs and RDCs in a single ensemble description

Ideally, PREs and RDCs would be combined in a single ensemble description to probe simultaneously local and long-range structure in the disordered state. In order to investigate the effect of long-range order on RDC profiles, we carried out Flexible-Meccano simulations of poly-valine chains of 100 amino acids with persistent long-range contacts between different parts of the chain (Figure 16). The simulations show that the presence of long-range contacts strongly modulates the RDC baselines, with RDCs being reinforced in the contacting regions and quenched in regions between the two contacts. The parameterization obtained of the RDC baseline in Eq. 4, therefore, needs to be modified in the case of persistent long-range order within the ensemble. Based on the simulations shown in Figure 16, we propose an empirical expression that combines the baseline with no specific contacts (Eq. 4) with Gaussian curves that take into account the modulation of the RDC baseline by the contact (red curves in Figure 16). Thus, for any length of protein and any persistent long-range contact the expected RDC baseline can be calculated analytically (see reference (*49*) for full analytical expression).

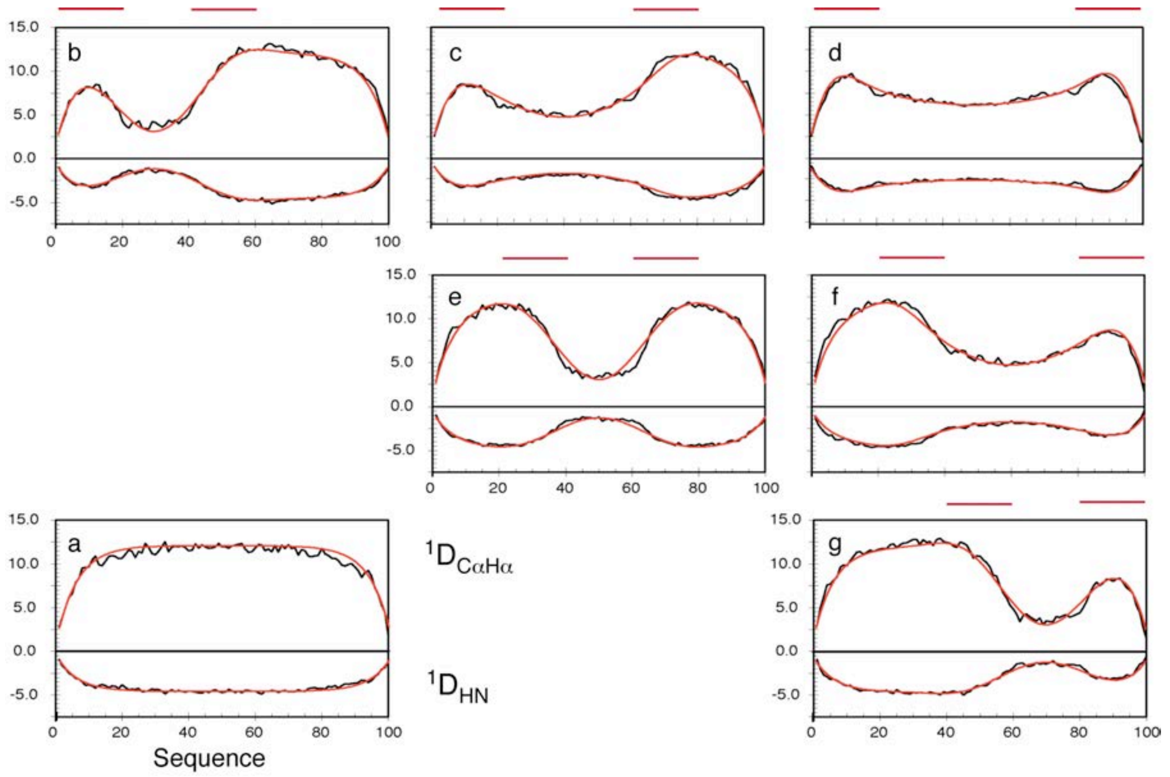

Figure 16: Simulations of ^{15}N-^{1}H^N and ^{13}Cα-^{1}Hα RDCs (black lines) in a poly-valine chain of 100 amino acids in the absence (a) and the presence (b-g) of different persistent long-range contacts: contact between 1-20 and 41-60 (b), 1-20 and 61-80 (c), 1-20 and 81-100 (d), 21-40 and 61-80 (e), 21-40 and 81-100 (f), 41-60 and 81-100 (g). RDCs were averaged over 100K conformers. The red lines correspond to the parameterization of the baseline with the contact positioned in the centre of each region.

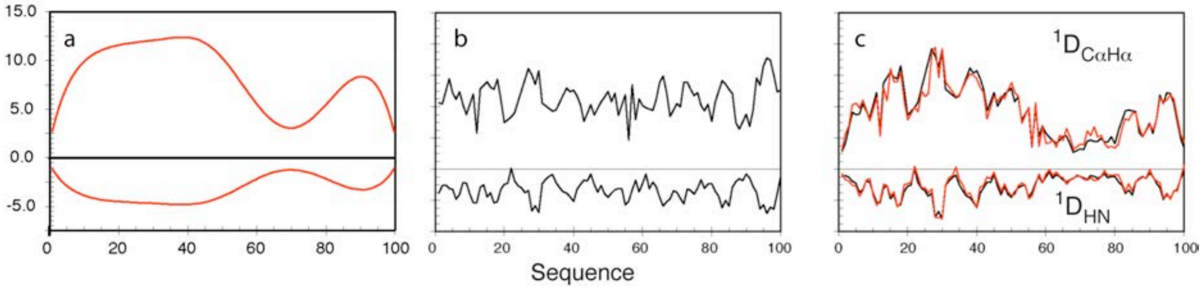

Figure 17: Example of the combination of analytically calculated baselines and RDCs averaged using the LAW approach for a model protein of 100 amino acids of arbitrary sequence. (a) Baseline contribution calculated analytically for contacts between regions centered on residue 50 and 90. (b) RDCs calculated using the LAW approach with windows of 15 amino acids in length. RDCs were averaged over 200 structures. (c) Combination of the baseline from (a) and the local RDCs from (b) (red curves) compared to the RDCs calculated using a global alignment tensor over 100K structures carrying a contact between 41-60 and 81-100 (black curves).

The combination of PREs and RDCs in a single ensemble description requires that both parameters converge over approximately the same number of structures. As described in the previous chapter, the combination (multiplication) of the empirical RDC baseline with RDCs calculated using the LAW approach dramatically reduces the number of structures needed to achieve convergence of the RDCs. As demonstrated in Figure 17, this approach can also be applied in the case of disordered ensembles carrying persistent long-range interactions. Thus, RDCs calculated using a global alignment tensor carrying a contact between the regions 41-60 and 81-100 are almost identical to the RDCs obtained by multiplication of the baseline corresponding to a contact between residues 50 and 90 with the RDCs calculated using the LAW approach (Figure 17).

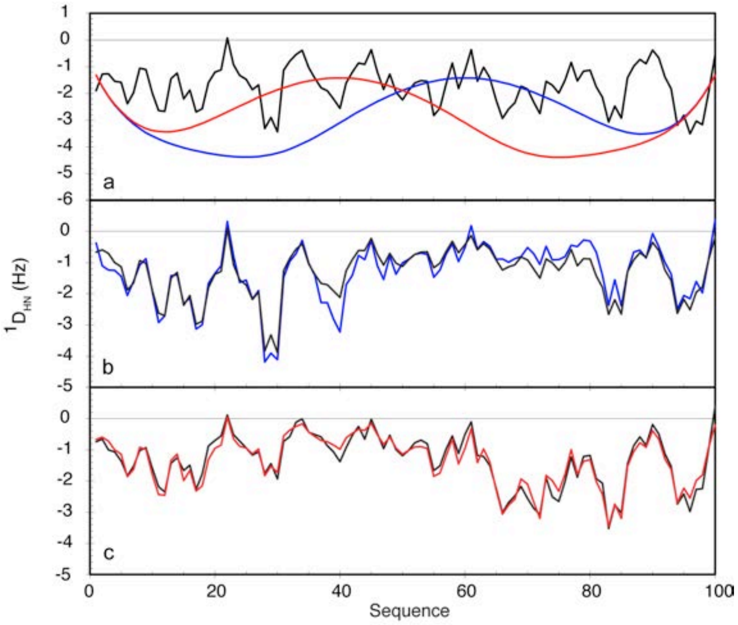

Figure 18: Combining PREs and RDCs in a single ensemble description in the context of simulated data. (a) Black: RDCs calculated using the LAW approach over an ensemble of 200 structures. Red: RDC baseline extracted from the contact map shown in Figure 15a (contact between 11-20 and 61-70). Blue: RDC baseline extracted from the contact map shown in Figure 15b (contact between 41-50 and 81-90). (b) Black: RDCs calculated using a global alignment tensor over an ensemble comprising 100K conformers containing a contact between 41-50 and 81-90. Blue: the combination of the RDCs calculated using the LAW approach and the blue baseline curve shown in (a). (c) Black: RDCs calculated using a global alignment tensor over an ensemble comprising 100K conformers containing a contact between 11-20 and 61-70. Red: combination of the RDCs calculated using the LAW approach and the red baseline curve shown in (a).

The strategy we propose for combining PREs and RDCs in a single ensemble description relies on the identification of long-range interactions using PREs, the calculation of the resulting baseline for the RDCs taking into account the long-range order, and finally the application of this baseline to RDCs calculated using the LAW approach. Although here described as a three-step process, the combination of PREs and RDCs can be done simultaneously during the evolution steps of the ASTEROIDS algorithm. We applied this strategy to the simulated PREs in Figure 14. The corresponding contact maps (Figure 15) were analyzed to determine the maximum of the difference between the PRE-derived ensembles and the reference ensembles containing no specific contacts. The RDC baselines were calculated for the two different simulations taking into account the PRE-derived long-range interactions. As shown in Figure 18, the combination of RDCs calculated using the LAW approach and baselines derived from the PRE contact maps nicely reproduces the RDCs calculated using a global alignment tensor for the ensembles used to obtain the synthetic PRE data sets. This shows that a decorrelation is possible of the local and long-range structure information inherent to RDCs allowing the combination of PREs and RDCs in a single ensemble description.

4.5. Application to PRE and RDC data of alpha-synuclein

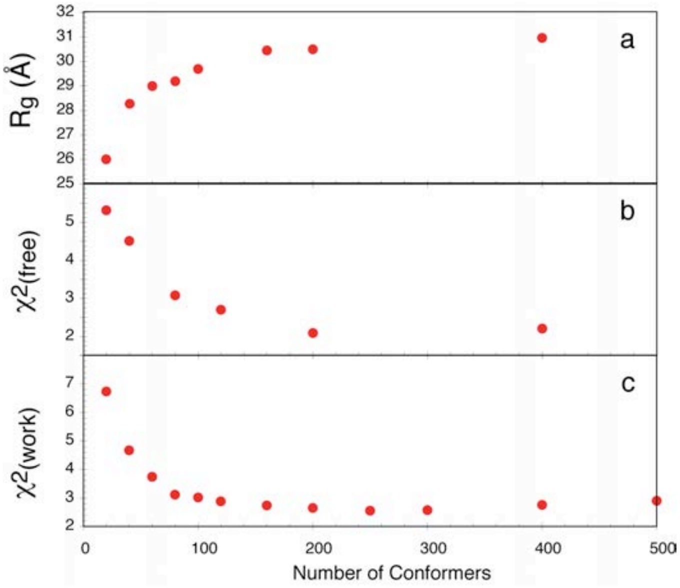

Figure 19: Selection of ensembles using ASTEROIDS on the basis of experimental PRE data of α-synuclein. ASTEROIDS ensemble characteristics are plotted as function of the selected ensemble size. (a) Average radius of gyration. (b) χ^2 for the "passive" data. The passive data in this case consist of the entire A76C data set. Only data from mutants A18C, A90C and A140C were used in the ensemble selection. (c) χ^2 for the "active" data. The active data in this case consist of PRE data from mutants A18C, A90C and A140C.

Following the successful application of ASTEROIDS to simulated PRE data sets, we targeted experimental PREs in α-synuclein. α-Synuclein is an IDP[1] of 140-amino acids found in human brain and strongly implicated in the onset of Parkinson's disease (*90*). The protein consists of three domains: the amphipathic N-terminus (residues 1-60), the hydrophobic self-aggregating non-Aβ component (NAC) (residues 61-95) and the acidic C-terminus (residues 96-140). PREs were previously measured for four different mutants A18C, A76C, A90C and A140C of α-synuclein with the aim of detecting long-range interactions in the protein (*54*). In order to determine the number of conformers necessary to describe the long-range interactions in α-synuclein, we carried out

---

[1] Recently, it was found that α-synuclein occurs physiologically as a folded, helical tetramer (*89*).

ASTEROIDS selections with different ensemble sizes. Ensembles containing too few structures are expected to be too compact, and the convergence of the radius of gyration towards a constant value can, therefore, be used to find the most appropriate ensemble size (Figure 19). In order to validate our selected ensembles, we also carried out ASTEROIDS selections on the basis of the PREs of A18C, A90C and A140C ("active" data) and used the PREs of A76C ("passive" data) for cross-validation. The convergence of the radius of gyration and the χ^2 of both the active and passive data indicate that 200 structures are appropriate to describe the experimental PREs of α-synuclein (Figure 19).

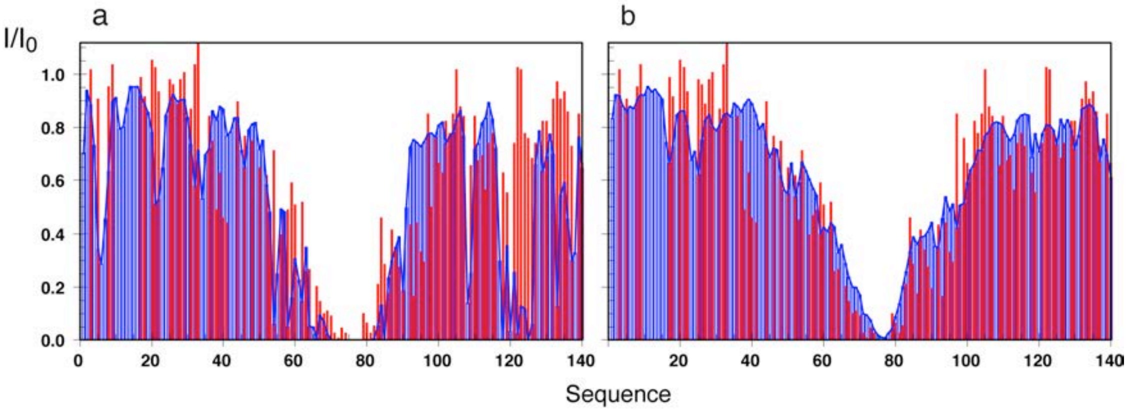

Figure 20: Reproduction of "passive" PREs of mutant A76C by an ASTEROIDS ensemble (200 structures) selected on the basis of PREs of A18C, A90C and A140C. The PREs were calculated using both a static (a) and dynamic (b) side chain of the MTSL spin label.

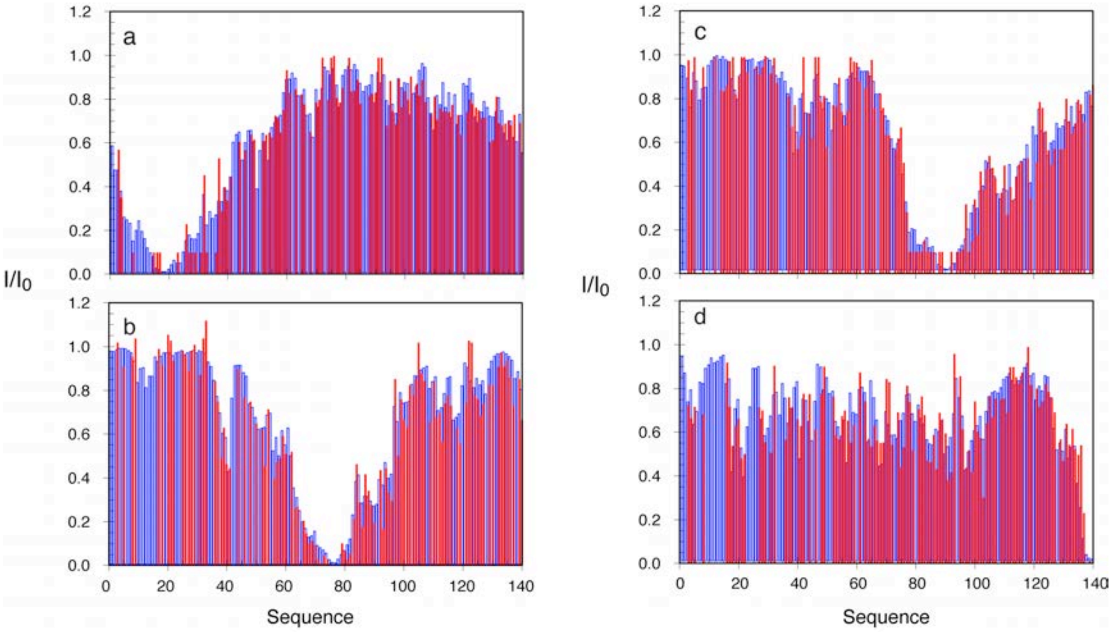

Figure 21: Agreement between experimental PREs (red) and back-calculated PREs (blue) from the selected ASTEROIDS ensemble comprising 200 structures. PREs are shown for the four different mutants A18C (a), A76C (b), A90C (c) and A140C (d). All experimental PREs were included in the ASTEROIDS selection.

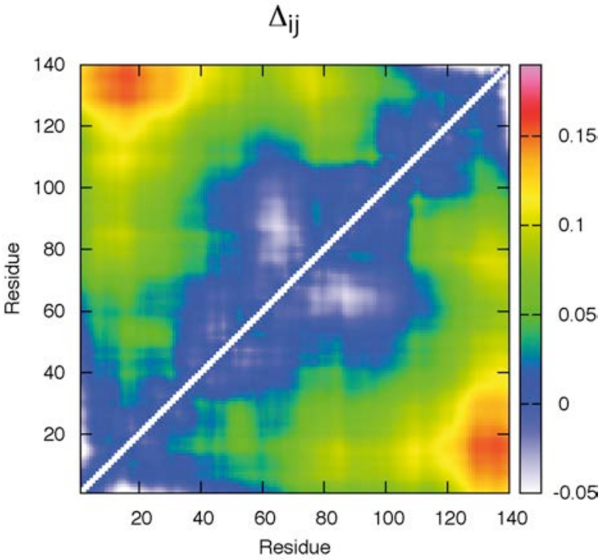

Figure 22: Contact map showing the long-range interactions in α-synuclein derived from the experimental PREs of the four mutants A18C, A76C, A90C and A140C.

In order to evaluate the effect on the selected ensembles of the imposed dynamics of the MTSL side chain, we carried out ASTEROIDS selections using both a static (Eq. 7) and a dynamic (Eq. 10) description of the MTSL side chain on the basis of the experimental PREs for the mutants A18C, A90C and A140C. The ensembles determined using both the static and the dynamic side-chain fit the experimental data from the three active mutants to within experimental uncertainty. A comparison of the back-calculated PREs from the selected ensembles to the experimental (passive) PREs of A76C, however, demonstrates that the dynamic side chain performs significantly better in the cross-validation than the static side chain (Figure 20).

An ASTEROIDS ensemble comprising 200 structures selected on the basis of the PREs from the four spin labels of α-synuclein nicely reproduces the experimental data (Figure 21) and, in agreement with previous studies of α-synuclein, the resulting contact map reveals a long-range interaction between the N- and C-terminus of the protein (Figure 22). This interaction has been proposed to protect the hydrophobic NAC domain from being exposed to the solvent and from subsequent fibrillation of the protein (*54*, *61*). As in the case of the simulated PRE data sets, we used the contact map to derive the RDC baseline corresponding to the most populated contact in the protein. The combination of this baseline with the RDCs calculated using the LAW approach significantly improves the reproduction of experimental RDCs in α-synuclein compared to an ensemble, where the presence of the long-range contact was not taken into account (Figure 23). This demonstrates that experimental RDCs are indeed affected by long-range contacts and an erroneous conformational sampling will inevitably be derived, if the RDCs are interpreted in terms of local structure only.

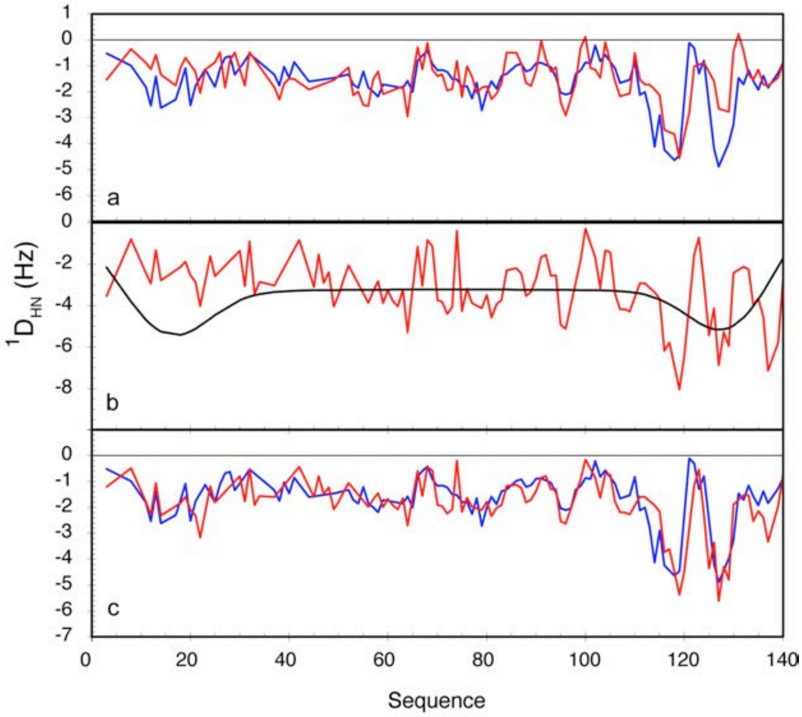

Figure 23: Combined analysis of PREs and RDCs in α-synuclein. (a) Comparison of experimental ^{15}N-$^1H^N$ RDCs obtained in PEG/hexanol liquid crystals (blue) with couplings calculated using a global alignment tensor over a Flexible-Meccano ensemble comprising 50K structures (red). The rmsd between the two sets of RDCs is 0.78 Hz. (b) RDCs predicted using the LAW approach (red) and effective baseline derived from the contact map shown in Figure 22 (black). (c) Combination of the curves shown in (b) (red) compared to the experimental ^{15}N-$^1H^N$ RDCs (blue). The rmsd between the two sets of RDCs is 0.52 Hz.

4.6. Conclusions

In this chapter, the use of ASTEROIDS was described for mapping long-range interactions in IDPs on the basis of extensive sets of PREs. The PREs were induced by MTSL spin labels attached at different positions in the protein, and a formalism was described for taking into account the motion of the MTSL side chain in the calculation of PREs from the structural ensembles. The approach was validated on synthetic PRE data sets and subsequently applied to experimental PREs in the IDP α-synuclein. Interestingly, a dynamic side chain of the MTSL spin label performed better than a static side chain in a cross-validation of “passive” PREs from the ASTEROIDS-selected ensemble of α-synuclein.

A mapping of both local and long-strange structure is essential for a complete description of IDPs, and the combination of PREs and RDCs is particularly attractive in this context. Here, we derived analytical expressions of the RDC baselines in the presence of long-range interactions allowing the combination of PREs and RDCs in a single ensemble description. We applied the approach to α-synuclein and showed that the inclusion of the PRE-derived long-range interaction between the N-and C-terminus of the protein improved the reproduction of experimental RDCs by a statistical coil ensemble.

One of the major challenges in characterizing long-range interactions in disordered proteins is the separation of the distance and population contributions to the measured PREs. Thus, a fully populated contact within a structural ensemble characterized by a long distance can in principle give rise to the same PRE profile as a less populated contact characterized by a short distance. The separation of populations and distances is only possible if enough spin labels are attached along the primary sequence of the protein. In the present analysis we have made no effort to quantify the different contacts determined by the ASTEROIDS analysis of PREs. Thus, the baselines that we apply correspond to a contact population of 100%. One way to overcome this problem in the future will be to apply a baseline for each member of the ASTEROIDS ensemble instead of applying a global baseline to the entire ensemble.

5. Defining conformational ensembles from chemical shifts

5.1. Introduction

Over the last decade, remarkable progress has been made in the prediction of chemical shifts from protein conformation (*91–95*) culminating recently with the successful structure determination of small proteins (< 150 amino acids) using only chemical shifts in combination with molecular mechanics force fields (*96*, *97*). Chemical shifts are very sensitive to backbone conformation and the possibility of exploiting this sensitivity to define conformational ensembles of IDPs is particularly attractive, especially considering that the determination of (backbone) chemical shifts is a prerequisite for any NMR study.

The advantage of chemical shifts as structural probes is that they depend in a different, complementary manner on the backbone dihedral angles and therefore, in principle, allows a site-specific mapping of the conformational sampling in disordered proteins. Here, we investigate this possibility by applying a combination of Flexible-Meccano, the state-of-the-art chemical shift prediction program SPARTA (*94*) and ASTEROIDS to target experimental chemical shifts in IDPs. The developed approach starts from a pool of random coil conformers and iteratively provides conformational ensembles with transiently populated secondary structures, if these are encoded in the experimental chemical shifts.

5.2. Chemical shifts and their dependence on backbone conformation

Chemical shift prediction using SPARTA relies on a database that contains 200 high-resolution crystal structures for which nearly complete sets of chemical shift assignments are available. SPARTA predicts chemical shifts by comparison of amino acid type, backbone and side chain conformation (φ, ψ and χ^1) of tri-peptides in the target protein to those of the SPARTA database. A scoring function is used to measure the similarity between the tri-peptides, and the predicted chemical shifts are obtained as averages over the 20 best performing tri-peptides in the SPARTA database. It is expected that prediction of chemical shifts of random coil conformers, such as those created by Flexible-Meccano, would largely be based on tri-peptides from the SPARTA database corresponding to loop regions of the crystal structures.

Initially, we investigated the information contents of the chemical shifts by performing calculations of an ensemble of small poly-alanine chains created by Flexible-Meccano. Figure 24 shows the distribution in Ramachandran space of the predicted chemical shifts by SPARTA of one of the alanine residues. While $^{13}C\alpha$, $^{13}C\beta$, $^{13}C'$ and $H\alpha$ chemical shifts depend strongly on φ and ψ, ^{15}N and $^{1}H^{N}$ chemical shifts show a more or less uniform dependence on the two dihedral angles. In addition, the $^{13}C\alpha$ and $^{13}C\beta$ chemical shifts display an almost inverse dependence on the φ/ψ distribution and therefore allows a precise determination of the populations of α-helix and β-sheet in the disordered state. This is the basis of the so-called SSP program developed by Forman-Kay and coworkers that converts $^{13}C\alpha$ and $^{13}C\beta$ chemical shifts into α– and β-secondary structure propensities (*26*).

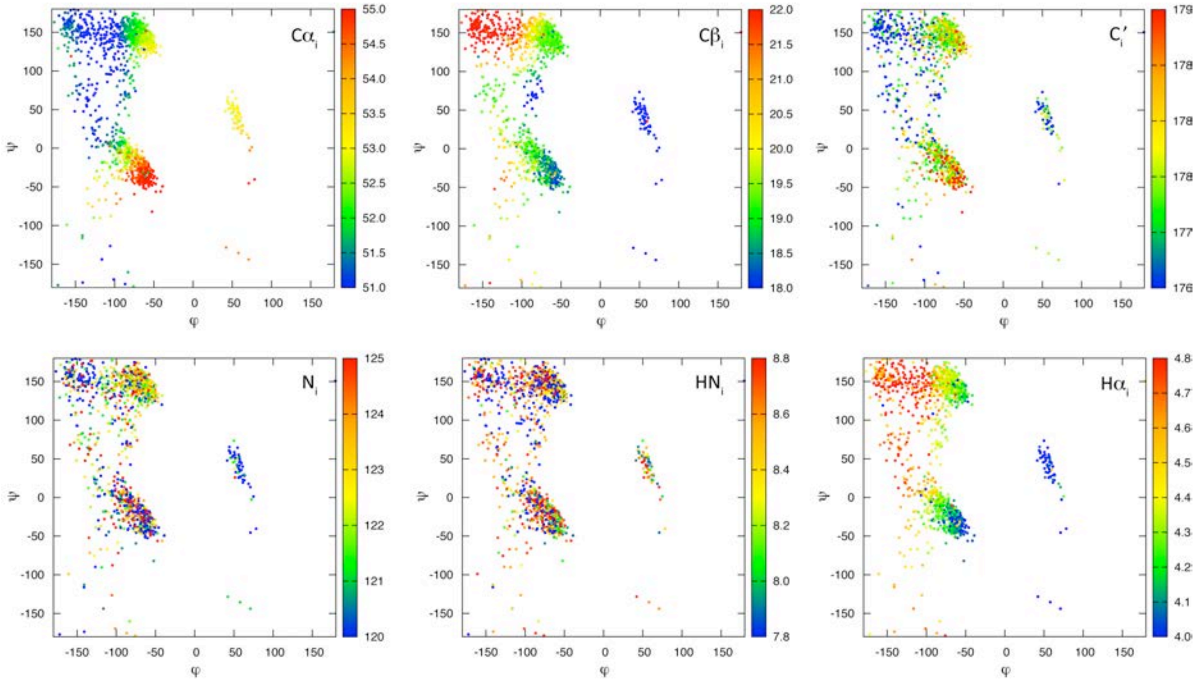

Figure 24: Distribution of predicted chemical shifts for the central residue *i* of an Ala-Ala_{i-1}-Ala_i-Ala_{i+1}-Ala sequence as function of the conformational sampling (φ and ψ) of residue *i*. The results for 2000 conformers are shown.

The determination of the populations in other regions of Ramachandran space is equally important. From complementary techniques such as vibrational spectroscopy and circular dichroism, it has been proposed that poly-proline II (PPII) is one of the dominant conformations in IDPs (*98–100*). A method that unequivocally maps the population of PPII conformations in disordered proteins on a residue specific basis is, therefore, of high interest. Almost equally sized chemical shifts are obtained in the PPII region and the upper left hand side of the α-helical region (we denote it the α'-helical region: $-125<\varphi<-75$ and $-50<\psi< 25$) making it difficult, on the basis of the chemical shifts shown in Figure 24, to map the precise populations in these regions. To a certain extent, this degeneracy can be overcome by considering the influence of the φ/ψ sampling on the chemical shifts of the neighbouring amino acids (Figure 25). In principle, $^{13}C'$ chemical shifts of the preceding residue, and especially ^{15}N and $^{1}H^{N}$ chemical shifts of the following residue allow to separate the populations in the PPII and the α'-helical region. A protocol that combines Flexible-Meccano and

ASTEROIDS to obtain site-specific conformational sampling on the basis of experimental chemical shifts, therefore, needs to take into account the influence of the φ/ψ sampling of one residue on the chemical shifts of the neighbouring residues.

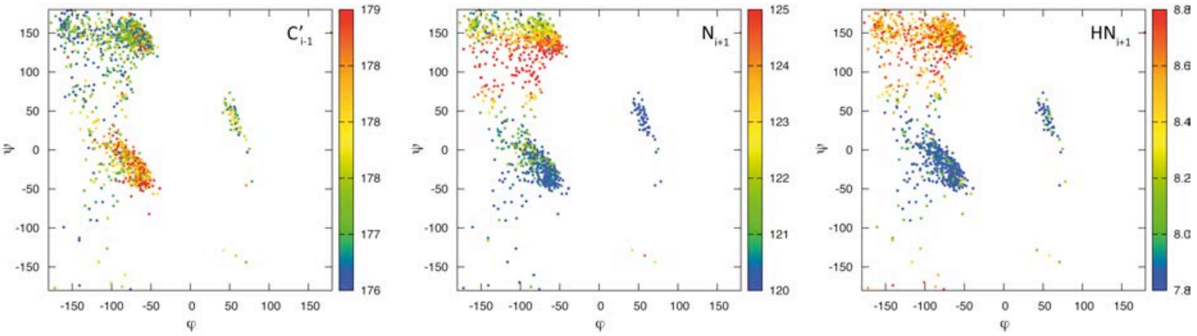

Figure 25: Predicted chemical shifts in the neighbouring amino acids (*i*-1 and *i*+1) of an Ala-Ala$_{i-1}$-Ala$_i$-Ala$_{i+1}$-Ala sequence as function of the conformational sampling (φ and ψ) of residue *i*. The results for 2000 conformers are shown. Only ^{13}C′ chemical shifts of residue *i*-1 (left), ^{15}N chemical shifts of residue *i*+1 (middle) and ^{1}H^{N} chemical shifts of residue *i*+1 (right) show a significant dependence on the conformational sampling of residue *i*. Other types of chemical shifts in the neighbouring amino acids display a more or less uniform dependence on φ and ψ.

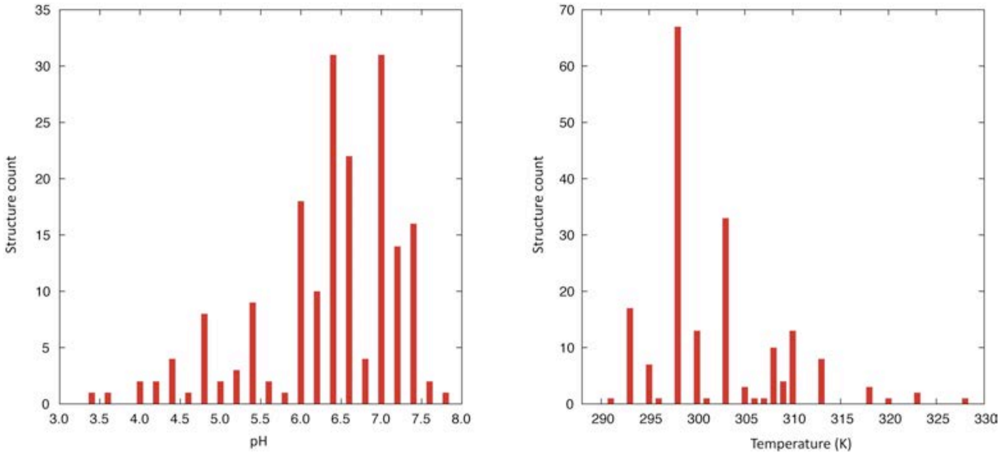

Figure 26: Distributions of pH and temperature in the NMR assignment experiments of the proteins constituting the SPARTA database.

A point to consider when using chemical shifts to map conformational sampling in disordered proteins is the experimental conditions under which the chemical shifts were measured. It is well known that chemical shifts depend on parameters such as pH and temperature, carbon chemical shifts being less dependent than N and HN chemical shifts (*101*). Figure 26 shows the distribution of pH and temperature used in the assignment experiments of the proteins constituting the SPARTA database. Thus, it is expected that prediction of chemical shifts using SPARTA will be most accurate close to neutral pH and around room temperature – excellent conditions for studies of IDPs by NMR. SPARTA holds less promise, however, for prediction of chemical shifts in pH-denatured proteins, where the change in protonation states significantly affects the chemical shifts of aspartic acids, glutamic acids and histidines. The same is true for chemically denatured proteins for which we are not yet able to take into account the effect of the denaturant on the chemical shifts.

5.3. Combining Flexible-Meccano and ASTEROIDS to target chemical shifts

The selection of sub-ensembles on the basis of chemical shifts using ASTEROIDS requires a diverse initial pool of structures. A simple selection of a sub-ensemble from a standard Flexible-Meccano pool (random coil conformers) on the basis of experimental chemical shifts is not likely to work in cases where the IDPs possess a significant amount of secondary structure. The protocol that we have developed to assemble structural ensembles on the basis of chemical shifts is outlined in Figure 27.

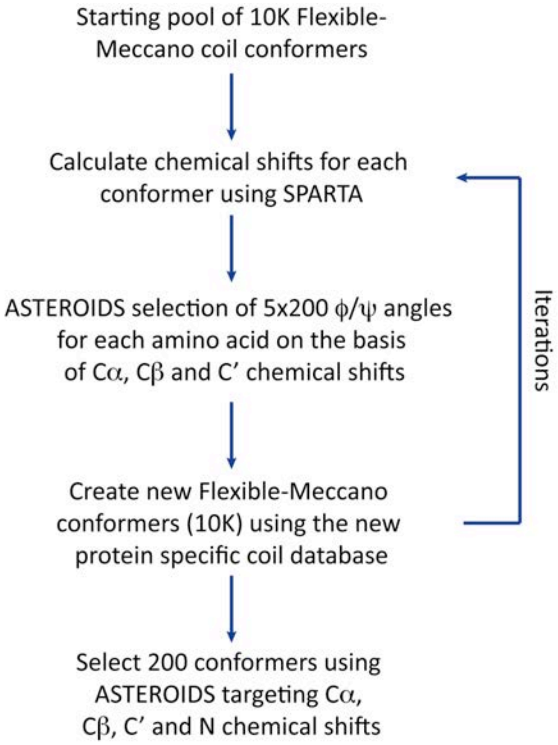

Figure 27: Flow-chart of Flexible-Meccano/ASTEROIDS selection of conformational ensembles on the basis of experimental chemical shifts of disordered proteins.

Initially, a large pool of random coil conformers (10K) is generated using Flexible-Meccano. For each conformer, side chains are added using the program SCCOMP (*102*) and the chemical shifts are calculated using SPARTA. The selection procedure involves two steps: an iteration step where each residue is treated independently, and a final step where full structures are selected.

The first iteration step consists of the selection of 200 ϕ/ψ values for each residue that are in agreement with the experimental $^{13}C\alpha$, $^{13}C\beta$ and $^{13}C'$ chemical shifts. This step is repeated five times to obtain 1000 ϕ/ψ values for each residue. A new ensemble of structures is created using Flexible-

Meccano, however, this time using the selected 1000 ϕ/ψ values for each residue. This results in a new pool of conformers specific to the protein under investigation. To avoid premature convergence of the protocol into a local minimum, we add 25% random coil Flexible-Meccano conformers to this new pool. ASTEROIDS is applied again for each residue independently to select 5 × 200 ϕ/ψ values from the new pool of conformers. This iterative procedure is repeated until no further improvement in the fitting of the chemical shifts of the individual residues can be obtained. In practice this depends on the amount of secondary structure present in the protein.

Step two of the selection procedure is then applied to $^{13}C\alpha$, $^{13}C\beta$, $^{13}C'$ and ^{15}N chemical shifts, where entire structures (typically 200 conformers) are selected from the pool generated during the previous iterations. The ^{15}N chemical shifts are used only in the second step of the selection procedure as they depend primarily on the conformational sampling of the neighbouring residues (Figure 24 and Figure 25) and, therefore in principle, do not contribute to the fitting of the individual amino acids in the iteration steps.

The convergence characteristics of the different types of chemical shifts over a structural ensemble were investigated. The range of chemical shifts predicted for members of a random coil ensemble is rather small, resulting in fast convergence of the chemical shifts compared to other NMR parameters such as RDCs. Typically, convergence is achieved around a few hundred structures and the selection of 200 conformers using ASTEROIDS, as proposed in the protocol above, seems to be reasonable and would eventually allow the simultaneous fitting of both chemical shifts and RDCs, assuming that the LAW approach is applied.

5.4. Testing ASTEROIDS on simulated chemical shift data

Initially, we applied the protocol in Figure 27 to simulated chemical shift data sets in order to investigate to which extent chemical shifts alone can reproduce a known conformational sampling. Ensembles of a model sequence (50 amino acids) were created using the standard ϕ/ψ database of Flexible-Meccano, an extended database sampling more β-sheet and PPII regions and a database sampling more α-helical conformations. The ensemble-averaged chemical shifts for the three ensembles were calculated over 10K conformers. The three sets of chemical shifts (standard, extended and helix) were then subjected to ASTEROIDS for selection of a sub-ensemble of 200 conformers from a pool of conformers generated using the standard database. The final selection of full structures was done with and without the inclusion of ^{15}N chemical shifts to evaluate the effect of these chemical shifts for distinguishing the populations in the PPII region and the α'-helical region.

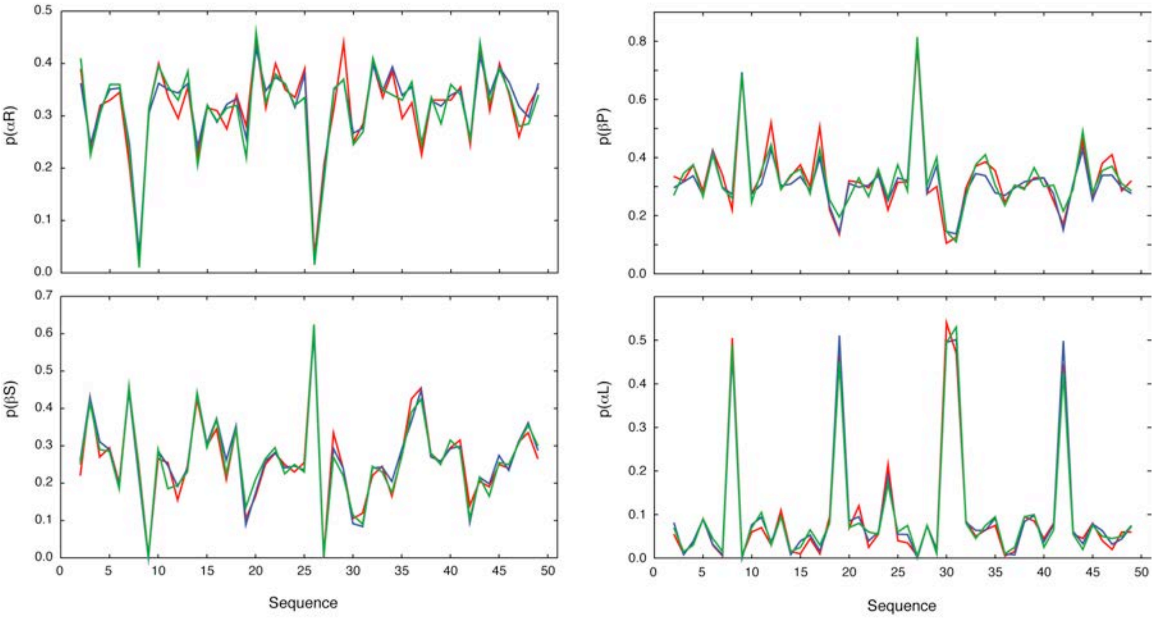

Figure 28: Reproduction of conformational sampling by an ASTEROIDS-selected ensemble comprising 200 conformers obtained by targeting the synthetic "standard" chemical shift data set. The initial pool from which the structures were selected was created using the standard coil library of Flexible-Meccano. The selection of full structures (second step of the protocol) was carried out using $^{13}C\alpha$, $^{13}C\beta$ and $^{13}C'$ (red) and $^{13}C\alpha$, $^{13}C\beta$, $^{13}C'$ and ^{15}N (green) and compared to the target conformational sampling (blue).

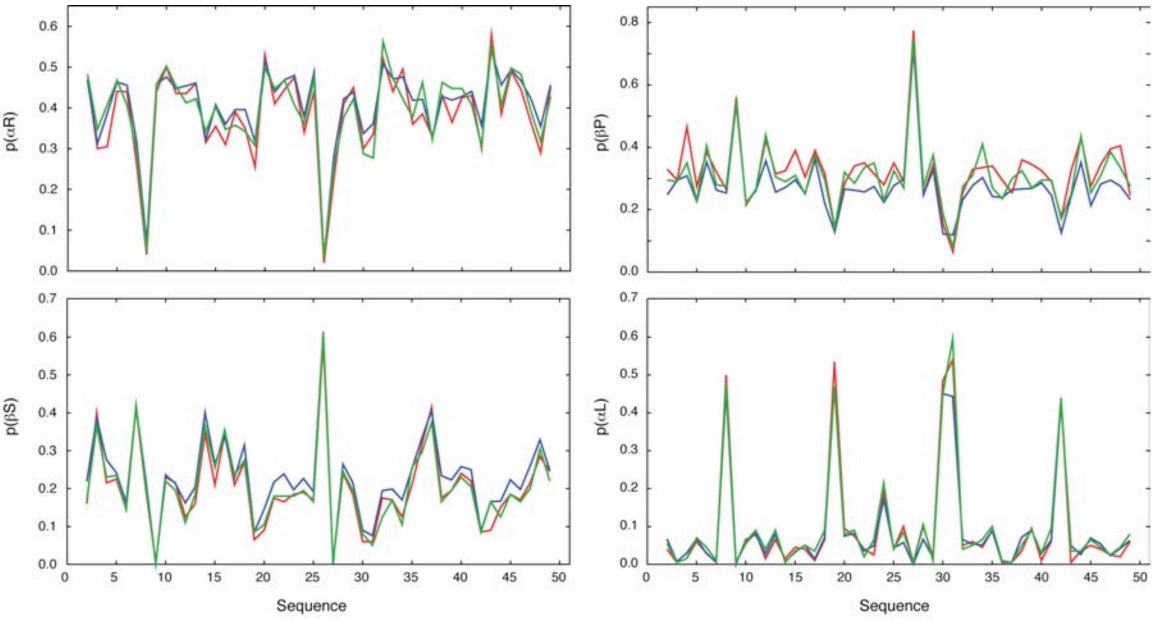

Figure 29: Reproduction of conformational sampling by an ASTEROIDS-selected ensemble comprising 200 conformers obtained by targeting the synthetic "helix" chemical shift data set. The initial pool from which the structures were selected was created using the standard coil library of Flexible-Meccano. The selection of full structures (second step of the protocol) was carried out using $^{13}C\alpha$, $^{13}C\beta$ and $^{13}C'$ (red) and $^{13}C\alpha$, $^{13}C\beta$, $^{13}C'$ and ^{15}N (green) and compared to the target conformational sampling (blue).

In all three cases (standard, extended and helix) the synthetic chemical shift data sets were well reproduced by the ASTEROIDS-selected ensembles (data not shown). The reproduction of the conformational sampling was investigated as described in the legend of Figure 7. The conformational sampling present in the standard and the more helical ensemble were well captured by the selected ensembles (Figure 28 and Figure 29), while slightly larger deviations were observed in the case of the more extended ensemble (Figure 30). Importantly, it is seen that the introduction of the ^{15}N chemical shifts in the final selection of full structures improves the reproduction of the conformational sampling, in particular in the case of the more extended ensemble. In general, the simulations demonstrate that it is possible to obtain a standard coil, more extended sampling or a more helical sampling directly from the chemical shifts to within 5% accuracy.

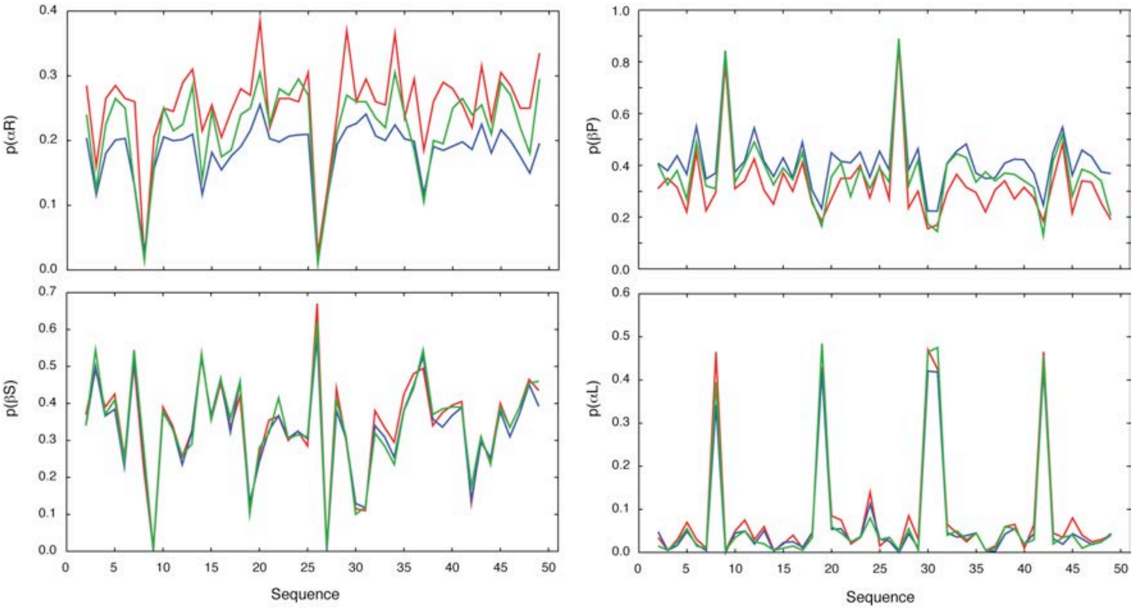

Figure 30: Reproduction of conformational sampling by an ASTEROIDS-selected ensemble comprising 200 conformers obtained by targeting the synthetic "extended" chemical shift data set. The initial pool from which the structures were selected was created using the standard coil library of Flexible-Meccano. The selection of full structures (second step of the protocol) was carried out using ^{13}Cα, ^{13}Cβ and ^{13}C′ (red) and ^{13}Cα, ^{13}Cβ, ^{13}C′ and ^{15}N (green) and compared to the target conformational sampling (blue).

5.5. Application to the C-terminal domain of Sendai virus nucleoprotein

The approach was then applied to map the site-specific conformational sampling of the intrinsically disordered C-terminal domain, N_{TAIL}, of Sendai virus nucleoprotein. We choose this particular example because N_{TAIL}, although intrinsically disordered, contains large amounts of α-helical propensity in its molecular recognition element (residues 475-495) and, therefore, constitutes a rather demanding test of the approach. An ASTEROIDS ensemble comprising 200 conformers of N_{TAIL} was selected on the basis of the experimental ^{13}Cα, ^{13}Cβ, ^{13}C′ and ^{15}N chemical shifts. Excellent agreement between the experimental chemical shifts and those back-calculated from the selected ensemble was obtained (Figure 31). The site-specific conformational sampling of N_{TAIL} derived from the chemical shifts shows that N_{TAIL} samples essentially random coil conformations with large amounts of helical propensity in the molecular recognition element (Figure 32 and Figure 33). An analysis of the distribution of the selected conformers outside the helical element reveals that the fully disordered regions of the protein have an overall tendency to sample less β-extended and more (on average 5%) PPII than is present in the standard random coil database.

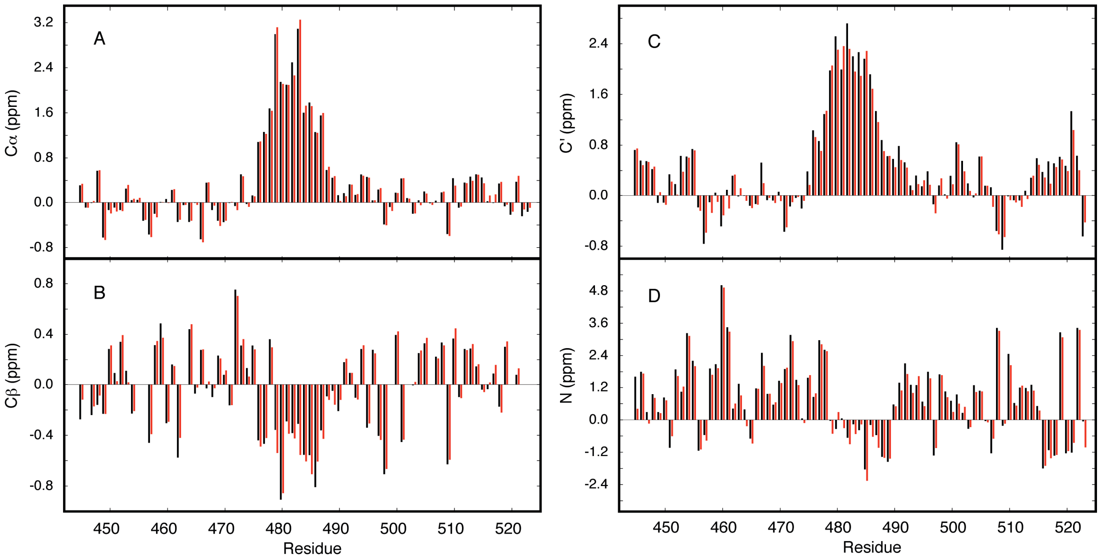

Figure 31: Agreement between experimental secondary chemical shifts (black) in N_{TAIL} and back-calculated secondary chemical shifts (red) from the ASTEROIDS-selected ensemble comprising 200 conformers.

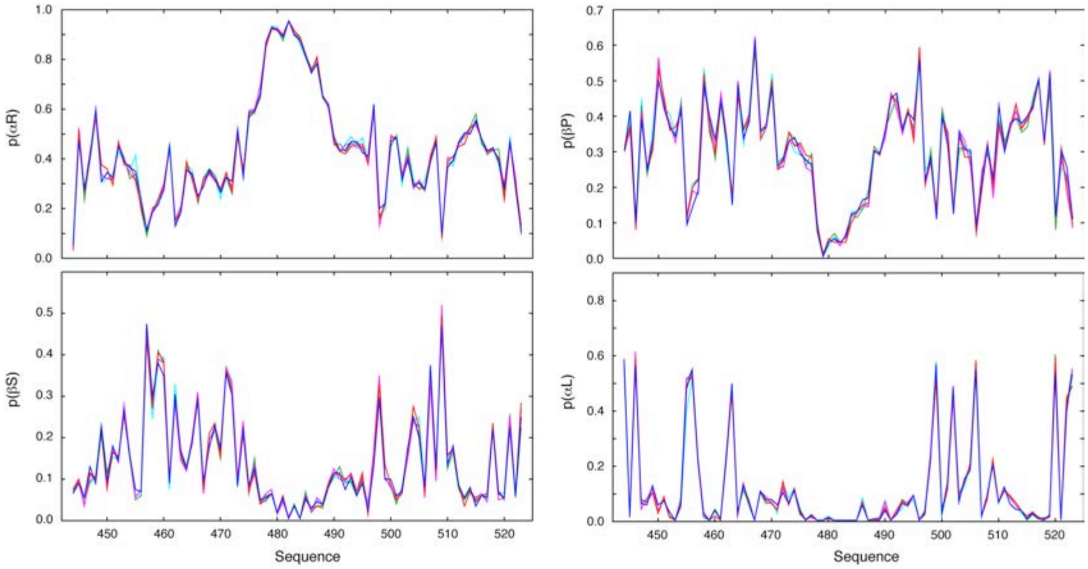

Figure 32: Conformational sampling of N_{TAIL} derived from ASTEROIDS-selected ensembles on the basis of experimental $^{13}C\alpha$, $^{13}C\beta$, $^{13}C'$ and ^{15}N chemical shifts. The selection of full structures (200 conformers) was repeated five times and the different colours indicate the conformational sampling obtained in each selection.

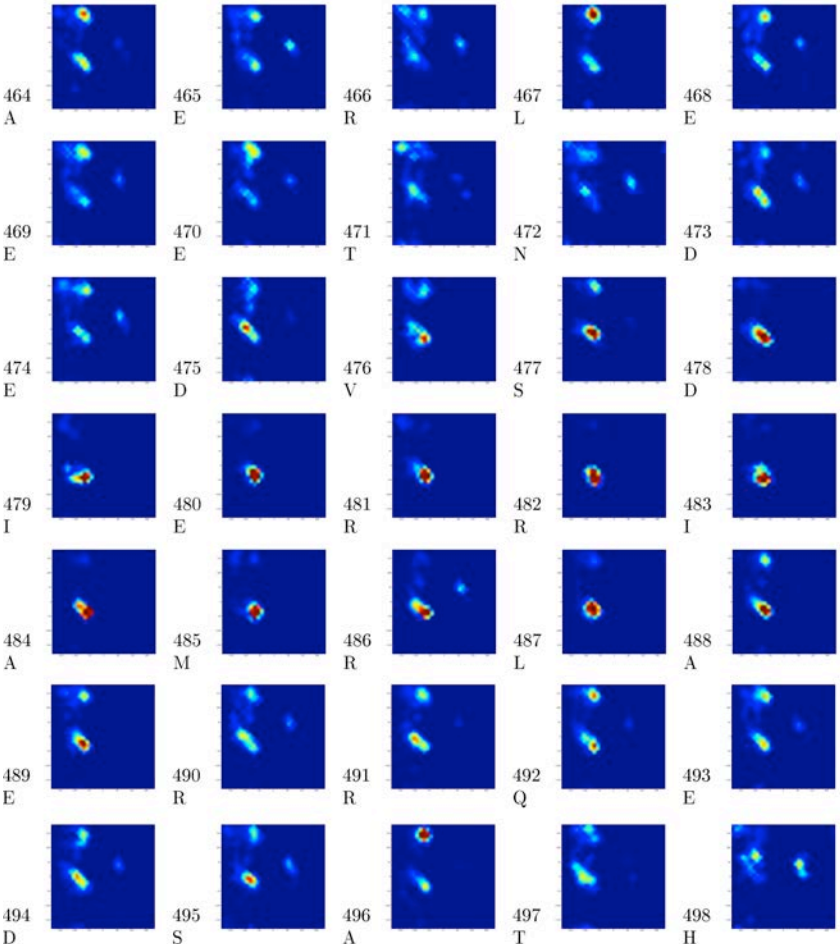

Figure 33: Site-specific conformational sampling (Ramachandran plots) obtained from the ASTEROIDS-selected ensembles of N_{TAIL} on the basis of experimental chemical shifts. An excerpt of the sequence is shown surrounding the molecular recognition element.

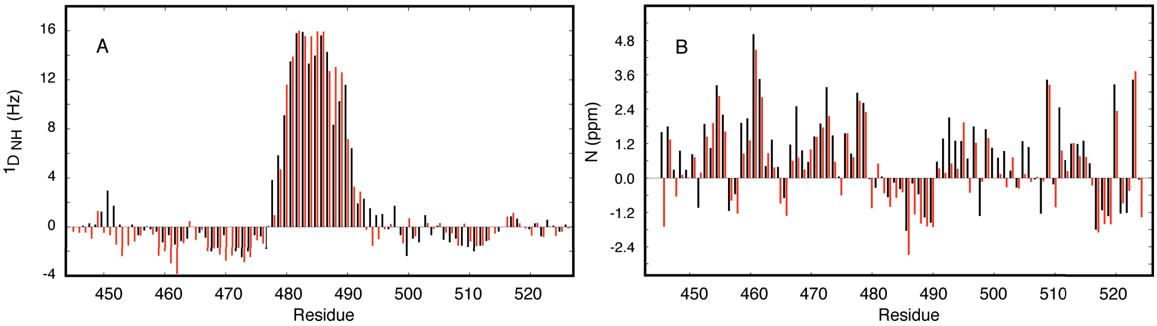

Figure 34: Reproduction of independent data not used in the ASTEROIDS selection. (A) Comparison of experimental ^{15}N-$^{1}H^{N}$ RDCs measured in a liquid crystal composed of poly-ethylene glycol and 1-hexanol (black) and RDCs back-calculated from a Flexible-Meccano ensemble of 50K conformers using the φ/ψ sampling obtained on the basis of the experimental $^{13}C\alpha$, $^{13}C\beta$, $^{13}C'$ and ^{15}N chemical shifts (red). (B) Comparison of experimental ^{15}N secondary chemical shifts (black) and secondary chemical shifts back-calculated from the ASTEROIDS ensemble selected on the basis of $^{13}C\alpha$, $^{13}C\beta$ and $^{13}C'$ (red). Random coil values from RefDB were used.

As described in the previous chapter, it is essential to be able to validate the selected ensembles for example using independent data not used in the ensemble selection (cross-validation). For this purpose we compared experimental ^{15}N-^{1}H^{N} RDCs measured in partially aligned N$_{TAIL}$ to those calculated using an ensemble obtained from the chemical shift derived conformational sampling (Figure 34). The agreement is striking, in both the partially folded and disordered regions of the protein, demonstrating the ability of ASTEROIDS to unambiguously interpret chemical shifts in terms of local conformational propensity. In a further test of consistency, we also compared the experimental ^{15}N secondary chemical shifts to those back-calculated from the ASTEROIDS-selected ensemble on the basis of ^{13}Cα, ^{13}Cβ and ^{13}C' only (Figure 34). Although this implies removing 25% of the data in the ASTEROIDS selection, the experimental secondary chemical shifts are still reasonably well reproduced providing a further validation of our approach.

5.6. Conclusions

The ability to describe conformational sampling on the basis of chemical shifts alone is important for the development of atomic resolution descriptions of IDPs. The approach presented here makes only simple assumptions concerning the true conformational properties of the molecule, starting with a standard statistical coil description of backbone conformational sampling, and refining this iteratively until convergence is reached compared to the experimental data. We show that this allows the identification and characterization of entire secondary structural elements and their associated populations, as well as providing indications of the subtle detail of local conformational sampling in disordered proteins. This approach raises the prospect of probing the conformational behaviour of disordered proteins under conditions where additional parameters cannot be easily measured, but where chemical shifts are still accessible, for example in crowded or cellular environments (*103*, *104*). Currently, we are applying ASTEROIDS to target experimental chemical shifts in many different IDPs in order to obtain a more elaborate picture of the conformational sampling in disordered proteins. Our preliminary results show, like in the case of N$_{TAIL}$, that the PPII conformation seems to be prevalent at the cost of β-extended conformations.

The simulations presented in this chapter showed that the populations in the different regions of Ramachandran space could be mapped with about 5% accuracy using the chemical shifts. However, the populations of the PPII region and the α'-helix region remain difficult to separate, because of the similar chemical shifts in these regions for most nuclei. In order to obtain a better separation of the populations in the Ramachandran space, we are currently combining the chemical shifts with other types of NMR parameters such as RDCs.

6. Quantitative description of helices in IDPs using RDCs

6.1. Introduction

The lack of stable, folded structures of IDPs confers numerous functional advantages, including, paradoxically, both binding promiscuity and high specificity in target interactions. Atomic resolution details of how IDPs are recognized by their partner proteins are currently lacking and in particular insight into which features control the kinetics of coupled folding-and-binding reactions. Some of the key questions that currently remain unanswered are to what extent the IDPs need to be structurally pre-configured prior to partner protein binding, and how the "right" conformation is stabilized in an otherwise disordered protein.

In order to provide insight into molecular recognition in disordered proteins, we have been studying IDPs that undergo α-helical folding upon binding to their partner proteins. One of these proteins is the intrinsically disordered C-terminal domain, N_{TAIL}, of Sendai virus nucleoprotein whose molecular recognition element undergoes α-helical folding upon binding to the phosphoprotein of the viral polymerase complex (*105*, *106*). In the previous chapter, the conformational sampling of this protein was investigated using ASTEROIDS targeting the experimental chemical shifts of N_{TAIL}, and it was found that the molecular recognition element samples large amounts of α-helix already in its pre-recognition state.

In this chapter, an approach is presented that, on the basis of measurements of multiple RDCs, allows a quantitative identification of the most populated helical elements in the pre-recognition state of N_{TAIL} – information that cannot be obtained from chemical shifts alone. The approach provides insight into early stages of molecular recognition in disordered proteins and reveals an essential role of the primary sequence for dictating and stabilizing this particular pre-recognition state.

6.2. Using dipolar waves to characterize helical elements in IDPs

RDCs are particularly sensitive to the presence of helical elements in disordered proteins. The reason for this is that the sign of ^{15}N-$^{1}H^{N}$ RDCs is expected to change when going from more extended (disordered) conformations to helical or turn conformations as illustrated in Figure 35 (*45*, *47*). To understand in more detail how RDCs can help us to characterize helices in terms of length, population and capping, we carried out Flexible-Meccano simulations of a poly-alanine sequence (100 amino acids), where helices with different lengths and positions were introduced by sampling the α-helical region of ϕ/ψ space. The alignment tensor for each conformer was calculated using PALES (*72*, *73*) and the RDCs were subsequently averaged over the entire ensemble.

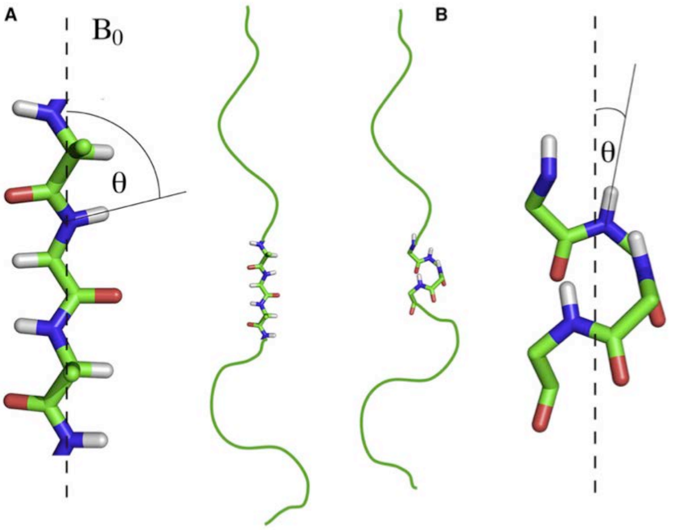

Figure 35: Figurative representation of effective angular averaging properties of $^1H^N$-^{15}N bond vectors in a disordered protein dissolved in a weakly aligning medium with the director along the magnetic field. (A) RDCs measured for $^1H^N$-^{15}N bond vectors in more extended conformations, as commonly found in disordered proteins, will have negative values (θ ~ 90°). (B) RDCs measured for $^1H^N$-^{15}N bond vectors in turns or helical conformations will have positive values (θ ~ 0°).

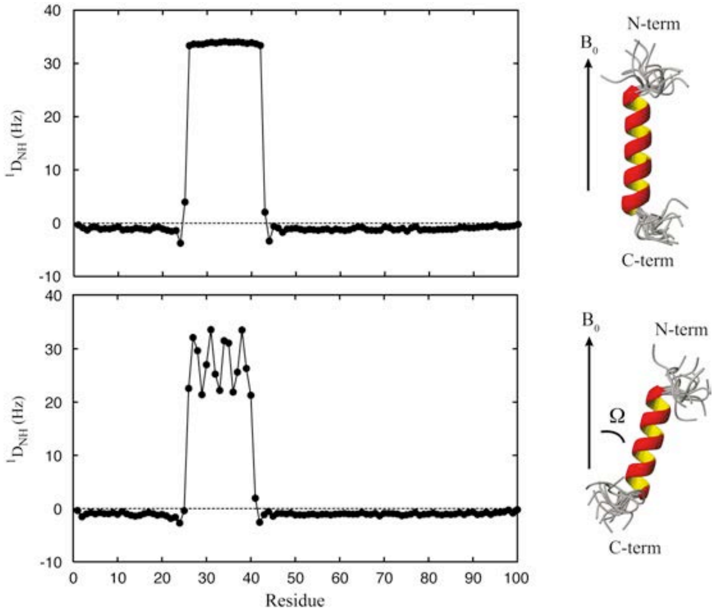

Figure 36: Ensemble-averaged RDCs (50K conformers) of a poly-alanine chain of 100 amino acids with helices at residues 26-41 (top) and 26-39 (bottom). The helices were introduced by sampling the α-helical region of ϕ/ψ space (-65°, -40°) using a Gaussian distribution with a width of 3°. In each case, 15 superimposed conformers from Flexible-Meccano show the directionality of the disordered chains projected from the helix caps. For clarity only the first four residues before and after the helix are shown. The effective tilt angle imposed by the helix capping is 1.3° (top) and 18° (bottom).

Comparing ^{15}N-$^1H^N$ RDCs from two of the simulations, where helices where introduced at residues 26-39 and 26-41, shows that a difference in length of only two amino acids induces distinctly different dipolar waves within the helical elements (Figure 36). Despite the high flexibility of the long N- and C-terminal chains, the appearance of periodicity in the RDC profile within the helix is strongly correlated with the direction of the unfolded chains as they are projected from the helix caps. If the dynamic chains are projected in the same direction (26-41), dipolar wave oscillations are small, because the effective orientation of the helix is close to parallel to the field. If the chains are projected in opposing directions (26-39), large oscillations are observed due to the large effective tilt of the helix (Figure 36). Fitting the simulated ensemble-averaged RDCs within the helical elements to a perfect helical structure in Module (*107*) using Eq. 3 shows that the rhombic part of the alignment tensor is very small and can essentially be ignored.

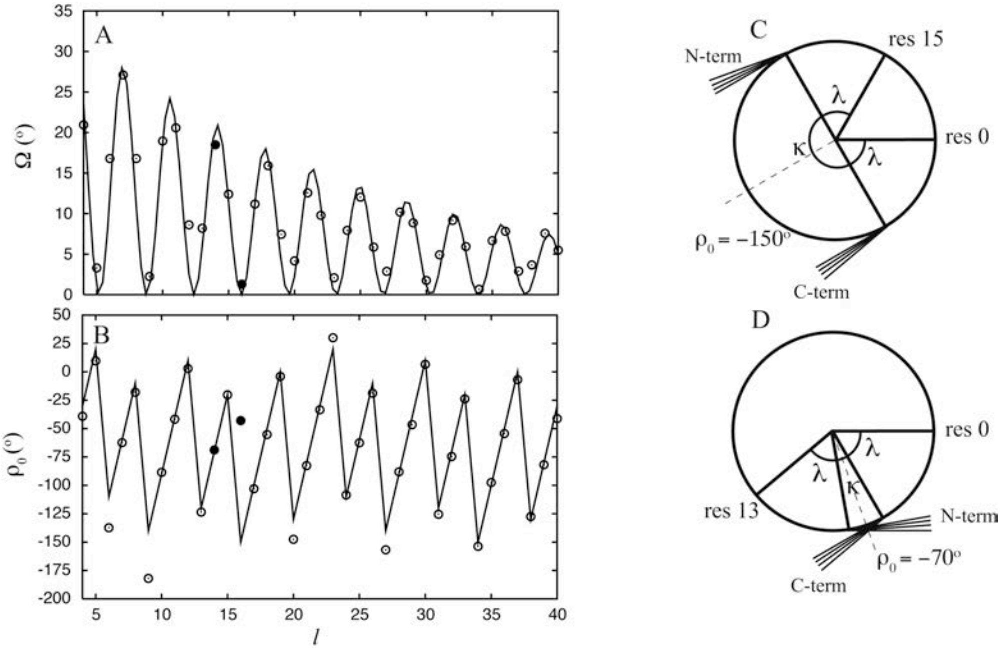

Figure 37: Dipolar waves in flexible chains. (A) Effective tilt angle, Ω, relative to magnetic field as a function of helix length l (in residues) introduced between sites 26 and 65 of a 100 amino acid poly-alanine chain. RDCs were averaged over 2000 conformers and fit to an ideal helix using Eq. 15. Line corresponds to the fit of simulated data points with $\Omega(l)=k_1\exp(-k_2l)\cos^2(\kappa(l)/2)$; κ is the angle between points of chain projection given by $\kappa(l)=360(l-1)/p+2\lambda$ (fitted values: $k_1 = 37.3$, $k_2=0.01$, p (periodicity) = 3.6 and $\lambda = 60$). Filled circles indicate the two simulations in Figure 36. (B) Helix polarity, ρ_0, as a function of l. ρ_0 was obtained by fitting simulated RDCs to Eq. 15 using an ideal helical structure. Line corresponds to the expected polarity calculated as the average position of the two points of chain projection. The polarity can only be accurately determined for $\Omega > 4°$. (C) Top view of helix of length 16 showing the position of the first (res 0) and last (res 15) residues as well as the positions of chain projection. Dashed line indicates expected ρ_0 value. (D) Top view of helix length 14. These helices correspond to the two simulations in Figure 36.

In agreement with previous studies of dipolar waves in folded proteins, the RDCs within helical elements of partially folded proteins can, therefore, be calculated according to (*108–112*):

$$D_{ij} = -\frac{\gamma_i\gamma_j\hbar\mu_0}{8\pi^2r^3}\left[A_a(3(\cos\Omega\cos\delta - \sin\Omega\sin\delta\cos(\rho-\rho_0))^2 - 1)\right] \qquad \text{(Eq. 15)}$$

Here, δ is the orientation of the inter-nuclear vector with respect to the helix axis and ρ ($=2\pi n/3.6$ with $n = 0, \dots, l - 1$) is the position of the individual residue along the helix. It turns out that the effective tilt (Ω) of the helix axis with respect to the alignment frame and the polarity, ρ_0, describing the angular position of the helix relative to this axis, depend in a periodic and parameterisable manner on the helix length (Figure 37). This allows the prediction of expected dipolar waves in partially folded proteins without the need for construction of explicit ensembles of the proteins.

From the observations made above, it is clear that the distribution of ^{15}N-^{1}H^{N} RDCs, the size of the couplings and the shape of the dipolar wave allow a determination of helix length, population and capping in partially folded proteins. This is in contrast to chemical shifts that only report on the helix population at a given residue. The measurement of several types of RDCs is expected to increase the accuracy of the determined helical parameters, as the only difference between different inter-nuclear bond vectors relies in the angle δ. The simulations presented above contained a single helical element with a well-defined dipolar wave. In disordered or partially folded proteins, several helices may be preferentially populated, and the observed dipolar wave has to be interpreted as the population-weighted sum of several dipolar waves as described below in the case of the molecular recognition element of Sendai virus N$_{\mathrm{TAIL}}$.

6.3. Application to the molecular recognition element of N$_{\mathrm{TAIL}}$

To quantitatively describe the α-helical sampling in the pre-recognition state of Sendai virus N$_{\mathrm{TAIL}}$, we measured several types of RDCs (^{15}N-^{1}H^{N}, ^{13}Cα-^{1}Hα, ^{13}Cα-^{13}C' and ^{13}C'-^{1}H^{N}) using polyethylene glycol (PEG)/1-hexanol liquid crystals. The ^{15}N-^{1}H^{N} RDCs measured within the molecular recognition element of N$_{\mathrm{TAIL}}$ are positive and clearly display a dipolar wave indicating that specific helices are preferentially being populated in solution (Figure 38). In order to exploit the available information in the experimental RDCs, we have developed a minimal ensemble approach that allows us to quantitatively characterize these helices (*113*). Using Flexible-Meccano, we generated conformational ensembles of N$_{\mathrm{TAIL}}$ with helices of different lengths and positions within the molecular recognition element (residues 476-495). Each ensemble consisted of 10K conformers in order to ensure convergence of the RDCs over the structural ensemble. A total of 153 different ensembles were necessary to simulate all possible helices with a minimum length of four amino acids to a maximum length of 20 amino acids. The approach serves to find the minimal ensemble that best fit the experimental RDCs by combining the predicted RDCs from the different ensembles and optimizing the populations of the helical elements:

$$D_{ij}^{\mathrm{calc}} = \sum_{k=1}^{N} p_k D_{ij}^{k} + \left(1 - \sum_{k=1}^{N} p_k\right) D_{ij}^{\mathrm{unfold}} \qquad \text{(Eq. 16)}$$

Here, p_k represents the populations of the N helical conformers for which D_{ij}^{k} are the individual predicted couplings between nuclei i and j, and D_{ij}^{unfold} are the predicted couplings from the completely disordered state. The calculated couplings are compared to the experimental data according to:

$$\chi^2 = \sum (D_{ij}^{\mathrm{calc}} - D_{ij}^{\mathrm{exp}})^2 / \sigma_{ij}^2 \qquad \text{(Eq. 17)}$$

All four measured types of couplings were included in the fitting procedure. The number of helical ensembles present in the equilibrium was increased (starting from N=1) and F-tests were applied to verify the statistical significance of the improvement of the fitting of the experimental data (Table 1). It was found that N=3 is the minimum number of helix lengths that are needed to describe the experimental RDCs. Figure 38 shows the agreement between experimental ^{15}N-^{1}H^{N} RDCs and back-calculated RDCs for the different models in Table 1, while Figure 39 shows the reproduction of all types of RDCs in the molecular recognition element of N_{TAIL} for the best fitting model (N=3).

Table 1: Reproduction of experimental RDCs in N_{TAIL} from models with different numbers of helical ensembles.

Number of helical ensembles (N)	χ^{2} [a]	Number of optimized parameters [b]	Helical conformers [c]	Population (%) [d]	Significance [e]
1	427	4	476-488	52	
2	227	7	476-488 479-484	37 45	P<0.0001
3	120	10	476-488 479-484 478-492	28 36 11	P<0.0001
4	113	13	476-488 479-484 478-495 479-492	31 24 6 5	P=0.16

[a] The target function (χ^2) included all 100 experimental data points (four types of RDCs).

[b] One helix implies the optimization of three parameters: starting amino acid, final amino acid and the population. In addition a scaling factor is optimized to take into account the absolute level of alignment for the RDCs.

[c] Range of the invoked helices.

[d] The population of the invoked helices. The remaining conformers are completely unfolded.

[e] Significance of the improvement of this model as compared to the simpler model calculated using a standard F-test.

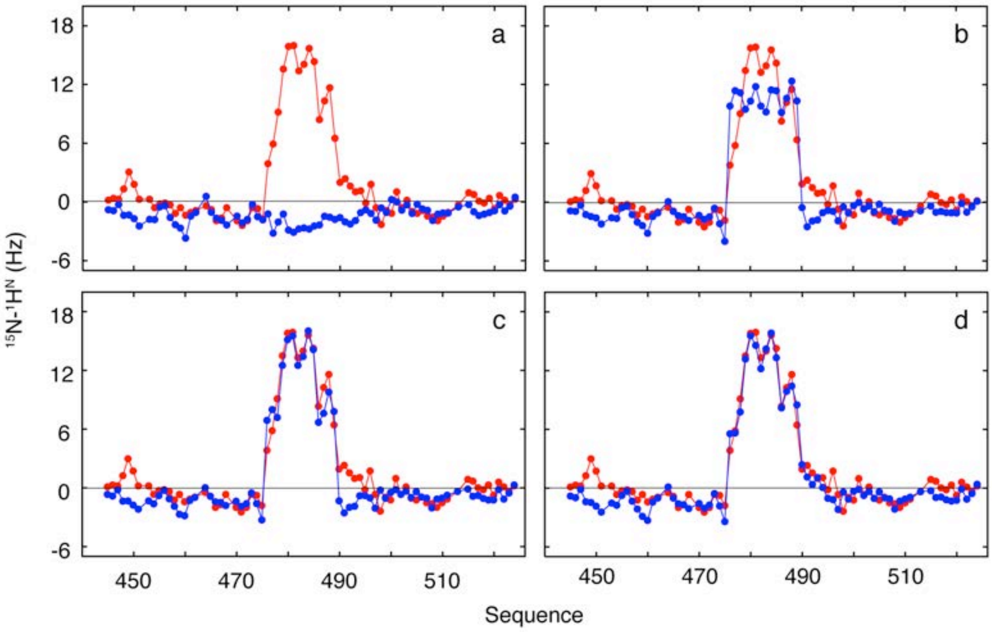

Figure 38: Reproduction of experimental ^{15}N-^{1}H^{N} RDCs in N_{TAIL} for models with an increasing number of helical ensembles: *N*=0 (a), *N*=1 (b), *N*=2 (c) and *N*=3 (d). Experimental RDCs are shown in red, while back-calculated RDCs from the different models are shown in blue.

The best fitting model of the molecular recognition element of N_{TAIL} comprises three helical elements, 479-484 populated to 36±3%, 476-488 populated to 28±1% and 478-492 populated to 11±1% in exchange with completely disordered conformers corresponding to 25±4% of all molecules (Figure 40). Noise-base Monte Carlo simulations were used to estimate the uncertainties in the helical populations. Interestingly, all the selected helices are preceded by aspartic acids or serines that are the most common N-capping residues in helices in folded proteins (Figure 40) (*114*–*116*). An N-capping residue stabilizes a helix by forming a hydrogen bond between its side chain and the backbone amides at position 2 or 3 in the helix (Figure 40) (*117*, *118*). Importantly, this indicates that the helices preferentially being populated in solution in N_{TAIL} are stabilized by N-capping interactions, and the helical formation is being promoted by strategically placed aspartic acids and serines in the primary sequence of the protein.

Sendai virus N_{TAIL} interacts with the C-terminal three-helix bundle domain, XD, of the phosphoprotein of the viral polymerase complex and thereby initiates transcription and replication of the virus. It has been proposed previously that electrostatic forces control this interaction, as a large negative patch is displayed on the surface of XD at the N_{TAIL} binding site (*106*). The molecular recognition element of N_{TAIL} contains several positively charged arginines spaced by 3-4 amino acids in the primary sequence. Thus, the formation of helices already in the pre-recognition state of N_{TAIL} places the arginines on the same side of the helical elements and, thereby, creates a positive "patch" complementary to the negative surface on the partner protein (Figure 40).

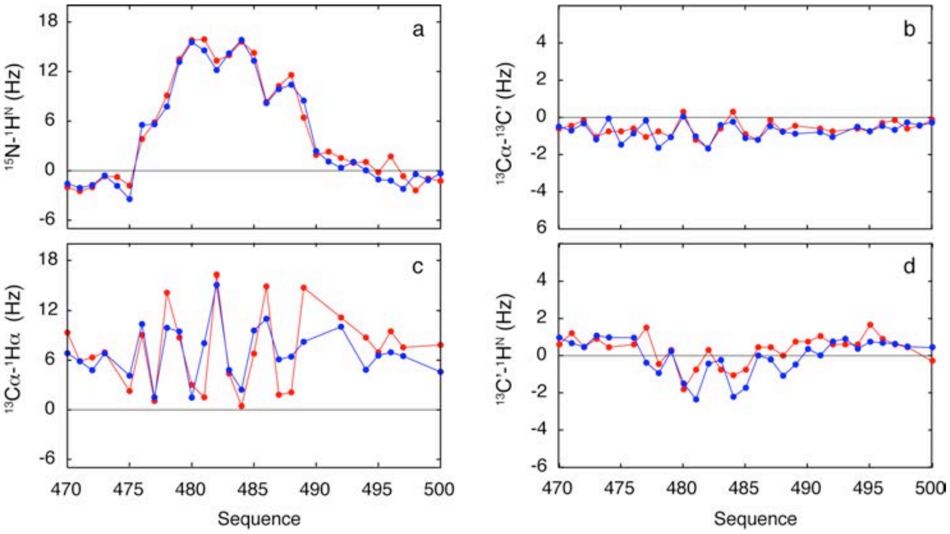

Figure 39: Reproduction of the four types of RDCs in the molecular recognition element of N_{TAIL} for the best-fitting model containing three helical ensembles (N=3). Experimental RDCs are shown in red, while back-calculated RDCs from the model are shown in blue.

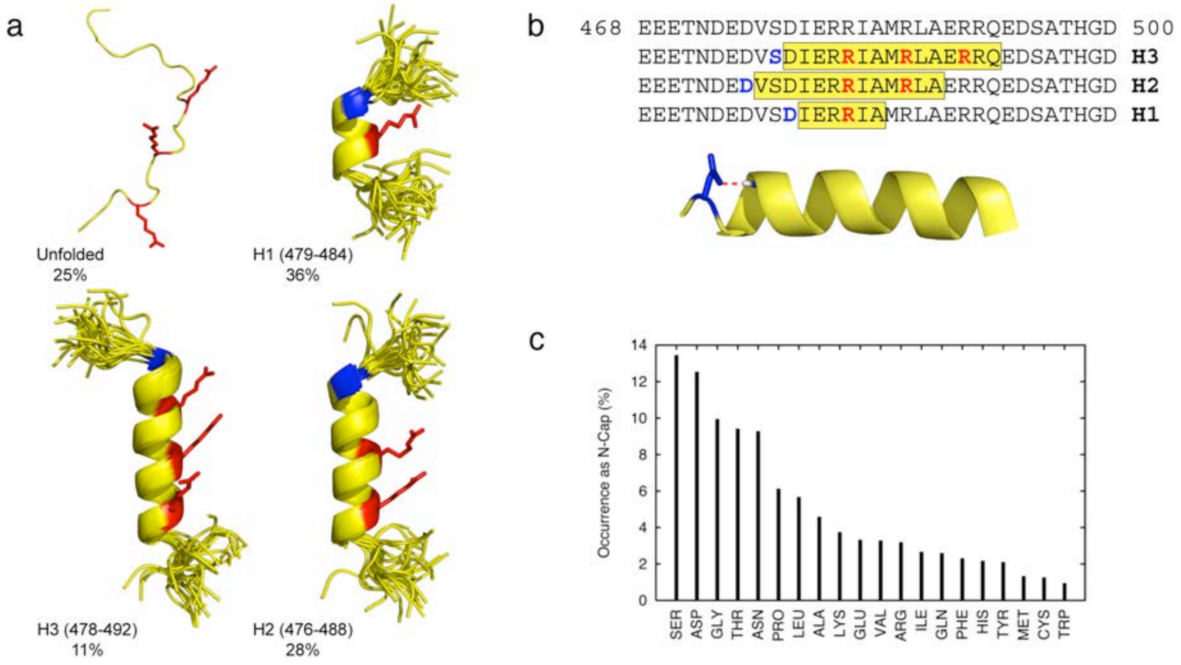

Figure 40: (a) Molecular representation of the proposed conformational equilibrium of the molecular recognition element of N_{TAIL} in solution. The four conformations are presented as a single structure for the completely disordered form and as twenty randomly selected conformers for the three helical conformations. The molecular recognition arginines are displayed in red, while N-capping residues are shown in blue. (b) The amino acid sequence of the molecular recognition element of N_{TAIL} showing the positions of the selected helices. The cartoon figure illustrates an N-capping aspartic acid side chain-backbone interaction. (c) The occurrence of different types of amino acids as N-capping residues in helices of folded proteins.

It is also noteworthy that the two most populated helices (H1 and H2) in N_{TAIL} differ by one turn in each end. As a consequence, the unfolded ends are projected in a specific direction in the majority of the helical conformers. This projection does not point away from the partner in the bound complex, as defined by the position of the arginines on one side of the helix, but rather in the same overall direction as the arginine side chains (Figure 40). This observation is only valid for the first few amino acids preceding and following the helix, beyond which the statistical coil sampling of the chain results in very diffuse directionalities, but may be important in terms of non-specific interactions away from the binding site, which facilitate formation of the complex (*119*).

6.4. Conclusions

The minimal ensemble approach presented in this chapter allows the identification of preferentially populated helical elements in partially folded proteins. This is particularly relevant for intrinsically disordered or partially folded proteins that bind to their partner proteins through helical regions, and for which an identification of the specific helical conformers provides new insight into molecular recognition. The limitation of the approach probably lies in the percentage of helix sampled, as the method is expected to be most accurate for larger amounts (>20%). In addition, if many helices are randomly sampled, the dipolar wave will be quenched and the applicability of the approach reduced.

Applying the approach to N_{TAIL}, that undergoes α-helical folding of the molecular recognition element upon binding to its partner protein, we find that N_{TAIL} samples specific helical conformers that are all stabilized by N-capping interactions. The high percentage of helix already in the pre-recognition state has the advantage of lowering the energy-barrier between free and bound forms of the protein, and probably controls the kinetics of the coupled folded-and-binding reaction. We are currently investigating the binding kinetics of N_{TAIL} and the role that the specific α-helical conformers play in this context by site-directed mutagenesis of the N-capping residues (see chapter 8).

7. Intrinsic disorder in Measles virus nucleocapsids

7.1. Introduction

IDPs are normally expressed and studied as isolated protein domains in vitro. However, in the cell, these proteins are often integral parts of much larger assemblies with many different protein domains – both folded and disordered. The possibility that a protein is disordered and highly flexible when studied in its isolated form in vitro, but folded or not flexible in situ, underlines the importance of studying intrinsic disorder under more physiological conditions.

One way to characterize IDPs in situ is to collect solution NMR data of the proteins in the context of the entire molecular assemblies. Although such objects are normally too big to be studied by NMR due to the slow re-orientation in solution, the signals of the IDPs are still detectable, if the domains retain their flexibility in the context of the full-length proteins. This approach allows us to study the conformational behaviour of disordered domains under conditions approaching the physiological and to compare the results with structural studies of the domains in their isolated forms.

Our first application of this approach concerned the disordered N-terminal transactivation domain (TAD) of the tumour suppressor p53 (*64*). We characterized this 93 amino acid disordered domain using RDCs, both in its isolated form and in the context of the 175 kDa full-length, tetrameric p53/DNA complex. It was shown that the TAD samples similar conformations in its isolated form and within the full-length protein, including 30% helical propensity stabilized by N-capping interactions in the MDM2 molecular recognition element located within the TAD (Figure 41).

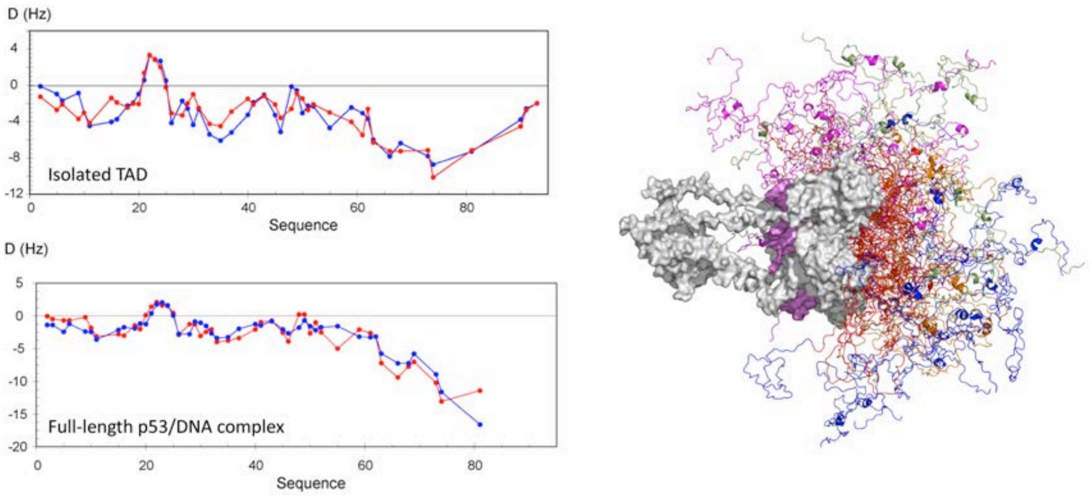

Figure 41: $^1H^N$-^{15}N RDCs measured in the isolated TAD of p53 (blue, top) and in the full-length tetrameric p53/DNA complex (bottom, blue) compared to predictions from Flexible-Meccano (red). In the predictions an amino acid-specific coil sampling was used with the presence of a single-turn helix at residues 22–24 populated at a level of 30% and an increased level of polyproline II sampling for the residues 58–91. The cartoon represents a model of the full-length, tetrameric p53/DNA complex including the flexible TAD. The tetramerization domain of p53 is shown in gray, while the DNA is shown in magenta.

Recently, we have pushed the limits significantly of this “in situ approach” by characterizing intrinsic disorder in intact Measles virus nucleocapsids – objects with sizes approaching 150MDa. Interestingly, as described in this chapter, we obtain new insight into the functional advantages of intrinsic disorder in the nucleocapsids – information that could not be obtained by studying the intrinsically disordered domain in isolation. In addition, we employ the power of “integrated structural biology” by combining our NMR data with SAXS and electron microscopy to provide an integral model of intact Measles virus nucleocapsids.

7.2. Measles virus

Measles virus belongs to the Paramyxovirus family comprising negative sense, single stranded RNA viruses. In paramyxoviruses, the viral genome is encapsidated by multiple copies of the nucleoprotein (N) forming a helical nucleocapsid. Transcription and replication of the viral RNA are initiated by an interaction between N and the polymerase complex, composed of phosphoprotein (P) and the RNA-dependent RNA polymerase (L) (*120*).

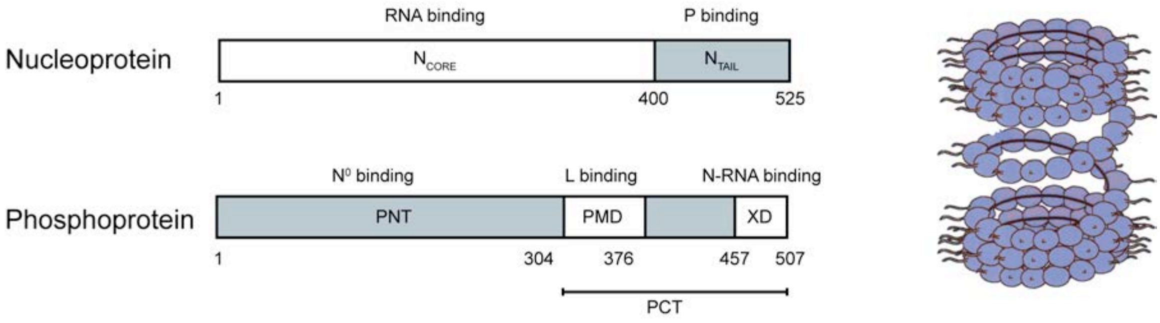

Figure 42: Domain organization of Measles virus nucleoprotein N and phosphoprotein P (left). Gray domains are predicted to be intrinsically disordered. Upon binding of the viral RNA, the nucleoprotein assembles into helical nucleocapsids (right). Sendai virus N and P proteins have similar domain organizations.

Nucleoprotein N consists of two domains: N_{CORE} (residues 1–400) of unknown structure, responsible for the interaction with the viral RNA and for maintaining the nucleocapsid structure, and an intrinsically disordered domain, N_{TAIL} (residues 401–525) serving as the anchor point for the polymerase complex (*105*, *121*) (Figure 42).

The phosphoprotein P comprises an intrinsically disordered N-terminal domain (PNT) through which P binds to RNA-free N (N^0), forming an N^0-P complex that prevents the polymerization of N and the non-specific encapsidation of host cell RNAs in the early stages of transcription and replication (*122–124*). The C-terminal domain of P (PCT) consists of two domains: the multimerization domain (PMD) that binds L and the C-terminal domain XD that associates with the N-RNA nucleocapsid template (Figure 42) (*125*). So far no structural characterization has been carried out of PNT and

PMD, while the three-helix bundle structure of XD has been determined by X-ray crystallography (*126*).

Sendai and Measles virus both belong to the Paramyxovirus family, and although the sequence homology is low between the two viruses, they share similar domain organizations of the N and P proteins. The structure of PMD of Sendai virus P has been determined by X-ray crystallography and adopts a tetrameric, coiled-coil conformation (*127*) (Figure 43). A similar structure is predicted for the PMD of Measles virus P. The XD domain of Sendai virus adopts a three-helix bundle structure and is, therefore, very similar to the Measles virus XD domain (*128*).

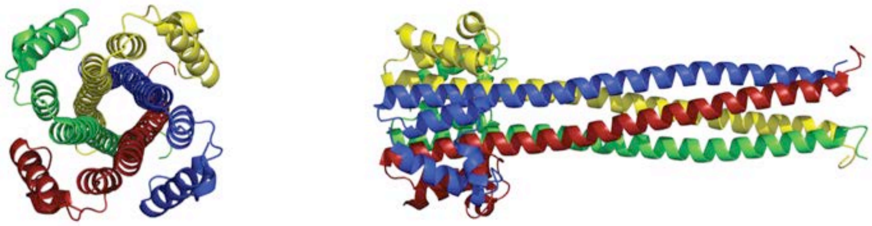

Figure 43: Structure of the tetramerization domain (PMD) of Sendai virus P determined by X-ray crystallography. Top view (left) and side view (right) are shown of the coiled-coil domain.

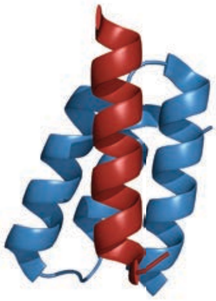

Figure 44: Crystal structure of the chimeric complex between Measles virus XD (blue) and a short construct of N_{TAIL} (red) encompassing the molecular recognition element (residues 485-502).

The molecular recognition element (residues 485–502) of Measles virus N_{TAIL} interacts with XD of P and thereby recruits the polymerase complex onto the nucleocapsid template (*129*, *130*). A crystal structure of the chimeric complex between XD and a construct of N_{TAIL} encompassing the molecular recognition element shows that N_{TAIL} folds into an α-helix in the complex (Figure 44) (*131*). The disorder present in N_{TAIL} is particularly intriguing, as the interaction site of N_{TAIL} is located approximately 80 disordered amino acids from the folded N_{CORE} domain that binds the RNA. In order to understand the mechanism by which this "remote" interaction between N_{TAIL} and XD initiates transcription and replication, it is necessary to have an atomic resolution model of N_{TAIL} in context of the intact nucleocapsids.

7.3. Structural characterization of isolated Measles virus N_{TAIL}

Initially, we performed NMR experiments of the isolated N_{TAIL} domain in order to obtain a detailed structural characterization. We assigned the backbone of the protein and measured several types of RDCs (^{15}N-$^{1}H^{N}$, $^{13}C\alpha$-$^{1}H\alpha$ and $^{13}C\alpha$-$^{13}C'$). An atomic resolution description of the conformational equilibrium of the molecular recognition element of N_{TAIL} was obtained using the minimal ensemble approach applied to the experimental RDCs and $C\alpha$ chemical shifts, in the same way as described in the previous chapter for Sendai virus N_{TAIL}. The $C\alpha$ chemical shifts were included in the analysis to obtain a better determination of the populations of the individual helical elements.

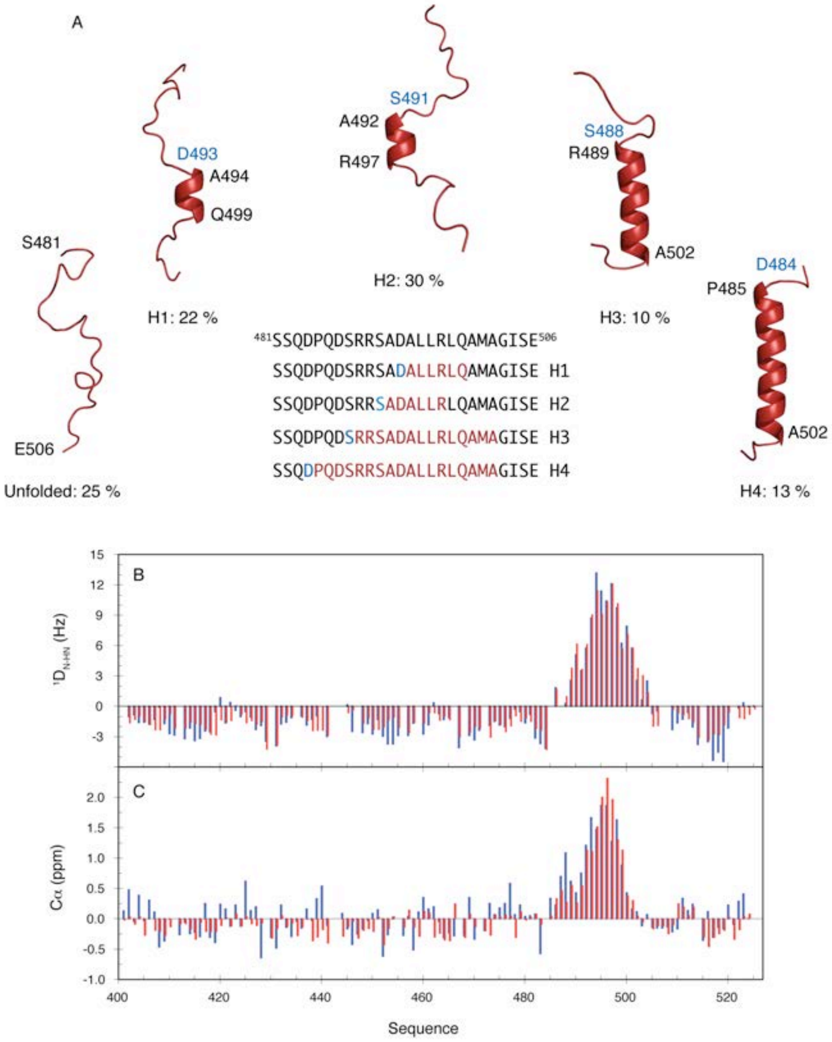

Figure 45: (A) Conformational equilibrium in the molecular recognition element of Measles virus N_{TAIL} obtained from the minimal ensemble approach applied to experimental RDCs and $C\alpha$ chemical shifts. The primary sequence shows the location of the helical elements preferentially being sampled in solution. All the selected helical elements are stabilized by N-capping interactions through aspartic acids or serine residues (blue residues). (B) Agreement between experimental ^{15}N-$^{1}H^{N}$ RDCs (blue) and back-calculated RDCs (red) from the conformational equilibrium shown in A. (B) Agreement between $C\alpha$ secondary chemical shifts (blue) and back-calculated secondary chemical shifts (red) from the conformational equilibrium shown in A. Only RDCs and chemical shifts of residues in the molecular recognition element were included in the fitting procedure.

Figure 45 shows the dynamic equilibrium adopted by the molecular recognition element of Measles virus N_{TAIL}. Four helices are preferentially being populated in solution in exchange with completely disordered conformations (25%). Two helices are rather short but highly populated (22% and 30%), while the longer helices are less populated (10% and 13%). All the selected helices in the minimal ensemble approach are preceded by aspartic acids or serines, again identifying N-capping interactions as important stabilizers of helices in IDPs and partially folded proteins. Excellent agreement is obtained between the experimental RDCs and chemical shifts and those back-calculated from the selected ensemble (Figure 45).

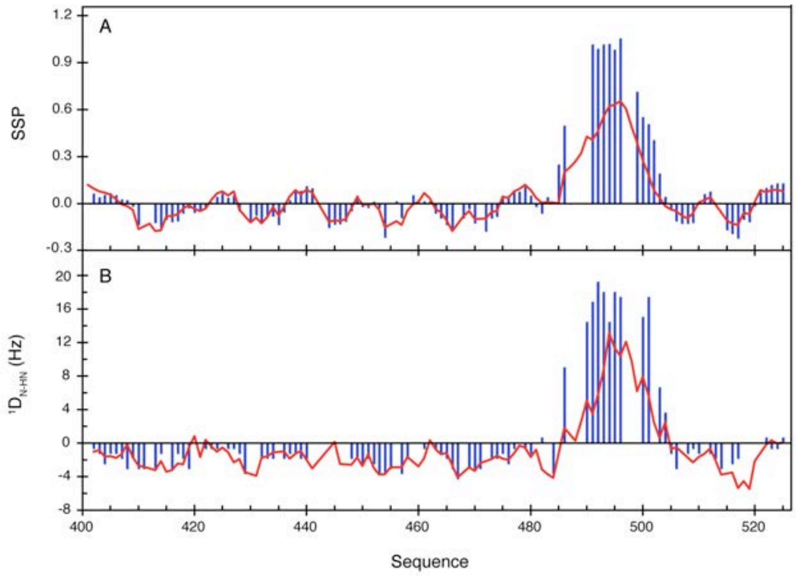

Figure 46: (A) Secondary structure propensity (SSP) of Measles virus N_{TAIL} obtained from experimental Cα and Cβ chemical shifts of N_{TAIL} alone (red) and in complex with XD (blue). (B) ^{15}N-^{1}H^{N} RDCs obtained in PEG/hexanol liquid crystals of N_{TAIL} alone (red) and in complex with XD (blue).

By comparing the conformational sampling of N_{TAIL} in its free state with the structure adopted in the complex with XD, interesting observations regarding molecular recognition in the N_{TAIL}/XD system can be made. The helices present in the molecular recognition element of N_{TAIL} in the free state ensure maximum helical propensity on the residues A494, L495, and L498. These residues are directly involved in the hydrophobic interaction with the partner protein XD as observed in the crystal structure (Figure 44). This demonstrates how helical propensity can be maximized for efficient interaction through strategically placed aspartic acids and serines in the primary sequence. Another interesting observation is that N_{TAIL} adopts a helix between residues Q486 and A502 in the crystal form of the complex. This helix corresponds to the longest helix sampled by N_{TAIL} already in its free pre-recognition state supporting conformational selection as a dominant molecular recognition mechanism (*132*, *133*).

We investigated the interaction between N_{TAIL} and XD by chemical shift titration experiments (*134*). As XD is added, line broadening and chemical shift changes are observed of the N_{TAIL} resonances in the molecular recognition element in agreement with a fast-to-intermediate exchange regime, and a dissociation constant of 10-50 μM was estimated on the basis of these measurements (*134*). Saturation of N_{TAIL} with XD allowed the partial assignment of the complex and the measurement of ^{15}N-^{1}H^{N} RDCs (Figure 46). The chemical shifts and RDCs show that N_{TAIL} undergoes α-helical folding upon binding to XD and that the first 80 amino acids of N_{TAIL} are not implicated in the interaction with XD. It is worth noting that even for a large excess of XD compared to N_{TAIL} (1:10), line broadening due to intermediate exchange persists for residues surrounding the two smallest helices (H1 and H2) selected by the minimal ensemble approach. This indicates that N_{TAIL} undergoes a coupled folding and binding reaction (induced fit) pointing towards a more dynamic complex than that represented by the crystal structure. Thus, both conformational selection and induced fit could play a role in molecular recognition in the N_{TAIL}/XD system.

7.4. Modelling N_{TAIL} in intact Measles virus nucleocapsids

The interaction between N_{TAIL} and XD did not reveal a functional role of the first 80 amino acids of N_{TAIL} and we therefore proceeded with a characterization in situ of the N_{TAIL} domain. Electron microscopy (EM) studies of the nucleocapsids show that they adopt a characteristic herringbone appearance (*135–138*). Both the structure and dynamics of the nucleocapsids appear to be strongly modulated by the disordered N_{TAIL}, where intact nucleocapsids have open and flexible structures that become more compact and rigid upon cleavage of the disordered tail by trypsin (*139*, *140*) (Figure 47).

In order to shed light on how N_{TAIL} modulates the morphology of the nucleocapsids and to probe the conformational behaviour and flexibility of N_{TAIL} in situ, we have recorded solution NMR spectra of ^{15}N, ^{13}C-labeled nucleocapsids (Figure 47). From EM we estimate the molecular mass distribution of the objects in the NMR sample to fall in a range between 2 to 50 MD that would normally preclude detection of solution state NMR signals of a folded, globular protein. A comparison of the ^{1}H-^{15}N HSQC spectra of the isolated N_{TAIL} domain and the intact nucleocapsids shows that the NMR resonances superimpose well, demonstrating that the conformational behaviour of the residues 450-525 of N_{TAIL} is retained in situ (Figure 47). This includes the helical sampling of the molecular recognition element determined by RDCs and chemical shifts in the isolated N_{TAIL} domain. An intensity profile of the nucleocapsid spectrum shows that the first 50 amino acids (residues 401-450) of N_{TAIL} are absent from the HSQC spectrum (Figure 48). The large variations in the intensities of the visible resonances indicate differential flexibility along the chain of N_{TAIL}.

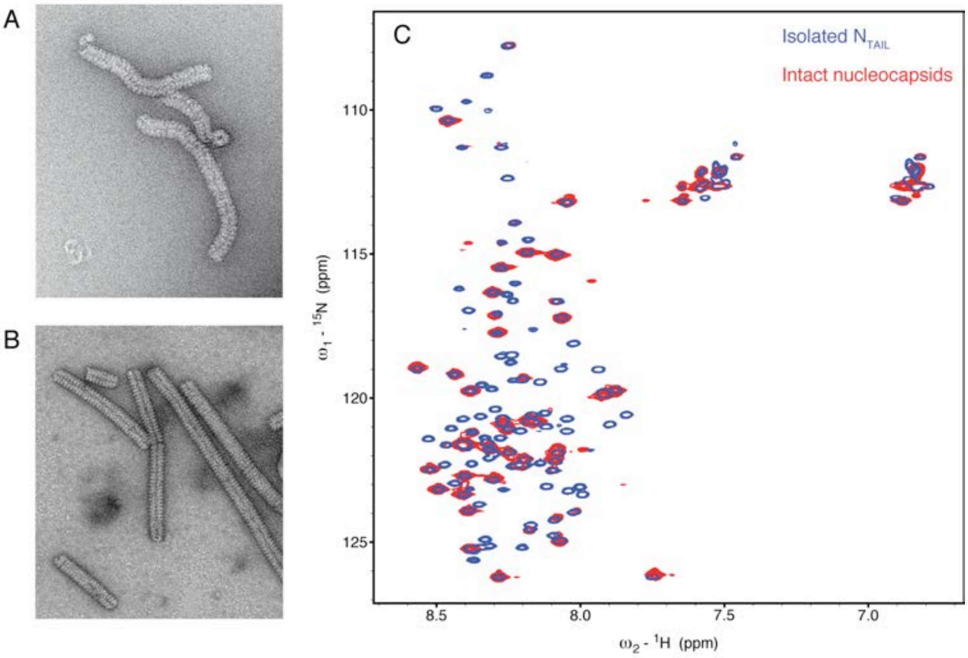

Figure 47: (A) Electron micrograph of the ^{13}C, ^{15}N-labeled intact Measles virus nucleocapsids used for NMR. (B) Electron micrograph of ^{13}C, ^{15}N-labeled cleaved nucleocapsids (no N_{TAIL}) (C) Superposition of the ^{1}H-^{15}N HSQC spectra of the isolated N_{TAIL} domain (blue) and intact nucleocapsids (red).

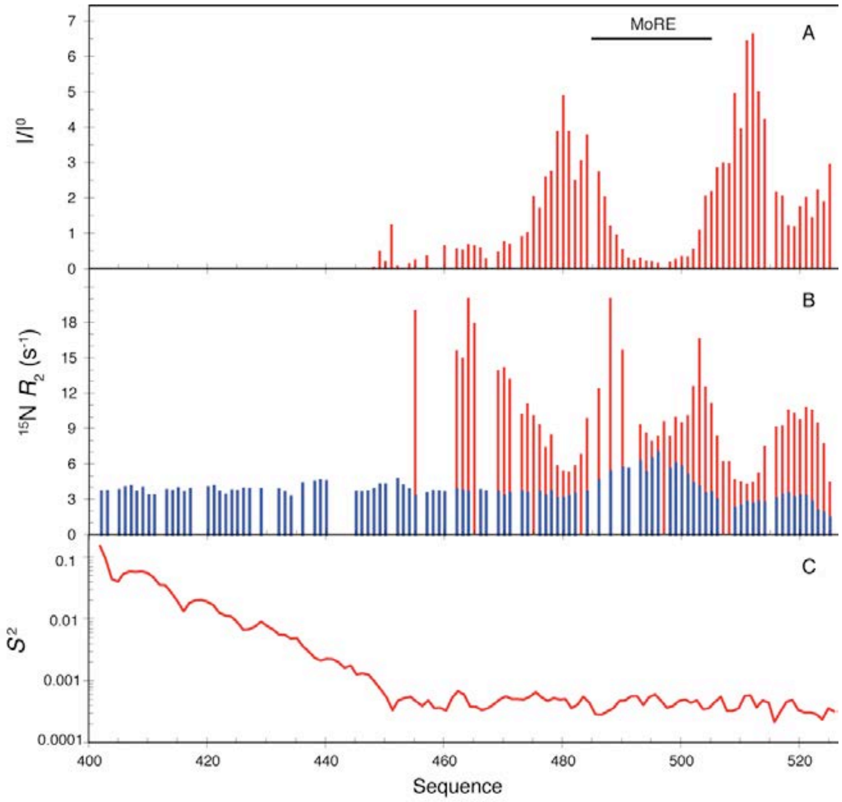

Figure 48: (A) Intensity profile of the ^{1}H-^{15}N HSQC spectrum of intact nucleocapsids calculated as the ratio of intensities in the nucleocapsid spectrum (I) and in the spectrum of the isolated N_{TAIL} domain (I^0). (B) ^{15}N R_2 relaxation rates in isolated N_{TAIL} (blue) and in situ (red). (C) Calculated order parameters of N_{TAIL} from the proposed model of intact nucleocapsids shown in Figure 50.

In order to further probe the conformational dynamics of N_{TAIL} in situ, we measured ^{15}N R_2 relaxation rates and compared them to the rates of the isolated N_{TAIL} domain (Figure 48). N_{TAIL} alone shows almost uniform relaxation rates except in the molecular recognition element where slightly elevated rates are observed due to the presence of residual helical structure. The relaxation rates of N_{TAIL} in situ increase significantly around residue 460 indicating a decrease in flexibility when going towards the folded domains forming the nucleocapsid. This is in agreement with the lack of signals in the ^{1}H-^{15}N HSQC spectrum for the first 50 amino acids of N_{TAIL}. Furthermore, the relaxation rates measured in the centre of the molecular recognition element (around residue 495) are almost identical in the isolated N_{TAIL} domain and in the nucleocapsids. This suggests that this part of the protein is in slow exchange on and off the surface of the nucleocapsid resulting in the large reduction in intensities for the residues in this region. It is estimated from the intensities that around 95% of the N_{TAIL} molecules are bound to the nucleocapsid surface via the molecular recognition element. On the border of the molecular recognition element, the rates measured of N_{TAIL} in situ are significantly larger than those measured in isolated N_{TAIL}, indicating that the exchange rate on and off the nucleocapsid surface appears faster (smaller chemical shift differences) for these sites. A final observation includes the C-terminus of N_{TAIL} that also appears to contact the nucleocapsid, either directly or by folding back onto the molecular recognition element as it interacts with the nucleocapsid surface.

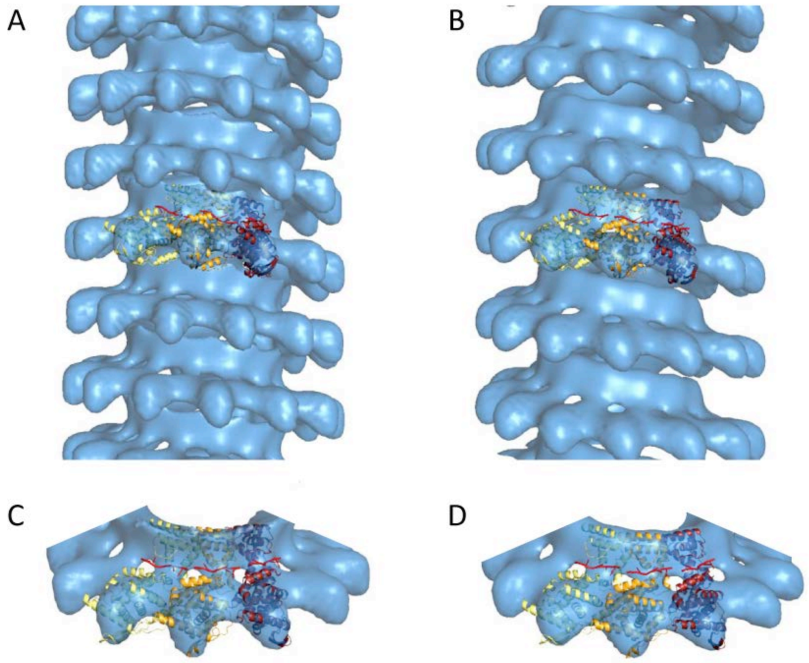

Figure 49: EM reconstructions of cleaved (A) and intact (B) Measles virus nucleocapsids. The atomic resolution crystal structure of RSV nucleoprotein was docked into the electron density of both cleaved (C) and intact (D) Measles virus nucleocapsids. The docking reveals that N_{TAIL} exits the nucleocapsid from the helix interior. The figure was adapted from Desfosses et al., J. Virol. (2011), 85, 1391-1395.

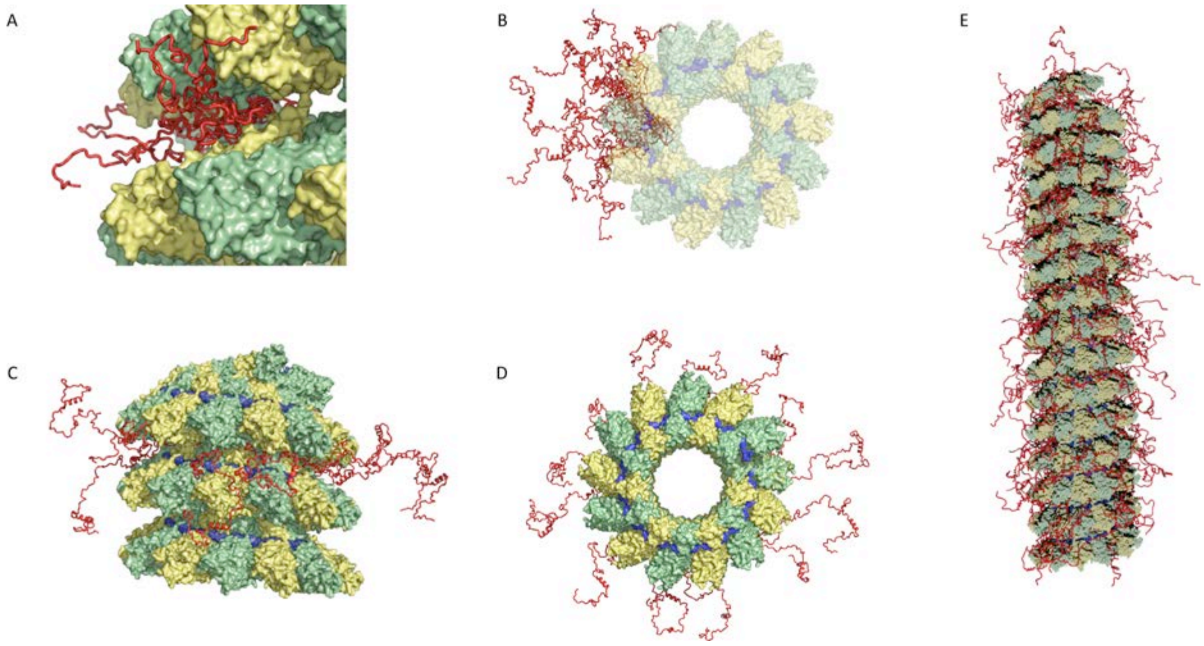

Figure 50: Proposed model of the location of N_{TAIL} in intact Measles virus nucleocapsids. (A) Representation of the conformational sampling of N_{TAIL} from a single N protein in the nucleocapsid. Different copies of the first 50 amino acids of N_{TAIL} (red) are shown to indicate the available volume sampling of the chain. (B) Representation of the conformational sampling of N_{TAIL} from a single N protein in the nucleocapsid, shown along the nucleocapsid axis. (C) Representation of the thirteen N_{TAIL} conformers from a single turn of the nucleocapsid. (D) Representation of the thirteen N_{TAIL} conformers from a single turn of the nucleocapsid shown along the helix axis. (E) Intact nucleocapsid showing the N_{TAIL} molecules in red. In the interests of clarity, (B–E) deliberately show more conformers outside the nucleocapsid, and fewer bound to the surface, than are probable at any one time. The position of the RNA is shown in blue.

In a recent study, EM reconstructions of both intact and cleaved nucleocapsids were obtained showing that each turn contains thirteen N subunits and that the nucleocapsid diameter decreases from 200 to 190 Å and the helical pitch shortens from 57.2 Å to 48.7 Å upon removal of N_{TAIL} (*141*) (Figure 49). In that study, no information could be obtained about the location and conformational state of N_{TAIL} in the nucleocapsids, because N_{TAIL} did not contribute coherently to the reconstructed density from EM. The atomic resolution structure of N_{CORE} is unknown, however the crystal structure of the N-RNA complex of another Paramyxovirus, namely respiratory syncytial virus (RSV), was solved recently (*142*). A docking of this structure into the reconstructed density of the Measles virus nucleocapsids surprisingly places the N-terminal end of N_{TAIL} at the interior of the nucleocapsid helix (*141*). Due to steric hindrance, the 13 copies of N_{TAIL} per turn cannot reside in the interior of the nucleocapsid and remain flexible enough to give rise to NMR signals.

In order to provide a model of N_{TAIL} in the context of the intact nucleocapsids, we used the EM reconstruction of the Measles virus nucleocapsids with the docked atomic resolution structure of the RSV nucleoprotein, to add conformers obeying random coil statistics of N_{TAIL} using Flexible-Meccano. A simple steric exclusion was imposed to avoid any contacts between N_{TAIL} and the folded N_{CORE} domains. The model building shows that there is enough space for N_{TAIL} to escape from the inside to the outside of the nucleocapsid via the interstitial space between successive N_{CORE} lobes (Figure 50). From the model we calculated an order parameter S^2 of N_{TAIL} reporting on the degree of orientational freedom of the unfolded chain. The order parameter decreases for the first 50 amino acids of the chain and then becomes approximately constant as N_{TAIL} exits from the interior of the nucleocapsid (Figure 48), providing a reasonable explanation for the lack of resonances for the first 50 amino acids of N_{TAIL} in the HSQC spectrum of the nucleocapsids.

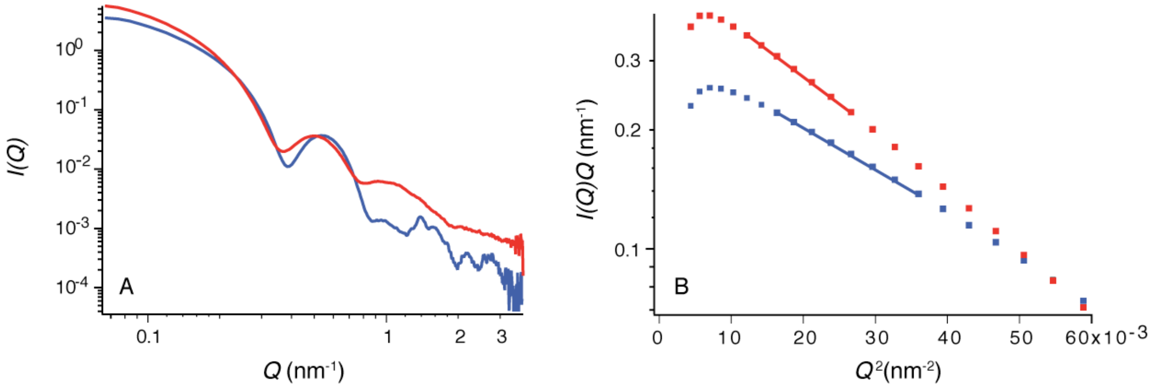

Figure 51: (A) Small angle X-ray scattering profiles of intact (red) and cleaved (blue) nucleocapsids. (B) Analysis of the scattering data according to Eq. 18 of the intact (red) and cleaved (blue) nucleocapsids.

As a validation of our model, we carried out SAXS experiments of cleaved and intact nucleocapsids with the aim of determining the cross-sectional radii of gyration (Figure 51). The scattering curves in both cases were analyzed in terms of rod-like particles where the length was assumed to be much larger than the diameter (*143*):

$$\ln(I(Q)Q) = \ln(I(0)Q) - \frac{1}{2}R_C^2Q^2 \qquad \text{(Eq. 18)}$$

The cross-sectional radius of gyration, R_C, was obtained to be (78.0 ± 0.6) Å and (69.5 ± 2.4) Å for intact and cleaved nucleocapsids, respectively (Figure 51). From our model we can explicitly calculate the expected cross-sectional radius of gyration using the radial coordinates r_i of N atoms in a unit sectorial element around the cylindrical axis of symmetry:

$$R_C^2 = \frac{1}{N}\sum_i r_i^2 \qquad \text{(Eq. 19)}$$

We obtain a value of 68.0 Å for the cleaved nucleocapsid, which is in excellent agreement with the value derived from the SAXS experiment. For the intact nucleocapsid, where N_{TAIL} was built using the Flexible-Meccano algorithm, we obtain a cross-sectional radius of gyration of 83.8 Å. By retaining only the Flexible-Meccano conformers where the molecular recognition element is within 8 Å of any N_{CORE} domain, a value of 78.4 Å is obtained. The SAXS experiments therefore substantiate our model and support the fact that the molecular recognition element of N_{TAIL} is not entirely free, but stays close to the surface of the nucleocapsid a large percentage of the time.

The presence of N_{TAIL} in the interstitial space between the helical lobes explains how N_{TAIL} modulates the structure and dynamics of the nucleocapsids as visualized by EM (Figure 47). N_{TAIL} appears to have two functional roles. It forms an articulated spacer that allows the molecular recognition element to escape from the inside to the outside of the nucleocapsid and at the same time provides access to the viral RNA by increasing the helical pitch. In addition, the molecular recognition element exchanges on and off the nucleocapsid surface, where approximately one N_{TAIL} molecule out of thirteen per helical nucleocapsid turn is free in solution at any given time. While we currently have no information about the position of the binding site, or whether this binding is specific, we can speculate that such a mode of action would provide an efficient mechanism by which N_{TAIL} could "catch" the viral polymerase complex when in free solution, and co-localize the complex on the nucleocapsid surface, thereby initiating transcription and replication of the viral RNA.

7.5. Conclusions

NMR studies of IDPs as isolated domains in vitro facilitate the characterization of the proteins at atomic resolution. The results presented in this chapter, however, underline the importance of investigating disordered domains in more physiological environments for example as integral parts of their high molecular weight assemblies. In the case of the disordered N_{TAIL} domain of the Measles virus nucleoprotein, we carried out a structural characterization of the isolated domain and compared it to that obtained in the context of the intact Measles virus nucleocapsids. The study in situ adds significantly to our understanding of the role of intrinsic disorder in transcription and replication of the virus. The disordered domain not only interacts with the viral polymerase complex, but also facilitates the access to the viral RNA by ensuring open and dynamic nucleocapsids.

The study of the intact Measles virus nucleocapsids presents a nice example of how several complementary techniques can be combined to resolve more complex problems. In this case, we used NMR for obtaining atomic resolution information about the conformational sampling and dynamics of N_{TAIL} in situ. We applied EM to probe the difference in appearance between cleaved and intact nucleocapsids and the corresponding reconstruction as a model for the folded domain of the nucleoprotein. SAXS was used to probe dimensions of the nucleocapsids in solution, in particular, the contribution to the scattering from the part of N_{TAIL} located on the outside of the nucleocapsids. Combining the different techniques provided us with an integral model of the intact nucleocapsids.

Future perspectives

8. The role of intrinsic disorder in Paramyxoviruses

8.1. Introduction

Understanding the role and functional implications of intrinsic disorder requires atomic resolution models of the proteins and their dynamic complexes. Due to the inherent flexibility of IDPs, NMR spectroscopy is the method of choice for characterizing both the structure and dynamics of the proteins at atomic resolution. The previous chapters of this habilitation have described the application of the ensemble selection algorithm ASTEROIDS for characterizing both local and long-range structure in IDPs from various NMR parameters such as RDCs, chemical shifts and PREs. Even for large molecular assemblies, as for example the viral nucleocapsids, NMR can be used to detect the signals of the flexible domains. NMR is therefore highly complementary to other techniques such as X-ray crystallography and electron microscopy that rarely observe these important, and often functional, domains. Future perspectives of the work described in this habilitation include the development of NMR methods for characterizing the structure, dynamics and kinetics of IDP complexes. In addition, we will apply the developed methods and ensemble descriptions to elucidate the role of intrinsic disorder in two novel biological systems, namely the Paramyxoviruses (this chapter) and the mitogen activated protein kinase (MAPK) cell signalling pathways (see chapter 9).

So far, we have made significant progress in understanding the role of intrinsic disorder in transcription and replication of Paramyxoviruses. We have characterized the isolated N_{TAIL} domains of Sendai and Measles virus nucleoproteins, and we have obtained a description of N_{TAIL} in the context of intact Measles virus nucleocapsids. In future studies, we will develop methods for characterizing the structure, dynamics and kinetics of the complex formed between N_{TAIL} and the C-terminal domain, XD, of the phosphoprotein. We will use relaxation dispersion measurements for probing the kinetics of the binding reaction, as well as for obtaining NMR parameters that describe the structure and dynamics of N_{TAIL} within the complex. We will characterize the N_{TAIL} domains of several members of the Paramyxovirus family, both as isolated domains and in context of intact nucleocapsids. In general, we aim at comparing the conformational behaviour of the disordered domains and their interactions across several members of the Paramyxovirus family such as Sendai, Measles, Hendra and Nipah viruses.

8.2. Structure, dynamics and kinetics of the N_{TAIL}/XD complex

One of the main difficulties encountered when studying protein-protein complexes involving IDPs is that the coupled folding-and-binding reaction and the dynamic behaviour of the proteins often result in exchange broadening of the NMR signals. This is also the case for the interaction between

N_{TAIL} and XD of both Sendai and Measles virus that show a fast-to-intermediate exchange regime, resulting in extreme line broadening of most of the NMR resonances in the molecular recognition element of N_{TAIL}. This exchange regime, however, holds great promise for relaxation dispersion measurements, where the complex kinetics and the bound form of N_{TAIL} are studied via measurements on the resonances of the free protein in the presence of sub-stoichiometric amounts of XD. We will attempt to tune the exchange rate to allow the observation of relaxation dispersion arising from the complex formation by variation of the concentrations of N_{TAIL} and XD as well as the temperature. In order to monitor more precisely the dynamics of N_{TAIL} in the complex, we will construct a chimeric N_{TAIL}/XD complex in order to identify the contribution from the binding event of N_{TAIL} to the relaxation dispersion curves. Using these methods, we will compare the binding kinetics and the dynamics of different N_{TAIL}/XD complexes within the Paramyxovirus family (Figure 52).

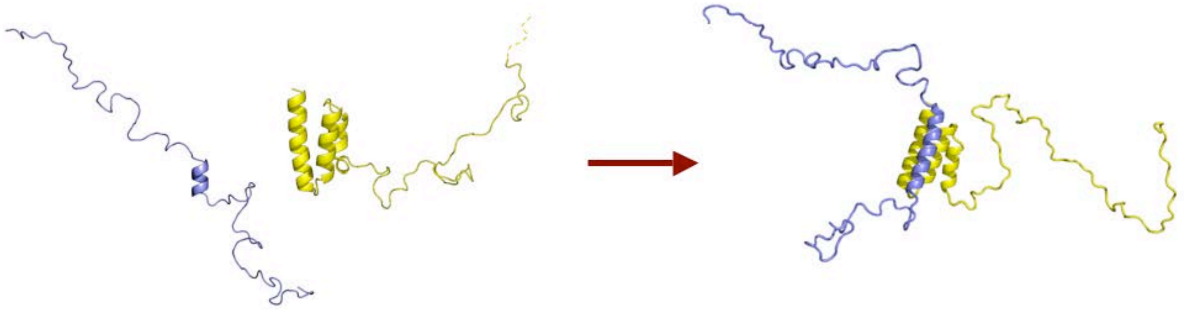

Figure 52: Model of the interaction between N_{TAIL} (blue) and XD (yellow) of Sendai, Measles, Hendra and Nipah viruses. The molecular recognition element of N_{TAIL} undergoes α-helical folding upon binding to XD. The binding kinetics and the structure and dynamics of N_{TAIL} in the complex will be studied through relaxation dispersion measurements.

The characterization of N_{TAIL} from Sendai and Measles virus using RDCs revealed that the molecular recognition element samples specific α-helical conformers stabilized by N-capping interactions. We will modify the pre-recognition state of N_{TAIL} by mutation of N-capping residues and, thereby, map the influence of specific helical conformers and their population on the kinetics of the complex formation. So far, we have performed a simultaneous mutation of two of the N-capping residues (D475A and D478A) in Sendai virus N_{TAIL} that are responsible for stabilizing more than 60% of the helical propensity (Figure 40). Our preliminary RDC data show that the mutations result in only a small decrease of helical propensity in the molecular recognition element, however, the conformational equilibrium changes such that the helical conformers are stabilized largely by S477 and to a smaller extent by D473. We will study by relaxation dispersion how this change in conformational equilibrium affects the binding kinetics of N_{TAIL}.

The intermediate exchange regime seems to be highly prevalent in interactions involving IDPs because of folding-upon-binding reactions and/or the existence of pervasive dynamics in the complexes. As discussed above, relaxation dispersion measurements are expected to be helpful in this context, however, we will also investigate the possibility of overcoming the intermediate exchange regime by studying IDP complexes at very low temperatures. We will pursue the study of the N_{TAIL}/XD complex in super-cooled water, where temperatures as low as -20°C can be accessed without freezing the protein solutions (*144–146*).

In addition to the studies proposed above for the N_{TAIL}/XD complex, we will explore the possibility of using paramagnetic metal ions, e.g. lanthanide ions, as structural probes. Most of the lanthanides are paramagnetic and have strong anisotropic magnetic susceptibilities inducing both pseudocontact shifts (PCSs), RDCs and PREs (*147*). The advantage of PCSs is that they are easy to measure and they depend not only on the distance to the paramagnetic metal ion, but also on the orientation of the metal-nucleus vector with respect to the magnetic susceptibility tensor (*148*). Paramagnetic lanthanide ions span a wide range of magnetic anisotropies providing the perspective of obtaining structural information in different distance ranges from the paramagnetic ion. For example, Ce^{3+} normally provides useful PCSs and relaxation enhancements in the 5-15 Å range from the metal ion, Yb^{3+} in the 9-25 Å range and Dy^{3+} in the 13-40 Å range (*149*). Thus, lanthanides are expected to extend the range of distances that can be measured, compared to for example MTSL spin labels, as well as introducing new independent orientational information in the IDPs.

The introduction of paramagnetic metal ions in IDPs requires a metal binding tag that can be attached to the side chain of cysteine residues. Several suitable, high-affinity lanthanide tags have been proposed that when attached to folded proteins induce sizable PCSs and RDCs (*150–158*). The feasibility of inducing PCSs and RDCs in IDPs via attachment of lanthanide binding tags has recently been demonstrated in urea-denatured ubiquitin, where a Dy^{3+}-loaded DOTA-derived metal binding tag attached at residue 6 induced sizable PCSs for residues 17-30, while residues 1-16 were broadened beyond detection due to strong PREs (*155*).

The application of paramagnetic metal ions for obtaining structural information is potentially even more powerful in studies of the structure and dynamics of IDP complexes. We will explore this possibility in the case of the N_{TAIL}/XD complex by attaching a lanthanide binding tag to XD and measure the paramagnetic effects in N_{TAIL}. In this way, we can increase the sensitivity of chemical shift titrations or relaxation dispersion type experiments by amplifying the chemical shift changes that occur upon complex formation due to the presence of an appropriately placed lanthanide ion.

8.3. Studies of nucleocapsids of Paramyxoviruses by NMR, EM and SAXS

The study of the conformational behaviour of N_{TAIL} in intact Measles virus nucleocapsids by solution NMR has opened up new possibilities to study the mechanism of transcription and replication of the virus. Our study revealed that the molecular recognition element of N_{TAIL} interacts with the surface of the nucleocapsid a large percentage of the time. We will determine the factors that govern this interaction by screening different solution conditions (salt, pH, buffer etc) with the aim of changing the affinity of the interaction. The molecular recognition element of both Sendai and Measles virus N_{TAIL} contains a number of positively charged arginines that could potentially be interacting with the negatively charged phosphate groups of the RNA. We will perform mutational studies of the molecular recognition element of N_{TAIL} in order to identify the residues that are responsible for this interaction.

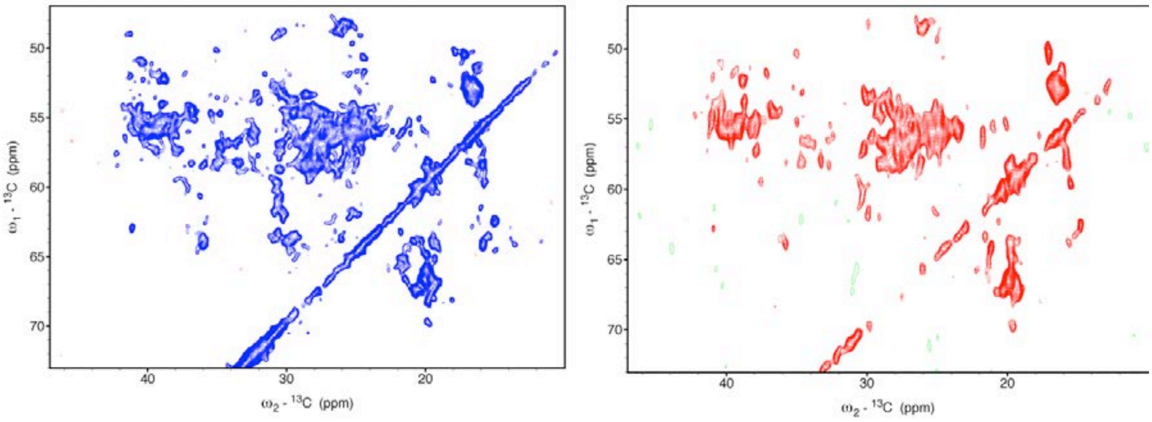

Figure 53: Expansions of the Cα-CX regions of the ^{13}C-^{13}C correlation spectra (DARR mixing of 25ms) recorded on trypsin-digested (blue) and intact (red) nucleocapsids. The experiments were acquired on a 1GHz spectrometer at 10kHz magic angle spinning.

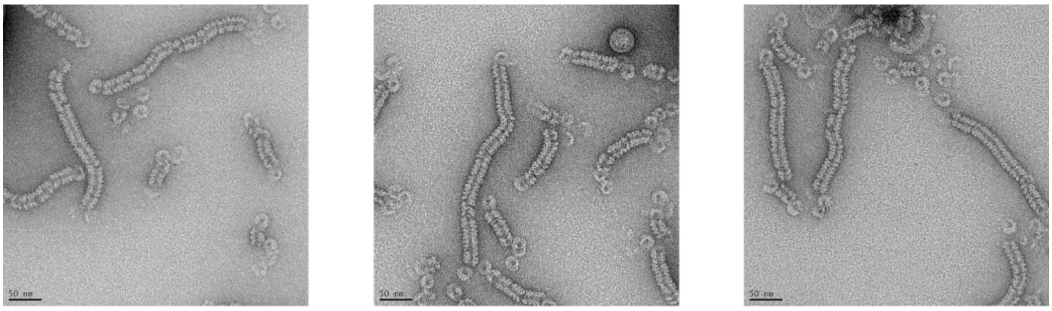

Figure 54: Electron micrographs of Sendai virus nucleocapsids revealing a striking similarity to Measles virus nucleocapsids (Figure 47).

To complement the solution NMR studies, we will carry out solid-state NMR of both intact and cleaved nucleocapsids (Figure 53). In the solid-state ^{13}C-^{13}C correlation spectra, the signals of the RNA (random cellular RNA) are visible, allowing us to investigate if the molecular recognition element of N_{TAIL} interacts with the RNA. This requires a comparison of the chemical shifts of the RNA in the cleaved and intact nucleocapsids, respectively.

In order to obtain further insight into the mechanism of transcription and replication of Measles virus, we will study the interaction between the nucleocapsids and XD (or full-length phosphoprotein), in order to probe which parts of N_{TAIL} stay flexible upon complex formation. In a more general picture, we will compare our findings in Measles virus nucleocapsids with other members of the Paramyxovirus family. So far, we have obtained nucleocapsids of Sendai and Hendra virus, and we are currently studying these by solution NMR and other complementary techniques (Figure 54).

9. Intrinsic disorder in MAPK cell signalling pathways

9.1. Introduction

Mitogen-activated protein kinases (MAPKs) are essential components of eukaryotic signal transduction networks that enable cells to respond appropriately to extracellular stimuli (mitogens). The MAPK signalling pathways feature three sequentially acting protein kinases making up a signalling module: an MKKK (MAPK kinase kinase) that phosphorylates and thereby activates an MKK (MAPK kinase), which then activates the MAPK by phosphorylation (*159*). In mammalian organisms four major MAPK cascades have been identified: ERK1/2 (ERK: extracellular-signal-regulated kinase), ERK5, p38 and JNK (*159*). The c-Jun N-terminal kinase (JNK) pathway primarily regulates stress and inflammatory responses, and the deregulation of this pathway has been shown to be implicated in the pathogenesis of many human diseases such as cancer (*160*), diabetes (*161*), Alzheimer’s (*162*) and Parkinson’s disease (*163*). Inhibition of JNK activity is considered as a possible therapy for many of these diseases.

Considering the large number of signalling pathways in mammalian organisms, one of the main questions is how signalling specificity is maintained allowing kinases to distinguish their correct substrates from a vast excess of incorrect substrates that contain similar target residues. As an example, we can consider a kinase that recognizes a substrate to be activated by phosphorylation at the target sequence Ser/Thr-Pro. This sequence is found in approximately 80% of all proteins and is, therefore, clearly insufficient to dictate whether or not a particular protein is a substrate of a given kinase (*164*).

Specificity in the MAPK signalling pathway appears to be largely controlled by intrinsically disordered regulatory domains of the kinases that selectively bind to the surfaces of the “right” substrates and/or by intrinsically disordered scaffold proteins that simultaneously bind the components of a specific pathway (*164–167*). Although many studies have mapped the large interaction network of each signalling pathway, a picture at the molecular level of the structural features governing signalling specificity is currently lacking. Here, we will focus on a single signalling module namely MLK3-MKK7-JNK1 and the associated scaffold protein JIP1. We will characterize the disordered regulatory domain of MKK7 and its interaction with JNK1, and we will investigate the role of JIP1 in mediating the interaction between MKK7 and JNK1. The overall aim of our study is to reveal the molecular basis for signalling specificity in the JNK pathway.

9.2. Specificity in the MKK7-JNK1 signalosome

Initially, we will focus on a single signalling module namely MLK3-MKK7-JNK1 with special emphasis on the interaction between MKK7 and JNK1. MKK7 is a 419 amino acid protein (isoform $\beta1$) that consists of the N-terminal regulatory domain (residues 1-100) and the protein kinase domain (residues 101-405). MKK7 is activated by phosphorylation at S271 and T275 and the structure of an activated mutant (S271D, T275D) of the kinase domain has been solved by X-ray crystallography (Figure 55). Very limited structural information is available for the regulatory domain of MKK7, and a prediction of the level of disorder in MKK7 using IUPRED (*168*) suggests that this domain is largely disordered (Figure 56). JNK1 is a 384 amino acid protein (isoform α1) comprising an N-terminal domain (residues 9-112 and 347-363) and a C-terminal domain (residues 113-337) that is activated by phosphorylation at T183 and Y185. A crystal structure has been solved of JNK1 in complex with a peptide from the scaffold protein JIP1 (Figure 55).

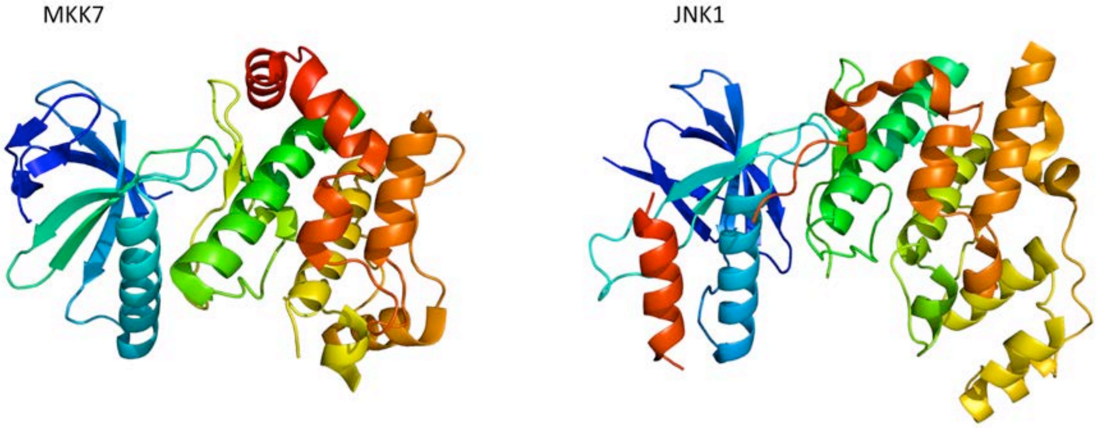

Figure 55: (Left) Crystal structure of human MKK7 activated mutant (S271D, T275D). The construct encompasses residues 101-405 of MKK7 (PDB: 2DYL). (Right) Crystal structure of the JNK1 α1 isoform in complex with a peptide from the JIP1 scaffold protein (peptide not shown in the representation) (PDB: 1UKH).

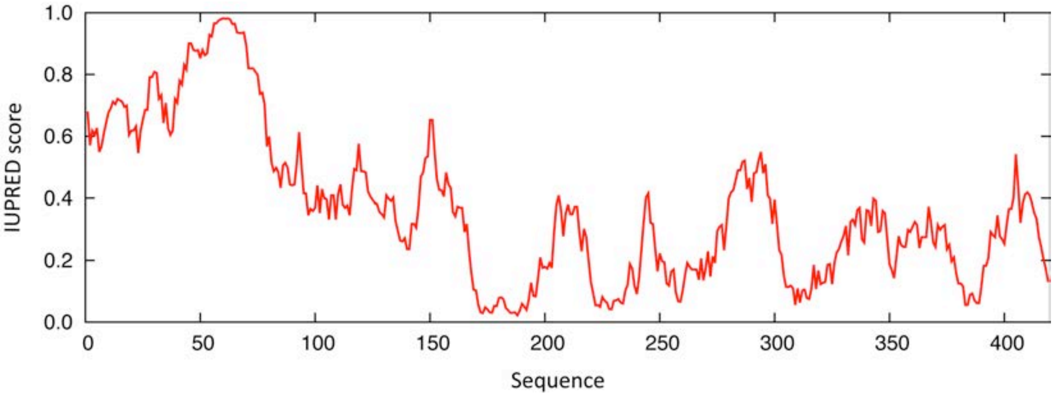

Figure 56: Disorder prediction of MKK7 using the IUPRED server. A score of 1 indicates disorder, while a score of 0 indicates order.

JNK1 is activated upon phosphorylation by MKK7, and signalling specificity is in some way controlled by the intrinsically disordered N-terminal regulatory domain of MKK7 that binds to a docking groove on the surface of JNK1 (Figure 57). In general, JNK1 interacts with proteins or substrates containing docking sites characterized by a core consensus sequence composed of a cluster of two to three basic residues, followed by a short spacer of 1-6 residues, and finally a hydrophobic-X-hydrophobic submotif ($\text{K/R}_{2\text{-}3}\text{-X}_{1\text{-}6}\text{-}\varphi\text{-X-}\varphi$) (*169*) (Figure 57). The hydrophobic residues are usually Leu, Ile or Val. The variability in the number and position of hydrophobic and basic residues within the docking site is known to determine specificity, however, the details at a molecular level are currently unknown (*170*, *171*).

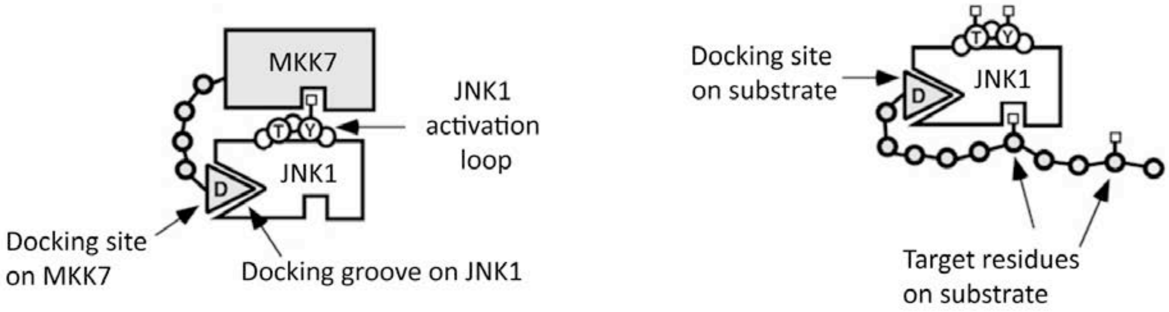

Figure 57: Specificity in MAPK signalling pathways is believed to be controlled by docking sites. (Left) The N-terminal regulatory domain of MKK7 docks onto the surface of its substrate JNK1, and the kinase domain of MKK7 phosphorylates the JNK1 activation loop. (Right) The substrates of JNK1 are recognized by docking sites in the substrates. Figure adapted from L. Bardwell, Biochem. Soc. Trans. (2006), 34, 837-841.

MKK7 ^{1}MAASSLEQKL ^{11}SRLEAKLKQE ^{21}NREARRRIDL ^{31}NLDISPQRPR ^{41}PTLQLPLAND
^{51}GGSRSPSSES ^{61}SPQHPTPPAR ^{71}PRHMLGLPST ^{81}LFTPRSMESI ^{91}EIDQKLQEIM

MKK4 ^{1}MAAPSPSGGG ^{11}GSGGGSGSGT ^{21}PGPVGSPAPG ^{31}HPAVSSMQGK ^{41}RKALKLNFAN
^{51}PPFKSTARFT ^{61}LNPNPTGVQN ^{71}PHIERLRTHS ^{81}IESSGKLKIS ^{91}PEQHWDFTAE

Figure 58: The sequence of the N-terminal regulatory domain of MKK7 and MKK4 (residues 1-100). Three docking sites are found within MKK7: D1: R25-L32, D2: R40-L47 and D3: R70-L77, while only one site is present in MKK4: K39-L46.

Three docking sites important for binding of JNK1 have been identified in the N-terminal regulatory domain of MKK7 (*172*) (Figure 58). The three sites appear to work in partial cooperation for the binding of JNK1, although the exact mechanism is not known. Two possibilities to consider are an additive mechanism versus a synergistic mechanism. In an additive mechanism, the different docking sites bind to the same docking groove of JNK1, and the effect is comparable with tripling the concentration of a single docking site. In a synergistic mechanism, the three docking sites bind to distinct docking grooves on JNK1 (*172*).

JNK1 is also regulated by another MAPK kinase namely MKK4. Optimal activation of JNK1 requires the activity of both kinases, although both are capable of dual phosphorylation of JNK1 at the activation loop threonine and tyrosine residues. MKK4 prefers the tyrosine and is primarily activated by environmental stresses, while MKK7 prefers the threonine and is primarily activated by cytokines (*173–175*). Only one docking site appears to be present within the regulatory domain of MKK4 (Figure 58). Interestingly, MKK4 can also activate the p38α kinase in the p38 signalling cascade (*176*) and, therefore, appears not to be particularly pathway specific.

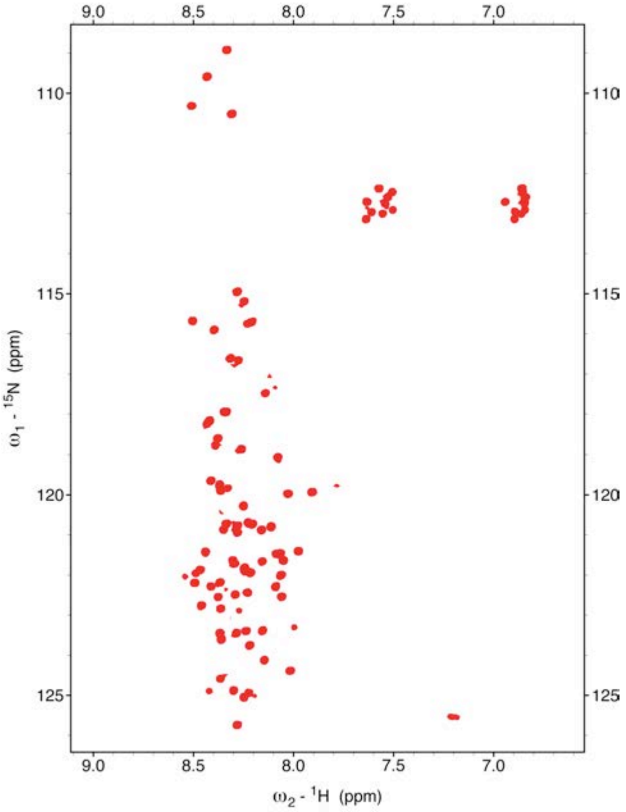

Figure 59: ^{1}H-^{15}N HSQC spectrum of a construct of MKK7 comprising residues 1-73. The limited ^{1}H spectral dispersion shows that the protein is disordered.

In order to obtain insight into how specificity is achieved using docking sites, we will study the N-terminal regulatory domains of MKK4 and MKK7 by NMR chemical shifts, RDCs and PREs and obtain representative ensemble descriptions of the domains. So far we have expressed and purified the construct 1-73 of MKK7 showing that this domain is indeed disordered (Figure 59). Assignment of a longer construct containing all three docking sites will follow. We will compare the conformational sampling of the different docking sites in MKK4 and MKK7 and perform the interaction with JNK1 using NMR titration experiments. In the case of MKK7 that contains three docking sites, we aim at determining the relative importance of the different docking sites, deciphering the binding events as well as determining the mechanism of interaction.

As a counterpart of the docking sites, JNK1 has a docking groove that regulates the docking specificity (Figure 57). The docking groove is located on the opposite side of the substrate recognition site. To obtain further insight into the docking specificity, we will explore the possibility of solution NMR studies of JNK1 using standard approaches such as per-deuteration combined with TROSY techniques or, if necessary, specific labelling of methyl groups combined with methyl TROSY NMR. We will monitor the binding of the regulatory domains of MKK4 and MKK7 to identify the surface of JNK1 involved in the binding as well as the residues controlling specificity within the different docking sites.

The crystal structure of JNK1 obtained in complex with the docking site of the scaffold protein JIP1 reveals that upon binding of the peptide, a change of about 15° occurs in the relative orientation of the N- and C-terminal domains of JNK1 (*177*). A recent crystal structure of the same complex (PDB: 3O17) shows another set of domain orientations. For this reason, it is not entirely clear how much domain motion occurs upon binding of different docking sites. We will attempt to study these domain motions using NMR relaxation and RDCs complemented by SAXS measurements of JNK1 in its free form and in complex with different docking sites. The domain motions are believed to be important for the activity of the kinase. The binding site for ATP is partially distorted upon domain reorganization leading to a decrease in the affinity of ATP binding to JNK1 with about a factor of 3 (*177*).

9.3. Characterizing the disordered scaffold protein JIP1 and its interactions

The JNK-interacting protein-1 (JIP1) was initially characterized as an inhibitor of the JNK signalling pathway following the observation that when over-expressed in cells it prevented JNK from phosphorylating its substrates such as c-Jun (*178*). This led to the development of peptide inhibitors of JNK based on the sequence of JIP1 (*177*, *179*). Subsequently, it was found that JIP1 also binds to MKK7 and members of the MLK (mixed-lineage kinase) family, and thereby could enhance JNK signalling when the scaffold protein was not in excess compared to the JNK components of the pathway (*180*). JIP1 is very likely an important regulator of JNK signalling being on one hand an inhibitor and on the other hand an activator depending on the protein levels in the cell.

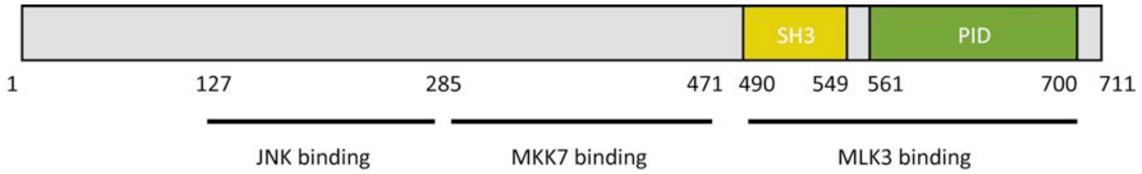

Figure 60: The domain organization of JIP1. Two domains have been identified, an SH3 domain and a PID, that are both located in the C-terminal part of the protein. The regions for the binding of the signalling module have been identified. JNK binds to JIP1 via the 127-285 region, while MKK7 and MLK3 bind to the central and C-terminal region of JIP1, respectively.

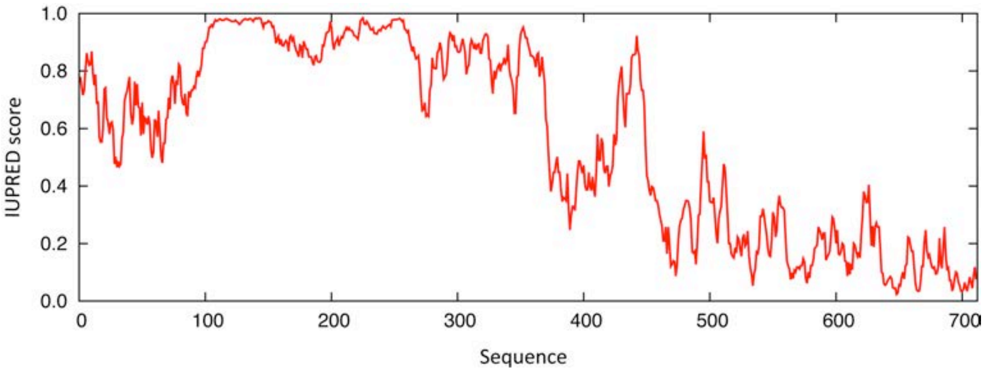

Figure 61: Disorder prediction of human JIP1 using the IUPRED server showing that the first 450 amino acids of JIP1 are largely disordered.

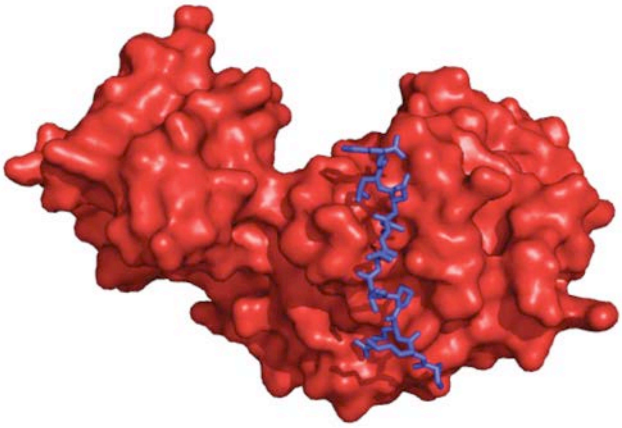

Figure 62: Crystal structure of JNK1 in complex with the docking site of JIP1. The surface of JNK1 is shown in red, while the docking site (sequence: PKRPTTLNLF) of JIP1 is shown in blue sticks.

Human JIP1 is a 711-residue protein with a domain organization as shown in Figure 60. The protein contains two well-defined domains: Src homology 3 (SH3) and a phosphotyrosine interaction domain (PID). Structural information of JIP1 is limited to the SH3 domain through which JIP1 has been shown to dimerize (*181*). A prediction of disorder using IUPRED reveals that the first 450 amino acids of JIP1 most likely are disordered (Figure 61). The interaction between JNK1 and JIP1 has been mapped to the region 127-285 of JIP1 with the interaction being mediated by a specific, high-affinity docking site located at residues 157-167 (*177*, *178*) (Figure 62). The interaction between MKK7 and JIP1 has been mapped to the region 285-471 of JIP1, while the last component of the signalling module MLK3 binds to the region of JIP1 containing the SH3 and PID domains (*178*). We will characterize JIP1 using NMR with special emphasis on the region possessing large amounts of intrinsic disorder. We will obtain ensemble descriptions on the basis of various NMR parameters, and we will compare the conformational sampling of the docking site of JIP1 with those of MKK7 and MKK4. We will perform interaction studies between JIP1 and JNK1 as well as between JIP1 and

MKK7 to locate more precisely the binding site for MKK7. MKK4 is reported not to bind to JIP1 (*182*) and we will verify this by NMR.

We will attempt to assemble the signalosome consisting of JIP1, JNK1 and MKK7 and determine the binding affinities as well as the order of binding of JNK1 and MKK7 using NMR. It is clear that with the presence of several docking sites within a single signalosome, the order of binding of the different kinases must be important. We will also focus on elucidating the role that the dimer formation of JIP1 plays in the assembly of the JIP1-JNK1-MKK7 signalosome. If a stable signalosome can be assembled we will investigate the overall architecture using SAXS.

10. References

1. Tompa, P. (2002) Intrinsically unstructured proteins, *Trends Biochem. Sci. 27*, 527-533.
2. Uversky, V. N. (2002) Natively unfolded proteins: a point where biology waits for physics, *Protein Sci. 11*, 739-756.
3. Dunker, A. K., Silman, I., Uversky, V. N., and Sussman, J. L. (2008) Function and structure of inherently disordered proteins, *Curr. Opin. Struct. Biol. 18*, 756-764.
4. Dyson, H. J., and Wright, P. E. (2005) Intrinsically unstructured proteins and their functions, *Nat. Rev. Mol. Cell Biol. 6*, 197-208.
5. Dyson, H. J. (2011) Expanding the proteome: disordered and alternatively folded proteins, *Q. Rev. Biophys.* 1-52.
6. Huang, A., and Stultz, C. M. (2009) Finding order within disorder: elucidating the structure of proteins associated with neurodegenerative disease, *Future Med. Chem. 1*, 467-482.
7. Uversky, V. N. (2009) Intrinsic disorder in proteins associated with neurodegenerative diseases, *Front. Biosci. 14*, 5188-5238.
8. Fink, A. L. (2005) Natively unfolded proteins, *Curr. Opin. Struct. Biol. 15*, 35-41.
9. Wang, J., Cao, Z., Zhao, L., and Li, S. (2011) Novel Strategies for Drug Discovery Based on Intrinsically Disordered Proteins (IDPs), *Int. J. Mol. Sci. 12*, 3205-3219.
10. Cheng, Y., LeGall, T., Oldfield, C. J., Mueller, J. P., Van, Y.-Y. J., Romero, P., Cortese, M. S., Uversky, V. N., and Dunker, A. K. (2006) Rational drug design via intrinsically disordered protein, *Trends Biotechnol. 24*, 435-442.
11. Uversky, V. N. (2011) Multitude of binding modes attainable by intrinsically disordered proteins: a portrait gallery of disorder-based complexes, *Chem. Soc. Rev. 40*, 1623-1634.
12. Sugase, K., Dyson, H. J., and Wright, P. E. (2007) Mechanism of coupled folding and binding of an intrinsically disordered protein, *Nature 447*, 1021-1025.
13. Demchenko, A. P. (2001) Recognition between flexible protein molecules: induced and assisted folding, *J. Mol. Recognit. 14*, 42-61.
14. Dyson, H. J., and Wright, P. E. (2002) Coupling of folding and binding for unstructured proteins, *Curr. Opin. Struct. Biol. 12*, 54-60.
15. Oldfield, C. J., Meng, J., Yang, J. Y., Yang, M. Q., Uversky, V. N., and Dunker, A. K. (2008) Flexible nets: disorder and induced fit in the associations of p53 and 14-3-3 with their partners, *BMC Genomics 9 Suppl 1*, S1.
16. Tompa, P., and Fuxreiter, M. (2008) Fuzzy complexes: polymorphism and structural disorder in protein-protein interactions, *Trends Biochem. Sci. 33*, 2-8.
17. Fisher, C. K., and Stultz, C. M. (2011) Constructing ensembles for intrinsically disordered proteins, *Curr. Opin. Struct. Biol. 21*, 426-431.
18. Eliezer, D. (2007) Characterizing residual structure in disordered protein States using nuclear magnetic resonance, *Methods Mol. Biol. 350*, 49-67.
19. Dyson, H. J., and Wright, P. E. (1998) Equilibrium NMR studies of unfolded and partially folded proteins, *Nat. Struct. Biol. 5 Suppl*, 499-503.
20. Dyson, H. J., and Wright, P. E. (2004) Unfolded proteins and protein folding studied by NMR, *Chem. Rev. 104*, 3607-3622.
21. Jensen, M. R., Markwick, P. R. L., Meier, S., Griesinger, C., Zweckstetter, M., Grzesiek, S., Bernadó, P., and Blackledge, M. (2009) Quantitative determination of the conformational properties of partially folded and intrinsically disordered proteins using NMR dipolar couplings, *Structure 17*, 1169-1185.
22. Mittag, T., and Forman-Kay, J. D. (2007) Atomic-level characterization of disordered protein ensembles, *Curr. Opin. Struct. Biol. 17*, 3-14.
23. Wishart, D. S., Sykes, B. D., and Richards, F. M. (1992) The chemical shift index: a fast and simple method for the assignment of protein secondary structure through NMR spectroscopy, *Biochemistry 31*, 1647-1651.

24. Wishart, D. S., and Sykes, B. D. (1994) Chemical shifts as a tool for structure determination, *Meth. Enzymol. 239*, 363-392.
25. Wishart, D. S., and Sykes, B. D. (1994) The 13C chemical-shift index: a simple method for the identification of protein secondary structure using 13C chemical-shift data, *J. Biomol. NMR 4*, 171-180.
26. Marsh, J. A., Singh, V. K., Jia, Z., and Forman-Kay, J. D. (2006) Sensitivity of secondary structure propensities to sequence differences between alpha- and gamma-synuclein: implications for fibrillation, *Protein Sci. 15*, 2795-2804.
27. Palmer, A. G., 3rd, Kroenke, C. D., and Loria, J. P. (2001) Nuclear magnetic resonance methods for quantifying microsecond-to-millisecond motions in biological macromolecules, *Meth. Enzymol. 339*, 204-238.
28. Hansen, D. F., Vallurupalli, P., and Kay, L. E. (2008) Using relaxation dispersion NMR spectroscopy to determine structures of excited, invisible protein states, *J. Biomol. NMR 41*, 113-120.
29. Pardi, A., Billeter, M., and Wüthrich, K. (1984) Calibration of the angular dependence of the amide proton-C alpha proton coupling constants, 3JHN alpha, in a globular protein. Use of 3JHN alpha for identification of helical secondary structure, *J. Mol. Biol. 180*, 741-751.
30. Ludvigsen, S., Andersen, K. V., and Poulsen, F. M. (1991) Accurate measurements of coupling constants from two-dimensional nuclear magnetic resonance spectra of proteins and determination of phi-angles, *J. Mol. Biol. 217*, 731-736.
31. Vuister, G. W., and Bax, A. (1993) Quantitative J correlation: a new approach for measuring homonuclear three-bond J(HNH.alpha.) coupling constants in 15N-enriched proteins, *J. Am. Chem. Soc. 115*, 7772-7777.
32. Smith, L. J., Bolin, K. A., Schwalbe, H., MacArthur, M. W., Thornton, J. M., and Dobson, C. M. (1996) Analysis of main chain torsion angles in proteins: prediction of NMR coupling constants for native and random coil conformations, *J. Mol. Biol. 255*, 494-506.
33. Mukrasch, M. D., Markwick, P. R. L., Biernat, J., Bergen, M. von, Bernadó, P., Griesinger, C., Mandelkow, E., Zweckstetter, M., and Blackledge, M. (2007) Highly populated turn conformations in natively unfolded tau protein identified from residual dipolar couplings and molecular simulation, *J. Am. Chem. Soc. 129*, 5235-5243.
34. Tjandra, N., and Bax, A. (1997) Direct measurement of distances and angles in biomolecules by NMR in a dilute liquid crystalline medium, *Science 278*, 1111-1114.
35. Hansen, M. R., Mueller, L., and Pardi, A. (1998) Tunable alignment of macromolecules by filamentous phage yields dipolar coupling interactions, *Nat. Struct. Biol. 5*, 1065-1074.
36. Clore, G. M., Starich, M. R., and Gronenborn, A. M. (1998) Measurement of Residual Dipolar Couplings of Macromolecules Aligned in the Nematic Phase of a Colloidal Suspension of Rod-Shaped Viruses, *J. Am. Chem. Soc. 120*, 10571-10572.
37. Torbet, J., and Maret, G. (1979) Fibres of highly oriented Pf1 bacteriophage produced in a strong magnetic field, *J. Mol. Biol. 134*, 843-845.
38. 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.
39. Tycko, R., Blanco, F. J., 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*, 9340-9341.
40. Rückert, M., and Otting, G. (2000) Alignment of Biological Macromolecules in Novel Nonionic Liquid Crystalline Media for NMR Experiments, *J. Am. Chem. Soc. 122*, 7793-7797.
41. 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. U.S.A 92*, 9279-9283.

42. 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.
43. Prestegard, J. H., Bougault, C. M., and Kishore, A. I. (2004) Residual dipolar couplings in structure determination of biomolecules, *Chem. Rev. 104*, 3519-3540.
44. Gebel, E. B., and Shortle, D. (2007) Characterization of denatured proteins using residual dipolar couplings, *Methods Mol. Biol. 350*, 39-48.
45. Fieber, W., Kristjansdottir, S., and Poulsen, F. M. (2004) Short-range, long-range and transition state interactions in the denatured state of ACBP from residual dipolar couplings, *J. Mol. Biol. 339*, 1191-1199.
46. Shortle, D., and Ackerman, M. S. (2001) Persistence of native-like topology in a denatured protein in 8 M urea, *Science 293*, 487-489.
47. Mohana-Borges, R., Goto, N. K., Kroon, G. J. A., Dyson, H. J., and Wright, P. E. (2004) Structural characterization of unfolded states of apomyoglobin using residual dipolar couplings, *J. Mol. Biol. 340*, 1131-1142.
48. Bernadó, P., Bertoncini, C. W., Griesinger, C., Zweckstetter, M., and Blackledge, M. (2005) Defining long-range order and local disorder in native alpha-synuclein using residual dipolar couplings, *J. Am. Chem. Soc. 127*, 17968-17969.
49. Salmon, L., Nodet, G., Ozenne, V., Yin, G., Jensen, M. R., Zweckstetter, M., and Blackledge, M. (2010) NMR characterization of long-range order in intrinsically disordered proteins, *J. Am. Chem. Soc. 132*, 8407-8418.
50. Gillespie, J. R., and Shortle, D. (1997) Characterization of long-range structure in the denatured state of staphylococcal nuclease. I. Paramagnetic relaxation enhancement by nitroxide spin labels, *J. Mol. Biol 268*, 158-169.
51. Gillespie, J. R., and Shortle, D. (1997) Characterization of long-range structure in the denatured state of staphylococcal nuclease. II. Distance restraints from paramagnetic relaxation and calculation of an ensemble of structures, *J. Mol. Biol. 268*, 170-184.
52. Solomon, I. (1955) Relaxation Processes in a System of Two Spins, *Phys. Rev. 99*, 559.
53. Kristjansdottir, S., Lindorff-Larsen, K., Fieber, W., Dobson, C. M., Vendruscolo, M., and Poulsen, F. M. (2005) Formation of native and non-native interactions in ensembles of denatured ACBP molecules from paramagnetic relaxation enhancement studies, *J. Mol. Biol. 347*, 1053-1062.
54. Bertoncini, C. W., Jung, Y.-S., Fernandez, C. O., Hoyer, W., Griesinger, C., Jovin, T. M., and Zweckstetter, M. (2005) Release of long-range tertiary interactions potentiates aggregation of natively unstructured alpha-synuclein, *Proc. Natl. Acad. Sci. U.S.A. 102*, 1430-1435.
55. Cho, M.-K., Nodet, G., Kim, H.-Y., Jensen, M. R., Bernado, P., Fernandez, C. O., Becker, S., Blackledge, M., and Zweckstetter, M. (2009) Structural characterization of alpha-synuclein in an aggregation prone state, *Protein Sci. 18*, 1840-1846.
56. Huang, J., and Grzesiek, S. (2010) Ensemble calculations of unstructured proteins constrained by RDC and PRE data: a case study of urea-denatured ubiquitin, *J. Am. Chem. Soc. 132*, 694-705.
57. Felitsky, D. J., Lietzow, M. A., Dyson, H. J., and Wright, P. E. (2008) Modeling transient collapsed states of an unfolded protein to provide insights into early folding events, *Proc. Natl. Acad. Sci. U.S.A. 105*, 6278-6283.
58. Lipari, G., and Szabo, A. (1982) Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 1. Theory and range of validity, *J. Am. Chem. Soc. 104*, 4546-4559.
59. Lipari, G., and Szabo, A. (1982) Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 2. Analysis of experimental results, *J. Am. Chem. Soc. 104*, 4559-4570.
60. Klein-Seetharaman, J., Oikawa, M., Grimshaw, S. B., Wirmer, J., Duchardt, E., Ueda, T., Imoto, T., Smith, L. J., Dobson, C. M., and Schwalbe, H. (2002) Long-Range Interactions Within a Nonnative Protein, *Science 295*, 1719 -1722.

61. Dedmon, M. M., Lindorff-Larsen, K., Christodoulou, J., Vendruscolo, M., and Dobson, C. M. (2005) Mapping long-range interactions in alpha-synuclein using spin-label NMR and ensemble molecular dynamics simulations, *J. Am. Chem. Soc. 127*, 476-477.
62. Esteban-Martín, S., Fenwick, R. B., and Salvatella, X. (2010) Refinement of ensembles describing unstructured proteins using NMR residual dipolar couplings, *J. Am. Chem. Soc. 132*, 4626-4632.
63. Bernadó, P., Blanchard, L., Timmins, P., Marion, D., Ruigrok, R. W. H., and Blackledge, M. (2005) A structural model for unfolded proteins from residual dipolar couplings and small-angle x-ray scattering, *Proc. Natl. Acad. Sci. U.S.A. 102*, 17002-17007.
64. Wells, M., Tidow, H., Rutherford, T. J., Markwick, P., Jensen, M. R., Mylonas, E., Svergun, D. I., Blackledge, M., and Fersht, A. R. (2008) Structure of tumor suppressor p53 and its intrinsically disordered N-terminal transactivation domain, *Proc. Natl. Acad. Sci. U.S.A. 105*, 5762-5767.
65. Nodet, G., Salmon, L., Ozenne, V., Meier, S., Jensen, M. R., and Blackledge, M. (2009) Quantitative description of backbone conformational sampling of unfolded proteins at amino acid resolution from NMR residual dipolar couplings, *J. Am. Chem. Soc. 131*, 17908-17918.
66. Choy, W. Y., and Forman-Kay, J. D. (2001) Calculation of ensembles of structures representing the unfolded state of an SH3 domain, *J. Mol. Biol. 308*, 1011-1032.
67. Marsh, J. A., and Forman-Kay, J. D. (2009) Structure and disorder in an unfolded state under nondenaturing conditions from ensemble models consistent with a large number of experimental restraints, *J. Mol. Biol. 391*, 359-374.
68. Mittag, T., Orlicky, S., Choy, W.-Y., Tang, X., Lin, H., Sicheri, F., Kay, L. E., Tyers, M., and Forman-Kay, J. D. (2008) Dynamic equilibrium engagement of a polyvalent ligand with a single-site receptor, *Proc. Natl. Acad. Sci. U.S.A. 105*, 17772-17777.
69. Mittag, T., Marsh, J., Grishaev, A., Orlicky, S., Lin, H., Sicheri, F., Tyers, M., and Forman-Kay, J. D. (2010) Structure/function implications in a dynamic complex of the intrinsically disordered Sic1 with the Cdc4 subunit of an SCF ubiquitin ligase, *Structure 18*, 494-506.
70. Marsh, J. A., Dancheck, B., Ragusa, M. J., Allaire, M., Forman-Kay, J. D., and Peti, W. (2010) Structural diversity in free and bound states of intrinsically disordered protein phosphatase 1 regulators, *Structure 18*, 1094-1103.
71. Meier, S., Grzesiek, S., and Blackledge, M. (2007) Mapping the conformational landscape of urea-denatured ubiquitin using residual dipolar couplings, *J. Am. Chem. Soc. 129*, 9799-9807.
72. Zweckstetter, M., and Bax, A. (2000) Prediction of sterically induced alignment in a dilute liquid crystalline phase: Aid to protein structure determination by NMR, *J. Am. Chem. Soc. 122*, 3791-3792.
73. Zweckstetter, M. (2008) NMR: prediction of molecular alignment from structure using the PALES software, *Nat. Protoc. 3*, 679-690.
74. Marsh, J. A., Baker, J. M. R., Tollinger, M., and Forman-Kay, J. D. (2008) Calculation of residual dipolar couplings from disordered state ensembles using local alignment, *J. Am. Chem. Soc. 130*, 7804-7805.
75. Chen, X., Sagle, L. B., and Cremer, P. S. (2007) Urea orientation at protein surfaces, *J. Am. Chem. Soc. 129*, 15104-15105.
76. Gabel, F., Jensen, M. R., Zaccaï, G., and Blackledge, M. (2009) Quantitative modelfree analysis of urea binding to unfolded ubiquitin using a combination of small angle X-ray and neutron scattering, *J. Am. Chem. Soc. 131*, 8769-8771.
77. Liepinsh, E., and Otting, G. (1994) Specificity of Urea Binding to Proteins, *J. Am. Chem. Soc. 116*, 9670-9674.
78. Modig, K., Kurian, E., Prendergast, F. G., and Halle, B. (2003) Water and urea interactions with the native and unfolded forms of a beta-barrel protein, *Prot. Sci. 12*, 2768-2781.
79. Pace, C. N. (1986) Determination and analysis of urea and guanidine hydrochloride denaturation curves, *Meth. Enzymol. 131*, 266-280.

80. Makhatadze, G. I., and Privalov, P. L. (1992) Protein interactions with urea and guanidinium chloride. A calorimetric study, *J. Mol. Biol. 226*, 491-505.
81. Schellman, J. A., and Gassner, N. C. (1996) The enthalpy of transfer of unfolded proteins into solutions of urea and guanidinium chloride, *Biophys. Chem. 59*, 259-275.
82. Möglich, A., Krieger, F., and Kiefhaber, T. (2005) Molecular basis for the effect of urea and guanidinium chloride on the dynamics of unfolded polypeptide chains, *J. Mol. Biol. 345*, 153-162.
83. Sezer, D., Freed, J. H., and Roux, B. (2008) Parametrization, molecular dynamics simulation, and calculation of electron spin resonance spectra of a nitroxide spin label on a polyalanine alpha-helix, *J. Phys. Chem. B 112*, 5755-5767.
84. Bruschweiler, R., Roux, B., Blackledge, M., Griesinger, C., Karplus, M., and Ernst, R. R. (1992) Influence of rapid intramolecular motion on NMR cross-relaxation rates. A molecular dynamics study of antamanide in solution, *J. Am. Chem. Soc. 114*, 2289-2302.
85. Iwahara, J., Schwieters, C. D., and Clore, G. M. (2004) Ensemble Approach for NMR Structure Refinement against 1H Paramagnetic Relaxation Enhancement Data Arising from a Flexible Paramagnetic Group Attached to a Macromolecule, *Journal of the American Chemical Society 126*, 5879-5896.
86. Sezer, D., Freed, J. H., and Roux, B. (2008) Simulating electron spin resonance spectra of nitroxide spin labels from molecular dynamics and stochastic trajectories, *J. Chem. Phys. 128*, 165106.
87. Iwahara, J., Tang, C., and Marius Clore, G. (2007) Practical aspects of (1)H transverse paramagnetic relaxation enhancement measurements on macromolecules, *J. Magn. Reson. 184*, 185-195.
88. Clore, G. M., and Iwahara, J. (2009) Theory, practice, and applications of paramagnetic relaxation enhancement for the characterization of transient low-population states of biological macromolecules and their complexes, *Chem. Rev. 109*, 4108-4139.
89. Bartels, T., Choi, J. G., and Selkoe, D. J. (2011) α-Synuclein occurs physiologically as a helically folded tetramer that resists aggregation, *Nature 477*, 107-110.
90. Spillantini, M. G., Schmidt, M. L., Lee, V. M., Trojanowski, J. Q., Jakes, R., and Goedert, M. (1997) Alpha-synuclein in Lewy bodies, *Nature 388*, 839-840.
91. Neal, S., Nip, A. M., Zhang, H., and Wishart, D. S. (2003) Rapid and accurate calculation of protein 1H, 13C and 15N chemical shifts, *J. Biomol. NMR 26*, 215-240.
92. Han, B., Liu, Y., Ginzinger, S. W., and Wishart, D. S. (2011) SHIFTX2: significantly improved protein chemical shift prediction, *J. Biomol. NMR 50*, 43-57.
93. Kohlhoff, K. J., Robustelli, P., Cavalli, A., Salvatella, X., and Vendruscolo, M. (2009) Fast and accurate predictions of protein NMR chemical shifts from interatomic distances, *J. Am. Chem. Soc. 131*, 13894-13895.
94. Shen, Y., and Bax, A. (2007) Protein backbone chemical shifts predicted from searching a database for torsion angle and sequence homology, *J. Biomol. NMR 38*, 289-302.
95. Shen, Y., and Bax, A. (2010) SPARTA+: a modest improvement in empirical NMR chemical shift prediction by means of an artificial neural network, *J. Biomol. NMR*.
96. Shen, Y., Lange, O., Delaglio, F., Rossi, P., Aramini, J. M., Liu, G., Eletsky, A., Wu, Y., Singarapu, K. K., Lemak, A., Ignatchenko, A., Arrowsmith, C. H., Szyperski, T., Montelione, G. T., Baker, D., and Bax, A. (2008) Consistent blind protein structure generation from NMR chemical shift data, *Proc. Natl. Acad. Sci. U.S.A. 105*, 4685-4690.
97. Cavalli, A., Salvatella, X., Dobson, C. M., and Vendruscolo, M. (2007) Protein structure determination from NMR chemical shifts, *Proc. Natl. Acad. Sci. U.S.A. 104*, 9615-9620.
98. Maiti, N. C., Apetri, M. M., Zagorski, M. G., Carey, P. R., and Anderson, V. E. (2004) Raman spectroscopic characterization of secondary structure in natively unfolded proteins: alpha-synuclein, *J. Am. Chem. Soc. 126*, 2399-2408.
99. Shi, Z., Chen, K., Liu, Z., and Kallenbach, N. R. (2006) Conformation of the backbone in unfolded proteins, *Chem. Rev. 106*, 1877-1897.

100. Woody, R. W. (2009) Circular dichroism spectrum of peptides in the poly(Pro)II conformation, *J. Am. Chem. Soc. 131*, 8234-8245.

101. Kjaergaard, M., Brander, S., and Poulsen, F. M. (2011) Random coil chemical shift for intrinsically disordered proteins: effects of temperature and pH, *J. Biomol. NMR 49*, 139-149.

102. Eyal, E., Najmanovich, R., McConkey, B. J., Edelman, M., and Sobolev, V. (2004) Importance of solvent accessibility and contact surfaces in modeling side-chain conformations in proteins, *J. Comput. Chem. 25*, 712-724.

103. Selenko, P., Frueh, D. P., Elsaesser, S. J., Haas, W., Gygi, S. P., and Wagner, G. (2008) In situ observation of protein phosphorylation by high-resolution NMR spectroscopy, *Nat. Struct. Mol. Biol. 15*, 321-329.

104. Serber, Z., and Dötsch, V. (2001) In-cell NMR spectroscopy, *Biochemistry 40*, 14317-14323.

105. Curran, J., Homann, H., Buchholz, C., Rochat, S., Neubert, W., and Kolakofsky, D. (1993) The hypervariable C-terminal tail of the Sendai paramyxovirus nucleocapsid protein is required for template function but not for RNA encapsidation, *J. Virol. 67*, 4358-4364.

106. Houben, K., Marion, D., Tarbouriech, N., Ruigrok, R. W. H., and Blanchard, L. (2007) Interaction of the C-terminal domains of sendai virus N and P proteins: comparison of polymerase-nucleocapsid interactions within the paramyxovirus family, *J. Virol. 81*, 6807-6816.

107. Dosset, P., Hus, J. C., Marion, D., and Blackledge, M. (2001) A novel interactive tool for rigid-body modeling of multi-domain macromolecules using residual dipolar couplings, *J. Biomol. NMR 20*, 223-231.

108. Mesleh, M. F., Lee, S., Veglia, G., Thiriot, D. S., Marassi, F. M., and Opella, S. J. (2003) Dipolar waves map the structure and topology of helices in membrane proteins, *J. Am. Chem. Soc. 125*, 8928-8935.

109. Mesleh, M. F., and Opella, S. J. (2003) Dipolar Waves as NMR maps of helices in proteins, *J. Magn. Reson. 163*, 288-299.

110. Mesleh, M. F., Veglia, G., DeSilva, T. M., Marassi, F. M., and Opella, S. J. (2002) Dipolar waves as NMR maps of protein structure, *J. Am. Chem. Soc. 124*, 4206-4207.

111. Marassi, F. M., and Opella, S. J. (2000) A solid-state NMR index of helical membrane protein structure and topology, *J. Magn. Reson. 144*, 150-155.

112. Mascioni, A., and Veglia, G. (2003) Theoretical analysis of residual dipolar coupling patterns in regular secondary structures of proteins, *J. Am. Chem. Soc. 125*, 12520-12526.

113. Jensen, M. R., Houben, K., Lescop, E., Blanchard, L., Ruigrok, R. W. H., and Blackledge, M. (2008) Quantitative conformational analysis of partially folded proteins from residual dipolar couplings: application to the molecular recognition element of Sendai virus nucleoprotein, *J. Am. Chem. Soc. 130*, 8055-8061.

114. Richardson, J. S., and Richardson, D. C. (1988) Amino acid preferences for specific locations at the ends of alpha helices, *Science 240*, 1648-1652.

115. Serrano, L., and Fersht, A. R. (1989) Capping and alpha-helix stability, *Nature 342*, 296-299.

116. Serrano, L., Sancho, J., Hirshberg, M., and Fersht, A. R. (1992) Alpha-helix stability in proteins. I. Empirical correlations concerning substitution of side-chains at the N and C-caps and the replacement of alanine by glycine or serine at solvent-exposed surfaces, *J. Mol. Biol. 227*, 544-559.

117. Doig, A. J., MacArthur, M. W., Stapley, B. J., and Thornton, J. M. (1997) Structures of N-termini of helices in proteins, *Protein Sci. 6*, 147-155.

118. Aurora, R., and Rose, G. D. (1998) Helix capping, *Protein Sci. 7*, 21-38.

119. Shoemaker, B. A., Portman, J. J., and Wolynes, P. G. (2000) Speeding molecular recognition by using the folding funnel: the fly-casting mechanism, *Proc. Natl. Acad. Sci. U.S.A. 97*, 8868-8873.

120. Curran, J., and Kolakofsky, D. (1999) Replication of paramyxoviruses, *Adv. Virus Res. 54*, 403-422.

121. Kingston, R. L., Baase, W. A., and Gay, L. S. (2004) Characterization of nucleocapsid binding by the measles virus and mumps virus phosphoproteins, *J. Virol. 78*, 8630-8640.

122. Curran, J., Marq, J. B., and Kolakofsky, D. (1995) An N-terminal domain of the Sendai paramyxovirus P protein acts as a chaperone for the NP protein during the nascent chain assembly step of genome replication, *J. Virol. 69*, 849-855.

123. Mavrakis, M., Méhouas, S., Réal, E., Iseni, F., Blondel, D., Tordo, N., and Ruigrok, R. W. H. (2006) Rabies virus chaperone: identification of the phosphoprotein peptide that keeps nucleoprotein soluble and free from non-specific RNA, *Virology 349*, 422-429.

124. Chen, M., Ogino, T., and Banerjee, A. K. (2007) Interaction of vesicular stomatitis virus P and N proteins: identification of two overlapping domains at the N terminus of P that are involved in N0-P complex formation and encapsidation of viral genome RNA, *J. Virol. 81*, 13478-13485.

125. Karlin, D., Ferron, F., Canard, B., and Longhi, S. (2003) Structural disorder and modular organization in Paramyxovirinae N and P, *J. Gen. Virol. 84*, 3239-3252.

126. Johansson, K., Bourhis, J.-M., Campanacci, V., Cambillau, C., Canard, B., and Longhi, S. (2003) Crystal structure of the measles virus phosphoprotein domain responsible for the induced folding of the C-terminal domain of the nucleoprotein, *J. Biol. Chem. 278*, 44567-44573.

127. Tarbouriech, N., Curran, J., Ruigrok, R. W., and Burmeister, W. P. (2000) Tetrameric coiled coil domain of Sendai virus phosphoprotein, *Nat. Struct. Biol. 7*, 777-781.

128. Blanchard, L., Tarbouriech, N., Blackledge, M., Timmins, P., Burmeister, W. P., Ruigrok, R. W. H., and Marion, D. (2004) Structure and dynamics of the nucleocapsid-binding domain of the Sendai virus phosphoprotein in solution, *Virology 319*, 201-211.

129. Bourhis, J.-M., Johansson, K., Receveur-Bréchot, V., Oldfield, C. J., Dunker, K. A., Canard, B., and Longhi, S. (2004) The C-terminal domain of measles virus nucleoprotein belongs to the class of intrinsically disordered proteins that fold upon binding to their physiological partner, *Virus Res. 99*, 157-167.

130. Longhi, S., Receveur-Bréchot, V., Karlin, D., Johansson, K., Darbon, H., Bhella, D., Yeo, R., Finet, S., and Canard, B. (2003) The C-terminal domain of the measles virus nucleoprotein is intrinsically disordered and folds upon binding to the C-terminal moiety of the phosphoprotein, *J. Biol. Chem. 278*, 18638-18648.

131. Kingston, R. L., Hamel, D. J., Gay, L. S., Dahlquist, F. W., and Matthews, B. W. (2004) Structural basis for the attachment of a paramyxoviral polymerase to its template, *Proc. Natl. Acad. Sci. U.S.A. 101*, 8301-8306.

132. Hammes, G. G., Chang, Y.-C., and Oas, T. G. (2009) Conformational selection or induced fit: a flux description of reaction mechanism, *Proc. Natl. Acad. Sci. U.S.A. 106*, 13737-13741.

133. Csermely, P., Palotai, R., and Nussinov, R. (2010) Induced fit, conformational selection and independent dynamic segments: an extended view of binding events, *Trends Biochem. Sci. 35*, 539-546.

134. Gely, S., Lowry, D. F., Bernard, C., Jensen, M. R., Blackledge, M., Costanzo, S., Bourhis, J.-M., Darbon, H., Daughdrill, G., and Longhi, S. (2010) Solution structure of the C-terminal X domain of the measles virus phosphoprotein and interaction with the intrinsically disordered C-terminal domain of the nucleoprotein, *J. Mol. Recognit.*

135. Fooks, A. R., Stephenson, J. R., Warnes, A., Dowsett, A. B., Rima, B. K., and Wilkinson, G. W. (1993) Measles virus nucleocapsid protein expressed in insect cells assembles into nucleocapsid-like structures, *J. Gen. Virol. 74 (Pt 7)*, 1439-1444.

136. Lund, G. A., Tyrrell, D. L., Bradley, R. D., and Scraba, D. G. (1984) The molecular length of measles virus RNA and the structural organization of measles nucleocapsids, *J. Gen. Virol. 65 (Pt 9)*, 1535-1542.

137. Finch, J. T., and Gibbs, A. J. (1970) Observations on the structure of the nucleocapsids of some paramyxoviruses, *J. Gen. Virol. 6*, 141-150.

138. Bhella, D., Ralph, A., Murphy, L. B., and Yeo, R. P. (2002) Significant differences in nucleocapsid morphology within the Paramyxoviridae, *J. Gen. Virol. 83*, 1831-1839.

139. Schoehn, G., Mavrakis, M., Albertini, A., Wade, R., Hoenger, A., and Ruigrok, R. W. H. (2004) The 12 A structure of trypsin-treated measles virus N-RNA, *J. Mol. Biol. 339*, 301-312.
140. Bhella, D., Ralph, A., and Yeo, R. P. (2004) Conformational flexibility in recombinant measles virus nucleocapsids visualised by cryo-negative stain electron microscopy and real-space helical reconstruction, *J. Mol. Biol. 340*, 319-331.
141. Desfosses, A., Goret, G., Farias Estrozi, L., Ruigrok, R. W. H., and Gutsche, I. (2011) Nucleoprotein-RNA orientation in the measles virus nucleocapsid by three-dimensional electron microscopy, *J. Virol. 85*, 1391-1395.
142. Tawar, R. G., Duquerroy, S., Vonrhein, C., Varela, P. F., Damier-Piolle, L., Castagne, N., MacLellan, K., Bedouelle, H., Bricogne, G., Bhella, D., Eleouet, J.-F., and Rey, F. A. (2009) Crystal Structure of a Nucleocapsid-Like Nucleoprotein-RNA Complex of Respiratory Syncytial Virus, *Science 326*, 1279-1283.
143. Porod, G. (1982) In small angle x-ray scattering / edited by O. Glatter and O. Kratky. Academic Press, New York.
144. Skalicky, J. J., Mills, J. L., Sharma, S., and Szyperski, T. (2001) Aromatic ring-flipping in supercooled water: implications for NMR-based structural biology of proteins, *J. Am. Chem. Soc. 123*, 388-397.
145. Mills, J. L., and Szyperski, T. (2002) Protein dynamics in supercooled water: the search for slow motional modes, *J. Biomol. NMR 23*, 63-67.
146. Szyperski, T., and Mills, J. L. (2011) NMR-based structural biology of proteins in supercooled water, *J. Struct. Funct. Genomics 12*, 1-7.
147. Otting, G. (2010) Protein NMR using paramagnetic ions, *Annu. Rev. Biophys. 39*, 387-405.
148. Kurland, R. J., and McGarvey, B. R. (1970) Isotropic NMR shifts in transition metal complexes: The calculation of the fermi contact and pseudocontact terms, *J. Magn. Reson. 2*, 286-301.
149. Allegrozzi, M., Bertini, I., Janik, M. B. L., Lee, Y.-M., Liu, G., and Luchinat, C. (2000) Lanthanide-Induced Pseudocontact Shifts for Solution Structure Refinements of Macromolecules in Shells up to 40 Å from the Metal Ion, *J. Am. Chem. Soc. 122*, 4154-4161.
150. Rodriguez-Castañeda, F., Haberz, P., Leonov, A., and Griesinger, C. (2006) Paramagnetic tagging of diamagnetic proteins for solution NMR, *Magn. Reson. Chem. 44 Spec No*, S10-16.
151. Wöhnert, J., Franz, K. J., Nitz, M., Imperiali, B., and Schwalbe, H. (2003) Protein alignment by a coexpressed lanthanide-binding tag for the measurement of residual dipolar couplings, *J. Am. Chem. Soc. 125*, 13338-13339.
152. Saio, T., Ogura, K., Yokochi, M., Kobashigawa, Y., and Inagaki, F. (2009) Two-point anchoring of a lanthanide-binding peptide to a target protein enhances the paramagnetic anisotropic effect, *J. Biomol. NMR 44*, 157-166.
153. Prudêncio, M., Rohovec, J., Peters, J. A., Tocheva, E., Boulanger, M. J., Murphy, M. E. P., Hupkes, H.-J., Kosters, W., Impagliazzo, A., and Ubbink, M. (2004) A caged lanthanide complex as a paramagnetic shift agent for protein NMR, *Chemistry 10*, 3252-3260.
154. Keizers, P. H. J., and Ubbink, M. (2011) Paramagnetic tagging for protein structure and dynamics analysis, *Prog. Nucl. Magn. Reson. Spectrosc. 58*, 88-96.
155. Häussinger, D., Huang, J., and Grzesiek, S. (2009) DOTA-M8: An extremely rigid, high-affinity lanthanide chelating tag for PCS NMR spectroscopy, *J. Am. Chem. Soc. 131*, 14761-14767.
156. Ikegami, T., Verdier, L., Sakhaii, P., Grimme, S., Pescatore, B., Saxena, K., Fiebig, K. M., and Griesinger, C. (2004) Novel techniques for weak alignment of proteins in solution using chemical tags coordinating lanthanide ions, *J. Biomol. NMR 29*, 339-349.
157. Graham, B., Loh, C. T., Swarbrick, J. D., Ung, P., Shin, J., Yagi, H., Jia, X., Chhabra, S., Barlow, N., Pintacuda, G., Huber, T., and Otting, G. (2011) A DOTA-amide lanthanide tag for reliable generation of pseudocontact shifts in protein NMR spectra, *Bioconjug. Chem, In press.*
158. Peters, F., Maestre-Martinez, M., Leonov, A., Kovačič, L., Becker, S., Boelens, R., and Griesinger, C. (2011) Cys-Ph-TAHA: a lanthanide binding tag for RDC and PCS enhanced protein NMR, *J. Biomol. NMR, In press.*

159. Chang, L., and Karin, M. (2001) Mammalian MAP kinase signalling cascades, *Nature 410*, 37-40.

160. Kennedy, N. J., Sluss, H. K., Jones, S. N., Bar-Sagi, D., Flavell, R. A., and Davis, R. J. (2003) Suppression of Ras-stimulated transformation by the JNK signal transduction pathway, *Genes Dev. 17*, 629-637.

161. Hirosumi, J., Tuncman, G., Chang, L., Görgün, C. Z., Uysal, K. T., Maeda, K., Karin, M., and Hotamisligil, G. S. (2002) A central role for JNK in obesity and insulin resistance, *Nature 420*, 333-336.

162. Yoshida, H., Hastie, C. J., McLauchlan, H., Cohen, P., and Goedert, M. (2004) Phosphorylation of microtubule-associated protein tau by isoforms of c-Jun N-terminal kinase (JNK), *J. Neurochem. 90*, 352-358.

163. Xia, X. G., Harding, T., Weller, M., Bieneman, A., Uney, J. B., and Schulz, J. B. (2001) Gene transfer of the JNK interacting protein-1 protects dopaminergic neurons in the MPTP model of Parkinson's disease, *Proc. Natl. Acad. Sci. U.S.A. 98*, 10433-10438.

164. Bardwell, L. (2006) Mechanisms of MAPK signalling specificity, *Biochem. Soc. Trans. 34*, 837-841.

165. Bardwell, A. J., Frankson, E., and Bardwell, L. (2009) Selectivity of docking sites in MAPK kinases, *J. Biol. Chem. 284*, 13165-13173.

166. Weston, C. R., Lambright, D. G., and Davis, R. J. (2002) Signal transduction. MAP kinase signaling specificity, *Science 296*, 2345-2347.

167. Goldsmith, E. J., Akella, R., Min, X., Zhou, T., and Humphreys, J. M. (2007) Substrate and docking interactions in serine/threonine protein kinases, *Chem. Rev. 107*, 5065-5081.

168. Dosztányi, Z., Csizmok, V., Tompa, P., and Simon, I. (2005) IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content, *Bioinformatics 21*, 3433-3434.

169. Sharrocks, A. D., Yang, S. H., and Galanis, A. (2000) Docking domains and substrate-specificity determination for MAP kinases, *Trends Biochem. Sci. 25*, 448-453.

170. Smith, J. A., Poteet-Smith, C. E., Lannigan, D. A., Freed, T. A., Zoltoski, A. J., and Sturgill, T. W. (2000) Creation of a stress-activated p90 ribosomal S6 kinase. The carboxyl-terminal tail of the MAPK-activated protein kinases dictates the signal transduction pathway in which they function, *J. Biol. Chem. 275*, 31588-31593.

171. Tanoue, T., Maeda, R., Adachi, M., and Nishida, E. (2001) Identification of a docking groove on ERK and p38 MAP kinases that regulates the specificity of docking interactions, *EMBO J. 20*, 466-479.

172. Ho, D. T., Bardwell, A. J., Grewal, S., Iverson, C., and Bardwell, L. (2006) Interacting JNK-docking sites in MKK7 promote binding and activation of JNK mitogen-activated protein kinases, *J. Biol. Chem. 281*, 13169-13179.

173. Lawler, S., Fleming, Y., Goedert, M., and Cohen, P. (1998) Synergistic activation of SAPK1/JNK1 by two MAP kinase kinases in vitro, *Curr. Biol. 8*, 1387-1390.

174. Fleming, Y., Armstrong, C. G., Morrice, N., Paterson, A., Goedert, M., and Cohen, P. (2000) Synergistic activation of stress-activated protein kinase 1/c-Jun N-terminal kinase (SAPK1/JNK) isoforms by mitogen-activated protein kinase kinase 4 (MKK4) and MKK7, *Biochem. J. 352 Pt 1*, 145-154.

175. Weston, C. R., and Davis, R. J. (2002) The JNK signal transduction pathway, *Curr. Opin. Genet. Dev. 12*, 14-21.

176. Brancho, D., Tanaka, N., Jaeschke, A., Ventura, J.-J., Kelkar, N., Tanaka, Y., Kyuuma, M., Takeshita, T., Flavell, R. A., and Davis, R. J. (2003) Mechanism of p38 MAP kinase activation in vivo, *Genes Dev. 17*, 1969 -1978.

177. Heo, J., Kim, S., Seo, C., Kim, Y., Sung, B., Lee, H., Lee, J., Park, S., Kim, J., Hwang, K., Hyun, Y., Jeon, Y., Ro, S., Cho, J., Lee, T., and Yang, C. (2004) Structural basis for the selective inhibition of JNK1 by the scaffolding protein JIP1 and SP600125, *EMBO J. 23*, 2185-2195.

178. Dickens, M., Rogers, J., Cavanagh, J., Raitano, A., Xia, Z., Halpern, J., Greenberg, M., Sawyers, C., and Davis, R. (1997) A cytoplasmic inhibitor of the JNK signal transduction pathway, *Science 277*, 693-696.

179. Bonny, C., Oberson, A., Negri, S., Sauser, C., and Schorderet, D. (2001) Cell-permeable peptide inhibitors of JNK novel blockers of beta-cell death, *Diabetes 50*, 77-82.

180. Whitmarsh, A., Cavanagh, L., Tournier, C., Yasuda, L., and Davis, R. (1998) Mammalian scaffold complex that selectively mediates MAP kinase activation, *Science 281*, 1671-1674.

181. Kristensen, O., Guenat, S., Dar, I., Allaman-Pillet, N., Abderrahmani, A., Ferdaoussi, M., Roduit, R., Maurer, F., Beckmann, J. S., Kastrup, J. S., Gajhede, M., and Bonny, C. (2006) A unique set of SH3-SH3 interactions controls IB1 homodimerization, *EMBO J. 25*, 785-797.

182. Yasuda, J., Whitmarsh, A. J., Cavanagh, J., Sharma, M., and Davis, R. J. (1999) The JIP group of mitogen-activated protein kinase scaffold proteins, *Mol. Cell. Biol. 19*, 7245-7254.

11. Curriculum Vitae

Personal information

Last name:	Jensen
First name(s):	Malene Ringkjøbing
Address:	Protein Dynamics and Flexibility (FDP) Institut de Biologie Structurale Jean-Pierre Ebel 41, rue Jules Horowitz 38027 Grenoble France
Telephone:	+33 438 784 780
E-mail:	malene.ringkjobing-jensen@ibs.fr
Date of birth:	5th of May 1978
Gender/Nationality:	Female/Danish

Education

05/2006	Ph.D. in Chemistry, University of Copenhagen, Denmark (supervision by Dr. Jens J. Led).
01/2003	M.Sc. in Chemistry, University of Copenhagen, Denmark.
08/2000	B.Sc. in Chemistry and Mathematics, University of Copenhagen, Denmark.

Professional experience

01/2010 – present	Associate Scientist (Chargée de Recherche, grade CR2), Centre National de la Recherche Scientifique (CNRS), Institut de Biologie Structurale, Grenoble, France.
01/2007 – 12/2009	Postdoctoral researcher in the group of Dr. Martin Blackledge, Institut de Biologie Structurale, Grenoble, France.
05/2006 – 12/2006	Postdoctoral researcher in the group of Dr. Jens J. Led, Department of Chemistry, University of Copenhagen, Denmark.

Awards and achievements

2009	National French competition for CNRS researchers: Classed 1st out of 81 auditioned candidates for CR2 positions in commission 16.
2008	Kirstine Meyer Prize (Denmark):

	Awarded to young promising scientists within physics or chemistry.
2006	Talent prize of the Lundbeck foundation (Denmark): Awarded to scientists below 30 years of age who have conducted particularly promising research in biomedicine or the natural sciences.
2005	Danish Young NMR Researcher Prize: Awarded to young researcher who has conducted particularly promising research within NMR spectroscopy.

Grants and stipends

03/2011– 02/2014	Grant from the French Agence National de la Recherche (ANR) (260.000 Euros) for the project: "Structural basis of molecular recognition in intrinsically disordered proteins"
01/2009 – 12/2009	Grant from the Lundbeck foundation, Denmark (59.900 Euros) for continued postdoctoral studies in the group of Dr. Martin Blackledge, Grenoble, France.
08/2007 – 07/2009	EMBO long-term fellowship for postdoctoral studies in the group of Dr. Martin Blackledge, Grenoble. France.
01/2007 – 12/2008	Grant from the Lundbeck foundation, Denmark (98.800 Euros) for postdoctoral studies in the group of Dr. Martin Blackledge, Grenoble. France.
05/2003 – 05/2006	Ph.D. stipend, University of Copenhagen, Denmark.

Co-supervision of Master and Ph.D. students (with Dr. Martin Blackledge)

09/2011 – 08/2014	Jaka Kragelj (Ph.D. student)
01/2011 – 07/2011	Mingxi Yao (Visiting Ph.D. student from National University of Singapore) *1 co-publication*
09/2010 – 08/2013	Guillaume Communie (Ph.D. student) *2 co-publications*
09/2009 – 08/2012	Valéry Ozenne (Ph.D. student) *3 co-publications*
09/2007 – 08/2010	Dr. Loïc Salmon (Ph.D. student) *7 co-publications, 2 as co-corresponding author*
02/2010 – 07/2010	Guillaume Communie (Master student)
02/2009 – 07/2009	Valéry Ozenne (Master student)
09/2007 – 12/2007	Dr. Jose Luis Ortega Roldan (Visiting Ph.D. student from Granada, Spain) *3 co-publications, 2 as co-corresponding author*
02/2007 – 07/2007	Nikolas Melcher (Master student)

Poster and oral presentations

09/2011	11th CCPN workshop “Experimental NMR”, Scarborough, England (invited speaker)
08/2011	European Magnetic Resonance Meeting (EUROMAR), Frankfurt, Germany (poster)
05/2011	Gordon Research Conference: Computational Aspects of Biomolecular NMR, Lucca, Italy (poster)
01/2011	Bio-NMR Annual User Meeting, Brno, Czech Republic (poster)
10/2010	Institute for Research in Biomedicine, Barcelona, Spain (invited speaker, host: Dr. Xavier Salvatella)
07/2010	European Magnetic Resonance Meeting (EUROMAR), Florence, Italy (poster)
05/2010	Workshop: Understanding Transient Molecular Interactions in Biology, Seville, Spain (poster)
03/2010	AFMB, Université d’Aix-Marseille, Marseille, France (invited speaker, host: Dr. Sonia Longhi)
01/2010	NMR: A Tool for Biology IX, Institut de Pasteur, Paris, France (poster)
07/2009	European Magnetic Resonance Meeting (EUROMAR), Göteborg, Sverige (poster)
06/2009	High-Field Solution and Solid-State Biomolecular NMR Workshop, Les Houches, France (invited speaker)
05/2009	Conference “Protein Folds in Infectious and Neurodegenerative Diseases”, Aussois, France (invited speaker)
03/2009	Experimental Nuclear Magnetic Resonance Conference (ENC), Asilomar, California, U.S.A. (poster)
01/2009	EU-NMR Annual User Meeting, Autrans, France (poster)
10/2008	Selskabet for Naturlærens Udbredelse, Copenhagen, Denmark (invited speaker, prize ceremony)
05/2008	Gordon Research Conference: Computational Aspects of Biomolecular NMR, Lucca, Italy (poster)
05/2008	Department of Chemistry, University of Copenhagen, Denmark (invited speaker, host: Dr. Leila Lo Leggio)
03/2008	Experimental Nuclear Magnetic Resonance Conference (ENC), Asilomar, California, U.S.A. (poster)
02/2006	Australian, New Zealand Magnetic Resonance Conference (ANZMAG), Murramarang, Australia (poster abstract upgraded to short oral presentation)
05/2005	Symposium of the Danish Instrument Center for NMR Spectroscopy of Biological Macromolecules, Carlsberg Research Center, Valby, Denmark (invited speaker, prize ceremony)
04/2005	Experimental Nuclear Magnetic Resonance Conference (ENC), Providence, Rhode Island, U.S.A. (poster)
09/2004	European Experimental Nuclear Magnetic Resonance Conference (EENC), Lille, France (poster)

06/2002 European Experimental Nuclear Magnetic Resonance Conference (EENC), Prague, Czech Republic (poster)

Publication list

* Corresponding author

(32) T. Tosi, N. N. Nickerson, L. Mollica, **M.R. Jensen**, M. Blackledge, B. Baron, P. England, A. P. Pugsley and A. Dessen*. Mol. Microbiol. (2011). In press.
"Pilotin-secretin recognition in the type II secretion system of Klebsiella oxytoca"

(31) R. Schneider, J.-R. Huang, M. Yao, G. Communie, V. Ozenne, L. Mollica, L. Salmon, **M.R. Jensen**, M. Blackledge*. Mol. Biosyst. (2011). In press.
"Towards a robust description of intrinsic protein disorder using nuclear magnetic resonance spectroscopy"

(30) **M.R. Jensen**, J.L. Ortega-Roldan, L. Salmon, N. van Nuland, M. Blackledge*. Eur. Biophys. J. (2011). In press.
"Characterizing weak protein-protein complexes by NMR residual dipolar couplings"

(29) F. Dupeux, J. Santiago, K. Betz, J. Twycross, S.-Y. Park, L. Rodriguez, M. Gonzales-Guzman, **M.R. Jensen**, N. Krasnogor, M. Blackledge, M. Holdsworth, S.R. Cutler, P.L. Rodriguez, J.A. Márquez*. EMBO J. (2011), 30, 4171-4184.
"A thermodynamic switch modulates abscisic acid receptor sensitivity"

(28) C. Leyrat, F. Yabukarski, N. Tarbouriech, E.A. Ribeiro Jr, **M.R. Jensen**, M. Blackledge, R.W.H. Ruigrok, M. Jamin*. PLoS Pathog. (2011), 7, e1002248.
"Structure of the vesicular stomatitis virus N^0*-P complex"*

(27) **M.R. Jensen**, G. Communie, E.A. Ribeiro Jr, N. Martinez, A. Desfosses, L. Salmon, L. Mollica, F. Gabel, M. Jamin, S. Longhi, R.W.H. Ruigrok, M. Blackledge*. Proc. Natl. Acad. Sci. U.S.A (2011), 108, 9839-9844.
"Intrinsic disorder in measles virus nucleocapsids"

(26) L. Salmon, J.L. Ortega-Roldan, E. Lescop, A. Licinio, N. van Nuland, **M.R. Jensen***, M. Blackledge*. Angew. Chem. (2011), 50, 3755-3759.
"Structure, dynamics, and kinetics of weak protein-protein complexes from NMR spin relaxation measurements of titrated solutions"

(25) J. Habchi, S. Blangy, L. Mamelli, **M.R. Jensen**, M. Blackledge, H. Darbon, M. Oglesbee, Y. Shu, S. Longhi*. J. Biol. Chem. (2011), 286, 13583-13602.
"Characterization of the interactions between the nucleoprotein and the phosphoprotein of Henipavirus"

(24) C. Leyrat, **M.R. Jensen**, E.A. Ribeiro Jr, F.C. Gerard, R.W.H. Ruigrok, M. Blackledge, M. Jamin*. Protein Sci. (2011), 20, 542-556.
"The N(0)-binding region of the vesicular stomatitis virus phosphoprotein is globally disordered but contains transient α*-helices"*

(23) L. Salmon, G. Nodet, V. Ozenne, **M.R. Jensen**, M. Zweckstetter, M. Blackledge*. J. Am. Chem. Soc. (2010), 132, 8407-8418.
"NMR characterization of long-range order in intrinsically disordered proteins"

(22) **M.R. Jensen**, P. Bernado, K. Houben, L. Blanchard, D. Marion, R. W. H. Ruigrok, M. Blackledge*. Protein and Peptide Letters (2010), 17, 952-960.
"Structural disorder within Sendai virus nucleoprotein and phosphoprotein: Insight into the structural basis of molecular recognition"

(21) **M.R. Jensen***, L. Salmon, G. Nodet, M. Blackledge*. J. Am. Chem. Soc. (2010), 132, 1270-1272.
"Defining conformational ensembles of intrinsically disordered and partially folded proteins directly from chemical shifts"

(20) S. Gely, D.F. Lowry, C. Bernard, **M.R. Jensen**, M. Blackledge, S. Costanzo, J.-M. Bourhis, H. Darbon, G. Daughdrill*, S. Longhi*. J. Mol. Recog. (2010), 23, 435-447.
"Solution structure of the C-terminal X domain of the measles virus phosphoprotein and interaction with the intrinsically disordered C-terminal domain of the nucleoprotein"

(19) G. Nodet, L. Salmon, V. Ozenne, S. Meier, **M.R. Jensen**, M. Blackledge*. J. Am. Chem. Soc. (2009), 131, 17908-17918.
"Quantitative description of backbone conformational sampling of unfolded proteins at amino acid specific resolution from NMR residual dipolar couplings"

(18) **M.R. Jensen**, P. Markwick, S. Meier, C. Griesinger, M. Zweckstetter, S. Grzesiek, P. Bernado, M. Blackledge*. Structure (2009), 17, 1169-1185.
"Quantitative determination of the conformational properties of partially folded and intrinsically disordered proteins using NMR dipolar couplings"

(17) M.-K. Cho, G. Nodet, H.-Y. Kim, **M.R. Jensen**, P. Bernado, C.O. Fernandez, S. Becker, M. Blackledge, M. Zweckstetter*. Protein Sci. (2009), 18, 1840-1846.
"Structural characterization of alpha-synuclein in an aggregation prone state"

(16) F. Gabel*, **M.R. Jensen**, G. Zaccai, M. Blackledge*. J. Am. Chem. Soc. (2009), 131, 8769-8771.
"Quantitative model-free analysis of urea binding to unfolded ubiquitin using a combination of small angle X-ray and neutron scattering"

(15) J.-L. Ortega-Roldan, **M.R. Jensen***, B. Brutscher, A. I. Azuaga, M. Blackledge*, N.A.J. van Nuland. Nucleic Acids Res. (2009), 37, e70.
"Accurate characterization of weak macromolecular interactions by titration of NMR residual dipolar couplings. Application to the CD2AP SH3-C:Ubiquitin complex"

(14) E.E. Büllesbach*, M.A.S. Hass, **M.R. Jensen**, D.F. Hansen, S.M. Kristensen, C. Schwabe, J.J. Led. Biochemistry (2008), 47, 13308-13317.
"Solution structure of a conformationally restricted fully active derivative of the human relaxin-like factor"

(13) **M.R. Jensen**, M. Blackledge*. J. Am. Chem. Soc. (2008), 130, 11266-11267.
"On the origin of NMR dipolar waves in transient helical elements of partially folded proteins"

(12) **M.R. Jensen**, K. Houben, E. Lescop, L. Blanchard, R.W.H. Ruigrok, M. Blackledge*. J. Am. Chem. Soc. (2008) 130, 8055-8061.
"Quantitative conformational analysis of partially folded proteins from residual dipolar couplings: Applications to the molecular recognition element of Sendai virus nucleoprotein"

(11) **M.R. Jensen**, S.M. Kristensen, C. Keeler, H.E.M. Christensen, M.E. Hodsdon, J.J. Led*. Proteins: Structure, Function and Bioinformatics (2008), 73, 161-172.
"Weak self-association of human growth hormone investigated by nitrogen-15 NMR relaxation"

(10) M. Wells, H. Tidow, T.J. Rutherford, P. Markwick, **M.R. Jensen**, E. Mylonas, D.I. Svergun, M.

Blackledge*, A.R. Fersht*. Proc. Natl. Acad. Sci. U.S.A. (2008) 105, 5762-5767.
"Structure of tumor suppressor p53 and its intrinsically disordered N-terminal transactivation domain"

(9) M.A.S. Hass, **M.R. Jensen**, J.J. Led*. Proteins: Structure, Function and Bioinformatics (2008) 72, 333-343.
"Probing electric fields in proteins in solution using NMR spectroscopy"

(8) **M.R. Jensen**, M.A.S. Hass, D.F. Hansen, J.J. Led*. Cell. Mol. Life Sci. (2007) 64, 1085-1104.
"Investigating metal binding in proteins by nuclear magnetic resonance"

(7) **M.R. Jensen**, S.M. Kristensen, J.J. Led*. Magn. Reson. Chem. (2007) 45, 257-261.
"Elimination of spin diffusion effects in saturation transfer experiments: Application to hydrogen exchange in proteins"

(6) **M.R. Jensen**, J.J. Led*. Biochemistry (2006) 45, 8782-8787.
"Metal-protein interactions: Structure information from Ni2+ induced pseudocontact shifts in a native non-metalloprotein"

(5) **M.R. Jensen**, D.F. Hansen, U. Ayna, R. Dagil, M.A.S. Hass, H.E.M. Christensen, J.J. Led*. Magn. Reson. Chem. (2006) 44, 294-301.
"On the use of pseudocontact shifts in the structure determination of metallo-proteins"

(4) **M.R. Jensen**, G. Petersen, C. Lauritzen, J. Pedersen, J.J. Led*. Biochemistry (2005) 44, 11014-11023.
"Metal binding sites in proteins: Identification and characterization by paramagnetic NMR relaxation"

(3) **M.R. Jensen**, J.J. Led*. J. Magn. Res. (2004) 167, 169-177.
"Determination of the electron relaxation rates in paramagnetic metal complexes: Applicability of available NMR methods"

(2) **M.R. Jensen**, C. Lauritzen, S.W. Dahl, J. Pedersen, J.J. Led*. J. Biomol. NMR (2004) 29, 175-185.
"Binding ability of a HHP-tagged protein towards Ni2+ studied by paramagnetic NMR relaxation: The possibility of obtaining long-range structure information"

(1) **M.R. Jensen**, D.F. Hansen, J.J. Led*. J. Am. Chem. Soc. (2002) 124, 4093-4096.
"A general method for determining the electron self-exchange rates of blue copper proteins by longitudinal NMR relaxation"

Book chapters:

(2) **M.R. Jensen**, V. Ozenne, L. Salmon, G. Nodet, P. Markwick, P. Bernado, M. Blackledge. Chapter in "Protein NMR Spectroscopy: Principal Techniques and Applications", G. Roberts and L.-Y. Lian (Eds.), Wiley-Blackwell (2011).
"Studying partially folded and intrinsically disordered proteins using NMR residual dipolar couplings"

(1) M. Blackledge, P. Bernado, **M.R. Jensen**. Chapter in "Instrumental analysis of intrinsically disordered proteins: Assessing structure and conformation", V. N. Uversky and S. Longhi (Eds.), John Wiley and Sons, New Jersey (2009).
"Atomic-level characterization of disordered protein ensembles using NMR dipolar couplings"