Nuclear Magnetic Resonance Studies of the Dynamics and Thermodynamics of Intrinsically Disordered Proteins

Anton Abyzov

Communauté
UNIVERSITÉ Grenoble Alpes

THÈSE

Pour obtenir le grade de

DOCTEUR DE LA COMMUNAUTÉ UNIVERSITÉ GRENOBLE ALPES

Spécialité : **PHYSIQUE POUR LES SCIENCES DU VIVANT**

Arrêté ministériel : 7 août 2006

Présentée par

Anton ABYZOV

Thèse dirigée par **Martin BLACKLEDGE** et
codirigée par **Malene Ringkjøbing JENSEN**

préparée au sein de l'équipe **Flexibilité et Dynamique des Protéines par RMN de l'Institut de Biologie Structurale**
dans **l'École Doctorale de Physique de Grenoble**

Nuclear Magnetic Resonance Studies of the Dynamics and Thermodynamics of Intrinsically Disordered Proteins

Thèse soutenue publiquement le **11 mars 2016**,
devant le jury composé de :

Prof. Franz BRUCKERT
Grenoble INP, Grenoble, Président

Dr. Carine VAN HEIJENOORT
Institut de chimie des substances naturelles, Gif sur Yvette, Rapporteur

Prof. Christian ROUMESTAND
Université de Montpellier, Montpellier, Rapporteur

Dr. Guido PINTACUDA
Institut des Sciences Analytiques, Lyon, Examinateur

Dr. Martin BLACKLEDGE
Institut de Biologie Structurale, Grenoble, Directeur de thèse

Remerciements

Je remercie d'abord mon directeur de thèse, Martin Blackledge, et mon co-directeur de thèse, Malene Ringkjøbing Jensen, pour leur investissement à mes côtés pour faire aboutir ce projet fondamental et complexe.

Je souhaite également exprimer ma reconnaissance aux membres de jury qui ont accepté d'évaluer mon travail de thèse. Je tiens à remercier Franz Bruckert pour avoir présidé mon jury. Je souhaite ensuite exprimer ma gratitude aux rapporteurs, à Christian Roumestand pour avoir analysé mon manuscrit en profondeur, et à Carine van Heijenoort pour son esprit critique. Je voudrais aussi dire un très grand merci à Guido Pintacuda, qui a fait émerger les discussions très intéressantes autour de la question de l'étude par la RMN de la dynamique des protéines lors de la soutenance de ma thèse.

Je remercie infiniment Nicola Salvi pour sa contribution à l'analyse et l'interprétation des données expérimentales, et Guillaume Bouvignies pour m'avoir aidé à différentes étapes de mon projet[1]. Un grand merci à Damien Maurin pour avoir assuré la production des protéines. Je remercie aussi tous les autres membres du groupe Flexibilité et Dynamique des Protéines, présents ou anciens, pour leur contribution, les discussions intéressantes ou tout simplement pour leur présence qui a marqué mon quotidien : Elise Delaforge, Stefaniia Ivashchenko, Sigrid Milles, Loïc Salmon, Benjamin Nemoz, Koji Umezawa, Robert Schneider, Jaka Kragelj, Guillaume Communie, Valéry Ozenne, Eric Condamine, Jie-Rong Huang, Paul Guerry, Luca Mollica, Pavel Srb, Emma Weiss, Katelyn Jackson…

Je tiens à remercier également tous mes amis thésards grenoblois pour les moments inoubliables que nous avons vécu ensemble ces années : Vitaly, Boris, Elena, Pavel, Petr… Je pense aussi à mes amis œnologues du Club Œnologie de Grenoble INP : Manu, Lauriane, Auriane, Anastasia, Polina et bien sur l'incroyable magicien Ferdi, qui a presque réussi un jour à faire disparaître mon vélo… Je remercie aussi mes voisins au 'château', Véronique, Filiz et François-Pierre, qui sont des personnes très intelligentes et gentilles avec qui je pouvais parler de tout et qui m'ont appris diverses choses sur l'histoire et la culture française. Et merci encore à Véronique pour avoir partagé avec moi ses œuvres gastronomiques !

[1] Je vous remercie aussi, Guillaume, Nicola et Loïc, pour les discussions très intéressantes sur les questions de société et d'actualité que nous avons eues ensemble autour d'un café.

Je souhaite aussi exprimer ma reconnaissance à la beauté de la nature autour de Grenoble, surtout dans les massifs du Vercors, de Chartreuse et de la vallée du Grésivaudan. Ils laissent des souvenirs inoubliables à ceux qui randonnent, ou à ceux qui, comme moi, montent les montagnes à vélo.

Bien sûr, je tiens à exprimer ma profonde gratitude à mes parents et à mon frère qui m'ont soutenu et conforté dans mes choix tout au long de mes études et de ma thèse. Sans vous, je n'aurais pu parcourir autant de chemin. Et enfin je ne pourrais jamais oublier l'enseignement de mon premier professeur de biologie, Ruslan Shalamov, qui a su me transmettre sa passion pour la biologie !

CONTENTS

Introduction

Understanding the role of proteins in life and how they function is a quest that began a long time ago but still seems only to be in its initial stages. In the 18th century, a French chemist Antoine de Fourcroy [1] along with other scientists, described a separate class of organic molecules that were subject to coagulation (denaturation) after being exposed to heat or acids. These molecules were given a name "proteins" only in 1838, when a Swedish chemist J. Berzelius and Dutch chemist G. Mulder, after an elemental analysis of 'albumins', discovered that they bore a strikingly similar composition [2]:

> "Le nom protéine que je vous propose pour l'oxyde organique de la fibrine et de l'albumine, je voulais le dériver de πρωτεῖος, parce qu'il paraît être la substance primitive ou principale de la nutrition animale"[1]
>
> — Letter from J. Berzelius to G. Mulder, 10 July 1838.

At the end of the 19th century most amino acids were discovered and at the beginning of the 20th century a German chemist Emil Fischer experimentally proved that proteins are composed of amino acids linked by CO-NH peptide bonds. However, at the time their role was unclear; many scientists including the Nobel laureate Richard Willstätter did not believe that enzymes were actually proteins [3]. Only in 1926 an American chemist James B. Sumner extracted and crystallized the enzyme urease, and then proved that it was a pure protein [4], thus establishing a central role for proteins in living organisms. Later he received the Nobel Prize for his work.

For a long period of time proteins were perceived as rigid molecules with a well-defined three-dimensional structure in their functional state. Intermolecular interactions were explained by the "lock and key" model developed by E. Fisher in 1894 for enzymes [5], which describes the necessity for geometrical and charge complementarity of interacting partners. In 1958, the first three-dimensional (3D) structure of a protein was obtained by X-ray crystallography [6]. The same year, some flexibility was introduced into the classical "lock and key" model by Daniel Koshland, who proposed an "induced fit" mechanism [7]: both the active site of an enzyme and the substrate can change their conformation in the course of binding. Yet, until quite recently, it was believed that the biological function of a protein is encoded in its native 3D structure, and the presence of functional domains lacking stable tertiary structure was not recognized [8]. Only at the end of the 20th and beginning of 21st century did it became apparent that a large fraction of the human genome codes for fully disordered proteins or for proteins containing large

[1] English translation: "The name protein which I propose to you for the organic oxide of fibrin and albumin, I wanted to derive it from πρωτεῖος, because it seems to be the original or principal substance of animal nutrition."

disordered domains and that these proteins play an important role in a number of human pathologies [9]. It became clear that understanding the role of the conformational flexibility of intrinsically disordered proteins (IDPs) would be necessary in order to understand their function, and eventually to develop pharmacological solutions targeting these proteins.

To study IDPs, an approach is necessary that takes into account the flexibility and dynamics of the protein chain. Nuclear Magnetic Resonance (NMR) spectroscopy has emerged as a technique particularly adapted to study motions in proteins at atomic resolution and over a wide range of time scales.

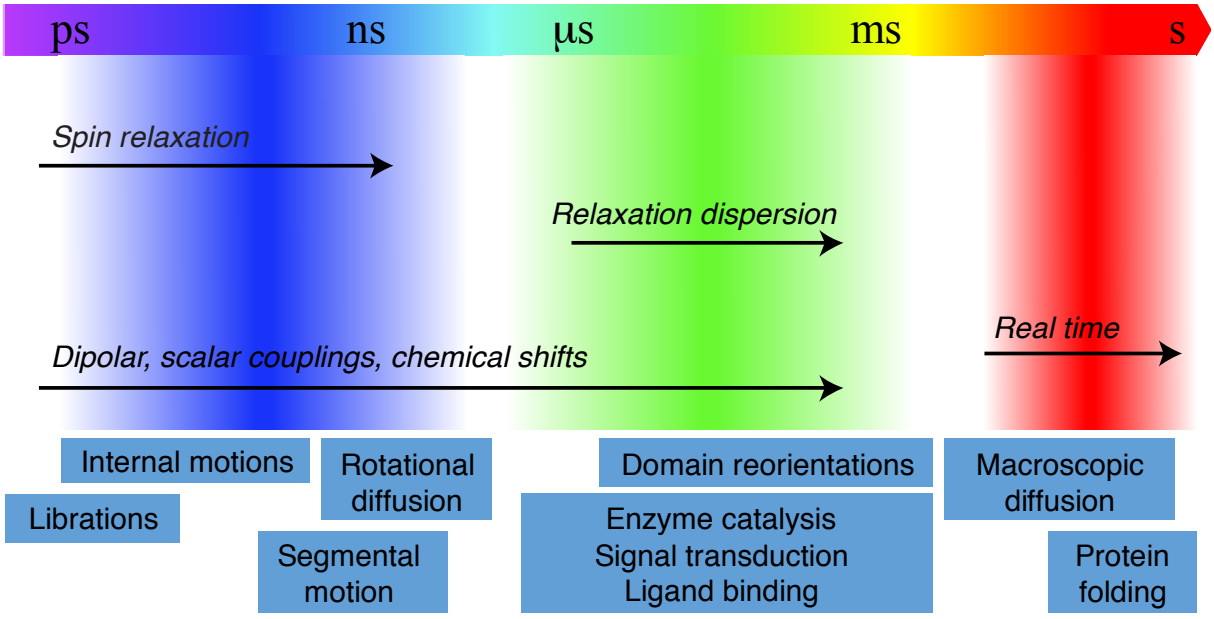


Fig. 0.1 Different timescales of motions in proteins together with the measurable NMR parameters that are sensitive to these time scales.

NMR gives us information about the behaviour of different nuclei in the protein that possess a non-zero spin. Nuclei with spin 1/2 (^{1}H, ^{13}C, ^{15}N) are the easiest to study. In theory, using appropriate NMR methods such as multidimensional experiments, as well as isotope labelling techniques to enrich the protein in ^{13}C and ^{15}N, it is possible to characterize the behaviour of each ^{1}H, ^{15}N or ^{13}C nucleus present in the molecule.

NMR is sensitive to motions on a broad range of timescales from picoseconds to seconds (Fig. 0.1). Of all the various NMR experiments that can be used to study protein dynamics, heteronuclear spin relaxation is particularly sensitive to protein backbone motions occurring on the pico- to nanosecond (ps-ns) timescale.

The main objective of this thesis is to study the dynamics of IDPs in terms of time scales and motional amplitudes, on the basis of multiple types of NMR spin relaxation rates

measured at different magnetic field strengths and temperatures. The thesis is organized as follows:

- Chapter 1 is a general introduction to the field of intrinsically disordered proteins. Particular emphasis is placed on available approaches for the study of IDPs and on their mechanisms of interaction with partners.

- Chapter 2 presents the context and the purpose of this work. It begins with the introduction of the protein under investigation, N_{TAIL} (the C-terminal domain of the nucleoprotein of *Sendai paramyxovirus*), which is a disordered protein containing transiently populated helical structures within its molecular recognition element.

- Chapter 3 presents the main aspects of NMR relaxation theory that are relevant to this thesis, including the relationship between relaxation rates and angular correlation functions of relaxation active interactions.

- In Chapter 4, studies of polymers are highlighted and diverse approaches developed to describe their dynamics are introduced. Transition is then made to NMR studies of IDPs followed by a presentation of different models used to describe the spectral density function (a Fourier transform of correlation functions). The "model-free" approach, used in this thesis, is discussed in more detail.

- In Chapter 5, we present briefly experimental aspects of this work, in particular the pulse sequences that were used to measure relaxation rates, and some aspects of the data analysis.

- Chapter 6 is devoted to the presentation of the work carried out in this thesis. The analysis of relaxation rates using an extended-model-free-like approach, the influence of chain length and temperature and the relevance of different timescales of motions extracted from relaxation rates are presented in the form of a submitted manuscript, which represents a large part of the chapter. Complementary results are presented that were not included in the manuscript, in particular the study of slower timescale motions in N_{TAIL}.

- Finally, the last chapter is a general conclusion to the thesis and perspectives.

Chapter 1. The world of intrinsically disordered proteins

This chapter introduces the field of intrinsic protein disorder. It starts with the prediction and proof of intrinsic disorder in proteins followed by a description of in-cell studies of IDPs. In the end, the question of the role of IDPs in biology and their interaction with other proteins is raised.

In the field of molecular biology, the second part of the 20th century was marked by a striking increase in solved and published structures of proteins. Seemingly, as the structure-function paradigm was predominant, the main purpose of structural biology was to express and purify proteins, refine crystallization conditions and solve high-resolution structures. Yet, in the end of the previous century the question arose of the relationship between internal dynamics in proteins and their functions [10]. Rapidly, not only the role of flexibility in folded proteins, but also the existence and importance of proteins and protein regions that are fully functional in the absence of a stable fold was generally gaining credibility. These intrinsically disordered proteins (IDPs) are found to be highly abundant in nature and exhibit numerous biological functions. In recent years, an enormous interest has emerged into how these proteins function at the molecular level as witnessed by the exponentially growing number of publications and citations concerning IDPs since 1995 (Fig. 1.1).

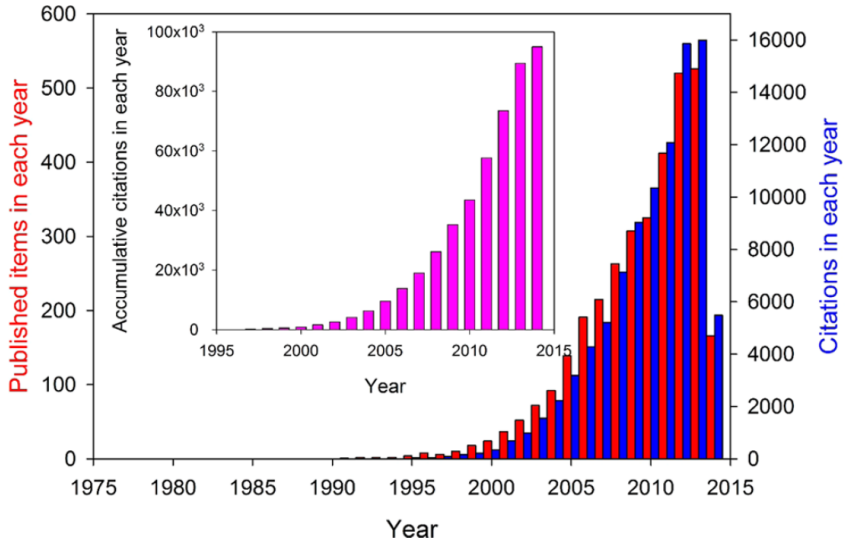


Fig. 1.1
Number of publications, number of citations about IDPs per year and in total (inner set). Taken from [11].

The pivotal point of the transformation of the "structure-function" paradigm was linked to the onset of bioinformatics. In 1998 Romero et al., published a study [12] in which they predicted, using neural network trained on primary sequences of a relatively small number of known disordered protein regions, that not less than 15000 proteins from the SwissProt database have disordered regions of 40 consecutive amino acids or longer. This study was completed by another study, in which the authors predicted that regions

coding for intrinsically disordered segments of more than 50 residues in length are well present in genomes of organisms from the three kingdoms of life, going up to 41% for Drosophila [13]. Around 75% of signaling proteins in mammals are predicted to contain disordered regions over 30 residues, and around 25% are predicted to be fully disordered [14].

However, even these discoveries were not sufficient to convince the scientific community of the importance of intrinsic disorder in nature: the idea that a structure-less protein may be functional was considered by some as a heresy [15]. In order to overcome this widespread scepticism it was necessary to show that IDPs are really disordered, that they may stay disordered even when interacting with partners, that they retain this disorder inside living cells, that they have important biological functions, and, finally, that it is possible and necessary to study their behaviour in order to develop effective drugs. The last question is still far from being solved.

1.1 Predicting disorder

There are several methods to predict the degree of disorder from a given sequence of amino acids. Some of them depend on a database of sequences known to be disordered (mostly, those that are lacking structure in the PDB database), others rely on "ab initio" prediction.

1.1.1 Amino acid composition of IDPs

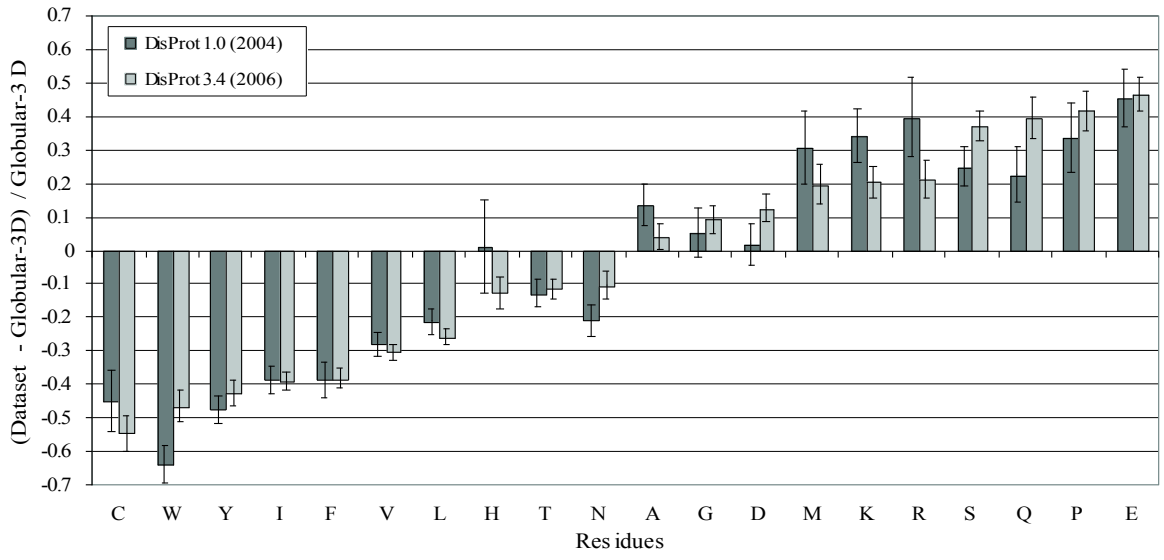


Fig. 1.2 The differences between the amino acid compositions of disordered data sets (DisProt 1.0 and Disprot 3.4) and that of an ordered data set (Globular-3D). Amino acids are arranged by the peak height as in DisProt 3.4. For details see [16].

Already in 2001 it was discovered [17] that disordered and folded proteins have a different content of certain amino acids. The amino acid composition of disordered protein segments relative to globular proteins (Fig. 1.2) show that IDPs are generally enriched in charged and structure-breaking residues (Pro and Gly), and depleted in hydrophobic and aromatic residues. Proteins with this amino acid composition have low mean hydrophobicity (which is a driving force of protein compaction) and high net charge (thus contributing to intramolecular repulsion), and thus are less likely to adopt a stable fold. IDPs also have a generally lower sequence complexity [18], in particular they abound in patterns containing Pro and certain charged amino acids (such as Glu) [19].

1.1.2 Prediction of protein disorder by computational tools

It has been shown that basic physico-chemical properties of a protein can predict its degree of disorder. On a CH plot (net charge versus mean hydrophobicity) natively folded and unfolded proteins tend to form two separate clusters. FoldIndex [20], PreLink [21] and GlobPlot [22] are predictors based on this approach (see [23]).

More complicated prediction tools use databases of experimentally characterized disordered proteins: DisProt [24], IDEAL [25], MobiDB [26]. They are based on neural networks trained on databases of folded and disordered proteins, comparing sequence and/or physico-chemical properties of different segments of a given protein to those in databases. For example, the PONDR predictor [18] calculates different attributes (such as hydropathy, fractional composition of particular amino acids or sequence complexity) of a sequence over windows from 9 to 21 residues for the proteins in the databases and uses these attributes to train its neural network. The same attributes are calculated over a given sequence and used as input for the neural network. Other examples of predictors using trained neural networks are DisEMBL [27], DISOPRED2[28] and RONN [29].

The drawback of this approach is that in those databases the regions of missing electron density in PDBs are overrepresented. These regions are typically short, as extensive disorder prevents crystallization, so the above-mentioned prediction tools are necessarily trained on short disordered regions and could be less efficient in predicting long disorder.

Another way of predicting disorder is based on the idea that different amino acids have different propensities to form contacts with each other and that amino acids that avoid each other are overrepresented in IDPs because they fail to form intramolecular contacts. The FoldUnfold tool calculates the expected number of contacts per amino acid residue from the protein sequence [30]. Another predictor IUPred [31] is based on the contribution to the fold stability of each amino acid in folded proteins.

A semi-automated method called HCA (Hydrophobic Cluster Analysis) is based on a two-dimensional graphic representation of protein sequence where hydrophobic residues forming clusters are marked (see Fig. 1.3). Disordered regions are recognizable as those depleted of hydrophobic clusters [32].

All these methods have their advantages and drawbacks. Some of them are better for short disordered regions (as DISOPRED2 and PreLink), while others (as IUPred) are more efficient in predicting long disordered segments. Finally some predictors, such as PONDR, GlobPlot, and FoldIndex, have been trained on both short and long disorder and provide a balanced performance. For this reason, to perform a reliable prediction of disordered regions, so-called "metapredictors" seek a consensus prediction based on scores obtained from different methods (PONDR-FIT is one example [33]). Another option is to apply different predictors independently and combine their output in a single analysis. The MeDor (MEtaserver of DisORder) server [34] combines the output of different disorder prediction servers and a HCA plot (see Fig. 1.3).

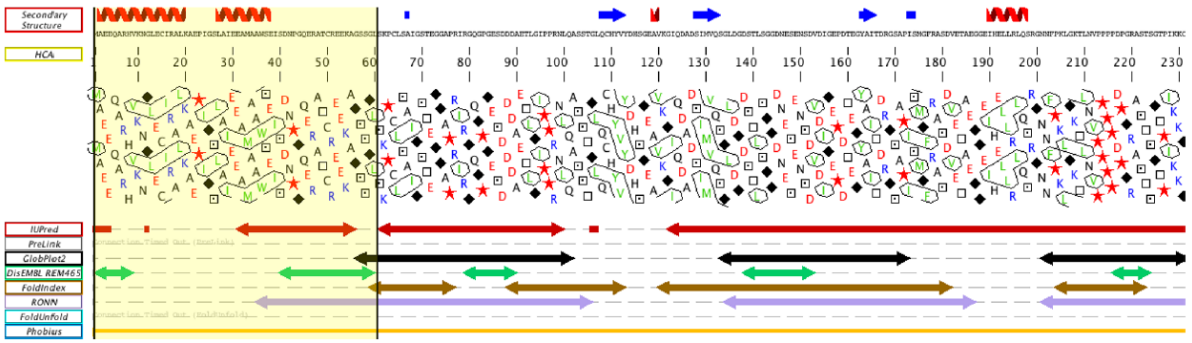


Fig. 1.3 MeDor output of the N-terminal domain of measles virus phosphoprotein. The upper line above the sequence shows secondary structure elements predicted by Pred2ary [35]. Below the sequence are shown the HCA plot and the output of different disorder predictors (disordered regions depicted as arrows).

At this stage of their development, disorder predictors do not in any way provide a proof of disorder in a given protein, they are rather a tool to find potentially disordered regions that merit further experimental studies. However, with the growing number of experimentally characterized IDPs, one can hope that in the future the sequence-based criteria of intrinsic disorder would be further established and experimentally validated.

1.2 Experimental characterization of intrinsically disordered proteins

One of the simplest methods to detect intrinsic disorder in proteins is **circular dichroism** (CD) **spectroscopy**. CD spectra provide estimates of secondary structure and allow for the distinction between different types of secondary structure elements (α-helices, β-sheets, turns, etc.) [36].

IDPs can also be recognized by their peculiar behaviour in SDS polyacrylamide gel electrophoresis, gel filtration or size-exclusion chromatography, where they have unusually high apparent molecular masses (mostly due to their amino acid composition and extension of their chains) [23]. They are also much more sensitive to protease digestion [17]. Finally, IDPs are much less sensitive to high temperatures and acidic treatment than folded proteins, which may precipitate at these conditions; this property is often exploited for purification purposes [23].

1.2.1 Nuclear magnetic resonance (NMR)

Already a one-dimensional proton NMR spectrum would give us an indication about protein disorder. In an IDP, inter-residue contacts are quasi-absent, and all residues have a similar chemical environment. As the chemical shifts of protons are more sensitive to the general structural context than to the local sequence composition [37], the dispersion of the resonances of IDPs in the proton dimension is limited. But the force of NMR comes from its possibility to study proteins at the residue-specific level, using high-field spectrometers and multidimensional NMR spectroscopy.

NMR is only sensitive to nuclei with nonzero spin. Nuclei with spin more that ½ give broad resonance lines due to electric quadrupole moment and hence are more difficult to study [38]. Therefore, the most important nuclei for biomolecular NMR are those with a spin of ½, as for example ^{1}H, ^{13}C and ^{15}N. In nature, the content of ^{13}C is 1.07% and that of ^{15}N is 0.364%. As this is not enough for most NMR experiments, especially for multidimensional experiments, due to limitations in accessible protein concentration, one of the most important requirements is to be able to express a protein heterologously, in a host that allows for uniform labelling of a protein. Most proteins for NMR studies are expressed in *E.coli*, which is the cheapest and the most simple expression system (which is still quite expensive, especially if not only ^{15}N but also ^{13}C labelling is required). However, there are some recent advances in expressing labelled proteins in insect cells [39, 40].

The starting point of each protein NMR study and one of the simplest spectra to run is the HSQC (heteronuclear single quantum coherence) experiment. This two-dimensional spectrum (^{15}N-^{1}H HSQC) correlates the backbone nitrogen with its directly attached amide proton and, therefore, displays a resonance at the position corresponding to the chemical shifts of the nitrogen and the amide proton. A set of three-dimensional experiments can be run in order to measure ^{13}Cα, ^{13}Cβ and ^{13}CO chemical shifts and link them to the ^{15}N and ^{1}HN shifts of the same and adjacent residue. As ^{15}N and especially ^{13}C chemical shifts are particularly sensitive to the type of amino acid residue, it allows us to establish a direct link between peaks and the protein sequence. This process is known as backbone resonance assignment and once completed, provides access to a variety of

information about the structure and dynamical behaviour of the protein at the residue-specific level, by running different types of NMR experiments (see also Fig. 1.1).

For unfolded and partially folded proteins, one of the most informative primary NMR observables is the **chemical shift**, because it is sensitive to polypeptide backbone torsion angles [23]. Secondary $^{13}C\alpha$, $^{13}C\beta$, ^{13}CO and $^{1}H\alpha$ chemical shifts are calculated as the deviation of the experimental chemical shifts from so-called random coil values [41]. The random coil values are amino acid specific and represent the expected chemical shifts for a protein devoid of secondary and tertiary structure. These values have traditionally been obtained by measuring the chemical shift values of each of the twenty amino acids within small peptides [41]. The experimental secondary chemical shifts are sensitive to the local structure and hence provide us with information about secondary structure content in the protein [23]. It should be mentioned that this approach is sensitive to the correct frequency referencing in the NMR experiment and to overcome this problem, $^{13}C\alpha$ and $^{13}C\beta$ secondary shifts can be used simultaneously to correct for a potential reference offset as the $^{13}C\alpha$ and $^{13}C\beta$ chemical shifts show an inverse dependence on α-helical propensity [42].

NMR relaxation is sensitive to backbone and side chain dynamics of IDPs. The most commonly used techniques involve the measurement of R_1, R_2 and heteronuclear NOEs for backbone resonances, and allow for a detailed characterization at the residue-specific level of contributions of different motions to the dynamics of the IDP chain on the ps-ns timescale. $^{15}N/^{13}C$ R_2 relaxation dispersion [43] and proton $R_{1\rho}$ relaxation dispersion [44] allow the characterization of exchange in proteins on the μs – ms timescale. NMR relaxation is more extensively discussed in the subsequent chapters of this thesis.

With NMR, we are not limited to the study of local structural propensities and local dynamics. **Paramagnetic relaxation enhancement** (PRE) is an excellent probe of long-range interactions in disordered proteins. This method relies on the introduction of a paramagnetic nitroxide spin-label at a specific position in the protein, usually via a conjugation with a solvent-exposed cysteine residue, which can be introduced by site-directed mutagenesis [45, 46]. The presence of an unpaired electron causes a broadening of the nuclear spin resonances within a radius of about 25 Å [47] due to the strong dipolar interaction [48]. HSQC spectra are typically recorded with the spin-label in its paramagnetic (oxidized) and diamagnetic (reduced) state. Differences in the line width, relaxation rates, or intensity in these two spectra give an estimate of the average distance between the spin label and the given residue. The limitations of this approach are linked to the introduction of a non-native residue in the protein, and the risk of perturbation of native long-range contacts by its side chain and/or the paramagnetic label. However, it offers a possibility to elucidate a fluctuating tertiary structure in a disordered protein that would be difficult to describe using other approaches [49].

^{1}H-^{1}H nuclear Overhauser enhancements (NOEs) were used to calculate structure of folded proteins since a long time [50]. In theory, NOEs can also be used to obtain more detailed information about local conformational sampling in IDPs [51]. However, quantitative interpretation of NOEs is complicated by their strong sensitivity on the range of dynamic timescales commonly encountered in unfolded proteins [49].

Another powerful method to assess local and global order in proteins is to measure **Residual Dipolar Couplings** (RDCs). Dipolar couplings between different spins (nuclei) are a very rich source of structural information because they depend only on the internuclear distance (r^{-3}), the orientation of the internuclear vector relative to the static magnetic field, and known physical constants. In an isotropic solution, however, all orientations are sampled with the same probability, and are therefore efficiently averaged to zero. But if the molecule is partially aligned in the magnetic field, one of the orientations is preferred and so-called *residual* dipolar couplings appear as additional contributions to the well-known scalar couplings. To obtain this partial alignment of a protein molecule, a number of approaches are available, most of them involve dissolving the protein in some kind of anisotropic medium such as solutions of liquid crystals [52], lipid bicelles [53], filamentous bacteriophages [54] or polyacrylamide gels [55, 56].

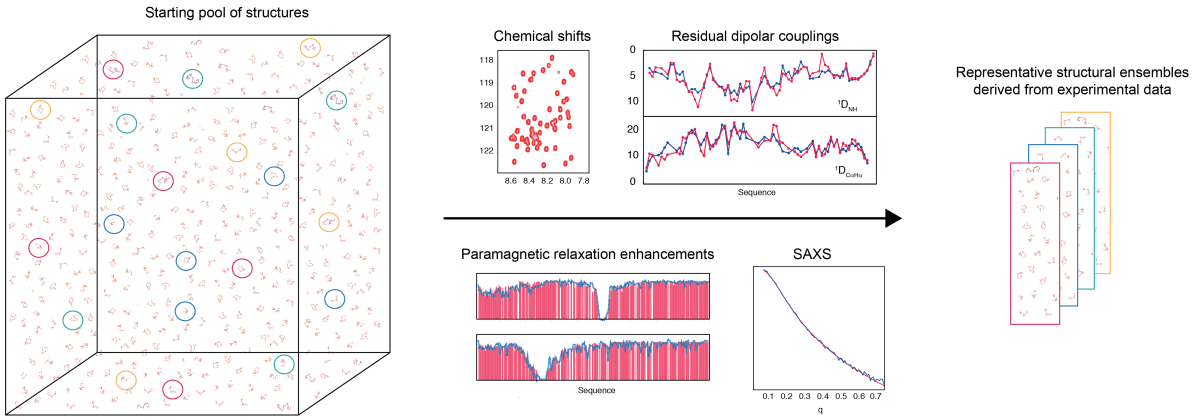


Fig. 1.4. Deriving representative ensembles of IDPs using a sample-and-select approach. Flexible-Meccano is used to generate a large pool of statistical coil conformers that is assumed to efficiently sample conformational space. Typically a pool of 20000 conformers is generated. The genetic algorithm ASTEROIDS is used to select a sub-ensemble in agreement with different complementary types of experimental data. The sizes of the sub-ensembles are adjusted depending on the protein investigated. Typical sub-ensembles contain 100-200 conformers.

Measuring RDCs has its own pitfalls; in particular, care should be taken to ensure that the protein (especially an IDP, which may have exposed hydrophobic groups) does not interact significantly with the alignment medium.

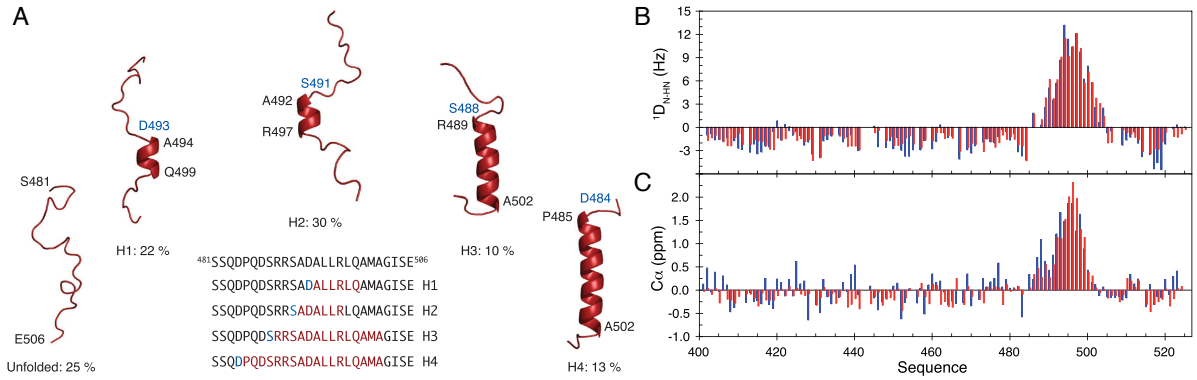


Fig. 1.5. A. Description of the conformational ensemble of the α-MoRE of the C-terminal domain, N_{TAIL}, of measles virus. B. Experimental (blue) and back-calculated from the ensemble (red) $D_{N\text{-}HN}$ RDCs for measles virus N_{TAIL}. C. Experimental (blue) and back-calculated C^{α} values.

RDCs are a valuable source of information on both structure and dynamics in disordered proteins [49, 57, 58]. An algorithm called ASTEROIDS (Fig. 1.4) was developed to use RDCs, chemical shifts and paramagnetic relaxation for the description of the conformational ensembles of IDPs. In this approach, an ensemble of random structures of an IDP is initially generated using a statistical coil generator, Flexible-Meccano [59, 60], on the basis of the primary sequence of the protein. If the experimental data are predicted correctly, then one can already conclude that a random-coil, or statistical-coil model, accurately described the conformational behaviour in solution. If there are significant deviations from coil behaviour, the genetic algorithm ASTEROIDS [61] can then be employed to select conformational ensembles that give the best fit between back-calculated and experimentally measured RDCs and chemical shifts. As an example, using the combination of Flexible-Meccano and ASTEROIDS, the conformational ensemble representing the α-MoRE (α-helical molecular recognition element) of the disordered C-terminal domain of measles virus nucleoprotein (Measles N_{TAIL}) was described (Fig. 1.5, [62]).

1.2.2 Small angle X-ray scattering

Small-angle X-ray scattering (SAXS) is a powerful biophysical method that allows the structural characterization of biological systems up to the resolution of several nm. It is known to provide basic geometrical information about the 3D shape and size of objects, such as biomolecules, in solution (e.g. radius of gyration, shape elongation), and has been used extensively to provide *ab-initio* reconstructions of the overall envelope of proteins [63, 64]. SAXS is now more and more used to study IDPs [65], in general as a complementary method to NMR studies. Already standard analyses of SAXS profiles, such as the Kratky representation and Guinier analysis give us insight into the presence of transiently folded structures inside IDPs. In addition, the application of SAXS *ab initio* reconstruction to complexes of IDPs and folded proteins can reveal the presence of flexible chains attached to globular particles, thus allowing the study of complex

formation [65]. Recently, different ensemble-selection tools were developed to describe protein flexibility by SAXS, of which the Ensemble Optimization Method (EOM, [66]) specifically addresses the question of structural ensemble descriptions of IDPs.

The major limitation of SAXS is that it requires a monodisperse sample, as the presence of oligomerization or aggregation would hamper the analysis of the shape or flexibility of macromolecules. However, if the aggregation is specific and the number of components in the system is limited (e.g. a monomer-dimer equilibrium) and the scattering profile of each oligomeric form can be predicted, it is possible to study the composition of the system [65].

1.2.3 Single-molecule approaches

One of the limitations of previously cited "in-bulk" approaches is that all conformations present in the sample at a given time are averaged. In single-molecule approaches, this issue is eliminated by detection of each individual molecule. Using **Förster resonance energy transfer** (FRET), which probes the distance between two fluorescent dyes attached to the protein at different positions, protein conformations as well as their time dependence down to the nanosecond time scale can be described using different experimental setups and analysis procedures [67].

The approaches based on atomic force microscopy (AFM) include high-speed AFM [68, 69], which enables the observation of dynamics in IDPs with sub-100ms to subsecond temporal resolution, as well as single-molecule force spectroscopy (SM-FS) [70]. The latter involves mechanically stretching the protein with an AFM tip and probes timescales from milliseconds to seconds and is particularly sensitive to the presence of elements of secondary structure in an IDP.

1.3 Intrinsically disordered proteins in living cells

The fact that IDPs exist and were characterized *in vitro* does not prove that intrinsic disorder is relevant in the context of a living cell. In the cell, a significant fraction of the total volume (20-30%) is occupied by macromolecules [71]. For example, inside an *E. coli* cell, the total concentration of proteins and RNA rise up to 400 mg/ml [72]. These excluded volume effects (crowding) may indeed have considerable impact on the conformational and dynamical properties of IDPs [73], and the possibility that they can force IDPs to become more extended, more compact or even fold needs to be assessed.
One possibility to answer this question is to mimic the crowding by studying an IDP in the presence of crowding agents. In order to approximate the crowding conditions in a cell as closely as possible, these agents should mimic the effect of volume exclusion [72], increased viscosity of local environment [74] as well as affecting the solvation/hydration

of proteins [75, 76]. High molecular weight polymers, such as dextran or Ficoll 70, increase the viscosity and generate the excluded volume effect. Osmolytes (such as sucrose, triflouroethanol (TFE), trimethylamine N-oxide (TMAO)), interact unfavorably with the protein backbone [77] and thereby stabilize secondary structure elements in IDPs [78].

Studies carried out on various IDPs under different conditions show that dextran and Ficoll 70 elicit some compaction but not folding of the proteins [79]. Osmolytes such as TFE or TMAO promote the formation of α-helical elements in some IDPs [80]. In a recent article [73], NMR relaxation experiments on different IDPs (ProTα, TC-1 and α-synuclein) in the presence of Ficoll 70 and Dextran 70 showed a limited impact of 160mg/ml of both agents on protein dynamics. At 400 mg/ml of Dextran 70, the chain dynamics of ProTα is slowed down, however, it still retains a certain level of flexibility. Ficoll 70 was equally found to induce a minor increase of helical propensity in a relatively structured region of the TC-1 protein [73]. In a more recent study [81], SM-FRET was used to study the behavior of different IDPs in the presence of polyethylene glycol (PEG). It was found that both increasing volume fraction and size of PEG lead to IDP compaction. However, other authors discourage the use of PEG as a crowding agent due to the promotion of attractive interactions between proteins [73]. In conclusion, the crowding effect does not fold IDPs but may promote the formation of local secondary structure elements [80].

Still, mimicking the intracellular environment using artificial crowders does not provide a full answer to the question of protein disorder in cells. One should directly assess the amount of disorder under in vivo conditions. One of the possibilities is to study proteasomal degradation of IDPs "by default" (i.e. ubiquitin-independent). This approach, applied to several IDPs, including p53 [82] and p21 [83], proved that the IDPs are degraded by the 20S proteaseome and thus supported the idea that the proteins stay disordered *in vivo*.

Another option is to study the protein by in-cell NMR [84]. This requires the deposition of labeled proteins within the cell for example by overexpressing them under labeling conditions, by covalently linking the proteins to cell-penetrating peptides [85], by electroporation [86, 87] or by microinjection into the cell cytoplasm [88]. Thus, a labeled protein is the only NMR observable macromolecule and allows investigation of the structural and dynamic properties of the protein under conditions close to those in vivo.

Different proteins were found to stay disordered even under those conditions. Tau protein was studied in *Xenopus oocytes* by in-cell NMR and was found to stay partially disordered [89]. In a recent article, Binolfi et al. [90] and Waudby et al. [91] studied the structure and dynamics of α-synuclein in E. coli cells. The results show that in the bacterial cytosol, α-synuclein populates a highly dynamic state, which has the same

characteristics as the disordered monomeric state in solution, despite being in a highly crowded environment. Very recently, α-synuclein was found to stay in disordered and highly dynamic monomeric state in mammalian cells [86].

1.4 Interactions of IDPs with their partners

Intrinsically disordered proteins play different roles in the cell. From the "mechanistic" point of view, i.e. from their mechanism of action at the molecular level, they can be classified into different classes [92]. In one such classification, the first class corresponds to "entropic chains", i.e. IDPs whose function does not depend on binding to other molecules: linkers, spacers, entropic springs, clocks and bristles, etc. The other four classes involve molecular recognition and hence binding to a target: RNA or DNA, another protein or a small ligand. "Effector" IDPs alter the action of their partner upon binding (e.g. inhibitors or activators). "Scavengers" store and/or neutralize small ligands, such as ions and organic compounds, either for disposal or for sequestration and later release upon the need of the organism. "Assembler" IDPs assemble, stabilize, and regulate large multiprotein complexes such as the ribosome, cytoskeleton, and chromatin. The fifth class is constituted from IDPs that act as display sites for the mediation of regulatory post-translational modifications (e.g. phosphorylation). Finally, with the finding that a high proportion of disordered segments in RNA and protein chaperons and the discovery of entirely disordered chaperons [23], the sixth class of "chaperone" IDPs can be added.

Thereby, IDPs often function by molecular recognition (as can be seen from this classification), and also often fold, at least locally, upon interaction with their partners [93]. In signalling and regulatory proteins, multiple conserved sequence motifs are thought to mediate the interaction with nucleic acids or other proteins [94, 95]. As estimated recently, the human proteome may contain over 100.000 short linear binding motifs that are located within intrinsically disordered regions [96]. These motifs are structurally polymorphous, and found to adopt different structures on binding to different targets [94]. Depending on whether we look at the binding motifs from a structural or sequence point of view, these motifs are denoted in the scientific literature as "molecular recognition elements/features" (MoRE [97] or MoRF [98]), "linear motifs" (LM) [99], "eukaryotic linear motifs" (ELM) [100] or "short linear motifs" (SLiM) [101]. They are called "linear" because they do not involve bringing distant segments of the protein together to create a recognizable unit [99].

The process of interaction of these disordered motifs with their targets and their subsequent folding is generically termed "coupled folding and binding" [95], and its mechanism is currently extensively studied by many theoretical, experimental and *in silico* approaches. One of the most interesting questions is whether the folding happens before, concomitantly to or upon binding, and this in turn relates to the importance of the

conformational behaviour of the protein in its free, unbound state. The "conformational selection" hypothesis suggests that IDPs populate, at least partly, the bound state prior to interaction, and that the partner of the IDP recognizes this folded state. This means that the folding occurs before binding. In the "induced fit" hypothesis, the folding occurs after binding and the binding process, or at least the proximity of the partner, induces the folded state. Experimental evidence from both NMR and stopped-flow kinetics shows support for both hypotheses and will be discussed further below.

In 2004, Fuxreiter et al. developed the concept of preformed structural elements (PSEs) in the unbound state of IDPs, and predicted that these PSEs would favour binding to other proteins [102]. In this study, structures adopted by recognition elements of IDPs in the bound form were compared to those observed (or predicted) in the free form. This comparison suggested that IDPs in the free form have an intrinsic preference for the conformations they attain in the bound state (especially concerning α-helical elements). This preference suggests that the PSEs serve as initial contact points for the binding.

But the experimental evidence does not uniformly support this concept. An example is the study of the binding of pKID to KIX (pKID is the phosphorylated kinase-inducible domain of cyclic AMP-responsive element-binding protein (CREB); KIX is a domain of CBP, CREB-binding protein). Using NMR relaxation dispersion experiments, it was found that pKID binds in a disordered state and folds on the surface of KIX [93] corresponding to an induced fit mechanism.

In another study, the binding of the IDP PUMA to the folded protein MCL-1 (myeloid leukaemia cell differentiation-1) was studied using time-resolved stopped-flow kinetics. PUMA is disordered in its free form but forms a single, contiguous α-helix upon binding to MCL-1. Using the helix-breaking residue proline, several mutant forms of PUMA were investigated in terms of binding to MCL-1. All these mutants bound to MCL-1, and while the binding was weaker than the wild-type protein, association rate constants were largely unaffected [103]. The authors conclude that no individual turn of helix needed to be preformed in the unbound ensemble of PUMA for the binding to occur, and the obtained results are inconsistent with a pure conformational selection mechanism. The authors did not investigate the modulation of the conformational behaviour of the mutants relative to the wild type protein, however. They could not rule out some kind of a 'mixed' mechanism, where a shorter helical segment of PUMA is selected by MCL-1 before binding, followed by an induced fit of the rest of the chain to the final helical conformation.

In the case of the interaction between the disordered activation domain of the co-activator ACTR (co-activator for thyroid hormone and retinoid receptors) with the molten globule NCBD domain of CBP, the initial association is extremely rapid (which is

advantageous for cell signalling rapidity) but is followed by slow conformational transitions, presumed to be associated with the folding process [104]. Although this corresponds more to the "induced fit" mechanism, an increase in the population of a pre-formed ACTR helix led to a modest increase in the association rate [105]. In other IDPs the stabilization of a pre-formed helix by amino acid substitution (to insert helix-favouring residues) and by covalent modifications to 'staple' the peptide in helical conformations was found to have a rather small positive effect on binding affinity, or even to destabilize the complex [106]. In conclusion, the role of preformed structural elements is not yet clear and needs more extensive studies.

Conformational selection vs induced fit

As mentioned above, it is of significant interest to understand the sequence of binding and folding. Even if conformational selection requires that the IDP samples conformations resembling the bound state, the presence of folded or transiently folded structures in IDPs is not an indicator itself of a conformational selection mechanism. The observation that IDPs can adopt different conformations on binding to different targets [97], and, more generally, the discovery of "fuzzy complexes" [107], where the bound state of the IDP is thought to remain disordered and to sample a number of conformations, complicates further the picture and reinforces the importance of the question of the role of structured elements in IDPs.

One of the major features of the coupled folding and binding reaction is the decoupling of binding affinity and specificity. When two folded proteins interact, the specificity and affinity are often coupled [97]. In the case of IDPs, the entropic cost of the disorder-to-order transition may reduce the binding strength and make highly specific interactions reversible. It is to limit these unfavourable entropic effects that lead to the proposition of the "conformational selection" model. Detailed kinetic and thermodynamic analyses of IDP interactions show that both models – conformational selection and induced fit – are possible.

It even appears that both models could be merged into a single "synergetic" model, to which a third mechanism, that of the "fly-casting" interaction, could be added. In the fly-casting mechanism, the IDP is assumed to have an enhanced capture radius by interacting with its partner at a larger distance through its disordered chain [108], thus increasing the k_{on} rate constant. In the "synergetic" model [109], this fly-casting interaction is assumed to fulfil the role of non-specific "reeling" of the IDP by its interaction partner. Once the IDP is close enough, preformed structural elements play an important role and the specific encounter is facilitated by conformational selection. In the last step, the induced fit mechanism comes into play where the interacting region of the IDP finalizes the complex formation.

Until recently, most descriptions of interactions involving IDPs with their partners have measured global thermodynamic, kinetic or structural parameters, and extrapolated the results to invoke specific binding modes. It was demonstrated by Sugase et al. [93] that NMR spectroscopy can provide remarkable insight into the different steps involved in such interactions, by actively probing the interconversion between free, partially and fully bound forms of the protein, using NMR exchange techniques such as relaxation dispersion. This provides unique site-specific information about structure, dynamics and kinetics. Schneider et al. [110] recently combined $^{13}C'$, ^{15}N and $^{1}H_N$ relaxation dispersion experiments, performed at multiple molar ratio admixture, to reveal a probable trajectory for formation of the complex between the disordered domain of Sendai virus nucleoprotein (N_{TAIL}) (whose dynamic nature is the subject of this thesis) and the C-terminal domain of the phosphoprotein (PX) involving initial funnelling of the existing free-state conformational equilibrium [111] via stabilization of one of the interconverting helices on the surface of PX, prior to specific binding of one side of the N_{TAIL} helix in an inter-helical groove on the surface of PX (see also Chapter 5). It was recently noted that this conformational funnelling may provide a general molecular recognition mechanism adopted by IDPs [112]. The measurement of effective association-rates of the interaction partners is until now interpreted in the complete absence of consideration of the timescales of dynamics of the IDP in its free form, which is the focus of this thesis, and which would be essential to develop a complete movie of the different stages of the interaction.

Seen from an alternative perspective, two models – conformational selection and induced fit – are proposed as pathways and limiting cases of a 'flux model'. In this model [113], both mechanisms are simultaneously possible, and the exact flux of the reaction depends on intrinsic and extrinsic factors. For example, Hammes et al. have suggested that higher ligand concentration increases flux through the induced fit pathway, and vice versa [114]. Greives et al. suggested that intrachain dynamics of an IDP can influence the binding mechanism, with slower conformational transitions favouring conformational selection and fast interconversion rates favouring induced fit pathways [115] – again reinforcing the need to study IDP conformational dynamics in order to better understand their mechanisms of interaction.

Finally it has been evoked that not all IDPs fold upon binding. In such "random complexes" [23] the IDP would behave in its bound form similarly to the free form, and would interact with its partner by transient contacts. By extension of the concept of "random" complexes, in "fuzzy" complexes (term introduced by Tompa and Fuxreiter [107]) the protein remains partially or fully disordered in its bound state, or adopts unrelated conformations binding the same partner. This "fuzziness" could provide many functional advantages, including interactions with alternative partners and simultaneous

interactions with different partners. Different structural categories of "fuzzy" complexes and dynamic recognition mechanisms are reviewed in detail in [116].

1.5 Role of IDPs in human diseases

Studies of IDPs do not have only a pure academic interest. A link between intrinsic disorder and various diseases (cancer, diabetes, amyloidosis, neurodegenerative and cardiovascular diseases) is supported by numerous bioinformatical analyses and detailed studies of individual proteins. According to these studies, not only protein misfolding is related to diseases, but also misidentification, missignaling, unnatural or non-native folding, leading to a disorder-in-disorder (D^2) concept [117].

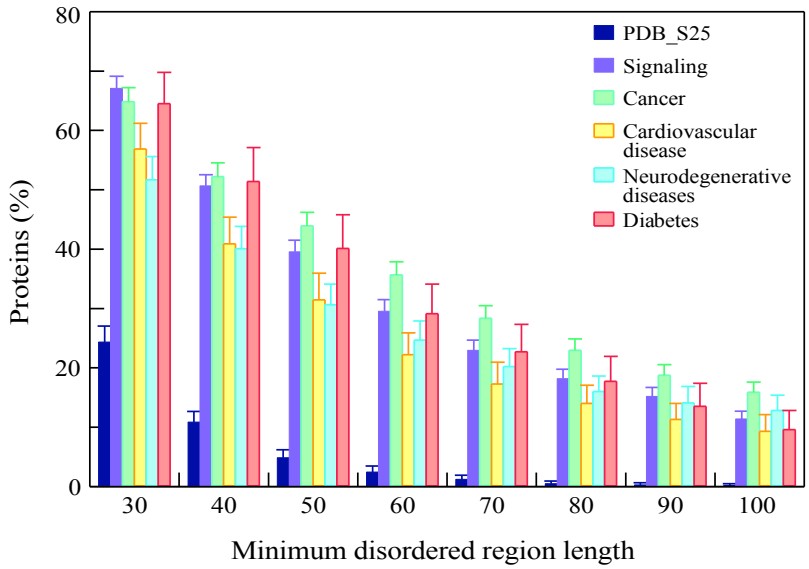


Fig. 1.6. Abundance of intrinsic disorder in disease-associated proteins: percentages of disease-associated proteins with ≥30 to ≥100 consecutive residues predicted to be disordered. Percentage of signalling and folded (i.e. PDB_S25) proteins with disordered segments of a given minimal length is shown for comparison. Taken from [117].

In Fig. 1.6, the percentage of segments predicted to be continuously disordered of a given minimal length is plotted, based on the analysis of sets of disease-related proteins including 1786, 487, 689 and 285 proteins for cancer, cardiovascular, neurodegenerative disease and diabetes, respectively [117].

A remarkable example of an IDP linked to signalling and cancer is p53, which has disordered N- and C-terminal domains. It plays a central role in regulating the expression of genes involved in the cell-cycle, in apoptosis induction, DNA repair and cellular response to stress [118].

Many human diseases, known as conformational diseases, or protein misfolding diseases, involve proteins that fail to adopt or keep their normal conformational state, lose their normal functions, gain toxic functions, aggregate and/or form fibrils. The prion protein plays a key role in transmissible spongiform encephalopathies (TSE); Aβ and tau proteins are at the centre of Alzheimer's pathogenesis, while α-synuclein and tau are central to

synucleinopathies including Parkinson's disease [119]. Recently the Alzheimer disease was found to be transmissible in specific cases [120], as if a link existed between it and prion-related diseases.

IDPs as drug targets

One of the strategies to hinder pathology-related protein interactions involving IDPs is to block the binding site of the folded binding partner. The binding partner often accommodates in its binding pocket a relatively short fragment of an IDP directly involved in the interaction. By rational drug-design approaches, if the 3D structure of the folded protein is known, or by random screening of compound libraries, it could be possible to find a small molecule that fits to the binding pocket instead of the IDP. One example is the inhibition of the interaction of p53 with Mdm2 [121]. p53 interacts with its partner through a short stretch of residues (residues 13-29) in its disordered N-terminal domain. NMR studies have shown that, although the N-terminal domain has no stable structure in solution, residues 18-25 transiently adopt a helical conformation (α-MoRE) [122]. In complex with Mdm2, this region forms an amphipatic α-helix binding to a deep groove on the Mdm2 surface. Several peptides and small molecules were created to mimic this α-MoRE motif of p53 an inhibit its binding to Mdm2 [123].

The alternative approach consists in targeting the IDPs themselves. Several small-molecule inhibitors were found to act against disordered targets, such as oncoproteins c-Myc [124], EWS-FLI1 [125] or Alzheimer's-related Aβ peptide [126, 127]. Tremendous efforts have been made to develop anti-aggregation drugs to block amyloid aggregation in neurogenerative diseases and to capture toxic misfolded forms. A small molecule phthalocyanines tetrasulfonated (PcTS) has been found to inhibit α-synuclein filament assembly and to promote the formation of nontoxic protein aggregates [128]. One of the possible design methods of anti-aggregation drugs is to modify the self-recognition interface of aggregating proteins in a way that they still can interact with native versions, further assembly into larger aggregates is inhibited. This method was successfully exploited to block diabetes-related islet amyloid polypeptide (IAPP) aggregation [129], and, interestingly, the modified version of IAPP was also found to effectively block Aβ assembly [130], underlining the possible link between different aggregation-related human pathologies.

Chapter 2. The context and the objectives of this work

In this chapter, I provide an introduction to the protein under investigation, namely the disordered C-terminal domain of the nucleoprotein of Sendai paramyxovirus (N_{TAIL}). I discuss the role that this domain plays in viral transcription and replication as well as the interaction mechanism between N_{TAIL} and its partner – the PX domain of the phosphoprotein. The conformational sampling of N_{TAIL} was previously described, and the interaction site of N_{TAIL} – the molecular recognition element (MoRE), has been shown to populate transient α-helical conformations. However, not much is known about the dynamics of N_{TAIL}, which plays an important role in this interaction, and that is the aim of this study.

Intrinsic disorder in proteins was shown to be highly abundant in eukaryotes and to be related to human diseases such as Alzheimer disease, Parkinson disease and numerous cancers (Chapter 1, section 2.4). High levels of intrinsic disorder has also been found in viruses [131–134]. In particular, proteins in RNA viruses display a high occurrence of disordered segments [135].

Sendai virus (SeV) is a member of the *paramyxoviridae* family, which includes non-segmented negative strand RNA viruses such as measles virus, mumps virus, and the emerging human pathogens Hendra and Nipah viruses. SeV causes bronchiolitis in mice and primates. In *paramyxoviridae*, the viral genome is packaged into a large helical assembly – the nucleocapsid – formed by multiple copies of the nucleoprotein (N) [136]. Transcription and replication are performed on this RNA-nucleoprotein complex by the large protein (L), which carries the polymerase activity, and its co-factor phosphoprotein (P) [137]. Little is known about the detailed molecular mechanisms of transcription and replication of these viruses. Both processes are believed to be initiated by the interaction between the nucleoprotein N (in form of nucleocapsids) and the phosphoprotein P associated with the polymerase L [138].

The protein N has two domains (Fig. 2.1), a core domain N_{CORE} of approximately 400 residues, and a C-terminal 'tail' domain (N_{TAIL} or NT) of about 125 residues, predicted to be almost entirely intrinsically disordered, with enhanced helical propensity in the C-terminal region that mediates the interaction with P [136]. Recently, a combination of solution NMR spectroscopy, small angle scattering and electron microscopy demonstrated that the first 50 amino acids of N_{TAIL} have reduced mobility in the nucleocapsids as N_{TAIL} exfiltrates from the inside to the outside via the interstitial space between successive turns of the helical nucleocapsid. The remaining part of N_{TAIL} remains dynamic and displays the same conformational behaviour as observed in the isolated form of N_{TAIL} in solution [62].

The phosphoprotein P contains a 300-470 residue long N-terminal disordered domain, followed by an oligomerization domain of 80-110 residues, a shorter unfolded domain of about 100 residues and finally a small (~50 residues) C-terminal domain known as PX that represents the N_{TAIL} binding site [136]. This PX domain was characterized by NMR (in the case of SeV) [139] and by X-ray crystallography for Measles [140, 141] and Hendra [142] viruses and adopts a structurally conserved three-helix bundle.

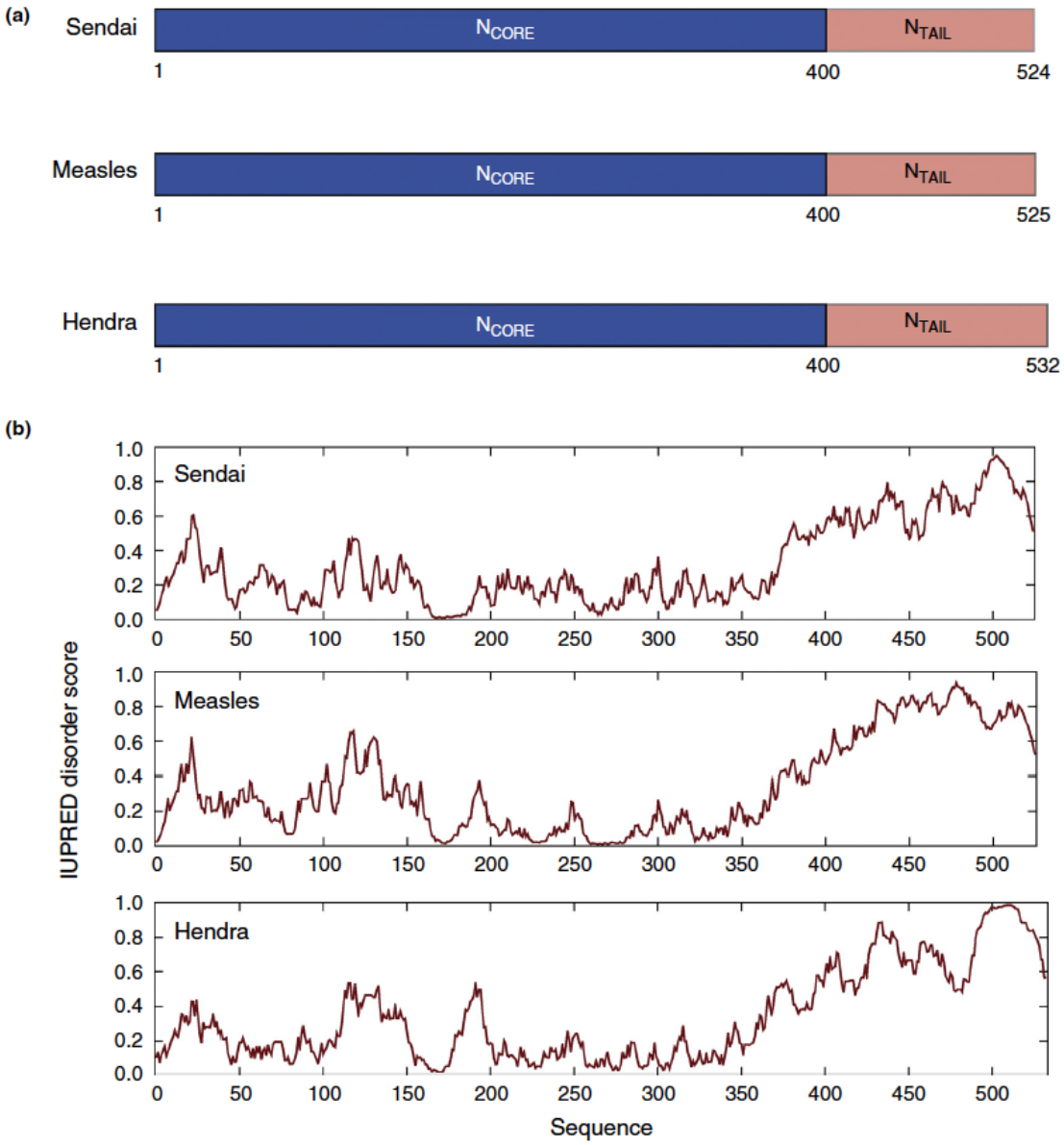


Fig. 2.1. a) Domains of nucleprotein predicted to be folded (blue) and unfolded (red) in three different viruses from the *paramyxoviridae* family. b) Disorder prediction in these nucleoproteins using IUPred [31].

The molecular recognition element (MoRE) of SeV N_{TAIL} was shown, on the basis of multiple types of residual dipolar couplings, to form three discrete transient helices of different populations in fast exchange (on the chemical shift timescale) in solution [111] (Fig. 2.2). All three identified helical segments are preceded by aspartic acids or serines, suggesting that the helical segments are stabilized by N-capping interactions involving

hydrogen bonds between sidechains and backbone amide moieties at position two or three in the helical elements. The interaction between the MoRE of N_{TAIL} and the PX domain of the phosphoprotein in SeV was recently characterized in detail by relaxation dispersion experiments [110].

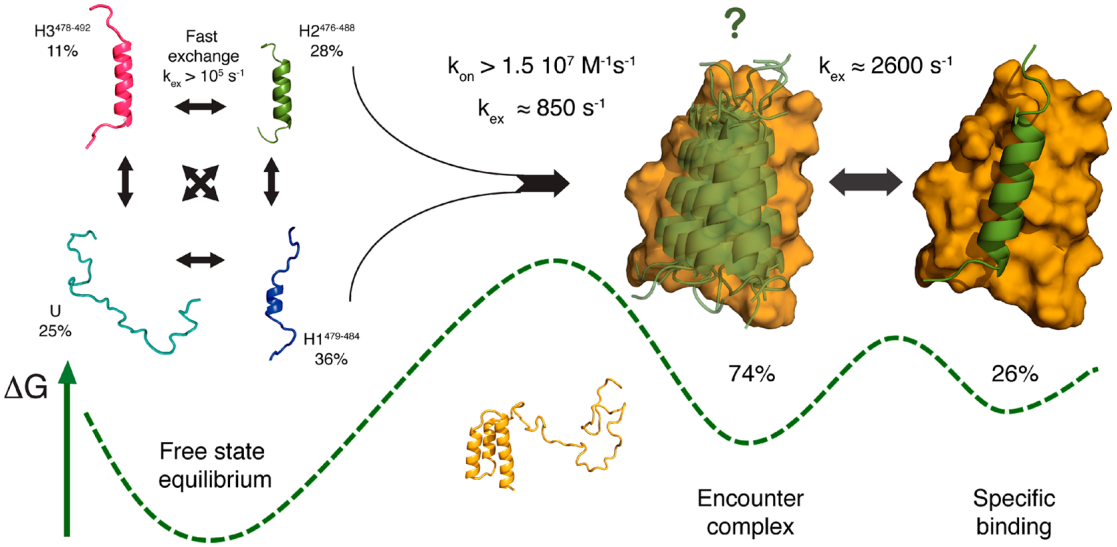


Fig. 2.2. Proposed mechanism of N_{TAIL} to PX binding in Sendai virus. H1, H2, H3: helical conformers predicted to be in the MoRE of N_{TAIL} in the free state in solution. U: unfolded population. Question mark on top of the encounter complex shows that the relative position of helical element with respect to PX is unknown. In the bottom part: a sketch of the N_{TAIL} free energy landscape.

A two-step model is proposed. Initially, N_{TAIL} binds to PX and they form a nonspecific and dynamic encounter complex, in which one of helical conformers present in the free state of N_{TAIL} (H2 on Fig. 2.2) is stabilized. The rate of this step is 850 s^{-1} as measured by relaxation dispersion NMR. The association rate constant k_{on} of this step is faster than the diffusion limit, suggesting that the encounter complex is electrostatically driven, which is supported by mutation studies [110]. Following the formation of the encounter complex, the stabilized helix in N_{TAIL} is engaged in the helical groove on the surface of PX in a second step at a rate of 2600 s^{-1}.

It is unclear whether N_{TAIL} with its MoRE in a helical state H2 is 'preselected' by PX and then stabilized upon formation of the encounter complex (conformational selection mechanism), or whether N_{TAIL} in any conformational state could form the encounter complex with PX and the stabilization of H2 is realized through an 'induced fit' from the other states (H1, H3, unfolded). However, the general pathway consists of 'funneling' of the initial conformational equilibrium in the MoRE of N_{TAIL} into a state resembling H2 in the encounter complex, in which the PX - N_{TAIL} interaction remains unspecific until the second binding event locks the MoRE into the binding groove on PX.

Despite the fact that the conformational sampling of N_{TAIL} has been characterized, still not much is known about dynamic timescales of the motion of the N_{TAIL} chain, which surely

play a role in the interaction kinetics, in particular in terms of the on-rate of the interaction. But our interest in the dynamical behavior or N_{TAIL} extends well beyond the context of its interaction with PX, because this 124 amino acid protein also provides a good model system, containing long unfolded domains with chain-like dynamics and regions with residual structure.

In this thesis, we study the dynamics of N_{TAIL} in the free state by NMR relaxation, and our goal is to answer the following questions:

- Can we describe the timescales of motions governing the dynamical behavior of N_{TAIL}?
- Can we assign described timescales to particular physical processes in a context of an intrinsically disordered protein?
- How these timescales depend on the temperature and chain length?
- How the local dynamics of a residue in the disordered chain is influenced by the behavior of its neighbours?
- Can we estimate the timescale of the interconversion of helical conformers in the MoRE region of N_{TAIL}?

In the next two chapters, we present the part of the NMR relaxation theory, which is relevant to this study, and we discuss previous studies where these – or similar - questions were addressed.

Chapter 3. NMR relaxation

NMR relaxation is intimately linked to molecular motions, and thus represents an important tool to study them. In this chapter, the theoretical basics of NMR relaxation, necessary to understand the work done in this thesis, are presented. It starts with a brief explanation of basic quantities and expressions used to describe NMR phenomena, continues with the semiclassical theory of spin relaxation, introducing different relaxation mechanisms of interest, and finally equations relating relaxation rates and dynamics in a molecule on picosecond-to-nanosecond timescale are derived and discussed. In the end of the chapter, the contribution to relaxation rates of the slower microsecond-to-millisecond motions, referred to as 'conformational exchange', is introduced, and typical relaxation rates encountered in folded and intrinsically disordered proteins are compared.

3.1 NMR basics

Elementary particles that compose atoms have mass and charge. They also have angular momentum. Two kinds of angular momentum are associated with particles: orbital momentum and spin. Orbital angular momentum arises when a particle follows a rotating or twisting trajectory. In some cases, though, as for example in the Stern-Gerlach experiment, it became evident that particles also possess an angular momentum that cannot be accounted for by orbital momentum: this momentum was called spin.

NMR spectroscopy is intrinsically linked to spin angular momentum of nuclei and its theory is essentially the quantum mechanics of nuclear spins [38]. The spin of a nucleus is characterized by the spin quantum number I, which is always integer or half-integer. Nuclei with odd mass numbers have half-integer spin numbers. Nuclei with even mass number and even atomic number have zero spin and are NMR inactive. Nuclei with even mass number and odd atomic number have integer spin numbers. As nuclei with spin more than ½ have non-spherical nuclear charge distribution, they also possess electric quadrupole moment, which makes the lifetime of their magnetic states in solution much shorter than those of ½ spin nuclei; consequently, their NMR resonance lines are broad and more difficult to study [38]. For the purposes of biomolecular NMR, the most important ½ spin nuclei ($I = 1/2$) are ^{1}H, ^{13}C, ^{15}N, ^{19}F and ^{31}P; the most important nucleus with $I = 1$ is ^{2}H (deuteron).

The spin angular momentum, $\hbar\mathbf{I}$, can be characterized using two operators:

$$\hat{\mathbf{I}}^2 = \hat{\mathbf{I}} \cdot \hat{\mathbf{I}} = I(I+1) \tag{3.1}$$

– the square of its norm, and

$$\hat{I}_Z = m \tag{3.2}$$

– its projection on one of the axes of Cartesian coordinate system. In quantum mechanics, observables are incompatible and cannot be measured together when their operators do not commute, and this is a case of $\hat{I}_X$, $\hat{I}_Y$ and $\hat{I}_Z$. By convention, the one measurable component of $\hat{\boldsymbol{I}}$ is its projection on axis *z*, which is defined as being parallel to the external static magnetic field. In the equation 3.2 *m* is the magnetic quantum number and has 2*I*+1 possible values: (–*I*, –*I*+1, …, *I*–1, *I*).

The nuclear magnetic moment $\hat{\boldsymbol{\mu}}$ is defined by:

$$\hat{\boldsymbol{\mu}} = \gamma\hbar\hat{\mathbf{I}} \tag{3.3}$$

where γ is the gyromagnetic ratio, a characteristic constant for a given nucleus that determines its receptivity. In the presence of a static magnetic field, the Hamiltonian of the interaction between spin and field is given by:

$$\mathcal{H} = -\hat{\boldsymbol{\mu}} \cdot \mathbf{B_0} = -\gamma\hbar\hat{\mathbf{I}} \cdot \mathbf{B_0} \tag{3.4}$$

When I = 1/2, two orientations of the spin exist (m=–1/2 and m=1/2) and two orientations of the magnetic moment are possible: parallel (α) and antiparallel to the field (β), with energies:

$$E_\alpha = -\frac{1}{2}\gamma\hbar B_0 \text{ and } E_\beta = \frac{1}{2}\gamma\hbar B_0. \tag{3.5}$$

As $\hat{I}_Z$ is always less than $\left|\hat{\mathbf{I}}\right| = \sqrt{\hat{\mathbf{I}}^2}$, magnetic moments cannot be parallel to the magnetic field, so they orient and organize themselves to form two cones around $\mathbf{B_0}$ direction, corresponding to α and β states, so that their projection follows the rule (3.2) and is equal to $\pm\frac{1}{2}\hbar$.

Magnetic moment is not static; all equivalent spins in the system precess around $\mathbf{B_0}$ with the same frequency $\boldsymbol{\omega_0}$, called Larmor frequency.

$$\boldsymbol{\omega_0} = (E_\beta - E_\alpha)/\hbar = -\gamma B_0 \tag{3.6}$$

Transitions between states α and β are observable at the Larmor frequency, and can also be stimulated by applied electromagnetic radiation with the same frequency.

At equilibrium, different energy states are unequally populated, because spin orientations with lower energy are more probable. The populations of α and β states are given by Boltzmann statistics:

$$\frac{N_\beta}{N_\alpha} = e^{-\frac{E_\beta - E_\alpha}{kT}} = e^{-\frac{\gamma\hbar B_0}{kT}} \approx 1 - \gamma\hbar\frac{B_0}{kT} \tag{3.7}$$

It is neither interesting nor possible to study an isolated spin. In any sample studied by NMR, we have to deal with a large number of identical spins, and only their macroscopic (bulk) parameters are measurable. The bulk magnetic moment is given by the vector sum of nuclear magnetic moments for each nucleus. At thermal equilibrium, the transverse components of magnetic moment of different nuclei are uncorrelated and sum to zero; the difference between the population of energy states gives rise to a bulk magnetization parallel (longitudinal) to the external static field $\mathbf{B_0}$.

However, the values of the gyromagnetic ratios of the nuclei are too small to create a sizeable difference in the energy level populations; $(N_\alpha - N_\beta)/N_\alpha \approx \gamma\hbar B_0/kT$ (see [3.7]), which gives around 10^{-5} for ^{1}H spins at the field 11.7T at room temperature. This explains why NMR is considered to be an insensitive spectroscopic technique, and why it is important to construct more and more powerful NMR magnets.

By applying external electromagnetic radiation to the sample, the bulk magnetization of spins can be rotated, e.g. to the direction opposite to $\mathbf{B_0}$, or perpendicular to it. In the latter case, transverse components of bulk magnetic moment are not zero, and it rotates around $\mathbf{B_0}$ with the corresponding Larmor frequency. However, over time, transverse components decorrelate back to zero, and the bulk magnetization vector regains its equilibrium value and orientation parallel to $\mathbf{B_0}$ field. This phenomenon is known as **spin relaxation**.

Pure state, mixed state, observables and density matrix

The understanding of the mechanisms of relaxation requires the introduction of some notions of quantum mechanics.

The Dirac notion is a compact formalism to represent the scalar product. For two wavefuctions, their scalar product will take form:

$$\langle\psi|\varphi\rangle = \int \psi^*\varphi d\tau \tag{3.8}$$

where $|\varphi\rangle$ represents the wavefunction in Dirac notion and $\langle\psi|$ represents ψ^*, the conjugate wavefunction for ψ. The decomposition of a wavefunction in its eigenfunctions, "basic kets", is written as

$$|\Psi\rangle = \sum_{n=1}^{N} c_n|n\rangle \tag{3.9}$$

A 'pure' quantum state is described by only one wavefunction. In such a state, the expectation value of some property is written as:

$$\langle A\rangle = \int \Psi^* \mathbf{A} \Psi d\tau = \langle \Psi|\mathbf{A}|\Psi\rangle = \sum\nolimits_{mn} c_m^* c_n \langle m|\mathbf{A}|n\rangle \tag{3.10}$$

The wavefunction of a macromolecule in an NMR sample is an extremely complicated function of all degrees of freedom of the molecule. It is neither possible nor necessary to determine the complete wavefunction of the molecule. Rather, the system is divided into two components: the spin system and the surroundings (all the other degrees of freedom), called *lattice* for historical reasons. Consequently, the spin wavefunctions for different molecules in the sample are no longer identical, because they depend upon 'hidden' lattice variables [38].

Such a system is called a 'mixed state', where the effects of the lattice are incorporated using statistical mechanics [38]. A 'mixed' state is represented as an ensemble of independent systems (or subensembles), each of which is a pure state with wavefuction Ψ. A probability density $P(\Psi)$ describes the contribution of each system (subensemble) to the mixed state. The expectation value for this mixed state is obtained by averaging over the probability distribution:

$$\begin{aligned}\langle \bar{A}\rangle = \int P(\Psi)\langle \Psi|\mathbf{A}|\Psi\rangle d\tau &= \sum\nolimits_{mn} \int P(\Psi) c_m^* c_n d\tau \, \langle m|\mathbf{A}|n\rangle \\ &= \sum\nolimits_{mn} \overline{c_m^* c_n} \langle m|\mathbf{A}|n\rangle \end{aligned} \tag{3.11}$$

In this equation, the ensemble average $\overline{c_m^* c_n}$ represent a matrix that is called *density matrix*. The corresponding operator is called *density operator* and is written as:

$$\sigma = \int P(\Psi)|\Psi\rangle\langle\Psi| d\tau \tag{3.12}$$

The expected average value of the A operator can then be shown to be given by the following equation:

$$\langle \bar{A}\rangle = \mathrm{Tr}(\sigma \mathbf{A}) \tag{3.13}$$

The time evolution of the density operator follows the Liouville-von-Neumann equation:

$$\frac{\mathrm{d}\sigma(t)}{\mathrm{d}t} = -i[\mathcal{H}, \sigma(t)] \tag{3.14}$$

Using (3.13) and (3.14), and having the Hamiltonian $\mathcal{H}$ of the system, we can describe the time evolution of any observable A (such as transversal or longitudinal magnetization, for example).

3.2. Origins of relaxation

It can be shown that neither spontaneous nor stimulated emission are effective relaxation mechanisms, in contrast to optical and UV spectroscopy, where they are important because the relevant photon frequencies are orders of magnitude bigger. Nuclear spin relaxation is rather a consequence of spin system being coupled to the aforementioned *lattice* (surroundings), which contains other degrees of freedom of the molecule as well as other molecules in the sample.

The lattice is assumed to have quasi-continuous energy levels with populations following the Boltzmann distribution, and to have an infinite heat capacity; thus, the lattice always stays in thermal equilibrium [38]. The spin system is weakly coupled to the lattice by local magnetic fields, which become time-dependent as the molecule participates in Brownian motion. Thus, the spin system is under the action of a time-dependent Hamiltonian that can be decomposed as follows:

$$\begin{aligned}\mathcal{H} &= \mathcal{H}_Z + \mathcal{H}_{local} = \mathcal{H}_Z + \mathcal{H}_{local}^{isotropic} + \mathcal{H}_{local}^{anisotropic}(t) \\ &= \mathcal{H}_Z + \mathcal{H}_{local}^{isotropic} + \mathcal{H}_{longitudinal}^{anisotropic}(t) + \mathcal{H}_{transverse}^{anisotropic}(t),\end{aligned} \tag{3.15}$$

where the local magnetic field is decomposed into an isotropic, i.e. rotationally invariant (and hence time-independent), and an anisotropic (time-dependent) components, with the time-dependent component being itself decomposed into parts parallel (longitudinal) and perpendicular (transverse) to the main static field. The time-dependent magnetic fields (and corresponding Hamiltonians) are stochastic; they average to zero in isotropic solutions, and correlations between them tend to zero for times $t \gg \tau_C$, where is τ_C is the correlation time of the stochastic process.

Relaxation processes by their nature are divided in two groups: *nonadiabatic*, when they require exchange of energy with the lattice, and *adiabatic* in the other case. Transverse components of the fluctuating local field may cause transitions between different energy levels of the spin system, if in their Fourier spectrum they have components at the corresponding frequency. As the spin system is coupled to the lattice and the total energy should be conserved, each transition of the spin system from the higher energy state to the lower state entails the lattice transition from the lower state to the higher one. Correspondingly, the opposite process is also possible (lattice transition to lower state and that of spin system to higher one). However, we make the approximation that the lattice always stays at thermal equilibrium, and the exchange of energy between the lattice and the spin system will inevitably bring the spin system back to thermal equilibrium, in the absence of external perturbations. Moreover, transitions between stationary states of the spin system reduce their lifetime and, as a consequence of the Heisenberg uncertainty relationship, broaden their energies, which leads to the loss of

phase coherence between different spins. In terms of the density matrix, its diagonal elements come back to their equilibrium values, and non-diagonal elements decay to zero.

Adiabatic processes are linked to the part of the fluctuating field that is parallel to z axis, the static field ($\mathbf{B_0}$) direction. As the total magnetic field in the z direction fluctuates in a random way, so do the energy levels of the spin system, and this leads to loss of spin phase coherences, as in the case of nonadiabatic processes. However, adiabatic processes do not contribute to the re-equilibration of energy state populations.

From the phenomenological point of view, one can distinguish:

- the *spin-lattice*, or *longitudinal* relaxation, which describes the return of energy level populations of the spin system to the equilibrium values, and consequently the recovery of the longitudinal magnetization (parallel to $\mathbf{B_0}$). This relaxation process is described by the rate constant R_1 and corresponding time constant $T_1 = 1/R_1$. Only nonadiabatic processes contribute to this relaxation.
- the *spin-spin*, or *transversal* relaxation, which describes the decay of transversal magnetization, linked to the decay of non-diagonal single-quantum coherence elements of the density matrix. This relaxation process is described by the rate constant R_2 and corresponding time constant $T_2 = 1/R_2$. Both adiabatic and non-adiabatic processes contribute to the transversal relaxation.

3.3 Bloch-Wangsness-Redfield (BWR) theory of relaxation

The semiclassical theory of relaxation, in which the spin system is described by quantum mechanics and the lattice is treated classically, was formulated by Bloch, Wangsness and Redfield. The main drawback of this theory is that in its description, the energy levels of the system are equally populated in equilibrium, which is formally correct only for an infinite spin temperature (high-temperature limit). For finite temperatures, corrections are required to ensure the Boltzmann distribution of populations at thermal equilibrium; however, at room temperature these corrections appear to be insignificant. The description of BWR theory in this thesis generally follows [38] with some aspect taken from [143].

3.3.1 The Master Equation

Resuming (3.15), in the absence of applied RF fields, we can decompose the Hamiltonian of the system in its static part $\mathcal{H}_0$ that acts only on the spin system, and a stochastic part $\mathcal{H}_1(t)$, which couples it to the lattice.

$$\mathcal{H}(t) = \mathcal{H}_0 + \mathcal{H}_1(t) \tag{3.16}$$

In terms of the perturbation theory, $\mathcal{H}_0$ is 'main' Hamiltonian, primarily due to external field $\mathbf{B_0}$, and $\mathcal{H}_1(t)$ is a small time-dependent 'perturbation', due to fluctuating local stochastic fields. By construction, $\overline{\mathcal{H}_1(t)} = 0$, and all local static contributions, i.e. those whose time average is not zero, are included in $\mathcal{H}_0$.

To study the evolution of the system under this Hamiltonian, we use the Liouville-von Neumann equation for the density operator (3.14):

$$\frac{\mathrm{d}\sigma(t)}{\mathrm{d}t} = -i[\mathcal{H}_0 + \mathcal{H}_1(t), \sigma(t)] = -i\{\hat{L}_0 + \hat{L}_1(t)\}\sigma(t), \tag{3.17}$$

where $\hat{L}(t) = [\mathcal{H}(t),]$ is the commutation Liouvillian superoperator. In order to get eliminate $\mathcal{H}_0$ from our calculations, we transform this equation to the so-called *interaction frame*, i.e. a rotating frame which rotates around the z-axis with the Larmor frequency corresponding to $\mathcal{H}_0$. This is achieved by variable replacement:

$$\sigma^T(t) = \exp\left(i\hat{L}_0 t\right)\sigma(t) = \exp(i\mathcal{H}_0 t)\,\sigma(t)\exp(-i\mathcal{H}_0 t) \tag{3.18}$$

In this frame the Liouville-von Neumann equation reduces to

$$\frac{\mathrm{d}\sigma^T(t)}{\mathrm{d}t} = -i\hat{L}_1^T(t)\sigma^T(t) = -i[\mathcal{H}_1(t), \sigma(t)], \tag{3.19}$$

where $\hat{L}_1^T(t) = [\mathcal{H}_1^T(t),]$ and

$$\mathcal{H}_1^T(t) = \hat{L}_0\mathcal{H}_1(t) = \exp(i\mathcal{H}_0 t)\,\mathcal{H}_1(t)\exp(-i\mathcal{H}_0 t) \tag{3.20}$$

The integral of (3.19) gives:

$$\sigma^T(t) = \sigma^T(0) - i\int_0^t [\mathcal{H}_1^T(t_1), \sigma^T(t_1)]\,dt_1 \tag{3.21}$$

We use the Picard-iteration variant of time-dependent perturbation theory to find the approximate solution of (3.21), following the approach described in [143]. The approximation of zero order would be $\sigma^T(t) \approx \sigma^T(0)$, and in the next step we replace the $\sigma^T(t_1)$ in the integral in (3.21):

$$\sigma^T(t) \approx \sigma^T(0) - i\int_0^t [\mathcal{H}_1^T(t_1), \sigma^T(0)]\, dt_1, \tag{3.22}$$

and for the second order approximation:

$$\sigma^T(t) \approx \sigma^T(0) - i\int_0^t [\mathcal{H}_1^T(t_2), \sigma^T(0)]\, dt_2 + i^2 \int_0^t dt_2 \int_0^{t_2} dt_1 \left[\mathcal{H}_1^T(t_2), [\mathcal{H}_1^T(t_1), \sigma^T(0)]\right] \tag{3.23}$$

Higher-order approximations turn out to be unnecessary, as the final result of the second order approximation coincides with the formalism not based on the perturbation theory (see [143]). The time derivative of (3.23) with the subsequent substitution $t_1 = t - \tau$ gives

$$\frac{\mathrm{d}\sigma^T(t)}{\mathrm{d}t} = -i[\mathcal{H}_1^T(t), \sigma^T(0)] - \int_0^t \left[\mathcal{H}_1^T(t), [\mathcal{H}_1^T(t-\tau), \sigma^T(0)]\right] d\tau \tag{3.24}$$

The further treatment of (3.24) requires the introduction of three assumptions in order to apply the ensemble average over all spins in the system. The first assumption is that the timescale studied satisfies $t \gg \tau_c$, where τ_c is the correlation time of time-dependent fluctuating interactions described by $\mathcal{H}_1^T(t)$. At this timescale $\mathcal{H}_1^T(t)$ and $\sigma^T(0)$ are uncorrelated and the ensemble average of the first term in equation 3.24? is zero, given $\overline{\mathcal{H}_1^T(t)} = 0$:

$$\overline{[\mathcal{H}_1^T(t), \sigma^T(0)]} = \left[\overline{\mathcal{H}_1^T(t)}, \overline{\sigma^T(0)}\right] = 0 \tag{3.25}$$

For the same reason $\overline{\left[\mathcal{H}_1^T(t), [\mathcal{H}_1^T(t-\tau), \sigma^T(0)]\right]} = 0$ for $t \gg \tau_c$, and the upper limit in the integral in (3.24) can be replaced with ∞. Thus the ensemble average of (3.24) gives:

$$\frac{\mathrm{d}\overline{\sigma^T(t)}}{\mathrm{d}t} = -\int_0^\infty \overline{\left[\mathcal{H}_1^T(t), [\mathcal{H}_1^T(t-\tau), \sigma^T(0)]\right]}\, d\tau \tag{3.26}$$

The second assumption was already made when we made the approximation of the zeroth order $\sigma^T(t) \approx \sigma^T(0)$: it holds only in vicinity of time $t = 0$, more precisely $t \ll T_1, T_2$, so that the density operator is still close to its initial value. The equation (3.26) can be relevant for any starting time by replacing back $\sigma^T(0)$ by $\sigma^T(t)$, however, the *range* of acceptable accuracy of (3.26) respects the condition

$$\tau_c \ll t \ll T_1, T_2 \tag{3.27}$$

This transposition in time is possible because the effect of fluctuating Hamiltonians in the integrand in (3.26) is independent on the absolute time; it is rather their correlation after an interval τ that matters.

Finally, the third assumption we make to ensure that the spin system relaxes toward thermal equilibrium is to replace $\sigma^T(t)$ with $\sigma^T(t) - \sigma_0$, where σ_0 is the equilibrium density operator.

Omitting the ensemble-averaging bar over the density operator, we can finally write:

$$\frac{\mathrm{d}\sigma^T(t)}{\mathrm{d}t} = -\int_0^\infty \overline{\left[\mathcal{H}_1^T(t), \left[\mathcal{H}_1^T(t-\tau), \sigma^T(t) - \sigma_0\right]\right]}\, d\tau \tag{3.28}$$

Now we need to transform equation (3.28) back to the laboratory frame. In order to do this, the stochastic Hamiltonian has to be decomposed as the sum of random function of spatial variables $F_k^q(t)$ and tensor spin operators $\mathbf{A}_k^q$, where k is the tensor rank:

$$\mathcal{H}_1(t) = \sum_{q=-k}^{k} (-1)^q F_k^{-q}(t) \mathbf{A}_k^q \tag{3.29}$$

The tensor spin operators $\mathbf{A}_k^q$ are chosen to be spherical tensor operators because of their convenient properties under rotational transformations. They are expanded in terms of their basis operators as followed:

$$\mathbf{A}_k^q = \sum_p \mathbf{A}_{kp}^q\,, \tag{3.30}$$

where $\mathbf{A}_{kp}^q$ are the eigenfunctions of the Hamiltonian commutation superoperator. They have following eigenfrequencies:

$$\hat{L}_0\{\mathbf{A}_{kp}^q\} = \left[\mathcal{H}_0, \mathbf{A}_{kp}^q\right] = \omega_p^q \mathbf{A}_{kp}^q \tag{3.31}$$

Equation (3.31) implies that:

$$\exp\left(i\hat{L}_0 t\right) \mathbf{A}_{kp}^q = \exp(i\mathcal{H}_0 t)\, \mathbf{A}_{kp}^q \exp(-i\mathcal{H}_0 t) = \exp\left(i\omega_p^q t\right) \mathbf{A}_{kp}^q, \tag{3.32}$$

and, if applied in the interaction frame, gives:

$$\mathbf{A}_k^{qT} = \exp(i\mathcal{H}_0 t)\,\mathbf{A}_k^q \exp(-i\mathcal{H}_0 t) = \sum_p \exp\left(i\omega_p^q t\right)\mathbf{A}_{kp}^q$$
$$\mathbf{A}_k^{-qT} = \sum_p \exp\left(-i\omega_p^q t\right)\mathbf{A}_{kp}^{-q} \tag{3.33}$$

where $\omega_p^{-q} = -\omega_p^q$. Substituting (3.33) and (3.29) in (3.28) yields

$$\frac{\mathrm{d}\sigma^T(t)}{\mathrm{d}t} = -\sum_{q,q'}\sum_{p,p'}(-1)^{q+q'}\exp\left\{i\left(\omega_p^q + \omega_{p'}^{q'}\right)t\right\}\left[\mathbf{A}_{kp'}^{q'},\left[\mathbf{A}_{kp}^q, \sigma^T(t) - \sigma_0\right]\right] \times \int_0^\infty \overline{F_k^{-q'}(t)F_k^{-q}(t-\tau)\exp\left(-\imath\omega_p^q\tau\right)}\,d\tau \tag{3.34}$$

Assuming that our sample is isotropic, there is no correlation among the polar coordinates r, φ, ϑ within the ensemble of spin systems, and, therefore, stochastic processes $F_k^q(t)$ and $F_k^{q'}(t)$ are uncorrelated unless $q' = -q$. Under this condition, the ensemble average in the integrand vanishes if $q' \neq -q$, thus

$$\frac{\mathrm{d}\sigma^T(t)}{\mathrm{d}t} = -\sum_{q=-k}^{k}\sum_{p,p'}\exp\left\{i\left(\omega_p^q - \omega_{p'}^q\right)t\right\}\left[\mathbf{A}_{kp'}^{-q},\left[\mathbf{A}_{kp}^q, \sigma^T(t) - \sigma_0\right]\right] \times \int_0^\infty \overline{F_k^q(t)F_k^{-q}(t-\tau)\exp\left(-\imath\omega_p^q\tau\right)}\,d\tau \tag{3.35}$$

Another simplification that is commonly used is to remove terms in (3.35) where we have $\left|\omega_p^q + \omega_{p'}^{-q}\right| = \left|\omega_p^q - \omega_{p'}^q\right| \gg 0$. These terms are so-called *non-secular* terms, which oscillate too rapidly on the timescale of relaxation and cancel on average. Furthermore, if no eigenfrequencies are degenerate, terms in (3.35) are secular and nonzero only if $p = p'$. Therefore,

$$\frac{\mathrm{d}\sigma^T(t)}{\mathrm{d}t} = -\sum_{q=-k}^{k}\sum_{p}\left[\mathbf{A}_{kp}^{-q},\left[\mathbf{A}_{kp}^q, \sigma^T(t) - \sigma_0\right]\right] \times \int_0^\infty \overline{F_k^q(t)F_k^{-q}(t-\tau)}\exp\left(-i\omega_p^q\tau\right)d\tau \tag{3.36}$$

Terms $\overline{F_k^q(t)F_k^{-q}(t-\tau)}$ are known as correlation functions; the real part of the integral in (3.36) is known as the power spectral density function,

$$j^q(\omega) = 2\mathrm{Re}\left\{\int_0^\infty \overline{F_k^q(t)F_k^{-q}(t-\tau)}\exp(-i\omega\tau)\,d\tau\right\}$$
$$= \mathrm{Re}\left\{\int_{-\infty}^\infty \overline{F_k^q(t)F_k^{-q}(t-\tau)}\exp(-i\omega\tau)\,d\tau\right\}$$
$$= \mathrm{Re}\left\{\int_{-\infty}^\infty \overline{F_k^q(t)F_k^{-q}(t+\tau)}\exp(-i\omega\tau)\,d\tau\right\} \quad (3.37)$$

The power spectral density function is real, even-valued function of τ and ω. The imaginary part of the integral, on the other hand, gives an odd function of ω.

$$k^q(\omega) = \mathrm{Im}\left\{\int_0^\infty \overline{F_k^q(t)F_k^{-q}(t-\tau)}\exp(-i\omega\tau)\,d\tau\right\}$$
$$= \mathrm{Im}\left\{\int_0^\infty \overline{F_k^q(t)F_k^{-q}(t+\tau)}\exp(-i\omega\tau)\,d\tau\right\} \quad (3.38)$$

Taking into account that $k^q\left(\omega_p^{-q}\right) = k^q\left(-\omega_p^q\right) = -k^{-q}\left(\omega_p^q\right) = -k^q\left(\omega_p^q\right),\ k^q(0) = 0$:

$$\sum_{q=-k}^{k}\sum_p \left[\mathbf{A}_{kp}^{-q}, \left[\mathbf{A}_{kp}^q, \sigma^T(t) - \sigma_0\right]\right] k^q\left(\omega_p^q\right) =$$
$$\sum_{q=0}^{k}\sum_p \left[\left[\mathbf{A}_{kp}^{-q}, \mathbf{A}_{kp}^q\right], \sigma^T(t) - \sigma_0\right] k^q\left(\omega_p^q\right) \quad (3.39)$$

In the eigenbase of the Hamiltonian operator, the matrix elements of the equilibrium density operator σ_0 are given by:

$$\sigma_{0,mn} = \left\langle m\left|e^{-\frac{\mathcal{H}}{k_BT}}\right|n\right\rangle\left\{\sum\nolimits_{i=1}^N \left\langle i\left|e^{-\frac{\mathcal{H}}{k_BT}}\right|i\right\rangle\right\}^{-1} = \delta_{m,n}e^{-\frac{\mathcal{H}}{k_BT}}\left\{\sum\nolimits_{i=1}^N \left\langle i\left|e^{-\frac{\mathcal{H}}{k_BT}}\right|i\right\rangle\right\}^{-1} \quad (3.40)$$

In the high-temperature limit, $E_n \ll k_BT$, it can be approximated:

$$\sigma_0 = e^{-\mathcal{H}/k_BT}/\mathrm{Tr}\left\{e^{-\mathcal{H}/k_BT}\right\} \approx (\mathbf{E} - \mathcal{H}/k_BT)/\mathrm{Tr}\{\mathbf{E} - \mathcal{H}/k_BT\} \quad (3.41)$$

Thus, $\sigma_0 \propto \mathcal{H}_0$. Because of (3.31), $\left[\left[\mathbf{A}_{kp}^{-q}, \mathbf{A}_{kp}^q\right], \sigma_0\right] = 0$, so we can simplify (3.36):

$$\frac{\mathrm{d}\sigma^T(t)}{\mathrm{d}t} = -\frac{1}{2}\sum_{q=-k}^{k}\sum_{p}\left[\mathbf{A}_{kp}^{-q},\left[\mathbf{A}_{kp}^{q},\sigma^T(t)-\sigma_0\right]\right]j^q\left(\omega_p^q\right) + i\sum_{q=0}^{k}\sum_{p}\left[\left[\mathbf{A}_{kp}^{-q},\mathbf{A}_{kp}^{q}\right],\sigma^T(t)\right]k^q\left(\omega_p^q\right) \tag{3.42}$$

By transforming (3.42) to the laboratory frame, we obtain the Liouville-von Neumann differential equation for the density operator:

$$\frac{\mathrm{d}\sigma(t)}{\mathrm{d}t} = -i[\mathcal{H}_0,\sigma(t)] - i[\Delta,\sigma(t)] - \hat{\Gamma}(\sigma(t)-\sigma_0) \tag{3.43}$$

where the *relaxation superoperator* $\hat{\Gamma}$ is

$$\hat{\Gamma} = \frac{1}{2}\sum_{q=-k}^{k}\sum_{p} j^q\left(\omega_p^q\right)\left[\mathbf{A}_{kp}^{-q},\left[\mathbf{A}_{kp}^{q},\right]\right] \tag{3.44}$$

and

$$\Delta = -\sum_{q=0}^{k}\sum_{p} k^q\left(\omega_p^q\right)\left[\mathbf{A}_{kp}^{-q},\mathbf{A}_{kp}^{q}\right] \tag{3.45}$$

is the dynamic frequency shift operator responsible for second-order frequency shifts of the resonance lines. It can be incorporated into the Hamiltonian by redefining $\mathcal{H}_0 + \Delta \rightarrow \mathcal{H}_0$, to obtain the final equation, called *master equation*:

$$\frac{\mathrm{d}\sigma(t)}{\mathrm{d}t} = -i[\mathcal{H}_0,\sigma(t)] - \hat{\Gamma}(\sigma(t)-\sigma_0). \tag{3.46}$$

For the master equation to be suitable for studying the auto- and cross-relaxation of longitudinal and transverse magnetization, it is better to expand the density operator in (3.46) in terms of its basis operators. In general, an arbitrary density operator can be represented as a linear combination of a complete set of its orthogonal basis operators $\mathbf{B}_k$:

$$\sigma(t) = \sum_{k=1}^{N} b_k(t)\mathbf{B}_k \tag{3.47}$$

where $b_k(t)$ are time-dependent complex coefficients and N depends on the dimensionality of the system; for the system of M ½-spins, $N = 4^M$. The orthogonality condition for the $\mathbf{B}_k$ operators in (3.47) is:

$$\mathrm{Tr}\{\mathbf{B}_j^\dagger \mathbf{B}_k\} = \langle \mathbf{B}_j | \mathbf{B}_k \rangle = \delta_{jk} \langle \mathbf{B}_k | \mathbf{B}_k \rangle \tag{3.48}$$

A normalized basis operator satisfies $\langle \mathbf{B}'_k | \mathbf{B}'_k \rangle = 1$; an unnormalized basis operator $\mathbf{B}_k$ can be normalized using relationship $\mathbf{B}'_k = \mathbf{B}_k / \langle \mathbf{B}_k | \mathbf{B}_k \rangle^{1/2}$. The expectation value of an operator $\mathbf{A}$ can be obtained by applying (3.47) to (3.13):

$$\langle A \rangle(t) = \mathrm{Tr}\{\sigma(t)\mathbf{A}\} = \mathrm{Tr}\left\{\sum_{k=1}^{N} b_k(t)\mathbf{B}_k\mathbf{A}\right\} = \sum_{k=1}^{N} b_k(t)\mathrm{Tr}\{\mathbf{B}_k\mathbf{A}\}. \tag{3.49}$$

The time evolution of the density operator, the Liouville-von Neumann equation can also be expressed in terms of basis operators:

$$\frac{\mathrm{d}\sigma(t)}{\mathrm{d}t} = -i[\mathcal{H}, \sigma(t)] = -i\sum_{k=1}^{N} b_k(t)[\mathcal{H}, \mathbf{B}_k]. \tag{3.50}$$

The choice of basis operators is not fixed. For a spin system of one spin, one can choose, for example, Cartesian set of basis operators $\mathbf{E}/2, I_x, I_y, I_z)$. These operators are Hermitian. Often another set of operators is used: $\mathbf{E}/2, I_z, I^+, I^-$, where $I^\pm = I_x \pm iI_y$ and $(I^\pm)^\dagger = I^\mp$. For a system of two spins, there are 16 operators, for example:

$$\frac{\mathbf{E}}{2}, I_x, I_y, I_z, S_x, S_y, S_z, 2I_xS_z, 2I_yS_z, 2I_zS_z, 2I_zS_x, 2I_zS_y, 2I_xS_x, 2I_yS_y, 2I_xS_y, 2I_yS_x \tag{3.51a}$$

$$\frac{\mathbf{E}}{2}, I^+, I^-, I_z, S^+, S^-, S_z, 2I^+S_z, 2I^-S_z, 2I_zS_z, 2I_zS^+, 2I_zS^-, 2I^+S^+, 2I^-S^-, 2I^+S^-, 2I^-S^+ \tag{3.51b}$$

The master equation is then rewritten in its matrix form:

$$\frac{\mathrm{d}b_r(t)}{\mathrm{d}t} = \sum_s \{-i\Omega_{rs}b_s(t) - \Gamma_{rs}(b_s(t) - b_{s0})\} \tag{3.52}$$

in which Ω_{rs} is a characteristic frequency:

$$\Omega_{rs} = \frac{\langle \mathbf{B}_r | [\mathcal{H}_0, \mathbf{B}_s] \rangle}{\langle \mathbf{B}_r | \mathbf{B}_r \rangle} \tag{3.53}$$

Γ_{rs} is the rate constant for the relaxation between operators $\mathbf{B}_r$ and $\mathbf{B}_s$:

$$\Gamma_{rs} = \frac{\langle \mathbf{B}_r | \hat{\Gamma} \mathbf{B}_s \rangle}{\langle \mathbf{B}_r | \mathbf{B}_r \rangle} = \frac{1}{2}\sum_{q=-k}^{k}\sum_p \left\{ \frac{\left\langle \mathbf{B}_r \middle| \left[\mathbf{A}_{kp}^{-q}, \left[\mathbf{A}_{kp}^{q}, \mathbf{B}_s\right]\right] \right\rangle}{\langle \mathbf{B}_r | \mathbf{B}_r \rangle} \right\} j^q(\omega_p^q) \tag{3.54}$$

For normalized basis operators, $\Gamma_{rs} = \Gamma_{sr}$. Finally $b_r(t)$ are the coefficients of the density operator decomposition, which could be obtained as:

$$b_r(t) = \langle \mathbf{B}_r | \sigma(t) \rangle / \langle \mathbf{B}_r | \mathbf{B}_r \rangle \tag{3.55}$$

When $\mathbf{B}_r$ is a normalized Hermitian operator, $\mathbf{B}_r^\dagger = \mathbf{B}_r$, and $\langle \mathbf{B}_r | \mathbf{B}_r \rangle = 1$:

$$b_r(t) = \langle \mathbf{B}_r | \sigma(t) \rangle = \mathrm{Tr}\{\mathbf{B}_r^\dagger \sigma(t)\} = \mathrm{Tr}\{\mathbf{B}_r \sigma(t)\} = \langle \overline{\mathbf{B}_r} \rangle (t) \tag{3.56}$$

In this case, $b_r(t)$ is the expectation value of the corresponding basis operator.

Relaxation rate constants Γ_{rs} are the elements of the relaxation matrix Γ. Diagonal elements Γ_{rr} are the rate constants for the auto-relaxation of operator $\mathbf{B}_r$; non-diagonal elements represent cross-relaxation between different operators $\mathbf{B}_r$ and $\mathbf{B}_s$. It can be shown that cross-relaxation is not possible between operators of different quantum order (for example, $2I_zS_z$ and S^+). Furthermore, for two operators of the same quantum order, if one of them involves transitions (as $2I^+S^-$and S_z), under the *secular hypothesis* it is deemed that if their frequencies are not degenerate, then cross-relaxation between these operators is not possible. Without secular hypothesis, it can be shown that cross-relaxation for two such operators $\mathbf{B}_r$ and $\mathbf{B}_s$ can be neglected as long as the difference between their eigenfrequencies $|\omega_r - \omega_s|$ is greater than the linewidth. Conversely, as $|\omega_r - \omega_s| \to 0$, cross-relaxation gradually becomes effective.

3.3.2 Spectral density function

For the relaxation in isotropic fluids in the high-temperature limit, as shown by Hubbard [144], only one spectral density function needs to be calculated:

$$j^q(\omega) = (-1)^q j^0(\omega) = (-1)^q j(\omega) \tag{3.57}$$

The relaxation mechanisms of interest – dipolar interaction, chemical shift anisotropy (CSA), can be described by tensor operators of rank 2. Random functions in (3.37) are factored to give

$$F_2^0(t) = c_0(t) Y_2^0[\Omega(t)] \tag{3.58}$$

and the spectral density function is rewritten as follows:

$$\begin{aligned} j(\omega) &= \mathrm{Re}\left\{\int_{-\infty}^{\infty} C(\tau) \exp(-i\omega\tau) d\tau\right\} \\ C(\tau) &= \overline{c_0(t) c_0(t+\tau) Y_2^0[\Omega(t)] Y_2^0[\Omega(t+\tau)]} \end{aligned} \tag{3.59}$$

where $C(\tau)$ is the *stochastic correlation function*. It contains $c_0(t)$, a function of physical constants and spatial variables (see Table 3.1), and $Y_2^0[\Omega(t)]$, a modified second-order spherical harmonic function of polar angles $\Omega(t)$ defining the orientation of principal direction of interaction in the laboratory reference frame.

Table 3.1 Spatial functions for relaxation mechanisms

Interaction	$c(t)$
Dipolar	$-\sqrt{6}(\mu_0/4\pi)\hbar\gamma_I\gamma_S r_{IS}(t)^{-3}$
CSA with axially symmetric chemical shift tensor, $\sigma_{zz}=\sigma_{\parallel}$, $\sigma_{xx}=\sigma_{yy}=\sigma_{\perp}, \Delta\sigma=\sigma_{\parallel}-\sigma_{\perp}$	$\Delta\sigma\gamma_I B_0/\sqrt{3}$

For the dipole-dipole interaction, this principal direction is the internuclear bond vector; for the CSA interaction with an axially symmetric chemical shift tensor, it is the axis of symmetry. Modified spherical harmonics are given in the Table 3.2.

Oscillating magnetic fields that cause relaxation are the result of the participation of the molecule in random motions, including rotational diffusion. The physical sense of the spectral density function is the measure of the contribution from motions with frequency components in the range $(\omega, \omega + d\omega)$ to the rotational dynamics of the molecule.

Table 3.2 Modified second-order spherical harmonics

q	Y_2^q	$Y_2^{-q}=(-1)^q Y_2^{q*}$
0	$(3\cos^2\theta-1)/2$	$(3\cos^2\theta-1)/2$
1	$-\sqrt{3/2}\sin\theta\cos\theta e^{i\phi}$	$\sqrt{3/2}\sin\theta\cos\theta e^{-i\phi}$
2	$\sqrt{3/8}\sin^2\theta e^{i2\phi}$	$\sqrt{3/8}\sin^2\theta e^{-i2\phi}$

Let us take an example of a rigid spherical molecule. In this case, spatial function is a constant, $c(t)=c_0$, and $j(\omega)=c_0^2 J(\omega)$, where $J(\omega)$ is the orientation spectral density function:

$$J(\omega)=\mathrm{Re}\left\{\int_{-\infty}^{\infty} C_{00}^2(\tau)\exp(-i\omega\tau)\,d\tau\right\} \tag{3.60}$$

in which $C_{00}(\tau)$ is the orientation correlation function:

$$C_{00}^2(\tau)=\overline{Y_2^0[\Omega(t)]Y_2^0[\Omega(t+\tau)]} \tag{3.61}$$

In the case of isotropic rotational diffusion of a rigid rotor, the orientation correlation function is given by [145], and has exponential form:

$$C_{00}^2(\tau) = \frac{1}{5}\exp\left(-\frac{\tau}{\tau_c}\right), \tag{3.62}$$

where τ_c, the *correlation time*, is approximately the average time for the molecule to rotate by 1 radian. In globular, rigid molecules, such as folded proteins, this depends on the size of the molecule, temperature and solvent viscosity, and generally is of the order of nanoseconds. The corresponding spectral density function has the form of a Lorentzian:

$$J(\omega) = \frac{2}{5}\frac{\tau_c}{(1+\omega^2\tau_c^2)} \tag{3.63}$$

In reality, a molecule is not a rigid body, and both overall rotational motion and relative motions of nuclei in the molecular reference frame are contributing to the stochastic modulation of local magnetic fields. In the case of isotropic rotational diffusion and under the approximation that intramolecular motions and overall rotational motion are statistically independent, then the total correlation function can be factored as:

$$C(\tau) = C_0(\tau)C_1(\tau), \tag{3.64}$$

where the correlation function for overall motion, $C_0(\tau)$, is given by equation (3.61) or (3.62), and the correlation function for internal motions, $C_1(\tau)$, has to be calculated as in (3.59) using a particular model of intramolecular motions and taking into account that the reference frame is that of the molecule. However, in 1982 G. Lipari and A. Szabo proposed a *model-free* approach, called now Lipari-Szabo model-free formalism [146, 147]. The local motion in this approach is characterized only by two parameters: the *generalized order parameter* $\mathcal{S}^2$ and the correlation time of internal motion τ_i. The generalized order parameter describes the 'degree of spatial restriction of the motion' [146]; $1-\mathcal{S}^2$ is the amplitude of the contribution of internal dynamics to the relaxation. It satisfies the inequality $0 \leq \mathcal{S}^2 \leq 1$. The correlation function for internal motion is written as:

$$C_1(\tau) = \mathcal{S}^2 + (1-\mathcal{S}^2)\exp(-\tau/\tau_i), \tag{3.65}$$

and the corresponding spectral density function is given by:

$$j(\omega) = \frac{2}{5}\overline{c_0^2}\left[\frac{\mathcal{S}^2\tau_c}{(1+\omega^2\tau_c^2)} + \frac{(1-\mathcal{S}^2)\tau_i'}{(1+\omega^2\tau_i'^2)}\right]; \quad \frac{1}{\tau_i'} = \frac{1}{\tau_c} + \frac{1}{\tau_i} \tag{3.66}$$

For example; if $\mathcal{S}^2 = 1$, the internal motion is totally restricted, and we fall back to the case of the rotating rigid body, (3.66) being equal to (3.63) (with $j(\omega) = c_0^2 J(\omega)$). The equation for the order parameter is:

$$S^2 = \left[\overline{c_0^2}\right]^{-1} \sum_{q=-2}^{2} \left|\overline{c_0 Y_2^q[\Omega]}\right|^2, \tag{3.67}$$

in which the bar indicates an ensemble average performed over the equilibrium distribution of orientations Ω in the molecular reference frame.

The application of the model-free formalism to the study of the dynamics of flexible and intrinsically disordered proteins is discussed in detail in the next chapter.

3.4 Relaxation mechanisms and equations for relaxation rates

There are a large number of physical interactions at the origin of spin relaxation. However, the ones that are generally important for macromolecules are: dipolar, quadrupolar and scalar interactions, chemical shift anisotropy and paramagnetic relaxation. The latter has the same Hamiltonian as the nuclear dipolar relaxation, but is caused by the interaction between a nucleus and an unpaired electron. For ½-spin nuclei in diamagnetic biological macromolecules, the dipolar interaction and chemical shift anisotropy are by far the most important mechanisms.

3.4.1 Dipolar relaxation

We will consider a two-spin system *IS* with $\omega_I \gg \omega_S$ and in the absence of scalar coupling between spins ($J_{IS} = 0$). The latter condition could be attained, if necessary, by decoupling the spins. Terms $\mathbf{A}_{2p}^q$ for the dipolar interaction are listed in the Table 3.3. Relaxation rates are then calculated using (3.54). For the longitudinal magnetization terms $I_z\, S_z$, the relaxation matrix will have 2x2 dimensions with individual elements Γ_{rs} giving the relaxation rate constant for the relaxation between I_z (for *r* or *s* = 1) and S_z (for *r* or *s* = 2).

Table 3.3 Tensor operators for the dipolar interaction

q	p	$\mathbf{A}_{2p}^q$	$\mathbf{A}_{2p}^{-q} = (-1)^q \mathbf{A}_{2p}^{q\dagger}$	ω_p^q
0	0	$\left(2/\sqrt{6}\right) I_z S_z$	$\left(2/\sqrt{6}\right) I_z S_z$	0
0	–1	$-1/\left(2/\sqrt{6}\right) I^- S^+$	$-1/\left(2/\sqrt{6}\right) I^+ S^-$	$\omega_S - \omega_I$
0	1	$-1/\left(2/\sqrt{6}\right) I^+ S^-$	$-1/\left(2/\sqrt{6}\right) I^- S^+$	$\omega_I - \omega_S$
1	0	$-(1/2)\, I_z S^+$	$(1/2)\, I_z S^-$	ω_S
1	1	$-(1/2)\, I^+ S_z$	$(1/2)\, I^- S_z$	ω_I
2	0	$(1/2)\, I^+ S^+$	$(1/2)\, I^- S^-$	$\omega_I + \omega_S$

Doing all necessary calculations, we obtain relaxation rates:

$$\begin{aligned}\Gamma_{11} &= R_{1I} = \frac{1}{24}[j(\omega_I - \omega_S) + 3j(\omega_I) + 6j(\omega_I + \omega_S)]\\ \Gamma_{22} &= R_{1S} = \frac{1}{24}[j(\omega_I - \omega_S) + 3j(\omega_S) + 6j(\omega_I + \omega_S)]\\ \Gamma_{12} &= \sigma_{IS} = \frac{1}{24}[-j(\omega_I - \omega_S) + 6j(\omega_I + \omega_S)]\end{aligned} \tag{3.68}$$

If the interspin distance r_{IS} is considered to be constant, then:

$$\begin{aligned}R_{1I} &= \frac{d_{00}}{4}[J(\omega_I - \omega_S) + 3J(\omega_I) + 6J(\omega_I + \omega_S)]\\ R_{1S} &= \frac{d_{00}}{4}[J(\omega_I - \omega_S) + 3J(\omega_S) + 6J(\omega_I + \omega_S)]\\ \sigma_{IS} &= \frac{d_{00}}{4}[-J(\omega_I - \omega_S) + 6J(\omega_I + \omega_S)]\\ d_{00} &= (\mu_0/4\pi)^2\hbar^2\gamma_I^2\gamma_S^2 r_{IS}^{-6}\end{aligned} \tag{3.69}$$

The evolution of the longitudinal operators I_z, S_z is described by equations:

$$\begin{aligned}\frac{d}{dt}(\langle I_z\rangle(t) - \langle I_z^0\rangle) &= -\Gamma_{11}(\langle I_z\rangle(t) - \langle I_z^0\rangle) - \Gamma_{12}(\langle S_z\rangle(t) - \langle S_z^0\rangle)\\ \frac{d}{dt}(\langle S_z\rangle(t) - \langle S_z^0\rangle) &= -\Gamma_{21}(\langle I_z\rangle(t) - \langle I_z^0\rangle) - \Gamma_{22}(\langle S_z\rangle(t) - \langle S_z^0\rangle)\end{aligned} \tag{3.70}$$

For the transverse relaxation, we need to consider the transverse magnetization operator I^+. After doing all the necessary calculations, we find out that no basis operator cross-relaxes with I^+. The same is valid for I^-, S^+, S^-. For the auto-relaxation:

$$\begin{aligned}R_{2I^+} = R_{2I^-} = R_{2I} &= \frac{1}{48}[4j(0) + j(\omega_I - \omega_S) + 3j(\omega_I) + 6j(\omega_S) + 6j(\omega_I + \omega_S)]\\ R_{2S^+} = R_{2S^-} = R_{2S} &= \frac{1}{48}[4j(0) + j(\omega_I - \omega_S) + 3j(\omega_S) + 6j(\omega_I) + 6j(\omega_I + \omega_S)]\end{aligned} \tag{3.71}$$

Again if r_{IS} is constant, then $d_{00} = (\mu_0/4\pi)^2\hbar^2\gamma_I^2\gamma_S^2 r_{IS}^{-6}$ and the relaxation rates are given by:

$$\begin{aligned}R_{2I} &= \frac{d_{00}}{8}[4J(0) + J(\omega_I - \omega_S) + 3J(\omega_I) + 6J(\omega_S) + 6J(\omega_I + \omega_S)]\\ R_{2S} &= \frac{d_{00}}{8}[4J(0) + J(\omega_I - \omega_S) + 3J(\omega_S) + 6J(\omega_I) + 6J(\omega_I + \omega_S)]\end{aligned} \tag{3.72}$$

3.4.2 Chemical shift anisotropy (CSA) relaxation

Chemical shift is the result of the modification of the magnetic field experienced by a given nucleus due to the local electronic environment. The resulting local field is anisotropic: therefore, the chemical shift should be rather described by a tensor. As a consequence of

the molecular motion, the reorientation of this tensor with respect to the laboratory frame induces a time-dependent random magnetic field that is experienced by the nucleus and gives rise to relaxation. The CSA relaxation is negligible for ^{1}H nuclei and is significant for ^{13}C, ^{15}N and ^{31}P; the rate constant has a quadratic dependence on the strength of the magnetic field. The $\mathbf{A}_{2p}^{q}$ terms for the CSA interaction are given in the Table 3.4. The CSA tensor in this calculation is considered to be axially symmetric, such that $\sigma_{zz} = \sigma_{\parallel},\ \sigma_{xx} = \sigma_{yy} = \sigma_{\perp}, \Delta\sigma = \sigma_{\parallel} - \sigma_{\perp}$.

Table 3.4 Tensor operators for the CSA interaction

q	p	$\mathbf{A}_{2p}^{q}$	$\mathbf{A}_{2p}^{-q} = (-1)^{q}\mathbf{A}_{2p}^{q\dagger}$	ω_{p}^{q}
0	0	$\left(2/\sqrt{6}\right)I_z$	$\left(2/\sqrt{6}\right)I_z$	0
1	1	$-(1/2)\ I^{+}$	$(1/2)\ I^{-}$	ω_I

Following all necessary calculations, we obtain longitudinal and transversal relaxation rates as follows:

$$
\begin{aligned}
R_{1I} &= d_{00}^{CSA,I}[J(\omega_I)], \qquad R_{1S} = d_{00}^{CSA,S}[J(\omega_S)] \\
R_{2I} &= \frac{d_{00}^{CSA,I}}{6}[4J(0) + 3J(\omega_I)] \\
R_{2S} &= \frac{d_{00}^{CSA,S}}{6}[4J(0) + 3J(\omega_S)]
\end{aligned}
\tag{3.73}
$$

where $d_{00}^{CSA,I} = (\Delta\sigma\gamma_I B_0)^2/3 = \Delta\sigma^2\omega_I^2/3, d_{00}^{CSA,S} = \Delta\sigma^2\omega_S^2/3$.

As shown later, relaxation rates in the presence of different relaxation mechanisms will be the sum of relaxation rates for each individual mechanism, if interference effects are not present or they are cancelled. So, for dipolar and CSA interaction the aforementioned relaxation rates will be written as (if r_{IS} is constant):

$$
\begin{aligned}
R_{1I} &= \frac{d^2}{4}[J(\omega_I - \omega_S) + 3J(\omega_I) + 6J(\omega_I + \omega_S)] + c_I^2[J(\omega_I)] \\
R_{1S} &= \frac{d^2}{4}[J(\omega_I - \omega_S) + 3J(\omega_S) + 6J(\omega_I + \omega_S)] + c_S^2[J(\omega_S)] \\
\sigma_{IS} &= \frac{d^2}{4}[-J(\omega_I - \omega_S) + 6J(\omega_I + \omega_S)] \\
R_{2I} &= \frac{d^2}{8}[4J(0) + J(\omega_I - \omega_S) + 3J(\omega_I) + 6J(\omega_S) + 6J(\omega_I + \omega_S)] + \frac{c_I^2}{6}[4J(0) + 3J(\omega_I)] \\
R_{2S} &= \frac{d^2}{8}[4J(0) + J(\omega_I - \omega_S) + 3J(\omega_S) + 6J(\omega_I) + 6J(\omega_I + \omega_S)] + \frac{c_S^2}{6}[4J(0) + 3J(\omega_S)] \\
d &= \sqrt{d_{00}} = \left(\frac{\mu_0}{4\pi}\right)\hbar\gamma_I\gamma_S r_{IS}^{-3}, \qquad c_I = \sqrt{d_{00}^{CSA,I}} = \frac{\Delta\sigma\omega_I}{\sqrt{3}}, \qquad c_S = \frac{\Delta\sigma\omega_S}{\sqrt{3}}
\end{aligned}
\tag{3.74}
$$

It is important to mention that the same spectral density function is used for dipolar and CSA interaction; it is possible to do so because we imposed the condition of motional isotropy in deriving the master equation (see commentary before (3.35)).

3.4.3 Nuclear Overhauser Effect

The nuclear Overhauser effect (NOE) is a manifestation of cross-relaxation between two dipolar-coupled spins. It is characterized either directly by the cross-relaxation rate constant, σ_{IS}, or by the steady-state NOE enhancement, η_{IS}. The latter is obtained in the so-called steady-state NOE experiment, which is described below.

Let us take a dipole-coupled two-spin system. If the I spin is irradiated by a weak RF field (that does not perturb the S spin) for a period of time sufficiently long compared to the longitudinal relaxation rates of both spins ($t \gg 1/R_{1I}, 1/R_{1S}$), then the average populations across the I spin transition are equalized. I spins are said to be saturated in this situation. The S spin magnetization evolves to its steady-state value $\langle S_z^{ss}\rangle$. Setting $d\langle S_z\rangle/dt = 0$ and $\langle I_z\rangle(t) = 0$ in (3.70), we obtain:

$$\frac{d}{dt}(\langle S_z\rangle(t) - \langle S_z^0\rangle) = -\sigma_{IS}(0 - \langle I_z^0\rangle) - R_{1S}(\langle S_z^{ss}\rangle - \langle S_z^0\rangle) = 0$$

$$\frac{\langle S_z^{ss}\rangle}{\langle S_z^0\rangle} = 1 + \frac{\sigma_{IS}\langle I_z^0\rangle}{[R_{1S}\langle S_z^0\rangle]} \tag{3.75}$$

Replacing $\langle I_z^0\rangle/\langle S_z^0\rangle$ by γ_I/γ_S gives:

$$\frac{\langle S_z^{ss}\rangle}{\langle S_z^0\rangle} = 1 + \eta_{IS}, \qquad \eta_{IS} = \frac{\sigma_{IS}\gamma_I}{R_{1S}\gamma_S} \tag{3.76}$$

The value of the NOE enhancement is measured using the steady-state NOE difference experiment, where 2 spectra are recorded: in one the I spins are saturated, in the other they are not (reference spectrum). The intensity of the S spin signal in the 'saturated' spectrum is proportional to $\langle S_z^{ss}\rangle$ while in the reference spectrum it is proportional to $\langle S_z^0\rangle$. Thus the ratio of these intensities gives us $1 + \eta_{IS}$.

3.4.4 Relaxation in the rotating frame

In the presence of an applied rf field (for example, in $R_{1\rho}$ or ROESY experiments), before the transformation into the interaction frame, we should remove the time dependence of rf Hamiltonian $\mathcal{H}_{rf}(t)$ so that it becomes "static"; this is achieved by applying a transformation into a rotating frame. If the Zeeman Hamiltonian is dominant in $\mathcal{H}_0$ and the scalar coupling Hamiltonian is ignored, then the interaction frame is equivalent to a

doubly rotating tilted frame (initial rotating frame to remove the rotation of $\mathcal{H}_{rf}$ around z axis, then tilt to align static $\mathcal{H}_{rf}$ on z axis, then second rotating (interaction) frame to remove the effect of the static Hamiltonian – as described before). As a consequence of the initial rotating-frame transformation, the frequencies used as arguments of the spectral density function $j^q(\omega_p^q)$ should be modified to $\omega_p^q + \omega_p^{q(rf)}$, where $\omega_p^{q(rf)}$ is defined by:

$$\sum_{i=1}^{K} \omega_{rf,i}\left[I_{zi}, \mathbf{A}_{kp}^q\right] = \omega_p^{q(rf)} \mathbf{A}_{kp}^q \tag{3.77}$$

where $\omega_{rf,i}$ is the frequency of the rotating frame for the *i*th spin and *K* is the number of irradiated spins. However, in most cases for macromolecules $\omega_1 \tau_c \ll 1$, where ω_1 is the strength of the applied rf field and τ_c is the rotational correlation time of the molecule; in that case, $j^q\left(\omega_p^q + \omega_p^{q(rf)}\right) \approx j^q\left(\omega_p^q\right)$. Under this approximation, the relaxation rate constants in the rotating frame are still calculated using the master equation (3.54) by transforming the operators $\mathbf{B}_r'$, $\mathbf{B}_s'$ in the tilted frame to the laboratory frame:

$$\Gamma_{rs}' = \left\langle \mathbf{U}^{-1}\mathbf{B}_r'\mathbf{U} \middle| \hat{\Gamma}\mathbf{U}^{-1}\mathbf{B}_s'\mathbf{U} \right\rangle / \langle \mathbf{B}_r' | \mathbf{B}_r' \rangle \tag{3.78}$$

If the applied RF field has x-phase, the transformation **U** is defined as a rotation around y axis:

$$\mathbf{U} = \exp\left\{ i \sum_{i=1}^{K} \theta_i I_{yi} \right\} \tag{3.79}$$

where $\theta_i = \omega_1/\Omega_i$ is the *tilt angle*. Thus, for example, the autorelaxation rate $R_{1i}(\theta_i)$, commonly called $R_{1\rho}$, is given by (i = *I*, *S*):

$$R_{1i}(\theta_i) = R_{1i}\cos^2\theta_i + R_{2i}\sin^2\theta_i \tag{3.80}$$

The effect of the tilted field is to average the relaxation rate constants of the laboratory (longitudinal) and rotating frames (transverse) by the projection of the spin operators onto the tilted reference frame.

3.4.5 Interference between relaxation mechanisms

In most cases, different Hamiltonians due to different relaxation mechanisms are acting on the spin system. It was mentioned before that, in the case of absence of interference, relaxation rates will sum up. Here this result will be obtained by calculation, and the impact of interference will be discussed.

The equation (3.29) is generalized for the case of multiple Hamiltonians:

$$\mathcal{H}_1(t) = \sum_m \sum_{q=-k}^{k} (-1)^q F_{mk}^{-q}(t) \mathbf{A}_{mk}^{q} \tag{3.81}$$

where the index m refers to different relaxation mechanisms. Following through all steps from (3.29) and (3.54) one could obtain the generalized form of the master equation for relaxation matrix:

$$\begin{aligned} \Gamma_{rs} &= \frac{1}{2}\sum_m \sum_q \sum_p \left\{ \frac{\left\langle \mathbf{B}_r \middle| \left[\mathbf{A}_{mkp}^{-q}, \left[\mathbf{A}_{mkp}^{q}, \mathbf{B}_s \right] \right] \right\rangle}{\langle \mathbf{B}_r | \mathbf{B}_r \rangle} \right\} j^q\left(\omega_p^q\right) \\ &+ \frac{1}{2}\sum_{\substack{m,n \\ m \neq n}} \sum_q \sum_p \left\{ \frac{\left\langle \mathbf{B}_r \middle| \left[\mathbf{A}_{mkp}^{-q}, \left[\mathbf{A}_{nkp}^{q}, \mathbf{B}_s \right] \right] \right\rangle}{\langle \mathbf{B}_r | \mathbf{B}_r \rangle} \right\} j_{mn}^q\left(\omega_p^q\right) \\ &= \frac{1}{2}\sum_m \Gamma_{rs}^m + \frac{1}{2}\sum_{\substack{m,n \\ m \neq n}} \Gamma_{rs}^{mn} \end{aligned} \tag{3.82}$$

where Γ_{rs}^m is the relaxation rate constant due to *m*th relaxation mechanism, and Γ_{rs}^{mn} is the relaxation rate constant due to the interference, or *cross-correlation* between different relaxation mechanisms. The latter contains a special, *cross-correlated* spectral density function:

$$j_{mn}^q(\omega) = \mathrm{Re}\left\{ \int_{-\infty}^{\infty} \overline{F_{mk}^q(t) F_{nk}^{-q}(t+\tau)} \exp(-i\omega\tau)\, d\tau \right\} \tag{3.83}$$

Unless random processes $F_{mk}^q(t)$ and $F_{nk}^q(t)$ linked to different relaxation mechanisms are correlated, the cross-correlated spectral density function is equal to zero. If the random processes related to distinct relaxation mechanisms are not correlated for a given spin system, then $\Gamma_{rs}^{mn} = 0$ and the contributions to relaxation of the different mechanisms simply sum up.

In the case of biological macromolecules, the most frequently encountered cases of cross-correlation between relaxation mechanisms are the dipole/dipole-CSA cross-correlation, and the cross-correlation between dipolar interactions of different pairs of spins. One of the best-known cases of the former kind is the cross-relaxation due to the cross-correlation between the ^{1}H-^{15}N dipolar coupling and ^{15}N chemical shift anisotropy.

One of the consequences of cross-correlation is the observation of cross-relaxation pathways that are “forbidden” in the presence of only one relaxation mechanism. This

effect is called *cross-correlated cross-relaxation*. The derivation of relaxation rates was demonstrated by Goldman in [148].

A two-spin system is considered, in which the S spin (e.g. ^{15}N) has a significant chemical shift anisotropy, and also has a dipolar relaxation with the spin I (e.g. ^{1}H). The evolution of the components of the density operator is described by equation:

$$\begin{aligned}
\frac{d\langle I_z\rangle(t)}{dt} &= -R_{1I}^{\mathrm{DD}}(\langle I_z\rangle(t) - \langle I_z^0\rangle) - \sigma_{IS}(\langle S_z\rangle(t) - \langle S_z^0\rangle) \\
\frac{d\langle S_z\rangle(t)}{dt} &= -\left(R_{1S}^{\mathrm{DD}} + R_{1S}^{\mathrm{CSA}}\right)(\langle S_z\rangle(t) - \langle S_z^0\rangle) - \sigma_{IS}(\langle I_z\rangle(t) - \langle I_z^0\rangle) - \eta_z\langle 2I_zS_z\rangle(t) \\
\frac{d\langle 2I_zS_z\rangle(t)}{dt} &= -\left(R_{1IS}^{\mathrm{DD}} + R_{1S}^{\mathrm{CSA}}\right)\langle 2I_zS_z\rangle(t) - \eta_z(\langle S_z\rangle(t) - \langle S_z^0\rangle)
\end{aligned} \tag{3.84}$$

for the longitudinal components, and:

$$\begin{aligned}
\frac{d\langle S^+\rangle(t)}{dt} &= -i\pi J_{IS}\langle 2I_zS^+\rangle(t) - \left(R_2^{\mathrm{DD}} + R_{2S}^{\mathrm{CSA}}\right)\langle S^+\rangle(t) - \eta_{xy}\langle 2I_zS^+\rangle(t) \\
\frac{d\langle 2I_zS^+\rangle(t)}{dt} &= -i\pi J_{IS}\langle S^+\rangle(t) - \left(R_{2IS}^{\mathrm{DD}} + R_{2S}^{\mathrm{CSA}}\right)\langle 2I_zS^+\rangle(t) - \eta_{xy}\langle S^+\rangle(t)
\end{aligned} \tag{3.85}$$

for the transverse components. R_{1IS}^{DD} and R_{2IS}^{DD} are the autorelaxation rates for terms $2I_zS_z$ and $2I_zS^+$, respectively, and are not discussed here. If we consider the chemical shift tensor to be axially symmetric, and the molecule to be a rigid sphere, we obtain the following cross-correlated cross-relaxation rate constants:

$$\eta_z = \sqrt{3}cd\left(\frac{3\cos^2\theta - 1}{2}\right)J(\omega_s) \tag{3.86}$$

for the longitudinal cross-correlated cross-relaxation, and:

$$\eta_{xy} = \frac{\sqrt{3}}{6}cd\left(\frac{3\cos^2\theta - 1}{2}\right)\{4J(0) + J(\omega_s)\} \tag{3.87}$$

for the transverse cross-correlated cross-relaxation, where $d = \left(\frac{\mu_0}{4\pi}\right)\hbar\gamma_I\gamma_S r_{IS}^{-3}$, $c = c_S = \Delta\sigma\omega_S/\sqrt{3}$ are the same as in (3.74), and θ is the angle between the IS interspin vector and the symmetry axis of the chemical shift tensor.

3.5 Discussion

We discussed in this chapter ^{15}N relaxation rates that are most used to study protein backbone dynamics. Let us summarize them, in the case of relaxation due to ^{15}N-^{1}H dipole-dipole interaction and ^{15}N chemical shift anisotropy interaction:

$$R_1 = \frac{d^2}{4}[J(\omega_H - \omega_N) + 3J(\omega_N) + 6J(\omega_H + \omega_N)] + c^2[J(\omega_N)] \quad (3.87)$$

$$R_2 = \frac{d^2}{8}[4J(0) + J(\omega_H - \omega_N) + 3J(\omega_N) + 6J(\omega_H) + 6J(\omega_H + \omega_N)] + \frac{c^2}{6}[4J(0) + 3J(\omega_N)] \quad (3.88)$$

$$NOE = 1 + \frac{d^2}{4}\frac{\gamma_H}{R_1\gamma_N}[-J(\omega_H - \omega_N) + 6J(\omega_H + \omega_N)] \quad (3.89)$$

$$\eta_{xy} = \frac{\sqrt{3}}{6}cd\left(\frac{3\cos^2\theta - 1}{2}\right)\{4J(0) + 3J(\omega_N)\} \quad (3.90)$$

$$\eta_z = \sqrt{3}cd\left(\frac{3\cos^2\theta - 1}{2}\right)J(\omega_N) \quad (3.91)$$

where $d = \left(\frac{\mu_0}{4\pi}\right)\hbar\gamma_I\gamma_S r_{IS}^{-3}$, $c = \Delta\sigma\omega_N/\sqrt{3}$, and θ is the angle between the NH bond vector and the symmetry axis of the chemical shift tensor; and, in addition:

$$R_{1\rho}(\theta_t) = R_1\cos^2\theta_t + R_2\sin^2\theta_t \quad (3.92)$$

where $\theta_t = \arctan(\omega_1/\Omega)$ is the tilt angle, ω_1 is the spin-lock field strength (in rad·s^{-1}), $\Omega = \omega - \omega_0$, ω_0 is the population-averaged ^{15}N chemical shift of the given resonance (in rad·s^{-1}), ω is the carrier frequency of the spin-lock field (in rad·s^{-1}).

Measured $R_{1\rho}$ and R_2 rates may have a contribution from slower µs-ms motions, which, when approximated as an instantaneous jump between discrete conformations, are referred to as "conformational exchange". If we consider a two-site $A \leftrightarrow B$ exchange, p_A and p_B are the populations of both sites, $k_{ex} = k_{AB} + k_{BA}$ is the exchange rate and $\Delta\omega = \omega_A - \omega_B$ is the difference between resonance frequencies of the two sites, then there are three possibilities:

- when $k_{ex} \ll \Delta\omega$, exchange is called 'slow' and two peaks corresponding to two sites A and B are present in the spectrum;
- when $k_{ex} \gg \Delta\omega$, exchange is called 'fast' and one peaks is present at $\omega = \omega_A p_A + \omega_B p_B$;
- when $k_{ex} \approx \Delta\omega$, exchange is called 'intermediate' and one or two broad peaks are present.

When the exchange is 'fast', measured R_2 value is given by [149, 150]:

$$R_2 = R_2^0 + \frac{p_A p_B \Delta\omega^2}{k_{ex}} \tag{3.93}$$

where R_2^0 is exchange-free R_2 given by (3.88). For $R_{1\rho}$ value, under the approximation that the effective field $\omega_{eff} = (\omega_1^2 + \Omega^2)$ is the same for both sites A and B, one obtains:

$$R_{1\rho} = R_{1\rho}^0 + \sin^2\theta \frac{p_A p_B \Delta\omega^2 k_{ex}}{k_{ex}^2 + \omega_{eff}^2} \tag{3.94}$$

where $R_{1\rho}^0$ is exchange-free $R_{1\rho}$, given by (3.92) for ^{15}N relaxation, but (3.94) itself is also valid for other kinds of $R_{1\rho}$ relaxation, including proton relaxation. If we apply a very high power spin lock field, $\omega_{eff} \to \infty$, then $R_{1\rho} \to R_{1\rho}^0$. This is the principle of the $R_{1\rho}$ relaxation dispersion: by measuring $R_{1\rho}$ at different spin-lock fields, one obtains its ω_{eff} dependency, which allows the estimation of the exchange rate k_{ex}.

Finally, one can show that exchange has no impact on other relaxation rates cited here and in particular on cross-correlated cross-relaxation rates η_{xy} and η_z: the scalar product in (3.82) with the exchange Hamiltonian (proportional to I_z and S_z) and $\mathbf{B}_r, \mathbf{B}_s$ operators $2I_zS^+, S^+$ (or $2H_zN_z, N_z$) will be zero. That is why η_{xy} and η_z are so precious in terms of obtaining correct values of $J(0)$; comparing $J(0)$ values obtained using auto-relaxation and cross-correlated cross-relaxation rates will provide information on the presence of exchange in the system.

We present here typical published values of R_1, R_2, NOE, η_{xy} and η_z obtained in a folded protein. Values on Fig. 3.1 were obtained at 600 MHz (14.1T) in the folded enzyme *Escherichia coli* ribonuclease H at 300K (27°C) [151].

Sequence profiles of this protein are remarkably flat, showing rather limited variability of dynamical modes along the sequence. High values of R_2 (around 12) and η_{xy} (around 10) indicate here the dominant contribution of slow motions of the order of several nanoseconds to the relaxation. High NOE values suggest that fast motions are restricted. These observations are in agreement with the dynamical behaviour of a folded protein, where the dominant contribution is that of the global rotational tumbling (with, in this case, the correlation time of about 10 ns).

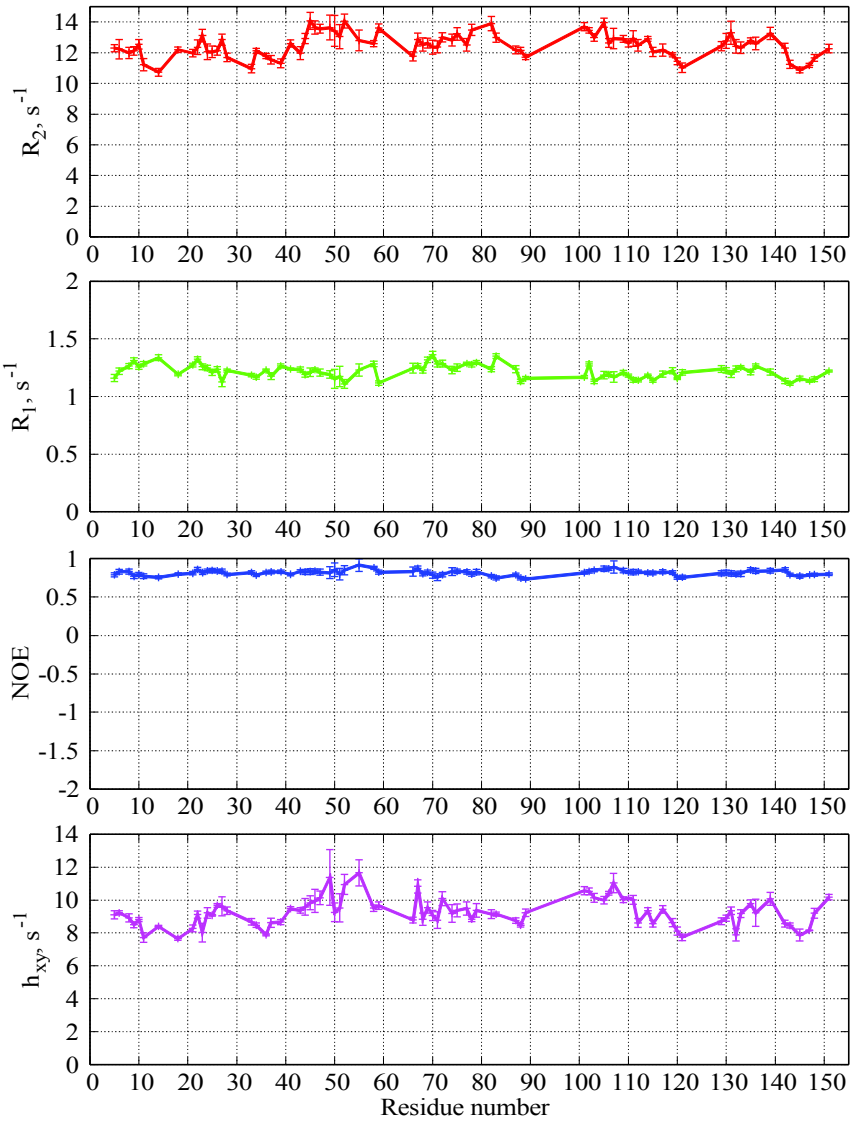


Fig. 3.1. Backbone ^{15}N relaxation rates for ribonuclease H at 14.1T, 300K. Plotted using original data from [151].

A completely different profile of relaxation rates is observed in an intrinsically disordered protein. On Fig. 3.2, we present for the sake of comparison the R_1, R_2, NOE and η_{xy} values obtained in the protein studied in this thesis, the disordered C-terminal domain of the Sendai virus nucleoprotein (N_{TAIL}), at 600 MHz, 298K (25°C). R_2 and η_{xy} are low for the whole chain, not exceeding 4.5 s^{-1} for R_2 and 3 s^{-1} for η_{xy}, except for the region 476-492 which is shown to possess a transient α-helical secondary structure (cf. Chapter 2). All NOEs are smaller than 0.6, and are even negative for a high proportion of residues in the sequence. This behaviour is in agreement with the dominant contribution of fast, picosecond-low nanosecond range motions, which would be expected for a flexible chain (discussed in Chapter 4).

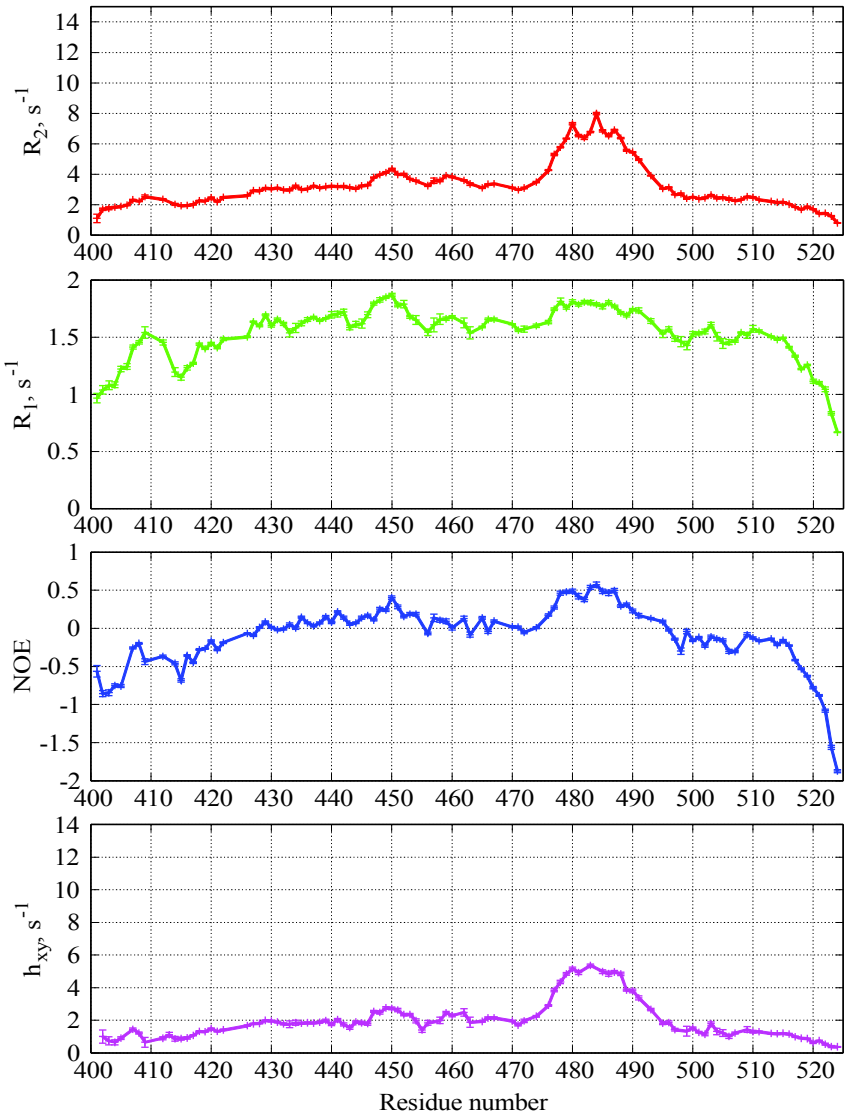


Fig. 3.2. Backbone ^{15}N relaxation rates for N_{TAIL} domain (401-524) of Sendai virus nucleoprotein at 14.1T, 298K. Residue numbers are shown relative to the sequence of the whole nucleoprotein (1-524).

Chapter 4. Unfolded protein dynamics and NMR studies of flexible proteins and IDPs

This chapter starts by the introduction of several models used in polymer physics to describe chain dynamics. Subsequently, I review and compare previous research carried out on the dynamics of flexible, unfolded and intrinsically disordered proteins, and I discuss different models proposed to describe protein dynamics. I present here the theory of the model-free approach to analyse relaxation data and the theory of the reduced spectral density mapping. First attempts to apply model-free analysis to study dynamics in disordered proteins and related problems (existence of the overall motion, timescale separability…) are then discussed. After this, I present different approaches used to avoid or to overcome these problems, including MD simulations, analysing the distribution of timescales of motion as well as adaptations of the model-free approach. I show that there is evidence that an extended model-free analysis can be usefully applied to disordered proteins and that the obtained correlation times may be physically meaningful. In the end of this chapter, I discuss the underlying chain-like nature of disordered proteins in the context of studies carried out by Schwalbe & co-workers, and the dependence of slow motions in disordered proteins on the temperature and the chain length. In the end of this chapter, I summarize the most relevant and important points.

4.1 Description of the chain dynamics by models from polymer physics

Intrinsically disordered proteins that have no structured elements and behave as a so-called 'random coil' can be regarded as heteropolymeric chains of amino acids. In a homopolymer chain, the large-scale properties are not very sensitive to the detailed stereochemical composition of monomers and their potential energy landscape [152]. Therefore, a coarse-grained description may be suitable to describe these large-scale properties, where all chemical specificity could be represented by a few characteristic parameters. This approach seems to be too approximate to describe the detailed dynamics of a disordered polypeptide chain where, in particular, the conformational propensity of each residue strongly depends on its type and that of its neighbours (see the review by M. Blackledge & co-workers [153]). However, it can still be useful to describe some basic properties common to all 'random-coil' polypeptide chains.

The most simple and popular model to describe the chain conformational sampling is the 'random walk' or 'freely joined' chain. The polymer is represented in this model by a sequence of so-called *Kuhn* segments of length b. Each Kuhn segment contains a sufficient number of monomers as to neglect any conformational restrictions on the relative orientation of two neighbouring segments. In a freely joined chain model, all relative

orientations are equiprobable and there is no orientational correlation. If $N >> 1$, then the end-to-end vector of the chain follows the Gaussian distribution, therefore, this model is also sometimes called a 'gaussian chain' model.

The Rouse model is a freely joined chain model in which each segment is represented by a 'bead', in which the segment mass is concentrated, and a massless 'spring', whose elasticity depends on the conformational entropy of the segment. The *Kuhn segment length b* is defined as the root mean square value

$$b \equiv \sqrt{\langle x^2 \rangle} \tag{4.1}$$

of the lengths x of all segments in a chain at a given moment. In the 'worm-like chain' polymer model (not discussed here), this parameter is equivalent to twice the persistence length [154].

The Flory radius R_F is the root mean square value of the chain end-to-end vector $\vec{R}_{end-end}$ in an ensemble of independent chains of N segments:

$$R_F \equiv \sqrt{\langle \vec{R}^2_{end-end} \rangle} = \sqrt{Nb^2} \tag{4.2}$$

where the second equality takes into account the Gaussian distribution of $\vec{R}_{end-end}$. If the sample is isotropic, and no external forces are applied (or they have no impact on the sample order), then naturally $\langle \vec{R}_{end-end} \rangle = 0$.

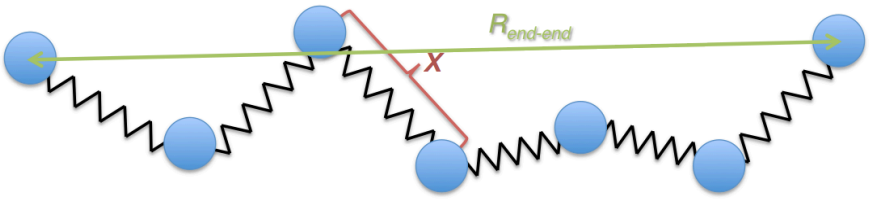


Fig 4.1. Freely joined chain in the Rouse model [152]. For long time and length scales, the polymer may be treated as a chain of beads connected by massless entropic springs. The values x and $\vec{R}_{end-end}$ used in definitions of Kuhn segment length (4.1) and of Flory radius (4.2) are depicted on this figure.

In the Rouse model, the bead-and-spring chain is considered to be in a viscous medium without hydrodynamic backflow and without excluded-volume effects. Its dynamics are described by the Langevin equation generalized for Brownian particles with internal degrees of freedom. Interactions between polymer segments are approximated by entropic harmonic interactions; interactions between different molecules and/or with medium are accounted for by friction and stochastic forces acting on segments. In the description of polymer dynamics using the Rouse model, we follow the account of [152].

The entropic spring constant and the Kuhn's segment friction coefficient are given by, respectively,

$$K = \frac{3k_B T}{b^2} \tag{4.3}$$

and

$$\zeta = 6\pi\eta a_h \tag{4.4}$$

in which k_B is Boltzmann's constant, T is the absolute temperature, η is the medium viscosity and a_h is the hydrodynamic radius of the segment (in the bead-spring model, the radius of the bead). If n is the number of the bead (in the range 0 .. N), $\vec{r}_n$ is the vector pointing to the nth bead centre, $\vec{p}_n(t)$ is its momentum and $\vec{f}_n^L(t)$ is the stochastic Langevin force acting on it, then we can write the equation of motion as:

$$\frac{\partial}{\partial t}\vec{p}_n(t) = K(2\vec{r}_n - \vec{r}_{n+1} - \vec{r}_{n-1}) - \zeta\frac{\partial \vec{r}_n}{\partial t} + \vec{f}_n^L(t) \tag{4.5}$$

In the continuum limit, we can treat n as a continuous variable, and rewrite (4.5) in approximate form as:

$$\frac{\partial}{\partial t}\vec{p}_n(t) \approx K\frac{\partial^2 \vec{r}_n}{\partial n^2} - \zeta\frac{\partial \vec{r}_n}{\partial t} + \vec{f}_n^L(t) \tag{4.6}$$

The acceptable range of n is still (0 .. N) and terminal segments have only a neighbour on one side. Thus, we write the boundary conditions:

$$\frac{\partial}{\partial t}\vec{r}_n(t)\Big|_{n=0,N} = 0 \tag{4.7}$$

In equation (4.6), the left-side term $\frac{\partial}{\partial t}\vec{p}_n(t)$ represents the time derivative of the segment's momentum, or 'inertial force'. It is only important for fast timescales ($t \ll m/\zeta$ (where m is the segment mass), in the picosecond range or faster in most cases) and is irrelevant to explain segmental and large-scale chain motions. Thus,

$$\zeta\frac{\partial}{\partial t}\vec{r}_n(t) \approx \frac{3k_B T}{b^2}\frac{\partial^2 \vec{r}_n(t)}{\partial n^2} + \vec{f}_n^L(t) \tag{4.8}$$

The differential operator $\frac{3k_B T}{b^2}\frac{\partial^2}{\partial n^2}$ represents entropically elastic forces. It has a discrete set of eigenfunctions, which, in the Rouse model, are called *Rouse normal coordinates*:

$$\vec{X}_p(t) = \frac{1}{N}\int_0^N dn \cos\left(\frac{\pi p}{N}\right)\vec{r}_n(t) \tag{4.9}$$

in which p is the normal mode number. In theory, this series of eigenfunctions is infinite; however, modes with $p \geq N$ result from considering non-integer values of n (the number of beads), and are irrelevant for the study of polymer dynamics at larger scale than that of one segment, provided that $N \gg 1$. Equation (4.8) can be solved, using (4.9), by introducing correlation functions $C_{pp'}(t) = \langle \vec{X}_p(t) \cdot \vec{X}_{p'}(0) \rangle$:

$$C_{pp'}(t) = \delta_{pp'} \frac{Nb^2}{2\pi^2 p^2} e^{-\frac{t}{\tau_p}} \tag{4.10}$$

where we used the following property of Langevin stochastic forces acting on the segments n and k at the moments of time t_1 and t_2:

$$\langle f_{\alpha n}^L(t_2) f_{\beta k}^L(t_1) \rangle = 2\zeta k_B T \delta_{ij} \delta(t_2 - t_1)\delta(n-k); \;\; i,j = x,y,z \tag{4.11}$$

Due to the Kronecker symbol $\delta_{pp'}$ in (4.10), $C_{pp'}(t)$ is non zero only if $p = p'$, so in reality for each mode we have only the autocorrelation function:

$$C_p(t) = \langle \vec{X}_p(t) \cdot \vec{X}_p(0) \rangle = \delta_{pp'} \frac{Nb^2}{2\pi^2 p^2} e^{-\frac{t}{\tau_p}} \tag{4.12}$$

The relaxation time of the mode p is defined by:

$$\tau_p = \tau_s \left(\frac{N}{p}\right)^2 \tag{4.13}$$

where $0 \leq p < N$, and τ_s is the relaxation time of the N^{th} Rouse mode, which is the shortest one in the frame of the Rouse model, and is called *segmental relaxation time*. It is given by:

$$\tau_s = \frac{b^2 \zeta}{3\pi^2 k_B T} \tag{4.14}$$

The longest relaxation time is associated with the mode number $p = 1$, which is called *terminal chain relaxation time*, is given by:

$$\tau_1 \equiv \tau_R = \tau_s N^2 \tag{4.15}$$

The relaxation time of the mode p can be redefined as:

$$\tau_p = \tau_1 p^{-2} \tag{4.16a}$$

where:

$$\tau_1 = \frac{b^2 \zeta N^2}{3\pi^2 k_B T} \tag{4.16b}$$

The Rouse model describes well the behaviour of polymer chains in polymer melts [155]. In the case of a dilute solution, however, the hydrodynamic interaction should be properly taken into account. This is done in the *Zimm model* (not discussed here). In the case of a solvent at the so-called Θ (theta) condition (when a polymer behaves in solution as an ideal chain), the distribution of relaxation times of different modes is not given by (4.16 a,b), but rather by the following equation:

$$\tau_p = \tau_1 p^{-\frac{3}{2}} \tag{4.17a}$$

where:

$$\tau_1 = \frac{\eta b^3 N^{\frac{3}{2}}}{\sqrt{3\pi} k_B T} \tag{4.17b}$$

In real solvents, the dependence of the relaxation times on the chain length and the mode number has a form of [155]:

$$\tau_p = \tau_1 p^{-3\nu},\ \tau_1 \sim N^{3\nu} \tag{4.18}$$

where ν may be situated somewhere in the range $1/2$ (Zimm model value) to $2/3$ (Rouse model value). Both Rouse and Zimm models predict one important property of polymers: as the chain gets longer, the distribution of relaxation modes includes slower and slower modes. This property is important in the dynamical behaviour of disordered proteins as discussed below.

4.2 Approaches to study dynamics in intrinsically disordered proteins

This subchapter reviews in more or less chronological order, previous publications on the analysis of protein dynamics by NMR relaxation, and in particular those that concern denatured, (partially) unfolded and intrinsically disordered proteins. The goal is to put the analytical methods used in this thesis in their context.

4.2.1 Theory of the model-free approach

Approaches to analyse NMR relaxation data were initially developed for folded proteins. In 1981, Lipari and Szabo addressed the question of studying side chain dynamics in folded proteins by ^{13}C NMR relaxation. In their article [146], they took a hypothetical example of a lysine side chain attached to a rigid protein backbone experiencing rotational diffusion in solution, and analysed which information on the motions of ^{13}C-H bond vector associated with side chain carbons could be extracted by measuring T_1 relaxation and heteronuclear NOEs at different magnetic fields.

The simplest description of side chain dynamics involves specifying the rate (inverse of time scale) and the spatial restriction of the motion of each ^{13}C-H bond vector, all in the reference frame of the protein backbone. In the first part of their article, the overall motion of the protein backbone is deemed to be isotropic.

Their first assumption is that overall protein motion and internal chain motions are statistically independent. Then the total correlation function (*vide supra*, eq. 3.59) can be rigorously factored as:

$$C(t) = C_O(t)C_I(t) \tag{4.19}$$

where $C_O(\tau)$ is the correlation function for overall motion in the isotropic case:

$$C_O(t) = \frac{1}{5}e^{-\frac{t}{\tau_M}} \tag{4.20}$$

where τ_M is the correlation time of the overall motion.

Because of intra- and interchain steric restrictions, the ^{13}C-H bond vector will not assume all possible orientations in the reference frame rigidly attached to the protein backbone; thus, its motion is restricted. The corresponding correlation function is:

$$C_I(t) = \langle P_2(\hat{\mu}(0) \cdot \hat{\mu}(t)) \rangle \tag{4.21}$$

where $\hat{\mu}$ is a unit orientation vector of the relevant relaxation-active interaction (dipolar or CSA), and $P_2(x) = (3x^2 - 1)/2$. It can be shown, that if the stochastic process governing the internal motion has the Markov property (i.e. it is a memoryless process), then $C_I(t)$ can be expressed as a series of exponentials:

$$C_I(t) = \sum_{i=0} a_i e^{-t/\tau_i} \tag{4.22}$$

where $\tau_0 = \infty, \tau_1 > \tau_2 > \tau_3 \ldots$ and $0 \leq a_i \leq 1$ for all *i*. The exact number of exponentials, amplitudes a_i and correlation times τ_i depend on the exact nature of the internal motion.

As A. Szabo explained later in his autobiography [156], by the time he and his graduate student G. Lipari were doing the work described in their article [146], it was common to use particular models (e.g. *N*-site jump, free or restricted rotational diffusion about an axis…) to describe the internal motion. Then G. Lipari found that he could analyse experimental data equally well using different physical models, and that the parameters of more complex models could not be uniquely determined. This observation led them to describe a number of properties of the internal correlation function, rather than using a particular model. That is why they called their approach "model-free".

The simplest property of the internal correlation function is:

$$C_I(0) = \langle P_2\big(\hat{\mu}(0) \cdot \hat{\mu}(0)\big)\rangle = 1 \tag{4.23}$$

Using the property of correlation functions that $\lim_{t\to\infty}\langle A(0)B(t)\rangle = \langle A\rangle\langle B\rangle$, and that $P_2(\hat{\mu}_1 \cdot \hat{\mu}_2) = \frac{4\pi}{5}\sum_{m=-2}^{2} Y_{2m}(\Omega_1)Y_{2m}^*(\Omega_2)$, we can write:

$C_I(\infty) = \mathcal{S}^2$, where

$$\mathcal{S}^2 = \frac{4\pi}{5}\sum_{m=-2}^{2}|\langle Y_{2m}(\Omega)\rangle|^2 = \int d\Omega_1 d\Omega_2 p_{eq}(\Omega_1)P_2\big(\cos\big(\widehat{\hat{\mu}_1\hat{\mu}_2}\big)\big)p_{eq}(\Omega_2) \tag{4.24}$$

in which $\langle \ldots \rangle$ is the equilibrium average and $p_{eq}(\Omega)$ is the normalized equilibrium distribution function of $\hat{\mu}$. Lipari and Szabo called the quantity $\mathcal{S}$ a generalized order parameter. According to the authors, $\mathcal{S}$ is a model-independent measure of the degree of spatial restriction of the internal motion; it satisfies the inequality $0 \leq \mathcal{S}^2 \leq 1$, is equal to 0 if internal motion is isotropic (and hence unrestricted), and is equal to 1 if the side-chain is fully rigid in the molecular frame. As will be discussed later, this definition may not be entirely correct in some cases [157].

The name "generalized order parameter" is justified by considering specific cases of the equation (4.24). If the internal motion is axially symmetric, i.e. p_{eq} is a function of theta only ($p_{eq}(\Omega) = p_{eq}(\theta)$), where θ is the angle between $\hat{\mu}$ and the symmetry axis, then the generalized order parameter reduces to the usual order parameter *S* defined as:

$$\mathcal{S} = S = \langle P_2(\cos(\theta))\rangle \tag{4.25}$$

Next, Lipari and Szabo proposed an approximation of the internal correlation function $C_I(t)$. A schematic representation of a correlation function fulfilling the conditions (4.23) and (4.24) is shown in the Fig. 4.2.

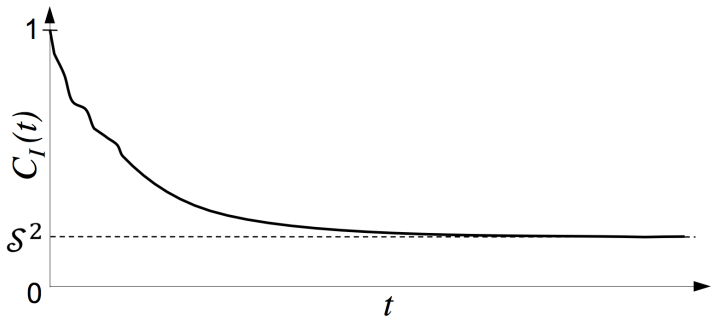


Fig 4.2. Schematic description of the internal motion correlation function.

Taking into account (4.22), the simplest approximation is that of the mono-exponential decay from $C_I(0)$ to $C_I(\infty)$:

$$C_I^{appr}(t) = \mathcal{S}^2 + (1 - \mathcal{S}^2)e^{-t/\tau_e} \tag{4.26}$$

The authors justify the choice of mono-exponential function for the simplest model by the fact that, for the diffusion in a cone model, where $C_I(t)$ is exactly given by infinite sum of exponentials, this approximation is very good when $\mathcal{S}^2$ is not too small [158, 159], and that the effective correlation time τ_e is determined by requiring the integrals from 0 to ∞ for $C_I^{appr}(t)$ and $C_I(t)$ to be equal:

$$\int_0^\infty \left(C_I^{appr}(t) - \mathcal{S}^2\right)dt = \int_0^\infty (C_I(t) - \mathcal{S}^2)dt \tag{4.27}$$

which gives:

$$\tau_e = \frac{\int_0^\infty (C_I(t) - \mathcal{S}^2)dt}{(1 - \mathcal{S}^2)} \tag{4.28}$$

It is emphasized that this approximation is not in any case a truncation of series (4.22) to two terms, but rather a simplest form of a time-dependent Padé approximant. If $C_I(t)$ is expanded as in (4.22), then τ_e can be calculated as:

$$\tau_e = \frac{1}{1 - \mathcal{S}^2}\sum_{i=1} a_i e^{-\frac{t}{\tau_i}}, \qquad a_0 = \mathcal{S}^2 \tag{4.29}$$

The “model-free” spectral density function, derived from the correlation function (4.26), is given by the formula:

$$J(\omega) = \frac{2}{5}\left(\frac{\mathcal{S}^2\tau_M}{1 + (\omega\tau_M)^2} + \frac{(1 - \mathcal{S}^2)\tau_e'}{1 + (\omega\tau_e')^2}\right) \tag{4.30}$$

where $\tau_e' = (\tau_M^{-1} + \tau_e^{-1})^{-1}$. In the extreme narrowing limit $(\omega\tau_e')^2 \ll 1, \tau_e \ll \tau_M$, one obtains:

$$J(\omega) = \frac{2}{5}\left(\frac{\mathcal{S}^2\tau_M}{1+(\omega\tau_M)^2} + (1-\mathcal{S}^2)\tau_e\right) \tag{4.31}$$

Lipari and Szabo show that the single exponential approximation of internal correlation function is exact (i.e. will fit the data) if the following conditions are fulfilled:

1) the overall motion is isotropic,
2) the internal motion is much faster than overall motion, $\tau_i \ll \tau_M$,
3) the extreme narrowing limit applies (in (4.22), $\forall i$: $(\omega\tau_i)^2 \ll 1$), and
4) τ_e is defined as in (4.28)

As reviewed later by B. Halle [157], if the internal motion is much faster than the overall motion ($\tau_i \ll \tau_M$), then these motions are statistically independent, and it is not necessary to require this condition explicitly (as it was required for equation (4.19)). However, the inverse affirmation is not always true.

To prove this, the equation (4.22) can be rewritten as

$$C_I(t) = \mathcal{S}^2 + \textstyle\sum_{i=1} a_i e^{-t/\tau_i} \tag{4.32}$$

then the corresponding spectral density function is:

$$J(\omega) = \frac{2}{5}\left(\frac{\mathcal{S}^2\tau_M}{1+(\omega\tau_M)^2} + \sum_{i=1}\frac{a_i(\tau_M^{-1}+\tau_i^{-1})^{-1}}{1+(\omega(\tau_M^{-1}+\tau_i^{-1})^{-1})^2}\right) \tag{4.33}$$

and in the extreme narrowing limit $(\omega\tau_i)^2 \ll 1, \tau_i \ll \tau_M$ it can be written as:

$$J(\omega) = \frac{2}{5}\left(\frac{\mathcal{S}^2\tau_M}{1+(\omega\tau_M)^2} + \sum_{i=1}\frac{a_i\tau_i}{1+(\omega\tau_i)^2}\right) = \frac{2}{5}\,\frac{\mathcal{S}^2\tau_M}{1+(\omega\tau_M)^2} + \frac{2}{5}\sum_{i=1} a_i\tau_i \tag{4.34}$$

Using the relationship (4.29), one obtains that (4.34) is equal to (4.30), thus proving that the single exponential approximation of internal correlation function is mathematically exact.

The practical meaning of this is that, regardless of the complexity of the internal motion and of the number of exponentials in the correlation function, if all relevant timescales

lay in the extreme narrowing limit, it is impossible to extract from experimental data individual amplitudes and correlation times, because the mono-exponential representation of internal correlation function is mathematically sufficient.

Lipari and Szabo also treated the case when some internal motions are much slower than the overall motion, for example, $\tau_1 \gg \tau_M$, but $\tau_2, \tau_3 \ldots \gg \tau_M$, then, using (4.33), one can formally obtain:

$$J(\omega) = \frac{2}{5}\frac{\mathcal{S}'^2\tau_M}{1+(\omega\tau_M)^2} + \frac{2}{5}\sum_{i=2} a_i\tau_i\,, \qquad \text{where } \mathcal{S}'^2 = \mathcal{S}^2 + a_1 \tag{4.35}$$

so the information on the slower motions is incorporated in $\mathcal{S}'^2$, which becomes a measure of the degree of internal motion restriction on a time scale determined by the overall motion.

However, if one of the timescales of the internal motion τ_i is close to τ_M, then the independence of internal and overall motion may be questioned (*vide supra*). Even if the separation of overall and internal motion still holds true, the mono-exponential form of $C_I(t)$ is not exact anymore, so the "model-free" approach as given by the equation (4.30) for the spectral density function may be not appropriate. By testing their approach on the relaxation data generated using different physical models, Lipari and Szabo summarize that $\mathcal{S}^2$ and τ_e are virtually exact if internal (but not overall) motions are close to the extreme narrowing limit ($\tau_i \ll (\omega_H + \omega_C)^{-1} = 350$ ps for a 360MHz NMR spectrometer), and are reasonably accurate (~25%) as long as $\omega\tau_e < 0.5$ and $\mathcal{S}^2 > 0.01$.

Lipari and Szabo also evoked previous approaches to analyse NMR relaxation data. In particular, in the works of Jardetzky et al. [160–162], it had been already proposed, assuming that motions in a protein are governed by some general Markov process, to express the spectral density function in terms of eigenfunctions and eigenvalues $-\lambda_n$ of the transition operator [160, 161] (in these articles, the $\frac{1}{5}$ factor in the spectral density function is incorporated in the equations for the relaxation rates):

$$J(\omega) = \frac{2}{5}\sum_n \frac{\alpha_n\lambda_n}{\lambda_n^2 + \omega^2} \tag{4.36}$$

The total correlation function is then written as:

$$C_I(t) = \sum_n \alpha_n e^{-\lambda_n t} \tag{4.37}$$

However, when the authors tried to implement this approach in practice, they obtained non-physical results, as pointed out by Lipari and Szabo, contributing to the lack of popularity of their approach compared to the model-free approach.

In a subsequent article by Clore and Szabo [163], an extension of the model-free analysis is discussed. A simple two-parameter model-free approach was unable to account for ^{15}N T_1, T_2 and heteronuclear NOE relaxation data simultaneously for some backbone amide groups in staphylococcal nuclease (SNase) and interleukin-1β. They explain this discrepancy by the nonexponential character of the internal correlation function, due to the presence of slower components not anymore in the extreme narrowing limit. They propose the simplest form of the internal correlation function that can describe such systems:

$$C_I(t) = \mathcal{S}^2 + A_f e^{-t/\tau_f} + A_s e^{-t/\tau_s}, \text{where} \quad \mathcal{S}^2 + A_f + A_s = 1 \qquad (4.38)$$

If the fast internal motion correlation time τ_f is at least an order of magnitude smaller that that of the slower internal motion, τ_s, then the correlation function $C_I(t)$ reaches a plateau value of $1 - A_f$ before levelling off at $\mathcal{S}^2$ due to slower motions (see Fig. 4.3) Indeed, as shown by Halle [157] (*vide infra*), motions that have very different correlation times contribute additively to the correlation function, and are in fact statistically independent. They interpret $1 - A_f$ as a generalized order parameter for fast motions, $\mathcal{S}_f^2$. Then one can write the internal correlation function as:

$$C_I(t) = \mathcal{S}^2 + (1 - \mathcal{S}_f^2)e^{-t/\tau_f} + (\mathcal{S}_f^2 - \mathcal{S}^2)e^{-t/\tau_s} \qquad (4.39)$$

and the spectral density function as:

$$J(\omega) = \frac{2}{5}\left(\frac{\mathcal{S}^2\tau_R}{1 + (\omega\tau_R)^2} + \frac{(\mathcal{S}_f^2 - \mathcal{S}^2)\tau_s'}{1 + (\omega\tau_s')^2} + \frac{(1 - \mathcal{S}_f^2)\tau_f'}{1 + (\omega\tau_f')^2}\right) \qquad (4.40)$$

where $\tau_s' = (\tau_R^{-1} + \tau_s^{-1})^{-1}, \tau_f' = (\tau_R^{-1} + \tau_f^{-1})^{-1}$, and τ_R is the overall rotational correlation time.

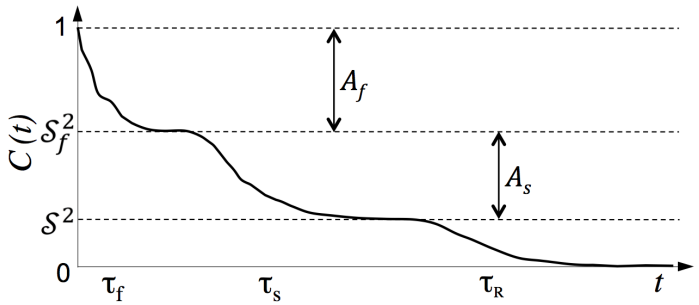


Fig 4.3. Schematic description of the extended model-free correlation function.

In the case when the fast internal motion is axially symmetric (and, consequently, $\mathcal{S}_f^2 = S_f^2$ (usual order parameter)), they also define the generalized order parameter for slow motion $\mathcal{S}_s^2$ by decomposing the total generalized order parameter as:

$$\mathcal{S}^2 = S_f^2 \mathcal{S}_s^2 \tag{4.41}$$

Finally, they also propose a simplified version of extended model-free approach, making the assumption that τ_f is sufficiently small and makes a negligible contribution to the spectral density function, so the third term in (4.40) is almost equal to zero:

$$J(\omega) = \frac{2}{5}\left(\frac{\mathcal{S}^2\tau_R}{1+(\omega\tau_R)^2} + \frac{(\mathcal{S}_f^2 - \mathcal{S}^2)\tau_s'}{1+(\omega\tau_s')^2}\right) \tag{4.42}$$

The advantage of this simplification is that it offers a more stable fit of other parameters by sacrificing τ_f.

By comparing parameters fitted using the original (4.30) and the simplified extended (4.42) model-free approach, they find that the order parameter $\mathcal{S}^2$ is essentially the same in two cases, but values of effective internal motion correlation time τ_e calculated by classic model-free approach (0.2-0.3 ns) are smaller by an order of magnitude than the values of the slower internal motion component τ_s (1-3 ns) calculated using (4.42). These slower motions are still faster than the overall motion (8-9 ns) and can have significant amplitude (their contribution is up to 30-40% in the studied case).

4.2.2 Theory of reduced spectral density mapping

Another classical method to analyse relaxation data is to use spectral density mapping. Instead of assuming a precise form of the spectral density function, as in the model-free approach, the idea is to extract (or map) the values of $J(\omega)$ at the frequencies to which relaxation rates are sensitive (see Fig. 4.4). The method is complicated by the fact that the 'classic' relaxation parameters (R_1, R_2, heteronuclear NOE) depend on $J(\omega)$ at five frequencies: $0, \omega_S, \omega_I, \omega_I - \omega_S, \omega_I + \omega_S$ (I is ^{1}H and S is ^{15}N for ^{15}N relaxation) so it is not mathematically possible to extract 5 independent values of $J(\omega)$ using 3 rates. In the original approach [164], the full spectral density mapping, this was resolved by measuring additional relaxation rates, antiphase coherence transverse relaxation $T_2(2H_ZN_{X,Y})$ and longitudinal two-spin order relaxation $T_1(2H_ZN_Z)$ (not related to cross-correlated relaxation rates at the bottom of Fig. 4.4). The problem is that first, the latter relaxation rates are not correctly described by a single exponential decay, as assumed in this approach, and second, calculations involve taking a difference between rates of similar magnitude and therefore are very sensitive to experimental errors.

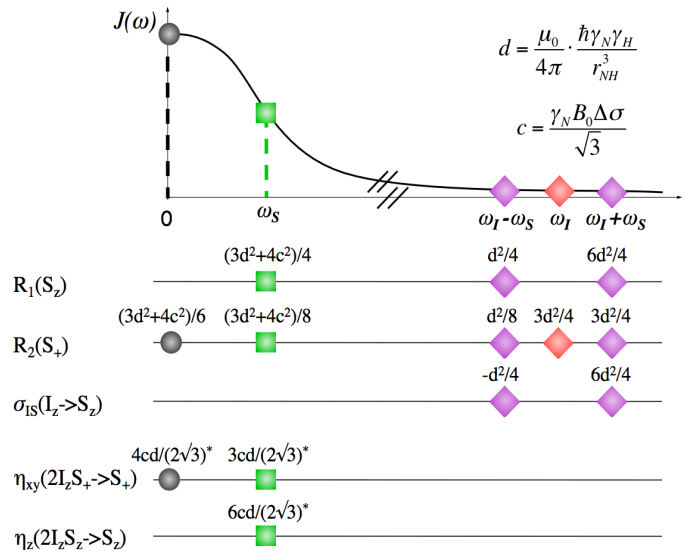


Fig 4.4. Dependency of relaxation rates used in this thesis on the spectral density function. *: cross-correlated relaxation rates, η_{xy} and η_z, depend on the cross-correlated spectral density function (see Section 3.4, *Interference between relaxation mechanisms*), which is different from the spectral density function used in equations for non-cross-correlated relaxation rates (rates R_1, R_2, σ_{IS}).

Taking into account that the value of the spectral density function $J(\omega)$ is expected to be similar for high frequencies $\omega_H - \omega_N, \omega_H, \omega_H + \omega_N$ in folded biological macromolecular in solution, Farrow et al. [10] proposed, instead of mapping $J(\omega)$ at 5 frequencies, to approximate the linear combination of high-frequency terms $J(\omega_H - \omega_N)$, $J(\omega_H), J(\omega_H + \omega_N)$ in the equations for R_1, R_2 and NOE with a single equivalent term, thus reducing the number of parameters to 3.

Farrow et al. proposed to take simply $J(\omega_H) \approx J(\omega_H \pm \omega_N)$ [165], or, more recently, to approximate high-frequency values of $J(\omega)$ as:

$$J(\omega) = \frac{\lambda_1}{\omega^2} + \lambda_2, \tag{4.43}$$

where, in the context of folded proteins, the first term represents slow overall motions and the second represents fast internal motions [10]. This expression is exact when $J(\omega)$ is a sum of Lorentzians

$$J(\omega) = \sum_i \frac{a_i \tau_i}{1 + (\omega\tau_i)^2}, \quad \text{and for all } \tau_i: \text{ either } \omega_H \tau_i \gg 1, \text{ or } \omega_H \tau_i \ll 1 \tag{4.44}$$

which is the case for example in folded proteins with slow overall motion and fast internal motions. The next step is to simplify the high frequency terms in the equation for the NOE (σ_{IS} rate) by requiring that:

$$6J(\omega_H + \omega_N) - J(\omega_H - \omega_N) = \mathrm{A}J(\omega_q) \tag{4.45}$$

where assuming (4.43), one obtains A = 5 and $\omega_q = 0.870\omega_H$. By applying the same assumption to simplify the linear combination of $J(\omega)$, high-frequency terms in the expressions for the relaxation rates represented in Fig. 4.4, one obtains:

$$6J(\omega_H + \omega_N) - J(\omega_H - \omega_N) = 5J(0.870\omega_H) \quad (4.46a)$$
$$J(\omega_H - \omega_N) + 6J(\omega_H + \omega_N) = 7J(0.921\omega_H) \quad (4.46b)$$
$$J(\omega_H - \omega_N) + 6J(\omega_H) + 6J(\omega_H + \omega_N) = 13J(0.955\omega_H) \quad (4.46c)$$

The equations for the relaxation rates and NOE can thus be written as:

$$NOE = 1 + \left(\frac{d^2}{4}\frac{\gamma_H}{\gamma_N}\right)[5J(0.870\omega_H)]\frac{1}{R_1} \quad (4.47)$$

$$R_1 = \frac{d^2}{4}[3J(\omega_N) + 7J(0.921\omega_H)] + c^2 J(\omega_N) \quad (4.48)$$

$$R_2 = \frac{d^2}{8}[4J(0) + 3J(\omega_N) + 13J(0.955\omega_H)] + \frac{c^2}{6}[4J(0) + 3J(\omega_N)] \quad (4.49)$$

Numerical simulations to test the range of validity of (4.43) when Lipari-Szabo model-free form of $J(\omega)$ is assumed (4.30), reveal that for the values typically obtained in folded proteins, 0.001 ns $\leq \tau_e \leq$ 1 ns for internal motion, $\tau_m \geq$ 2 ns for overall motion and $0.1 \leq \mathcal{S}^2 \leq 1$, the two sides of Eq.4.47-4.49 differ always by less than 8%.

The value of $J(0.870\omega_H)$ can be directly extracted from (4.47). The values of $J(0.921\omega_H)$ and $J(0.955\omega_H)$ can be estimated from $J(0.870\omega_H)$ using one of the methods below. Once these values are known, it is straightforward to extract $J(\omega_N)$ and $J(0)$ using (4.48) and (4.49).

To estimate values of $J(0.921\omega_H)$ and $J(0.955\omega_H)$, the authors proposed the use of the original approximation for $J(\omega)$ at high frequencies (4.43) in its two extreme cases: $J(\omega) = const$ (Method 1) and $J(\omega) \propto \frac{1}{\omega^2}$ (Method 2), corresponding to the slowest and the fastest rate of decay of $J(\omega)$, respectively, consistent with a Lorentzian form of the spectral density function. In a third method, if data sets (R_1 and NOE at least) are obtained at two different fields, they propose to extrapolate the values of $J(0.921\omega_H)$ and $J(0.955\omega_H)$ using the following equation:

$$J(\varepsilon\omega_H) \approx J(0.870\omega_H) + (\varepsilon - 0.870)\omega_H \left.\frac{dJ(\omega)}{d\omega}\right|_{\omega=0.870\omega_H} \quad (4.50)$$

where $\left.\frac{dJ(\omega)}{d\omega}\right|_{\omega=0.870\omega_H}$ can be approximated from measurements at two field strengths:

$$\left.\frac{dJ(\omega)}{d\omega}\right|_{\omega=0.870\omega_H} = \frac{J(0.870\omega_H) - J(0.870\omega_H')}{0.870(\omega_H - \omega_H')} \tag{4.51}$$

where ω_H' is the Larmor frequency of proton at the second field strength.

Farrow et al. applied their approach to study dynamics in a 59-amino acid N-terminal SH3 domain of the *Drosophila* signal transduction protein drk (drkN SH3). Studies were carried out with drkN SH3 dissolved in 2M guanidine hydrochloride, in which the protein is completely denatured, and in 0.4M sodium sulphate, which stabilizes the folded form of the protein. Values of $J(0.87\omega_H)$ are positive and $J(\omega)$ decreases with ω, as expected. Moreover, in the disordered form of drkN SH3, $J(\omega)$ is higher at $0.87\omega_H$ and decreases faster with ω than in the folded form, where it is already low and does not decrease as rapidly. That is consistent with the more significant contribution of faster timescale motions in the disordered form of drkN SH3.

Simulation of R_1, R_2 and NOE data generated using different motional regimes of the model-free form of $J(\omega)$ ($\tau_m = 5$ ns; $\mathcal{S}^2 = 1, \tau_e = 0$; $\mathcal{S}^2 = 0.8, \tau_e = 50$ ps; $\mathcal{S}^2 = 0.5, \tau_e = 500$ ps), with measurement errors of R_1, R_2 and NOE assumed to be 2%, 2% and 5% respectively, the authors found that the difference between the spectral density function values extracted using different models (1 or 2, *vide supra*) is not bigger than the uncertainty due to experimental errors. Moreover, by comparing $J(\omega_N)$ and $J(0)$ values extracted using different methods (1, 2 and 3) from relaxation data of the folded and unfolded form of drkN SH3, the authors found that, despite different dynamics observed in these forms, the method of the estimation of $J(0.921\omega_H)$ and $J(0.955\omega_H)$ has little impact on the accuracy of $J(\omega_N)$ and $J(0)$ determination.

Conformational exchange on 100ns-ms timescale may contribute to the measured R_2 value (4.53). Thus extracted $J(0)$ values may be wrong, because one should use 'pure' R_2 in (4.49), $R_2^0 = R_2 - R_2^{ex}$. However, in the 'fast' exchange regime, R_2^{ex} is proportional to the square of magnetic field B_0^2, and it is possible to extract the value of its contribution by measuring R_2 at different fields and correct the $J(0)$ values. Including R_2^{ex} of 0.5 s^{-1} in simulated relaxation data results in overestimated $J(0)$ values, up to 15%, however by using a two-field R_2 correction for $J(0)$, its value can be measured to within 1% of error.

In summary, Farrow et al. proposed an alternative approach to model-free analysis, in which no assumption is required on the explicit form of the spectral density function nor on the isotropicity of the overall tumbling. By comparing obtained values of $J(\omega)$, we can obtain some insight into the timescale range of motions that are dominating the relaxation. If nanosecond-range motions are dominant, one observes a higher $J(0)$ value, a rapid drop of the spectral density function between $J(0)$ and $J(\omega_N)$, and very low $J(\omega_H)$ values. If faster picosecond-range motions dominate, $J(0)$, $J(\omega_N)$, and $J(\omega_H)$ values are

similar. However, a significant drawback of this approach is that it does not provide any physical insight, such as timescales and relative amplitude of motions.

4.2.3 Application of classic or extended model-free approaches to study dynamics in unfolded and disordered proteins

One of the first studies where the model-free approach was applied to an unfolded protein is that of a 131 residue fragment of the staphylococcal nuclease (SNase) protein, Δ131Δ by Alexandrescu and Shortle [166], shown previously to possess little residual structure even in the absence of denaturants. Three relaxation parameters were measured at 14.1T (600MHz) and at 32°C: ^{15}N T_1, T_2 and ^{15}N-^{1}H heteronuclear NOE.

The relaxation data confirm the disordered nature of Δ131Δ (R_2 values around 6s^{-1} and negative NOEs), except one contiguous segment between residues 79 and 115, which gives almost entirely positive NOEs and R_2 values above 6s^{-1}. This region is called a 'restricted segment' in the article.
Three models were proposed to fit the relaxation data. In the first model the correlation time $\tau_{m,loc}$ is fitted for each residue independently (formally, as if segments were one-residue length), but it still represents an effective correlation time of the bead (segment) to which belongs the residue. The spectral density function is then given by the equation

$$J(\omega) = \frac{2}{5}\frac{\tau_{m,loc}}{1+\left(\omega\tau_{m,loc}\right)^2} \tag{4.52}$$

Most residues had an effective correlation time $\tau_{m,loc}$ around 1 ns, which was compared to 0.5 ns found in previous publications in temperature-denatured collagen and acid-denatured ribonuclease, and, more importantly, to 0.6 ns segmental correlation time in atactic polystyrene [167]. Residues in the 'restricted segment' had effective correlation times from 4 to 7.5 ns, suggesting greater structural organization in this section than that of a random coil. The major problem of this model is that it reproduced experimental data very poorly, indicating that although unfolded, the protein could not be characterised by an ultra-simplistic motional model. This result was subsequently reproduced in essentially all characterisation of disordered proteins by NMR relaxation.

In order to satisfactorily fit their experimental data, the authors found it necessary to use either the extended model-free approach (Eq. (4.42)) introduced by Clore et al. for folded proteins, or the standard model-free approach with an additional parameter accounting for an exchange contribution to the R_2 rate, although the available experimental data did not allow them to determine the statistical significance of this model.

The authors note that in unfolded proteins, slower internal motions on the order of 1 ns make a greater contribution to relaxation than in folded proteins; in the latter, however, these motions could still be found to be significant in disordered segments, such as loops and termini.

Finally, they show a weak correlation between S^2, hydrophobicity and helix propensity. They explain this correlation by assuming that in a disordered protein, residues with higher propensity for hydration are likely to spend more time in flexible solvent-exposed conformations, and, by contrast, residues involved in hydrophobic regions of the chain are more likely to be involved in more compact chain conformations with more restricted motions giving rise to higher S^2 values.

This pioneering study, although using minimal data sets and appearing somewhat naïve in interpretation, actually crystallizes many of the recurrent problems that have plagued the interpretation of relaxation from disordered proteins, and that remain even when measuring multiple relaxation rates from many different magnetic field strengths.

Frank et al. [168] studied in 1995 the dynamics in the urea-denatured B1 immunoglobulin binding domain of streptococcal protein G (GB1) by model-free analysis. ^{15}N relaxation rates R_1, R_2 and heteronuclear NOE were measured at 500 and 600MHz at 25°C. Model-free parameters were quite uniform and did not show any evident segments of restricted motions. The average value of S^2 was 0.49±0.04, in sharp contrast with the range 0.05-0.8 obtained in Δ131Δ by Alexandrescu and Shortle. Tge authors conclude that urea-denatured GB1 is a closer example of a random-coil unfolded protein than Δ131Δ.

The data were also analysed using an extended model-free approach, with three correlation times – one for slower motions and two for faster motions – fitted as global values for all residues to values of 2.92±0.06 ns, 75±20 ps and 400±80 ps, respectively. Values of S^2 are very close to those obtained using the simple model-free approach (0.46±0.03 on average), and the values of S_f^2 are relatively uniform with an average value of 0.59±0.07. This result is consistent with the proposition by Alexandrescu and Shortle that in unfolded proteins a significant contribution to measured relaxation comes from motions on timescales around 1 ns.

Another study of a partially unfolded protein is presented in the article of Brutscher et al. [169] on the partially folded A state of ubiquitin in 60%/40% methanol/water mixture at pH 2 and 27°C. Different relaxation parameters were measured: R_1, R_2 and NOE at 600 MHz, R_1 and NOE at 800 MHz and R_1 and R_2 at 400 MHz.
This more extensive set of data allowed the authors for the first time to compare the classical model-free analysis and the extended model-free analysis on a residue specific basis in a disordered protein. The comparison of fit quality yielded a clear preference for

the extended model-free analysis with all three correlation times being fitted. The correlation function used by the authors is given by

$$C(t) = e^{-|t|/\tau_c}\left\{\mathcal{S}^2 + A_1 e^{-|t|/\tau_1} + A_2 e^{-|t|/\tau_2}\right\}, \mathcal{S}^2 = 1 - A_1 - A_2 \qquad (4.54)$$

Internal motions τ_1, τ_2 are fitted with their amplitudes of contribution A_1, A_2, instead of using specific order parameter for internal motions ($\mathcal{S}_f^2 = 1 - A_2$).

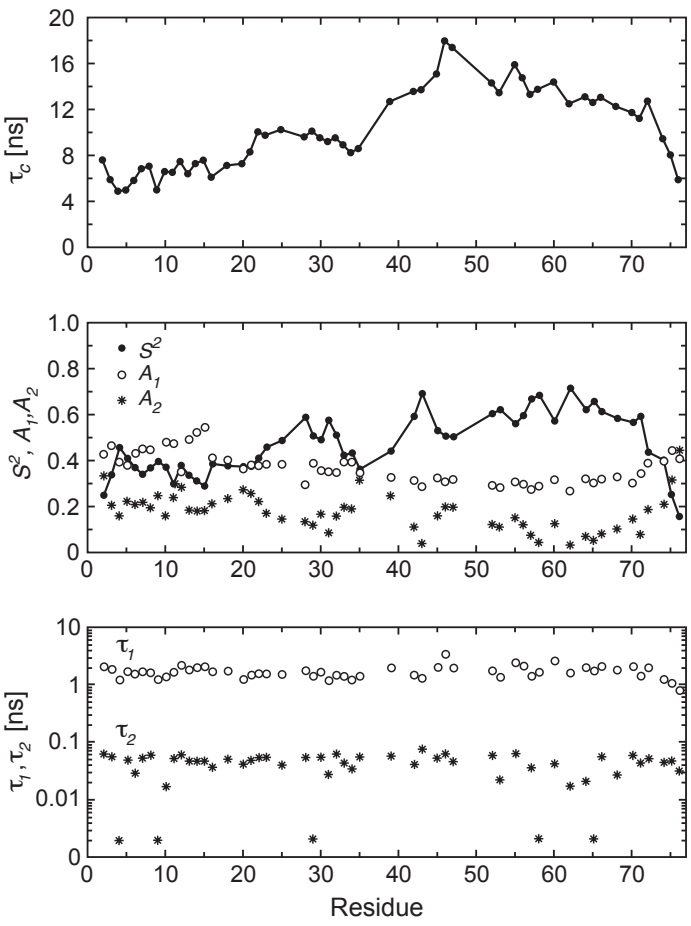


Fig. 4.5. Model-free analysis of dynamics in A state of ubiquitin.

Obtained parameters are shown in the Fig. 4.5. Order parameter $\mathcal{S}^2$ was found to be unusually low (0.4-0.6) compared to values typically observed in folded proteins (0.8 to 0.85). The sequence profile of global motion correlation time τ_c clearly distinguishes three separate secondary structure elements. The authors propose that these protein segments are flexibly linked and their dynamics is governed not by the rotational tumbling of the whole molecule, but by the rotational tumbling of a segment.

In contrast, the sequence profile of timescales and amplitudes of internal motions is quite flat. Two internal motion timescales are separated by more than 1 order of magnitude: fast internal motions (τ_2) lie in the range 1-60 ps with an average of 40 ps, and their amplitude A_2 range between 0.1 and 0.2. Slower internal motions (τ_1) have amplitudes A_1 in the range 0.3–0.5 and the corresponding correlation time is in the range 1–3 ns, similar to that found in the most unstructured regions of TFE-unfolded lysozyme (study by Buck et al. [170]). The authors attribute the fastest timescale τ_2 to internal motions, and τ_1 to slow internal motion, that may be influenced by strong tumbling anisotropy of protein segments.

It was noted that inclusion of a second internal correlation time yielded an increase in the overall correlation time τ_c and a decrease in the generalized order parameter S^2, compared to the classical model-free approach. As will be discussed later in Section 4.2.5, this observation may be explained by an underlying non-discrete distribution of motional correlation times, and fitting two or three correlation times would lead to different sampling of this distribution.

Almost simultaneously, Farrow et al. compared relaxation data (R_1, R_2, and heteronuclear NOE) collected at 500 and 600 MHz at 30°C and 14°C in denatured (in 2M guanidine hydrochloride) and folded (in 0.4M Na_2SO_4) drkN SH3. This allowed the authors to study the dependence of the values obtained from reduced spectral density mapping and model-free analysis on the temperature for both folded and unfolded forms of drkN SH3.

For both states of the protein, $J(0)$ at 30°C is 0.7 times smaller than that determined at 14°C. Measurements of solvent viscosity of both protein samples reveal that an increase of temperature from 14°C to 30°C is accompanied by a reduction in viscosity by a factor of 0.7 suggesting that $J(0)$-related slow dynamics in the unfolded form are also viscosity-dependent.

Model-free analysis with a global overall correlation time τ_m, residue-specific order parameter S^2 and residue-specific internal motion correlation time τ_e was carried out on the relaxation data of both drkN SH3 forms at both temperatures. The values of S^2 for the folded form are almost temperature-independent. However, for the denatured state, the average S^2 value falls from 0.51 to 0.41 upon increasing the temperature from 14°C to 30°C. The authors propose that in the folded state the structure is stabilized by many interactions. As the temperature of the folded sample is increased some of the interactions will likely be lost; however, those remaining will suffice to maintain the folded structure of the protein and therefore limit the dynamics. This cooperative stabilization is not present in the unfolded state and therefore increased temperatures is expected to result in higher levels of motions within the protein.

Similarly, Yang et al. [171] observed the same temperature-dependence of order parameters in folded staphylococcal nuclease SNase (order parameters constant upon temperature changes) and in its disordered segment Δ131Δ (they decrease with temperature). They suggest that this difference is due to the heat capacity of the folded and unfolded chains.

Previously, the authors had developed a theory relating conformational entropy to S^2 [172]. They remark that previously 'entropy-order parameter' profiles calculated from a 1 ns molecular dynamics trajectory of (folded) RNase H1 were remarkably well fit assuming a model in which individual bond vectors diffuse in a cone, and use the same

relationship to describe the difference of $\mathcal{S}^2$ behaviour with temperature in folded and unfolded drkN SH3, SNase and Δ131Δ. If $\Delta S_p(q)$ is the change in conformational entropy of a bond vector q between states a and b, $S_a = \sqrt{S_a^2}$ is the order parameter of the bond vector in state a and S_b corresponds to the state b, then:

$$\Delta S_p(q) = k \cdot ln \frac{3 - \sqrt{1 + 8S_b}}{3 - \sqrt{1 + 8S_a}}. \tag{4.55}$$

The change of entropy with temperature is described by

$$\Delta S = \int C_p \, \mathrm{dln}T, \tag{4.56}$$

from which one obtains the heat capacity:

$$C_p = \frac{\Delta S}{\ln\left(\frac{T^b}{T^a}\right)}. \tag{4.57}$$

Thus, the observed difference in the temperature behaviour of $\mathcal{S}^2$ in folded and denatured proteins suggests that denatured proteins have a larger heat capacity than folded proteins. However, the authors remark that the obtained difference is bigger than what is observed experimentally by calorimetry. Indeed, ^{15}N relaxation provides only information on the backbone dynamics, so the obtained difference is only indirectly related to the total heat capacity. Based on side chain relaxation studies in folded drkN SH3, they suggest that side chain dynamics contribute more to the heat capacity of this protein than backbone motions, at least in the range from 14°C to 30°C.

Early model-free studies of dynamics in proteins involved measuring NMR relaxation at one - three fields. Roumestand & co-workers in 2001 measured ^{15}N R_1, R_2 relaxation rates and ^{15}N-^{1}H heteronuclear NOEs at five different fields and analysed these data to characterize dynamics in C12A-p8$^{\mathrm{MTCP1}}$, a protein that contains disordered termini [173]. The authors fitted the relaxation data using the following form of the spectral density function:

$$J(\omega) = A + \frac{B}{1 + \omega^2 \tau_s^2} \tag{4.58}$$

in which parameter B is related to model-free parameters as $B = 2\mathcal{S}^2\tau_s^2$, parameter A represents fast internal motions in the extreme narrowing limit and can be expressed as $A = 2(1 - \mathcal{S}^2)\tau_e^2$, where τ_e is the effective correlation time of internal motions in the

model-free approach. τ_s represents an effective correlation time describing slow motions sensed by a NH vector and it was fitted individually for each residue.
The authors noted that measurement of such an abundant set of relaxation data at multiple fields should make possible the determination of ^{15}N chemical shift anisotropy ($\Delta\sigma$) values for each residue. The resulting variation along the sequence is very small and almost within the experimental uncertainty, with optimal values in the range $-170..-180$ ppm.

In a study of the unfolded D2 domain of annexin I [174] by ^{15}N NMR relaxation, Ochsenbein et al. analysed in detail the consequences of applying model-free approaches to unfolded proteins. The authors remark that the longest correlation time fitted by a model-free approach evoking two separate time scales does not correspond to overall tumbling but rather represents a *segmental motion* for which values between 1 and 3 ns were found for each residue. Model-free analysis using two or three time scales fit the experimental data equally well, but the obtained timescales of motion are quite different (see above). The authors suggest that these values do not represent discrete timescales in flexible proteins, but rather an averaging of motions on several timescales. For each model, this averaging is done in a different way. That is why different correlation times can be obtained, and, in particular, the two timescales of motion of the standard model-free analysis lie in between the three timescales of motion of the extended model-free fit.

In a remarkable paper [157], Halle argues that, in order to separate global and internal motions, i.e. (cf. Equation 4.19)

$$C_{total}(t) = C_{global}(t) \cdot C_{internal}(t),$$

not only the internal motions have to be statistically independent from the global reorientation, but also the latter has to be isotropic.

However, the model-free form of the spectral density function with two additive contributions from overall and internal motion would still be valid provided that the timescale of internal motion are much faster than that of the overall motion, But the underlying mathematical basis for this is different. Referring to the previous work by Furó [175], Halle remarks that if internal motions are fast enough compared to the overall motion so that the *adiabatic approximation* holds, the total correlation function would have a form of:

$$C_{total}(t) = C_{global}(t) + C_{internal}(t) \tag{4.59}$$

The *adiabatic approximation* requires that any correlation would be almost lost in the internal motion on a timescale on which the correlation function of the global motion is

still close to its original value. If the internal correlation function is given or approximated by an exponential, then the resulting formula for the spectral density function would be identical to the equation (4.30) of the classical model-free approach, as $\tau_e' = (\tau_M^{-1} + \tau_e^{-1})^{-1} \simeq \tau_e$. Halle also shows that the timescale separation under adiabatic approximation always implies the statistical independence of motions, but the opposite affirmation is not true.

In conclusion, the model-free approach can be used either when the timescales of motions are of different orders of magnitude, or when the internal motion is statistically independent from the global motion and the latter is the isotropic overall tumbling. Initially, it was conceived to analyse relaxation data in folded proteins, where the slowest relaxation-active timescale of motion is that of the overall rotation of the molecule, and internal motions lay on a timescale that is one or two orders of magnitude shorter than that. As we discussed here, in disordered proteins the slowest fitted correlation time does not seem to be related to the overall reorientation of the molecule; and, besides, one could hardly imagine that such a process has any pertinence in the description of the dynamics of a flexible chain. Moreover, it was proposed that there is a distribution of timescales of motions in disordered proteins, and it was not known if there is any timescale separation of these motions.

As it is not clear if the premises for model-free analysis are fulfilled in the case of disordered proteins, a lot of effort has been made to study these proteins by molecular dynamics (MD) simulations, to adapt or justify the model-free approach, and to develop alternative analytical approaches for studying them. We will discuss these in the following sections.

4.2.4 Molecular dynamics studies of flexible and disordered proteins

Prompers and Brüschweiler published several articles on the dynamics of folded and unfolded ubiquitin, based on a principal component analysis of covariance matrices extracted from molecular dynamics (MD) trajectories [176–178]. In the article [178], they measured relaxation data and ran MD simulations on folded ubiquitin and on its partially folded A state. In the method called isotropic reorientational eigenmode dynamics (iRED), either Cartesian coordinates of directions of N-H bond vectors (iRED rank 1) or associated spherical harmonics $Y_{LM}(\Omega_j)$ (iRED rank 2) are used to extract covariance matrices from MD trajectories. As pointed out by authors, MD simulations were too short in 2002 (and are still so with a few rare exceptions) to produce an isotropic distribution of conformers. So they enforced an isotropic average of the obtained covariance matrix, so that each MD snapshot was represented in the matrix by an infinite number of isotropically reoriented replicas. Subsequently, a principal component analysis was done on the averaged covariance matrix to obtain reorientational eigenmodes of a protein with

their eigenvalues λ_m. These eigenmodes describe correlated modulations of different spin interactions, and their eigenvalues correspond to the variances of the amplitude fluctuations of the trajectory along the eigenmode, or, in other words, concerted motions of N-H bond vectors contributing to protein dynamics and amplitudes of such motions.

Information on correlation times τ_m associated with individual modes m is not present in the covariance matrix, as it is not time-dependent, so correlation times τ_m were obtained from the fit to experimental relaxation data.

The authors plotted mode collectivity (percentage of interactions, or N-H vectors, significantly affected by a given mode) versus eigenvalue (amplitude) for the isotropic RED analysis (Fig. 4.6).

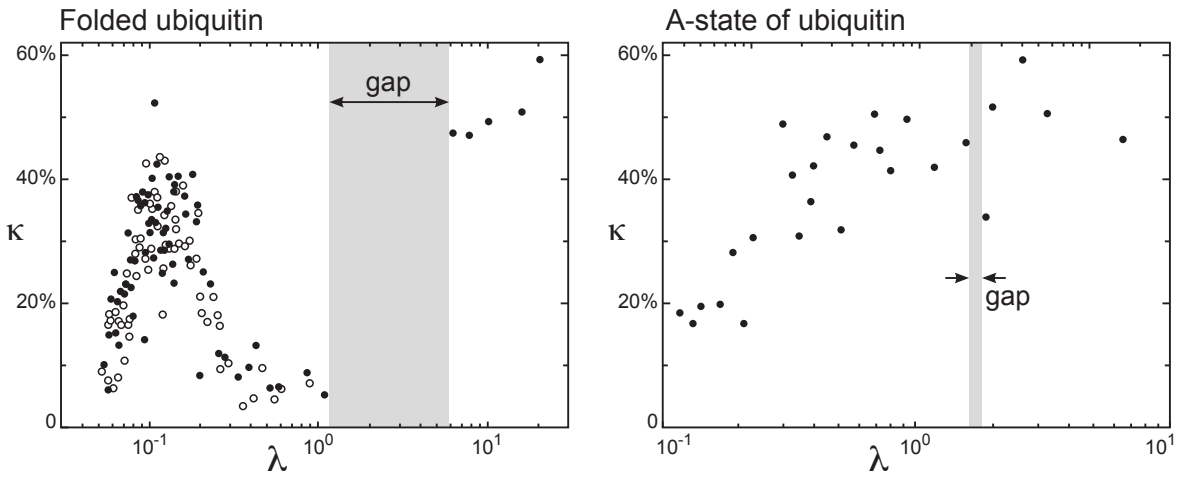


Fig. 4.6. Mode collectivities κ versus eigenvaules λ for iRED analysis (black circles) on folded ubiquitin and it's A-state. Open circles are for the results obtained using standard RED analysis [177] not discussed here.

In the case of folded ubiquitin, there is a clear gap between overall rotational modes and internal modes (Fig. 4.6, left plot). However, no clear gap is present between the 5 largest amplitude modes and other modes, suggesting that separability between overall and internal motions is not fulfilled in the case of the partially folded A-state of ubiquitin.

Mode-specific correlation times and order parameters obtained in this study were compared with correlation times, order parameter and internal motion amplitudes obtained in the previously discussed study of ubiquitin A-state by Brutscher et al. [169]. They conclude that both analyses cover a similar range of correlation times, but the iRED analysis provides more detailed dynamic parameters. In particular, the iRED analysis suggests a presence of multiple slow correlation times in the A-state of ubiquitin.

The authors conclude that iRED provides a general framework for the characterization of macromolecular dynamics on nano- and subnanosecond timescales without any

assumption about the separability of overall and internal motions, and is applicable to a variety of systems, including both folded and unfolded proteins.

Intrinsically disordered proteins were not extensively studied by molecular dynamic simulations until quite recently. Their available conformational space is so vast that dedicated hardware and/or special sampling techniques are required to correctly sample it.

The extended nature of the polypeptide chain in IDPs raises the question of correctly accounting for solvation effects. Ideally, they should be modelled using all-atom explicit solvent MD simulation. However, urea-denatured proteins were often found too compact if not structured in such trajectories [179]. Another artefact often found in MD simulations of disordered proteins is the excessive helical propensity [180]. All these effects make predicting the dynamical behaviour of disordered proteins, and in particular reproducing relaxation rates, more difficult.

In 2001, Xue and Skrynnokiv studied the unfolded state of ubiquitin by a high-temperature (500K) 1 μs MD simulation with implicit solvent and the amber ff99SB force field [181]. High temperature was chosen to simulate the chemical denaturation of protein in an acid/urea environment. Relaxation rates (^{15}N R_1, R_2, heteronuclear NOEs) were measured at 600 MHz and 5°C. The obtained MD trajectory was then 'rescaled' to match the slower experimental dynamics measured at 278K.

Calculated values of the radius of gyration were very close to those observed in SAXS/SANS studies. Calculated R_1 rates were also very close to measured values, while R_2 rates showed more variability due to the transient formation of secondary structure. Simulated NOE values were on average lower (0.15) than observed experimentally (0.28), and a larger residue-by-residue variation was observed. However, NOEs were not included in the fit of the rescaling factor, and the observed difference was small enough to be viewed as an independent validation of the MD model.

The analysis of NH bond vector correlation functions allowed the identification of three exponential components with averaged amplitudes and correlation times of 0.30/44 ps, 0.42/1.4 ns and 0.28/9.4 ns. Importantly, the first mode was attributed to fast internal librations, the second to local rearrangements and the third to global conformational transitions.

A more recent article by Lindorff-Larsen et al. [182] presented a 200μs-long MD simulation of the acid-unfolded state of the bovine acyl-coenzyme A binding protein (ACBP). It was performed on a supercomputer named Anton, designed specifically for MD simulations [183] at D. E. Shaw Research.

ACBP was chosen because its partially folded state is stabilized without the use of high denaturant concentrations, thus not requiring the use of a mixed solvent. This form of ACBP has been shown to contain four transiently populated helices and so could be considered to be closer to a folded than an unfolded protein.

Helices 2 and 4 were estimated to form and disrupt on timescales of about 5μs, and helix 1 on a considerably shorter timescale about 600 ns. The detailed study of contacts formed in the trajectory revealed that residues in helices 2 and 4 form long-range contacts, a property that were also observed experimentally. The authors suggest that faster dynamics in the helix 1 are caused by its less involvement in contact formation, in agreement with the observation that secondary structure elements are stabilized by long-range contacts in disordered proteins (Section 4.2.7).

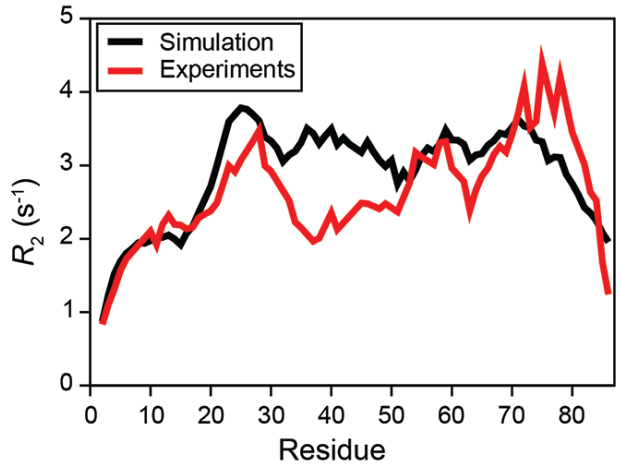


Fig. 4.7. R_2 relaxation rates calculated from correlation functions extracted from trajectory (black line) and measured experimentally (red line).

A comparison of experimental and MD-derived R_2 relaxation rates is shown in Fig. 4.7. A reasonable agreement with experiment was obtained, taking into account that no fitting to experimental data was involved.

The authors conclude that there is great promise for MD simulations in dynamical studies of disordered proteins, however, there is still considerable work to be done on the force fields to improve the local conformational sampling.

4.2.5 Analysing the distribution of correlation times in disordered proteins

A number of groups have attempted to analytically describe the continuous distribution of motions occurring on the ps-ns timescale in unfolded proteins. A simple solution is to introduce a continuous distribution of correlation times in the autocorrelation function. Such functions were proposed in polymer physics [184] to explain the dynamic behaviour of synthetic polymers, which share some properties with unfolded proteins.

The general form of a spectral density function in which the distribution of correlation times is integrated is given by:

$$J(\omega) = \int_0^\infty F(\tau) J(\omega, \tau) d\tau \,, \quad \text{where}$$
$$\int_0^\infty F(\tau) d\tau = 1 \tag{4.60}$$

A Cole-Cole distribution, previously used to analyse R_1 and NOE data for polymers [184], was applied to study dynamics in the unfolded pro-peptide of subtilisin [185, 186]. It is given by the following equation:

$$F(\tau) = \frac{1}{2\pi} \frac{\sin(\epsilon\pi)}{\cosh\left(\epsilon \ln\left(\frac{\tau}{\tau_0}\right)\right) + \cos(\epsilon\pi)} \tag{4.61}$$

where τ_0 is the mean value and ϵ $(0 < \epsilon < 1)$ is the width of the distribution. A drawback of the Cole-Cole distribution and in general of distributions proposed in polymer physics, is that the derived spectral density function usually diverges when $\omega \to 0$ (i.e., $(J(0) \to \infty)$), so they are only adapted to fit R_1 and heteronuclear NOEs and are not suitable for analysing R_2.

Ochsenbein et al. [187] proposed to use a lorentzian distribution of correlation times:

$$F(\tau) = K \frac{\Delta}{\Delta^2 + (\tau - \tau_0)^2} \quad \text{for } 0 \le \tau \le \tau_{max}$$
$$F(\tau) = 0 \quad \text{for } \tau > \tau_{max} \,, \tag{4.62}$$

in which K is a normalization constant to satisfy $\int_0^\infty F(\tau) d\tau = 1$:

$$\frac{1}{K} = \text{arctg} \frac{\tau_{max} - \tau_0}{\Delta} + \text{arctg} \frac{\tau_0}{\Delta} \tag{4.63}$$

The distribution $F(\tau)$ is characterized by its center τ_0 (i.e., the timescale of the most probable motions) and width Δ (i.e. a measure of the dynamical heterogenity), and is cut after τ_{max} to insure the non-divergent behaviour of $J(0) = \int_0^\infty F(\tau) J(0, \tau) d\tau$.

If the spectral density function of each mode is given by:

$$J(\omega, \tau) = \frac{2}{5} \frac{\tau}{1 + (\omega\tau)^2} \tag{4.64}$$

then the spectral density function for a lorentzian distribution of modes is given by:

$$J_{LD}(\omega) = \frac{1}{5D}\left\{\begin{matrix} 2\tau_0[1+S\omega^2] - 4K\Delta\omega\tau_0 \mathrm{arctg}(\omega\tau_{max}) \\ +K\Delta[1-S\omega^2]\ln\dfrac{\Delta^2+(\tau_{max}-\tau_0)^2}{S(1+\omega^2{\tau_{max}}^2)} \end{matrix}\right\}, \quad \text{where}$$
$$S = (\Delta^2+\tau_0)^2, D = \{1-2(\Delta^2-\tau_0)^2\omega^2 + S^2\omega^4\} \tag{4.65}$$

This approach was applied to the unfolded D2 domain of the annexin [174]. The model containing only a lorentzian distribution of correlation times (Eq. 4.65) was unable to correctly reproduce the experimental data. The contribution of very fast motions at the picosecond timescale, corresponding to librational motions, was then introduced. It was assumed to be independent from the nanosecond-range motions in $J_{LD}(\omega)$, thus contributing linearly to the spectral density function:

$$J_{LD2}(\omega) = A_{LD}J_{LD}(\omega) + \frac{2}{5}(1-A_{LD})\frac{\tau_f}{1+\left(\omega\tau_f\right)^2} \tag{4.66}$$

where the second term was considered to be negligible (as in the simplified extended model-free analysis (4.42)), so the contribution of very fast motions was implicitly taken into account by the term A_{LD}. This term is analogous to the order parameter $\mathcal{S}_f^2$ in the model-free approach, $1-A_{LD}$ or $1-\mathcal{S}_f^2$ representing the contribution of fast motions on ps timescale.

Comparing parameters obtained using a lorentzian distribution of correlation times with previous results of a conventional model-free analysis (Fig. 4.8), the authors note that the two correlation times obtained from local model-free approach appear to be located on each side of the distribution maximum, while in the extended model-free analysis the correlation time of slow internal motions τ_s closely matches the centre of the distribution.

The authors suggest that the 'overall' correlation time is associated with the slowest timescale of motions in the protein that are not taken into account by τ_s and $\mathcal{S}_s^2$. A good correlation between sequence profiles of the slow motional order parameter $\mathcal{S}_s^2 = \mathcal{S}^2/\mathcal{S}_f^2$ and the distribution width Δ suggests that $\mathcal{S}_s^2$ is related rather to dynamical heterogeneity of motions rather than to their amplitude.

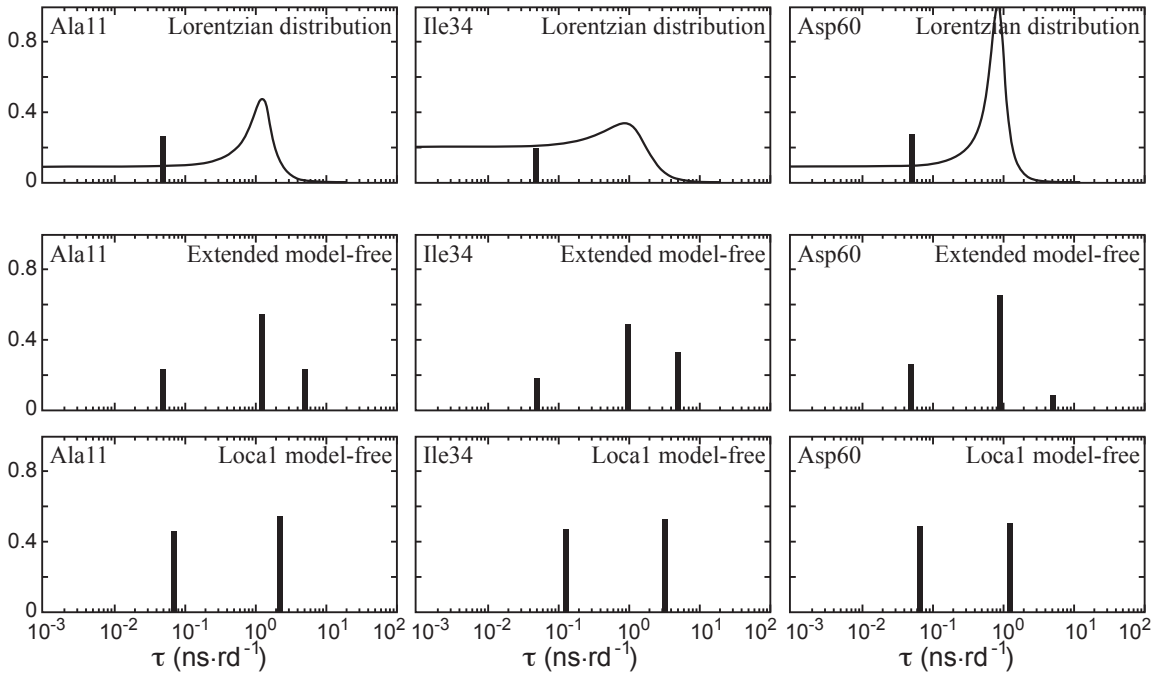


Fig. 4.8. Comparison of distributions of correlation times for three different residues of annexin I domain 2. Continuous line is used for lorentzian distribution, and discrete correlation times obtained by model-free and extended model-free analysis are represented by vertical bars with a height corresponding to their amplitudes (from left to right: $1 - \mathcal{S}_f^2; \mathcal{S}_f^2 - \mathcal{S}^2; \mathcal{S}^2$). On the lorentzian distribution plots, a vertical bar corresponds to the amplitude of fast internal motions $1 - A_{LD}$.

A similar conception of correlation times distribution was later used by Modig and Poulsen in 2008 to describe motions in the acid-unfolded state of acyl-coenzyme A binding protein (ACBP) [188]. Calling their approach 'model-independent correlation (MIC) time distribution', they concentrate on calculating the statistical properties of the distribution, namely the mean correlation time $\langle \tau_c \rangle$, standard deviation σ_τ, skewness γ_τ. They also calculated some fundamental properties of the resulting correlation function: the half-time $\tau_{1/2}$ and a ratio of the integral of $C(\tau)$ from $\tau_{1/2}$ to ∞ and its full integral from 0 to ∞, which measures the amount of correlation left after $\tau_{1/2}$. However, the profiles of almost all obtained parameters are either correlating or anti-correlating with each other, and the information content of each parameter is unclear. As remarked by Khan et al. [189] little information is provided on the actual number of modes in the distribution.

4.2.6 Recent adaptation of the model-free approach to study dynamics in disordered proteins

A recent approach by Khan et al. [189], called IMPACT (Interpretation of Motions by a Projection onto an Array of Correlation Times), involves fitting the spectral density function using a sum of Lorentzians with *fixed* correlation times and variable amplitudes.

The number and values of correlation times are adjusted in a way to obtain a minimal set of timescales and amplitudes that reproduce the experimental data.

The IMPACT approach was originally conceived to study backbone dynamics in the partially disordered transcription factor Engrailed 11 fragment. This fragment contains a N-terminal 54-residue disordered region and the folded homeodomain (residues 200-259). In the homeodomain, a clear gap between slow ns motions (A_1-A_3) and very fast ps motions (A_6) was observed, with A_4 and A_5 being very close to zero. This gap is similar to the gap between overall and internal motions observed by Prompers and Brüschweiler (Section 4.2.4) in folded ubiquitin [178].

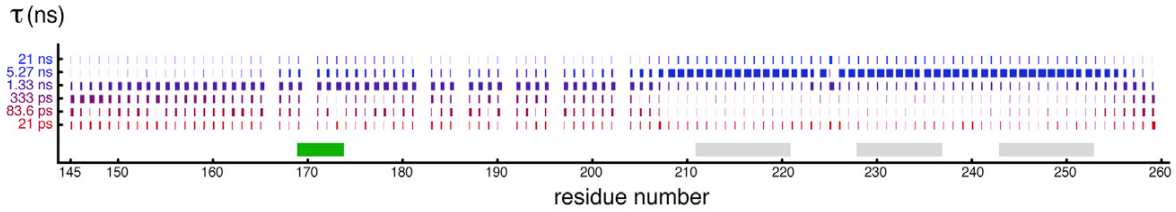


Fig. 4.9. Graphical "barcode" representation of the IMPACT analysis. The green bar shows the position of the hexapeptide in the disordered domain. Grey bars mark the positions of the helices in the homeodomain.

In the disordered region, the maximal contribution is from motions around τ_3 (1.33 ns), with A_3 values being around 0.4, but the observed distribution is quite broad. The N-terminal region shows significant contributions from motions in the ps – low ns range and almost no contribution from motions slower than $\tau_2 = 5.27$ ns. The C-terminal half of the disordered region (between the hexapeptide and the homeodomain) however shows a more significant contribution from slower ns motions with A_2 being around 0.15. The authors do not assign these slower motions to any particular process, saying that a better characterization of the conformational space of the protein would be needed. The barcode representation, shown in Fig. 4.9, clearly shows that motions are shifted from the subnanosecond range to the suprananosecond range when moving from the unfolded to the folded and partially folded domains, with nonzero contribution of motions in the 5.27 ns – 21 ns range. The authors compare their analysis to conventional model-free analyses.

The advantages of fixing the numerical values of common correlation times for the entire protein, compared to determining their values directly from experimental relaxation for each site, is unclear (LeMaster proposed to fix correlation times to the inverse of the relaxation active Larmor frequencies LeMaster 1995 JBNMR), although the authors correctly note that effective correlation times determined from conventional modelfree approaches do not necessarily report on physically meaningful motional modes. In this context it is also important to note that the grid of fixed correlation times must be pre-determined from an initial analysis of the entire protein, complicating statistical comparison with modelfree approaches.

Addressing the context of the validity of applying modelfree approaches to disordered proteins, Gill et al. [190] recently studied the dynamics of the 58-residue basic leucine-zipper (bZip) domain of the *Saccharomyces cerevisiae* transcription factor GCN4. In the absence of the DNA substrate, its N-terminal region is intrinsically disordered but contains significant residual helicity, while the coiled-coil region remains ordered and dimeric. ^{15}N relaxation rates R_1, R_2 and steady-state ^{15}N-^{1}H NOEs were measured at 300K at 600, 700, 800 and 900 MHz. To obtain insight into its dynamic properties, the authors applied a modified version of the spectral density mapping approach. They initially obtain a set of parameters for each field:

$$\Gamma_{auto} = R_2 - 0.5R_1 - 0.454\sigma_{HN} = J(0)(3d^2 + 4c^2)/6 \quad (4.67.1)$$
$$J(\omega_N) = \{R_1 - 1.249\sigma_{HN}\}/(3d^2/4 + c^2) \quad (4.67.2)$$
$$J(0.870\omega_H) = 4\sigma_{HN}/5d^2 \quad (4.67.3)$$

Then $J(0)$ is obtained as a slope of a linear fit through the origin of Γ_{auto} vs. $(3d^2 + 4c^2)/6$. Assuming an extended model-free form for the spectral density functions, the motional parameters can be extracted from linear fits to:

$$J(0.870\omega_H) = m_H(0.870\omega_H)^{-2} + b_H \quad (4.68)$$

and

$$J(\omega_N) = m_N{\omega_N}^{-2} + b_N \quad (4.69)$$

in which

$$m_H = \frac{2}{5}\left(\frac{\mathcal{S}^2}{\tau_M} + \frac{(\mathcal{S}_f^2 - \mathcal{S}^2)}{\tau_s'}\right)$$
$$b_H = \frac{2}{5}(1 - \mathcal{S}_f^2)\tau_f'$$
$$m_N = \frac{2}{5}\mathcal{S}^2/\tau_M$$
$$b_N = \frac{2}{5}\{(\mathcal{S}_f^2 - \mathcal{S}^2)\tau_s' + (1 - \mathcal{S}_f^2)\tau_f'\} \quad (4.70)$$

The authors compared the obtained results with those obtained using the extended Lipari-Szabo approach. Values obtained by the two approaches were very similar. However, the spectral density mapping-based approach failed to extract the two faster timescales (τ_s and τ_f) in the folded coiled-coil region, as this involves evaluating ratios $b_H/(1 - \mathcal{S}_f^2)$ and $(b_N - b_H)/(m_H - m_N)$ where both divisor and dividend are close to zero.

Despite this stability drawback, the method has an advantage: for the linear relationships in (4.68) and (4.69) to hold it is required that $\tau_M \gg \tau_s \gg \tau_f$. Thus the method provides a direct way of evaluating whether the requirements for using an extended model-free approach are met (cf. end of Section 4.2.3), demonstrating that in the case of the disordered tail of GCN4, this is the case.

4.2.7 Disordered proteins as polymer chains and segmental motions

Schwalbe et al. noticed that the behaviour of R_2 in unfolded proteins in the so-called 'random coil' state devoid of any structured elements has a quite simple pattern: in the centre of the chain it approaches a 'plateau', whether in the terminal parts it descends as the residue is getting closer to the end of the chain. This would imply that the relaxation properties of a given amide depend less on the specific nature of its neighbours but rather reflect the fact that it is a member of a polypeptide chain. Schwalbe et al. thus suggest that in polymers the correlation time should be largely independent of the overall tumbling rate and be affected only by segmental motion. They propose [191] to fit R_2 rates to a model describing the protein as a heteropolymer consisting of 20 types of monomers, in which a) the contribution of the neighbouring residues to the relaxation rate of a given residue decays exponentially with the distance in the sequence and b) the influence of neighbouring residues is independent of side chain volume or hydrophobicity. The fit equation is given by:

$$R_2^{rc}(i) = R_{int} \sum_{j=1}^{N} e^{\frac{|i-j|}{\lambda_0}} \tag{4.71}$$

where R_{int} is the intrinsic relaxation rate, which depends also on the temperature and the viscosity, λ_0 is the persistence length of the polypeptide chain, and N is the total chain length in residues.

One of the most recent and convincing applications of this model is found in the article by Wirmer et al. [192] about motional properties of unfolded ubiquitin. Relaxation data (R_1, R_2, heteronuclear NOE) were measured at 25°C at 600 MHz.

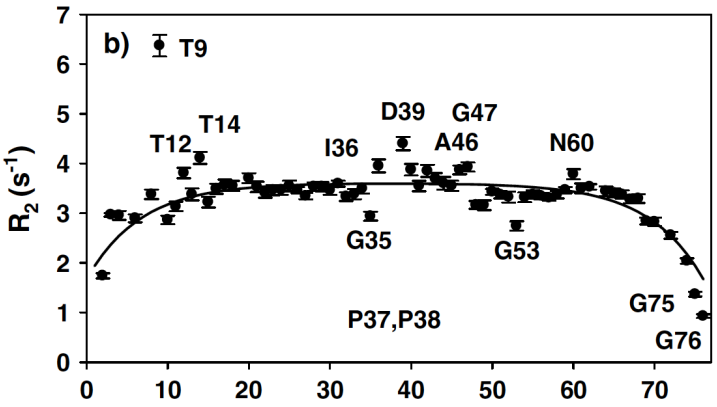


Fig. 4.10. Experimental R_2 relaxation in unfolded ubiquitin rate fitted with the segmental motion model for the dynamics of an unbranched polymer chain.

R_2 rates were fitted using equation (4.71) and the result is shown in Fig. 4.10. The best fit values are $R_{int} = 0.27\text{s}^{-1}$ and $\lambda_0 = 6.67$. This persistence length agrees with the

persistence length $\lambda_0 = 7$ found in other unfolded proteins [191, 193, 194], leading the authors to propose that the persistence length of seven residues is a general property of unfolded polypeptide chains.

Residual structure in disordered proteins

However, in many other intrinsically disordered and partially denatured proteins, residual structure has been detected. In the article by Klein-Seetharaman et al. [193] the authors remark that residual structure in denatured and unfolded proteins tends to reside in hydrophobic clusters, in which tryptophan or histidine residues are surrounded by other hydrophobic side chains. They measured R_2 in reduced hen lysozyme and fitted with optimal parameters $R_{int} = 0.20\text{s}^{-1}$ and a persistent length λ_0 of 7 residues. Several regions of the polypeptide chain had R_2 values much bigger than those predicted by the model (Fig. 4.11 B). These differences were attributed to the presence of clusters of hydrophobic side chains. To account for the presence of these hydrophobic clusters, a modification of the random coil model was proposed, that contains a baseline corresponding to random-coil behaviour and an additive Gaussian-form contribution to R_2 from these clusters:

$$R_2^{rc+clust}(i) = R_2^{rc}(i) + R_{clust} \sum_{clust} e^{\frac{|i-x_{clust}|^2}{\lambda_{clust}}} \tag{4.72}$$

where x_{clust} is the centre of the cluster, R_{clust} and λ_{clust} are the intrinsic relaxation rate and persistence length of the cluster. This model fits the relaxation rates well in all cases (Fig. 4.11).

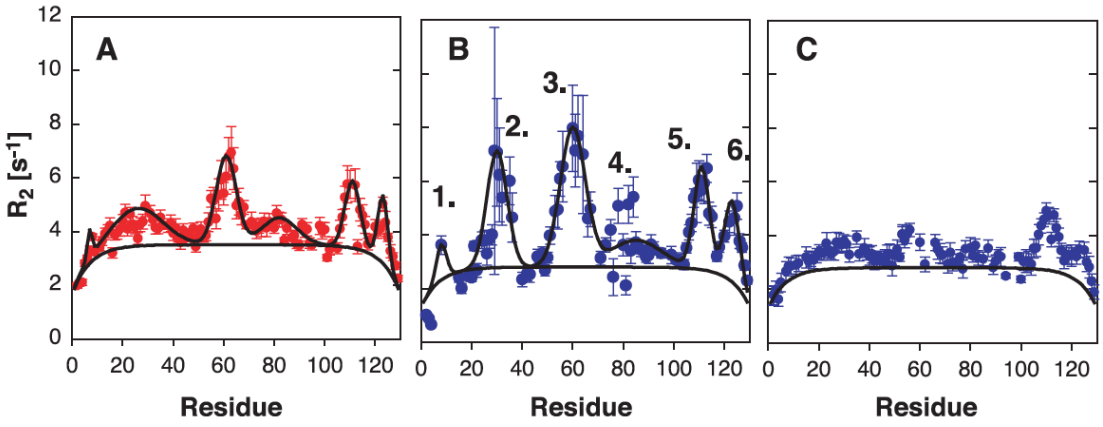


Fig. 4.11. R_2 relaxation rates in reduced methylated hen lysozyme. A: Wild-type lysozyme in 8M urea, pH2. B: Wild-type lysozyme in water, pH2. Numbers indicate hydrophobic clusters. C: W62G mutant in water, pH2. Lower black line: fit with segmental motion model. Higher black line: fit with segmental model with clusters.

4.2.8 Temperature and chain length dependence of segmental motions in IDPs

In the article by Parigi et al. [195], the authors discuss the general belief that backbone reorientation in IDPs is predominantly governed by internal segmental motions independent of the overall tumbling of the molecule, and involve concerted motions of local clusters or entire subdomains on the timescale of several nanoseconds. However, high-field relaxation analysis is not ideally suited to study such slow motions and the authors therefore use fast field cycling (FFC) relaxometry, allowing the measurement of average ^{1}H R_1 values from very low fields up to 50 MHz.

Relaxation rates were measured in different intrinsically disordered proteins at 298K from 0.02 to 45 MHz. Relaxation dispersion curves were fitted with a lorentzian-shaped spectral density function with correlation time τ_R and an order parameter/amplitude S_C^2. Residual contributions from faster motions (assumed to be in extreme narrowing limit) were fitted using an additive parameter α.

By comparing data measured on 99-residue Tau K19 fragment and on a longer 441-residue hTau40 form, the authors showed that the correlation time of slow motions in IDPs increase with the chain length. They also found it to decrease with increasing temperature, with a good correlation with η/T term derived from Stokes law (Fig. 4.12)

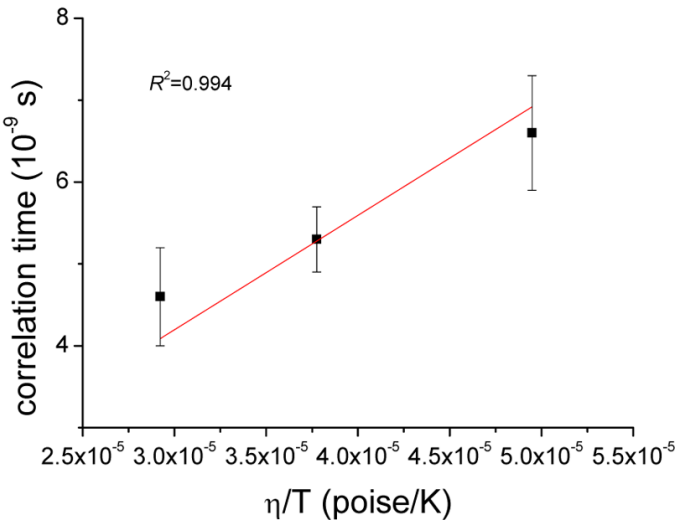


Fig. 4.12 Temperature dependence of correlation times in Tau K19. Red line shows correlation time obtained at 288K rescaled to different η/T values.

In earlier studies, a dependence of the overall correlation time on chain length was observed in the context of a globular domain with a disordered chain connected to it. For example, in a study [196] of a Syrian hamster prion protein (ShPrP), two fragments of different length were considered. Both contain a globular domain and largely unstructured N-terminal domains of different length. As measured by the ratio of R_2/R_1, the τ_C of the shorter form is 10.5 ns, whereas longer form has a τ_C value of 12.3 ns. This shows that the globular domain 'feels' the presence of unfolded domains of increasing length.

4.3 Summary and discussion

It has long been known that any Markovian motion (a process where probabilities of going to each particular next state depend only on the current state of the system and not on its state history) may be described by a correlation function expressed as a series of exponentials. This property was used by King and Jardetzky [160] to justify the fit of relaxation rates in a protein using a spectral density function represented as a sum of Lorentzians. Then Lipari and Szabo formulated their 'model-free' approach, conceived to fit relaxation data in folded proteins assuming that two types of motions are contributing to relaxation: overall tumbling of the molecule and internal motions close to the extreme narrowing limit. The spectral density in this approach is fitted with an equation containing two lorentzian terms representing each kind of motion. Clore et al. [163] extended this model-free approach to the cases where slow internal motions are also present, fitting a spectral density function with three Lorentzians. Several authors tried then to apply model-free or extended model-free approach to fit relaxation data in partially unfolded or disordered proteins [165, 168–170, 174, 197].

One of the main problems was assigning the fitted correlation times to physical motions, as fits with two or three Lorentzians were giving different correlation times [174]. The model-free approach was originally conceived to analyse relaxation data in folded proteins, where internal motions are superimposed on an overall isotropic rotation of the molecule. This assumption does not seem to be valid in the context of unfolded proteins, where difficulties in assigning the slowest fitted correlation time to an overall motion were pointed out [187]. Different studies proposed fitting relaxation data in disordered proteins with a distribution of correlation times, instead of a discrete set of the latter [185–188].

An article by Halle [157] shows that the model-free approach is also valid when, in particular, timescales of motions in a protein are separated. A recent article [190] shows that the timescale separation holds in a disordered protein, thus justifying the application of an extended model-free approach with three timescales to these proteins.

The chain-like behaviour of disordered proteins is underlined in the works of Schwalbe [192, 193, 198], where it is proposed to fit R_2 relaxation rates in terms of a polymer chain with a certain persistence length. The consensus value of the latter was found to be 7 residues. When the protein contained regions of compactness and secondary structure elements, the underlying behaviour was still assumed to be that of a random-coil, and the regions of restrained mobility, 'clusters', were assumed to give an additive contribution to the random-coil R_2 rate.

Different models used in polymer physics, as for example Rouse and Zimm model discussed in the subchapter 4.1, predict that a polymer chain has a distribution of motional modes, and longer chains have motional modes with slower correlation times. Fast-field cycling relaxometry studies [195] of a set of disordered proteins show that slower ($\geq$ 5 ns) motions are present in IDPs. Their correlation time seem to depend on the chain length, as would be predicted by polymer models, but also on the temperature, with a characteristic Stokes-law η/T-proportionality. The authors thus propose that this timescale of motions is related to segmental motion in disordered proteins. Similar conclusion on the nature of these slow motions are made by MD studies [181].

We can conclude that disordered proteins are extremely complex dynamic systems with different timescales of motions contributing to NMR relaxation; however, many peculiarities of their behaviour come from the underlying chain-like behaviour as in homopolymers. There is more and more evidence that slowest ($\geq$ 5 ns) motions that contribute to NMR relaxation can be assigned to segmental motions or large-scale reorientations of the protein chain. Finally, there is evidence that disordered proteins can be usefully studied using an extended model-free approach with three timescales that describes the principal dynamic modes of the protein.

Chapter 5. Materials and methods

5.1. Sample preparation

Three different constructs of the disordered C-terminal N_{TAIL} domain of the nucleoprotein of Sendai virus strain Harris were expressed and purified. The sequence corresponds to that of the N_{TAIL} of Fushimi strain (UniProtKB No Q07097) except for one mutation E410K. The longest construct $N_{TAIL,L}$ or NT_L comprises the entire C-terminal domain of the nucleoprotein (residues 401-524), whereas the medium length $N_{TAIL,M}$/NT_M and the short length $N_{TAIL,C}$/NT_C constructs correspond to residues 443-524 and 443-501 respectively. A double mutant of $N_{TAIL,L}$ W441A W449A ($N_{TAIL,WW}$) was also expressed and purified. Expression and purification of these constructs were performed by DAMIEN MAURIN as described previously [111, 199]. Samples for NMR measurements were prepared in 50 mM sodium phosphate buffer at pH6 with 500 mM NaCl and 10% D_2O (v/v), and concentrated by centrifugal filtration up to the final concentration of 500 µM. Samples were kept frozen in 3mm NMR tubes between measurements. Before each use and after thawing, the sample was mixed by shaking to insure its homogeneity.

5.2. NMR measurements

Backbone resonances from the longest $N_{TAIL,L}$ construct was assigned at 298K using a set of BEST-type triple resonance spectra: HNCO, intra-residue HN(CA)CO, HN(CO)CA, intra-residue HNCA, HN(COCA)CB and intra-residue HN(CA)CB [200], acquired on a Varian spectrometer operating at a ^1H frequency of 600 MHz equipped with a cryoprobe. All spectra were processed using NMRPipe [201] and analysed in Sparky [202] and CCPN [203]. MARS [204] was used for automatic assignment of spin systems followed by a manual verification in CCPN.
The $N_{TAIL,L}$ assignment was transferred to other temperatures (258, 263, 268, 273, 278, 283, 288, 293, 298, 303K) by performing a series of ^1H-^{15}N HSQC spectra. The obtained assignment of the HSQC spectra was verified at 278K and 288K using HNCO, intra-residue HN(CA)CO, HN(CO)CA and intra-residue HNCA spectra. These spectra were processed using NMRPipe and analysed in CCPN.

$N_{TAIL,M}$ [111] and $N_{TAIL,C}$ [199] assignments at 298K were reported previously. At 278K, most peaks in these shorter forms were identified from the $N_{TAIL,L}$ assignment at this temperature; unassigned peaks were assigned using HNCO, HN(CO)CA and intra-residue HNCA spectra for $N_{TAIL,C}$ and in addition HN(COCA)CB and intra-residue HN(CA)CO spectra for $N_{TAIL,M}$, acquired on a Varian spectrometer operating at a ^1H frequency of 600 MHz equipped with a cryoprobe. $N_{TAIL,M}$ assignment at 288K was obtained by comparing $N_{TAIL,M}$ assignments at 278K and 298K, and the $N_{TAIL,L}$ assignments at 288K.

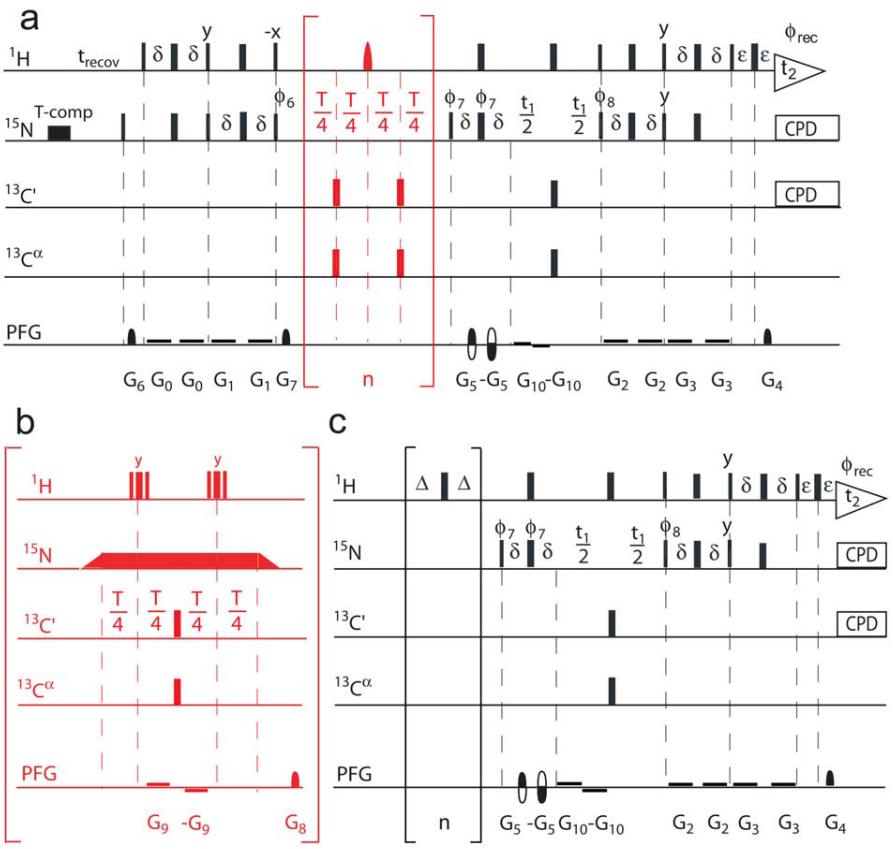


Fig. 5.1. (Reproduced from [205]) Pulses sequences for measuring ^{15}N relaxation rates through sensitivity-enhanced HSQC ^{1}H detection. (**a**) Scheme for measurement of ^{15}N R_1. Substitution of the element shown in (**b**) for the red-bracketed element in (**a**) converts the experiment to a ^{15}N $R_{1\rho}$ measurement. (**c**) Pulse scheme for measurement of ^{15}N-{^{1}H} NOE. Narrow and wide pulses correspond to 90° and 180° flip angle pulses, respectively, with phase x unless otherwise indicated. For application to ^{13}C labeled samples: durations of ^{13}C pulses (all 180°) are equal to $\frac{\sqrt{3}}{2\Omega}$ (47.4 μs at 600 MHz), where Ω is the frequency difference between ^{13}C$^{\alpha}$ and ^{13}C'. Delay durations: δ = 2.65 ms (≈¼ $^{1}J_{NH}$); ε corresponds to the duration of the decoding gradient G_4 (201 μs) plus a short (250 μs) recovery delay. Gradients G_0 (2.65 ms; 2.1 G/cm), G_1 (2.65 ms; 1.4 G/cm), G_2 (2.65 ms; -3.5 G/cm), G_3 (2.65 ms; -0.35 G/cm), G_9 (T/4; 0.35 G/cm), G_{10} (t_1/4; 1.05 G/cm) are rectangular shaped. Gradients G_4 (201 μs; 28 G/cm), G_5 (1 ms; 28 G/cm), G_6 (1 ms; 21 G/cm), G_7 (200 μs; -35 G/cm) and G_8 (300 μs, 28 G/cm) are sine-bell shaped. Phase cycling: φ_6 = 2(y), 2(-y); φ_7 = y,-y; φ_8 = x; φ_{rec} = y,-y,-y,y. Quadrature detection is implemented using the gradient-enhanced echo/anti-echo scheme [206] with the polarity of gradients G_5 and $-G_5$ inverted, and phase φ_8 = -x for the second FID, generated for each quadrature pair. (**a**) The 180° ^{1}H pulse applied at the midpoint of T (40 ms) is of the IBURP2 type and has a 2 ms duration (at 600 MHz), and the loop is repeated an even number of times (n=0, 2, ...). ^{13}C 180° pulses serve to eliminate cross correlation effects resulting from ^{15}N-^{13}C dipolar interaction in samples that include ^{13}C labeling. To ensure that the same RF heating applies in the R_1 and $R_{1\rho}$ experiments, immediately following data acquisition, a ^{15}N temperature compensation pulse [207] is applied that corresponds to the longest spin-lock time of the $R_{1\rho}$ experiment (see below). (**b**) The triangle shaped RF fields preceding and following the spin-lock period are adiabatic half passage pulses (see [205]). To eliminate the effect of cross-correlated relaxation, RF-inhomogeneity and offset-compensated 90_x-210_y-90_x ^{1}H pulses are applied at T/4 and 3T/4. H_2O radiation damping between the two composite pulses is prevented by the very weak (0.35 G/cm) gradient G_9. For ^{13}C-enriched samples, 180° ^{13}C$^{\alpha}$ and ^{13}C' pulses are applied to cancel possible cross-correlated relaxation effects related to the ^{15}N-^{13}C dipolar coupling.

Temperature compensation (50 kHz off-resonance) is applied at the RF field strength used during T, for a duration that equals the difference of the current and the maximum spin-lock durations. (**c**) ^{1}H saturation is achieved by *n* repetitions of the symmetric (Δ-180°-Δ) unit [208], with the ^{1}H carrier switched to 8.6 ppm and Δ equal to ≈11 ms which corresponds to $1/J_{NH}$.

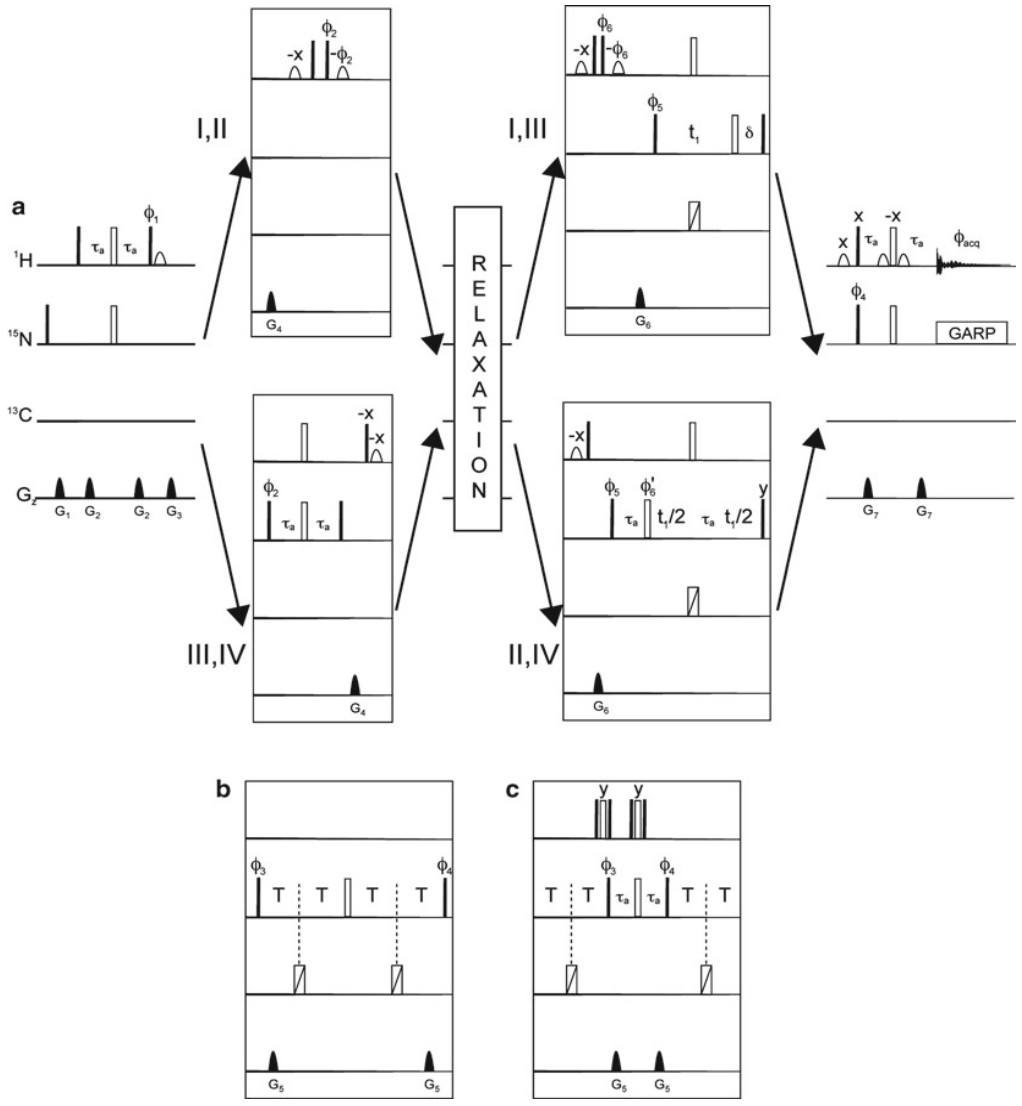


Fig. 5.2. (Reproduced from [209]) Pulses sequences for measuring CSA/DD cross-correlated cross-relaxation rates. (**a**) General scheme. (**b**) Relaxation block for measuring the transverse cross-relaxation rate. (**c**) Relaxation block for measuring the longitudinal cross-relaxation rate. Only specific elements are detailed here. *T* is the relaxation delay. The delay δ is equal to the shortest value of t_1 so that the first effective value of t_1 is zero. The phase cycles were: ϕ_1 = {y,-y}; ϕ_2 = {x,x,x, x, -x,-x,-x,-x}; ϕ_3 = {x,x,x,x, x,x,x,x, -x,-x,-x,-x, -x,-x,-x,-x}; ϕ_4 = { x,x,x,x, x,x,x,x, x,x,x,x, x,x,x,x, -x,-x,-x,-x, -x,-x,-x,-x, -x,-x,-x,-x, -x,-x,-x,-x}; ϕ_5 = {x}; ϕ_6 = {x,x,-x,-x }; ϕ'_6 = {x,x,y,y}; ϕ_{acq} = {x,-x,-x,x, -x,x,x,-x, -x,x,x,-x, x,-x,-x,x, -x,x,x,-x, x,-x,-x,x, x,-x,-x,x, -x,x,x,-x}. Gradient durations and peak amplitudes over the *x*, *y* and *z* orientations were, respectively: G_1; 600 ms, 15 G/cm, 15 G/cm, 0; G_2; 600 ms, 6.5 G/cm, 0, 0; G_3; 2 ms, 0, 0, 40 G/cm; G_4; 1 ms, – 9.5 G/cm, 9.5 G/cm, 0 G/cm; G_5; 600 ms, 3.5 G/cm, 0 G/cm, 14.5 G/cm; G_6; 1 ms, – 35 G/cm, – 35 G/cm, – 35 G/cm; G_7; 1 ms, 35 G/cm, 35 G/cm, 35 G/cm. Frequency sign discrimination was performed using States-TPPI.

The assignment of the double mutant $N_{TAIL,WW}$ at 298K was obtained using the same set of six spectra as for the wild type form. These spectra were processed using NMRPipe and

analysed in Sparky and CCPN. Assignment was transferred to 278K by performing a series of ^{1}H-^{15}N HSQC spectra at 298.2, 293.2, 288.2, 283.2, 280.7 and 278.2K.

Most NMR relaxation experiments were performed on Bruker spectrometers operating at ^{1}H frequencies of 600, 700, 850 and 950 MHz equipped with a cryoprobe. ^{15}N longitudinal relaxation (R_1), ^{15}N longitudinal relaxation in the rotating frame ($R_{1\rho}$) and {^{1}H}-^{15}N steady state heteronuclear NOEs were measured as described by Lakomek et al. [205]. The corresponding pulse sequences are shown in 5.1. A spin-lock field strength of 1500 Hz was used for measuring $R_{1\rho}$. The ^{15}N transverse relaxation rates R_2 were derived from the $R_{1\rho}$ and R_1 rates using the relationship:

$$R_2 + R_{ex} = \frac{R_{1\rho} - R_1 \cos^2 \theta}{\sin^2 \theta} \tag{5.1}$$

where $\theta = \arctan\left(\omega_1/(\omega - \omega_0)\right)$ is the 'tilt' angle (cf. Eq. 3.79), ω_1 is the spin-lock field strength (in rad·s^{-1}), ω_0 is the population-averaged ^{15}N chemical shift of the given resonance (in rad·s^{-1}), ω is the carrier frequency of the spin-lock field (in rad·s^{-1}). The effective spin-lock field strength for a resonance is given by $\omega_e = (\omega_1^2 + (\omega - \omega_0)^2)$. On the basis of our experimental results (*vide infra*), R_{ex} exchange contributions to the transverse relaxation were assumed to be negligible. ^{15}N-^{1}H CSA/DD cross-correlated transverse (η_{xy}) and longitudinal (η_z) relaxation rates were measured as described by Pelupessy et al. [210, 211] and the corresponding pulse sequences are shown in Fig. 5.2.

At 268K, all experiments were acquired as described above on a Bruker spectrometer using an Oxford magnet operating at a ^{1}H frequency of 600 MHz with a room temperature probe. The use of a 3mm NMR tube and high salt concentration (500 mM) in the buffer prevented the sample from freezing at this temperature.

Some R_1, R_2 and NOE experiments, in particular ^{15}N R_2 relaxation in $N_{TAIL,L}$ at 258K and some relaxation rates in the shorter forms $N_{TAIL,M}$ and $N_{TAIL,C}$ (as shown in the Results section) were run on a Varian NMR Direct-Drive Systems spectrometer operating at a ^{1}H frequency of 600 or 800 MHz. ^{15}N-R_1, ^{15}N-R_2 and {^{1}H}-^{15}N heteronuclear NOE experiments were carried out and the pulse sequence details can be found in reference [212]. ^{15}N R_2 relaxation in $N_{TAIL,L}$ at 258K was measured using a bundle of 1mm glass capillaries fitted into a 5mm NMR tube to prevent the sample from freezing [213].

In all experiments, an interscan delay of 1.5-3 s and a high number (64-128) of dummy scans in the beginning of experiment was used for measuring $R_{1\rho}$, R_1, R_2, η_{xy}, and η_z rates. For {^{1}H}-^{15}N heteronuclear NOEs a saturation delay of 4-6 s was used, and an interscan delay of 1 s in addition. To reduce experimental time or to increase the number of scans for the same experimental time, a reduced ^{15}N sweep width was used and hence a reduced

number of t_1 increments. The ^{15}N sweep width was minimized to a value for which no overlap occurs between folded and non-folded resonances in the spectra.

The typical set of relaxation delays was [0.07, 0.13, 0.02, 0.09, 0.23, 0.21, 0.05, 0.17, 0.07, 0.001 s] for $R_{1\rho}$ experiment and [0, 0.6, 0.08, 1.6, 0.4, 1.8, 1.04, 0.8, 0.2, 0.6] for R_1 experiment. Shorter delays were used for measuring $R_{1\rho}$ and R_1 relaxation rates at 268K. For the transversal cross-correlated cross-relaxation rate η_{xy}, a relaxation delay of 50 ms was used. For the longitudinal cross-correlated cross-relaxation rate η_z, a relaxation delay of 80 ms was used.

$R_{1\rho}$, R_1, and R_2 rates were obtained using exponential fitting of peak intensities in CCPN (Fig. 5.3). Uncertainties were estimated in CCPN using a bootstrap estimate of standard error [214] (which relies on repeated sampling with replacement), as described in CCPN documentation.

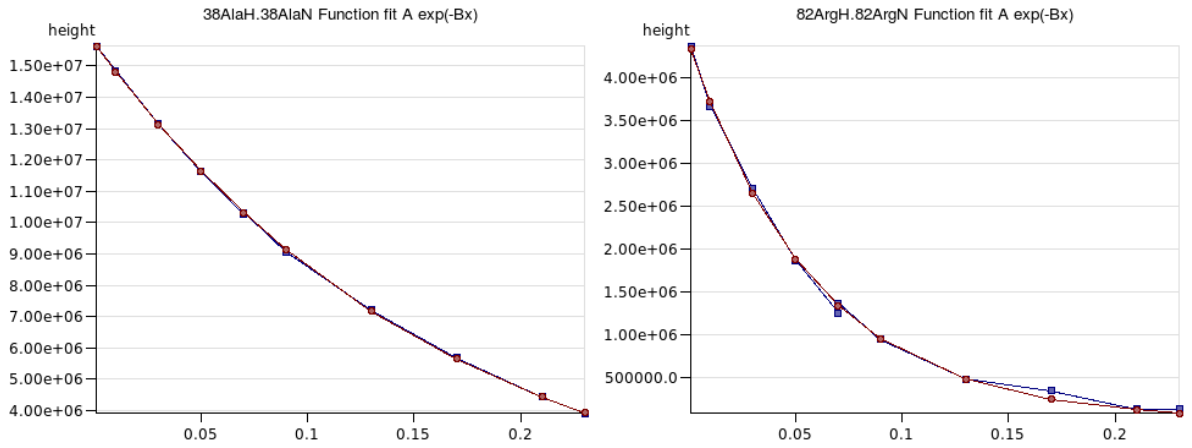


Fig. 5.3. Example of exponential peak intensity fits for $R_{1\rho}$ relaxation rate measured in $N_{TAIL,L}$ at 700 MHz at 278K. Residue 38 is located in the disordered part of the protein, residue 84 – in the α-helical MoRE.

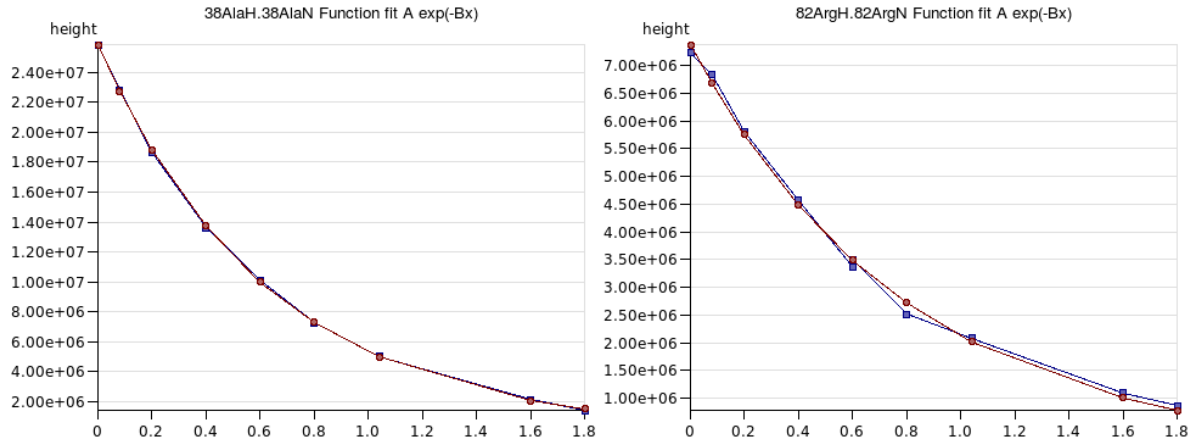


Fig. 5.4. Example of exponential peak intensity fits for R_1 relaxation rate measured in $N_{TAIL,L}$ at 700 MHz at 278K.

NOEs were calculated as the ratio of peak intensities in saturation (I_{sat}) and reference (I_{ref}) ^{1}H-^{15}N HSQC planes. Uncertainties were estimated using the equation:

$$\delta_{NOE} = \delta\left(\frac{I_{sat}}{I_{ref}}\right) = \sqrt{\frac{N_{sat}^2}{I_{ref}^2} + \frac{I_{sat}^2 N_{ref}^2}{I_{ref}^4}} \tag{5.2}$$

where N_{sat} and N_{ref} are the root mean square noise estimations in saturation and reference HSQC planes.

Transversal and longitudinal cross-correlated cross-relaxation rates were calculated from peak intensities as described in [209]. Uncertainties were estimated using Monte-Carlo simulations based on root mean square noise estimations in HSQC planes.

Temperature was individually calibrated using methanol-d_4 sample, by measuring chemical shift difference between OH and CH_2 peaks in the 1D ^{1}H spectrum and relating it to the sample temperature using equation [215]:

$$T = -16.7467 \cdot (\Delta\delta)^2 - 52.5130 \cdot \Delta\delta + 419.1381 \tag{5.3}$$

where $\Delta\delta$ is the chemical shift difference in ppm.

High power proton $R_{1\rho}$ relaxation dispersion experiments on $N_{TAIL,L}$ at 268K were performed on a Bruker spectrometer operating at ^{1}H frequency of 700 MHz equipped with a cryoprobe, as described in [216]. The $N_{TAIL,L}$ sample for this experiment was expressed and purified as described in section 5.2.1, except that the protein was grown in a medium containing only D_2O as water sourse, to deuterate the majority of non-exchangeable protons. Sample deuteration is desirable because it reduces pseudo-dispersion effects [44] and enhances resolution of the spectra by minimizing the $H_N - H^{\alpha}$ scalar coupling [217]. Very high spin-lock fields were used, up to 30 kHz, however, due to observed artefacts, only results obtained with spin-lock fields up to 20 kHz are discussed. Data were analysed as described in [44].

5.3. Data analysis

Relaxation rates were treated and analysed as described in the manuscript (Chapter 6, section 6.2). We used a value of $r_{NH} = 1.02$ Å For the distance between amide proton and ^{15}N nucleus in agreement with previous publication and works [38, 189, 190]. Assuming that the chemical shift anisotropy (CSA) tensor σ is axially symmetric, the value of $\Delta\sigma = \sigma_{\parallel} - \sigma_{\perp} = -172$ ppm was used in agreement with previous work [190] and in agreement with observations by Canet et al. that the optimal value is around –170 ppm [173].

Before applying the model-free analysis, experimental uncertainties in measured at 278K and 298K relaxation parameters were scaled by a factor 1.4. If the uncertainty value was still below 2% of the relaxation parameter, it was increased to this level.

In order to improve the reproduction of the cross-correlated cross-relaxation rates in the model-free analysis, the θ value in the term $\frac{3\cos^2\theta-1}{2}$ in equations 3.90 and 3.91 for η_{xy} and η_z rates was optimized. θ represents here the effective value that arises due to the motional averaging of the $\frac{3\cos^2\theta-1}{2}$ term. Note that the relaxation interactions were otherwise assumed to describe the same motion, which is thereby assumed to be isotropic. While this is probably not strictly true, it is not expected to compromise the analysis significantly.

Chapter 6. Results

In this chapter, the conformational sampling and dynamics of the disordered C-terminal domain of the nucleoprotein of Sendai paramyxovirus (N_{TAIL} or NT) is described. Timescales of motions and their amplitudes were studied by the extended model-free approach on the basis of an extensive set of relaxation rates measured at multiple magnetic fields, multiple temperatures and in three different length constructs of N_{TAIL}. Reasoned proposals are made concerning the physical nature of the described timescales of relaxation-active motions. Associating slow motional timescale with chain-like motions, we propose an extension of the model-free approach that takes into account common motional modes in the chain. Finally, the measurement of high-power proton $T_{1\rho}$ relaxation and the comparison of measured relaxation rates to previously obtained RDCs in the molecular recognition element (MoRE) of N_{TAIL} allows us to estimate a range for the timescale of conformational exchange between different helical conformers in the MoRE of N_{TAIL}.

6.1 Conformational sampling of $N_{TAIL,L}$ from chemical shifts

Ensembles of $N_{TAIL,L}$ conformers were generated and selected from $^{1}H_N$, ^{15}N, $^{13}C'$, $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ chemical shifts using an iterative algorithm based on Flexible-Meccano and ASTEROIDS (cf. Chapter 1) and described by Jensen et al. [218]. A final ensemble of 200 conformers, for which ensemble-averaged calculated chemical shifts reproduced experimentally measured values (6.14) was selected after 5 iterations, necessary for the convergence of the fit.

Conformational sampling preferences of each residue are shown on Fig. 6.1 B,C. Not surprisingly, residues in the MoRE element of $N_{TAIL,L}$ are predicted to preferentially sample the α-helical region of the Ramachandran space, in agreement with the prediction based on RDC measurements in $N_{TAIL,M}$ [111]. Most residues outside the helical region were found to sample α_R, β_S and β_P regions of Ramachandran space, with slightly increased preference for the former and less preference for the latter. α_L regions are almost exclusively sampled by Gly residues, and, to a lesser extent, by His residues.

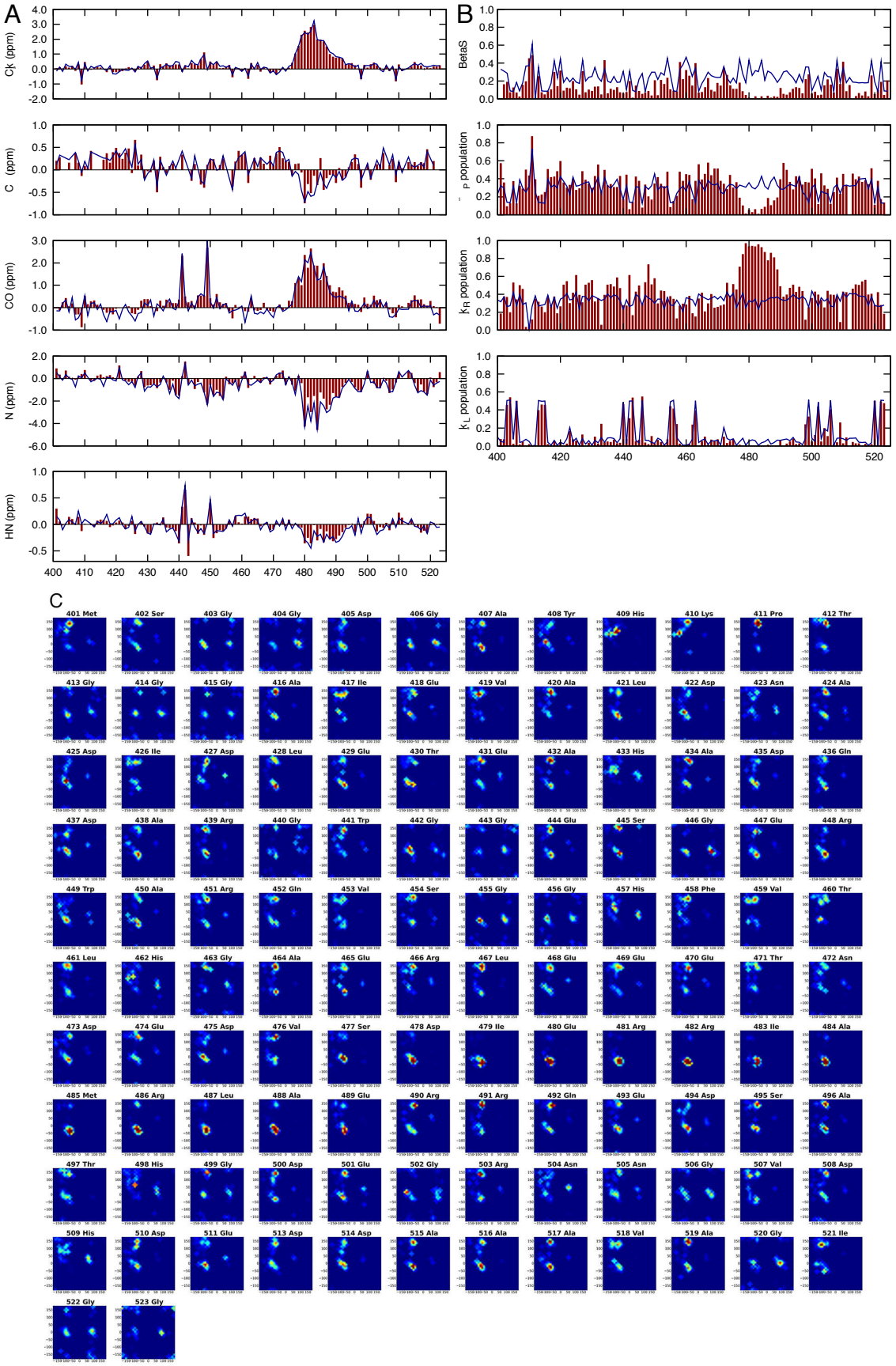


Fig. 6.1. Left column (A): reproduction of experimental chemical shifts in $N_{TAIL,L}$ (blue line) with the ensemble selected by ASTEROIDS (red bars). Right column (B): population of different regions of Ramachandran space from the selection (red bars) and from the starting statistical coil database (blue lines). Bottom plot (C): Amino-acid specific Ramachandran distribution of $N_{TAIL,L}$.

6.2. Initial characterisation of N_{TAIL} at different temperatures

A series of ^{1}H-^{15}N HSQC spectra of $N_{TAIL,L}$ were measured at 303, 298, 293, 288, 283, 278, 273, 268, 263 and 258K (Fig. 6.2). For most residues, with lower temperature, peaks shift towards higher values of ^{1}H and ^{15}N chemical shifts δ_{HN} and δ_N at a constant rate. However, for some peaks a different behaviour is observed: they shift in a different direction, and with a slower rate with respect to temperature.

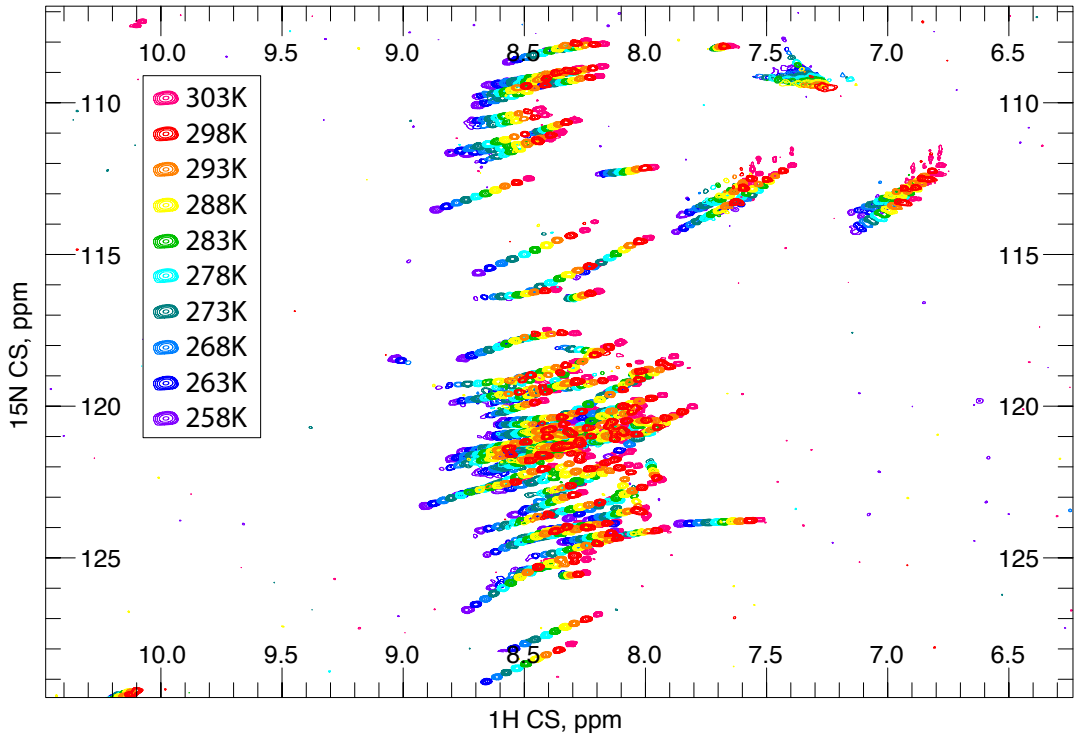


Fig. 6.2. ^{15}N-1H HSQC spectra of $N_{TAIL,L}$ at different temperatures.

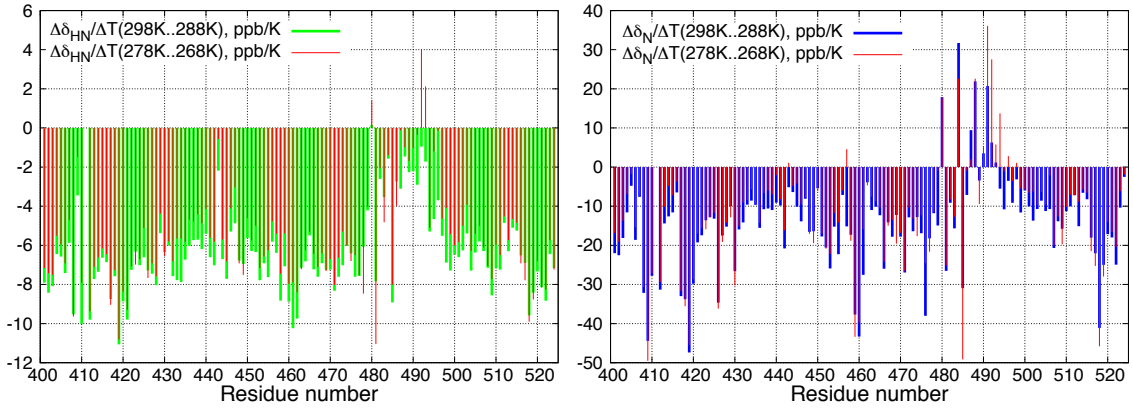


Fig. 6.3. Evolution of δ_{HN} and δ_N chemical shift temperature coefficients with temperature. $\Delta\delta_{HN}/\Delta T$ were estimated as $\delta_{HN}^{T_A} - \delta_{HN}^{T_B}/(T_A - T_B)$, where the temperatures $(T_A..T_B)$ are indicated on the plot. The same estimation was made for $\Delta\delta_N/\Delta T$.

Amide proton and nitrogen temperature coefficients were calculated from the displacement of their chemical shifts in the range 298 – 288K and in the range 278 – 268K (Fig. 6.3). In the molecular recognition element (MoRE) region of $N_{TAIL,L}$ (residues 476-492) low negative or even positive temperature coefficients were observed. Studies of folded proteins with known structures showed that an amide proton temperature

coefficient $\Delta\delta_{HN}/\Delta T > -4.6$ ppb/K is indicative of the participation of that proton in the formation of intramolecular hydrogen bonds [219, 220]. Thus, our observation is in agreement with the predicted transient helical population of the $N_{TAIL,L}$ MoRE.

	Field				
	600 MHz	700 MHz	800 MHz	850 MHz	950 MHz
$N_{TAIL,C}$					
5°C	$R_{1\rho}$, R_1, *NOE*	$R_{1\rho}$, R_1, *NOE*		$R_{1\rho}$, R_1, *NOE*, η_{xy}	η_{xy}
25°C	R_2, R_1, *NOE*		R_2, R_1		
$N_{TAIL,M}$					
-5°C	$R_{1\rho}$, R_1, η_{xy}				
1°C	$R_{1\rho}$, R_1, *NOE*	$R_{1\rho}$, R_1, η_{xy}			
5°C	$R_{1\rho}$, R_1, *NOE*	$R_{1\rho}$, R_1, *NOE*		$R_{1\rho}$, R_1, *NOE*, η_{xy}, η_z	$R_{1\rho}$, R_1, η_{xy}, η_z
15°C	$R_{1\rho}$, R_1, *NOE*, η_{xy}				$R_{1\rho}$, R_1, *NOE*, η_{xy}
25°C	$R_{1\rho}$, R_1, *NOE*, η_{xy}	$R_{1\rho}$, R_1, *NOE*, η_{xy}, η_z		$R_{1\rho}$, R_1, *NOE*, η_{xy}, η_z	$R_{1\rho}$, R_1, *NOE*, η_{xy}, η_z
$N_{TAIL,L}$					
-15°C	R_2				
-5°C	$R_{1\rho}$, R_1, η_{xy}	^1H $R_{1\rho}$ dispersion			
1°C	$R_{1\rho}$, R_1, *NOE*, η_{xy}	$R_{1\rho}$, R_1, *NOE*, η_{xy}		$R_{1\rho}$, R_1, *NOE*, η_{xy}	
5°C	$R_{1\rho}$, R_1, *NOE*, η_{xy}	$R_{1\rho}$, R_1, *NOE*		$R_{1\rho}$, R_1, *NOE*, η_{xy}, η_z	$R_{1\rho}$, R_1, *NOE*, η_{xy}
15°C	$R_{1\rho}$, R_1, *NOE*	$R_{1\rho}$, R_1, *NOE*, η_{xy}, η_z		$R_{1\rho}$, R_1, *NOE*, η_{xy}	
25°C	$R_{1\rho}$, R_1, *NOE*, η_{xy}	$R_{1\rho}$, R_1, *NOE*, η_{xy}		$R_{1\rho}$, R_1, *NOE*, η_{xy}, η_z	$R_{1\rho}$, R_1, *NOE*, η_{xy}, η_z
30°C	$R_{1\rho}$, R_1, *NOE*, η_{xy}	$R_{1\rho}$, R_1, *NOE*, η_{xy}			

Table 6.1. Relaxation data measured on $N_{TAIL,L}$, $N_{TAIL,M}$, and $N_{TAIL,C}$. η_{xy}, η_z are, respectively, the transversal and the longitudinal ^{15}N-^1H dipolar – ^{15}N CSA cross-correlated cross-relaxation rates. Rates in black were measured on a Bruker spectrometer with cryoprobe, in blue – on a Bruker spectrometer without cryoprobe, in orange – on a Varian spectrometer (with cryoprobe, and for ^{15}N R_2 at -15°C, without cryoprobe).

As discussed previously, ^{15}N relaxation is particularly sensitive to the protein backbone motions in ps-ns timescale range. To investigate in detail the dynamical behaviour of an intrinsically disordered protein N_{TAIL}, over 100 different relaxation rates were measured at different temperatures, 5 magnetic fields and for N_{TAIL} constructs of different length (Fig. 6.4). The list of all measured rates is presented in Table 6.1.

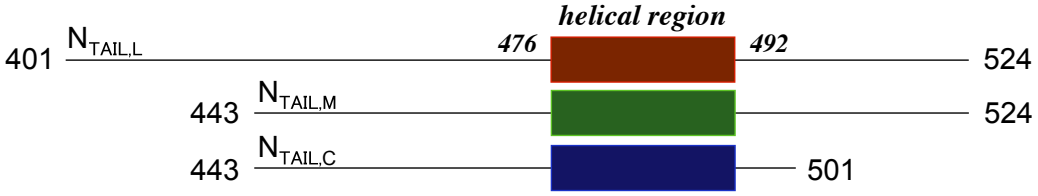


Fig. 6.4. N_{TAIL} constructs of different length. All three constructs share the region with helical population corresponding to the MoRE of N_{TAIL}.

The largest part of our work is presented in an article[2] which is reproduced in the next section of this thesis.

6.3. Article: Identification of Dynamic Modes in an Intrinsically Disordered Protein Using Temperature-Dependent NMR Relaxation

[2] A. Abyzov, N. Salvi, R. Schneider, D. Maurin, R.W.H. Ruigrok, M.R. Jensen and M. Blackledge, *J. Am. Chem. Soc.* **138**, 6240–6251 (2016)

JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

Article

pubs.acs.org/JACS

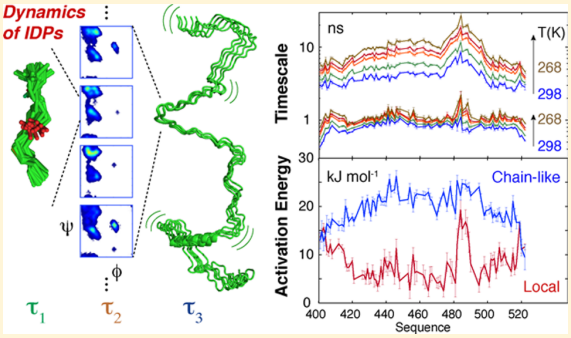
Identification of Dynamic Modes in an Intrinsically Disordered Protein Using Temperature-Dependent NMR Relaxation

Anton Abyzov,[‡] Nicola Salvi,[‡] Robert Schneider,[†] Damien Maurin, Rob W.H. Ruigrok, Malene Ringkjøbing Jensen, and Martin Blackledge*

Institut de Biologie Structurale (IBS), CEA, CNRS, Université Grenoble Alpes, 38044 Grenoble, France

Supporting Information

ABSTRACT: The dynamic modes and time scales sampled by intrinsically disordered proteins (IDPs) define their function. Nuclear magnetic resonance (NMR) spin relaxation is probably the most powerful tool for investigating these motions delivering site-specific descriptions of conformational fluctuations from throughout the molecule. Despite the abundance of experimental measurement of relaxation in IDPs, the physical origin of the measured relaxation rates remains poorly understood. Here we measure an extensive range of auto- and cross-correlated spin relaxation rates at multiple magnetic field strengths on the C-terminal domain of the nucleoprotein of Sendai virus, over a large range of temperatures (268−298 K), and combine these data to describe the dynamic behavior of this archetypal IDP. An Arrhenius-type relationship is used to simultaneously analyze up to 61 relaxation rates per amino acid over the entire temperature range, allowing the measurement of local activation energies along the chain, and the assignment of physically distinct dynamic modes. Fast ($\tau \leq 50$ ps) components report on librational motions, a dominant mode occurs on time scales around 1 ns, apparently reporting on backbone sampling within Ramachandran substates, while a slower component (5−25 ns) reports on segmental dynamics dominated by the chain-like nature of the protein. Extending the study to three protein constructs of different lengths (59, 81, and 124 amino acids) substantiates the assignment of these contributions. The analysis is shown to be remarkably robust, accurately predicting a broad range of relaxation data measured at different magnetic field strengths and temperatures. The ability to delineate intrinsic modes and time scales from NMR spin relaxation will improve our understanding of the behavior and function of IDPs, adding a new and essential dimension to the description of this biologically important and ubiquitous class of proteins.



INTRODUCTION

A significantly high proportion of genomes from all domains of life is predicted to code for proteins that are disordered in their functional state.[1−3] The role of the primary sequence of such intrinsically disordered proteins (IDPs) is not to ensure an energetically stable three-dimensional conformation, as in the case of many folded proteins, but rather to sample a continuum of very different conformations on a broad, flat free-energy surface. The observation of such high levels of functionally important disorder has motivated considerable interest in understanding the link between primary sequence composition, the nature of the Boltzmann ensemble sampled at equilibrium, and eventually biological function. Nuclear magnetic resonance (NMR) spectroscopy offers a particularly powerful set of tools for the atomic resolution characterization of disordered proteins in solution, providing population-weighted averages over all conformations present in the ensemble.[4−7] Chemical shifts, and scalar and dipolar couplings average on time scales faster than 100 μs and can be used to describe the conformational space spanned by the conformational ensemble.[8−17] Crucially, however, little is known about the time scales of the intrinsic dynamics of IDPs.

This lack of information about dynamic time scales is important for a number of reasons. The flexible nature of IDPs strongly suggests that their activity is related to their dynamic behavior, in terms of both time scales of motions and extent of conformational excursions. These descriptors define the conformational behavior of the protein and are necessarily correlated with its function. Therefore, our limited understanding of IDP dynamics, even in the free state, is reflected in the limited understanding of their functional mechanisms. Many theoretical and experimental studies have proposed correlations between intrinsic conformational propensities in free and bound states of IDPs, invoking for example conformational selection from the free-state ensemble, or induced fit, as key binding mechanisms.[18−26] Such considerations however tend to ignore intrinsic interconversion rates between free-state backbone conformations, which may

Received: March 6, 2016
Published: April 26, 2016

DOI: 10.1021/jacs.6b02424
J. Am. Chem. Soc. 2016, 138, 6240−6251

influence molecular association. Similarly, chain-like dynamics will be expected to affect the efficiency of transient secondary interactions, for example via fly-casting mechanisms.[27] Finally, although in principle molecular dynamics (MD) simulation can provide detailed insight into the conformational dynamics of IDPs,[8,28−30] current force fields generally fail to reproduce their macroscopic properties in solution, probably due to the increased importance of protein−solvent interactions.[31−34] The experimental characterization of dynamic time scales is therefore essential for the conception and improvement of force fields that can simulate the behavior of IDPs.

^{15}N spin relaxation[35] offers probably the most powerful and generally accessible tool for the study of the dynamics of IDPs, providing sensitive probes of the motional properties of bond vectors throughout the protein.[36−54] Not surprisingly, spin relaxation is regularly measured, in both free and bound forms of IDPs, and under increasingly relevant physiologically conditions. The presence of segmental motions has been evoked on the basis of the commonly observed bell-shaped dependence of transverse relaxation components, with modulations along the sequence interpreted as a measure of residual order,[36,39,55] for example due to local hydrophobicity,[36,40,45,48,56] or electrostatics.[57,48] Segmental motions in α-synuclein have been proposed on the basis of global ^{1}H relaxometry measurements,[58] while solvent-induced "drag" of long disordered chains has been predicted to affect rotational diffusion properties of partially folded proteins.[59−61] However, despite the abundance of NMR relaxation studies, the physical origin of the measured relaxation rates remains poorly delineated. The aim of this study is to investigate the molecular origin of spin relaxation rates measured in IDPs and, equally importantly, to provide new physical insight into the molecular behavior of IDPs in terms of these essential conformational modes.

In order to achieve this we study the dynamic behavior of the C-terminal domain of the nucleoprotein (NT) of Sendai virus (SeV), a 124 amino acid archetypal IDP, comprising long, unfolded domains and a short molecular recognition motif that interacts with the PX domain of the phosphoprotein.[62,63] Auto- and cross-correlated spin relaxation rates are measured at four magnetic field strengths (600 to 950 MHz ^{1}H frequency) over a large range of temperatures (268−298 K). Up to 61 relaxation rates per site are simultaneously analyzed using an Arrhenius-type relationship relating dynamic time scales over the range of temperatures,[64] allowing assignment of local activation energies of distinct modes along the chain, and identification of the physical origin of the different motional contributions. A fast ($\tau \leq 50$ ps) component reports on librational motions, local backbone sampling of the peptide chain occurs on time scales in the nanosecond range, while the slowest component reports on segmental dynamics dominated by the chain-like nature of the protein. Extending the study to constructs that vary the chain dimensions significantly, supports the assignment of these dominant modes.

THEORETICAL CONSIDERATIONS

The five measured relaxation rates are given by the following functions:[35]

$$R_1 = \frac{1}{10}\left(\frac{\mu_0 \hbar \gamma_H \gamma_N}{4\pi r_{NH}^3}\right)^2 (J(\omega_H - \omega_N) + 3J(\omega_N) + 6J(\omega_H + \omega_N)) + \frac{2}{15}\omega_N^2(\sigma_\parallel - \sigma_\perp)^2 J(\omega_N) \quad (1)$$

$$R_2 = \frac{1}{20}\left(\frac{\mu_0 \hbar \gamma_H \gamma_N}{4\pi r_{NH}^3}\right)^2 (4J(0) + J(\omega_H - \omega_N) + 3J(\omega_N) + 6J(\omega_H + \omega_N) + 6J(\omega_H)) + \frac{1}{45}\omega_N^2(\sigma_\parallel - \sigma_\perp)^2 (4J(0) + 3J(\omega_N)) \quad (2)$$

$$\sigma_{NH} = \frac{1}{10}\left(\frac{\mu_0 \hbar \gamma_H \gamma_N}{4\pi r_{NH}^3}\right)^2 (6J(\omega_H + \omega_N) - J(\omega_H - \omega_N)) \quad (3)$$

$$\eta_{xy} = \frac{1}{15}P_2(\cos\theta)\left(\frac{\mu_0 \hbar \gamma_H \gamma_N}{4\pi r_{NH}^3}\right)(\sigma_\parallel - \sigma_\perp)\omega_N (4J(0) + 3J(\omega_N)) \quad (4)$$

$$\eta_z = \frac{1}{15}P_2(\cos\theta)\left(\frac{\mu_0 \hbar \gamma_H \gamma_N}{4\pi r_{NH}^3}\right)(\sigma_\parallel - \sigma_\perp)\omega_N (6J(\omega_N)) \quad (5)$$

where $J(\omega)$ is the angular spectral density function at frequency ω, and θ is the angle between the principal axis of the chemical shift anisotropy (CSA) tensor (assumed axially symmetric with anisotropy $\sigma_\parallel - \sigma_\perp = -172$ ppm) and the N−H dipole−dipole interaction. r_{NH} is the N−H internuclear distance (assumed to be 1.02 Å), γ_H and γ_N are the gyromagnetic ratio of ^{1}H and ^{15}N nuclei respectively, μ_0 is the permittivity of free space, and $\hbar$ is Planck's constant divided by 2π.

The spectral density function of each relaxation-active vector in highly disordered systems can be mapped by solving for the different characteristic $J(\omega)$ contributing to the different relaxation rates.[65] Reduced spectral density mapping[66−68] is often used to determine $J(0)$, $J(\omega_N)$, and $J(0.87\omega_H)$, where the latter is an effective value representing J at higher frequencies.

In folded proteins, relaxation rates are commonly interpreted using the model-free approach, where contributions from local and global motions to the autocorrelation function are analytically separated,[69−72] the latter being treated as a common contribution for all sites. The concept of common overall motion has little relevance for IDPs, prompting the development of alternative approaches, evoking either a distribution of correlation times[41,44,51,73] or MD-based frameworks for identifying relaxation-active correlation times.[74] Alternatively, the model-free approach can be applied in a general form for each site, with different contributions to the correlation function/spectral density function representing independent motions occurring on distinct time scales. Modeling the distribution of motions with a limited number of contributions (normally two or three) surely oversimplifies the large number of possible modes, but nevertheless provides valuable insight into the variation of dynamic behavior along the chain.[75,76]

As noted by Halle,[72] provided that the time scales are sufficiently separated, the adiabatic approximation holds and, assuming all motions to be isotropic, the correlation function $C(t)$ can be modeled by the sum of exponentially decaying terms:

DOI: 10.1021/jacs.6b02424
J. Am. Chem. Soc. 2016, 138, 6240−6251

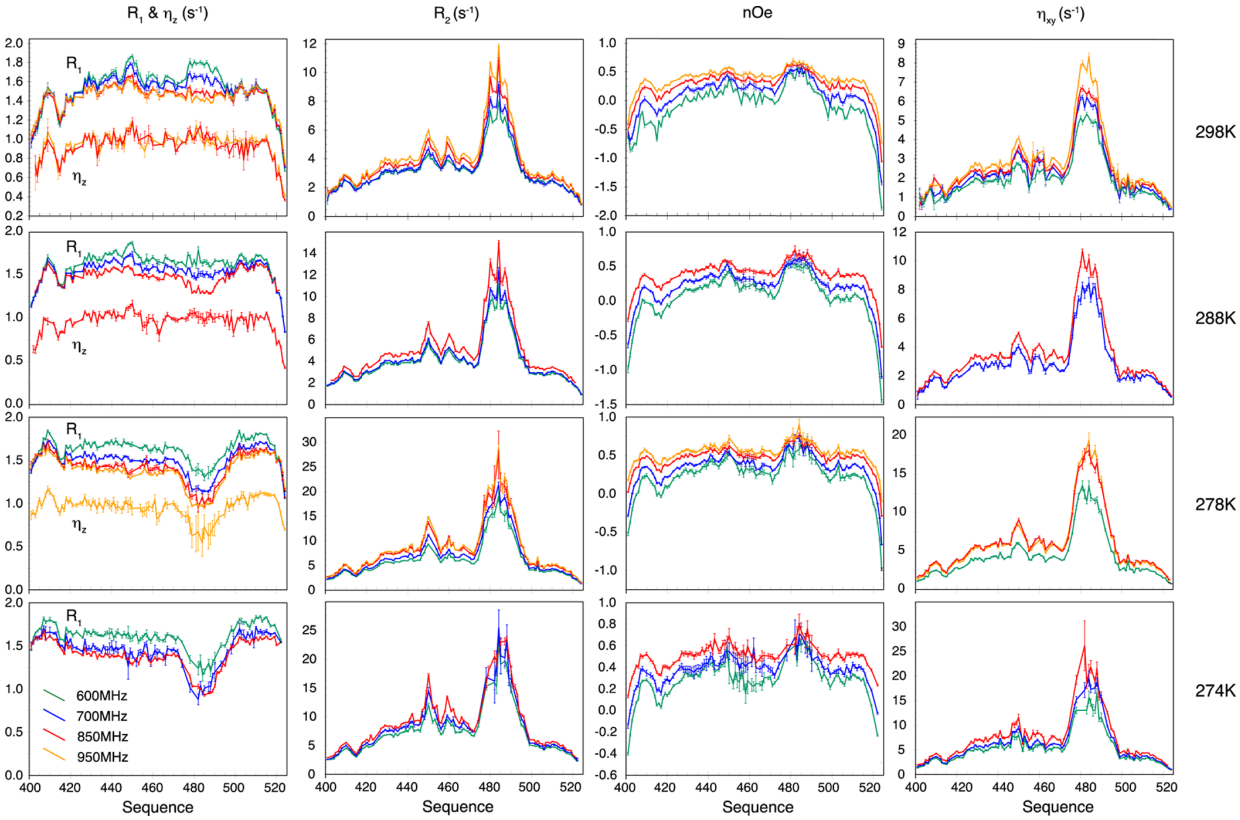


Figure 1. Experimental relaxation rates of NT_L (residues 401−524), measured at different magnetic field strengths (green, 600 MHz ^{1}H frequency; blue, 700 MHz; red, 850 MHz; orange, 950 MHz) and temperatures (top row, 298 K; second row, 288 K; third row, 278 K; bottom row, 274 K). Longitudinal relaxation R_1 and η_z are shown in the same panels; in this case, the lower curves correspond to the cross-correlated relaxation rates. Data were also measured at 268 K (Supporting Information Figure S2).

$$C(t) = \sum_k A_k e^{-t/\tau_k} \tag{6}$$

The requirement of separation of time scales implies that faster contributions decay completely before the slower contributions significantly affect the correlation function (Figure S1). The definition of the amplitudes A_k implies that $\sum_k A_k = 1$.

The Fourier transformation of $C(t)$ yields the spectral density function

$$J(\omega) = \sum_k \frac{A_k \tau_k}{1 + \omega^2 \tau_k^2} \tag{7}$$

In this study, we use an Arrhenius analysis to relate individual dynamic contributions at different temperatures via the following function:

$$\tau_k(T) = \tau_{k,\infty} e^{E_{k,a}/RT} \tag{8}$$

where $E_{k,a}$ represents the local activation energy (Gibbs free energy or free enthalpy) of the process k associated with the different components, and $\tau_{k,\infty}$, the prefactor that represents the correlation time at infinite temperature. This approach allows us to simultaneously analyze relaxation rates measured at multiple fields and temperatures to extract local activation energies, while at the same time increasing the robustness of the fitting procedure involving multiple time scales.

RESULTS

Site-Specific Dynamics of Intrinsically Disordered NT As a Function of Temperature. Longitudinal and transverse auto- and cross-correlated ^{15}N relaxation rates were measured at 600, 700, 850, and 950 MHz (^{1}H frequency) from NT at four different temperatures (Figure 1). Relaxation rates were also collected at 268 K at 600 MHz (Figure S2). Distinct conformational properties of different regions of the protein are immediately discernible. The helical element (476−492) exhibits relaxation rates typical of slower correlation times, with additional local structure apparent around two tryptophans at positions 443 and 451, while a clear dip appears in most relaxation rates around residue 415, indicative of increased flexibility coincident with three consecutive glycine residues. The temperature dependence of the different rates is more explicitly shown in Figure 2 for experiments performed at 850 MHz.

Reduced spectral density mapping was performed for data obtained at each magnetic field strength and at each individual temperature. Comparison of $J(0)$ derived from analysis of relaxation data from different field strengths using autocorrelated relaxation rates (R_1, R_2, and σ_{NH}) and cross-correlated rates (η_{xy}, η_z) indicates that exchange contributions to R_2 are negligible and reveals a high level of consistency between the different data sets (Figure S3). Cross-correlated transverse relaxation rates exhibit similar periodicity along the helical element as residual dipolar couplings (RDCs) measured in NT

DOI: 10.1021/jacs.6b02424
J. Am. Chem. Soc. 2016, 138, 6240−6251

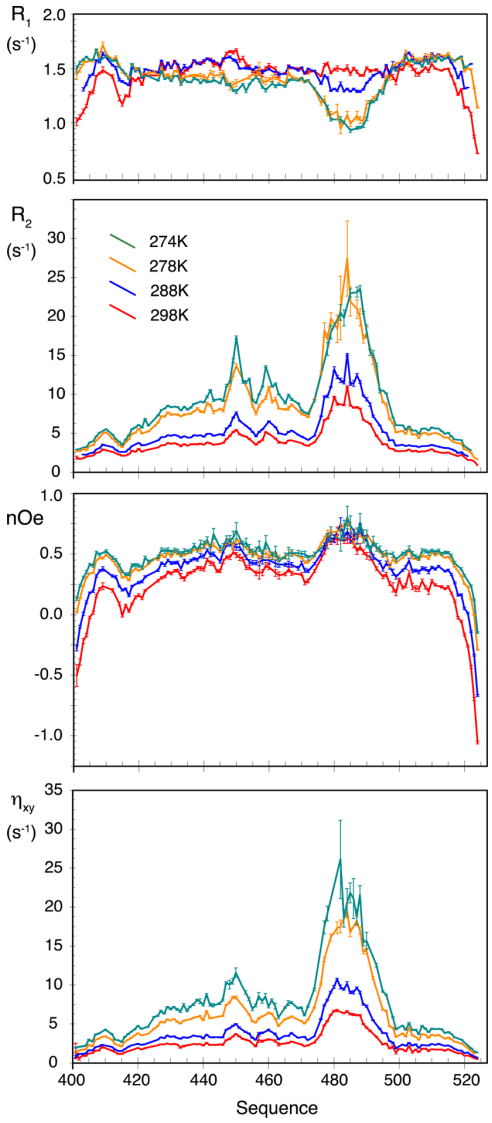


Figure 2. Experimental relaxation rates (R_1, R_2, NOE, and η_{xy}) of NT_L (residues 401−524), measured at 850 MHz ^{1}H frequency (green, 274 K; orange, 278 K; blue, 288 K; red, 298 K).

(Figure S4), which report on distinct helical substates in fast exchange.[77,78] Anisotropic rotational diffusion induces similar orientational dependence in ^{15}N relaxation rates as steric alignment induces in dipolar couplings.[79] Although the time scale of interconversion of the helical substates is unknown, this similar periodicity is consistent with exchange time scales that are too slow to be relaxation-active (but fast on the chemical shift time scale). Our subsequent analysis assumes that the measured rates therefore report on averages over relaxation occurring in each helical state.

We note that the backbone conformational propensities remain stable with temperature in the unfolded regions of NT, while the predominantly helical region (residues 476−492) shows a slight increase in helical propensity (around 5−7%) at lower temperatures as judged from temperature-corrected[80] $^{13}C^{\alpha}$ chemical shifts (Figure S5).

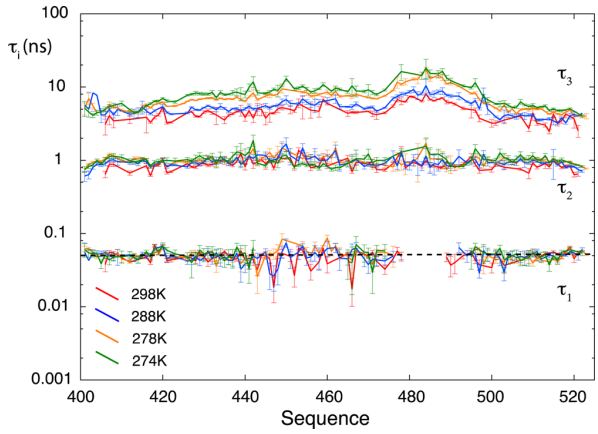


Figure 3. Sequence dependence of the correlation time of the fast (τ_1), intermediate (τ_2), and slow (τ_3) correlation times in NT_L for a three-exponential model of the autocorrelation function using the model-free approach with the fast motional time scale as a free parameter, except in the helix where τ_1 is set to 0. Analysis is shown for data measured at 298 K (red), 288 K (blue), 278 K (orange), and 274 K (green). The dashed line indicates a value of 50 ps. No obvious progression of the fast component is seen as a function of temperature. The *y*-axis is shown on a logarithmic scale.

Identification of Appropriate Dynamic Models of the Spectral Density Function. Data measured at 274, 278, 288, and 298 K were independently analyzed by minimizing eq 10, assuming one, two, or three exponential contributions to auto- and cross-correlation functions (eqs 6,7). Statistical analysis (Figure S6) is consistent with the use of a six-parameter (A_2, A_3, τ_1, τ_2, τ_3, and θ) model throughout the protein and a five-parameter (A_2, A_3, τ_2, τ_3, and θ with $\tau_1 = 0$) model in the helix (residues 476 to 492). The time scales of the three different contributions resulting from this analysis are shown in Figure 3 for the four temperatures (274, 278, 288, and 298 K) for which (12, 14, 14, and 18) different relaxation rates were measured, respectively. Outside the helical region, the faster time scale motion is closely distributed around 50 ps at all four temperatures. Indeed fixing this component to 50 ps has negligible impact on the motional amplitudes (Figures 4 and S7).

The slower time scale motion varies through a range from around 5 to 25 ns over the whole temperature range, with clear maxima in the helical region, while the intermediate time scale motion ranges from 0.5 to 1.3 ns. Not surprisingly the relative contribution of the slowest component increases as temperature drops, with a concomitant reduction in the faster component, in particular in the helical element. In order to investigate these effects in more detail we have analyzed data from all four temperatures simultaneously (*vide infra*). The derived site-specific dipole−dipole/CSA interaction angle θ corresponds to an effective value reproducing the ensemble-averaged $P_2(\cos\theta) = \frac{3\cos^2\theta - 1}{2}$ in eqs 4 and 5, which may absorb a number of effects, for example local anisotropies that are not accounted for in our analysis. θ values derived at 298, 288, and 278 K generally fluctuate around 22°−25° in the unfolded region (Figure S8), with lower values occurring in the helical region, in line with studies of folded proteins.[81−83] Repeating the analysis with fixed values of θ (22°) results in essentially identical motional parameters, inducing small

DOI: 10.1021/jacs.6b02424
J. Am. Chem. Soc. 2016, 138, 6240−6251

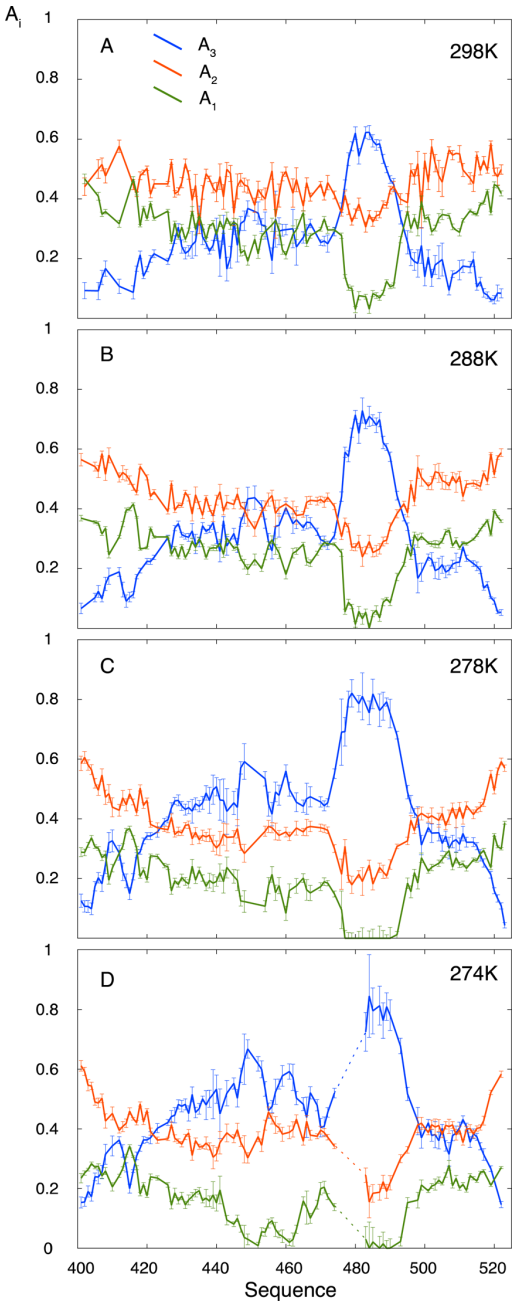


Figure 4. Amplitude of the motional contributions A_1 (fast contribution, shown in green), A_2 (intermediate time scale contribution, shown in orange), and A_3 (slowest contribution, shown in blue) in NT_L at the four temperatures, 298 K (A), 288 K (B), 278 K (C), and 274 K (D). The fast component clearly falls in amplitude with decreasing temperature, while the slow contribution increases. The dashed line in Figure D corresponds to residues for which insufficient data were available to characterize the dynamics with sufficient accuracy.

discrepancies between predicted and calculated η_{xy} and R_2 values. No correlation was found between θ and the motional parameters A_i and τ_i. The combined analysis of data at multiple temperatures assumes a single value for θ for each site across all temperatures.

Site-Specific Temperature Dependence of the Spectral Density Function. All data from all temperatures were fitted using a single function with three motional components, characterizing intermediate and slow time scale motions with Arrhenius-type dependences for each site (see Theoretical Considerations). This requires the determination of $\tau_{2,\infty}$, $\tau_{3,\infty}$, $E_{2,a}$, and $E_{3,a}$ where 2 and 3 refer to the two components respectively, θ for each site, and A_2, A_3 for each temperature. The fast time scale motion, which shows negligible temperature dependence (see above), was fixed to 50 ps throughout the protein and 0 ps in the helical region at all temperatures. A total of 58 relaxation rates were fitted to these 13 parameters for 86 resonances that were resolved at all temperatures. The results are summarized in Figure 5, with data reproduction shown in Figure S9. θ values show a smooth distribution along the primary sequence with limited variance ($23.5 \pm 2.5°$) and lower tendencies in the helical region (Figure S10).

The time scale of the slower contribution exhibits a distinct bell-shaped distribution, while the intermediate time scale is overall flatter, with deviations to slower time scales in the helical region. The steeper temperature dependence of the slower time scale motion is reflected in the higher effective activation energy (reaching 20−25 kJ mol^{-1} in the center of the chain) compared to the intermediate time scale motion. The activation energy of the slower time scale motion also exhibits a broad bell-shaped sequence dependence, encompassing the helical region with only a minor increase in the range 476−492, while the intermediate motion in the region 425−500 shows a flatter distribution (around 5 kJ mol^{-1}) that is clearly interrupted by the helical element, where activation energies are locally increased to 15 kJ mol^{-1}. Motional amplitudes are very similar to those determined from the individual temperature analysis. Interestingly the intermediate time scale motion has a maximum contribution at 288 K. Although only three rates could be accurately measured at 268 K, the five-temperature Arrhenius analysis (determination of 15 parameters from 61 relaxation rates) is fully consistent with the four-temperature analysis, resulting in very similar trends for A_1, A_2 and A_3 as well as τ_2 and τ_3 (Figure S11).

We note that the temperature dependence of tabulated values of the viscosity for 500 mM NaCl solution[84] compare closely with the observed temperature dependence of the time scale of the slow component (Supporting Information Figure S12). The effective temperature coefficient is predicted to be 20.5 kJ mol^{-1}, suggesting that the slower component detected in the relaxation rates is dominated by solvent viscosity. Intermediate motions, whose activation energy is much lower than the solvent-dominated temperature coefficient, are more convincingly associated with internal modes.

Robustness of the Analysis. In order to determine the robustness of this analysis we have performed the following cross-validations: 10% of all experimental data were randomly removed, and the analysis was repeated in the absence of these "passive" data points, whose values were predicted on the basis of the analysis of the active data. The results, shown in Figure 6A for three separate random selections, give a remarkably good reproduction of the passive data (Pearson's correlation coefficient $r > 0.995$). An additional plot showing the close reproduction of the rates across the sequence is included in the Supporting Information (Figure S13). Second, we have tested the dependence of the analysis on available magnetic field

DOI: 10.1021/jacs.6b02424
J. Am. Chem. Soc. 2016, 138, 6240−6251

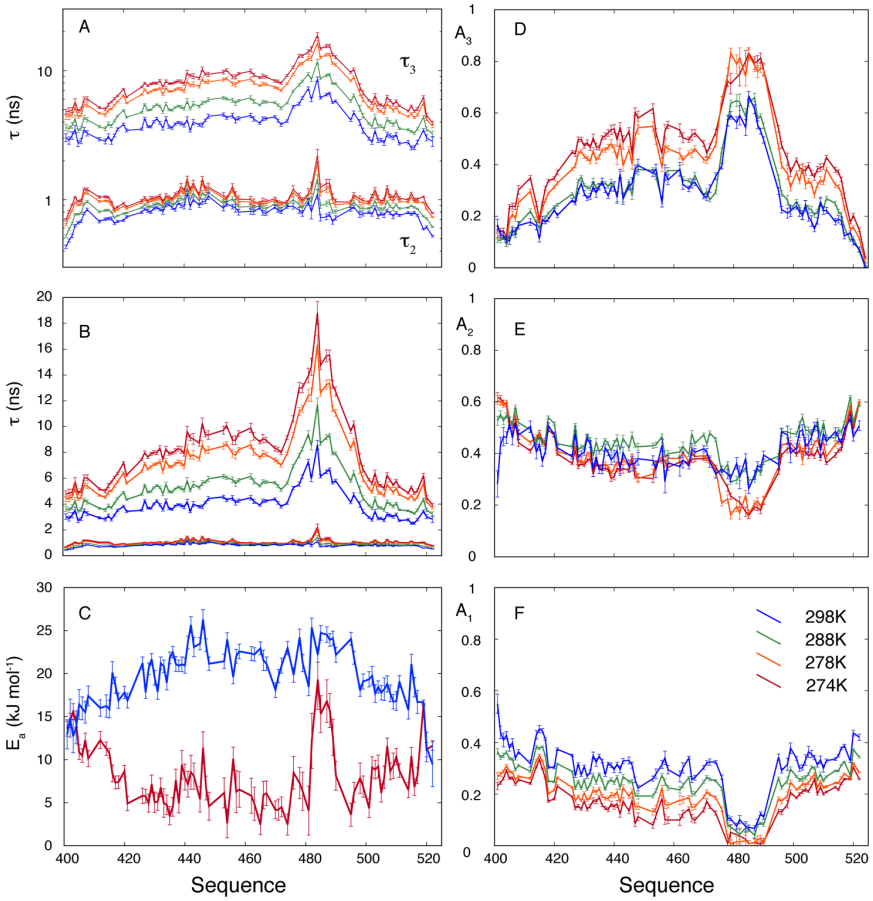


Figure 5. Simultaneous analysis of relaxation data measured over a range (274−298 K) using a site specific Arrhenius type relationship to model the characteristic correlation times of intermediate and slow motional modes in NT_L. (A, B) Intermediate (τ_2) and slow (τ_3) correlation times at 298 K (blue), 288 K (green), 278 K (orange), and 274 K (red). τ_3 exhibits a clear bell-shaped distribution, while τ_2 is much flatter over the sequence. The same data are shown in both plots, in (A) a logarithmic scale is used. (C) Effective activation energies describing the temperature coefficient of the motional time scales for each amino acid. Blue and red curves show the temperature coefficients of the slow (τ_3), and intermediate (τ_2) components, respectively. (D−F) Amplitude of the motional contributions A_3 (D, slowest contribution) A_2 (E, intermediate time scale contribution) and A_1 (F, fast contribution) at the four temperatures (color scheme as that in A, B). Error bars in all fits are derived from 200 noise based Monte Carlo simulations. Figure S11 shows the same fitting procedure applied to data from five temperatures simultaneously (268−298 K).

strengths by repeating the analysis with all data removed (at all temperatures) for each magnetic field strength independently. Time scales, amplitudes, and effective activation energies determined from only three fields reproduce values determined from the full analysis remarkably well, irrespective of the field that is removed from the analysis (Figure S14). Data from the removed field strength are again accurately predicted (Figure 6B) by the analysis using only three fields (Pearson's correlation coefficient $r > 0.996$ in all cases). This robustness is also remarkable because, in the latter case, up to 31% of the data are removed compared to the original analysis.

Dependence of Relaxation Rates on Chain Length. To investigate the segmental, or chain-like, motions further we have measured relaxation rates for two shorter constructs of the protein, lacking the first 43 amino acids (NT_M), and the first 43 and the last 23 amino acids respectively (NT_S) compared to the full length (NT_L). Typical rates measured at 278 K are compared in Figure 7, showing strong similarity between rates in constructs with common termini, from the helical region to the C-terminus in NT_M and NT_L and from the N-terminus to the helical region in NT_S and NT_M. Applying the same motional model to data sets measured at 278 K for all three constructs (10 rates for NT_S, 12 rates for NT_M, and 14 rates for NT_L, Figure 8) reveals a similar distribution of the different amplitudes in the common regions of NT_S, NT_M, and NT_L, in particular in the helical element, accompanied by a close reproduction of the intermediate time scale motion τ_2. The slower motional time scale τ_3 exhibits a clear length-dependence on the helical element, with NT_L displaying the slowest motions and NT_S showing components that are approximately 5 ns faster than the same residues in NT_L. Although error bars are large, the data consistently show the same features throughout the helical element.

DISCUSSION

The dynamic modes and time scales sampled by intrinsically disordered proteins (IDPs) define their physical nature and therefore their activity. For this reason the establishment of a framework to describe the conformational dynamics of IDPs remains a key challenge for understanding their mechanistic

DOI: 10.1021/jacs.6b02424
J. Am. Chem. Soc. 2016, 138, 6240−6251

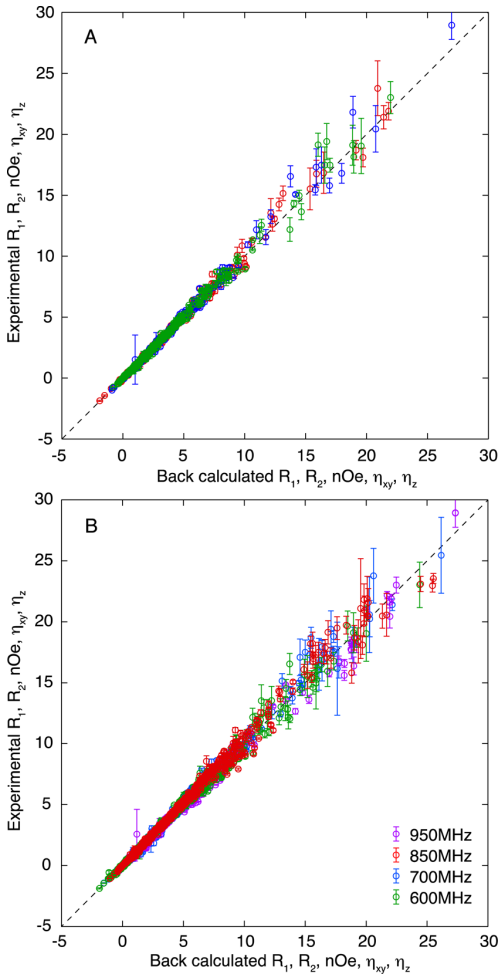


Figure 6. Cross-validation of the Arrhenius analysis of the multifield/multitemperature relaxation data. (A) Reproduction of experimental relaxation rates/enhancements using a cross-validation approach whereby 10% of all data were randomly removed and predicted from an Arrhenius analysis treating only the remaining 90%. Three random selections were performed, and results from all three are shown for the points removed in each case (red, green, and blue circles, $r > 0.995$ in all cases). Sequence dependence is shown in Figure S13. (B) Reproduction of experimental relaxation rates/enhancements using a cross-validation approach whereby all data points from each magnetic field strength (green, 600 MHz comprising 1230 relaxation measurements; blue, 700 MHz comprising 1402 relaxation measurements; red, 850 MHz comprising 1480 relaxation measurements; purple, 950 MHz comprising 787 relaxation measurements) were independently removed from the analysis and predicted from the fit of the remaining data using the Arrhenius analysis. $r > 0.996$ in all cases. Time scales, amplitudes, and effective activation energies derived from these analyses are very similar to those from the analysis using all available data (Figure S14).

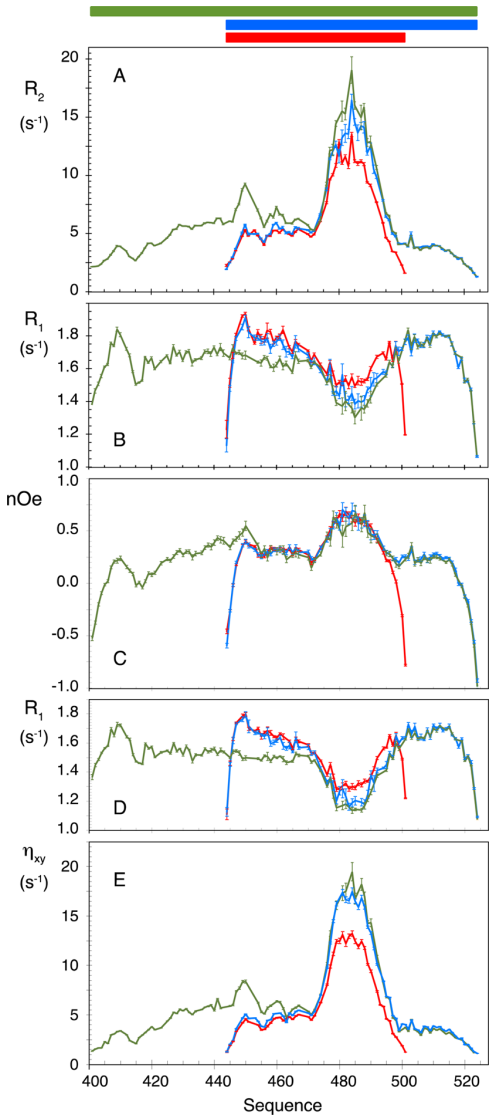


Figure 7. Length dependence of relaxation in NT. Comparison of representative relaxation rates measured at 278 K for NT_L (residues 401−524, green lines) NT_M (residues 444−524, blue lines) and NT_S (residues 444−501, red lines). (A) R_2 measured at 600 MHz. (B) R_1 measured at 600 MHz. (C) Heteronuclear NOE measured at 600 MHz. (D) R_1 measured at 700 MHz. (E) η_{xy} measured at 700 MHz.

and functional behavior. Although NMR relaxation represents the most accessible atomic resolution probe of protein dynamics, the assignment of the physical origin of the observed relaxation rates in IDPs remains obscure. In this study we have measured an extensive set of auto- and cross-correlated relaxation rates that offer complementary probes of different combinations of $J(\omega)$. These parameters were measured at multiple temperatures, thereby providing access to parameters that are sensitive, not only to kinetics, but also to the thermodynamic origin of the motions. Analysis of NT_L data measured at individual temperatures demonstrates that three independent contributions to the spectral density function are necessary to reproduce the experimental data. The time scale of the fastest component (characterized by τ_1) shows essentially no temperature dependence (Figure 3), with a pervasive contribution around 50 ps throughout the protein, except for the helical element where the fastest component is too fast to be determined. Lowering the temperature reduces the efficiency of relaxation from faster motions. The evolution of the parameters describing motion on time scales characterized by the intermediate (τ_2) and slow (τ_3) components is more complex, and in order to better understand this evolution we

DOI: 10.1021/jacs.6b02424
J. Am. Chem. Soc. 2016, 138, 6240−6251

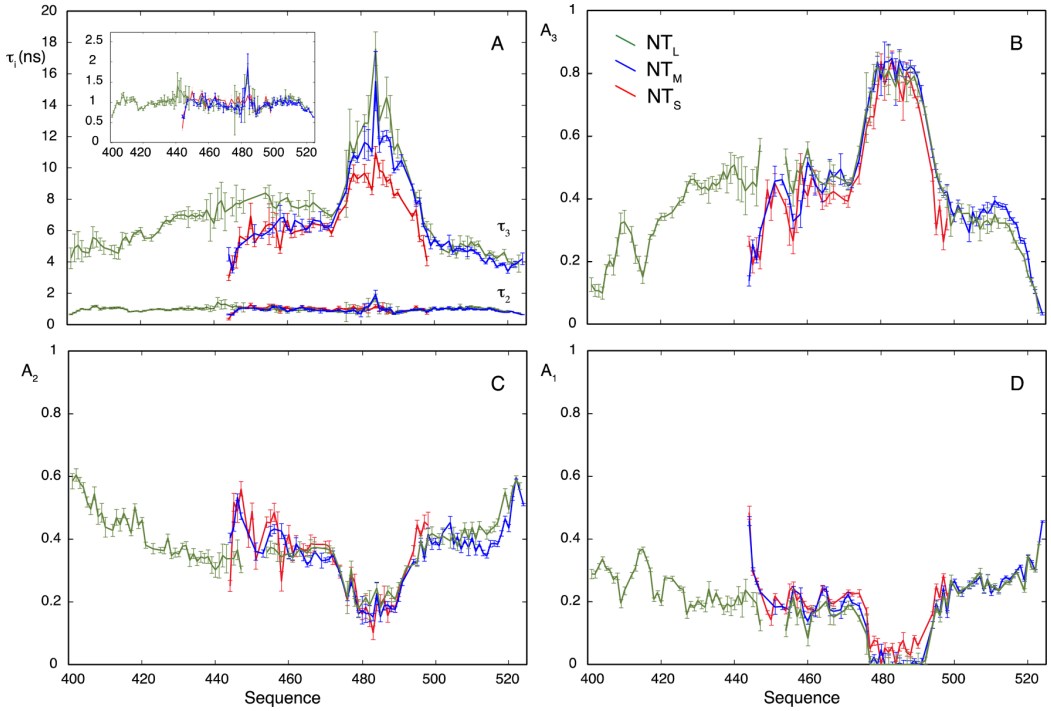


Figure 8. Comparison of the motional characterization of relaxation data measured for NT_L (green), NT_M (blue), and NT_S (red) with the three-exponential model of the autocorrelation function using the model-free approach at 278 K. (A) Characteristic intermediate (τ_2) and slow (τ_3) correlation times as a function of sequence. Inset shows only (τ_2) on a logarithmic scale. Error bars are derived from 200 noise based Monte Carlo simulations. (B–D) Amplitude of the motional contributions A_3 (B), A_2 (C), and A_1 (D) for the three chain lengths (color scheme as in A). Error bars again are derived from 200 noise based Monte Carlo simulations.

have simultaneously analyzed data for all temperatures at each site in the protein.

Site-specific Arrhenius analysis convincingly reproduces all 58 rates using only 13 fitted parameters (Figure S9) and provides considerable insight into the evolution of the different contributions. Remarkable correlation is seen between the amplitudes derived at the different temperatures (Figure 5). For example the GGG sequence at residue 415 shows increased amplitude fast motions (A_1), and the region between residue 424 and 438 exhibits a clear oscillation of contributions from both fast and intermediate time scale components (A_1, A_2) along the sequence (Figure 5E,F). This stretch corresponds to the primary sequence ^{424}ADIDLETEAHADQDA^{438}, where a repetitive alternation between hydrophobic and charged side chains dominates. The distribution of charged side chains, and their effect on the dynamics and function of IDPs, is a subject of current interest.[85] Here we appear to detect dynamic behavior that depends directly on the composition of the primary sequence.

In order to assess the stability of the approach, it is important to determine how accurately the model can predict independent data. This kind of cross-validation is rarely applied to spin relaxation measured in proteins; however, in view of the volume of available data at our disposal it is possible to rigorously apply such approaches, allowing us to estimate the predictive nature of the approach. The simultaneous Arrhenius analysis is shown to be remarkably robust with respect to removal of data and their back-prediction from fits to the remaining data points (Figure 6), both for random selections of data and the entire removal of data from any single magnetic field (up to 31% of the total data set).

The Arrhenius analysis clearly separates slow and intermediate time scale contributions on the basis of their effective activation energies (Figure 5C). The slower time scale motion exhibits the strongest temperature dependence that coincides with expected values for viscosity-dominated rotational diffusion (Figure S12) and is consistently bell-shaped with respect to both time scale (Figure 5A,B) and contribution to the spectral density function (A_3) (Figure 5D). The intermediate time scale motions sample a more restricted range around 0.5–1.5 ns, which account for approximately 50% of the angular decorrelation in the disordered regions (Figure 5E), and are much flatter across the protein at all five temperatures, with the exception of the helical element. Intriguingly, in all nonhelical regions, A_2 exhibits a clear maximum at 288 K.

Putting these observations together, we can develop the following description. The fastest motional contribution occurs on a time scale of tens of picoseconds and exhibits negligible temperature dependence. These motions appear to be essentially librational, reporting on diffusion in an effectively flat potential, whose geometric limits apparently increase with increasing temperature, but which contains no detectable conformational energy barrier.

The flatness of both time scale and amplitude of the intermediate motions over the primary sequence, with the exception of the helical element and the termini, suggests that this component is largely independent of the chain-like nature of the protein and rather reports on local backbone sampling

DOI: 10.1021/jacs.6b02424
J. Am. Chem. Soc. 2016, 138, 6240–6251

dictated by the amino acid of interest and its immediate neighbors. The sequence-dependent activation energies of the intermediate time scale contribution fall in the range of around 5 kJ mol^{-1}, which coincides with values derived from both theoretical and UV-based studies of the free energy landscape of polypeptides.[86,87] Based on theoretically expected barriers between substates (on the order of 15−20 kJ mol^{-1}),[32,34] our observations appear to report on diffusion within distinct Ramachandran minima (e.g., α-helical, polyproline II, β-strand) in the central part of the protein (420−510). Within 15 amino acids of the termini, values gradually increase. It seems likely that this reflects more complex sampling of the backbone in the presence of increased degrees of freedom at the chain termini. How exactly this impacts the effective activation energy of nanosecond motions is not clear and will probably require the use of MD simulation to propose a molecular origin of the phenomenon. The increase in $E_{a,2}$ is accompanied by a reduction in $E_{a,3}$ at both N- and C-termini, which becomes most noticeable beyond ^{414}GGG^{416} and ^{520}AG^{521}, suggesting that the phenomenon reports on segments of the chain that are gradually decoupled from the rest of the chain, until $E_{a,2}$ and $E_{a,3}$ converge at the termini, indicating that when the observed site is no longer part of a segmental chain, the two distinct contributions are no longer necessary. Importantly, the activation energy of the intermediate motions increases abruptly from 5 to 15−20 kJ mol^{-1} in the helical region, probably due to the increased conformational restriction in the helix, with higher activation energies associated with partial unfolding or constrained motion within the secondary structural element.[88,32,89,34] The average activation energies of backbone dynamics in a folded crystalline protein[64] that would be expected to be higher than those in a fluctuating helix in an IDP, were assigned values around 30 kJ mol^{-1}, placing the currently estimated values in the expected range.

Finally the slowest time scale component is assigned to segmental or chain-like motions: the dynamic parameters are bell-shaped, indicating a stronger dependence on segments and longer persistence lengths. The temperature dependence reproduces expected solvent viscosity behavior. End-effects extend to around 15 amino acids from the termini, supporting the idea that the slower component mainly reflects segmental motions within which modes are coupled, both to local residues and to the immediate solvation shell. In the helical element we observe a small but systematic dependence of the time scale of this component on construct length (Figure 8), while all other descriptors remain the same. It seems likely that this effect is only visible in the helical regions and not elsewhere, because the autocorrelation functions in the unfolded regions are more efficiently quenched by extensive faster time scale motions. The viscosity-dominated drag effect is probably related to the observation of longer effective rotational correlation times of folded domains when disordered domains are present, making them apparently more "sluggish" in solution.[59]

The sensitivity of spin relaxation to slower motions in IDPs is entirely dependent on the efficiency of fast motions to quench the angular correlation function, so that in the helix, where this sensitivity is maximal, it is possible to detect motions as slow as 20 ns at 274 K (and 25 ns at 268 K). In this respect, engineering short stable helices into long IDPs may provide a sensitive approach to the detection of such slower motions. These considerations are also important for example for the study of relaxation of IDPs in more complex environments, where slower chain-like motions may affect effective relaxation characteristics of prestructured linear motifs, even in the absence of direct interaction with partner proteins.

CONCLUSIONS

In conclusion, the measurement of an extensive set of relaxation rates not only at multiple magnetic fields but also at multiple temperatures and in three different length constructs of the same IDP has allowed us to characterize the dynamic nature of SeV NT in unprecedented detail. Simultaneous analysis of up to 61 relaxation rates per amino acid assigns local activation energies along the chain and identifies three physically distinct dynamic modes, reporting on librational motions, local backbone sampling, and chain motions. The analysis is shown to be statistically highly robust and to accurately predict relaxation data measured at different magnetic field strengths and temperatures, providing a novel framework for the description of the dynamic properties of IDPs. The ability to identify intrinsic modes and time scales at atomic resolution from NMR spin relaxation will contribute significantly to our understanding of the behavior and function of IDPs, while providing a new and essential dimension to our description of this biologically important and ubiquitous class of proteins.

EXPERIMENTAL SECTION

Sample Preparation. Three constructs of the NT domain, comprising residues 401−524 (NT_L), 444−524 (NT_M), and 444−501 (NT_S) of the nucleoprotein of Sendai virus strain Harris, whose sequence corresponds to that of SeV Fushimi strain NT (UniProtKB accession number Q07097) except for mutation E410 K, were expressed and purified as described previously.[63,77] NMR was performed at protein concentrations of 500 μM in 50 mM sodium phosphate buffer at pH 6.0 with 500 mM NaCl containing 10% D_2O (v/v).

NMR Spectroscopy. The assignment of the longest construct of NT (residues 401−524) was obtained at 298 K using a set of six BEST-type triple resonance[90] correlating Cα, Cβ, and C′ acquired on a Varian spectrometer operating at a ^{1}H frequency of 600 MHz. Spectra were processed in NMRPipe[91] and analyzed in Sparky,[92] and MARS[93] was used to obtain sequential connectivities. The assignment of this construct was also obtained at 278 K using similar experiments and procedures. The assignments at 298 K of the shorter constructs of NT were obtained previously.[63,77]

NMR relaxation experiments were performed on Bruker spectrometers operating at ^{1}H frequencies of 600, 700, 850, and 950 MHz. Relaxation rates were measured for NT_L, NT_M, and NT_S, in identical buffer, at 298, 288, 278, 274, and 268 K. Spectral assignments at temperatures between 278 and 298 K were obtained by following the evolution of ^{1}H−^{15}N HSQC spectra with temperature (steps of 5 K). ^{15}N longitudinal relaxation (R_1) and {^{1}H}-^{15}N heteronuclear NOE and $R_{1\rho}$ (using a spin lock of 1.5 kHz) were measured as described by Lakomek et al.[94] ^{15}N transverse relaxation (R_2) was determined from $R_{1\rho}$ and R_1, taking into account resonance offset.[95] ^{15}N−^{1}H CSA/DD cross-correlated transverse (η_{xy}) and longitudinal (η_z) relaxation rates were measured as described by Pelupessy et al.[96,97]

An interscan delay of 1.5−3 s and 64−128 dummy scans were used for measuring $R_{1\rho}$, R_1, η_{xy}, and η_z rates. The typical set of relaxation delays included [1, 20, 50, 70, 90, 130, 170, 210, 230] ms for $R_{1\rho}$ experiments and [0, 0.6, 0.08, 1.6, 0.4, 1.8, 1.04, 0.8, 0.2, 0.6] s for the R_1 experiment. Shorter delays, [1, 10, 20, 35, 50, 70, 90, 120] ms, were used for measuring the $R_{1\rho}$ relaxation rate at 268 K. For η_{xy} and η_z, relaxation delays of 50 and 100 ms were used, respectively.

Temperature was individually calibrated on each spectrometer using a 99.8% methanol-d_4 sample by measuring the chemical shift difference $\Delta\delta$ between OH and CH_2 peaks in the ^{1}H spectrum and using the following equation:

$$T = -16.7467\cdot(\Delta\delta)^2 - 52.5130\cdot\Delta\delta + 419.1381 \quad (9)$$

DOI: 10.1021/jacs.6b02424
J. Am. Chem. Soc. 2016, 138, 6240−6251

where $\Delta\delta$ is the chemical shift difference in ppm. Significant sample heating during the relaxation experiments was excluded by monitoring the position of the cross peaks.

All spectra were processed in NMRPipe[91] and analyzed in CCPN.[98] The data set measured at supercooled temperatures (268 K) was obtained by employing 3 mm NMR tubes. The high salt concentration (500 mM) prevents the solution from freezing at this temperature.

Fitting Procedure. A Levenberg−Marquardt algorithm was used to fit the data to the expressions in eqs 1−8 using the definition of the spectral density function in eq 7, by minimizing the following function for each residue i:

$$\chi_i^2 = \sum_{n=1}^{5}\sum_{m=1}^{N}\left\{\frac{(R_{n,\mathrm{expt}}^m - R_{n,\mathrm{calcd}}^m)}{\sigma_{n,\mathrm{expt}}^m}\right\}^2 \quad (10)$$

where n identifies the different rates and m identifies the different temperatures used in the fit (one, four, or five temperatures). When only one temperature is considered, (τ_1, τ_2, A_2, and θ), (A_2, A_3,τ_1, τ_2, τ_3, and θ), or (A_2, A_3, τ_2, τ_3, and θ) were optimized using a nonlinear least-squares fitting approach. When multiple temperatures are analyzed simultaneously, A_2 and A_3 are optimized for each temperature, together with $\tau_{2,\infty}$, $\tau_{3,\infty}$, $E_{a,2}$, $E_{a,3}$, and θ.

Several models were tested when analyzing relaxation at each temperature: a model containing three motional time scales with six fitting parameters (A_2, A_3, τ_1, τ_2, τ_3, and the angle θ; $A_1 = 1 - A_2 - A_3$); a model with three motional time scales in which τ_1 was either fixed at 50 ps (for the unfolded chain) or set to 0 (for the helical region), resulting in five parameters (A_2, A_3, τ_2, τ_3, and θ); a model with two motional modes and four parameters (τ_2, τ_3, A_3, and θ, while $A_2 = 1 - A_3$), and finally a model with one motional mode and two parameters (τ and θ).

Statistical Analysis. Errors were estimated using explicit noise-based Monte Carlo simulations in all cases. The results were concordant. Experimental errors were estimated from repeat measurements. In the case of data from 298 and 278 K, experimental errors were scaled by a factor of 1.4, so that the total target function for each individual temperature was within 95% confidence limits. The results of the fit to different models were compared using the corrected Akaike Information Criterion.[99]

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/jacs.6b02424.

Supporting figures showing comparison of secondary structure as a function of temperature, relaxation rates measured at 268 K, results of reduced spectral density mapping, statistical analysis of model-free treatment, three-component analysis with variable τ_1, analysis of θ as a function of sequence, Arrhenius analysis of data from 268 to 298 K, viscosity dependence of 500 mM saline solution as a function of temperature (PDF)

■ AUTHOR INFORMATION

Corresponding Author

*martin.blackledge@ibs.fr

Present Address

†Université Lille 1, UMR CNRS 8576, 50 avenue du Halley, 59650 Villeneuve d'Ascq, France.

Author Contributions

‡A.A. and N.S. contributed equally.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was funded by the Agence Nationale de Recherche under ComplexDynamics (SIMI7 - M.B.) and NMRSignal (JCJC - M.R.J.) and by the Human Frontier Science Program (long-term fellowship LT000322/2011-L to R.S.). N.S. acknowledges funding from a Swiss National Science Foundation Early Postdoc Mobility Fellowship P2ELP2_148858. This work used the platforms of the Grenoble Instruct Center (ISBG: UMS 3518 CNRS-CEA-UJF-EMBL) with support from FRISBI (ANR-10-INSB-05-02) and GRAL (ANR-10-LABX-49-01) within the Grenoble Partnership for Structural Biology (PSB).

■ REFERENCES

(1) Uversky, V. N. *Protein Sci.* **2002**, *11*, 739−756.
(2) Dyson, H. J.; Wright, P. E. *Nat. Rev. Mol. Cell Biol.* **2005**, *6*, 197−208.
(3) Tompa, P.; Davey, N. E.; Gibson, T. J.; Babu, M. M. *Mol. Cell* **2014**, *55*, 161−169.
(4) Dyson, H. J.; Wright, P. E. *Chem. Rev.* **2004**, *104*, 3607−3622.
(5) Forman-Kay, J. D.; Mittag, T. *Structure* **2013**, *21*, 1492−1499.
(6) Jensen, M. R.; Ruigrok, R. W.; Blackledge, M. *Curr. Opin. Struct. Biol.* **2013**, *23*, 426−435.
(7) Rezaei-Ghaleh, N.; Blackledge, M.; Zweckstetter, M. *ChemBioChem* **2012**, *13*, 930−950.
(8) Sgourakis, N. G.; Yan, Y.; McCallum, S. A.; Wang, C.; Garcia, A. E. *J. Mol. Biol.* **2007**, *368*, 1448−1457.
(9) Wu, K.-P.; Weinstock, D. S.; Narayanan, C.; Levy, R. M.; Baum, J. *J. Mol. Biol.* **2009**, *391*, 784−796.
(10) Zhang, W.; Ganguly, D.; Chen, J. *PLoS Comput. Biol.* **2012**, *8*, e1002353.
(11) Narayanan, C.; Weinstock, D. S.; Wu, K.-P.; Baum, J.; Levy, R. M. *J. Chem. Theory Comput.* **2012**, *8*, 3929−3942.
(12) Mittal, J.; Yoo, T. H.; Georgiou, G.; Truskett, T. M. *J. Phys. Chem. B* **2013**, *117*, 118−124.
(13) Marsh, J. A.; Forman-Kay, J. D. *J. Mol. Biol.* **2009**, *391*, 359−374.
(14) Salmon, L.; Nodet, G.; Ozenne, V.; Yin, G.; Jensen, M. R.; Zweckstetter, M.; Blackledge, M. *J. Am. Chem. Soc.* **2010**, *132*, 8407−8418.
(15) Fisher, C. K.; Huang, A.; Stultz, C. M. *J. Am. Chem. Soc.* **2010**, *132*, 14919−14927.
(16) Jensen, M. R.; Zweckstetter, M.; Huang, J.; Blackledge, M. *Chem. Rev.* **2014**, *114*, 6632−6660.
(17) Jensen, M. R.; Blackledge, M. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, E1557−1558.
(18) Csermely, P.; Palotai, R.; Nussinov, R. *Trends Biochem. Sci.* **2010**, *35*, 539−546.
(19) Wright, P. E.; Dyson, H. J. *Curr. Opin. Struct. Biol.* **2009**, *19*, 31−38.
(20) Sugase, K.; Dyson, H. J.; Wright, P. E. *Nature* **2007**, *447*, 1021−1025.
(21) Hilser, V. J.; Thompson, E. B. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 8311−8315.
(22) Kiefhaber, T.; Bachmann, A.; Jensen, K. S. *Curr. Opin. Struct. Biol.* **2012**, *22*, 21−29.
(23) Iesmantavicius, V.; Dogan, J.; Jemth, P.; Teilum, K.; Kjaergaard, M. *Angew. Chem., Int. Ed.* **2014**, *53*, 1548−1551.
(24) Rogers, J. M.; Wong, C. T.; Clarke, J. *J. Am. Chem. Soc.* **2014**, *136*, 5197−5200.
(25) Schneider, R.; Maurin, D.; Communie, G.; Kragelj, J.; Hansen, D. F.; Ruigrok, R. W. H.; Jensen, M. R.; Blackledge, M. *J. Am. Chem. Soc.* **2015**, *137*, 1220−1229.
(26) Gianni, S.; Dogan, J.; Jemth, P. *Curr. Opin. Struct. Biol.* **2016**, *36*, 18−24.
(27) Shoemaker, B. A.; Portman, J. J.; Wolynes, P. G. *Proc. Natl. Acad. Sci. U. S. A.* **2000**, *97*, 8868−8873.

DOI: 10.1021/jacs.6b02424
J. Am. Chem. Soc. 2016, 138, 6240−6251

(28) Fawzi, N. L.; Phillips, A. H.; Ruscio, J. Z.; Doucleff, M.; Wemmer, D. E.; Head-Gordon, T. *J. Am. Chem. Soc.* **2008**, *130*, 6145−6158.
(29) Lindorff-Larsen, K.; Trbovic, N.; Maragakis, P.; Piana, S.; Shaw, D. E. *J. Am. Chem. Soc.* **2012**, *134*, 3787−3791.
(30) Robustelli, P.; Trbovic, N.; Friesner, R. A.; Palmer, A. G. *J. Chem. Theory Comput.* **2013**, *9*, 5190−5200.
(31) Vitalis, A.; Pappu, R. V. *J. Comput. Chem.* **2009**, *30*, 673−699.
(32) Best, R. B.; Hummer, G. *J. Phys. Chem. B* **2009**, *113*, 9004−9015.
(33) Piana, S.; Donchev, A. G.; Robustelli, P.; Shaw, D. E. *J. Phys. Chem. B* **2015**, *119*, 5113−5123.
(34) Rauscher, S.; Gapsys, V.; Gajda, M. J.; Zweckstetter, M.; de Groot, B. L.; Grubmüller, H. *J. Chem. Theory Comput.* **2015**, *11*, 5513−5524.
(35) Palmer, A. *Chem. Rev.* **2004**, *104*, 3623−3640.
(36) Alexandrescu, A.; Shortlet, D. *J. Mol. Biol.* **1994**, *242*, 527−546.
(37) Farrow, N.; Zhang, O.; Formankay, J.; Kay, L. *Biochemistry* **1995**, *34*, 868−878.
(38) Frank, M.; Clore, G.; Gronenborn, A. *Protein Sci.* **1995**, *4*, 2605−2615.
(39) Brutscher, B.; Brüschweiler, R.; Ernst, R. R. *Biochemistry* **1997**, *36*, 13043−13053.
(40) Schwalbe, H.; Fiebig, K. M.; Buck, M.; Jones, J. A.; Grimshaw, S. B.; Spencer, A.; Glaser, S. J.; Smith, L. J.; Dobson, C. M. *Biochemistry* **1997**, *36*, 8977−8991.
(41) Buevich, A. V.; Baum, J. *J. Am. Chem. Soc.* **1999**, *121*, 8671−8672.
(42) Yang, D. W.; Mok, Y. K.; Muhandiram, D. R.; Forman-Kay, J. D.; Kay, L. E. *J. Am. Chem. Soc.* **1999**, *121*, 3555−3556.
(43) Yao, J.; Chung, J.; Eliezer, D.; Wright, P. E.; Dyson, H. J. *Biochemistry* **2001**, *40*, 3561−3571.
(44) Ochsenbein, F.; Neumann, J. M.; Guittet, E.; Van Heijenoort, C. *Protein Sci.* **2002**, *11*, 957−964.
(45) Klein-Seetharaman, J.; Oikawa, M.; Grimshaw, S. B.; Wirmer, J.; Duchardt, E.; Ueda, T.; Imoto, T.; Smith, L. J.; Dobson, C. M.; Schwalbe, H. *Science* **2002**, *295*, 1719−1722.
(46) Yang, D. W.; Kay, L. E. *J. Mol. Biol.* **1996**, *263*, 369−382.
(47) Wirmer, J.; Peti, W.; Schwalbe, H. *J. Biomol. NMR* **2006**, *35*, 175−186.
(48) Le Duff, C. S.; Whittaker, S. B.-M.; Radford, S. E.; Moore, G. R. *J. Mol. Biol.* **2006**, *364*, 824−835.
(49) Houben, K.; Blanchard, L.; Blackledge, M.; Marion, D. *Biophys. J.* **2007**, *93*, 2830−2844.
(50) Ebert, M.-O.; Bae, S.-H.; Dyson, H. J.; Wright, P. E. *Biochemistry* **2008**, *47*, 1299−1308.
(51) Modig, K.; Poulsen, F. M. *J. Biomol. NMR* **2008**, *42*, 163−177.
(52) Silvers, R.; Sziegat, F.; Tachibana, H.; Segawa, S.; Whittaker, S.; Günther, U. L.; Gabel, F.; Huang, J.; Blackledge, M.; Wirmer-Bartoschek, J.; Schwalbe, H. *J. Am. Chem. Soc.* **2012**, *134*, 6846−6854.
(53) Sziegat, F.; Silvers, R.; Hähnke, M.; Jensen, M. R.; Blackledge, M.; Wirmer-Bartoschek, J.; Schwalbe, H. *Biochemistry* **2012**, *51*, 3361−3372.
(54) Kurzbach, D.; Schwarz, T. C.; Platzer, G.; Hoefler, S.; Hinderberger, D.; Konrat, R. *Angew. Chem., Int. Ed.* **2014**, *53*, 3840−3843.
(55) Cho, M.-K.; Kim, H.-Y.; Bernado, P.; Fernandez, C. O.; Blackledge, M.; Zweckstetter, M. *J. Am. Chem. Soc.* **2007**, *129*, 3032−3033.
(56) Bussell, R.; Eliezer, D. *J. Biol. Chem.* **2001**, *276*, 45996−46003.
(57) Bertoncini, C. W.; Jung, Y.-S.; Fernandez, C. O.; Hoyer, W.; Griesinger, C.; Jovin, T. M.; Zweckstetter, M. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 1430−1435.
(58) Parigi, G.; Rezaei-Ghaleh, N.; Giachetti, A.; Becker, S.; Fernandez, C.; Blackledge, M.; Griesinger, C.; Zweckstetter, M.; Luchinat, C. *J. Am. Chem. Soc.* **2014**, *136*, 16201−16209.
(59) Bae, S.-H.; Dyson, H. J.; Wright, P. E. *J. Am. Chem. Soc.* **2009**, *131*, 6814−6821.
(60) Amoros, D.; Ortega, A.; Garcia de la Torre, J. *J. Chem. Theory Comput.* **2013**, *9*, 1678−1685.
(61) Rezaei-Ghaleh, N.; Klama, F.; Munari, F.; Zweckstetter, M. *Angew. Chem., Int. Ed.* **2013**, *52*, 11410−11414.
(62) Blanchard, L.; Tarbouriech, N.; Blackledge, M.; Timmins, P.; Burmeister, W. P.; Ruigrok, R. W. H.; Marion, D. *Virology* **2004**, *319*, 201−211.
(63) Houben, K.; Marion, D.; Tarbouriech, N.; Ruigrok, R. W. H.; Blanchard, L. *J. Virol.* **2007**, *81*, 6807−6816.
(64) Lewandowski, J. R.; Halse, M. E.; Blackledge, M.; Emsley, L. *Science* **2015**, *348*, 578−581.
(65) Peng, J. W.; Wagner, G. *Biochemistry* **1992**, *31*, 8571−8586.
(66) Farrow, N.; Zhang, O.; Szabo, A.; Torchia, D.; Kay, L. *J. Biomol. NMR* **1995**, *6*, 153−162.
(67) Ishima, R.; Nagayama, K. *Biochemistry* **1995**, *34*, 3162−3171.
(68) Kaderavek, P.; Zapletal, V.; Rabatinova, A.; Krasny, L.; Sklenar, V.; Zidek, L. *J. Biomol. NMR* **2014**, *58*, 193−207.
(69) Halle, B.; Wennerström, H. *J. Chem. Phys.* **1981**, *75*, 1928−1943.
(70) Lipari, G.; Szabo, A. *J. Am. Chem. Soc.* **1982**, *104*, 4546−4559.
(71) Clore, G.; Szabo, A.; Bax, A.; Kay, L.; Driscoll, P.; Gronenborn, A. *J. Am. Chem. Soc.* **1990**, *112*, 4989−4991.
(72) Halle, B. *J. Chem. Phys.* **2009**, *131*, 224507.
(73) Cole, K. S.; Cole, R. H. *J. Chem. Phys.* **1941**, *9*, 341−351.
(74) Prompers, J. J.; Brüschweiler, R. *J. Am. Chem. Soc.* **2002**, *124*, 4522−4534.
(75) Khan, S. N.; Charlier, C.; Augustyniak, R.; Salvi, N.; Dejean, V.; Bodenhausen, G.; Lequin, O.; Pelupessy, P.; Ferrage, F. *Biophys. J.* **2015**, *109*, 988−999.
(76) Gill, M. L.; Byrd, R. A.; Arthur, G.; Palmer. *Phys. Chem. Chem. Phys.* **2016**, *18*, 5839−5849.
(77) Jensen, M. R.; Houben, K.; Lescop, E.; Blanchard, L.; Ruigrok, R. W. H.; Blackledge, M. *J. Am. Chem. Soc.* **2008**, *130*, 8055−8061.
(78) Jensen, M. R.; Blackledge, M. *J. Am. Chem. Soc.* **2008**, *130*, 11266−11267.
(79) de Alba, E.; Baber, J.; Tjandra, N. *J. Am. Chem. Soc.* **1999**, *121*, 4282−4283.
(80) Kjaergaard, M.; Brander, S.; Poulsen, F. M. *J. Biomol. NMR* **2011**, *49*, 139−149.
(81) Loth, K.; Pelupessy, P.; Bodenhausen, G. *J. Am. Chem. Soc.* **2005**, *127*, 6062−6068.
(82) Hall, J. B.; Fushman, D. *J. Am. Chem. Soc.* **2006**, *128*, 7855−7870.
(83) Yao, L.; Grishaev, A.; Cornilescu, G.; Bax, A. *J. Am. Chem. Soc.* **2010**, *132*, 4295−4309.
(84) Kestin, J.; Khalifa, H. E.; Correia, R. J. *J. Phys. Chem. Ref. Data* **1981**, *10*, 71−88.
(85) Das, R. K.; Pappu, R. V. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 13392−13397.
(86) Young, W. S.; Brooks, C. L., III *J. Mol. Biol.* **1996**, *259*, 560−572.
(87) Mikhonin, A. V.; Asher, S. A. *J. Am. Chem. Soc.* **2006**, *128*, 13789−13795.
(88) Huang, C.-Y.; Getahun, Z.; Zhu, Y.; Klemke, J. W.; DeGrado, W. F.; Gai, F. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 2788−2793.
(89) Best, R. B.; Mittal, J. *J. Phys. Chem. B* **2010**, *114*, 8790−8798.
(90) Lescop, E.; Schanda, P.; Brutscher, B. *J. Magn. Reson.* **2007**, *187*, 163−169.
(91) Delaglio, F.; Grzesiek, S.; Vuister, G. W.; Zhu, G.; Pfeifer, J.; Bax, A. *J. Biomol. NMR* **1995**, *6*, 277−293.
(92) Goddard, T.; Kneller, D. G. *SPARKY 3*; University of California—San Francisco.
(93) Jung, Y.-S.; Zweckstetter, M. *J. Biomol. NMR* **2004**, *30*, 11−23.
(94) Lakomek, N.-A.; Ying, J.; Bax, A. *J. Biomol. NMR* **2012**, *53*, 209−221.
(95) Akke, M.; Palmer, A. G. *J. Am. Chem. Soc.* **1996**, *118*, 911−912.
(96) Pelupessy, P.; Espallargas, G. M.; Bodenhausen, G. *J. Magn. Reson.* **2003**, *161*, 258−264.
(97) Pelupessy, P.; Ferrage, F.; Bodenhausen, G. *J. Chem. Phys.* **2007**, *126*, 134508.

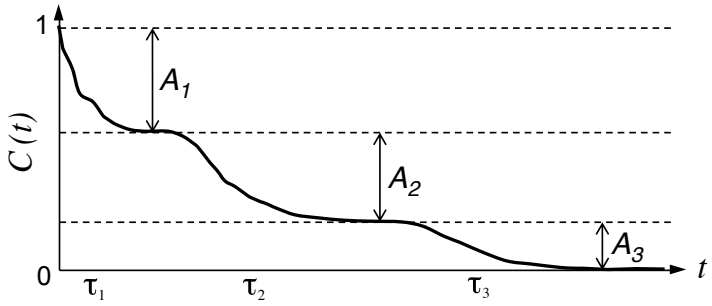
DOI: 10.1021/jacs.6b02424
J. Am. Chem. Soc. 2016, 138, 6240−6251

(98) Vranken, W. F.; Boucher, W.; Stevens, T. J.; Fogh, R. H.; Pajon, A.; Llinas, M.; Ulrich, E. L.; Markley, J. L.; Ionides, J.; Laue, E. D. *Proteins: Struct., Funct., Genet.* **2005**, *59*, 687−696.
(99) Burnham, K. P.; Anderson, D. R. *Sociol. Methods Res.* **2004**, *33*, 261−304.

DOI: 10.1021/jacs.6b02424
J. Am. Chem. Soc. 2016, 138, 6240−6251

SUPPORTING INFORMATION

Figure S1



Sketch of the correlation function used in the manuscript (see equation 6).

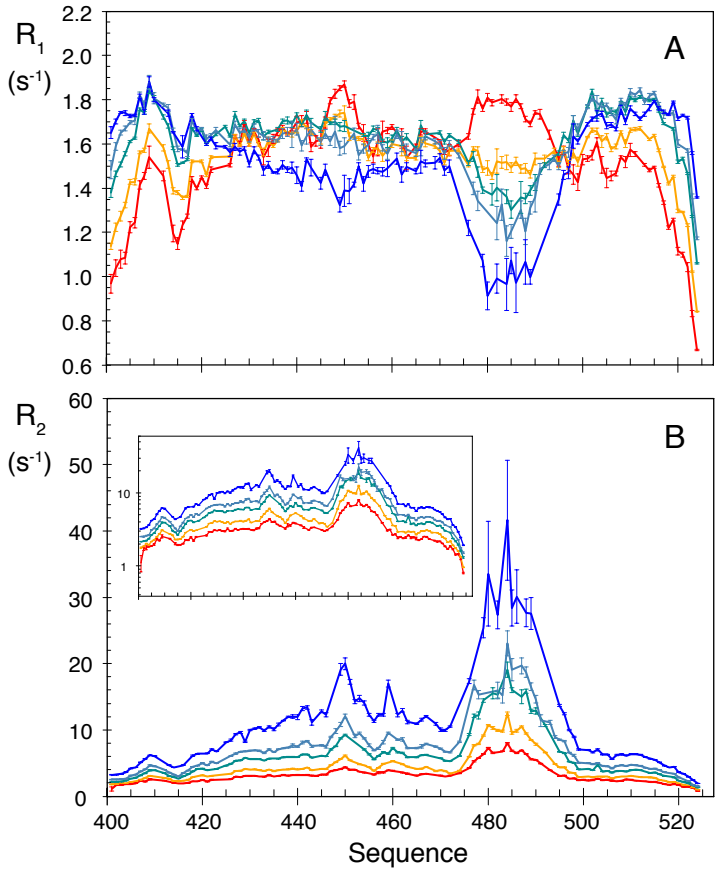


Figure S2

Experimental relaxation rates of NT_L (residues 401–524), measured at ^1H frequency of 600 MHz at 268K (blue), 274K (blue/grey), 278K (green), 288K (orange) and 298K (red). A - longitudinal (R_1) relaxation. B - transverse (R_2) relaxation (inset shows logarithmic representation).

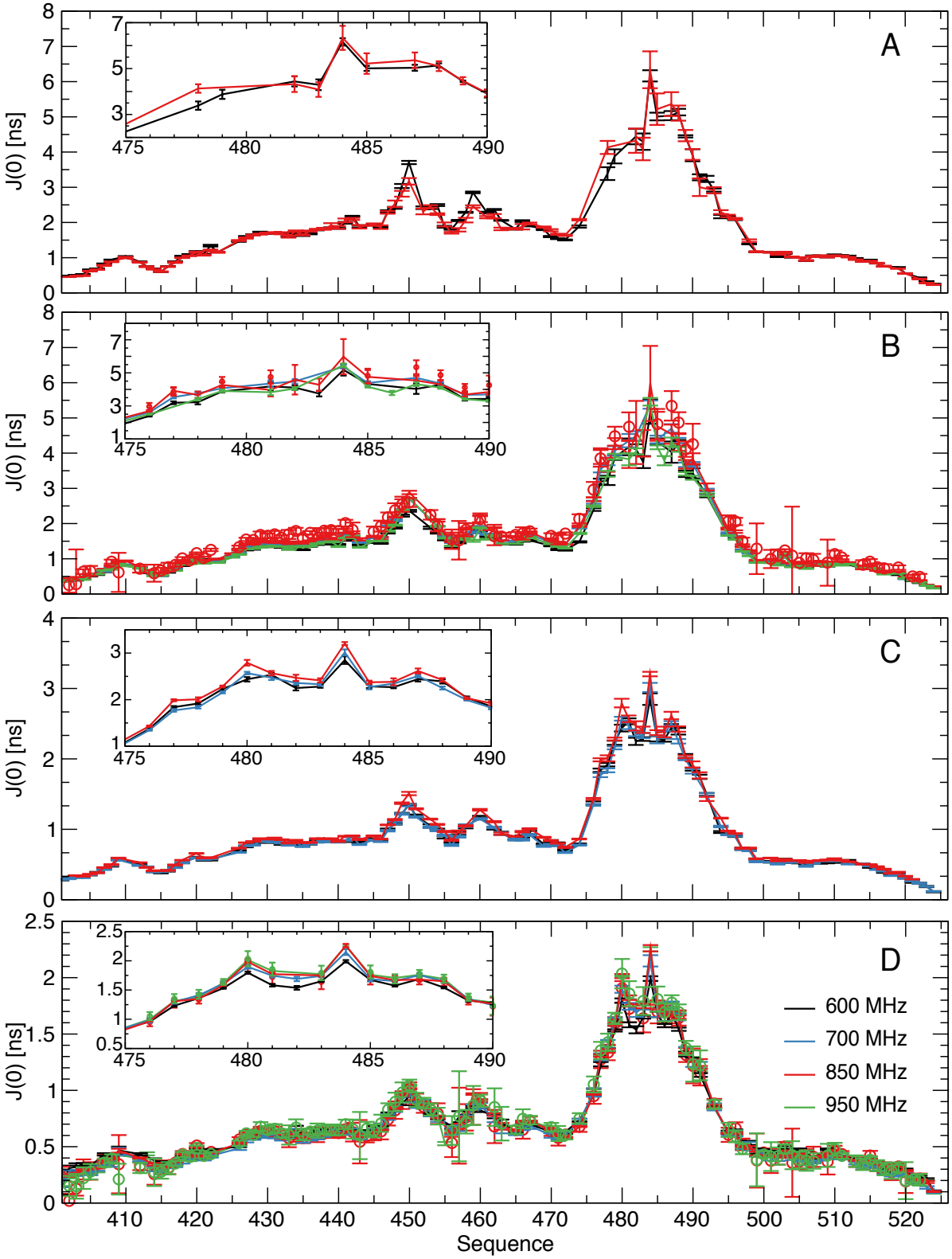


Figure S3

Values of $J(0)$ derived from reduced spectral density mapping using either uniquely auto- (lines) or cross- (open circles) correlated relaxation rates at the four different magnetic field strengths. The independent analyses show very similar values, both in the helical and unfolded regions, indicating that exchange contributions to transverse relaxation (R_2) are not significant. Color code in D applies to all panels.

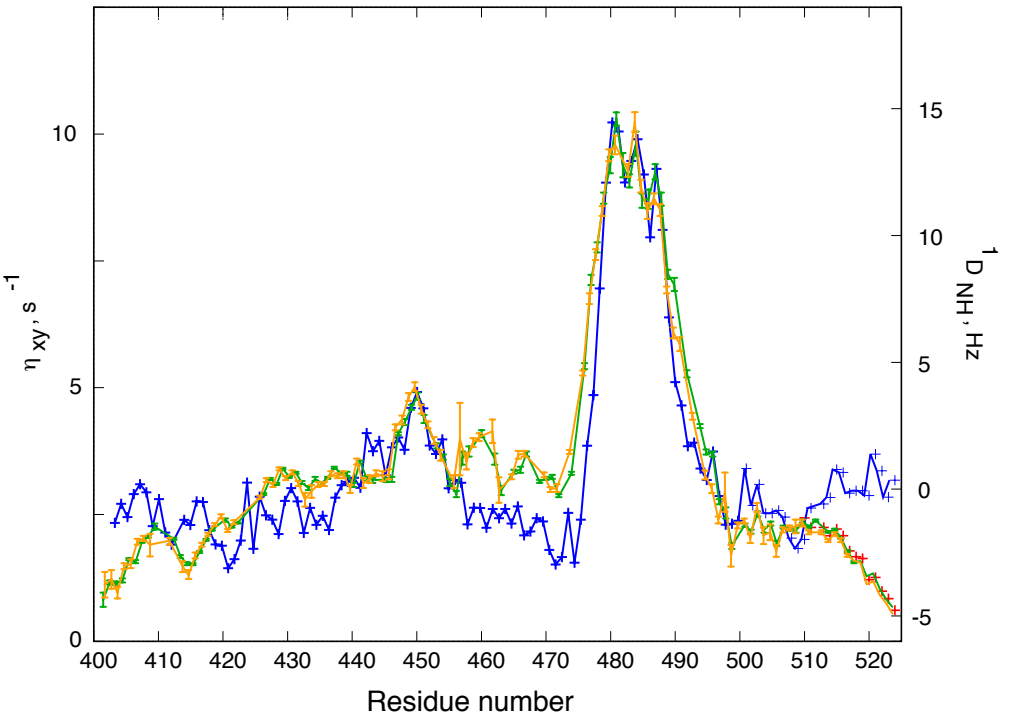


Figure S4

Comparison of residual dipolar couplings measured in NT aligned in PEG-hexanol alignment medium (blue) and η_{xy} measured at 950MHz (orange) and 298K and 850MHz and at 288K (green – scaled such that the maximum value is the same as the maximum value measured at 298K). The so-called dipolar wave observed in the dipolar coupling profile along the helix is also observed in the cross correlated relaxation. Anisotropic rotational diffusion induces similar orientational dependence in ^{15}N relaxation rates as steric alignment induces in dipolar couplings, so the similarity of the dipolar waves is consistent with helical inter-conversion occurring on timescales that are too slow to be relaxation-active. In this case the characteristics of the helical element report on an average over the relaxation occurring in the helical states and the 20-25% unfolded state, in a similar way to the averaging of dipolar couplings that are also averaged over the different sub-states.

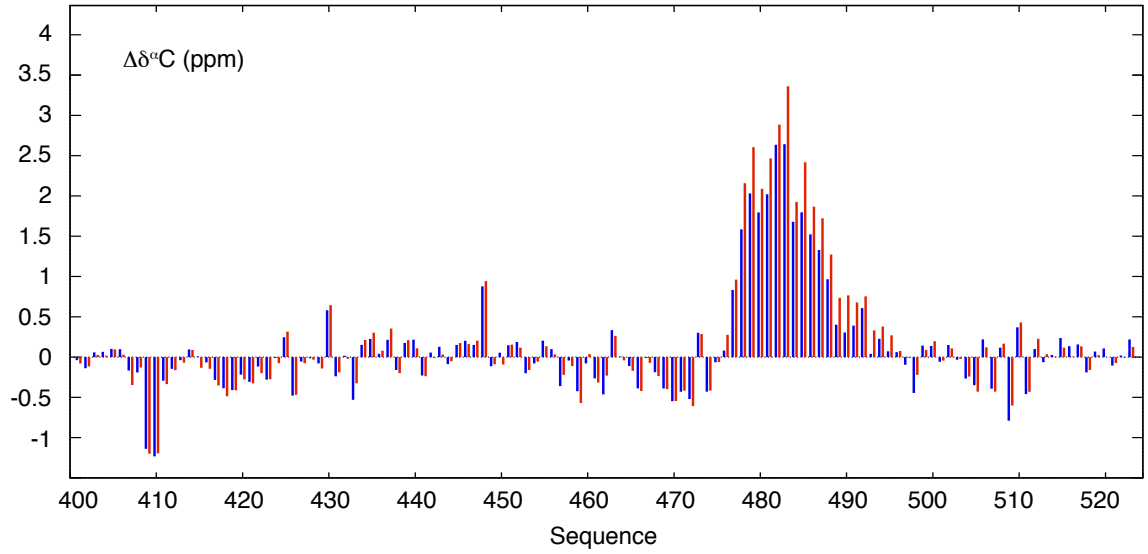


Figure S5

Secondary $^{13}C\alpha$ chemical shifts measured at 278K (red) and 298K (blue). Experimental chemical shift values were subtracted from random coil values and corrected using published temperature coefficients. Simple model calculations demonstrate that the Arrhenius analysis is weakly dependent on this small change of population (less than 5% difference in activation energy and between 1 and 3 % in the populations A_1, A_2 and A_3).

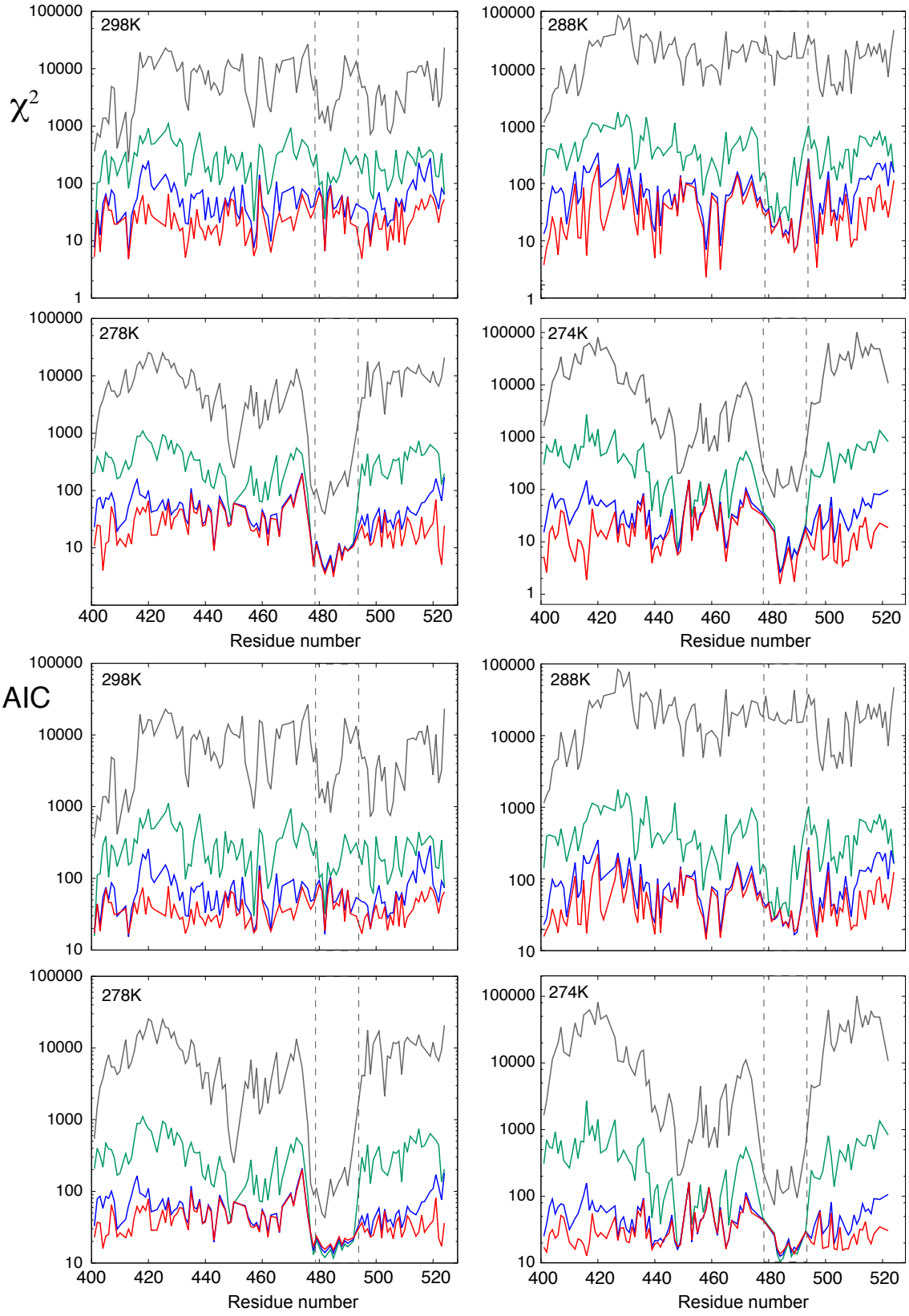


Figure S6 (description on the next page)

Figure S6

Statistical analysis of 2 (τ_1 and θ), 4 (τ_1, τ_2, A_2, and θ), 5 (A_2, A_3, τ_2, τ_3 and θ) or 6 (A_2, A_3, τ_1, τ_2, τ_3 and θ), parameter fits.

Red: 6-parameter fit - three contributions, two amplitudes, three explicit timescales
Blue: 5-parameter fit - three contributions, two amplitudes, two explicit timescales, fast timescale set to 0 (too fast to contribute to spin relaxation)
Green: 4-parameter fit - two contributions, one amplitude and two timescales are fitted
Grey: 2-parameter fit - one contribution, one timescale is fitted

χ^2 and AIC parameters are shown in the figure.
AIC statistics demonstrate that the 6-parameter fit is superior throughout the protein, with the exception of the helical region.

Here two cases can be distinguished:

- For 298 and 288K, the 5-parameter fit is justified over the 4-parameter fit, but the 6-parameter fit does not provide any improvement in the fit (χ^2 is equivalent, and therefore AIC less favourable). Inspection of the fitted parameters reveals that the fast timescale τ_1 is in fact equal to zero for the helical residues, so that the spectral density function of the 6-parameter model is equivalent to the 5-parameter model.

- For 278K and 274K, the 4-parameter fit is justified. Neither the 5- or the 6-parameter fits are justified. Again inspection of the fitted parameters clarifies the situation: in this case not only is the timescale τ_1 equal to zero, but the associated amplitude is also equal to zero. This is the case for the helical, as well as some residues in the 450-460 region at 274K.
  In this case the spectral density function of models 5 and 6 are identically equivalent to that of model 4.

The statistical analysis is therefore fully consistent with the use of the 6-parameter model throughout the protein, and the 5-parameter model in the helix (at some temperatures this becomes automatically the 4-parameter model).
This combination can therefore also be used for the combined Arrhenius analysis of all temperatures.

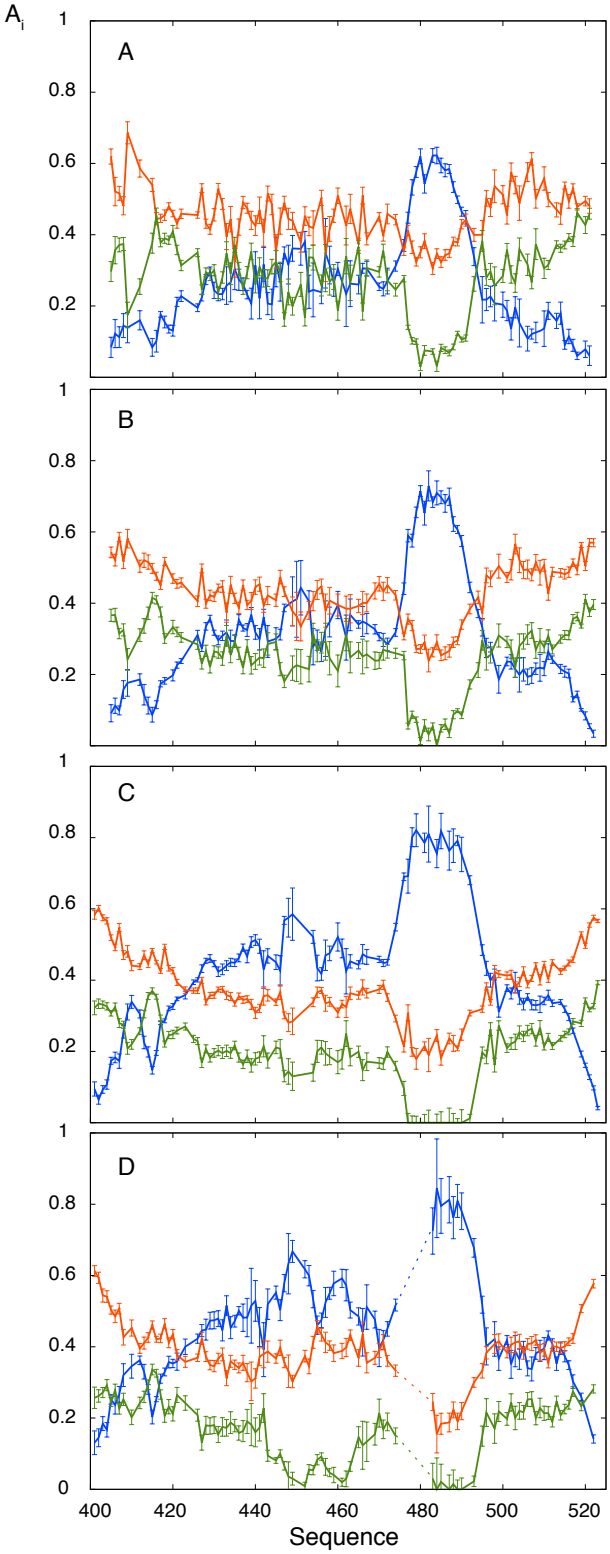


Figure S7

Amplitude of the motional contributions A_1 (fast contribution, green), A_2 (intermediate timescale contribution, orange) and A_3 (slowest contribution, blue) in NT_L at the four temperatures, 298K (A), 288K (B), 278K (C) and 274K (D) when the fast timescale (τ_1) is fitted simultaneously. The amplitudes are effectively indistinguishable from Figure 4, where the fast timescale is fixed.

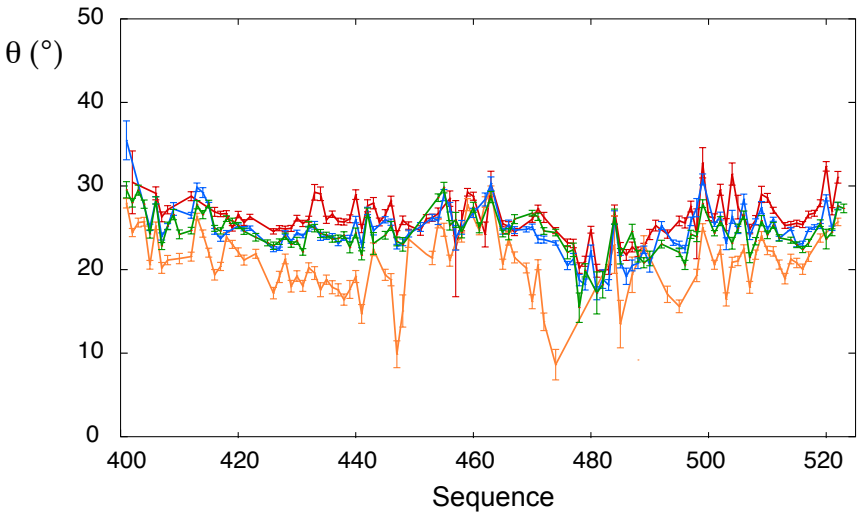


Figure S8

Site-specific values of the dipole-dipole/CSA interaction angle θ derived at 298K (red), 288K (green), 278K (blue) and 274K (orange). These values were optimized simultaneously to amplitude and timescale parameters (Figure S7 and Figure 3).

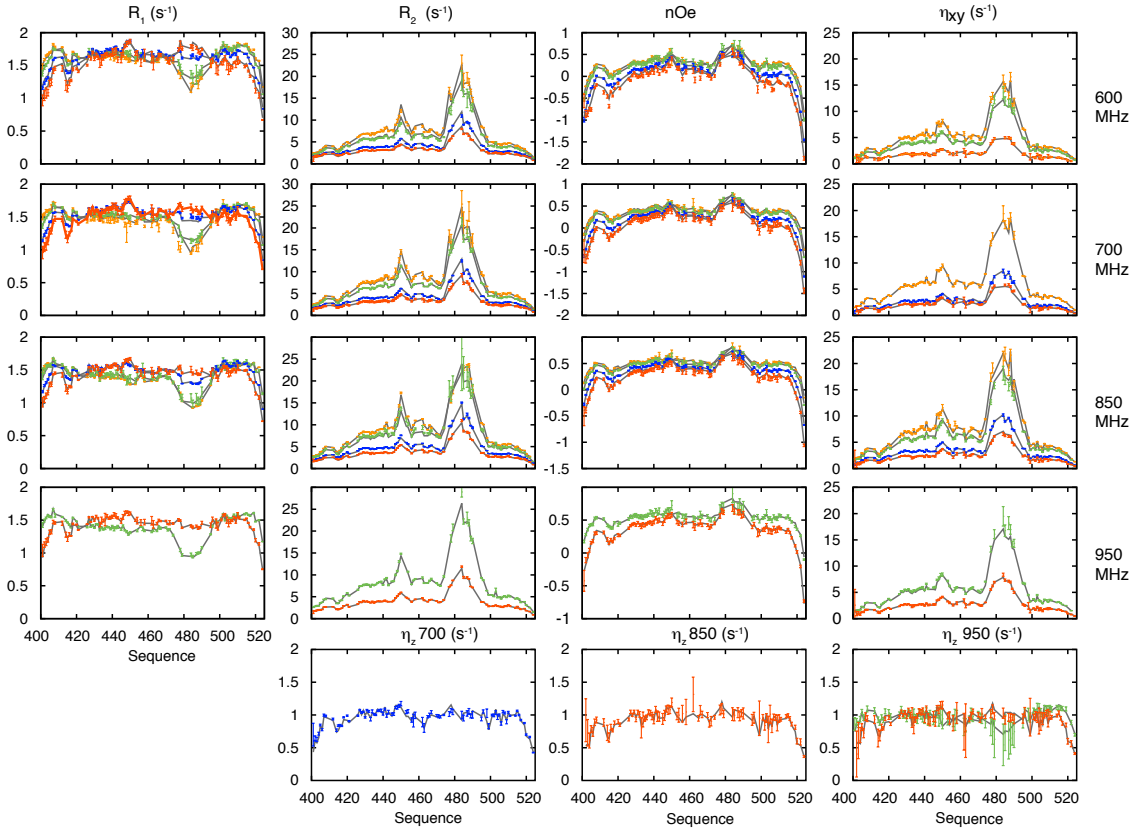


Figure S9

Data reproduction of the 58 relaxation rates present in the fit shown in Figure 5. Experimental points are shown in Red (298K), blue (288K), green (278K), orange (274K), while fitted values are shown as solid (grey) lines.

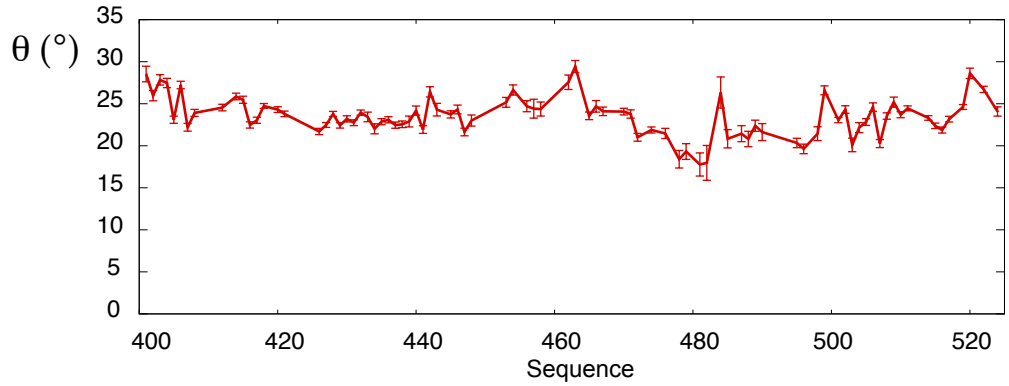


Figure S10

Site-specific values of the dipole-dipole/CSA interaction angle θ derived simultaneously to amplitude and timescale parameters (Figure 5) during the multi-temperature Arrhenius analysis. Errors are estimated from noise-based Monte Carlo simulations.

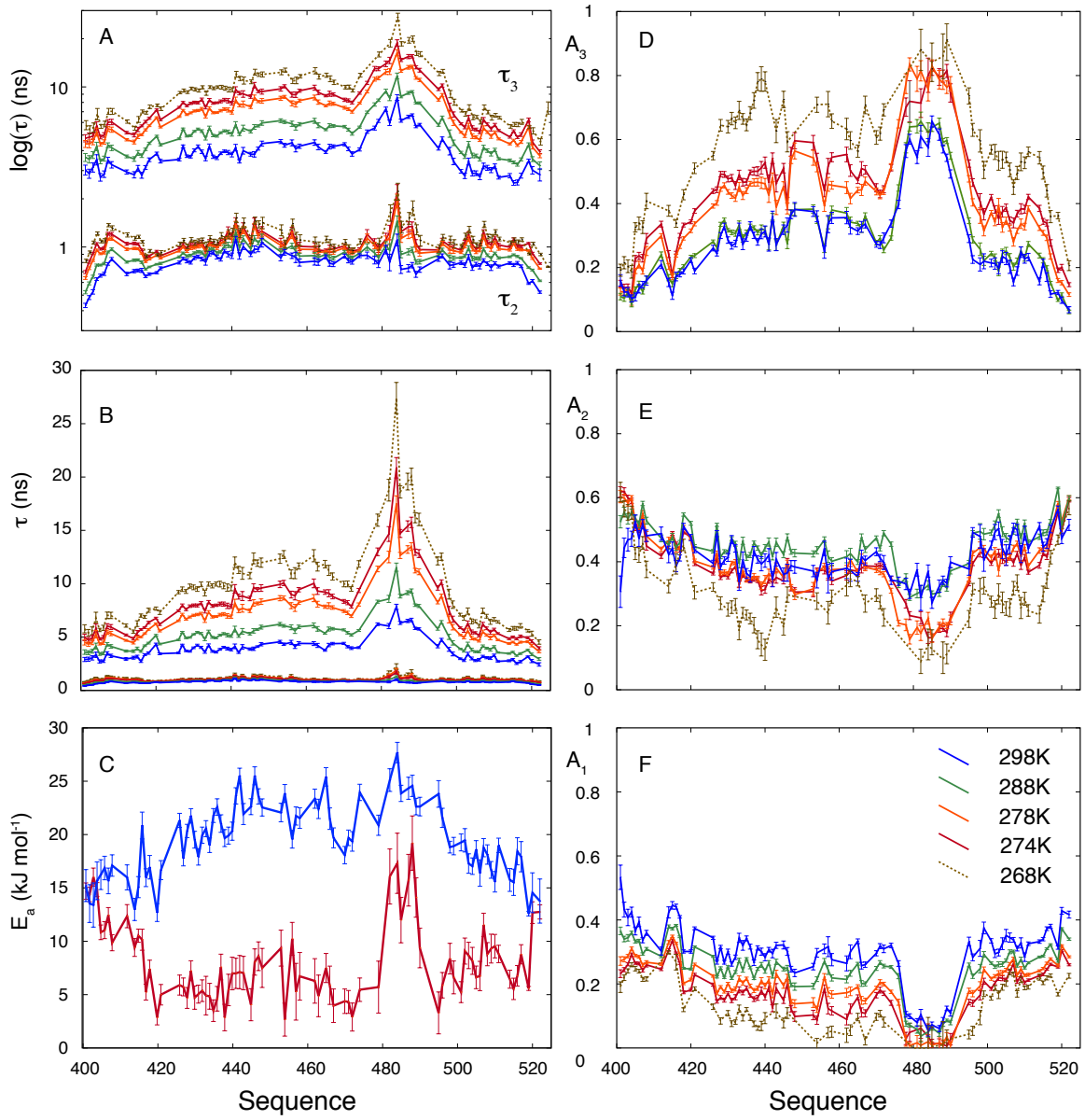


Figure S11

Simultaneous analysis of relaxation data measured over a range (268-298)K using a site specific Arrhenius type relationship to model the characteristic correlation times of intermediate and slow motional modes in NT_L.
(A, B) Intermediate (τ_2) and slow (τ_3) correlation times at 298K (blue), 288K (green), 278K (orange), 274K (red) and 268K (dashed gold). The same data are shown in both plots, in (A) a logarithmic scale is used.
(C) Effective activation energies describing the temperature coefficient of the motional timescales for each amino acid. Blue and red curves show the temperature coefficients of the slow (τ_3), and intermediate (τ_2) components respectively.
(D-F) Amplitude of the motional contributions A_3 (D - slowest contribution) A_2 (E - intermediate timescale contribution) and A_1 (fast contribution) at the five temperatures (color scheme as in A, B).

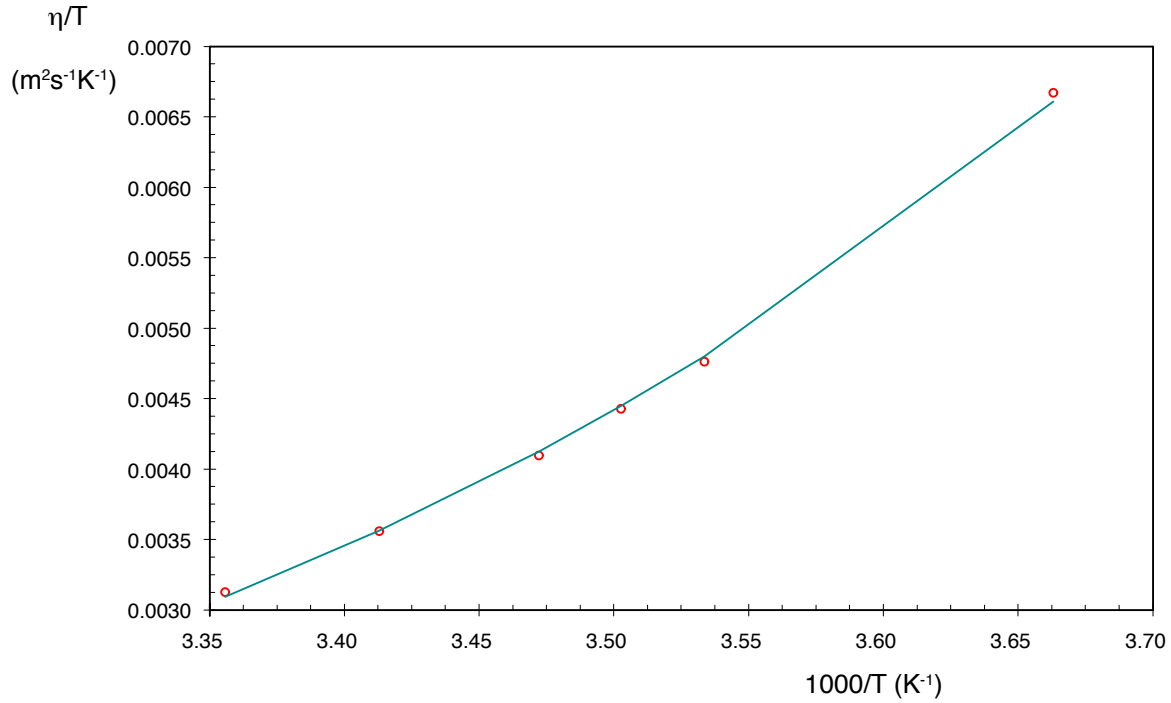


Figure S12

Temperature dependence of tabulated values (article reference 84) of the viscosity of 500mM NaCl solution (red points). The data points can be fitted to an Arrhenius relationship resulting in an effective temperature coefficient of 20.5kJmol^{-1} (green line).

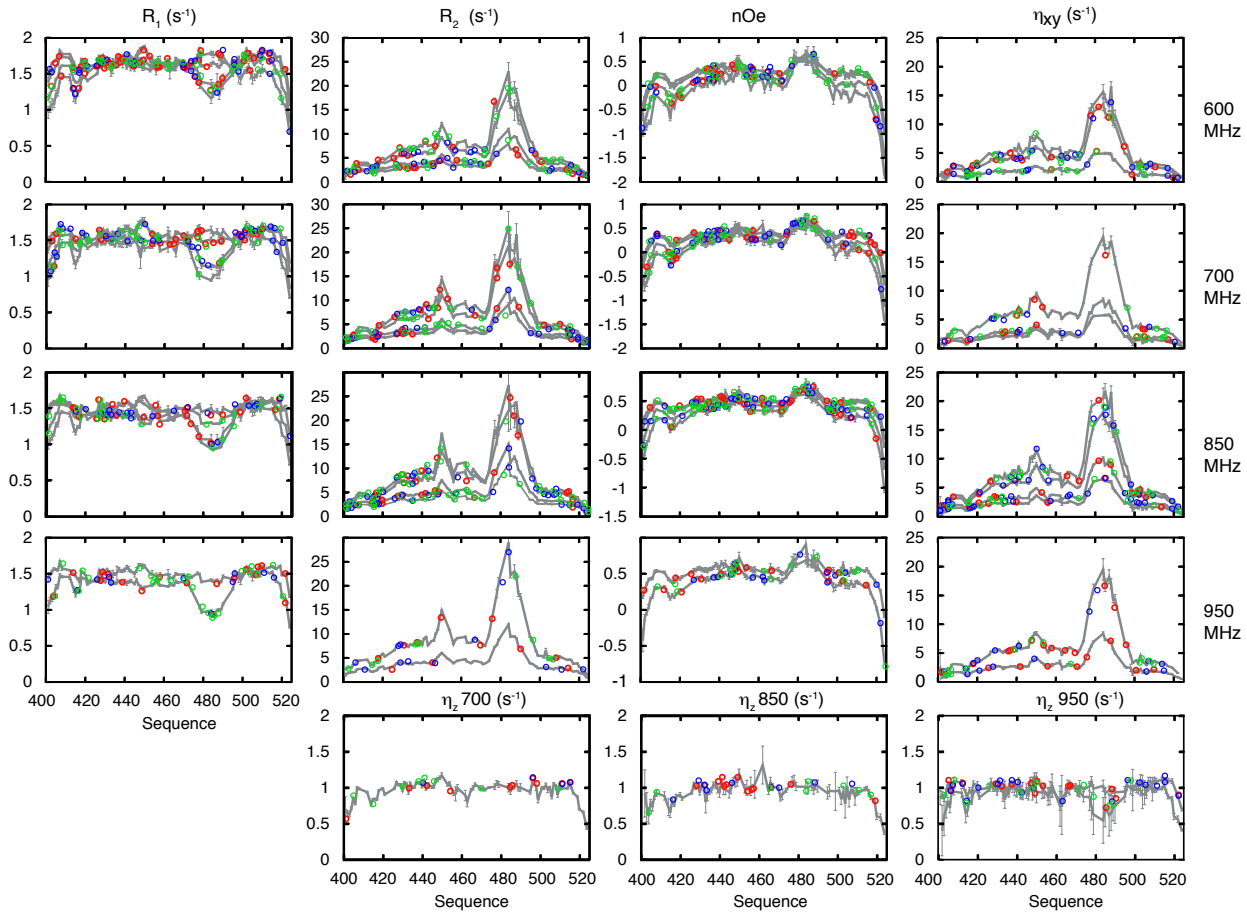


Figure S13

Reproduction of experimental relaxation rates (shown in grey) using a cross-validation approach whereby 10% of all data points were randomly removed and predicted from an analysis treating only the remaining 90%. Three random permutations were performed and the results from all three are shown here for the 10% of points removed in each case (red, green and blue circles). Data for multiple temperatures are shown in each panel.

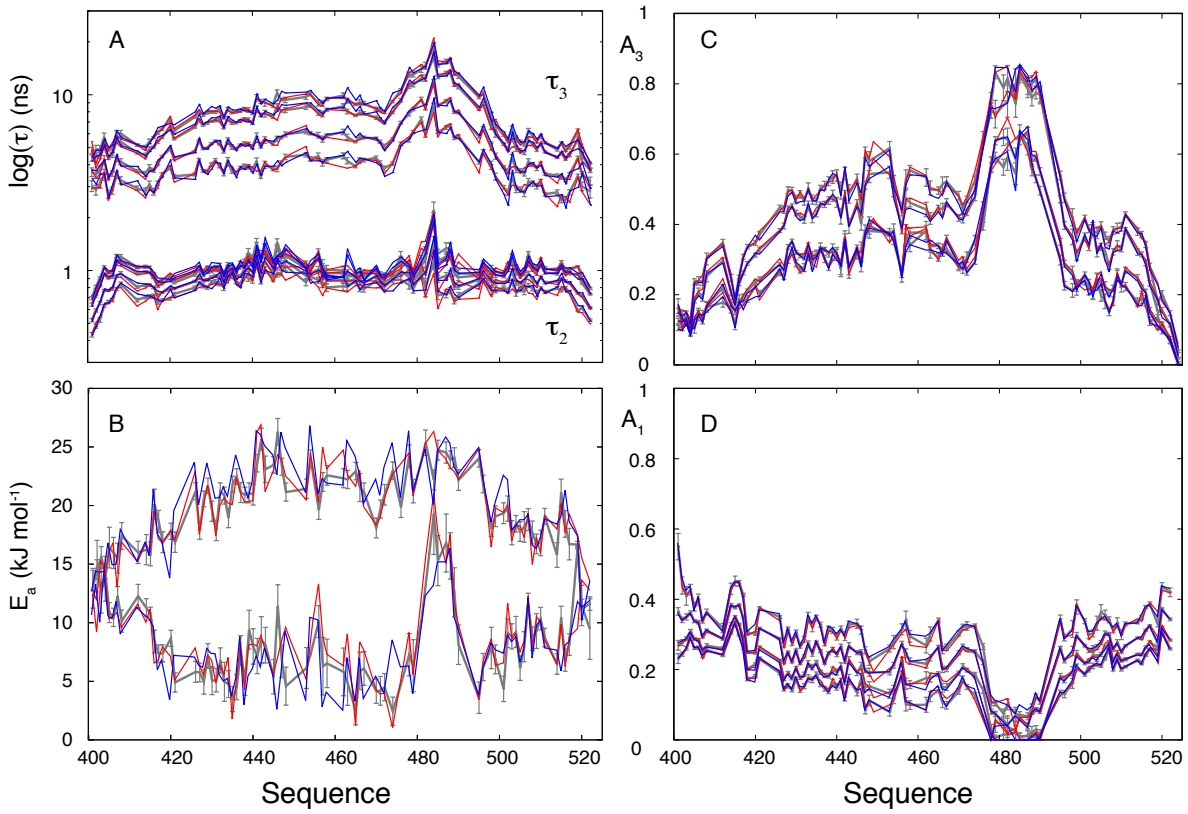


Figure S14

Timescales, amplitudes and effective activation energies derived from Arrhenius analyses. The simultaneous analysis using all data measured at 600Mz, 700MHz, 850MHz and 950MHz are shown in grey (as in Figure 5). Parameters shown in red and blue represent the analyses performed in the absence of all data from 700MHz and 950MHz respectively.

6.4. Complementary results and discussion

6.4.1 Evolution of conformational sampling of $N_{TAIL,L}$ with temperature

As discussed in the manuscript, the temperature dependence of the backbone conformation sampling of $N_{TAIL,L}$ was investigated by comparing $^{13}C\alpha$ secondary chemical shifts at 298K and 278K (Fig. 6.5, left panel). A slight increase in the helical propensity was observed in the α-helical MoRE region while for unfolded regions of the protein, the backbone conformational sampling remained the same. In addition, we compared $^{13}C'$ secondary chemical shifts at these temperatures confirming our conclusion (Fig. 6.5).

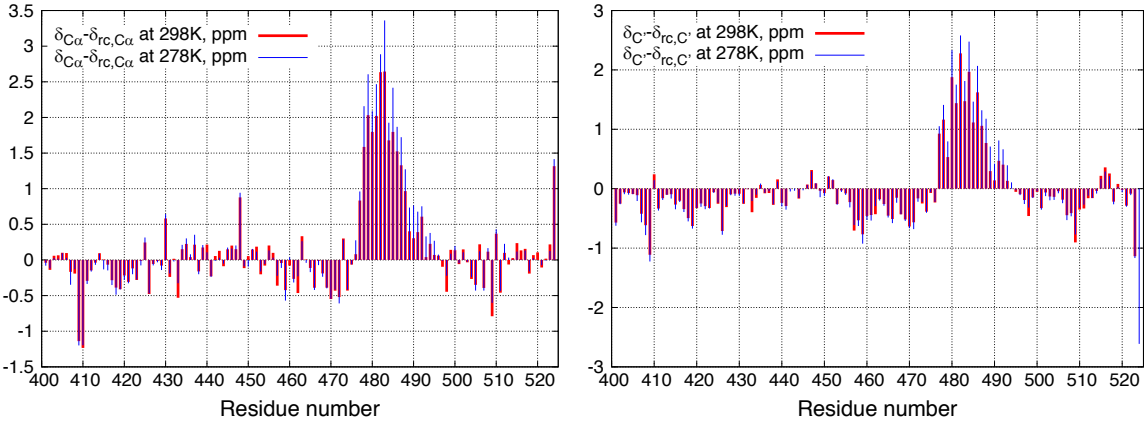


Fig. 6.5. Evolution $^{13}C^{\alpha}$ and $^{13}C'$ secondary chemical shifts with temperature, calculated using a published table of random coil chemical shifts δ_{rc} with temperature and sequence corrections [221].

6.4.2 Temperature dependence of $N_{TAIL,L}$ relaxation rates in the 258K..303K range

The most complete set of relaxation rates of $N_{TAIL,L}$ at different temperatures was measured at the magnetic field corresponding to ^{1}H frequency of 600 MHz (Fig. 6.6). Auto-relaxation rates R_1 and R_2, heteronuclear NOEs were measured at 268, 274, 278, 288, 298 and 303K, and cross-correlated transversal cross-relaxation rate (η_{xy}) at 268, 274, 278, 298 and 303K. The R_2 relaxation rate in particular was measured at 7 temperatures, including 258K (-15°C), a temperature that is not often accessible to solution NMR studies of proteins. At 258K, only rates extracted from the most intense peaks corresponding to more flexible N- and C-terminal regions with R_2 values below 20 s^{-1} are shown. Other residues have larger R_2 rates and give rise to weak and broad peaks in the spectrum. This is related to the use of glass capillaries inside 5mm NMR tube to prevent sample freezing: in such a complex system, it is difficult to attain high levels of magnetic field homogeneity, and hence peaks are even more broader due to T_2* effect and become undetectable.

At 268K, glass capillaries were no longer used and all rates were measured with the sample directly in a standard 3mm NMR tube. Apparently, at this diameter capillary

forces start to play their role in protecting the sample from freezing, and the high concentration of salt (500 mM NaCl) also contributed to keeping the sample in the liquid state. We successfully lowered the temperature to 263K and the sample stayed in the liquid state overnight in the spectrometer; we did not try going further down, but it is possible that capillaries are not necessary even at 258K.

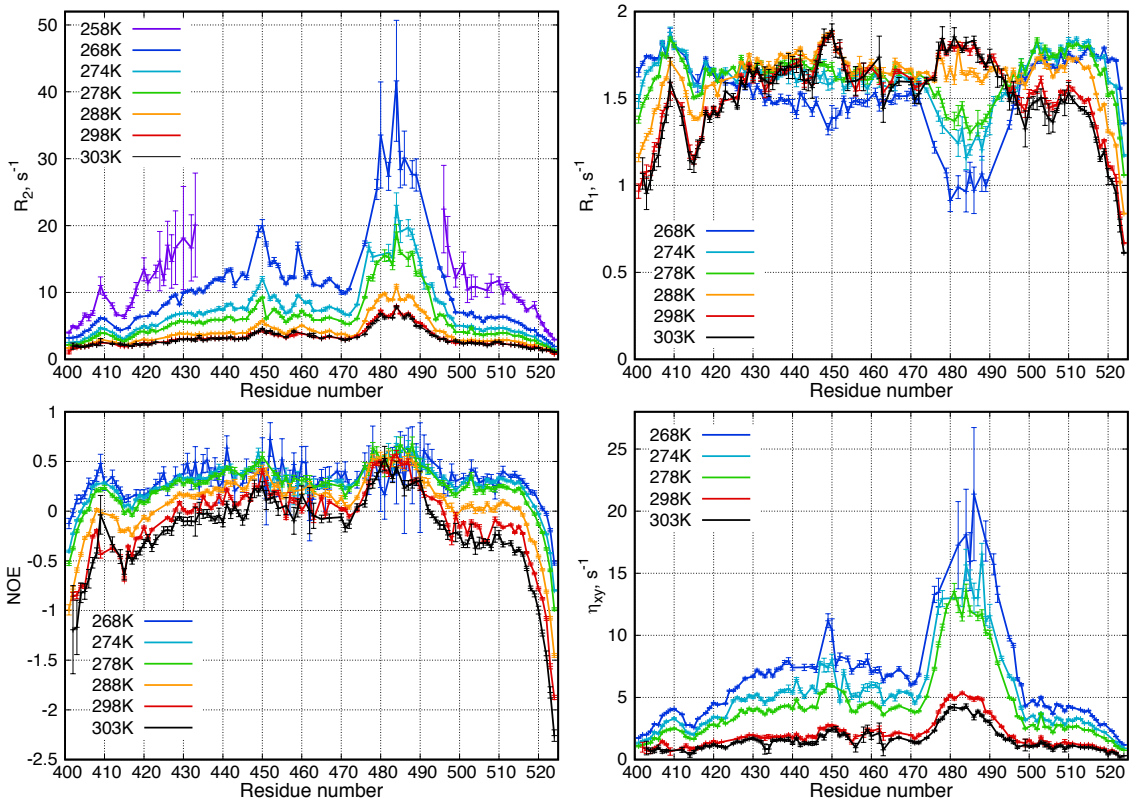


Fig. 6.6. Temperature dependence of relaxation rates of $N_{TAIL,L}$, (R_1, R_2, NOE and η_{xy}) measured at 600 MHz ^{1}H frequency. Points with too big uncertainties (over 50%) were removed from R_2 plot at 258K.

As temperature decreases, transverse R_2 and η_{xy} relaxation rates increase throughout the protein. The sequence profile of NOEs moves to higher values with lower temperature, but conserves its overall shape. That is not the case of R_1 rate: it goes up faster in helical region with temperature, slower in the central region and even goes down with temperature in N- and C-terminal tails. This is an evidence of a broad range of timescales of motions: R_1 timescale dependence has a characteristic two-hill profile with a maximum, and, depending on which side of this maximum the timescale of motion is located, R_1 can increase with temperature if motions are slow (as in the helical region), and decrease for fast motions (as in flexible tails). For motions on intermediate timescale, R_1 first increase and then decrease with temperature; this behavior could be observed in residues 425-445 in the central region of $N_{TAIL,L}$.

The region between residue 443 and the helical element has some motional restriction, as we observe elevated values of R_2, η_{xy} rates and NOEs, and the R_1 temperature

dependence in the region around residue 450 is the same as in the helical part, consistent with slower motions. As we discussed in the manuscript, some local compactness and structuration is caused in this region by the presence of two tryptophan residues – W443 and W449 – possessing hydrophobic side chains. This observation is in agreement with previous studies of disordered proteins, and in particular with [193] (cf. section 4.2.7). Interestingly, the motional restriction in this region seems to be centered on the second tryptophan region. This could be explained by the fact that the first tryptophan, W441, has a specific position in the sequence: it is surrounded by one glycine on N-terminal side and two glycines on C-terminal side.

6.4.3 Conformational sampling and dynamics in N_{TAIL} constructs of different length

Three forms of N_{TAIL} – full domain $N_{TAIL,L}$ (401-524), and shorter constructs $N_{TAIL,M}$ (443-524) and $N_{TAIL,C}$ (443-501) share the same helical molecular recognition element (MoRE) (residues 476-492), but have different disordered chain length, which makes them a pertinent system to study the role of chain motions in the dynamics of IDPs. These constructs also possess different number of tryptophan residues – whereas $N_{TAIL,L}$ has two (W441 and W449), $N_{TAIL,M}$ and $N_{TAIL,C}$ has only one (W449). As this might have potential impact on their conformational sampling, and in particular on the behaviour of the helical MoRE region (by, for example, an hydrophobic interaction between the latter and the tryptophans in the chain), we compared C^{α} secondary chemical shifts in these constructs, using the same procedure as in section 6.3.1.

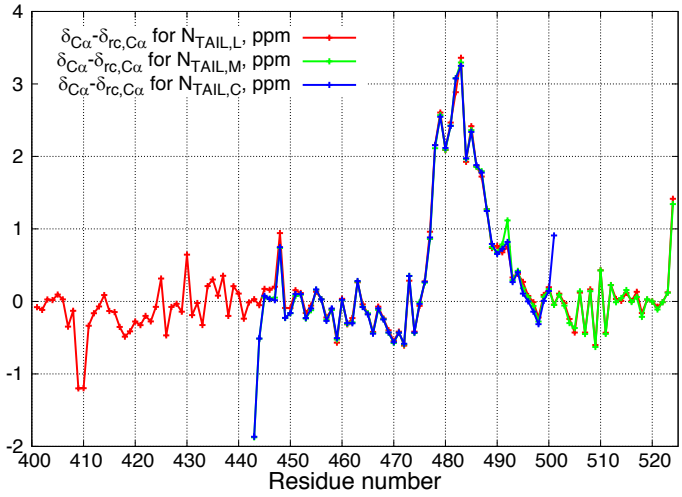


Fig. 6.7. $^{13}C^{\alpha}$ secondary chemical shifts with temperature, calculated for N_{TAIL} constructs of different length.

As shown on Fig. 6.7, the secondary chemical shifts are almost identical for the common parts of three constructs, except for a couple of residues located on N-terminal end of $N_{TAIL,C}$ and on C-terminal end of $N_{TAIL,M}$, which suggests that the conformational sampling is the same in these constructs. However, this does not preclude them from having different dynamics as we discussed in the manuscript and as we will continue to discuss

here. However, to ensure that any difference in the dynamics of MoRE region does not stem from the different number of tryptophans, the wild type form of $N_{TAIL,L}$ was compared to the double mutant, W441A W449A, in which both tryptophans are replaced by alanine (Fig. 6.8).

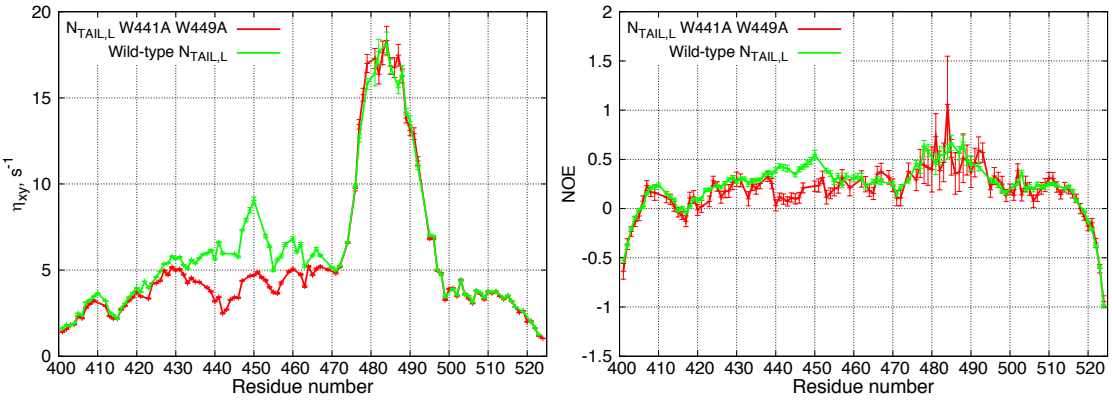


Fig. 6.8. Comparison of relaxation rates in the wild-type $N_{TAIL,L}$ and in the double mutant W441A and W449A.

The mutation disrupts the local structure formed by tryptophan residues and the whole region between residues 420 and 470 becomes more flexible. However, this seems to have negligible effect on helical MoRE and N-terminal region of the proteins, as the rates coincide within experimental uncertainty. We also note that even if the R_2 rate is larger in central region of the mutant form of the protein due to increased flexibility, it has the same local minima and maxima as in the wild-type protein, which are defined by the protein sequence. In particular, the local maximum around residue 410 corresponds to the proline at position 411, a local minimum between residues 410 and 420 corresponds to three glycines (413-415), a local minimum after the residue 440 corresponds to two glycines 442 and 443, and there is another R_2 minimum at residues 455 and 456 which are also glycines. It is also possible that the Tryptophans, in particular W449, mediates a weak interaction with the helix, that may have the effect of raising the effective rigidity of the intervening strand.

Model-free analysis of different N_{TAIL} constructs

In the manuscript, we discussed the model-free analysis of three N_{TAIL} forms at 278K (Figures 8 and S10 in the article). Here, we complement these results by showing the analysis of $N_{TAIL,L}$ and $N_{TAIL,M}$ forms at 298K (Fig. 6.9). For the shortest $N_{TAIL,C}$ form, we did not have enough relaxation rates (cf. Table 6.1) to obtain stable parameters using the 3-component form of $J(\omega)$. In the $N_{TAIL,L}$ analysis, the fast motion correlation time τ_1 was fitted outside the helical region. In the $N_{TAIL,M}$ analysis, we were not able to obtain stable parameters when fitting τ_1, and then we decided to fix it at its average value of 50 ps obtained in this fit. This value was for observed in $N_{TAIL,L}$ analyses. In the helical region (476-492) it was fixed to 0 ps in both cases.

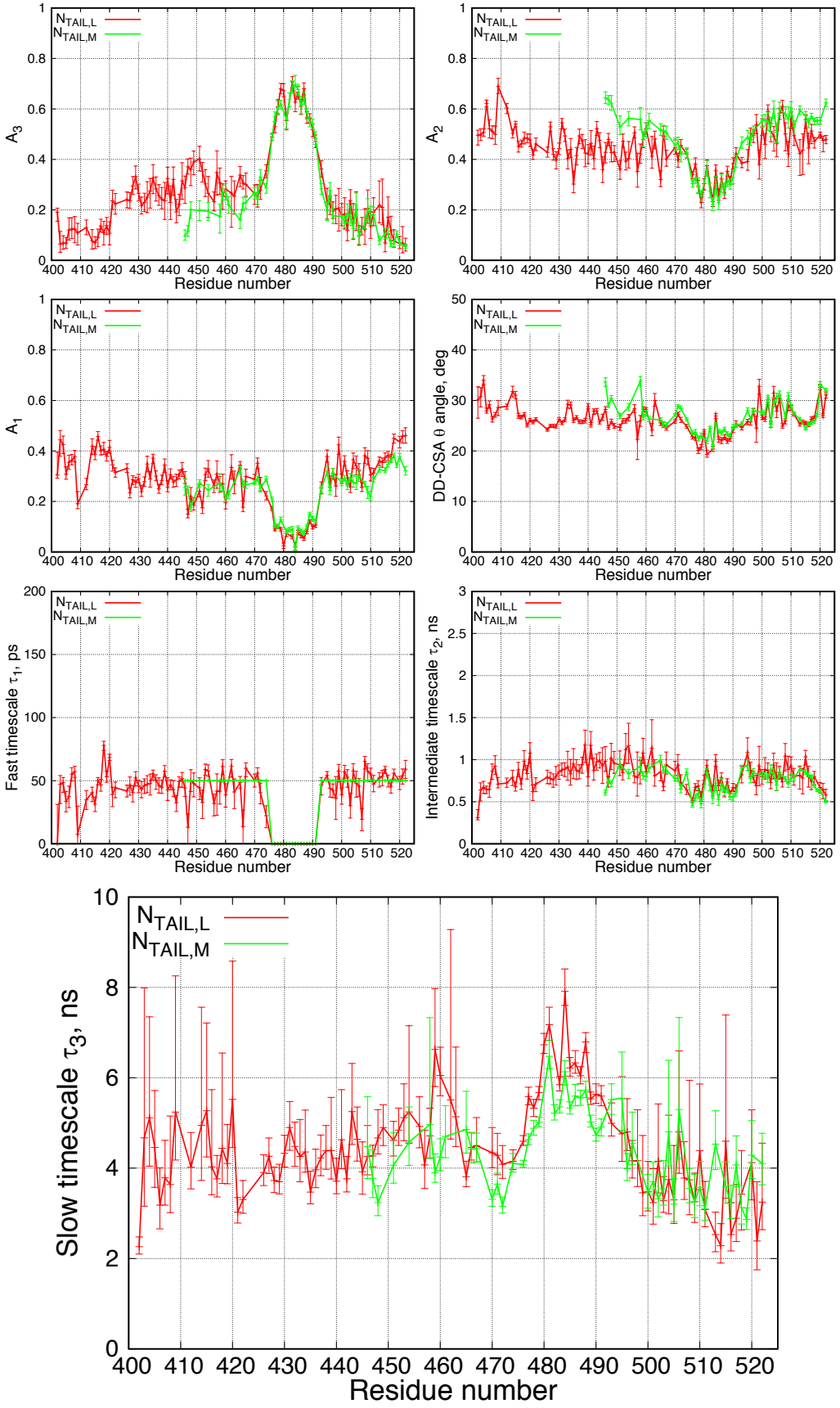


Fig. 6.9. Model-free analysis of $N_{TAIL,L}$ and $N_{TAIL,M}$ forms at 298K. A_1, A_2, A_3 represent amplitudes and τ_1, τ_2, τ_3 are the correlation times of fast, intermediate and slow-timescale motions.

Both at 278 and 298K, the motional amplitudes (A_1, A_2, A_3) are highly similar between $N_{TAIL,M}$ and $N_{TAIL,L}$ in the helical and N-terminal regions. In the helical element, even the local maxima and minima coincide in two forms of N_{TAIL}, reinforcing the conclusion that the conformational sampling is almost identical in these forms. The correlation time of the intermediate timescale motion τ_2 is identical within the experimental uncertainty on the whole length of $N_{TAIL,M}$, as was the correlation time of fast motions τ_1.

At 278K, we observed that the correlation time of slowest motion is the same in the flexible N-terminal region, but is different in the helical element and clearly increases with the chain length. At 298K, we would also make this observation, however, the sequence profile of τ_3 is too noisy, and that precludes us from comparing the dynamics in the N-terminal region: the values are the same within the uncertainty of the fit. However the τ_3 values in the helical element do appear to differ more than the uncertainty value, although this effect is clearly less evident than at 278K, where the longer correlation time amplifies this effect.

In the following section, we discuss an approach that takes into account the chain nature of IDPs when fitting relaxation data, and in particular allows us to stabilize fitted parameters by reducing their number.

6.4.4 Sliding window model-free analysis of N_{TAIL} constructs

As seen from the profiles of the correlation time of slow motions (Figures 3, 5, 8 in the manuscript), its distribution along the sequence appears smooth. The reason for this may be the underlying random-coil polymer behaviour (cf. section 4.2.7). Here we propose to introduce this concept into an analytical approach of the experimental data to treat the correlation time of the slow motion τ_3.

The basis of our approach extends the extended model-free form of the spectral density function (6.1b) (with fitted or fixed τ_1). Whereas, the correlation time of the slow motion τ_3 is normally fitted separately for each residue and independently on the neighbours. If segmental motions are at the origin of the slowest contribution, then τ_3 values should be shared over the segment. We have experimented with imposing this similarity of τ_3 values for neighboring residues in the following manner (demonstrated on Fig. 6.10):

- For each residue *i*, instead of treating it as an isolated system, a window (or a segment) of *n* residues in the sequence centred on *i* is treated.

- Relaxation data for all residues in this window are treated simultaneously, with all fit parameters being specific for each residue in this window, except the correlation time of slow motion τ_3, for which a common value is fitted for all residues in the window.

- Only fit parameters for residue *i* in the segment are finally presented.

Only the slow timescale is common, with all amplitudes and faster timescales treated as in the previous three-component analysis.

If there were no relaxation data for a particular residue in the sliding window, or the window itself was not entirely in the protein sequence (when analysing residues that were at the ends of the chain), the effective size of the window (number of residues for which relaxation data were actually fitted) was smaller.

$$J(\omega) = A_3 \frac{\tau_3}{1+\tau_3^2\omega^2} + A_2 \frac{\tau_2}{1+\tau_2^2\omega^2} + A_1 \frac{\tau_1}{1+\tau_1^2\omega^2}$$

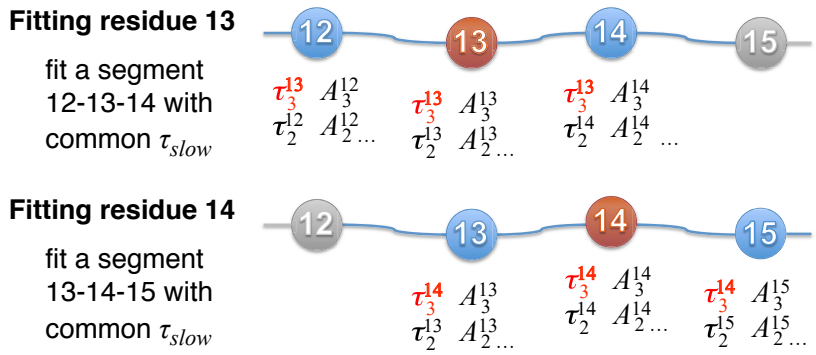


Fig. 6.10. Illustration of the sliding window model-free analysis. In this case, *n*=3

Clearly the value *n* of window length should be optimised. Initially using *n*=7 residues, a value that was proposed as a persistence length of a disordered chain by Schwalbe & co-workers (section 4.2.7), to fit relaxation data in $N_{TAIL,L}$ at multiple fields at 278K. Results are shown on Fig. 6.11.

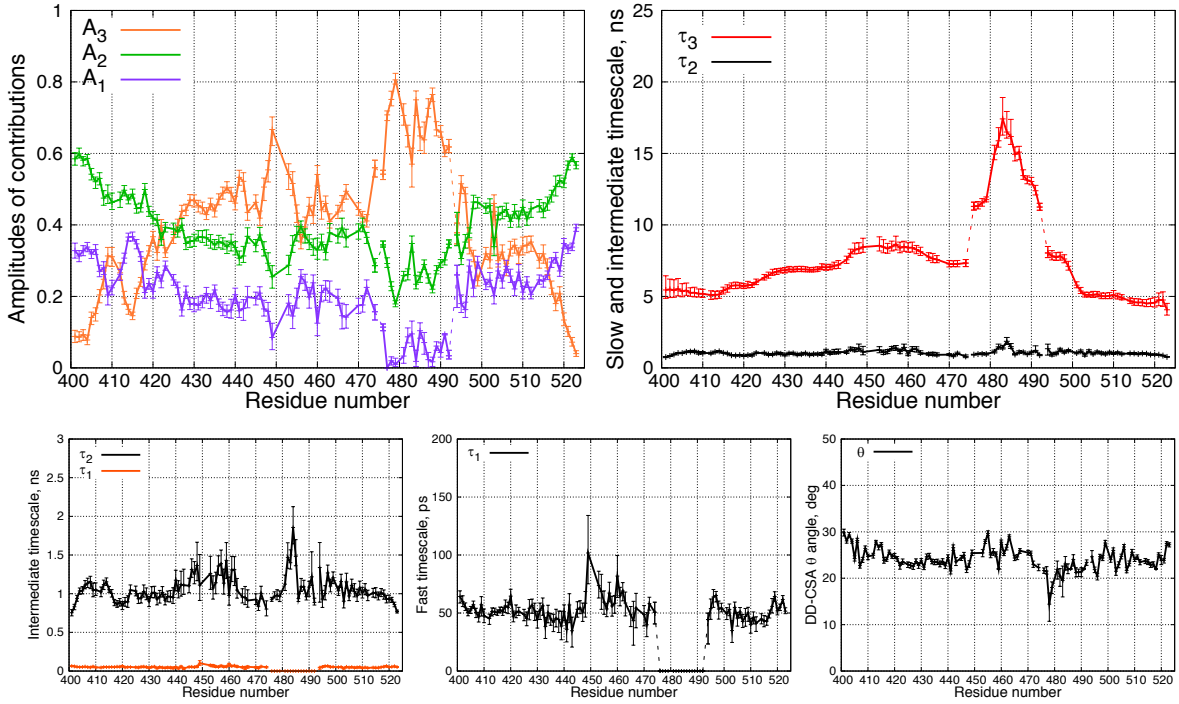


Fig. 6.11. Fitting relaxation data for $N_{TAIL,L}$ at 278K using model-free approach with τ_{slow} window size of 7 residues.

In order to avoid overlapping segments containing helical and non-helical regions, the helical region of $N_{TAIL,L}$ was analysed separately and the strands outside this region as separate chains, and then combined. On Fig. 6.11-6.13 the link between the last analysed residue of the non-helical part and first analysed residue of the helical region (and vice versa) is shown by the dashed line. The correlation time of fast motion was fitted outside helical region and fixed at 0 ps inside it. Comparing Fig. 6.11 (sliding window model-free analysis) and Figure 8 and S10 in the manuscript (model free analysis of individual residues), we observe a high similarity in the sequence profile of all parameters, however, not surprisingly, experimental uncertainties, in particular of τ_3 values, are smaller in the sliding window analysis. The sequence profile of fitted parameters outside the helical region is virtually identical; in the helical region, the overall shape of τ_3 is conserved in the sliding window analysis.

Analyses with windows of 3 and 11 residues are shown on Fig. 6.12. The overall shape of fitted parameters outside helical region is preserved compared to individual-residue analysis. When comparing reduced χ^2 values (that is, sum of squares of fit residuals divided by (number of fit conditions – number of fit parameters)), we observe that the fit is better in the case of 7-residue window as compared to 3-residue and 11-residue window. Finally, the value of 7 residues is in agreement with cited previously studies of R_2 rates in disordered protein. We are currently investigating the possibility of optimizing the window length for different regions of protein and its dependence on sequence composition, and this will be discussed more in detail in the end of this chapter.

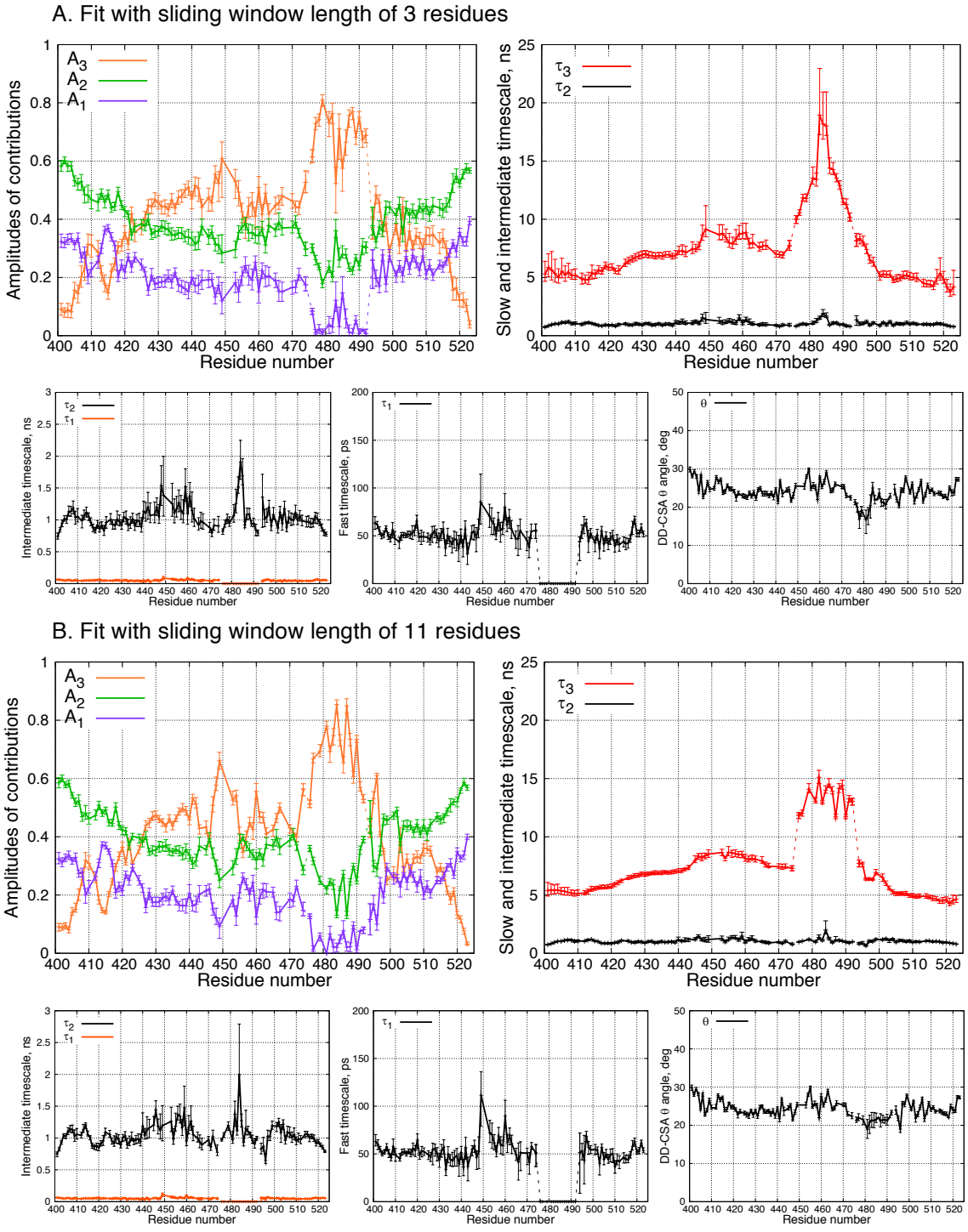


Fig. 6.12. Fitting relaxation data for $N_{TAIL,L}$ at 278K with alternative window sizes

Application of the 7-residue sliding-window model-free approach to the analysis of $N_{TAIL,L}$ and $N_{TAIL,M}$ relaxation at 298K is shown on Fig. 6.14. Again, the sequence profile of fitted parameters is very similar in two approaches (with and without sliding window), but the improvement of the τ_3 uncertainty is remarkable: now we clearly see that two constructs share the same dynamical properties in the flexible N-terminal part right to the helix, but in the helix the difference in the correlation motion of slow motion is evident, and the dependence on the chain length is clearly visible, substantiating the results presented in the manuscript, based on analysis of individual residues.

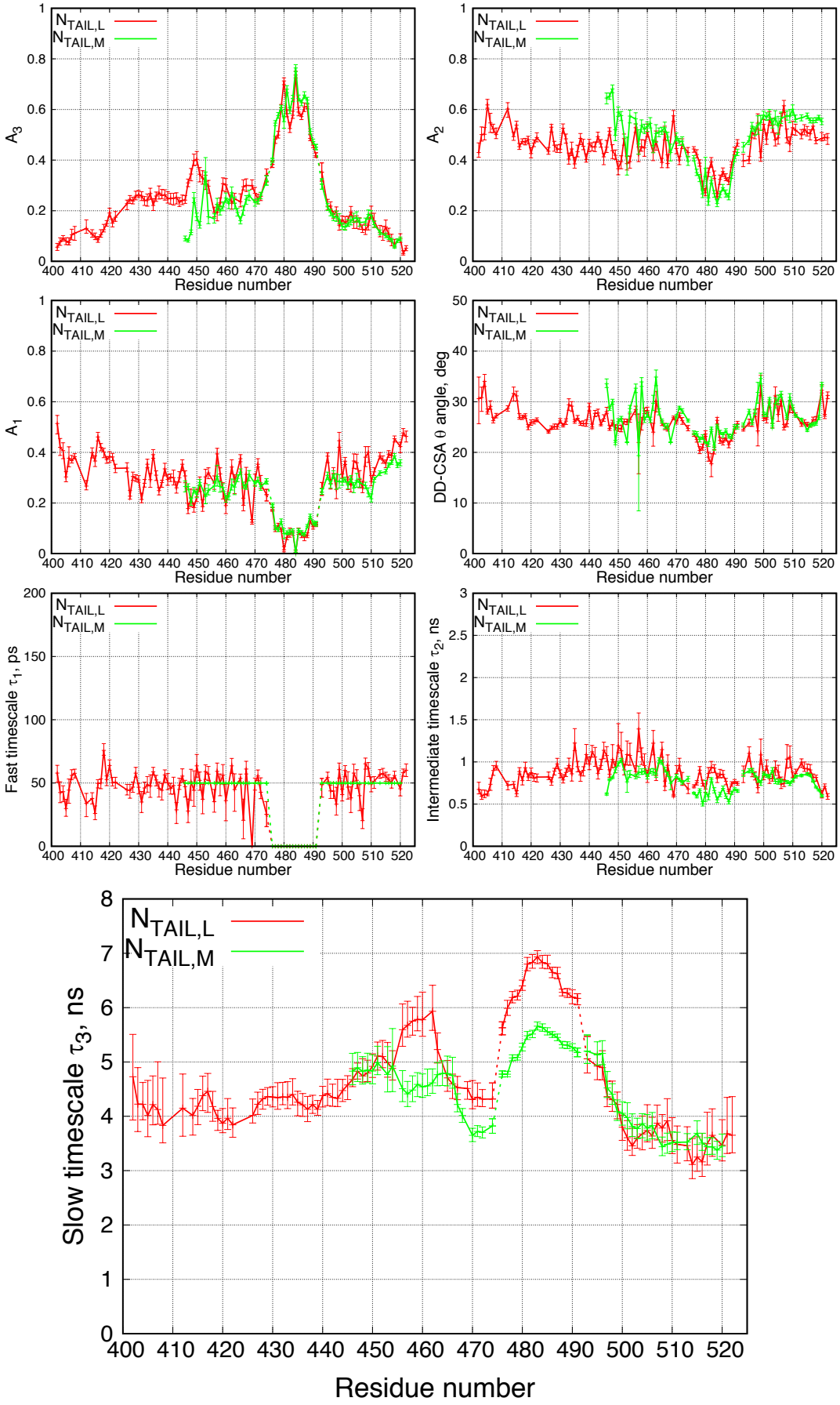


Fig. 6.13. Model-free analysis of $N_{TAIL,L}$ and $N_{TAIL,M}$ forms at 298K with a sliding window of 7 residues.

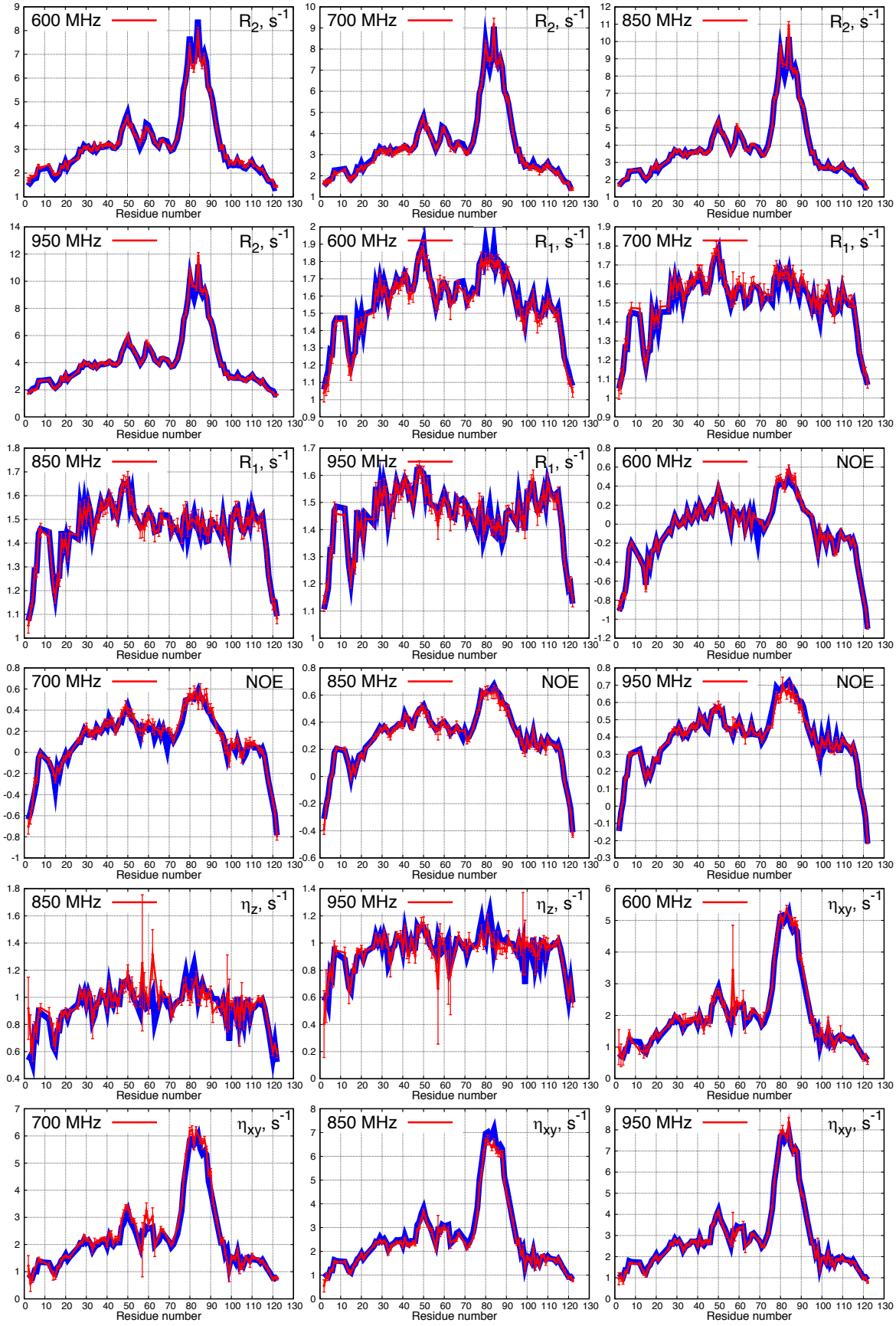


Fig. 6.14. Quality of data reproduction by the seven-residue sliding window model-free analysis of $N_{TAIL,L}$ at 298K.

6.4.5 Relaxation dispersion studies of conformational exchange in $N_{TAIL,L}$

On the basis of the comparison of $J(0)$ values extracted from exchange-prone auto-relaxation rates and from exchange-free cross-correlated cross-relaxation rates (Figures S3 in the Article), we can already conclude that there is negligible exchange on timescales to which the $R_{1\rho}$ experiment with B_1 field of 1500 Hz would be sensitive (hundreds of microseconds or slower). To probe faster timescales, high-power ^{1}H $R_{1\rho}$ relaxation dispersion experiments [44, 216] were carried out at 268K and 700MHz, in collaboration with the Griesinger group a the MPI in Göttingen where specific cryo-probe equipment is available that can withstand high levels of power deposition. Relaxation rates at different spin-lock and effective fields are shown on Fig. 6.15.

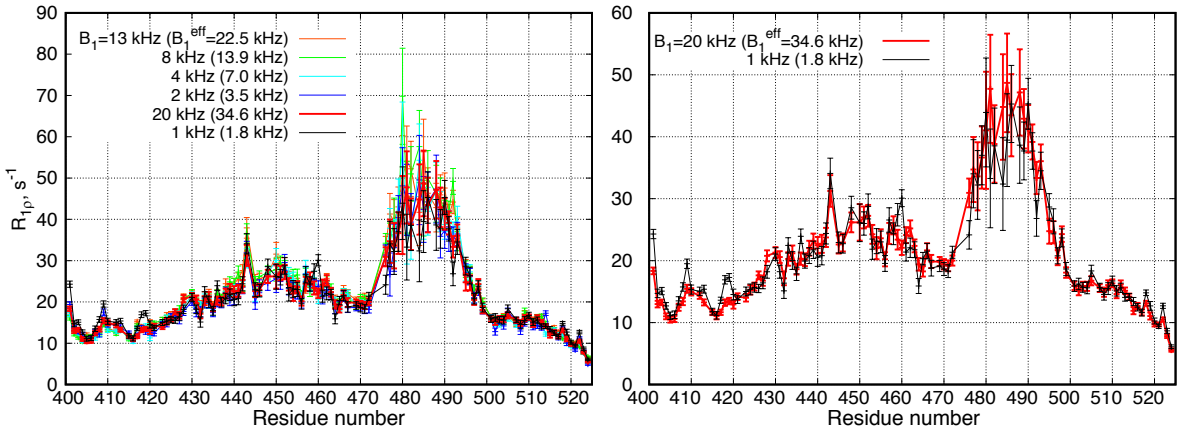


Fig. 6.15. ^{1}H $R_{1\rho}$ relaxation dispersion in $N_{TAIL,L}$ at 268K. For each spin-lock field strength B_1, we show the associated values of effective field B_1^{eff}. Left plot: $R_{1\rho}$ relaxation profile for all spin-lock fields. Right plot: $R_{1\rho}$ relaxation profile only for the lowest and highest spin-lock fields.

$R_{1\rho}$ relaxation rates at spin-lock fields up to 20 kHz were measured, which corresponded to effective field of almost 35 kHz, and no significant change in $R_{1\rho}$ rates was observed. This observation reveals that if the conformational exchange is present in $N_{TAIL,}$ it should be on timescales shorter than $1/B_{1,max}^{eff} = 30$ µs. As discussed in Chapter 2, the MoRE region of N_{TAIL} is predicted to have three exchanging helical conformers, putting the upper-limit of 30 µs on the timescale of this exchange.

6.4.6 Discussion about the timescale of conformational exchange in MoRE of $N_{TAIL,L}$

^{1}H $R_{1\rho}$ relaxation dispersion experiments on N_{TAIL} allowed us suggesting the upper limit for the timescale of interconversion between helical conformers in $N_{TAIL,L}$. Here we present an observation that suggests a potential lower limit for this interconversion timescale.

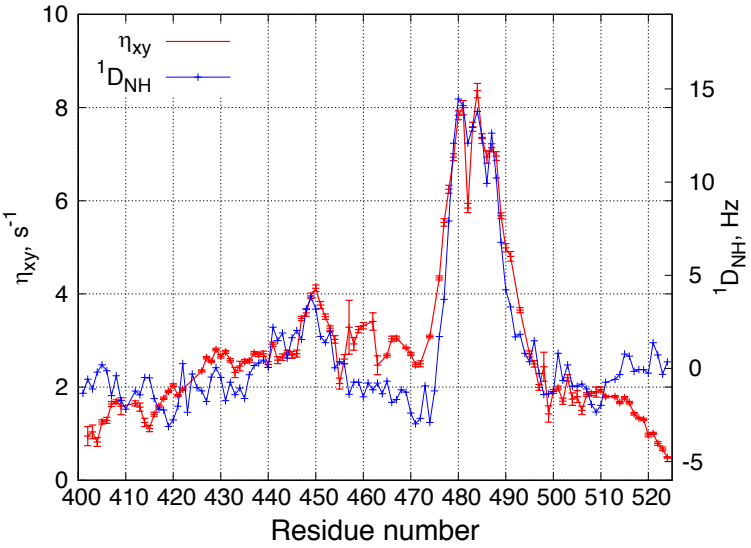


Fig. 6.16. Comparison of the ^{15}N-^{1}H dipole-dipole – ^{15}N CSA cross-correlation transversal cross-relaxation rate, η_{xy} (red line), and $^{1}D_{NH}$ RDC (blue line), both measured in $N_{TAIL,L}$ at 298K. η_{xy} rate was measured at 950 MHz. $^{1}D_{NH}$ RDC values were scaled in order to reveal the identical positions of local maxima in the helical region.

We noticed that the sequence profiles of the transversal relaxation rates – R_2 and η_{xy}, have the same local maxima in the helical region as the sequence profile of $^{1}D_{NH}$ residual dipolar couplings (measured in PEG-hexanol alignment medium). Here we compare $^{1}D_{NH}$ and the cross-correlated cross-relaxation rate η_{xy} (which has no potential exchange contribution). Remarkably, their profiles in the helical region (476-492) share all maxima and mimina. In RDCs, these maxima and minima were predicted to stem from the population averaging of three different helical conformers in the MoRE region of N_{TAIL} [111], interconverting on a timescale faster than millisecond. Alba et al. [222] proposed that the alignment tensor and diffusion tensor are both influenced by the molecular shape, and thus may have similar orientations; if they do have similar orientations, then a correlation in R_2/R_1 ratio and RDC values would be observed. In the case of interconverting structural elements as in the MoRE of N_{TAIL}, we can extend this reasoning by adding a condition that also the averaging properties should be similar; that is, if RDCs are defined by population-averaged alignment tensors of individual conformers, then relaxation rates should also be defined by a population average of diffusion tensors of these conformers, for the correlation to be observed. However, relaxation rates depend on a population average of diffusion tensors only when the interconversion of individual conformers is much slower than timescales of relaxation-active motions, otherwise the interconversion would contribute itself to the relaxation. Thus, we suggest that, even if the correlation observed on Fig. 6.16 in the helical region is not in any case an evidence of interconversion being in the 100s ns – 30 μs timescale range, it is in agreement with such hypothesis.

It is difficult to study dynamics in this (hundreds of ns – tens of μs) timescale range, because these motions are too slow to contribute to relaxation, and still too fast to give any relaxation dispersion at currently available spin-lock fields. One of possible options to study these motions are MD simulations. Remarkably, as we discussed in section 4.2.4, Lindorff-Larsen et al. found that the timescales of formation and breaking of helices in their study are situated in the range 600 ns – 5 μs.

Chapter 7. Conclusion and perspectives

Intrinsically disordered proteins play a crucial role in our cells, being involved in signal transduction, molecular recognition, transcription, translation and other processes, and represent potential drug targets against cancer and neurodegenerative diseases. Their conformational sampling is the subject of detailed studies, but still not much is known about their dynamics.

^{15}N NMR relaxation offers the most powerful probe of protein dynamics, as it provides site-specific information on a broad range of timescales (picoseconds to nanoseconds). In folded proteins, relaxation rates were traditionally analysed by the model-free approach, assuming that the only relaxation-active motions are the slow overall tumbling, and fast internal motions. It was recently pointed out that in an IDP the underlying behaviour is that of a random-coil chain [181], which describes the protein as a polymer consisting of 20 different monomers [192]; so that polymer models may also be relevant [185].

While such a complex object as an IDP cannot be reduced to a simple homopolymer chain, we should also not underestimate the extent to which chain-like motions may govern the dynamics of IDPs. In particular, Rouse and Zimm models predict that there is a distribution of timescales of motions – or 'relaxation modes' – in polymer chains; and for longer chains slower motions would be observed. In IDPs, the distribution of correlation times was equally proposed [178, 187], and a dependence of slow motions on chain length was recently observed by relaxometry [195].

We studied here an intrinsically disordered protein – N_{TAIL} – by ^{15}N relaxation. Apart from being important and interesting in the context of viral replication, this protein represents a particularly relevant model for IDP studies, as it contains a long mostly disordered chain and a small region with transient helical population. Over 100 different relaxation rates were measured in total at multiple magnetic fields, multiple temperatures and for three N_{TAIL} constructs of different length. We analysed relaxation data simultaneously at multiple temperatures using an extended model-free analysis and an Arrhenius-like relationship for the temperature dependence of correlation times. The obtained values of correlation times and of activation energies were in agreement with the following physical description of fitted timescales of motions: fast motions on 10s ps timescale report on local librational motions; intermediate motions report on local Ramachandran sampling of each backbone residue and do not stem from the chain-like nature of IDPs; finally, the slowest component reflects segmental (or chain-like) motions within which modes are coupled. The observed dependence of the slowest component on the chain length reinforced its assignment to chain-like motions. The chain-length dependence was

much less evident in the disordered parts of the protein, compared to the helical region; we propose that slower motional modes that depend on chain length (or that appear in longer chains) are quenched in the more flexible parts by faster motions, but they are more pronounced in helix because these faster motions are restricted. A possible use of this observation would be to introduce artificial helical elements in disordered proteins as 'probes' of slow motions.

Evidently, not all dynamic processes in IDPs are contributing to NMR relaxation: based on high-power proton $R_{1\rho}$ relaxation measurements and on comparison of ${}^{1}D_{\mathrm{NH}}$ RDCs and η_{xy} values in helical region, we propose that the timescale of interconversion between three helical constructs in this region is located in 100s ns – 30 μs range, which is compatible with values of 600ns and 5 μs found recently by MD simulations in another system [182]

Even though our Arrhenius-equation model-free analysis allowed the identification of different motional modes in an IDP, and chain-like segmental motions in particular, it was still a residue-specific analysis, that did not actively invoke that possible coupling of motional modes of different residues in the chain. In such a model, residues that are close in the sequence should generally have similar dynamic parameters, and in particular similar values of slow motion correlation times, which is in fact what is observed experimentally.

We therefore discuss a quite naïve approach aimed at exploiting this property of chain dynamics, which we call 'sliding window model-free', whereby values of slow motional correlation times are constrained to be common within any window (or segment) of a given length. Despite its simplicity, this model, that accounts for a common slow motional timescale, yielded good results when fitting relaxation data at 278K (fitted parameters were highly similar to those fitted in the standard model-free analysis), and, moreover, clarified the analysis of two $\mathrm{N_{TAIL}}$ constructs of different length at 298K. From the parameters fitted by standard model-free analysis, we were unable to describe properly the influence of chain length on the correlation time of slow motion at 298K; and after applying the 'sliding window', the difference in the dynamics of two constructs become clear.

In this 'sliding window' or segment analysis, we used segment size of 7 residues, which gave statistically better values than 3 and 11 residues. However, we are actually working on a more systematic segment analysis that would involve, for example, optimizing a set of non-overlapping segments along the sequence using a genetic algorithm.

Ideally, this segment analysis should take into account the physical properties of different amino acids in the sequence, and in particular the side chain bulkiness. As a very simple

example, we've seen on Fig. 6.8 that local minimums in the transversal relaxation rate correspond to the presence of pairs and triplets of glycines. In the future, the description of motional modes in IDPs should not only include relaxation rates, which report mainly on local dynamics, but also other measurements that report on long-range interactions, such as paramagnetic relaxation enhancements (PREs) and fluorescence resonance energy transfer (FRET) measurements. We place this work in the context of a new era of IDP dynamics studies where NMR relaxation contributes to an integrated dynamic approach to characterizing the dynamics of IDPs.

8. Résumé en français

La thèse commence par l'introduction, où nous présentons un court rappel de l'histoire de l'étude des protéines en général et des protéines désordonnées en particulier. Ensuite nous expliquons les avantages de la résonance magnétique nucléaire (RMN) pour l'étude des protéines intrinsèquement désordonnées, et enfin nous présentons les différents chapitres de la thèse.

8.1 Le monde des protéines intrinsèquement désordonnées

Le premier chapitre commence par la présentation des approches de prédiction du désordre à partir de la séquence primaire d'une protéine. Nous discutons ensuite les méthodes expérimentales pour étudier les protéines désordonnées, telles que le dichroïsme circulaire, la diffusion aux petits angles, les approches à molécule unique, et en particulier la RMN. Comme les études de ces protéines ont majoritairement été réalisées *in vitro*, la question de la présence du désordre dans le contexte cellulaire s'est posée, et nous présentons les études de protéines intrinsèquement désordonnées *in cell*. Nous discutons ensuite la physico-chimie des interactions de ces protéines avec les partenaires cellulaires, et enfin leur rôle dans les maladies humaines.

8.2 Le contexte et l'objet du travail

Dans le deuxième chapitre, nous introduisons l'objet de notre étude – le domaine C-terminal intrinsèquement désordonné (N_{TAIL}) de la nucléoprotéine du virus de Sendai. Dans le cadre de la réplication virale, N_{TAIL} interagit avec le domaine PX de la phosphoprotéine du même virus. Il a été montré dans les études précédentes que le site d'interaction de N_{TAIL} échantillonne trois conformations hélicoïdales transitoires. Dans une étude plus récente, le mécanisme d'interaction entre N_{TAIL} et PX a été décrit en détail, et il a été montré qu'un complexe « de rencontre » non-spécifique se forme, stabilisant l'une des conformations hélicoïdales du site d'interaction, avant l'amarrage de cette conformation hélicoïdale à la surface de PX. La dynamique de la chaîne désordonnée de N_{TAIL} joue certainement un rôle important dans la formation du complexe de rencontre, en particulier au niveau de la constante de vitesse d'association. Ce domaine de 124 acides aminés est aussi une protéine intrinsèquement désordonnée modèle contenant à la fois de longs domaines dépliés et des régions de structure résiduelle, dont l'échantillonnage conformationnel a été caractérisé, mais dont la dynamique reste encore mal comprise. Dans cette thèse, nous étudions la dynamique de N_{TAIL} en solution par la relaxation RMN, dans le but de répondre à des questions suivantes :

- Est-il possible de caractériser les échelles de temps des mouvements de N_{TAIL} ?
- Auxquels processus physiques correspondent ces échelles de temps?
- Comment ces échelles de temps de mouvements dépendent de la longueur de la chaîne et de la température ?
- Comment la dynamique locale d'un résidu dans la chaîne est influencée par le comportement des résidus voisins ?
- Est-ce que nous pouvons estimer l'échelle de temps de l'interconversion des conformations hélicoïdales dans le site d'interaction de N_{TAIL} ?

8.3 La relaxation RMN

Le chapitre 3 est consacré à des aspects théoriques de la relaxation RMN utiles pour la compréhension du travail réalisé dans le cadre de cette thèse. Nous commençons par l'introduction des quantités essentielles et des équations utilisées pour la description du phénomène de la résonance magnétique nucléaire. La relaxation RMN, ses origines et le lien entre les taux de relaxation et la fonction de corrélation angulaire sont expliqués dans le cadre de la théorie semi-classique de relaxation. Nous introduisons les différents mécanismes de relaxation, la corrélation croisée entre plusieurs mécanismes de relaxation et leur contribution aux différents taux de relaxation mesurés dans notre étude. Enfin, nous résumons les équations pour tous les taux de relaxation utilisées dans cette thèse, introduisons la contribution de l'échange conformationnel pour chaque taux de relaxation, et comparons les taux de relaxation caractéristiques d'une protéine repliée et les taux de relaxation mesurés pour la protéine d'étude, N_{TAIL}.

8.4 La dynamique des protéines dépliées et les études RMN des protéines flexibles et intrinsèquement désordonnées

Nous commençons le chapitre 4 par la présentation de quelques modèles physiques simples décrivant la dynamique de chaînes des polymères. Ensuite, nous nous intéressons à des publications précédentes qui représentent un intérêt pour l'étude de la dynamique des protéines intrinsèquement désordonnées. Il était connu depuis longtemps que tout mouvement markovien peut être décrit par une fonction de corrélation en forme d'une somme des contributions exponentielles. L'approche dite « model-free », dans laquelle on considère qu'il existe une réorientation globale de la molécule dont les mouvements internes sont statistiquement séparés, repose sur cette propriété. La fonction de corrélation dans cette approche contient deux contributions exponentielles – l'une pour la réorientation globale, l'autre pour les mouvements internes rapides, et les paramètres dynamiques sont ajustés pour avoir une reproduction optimale des taux de relaxation mesurés. Cette approche était subséquemment adaptée pour les cas ou les mouvements internes peuvent avoir deux échelles de temps – l'une plus rapide, de quelques dizaines ou centaines de picosecondes, et l'autre plus proche des nanosecondes. Dans cette approche dite « extended model-free » ou « model-free » étendu, trois contributions exponentielles à la fonction de corrélation sont nécessaires.

Plusieurs tentatives d'appliquer l'approche « model-free » à l'étude de la dynamique dans les protéines dépliées ou intrinsèquement désordonnées sont publiées ; mais dans ces systèmes hautement flexibles les mouvements les plus lents contribuant à la relaxation ne peuvent pas être associés à la réorientation globale, et la supposition initiale de l'analyse « model-free » n'est plus applicable. Plusieurs approches adaptant l'analyse « model-free » à des protéines intrinsèquement désordonnées ont alors été proposées, tenant compte notamment de la distribution des échelles de temps de mouvements dans ces protéines. Les études les plus récentes montrent que l'approche « model-free » est également valable quand les échelles de temps des mouvements sont séparées, même s'il n'y a pas de réorientation globale, et que cette séparation des échelles de temps existe dans les protéines intrinsèquement désordonnées.

Nous discutons également les études où la dynamique de la chaîne a été mise en avant dans la description du comportement des protéines intrinsèquement désordonnées. D'abord, nous présentons l'analyse des taux de relaxation transversale qui traite une protéine désordonnée comme une chaîne ayant une certaine longueur de persistance. Ensuite, nous discutons les études de ces protéines par la relaxométrie RMN et les simulations de la dynamique moléculaire, qui attribuent les mouvements lents (à l'échelle de temps de nanosecondes) à la dynamique de la chaîne désordonnée, ayant une dépendance caractéristique à la longueur de la chaîne et à la viscosité.

Nous concluons que les protéines désordonnées ont un comportement complexe, qui comporte certains aspects de la dynamique de la chaîne polymère, et qui peut être analysé, malgré sa complexité, par une approche « model-free » à trois contributions exponentielles.

8.5 Matériels et méthodes

Dans ce chapitre nous présentons les aspects expérimentaux de notre étude, tels que la préparation de l'échantillon, l'attribution du squelette peptidique de N_{TAIL} et de ses formes plus courtes, la calibration du spectromètre RMN, la mesure des taux de relaxation RMN, les séquences d'impulsions utilisées, et quelques aspects du traitement et de l'analyse des données expérimentales.

8.6 Résultats

Le chapitre 6 présente tous les résultats que nous avons obtenus dans notre travail.

D'abord, nous décrivons l'échantillonnage conformationnel de N_{TAIL} à l'aide de l'approche ASTEROIDS qui se base sur la sélection d'ensembles de conformations étant en accord avec les déplacements chimiques. La région MoRE de N_{TAIL} échantillonne presque exclusivement la partie hélicoïdale (α_L) de l'espace de Ramachandran, tandis que les résidus dans les chaines désordonnées échantillonnent plusieurs régions de cet espace (α_L, β_S, β_P) à la fois.

Les résidus dans les régions désordonnées et dans le MoRE de N_{TAIL} ont également un comportement distinct à l'égard des déplacements chimiques dans les spectres ^{1}H-^{15}N HSQC. Si, avec la température, les signaux correspondant aux résidus désordonnés se déplacent dans la même direction avec la même vitesse, ce n'est pas le cas pour les résidus du MoRE, dont les signaux se déplacent à une vitesse inferieure et changent la direction avec la température. Les coefficients de température de déplacements chimiques sont très peu négatifs, voire positifs pour les résidus du MoRE, en accord avec sa structure hélicoïdale impliquant la formation de liaisons hydrogènes.

Article : Identification des modes dynamiques d'une protéine intrinsèquement désordonnée à partir de la dépendance à la température de la relaxation RMN

La majeure partie des résultats de notre travail est décrite dans l'article qui a été publié récemment.

Introduction

Le rôle de la séquence primaire des protéines intrinsèquement désordonnées n'est pas de stabiliser une conformation tridimensionnelle, mais de permettre l'échantillonnage d'un grand nombre de conformations sur une vaste surface plate d'énergie libre. Les méthodes de la spectroscopie RMN, telles que la mesure des déplacements chimiques et des couplages dipolaires résiduels, en combinaison avec des techniques d'échantillonnage et de sélection (Flexible-Meccano, ASTEROIDS) ou avec les simulations de dynamique moléculaire, nous donnent beaucoup d'informations sur l'espace conformationnel et sur la nature des minima d'énergie libre de ces protéines. Cependant, peu d'informations sont disponibles à propos des échelles de temps de la dynamique des protéines intrinsèquement désordonnées. Parmi les méthodes qui caractérisent la dynamique locale dans une protéine, la relaxation RMN est sans doute l'une des plus puissantes, car elle nous donne l'information à l'échelle de chaque résidu dans le squelette peptidique.

Dans les protéines repliées, les taux de relaxation sont fréquemment analysés en utilisant l'approche « model-free ». Dans cette approche, il existe deux types de mouvements contribuant à la relaxation: la réorientation globale de la molécule et les mouvements internes qui en sont statistiquement séparés. Mais le concept de la réorientation globale est peu pertinent dans le contexte des protéines désordonnées. Différentes adaptations de l'analyse « model-free » pour les protéines désordonnées ont été proposées pour mieux prendre en compte le fait que ces protéines sont des polymères hautement flexibles, et incluant notamment l'hypothèse de la distribution des temps de corrélation. Très récemment, il a été montré que l'approche classique dite « model-free » étendue, incluant trois échelles de temps dans la fonction de corrélation, nous permet de saisir et d'interpréter de façon très exacte l'information présente dans les taux de relaxation mesurés à plusieurs champs.

Malgré tout, il reste difficile d'attribuer les différentes échelles de temps des mouvements, qui contribuent à la relaxation à des processus physiques. Afin d'avoir une meilleure compréhension de la nature des mouvements dans les protéines intrinsèquement désordonnées, nous étudions N_{TAIL} (ou NT), le domaine C-terminal intrinsèquement désordonné de la nucléoprotéine du virus de Sendai. N_{TAIL} contient 124 acides aminés (résidus 401-524 de la nucléoprotéine) formant une chaine quasi totalement désordonnée, à l'exception d'une courte région MoRE (motif de reconnaissance moléculaire) qui interagit avec le domaine PX de la phosphoprotéine. Cette interaction présente un intérêt dans le cadre d'étude de la réplication virale, où la dynamique de la chaîne de N_{TAIL} joue un rôle important, en particulier en ce qui concerne la constante de vitesse d'association en particulier. Mais N_{TAIL} est surtout une protéine intrinsèquement désordonnée modèle, contenant à la fois une longue chaine désordonnée et une courte région d'interaction (MoRE), qui, comme cela a été montré dans les études précédentes, échantillonne trois conformations hélicoïdales à une échelle de temps inférieure à 100 µs.

Résultats

Nous avons comparé l'échantillonnage conformationnel de N_{TAIL} à 278K et 298K en analysant des déplacements chimiques secondaires $^{13}C^{\alpha}$. La propension à former des structures secondaires reste stable dans les régions désordonnées de la protéine ; dans

la partie MoRE hélicoïdale (résidus 477-492), la propension à former une hélice est un peu plus forte à 278K (5-7%). Les taux de relaxation dans le MoRE sont en accord avec la formation des éléments de structure secondaire. Les mesures de relaxation révèlent aussi qu'une structure locale est également présente autour des deux tryptophanes, W443 et W449, et que la chaine est très flexible autour du résidu 415, en accord avec la présence d'un triplet de glycines.

Pour exclure une contribution de l'échange conformationnel à la relaxation transversale R_2, nous avons comparé les valeurs de la fonction de densité spectrale à plusieurs champs, et les valeurs étaient les mêmes, qu'elles soient obtenues à partir des taux d'autorelaxation ou à partir des taux de cross-relaxation cross-corrélée. Ce résultat a été confirmé pour chaque température à laquelle les taux de relaxation étaient mesurés, confirmant ainsi que les données sont mutuellement cohérentes.

Nous avons analysé les données de relaxation pour chaque température séparément, en utilisant l'approche « model-free » avec la fonction de corrélation ayant deux ou trois contributions exponentielles indépendantes. Les paramètres étaient ajustés pour chaque résidu séparément. En utilisant le test de Ficher ou le critère d'information d'Akaike, nous avons mis en évidence que la fonction de corrélation avec trois contributions permet la meilleure reproduction des données expérimentales pour tous les résidus de N_{TAIL}, sauf ceux de la partie MoRE hélicoïdale, ou deux contributions exponentielles étaient suffisantes. Le temps de corrélation de la contribution la plus rapide a été ajusté à une valeur très proche de 50 ps pour toutes les températures. Pour la contribution de mouvements à l'échelle de temps intermédiaire, nous avons trouvé un temps de corrélation entre 0.5 et 1.3 ns, et pour la contribution des mouvements les plus lents entre 5 et 25 ns, avec les maxima dans la région hélicoïdale de N_{TAIL}.

Ensuite, nous avons analysé les données pour toutes les températures simultanément (274K, 278K, 288K et 298K), en liant les temps de corrélation des échelles de temps des mouvements « lents » et « intermédiaires » par la relation d'Arrhenius, ce qui a impliqué, en particulier, l'ajustement des énergies d'activation effectives pour ces mouvements. Le temps de corrélation et l'énergie d'activation de la contribution des mouvements lents ont une distribution « en cloche » le long de la séquence. L'échelle de temps des mouvements lents a une plus forte dépendance à la température que l'échelle de temps intermédiaire, ce qui est reflété par une plus haute énergie d'activation associée, allant jusqu'à 20-25 kJ·mol^{-1} au centre de la chaîne. Les mouvements « intermédiaires » ont une distribution de l'énergie d'activation plus plate le long de la séquence, autour de 5 kJ·mol^{-1} pour les résidus 425-500, sauf la partie hélicoïdale, où elle monte jusqu'à 15 kJ·mol^{-1}. L'amplitude de la contribution des mouvements lents devient plus importante et l'amplitude de la contribution des mouvements rapides se réduit quand la température descend. La contribution des mouvements à l'échelle de temps intermédiaire a un maximum d'amplitude à 288K.

En addition, nous avons analysé les taux de relaxation à 278K pour deux formes de N_{TAIL} plus courtes : $N_{TAIL,M}$ (443-524) et $N_{TAIL,C}$ (443-501). Nous observons que les amplitudes des contributions et l'échelle de temps intermédiaire sont très similaires pour les trois formes (incluant N_{TAIL} complet ou $N_{TAIL,L}$) tout au long de la séquence. Par contre, le temps de corrélation de la contribution associée aux mouvements les plus lents augmente clairement dans la partie hélicoïdale de N_{TAIL} avec la longueur de la chaîne.

Discussion

Notre analyse suit l'approche établie auparavant, où la fonction de la densité spectrale est modélisée par la somme des contributions lorentziennes, ce qui suppose implicitement la modélisation des fonctions d'auto- et de cross-corrélation par une combinaison des contributions exponentielles ayant des échelles de temps bien séparées. Nous avons mesuré jusqu'à 5 taux de relaxation par intensité de champ magnétique et à 5 températures différentes, ayant au total 61 taux de relaxation pour $N_{TAIL,L}$, et 122 taux de relaxation au total si l'on prend en compte les formes $N_{TAIL,M}$ et $N_{TAIL,C}$. L'analyse des données mesurées a montré qu'il faut trois contributions à la fonction de corrélation pour reproduire les données expérimentales.

Le temps de corrélation de la contribution la plus rapide est autour de 50 ps pour tous les résidus de la chaîne de N_{TAIL}, sauf pour sa partie hélicoïdale, où le temps de corrélation est trop rapide pour être déterminé et était fixé à 0 ps. La dépendance à la température de cette contribution est négligeable. Nous attribuons cette contribution à des mouvements essentiellement librationnels, représentant une diffusion dans un potentiel plat de l'énergie libre, dont les limites géométriques augmentent avec la température, mais qui n'a pas de barrières énergétiques détectables.

L'amplitude des contributions « rapides » et « intermédiaires » oscille avec l'alternance des résidus chargés et hydrophobes dans la région de N_{TAIL} entre les résidus 424 et 438, et il semble que ces contributions représentent les processus dont la dynamique dépend directement de la séquence primaire. Les mouvements « intermédiaires » ont un temps de corrélation à l'échelle de temps de la nanoseconde, et l'amplitude de leur contribution à la perte de corrélation est très importante, allant jusqu'à 50% à 288 et 298K. Cette amplitude a un maximum à 288K (sauf pour la région hélicoïdale). Nous supposons qu'en changeant la température, nous déplaçons la distribution des modes de mouvements à travers le diapason des fréquences auxquelles la relaxation RMN est le plus sensible, et que le maximum de cette distribution se trouve autour de quelques nanosecondes. Le temps de corrélation et l'amplitude de la contribution des mouvements « intermédiaires » ont une distribution plate le long de la séquence, sauf pour la région hélicoïdale et les terminus de N_{TAIL}. Les valeurs de l'énergie d'activation de cette contribution sont comparables avec les valeurs obtenues dans les études précédentes de l'échantillonnage conformationnel des polypeptides. Toutes ces observations nous suggèrent que cette contribution « intermédiaire » nous informe sur l'échantillonnage local des différentes conformations dans l'espace de Ramachandran, même s'il est moins clair qu'il s'agisse de la diffusion conformationelle dans un seul minimum du paysage de l'énergie libre associé à cet espace ou entre plusieurs minima. En ce qui concerne la partie hélicoïdale, les valeurs élevées de l'énergie d'activation de la contribution « intermédiaire » sont en accord avec les valeurs obtenues dans l'étude précédente des énergies d'activation des mouvements dans l'hélice d'une protéine repliée.

Finalement, nous associons les mouvements les plus lents à des mouvements des segments de la chaîne désordonnée de N_{TAIL}. Leur dépendance à la température, et en particulier du rapport viscosité-température, est très semblable à celle prédite par la loi de Stokes, ce qui suggère que cette contribution « lente » reflète les mouvements de la chaîne désordonnée dominés par la viscosité. Les paramètres dynamiques de la contribution de ces mouvements (temps de corrélation, amplitude, énergie d'activation) ont un profil le long de la séquence plutôt « en cloche » que plat, avec les effets de

terminus se faisant sentir jusqu'à 15-20 résidus de l'extrémité de la chaine, en comparaison avec ~5 résidus pour la contribution « intermédiaire », soutenant ainsi l'idée que cette contribution représente les mouvements des segments dont les modes dynamiques sont liés l'un à l'autre et à la couche de solvatation locale. De plus, nous observons la dépendance de l'échelle de temps de cette contribution dans la partie hélicoïdale à la longueur de la chaine de N_{TAIL}, ce que nous associons à l'effet de la traînée moléculaire liée à la viscosité intrinsèque de la chaîne. Il semble que nous ne puissions observer cet effet que dans la partie hélicoïdale de la protéine car dans les parties désordonnées de N_{TAIL}, la fonction de corrélation est presque à zéro à des échelles de temps correspondant à ces mouvements, à cause des mouvements plus rapides. Dans la partie hélicoïdale de N_{TAIL}, les mouvements rapides sont restreints, et la relaxation RMN est donc plus sensible à ces mouvements de la chaîne qui dépendent de sa longueur. A cet égard, nous proposons l'insertion des éléments hélicoïdaux artificiels dans les protéines intrinsèquement désordonnées pour mieux étudier leur dynamique associée à des mouvements de la chaîne.

Conclusions

Cette étude, où nous avons analysé 61 taux de relaxation du N_{TAIL} du virus de Sendai, nous a permis d'attribuer des énergies d'activation locales à chaque résidu et d 'identifier trois modes dynamiques différents, liés à des librations, à l'échantillonnage conformationnel local et à des mouvements de la chaîne. La capacité de délimiter ces modes dynamiques distincts nous permettra de mieux comprendre le comportement et la fonction des protéines intrinsèquement désordonnées.

Résultats complémentaires et discussion

Dans cette partie du chapitre, nous présentons les résultats importants des études réalisées dans le cadre de cette thèse, mais qui n'ont pas été inclus dans le manuscrit.

Nous commençons par l'analyse de l'évolution de l'échantillonnage conformationnel de $N_{TAIL,L}$ avec la température, que nous complétons par la comparaison des déplacements chimiques secondaires $^{13}C'$ à 298K et 278K. Ensuite, nous discutons plus en détail la dépendance à la température des taux de relaxation de $N_{TAIL,L}$, en rajoutant notamment qu'il est possible d'étudier la dynamique des protéines désordonnées même à 258K en état soluble. Nous remarquons la formation d'une région au milieu de la chaîne désordonnée ou les mouvements sont plus restreints, et qui est liée à la présence de deux tryptophanes, qui peuvent aussi interagir avec la partie hélicoïdale de N_{TAIL}. La mutation de ces tryptophanes en alanines conduit à la disparition de cette restriction de mouvements. Cependant, le profil ne devient pas plat et l'on continue d'observer les minima et maxima des taux de relaxation dans cette région, que nous associons à des particularités locales de la séquence primaire de N_{TAIL}. D'autre part, la mutation a une influence négligeable sur la dynamique de la partie hélicoïdale de N_{TAIL}, ce qui nous permet de conclure que la présence des tryptophanes est moins importante pour l'étude de la dynamique de la région hélicoïdale, et de justifier l'étude des formes plus courtes de N_{TAIL} ($N_{TAIL,M}$ et $N_{TAIL,C}$) ne contenant qu'un seul tryptophane. Nous comparons également l'échantillonnage conformationnel des trois formes de N_{TAIL}, concluant qu'il est le même pour leur partie commune.

Cependant, comme nous l'avons remarqué dans le manuscrit, la dynamique des trois formes de N_{TAIL} est différente. Nous présentons donc la comparaison de la dynamique des deux formes de N_{TAIL} ($N_{TAIL,L}$ et $N_{TAIL,M}$) à 298K, en utilisant l'analyse « model-free » avec trois contributions. L'imprécision des paramètres obtenus nous conduit à une réflexion sur la question de la « stabilisation » de l'analyse « model-free » en prenant en compte le fait que les mouvements lents impliquent plusieurs résidus à la fois dans un « segment ». Finalement, nous discutons les mouvements dans la partie hélicoïdale de N_{TAIL} qui sont plus lents que les mouvements contribuant à la relaxation RMN. Ces résultats sont également discutés dans le chapitre « Conclusions et perspectives » dont le résumé est présenté ci-dessous.

8.7 Conclusions et perspectives

Dans le chapitre 7, nous résumons notre travail et présentons les pistes potentielles pour les futures études.

Les protéines intrinsèquement désordonnées sont impliquées dans un bon nombre de processus cellulaires et représentent des cibles potentielles pour les traitements anticancéreux et contre les maladies neurodégénératives. Leur échantillonnage conformationnel a fait récemment l'objet d'études détaillées, mais on connaît moins bien leur dynamique.

L'une des meilleures approches pour étudier la dynamique dans les protéines est sans doute la relaxation RMN, qui permet d'accéder à une information à l'échelle atomique dans la dimension spatiale, et à l'échelle des picosecondes – nanosecondes dans sa dimension temporelle. En ce qui concerne les protéines repliées, la méthode dite « model-free », qui suppose la séparation de la réorientation globale d'une molécule et des mouvements internes, était fréquemment appliquée. Or, dans les protéines intrinsèquement désordonnées, le comportement sous-jacent est différent et défini par leurs propriétés en tant que polymères. Dans la physique des polymères, des modèles simples – tels que les modèles de Rouse ou de Zimm – prédisent la distribution des échelles de temps de mouvements (ou des 'modes' de relaxation) dont les mouvements lents dépendent de la longueur de la chaine, et des observations similaires ont été faites récemment dans les protéines intrinsèquement désordonnées.

Dans ce travail, nous étudions une protéine intrinsèquement désordonnée – N_{TAIL} – par la relaxation RMN ^{15}N. Cette protéine représente un intérêt dans le cadre de la réplication virale, mais surtout en tant que protéine modèle, étant à la fois largement désordonnée mais contenant une petite région hélicoïdale. Nous avons mesuré plus de 100 taux de relaxation à des températures différentes pour trois formes de N_{TAIL} de longueur différente, ce qui nous a permis de proposer une description physique des échelles de temps de mouvements observés dans le cadre d'analyse « model-free » étendu. Dans cette description, les mouvements les plus lents correspondent au comportement de la protéine en tant que chaîne polymère. Nous observons à 278K que ces mouvements dépendent de la longueur de la chaîne, étant plus rapides pour les chaînes plus courtes.

Bien évidemment, il existe des mouvements dans les protéines désordonnées qui ne contribuent pas à la relaxation RMN : dans la région hélicoïdale (MoRE) de N_{TAIL}, grâce à

la dispersion de relaxation ^{1}H et à la comparaison des comportements de taux de relaxation cross-corrélée et de couplages dipolaires résiduels, nous avons émis l'hypothèse que les trois conformations hélicoïdales, qui ont été décrites auparavant, s'échangent à l'échelle de temps entre des centaines de nanosecondes et 30 microsecondes.

Nous remarquons que notre analyse « model-free », même si elle a permis de décrire différents mouvements dans notre protéine, incluant les mouvements lents impliquant plusieurs résidus à la fois, a l'inconvénient de traiter chaque résidu séparément. Nous proposons donc une approche assez simple qui tient compte du fait que les temps de corrélation de mouvements lents ne présentent pas de variation importante entre résidus voisins – ce que nous observons d'ailleurs en pratique. Dans cette approche nous définissons une « fenêtre » ou un « segment » d'un nombre prédéfini de résidus, où ce temps de corrélation « lent » ne doit pas varier de façon importante. Avec un segment de 7 résidus les résultats sont statistiquement meilleurs qu'avec 3 ou 11 résidus, et nous permettent de mieux analyser l'influence de la longueur de la chaîne sur les mouvements lents à une température plus haute, 298K. Cependant, nous travaillons en ce moment sur une analyse plus complète qui définit des segments non-chevauchants tout au long de la séquence en utilisant un algorithme génétique. Idéalement, vu l'impact d'une paire de glycines voisines sur le taux de relaxation transversale dans N_{TAIL}, cette analyse doit inclure les propriétés physiques de chaque acide aminé, en en particulier son encombrement stérique.

Dans le futur, la description de la dynamique des protéines intrinsèquement désordonnées devrait lier la relaxation RMN, qui nous informe surtout sur le comportement local, à d'autres techniques plus sensibles aux interactions à longue distance, telles que l'augmentation paramagnétique de la relaxation (PRE) et le transfert d'énergie entre molécules fluorescentes (FRET). Nous plaçons ce travail dans le contexte d'une nouvelle ère dans l'étude des protéines intrinsèquement désordonnées où la relaxation RMN va contribuer à une approche intégrale pour étudier leur dynamique.

Bibliography

1. W. A. Smeaton, *Fourcroy - Chemist and Revolutionary* (W. Heffer & Sons, First Edition edition., 1962).
2. G. J. Mulder, Zusammensetzung von Fibrin, Albumin, Leimzucker, Leucin u. s. w. *Ann. Pharm.* **28**, 73–82 (1838).
3. R. Willstätter, Faraday lecture. Problems and methods in enzyme research. *J. Chem. Soc. Resumed*, 1359–1381 (1927).
4. N. H. Barton, D. E. G. Briggs, J. A. Eisen, D. B. Goldstein, N. H. Patel, *Evolution* (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y, 1st edition., 2007).
5. E. Fischer, Einfluss der Configuration auf die Wirkung der Enzyme. *Berichte Dtsch. Chem. Ges.* **27**, 2985–2993 (1894).
6. J. C. Kendrew *et al.*, A three-dimensional model of the myoglobin molecule obtained by x-ray analysis. *Nature*. **181**, 662–666 (1958).
7. D. E. Koshland, Application of a Theory of Enzyme Specificity to Protein Synthesis. *Proc. Natl. Acad. Sci. U. S. A.* **44**, 98–104 (1958).
8. P. E. Wright, H. J. Dyson, Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. *J. Mol. Biol.* **293**, 321–331 (1999).
9. R. Schneider *et al.*, Towards a robust description of intrinsic protein disorder using nuclear magnetic resonance spectroscopy. *Mol. Biosyst.* **8**, 58–68 (2011).
10. N. A. Farrow, O. Zhang, A. Szabo, D. A. Torchia, L. E. Kay, Spectral density function mapping using 15N relaxation data exclusively. *J. Biomol. NMR*. **6**, 153–162 (1995).
11. V. N. Uversky, Introduction to Intrinsically Disordered Proteins (IDPs). *Chem. Rev.* **114**, 6557–6560 (2014).
12. P. Romero *et al.*, Thousands of proteins likely to have long disordered regions. *Pac. Symp. Biocomput. Pac. Symp. Biocomput.*, 437–448 (1998).
13. A. K. Dunker, Z. Obradovic, P. Romero, E. C. Garner, C. J. Brown, Intrinsic protein disorder in complete genomes. *Genome Inform. Workshop Genome Inform.* **11**, 161–171 (2000).
14. A. K. Dunker, I. Silman, V. N. Uversky, J. L. Sussman, Function and structure of inherently disordered proteins. *Curr. Opin. Struct. Biol.* **18**, 756–764 (2008).
15. V. N. Uversky, A decade and a half of protein intrinsic disorder: biology still waits for physics. *Protein Sci. Publ. Protein Soc.* **22**, 693–724 (2013).
16. A. K. Dunker *et al.*, The unfoldomics decade: an update on intrinsically disordered proteins. *BMC Genomics*. **9 Suppl 2**, S1 (2008).
17. A. K. Dunker *et al.*, Intrinsically disordered protein. *J. Mol. Graph. Model.* **19**, 26–59 (2001).
18. P. Romero *et al.*, Sequence complexity of disordered protein. *Proteins*. **42**, 38–48 (2001).

19. S. Lise, D. T. Jones, Sequence patterns associated with disordered regions in proteins. *Proteins*. **58**, 144–150 (2005).

20. J. Prilusky *et al.*, FoldIndex©: a simple tool to predict whether a given protein sequence is intrinsically unfolded. *Bioinformatics*. **21**, 3435–3438 (2005).

21. K. Coeytaux, A. Poupon, Prediction of unfolded segments in a protein sequence based on amino acid composition. *Bioinformatics*. **21**, 1891–1900 (2005).

22. R. Linding, R. B. Russell, V. Neduva, T. J. Gibson, GlobPlot: exploring protein sequences for globularity and disorder. *Nucleic Acids Res.* **31**, 3701–3708 (2003).

23. J. Habchi, P. Tompa, S. Longhi, V. N. Uversky, Introducing protein intrinsic disorder. *Chem. Rev.* **114**, 6561–6588 (2014).

24. M. Sickmeier *et al.*, DisProt: the Database of Disordered Proteins. *Nucleic Acids Res.* **35**, D786–D793 (2007).

25. S. Fukuchi *et al.*, IDEAL: Intrinsically Disordered proteins with Extensive Annotations and Literature. *Nucleic Acids Res.* **40**, D507–D511 (2012).

26. T. D. Domenico, I. Walsh, A. J. M. Martin, S. C. E. Tosatto, MobiDB: a comprehensive database of intrinsic protein disorder annotations. *Bioinformatics*. **28**, 2080–2081 (2012).

27. R. Linding *et al.*, Protein Disorder Prediction: Implications for Structural Proteomics. *Structure*. **11**, 1453–1459 (2003).

28. J. J. Ward, L. J. McGuffin, K. Bryson, B. F. Buxton, D. T. Jones, The DISOPRED server for the prediction of protein disorder. *Bioinformatics*. **20**, 2138–2139 (2004).

29. Z. R. Yang, R. Thomson, P. McNeil, R. M. Esnouf, RONN: the bio-basis function neural network technique applied to the detection of natively disordered regions in proteins. *Bioinformatics*. **21**, 3369–3376 (2005).

30. S. O. Garbuzynskiy, M. Y. Lobanov, O. V. Galzitskaya, To be folded or to be unfolded? *Protein Sci. Publ. Protein Soc.* **13**, 2871–2877 (2004).

31. Z. Dosztányi, V. Csizmók, P. Tompa, I. Simon, The pairwise energy content estimated from amino acid composition discriminates between folded and intrinsically unstructured proteins. *J. Mol. Biol.* **347**, 827–839 (2005).

32. I. Callebaut *et al.*, Deciphering protein sequence information through hydrophobic cluster analysis (HCA): current status and perspectives. *Cell. Mol. Life Sci. CMLS*. **53**, 621–645 (1997).

33. B. Xue, R. L. Dunbrack, R. W. Williams, A. K. Dunker, V. N. Uversky, PONDR-FIT: a meta-predictor of intrinsically disordered amino acids. *Biochim. Biophys. Acta*. **1804**, 996–1010 (2010).

34. P. Lieutaud, B. Canard, S. Longhi, MeDor: a metaserver for predicting protein disorder. *BMC Genomics*. **9 Suppl 2**, S25 (2008).

35. J.-M. Chandonia, M. Karplus, New methods for accurate prediction of protein secondary structure. *Proteins Struct. Funct. Bioinforma.* **35**, 293–306 (1999).

36. G. D. Fasman, Ed., *Circular Dichroism and the Conformational Analysis of Biomolecules* (Springer US, Boston, MA, 1996; http://link.springer.com/10.1007/978-1-4757-2508-7).

37. J. Yao, H. J. Dyson, P. E. Wright, Chemical shift dispersion and secondary structure prediction in unfolded and partly folded proteins. *FEBS Lett.* **419**, 285–289 (1997).

38. J. Cavanagh, W. J. Fairbrother, A. G. Palmer III, N. J. Skelton, *Protein NMR spectroscopy: principles and practice* (Academic Press, 1995).

39. A. Meola *et al.*, Robust and low cost uniform 15N-labeling of proteins expressed in Drosophila S2 cells and Spodoptera frugiperda Sf9 cells for NMR applications. *J. Struct. Biol.* **188**, 71–78 (2014).

40. L. Skora, B. Shrestha, A. D. Gossert, Isotope Labeling of Proteins in Insect Cells. *Methods Enzymol.* **565**, 245–288 (2015).

41. S. Schwarzinger *et al.*, Sequence-Dependent Correction of Random Coil NMR Chemical Shifts. *J. Am. Chem. Soc.* **123**, 2970–2978 (2001).

42. J. A. Marsh, V. K. Singh, Z. Jia, J. D. Forman-Kay, Sensitivity of secondary structure propensities to sequence differences between α- and γ-synuclein: Implications for fibrillation. *Protein Sci.* **15**, 2795–2804 (2006).

43. A. G. Palmer III, C. D. Kroenke, J. Patrick Loria, in *Methods in Enzymology*, V. D. and U. S. Thomas L. James, Ed. (Academic Press, 2001; http://www.sciencedirect.com/science/article/pii/S0076687901393151), vol. 339 of *Nuclear Magnetic Resonance of Biological Macromolecules - Part B*, pp. 204–238.

44. C. Eichmüller, N. R. Skrynnikov, A new amide proton R1rho experiment permits accurate characterization of microsecond time-scale conformational exchange. *J. Biomol. NMR*. **32**, 281–293 (2005).

45. J. L. Battiste, G. Wagner, Utilization of Site-Directed Spin Labeling and High-Resolution Heteronuclear Nuclear Magnetic Resonance for Global Fold Determination of Large Proteins with Limited Nuclear Overhauser Effect Data†. *Biochemistry (Mosc.)*. **39**, 5355–5365 (2000).

46. V. Gaponenko *et al.*, Protein global fold determination using site-directed spin and isotope labeling. *Protein Sci.* **9**, 302–309 (2000).

47. J. R. Gillespie, D. Shortle, Characterization of long-range structure in the denatured state of staphylococcal nuclease. I. paramagnetic relaxation enhancement by nitroxide spin labels1. *J. Mol. Biol.* **268**, 158–169 (1997).

48. P. A. Kosen, B.-M. in Enzymology, Ed. (Academic Press, 1989; http://www.sciencedirect.com/science/article/pii/0076687989770075), vol. 177 of *Nuclear Magnetic Resonance Part B Structure and Mechanism*, pp. 86–121.

49. M. R. Jensen *et al.*, Quantitative Determination of the Conformational Properties of Partially Folded and Intrinsically Disordered Proteins Using NMR Dipolar Couplings. *Structure*. **17**, 1169–1185 (2009).

50. K. Wüthrich, *NMR of Proteins and Nucleic Acids* (Wiley, 1986).

51. S. Macura, R. R. Ernst, Elucidation of cross relaxation in liquids by two-dimensional N.M.R. spectroscopy. *Mol. Phys.* **41**, 95–117 (1980).

52. M. Rückert, G. Otting, Alignment of Biological Macromolecules in Novel Nonionic Liquid Crystalline Media for NMR Experiments. *J. Am. Chem. Soc.* **122**, 7793–7797 (2000).

53. N. Tjandra, A. Bax, Direct Measurement of Distances and Angles in Biomolecules by NMR in a Dilute Liquid Crystalline Medium. *Science*. **278**, 1111–1114 (1997).

54. G. M. Clore, M. R. Starich, A. M. Gronenborn, 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 (1998).

55. H.-J. Sass, G. Musco, S. J. Stahl, P. T. Wingfield, S. Grzesiek, 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 (2000).

56. R. Tycko, F. J. Blanco, Y. Ishii, 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 (2000).

57. D. Shortle, M. S. Ackerman, Persistence of Native-Like Topology in a Denatured Protein in 8 M Urea. *Science*. **293**, 487–489 (2001).

58. M. Louhivuori *et al.*, On the Origin of Residual Dipolar Couplings from Denatured Proteins. *J. Am. Chem. Soc.* **125**, 15647–15650 (2003).

59. P. Bernadó *et al.*, 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 (2005).

60. V. Ozenne *et al.*, Flexible-meccano: a tool for the generation of explicit ensemble descriptions of intrinsically disordered proteins and their associated experimental observables. *Bioinformatics*. **28**, 1463–1470 (2012).

61. G. Nodet *et al.*, 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 (2009).

62. M. R. Jensen *et al.*, Intrinsic disorder in measles virus nucleocapsids. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 9839–9844 (2011).

63. D. I. Svergun, Restoring Low Resolution Structure of Biological Macromolecules from Solution Scattering Using Simulated Annealing. *Biophys. J.* **76**, 2879–2886 (1999).

64. D. Franke, D. I. Svergun, DAMMIF, a program for rapid ab-initio shape determination in small-angle scattering. *J. Appl. Crystallogr.* **42**, 342–346 (2009).

65. P. Bernadó, D. I. Svergun, Structural analysis of intrinsically disordered proteins by small-angle X-ray scattering. *Mol. Biosyst.* **8**, 151–167 (2012).

66. P. Bernadó, E. Mylonas, M. V. Petoukhov, M. Blackledge, D. I. Svergun, Structural characterization of flexible proteins using small-angle X-ray scattering. *J. Am. Chem. Soc.* **129**, 5656–5664 (2007).

67. E. Sisamakis, A. Valeri, S. Kalinin, P. J. Rothwell, C. A. M. Seidel, in *Methods in Enzymology*, N. G. Walter, Ed. (Academic Press, 2010; http://www.sciencedirect.com/science/article/pii/S0076687910750187), vol. 475 of *Single Molecule Tools, Part B:Super-Resolution, Particle Tracking, Multiparameter, and Force Based Methods*, pp. 455–514.

68. T. Ando, N. Kodera, in *Intrinsically Disordered Protein Analysis* (Springer, 2012; http://link.springer.com/protocol/10.1007/978-1-4614-3704-8_4), pp. 57–69.

69. A. Miyagi *et al.*, Visualization of Intrinsically Disordered Regions of Proteins by High-Speed Atomic Force Microscopy. *ChemPhysChem*. **9**, 1859–1866 (2008).

70. M. Sandal *et al.*, Conformational Equilibria in Monomeric α-Synuclein at the Single-Molecule Level. *PLoS Biol*. **6**, e6 (2008).

71. R. J. Ellis, Macromolecular crowding: obvious but underappreciated. *Trends Biochem. Sci.* **26**, 597–604 (2001).

72. S. B. Zimmerman, S. O. Trach, Estimation of macromolecule concentrations and excluded volume effects for the cytoplasm of Escherichia coli. *J. Mol. Biol.* **222**, 599–620 (1991).

73. E. A. Cino, M. Karttunen, W.-Y. Choy, Effects of molecular crowding on the dynamics of intrinsically disordered proteins. *PloS One*. **7**, e49876 (2012).

74. N. Kozer, G. Schreiber, Effect of Crowding on Protein–Protein Association Rates: Fundamental Differences between Low and High Mass Crowding Agents. *J. Mol. Biol.* **336**, 763–774 (2004).

75. M. Feig, Y. Sugita, Variable Interactions between Protein Crowders and Biomolecular Solutes Are Important in Understanding Cellular Crowding. *J. Phys. Chem. B*. **116**, 599–605 (2012).

76. S. K. Mukherjee, S. Gautam, S. Biswas, J. Kundu, P. K. Chowdhury, Do Macromolecular Crowding Agents Exert Only an Excluded Volume Effect? A Protein Solvation Study. *J. Phys. Chem. B*. **119**, 14145–14156 (2015).

77. D. W. Bolen, I. V. Baskakov, The osmophobic effect: natural selection of a thermodynamic force in protein folding. *J. Mol. Biol.* **310**, 955–963 (2001).

78. V. Receveur-Bréchot, J.-M. Bourhis, V. N. Uversky, B. Canard, S. Longhi, Assessing protein disorder and induced folding. *Proteins*. **62**, 24–45 (2006).

79. S. L. Flaugh, K. J. Lumb, Effects of Macromolecular Crowding on the Intrinsically Disordered Proteins c-Fos and p27Kip1. *Biomacromolecules*. **2**, 538–540 (2001).

80. P. Tompa, A. Fersht, *Structure and function of intrinsically disordered proteins* (CRC Press, 2009).

81. A. Soranno *et al.*, Single-molecule spectroscopy reveals polymer effects of disordered proteins in crowded environments. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 4874–4879 (2014).

82. G. Asher, P. Tsvetkov, C. Kahana, Y. Shaul, A mechanism of ubiquitin-independent proteasomal degradation of the tumor suppressors p53 and p73. *Genes Dev.* **19**, 316–321 (2005).

83. X. Chen, L. F. Barton, Y. Chi, B. E. Clurman, J. M. Roberts, Ubiquitin-Independent Degradation of Cell-Cycle Inhibitors by the REGγ Proteasome. *Mol. Cell*. **26**, 843–852 (2007).

84. F.-X. Theillet *et al.*, Physicochemical Properties of Cells and Their Effects on Intrinsically Disordered Proteins (IDPs). *Chem. Rev.* **114**, 6661–6714 (2014).

85. T. Takeuchi *et al.*, Direct and Rapid Cytosolic Delivery Using Cell-Penetrating Peptides Mediated by Pyrenebutyrate. *ACS Chem. Biol.* **1**, 299–303 (2006).

86. F.-X. Theillet *et al.*, Structural disorder of monomeric α-synuclein persists in mammalian cells. *Nature*. **530**, 45–50 (2016).

87. A. Binolfi *et al.*, Intracellular repair of oxidation-damaged α-synuclein fails to target C-terminal modification sites. *Nat. Commun.* **7**, 10251 (2016).

88. K. Inomata *et al.*, High-resolution multi-dimensional NMR spectroscopy of proteins in human cells. *Nature*. **458**, 106–109 (2009).

89. J.-F. Bodart *et al.*, NMR observation of Tau in Xenopus oocytes. *J. Magn. Reson. San Diego Calif 1997*. **192**, 252–257 (2008).

90. A. Binolfi, F.-X. Theillet, P. Selenko, Bacterial in-cell NMR of human α-synuclein: a disordered monomer by nature? *Biochem. Soc. Trans.* **40**, 950–954 (2012).

91. C. A. Waudby *et al.*, In-cell NMR characterization of the secondary structure populations of a disordered conformation of α-synuclein within E. coli cells. *PloS One*. **8**, e72286 (2013).

92. P. Tompa, Intrinsically unstructured proteins. *Trends Biochem. Sci.* **27**, 527–533 (2002).

93. K. Sugase, H. J. Dyson, P. E. Wright, Mechanism of coupled folding and binding of an intrinsically disordered protein. *Nature*. **447**, 1021–1025 (2007).

94. H. J. Dyson, P. E. Wright, Intrinsically unstructured proteins and their functions. *Nat. Rev. Mol. Cell Biol.* **6**, 197–208 (2005).

95. H. J. Dyson, P. E. Wright, Coupling of folding and binding for unstructured proteins. *Curr. Opin. Struct. Biol.* **12**, 54–60 (2002).

96. P. Tompa, N. E. Davey, T. J. Gibson, M. M. Babu, A million peptide motifs for the molecular biologist. *Mol. Cell*. **55**, 161–169 (2014).

97. C. J. Oldfield *et al.*, Coupled folding and binding with alpha-helix-forming molecular recognition elements. *Biochemistry (Mosc.)*. **44**, 12454–12470 (2005).

98. V. Vacic *et al.*, Characterization of Molecular Recognition Features, MoRFs, and Their Binding Partners. *J. Proteome Res.* **6**, 2351–2366 (2007).

99. V. Neduva, R. B. Russell, Linear motifs: evolutionary interaction switches. *FEBS Lett.* **579**, 3342–3345 (2005).

100. P. Puntervoll *et al.*, ELM server: a new resource for investigating short functional sites in modular eukaryotic proteins. *Nucleic Acids Res.* **31**, 3625–3630 (2003).

101. K. Van Roey *et al.*, Short Linear Motifs: Ubiquitous and Functionally Diverse Protein Interaction Modules Directing Cell Regulation. *Chem. Rev.* **114**, 6733–6778 (2014).

102. M. Fuxreiter, I. Simon, P. Friedrich, P. Tompa, Preformed structural elements feature in partner recognition by intrinsically unstructured proteins. *J. Mol. Biol.* **338**, 1015–1026 (2004).

103. J. M. Rogers, C. T. Wong, J. Clarke, Coupled folding and binding of the disordered protein PUMA does not require particular residual structure. *J. Am. Chem. Soc.* **136**, 5197–5200 (2014).

104. J. Dogan, T. Schmidt, X. Mu, Å. Engström, P. Jemth, Fast association and slow transitions in the interaction between two intrinsically disordered protein domains. *J. Biol. Chem.* **287**, 34316–34324 (2012).

105. V. Iešmantavičius, J. Dogan, P. Jemth, K. Teilum, M. Kjaergaard, Helical propensity in an intrinsically disordered protein accelerates ligand binding. *Angew. Chem. Int. Ed Engl.* **53**, 1548–1551 (2014).

106. P. E. Wright, H. J. Dyson, Intrinsically disordered proteins in cellular signalling and regulation. *Nat. Rev. Mol. Cell Biol.* **16**, 18–29 (2015).

107. P. Tompa, M. Fuxreiter, Fuzzy complexes: polymorphism and structural disorder in protein-protein interactions. *Trends Biochem. Sci.* **33**, 2–8 (2008).

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

109. L. M. Espinoza-Fonseca, Reconciling binding mechanisms of intrinsically disordered proteins. *Biochem. Biophys. Res. Commun.* **382**, 479–482 (2009).

110. R. Schneider *et al.*, Visualizing the Molecular Recognition Trajectory of an Intrinsically Disordered Protein Using Multinuclear Relaxation Dispersion NMR. *J. Am. Chem. Soc.* **137**, 1220–1229 (2015).

111. M. R. Jensen *et al.*, 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 (2008).

112. S. Gianni, J. Dogan, P. Jemth, Coupled binding and folding of intrinsically disordered proteins: what can we learn from kinetics? *Curr. Opin. Struct. Biol.* **36**, 18–24 (2016).

113. E. B. Gibbs, S. A. Showalter, Quantitative biophysical characterization of intrinsically disordered proteins. *Biochemistry (Mosc.)*. **54**, 1314–1326 (2015).

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

115. N. Greives, H.-X. Zhou, Both protein dynamics and ligand concentration can shift the binding mechanism between conformational selection and induced fit. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 10197–10202 (2014).

116. M. Fuxreiter, Fuzziness: linking regulation to protein dynamics. *Mol. Biosyst.* **8**, 168–177 (2012).

117. V. N. Uversky, C. J. Oldfield, A. K. Dunker, Intrinsically disordered proteins in human diseases: introducing the D2 concept. *Annu. Rev. Biophys.* **37**, 215–246 (2008).

118. K. H. Vousden, X. Lu, Live or let die: the cell's response to p53. *Nat. Rev. Cancer*. **2**, 594–604 (2002).

119. N. Rezaei-Ghaleh, M. Blackledge, M. Zweckstetter, Intrinsically disordered proteins: from sequence and conformational properties toward drug discovery. *Chembiochem Eur. J. Chem. Biol.* **13**, 930–950 (2012).

120. Z. Jaunmuktane *et al.*, Evidence for human transmission of amyloid-β pathology and cerebral amyloid angiopathy. *Nature*. **525**, 247–250 (2015).

121. L. T. Vassilev, Small-Molecule Antagonists of p53-MDM2 Binding: Research Tools and Potential Therapeutics. *Cell Cycle*. **3**, 417–419 (2004).

122. M. Wells *et al.*, Structure of tumor suppressor p53 and its intrinsically disordered N-terminal transactivation domain. *Proc. Natl. Acad. Sci.* **105**, 5762–5767 (2008).

123. L. T. Vassilev *et al.*, In vivo activation of the p53 pathway by small-molecule antagonists of MDM2. *Science*. **303**, 844–848 (2004).

124. D. I. Hammoudeh, A. V. Follis, E. V. Prochownik, S. J. Metallo, Multiple Independent Binding Sites for Small-Molecule Inhibitors on the Oncoprotein c-Myc. *J. Am. Chem. Soc.* **131**, 7390–7401 (2009).

125. H. V. Erkizan *et al.*, A small molecule blocking oncogenic protein EWS-FLI1 interaction with RNA helicase A inhibits growth of Ewing's sarcoma. *Nat. Med.* **15**, 750–756 (2009).

126. S. J. Metallo, Intrinsically disordered proteins are potential drug targets. *Curr. Opin. Chem. Biol.* **14**, 481–488 (2010).

127. T. L. Kukar *et al.*, Substrate-targeting γ-secretase modulators. *Nature*. **453**, 925–929 (2008).

128. G. R. Lamberto *et al.*, Structural and mechanistic basis behind the inhibitory interaction of PcTS on alpha-synuclein amyloid fibril formation. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 21057–21062 (2009).

129. L.-M. Yan, M. Tatarek-Nossol, A. Velkova, A. Kazantzis, A. Kapurniotu, Design of a mimic of nonamyloidogenic and bioactive human islet amyloid polypeptide (IAPP) as nanomolar affinity inhibitor of IAPP cytotoxic fibrillogenesis. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 2046–2051 (2006).

130. L.-M. Yan, A. Velkova, M. Tatarek-Nossol, E. Andreetto, A. Kapurniotu, IAPP mimic blocks Abeta cytotoxic self-assembly: cross-suppression of amyloid toxicity of Abeta and IAPP suggests a molecular link between Alzheimer's disease and type II diabetes. *Angew. Chem. Int. Ed Engl.* **46**, 1246–1252 (2007).

131. G. K.-M. Goh, A. K. Dunker, V. N. Uversky, Protein intrinsic disorder toolbox for comparative analysis of viral proteins. *BMC Genomics*. **9**, 1–14 (2008).

132. B. Xue *et al.*, Viral Disorder or Disordered Viruses: Do Viral Proteins Possess Unique Features? *Protein Pept. Lett.* **17**, 932–951 (2010).

133. C. Alves, C. Cunha, Order and disorder in viral proteins: new insights into an old paradigm. *Future Virol.* **7**, 1183–1191 (2012).

134. R. Pushker, C. Mooney, N. E. Davey, J.-M. Jacqué, D. C. Shields, Marked Variability in the Extent of Protein Disorder within and between Viral Families. *PLoS ONE*. **8**, e60724 (2013).

135. N. Tokuriki, C. J. Oldfield, V. N. Uversky, I. N. Berezovsky, D. S. Tawfik, Do viral proteins possess unique biophysical features? *Trends Biochem. Sci.* **34**, 53–59 (2009).

136. G. Communie, R. W. Ruigrok, M. R. Jensen, M. Blackledge, Intrinsically disordered proteins implicated in paramyxoviral replication machinery. *Curr. Opin. Virol.* **5**, 72–81 (2014).

137. J. Curran, D. Kolakofsky, Replication of paramyxoviruses. *Adv. Virus Res.* **54**, 403–422 (1998).

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

139. L. Blanchard *et al.*, Structure and dynamics of the nucleocapsid-binding domain of the Sendai virus phosphoprotein in solution. *Virology*. **319**, 201–211 (2004).

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

141. K. Johansson *et al.*, 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 (2003).

142. G. Communie *et al.*, Atomic Resolution Description of the Interaction between the Nucleoprotein and Phosphoprotein of Hendra Virus. *PLoS Pathog*. **9**, e1003631 (2013).

143. R. Kimmich, *NMR: tomography, diffusometry, relaxometry* (Springer Science & Business Media, 2012).

144. P. S. Hubbard, Some Properties of Correlation Functions of Irreducible Tensor Operators. *Phys. Rev.* **180**, 319–326 (1969).

145. L. G. Werbelow, D. M. Grant, Intramolecular dipolar relaxation in multispin systems. *Adv Magn Reson*. **9**, 34 (1977).

146. G. Lipari, A. Szabo, 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 (1982).

147. G. Lipari, A. Szabo, 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 (1982).

148. M. Goldman, Interference effects in the relaxation of a pair of unlike spin-12 nuclei. *J. Magn. Reson. 1969*. **60**, 437–452 (1984).

149. D. E. Woessner, Nuclear Transfer Effects in Nuclear Magnetic Resonance Pulse Experiments. *J. Chem. Phys.* **35**, 41–48 (1961).

150. J. S. Leigh, Relaxation times in systems with chemical exchange: Some exact solutions. *J. Magn. Reson. 1969*. **4**, 308–311 (1971).

151. C. D. Kroenke, M. Rance, A. G. Palmer, Variability of the 15N Chemical Shift Anisotropy in Escherichia coli Ribonuclease H in Solution. *J. Am. Chem. Soc.* **121**, 10119–10125 (1999).

152. R. Kimmich, N. Fatkullin, in *NMR • 3D Analysis • Photopolymerization* (Springer Berlin Heidelberg, 2004;

http://link.springer.com.gate1.inist.fr/chapter/10.1007/978-3-540-40000-4_1), *Apvances in Polymer Science*, pp. 1–113.

153. M. R. Jensen, M. Zweckstetter, J. Huang, M. Blackledge, Exploring free-energy landscapes of intrinsically disordered proteins at atomic resolution using NMR spectroscopy. *Chem. Rev.* **114**, 6632–6660 (2014).

154. G. R. Strobl, *The Physics of Polymers: Concepts for Understanding Their Structures and Behaviour* (Springer Science & Business Media, 1997).

155. M. Doi, S. F. Edwards, *The Theory of Polymer Dynamics* (Clarendon Press, 1988).

156. Autobiography of Attila Szabo. *J. Phys. Chem. B*. **112**, 5883–5886 (2008).

157. B. Halle, The physical basis of model-free analysis of NMR relaxation data from proteins and complex fluids. *J. Chem. Phys.* **131**, 224507 (2009).

158. G. Lipari, A. Szabo, Effect of librational motion on fluorescence depolarization and nuclear magnetic resonance relaxation in macromolecules and membranes. *Biophys. J.* **30**, 489–506 (1980).

159. K. Kinosita, S. Kawato, A. Ikegami, A theory of fluorescence polarization decay in membranes. *Biophys. J.* **20**, 289–305 (1977).

160. R. King, O. Jardetzky, A general formalism for the analysis of NMR relaxation measurements on systems with multiple degrees of freedom. *Chem. Phys. Lett.* **55**, 15–18 (1978).

161. R. King *et al.*, Magnetic relaxation analysis of dynamic processes in macromolecules in the pico- to microsecond range. *Biophys. J.* **24**, 103–117 (1978).

162. A. A. Ribeiro, R. King, C. Restivo, O. Jardetzky, An approach to the mapping of internal motions in proteins. Analysis of carbon-13 NMR relaxation in the bovine pancreatic trypsin inhibitor. *J. Am. Chem. Soc.* **102**, 4040–4051 (1980).

163. G. M. Clore *et al.*, Deviations from the simple two-parameter model-free approach to the interpretation of nitrogen-15 nuclear magnetic relaxation of proteins. *J. Am. Chem. Soc.* **112**, 4989–4991 (1990).

164. J. W. Peng, G. Wagner, Mapping of spectral density functions using heteronuclear NMR relaxation measurements. *J. Magn. Reson. 1969*. **98**, 308–332 (1992).

165. N. A. Farrow, O. Zhang, J. D. Forman-Kay, L. E. Kay, Comparison of the Backbone Dynamics of a Folded and an Unfolded SH3 Domain Existing in Equilibrium in Aqueous Buffer. *Biochemistry (Mosc.)*. **34**, 868–878 (1995).

166. A. T. Alexandrescu, D. Shortlet, Backbone Dynamics of a Highly Disordered 131 Residue Fragment of Staphylococcal Nuclease. *J. Mol. Biol.* **242**, 527–546 (1994).

167. A. Allerhand, R. K. Hailstone, Carbon-13 Fourier Transform Nuclear Magnetic Resonance. X. Effect of Molecular Weight on 13C Spin-Lattice Relaxation Times of Polystyrene in Solution. *J. Chem. Phys.* **56**, 3718–3720 (1972).

168. M. K. Frank, G. M. Clore, A. M. Gronenborn, Structural and dynamic characterization of the urea denatured state of the immunoglobulin binding domain of streptococcal protein G by multidimensional heteronuclear NMR spectroscopy. *Protein Sci. Publ. Protein Soc.* **4**, 2605–2615 (1995).

169. B. Brutscher, R. Brüschweiler, R. R. Ernst, Backbone Dynamics and Structural Characterization of the Partially Folded A State of Ubiquitin by 1H, 13C, and 15N Nuclear Magnetic Resonance Spectroscopy. *Biochemistry (Mosc.)*. **36**, 13043–13053 (1997).

170. M. Buck, H. Schwalbe, C. M. Dobson, Main-chain Dynamics of a Partially Folded Protein:15N NMR Relaxation Measurements of Hen Egg White Lysozyme Denatured in Trifluoroethanol. *J. Mol. Biol.* **257**, 669–683 (1996).

171. D. Yang, Y.-K. Mok, J. D. Forman-Kay, N. A. Farrow, L. E. Kay, Contributions to protein entropy and heat capacity from bond vector motions measured by NMR spin relaxation1. *J. Mol. Biol.* **272**, 790–804 (1997).

172. D. Yang, L. E. Kay, Contributions to Conformational Entropy Arising from Bond Vector Fluctuations Measured from NMR-Derived Order Parameters: Application to Protein Folding. *J. Mol. Biol.* **263**, 369–382 (1996).

173. D. Canet, P. Barthe, P. Mutzenhardt, C. Roumestand, A Comprehensive Analysis of Multifield 15N Relaxation Parameters in Proteins: Determination of 15N Chemical Shift Anisotropies. *J. Am. Chem. Soc.* **123**, 4567–4576 (2001).

174. F. Ochsenbein *et al.*, 15N NMR relaxation as a probe for helical intrinsic propensity: The case of the unfolded D2 domain of annexin I. *J. Biomol. NMR*. **19**, 3–18 (2001).

175. I. Furó, NMR spectroscopy of micelles and related systems. *J. Mol. Liq.* **117**, 117–137 (2005).

176. J. J. Prompers, C. Scheurer, R. Brüschweiler, Characterization of NMR relaxation-active motions of a partially folded A-state analogue of ubiquitin1. *J. Mol. Biol.* **305**, 1085–1097 (2001).

177. J. J. Prompers, R. Brüschweiler, Reorientational Eigenmode Dynamics: A Combined MD/NMR Relaxation Analysis Method for Flexible Parts in Globular Proteins. *J. Am. Chem. Soc.* **123**, 7305–7313 (2001).

178. J. J. Prompers, R. Brüschweiler, General Framework for Studying the Dynamics of Folded and Nonfolded Proteins by NMR Relaxation Spectroscopy and MD Simulation. *J. Am. Chem. Soc.* **124**, 4522–4534 (2002).

179. A. Caflisch, M. Karplus, Structural details of urea binding to barnase: a molecular dynamics analysis. *Structure*. **7**, 477-S2 (1999).

180. R. B. Best, N.-V. Buchete, G. Hummer, Are Current Molecular Dynamics Force Fields too Helical? *Biophys. J.* **95**, L07–L09 (2008).

181. Y. Xue, N. R. Skrynnikov, Motion of a Disordered Polypeptide Chain as Studied by Paramagnetic Relaxation Enhancements, 15N Relaxation, and Molecular Dynamics Simulations: How Fast Is Segmental Diffusion in Denatured Ubiquitin? *J. Am. Chem. Soc.* **133**, 14614–14628 (2011).

182. K. Lindorff-Larsen, N. Trbovic, P. Maragakis, S. Piana, D. E. Shaw, Structure and Dynamics of an Unfolded Protein Examined by Molecular Dynamics Simulation. *J. Am. Chem. Soc.* **134**, 3787–3791 (2012).

183. D. E. Shaw *et al.*, in *Proceedings of the Conference on High Performance Computing Networking, Storage and Analysis* (2009), pp. 1–11.

184. F. A. Bovey, P. A. Mirau, *NMR of Polymers* (Academic Press, 1996).

185. A. V. Buevich, J. Baum, Dynamics of Unfolded Proteins: Incorporation of Distributions of Correlation Times in the Model Free Analysis of NMR Relaxation Data. *J. Am. Chem. Soc.* **121**, 8671–8672 (1999).

186. A. V. Buevich, U. P. Shinde, M. Inouye, J. Baum, Backbone dynamics of the natively unfolded pro-peptide of subtilisin by heteronuclear NMR relaxation studies. *J. Biomol. NMR*. **20**, 233–249 (2001).

187. F. Ochsenbein, J.-M. Neumann, E. Guittet, C. Van Heijenoort, Dynamical characterization of residual and non-native structures in a partially folded protein by 15N NMR relaxation using a model based on a distribution of correlation times. *Protein Sci.* **11**, 957–964 (2002).

188. K. Modig, F. M. Poulsen, Model-independent interpretation of NMR relaxation data for unfolded proteins: the acid-denatured state of ACBP. *J. Biomol. NMR*. **42**, 163–177 (2008).

189. S. N. Khan *et al.*, Distribution of Pico- and Nanosecond Motions in Disordered Proteins from Nuclear Spin Relaxation. *Biophys. J.* **109**, 988–999 (2015).

190. M. L. Gill, R. A. Byrd, I. I. I. Arthur G. Palmer, Dynamics of GCN4 facilitate DNA interaction: a model-free analysis of an intrinsically disordered region. *Phys. Chem. Chem. Phys.* **18**, 5839–5849 (2016).

191. H. Schwalbe *et al.*, Structural and Dynamical Properties of a Denatured Protein. Heteronuclear 3D NMR Experiments and Theoretical Simulations of Lysozyme in 8 M Urea. *Biochemistry (Mosc.)*. **36**, 8977–8991 (1997).

192. J. Wirmer, W. Peti, H. Schwalbe, Motional properties of unfolded ubiquitin: a model for a random coil protein. *J. Biomol. NMR*. **35**, 175–186 (2006).

193. J. Klein-Seetharaman *et al.*, Long-Range Interactions Within a Nonnative Protein. *Science*. **295**, 1719–1722 (2002).

194. J. Wirmer *et al.*, Modulation of Compactness and Long-Range Interactions of Unfolded Lysozyme by Single Point Mutations. *Angew. Chem.* **116**, 5904–5909 (2004).

195. G. Parigi *et al.*, Long-Range Correlated Dynamics in Intrinsically Disordered Proteins. *J. Am. Chem. Soc.* **136**, 16201–16209 (2014).

196. J. H. Viles *et al.*, Local Structural Plasticity of the Prion Protein. Analysis of NMR Relaxation Dynamics†. *Biochemistry (Mosc.)*. **40**, 2743–2753 (2001).

197. A. T. Alexandrescu, C. Abeygunawardana, D. Shortle, Structure and Dynamics of a Denatured 131-Residue Fragment of Staphylococcal Nuclease: A Heteronuclear NMR Study. *Biochemistry (Mosc.)*. **33**, 1063–1072 (1994).

198. J. Wirmer, H. Berk, R. Ugolini, C. Redfield, H. Schwalbe, Characterization of the unfolded state of bovine α-lactalbumin and comparison with unfolded states of homologous proteins. *Protein Sci.* **15**, 1397–1407 (2006).

199. K. Houben, D. Marion, N. Tarbouriech, R. W. H. Ruigrok, L. Blanchard, 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 (2007).

200. E. Lescop, P. Schanda, B. Brutscher, A set of BEST triple-resonance experiments for time-optimized protein resonance assignment. *J. Magn. Reson.* **187**, 163–169 (2007).

201. F. Delaglio *et al.*, NMRPipe: A multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR*. **6**, 277–293 (1995).

202. T. D. Goddard, D. G. Kneller, *San Francisco: University of California* (SPARKY, 3).

203. W. F. Vranken *et al.*, The CCPN data model for NMR spectroscopy: Development of a software pipeline. *Proteins Struct. Funct. Bioinforma.* **59**, 687–696 (2005).

204. Y.-S. Jung, M. Zweckstetter, Mars - robust automatic backbone assignment of proteins. *J. Biomol. NMR*. **30**, 11–23 (2004).

205. N.-A. Lakomek, J. Ying, A. Bax, Measurement of 15N relaxation rates in perdeuterated proteins by TROSY-based methods. *J. Biomol. NMR*. **53**, 209–221 (2012).

206. L. Kay, P. Keifer, T. Saarinen, Pure absorption gradient enhanced heteronuclear single quantum correlation spectroscopy with improved sensitivity. *J. Am. Chem. Soc.* **114**, 10663–10665 (1992).

207. A. C. Wang, A. Bax, Minimizing the effects of radio-frequency heating in multidimensional NMR experiments. *J. Biomol. NMR*. **3**, 715–720 (1993).

208. F. Ferrage, A. Reichel, S. Battacharya, D. Cowburn, R. Ghose, On the measurement of 15N–{1H} nuclear Overhauser effects. 2. Effects of the saturation scheme and water signal suppression. *J. Magn. Reson.* **207**, 294–303 (2010).

209. F. Ferrage, in *Protein NMR Techniques* (Springer, 2012; http://link.springer.com/protocol/10.1007/978-1-61779-480-3_9), pp. 141–163.

210. P. Pelupessy, G. M. Espallargas, G. Bodenhausen, Symmetrical reconversion: measuring cross-correlation rates with enhanced accuracy. *J. Magn. Reson.* **161**, 258–264 (2003).

211. P. Pelupessy, F. Ferrage, G. Bodenhausen, Accurate measurement of longitudinal cross-relaxation rates in nuclear magnetic resonance. *J. Chem. Phys.* **126**, 134508 (2007).

212. N. A. Farrow *et al.*, Backbone Dynamics of a Free and a Phosphopeptide-Complexed Src Homology 2 Domain Studied by 15N NMR Relaxation. *Biochemistry (Mosc.)*. **33**, 5984–6003 (1994).

213. J. J. Skalicky, D. K. Sukumaran, J. L. Mills, T. Szyperski, Toward Structural Biology in Supercooled Water. *J. Am. Chem. Soc.* **122**, 3230–3231 (2000).

214. B. Efron, R. Tibshirani, Bootstrap Methods for Standard Errors, Confidence Intervals, and Other Measures of Statistical Accuracy. *Stat. Sci.* **1**, 54–75 (1986).

215. M. Findeisen, T. Brand, S. Berger, A 1H-NMR thermometer suitable for cryoprobes. *Magn. Reson. Chem.* **45**, 175–178 (2007).

216. D. Ban *et al.*, Exceeding the limit of dynamics studies on biomolecules using high spin-lock field strengths with a cryogenically cooled probehead. *J. Magn. Reson.* **221**, 1–4 (2012).

217. A. S. Maltsev, J. Ying, A. Bax, Deuterium isotope shifts for backbone 1H, 15N and 13C nuclei in intrinsically disordered protein α-synuclein. *J. Biomol. NMR*. **54**, 181–191 (2012).

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

219. T. Cierpicki, J. Otlewski, Amide proton temperature coefficients as hydrogen bond indicators in proteins. *J. Biomol. NMR*. **21**, 249–261 (2001).

220. T. Cierpicki, I. Zhukov, R. A. Byrd, J. Otlewski, Hydrogen Bonds in Human Ubiquitin Reflected in Temperature Coefficients of Amide Protons. *J. Magn. Reson.* **157**, 178–180 (2002).

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

222. E. de Alba, J. L. Baber, N. Tjandra, The Use of Residual Dipolar Coupling in Concert with Backbone Relaxation Rates to Identify Conformational Exchange by NMR. *J. Am. Chem. Soc.* **121**, 4282–4283 (1999).

Abstract

Intrinsically disordered proteins (IDPs) are highly flexible heteropolymers, implicated in important cellular activities (signal transduction, molecular recognition, transcription, translation, etc.) and representing potential drug targets against cancer and neurodegenerative diseases, whose dynamic modes define their biological function. Although the conformational states sampled by IDPs are relatively well understood, essentially nothing is known about the associated dynamic timescales. In this study we investigate the conformational behavior of the intrinsically disordered C-terminal domain of the nucleoprotein of Sendai virus (N_{TAIL}), which interacts with the PX domain of the phosphoprotein. The interaction site has been shown to sample an equilibrium of discrete helices in the free state, which forms an encounter complex implicating the stabilization of one of the helical conformers upon interaction with PX, prior to diffusing on the surface of PX and engaging in the actual binding site. However, very little is known about the timescales of chain motions, which surely play a role in the interaction kinetics, in particular in terms of the on-rate of the interaction. This 124 amino acid protein also provides a good model system, containing long unfolded domains with chain-like dynamics and regions with residual structure. The measurement of extensive set of coherent relaxation rates at multiple magnetic fields, multiple temperatures and in three different length constructs of the same IDP has allowed us to characterize the dynamic nature of N_{TAIL} in unprecedented detail. By analyzing the relaxation data using extended model-free approach, we show that fast (≤ 50 ps) components of the correlation function report on librational motions. A dominant mode occurs on timescales around one nanosecond, apparently reporting on backbone sampling within Ramachandran sub-states, while a slower component (5-25 ns) reports on segmental dynamics dominated by the chain-like nature of the protein. The ability to delineate intrinsic modes and timescales will improve our understanding of the behavior and function of IDPs.

Résumé

Les protéines intrinsèquement désordonnées sont des hétéropolymères très flexibles, impliqués dans des activités cellulaires importantes (transduction du signal, reconnaissance moléculaire, traduction etc.), représentant des cibles potentielles de médicaments contre les maladies neurodégénératives et cancers, et dont les modes dynamiques définissent leur fonction biologique. Même si les états conformationnels qu'elles échantillonnent sont relativement bien connus, ce n'est pas le cas des échelles de temps de la dynamique associée. Dans ce travail nous étudions le comportement conformationnel du domaine C-terminal intrinsèquement désordonné de la nucléoprotéine de virus de Sendai (N_{TAIL}), qui interagit avec le domaine PX de la phosphoprotéine. Des études précédentes montrent que le site d'interaction échantillonne un équilibre entre trois hélices discrètes dans l'état libre, et que l'interaction avec PX passe d'abord par la formation d'un pré-complexe, où l'une des conformations hélicoïdales de N_{TAIL} est stabilisée, puis par sa diffusion sur la surface de PX, et enfin sa rétention sur le site de liaison. Cependant, aucun renseignement n'existe sur les échelles de temps de mouvements de la chaine de N_{TAIL}, qui influencent certainement la cinétique de cette interaction, en particulier sa constante de vitesse d'association. Cette protéine de 124 acides aminés représente aussi un système modèle pertinent contenant à la fois de longs domaines dépliés et des régions de structure résiduelle. La mesure d'un vaste et cohérent ensemble de taux de relaxation à différents champs magnétiques et différentes températures nous a permis de caractériser la dynamique de N_{TAIL} à un niveau de détail sans précèdent. A l'aide d'analyse « model-free » étendu, nous avons montré que les composants rapides de la fonction de corrélation nous informent sur les librations. Le mode dominant se situe à des échelles de temps autour d'une nanoseconde et est lié à l'échantillonnage de l'espace de Ramachandran par le squelette peptidique. Enfin, le composant lent (5-25 ns) nous informe sur les mouvements de segments de la chaine peptidique. La description des mouvements intrinsèques des protéines désordonnées et leurs échelles de temps contribuera à notre compréhension du comportement et des fonctions de ces protéines.