NMR methods for intrinsically disordered proteins : application to studies of NS5A protein of hepatitis C virus

Zsofia Burkart-Solyom

To cite this version:

Zsofia Burkart-Solyom. NMR methods for intrinsically disordered proteins : application to studies of NS5A protein of hepatitis C virus. Structural Biology [q-bio.BM]. Université de Grenoble, 2014. English. NNT : 2014GRENY077 . tel-01343993

HAL Id: tel-01343993
https://theses.hal.science/tel-01343993

Submitted on 11 Jul 2016

UNIVERSITÉ DE GRENOBLE

THÈSE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ DE GRENOBLE

Spécialité : **Physics for life sciences**

Arrêté ministériel : 7 août 2006

Présentée par

Zsófia BURKART-SÓLYOM

Thèse dirigée par **Bernhard BRUTSCHER**

préparée au sein **Biomolecular NMR spectroscopy group, Institut de Biologie Structurale**
et de **Ecole Doctorale de Physique**

NMR methods for intrinsically disordered proteins – Application to studies of NS5A protein of hepatitis C virus

Thèse soutenue publiquement le **06/11/2014**,
devant le jury composé de :

Dr. Guy LIPPENS
DR-CNRS, Institut Pasteur de Lille, France, Président
Prof. Robert KONRAT
Professeur, University of Vienna, Austria, Rapporteur
Prof. Isabella FELLI
Professeur, University of Florence, Sesto Fiorentino, Italy, Examinatrice
Prof. Dieter WILLBOLD
Professeur, University of Düsseldorf, Germany, Examinateur
Dr. Malene RINGKJØBING-JENSEN
CR-CNRS, IBS, Grenoble, France, Examinatrice
Dr. Bernhard BRUTSCHER
DR-CEA, IBS, Grenoble, France, Directeur de thèse

Preface

The research described in this PhD thesis was carried out in the framework of the FP7 IDPbyNMR Marie Curie intensive training network. This network provided us, the eleven PhD students and three postdoctoral researchers, a great environment to work on the expanding field of intrinsically disordered proteins (IDPs). Part I of this thesis is a general introduction to IDPs and the introductory and theoretical part of Part II introduces the concepts of NMR spectroscopy of IDPs. Parts of this introduction at the beginning of Part II is going to be published in a book chapter on NMR spectroscopy of IDPs, co-authored by the author of this thesis.

One of the objectives of this PhD project was defined by the role of the partner laboratory in the network – development of fast NMR methods for studies of IDPs. Development of fast NMR methods has a long history in the laboratory, a large toolbox of time- and sensitivity-optimized methods were developed over the past few decades. In the course of this PhD project BEST-TROSY-type experiments were introduced for time-efficient studies of IDPs. 3D BEST-TROSY sequences were introduced for assignment of large IDPs. These experiments provide high resolution and sensitivity in a short overall experimental time. Furthermore, a Pro-edited experiment was introduced, for aiding backbone assignment for IDPs, that are typically rich in proline residues. Thirdly, a BEST-TROSY-based approach was proposed for fast measurements of amide proton solvent exchange rates. In addition, in the framework of a collaboration with CERM (Florence, Italy, the coordinators of the IDPbyNMR project) the $\mathrm{H}^{\mathrm{N-BEST}}$CON experiment was developed. This experiment combines the advantages of the larger signal dispersion provided by carbon detection, with the signal enhancement resulting from selective pulsing on amide protons. The results on NMR method developments are presented in Part II of this thesis.

During this PhD research project I had the opportunity to visit three of the partner laboratories of this intensive training network. I spent three weeks in CERM in Florence with Profs. Isabella Felli and Roberta Pierattelli, working on development of the $\mathrm{H}^{\mathrm{N-BEST}}$CON experiment, as mentioned above. I spent one week at the EMBL Hamburg partner in the laboratory of Dr. Dmitri Svergun, measuring and analyzing SAXS data, presented in Chapter 8. Furthermore, I also had the chance to go for a week-long visit at Cambridge University in the laboratory of Prof. Michele Vendruscolo, setting up replica-averaged chemical shift restrained molecular dynamics simulation of NS5A D2D3. However, unfortunately, because of the too large size of the protein the calculations had to be stopped before yielding results.

Parallel to method developments for IDPs, the other part of the project was to apply these methods for studies of a biologically important IDP. The structural and functional NMR studies of a particular IDP, non-structural 5A (NS5A) protein of hepatitis C virus originated from a project that was started in Forschungszentrum (FZ) Jülich, in the laboratory of Prof. Dieter Willbold and continued at IBS by a postdoctoral researcher, Sophie Feuerstein in collaboration with FZ Jülich. I started working on the project during a three-months internship in 2010, about a year before this thesis started. In the meantime, yet another postdoc worked on the project, Melanie Schwarten with whom I continued this work in the course of this thesis project.

The main questions addressed during this thesis concerning NS5A were as follows. Firstly, characterization of structure and dynamics of the entire intrinsically disordered region of the protein. Secondly, a detailed characterization of its binding modes to the SH3 domain of Bin1 protein, as a follow-up of the first studies conducted by Sophie Feuerstein. Thirdly, an NMR characterization of the phosphorylation of the protein and its impact on the structural ensemble of this protein fragment. The research carried out on NS5A is described in Part III.

Acknowledgements

During these three years at IBS I had the opportunity to learn a lot, not only about the exciting field of protein NMR spectroscopy, but also about the life of a scientist. These three years were a marking transition period, from life as a student to life as a researcher. I would like to thank here all the people who played an important role in this process, and all those, who contributed to the research described in this thesis.

I would like to acknowledge my advisor, Bernhard Brutscher, for giving me the opportunity to work in the NMR group at IBS and to become an IDPbyNMR fellow. Thank you for giving me a lot of freedom in this research project, and for the discussions about research and science, teaching me the importance of working independently. Thank you for the time you invested in showing me the practices of NMR spectroscopy and for showing at the spectrometer how a "fast NMR" spectroscopist works with extreme efficiency. I would also like to thank you for organizing the purchase of all the new Bruker spectrometers for the new IBS building as the former head of the NMR group, resulting in more NMR time available for our research and opportunity to work on these high-performance new instruments. I would also like to thank Jerome Boisbouvier, who took over the task of leading the NMR group from Bernhard, for managing this increasing group in a very efficient way and for the setup of the life of the group in the new building.

I would like to acknowledge my closest co-workers during this PhD project. Sophie, who was there at the beginning and whose laughing and eternal good mood is unforgettable. Melanie, from whom I learned a lot about working in the wet lab and research in general. Enrico, thank you for being that great office mate and colleague, patiently teaching me how to write Python scripts as well as for the instructive days spent at the spectrometer, optimizing the HET^{ex}-BEST-TROSY experiment. I learned a lot from you about NMR spectroscopy, scientific rigor and sarcastic Enrico jokes.

Adrien and Sergi for taking care of the spectrometers in the old building and in the new, and making it possible for everyone to use them without problem. Merci, Isabel, Elodie, Audrey and Lionel, for organizing the life of the wetlab. Thank you Isa and Audrey, for the help in the wetlab, and for Isa, for producing the Core protein of pestivirus.

I would like to acknowledge the people in the IDPbyNMR network for

all the interesting training courses and all the opportunities this network gave us. I thank Isabella Felli and Roberta Pierattelli for setting up and coordinating this network. I also thank all the supervisors and fellows for organizing and participating in the training courses, where we spent nice and instructive time together. The people whom I had the chance to meet and to shortly work with during the short stays at other partner laboratories. Giancarlo Tria and Clement Blanchet at EMBL Hamburg, Tomas Hosek and Sergio Gil at CERM in Florence, and Carlo Camilloni and Predrag Kukic in Cambridge.

I would like to acknowledge all the senior scientists in the group, for interesting scientific discussions in the group seminar and in general. Especially Dominique Marion and Paul Schanda, for organizing the seminars in NMR theory and sharing with us their big knowledge in NMR in the form of interesting discussions.

I thank the PhD student and postdocs friends in the laboratory, who contributed also to the non-scientific life during these three years were very important. My co-IDPbyNMR fellow in Grenoble, Jaka, with whom we shared nice events not only in Grenoble, but also during traveling to and at the numerous IDPbyNMR training courses. The friends from the lab, older ones and recent ones as well. The composition of the group of young people is continuously changing at IBS, but the good atmosphere and the wish for some fun and to go to the mountains always remained. Thank you Mathieu, Guillaume, Lauriane, Nicolas, Rime, Mathilde, Jens, Audrey, Katharina, Meike, Melanie, Enrico, Sophie, Valery, Robert and Luca.

I would like to thank those people who read through parts, or the entire length of this manuscript. Thank you Bernhard for your critical comments, suggestions and corrections. Thanks for Jürgen, Gitte, Katharina and Johannes for proofreading.

I would also like to thank Robert Konrat and Guy Lippens for accepting reviewing this thesis.

And last, but most importantly, my parents, who always supported me in doing what I wanted to do, even in moving so far away from them for enrolling in this PhD program. And Johannes...I do not attempt to express the unexpressable.

Contents

Part I

Intrinsically disordered proteins: history, roles, viral IDPs

Chapter 1

Intrinsically Disordered Proteins

1.1 A short historical overview

Intrinsically disordered proteins (IDPs) are a class of proteins that have no permanent secondary or tertiary structure and are functional as such. This definition contrasts the central dogma of structural biology that has emerged since the first protein structure of myoglobin was solved and more and more protein structures have been elucidated since then and deposited in the Protein Database (PDB). This central dogma states that the three dimensional structure of the protein determines its function and is absolutely required for it. Certainly, once the structure of a protein is known, important insights can be gained into the mechanism of its function. However, the deterministic view on the function of proteins has been challenged by the expanding field of intrinsically disordered proteins.

The emergence of the concept of IDPs was a relatively slow procedure. It has been recognized for certain classes of proteins that large fragments of the protein sequence are characterized by low sequence complexity, that are unlikely to fold. For example, in case of DNA binding proteins this fact was described already in the 1980s [1]. At that time it was also increasingly recognized that certain, functionally important segments of proteins are missing from the X-ray structures and shortly afterwards NMR data of functional but disordered proteins were presented, clearly showing that they are unfolded [2]. A study of the Swiss protein database in 1994 showed that half of the protein sequences contain disordered regions [3]. At the end of the 1990's the first disorder predictor, PONDR®, was developed [4].

A major milestone in changing the view on the structure-function paradigm was the publication of the review of Wright and Dyson in 1999 [5]. They presented typical classes of disordered proteins and the data available at that time confirming their existence. They also presented a thorough characterization of an IDP by NMR spectroscopy (at that time termed IUP–Intrinsically unstructured protein), a fragment of the large, modular CREB-binding protein (CBP) and they described the folding of this IDP upon interaction with its partner.

At the beginning of the 21$^{\text{st}}$ century the interest for IDPs continued growing and the number of examples was increasing as well. This led to the appearance of several reviews of the new field [2, 6] and also to the development of sequence-based disorder predictors that will be briefly discussed in the next section. In order to allow for bioinformatics studies of IDPs, a database of intrinsic disorder was introduced (DisProt) [7].

Another important step in the history of IDP research was the in-cell NMR specta of τ protein injected in *Xenopus Laevis* oocytes and α-synuclein overexpressed in *E. Coli.* These were important results against the arguments that disorder is only an *in vitro* artifact resulting from the lack of macromolecular crowding or the absence of the physiological binding partner. The result cited here demonstrated that disorder persists in the crowded environment of a cell. However, it has to be pointed out that as the proteins are either overexpressed or injected in larger than physiological concentrations these studies did not disprove the second concern about the absence of physiological binding partners (available at appropriate concentrations). Current efforts thus aim at introducing proteins at close to physiological concentrations in their physiological cell type.

The existence and functional relevance of IDPs is nowadays generally accepted. Focus is shifting from the demonstration of biological relevance of disorder towards the description of their interaction modes and also towards the description of their conformational free energy hypersurface with ensembles of conformers derived from Monte Carlo or MD simulations. These approaches will be discussed in later sections of this thesis.

A recent advance in the field was the opening of the pE-DB (protein ensemble database) [8] for deposition of structural ensembles of IDPs. Let us note, that there is also an Ensemble Protein Database (EPDB) website hosted by the University of Pittsburgh, but it is not opened for deposition by external users.

1.2 Disorder predictors

Two types of disorder, long-range and short-range disorder, can be distinguished. Long-range disorder is typically defined as a disordered region longer than 40 amino acids. Several predictors focus on identification of protein fragments of long disorder, predictors of short-range disorder are less frequent. The first disorder predictor, that was developed towards the end of the 90's is PONDR® [9]. It is based on the fact that amino acid composition of IDPs is different from that of folded proteins. Amino acids are classified as disorder-promoting or order-promoting and each amino acid is given a score according to the amino acid composition of the residues around it in a defined window. Several factors are considered in the scoring algorithm, including hydrophobicity, charge etc. [10].

Another commonly used disorder predictor that we also applied for our studies is IUPRED [11,12]. IUPRED evaluates possible interactions between the amino acids in the sequence. Pairwise interaction energies of amino acids are taken from globular protein structures. Based on these interaction energies a score, named disorder tendency, is given to each amino acid based on its probability to establish contacts with other amino acids within the protein sequence.

The number of disorder predictors available is still increasing. For a comprehensive review on disorder predictors see Ref [13].

1.3 Roles of disorder

Here, a non-comprehensive overview of the so-far identified most frequent roles of disorder will be given. As already mentioned, the first class of proteins, where prevalence of disorder was recognized, is the class of DNA-binding proteins. Their folding upon binding to DNA has been described, as reviewed by [5].

In order to understand the roles of disorder it is instructive to consider the advantages that flexibility provides as these proteins evolved for specific functions where conformational flexibility is an advantage.

As a first example, the lack of a stable 3D structure allows IDPs to interact with multiple partners. Therefore, IDPs are often found as hub proteins in protein interaction networks [14]. This is advantageous for playing roles in regulation or for proteins of pathogens, interfering with the host cell defense mechanism.

A more direct example for the advantage of flexibility for function are entropic chains that act by entropic exclusion effects. The disordered protein regions in the nucleoporin complex represent a major example for this function of disorder. They exclude substrates passing the membrane that are too large and should not enter the nucleus.

Often, sites important in signaling, such as sites for phosphorylation or ubiquitination are found in disordered regions [15,16]. In these cases, flexibility makes the sites accessible for post-translational modifications, providing a transient but specific interaction with the enzyme. In addition, as interactions of IDPs often occur via short linear motifs, these motifs can be tuned by post-translational modifications eg. phosphorylation.

Transient binding is also necessary for chaperoning activity, as it allows non-specific and week interactions with multiple partners. Again related to the ability of interacting with multiple partners, the large hydrodynamic

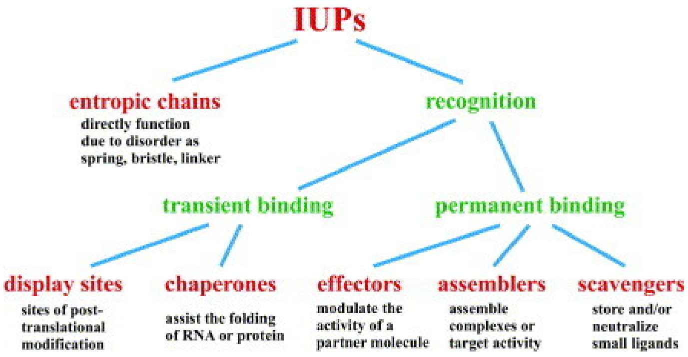


Figure 1.1: Most common functions of IDPs. The figure has been adapted from Ref [18]

radius of IDPs compared to folded proteins is advantageous for establishing contacts with their binding partners. This is the so called “fly-casting” mechanism of IDPs.

Furthermore, conformational disorder has also been shown to be important among RNA and protein chaperones as reviewed by Tompa [17].

Typical functions of IDPs have been classified into six classes in a review by Tompa [18] based on the role of disorder in exerting their function. This classification is shown in Fig. 1.1.

Figure 1.2 adapted from the review by He *et al.* illustrates the functions of known IDPs in 2009 as seen in the DisProt database [13].

1.4 Viral IDPs

Disorder also plays an important role for the function of many viral proteins. Viruses generally have small genomes, encoding only a few proteins, usually helicases, proteases, DNA/RNA polymerases and structural proteins. In addition, many viral proteins are multifunctional and responsible for “hijacking” the host cell defense mechanisms by interfering with the regulation of key cellular processes. It is becoming more and more clear that eukaryotes use short linear motifs (SLiMs) extensively in regulation [19]. Some of these motifs can be turned on and off by signaling events such as phosphorylation. Viral genomes evolved to use similar short linear motifs much like the host regulatory proteins, a phenomenon termed motif mimicry [20]. Intrinsic disorder is thus advantageous for this class of viral proteins as it allows the motifs to be available for interaction when the partner is approached. These interactions should be transient, of low-affinity, and characterized by high off-rates, as the protein has to interact with several partners as known from their interaction networks.

Another advantage of intrinsic disorder for viral proteins is that it supports fast evolution allowing for high mutation rates that is a key feature of

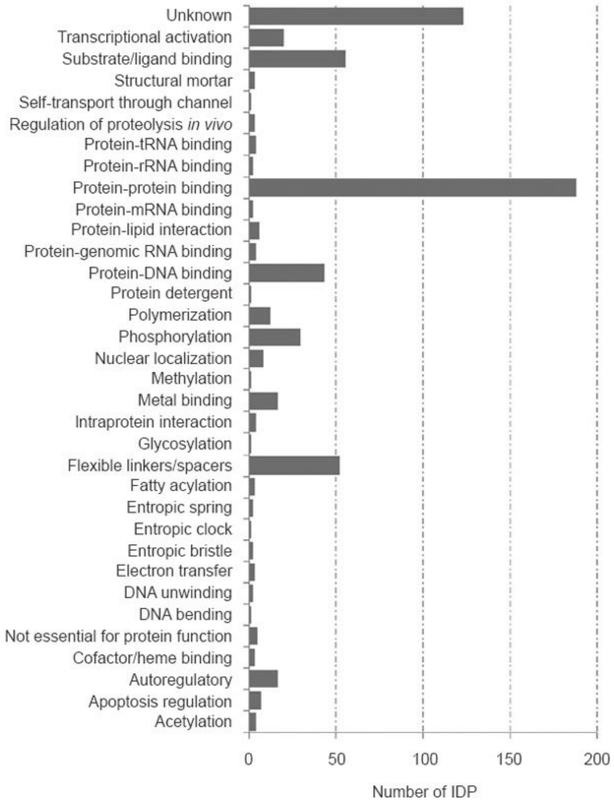


Figure 1.2: Known functions of IDPs and number of IDPs exerting these roles as described in the DisProt database as of 2009. The figure has been adapted from Ref. [13]

viruses. In case of IDPs only the residues in the interaction motifs are highly conserved, the other residues are able to mutate freely, as these mutations will not result in misfolding of a structure and thus cause the protein to be unfunctional. This high adaptability of viral proteins often results in drug resistance.

Many viral infections are known to provoke cancer development, such as Epstein-Barr virus (EBV), human papilloma virus (HPV) or hepatitis C virus (HCV). HIV does not directly cause cancer, but, by weakening the immune system, cancer development is more likely than in healthy individuals. The reason for this is likely to be the deregulation of cellular pathways, caused by the motif mimicry mentioned above. Thus, cancer development resulting from viral infection, according to the emerging consensus is an unfortunate side effect, as it has no direct advantage for the virus [21]. Several examples of viral IDPs are known to be oncoproteins, such as E7 from HPV or Adenovirus E1A. NS5A of HCV is also likely to have oncogenic activity. Its interaction with p53 has been shown to exert an anti-apoptotic effect as reviewed by Tsai and Chung [22].

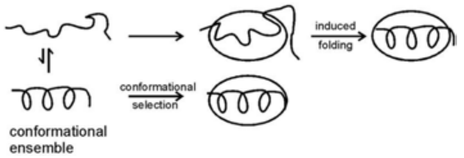


Figure 1.3: Schematic mechanism for folding upon binding and induced fit scenarios. The figure has been adapted from Ref [23]

1.5 Disorder to order transition

Intrinsically disordered proteins are far from being random coil-like. Preformed structural elements, termed molecular recognition elements (MoREs) are often required for interactions of this kind of proteins. Thus identification of these preformed elements may give a hint on the bound form conformation, as these structures are often stabilized by binding. Several studies aimed at understanding the detailed mechanism of this phenomenon of folding upon binding. Basically there are two extreme cases for the folding mechanism, binding-induced folding and conformational selection. The difference between the two scenarios is illustrated in Figure 1.3. In conformational selection the bound conformation pre-exists in the structural ensemble, while in the binding-induced folding scenario only proximity of the ligand induces changes of the conformational energy landscape in a way that the protein will fold into the bound conformation. In reality, usually a combination of these two extreme scenarios occurs, but there are examples for cases closer to one or to the other extreme as well, as reviewed by Wright and Dyson [23]. The idea of conformational selection is supported by a recent study showing that a hub protein interacting with a large number of disordered proteins binds those substrates with the highest affinity, where the bound conformation is preformed in the structural ensemble as shown by ITC measurements and MD simulations of all the binding partners [24].

There are some examples of interactions of IDPs where no structuring upon binding is observed and the protein remains disordered in the complex. Tompa and Fuxreiter termed these interactions “fuzzy” [25, 26]. Usually electrostatic interactions govern formation of this kind of complexes. An example of such an interaction is the binding of Sic1 to Cdc4 [27].

Chapter 2

Production of IDP samples

2.1 Expression and purification

In order to be able to study a protein with NMR or other biophysical methods, the protein should be expressed and purified. The most common way is to express the protein recombinantly in *E. Coli.* In this case, the synthetic gene coding for the protein of interest can be purchased from a manufacturer. Alternatively, the gene coding for the sequence of interest can be isolated from the original organism. This, however, would result in significantly reduced yields, because of the lack of codon optimization. Synthetic genes are codon-optimized to match the tRNAs that are most frequent in *E. Coli.* The next step is to clone this gene into an expression vector. This step can usually also be performed by specialized companies. The expression vector encodes for a sequence that will allow purification of the protein. This can be for example a His-tag, GST-tag or other fusion proteins.

The next step is to find the optimal bacterial strain for expression, and to test for the optimal induction time, temperature and duration of overexpression after induction. After all these parameters have been optimized, the expressed protein can be purified with the help of the affinity tag, and one or two additional chromatographic steps, eventually with cleavage of the purification tag. This is a critical step in case of IDPs, because they are very sensitive to proteases thus the time they are exposed to them should be kept minimal. Therefore, purification has to be performed as fast as possible.

In a recent review, Uversky pointed out that the potential states of proteins should not be considered as simply folded or disordered, instead as a continuum of states from order to disorder, this was termed the flavors of disorder [10]. Some IDPs, especially those with a “flavor” about halfway between the ordered and disordered states (those with exposed hydrophobic residues) are prone to aggregation. A new technique for resolving aggregation problems of folded proteins has recently been introduced. The key idea is to express the aggregation-prone protein fused to a highly charged IDP that is highly soluble [28].

Some IDPs can be purified simply by boiling the bacterial cell lysates, an example for this is α-synuclein. In this case all the other proteins, including proteases get denatured and coagulate, thus the protein can be obtained by simply taking the supernatant after centrifugation.

A problem may occur if the IDP turns out to be toxic for the expressing cell. In this case the recently emerged technique of cell-free protein expression can be applied that uses the cell expression machinery, but in the form of an extract. Thus there are no living cells, so toxicity is no longer an issue. In this case however, one has to deal with the problem of proteases present in the cell lysate which in a living cell are found in different compartments than where expression takes place. Therefore, appropriate protease inhibitors need to be used.

The bacterial cytosol is a reductive environment, thus cysteins of the expressed IDPs will not form disulfide bonds. However, as soon as the cells are lysed, oxidation might occur. This can result in the formation intermolecular disulfide bridges. This can be avoided by including strong reducing agents in every buffer throughout the purification procedure starting from the cell lysis.

2.2 Choice of sample conditions

As will be discussed in detail in later chapters of this thesis, sample conditions can be extremely important in studying IDPs. In case of a folded protein changes in sample conditions might for example change aggregation tendencies, but in general it is not expected that the overall conformation of the protein would change, until the point which results in denaturation. Therefore, the experimenter can adjust temperature and pH in a relatively broad ranges (of course the pI of the protein should be respected) limited only by denaturation or loss of activity of the protein.

For IDPs the situation is completely different. Multiple studies have shown, that sample conditions have a major impact on the conformational ensemble of intrinsically disordered proteins. Long range interactions of IDPs are largely mediated by transient electrostatic interactions causing collapse of highly charged IDPs in aqueous solution in spite of the lack of a hydrophobic core [29]. As this collapse is mediated by electrostatics it can be very sensitive to ionic strength. Extreme acidity – or basicity– induced compaction has also been observed for IDPs (probably resulting in introduction/neutralization of charges and consequently changes in the aforementioned electrostatic effect). Furthermore, a recent study showed that α-helicity increased when decreasing the pH for BASP1 protein. The authors observed that the increasingly helical region was rich in glutamic acid residues that are neutralized at low pH which can account for reduced electrostatic repulsion between the side chains in the context of an α-helix [30]. On the other hand, temperature induced structuring has been reported for several IDPs, as manifested by higher amount of α-helicity seen in the CD spectra. This was explained by increasing strength of hydrophobic interactions with increasing temperature (Ref [10] and references therein). In another, NMR and SAXS-based study a different explanation was suggested, namely loss of PPII conformation, that results in similar CD signature, in addition to loss of α-helicity [31]. The authors found radii of gyration of two IDPs to increase with increasing temperature. However, our SAXS data on NS5A protein contrast this finding, as will be discussed in Chapter 8 (Table 8.1).

A general conclusion from the above presented properties of IDPs that the result of their biophysical characterization might largely depend on the chosen sample conditions that have to be selected very carefully. Consequently, in case of functional studies it is important to be as close to physiological conditions as possible (note, that the relevant conditions for function are often unknown, especially if exact cellular localization of the protein is not known).

2.3 Aggregation of IDPs

Some IDPs are prone to aggregation because of their exposed hydrophobic residues. The most prevalent examples are amyloidogenic proteins, that became well-known and well-studied because of their roles in amyloidogenic diseases, that are neurodegenerative diseases and are caused by formation of amyloids. Examples of such IDPs are α-synuclein that plays a role in the development of Parkinson disease, τ protein and amyloid-β proteins that play a role in Alzheimer's disease.

IDPs that are prone to aggregation should be studied at low sample concentrations in order to minimize aggregation. Ideally, before undertaking further biophysical studies monodispersity of the sample should be reached by optimizing sample conditions. Monodispersity can be checked by multi-angle laser light scattering (MALLS) or analytical ultracentrifugation.

2.4 Proteolytic degradation

In the living cell, IDPs are located in a well-defined compartments, where protease activity is highly regulated. Disorder provides sites for signaling events, such as ubiquitination [32], that targets them for proteolytic degradation. There are controversial studies, but most data point to the conclusion that the *in vivo* half life of IDPs seems not to be significantly shorter than that of folded proteins [33]. According to the “nanny model” of IDPs they are prevented from degradation by complex formation [34]. However, when preparing the cell lysate, all the proteins present at different locations in the cell are mixed and even traces of proteases are sufficient to degrade an IDP. They are readily degraded by small, broad-specificity proteases simply because of the exposed nature of their peptide bonds.

The problem of degradation can be handled either by heat stabilization, by cell-free expression or by application of protease inhibitors (however, according to our experience even broad-range protease inhibitor cocktails do not necessarily provide sufficient protection against degradation).

In case of cell-free protein expression, the above mentioned protection of the highly regulated protein degradation machinery of a living cell is no longer valid, so the proteases responsible for degradation should be identified and suitable protease inhibitors should be used [35].

Chapter 3

Biophysical methods for characterization of IDPs

In this chapter the most common biophysical methods for the studies of IDPs will be reviewed. They were applied in the course of this thesis with the exception of circular dichroism (CD) spectroscopy that is included for the sake of completeness as it was very important for early studies of IDPs, and is still frequently used for quick sample characterization.

3.1 CD spectroscopy

Circular dichroism (CD) spectroscopy is a frequently used method for quick characterization of secondary structure content in proteins. It gives information on the sum of secondary structural elements in the whole protein, as it is a spatially low resolution method. It relies on measurement of differential absorbance of right and left circularly polarized light in the UV frequency range. Optically active molecules give a signal, if the mixture is not racemic. As proteins are optically active (as in nature only L-amino acids can be found) the method can be used for their characterization. CD spectra are measured at frequencies near the absorption bands of interest. Most of the secondary structural conformations give characteristic signatures in their CD spectrum. As CD spectroscopy is not an atomic resolution method it is not useful for detailed structural characterization of biomolecules. However, it can be very useful for monitoring changes of secondary structure for various perturbations, as well as for monitoring effects of changing temperature, pH, ionic strength, etc. on secondary structure content. This is why CD spectroscopy played an important role in early biophysical characterization studies of IDPs and is still used for screening of sample conditions. An example of a CD study of varying pH and monitoring the effect on the structure of α-synuclein is shown in Figure 3.1.

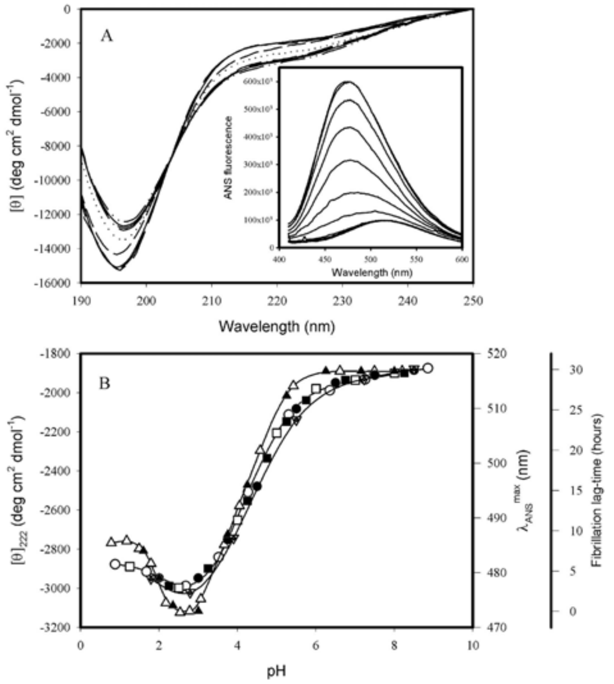


Figure 3.1: (A) Far-UV CD spectra of α-synuclein at multiple pH values. (B) Comparison of the effect of pH on far-UV CD (*circles* and *squares*) and ANS fluorescence spectra (*triangles*). The figure was adapted from Ref [36].

3.2 Static and dynamic light scattering

Light scattering measurements give very useful information on the dimensions and the molecular mass of IDPs. Static light scattering (SLS) and dynamic light scattering (DLS) will be briefly discussed here. Static light scattering gives information on the molecular mass. Multi-angle laser light scattering (MALLS) is nowadays the most frequently used static light scattering technique. The major information that can be extracted from the data is the molecular mass. It is a non-invasive method and the molecular mass, in contrast to mass spectrometry, can be determined in the solution state at the sample conditions optimized for the biophysical characterization of the IDP. Usually MALLS is coupled to a size exclusion chromatography (SEC) column, because in this way a monodisperse solution gets into the cuvette (at least on the SEC timescale) . In addition, it is also coupled to a UV-VIS spectrometer, in order to measure the sample concentration, and to a refractometer. The experimental setup of the light scattering experiment is shown in Figure 3.2.

The technique is based on the fact that light is elastically scattered (termed Rayleigh scattering) on the macromolecules in the solution. Instead of the scattering of molecules in vacuum, excess scattering of the

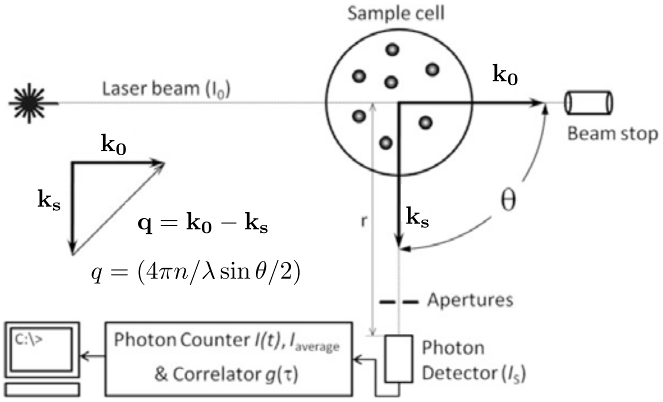


Figure 3.2: Experimental setup of the light scattering experiments. The figure was adapted from Ref [37].

macromolecules compared to the solution is measured. As shown in Figure 3.2 the intensity difference between the incident beam and the scattered beam is measured. The incident light is described by an electric field $\mathbf{E}(t, \mathbf{r}) = \mathbf{E_0} \exp(-j(\omega t - \mathbf{k_0 r}))$. The intensity of the incident beam is $I_0 \propto |\mathbf{E}^2|$. The incident beam is elastically scattered on the macromolecule, or more accurately, on a scattering element of the macromolecule. The dipole moment that is created in the scattering element of the macromolecule is equal to $\mathbf{p}(t) = \alpha \mathbf{E}(t)$ where α is the polarizability of the scattering element.

The scattered intensity at distance $\mathbf{r}$ is

$$I_s = \frac{16\pi^4}{\lambda^4} \frac{\alpha^2}{\mathbf{r}^2} I_0. \tag{3.1}$$

However, the scattered light from a scattering element will have a phase $phi_i = \mathbf{q}\,\mathbf{r_i}$, that depends on the scattering vector $\mathbf{q} = \mathbf{k_0} - \mathbf{k_s}$. The amplitude of $\mathbf{q}$ is $q = (4\pi n/\lambda) \sin\vartheta/2$. The scattered light with different phases from the differently oriented scattering elements i will produce, after their superposition, the total instantaneous intensity that can be measured behind the aperture of the detector.

This intensity fluctuates in time, as the scattering elements are moving by translational diffusion that changes their location $\mathbf{r_i}$. In SLS the time average of the scattering is measured while in DLS this time fluctuation is measured.

In SLS the scattering can be calculated based on Eq. 3.1. In order to be able to measure the molecular weight of the particle the polarizability (α) of the scattering element is expressed as the difference between the refractive index of the solution (n) and the solvent (n_0). This relation is called the Clausius-Mosotti equation and it gives the relation between the microscopic quantity, the polarizability and the resulting macroscopic one, the refraction index.

$$\alpha = \frac{n^2 - n_0^2}{4\pi N'} \tag{3.2}$$

where N' is the number of particles per unit volume $N' = N_A c/M$. N_A is the Avogadro number, M is the molecular weight (we want to determine) and c is the weight concentration of the solute.

From equations 3.1 and 3.2 we get the dependence of the excess scattered intensity ($I_{ex} = I_s - I_{solvent}$) on the molecular weight as

$$I_{ex} = \frac{4\pi^2 n_0^2 (\partial n/\partial c)^2}{\lambda^4 N_A} \frac{\nu}{r^2} cMI_0 = H\frac{\nu}{r^2} cMI_0. \tag{3.3}$$

The method allows highly accurate measurements of molecular weights in the range of $10^3 - 10^8$ Da. In MALLS the scattering is measured at multiple angles in order to get more accurate measure of the molecular weight. The MALLS experiments we conducted during this thesis are discussed in Chapter 8.

3.3 Small-angle X-ray scattering

Small angle X-ray scattering (SAXS) studies of biological macromolecules are usually conducted at synchrotrons, but there are also examples of studies with laboratory X-ray sources. It is a solution technique, the sample is irradiated with X-ray radiation and the scattering at low angles is measured. The experimental setup of a SAXS experiment is shown in Figure 3.3. Before the measurement, monodispersity of the sample has to be verified and the sample must be free from aggregates. The principle of the technique is similar to that of static light scattering presented in the previous section as the principles of elastic scattering also account for the physical basis of the SAXS experiment. SAXS is also an SLS technique, the difference to MALLS being the wavelength of the light and the scattering angle at which the intensity is measured, that is much smaller than in the case of MALLS (see the difference of the experimental setup: the SAXS detector is positioned at $0\,^\circ$ with respect to the incident beam. The scattering is measured at small angles as function of the scattering angle ϑ, while in MALLS the scattering is measured at multiple, but larger angles).

The scattering intensity is recorded as a function of the amplitude of the scattering vector **s** (that was denoted with **q** in the previous paragraph) and radially integrated. The amplitude of the scattering vector equals $s = 4\pi/\lambda \sin\vartheta/2$ where λ is the wavelength of the X-ray radiation and ϑ is the angle enclosed between the incident beam and the scattered light. The

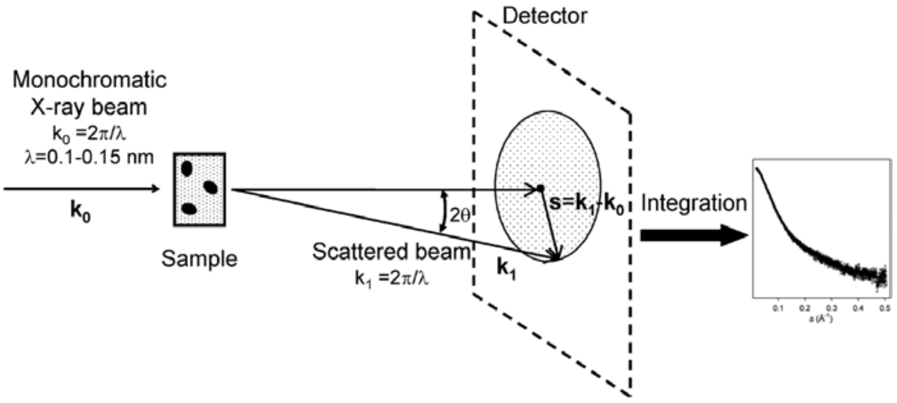


Figure 3.3: Experimental setup of the SAXS experiments. The figure was adapted from Ref. [38].

scattering depends on the excess electron density, that is relatively small in case of biomolecules as they are composed of light atoms. The scattering intensity of the buffer is subtracted from the scattering intensity of the solution containing the macromolecule. This difference yields the scattering of the macromolecule alone. The intensity of the scattered radiation is

$$I(s) = \langle I(s) \rangle_{\Omega} = \langle A(s) A^{*}(s) \rangle_{\Omega} \tag{3.4}$$

where $\langle I(s) \rangle_{\Omega}$ is the spherical average of the scattering intensity and $A(s)$ is the scattering amplitude that in turn depends on the Fourier transform of the excess electron density $\rho(\mathbf{r})$,

$$A(s) = \mathcal{F}[\rho(\mathbf{r})] = \int \Delta\rho(\mathbf{r}) \exp{(js\mathbf{r})} \mathrm{d}\mathbf{r}. \tag{3.5}$$

The scattering of a polydisperse system, and IDPs should be regarded as such on the timescale of X-ray scattering, is given by the sum of the scattering of the individual components:

$$I(s) = \frac{1}{N} \sum_{n=1}^{N} I_n(s), \tag{3.6}$$

where $I_n(s)$ is the intensity contribution to the scattering by the n-th conformer. N is the total number of conformers present, that is for IDPs unknown. When recording the SAXS curve, we measure the intensity of the scattered beam $I(s)$ as a function of the scattering angle.

The first information we can get from the recorded scattering curve is whether the sample is aggregated or not. The radius of gyration (R_G) can be easily extracted from the SAXS data, by the so-called Guinier analysis, that is based on the fact, that at small angles the dependence of $\ln I(s)$ on

s^2 is linear and therefore applying linear regression to this region allows for extrapolation to the scattering at zero angle, $I(0)$. The radius of gyration can be extracted, as

$$\ln I(s) = \ln I(0) - \frac{R_G^2}{3} s^2. \tag{3.7}$$

Of course, in the case of a polydisperse system, such as IDPs, the measured R_G is an ensemble average of the radii of gyration of the different conformers. Ensemble fitting of IDP structures to the scattering data, a method called Ensemble optimization method (EOM) allows for extraction of the distribution of radii of gyration of the conformations of the IDP that fit the SAXS data versus the conformations of a randomly generated pool of conformers [39]. For a recent review on SAXS of IDPs, see ref [38]. Small angle X-ray data measured during this thesis and their analysis, including application of the Guinier approximation to extract R_Gs and EOM fit will be shown and discussed in Chapter 8.

3.4 Isothermal titration calorimetry

Interaction studies involve determination of the dissociation constant as this gives an important information on the strength and in some case also the relevance of the interaction. Isothermal titration calorimetry (ITC) is an experimental thermodynamic approach which allows determination of the dissociation constant of the interaction by measuring the heat emitted or absorbed by the system (ΔH).

The experimental setup is shown in Figure 3.4. The setup consists of two, thermally well isolated cells, whose temperatures are equilibrated. One is the reference cell, the other one is the measurement cell containing the solution of the macromolecule. There is a syringe connected to the second cell that allows addition of small quantities of the ligand, thereby performing the titration. Both cells are kept at a constant temperature by a temperature control unit. When the ligand is titrated into the solution the temperature of the solution is changing, because of the reaction heat. The difference in temperature between the reference cell and the measurement cell is sensed and a power is applied that cools or heats the measurement cell to the temperature of the reference cell. The power that is applied is measured (in μJ/s) and recorded as a function of time, as shown in Figure 3.4. The area under the measured curve is then integrated and the result is the reaction heat corresponding to the fraction of ligand that is bound to the target. Knowledge of the exact concentrations of the ligand and the target is imperative as well as a perfect buffer match between the two solutions. In addition, a blank titration also has to be performed with the buffer alone to measure the eventual dilution heat that has to be subtracted from the measured reaction heat.

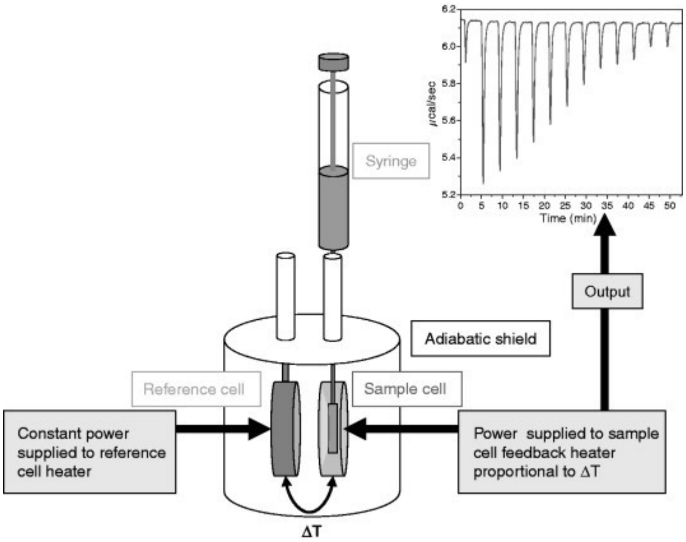


Figure 3.4: Schematic experimental setup of the power compensated ITC experiment. The figure was adapted from Ref [40].

When a curvature can be observed in the ITC curve the data is of good quality and can be fitted with a model including the stochiometry, ΔH, ΔS and K_D. For the ITC experiment performed during this thesis see section 9.3.2.

3.5 NMR spectroscopy

The potential of NMR spectroscopy for studies of IDPs will be discussed in detail throughout Parts II and III of this thesis. NMR has been used since the early days of IDP research for their characterization at atomic resolution. The methods that have been developed for studies of unfolded proteins have been applied to IDPs as well. NMR spectroscopy is capable of characterizing residual secondary structure of IDPs by measurements of secondary chemical shifts and residual dipolar couplings. Long range interactions, in other words residual tertiary structure, is most often characterized by measurement of paramagnetic relaxation enhancement. Based on the combination of NMR data, eventually with SAXS and other biophysical data, structural ensembles of IDPs can be calculated that fit the observed parameters.

By means of NMR, one can also characterize interactions of IDPs either with other IDPs, with folded proteins, or with small molecules. The binding interface can be mapped and in case of fast exchange, the dissociation constants can also be extracted from the NMR titration data.

Part II

New tools for NMR spectroscopic studies of IDPs

Chapter 4

Brief overview of NMR spectroscopy of IDPs

4.1 Introduction to NMR spectroscopy

4.1.1 NMR spectroscopy: a tool for structural biology

The phenomenon of nuclear magnetic resonance was discovered by Purcell et al. [41] and Bloch et al. [42] independently. Shortly afterwards they were awarded the Nobel Prize in Physics *"for their development of new methods for nuclear magnetic precision measurements and discoveries in connection therewith"*. Nuclear magnetic resonance spectroscopy, beyond its fundamental role in physics, is a very powerful tool in chemistry and biology for structural studies of small organic and inorganic compounds as well as large systems, including polymers and biomolecules. Besides X-ray crystallography and electron microscopy, it is the only method capable of resolving the structure of biological macromolecules at atomic resolution. Its important advantage compared to X-ray crystallography is that it allows studies of biological macromolecules in the liquid state – meaning that crystallization of the molecule is not required. This enables structural studies of highly dynamic systems such as intrinsically disordered proteins (IDPs). NMR spectroscopy can also be employed for studies of interactions of biomolecules with small organic ligands as well as other biological macromolecules, with no need for co-crystallization. The binding interface can easily be mapped, moreover, the binding affinity can be extracted from the NMR observables.

Recently, the importance of regarding biomolecules as dynamic entities, often insufficiently described by a single static structure, has become widely recognized [43, 44]. NMR spectroscopy is a unique tool to access dynamics information of biomolecules at atomic resolution, from the picosecond timescale to slow exchange processes on the second timescale. This is especially important for intrinsically disordered proteins that are often charac-

Isotope	Spin I	gyromagnetic ratio γ_n ($10^6\,\mathrm{rad\,s^{-1}\,T^{-1}}$)	gyromagnetic ratio/2π $\gamma_n/(2\pi)$ ($\mathrm{MHz\,T^{-1}}$)	natural abundance isotope %
^{1}H	1/2	267.513	42.576	99.98
^{2}H	1	41.065	6.536	0.015
^{13}C	1/2	67.262	10.705	1.108
^{15}N	1/2	-27.116	-4.316	0.37
^{31}P	1/2	108.291	17.235	100.0

Table 4.1: Gyromagnetic ratios and natural abundances of nuclei important for biomolecular NMR

terized by heterogeneous dynamics, that give hints about their tendencies to form transient structures or on their tendencies to aggregate.

4.1.2 Basics of nuclear magnetic resonance

NMR spectroscopy studies the interaction of atomic nuclei with electromagnetic waves in the radio frequency range, that induce spin transitions of the atomic nuclei in an external magnetic field. Here, only a brief introduction is given, with major focus on the factors determining sensitivity and resolution, as the methodological developments presented in this thesis mainly address these issues. For further details on the physical background of the technique the reader is directed to textbooks [45–47].

An atomic nucleus, in order to be observable by NMR spectroscopy, has to possess a non-zero spin quantum number I. The angular momentum associated with the nuclear spin is quantized. Its z components are determined by the magnetic quantum number m, that can take values of $-I$, $-I+1$, $-I+2$, ... $+I-2$, $+I-1$, $+I$, in total $2I+1$ states are possible. This leads to the magnetic momenta μ_z:

$$\mu_z = \gamma m \hbar \tag{4.1}$$

where γ is the gyromagnetic ratio and $\hbar$ is the reduced Planck constant. Gyromagnetic ratios and natural abundances of the nuclear isotopes important in biomolecular NMR are listed in Table 4.1.

When the nuclei are placed into a magnetic field B_0, that is by convention aligned along the z axis, the nuclear magnetic momenta interact with the magnetic field. This leads to a splitting of the system into $2I+1$ energy levels, with energies given by Eq. 4.2

$$E = -\mu_z B_0 = \gamma m \hbar B_0 \tag{4.2}$$

4.1.3 Sensitivity in NMR spectroscopy

Equation 4.2 has important consequences when considering the sensitivity of the NMR experiments. The energy difference between the two states of spin $I = 1/2$ nuclei is

$$\Delta E = \gamma \hbar B_0 \tag{4.3}$$

At thermal equilibrium at temperature T the ratio of spins in the lower and in the higher energy state follows the Boltzmann distribution:

$$\frac{N_\beta}{N_\alpha} = e^{-\frac{\Delta E}{kT}}. \tag{4.4}$$

Calculating a typical value illustrates the reason for the intrinsically low sensitivity of NMR spectroscopy. The NMR signal is proportional to the magnetization that in turn is dependent on the polarization (P)

$$P = \frac{N_\alpha - N_\beta}{N_\alpha + N_\beta} \tag{4.5}$$

that is the population difference between the two states divided by the total number of spins. In case of the proton nucleus, that has the highest gyromagnetic ratio, the ratio of spins in the upper energy state versus the lower energy state is 0.999872 at room temperature (298 K) in a magnetic field of 18.8 T. This means that only a very small fraction of the spins present in the sample give rise to the observable NMR signal. This is the case for conventional NMR spectroscopy, but in the rapidly evolving field of dynamic nuclear polarization NMR spectroscopy the much larger spin polarization of the electrons is transferred to the nuclei, resulting in dramatically increased sensitivity. [48, 49].

Furthermore, in case of the ^{13}C and ^{15}N nuclei, that are so important for biomolecular NMR, the natural abundance of NMR-active nuclei is very low, as shown in Table 4.1. In order to have high enough sensitivity for these nuclei, to be able to perform multidimensional experiments, isotopic enrichment is required, as it is also discussed in section 8.2.

Sensitivity of the experiment can be quantified by the signal to noise ratio (SNR) per unit time. It is instructive to consider the factors that determine the SNR, as shown in Eq. 4.6, as they give an overview of the options for improvements.

$$\mathrm{SNR} \propto \frac{N}{V} \gamma_{exc} \gamma_{det}^{3/2} B_0^{3/2} n_{scans}^{1/2} \frac{1}{\sqrt{R_s(T_a T_s) + R_c(T_a T_c)}} \tag{4.6}$$

SNR, similarly to other spectroscopic methods, is proportional to the number of spins N present in the sample. Thus it is desirable to use highly concentrated samples for NMR spectroscopy. However, for biomolecules, the maximum concentration of the sample is often limited by protein aggregation. Furthermore, obtaining a large quantity of labeled protein can be an expensive and time-consuming process.

Multidimensional NMR experiments usually employ proton excitation and detection, to take advantage of the high gyromagnetic ratio of ^{1}H nuclei. For intrinsically disordered proteins, ^{13}C-detected experiments are often used. It is worth noting at this point that in case of ^{13}C excitation as well as detection there is an intrinsic 32-fold decrease in sensitivity. In case of proton start and carbon detected experiments the intrinsic sensitivity loss is only 8-fold. However, different relaxation effects should also be considered when comparing sensitivity of ^{1}H-detected and heteronuclear-detected experiments [50], as well as adverse effects of high salt concentration in case of cryoprobes. Probes optimized for detection of heteronuclei are less affected by the latter [51].

The proportionality of SNR to $B_0^{3/2}$ gives the explanation for the well-known race for higher magnetic fields. To date, the highest field commercial magnet corresponds to 1 GHz proton frequency, and 1.2 GHz magnets are under construction.

The next factor in the expression represents the fact, that SNR increases with the square root of the number of scans n_{scans}. This is due to the fact that the intensity of the signal increases linearly, but the noise increases with the square root of the number of scans.

A recent development that greatly enhanced the sensitivity of NMR spectroscopy was the introduction of cryogenically cooled probes. The last factor in expression 4.6 is related to the electronics of the probe. The SNR of the experiment increases with decreasing the temperature of the preamplifier (T_a), the sample (T_s), the coil (T_c) as well as with resistance of the sample R_s and the coil R_c. This is why samples with high ionic strengths degrade the beneficial effects of cryogenically cooled probes on experimental sensitivity.

4.1.4 Basic NMR observables

After having discussed the factors influencing the sensitivity of the NMR experiment we can continue with discussion of the observables of our NMR experiment. Peak positions in the spectrum are dependent on the resonance frequencies of the observed nuclei. Based on Eq. 4.2 one would expect to see a single line in the NMR spectrum for each nucleus of a different type at the transition frequency determined by their gyromagnetic ratios. This is only true in the hypothetical experiment of observing “naked” nuclei. Nuclei resonate at slightly distinct frequencies if they are in chemically different environments within the molecule. The reason is that the nuclei experience a net B field that is the sum of the static B_0 field and secondary shielding fields induced by the motions of the electrons surrounding them. The net field can be written as

$$B = (1 - \sigma)B_0 \tag{4.7}$$

where σ is the isotropic, average shielding factor. Chemical shifts can be very characteristic for atoms in different chemical moieties and they give the first useful information that we can obtain from the spectrum. Chemical shifts are measured relative to chemical shift of a standard according to Eq. 4.8

$$\delta = \frac{\Omega - \Omega_{ref}}{\omega_0} \times 10^6 \tag{4.8}$$

where ω_0 is the operating frequency of the magnet in MHz, Ω is the offset of the nucleus of interest in Hz and Ω_{ref} is the offset of the standard in Hz, chemical shifts are expressed in ppm. Being relative, normalized quantities, measured chemical shift values are independent of the field of the instrument. The chemical shift makes NMR an atomic resolution technique: once we assign a resonance to a given nucleus, we can follow the response of this resonance to further manipulations. However, the difference of the chemical shifts of certain nuclei in only slightly differing chemical environments can be very small. We need sufficient resolution and signal to noise ratio in the spectrum to be able to distinguish them.

Spectral resolution is dependent on the linewidth of the peaks, that is determined either by their natural linewidth or the digital resolution of the spectrum, which in turn is the frequency range of the spectrum (known as spectral width *sw*) divided by the number of points defining the spectrum. The resolution of a 1D ^{1}H spectrum is in general not sufficient to distinguish the large number of not well-dispersed resonances arising from a protein sample as it can be seen from Figure 4.2. This problem can be tackled by means of multidimensional NMR experiments, that are also essential for assigning the resonances of the protein and obtaining chemical shifts of heteronuclei carry information about the conformations of the polypeptide chain.

Backbone heteronuclear chemical shifts are informative on the secondary structures the protein adopts. For a ^{13}C and ^{15}N labeled protein one can measure amide ^{15}N, ^{13}CO, ^{13}C$^{\alpha}$ and ^{13}C$^{\beta}$ chemical shifts, all of which are indicative of secondary structures. Information on the secondary structure may be obtained by comparing the measured chemical shifts with predicted random coil values as will be discussed and shown in section 8.5. Chemical shifts are also sensitive to changes in temperature and pH. In recent random coil chemical shift prediction algorithms these effects are taken into account.

Peak intensities are proportional to the number of nuclear spins giving rise to the peak. Importantly for IDPs, they also depend on the T_2 relaxation time constant of the spin, that depends on the motions of the molecule. The T_2 relaxation time constant determines the homogeneous linewidth of the Lorentzian peak $\Delta\nu = R_2/\pi$, where $\Delta\nu$ is the full width at half maximum. The experimentally observed linewidth is always larger than the homogeneous linewidth because of instrumental imperfections and temperature gradients. In case of IDPs, peak intensities in the spectrum are often heterogeneous, because relaxation properties of the spins vary, depending on the extent of transient structure in different parts of the molecule. Further intensity variations can be the result of differential solvent exchange rates, depending on the residue type, temperature, pH and the extent of the solvent-exposition of the amide proton. Peak intensities can be used for measurements of relaxation rates, PREs, solvent exchange rates, NOEs and to measure sample concentrations. Usually, the NMR analysis software determines the peak intensities in the spectrum during the procedure of peak picking. When the peak is picked the program fits it with a function – lorentzian, gaussian or a polynomial – according to the choice of the user, and the maximum corresponds to the peak intensity.

Interactions of the nuclear spin with nearby spins give rise to spin couplings. If the interaction is mediated by the electrons in a covalent bond, the effect causes splitting of the peak into $n+1$ peaks, where n is the number of the chemically equivalent nuclei involved in the interaction. The values of ^{3}J couplings in a polypeptide depend on the dihedral angle, thus they yield information on the geometry and secondary structure, following the Karplus relation. ^{3}J couplings can be determined by measuring the chemical shift difference between the split signals in Hz.

Magnetic dipoles placed into a strong magnetic field B_0 align parallel to it and interact through space with the interaction energy given by

$$E_{\mu_1\mu_2} = \frac{\mu_0}{4\pi}\frac{1}{r_{12}^3}\mu_1\mu_2\left(1 - 3\cos^2\theta\right) \tag{4.9}$$

where μ_1 and μ_2 are the magnetic moments of the two nuclei involved in the interaction, r_{12} is the distance between them, and θ is the angle the vector connecting them encloses with the B_0 magnetic field.

In NMR spectroscopy, the relevant magnetic dipoles are the nuclear dipoles, and the interaction energy is determined by the dipolar interaction Hamiltonian, where the secular term is

$$H_D = -\frac{\gamma_I\gamma_S\hbar\mu_0}{4\pi r^3}(3\cos^2\theta - 1)I_zS_z \tag{4.10}$$

that in isotropic media averages to zero,

$$\int_0^\pi (1 - 3\cos^2\theta)\sin\theta \mathrm{d}\theta = 0 \tag{4.11}$$

as all orientations are sampled with equivalent probability. Thus the information content about relative angles of the spins and their distances is lost. However, it is possible to reintroduce this coupling effect by weakly aligning the molecules in liquid crystalline media, and thus measure residual dipolar

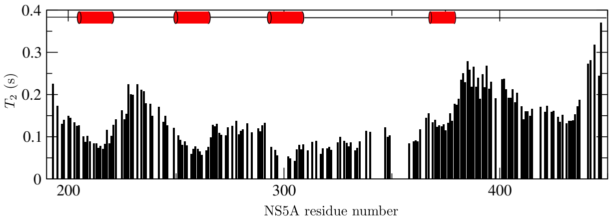


Figure 4.1: ^{15}N T_2 relaxation time constants of NS5A D2D3 measured at 800 MHz at 5 °C. Transiently helical regions are indicated by red cylinders.

couplings (RDCs). They can be measured in the same way as scalar couplings. The measured couplings in a partially aligned sample are the sum of the scalar couplings and residual dipolar couplings.

4.2 Challenges for NMR spectroscopy of IDPs

4.2.1 Experimental sensitivity

IDPs are by definition very flexible macromolecules. From a spectroscopic point of view this flexibility has both advantages and disadvantages, but studies of IDPs might require different techniques than those of globular proteins. One great spectroscopic advantage of their flexibility is that highly flexible, random coil-like regions lacking transient structure have ^{15}N T_2 relaxation time constants on the order of several hundreds of ms. This produces sharp lines in the amide proton-nitrogen correlation spectra and increased sensitivity. However, in regions with significant amount of transient structure, this is no longer the case, as illustrated by amide ^{15}N T_2 relaxation time constants of the NS5A protein of HCV (Figure 4.1). There are regions that have T_2 relaxation time constants four times shorter than other regions, leading to a large dynamic range of peak intensities.

Sensitivity can also be reduced because of aggregation. Some IDPs have tendencies to aggregate, as there might be exposed hydrophobic patches. Aggregation is governed by the same principles as protein folding, by interactions of these exposed hydrophobic patches. Some IDPs are involved in the development of neurodegenerative diseases such as α-synuclein or τ or prion proteins. Tendencies for aggregation can hamper NMR studies of these proteins, because it limits the maximal NMR sample concentrations.

Often problems are encountered because of proteolytic degradation of IDPs. When the flexibility of the protein permits sample stabilization by boiling, the proteases can be deactivated by heat denaturation. However, in case of exposed hydrophobic patches, thermal denaturation would result in irreversible changes of the residual structure and could cause precipitation. In these cases, the NMR spectra have to be acquired before sample degradation occurs, in a short experimental time.

As it will be shown in Part III transient structure of IDPs is highly sensitive to the sample and measurement conditions (pH, temperature). Thus many studies on IDPs attempt to mimic physiological conditions by using elevated pH and temperature. Furthermore, studies of enzymatic reactions, for example post-translational modifications, also require elevated pH and temperature in order to ensure optimal enzyme activity. Last, but not least, in the field of in-cell NMR spectroscopy sample conditions are obviously physiological. In-cell NMR has been especially important in the field of IDPs for showing that they remain flexible *in vivo* and disorder is not an artifact of sample conditions [52].

From an NMR spectroscopic point of view, the major difficulty is, in case of elevated pH and temperature, the fast exchange of amide protons with the solvent water protons. Exchange rates increase rapidly with increasing temperature and pH. This leads to broadening of amide resonances in an amide proton nitrogen correlation spectrum. Amide proton exchange rates depend largely on the residue type by as much as two orders of magnitude. This also leads to large dynamic range of peak intensities in the fingerprint NMR spectrum.

4.2.2 Spectral resolution

As already mentioned in the previous section 4.2.1, large T_2 relaxation time constants result in sharp lines in IDP spectra. However, major difficulties of IDP studies are caused by the very low frequency dispersion, especially in the ^{1}H dimension as illustrated by Figure 4.2. As there is no structuring, there is no hydrogen bond network that would result in proton chemical shifts different from random coil values in contrast to folded proteins. It has been recognized, that amino acid sequences of IDPs are often very repetitive. Because of lack of structuring, chemical environments of the corresponding nuclei in the repetitive regions are very similar, so they resonate at nearly identical frequencies. Thus, it is often not only impossible to distinguish them in the 2D fingerprint spectra, but there might be overlaps even in higher dimensional experiments.

4.2.3 Further difficulties, breakpoints in sequential assignment

In addition to problems with sensitivity and resolution, further problems can occur because of the typical large proline content of IDPs. The sequences are usually biased towards disorder-promoting residues and prolines are part of these [53]. As prolines do not have backbone amide protons, they are not present in amide proton-nitrogen correlation spectra. ^{13}C direct-detected experiments allow their observation, but as has been mentioned already,

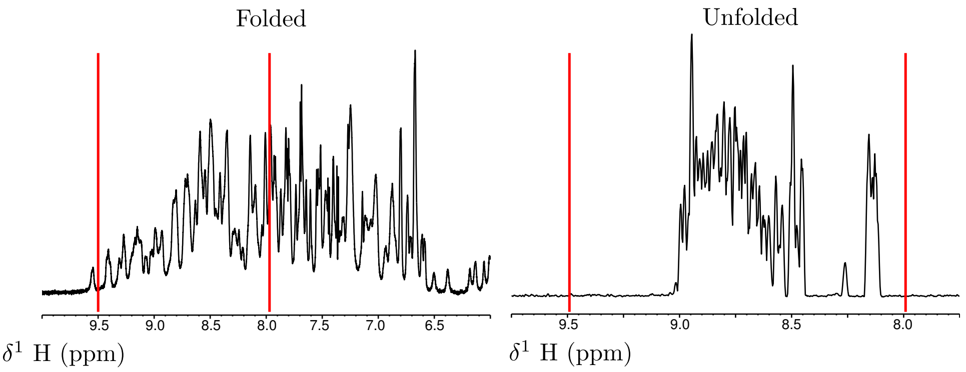


Figure 4.2: 1D spectra of a folded protein (Ubiquitin) and an IDP (NS5A D2D3). The vertical red lines indicate the same spectral region in both cases.

these experiments suffer from intrinsically lower sensitivity. Therefore, proline residues result in a breakpoint in the sequential assignment procedure.

4.3 NMR observables for IDPs

Once the difficulties discussed in the previous section are overcome and resolution and sensitivity are appropriate for undertaking further studies, a variety of NMR observables can be obtained that are useful for characterization of IDPs in terms of structure and dynamics. These observables will be briefly introduced in this section.

4.3.1 Chemical shifts

Chemical shifts are sensitive to the local chemical environment of the nuclei, as discussed in section 4.1.4. Therefore, structural information can be extracted from differences of chemical shifts.

Proton chemical shifts can be calculated by taking into account electric field effects, bond magnetic anisotropy and chemical shifts changes due to ring current effects. Hydrogen bonding affects the bond magnetic anisotropy. In addition to ^{1}H, also ^{15}N chemical shifts are very sensitive to hydrogen bonding of the amide proton. Buffer effects and temperature influence these chemical shifts, especially for solvent exposed residues, because of perturbations of hydrogen bonding within the protein as well as with the solvent water molecules. ^{1}H and ^{15}N chemical shifts change linearly with temperature as long as there are no major conformational changes, and the temperature coefficients are usually independent of the pH too. Thus, temperature coefficients provide complementary information about hydrogen bonding, in addition to hydrogen exchange rates [54].

Heteronuclear chemical shifts can be described as the sum of the random coil chemical shift and the secondary chemical shift ($\Delta\delta$)resulting from increased sampling of a certain conformation with respect to random coil. $^{13}C^{\alpha}$ chemical shifts are very sensitive to increased α-helical sampling, and slightly less, but also sensitive to excess of β-conformations.

For calculation of secondary chemical shifts, the influence of the environment on the chemical shift, such as buffer composition, temperature and pH, should be taken into account. Chemical shifts have also been successfully used for structure calculation of folded molecules [55]. They are also implemented as restraints in ensemble calculation programs for IDPs that will be discussed in section 4.3.5, in ENSEMBLE as well as in Flexible Meccano.

Secondary chemical shifts give information on the backbone conformational preferences of the residue i as they are dependent on the backbone dihedral angles. The chemical shifts of the backbone nuclei are analyzed with respect to random coil chemical shifts, termed perhaps more accurately statistical coil chemical shifts. Random coil chemical shifts are the theoretical chemical shifts of a random coil polypeptide of the same amino acid sequence as the studied protein. In order the understand what these chemical shifts mean, first we have to understand what is a random coil. A random coil is characterized by lack of long range, meaning further than the neighboring amino acids, cooperative secondary structure stabilization effects. The conformational sampling of a polypeptide is never random in the sense that all dihedral angles would be adopted with equal probability, because of steric effects and chemical interactions between the neighboring sidechains. The distribution of the sampled dihedral angles is determined by the conformational Gibbs free energy hypersurface, that will, in turn, depend on pH, temperature and solvent effects, such as ionic strength. Therefore, the term statistical coil is probably more accurate. Nevertheless, in this thesis the term random coil is used, because it is more commonly used in the literature.

Accuracy of secondary chemical shift measurements depends on the accuracy of random coil chemical shifts. For understanding the models used for calculating random coil chemical shifts, first let us discuss how the concept of random coil developed. This concept has been discussed thoroughly in the review by Smith *et al.* [56]. In polymer chemistry the first model concerning the distribution of conformations of a polymer was the random flight model. In this model, the monomer units are regarded as segments and the motions of these segments are uncorrelated the length of a segment being the unit length and all spherical angles were adopted with equal probabilities. This leads to a gaussian distribution of radii of gyration. This model has been refined by Flory and co-workers, by taking into account the excluded volume effect [57] and the energetics of the different conformations [58]. Flory defined the random coil for polypeptides such, that the ϕ, ψ pairs of angles for a given residue are sampled independently from the conformations of all other residues (isolated pair hypothesis). This definition, however, is disputed and has given rise to several further studies into properties of denatured states and short peptides [59].

Protein regions that have no secondary or tertiary structure, as common in intrinsically disordered proteins sample the dihedral angles characteristic for α-helical, β-sheet and PPII conformations as determined by their conformational Gibbs free energy hypersurface, but the sampling of residue i is not correlated to the sampling of the neighboring residues. The random coil model used for calculation of the random coil chemical shifts thus has a major effect on the accuracy of the extracted secondary structural propensities. An ideal random coil model would take into account all the neighboring residue effects resulting from the chemical nature of the polypeptide chain.

There are two major types of models that allow calculation of random coil, or statistical coil, chemical shifts. One is based on chemical shift measurements on small peptides. These models have the advantage of exact control of the effects of temperature and pH, but they are perturbed by other artefacts, resulting of the bias introduced by selections of the residue types for the model [60]. The most commonly used data set was measured by Wishart *et al.* [61]. Recently, temperature and pH correction factors have been introduced from a peptide based study, where instead of taking glycine based peptides as Wishart *et al.*, glutamine peptides were used, as their conformational sampling is considered to be more representative [62]. The other type of reference data can be obtained from chemical shifts from databases. This approach has the disadvantage that pH and temperature vary between the different entries, furthermore, the neighbor residue and other chemical effects will be also biased depending on the composition of entries. The most commonly used one is refDB by Zhang *et al.* [60]. Recently, an IDP chemical shift based random coil database has also been introduced [63]. Neighbor corrections are usually required to account for changes in conformational sampling because of steric clashes and other chemical effects such as electrostatics or ring current effects from the neighbors [61, 64, 65].

A detailed analysis of secondary chemical shifts of the NS5A D2D3 protein is presented in section 8.5.

4.3.2 Residual dipolar couplings

Residual dipolar couplings (RDCs) are measured in an alignment medium in order to reintroduce anisotropy, as in isotropic media the dipolar couplings average out. RDCs are sensitive to local conformations, as they depend on the angle the vector connecting the two nuclei encloses with the static magnetic field. Because of this property RDCs are also used to determine relative orientations of domains, for instance. RDCs can be calculated from structural models and therefore are useful as structural restraints. RDCs have been used for the first time for calculation of a structural ensemble by Bernado *et al.* [66] and shown to be sensitive to the conformational sampling of the residue and its neighbors.

It has been shown that RDCs are sensitive to residual α-helical regions and β-turn regions in IDPs. However, with a combination of a larger set of RDCs and chemical shifts it is also possible to determine populations of all the secondary structural conformations of the Ramachandran space, including the PPII which is degenerate when only chemical shifts or only RDCs are used [67].

4.3.3 ^{15}N relaxation

NMR methods give access to information on the dynamics of a macromolecule on a large range of timescales from ps timescale motions to motions on timescales up to seconds or eventually even longer times. ^{15}N relaxation measurements are informative on fast correlated motions present in the macromolecule from the ps up to ns timescales and on slower motions on the μs to ms timescales as well. They also yield information on the rotational tumbling correlation time, with values up to a few tens on nanoseconds [43].

For IDPs, ^{15}N relaxation measurements are conducted routinely as part of their structural characterization. Usually three parameters are measured: T_1 longitudinal relaxation time constant, T_2 transverse relaxation time constant and ^{1}H-^{15}N heteronuclear NOEs. They are informative on local structuring, rigidity and, eventually, chemical exchange phenomena.

Relaxation in NMR is caused by oscillating magnetic fields due to the stochastic rotating motions of the macromolecule and non-oscillating fields, that contribute to dephasing of the coherence resulting in T_2 relaxation. The distribution of the frequencies of the created oscillating magnetic fields depends on the stochastic motions of the molecule and is described by the power spectral density function $J(\omega)$, that measures the contribution of a motion with frequency components in the range of ω and $\omega + \mathrm{d}\omega$ to the dynamics of the molecule. The power spectral density function is the Fourier transform of the autocorrelation function of the stochastic motions. Assuming the simplest case of an autocorrelation function of an exponential decay $C = \exp|-t/\tau_C|$ the power spectral density function equals its Fourier transform

$$J(\omega) = \frac{2\tau_C}{1 + \tau_C^2\omega^2} \tag{4.12}$$

Motions on the timescale of the spin transitions contribute to relaxation. Chemical shift anisotropy and dipolar interactions dominate the relaxation of $I = 1/2$ spins. In a scalar coupled two spin system (^{1}H-^{15}N), relaxation is determined by the spectral densities at five frequencies, $J(0)$, that corresponds to the non-oscillating fields and $J(\omega_H)$, $J(\omega_N)$, $J(\omega_H + \omega_N)$, and $J(\omega_H - \omega_N)$ corresponding to the transition frequencies of the scalar coupled two spin system. The most commonly measured three relaxation parameters expressed as function of the power spectral densities read

$$\mathrm{NOE} = 1 + \frac{d^2}{20}\frac{\gamma_H}{\gamma_N}[6J(\omega_H + \omega_N) - J(\omega_H - \omega_N)]T_1 \tag{4.13}$$

$$\frac{1}{T_1} = \frac{d^2}{20}[J(\omega_H - \omega_N) + 3J(\omega_N) + 6J(\omega_H + \omega_N)] + \frac{c^2}{15}J(\omega_N) \tag{4.14}$$

$$\frac{1}{T_2} = \frac{d^2}{40}[4J(0) + J(\omega_H - \omega_N) + 3J(\omega_N) + 6J(\omega_H) + 6J(\omega_H + \omega_N)] + \frac{c^2}{90}[4J(0) + 3J(\omega_N)] + R_{ex} \tag{4.15}$$

where d is the dipolar coefficient $d = (\mu_0/4\pi)\gamma_H\gamma_N(h/2\pi)r_{NH}^{-3}$ and $c = \omega_N(\sigma_{\|} - \sigma_{\perp})$ is the ^{15}N chemical shift anisotropy.

The standard analysis of relaxation data for globular proteins uses the model-free formalism introduced by Lipari and Szabo [68] where overall motion is separated from internal motions. This separation is justified by the difference of the time scales of these motions in case of globular proteins. However, as pointed out in their paper, this model is not rigorous for random coil-like polymers. However, calculation of a pseudo-τ_c from the T_1/T_2 ratio according to ref. [69] gives a quick way of obtaining information on the local rigidity.

For IDPs, the reduced spectral density mapping approach [70] can be used to sample the spectral density function. The method exploits the near 10-fold difference in the gyromagnetic ratios of ^{1}H and ^{15}N, and substitutes $J(\omega_H)$, $J(\omega_H - \omega_N)$ and $J(\omega_H + \omega_N)$ with an effective $J(0.87\omega_H)$ value.

$$\mathrm{NOE} = 1 + \frac{d^2}{20} \cdot (\gamma_H/\gamma_N)[5J(0.87\omega_H)]T_1 \tag{4.16}$$

$$1/T_1 = \frac{d^2}{20} \cdot [3J(\omega_N) + 7J(0.87\omega_H)] + \frac{c^2}{15}J(\omega_N) \tag{4.17}$$

$$\frac{1}{T_2} = \frac{d^2}{40}[4J(0) + 13J(0.87\omega_H) + 3J(\omega_N) + \frac{c^2}{90}[4J(0) + 3J(\omega_N)] + R_{ex} \tag{4.18}$$

The validity of the approach has been recently studied for very flexible IDPs and a slightly optimized protocol was suggested including removal of exchange contributions by measurement of cross-correlated relaxation rates [71].

Recently, measurement of cross-correlated relaxation rates was suggested for IDPs together with introduction of a new 3D pulse sequence for their measurement optimized for IDPs to resolve spectral overlap [72]. They can be helpful for distinguishing β-turn type I and type II conformations and PPII helices that do not have a characteristic chemical shift signature.

The analysis of relaxation data obtained during this thesis is presented in section 8.6.

4.3.4 Paramagnetic relaxation enhancement

Attachment of a paramagnetic label on a residue results in broadening due to the paramagnetic relaxation enhancement (PRE) effect. The effect is proportional to the inverse sixth power of the distance between the observed nucleus and the spin label. Usually the MTSL label is used. A major limitation of this technique for IDPs is that a time and ensemble average of the effect is observed. In other words, the same PRE effect is produced between two residues whether they are very close in space for a very short fraction of time or at a medium distance over a long period of time. Because of this, a rigorous analysis of PRE data for IDPs is not straightforward [73]. However, useful information has been extracted from PREs for a large number of IDPs by comparing the PRE profile to that expected for a random coil. In a random coil, the effect is expected to extend to about 15 residues. Observable effects further than 15 amino acids from the label indicate contacts of the spin label with the affected region.

4.3.5 Ensemble modeling

As IDPs cannot be characterized by a single representative structure, an ensemble of structures that collectively fit the experimental data is frequently used to characterize them. However, the full solution to their structural description would probably be the determination of their conformational energy landscape, as this would give full information on the transition states, on the interconversion rates, furthermore the entropic and enthalpic contributions of the different states. Such models are currently not available, but steps are taken towards them by replica-averaged molecular dynamics simulations. The main limitation of this technique is the size of the molecule, as the simulations are computationally expensive.

The more frequently used and computationally much more affordable methods are based on Monte Carlo simulations, without size limitations. Currently two programs are publicly available, Flexible Meccano [66,67] (for creation of the ensemble of structures for the random coil pool) combined with ASTEROIDS [74] (for selection of conformers from the pool that fit the experimental data, not yet available publicly) and ENSEMBLE (ensemble creation and selection is included) [75], [76]. Most of the NMR observables, such as chemical shifts, residual dipolar couplings and paramagnetic relaxation enhancement data, can be used as input in the selection. ^{15}N T_2 relaxation time constants can be used in the ensemble modeling software

ENSEMBLE. Complementary to NMR data SAXS data can be used, and in ENSEMBLE hydrodynamic radii as well.

As mentioned before, the obtained ensemble of structures can be deposited in the pE-DB database, allowing to extend the range of bioinformatic studies of IDPs.

Chapter 5

Longitudinal relaxation optimization for IDPs

As discussed in Chapter 4, NMR spectroscopy of biomolecules is often troubled by sensitivity issues. Recent advances in hardware development, that have greatly increased available sensitivity over the past decades, have also been discussed therein. Further developments have been made by the design of optimized pulse sequences, as well as introduction of new improved data acquisition and data sampling techniques, that allow to increase experimental sensitivity. In this chapter, longitudinal relaxation optimization techniques are going to be discussed. Further techniques will be mentioned, some of which can be combined with longitudinal relaxation optimization.

5.1 Sensitivity and longitudinal relaxation

The minimum duration of a multidimensional NMR experiment is determined by the number of scans that have to be acquired to complete the phase cycling, and by the number of increments in the indirect dimension(s) required to reach sufficient resolution. In addition, if the signal to noise ratio is not sufficient, further increase of the number of scans might be necessary.

As illustrated by Figure 5.1, the duration of a scan is the sum of the duration of the pulse sequence itself, including all the pulses and the delays (usually about 100 ms), the acquisition time and the recovery delay. For the most frequently employed ^{1}H-detected experiments the duration of the acquisition time is usually about 70-100 ms. Its duration depends on the required resolution and is limited by the ^{1}H T_2 relaxation time constant. For the less common, but for IDPs often employed ^{13}C-detected experiments the acquisition time can be longer, because of the typically larger ^{13}C T_2 relaxation time constants. The following recovery delay is required for relaxation of the system, to restore sufficient polarization to restart the experiment. Its duration depends on the longitudinal relaxation time constant T_1 that

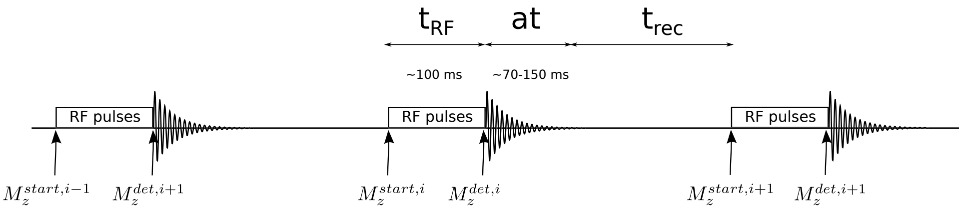


Figure 5.1: Schematic NMR experiment, with major times to be considered in view of the duration of the experiment.

will be introduced in detail in section 5.2. In the following, the dependence of the experimental sensitivity on the recovery delay and on the excitation angle will be discussed. The amount of magnetization (or the polarization) available at the beginning of the scan determines the signal intensity that can be detected. The detected magnetization is in the xy plane. Neglecting relaxation losses and imperfections, the effect of the pulse sequence on the magnetization can be considered as an overall rotation of the M_z magnetization by an angle β. Thus the magnetization in the xy plane at the end of the pulse sequence is $M_0 \sin\beta$, while the magnetization left along the z axis is $M_0 \cos\beta$. However, at the beginning of the NMR experiment dummy scans are applied to ensure steady-state conditions in order to avoid spectral artifacts. The magnetization along the z axis at the beginning of pulsing is denoted with M_z^{start} and M_z^{det} at the beginning of the detection at scan i, and $M_z^{start,ss}$ and $M_z^{det,ss}$ in the steady state, respectively. As shown in a review of fast pulsing techniques by Schanda [77] we can write the Bloch equation for the longitudinal magnetization of the system as

$$M_z^{start,i} = M_z^0 - \left(M_z^0 - M_z^{det,i-1} \exp\left(-\frac{at + t_{rec}}{T_1} \right) \right). \tag{5.1}$$

In the steady state

$$M_z^{start,i-1} = M_z^{start,i} = M_z^{start,ss} \tag{5.2}$$

and

$$M_z^{det,i-1} = M_z^{det,i} = M_z^{det,ss} \tag{5.3}$$

and taking into account that $M_z^{det,ss} = \cos\beta M_z^{start,ss}$, the longitudinal magnetization at the end of the pulse sequence is

$$M_z^{start,ss} = M_z^0 \frac{1 - \exp(-T_{rec}/T_1)}{1 - \exp(-T_{rec}/T_1)\cos\beta}. \tag{5.4}$$

In case of $\beta = 90\,^\circ\mathrm{C}$ the expression simplifies to

$$M_z^{start,ss} = M_z^0 \left(1 - \exp\left(-\frac{T_{rec}}{T_1} \right) \right). \tag{5.5}$$

The SNR is proportional to the transverse magnetization available at the start of detection and scales with the square-root of the number of scans. This can be summarized as

$$\mathrm{SNR} \propto \frac{(1-\exp(-T_{rec}/T_1))\sin\beta}{1-\exp(-T_{rec}/T_1)\cos\beta}\frac{1}{\sqrt{T_{scan}}}. \tag{5.6}$$

The maximum of the expected SNR ($\mathrm{SNR}_{\max}$) can be calculated by differentiating the above expression with respect to T_{rec} for a given pulse sequence length.

$$\mathrm{SNR}_{\max} \propto 0.71/\sqrt{T_{rec}^{opt}} \sim 0.64\sqrt{1/T_1} \tag{5.7}$$

Rearranging this expression yields

$$T_{rec}^{opt} \sim 1.25T_1 \tag{5.8}$$

that gives the approximate value of the optimal recycle delay that should be used in order to reach maximum sensitivity. It has been shown by Ernst and Anderson [78] that the maximum signal intensity for a given recovery delay can be expected at the *Ernst angle* β_{Ernst} in case of a single pulse experiment (Eq. 5.9)

$$\cos\beta_{Ernst} = \exp(-T_{rec}/T_1) \tag{5.9}$$

Implementation of Ernst angle excitation in multidimensional experiments is not straightforward, as application of multiple pulses usually does not leave the magnetization remaining along the z axis after the first pulse. However, in the HMQC sequence its application is possible, as has first been proposed in the Fast-HMQC experiment [79]. It is also exploited in the SOFAST-HMQC technique [80], that combines it with selective pulsing, as will be discussed in section 5.4.

5.2 Mechanisms of longitudinal relaxation

Longitudinal relaxation is also termed spin-lattice relaxation. It is a non-adiabatic process, the energy emitted or absorbed by the transition is absorbed by the lattice, which can be considered to have infinite heat capacity. It is caused by transverse fluctuating magnetic fields generated by the interaction of the lattice with the biomolecule that are determined by the stochastic motions of the biomolecule. Longitudinal relaxation restores the initial spin polarization and returns the system to thermal equilibrium. The main mechanism governing longitudinal relaxation is the dipolar interaction of the relaxing spin with the surrounding spins. For a basic theoretical treatment of longitudinal relaxation we will consider a dipolar coupled system of two spins, that have four energy eigenstates, as illustrated in Figure 5.2.

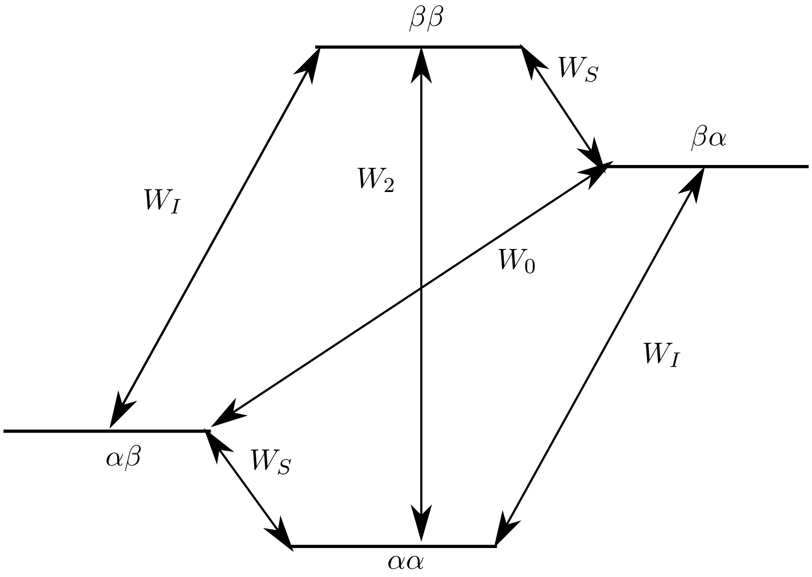


Figure 5.2: Transitions in a two spin system

There are four possible transition paths between the energy levels. W_0 denotes the transition where both spins are flipped in opposite senses (flip-flop transition), W_I is the one where spin I is flipped, W_S is the one where spin S is flipped and in the course of W_2 both spins are flipped in the same senses (flip-flip transition). The angular frequencies corresponding to the energies required for these transitions (generated by the fluctuating fields) can be read from the energy diagram. They are $\omega_I - \omega_S$ for W_0, ω_I for W_I, ω_S for W_S and $\omega_I + \omega_S$ for W_2, respectively. Thus, transitions at these frequencies depend on the magnitude of the power spectral density of the stochastic fluctuating fields at these angular frequencies. The concept of power spectral density is introduced in detail in standard textbooks [47] and spectral densities will be discussed in Part III when presenting dynamics studies of NS5A.

The time dependence of the populations can be described using the defined transition probabilities by inspection of the figure, such as

$$\frac{\mathrm{d}P_{\alpha\alpha}}{\mathrm{d}t} = W_S(P_{\alpha\beta} - P_{\alpha\alpha}) + W_I(P_{\beta\alpha} - P_{\alpha\alpha}) + W_2(P_{\beta\beta} - P_{\alpha\alpha}) + \text{const.}, \quad (5.10)$$

and similarly for the other eigenstates. The value of const. can be determined using the boundary condition of constant population in equilibrium, $\frac{\mathrm{d}P_{\alpha\alpha}}{\mathrm{d}t} = 0$. From this follows, that the differential equation governing the time dependence of the deviations from the equilibrium populations can be written as

$$\frac{\mathrm{d}\Delta P_{\alpha\alpha}}{\mathrm{d}t} = W_S(\Delta P_{\alpha\beta} - \Delta P_{\alpha\alpha}) + W_I(\Delta P_{\beta\alpha} - \Delta P_{\alpha\alpha}) + W_2(\Delta P_{\beta\beta} - \Delta P_{\alpha\alpha}), \quad (5.11)$$

where the $\Delta P_{\gamma\delta}$s are the deviations of the populations from equilibrium. From this equation, the relation to the I_z and S_z macroscopic magnetizations can be seen, as they are determined by populations of the different eigenstates multiplied by the trace of their spin density matrices (Eq. 5.12).

$$I_z = \frac{1}{2}\begin{bmatrix}1 & 0 & 0 & 0\\ 0 & 1 & 0 & 0\\ 0 & 0 & -1 & 0\\ 0 & 0 & 0 & -1\end{bmatrix}, \quad S_z = \frac{1}{2}\begin{bmatrix}1 & 0 & 0 & 0\\ 0 & -1 & 0 & 0\\ 0 & 0 & 1 & 0\\ 0 & 0 & 0 & -1\end{bmatrix} \tag{5.12}$$

Thus the deviation of the I_z and S_z magnetizations from their equilibria are $\Delta I_z = \Delta P_{\alpha\alpha} + \Delta P_{\alpha\beta} - \Delta P_{\beta\alpha} - \Delta P_{\beta\beta}$ and $\Delta S_z = \Delta P_{\alpha\alpha} - \Delta P_{\alpha\beta} + \Delta P_{\beta\alpha} - \Delta P_{\beta\beta}$. From this and from Eq. 5.11 follow the Solomon equations for a two spin system:

$$\begin{aligned}\frac{\mathrm{d}\Delta I_z}{\mathrm{d}t} &= -\rho_I \Delta I_z - \sigma_{IS}\Delta S_z,\\ \frac{\mathrm{d}\Delta S_z}{\mathrm{d}t} &= -\rho_S \Delta S_z - \sigma_{IS}\Delta I_z,\end{aligned} \tag{5.13}$$

where

$$\begin{aligned}\rho_I &= W_0 + 2W_I + W_2,\\ \rho_S &= W_0 + 2W_S + W_2,\\ \sigma_{IS} &= W_2 - W_0.\end{aligned} \tag{5.14}$$

A major conclusion from the Solomon equations of the two spin system is that the longitudinal relaxation of spin I depends not only on the deviation of the spin itself from its equilibrium polarization ΔI_z, but also on the deviation of spin S from its equilibrium polarization ΔS_z. The first effect is termed auto-relaxation, and the corresponding relaxation rate constants are denoted by ρ_I and ρ_S. The second one is termed cross-relaxation, and its rate constant is denoted by σ_{IS}. Thus, even in the case of two spins, longitudinal relaxation is not necessarily a monoexponential process. The Solomon equations can be extended to the case of n spins. The time evolution of the longitudinal magnetization of spin I_n surrounded by $n-1$ spins I_k depends on its auto-relaxation rate constant and its cross relaxation rate constants with all the surrounding $n-1$ spins. This relation can be described by the matrix form of the Solomon equations as

$$-\frac{\mathrm{d}}{\mathrm{d}t}\begin{pmatrix}\Delta I_{1z}\\ \Delta I_{2z}\\ \vdots\\ \Delta I_{nz}\end{pmatrix} = \begin{pmatrix}\rho_1 & \sigma_{12} & \cdots & \sigma_{1n}\\ \sigma_{21} & \rho_2 & \cdots & \sigma_{2n}\\ \vdots & \vdots & \ddots & \vdots\\ \sigma_{n1} & \cdots & \cdots & \rho_n\end{pmatrix}\begin{pmatrix}\Delta I_{1z}\\ \Delta I_{2z}\\ \vdots\\ \Delta I_{nz}\end{pmatrix} \tag{5.15}$$

where

$$\rho_n = \sum_{n\neq j}\rho_{nj} \tag{5.16}$$

In case of biomolecules the main mechanism governing longitudinal relaxation is the dipolar relaxation mechanism. As shown in standard textbooks, the auto- and cross-relaxation rate constants of the IS two spin system depend on the spectral densities at the transition frequencies, the distance of the two spins (r) and their gyromagnetic ratios (γ_I and γ_S)according to the following relations:

$$\begin{aligned} \rho_I &= (d^2/4)J(\omega_I - \omega_S) + 3J(\omega_I) + 6J(\omega_I + \omega_S) \\ \rho_S &= (d^2/4)J(\omega_I - \omega_S) + 3J(\omega_S) + 6J(\omega_I + \omega_S) \\ \sigma_{IS} &= (d^2/4) - J(\omega_I - \omega_S) + 6J(\omega_I + \omega_S) \end{aligned} \tag{5.17}$$

where

$$d = (\mu_0/4)\hbar\gamma_I\gamma_S r_{IS}^{-3}. \tag{5.18}$$

It is very important to note that the relaxation of the spins depends on the inverse sixth power of the distance between the two spins. It can be seen from the Solomon equations, that selective excitation of the spin of interest will affect recovery of the spin magnetization because of cross-relaxation to the nearby spins. If $\sigma_n, k < 0$, as it is the case for macromolecules with tumbling correlation times in the ns range, selective excitation of a set of spins speeds up their longitudinal relaxation with respect to non-selective excitation [81], [77]. This explains the underlying principle of longitudinal relaxation optimized techniques for macromolecules, that is: selective excitation of the spins of interest speeds up their longitudinal relaxation. In case of proteins, usually a 2D amide ^{1}H-^{15}N correlation spectrum is recorded in order to provide site-resolved information. Fortunately, the amide protons resonate at a different frequency range than aliphatic protons, so their selective excitation is feasible, and thus their longitudinal relaxation can be enhanced.

A further very important relaxation mechanism for longitudinal relaxation optimized techniques occurs via chemical exchange of water protons with exposed exchangeable protons. These exchangeable protons can be the observed amide protons, or can be labile sidechain protons. As a result of the exchange, the exchangeable proton directly assumes the magnetization of water. Longitudinal relaxation of water is rather slow compared to that of biomolecules, T_1 is on the range of 2-4 s under the conditions used in biomolecular NMR spectroscopy. Because of spin diffusion, even in case of globular proteins because of the exposed side chains, this mechanism can contribute significantly to longitudinal relaxation of the system [77]. In case of IDPs, this mechanism is very efficient, as all of the amide protons are solvent exposed at least for some part of the time. When approaching physiological conditions, amide exchange with water is becoming very fast.

The Solomon equation of the n-spin system can be expanded with the magnetization of water to include magnetization originating from solvent exchange, this is called the Bloch-McConnell equation.

$$
-\frac{\mathrm{d}}{\mathrm{d}t}\begin{pmatrix} I_W - I_W^0 \\ I_{1z} - I_{1z}^0 \\ I_{2z} - I_{2z}^0 \\ \vdots \\ I_{nz} - I_{nz}^0 \end{pmatrix} = \begin{pmatrix} \rho_W & 0 & 0 & 0 & 0 \\ 0 & \rho_1 & \sigma_{12} & \cdots & \sigma_{1n} \\ 0 & \sigma_{21} & \rho_2 & \cdots & \sigma_{2n} \\ 0 & \vdots & \vdots & \ddots & \vdots \\ 0 & \sigma_{n1} & \cdots & \cdots & \rho_n \end{pmatrix} \begin{pmatrix} I_W - I_W^0 \\ I_{1z} - I_{1z}^0 \\ I_{2z} - I_{2z}^0 \\ \vdots \\ I_{nz} - I_{nz}^0 \end{pmatrix} + \\
\begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ -k_{ex,1} & k_{ex,1} & 0 & 0 & 0 \\ -k_{ex,2} & 0 & k_{ex,2} & 0 & 0 \\ \vdots & 0 & 0 & \ddots & 0 \\ -k_{ex,n} & 0 & 0 & 0 & k_{ex,n} \end{pmatrix} \begin{pmatrix} I_W \\ I_{1z} \\ I_{2z} \\ \vdots \\ I_{nz} \end{pmatrix} \quad (5.19)
$$

The first part of the equation is the same as in Eq. 5.15, but it is extended with the magnetization of the bulk water, that is considered to relax with the auto-relaxation rate constant ρ_W and does not cross-relax with the protein. The reason for not including cross-relaxation in the equation is that the residence time of water around the protein molecule has been shown to be on the subnanosecond time scale [82]. The second term on the right hand side of the equation takes into account the chemical exchange rates of the solvent exposed protons with the bulk water. It will be shown in the followings that solvent exchange with water can be the main factor determining the longitudinal relaxation enhancement that can be achieved by selective pulsing for IDPs.

5.3 Selective excitation

As discussed in the previous section, selective excitation of the observed spins results in efficient enhancement of their longitudinal relaxation. There are two viable techniques applied for selective excitation.

The first is that excitation is performed with non-selective pulses, and unobserved protons are flipped back, using their different J-couplings. For example, when the observed spins are coupled to ^{15}N, all other aliphatic spins that are coupled to ^{13}C, or not coupled to heteroatoms (like in water) are flipped back. Examples of such sequences are shown in references [81,83]. However, the drawback of these techniques is that, as excitation of the flipped back spins takes place before flipping them back, they are involved in coherence transfer pathways, during which T_2 relaxation takes place and therefore their flip-back is never complete. This approach has been applied for example to the implementation of the 'H-flip'-type of longitudinal relaxation optimized ^{13}C-detected experiments [84].

The second technique, introduced first by Pervushin *et al.* [81] and later by Schanda and Brutscher [80], is to apply selective excitation of the observed proton spins, when their frequency separation makes it possible, by use of band-selective pulses. The disadvantage of band-selective pulsing can be that certain amino acid sites of interest can have chemical shifts outside of the excitation bandwidth of the selective pulse, for example due to ring-current effects resulting in extreme chemical shifts. They are not excited by the band-selective pulses, therefore they do not show up in the spectrum. This problem is rarely encountered for IDPs, as they are characterized by a very narrow spectral dispersion in the ^{1}H dimension (4.2.2).

Ideally, the selective pulses used in longitudinal relaxation optimized experiments should have clean, so-called “top hat” profiles (such as shown in Figure 6.14), meaning that for spins resonating in the selected frequency range the excitation or inversion should result in the desired effect, while for spins lying outside of this bandwidth the effect of the pulse should be zero. The pulses can be classified into excitation, flip-back, refocusing and inversion pulses. Some shaped pulses have been optimized for a particular purpose, eg. inversion ($I_z \longrightarrow -I_z$). These pulses optimized for a particular purpose do not yield clean profiles for other purposes.

For the BEST experiments, discussed in this thesis, PC9 pulses are used for excitation, REBURP for refocusing, and EBURP-2 for flip-back. For easy implementation of selective pulses in experiments, a binary scheme has been proposed [85], that shows for which fraction of a selective pulse scalar coupling and chemical shift evolution are active or inactive. A special feature of the PC9 pulse is that is also performs well for excitation angles other than $90\,^\circ$, a feature exploited in the SOFAST-HMQC experiment discussed in the next section.

5.4 The SOFAST-HMQC experiment

In section 5.2, we discussed the underlying principles of longitudinal relaxation optimization. The SOFAST-HMQC experiment [80, 86] is a technique capable of yielding fingerprint $^1\mathrm{H} -^{15}\mathrm{N}$ correlation spectra in a few seconds. The experiment is based on the combination of two techniques of longitudinal relaxation optimization: Ernst angle excitation and selective pulsing on amide protons.

Ernst angle excitation has been originally introduced for a single pulse experiment. It cannot be implemented in 2D spectroscopy employing INEPT blocks, because the magnetization left along the z axis as a result of partial excitation is flipped to the xy-plane by the concluding $\pi/2$ pulse of the INEPT block. The HMQC experiment, however, does not include INEPT blocks, the sequence employs only two proton pulses. Therefore, implementation of Ernst-angle excitation is possible, as shown by Ross and co-workers [79]. The longitudinal magnetization after the excitation pulse (of angle α)is flipped by $180\,^\circ$ by the refocusing proton π pulse, so the magnetization along the z-axis after the π pulse is proportional to $\cos \beta_{Ernst}$ where $\beta_{Ernst} = 180^\circ - \alpha$.

The second source of sensitivity gain results from enhancement of longitudinal relaxation of the amide proton spins due to selective pulsing. As Ernst angle excitation is used, the 90° selective proton pulse has to be robust enough against change of the excitation angle. The polychromatic PC9 pulse keeps a clean "top hat" excitation profile for a wide range of excitation angles and therefore is applied in the SOFAST-HMQC experiment. The second proton pulse was applied originally with an r-SNOB pulse shape [86], but shortly afterwards it was shown by the same authors that the REBURP pulse shape performs better [80]. Sensitivity curves show, that in case of need for fast pulsing, the sensitivity gain reached by the SOFAST-HMQC experiment over conventional experiments is dramatic, but even in the optimal delay regime there is a significant sensitivity gain while using the SOFAST experiment

Since its introduction in 2005, the SOFAST-HMQC experiment has found many applications. It has become routinely employed for quick recording of fingerprint spectra of proteins, it is often used for studies in the field of in-cell NMR spectroscopy [87], its methyl-selective version is used for studies of large protein assemblies [88] and it is employed in real-time NMR studies of protein folding [89] and kinetics studies of post-translational modifications either in-cell or *in vitro*.

The disadvantage of the SOFAST-HMQC experiment, especially for IDPs, is the relatively low resolution that can be achieved. The reason for this is that only relatively short t_1 periods can be applied, because relaxation of multiple quantum coherences (eg. H^+N^+) is fast, therefore the resolution in the indirect dimension is limited.

5.5 The BEST experiments

In the previous section, we discussed advantages of the SOFAST-HMQC experiment for quick acquisition of 2D $^1\mathrm{H} - ^{15}\mathrm{N}$ fingerprint spectra of proteins. The drawback of its relatively low resolution has also been mentioned. The most straightforward way of increasing resolution is to increase dimensionality of the experiment, in order to resolve overlaps. However, extensions of HMQC sequences to higher dimensions would not be advantageous, because their sensitivity would be very low due to fast relaxation of multiple quantum terms. Instead, usually HSQC type sequences are extended to higher dimensions. As already discussed in the previous section, Ernst-angle excitation cannot be advantageously exploited for experiments employing INEPT building blocks. However, selective proton pulsing for longitudinal relaxation optimization has been introduced for 3D sequences shortly after introduction of the SOFAST-HMQC experiment [90]. This class of experiments are termed the BEST-type experiments, BEST standing for Band-selective Excitation Short-Transient. First in ref. [90] the BEST-HNCO and BEST-HNCA experiments have been introduced. Later, the BEST-concept has been extended to a large set of 3D experiments [91], that allow fast acquisition of 3D data for sequential assignment or measurement of spin coupling constants [92].

However, it is not evident how well the BEST-concept is applicable for IDPs, as there is a much lower proton spin density around an amide proton than in globular proteins, as pointed out by Marion [93]. This lower spin density is expected to reduce the efficiency of longitudinal relaxation enhancement achievable by selective pulsing. On the other hand, amide proton exchange with water can speed up longitudinal relaxation, when perturbation of water is avoided. In the same paper, Marion extended the BEST experiment to four dimensions and applied it, combined with non-uniform sampling, for an unfolded state of the HBXIP protein.

5.6 Longitudinal relaxation of IDPs

In order to evaluate the contribution of solvent exchange and of dipolar cross-relaxation with unperturbed aliphatic proton spins for IDPs, we investigated the effect of selective perturbation of amide spins and water flip-back schemes on amide proton longitudinal relaxation.

As IDPs lack stable secondary and tertiary structures, all of their amide protons are exposed to the solvent, thus they are involved in chemical exchange with the bulk water. It has to be noted that the exchange rates are very strongly affected by the sample and measurement conditions. This is something to keep in mind when designing NMR studies of IDPs.

The theoretical exchange rates k_{ex}^{theo} for an unprotected, completely random polypeptide can be predicted by bioinformatic algorithms, for example the SPHERE predictor [94]. As it will be shown in the following, k_{ex}^{theo} depends on the amino acid type of the residue and of the neighboring amino acids, the temperature and the pH. The experimental exchange rates depend on the solvent accessibility of the given site. The protection factor of a given amide site is P, it is used to quantitatively characterize the solvent accessibility of the site.

$$P = \frac{k_{ex}^{theo}}{k_{ex}^{obs}} \tag{5.20}$$

where k_{ex}^{obs} is the experimentally determined exchange rate. $P = 1$ if the amide site is fully exposed and the polypeptide behaves like a random coil. $P < 1$ is not meaningful, it would mean that the amide site is less protected than in a random coil. Protection factors are frequently used in protein folding studies. Sometimes they are also used for characterization of IDPs.

Assuming a transient helix is formed in 50% of the time, one would expect a protection factor $P = 2$, and therefore a reduced exchange rate.

Cross-relaxation to other sites also depends on the compaction of the IDP and on how much time the IDP spends in the compact conformation, or when considering an ensemble, what percentage of the molecules of the ensemble at the given time adopt the compact conformation (ergodicity). These two longitudinal relaxation mechanisms contribute in a complex and complementary manner to relaxation enhancement of amide protons in IDPs.

We studied contributions of these mechanisms to ^{1}H longitudinal relaxation experimentally by inversion recovery experiments on several IDP samples at different conditions. We measured apparent T_1 relaxation time constants (Figure 5.3) with the inversion blocks shown in panel a). We measured the reference proton T_1 time constants by inversion of all proton spins with a hard 180 ° pulse (*black*), denoted in the text by $T_{1,hard}$. In order to measure the contribution of solvent exchange only to longitudinal relaxation enhancement, we used a water-flip-back block for inversion and thereby measured the apparent T_1 relaxation time constants (shown in *green*). This pulse sequence block is based on radiation damping, therefore it requires a cryo-probe to be applicable. A 179 ° non-selective pulse is applied, that leaves some water magnetization in the transverse plane. Radiation damping then flips back the water magnetization very selectively to the z axis during a short delay. Then follows the relaxation delay, during which a low power gradient is applied. The effect of selective inversion of amide protons on the T_1 relaxation time constants is measured with the third version of the experiment (shown in *red*). The resulting apparent T_1 time constants are shown in panel b). The impact of sample conditions for NMR studies of IDPs can be appreciated from the differences in the longitudinal relaxation enhancement mechanisms that can be observed for the four IDPs we studied. It is interesting to estimate the theoretical exchange rates for the different proteins under the experimental conditions used in the NMR experiments. The predictions have been performed with the SPHERE predictor and the results are summarized in Table 5.1 Several observations can be made from the plots concerning the mechanisms of longitudinal relaxation enhancement for IDPs.

The first important observation is that even in case of the highly dynamic BASP1 protein, where conditions do not favor exchange of amide protons, apparent longitudinal relaxation time constants are reduced to an average of $\overline{T}_{1,sel} = 0.31$ s, from $\overline{T}_{1,hard} = 0.91$ s in the non-selective case, while there is only minimal contribution of the exchange with water to longitudinal relaxation enhancement, as it can be seen from the water flip-back $T_{1,\mathrm{WFB}}$ values, that are very similar to the non-selective ones. It is well known, that long range NOEs usually can not be observed in the NOESY spectra of IDPs, but nearby spins give rise to NOE cross peaks. Our data suggests that

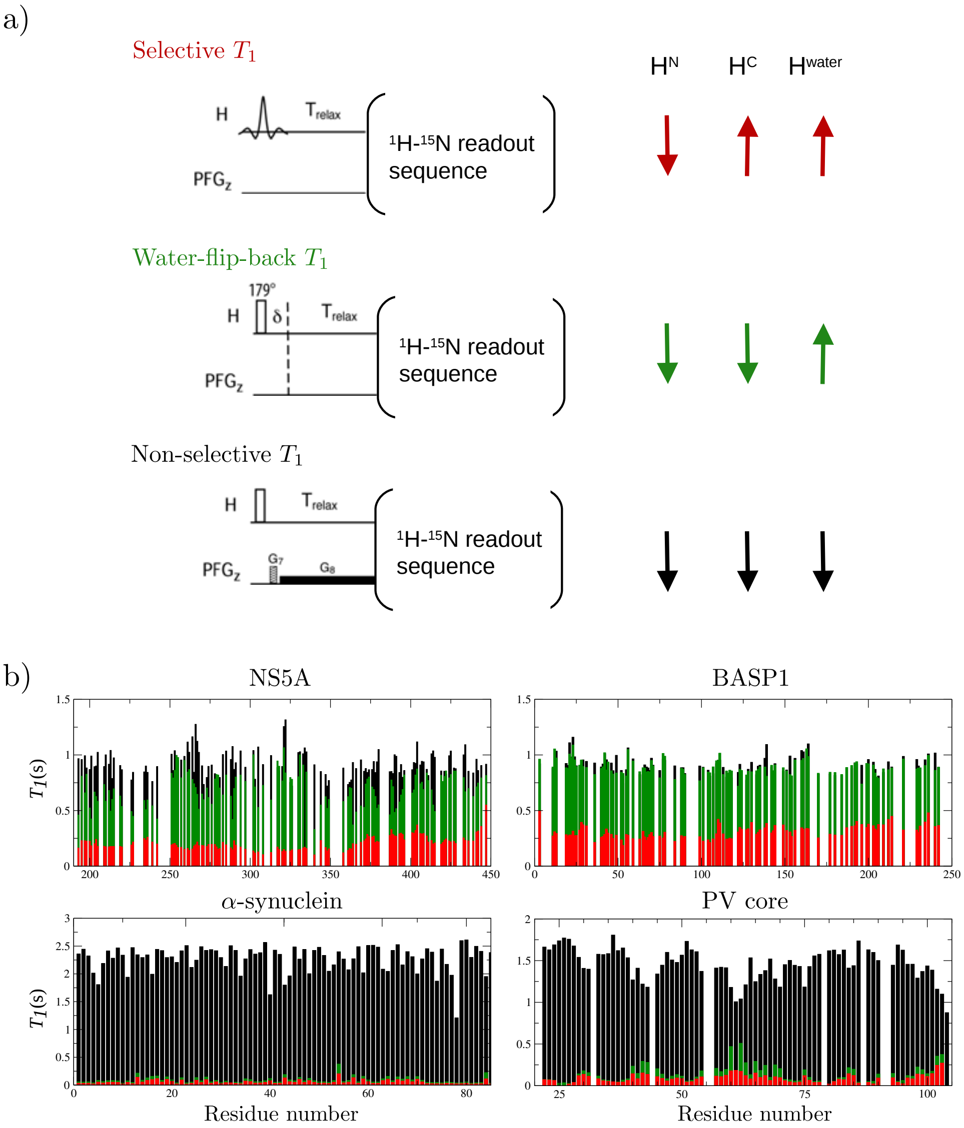


Figure 5.3: Apparent ^{1}H T_1 relaxation time constants measured by inversion recovery experiments. Each sequence consists of a recycle delay of 6s and an inversion pulse followed by a relaxation delay. For readout, a BEST-HSQC sequence was used. The inversion blocks are shown in panel a) for the selective (*red*), water-flip-back (*green*) and non-selective (*black*) scenarios. For amide proton-selective inversion (*red*) an iSNOB-5 pulse shape was applied with a bandwidth of 4.0 ppm and centered at 9.0 ppm; for water-flip-back inversion a 179 ° inversion pulse is followed by a short delay $d \sim$ 10ms during which radiation damping brings the water magnetization back to equilibrium; for non-selective ^{1}H inversion a broad-band inversion pulse is followed by a strong pulsed field gradient to spatially defocus residual water transverse magnetization. A low-power gradient is applied during the entire relaxation delay (T_{relax}) to avoid radiation-damping effects. Panel b) shows the measured apparent T_1 time constants under different sample conditions in corresponding colors. The four samples and the conditions were NS5A D2D3 protein (pH 6.5, 5 °C), BASP1 (pH 6.5, 5 °C), α-synuclein (pH 7.4, 15 °C) and PV core (pH 7.5, 5 °C).

	NS5A	BASP1	α-synuclein	PV Core
Conditions				
pH	6.5	2.0	7.4	7.5
T (°C)	5	5	15	5
Predictions				
range of k_{ex} (s^{-1})	7.6×10^{-2}– $1.2\times10^{+2}$	1.9×10^{-4}– 1.3×10^{-3}	1.8×10^{-1}– $2.1\times10^{+2}$	8.7×10^{-2}– $1.6\times10^{+2}$
$\overline{k}_{ex}$ (s^{-1})	1.5	3.7×10^{-4}	$1.7\times10^{+1}$	$1.1\times10^{+1}$
NMR data				
$\overline{T}_{1,hard}$ (s)	0.92	0.91	2.27	1.46
$\overline{T}_{1,\mathrm{WFB}}$ (s)	0.70	0.89	0.09	0.14
$\overline{T}_{1,sel}$ (s)	0.21	0.31	0.06	0.09

Table 5.1: Sample conditions (pH and temperature) used in our study. Ranges of predicted amide proton solvent exchange rates (k_{ex}) (predictions were performed with SPHERE [94]) and their average over all residues $\overline{k}_{ex}$ are listed as well as the ranges and averages of the apparent T_1 relaxation time constants measured by the sequences shown in Fig. 5.3.

the short range dipolar cross relaxation effects are sufficient for enhancing longitudinal relaxation of the amide spins of the IDP.

Furthermore, for both NS5A and BASP1 the major relaxation enhancement mechanism is dipolar interaction, while this situation changes as exchange becomes faster. In Table 5.1 the predicted exchange rates are listed. For NS5A the average predicted k_{ex} over the residues is 1.5 s^{-1}, while for PV Core it is 11 s^{-1}. This order of magnitude increase of the exchange rates makes solvent exchange the dominant mechanism. The reason can be seen by examining the Bloch-McConnell equation (Eq. 5.19). In the case of NS5A the exchange rates are in the range of the relaxation rates $\overline{R}_{1,hard} = 1/\overline{T}_{1,hard}$. In case of PV Core $\overline{k}_{ex}$ is one order of magnitude larger than $\overline{R}_{1,hard}$, so the dominating rate constant becomes that of the solvent exchange. An important consequence is the increasing average non-selective $\overline{T}_{1,hard}$ relaxation rates that can be observed under conditions that favor rapid solvent exchange. The extreme case is observed for α-synuclein where the $T_{1,hard}$ relaxation time constants are very close to that of water. The measured value of $\overline{T}_{1,hard}$ is 2.27 s, which is very close to the expected value for the relaxation time constant of water that is 2.55 s at 15 °C [95]. This means that when using non-selective pulses, the recycle delay that should be used to reach optimal sensitivity is becoming longer as the conditions come closer to physiological. For PV Core and NS5A, an increase of non-selective $\overline{T}_{1,hard}$ can be observed at the least protected regions.

The last consequence that should be pointed out here is that the enhanced longitudinal relaxation in the case of α-synuclein result in an average selective $\overline{T}_{1,sel}$ of as short as 60 ms, that is a 38-fold reduction. This means that, in case of selective pulsing, the optimal recycle delay is $\sim 1.25 \cdot T_1 = 1.25 \cdot 60\,\mathrm{ms} = 75\,\mathrm{ms}$. On the contrary, in case of non-selective excitation, very long recycle delays of 2.7 s should be used to reach optimal sensitivity. We can also estimate the sensitivity increase that can be associated with selective pulsing, when using optimal recycle delays. As T_1 is reduced 38-fold, the expected $\mathrm{SNR}_{\mathrm{max}}$ for optimized recycle delays is increased $\sqrt{38} \sim 6$-fold, according to Eq. 5.7. Therefore, when near physiological conditions are required for studies of IDPs, the dramatically enhanced longitudinal relaxation of these amide sites by selective pulsing can be exploited by use of short recycle delays to gain sensitivity for the exchange-broadening signals. To summarize, it is very important in NMR spectroscopic studies of IDPs to take care of the exact conditions used, estimate the solvent exchange rates, and set the recycle delay accordingly in order to optimize sensitivity.

5.7 Paramagnetic relaxation enhancement agents

Longitudinal relaxation has also been enhanced by application of paramagnetic agents [96–98]. In the first two references the paramagnetic compound was used for directly enhancing longitudinal relaxation of the nuclei of the protein while in the third the effect was exploited indirectly by enhancing the T_1 relaxation of water, that enhances the relaxation of the protein spins by chemical exchange and cross-relaxation. This principle has also been applied to fast-pulsing experiments on IDPs in two works.

In the first one [99], it has been demonstrated that adding Ni-DO2A paramagnetic relaxation agent to the sample is beneficial even in the case of the SOFAST experiment. Ni-DO2A belongs to the family of paramagnetic relaxation agents that act directly on the longitudinal relaxation rates of amides. They found the relaxation enhancement effect to be greater for residues that exchange slower with water. For faster exchanging residues, they found that relaxation enhancement by selective pulsing is already efficient, thus there is very small additional advantage resulting from the effect of Ni-DO2A.

In the other work, the effect of a gadolinium chelate, Gd(DTPA-BMA), was studied on BEST-type experiments [100]. Gd(DTPA-BMA) enhances the longitudinal relaxation of water and thus it enhances longitudinal relaxation of the protein. Surprisingly, the authors found that even in case of selective pulsing, reducing T_1 of the bulk water resulted in increased sensitivity of the BEST-type experiments. This observation could be explained either by partial excitation of water in spite of selective pulsing, or by direct dipolar cross-relaxation to the bulk water.

We, too, investigated the effect of a paramagnetic agent for the BEST-TROSY fast pulsing experiment in the course of this thesis. The results are presented and discussed in Chapter 6.

5.8 Longitudinal relaxation optimization and ^{13}C detection

^{13}C-detected experiments offer the advantage of larger frequency dispersion per linewidth of the ^{13}CO resonances compared to ^{1}H resonances, that are severely overlapping for most IDPs because of the lack of structuring. As mentioned before, ^{13}C-detected experiments suffer from the inherently lower sensitivity, due to the smaller γ of the ^{13}C nucleus. To partially circumvent this problem, ^{1}H-start versions of the CON and several other carbon-detected experiments have been developed. Longitudinal relaxation optimization has also been applied to these sequences, with the flip-back method termed "H-flip" rather than use of selective pulses [84]. In this experiment, aliphatic protons are flipped back resulting in enhanced longitudinal relaxation of the amide proton spins.

Later, in the course of this thesis and within the framework of a collaboration with CERM (Florence), we implemented an amide proton start version of the sequence with selective proton pulses called H^{N-BEST}CON as shown in Figure 5.4 in order to improve longitudinal relaxation optimization in proton-start ^{13}C-detected experiments. This experiment is similar to a BEST-HNCO, but instead of transferring the magnetization back to the amide proton, the magnetization is detected on carbon and therefore the proton dimension is omitted. This experiment, similarly to the CON yields 2D fingerprint spectra where one correlation can be seen per amino acid.

We recorded sensitivity curves in order to characterize the performance of this experiment compared to the $H^{\text{N-start}}$CON version of the experiment or compared to the classical (carbon start) CON experiment. Sensitivity curves provide means for characterizing experimental sensitivity as a function of the duration of a scan (T_{scan}). A series of spectra was recorded with changing recycle delay, and the peak intensities in the resulting spectra were normalized by the duration of the scan. The T_{scan}-value where the maximum normalized intensity is observed accounts for the optimal recycle delay, while the observed maxima provide means for comparison of the different experiments at their optimal recycle delays.

Our data shows that selective pulsing results in a significant sensitivity gain compared to the H-flip version of the experiment or compared to the CON experiment (lower panel of Figure 5.4). The sensitivity gain varies from residue to residue, depending on the solvent exchange rates of the residue

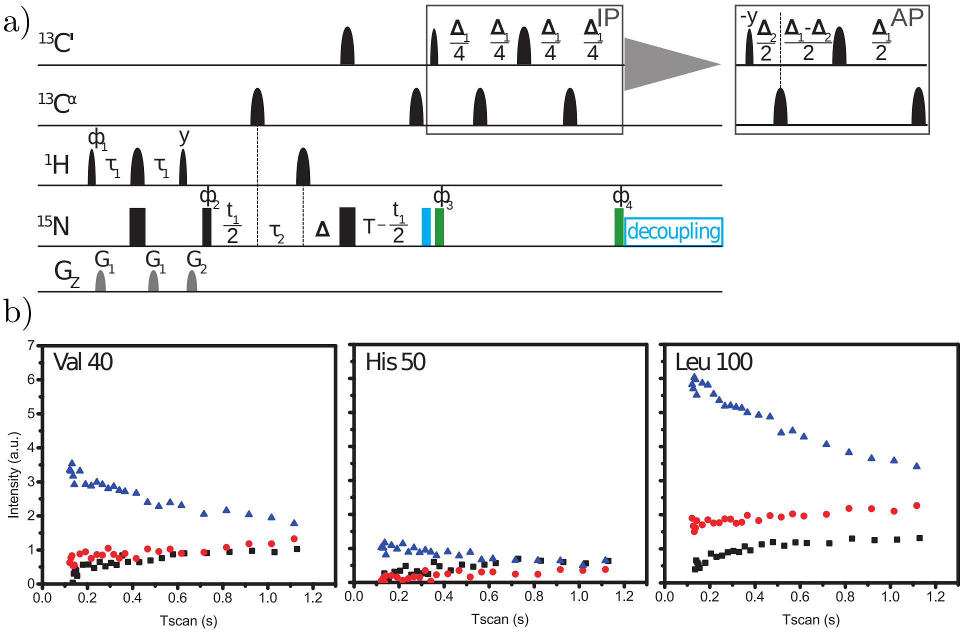


Figure 5.4: a)H$^{\text{N}-\text{BEST}}$CON pulse sequence. The delays are: $\tau_1 = 1/(4J_{NH} - 0.5\delta_1 - 0.5\delta_2)$, $\tau_2 = 1/(4J_{NH} - 0.5\delta_2)$, $\Delta_1 = 1/(2J_{CON})$, 33 ms, $\Delta_2 = 1/(2J_{COC_\alpha})$, 9 ms, $\Delta = 1/(4J_{CON}) - 1/(4J_{NH}) - t_1/2, T = 1/(4J_{CON})$. The delays δ_1 and δ_2 correspond to the lengths of the PC9 and REBURP pulses that are used for 90 ° and 180 ° pulses, respectively. The phase cycle is $\phi_1 = x, -x; \phi_2 = 2(x), 2(-x)$; $\phi_3 = 4(x), 4(-x)$; $\phi_4 = 4(y), 2(-y)$; $\phi_{rec} = x, -x, x, -x$. Quadrature detection was obtained by incrementing phase $\phi_2(t_1)$ in States-TPPI manner. For the ^{15}N virtual decoupling version, two 90 ° ^{15}N radio-frequency (rf) pulses are applied (white rectangles) instead of the one 90 ° ^{15}N rf pulse preceding the IPAP block and ^{15}N decoupling during acquisition. b) Intensities normalized by time T_{scan}, for CON (*black*), H$^{\text{N}-\text{START}}$CON (*red*), and H$^{\text{N}-\text{BEST}}$CON (*blue*) for selected amino acids of α-synuclein at pH 7.4 and 295.5 K.

(largely dependent on its residue type) and on the relaxation time constants of the ^{1}H and ^{13}C spins. H$^{\text{N-start}}$CON does not result in large sensitivity gains, as under conditions of elevated pH and temperature amide proton spins relax with similar T_1 relaxation time constant as the solvent protons, as discussed in the previous section. Therefore, the maximal sensitivity is expected at large recycle delays and overall sensitivity is comparably low.

A particular advantage of combination of ^{13}C detection with longitudinal relaxation optimization is that recovery of the ^{1}H z-magnetization starts after the last pulse on proton. During the pulsing on carbons, the magnetization of protons is not perturbed, therefore even shorter recovery delays can be used. Actually, as the sensitivity curves show, because of very fast recovery of longitudinal magnetization due to solvent exchange, and because of the fact that recovery can start well before the end of the pulse sequence, the maximum of the sensitivity curve is not in the sampled region. It would probably occur at negative $d1$ delays. In order to avoid using conventional decoupling of ^{15}N during detection virtual decoupling was implemented in the experiment.

The $\mathrm{H^{N-BEST}}$CON can also be used to follow real-time kinetics of a reaction, such as phosphorylation of NS5A, as will be described in section 10.5. In this way, changing CO shifts can be monitored. They report directly on changes of secondary structure, that can occur due to phosphorylation-induced changes of the structural ensemble of the protein. The details of this experiment are going to be discussed in Chapter 10. However, the experiment has certain limitations, compared to the conventional CON. The largest sensitivity gains compared to the CON experiment can be observed in the regions with transient structure, because the relatively long C-N transfer only takes place once. In regions of shorter ^{15}N T_2 relaxation times this leads to important sensitivity gain. The advantage is also large in regions of fast, but not too fast, solvent exchange. However, certain residues with the largest exchange rates can not be observed in the $\mathrm{H^{N-BEST}}$CON experiment. Thus, in case of very high exchange rates, use of the conventional CON experiment is advisable.

Chapter 6

BEST-TROSY experiments for IDPs

The advantages of the BEST-type longitudinal relaxation optimization techniques were discussed in the previous chapter. In this chapter we discuss how transverse relaxation optimized spectroscopy (TROSY) can be combined with longitudinal relaxation optimization and how these types of BEST-TROSY experiments can be exploited to acquire high resolution spectra of IDPs with optimized sensitivity.

6.1 The 2D BEST-TROSY experiment

Transverse relaxation-optimized spectroscopy (TROSY) was introduced by Pervushin *et al.* in 1997 [101]. The technique makes use of the interference of dipole-dipole coupling and chemical shift anisotropy that results in line narrowing of the TROSY-component at high magnetic fields. Longitudinal relaxation-optimization techniques have also been applied to the TROSY sequence first by Pervushin et al. [81] in their original paper on usage of selective pulses for longitudinal relaxation optimization. After the introduction of the SOFAST- and BEST-type experiments, the longitudinal relaxation optimized BEST-TROSY experiment has been applied to studies of nucleic acids [102]. Later an improved version of the sequence (shown in Figure 6.1) was introduced and applied to folded proteins [103].

A special feature of the BEST-TROSY experiment is that the H_z polarization that builds up due to relaxation during the t_1-editing period is converted to $-N_z$ polarization by the ST2-PT sequence that is inverted to $+N_z$ polarization by an additional 180 ° pulse after the detection period [103]. In a conventional TROSY sequence this additional polarization would be lost due to ^{15}N T_1 relaxation during the recovery delay that restores equilibrium Boltzmann polarization, but with the short optimal recycle delays of the BEST-type experiments it can be used for increasing sensitivity of

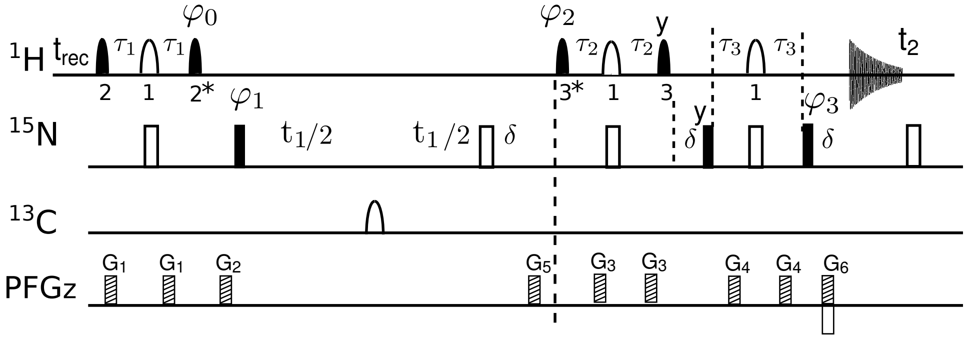


Figure 6.1: Pulse sequence of the 2D gradient selected BEST-TROSY experiment [103]. *Filled and open symbols* indicate 90 ° and 180 ° rf pulses, respectively. Pulses are applied with x phase unless indicated. t_1 denotes the incremented frequency-editing delay. Selective amide ^{1}H pulses cover a bandwidth of 3.5 ppm (centered at 8.5 ppm), with shapes and durations as follows (1) REBURP [104], δ_1, (2) PC9 [105], δ_2, and (3) EBURP-2 [104], 1.4 ms δ_3. An asterisk indicates time and phase reversal of the corresponding pulse shape. The transfer delays are set to account for spin evolution during the selective pulses: $\tau_1 = 1/(4J_{NH}) - 0.5\delta_1 - 0.5\delta_2$, $\tau_2 = 1/(4J_{NH}) - 0.5\delta_1 - \kappa\delta_3$ where $\kappa \sim 0.7$ (to be optimized), $\tau_1 = 1/(4J_{NH})$. The 180 ° pulse on ^{15}N after the acquisition period converts the $-N_z$ polarization originating from proton relaxation to $+N_z$, the phase φ_0 should be set to y or $-y$ depending on the spectrometer, to add the magnetization originating from ^{1}H and ^{15}N. The phase cycling is: $\varphi_1 = -x, x$; $\varphi_2 = -y$; $\varphi_3 = -x$;$\varphi_{acq} = -x, x$. G_1-G_6 denote pulsed field gradients along the z axis.

the next scan. For IDPs, the required high resolution in the ^{15}N-dimension calls for a long t_1-editing period. Therefore, this feature of the experiment is especially advantageous for IDPs, especially if very short recycle delays can be used. This is the case for near physiological conditions as discussed in Section 5.6. For globular proteins deuteration is crucial in order to reach optimal TROSY line narrowing, because of remote $^1H -^1 H$ dipolar interactions, that determine the linewidth of the TROSY component in case of perfect canceling of dipole-dipole coupling and chemical shift anisotropy. In case of intrinsically disordered proteins, there is a lower amount of remote proton spins, essentially they come from the H_α and the sidechain protons. This results in the fact that even without perdeuteration significant line narrowing can be observed for them as we will show in the next section (Figure 6.3). Summarizing the advantages of BEST-TROSY expected for IDPs: *i)* a significant line narrowing can be observed, due to the TROSY effect especially in the transiently structured regions with increased local tumbling correlation time, *ii)* a beneficial sensitivity gain due to longitudinal relaxation-optimization and finally *iii)* the conversion of H_z polarization building up due to relaxation into $+N_z$ polarization is also expected to contribute to sensitivity gain especially for nuclei characterized by small ^{1}H T_1 time constants and large ^{15}N T_1 time constants (as discussed above).

The question rises, how to profit maximally from the TROSY-effect in the case of IDPs. On the one hand, it has been shown by Maltsev *et al.* [106] that even in case of IDPs deuteration can be beneficial for the quality of the recorded TROSY spectra. In their work they used acquisition times as long as 206 ms and they have shown very well resolved spectra in the ^1H dimension. On the other hand, complete deuteration results in having no effect of longitudinal relaxation enhancement from the dipolar relaxation mechanism. The answer is therefore, that the optimal choice depends on the protein. Deuteration can be expensive and results in significantly lower expression yields. Therefore, the approach of Maltsev *et al.* can be more useful for IDPs that give high expression yields and whose spectra would suffer from extensive overlaps in the ^{1}H dimension and are studied at conditions relatively far from the physiological pH and temperature, therefore not favoring solvent exchange. The BEST-TROSY experiment can be more advantageous for IDPs that express less well and/or are studied at near physiological pH and temperature (as indicated by the data in Figure 5.3 in case of near physiological conditions longitudinal relaxation enhancement resulting from exchange with the relaxed water protons is highly beneficial) and if slightly larger linewidths in the ^1H dimension can be tolerated compared to the deuterated case.

We applied the 2D BEST-TROSY sequence to two IDPs of relatively large size ($\sim$ 270 residues). In order to reach high resolution and to profit from the TROSY effect the spectra were recorded at high field (800 MHz) and in the indirect ^{15}N dimension 150 increments were recorded with a spectral width of 2000 Hz in order to allow for high resolution. The 2D BEST-TROSY spectra recorded for BASP1 and D2D3 fragment of NS5A (NS5A) are shown in Figure 6.2. The figure shows that relatively high signal dispersion could be reached in spite of the large size of these two IDPs.

In the following section we will discuss in further details and illustrate the advantage of the BEST-TROSY experiments for IDPs, also when extended to higher dimensional experiments. We undertook studies to characterize the gain in sensitivity and resolution experimentally on the example of these two large IDPs.

We applied in the course of this thesis the 2D BEST-TROSY sequence for various applications, for mapping interactions, measurements of RDCs, sample quality checks, kinetics measurements etc. as will be shown in Part III.

6.2 Performance of BEST-TROSY compared to BEST-HSQC

NMR studies of proteins are usually based on a “fingerprint” spectrum that should be sufficiently well resolved to give a correlation peak per amino acid

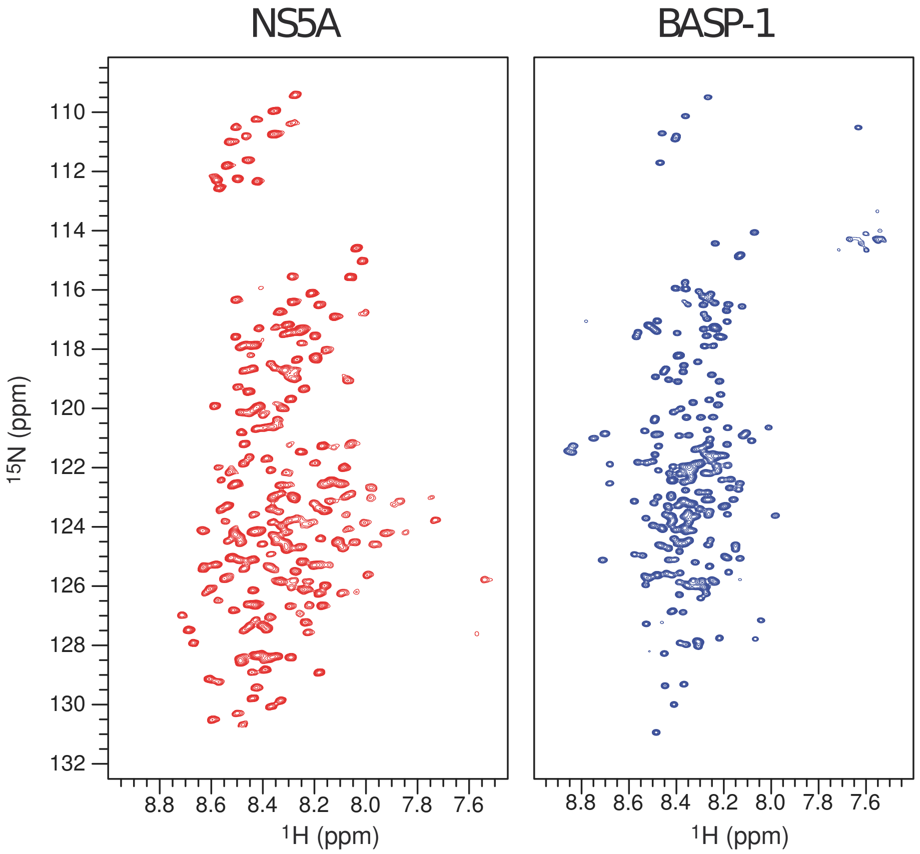


Figure 6.2: 2D BEST-TROSY spectra of NS5A D2D3 (pH 6.5) *left* and BASP1 (pH 2.0) *right* recorded at 800 MHz field at 5 °C.

site in the protein. Usually this "fingerprint" spectrum is a 2D amide ^{1}H-^{15}N correlation spectrum. Sequential backbone assignment is a crucial step when undertaking studies of proteins, because we need to know which correlation peaks belong to which amino acid in the "fingerprint" spectrum to be able to pursue with subsequent studies, to identify where changes occur in the protein as a consequence of an interaction, chemical modification, mutation, changes of conditions or any other perturbation that one wishes to study. 3D BEST-type sequences have been introduced by Lescop *et al.* [91]. Here we will show that for IDPs usage of the BEST-TROSY-type 3D sequences are more advantageous. The advantages of the 2D BEST-TROSY experiment for IDPs, discussed above, with respect to sensitivity and resolution, can be maintained in higher dimensional versions as well. Extension to a third, ^{13}C-dimension (or alternatively to another ^{15}N-dimension) helps in resolving overlaps. Furthermore, sequential correlations can be provided, either by recording pairs of intraresidual- and sequential versions of the experiment or recording a single experiment with intraresidual- and sequential correlation peaks showing up simultaneously in the spectrum (eg. HNCA). The high resolution in the ^{15}N-dimension can be maintained by the use of semi-constant time frequency editing in the 3D versions of the experiment.

To assess the sensitivity enhancement provided by BEST-TROSY we recorded 2D BEST-TROSY HNco and BEST-HSQC HNco spectra of the proteins BASP1 and NS5A D2D3, with all possible acquisition parameters set identical or as similar as possible. The pairs of BT-HNco and BH-HNco spectra of BASP1 and the NS5A proteins are shown in Figure 6.3. It can be appreciated from the figure that the resolution in the BEST-TROSY versions is increased due to the TROSY line narrowing, the extent of which, however, is different for the two proteins. The different characteristics of the structural ensembles of the two proteins result in larger TROSY line narrowing effect for NS5A, as it has more transiently structured regions that result in larger local tumbling correlation times and presumably more dipolar interaction with remote protons in the ensemble of structures it adopts (as shown by the ^{15}N relaxation data in panel c) of Figure 6.3), but the effect is significant for BASP1 as well.

Furthermore, our data shows (Fig. 6.3 b) that as expected from the considerations discussed in the previous sections, peak intensities in the BEST-TROSY spectra are increased by 80% on average for NS5A and 20% for BASP1 compared to the BEST-HSQC experiment. Peak intensity ratios are plotted against peak intensities in the BEST-HSQC (Fig. 6.3 panel d)). This highlights the fact, that in general peaks that are the least intense in BEST-HSQC are more enhanced in the BEST-TROSY. There can be two reasons for this, as there are two reasons also for peaks being weaker than others. The first one is solvent exchange, that is only significant for certain residues of NS5A (BASP1 being measured at pH 2.0). As discussed before, when pulsing selectively, solvent exchange results in a reduced effective proton T_1. This in turn causes more H_z polarization building up during the $N \longrightarrow CO$-transfer delay that consequently results in a larger N_z polarization at the end of the sequence after the $180\,^\circ$ pulse. The second reason can be locally more restricted dynamics resulting in slower tumbling, therefore shorter T_2 relaxation time constants. As can be appreciated from comparison of panel (b) and panel (c) of Figure 6.3 the regions that are more rigid as seen by ^{15}N relaxation data profit more from the TROSY-effect as well.

These two effects result in a more uniform distribution of peak intensities in the spectrum. It should be noted, that signal enhancement is also less efficient for BASP1 than for NS5A, with a few residues even experiencing slight sensitivity loss compared to the HSQC-version. This can be attributed to the different sample conditions used for the two proteins and the different structural ensembles they adopt. This difference in their structural ensemble is illustrated by the ^{15}N relaxation data. The different sample conditions result in shorter proton T_1 values for NS5A than for BASP1, therefore the optimal recycle delay is shorter in the case of NS5A, which results in more efficient recovery of the polarization lost during the $N \longrightarrow CO$-transfer delay. Therefore we can conclude from the comparison of the performance of BEST-TROSY for NS5A and BASP1, that the BEST-TROSY experiments are especially advantageous for IDPs that have a significant amount of residual structure, as peak intensities will be more homogeneous in the spectrum along with an enhanced sensitivity and resolution.

6.3 3D BEST-TROSY experiments for assignment of IDPs

In this section 3D BEST-TROSY experiments will be discussed that we recommend for time efficient sequential assignment of IDPs. In the previous section the advantage of BEST-TROSY over its BEST-HSQC counterpart in terms of sensitivity and resolution has been shown on the example of the HNco experiment. It has been shown before [107] that the expected frequency dispersion for IDPs is the largest in the N and CO dimensions, while in the ^{1}H dimension the signal dispersion is limited. This is why ^{13}C-detected experiments for IDPs can be advantageous, but they are compromised by the intrinsic lower sensitivity of the ^{13}C nucleus compared to ^{1}H. Therefore, on samples like NS5A with limited solubility (allowing maximal sample concentrations of 150 μM) experimental time would be much too long for ^{13}C-detected experiments. Instead, the 3D BEST-TROSY experiments employ excitation and detection of ^{1}H, but also benefit in the indirect dimensions from the larger frequency dispersion of ^{13}C and ^{15}N. The suite of the 3D BEST-TROSY pulse sequences is shown in Figure 6.4. They all employ semi-constant time frequency editing that allows maintaining the high resolution in the ^{15}N dimension also in the 3D experiments. Typically four experiments are required to achieve sequential assignment for IDPs and to get a complete set of backbone chemical shifts that yield information on the secondary structural conformations. The pulse sequences of these four experiments are shown in Figure 6.4. BT-hNcocaNH experiments help in quick establishment of sequential connectivities, complemented by the BT-hnCOcaNH experiment. The BT-hNcocaNH experiment gives connectivities from residue i to residue $i+1$, while the BT-hnCOcaNH experiment gives connectivities from residue i to residue $i-1$, like usual sequential correlation experiments. A drawback for these sequences is the long transfer delays as they might be insensitive for transiently structured regions characterized by small T_2 relaxation time constants. Nevertheless, the highly flexible regions can be assigned easily and time-efficiently with these experiments. The further two experiments are the BT-iHNCACB and BT-HNcoCACB experiments, that also help establishing sequential connectivites in the $i-1$ direction. These experiments are helpful for the transiently structured regions, because the signal dispersion of C$^\alpha$ and C$^\beta$ are larger and, as the

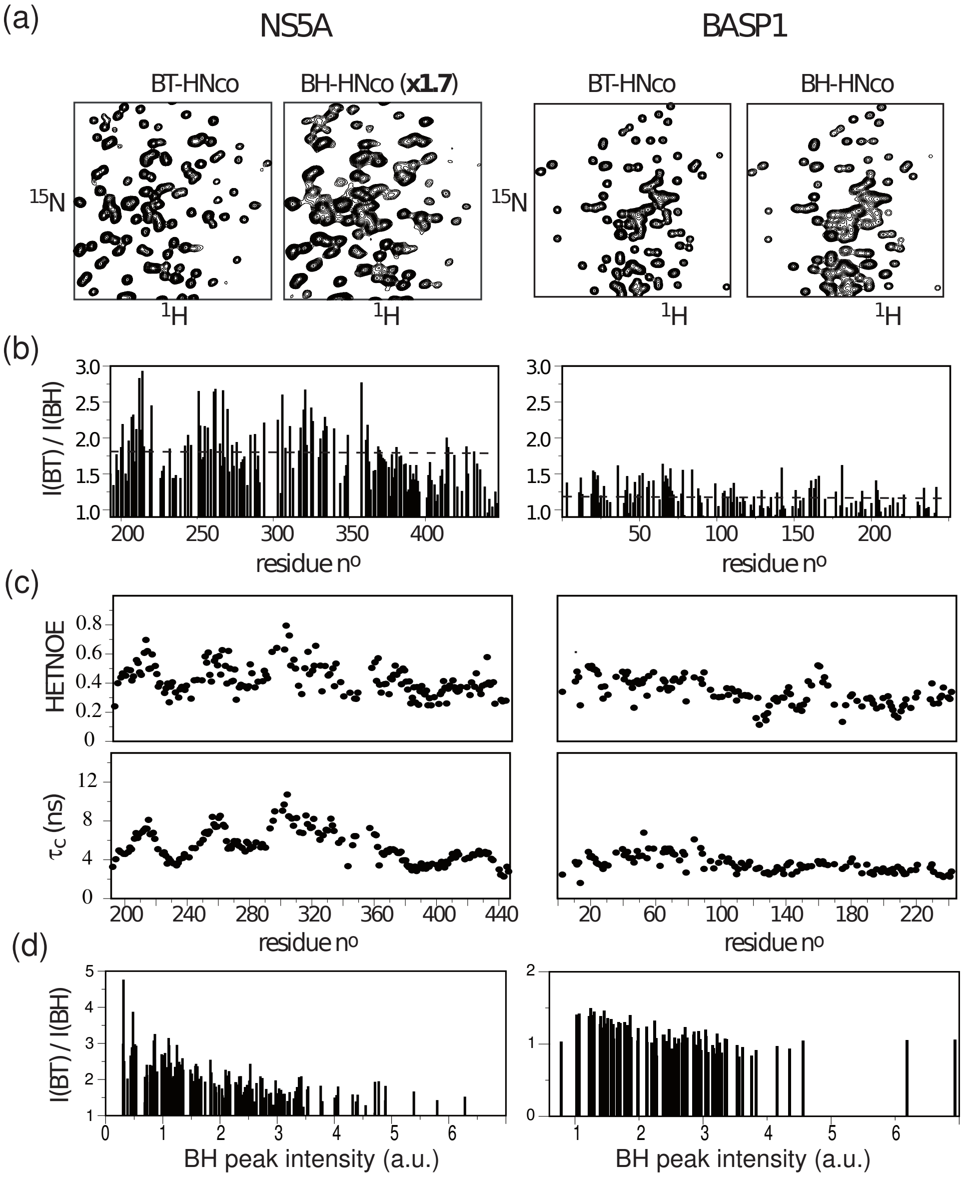
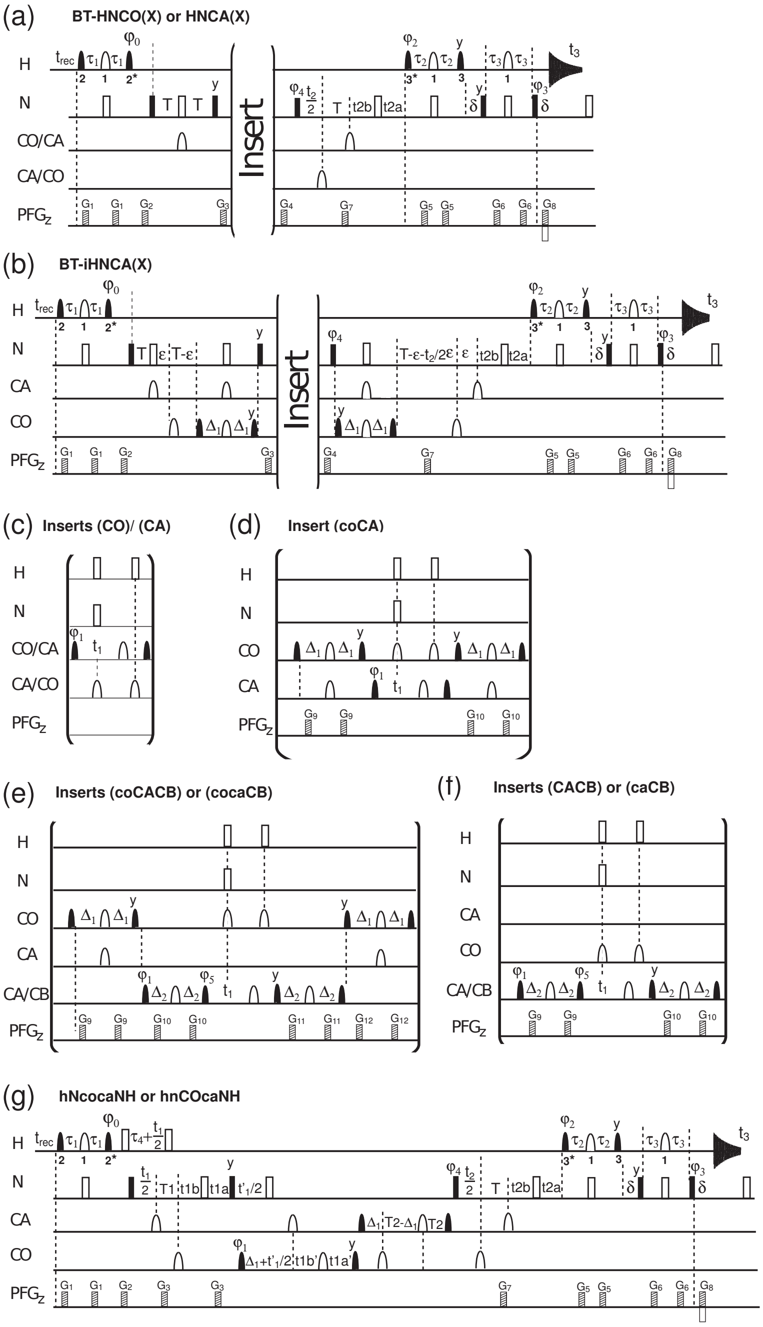


Figure 6.3: BEST-TROSY (BT) and BEST-HSQC (BH) version of the HNco experiment are compared in terms of sensitivity and resolution for two IDPs differing in their dynamics properties. a) Gain in resolution in the BEST-TROSY (BT) compared to the BEST-HSQC (BH) version of the HNco experiment. b) Residue resolved intensity gain (BT vs BH) for BASP1 and NS5A D2D3. Comparison with the ^{15}N relaxation data (panel c) shows the correlation of the intensity gain to the dynamics properties. In panel d) intensity gains in BT versus BH are plotted as function of the original peak intensities in the BH spectrum, illustrating that the BT results in more homogeneous peak intensity distribution of the spectrum.


(a) BT-HNCO(X) or HNCA(X)

(b) BT-iHNCA(X)

(c) Inserts (CO)/ (CA)

(d) Insert (coCA)

(e) Inserts (coCACB) or (cocaCB)

(f) Inserts (CACB) or (caCB)

(g) hNcocaNH or hnCOcaNH

Figure 6.4: BEST-TROSY sequences: (a+c) HNCO or HNCA, (a+d) HNcoCA, (a+e) HNcoCACB or HNcocaCB, (a+f) HNCACB or HNcaCB, (b+c) iHNCA, (b+f) iHNCACB, (g) hNcocaNH and hnCOcaNH. Experimental details are similar to those of the 2D BEST-TROSY (Figure 6.1). In addition to the notations defined there, open squares on ^{1}H indicate BIP-720-50-20 broadband inversion pulses [108]. CO pulses have the shape of the center lobe of a $\sin x/x$ function, whereas CA and CA/CB pulses are applied with a rectangular shape and zero excitation at the CO frequency. For further experimental details see the caption of the original paper [109] in Appendices .

transfer delays are shorter, signals can also be observed for the transiently structured residues. As C^{α} and C^{β} chemical shifts are sensitive for transient secondary structure, usually we are also not facing the problem of signal overlap in these regions. The combination of these four experiments is sufficient, according to our experience, for assignment of relatively large IDPs. For obtaining CO chemical shifts for residues in the transiently structured regions an additional BT-HNCO experiment might be necessary.

The application of the 3D BEST-TROSY experiments will be shown on the example of the assignment of a protein, core protein of Pestivirus, an IDP of 104 residues length, that is known to have RNA chaperoning activity.

The 3D BEST-TROSY experiments introduced in Section 6.3 were used for time efficient establishment of backbone connectivities. Example strips of the BEST-TROSY spectra show how these sequences were used for assignment of PV-core. The first spectrum used for quick establishment of sequential connectivities was the BT-hNcocaNH experiment, its use is illustrated in Figure 6.5. For confirming the connectivities established with the BT-hNcocaNH experiment the connectivities were checked in the BT-hnCOcaNH and the pair of BT-HNcoCACB and BT-iHNCACB experiments. Example strips are shown for a stretch of residues of PV core (residues 18-22) showing the BT-hnCOcaNH and the pair of BT-HNcoCACB and BT-iHNCACB experiments (Figure 6.6). These experiments allow assignment of an IDP such as PV core, of a sample concentration of $\sim 150\mu$M with overall acquisition time of less than a weekend. This set of experiments allowed obtaining all backbone chemical shifts (except H^{α}).

However, in the course of establishing connectivities sometimes there is a break point. This can be caused either by a proline residue, or too low sensitivity of the amide site due to solvent exchange or relaxation properties due to transient structure as discussed previously. Proline residues are very abundant in intrinsically disordered proteins [110]. They result in a break point in the assignment procedure, as they do not have an amide proton. In the course of the sequential assignment procedure especially in case of IDPs with possibly large signal intensity variations in their spectra it is difficult to identify why no connection shows up to the neighboring residue. In the course of this thesis we developed an experiment that allows identification of proline-neighboring residues.

The experiment is based on the fact that prolines have distinct amide ^{15}N chemical shifts. The Pro-BT-HNcocan and Pro-BT-iHNcan experiments exploit this characteristic chemical shift range of proline amide nitrogens, that is illustrated in Figure 6.7. These entries were extracted from IDP chemical shifts deposited in the Biological Magnetic Resonance Data Bank (BMRB).

The experiment belongs to the class of spin-echo difference experiments. The pulse sequence of the Pro-BT-HNcocan is shown in Figure 6.8. This experiment yields correlation peaks for the residues following prolines. The

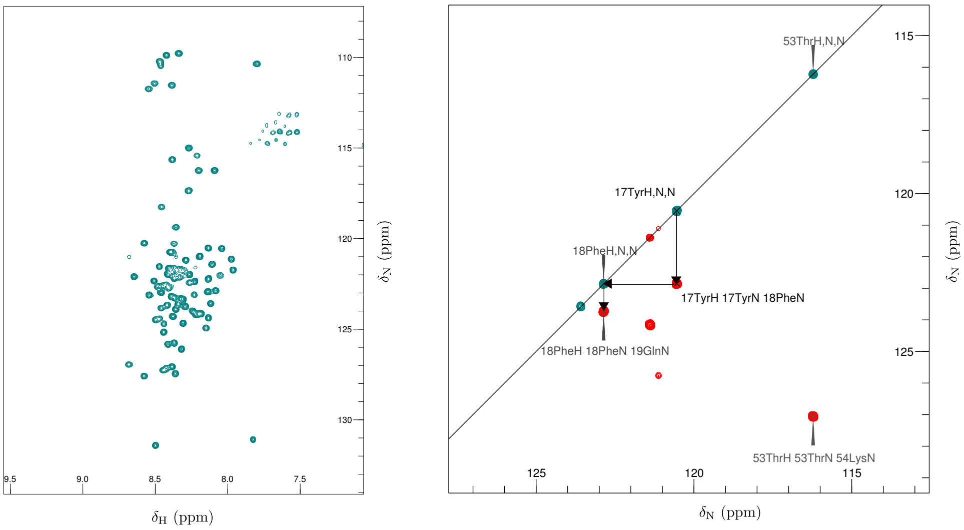


Figure 6.5: The establishment of sequential correlations is illustrated using the BT-hNcocaNH experiment. For correlating amide peaks in the 2D-BT spectrum (*left*) one can navigate to the 2D N-N plane of the BT-hNcocaNH spectrum at the corresponding ^{1}H frequency (*right*) and search for the correlation peak that gives the ^{15}N chemical shift of residue $i + 1$.

experiment can also be performed in the iHNcan version, where the residues preceding prolines are detected selectively. The spectrum, where only amide resonances of Pro-neighboring residues show up, is formed by a difference spectrum of the *reference* and *transfer* experiments. The experiment is an extension of the BT-HNcoca experiment with an additional INEPT transfer to the nitrogen of the neighboring residue. In the transfer experiment a proline nitrogen selective pulse is applied, and thus magnetization is transferred to the Pro-nitrogen, while coupling evolution is refocused for all other residue types. In the reference experiment the pulse on the amide nitrogen is omitted, therefore coupling evolution is refocused for all residues including prolines. The selective amide nitrogen pulse is applied at the nitrogen frequency at 138 ppm, as we found this value to be the average of the chemical shifts of proline nitrogens of IDP data sets deposited in the BMRB with a bandwidth of 7 ppm that covers the shifts of all the proline amide nitrogens found there (6 entries, 79 chemical shifts). It has to be noted, that in case of extreme chemical shifts, resulting for example from ring current effect in case of partially structured IDPs, false positives might show up in the spectra. This experiment is expected to be applicable, however, to most IDPs, as such extreme chemical shifts are rare for IDPs.

In Figure 6.8 application of this experiment is shown for the NS5A D2D3 protein. For the assignment of PV core, as it does not contain high amounts of prolines, this experiment was not necessary.

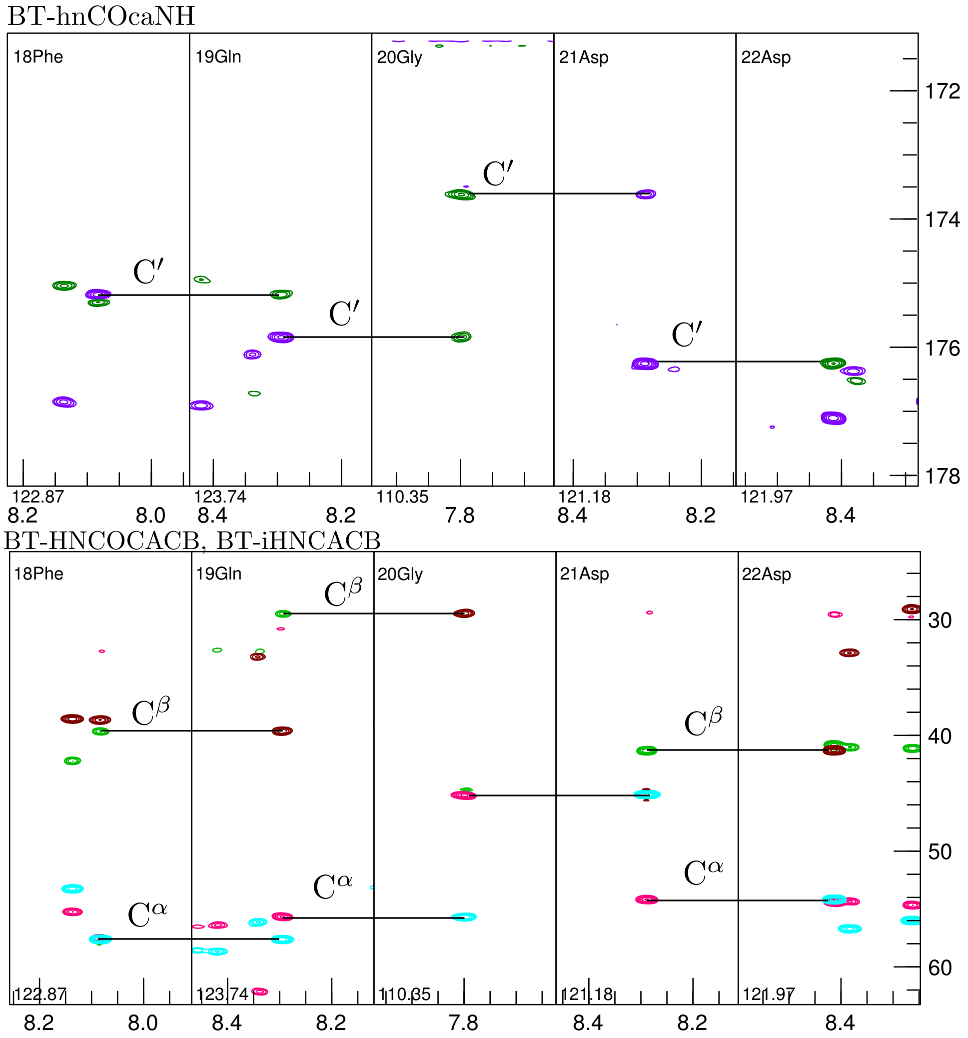


Figure 6.6: Example strips of the BT-hnCOcaNH and the pair of BT-HNcoCACB and BT-iHNCACB experiments illustrating their use for sequential resonance assignment. The intraresidual (*purple*) and sequential (*green*) peaks in the BT-hnCOcaNH spectrum show up with opposite sign. The intraresidual peaks in the BT-HNcoCACB are shown in *brown* for C^{β} and *blue* for C^{α}. The peaks in the BT-iHNCACB spectra are shown in *green* and *magenta* for C^{β} and C^{α}, respectively.

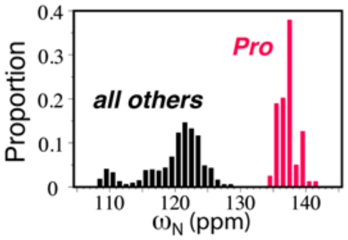


Figure 6.7: Statistics of ^{15}N chemical shifts of intrinsically disordered protein data sets deposited in the BMRB. Proline nitrogen chemical shifts extracted from 6 IDP chemical shift data sets (*pink*) are plotted against chemical shifts of amide nitrogens of all other residues from 14 IDP chemical shift data sets (*black*).

In case of doubts concerning the position of the fragments with sequential connectivities, the HADAMAC [111] and iHADAMAC [112] experiments can be recorded, like it was done in the course of the assignment of NS5A.

6.4 Saturation of water in the BEST-TROSY experiments

In Chapter 5 we discussed the advantages of selective pulsing. There we assumed a perfect “top hat” excitation profile and we did not consider problems resulting from eventual pulse imperfections. In this section we discuss the consequences of pulse imperfections that result in partial excitation at the water frequency.

The equation describing the effect of partial excitation on the longitudinal magnetization of water is the same as Eq. 5.4 but instead of the β excitation angle the f factor is introduced that denotes the remaining fraction of the polarization of water after one scan:

$$M_z^{W,start,ss} = M_z^{W,0} f \frac{1 - \exp(-T_{rec}/T_{1,W})}{1 - f \exp(-T_{rec}/T_{1,W})}. \tag{6.1}$$

The values of $T_{1,W}$ for a range of temperatures have been reported for oxygen-free water (Table 6.1) [95]. As it can be appreciated from the table, the longitudinal relaxation time constant of water increases with increasing temperature. We remeasured the longitudinal relaxation time constants of water at two temperatures and the result is in excellent agreement with the data measured 60 years ago (as shown in Fig. 6.9 a). Theoretical curves calculated from Eq. 6.1 show that in the regime of short recycle delays as used in fast NMR, if the water polarization is even very slightly perturbed, the steady state water polarization ($M_z^{W,start,ss}$)decreases dramatically (f) as illustrated by Figure 6.9 b.

Considering the Bloch-McConnell equation (Equation 5.19) the consequences of this reduced magnetization of water can be seen: instead of the

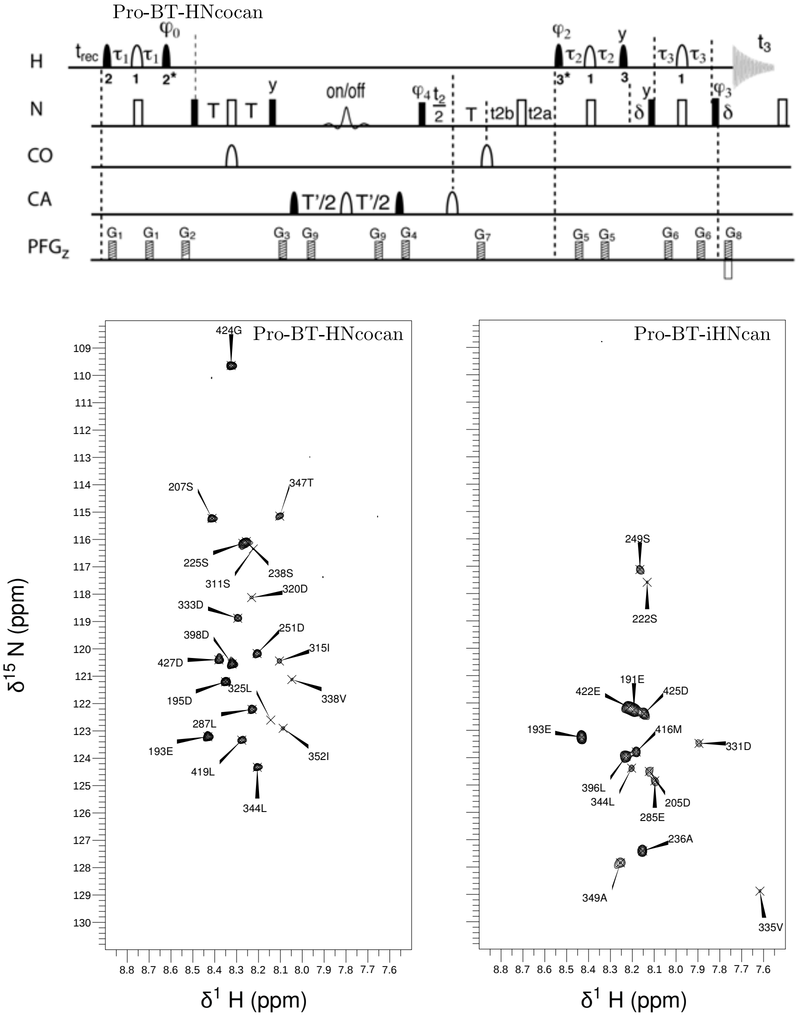


Figure 6.8: Pulse sequence of the Pro-BT-HNcocan experiment and its application to the D2D3 fragment of NS5A. Transfer delays and other experimental details are similar to those of the BEST-TROSY 3D experiments (Fig. 6.4). The spectra are obtained by subtracting two data sets recorded with and without the Pro ^{15}N refocusing pulse. This pulse of REBURP pulse shape is centered at 138 ppm and covers a band width of 7 ppm (corresponding to 8.6 ms pulse length at 80 MHz). The T' filter delay is set to 58 ms. In *bottom*, the Pro-BT-HNcocan *left* and Pro-BT-iHNcan *right* spectra are shown, where residues following, and residues preceding prolines are detected, respectively.

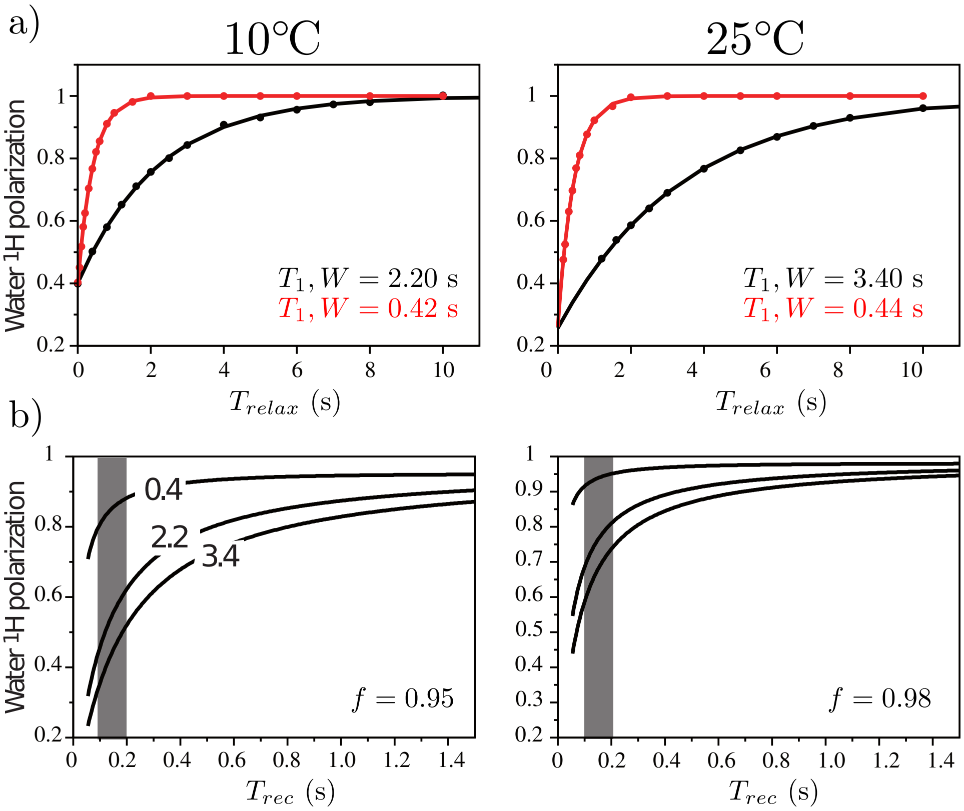


Figure 6.9: a) T_1 relaxation curves of water are shown at 10 °C (*left*) and 25 °C (*right*) in the absence (*black*) and in the presence (*red*) of gadodiamide. Polarization of water recovering from saturation measured as function of the relaxation delay. The data was measured at 600 MHz field at a spectrometer equipped with a cryoprobe by taking the intensity of the first point of the FID (intensities would be affected by radiation damping). b) Theoretical curves of recovery of the water polarization as function of the recycle delay for the T_1 values of water shown in panel a) in case of slight perturbation of water (f=0.95 *left*, f=0.98 *right*). The range of recycle delays typically used in fast pulsing experiments is highlighted in *gray*.

T (°C)	$T_{1,W}$ (s)
0	1.59
5	1.88
10	2.20
15	2.55
20	2.95
25	3.37
30	3.82
35	4.30
40	4.76

Table 6.1: Longitudinal relaxation time constants of oxygen-free water [95]

Boltzmann water magnetization, the reduced steady state water magnetization ($M_z^{W,start,ss}$)is transferred to the amide protons by chemical exchange, therefore for fast exchanging protons the experimental sensitivity is severely reduced.

Therefore, paramagnetic relaxation agents can be helpful if the water polarization is perturbed (eg. by imperfections of selective pulses) as they speed up longitudinal relaxation of bulk water (as shown in Fig.6.9 b). Motivated by Ref. [100] where the authors report increase of experimental sensitivity for BEST-type experiments for globular as well as for disordered proteins by adding a paramagnetic relaxation agent (gadodiamide) to the sample, we wanted to test the described effect on BEST-TROSY experiments with keeping in mind that if the selective pulses perform perfectly, the polarization of water should not be perturbed. Therefore, we should not be able to observe an effect on the peak intensities. We also chose gadodiamide, as in the publication cited above, for this purpose as it used to be widely used as an MRI contrast agent and as it is not supposed to interact with proteins at low concentrations. To assess its potential of speeding up longitudinal relaxation of water and to optimize for the necessary concentration we studied its effect on the T_1 of the bulk water. We measured the T_1 relaxation time constants as a function of the concentration of the compound (c_{Gd}). The results are shown in Table 6.2.

We found that 0.5 mM gadodiamide concentration sufficiently enhances the longitudinal relaxation of water. Its effect on the T_1 relaxation time constant of water (*red* curve), furthermore its theoretical effect of speeding up recovery of the water polarization in case of slight perturbation of the water polarization by the pulse sequence are shown in Fig.6.9. We pursued to test the effect of the compound on peak intensities in BEST-TROSY 2D experiments by adding the gadodiamide paramagnetic agent to the sample. The two samples were identical except the added gadodiamide (dilution effect for the protein is negligible, as 1 M gadodiamide solution was used).

c_{Gd} (mM)	$T_{1,W}$ (s)
0	3.40
0.25	0.63
0.50	0.44
0.75	0.27

Table 6.2: Longitudinal relaxation time constants of water measured for different gadodiamide concentrations (c_{Gd}) at 25 °C

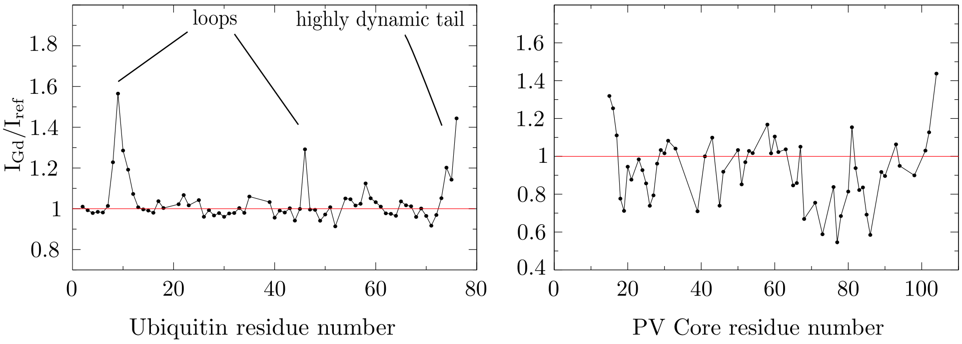


Figure 6.10: Peak intensity ratios in 2D BEST-TROSY spectra recorded in the absence (I_{ref}) and in the presence of 0.5 mM gadodiamide (I_{Gd}) of a globular protein, ubiquitin (*left*) and of an IDP, PV Core (*right*).

We found that indeed, addition of 0.5 mM gadodiamide to the sample results in enhanced peak intensities for the loop region of ubiquitin that is involved in solvent exchange because of being exposed, as shown in Figure 6.10 left panel. The buried regions of ubiquitin are unaffected by addition of the compound (as expected). These findings indicated that imperfections of the BEST-TROSY pulse sequence perturb the water polarization.

For the intrinsically disordered PV core protein (Fig. 6.10 right panel) we found an overall loss in sensitivity that is supposedly due to reduction of the T_2 relaxation time constants of the protein due to the presence of gadodiamide. This could be caused by an interaction between gadodiamide and the protein.

In order to test which pulse is responsible for the perturbation of the magnetization of water we performed the following measurements. We measured the polarization of water protons in the steady state for the selective pulses used in the BEST-type experiments and for some other pulse shapes, as a function of the offset of the center of the shaped pulse. The measurement of the water polarization was performed by application of the selective pulse and then a recovery delay of 100 ms repeated 32 times in order to reach steady-state conditions and after the last pulse the H_z magnetization of water was read out with a non-selective small flip-angle pulse ($\sim 1\,°$).

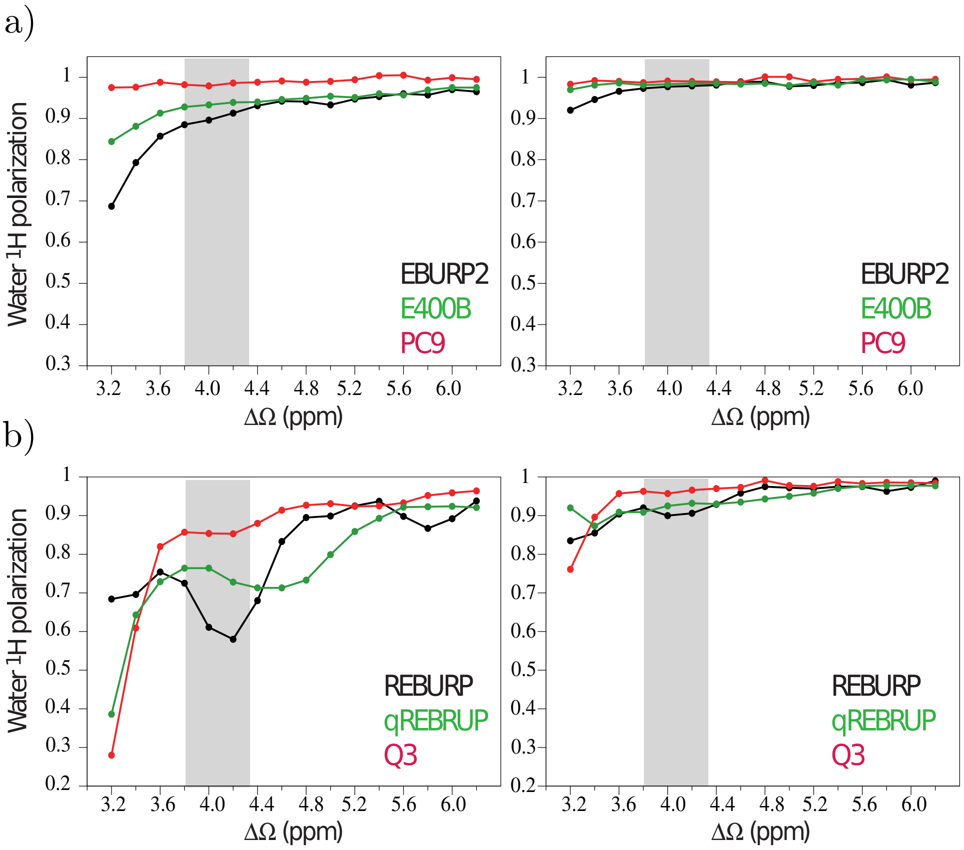


Figure 6.11: Steady state water polarization after 32 repetitions of excitation with a shaped pulse followed by a 200 ms recycle delay, as function of the offset of a) 90 ° shaped pulses and b) 180 ° shaped pulses as measured at 600 MHz Agilent spectrometer equipped with a cryo-probe in the absence (*left*) and in the presence (*right*) of gadodiamide.

On spectrometers equipped with a cryo-probe the resulting free induction decay (FID) is affected by radiation damping, therefore the magnitude of the first point of the FID was measured. The reference value was measured in the same way, but without applying the shaped pulse. Then the offset of the shaped pulse was varied between 3.2 and 6.2 ppm with keeping the band width constant. The resulting data on the effect of the different selective pulses on the water polarization are shown in Figure 6.11. The region where the pulses are usually centered in the BEST-TROSY experiments is highlighted in gray.

The results show that the 180 ° refocusing REBURP pulse, significantly perturbs the water polarization if the pulses are centered at around 4 ppm offset, as necessary for selective excitation of the amide protons. The other tested pulses, qREBURP (E. Kupce, unpulished) and Q3 performed slightly better with respect to perturbation of the water polarization, with Q3 per-

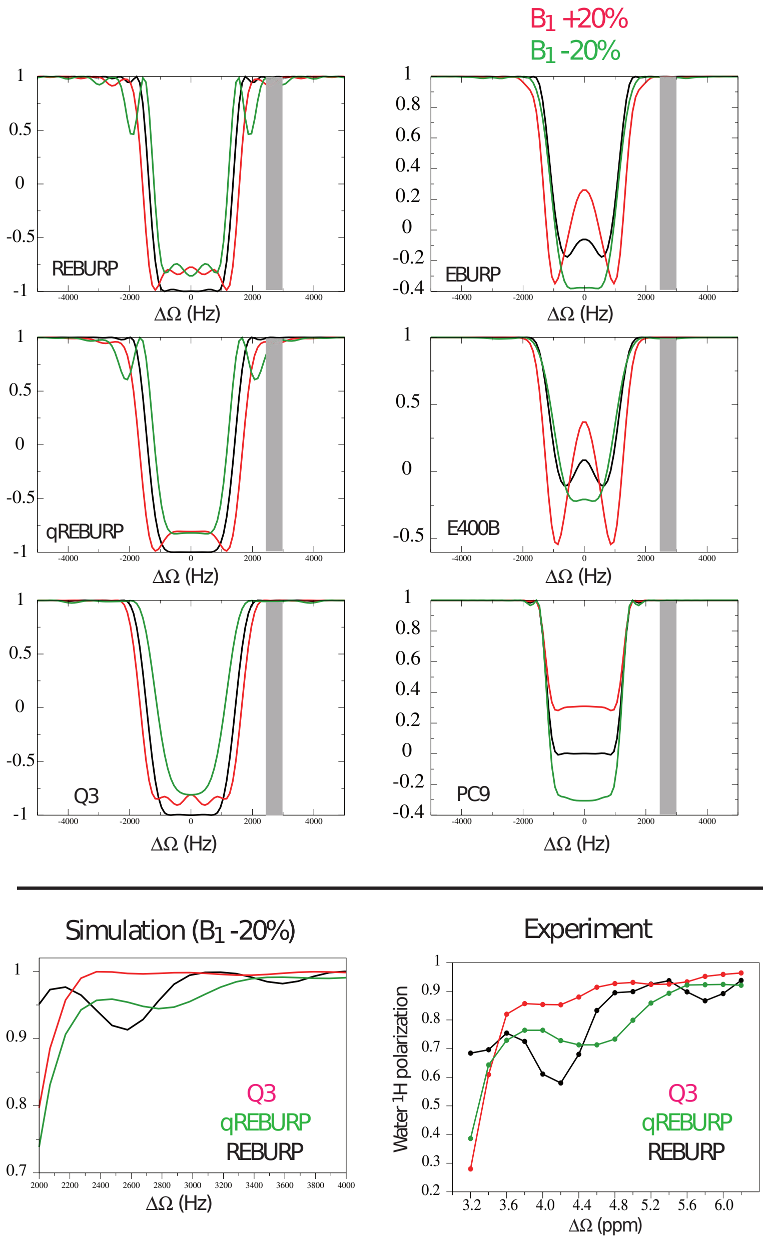


Figure 6.12: Simulated excitation profiles of the tested pulse shapes with the optimal $\mathbf{B_1}$ field (*black*) and -20% (*green*) and +20% (*red*) $\mathbf{B_1}$ fields. In the bottom panel simulated steady state water polarization as function of the offset of the REBURP pulse is shown (*left*) as compared to the experimentally measured data (*right*).

forming best under these conditions.

Simulations of the excitation profiles of the tested pulse shapes are shown in Fig. 6.12. The $\mathbf{B_1}$ was varied in order to try to reproduce the experimen-

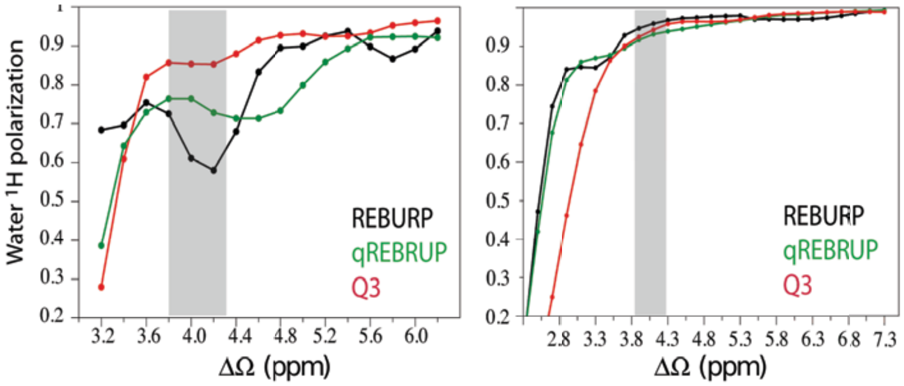


Figure 6.13: Steady state water polarization as function of the offset of the tested shaped pulses (REBURP *black*, qREBURP *green* and Q3 *red*) as measured at 600 MHz Agilent spectrometer equipped with a cryoprobe (*left*) and measured on a Bruker 600 MHz spectrometer with a cryoprobe (*right*).

tally measured data. In the simulated excitation profile of the REBURP pulse shape only small wiggles are seen at the water frequency that do not explain the observed large perturbation of the water polarization. However, when the simulation is performed with a 20 % reduced $\mathbf{B_1}$ field in order to account for eventual detuning of the probe these wiggles increased, resulting in the observed large perturbations of the water polarization.

The simulations showed that most likely our data can be explained by detuning of the probe under fast pulsing conditions. When new spectrometers arrived at the new institute building of IBS with brand new cryoprobes, we tried to reproduce these results. When remeasuring the same data at the new spectrometers we did not see a similar water perturbation profile as we had measured with the old probes. Actually, the REBURP pulse performed best among the tested pulses. The comparison of the results obtained on the new probe and the old probe is shown in Fig. 6.13.

The conclusion was, therefore, is that the observed water perturbation profiles probably resulted from detuning of the probe under fast pulsing conditions. With the new probe this was not the case, the REBURP pulse performed experimentally as simulated, so probably addition of the paramagnetic agent would not result in increased sensitivity (we have not yet carried out this experiment). However, in case a probe gets detuned while pulsing fast, another option to improve sensitivity would be the use of selective pulses optimized for a range of $\mathbf{B_1}$ fields. Our collaborators, Stefan Glaser and Steffen Düwel (TU Münich) optimized asymmetric selective pulses for keeping a clean excitation profile at the water frequency when detuned, while large wiggles are allowed to appear towards higher than the amide proton resonance frequencies. The simulated excitation profiles of their first resulting pulse at optimal $\mathbf{B_1}$ and in case of reduced $\mathbf{B_1}$ due to detuning are

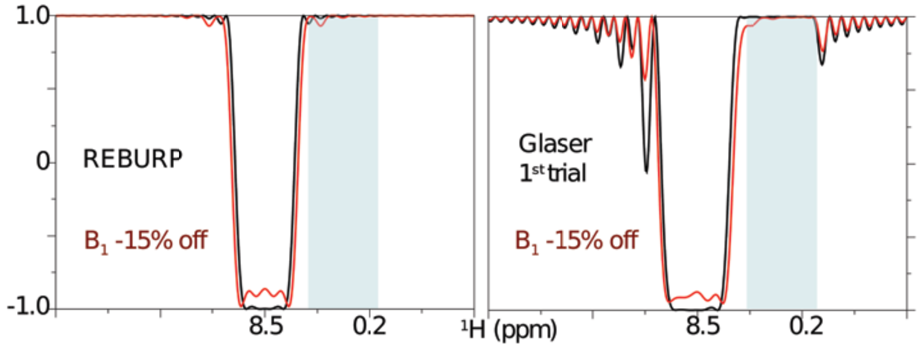


Figure 6.14: Simulated excitation profiles of REBURP pulse and the 'Glaser' pulse. The profiles are shown for siulations conducted with the optimal $\mathbf{B_1}$ fields (*black*) and with $\mathbf{B_1}$ 15% off (*red*). The frequency region where the pulse is expected to result in no perturbation of the spins is highlighted in *light blue*.

shown in Figure 6.14. The alternative solution to increase sensitivity would be the use of the PRE agent, but as it results for IDPs in reduction of T_2 of the protein because of the PRE effect, its use is questionable. Optimized pulse sequences do not result in adverse effects on relaxation. As we have shown, the BEST-TROSY pulse sequence avoids perturbation of water on well-performing probes, while on other probes there is place for improvement by further optimized pulse shapes that provide a clean profile at the water frequency also at lower $\mathbf{B_1}$ fields.

6.5 Measurement of solvent exchange rates with BEST-TROSY

Solvent exchange rates yield information on solvent exposure of the amide protons and thus they can be used for structural characterization of proteins. Often, measurement of solvent exchange rates is used for identification of interacting residues, as solvent exposed parts of the protein become buried upon interaction. It can also be used to observe unfolding or folding induced changes in solvent accessibility. However, presently available techniques are not rapid enough to monitor for example folding events in real-time.

Several methods exist for measurements of solvent exchange rates. One that allows measurement of k_{ex} in relatively slow exchange regime is H/D exchange [113], that is based on dissolving a protonated protein in D_2O and measuring a series of amide ^{1}H-^{15}N correlation spectra to monitor real-time the disappearance of the proton signals. The k_{ex} values are then obtained from exponential fit to the intensities. With fast acquisition techniques, exchange rates up to $k_{ex} < 0.1\,\mathrm{s}^{-1}$ can be measured with this method, provided sensitivity is very high. [92,114]. For measurement of smaller exchange rates a number of techniques have been proposed that are based on transfer of the water magnetization to the amide protons of the protein. [115–119]. For characterization of intrinsically disordered proteins the most used method is the CLEANEX experiment [117]. The key feature of the experiment compared to previous techniques is the CLEANEX-PM spin-locking sequence, that suppresses to a large extent intramolecular and exchange-relayed NOEs. However, this experiment requires long acquisition times. Exchange rates in the range of $1 < k_{ex} < 100\,s^{-1}$ can be measured with magnetization transfer techniques. The upper limit is imposed by line broadening of the correlation peaks due to exchange, while the lower limit results from longitudinal relaxation limiting the mixing time that can be used. A recently introduced method is SOLEXSY [120], that is performed in 50% D_2O and 50% H_2O. It uses ^{15}N as a probe and monitors the conversion of deuterated species to protonated species. This experiment is also characterized by relatively low sensitivity.

The HET$^{\text{ex}}$-SOFAST experiment was presented in the literature some years ago [121]. It allows obtaining quick qualitative information about the buried or exposed nature of the amide protons along the polypeptide chain. The experiment is basically a SOFAST-HMQC experiment, with selective saturation of water in the *sat* and without saturation in the *ref* experiment. The saturation of water is transferred via chemical exchange to the observed amide protons. Therefore, for exposed amides the intensity ratio I_{sat}/I_{ref} decreases while for completely buried amide protons it remains near one.

We addressed the question whether an optimized version of the experiment could be used for quantitative measurement of amide solvent exchange rates. We modified the experiment and implemented a BEST-TROSY version in order to profit from the higher resolution due to the TROSY-effect at high magnetic fields. This experiment profits from the high sensitivity thanks to longitudinal relaxation enhancement achieved by selective pulsing. First the theoretical model for extracting the exchange rates from the data is going to be discussed.

As already mentioned in the previous chapter, the evolution of the longitudinal magnetization is described by the Bloch-McConnell equation (Eq. 5.19). For the present discussion we only consider evolution of the magnetization of one amide proton spin and the bulk water magnetization. We ignore cross-relaxation to the other proton spins. In this case the Bloch-McConnel equation for the system is:

$$-\frac{\mathrm{d}}{\mathrm{d}t}\begin{pmatrix} I_W - I_W^0 \\ I_{HN} - I_{HN}^0 \end{pmatrix} = \begin{pmatrix} \rho_W & 0 \\ 0 & \rho_{HN} \end{pmatrix}\begin{pmatrix} I_W - I_W^0 \\ I_{HN} - I_{HN}^0 \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ -k_{ex} & k_{ex} \end{pmatrix}\begin{pmatrix} I_W \\ I_{HN} \end{pmatrix} \tag{6.2}$$

where I_W is the polarization of the bulk water and I_{HN} is the polarization of the proton.

A reference and a saturation experiment are recorded in the HET$^{\text{ex}}$-BEST-TROSY experiment in order to extract information on solvent exchange. The boundary conditions for solving equation 6.2 are as follows.

At time $t = 0$ in the reference experiment the amide polarization $I_{HN,ref}(0) = 0$ and the water polarization is assumed to be at the equilibrium polarization $I_{W,ref}(0) = I_W^0$.

In the saturation experiment water is assumed to be saturated all the time, therefore $I_{W,sat}(t) = 0$ and the amide polarization at time $t = 0$ is $I_{HN,sat}(0) = 0$.

The solutions of equation 6.2 for these boundary conditions as function of the relaxation delay (d_{relax}) are given by:

$$I_{HN,ref}(d_{relax}) = I_{HN}^0 \frac{\rho_{HN} + k_{ex} I_{W,ref}(d_{relax})}{\rho_{HN} + k_{ex}} \left[1 - \exp\left\{-\left(\rho_{HN} + k_{ex}\right) d_{relax}\right\}\right] \tag{6.3}$$

for the reference experiment and

$$I_{HN,sat}(d_{relax}) = I_{HN}^0 \frac{\rho_{HN}}{\rho_{HN} + k_{ex}} \left[1 - \exp\left\{-\left(\rho_{HN} + k_{ex}\right) d_{relax}\right\}\right] \tag{6.4}$$

for the saturation experiment.

Note, that this result was not considering presence of nearby protons that might also exchange with water and thus influence significantly the evolution of the magnetization. These relaxation pathways are not accounted for, but as will be shown below, this model explains sufficiently well the evolution of the magnetization for highly dynamic IDPs. In order to fit for k_{ex} (that requires fits also for ρ_{HN} and I_{HN}^0), measurement of $I_{HN,ref}(d_{relax})$, $I_{HN,sat}(d_{relax})$ and $I_{W,ref}(d_{relax})$ is required at least at two d_{relax} relaxation delays. However, when the data is only available at a single d_{relax} time, the $HET^{ex}ratio$ can be computed:

$$HET^{ex}ratio = \frac{\left(I_{HN,ref}(d_{relax})/I_{HN,ref}(d_{relax}) - 1\right)}{I_{W,ref}(d_{relax})} = \frac{k_{ex}}{\rho_{HN}} \tag{6.5}$$

.

The pulse sequence of the BEST-TROSY version of the experiment is presented in Figure 6.15. The experiment, apart from the saturation block, is very similar to the 2D BEST-TROSY experiment (Fig.6.1) with the difference that here the last 180 ° pulse on ^{15}N after the acquisition period is omitted. In the reference experiment even minimal perturbation of water should be avoided. The considerations in the previous section showed that the BEST-TROSY pulse sequence at a well-performing probe ensure only minimal perturbation of water. In the saturation experiment the aim is to reach a nearly complete saturation of water. Almost constant saturation of water is reached by applying the water-selective saturating pulse at a constant frequency. Therefore, the duration of the relaxation delay (d_{relax}) has

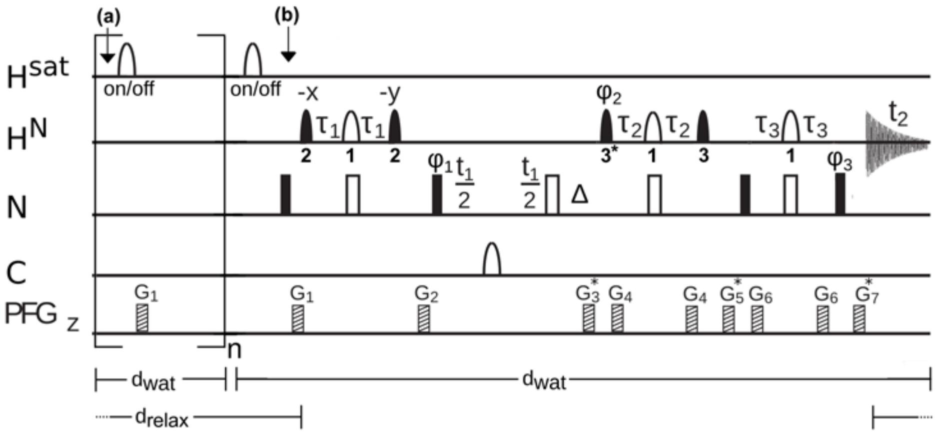


Figure 6.15: Pulse sequence of the HET^{ex}-BEST-TROSY experiment. The notation and the delays used are the same as in Figure 6.1. Water saturation is reached by series of on-resonance ^{1}H pulses of ISNOB-5 shape (150 ms at 600 MHz). The repetition time d_{wat} equals the duration of the sequence. The water magnetization can be read out at points (a) and (b) of the pulse sequence by application of a small-angle (1 °) read-out pulse preceded by a strong pulsed field gradient pulse.

to be integer multiple of the duration of the pulse sequence and the pulse on water is applied repeatedly during the relaxation delay. We found that an ISNOB-5 pulse of ∼ 150 ms achieves good saturation of water over a narrow bandwidth.

Before each experiment on a new sample, the bandwidth of the water selective pulse has to be calibrated in order to ensure minimal perturbation of ^{1}H$^\alpha$ protons, that resonate near the water frequency, but good saturation of the water protons (Figure 6.16, right panel). The saturation of water can to be measured at two points in the pulse sequence for obtaining information on the extent of saturation of water in the *sat* experiment and also it yields a measure of $I_{W,ref}(d_{relax})$ that is required for fitting of k_{ex} (Figure 6.16, left panel).

The accuracy of the measured exchange rates was cross-validated by measurement with the CLEANEX experiment. We used α-synuclein as the IDP example for validation of our approach. The CLEANEX experiment has been used to measure reference exchange rates. As can be observed from Figure 6.17 the correlation of the HET^{ex}-BEST-TROSY with CLEANEX measured exchange rates is very good. The reason is that NOEs are weak due to the flexible nature of the protein, thus the model used for fitting of the exchange rates is applicable. (As described in the manuscript found in Appendices the correlation for ubiquitin is very bad. The reason is the much higher amount of cross-relaxation in a folded protein compared to an IDP.)

The experiment has also been applied for measurement of exchange rates for the NS5A D2D3 protein, that possesses residual structure as will be

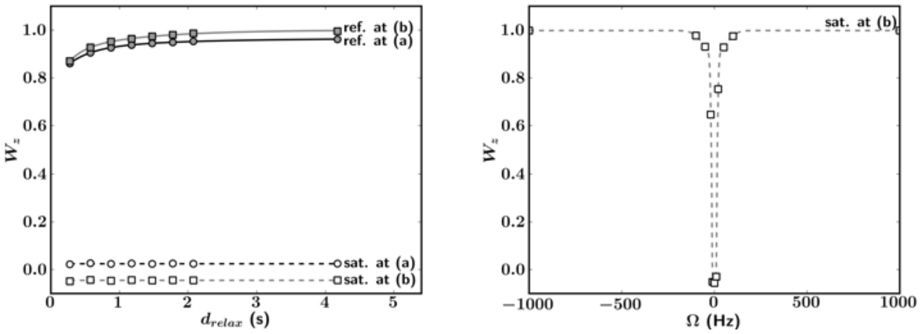


Figure 6.16: Normalized water polarization measured at points (a) and (b) of the pulse sequence (Fig. 6.15) in the *ref* (*filled symbols*) and in the *sat* (*open symbols*) experiments as function of the relaxation delay (*left*), and measured at point (b) in the *sat* experiment as function of the offset of the H^{sat} pulses (*right*).

shown in Part III. First, we were curious to see if the experiment allows measurement of exchange rates for a large IDP like NS5A with significant amount of residual structure. The second motivation was to find out whether the exchange rates obtained in this fashion can be used for identification of the transiently structured regions. CLEANEX measurement of the exchange rates was also attempted for cross-validation but even an experiment over a weekend did not provide sufficient sensitivity at the sample concentration we have to work with in the case of NS5A (150 μM), the data was too noisy to obtain good fits. However, the high sensitivity of the HET^{ex}-BEST-TROSY experiment allowed site-resolved measurement of solvent exchange rates.

$I_{HN,ref}(d_{relax})$, $I_{HN,sat}(d_{relax})$ and $I_{W,ref}(d_{relax})$ were measured at four d_{relax} values in an overnight experiment. Fits were good for residues with exchange rates up to $k_{ex} < 10$ s^{-1}, as shown in Figure 6.18. The reason for this upper limit is that for residues exchanging faster, recovery of the magnetization is too fast to be sampled even with the shortest recovery delay that we can use (˜200 ms). An important point to consider when measuring exchange rates is that depending on the amino acid type of the residue and of the neighboring residues, exchange rates can change by up to two orders of magnitude as illustrated by predicted exchange rates for NS5A D2D3 in Figure 6.18 upper panel, red curve. Therefore, the conditions for measurement of exchange rates have to be selected carefully so that it is in the optimal regime for the type of experiment used for the measurement. This can be a problem in case all the assignment experiments have been conducted at a temperature and/or pH which is not favorable for the measurement.

The measured and predicted exchange rates (k_{ex}) are plotted in Figure 6.18 . The plot of predicted versus measured exchange rates illustrates an issue of using exchange rates for structural characterization of IDPs. The exchange rates, because of the chemical properties of the amino acids vary

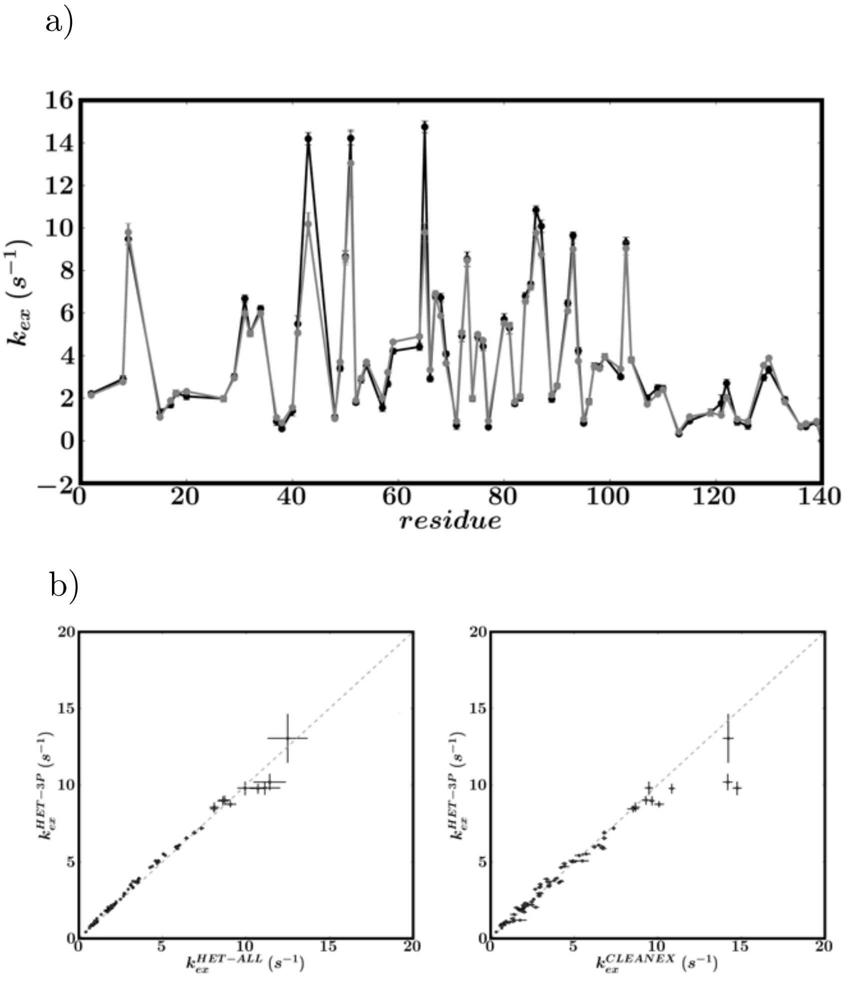


Figure 6.17: a) Measured exchange rates (k_{ex}) obtained from fit to HET$^{\text{ex}}$-BEST-TROSY data measured at three d_{relax} values (*gray*) and k_{ex} obtained from fits to CLEANEX data (*black*) are compared. b) Correlation of the extracted exchange rates obtained from measurements at the set of three d_{relax} to all d_{relax} delays (in range 0.2-4.2 s) HET$^{\text{ex}}$-BT data(*left*) and correlation of k_{ex} from three points HET$^{\text{ex}}$-BT data set to k_{ex} from CLEANEX data set are shown (*right*).

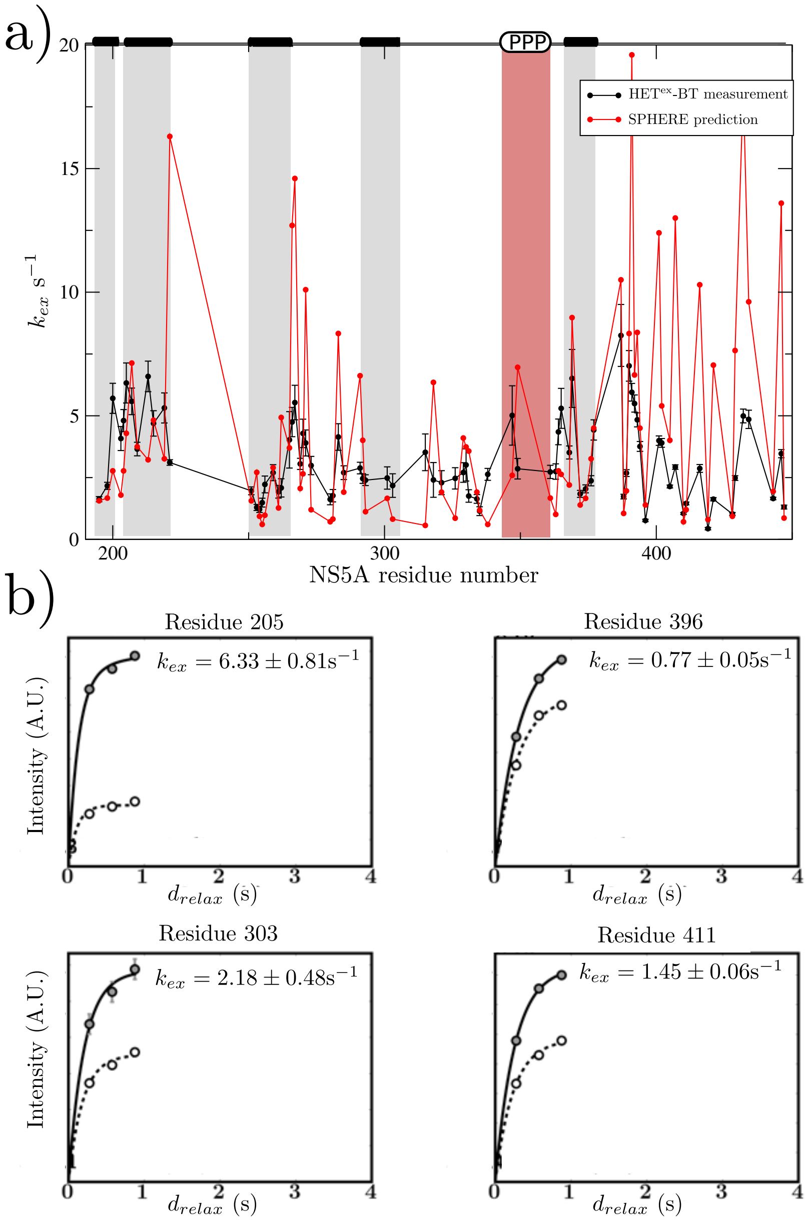


Figure 6.18: a) Solvent exchange rates for NS5A D2D3 (pH 6.5, 25 °C) obtained from fits to HET^{ex}-BEST-TROSY experiment measured at four d_{relax} delays (*black*) compared to predicted solvent exchange rates for the same conditions by the SPHERE predictor (*red*). Predictions are only shown for sites where we were able measure k_{ex} experimentally, for better visibility (very fast exchanging sites are not shown). The four transiently structured helical regions in NS5A are highlighted and indicated by cylinders on top. The polyproline region is higlighted in *red.* b) Representative fits to the data are shown.

over two orders of magnitudes. In case of a transiently formed helix that is, let us assume, formed in 50% of the time the exchange rate will drop about 50% for the residues of the helix, compared to the theoretical random coil exchange rates. In principle, all the measured exchange rates should be lower than the predicted values, as the scenario of the sample being less protected than the random coil is not meaningful. Apparently, the data shows, as judged from the large differences between measured and predicted exchange rates, the error of the prediction is larger than 50%, only the order of magnitude seems to be right for the different amino acids (note that for the faster than 10 s^{-1} exchanging residues the errors of the measurement were very large, so they were removed from the figure for clarity).

Measurement of exchange rates, however, can be used for identification of interacting regions that become buried. We have not yet performed this measurement, but comparison of exchange rates in the free IDP and in the complex is expected to give useful information on the interaction surface. It has to be emphasized, that for all such measurements it is extremely important to use exactly identical conditions (temperature, pH) as exchange rates are very sensitive to them.

6.6 Conclusions

In this Part of the thesis NMR method developments for IDPs were presented. These fast NMR methods all rely on longitudinal relaxation optimization as a common feature, in order to enhance experimental sensitivity and therefore reduce overall acquisition times.

The applicability of the BEST approach for longitudinal relaxation optimization for IDPs has been shown (Section 5.6) for different sample conditions. The two mechanisms of solvent exchange-induced reduction of apparent amide proton T_1 and the dipolar relaxation mechanism have both been shown to contribute to longitudinal relaxation enhancement by selective pulsing for this class of proteins. The contribution of the two mechanisms has been shown to depend on the sample conditions. In case of low pH the contribution from solvent exchange is negligible, nevertheless, amide proton T_1s still get significantly reduced via the dipolar relaxation mechanism. Therefore overall experimental sensitivity can be increased by fast pulsing. When conditions favor fast exchange (elevated pH and temperature) amide proton T_1s get reduced dramatically, a fact that can be exploited to boost experimental sensitivity for fast exchanging residues. However, in case of too rapid solvent exchange ($k_{ex} > 100\,\mathrm{s}^{-1}$) line broadening for the exchanging residues may so severely reduce sensitivity that the resonances cannot be detected. Therefore, with careful selection of conditions, longitudinal relaxation optimization allows working at elevated pH and temperature (*i.e.* near physiological conditions) with high experimental sensitivity.

The $H^{N-BEST}CON$ experiment (Section 5.8) developed in the course of this thesis combines the advantages of the above discussed longitudinal relaxation optimization by selective pulsing on amide protons and the larger signal dispersion for IDPs by ^{13}C-detection. We showed that this experiment is significantly more sensitive than the ^{13}C-start and -detected CON experiment as it profits from the larger gyromagnetic ratio of proton as polarization source and employs longitudinal relaxation optimization. This higher sensitivity comes at the expense of not detecting proline residues and limited applicability with respect to very high pH and temperature. We did not make a thorough comparison quantifying the extent, but its sensitivity remains lower than that of longitudinal relaxation optimized 2D ^{1}H-^{15}N correlation experiments such as the BEST-TROSY. The reason is its intrinsically lower sensitivity because of the lower gyromagnetic ratio of the ^{13}C nucleus. This experiment is expected to be more advantageous than for example the 2D BEST-TROSY when significantly more resonances are resolved due to the larger frequency dispersion of the ^{13}CO resonances. In order to be sure about using the most adapted experiment, both of them should be tested and optimized for the given protein sample and then compared with respect to sensitivity and resolution.

We showed in Section 6.2 that the BEST-TROSY experiments provide superior resolution and sensitivity than their BEST-HSQC counterparts. Furthermore, we showed on the example of two large IDPs, that BEST-TROSY provides high resolution allowing time-efficient resonance assignment with a set of 3D correlation experiments (Section 6.3). Using these experiments, comparably high resolution can be reached thanks to the TROSY-effect even without need for expensive deuteration. The BEST-TROSY-effect also provides increased sensitivity especially for transiently structured regions and results in a more homogeneous distribution of intensities. Therefore this class of experiments can be useful in NMR studies of IDPs characterized by large peak intensity distribution caused by broadening resulting from residual structure or solvent exchange.

Furthermore, the Pro-BT-HNcocan and Pro-BT-iHNcan experiments have been introduced for selective identification of residues neighboring proline. These experiments also profit from all the advantages of BEST-TROSY. They can be helpful for assignment of IDPs with large content of prolines (typical for IDPs). Identification of these residues provides starting points for amino acid-type anchoring of fragments and prevents loosing time searching for correlation peaks next to proline residues. The information provided by the experiment could also be included into programs for automatic backbone assignment to resolve ambiguities. These experiments provide complementary information to the amino acid-type information that can be obtained from the HADAMAC and iHADAMAC experiments, as there prolines belong to the "rest" group, and so are not identified selectively.

We showed that time-efficient determination of solvent exchange rates is possible with the HET$^{\mathrm{ex}}$-BEST-TROSY experiment (an optimized version of the HET$^{\mathrm{ex}}$-SOFAST experiment [121]), for IDPs. This experiment allows accurate measurement of k_{ex} in the range of $0.1 < k_{ex} < 10$ s^{-1}. The exchange rates should be interpreted with care, for example comparison with predicted values is not sufficiently accurate to conclude about protection factors for IDPs. However, measurements of k_{ex} can yield useful information for monitoring changes in the structure due to chemical modification, interaction or other types of perturbation.

Part III

Biophysical studies of NS5A protein of HCV

Chapter 7

Introduction, molecular biology of HCV

7.1 Hepatitis C disease

Hepatitis C disease currently affects about 130-170 million people around the world. Non-A, non-B hepatitis was described for the first time in 1975. A few years later interferon-based therapy was introduced for hepatitis C patients. Acute hepatitis C infection results in clearance from the virus by natural immune response only for about 30% of the affected individuals. Thus, for the majority of the patients the infection persists and develops into chronic hepatitis C disease, in which stage spontaneous clearance of the virus is extremely rare [122].

Within a few decades after the infection patients start to develop liver cirrhosis, that in some cases results in hepatocellular carcinoma. These in turn result in liver failure and only liver transplantation can save the patient. The traditional treatment of hepatitis C disease is interferon-based therapy, by serum introduction of recombinant interferon-α (IFNα).

However, IFNα therapy has limited efficiency, and research efforts focused on understanding determinants on IFNα-sensitivity and viral escape. Later on, the first atomic resolution structures of HCV proteins have been solved (for details see Section 7.3). Since then, more detailed understanding of the viral life cycle led to the development of new antivirals that target directly the viral proteins, and other drugs targeting host factors essential for the life cycle of hepatitis C virus.

Despite these new therapies, there is still no generally effective treatment for HCV, neither a vaccine available. Therefore, in addition to fundamental interest, a more detailed, molecular level understanding of the viral life cycle is also necessary for clinical applications.

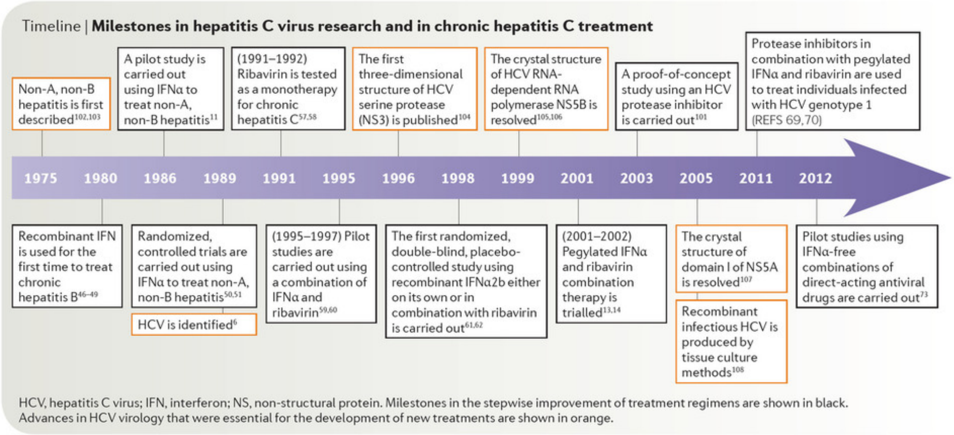


Figure 7.1: Milestones in hepatitis C research and treatment. Adapted from Ref. [122]

7.2 Hepatitis C virus

Hepatitis C virus has been isolated in 1989 [123]. The virus has been classified into the *Flaviviridae* family of viruses, in the *Hepacivirus genus*. It is a positive stranded RNA virus, meaning that the RNA found in the virion is positive-sense RNA. This means that (in contrast to negative stranded RNA viruses) its RNA can be directly translated after entry into the host cell, because the positive sense viral RNA is similar to mRNA. Replication of the viral RNA requires viral proteins and host cofactors. Replication takes place in HCV-infected cells at a very high copy rate resulting in many errors and thus high genetic variability. This high replication rate is necessary because lifetime of the virus is short. To date, six genotypes have been distinguished with several subtypes. The virus, after entry into the cell, uses the host machinery to express its single viral polyprotein that is then further processed by viral and host proteases. The life cycle of HCV takes place on membranes, all the viral proteins are membrane associated. Viral replication takes place on the so-called membranous web, formed from the endoplasmic reticle (ER) of the host cell. On the other hand, viral particle assembly takes place on lipid droplets.

7.3 HCV proteins

The HCV genome encodes ten proteins, three structural and seven non-structural proteins as illustrated in Figure 7.2. The structural proteins of HCV, that build up the virion, are core protein, E1 and E2 envelope glycoproteins. A host-derived membrane also plays a role in building up the virion. To date, there is no data on the structure of the virion. Core protein

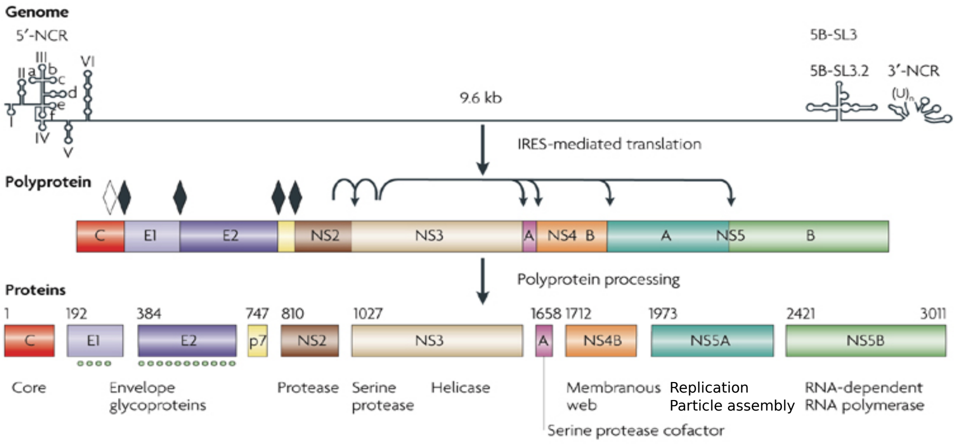


Figure 7.2: Genome organization of the viral strucutral and non-structural proteins. The figure has been adapted from Ref. [124]

has been relatively well characterized. Mature core protein consists of two domains. Domain 1 is intrinsically disordered when expressed alone, containing many positively charged residues. Similarly to the other core proteins of Flaviviridae it has RNA chaperoning activity [125], [126]. It is also involved in interactions with numerous host factors [127]. The C-terminal domain 2 is largely hydrophobic and dimeric, via disulfide bond formation by C128. Core has also been reported to interact with domain 3 of NS5A [128]. This interaction has been shown to be dependent on phosphorylation of certain residues in domain 3 by mutation analysis. E1 and E2 contain intrinsically disordered regions, an atypical feature for viral structural proteins. They play roles in viral entry and, evidently, in building up the virion.

The other proteins of HCV are non-structural proteins. p7 is a viroporin, an oligomeric membrane protein that has been shown to function as a cation channel and is necessary for productive infection. In addition to its role as an ion channel it has been reported to interact with Core and to play a role in particle assembly. Its structure has been recently solved by liquid-state NMR spectroscopy [129], giving an impulse for drug design studies to find inhibitors. NS2-NS3 acts as an autoprotease, after their self-cleavage they fulfill their roles individually. The structure of the catalytic domain of NS2 has been solved. It is a cystein protease, dispensable for RNA replication but crucial for formation of free NS3. NS3 protein is multifunctional, its structure has been solved. Its N-terminus is a serine protease, and its C-terminus is the ATP-dependent RNA helicase forming a complex with NS4A. NS4B is an integral membrane protein. It plays a role in the formation of the membraneous web, but the exact mechanism is unknown. NS5A we will discuss in detail in the next section, and NS5B is the RNA polymerase. The

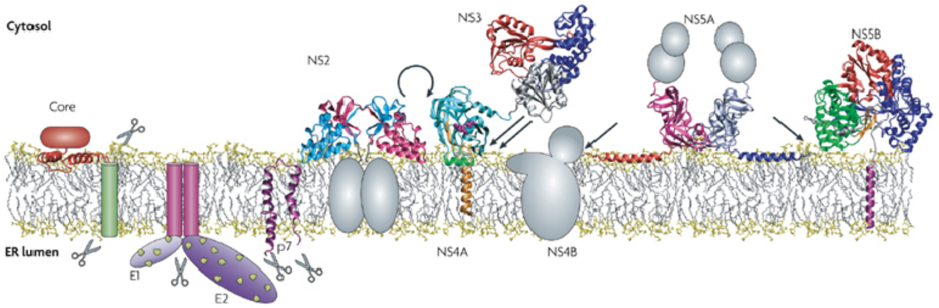


Figure 7.3: Illustration of membrane organization of the HCV proteins. The figure has been adapted from Ref [124]

latter is anchored to the membrane by its C-terminal integral membrane protein domain. Deletion studies showed that from all viral proteins, NS3 – NS5B are required, in addition to host factors. Figure 7.3 shows membrane organization of the HCV proteins and the available structural information, with the exception of p7. For p7 the figure shows the structure of the monomer, but not the high resolution structure of the oligomeric ion channel that has only been solved by NMR spectroscopy very recently [129].

7.4 Roles of NS5A

NS5A is a multifunctional phosphoprotein, existing in two differentially phosphorylated forms, termed basal and hyperphosphorylated. This differential phosphorylation has been suggested to act as a switch between its functions in RNA replication and viral particle assembly. Basal phosphorylation has been located to residues mainly in domain 3 and some in domain 2. Hyperphosphorylation has been mapped to residues of the cluster of serines in domain 2, for a more detailed discussion see Section 10.3. It has been recognized, that mutations that impair hyperphosphorylation of residues in the serine cluster of the LCSI region result in increase of replication efficiency [130]. This and other observations led to the hypothesis that NS5A is a key regulator of the HCV life cycle and the differential phosphorylation of its two forms mediates the switch of functions between replication and particle assembly. Furthermore, NS5A interacts with the largest number of host factors out of all the HCV proteins [127] in addition to its interactions with the other viral proteins. Presumably its flexible, intrinsically disordered domains account for this ability of binding to many partners. A widely known advantage of intrinsic disorder is to allow binding to a multitude of binding partners often mediated by preformed structural elements or short recognition motifs. All these elements are exposed due to conformational disorder. One example of its interaction with a host factor is that with cyclophilin A, which is crucial for RNA replication of the virus [131]. This interaction already gave rise to an anti-HCV drug under development targeting cyclophilin A.

NS5A has been divided into three domains based on low complexity sequences identified by bioinformatic analysis. This model was confirmed by limited proteolysis and mass spectrometry [132]. Proteolytic cleavage of NS5A yielded two major cleavage sites in the low complexity sequences and two major protease-resistant products were identified by mass spectrometric analysis, 25-215 and 25-355 that correspond to domain 1 and domains 1+2, respectively. The amphipathic α helix that is semi-embedded in the membrane precedes domain 1. The structure of domain 1 has been solved by X-ray crystallography, revealing a dimeric structure and zinc-finger motif [133]. The dimeric interface has been suggested to form an RNA-binding groove that could accommodate single- or double-stranded RNA as well. NS5A has been shown to be an RNA-binding protein [134], thus this finding gave rise to further speculations on the role of NS5A. In one scenario pairs of NS5A dimers form a basic surface at the membrane that serves like a 'railway' to transport viral RNA [135]. In another hypothesis NS5A serves for translocating RNA from the site of viral replication at the membrane the the site of particle formation at the lipid droplets. According to recent findings this translocation of NS5A and possibly the bound RNA to the lipid droplets is mediated via interaction of NS5A with DGAT1 [136]. A few years later another crystal structure of domain 1 has been solved, with a different dimerization interface [137]. Electrostatic mapping of this structure shows no clear possible RNA binding site [134]. The role of these different dimeric structures remains unclear. Domains 2 and 3 of NS5A have been reported by us and by others to be intrinsically disordered [112], [138], [139], [140].

Deletion studies have shown that domains 1 and 2 play a role in RNA replication while domain 3 can be completely deleted without major effect on replication [130]. Domain 3 on the other hand has been shown to be required for particle assembly, and it mediates the interaction to the core protein, where phosphorylation of S457 seems to play a role in the interaction [128].

As major topic of this PhD we conducted studies to characterize the intrinsically disordered domains of NS5A of HCV, their interaction with Bin1 SH3 domain and their phosphorylation, to gain further understanding about the structure and functions of this multifunctional protein. This will be presented in the following chapters.

Chapter 8

Structure and dynamics of the intrinsically disordered region of NS5A

In this chapter, we present our results on the characterization of structure and dynamics of the intrinsically disordered region of NS5A. The study was carried out by NMR spectroscopy and SAXS in the course of this thesis.

8.1 Domain organization of NS5A and our constructs

As mentioned in the previous Section, NS5A protein of HCV has been divided into three domains by bioinformatic analysis and limited proteolysis as shown in Figure 8.1. The protein is anchored to the ER membrane by an amphipatic helix of ~ 30 residues. Three domains follow (D1 D2 and D3) that are separated by low complexity sequences 1 and 2 (LCS1 and LCS2). LCS2 contains two class 1 and one class 2 polyproline motifs (PxxP motifs).

The IUPRED disorder prediction is shown here for the full-length protein. As shown in Figure 8.1 IUPRED predicts low disorder score for D1, but from residue 180 the score increases. The construct of domain 1 that yielded the crystal structure [133] ends at residue 198 and in the crystal structure the final residue of the last (B9) β-strand is Y182. The following residues 182-198 in the structure have no well-defined secondary structure, consistently with the increasing disorder score. An interesting feature from the crystal structure is the disulfide bond that connects the conserved cystein C190 to C142 that is also located in a loop. The N-terminal residue of our constructs is E191.

IUPRED predicts disorder scores around 0.5 for D2 with small variations and higher disorder scores for D3 except for a short region at the N-terminal

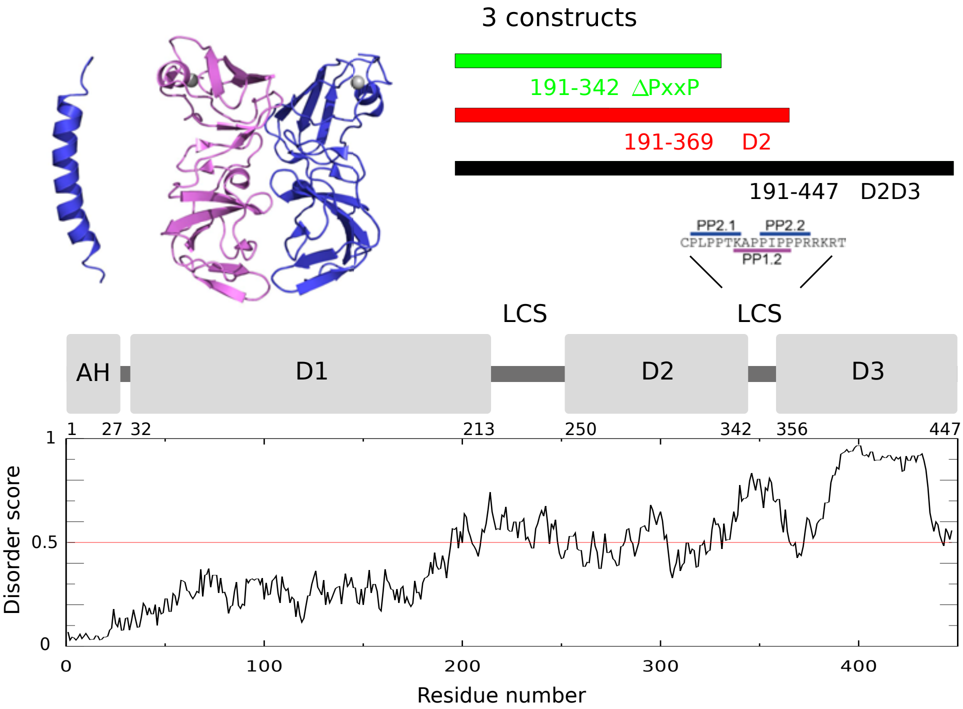


Figure 8.1: Domain organization of NS5A protein of HCV according to Tellinghuisen *et al.* is shown [132]. The domains, low complexity sequences and polyproline motifs are highlighted. Disorder scores are shown underneath as predicted by IUPRED. The structures of the N-terminal amphipatic helix (PDB entry 1R7C) and of domain 1 are shown (PDB entry 1ZH1). Our three constructs are shown above the domain organization with the residue numbers.

part of the domain which is shown by inspection of the sequence to be hydrophobic.

The D2 construct was studied in the group prior to this thesis and at the beginning of this thesis, before the D2D3 construct has been expressed and purified. The data on that construct was published in ref [112] that is included in Appendices. The construct was designed to include the polyproline motifs of LCS2 in order to allow study of its interactions with SH3 domains (Chapter 9).

A shorter construct, ΔPxxP was designed to study in detail the non-canonical interactions we found by interaction studies with SH3 domains.

The largest construct D2D3 comprises the entire intrinsically disordered region of the protein starting from residue E191.

The background of disorder predictors was discussed briefly in Chapter 1. We performed predictions for our largest construct, D2D3 with IUPRED and PONDR®. Both of them are available on web servers. The result of the predictions for the D2D3 construct is shown in Figure 8.2. At first glance

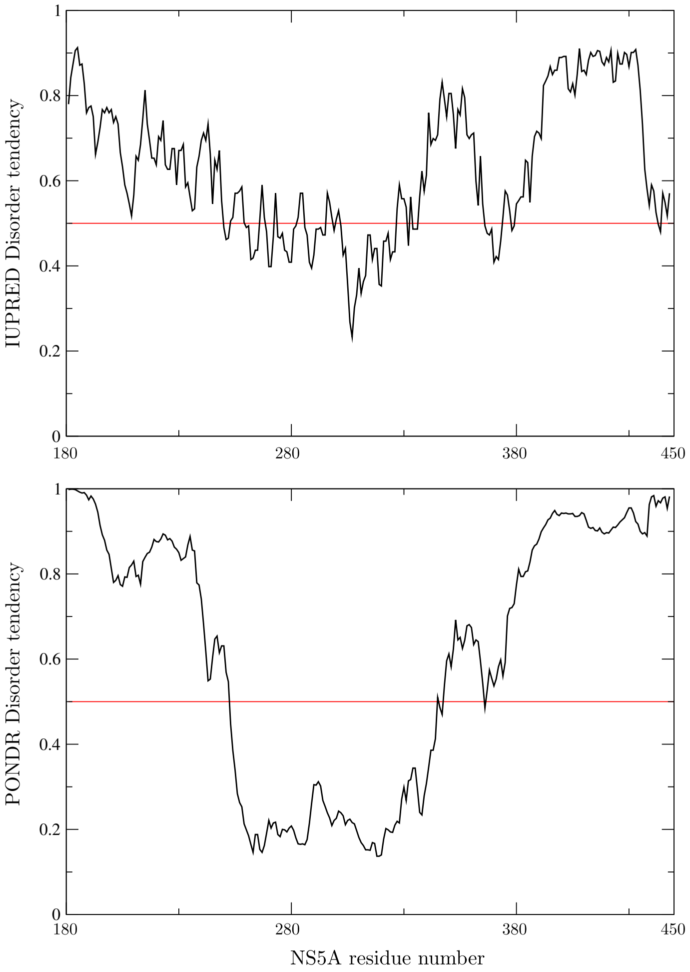


Figure 8.2: The results of the IUPRED and PONDR® sequence-based disorder predictors for the D2D3 construct of NS5A. IUPRED is shown on top panel, PONDR® on the lower panel. Scores below the *red line* (lower than 0.5) indicate tendency for order and above the *red line* tendency for disorder.

the two outputs seem similar, D3 is predicted to be highly disordered while D2 has lower disorder tendency. As predicted by IUPRED, the disorder tendency is around 0.5 for D2, while PONDR® suggest an even lower disorder tendency. As will be seen later, these predictions are consistent with our findings about the differences in dynamics and the extent of residual structure in the disordered regions of NS5A. This will be shown and discussed in detail in Sections 8.5 and 8.6.

8.2 Expression and purification of NS5A constructs

The amino acid sequence of the subject of our studies is from HCV genotype 1b HC-J4 strain deposited in the European HCV database. We expressed and purified three constructs, D2, D2D3 and ΔPxxP.

The plasmid of D2 came from our collaborators (Dieter Willbold at FZ Juelich). Expression, purification and sample conditions were optimized by Amine Aladag and Sophie Feuerstein prior to this thesis. The protocol is described in reference [112]. Here it suffices to note that the protein is expressed in *E. Coli* and that a GST-fusion allowed purification by affinity chromatography and a further purification step consisted of size exclusion chromatography.

The ΔPxxP construct was expressed and purified by Melanie Schwarten as described in Ref. [141].

The D2D3 construct was also first ordered as a GST-fusion and expressed and purified according to the protocol for D2. However, yields of the GST-tagged D2D3 construct were unreproducible and very low, under 1 mg protein from a liter of culture. Thus we decided to further optimize the construct and we ordered a His-tagged gene. However, we found from NMR studies that the active protein concentration was lower than measured by UV-Vis spectroscopy. We concluded that this must be due to the His-tag as this was the only change in the construct compared to the GST-tagged one. Thus we decided to subclone the gene into a vector with a cleavable His-tag.

Expression tests were conducted at four different temperatures and different expression times in four different *E. Coli* strains: BL21(DE3), BL21(DE3) RIL, BL21(DE3) STAR and BL21(DE3) pLys. The expressed NS5A protein on the expression gels is not visible, as it migrates together with a major bacterial protein. Thus we detected it for quantification of the expression efficiency with Western Blotting, using a His-tag antibody. The resulting optimized protocol can be found in Appendices.

As we had severe problems with quick degradation of the sample, its stabilization by heat denaturation of the contaminant proteins was tested. IDPs are known to resist high temperatures, several IDP purification protocols exploit this property [142]. After 20 min heating at 90 °C a 2D BEST-TROSY spectrum was recorded and compared with the reference spectrum. Small changes could be observed around the cystein residues, that suggested oxidation of the sample had occurred. Presumably DTT did not resist the high temperature. Fresh DTT was added to the sample and a 2D BEST-TROSY spectrum was recorded afterwards. The spectrum was found to be identical with the spectrum we had recorded previous to the heating as shown of Figure 8.3

Later on, further improvement was made to the protocol. In the original protocol β-mercaptoethanol was used as reducing agent. This was replaced

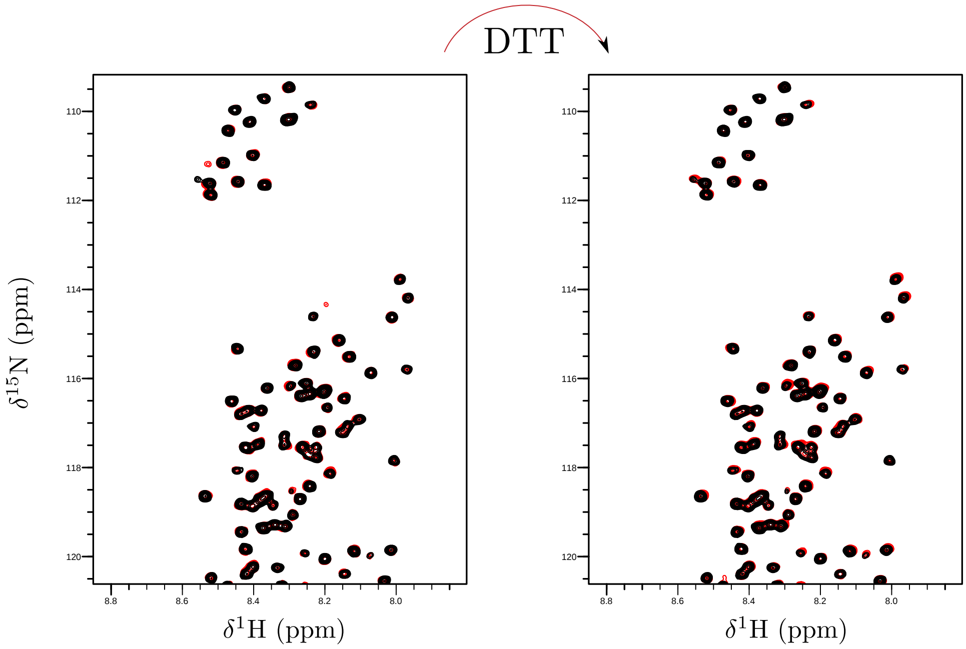


Figure 8.3: BEST-TROSY spectra recorded at 800 MHz field and 5 °C on the NS5A D2D3 sample for testing its thermal stability. The reference spectrum is shown in *black*, the spectra recorded after heating are shown in *red*. On the left side before, on the right side after addition of fresh DTT.

by DTT, which is a stronger reducing agent, as we noticed oxidation of the sample. In the course of subsequent NMR studies even the samples containing DTT got oxidized relatively fast. This is due to the fact that at pH 6.5 DTT is not efficient as a reducing agent, because the thiol groups are protonated at acidic pH. Thus the reducing agent was replaced by TCEP, which is also active at acidic pH. As the sample was oxidized very readily a purification was performed with DTT in the lysis buffer (in contrast to the original protocol that contained no reducing agent at that stage). This modification improved the yield four-fold, instead of 1.25 mg/l culture, 5 mg/l culture protein could be purified.

The buffer conditions for the NS5A studies were chosen before the start of this project. For the sake of comparability with data on the other constructs the conditions were chosen to be identical with those of previous studies.

8.3 Assignment of the NS5A constructs

The three NS5A constructs have been assigned at two temperatures, 5 °C and 25 °C. The 2D BEST-TROSY spectrum, serving as the basis of the assignment, was shown in Figure 6.2. Sequential connectivities were then

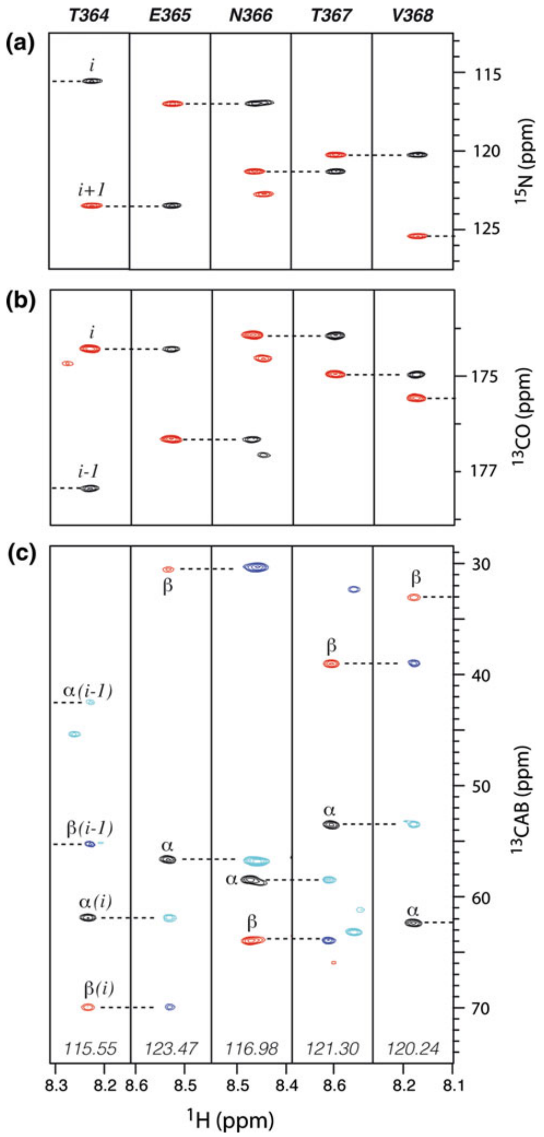


Figure 8.4: Establishment of sequential connectivities for NS5A is illustrated on residues 364-368 based on a) BT-3D hNcocaNH, b) BT-3D hnCOcaNH, and c) BT-3D HNcoCACB and iHNCACB spectra, recoded at 800 MHz field and 5 °C temperature. The spectra were recorded with parameter settings as described in Ref. [109]

identified using the 3D BEST-TROSY experiments presented in Section 6.3. Example strips of the spectra used for assignment are shown in Figure 8.4

After sequential connectivities had been identified, HADAMAC [111] and iHADAMAC [143] experiments were used to facilitate amino acid-type anchoring. The HADAMAC spectra are composed of seven subspectra. In each of them correlation peaks show up depending on the amino acid class of the protein. The seven classes and the bases or their distinction together with an illustration of the use of the experiment are shown in Figure 8.5 a. The HADAMAC and iHADAMAC spectra of NS5A D2D3 are shown in Figure 8.5 b.

Nearly complete backbone resonance assignment has been obtained for all the three constructs at 5 °C and for D2D3 at 25 °C. The assignment of D2 has been published [144] and deposited in the BMRB.

8.4 Characterization of residual structure in the D2 construct

Most of the work that is presented in this section on the D2 construct has been done by Sophie Feuerstein before this PhD project started (with some contribution during my internship at IBS at 2010 at at the beginning of this thesis). It is included here because it was the state of the project when this PhD project started.

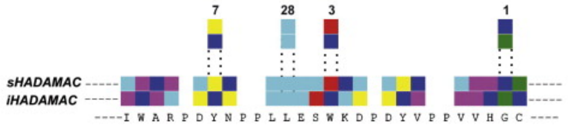
Residual secondary structure was characterized by secondary chemical shifts and ^{15}N relaxation data (Fig.8.6(a,b,c)). Three transiently structured helices were identified in this D2 construct (H1, H2 and H3) from chemical shifts, a finding supported by ^{15}N relaxation data showing higher rigidity for these regions. Secondary structural propensities were estimated by the SSP score (positive values account for α-helical propensities while negative values for β-strand or extended propensities) to be $\sim$40% for H1 and H2 and $\sim$50% for H3. The resulting propensities were compared to propensities predicted by the AGADIR [145] program (Fig.8.6 f). AGADIR predicts helical propensities taking into account pH, temperature and ionic strength. The AGADIR output shows significantly lower helicity for H2 and H3 than what was found experimentally. This may be explained by transient tertiary interactions stabilizing these helices. PRE data that can be found in the publication supports this view, although, as discussed in Section 4.3.4, interpretation of these data for IDPs is not straightforward, and more spin labels would be necessary to get a more detailed picture about the structural ensemble this IDP adopts.

Having characterized the D2 construct we continued the project with expression, purification and subsequent structural characterization of the D2D3 construct during this PhD thesis.

8.5 Secondary chemical shifts

As it was discussed in Section 4.3.1 chemical shifts are sensitive for secondary structural conformations and therefore they can be used to analyze residual secondary structure in IDPs.

a)



Amino-acid "band"	*Gly*	*Val-Ile/Ala*	*Thr*	*Asn-Asp*	*Arom-Cys*	*Ser*	*Rest*
Topology class	1	2 / 3	4	5	6	6	7
H-multiplicity (C^α)	2	1	1	1	1	1	1
H-multiplicity (C^β)	-	1 / 3	2	2	2	2	2
carbonyl C^γ	0	0	0	1	0	0	0
aliphatic C^γ	0	2 / 0	1	0	0	0	1
C^β (60-80 ppm)	no	no	yes	no	no	yes	no

b)

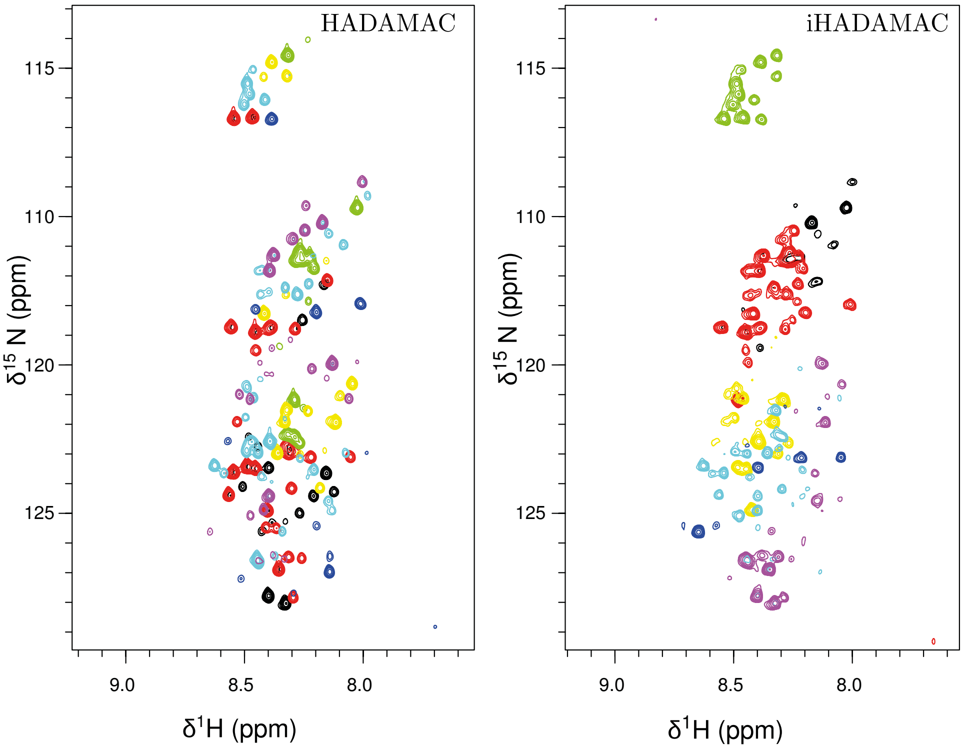


Figure 8.5: a) Illustration of the amino acid-type anchoring with the color-codes used for amino acid-type identification of the ^{1}H-^{15}N correlation peaks. Amino acid-type color-codes and the seven subclasses are shown below. b) HADAMAC and iHADAMAC spectra of NS5A D2D3 recorded at 5 °C at 800 MHz field.

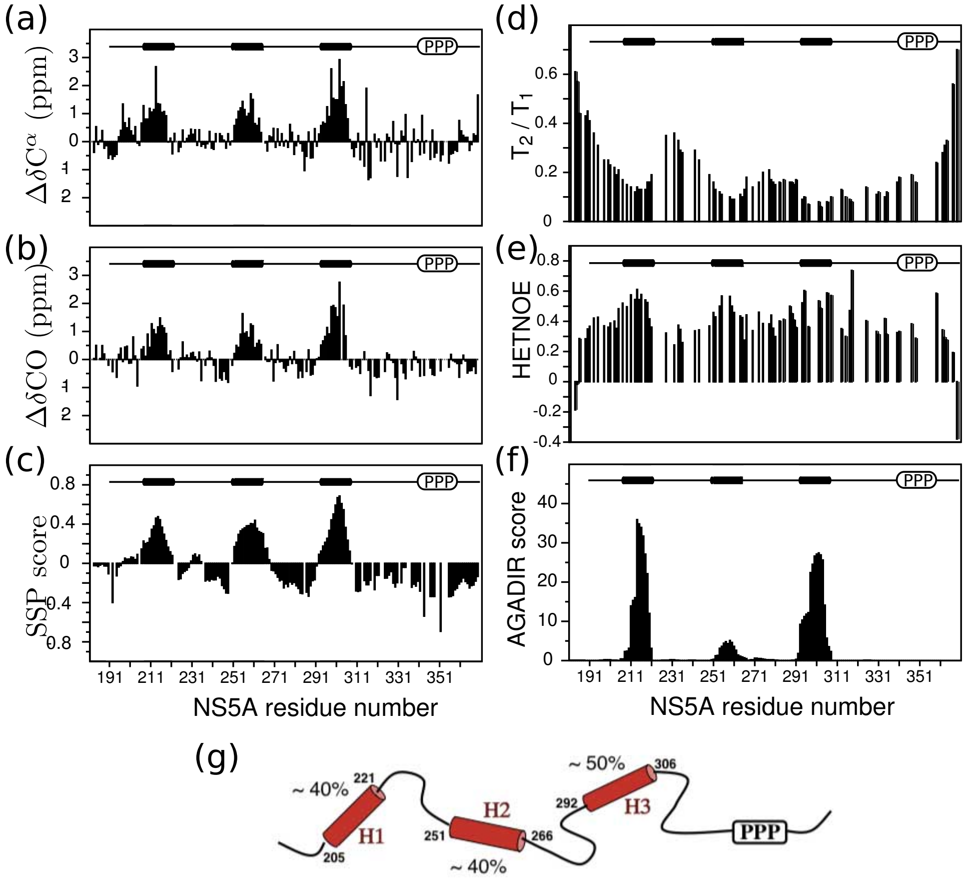


Figure 8.6: "Structural and dynamic characterization of NS5A(191369) (D2) at 5 °C from NMR chemical shift, and ^{15}N relaxation data. (a) C$^{\alpha}$ and (b) CO secondary chemical shifts calculated as the difference between measured and random-coil chemical shifts from the refDB database [60] and corrected for next-neighbor effects [64]. (c) Structural propensities calculated with the SSP program [146]. All available chemical shifts (HN, N, C$^{\alpha}$, C$^{\beta}$, and CO) have been used as input data for the SSP calculation. Positive score values indicate the amount of α-helical conformation present along the sequence, while negative values are indicative of extended or β-strand conformations. (d) Ratios of measured transverse and longitudinal ^{15}N relaxation times (T2/T1) providing an estimate of the local rotational correlation times τ_c along the peptide chain. (e) $\{^1\text{H}\}^{15}$N NOE (HETNOE) values indicative of the amplitude (order parameter) of local subnanosecond motions (high values: restricted motion, low values: high-amplitude motions). (f) Prediction of helical propensities based on the physical chemistry of the polypeptide chain using the program AGADIR [145]. (g) Structural model derived from the recorded NMR data showing three α-helical segments of similar length that are formed with propensities of ∼40% for H1 and H2 and ∼50% for H3, connected by flexible linker regions." As published in Ref. [112].

In this work three different ways were used to calculate secondary chemical shifts for NS5A D2D3. The first one is the web server at NIH for calculating random coil chemical shifts according to the studies by Kjaergaard and Poulsen. The correction factors here are based on peptide studies [65], [62] yielding glutamine-based neighbor correction factors and temperature and pH correction factors. The second one is using the IDP-based database of Tamiola *et al.* [63]. The third one is using glycine-based neighbor-correction factors and the average chemical shifts of refDB (from coil regions of protein structures) that was used for characterization of D2, presented in the previous section.

The results obtained by the three different random coil chemical shifts are shown in Figure 8.7. Several regions display tendencies for residual secondary structure as shown by the same sign of secondary chemical shifts over a stretch of residues. Comparing the three profiles, the one shown on the bottom panel (Kjaergaard and Poulsen) shows the smallest deviations from zero for regions not characterized by residual secondary structure, therefore we decided to use this data set for subsequent analysis. All the measured secondary chemical shifts using the random coil data according to Kjaergaard and Poulsen are plotted in Figure 8.8.

Now we continue with a detailed analysis of the secondary chemical shifts of NS5A measured at 25 °C based on Fig. 8.8. The secondary chemical shifts against the random coil values calculated from the C^α and C' secondary chemical shifts are more sensitive to α-helical conformations, as they result in larger maximum shifts. The average C^α secondary chemical shift for a fully formed helix is around +2.5 ppm, while for β-strand it is around -1.4 ppm. C' secondary chemical shifts for α-helix are on average +2 ppm, for β-strand around +1.5 ppm On the other hand C^β and H^α shifts are more sensitive to extended conformations, with values of about +0.4 ppm H^α secondary chemical shift and +2.2 ppm C^β secondary chemical shifts for β-strand conformations and -0.4 ppm H^α secondary chemical shift and also about -0.4 ppm C^β secondary chemical shift for α-helical conformation.

Starting from the N terminus, inspection of the secondary chemical shifts leads to the following conclusions about transient secondary structure. Higher than random coil helical sampling starts from residue D195, which is an aspartate, typical N-capping residue. N-capping residues have a characteristic C^α and C^β signature: C^α shifts 1-2 ppm upfield and C^β shifts 1-4 ppm downfield [147] with the following amino acids showing the typical secondary α-helical chemical shifts. Because of the presence of the proline at residue 206, the helix either breaks at residue S205 or there is a kink in the helix. Based on the secondary chemical shift data it seems more likely that the helix breaks. This first region, with helical propensity, we denote with H1a. From residue 206-221 the secondary chemical shifts report on increased α-helical propensity (H1b).

A region with random-coil secondary chemical shifts follows transient

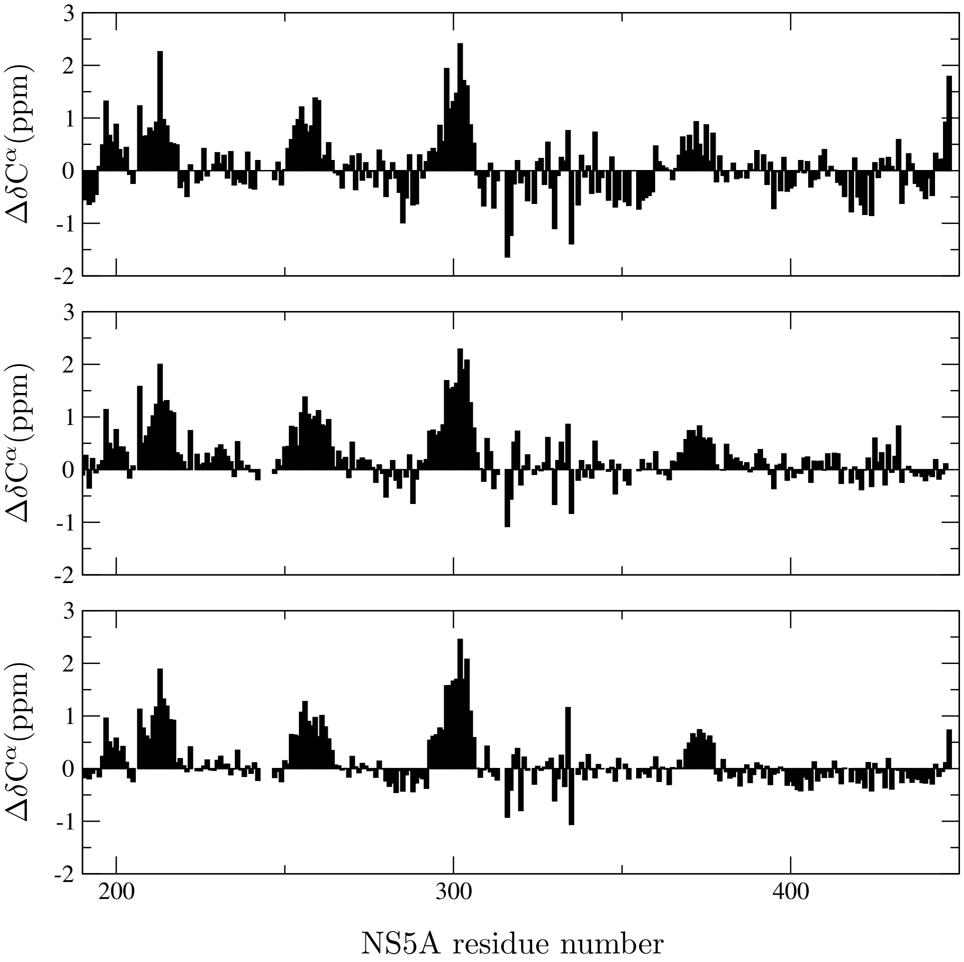


Figure 8.7: C^{α} secondary chemical shifts of NS5A D2D3. The difference of the measured chemical shifts and random coil values are plotted from top to bottom according to references ([60] and [64], [63], [65] and [62])

H1 until residue 251, which has increased helical propensity as well. This transiently populated helical region 2 (helix 2 or H2) ends at residue 266.

Between residues 279-291 the secondary chemical shifts indicate a slight tendency for β-strand formation (S1).

Residue D292 is characterized by a typical N-capping chemical shift signature. Inspection of the data for this transiently helical region 3 (helix 3 or H3) shows that at S297 there is an overall drop of helical propensity. It is very probable, that a shorter helix is formed with a N-capping by S297. However, in this case the N-capping signature is not visible, because of the effect of the conformations of the longer helix on the chemical shifts. Residue 299 is a proline. A proline residue because of steric clash always causes a kink in the helix or breaks the helix. In this case chemical shifts show a

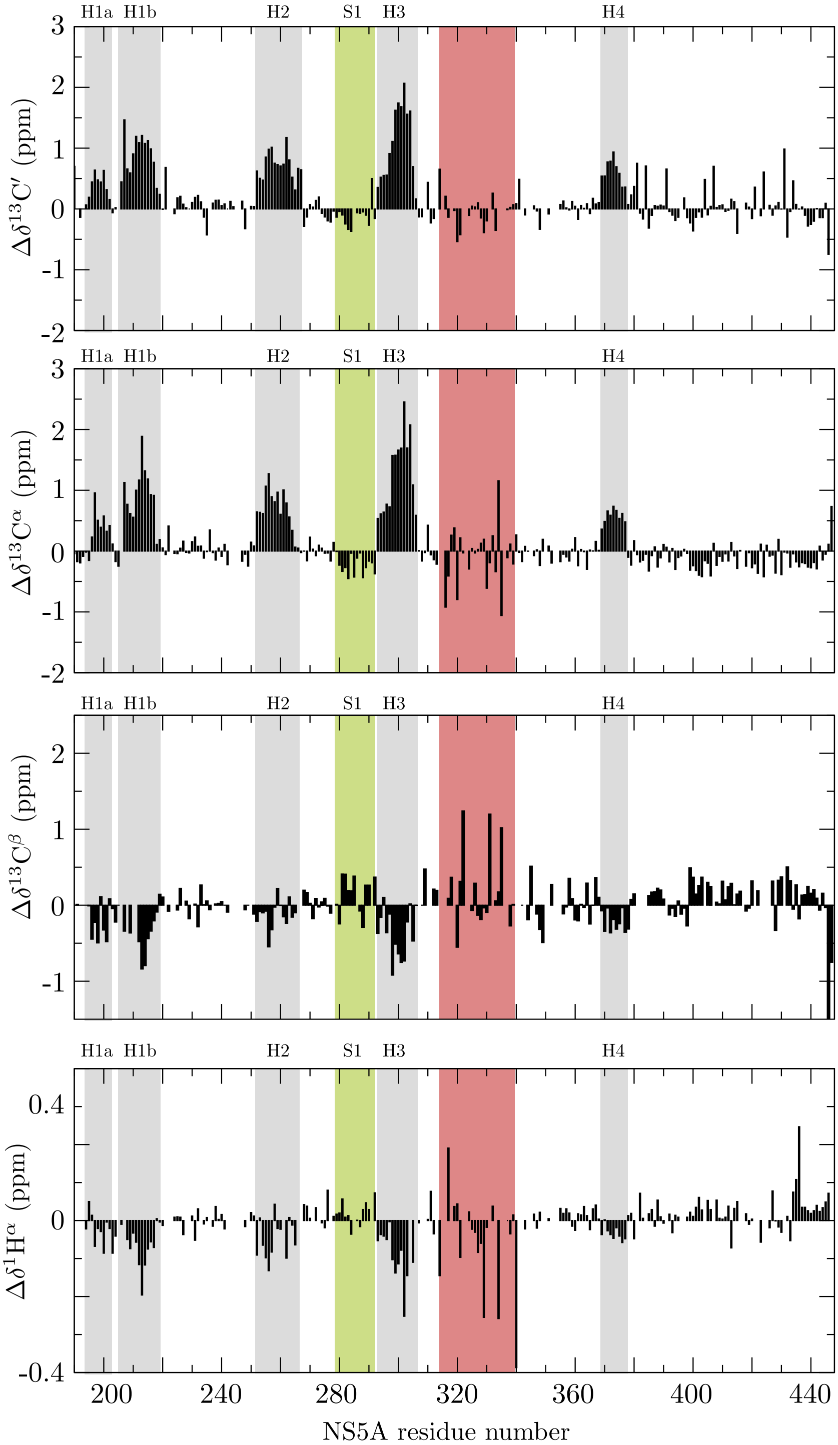


Figure 8.8: C^{α} C^{β} C' and H^{α} secondary chemical shifts of NS5A D2D3. The difference of the measured chemical shifts and the predicted random coil values (according to Kjaergaard and Poulsen [62, 65]) are plotted. The regions with transient secondary structure are highlighted. Transient α-helices in *gray*, a transient β-strand in *green* and the region showing a particular secondary chemical shift pattern is highlighted in *red*.

continuous stretch of helical secondary shifts, thus it is more likely that a kink is present in the helix. The end of H3 is residue 306.

H3 is followed by a region with a particular secondary chemical shifts pattern, highlighted red in the figure. As we will see from the dynamics data, this region is also characterized by increased rigidity. Actually, a part of this region (315-320) is the one that has been reported to interact with cyclophilin A (CypA) and with NS5B [148]. This region contains proline residues whose cis-trans isomerization is catalyzed by CypA. The region (residues 315-334) characterized by this particular chemical shift signature will be denoted in this thesis with CypAIR (for CypA interaction region). However, in our construct we only observe one set of peaks for each residue, therefore corresponding only to one isomer.

The polyproline region probably adopts PPII conformations, but this could only be identified based on ^{15}N chemical shifts, combined with RDC data. This region is followed by domain 3. At the beginning of domain 3 there is a region of 8 residues (residues 370-378) with chemical shifts characteristic for increased α-helical sampling (H4). This stretch of amino acids is largely hydrophobic, in contrast to the other transiently populated helices. The N-capping residue is a serine, again a typical N-capping residue while C-capping is probably the lysine, that is favorable because of its positive charge. The rest of D3 possesses no residual secondary structure, which could be clearly identified from secondary chemical shifts.

To be able to compare the data with that of D2 (Fig. 8.6) and to obtain a quantitative estimation of structural propensities, the SSP score has been calculated from all available chemical shifts [146] and plotted with the CO and C$^{\alpha}$ secondary chemical shifts using the refDB random coil shifts as was the case for D2. (Let us note, that the SSP program uses the refDB chemical shifts for random coil values that are not the best adapted for IDPs. In addition, the 5 residue-frame smoothing might cause artifacts so the data should be interpreted with care, especially concerning minor structural propensities.) Comparing the secondary chemical shifts in Fig. 8.8 for the helices yields relative propensities. The SSP score allows approximate estimation of absolute propensities (data shown in red in Fig. 8.9). Combining the available information, the propensities of the transiently formed helices are estimated to be approximately 15% for H1a, 40% for H1b, 35% for H2, 60% for H3 and 20% for H4.

The assignment, as mentioned before, has been carried out at two temperatures, at 5 °C and 25 °C. The differences of secondary chemical shifts at the two temperatures have also been analyzed (Fig. 8.9). The AGADIR prediction of helical propensities shows that by increasing the temperature to 25 °C from 5 °C the propensities are expected to drop by ∼30% of their propensities at 5 °C. Our secondary chemical shift data show only a slight drop in helical propensities in contrast to that expected by the prediction. This might be explained by transient long range interactions (residual tertiary structure) stabilizing the helical folds.

For the transiently structured helices helical wheel representations have been plotted in Fig. 8.10. This representation shows H4 to differ from the other transient helices in its amino acid composition. The first three helices contain several charged residues, thus they are more likely to be implicated in electrostatically driven interactions. The fourth transient helix is formed almost exclusively by hydrophobic residues. Perhaps it is involved in an interaction with the lipid droplets (where NS5A plays a role in particle assembly).

8.6 ^{15}N relaxation and dynamics

After having analyzed the secondary structural propensities from the chemical shifts, ^{15}N relaxation data was recorded at two temperatures (5 °C and 25 °C) and two magnetic field strengths (700 MHz and 800 MHz), to characterize the dynamics of D2D3 fragment of NS5A. In addition, ^{15}N relaxation data was recorded at 800 MHz at 5 °C also for the D2 and ΔPxxP constructs. The transiently structured regions, as identified by secondary chemical shifts, are expected to show more rigidity in the relaxation measurements.

As mentioned in Section 4.3.3 calculation of a pseudo-τ_c from the T_1/T_2 ratio according to ref. [69] gives a quick way of obtaining information on the local rigidity. Plotting this data and the $\{^1H\}$-^{15}N heteronuclear NOEs (hetNOEs) allows rapid comparison of motions on the ps-ns timescale (hetNOEs) with motions on slower timescales (pseudo-τ_c) We used this approach for comparing relaxation data of our three NS5A constructs. Differences may indicate long-range interactions that are altered between the constructs, due to absence of an interacting region. The data is shown in Fig. 8.11. The first observation that can be made from the relaxation data is that the regions with higher pseudo-τ_c values and higher heteronuclear NOE values correlate well with the regions characterized by residual secondary structure, as identified from the secondary chemical shift analysis. The relaxation data allows identification of two sub-regions (315-322 and 330-335) within the CypAIR that show increased rigidity with respect to the center of the CypAIR region. A further observation is that the data on the highly dynamic part of D3 also displays differences in local rigidity, with elevated hetNOE and τ_c values in the region between residues 412-436, compared to the region between the end of H4 and residue 412. The reason for this increased rigidity, observed from our data, is yet unclear.

The relaxation data for the three constructs at first glance seem to be very similar. However, careful inspection shows that the CypAIR is more flexible in the ΔPxxP construct compared to the D2 and D2D3 constructs. In addition, the PxxP region seems to be more rigid in the D2D3 construct

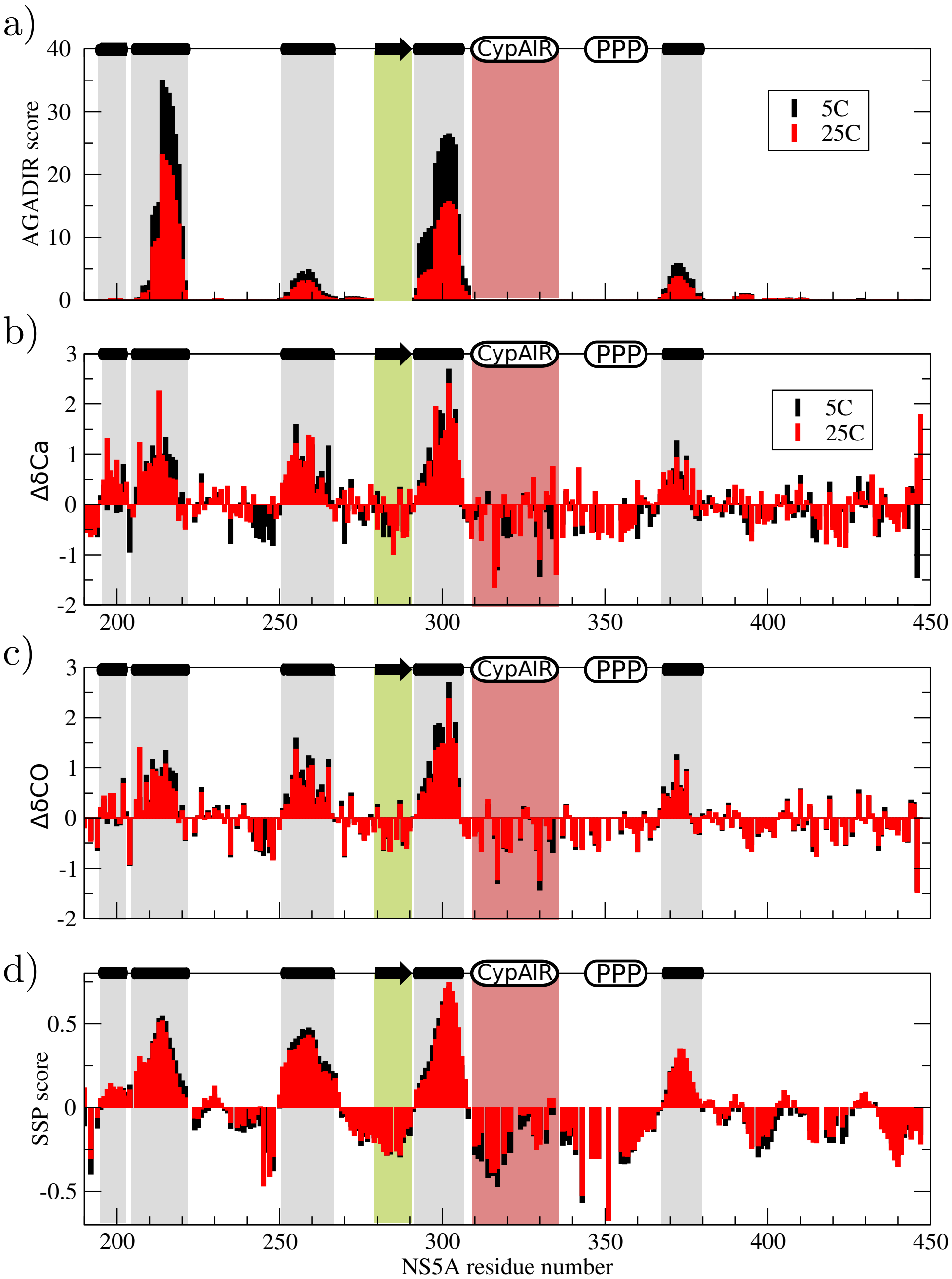


Figure 8.9: Characterization of residual structure in the D2D3 NS5A construct is shown, performed in the same manner as the characterization of D2 in Fig. 8.6, at two temperatures 5 °C (*black*) and 25 °C (*red*). The transiently structured regions identified by the chemical shift analysis (Fig. 8.8) are highlighted. a) AGADIR prediction of helical propensities. b) Secondary C$^{\alpha}$ chemical shifts. c) Secondary CO chemical shifts. d) SSP score calculated from all the available chemical shifts (HN, N, CO, CA, CB)

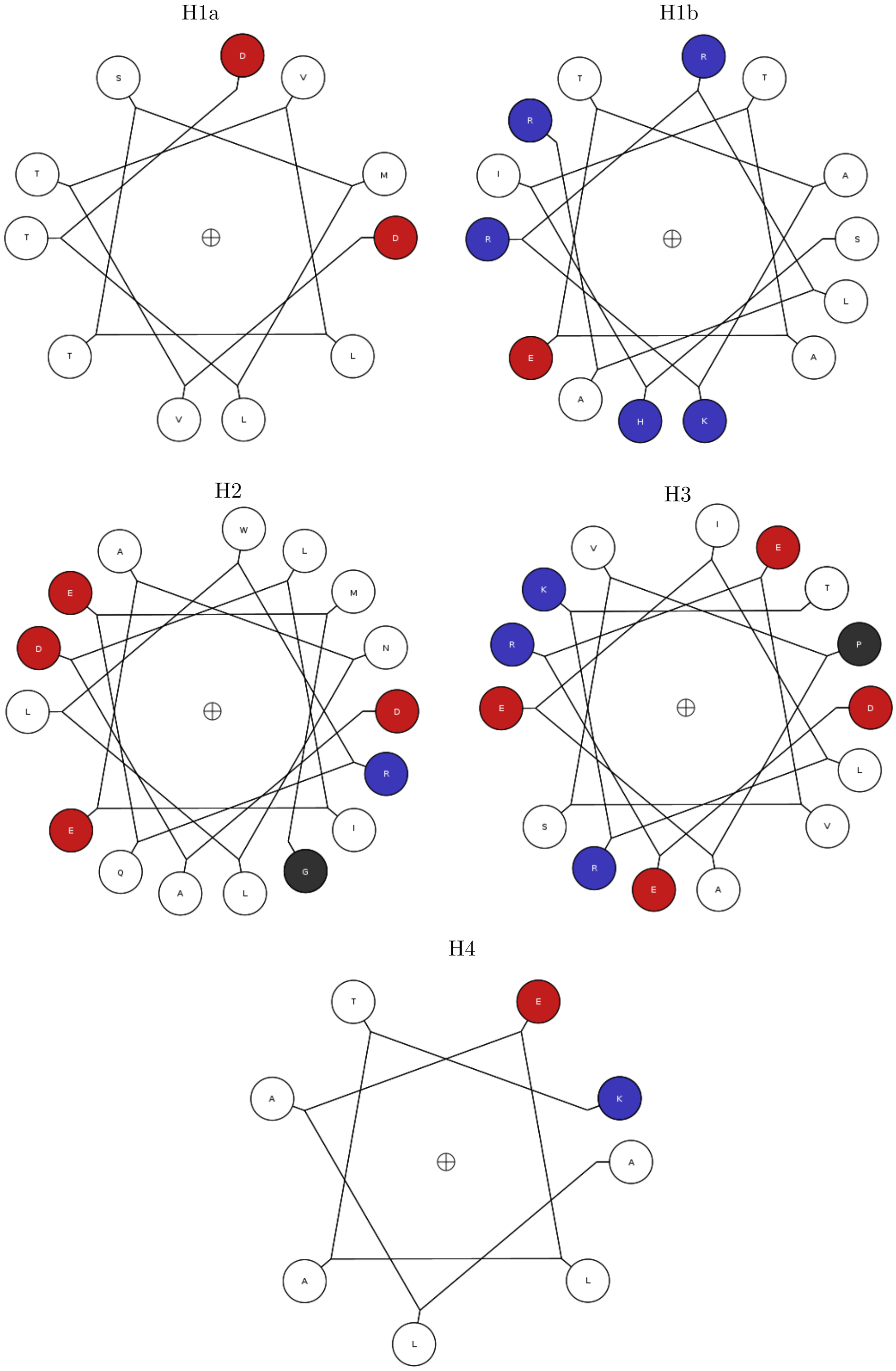


Figure 8.10: Helical wheel representation of the four transiently structured helices. H1a and H1b are plotted separately for better visibility, instead of H1. Positively charged residues are represented by *blue circles*, negatively charged residues by *red circles* and hydrophobic and neutral residues by *white circles*, while glycines and prolines that do not favor helical conformations are represented with *black circles*.

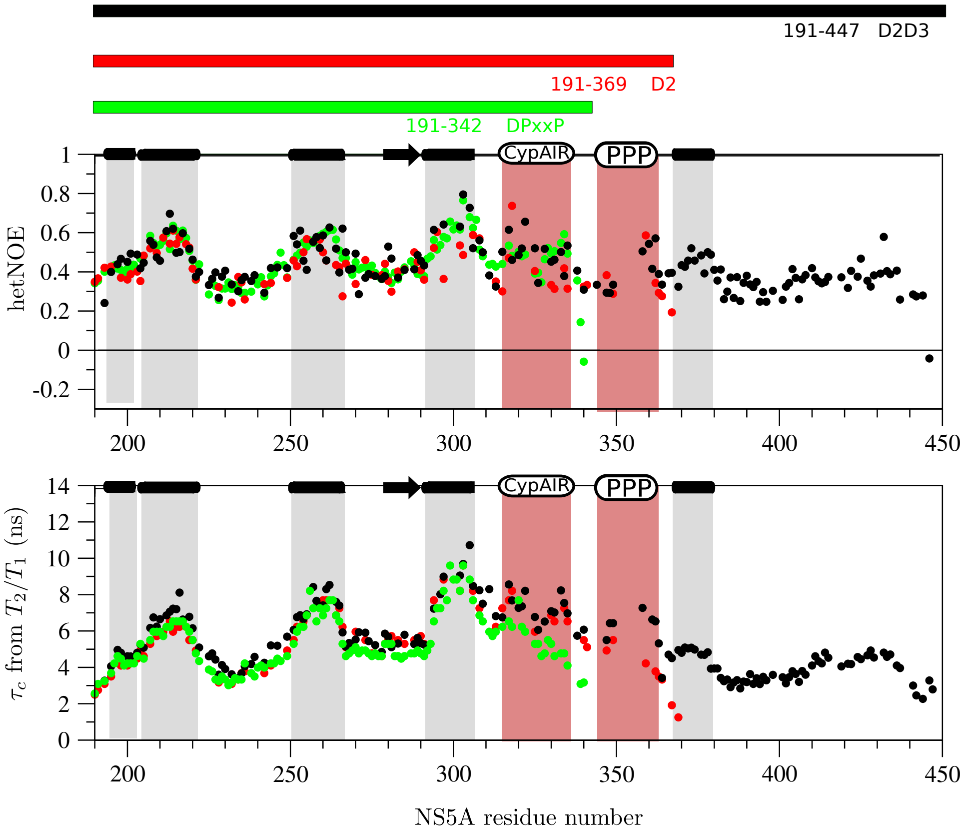


Figure 8.11: ^{15}N relaxation data of the three NS5A constructs, D2D3 (*black*), D2 (*red*) and ΔPxxP (*green*) acquired at 800 MHz at 5 °C. The transiently helical regions are highlighted (*gray*) and the regions where differences can be seen between the constructs are highlighted (*mauve*), namely the PxxP region and the CypA interacting region.

than in the D2 construct. However, in the D2 construct the PxxP region is near the C-terminus. Therefore it cannot be excluded that this difference is a result of the proximity of the C-terminus and not of a structural difference between the D2 and the D2D3 constructs. Furthermore, slight differences can be observed in the rigidity of the three transiently structured helices as well.

Motivated by these data, we dialysed the ΔPxxP and the D2D3 constructs to identical buffers in order to compare chemical shifts. The observed small chemical shift differences between the D2D3 and ΔPxxP constructs are shown in Figure 8.12. These data, similarly to differences in relaxation data, also indicate that long-range interactions influence the structure of these protein fragments. The largest differences are observed at H1, H3 and at the CypAIR. In a future study this experiment should be performed with all the three constructs.

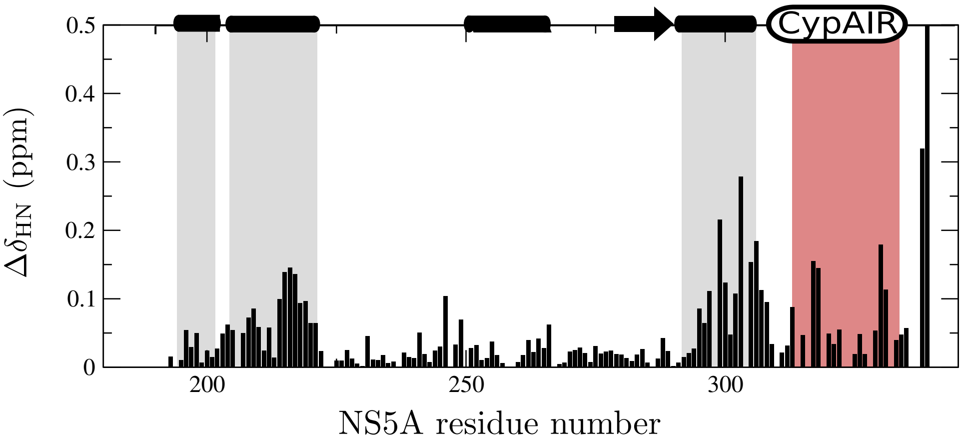


Figure 8.12: ^{1}H-^{15}N summed chemical shift differences between the D2D3 and ΔPxxP constructs calculated by measuring chemical shifts in the ^{1}H and ^{15}N dimensions, summed according to $((10\delta(^1\mathrm{H}))^2 + (\delta(^{15}\mathrm{N}))^2)^{1/2}$. The regions showing the largest differences are highlighted. Transiently structured regions are indicated on the top of the plot by cylinders (α-helical) and arrow (β-strand).

We measured relaxation data at two fields (700 MHz and 800 MHz). The extracted T_1, T_2 and hetNOEs are shown in Fig. 8.13. The field dependence of the relaxation parameters can be observed from the data. Our results show that relaxation properties of NS5A D2D3 are very heterogeneous along the polypeptide chain. In order to get a more detailed insight into the dynamics we decided to extract spectral densities from the data.

As mentioned in Section 4.3.3 a recent study suggests a slightly improved spectral density mapping protocol for IDPs including removal of exchange contributions by measurement of cross-correlated relaxation rates [71]. However, the studied NS5A fragments are characterized by very heterogeneous dynamics resulting in relatively large errors of the relaxation parameters in the more structured parts of the protein due to low signal to noise ratios. Therefore we decided to stay with the original approach of Farrow *et al.* [70]. Calculations of spectral densities from relaxation data were performed with a Python script written for this purpose, by solving equations 4.16-4.18.

Reduced spectral density mapping has been performed from the relaxation data recorded at different fields as shown in Figure 8.14. No global fit was applied, the power spectral densities were mapped individually for each residue. The calculated $J(0)$ values from the two data sets show very good agreement for most of the protein. Overall discrepancy in the calculated $J(0)$ would be a sign for measurement errors. Local differences would indicate conformational exchange phenomena on the μs-ms timescale for the affected residues. The data shows small differences in the calculated J(0) in the H3 region. Motivated by this observation we performed CPMG relaxation dispersion measurements, but we were unable to see any dispersion

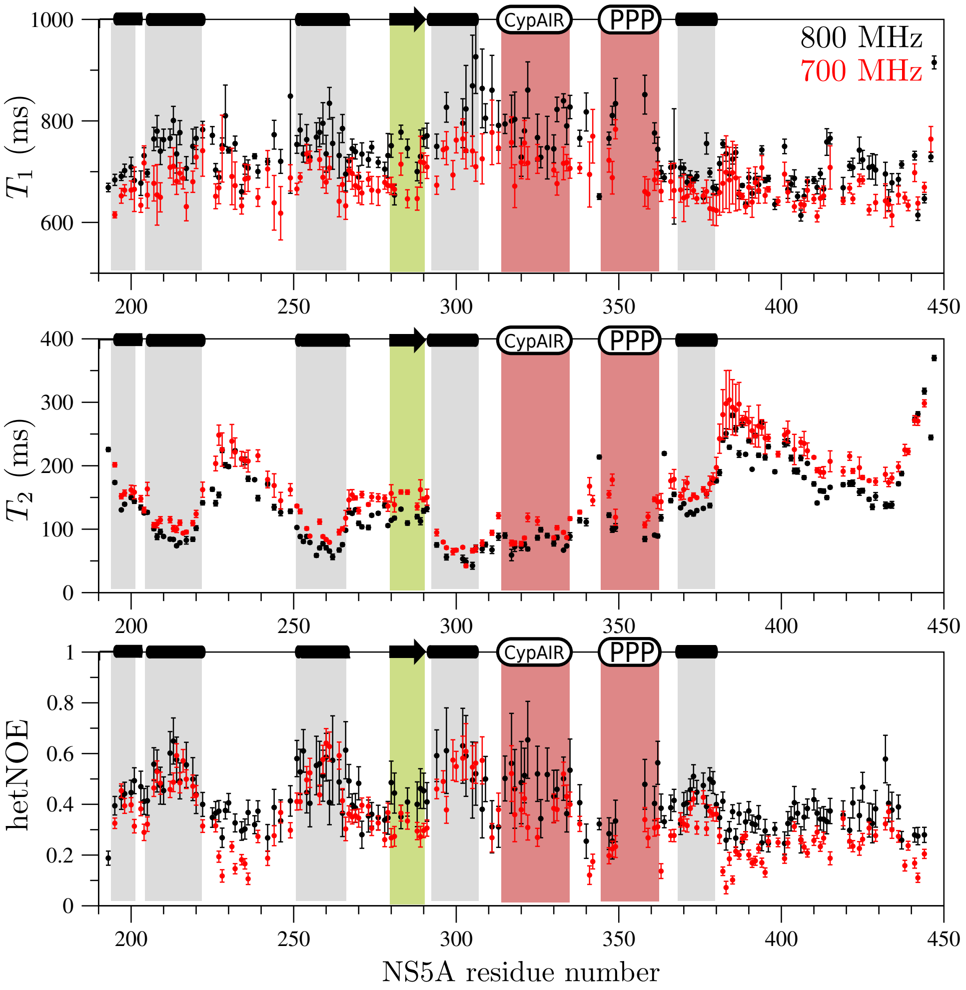


Figure 8.13: ^{15}N relaxation data, T_1, T_2 and ^{1}H-^{15}N NOE (hetNOE) of NS5A D2D3 measured at 700 MHz (*red*) and 800 MHz (*black*) at 5 °C.

when changing the CPMG field in the 100 Hz-1000 Hz range. To probe for slightly faster motions, in the μs – ms timescale, we also recorded $R1\rho$ data at a 3000 Hz spin lock field. From the measured data T_2s were calculated. Differences in T_2 at a high frequency spin lock field would indicate exchange too. We found differences everywhere along the protein, probably resulting from systematic errors of the measurement. Thus it is not yet excluded, that conformational exchange processes contribute to signal broadening. Further experiments will be necessary to decide on this question. The $J(0)$ values, in case of rigid spherical molecule are proportional to τ_C. For IDPs the situation is more complex, but nevertheless we expect a correlation of $J(0)$ and τ_C. As expected, for regions with transient secondary structure the observed $J(0)$ values are higher, as they are tumbling slower, while for flexible

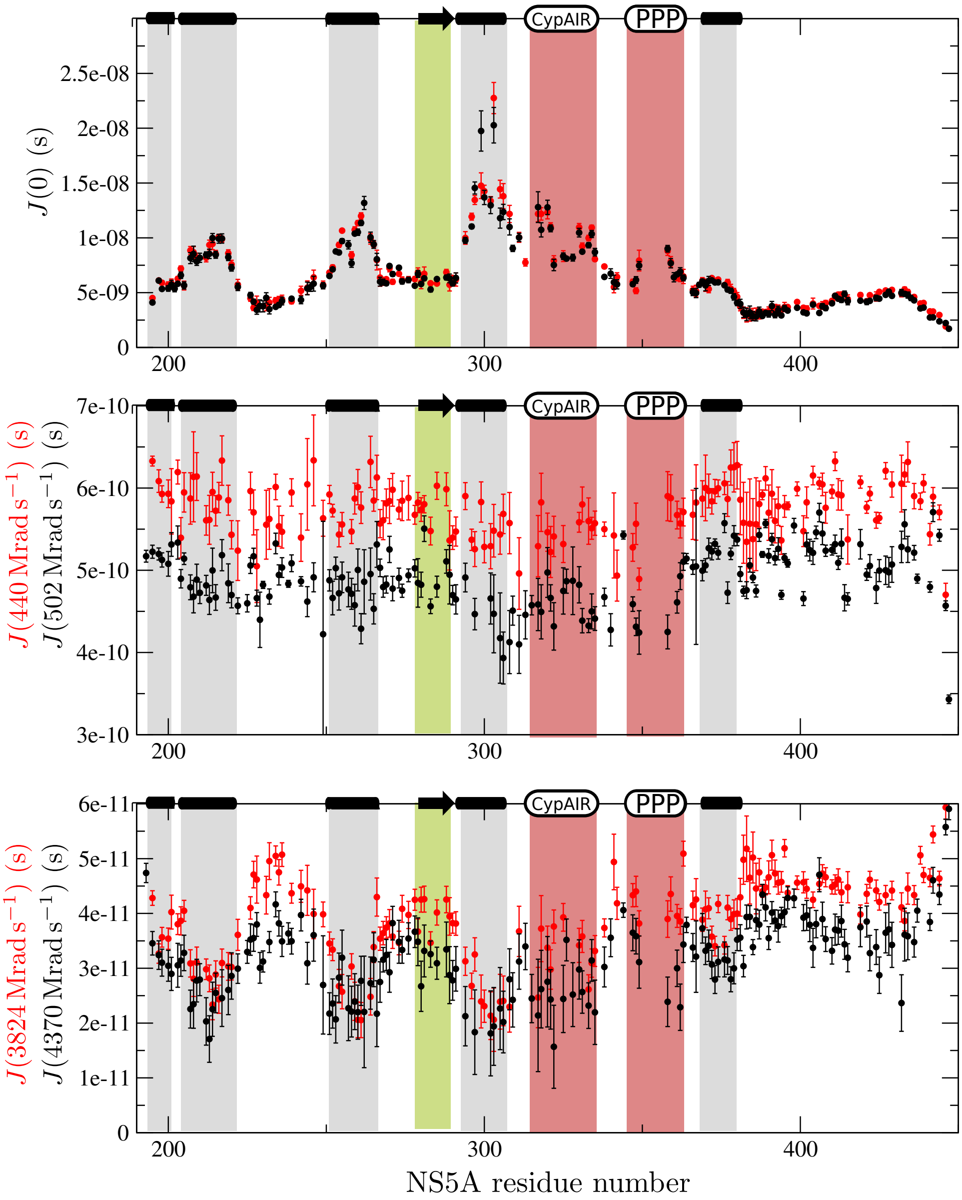


Figure 8.14: Power spectral densities calculated from ^{15}N relaxation data, measured at 700 MHz (*red*) and 800 MHz (*black*) fields at 5 °C. Transiently structured regions are highlighted.

regions they are smaller.

The power spectral densities $J(\omega_N)$ and $J(0.87\omega_H)$ also show heterogeneity along the protein. The power spectral densities obtained at different fields for $J(\omega_N)$ and $J(0.87\omega_H)$, respectively, are plotted together, because their differences yield information about the shape of the function. Power spectral densities decrease more rapidly between 70 and 80 MHz in the transiently structured regions than in flexible regions. $J(0.87\omega_H)$ also shows the heterogeneity in the dynamics of the flexible D3 as it was observed from $J(0)$. $J(\omega_N)$ carries contributions from high frequency motions and low frequency motions as well, therefore it is not possible to tell what types of motions result in the differences that can be observed along the protein. Overall, these data also allow to confirm the presence of the transiently structured regions, identified from secondary chemical shifts. In addition, it highlights the heterogeneous dynamics in D3, although secondary chemical shifts do not show presence of regions with residual secondary structure.

8.7 Effect of temperature on structure and dynamics

Dynamics data and chemical shift data as well as SAXS data discussed in the following Section were measured at two temperatures. Secondary chemical shifts at the two temperatures are plotted in Figure 8.9. As discussed before, the data show that in the studied range there are no major structural changes in the regions of transient secondary structure.

However, dynamics change significantly with temperature. The power spectral densities, extracted from relaxation data measured at two temperatures (5 °C and 25 °C) show interesting features. The plots are shown in Fig. 8.15.

The differences of $J(0)$ at the two temperatures correlate with the extent of transient structure, as it is dependent on the tumbling, that becomes faster with increasing temperature. The largest changes are observed for regions with high extent of secondary structure. Namely, $J(0)$ changes the most in H3, while the most flexible region of the protein, following H4, does not display significant change.

Changes in $J(0)$ show slight differences in dynamics between the transiently structured helices, especially H3 behaves differently from the other transient helices. At 5 °C $J(0)$ is increasing towards the end of the helical region, while at 25 °C there are no large differences within the helical region. In case there are contributions of conformational exchange to T_2 in H3 (as possible based on the observed field dependence of $J(0)$) such changes could also be observed. Therefore it is possible, that this change of $J(0)$ with temperature results from changes of conformational exchange rates with the temperature. Temperature dependence of $J(0)$ near the C-terminus (residues 430-440) indicate that they are more rigid than other regions of D3. Actually, at 5 °C a larger difference can be observed in the tumbling properties of this region and the rest of D3. Chemical shifts did not indicate residual secondary structure for this region. It may be, that it is involved in long range interactions that are stronger at lower temperature resulting in slower tumbling.

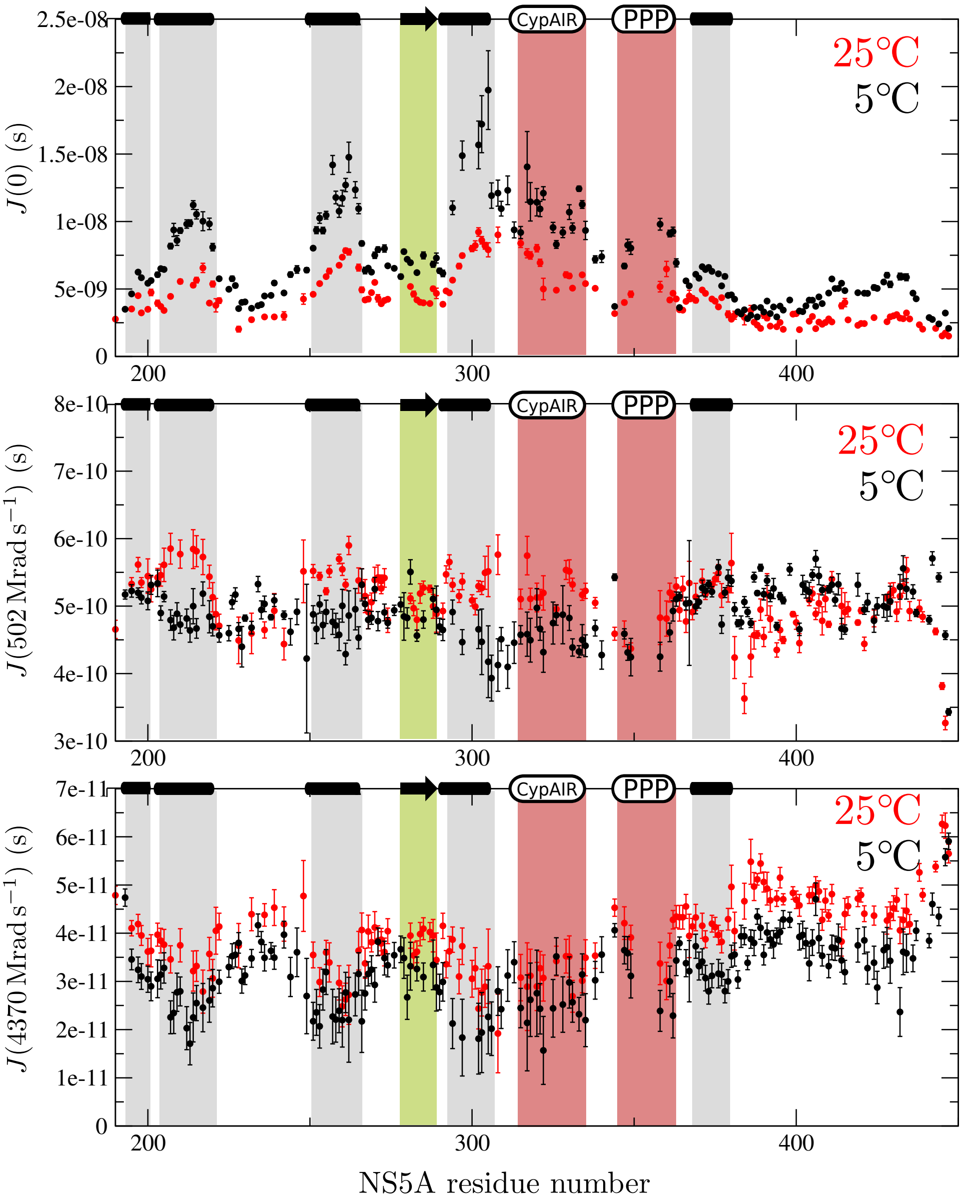


Figure 8.15: Power spectral densities of NS5A D2D3 at 5 °C (*black*) and 25 °C (*red*) from relaxation data (T_1, T_2, HETNOE) measured at 800 MHz. The transiently structured regions are highlighted.

Increasing contribution of fast timescale motions is rather uniform along the sequence. However, noise is in the order of magnitude of the difference as shown by the error bars in Fig. 8.15, therefore it is not possible to extract useful information from this data.

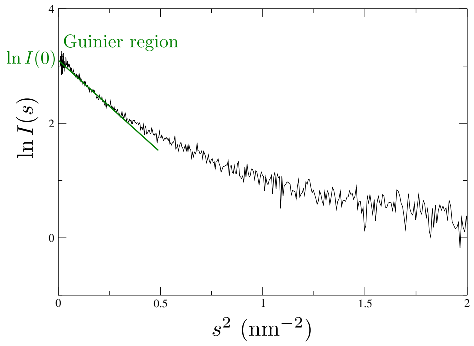


Figure 8.16: Guinier plot of NS5A D2 measured at 20 °C. A linear fit is shown for the Guinier-region.

8.8 SAXS measurements

Information from Small Angle X-ray Scattering (SAXS) experiments nicely complement the information content of the NMR data. NMR measurements yield atomic resolution information, but often it is difficult or even impossible to reconstitute information about the global shape of the molecule without solving the structure (which is only possible in case of globular proteins, not for IDPs). This kind of low resolution information (such as radius of gyration R_G)is available from SAXS measurements.

We decided to collect SAXS data on the studied NS5A protein fragments, D2, D2D3 and ΔPxxP in order to obtain additional information about their overall structural compactness, and to complement the NMR data on the temperature dependence of their structure. SAXS measurements were performed at DESY in Hamburg, at EMBL at the BioSAXS beamline of Doris that has been closed since then. Unfortunately, D2D3 has been found to contain large aggregates that could not be removed by extensive centrifugation, so no analysis was performed on these scattering data.

The radius of gyration, R_G can be easily extracted from SAXS data using the Guinier approximation yielding fast information on the overall compaction of the protein. In the Guinier approximation, $\ln I(s)$ is plotted against s^2. The Guinier region is linear, and thus the scattering at zero angle can be extrapolated, that is approximately proportional to the radius of gyration, as illustrated in Fig. 8.16 according to Eq. 3.7. Guinier analysis of the concentration series (c1, c2, c3, shown in Table 8.1) gives the first criterium for the quality of the data. For good quality data (sample free from aggregates) no concentration dependence of the Guinier-measured R_G should be observed. For NS5A D2 and ΔPxxP this was the case, the differences of R_Gs in the concentration series were within the error of the

	R_G 5 °C (nm)	R_G 20 °C (nm)	R_G 37 °C (nm)
D2			
c_1	3.85 ± 0.06	3.59 ± 0.06	3.52 ± 0.06
c_2	3.94 ± 0.11	3.29 ± 0.09	3.39 ± 0.07
c_3	3.81 ± 0.19	3.45 ± 0.16	too noisy
ΔPxxP			
c_1	3.76 ± 0.06	3.65 ± 0.05	3.35 ± 0.05
c_2	3.80 ± 0.10	3.59 ± 0.10	3.26 ± 0.06
c_3	3.60 ± 0.17	3.60 ± 0.13	too noisy

Table 8.1: Radii of gyration for the NS5A D2 and DPxxP constructs, determined by Guinier analysis of small angle scattering curves. The data obtained for the three sample concentrations (c_1, c_2, c_3) are shown.

measurement. SAXS measurements were conducted at three temperatures, 5 °C, 20 °C and 37 °C. Radii of gyration show that at higher temperature slightly more compact conformations are sampled, this finding is different from the finding of [31], who measured SAXS data at two temperatures and found increase of R_Gs at higher temperature.

The next step in the analysis of scattering data of IDPs consists of plotting the Kratky plot, which is just a different representation of the data, but offers a quick indication of residual structure in the IDP. A bump on the Kratky plot indicates residual structure, while for a Gaussian chain the profile would be perfectly flat. On the Kratky plot of D2 a small bump can be observed, while in the case of ΔPxxP this bump is less visible (Figure 8.17). Comparing this information with differences in ^{15}N relaxation data between the D2 and ΔPxxP constructs (Fig. 8.11) gives a possible explanation for this finding. The presence of the PxxP motif results in a more rigid CypAIR region. This could cause the additional minor compaction of D2 compared to ΔPxxP.

Further analysis of the obtained SAXS data was done with the EOM software. The EOM acronym stands for Ensemble Optimization Method [39]. In the first step of the analysis an ensemble of conformations is created, that was 10000 conformers in our case. These conformers should cover the conformational space available to the IDPs, in other words they are supposed to represent random coil sampling. (Usually $M = 10000$ conformers are created.)

As the next step, N conformers are selected randomly out of this pool of M conformers. This small ensemble is called the chromosome. The scattering curves of the individual conformers are computed by the Crysol software [149] and are summed according to Eq. 3.6 and a fitness parameter is calculated based on the deviation of the calculated scattering curve from

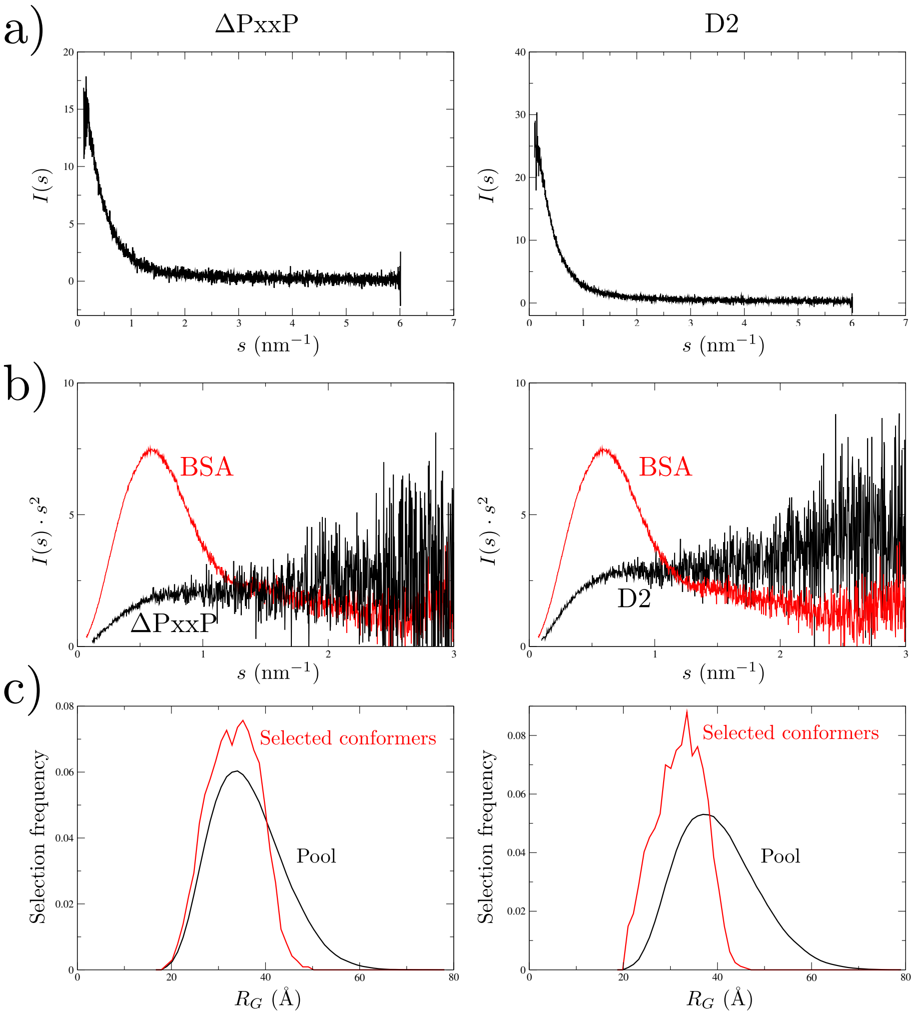


Figure 8.17: Small angle X-ray scattering data of NS5A D2 and ΔPxxP at 20 °C. Panel a) shows the scattering data for the two constructs. The Kratky plots are shown in panel b), for the sake of comparison to a Kratky plot of a globular protein, the Kratky plot of BSA (*red*) used as a standard for checking the instrument is also shown. Distribution of radii of gyration for an ensemble of structures resulting from EOM, whose back-calculated scattering fits the data, are shown (*red*) on panel c) versus the radii of gyration of the random coil pool ensemble (*black*).

experimental data. Then a genetic algorithm optimizes the ensemble that will fit the scattering curve by minimization of the fitness parameter. Then the radii of gyration of the pool and the selected ensemble are plotted in a histogram.

It is important to note that the conformations in the final selected ensemble are not meaningful in the sense that they would represent real conformations present in the ensemble of IDPs. The result of such an ensemble optimization is a histogram of number of selected conformers versus the radius of gyration. The radius of gyration distribution of the pool is plotted and the radius of gyration distribution of the selected conformers as well. This plot indicates whether the conformational ensemble of the sample is more compact or more extended than that of the pool.

The results for NS5A D2 and ΔPxxP are shown in Figure 8.17 c. It is apparent from the data that the conformations NS5A D2 adopts are more compact than the average of the pool, while the average radius of gyration of conformers of ΔPxxP is similar to that of the pool. This finding may be explained by the additional compaction resulting from a long range interaction of the PxxP region with another region of the protein, possibly with CypAIR or D3 as indicated by differences in the relaxation data and by chemical shift differences between the three NS5A constructs.

8.9 RDC measurement attempts

To gain further insight into the structure of NS5A and to be able to collect sufficient amount of data for ensemble modeling, we attempted to measure residual dipolar couplings. The available information from RDCs was detailed in Section 8.9, as well as the principles of the alignment media.

The first attempt to align NS5A D2D3 was with the most common alignment medium used for IDPs, with the PEG-hexanol liquid crystalline phase. First a test was done with buffer only, then with the protein. D_2O splitting of 16 Hz could be observed for the buffer alone, so the experiment was repeated with the protein solution. Again, D_2O splitting could be observed, and a BEST-TROSY spectrum was recorded, it is shown in Figure 8.18 left panel. It is visible from the spectrum, that a part of the protein interacts with the alignment medium and thus peaks from D2 broaden beyond detection.

Next, alignment in a compressed polyacrylamide gel was tested. The experiment was performed using the apparatus of Chou and Bax [150]. Compressed gel was generated, including the protein sample. Unfortunately, the protein interacted with this medium as well, as shown in Figure 8.18 right panel. The intensity distribution in the spectra hints that NS5A D2D3 interacts with the media via its transiently structured D2 domain.

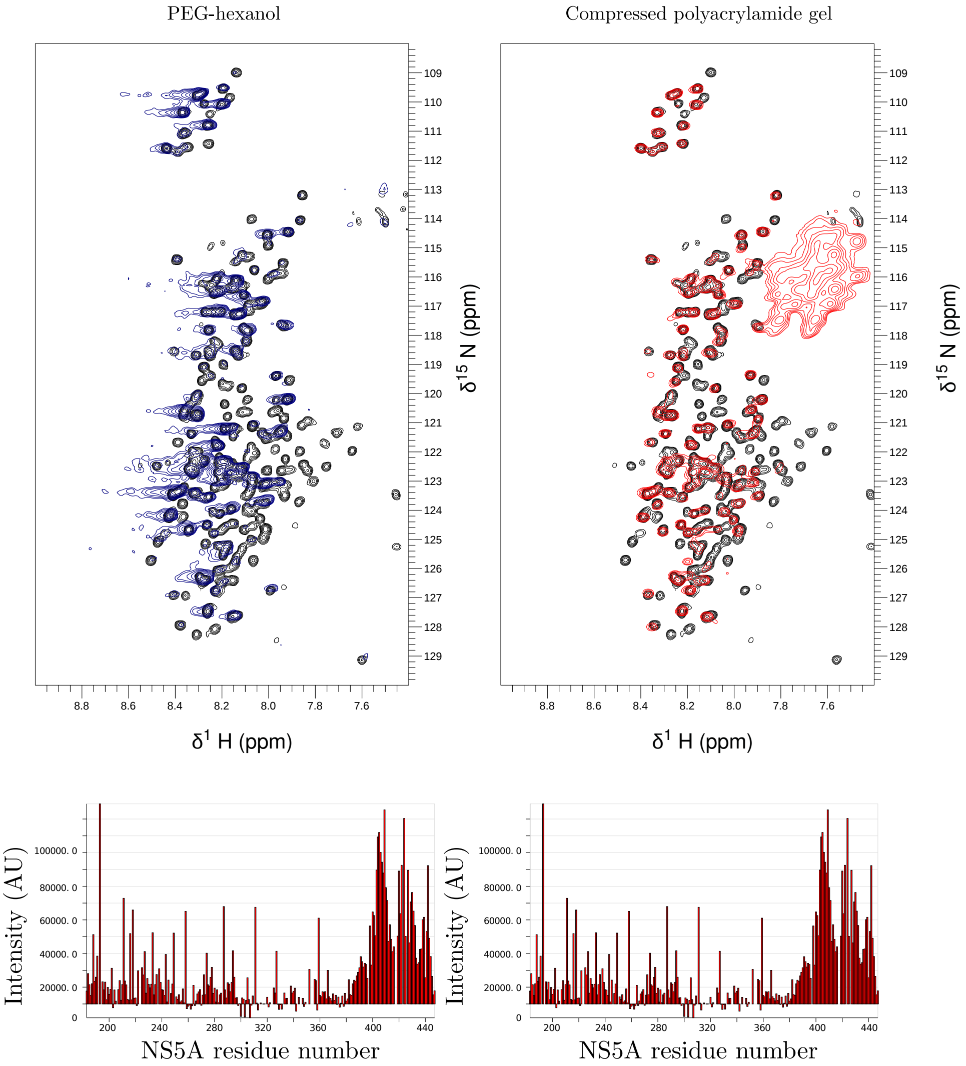


Figure 8.18: BEST-TROSY spectra of NS5A D2D3 in the tested alignment media, PEG-hexanol liquid crystalline phase on the left panel and radially compressed polyacrylamide gel on the right panel. The reference spectra are shown in *black*, the spectrum recorded of NS5A aligned in PEG-hexanol in *blue* and that of NS5A aligned in compressed polyacrylamide gel in *red*. Below the spectra peak intensities are shown as function of the sequence, showing that most of the residues in D2 are broadened, while correlation peaks of D3 residues remain intense.

At this point we stopped our attempts to align NS5A D2D3, as there are no more steric alignment media available, and electrostatic alignment media are not suitable for IDPs. They may perturb the structural ensemble significantly, as long-range interactions of IDPs are often mediated by exposed positively and negatively charged stretches.

8.10 Conclusions and perspectives

Our results on the characterization of structure and dynamics in different NS5A constructs were presented. The D2D3 construct of NS5A, although being intrinsically disordered, contains large amount of transiently structured regions. In addition, our data indicates the presence of long range interactions that cause compaction of the structural ensemble. Furthermore, significant differences in dynamics between D2 and D3 are observed.

Transient secondary structure of the D2D3 construct of NS5A was characterized based on secondary chemical shifts. In our previous study on the D2 construct, a different database was used for obtaining random coil chemical shifts. Analysis of the chemical shifts of NS5A with a temperature- and pH corrected database and completion of the chemical shift data set with C^{β} and H^{α} allowed identification of further residual structural regions compared to our previous analysis of the D2 construct. A short transiently helical region H1a was identified preceding the larger helix H1b. In addition, a short region was identified with small β-strand propensity between H2 and H3. This might give an indication on why the linker connecting H2 and H3 is more rigid than the one connecting H1 and H2 as seen by relaxation data. In the structural ensemble there might be some compact conformations present where H2, S1 and H3 all stabilize each other. Furthermore, as our D2D3 construct comprises domain 3, a small hydrophobic, low populated (20%) helical region was identified in D3.

The presence of transient long range interactions were first indicated by PRE data of D2 that was published at the beginning of this thesis [112]. Comparison of theoretically predicted versus experimentally derived helical propensities indicated that some additional interactions stabilize the helical structures. Furthermore, differences in relaxation data and chemical shifts between our three constructs also showed slight localized differences that indicate the presence of transient long range interactions. One such long range interaction is indicated by the difference between the D2 and the D2D3 constructs, namely that the PxxP region was found to be more rigid in D2D3 than in D2. We may speculate that the region in D3 between residues 430-435 that shows increased rigidity is the one involved in this long range interaction. The other long range interaction is indicated by that in ΔPxxP the CypAIR seems to be more flexible than in D2 and D3. The only difference between D2 and ΔPxxP being the PxxP motif, it is logical to assume

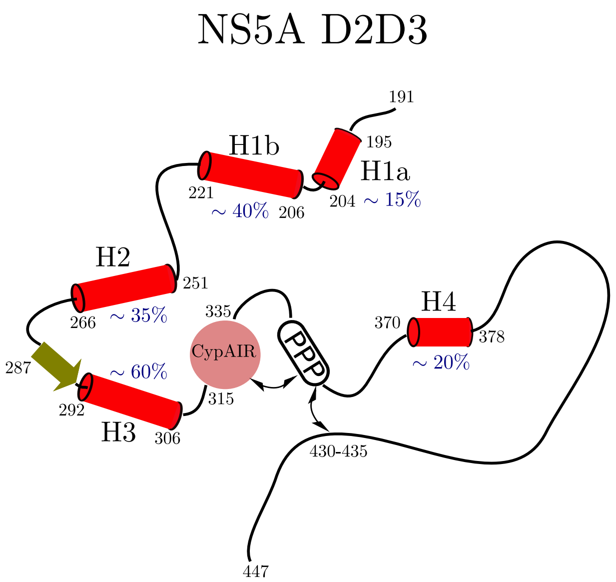


Figure 8.19: A model of the residual secondary structure and transient long-range interactions identified in NS5A D2D3. Red cylinders indicate transiently structured helical regions, their propensities are shown next to them in *dark blue*. A probable, transiently populated β-strand in shown in *green* with unknown, low propensity. CypAIR and PPP indicate regions with increased rigidity as observed from relaxation data with unknown secondary structure propensities. Transient long range interactions are indicated by *double-headed arrows*.

that this is caused by an interaction between the PxxP motif and CypAIR. Chemical shift differences between the D2D3 and ΔPxxP constructs are also consistent with this model.

We characterized extensively the temperature dependence of structure and dynamics. Slight changes in secondary chemical shifts could be identified. However, much larger changes would be expected based on the physicochemical properties as predicted by AGADIR. This difference also indicates presence of long range interactions that stabilize the secondary structure. Such long-range interactions in IDPs are often mediated by electrostatics. We showed by helical wheel representations, that H1-H3 helical regions contain large amounts of charged residues, therefore such interactions could easily be present among them.

A model is emerging from our data on the transient long range interactions present at some time points in the structural ensemble of NS5A D2D3 as shown in Fig. 8.19. It is very important to point out that this model should not be regarded as a static structure, spatial proximities are meant to indicate transient contacts that do not necessarily occur at the same time. NMR observations are ensemble- and time-averaged, therefore it is not possible to resolve the time dependence of these interactions.

What may be the role of these transient long range interactions? In the context of an interaction with other viral proteins or host factors, some of these structures might get stabilized. Furthermore, it is also possible, that NS5A binds several partners simultaneously and these dynamic interactions help in directing the binding partners towards each other or towards other binding sites on NS5A.

We may even go into further speculations about the transient structure in the D2D3 region of NS5A. As discussed by Tellinghuisen *et al.* D1, though being a folded domain, contains large amounts of long loop regions [132]. They point out an interesting feature in the crystal structure of domain 1, that is a disulfide bridge formed by two conserved cysteins, C142 and C190. This disulfide bond connects the disordered region at the C-terminus of the domain with a large loop. This bond may be essential for ensuring the correct fold of D1. This data on the structure of D1 rises the possibility of some further structuring that would require correctly formed disulfide bonds. Our construct starts with E191, right after the conserved disulfide-forming C190. Cystein residues in D2, C243 and C342 are conserved between HCV genotypes as shown by the multiple sequence alignment of the region (Fig. 10 in Appendices). As NS5A is located at ER derived membrane alterations it is possible that oxidation of these cysteins occurs in a regulated manner resulting in disulfide bonds. This could result in spatial proximity of the transiently structured regions. No studies have been published so far about the oxidation state of these cysteins. A possible perspective of the project is a study of structural differences between oxidized and reduced states of the protein.

A further conclusion from our data is that D3 is more flexible than D2, but different degrees in flexibility can also be observed within D3 as indicated by our data. In particular, we showed that temperature dependence of dynamics data can be a sensitive indicator of such slight differences in flexibility, that may not be identified at a single temperature alone.

Furthermore, the temperature dependence of the relaxation data raises the question: what is the ideal temperature that should be used for characterization of such proteins? There might be some conditions where certain structural features are amplified while in other conditions they are not identifiable. As the exact local physiological conditions for proteins are often not known, probably the best is to test the dependence of residual structure on sample conditions. Persistence of residual structure under different conditions would indicate sufficiently, that the structuring may be relevant *in vivo*. Sequence conservation of the transiently structured regions indicates their possible relevance.

Another perspective is modeling the structural ensemble of this construct. In the course of this thesis we attempted modeling this construct by replica averaged chemical shift restrained molecular dynamics simulations, but its large size made the calculations too expensive. As higher computational resources become available, the simulation may be performed.

SAXS data could also be included in such simulation as a restraint. As NS5A D2D3 shows oligomerization tendencies this would require a SEC purification directly preceding the measurement.

The structural ensemble could be modeled by Monte Carlo based ensemble calculation methods such as ENSEMBLE or Flexible Meccano. However, this would require further data for selection, such as RDCs. However, in spite of all our efforts, NS5A D2D3 could not be aligned in steric alignment media. PRE data would also suffer from the oligomerization tendencies of the construct, probably intra- and intermolecular PREs would be observed simultaneously. Therefore, MD-based modeling is likely to be better suited for modeling such a system.

Chapter 9

Interaction of NS5A with Bin1 SH3 domain

9.1 Introduction: Interaction of NS5A with various SH3 domains

Some viral proteins, such as NS5A, interact with a multitude of host proteins and interfere with signaling processes of the host cell. NS5A has been reported to interact with various SH3 domains. SH3 domains are small protein domains of β-barrel fold in modular proteins that mediate protein-protein interactions. They recognize the PxxP motif and bind it with various affinities. The determinants of specificity and affinity of SH3 domain interactions have been subject to research, some examples can be found in references [151], [152], [153], [154]. For a recent review on interaction of SH3 domains and viral proteins see reference [155].

Viral proteins have been recognized to frequently use linear motifs for their interactions. For a database of eukaryotic linear motifs see the ELM server (elm.eu.org [19]). The ELM server can be used to identify linear motifs in amino acid sequences. NS5A protein of HCV contains two class-I and one class-II PxxP motifs in the polyproline region. NS5A protein has been reported to interact with a large number of SH3 domains via these polyproline motifs located in the PxxP region. Examples are Grb2 [156], Fyn, Lyn, Hck, Lck [157], [158] as seen by co-immunoprecipitation and in-vitro assays.

A few studies have addressed the biological significance of binding of NS5A to Bin1 SH3. This interaction was first studied by Nanda et al [159]. The interaction was identified *in vivo* by co-immunoprecipitation and by *in vitro* binding assays as well. They found that NS5A might be implicated in deregulating Bin1-mediated apoptosis and thus the interaction contributes to pathogenesis of HCV. An other study identified that phosporylation of NS5A is reduced *in vivo* in cells cotransfected with NS5A and Bin1 but not in cells co-transfected with NS5A and a Bin1 deletion mutant lacking the SH3 domain [160].

Interaction of Bin1 SH3 domain with the disordered Myc oncoprotein has been subject to a recent NMR study during the course of this thesis [161]. The authors found that Myc interacts with Bin1 SH3 in a multivalent fashion and interaction with Bin1 perturbs the long range interactions present in the conformational ensemble of Myc.

More recently, the results from our collaborators in Forschungszentrum Juelich have been published, that motivated our detailed study. They found that an NS5A peptide comprising the PxxP motif is able to displace Myc oncoprotein from binding the tumor suppressor Bin1 *in vitro*. NS5A has much higher K_D than Myc for binding Bin1 SH3.

9.2 Titration of NS5A D2D3 with Bin 1 SH3 domain

For studies of interaction of NS5A with SH3 domains, NMR titration was our major tool. Before the start of this PhD thesis, several titrations of various SH3 domains to NS5A D2 were performed by Sophie Feuerstein to characterize their interactions.

The principle of an NMR titration is to observe changes in the spectrum of the protein upon gradual addition of the ligand at predetermined ratios. Chemical shifts and/or intensities can change when the protein is binding the ligand. The changes of chemical shifts are termed chemical shift perturbations. Provided the assignment of the fingerprint spectrum is available, these changes can be mapped onto the sequence, or provided a structure is available, on the structure. This allows identification of regions of the protein affected by binding. Chemical shift perturbations larger that the noise are observed for a residue either if its chemical environment changes because of proximity of the ligand, or if the structure changes due to binding of the ligand at a remote site (allosteric effect). If the exchange rate between bound and free form is significantly larger than the difference of their resonance frequencies ($k_{ex} >> \Delta\omega$), a single peak is detected at the frequency corresponding to the concentration-weighted average of the two forms. This exchange regime is called fast exchange and mapping of chemical shift perturbations allows for calculation of dissociation constants (K_Ds) from the data, as detailed in Section 9.4. If the exchange rate is significantly smaller than the difference of their resonance frequencies ($k_{ex} << \Delta\omega$), two peaks are detected corresponding to the resonance frequencies of the free and bound forms with intensities corresponding to their concentrations. In the scenario when the exchange rate is near the frequency difference ($k_{ex} \sim \Delta\omega$) the peak is broadening. This is the intermediate exchange regime. Our titra-

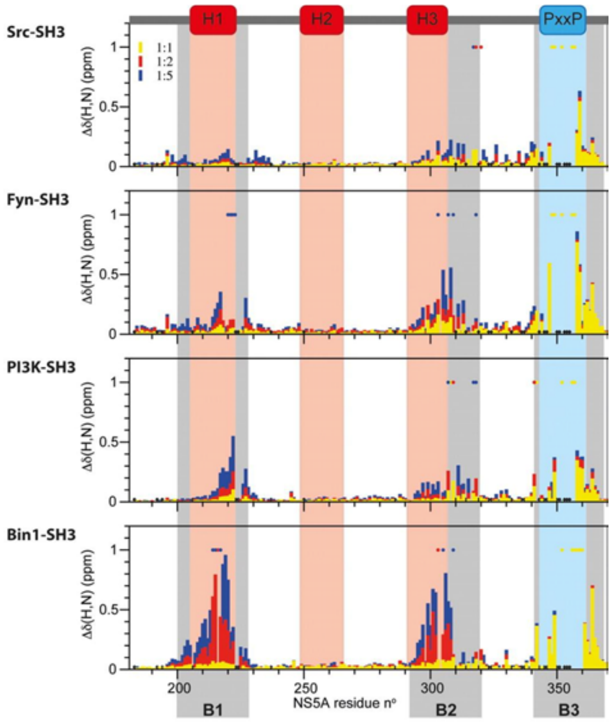


Figure 9.1: Titrations of NS5A D2 with SH3 domains. Chemical shift perturbations are shown at different NS5A:Bin1 SH3 ratios as shown on the label in the figure (1:1 *yellow*, 1:2 *red*, 1:5 *blue*). The secondary structural regions of NS5A are highlighted. The binding regions are denoted by B1, B2 and B3.

tions provided examples for both slow and fast exchange phenomena, as can be appreciated from Fig. 9.2.

The chemical shift perturbations observed during the aforementioned titrations of NS5A D2 with several SH3 domains are shown in Fig. 9.1. These data were published in Refs. [112] and [141]. The chemical shift perturbations showed that up to 1:1 ratio of SH3 domain:NS5A, chemical shift changes are mostly localized to the PxxP motifs (B1 region). The data recorded at different sub-stoichiometric ratios showed that the exchange regime was slow exchange, as two separate peaks, corresponding to the bound and free form, could be observed simultaneously. When the ratio was changed the intensities changed, but peak positions did not. However, minor perturbations could be observed near H1 and H3 as well, in the regions denoted binding regions B1 and B2 at small excess of the SH3 domain. When the titration was continued with larger excess of the SH3 domain, these chemical shift perturbations at B1 and B2 increased gradually, showing that the exchange is in the fast exchange regime. These data can be explained by a high affinity binding site at the PxxP motif that gets saturated first. At an excess of SH3 domain the lower affinity sites compete for binding. Among the studied SH3 domains the largest chemical shift changes were observed for Bin1 SH3 domain. Therefore, in the course of this thesis we focused on the Bin1 SH3 domain–NS5A interaction. We aimed at a more detailed investigation of the binding mechanisms both of the high affinity and of the low affinity binding. In order to investigate the low affinity binding in more detail without the disturbing presence of the PxxP region, we expressed and purified the ΔPxxP construct of NS5A (Fig. 8.1). We decided to study the high affinity binding on the D2D3 construct as well.

As the longest intrinsically disordered construct, D2D3 was purified, we performed a titration with Bin1 SH3 at 5 °C, as earlier titrations were also performed at this temperature. The chemical shift perturbations are shown in Fig. 9.2, panel a. The chemical shift perturbations upon the titration resemble the titration with NS5A D2. At NS5A D2D3:Bin1 SH3 ratios smaller than 1:1 the largest chemical shift changes were observed for residues surrounding the PxxP motifs. The exchange regime was found to be slow exchange, as seen from the D2 titration, and as expected from the nanomolar K_D of the interaction of an NS5A PxxP peptide with Bin1 SH3 (as measured by SPR, Amine Aladag) [112]. However, an interesting observation was made from the data. Small chemical shift perturbations could be detected for nearly all residues in D3 of NS5A with slow exchange. To identify the reason for these small chemical shift perturbations, further experiments were performed. Details are described in Section 9.3. The residues located in the transient helices H1 and H3 behave similarly as in the case of the D2 construct. To characterize in further detail the low affinity interactions we studied the construct of D2 lacking the PxxP motif (ΔPxxP).

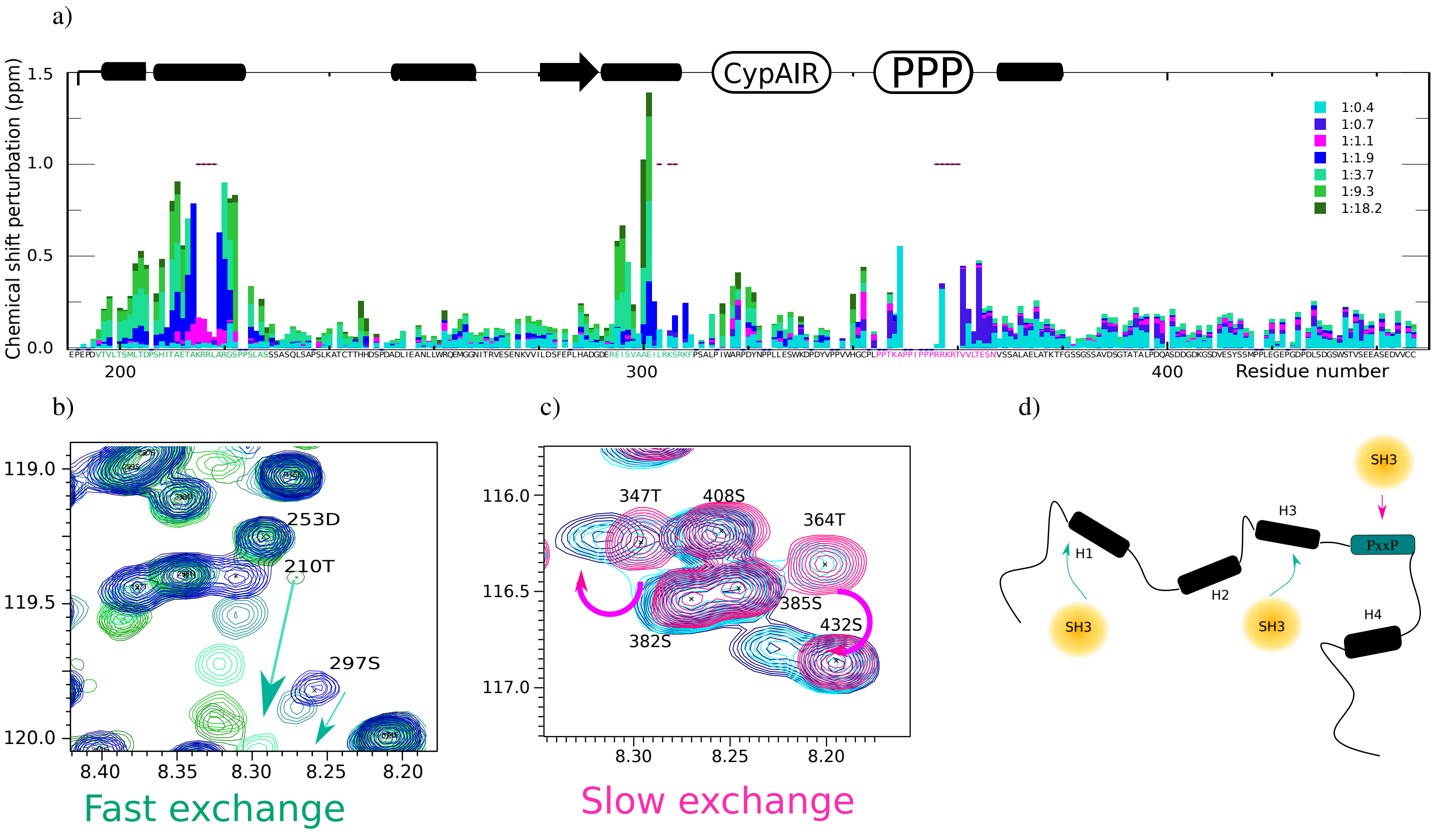


Figure 9.2: a) Chemical shift perturbations at different ratios (color-coding as shown on label) of NS5A D2D3:Bin1 SH3 as measured in BEST-TROSY spectra, calculated by measuring chemical shifts in the ^{1}H and ^{15}N dimensions, summed according to $((10\delta(^1\mathrm{H}))^2 + (\delta(^{15}\mathrm{N}))^2)^{1/2}$. b) Excerpt of the ^{1}H-^{15}N BEST-TROSY spectra of the titration series from ratios 1:1.9 to 1:18.2 in colors as shown on the label in a) and the reference specrum (*black*) are shown to illustrate low affinity binding in the fast exchange regime. c) Excerpt showing chemical shift perturbations corresponding to the high affinity binding in the slow exchange regime, BEST-TROSY spectra are shown from ratios 1:0 to 1:1.1 in colors as shown on the label in panel a).

	Bin1 SH3 n (nmol)	c (μM)	NS5A D2 n (nmol)	c (μM)	V_{total} (μl)
ref	12	100	-	-	120
5:1	12	89.5	2.4	17.9	134.1
2:1	12	77.3	6	38.6	155.3
1:1	12	63.0	12	63.0	190.6

Table 9.1: Titration of [^{15}N]-Bin1 SH3 with NS5A D2

9.3 Interaction with high affinity

As mentioned before, the chemical shift changes upon titration occurred in the vicinity of the polyproline region and small chemical shift perturbations, resulting from the high affinity binding were observed throughout D3. All these chemical sift changes indicated slow exchange regime. However, resonances in the center of the binding site showed broadening. This could be a consequence either of i) increased tumbling of the complex, ii) dynamics within the binding region in the bound form, iii) intermediate exchange. The third possibility can be excluded, as residues near the binding site indicate that the overall exchange rate of the complex formation is slow. To characterize this interaction in detail and to explain the previously mentioned observations, we performed the experiments presented in this section.

9.3.1 Titration of NS5A D2 to [^{15}N]-Bin1 SH3

To map the binding site of NS5A on the structure of Bin1 SH3 domain a titration was performed by titrating the NS5A D2 construct to ^{15}N-labeled Bin1 SH3. The titration was performed according to Table 9.1. The BEST-TROSY spectra recorded at each point of the titration are shown in Fig. 9.3. The spectra were recorded with identical parameter settings and are plotted at the same contour levels at 5 °C. The data shows, that all residues of Bin1 SH3 domain are broadening as NS5A D2D3 is titrated to it, except the three flexible C-terminal residues. At a 1:1 Bin1 SH3:NS5A D2 ratio the resonances of Bin1 SH3 are broadened beyond detection. But at lower ratios peak shifts can be observed. This can be explained as follows. As there are both high affinity and low affinity binding sites present in NS5A, as long as there is excess of Bin1 SH3, an equimolar amount of Bin1 SH3 is involved in the high affinity interaction with NS5A. The excess of Bin1 participates in the low affinity interaction, resulting in changing shifts at different ratios, as the low affinity interaction is in the fast exchange regime. Therefore, on the spectra we can observe the superposition of broadening from the high affinity interaction and shifting peaks resulting from the low affinity binding. At 1:1 ratio we observed broadening of most of the Bin1 SH3 resonances beyond

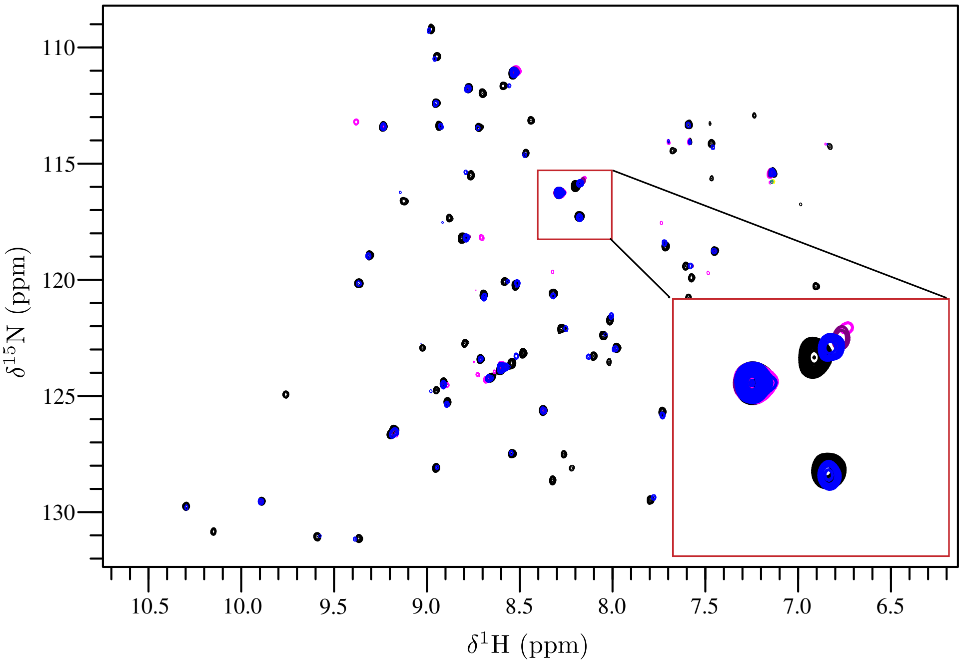


Figure 9.3: ^{1}H-^{15}N BEST-TROSY spectra recorded at 800 MHz at 5 °C field strength of the titration of [^{15}N]-Bin1 SH3 domain with NS5A D2 at different ratios of Bin1 SH3 to NS5A as follows. Bin1 SH3:NS5A 1:0 (*black*), 5:1 (*blue*), 2:1 (*purple*), 1:1 (*magenta*).

detection. This is why we decided to increase the temperature in order to see whether the dynamics resulting in broadening is temperature dependent. We found that, though with low intensities, resonances belonging to the bound form of Bin1 SH3 appeared in the spectrum recorded at 25 °C.

Therefore we repeated the titration of Bin1 SH3 to NS5A D2D3 at 25 °C, this time focusing on the high affinity interaction. Both proteins were ^{15}N-labeled. The last step of the titration was at 1:1 ratio of NS5A D2D3 and Bin1 SH3. At this temperature (25 °C) is was possible to detect and assign Bin1 SH3 in the complex, by superposition with the assigned spectra of Bin1 SH3 in complex with an NS5A peptide corresponding to the high affinity binding site, recorded by Amine Aladag at Forschungszentrum Juelich [112]. Chemical shift perturbations could be mapped onto the surface of Bin1 SH3 as illustrated in Fig. 9.4. The data shows that chemical shifts are perturbed at the canonical PxxP binding site of the domain, as expected. As the resulting peak intensities in the complex were very low, we decided not to go on attempting structure determination of the complex. The reason for the line broadening in the complex is still unclear. The fact that with increasing temperature the resonances became observable could result from faster tumbling with higher temperature as well as from changing exchange regime. As mentioned before, two class-I and one class-II PxxP motifs can be found in

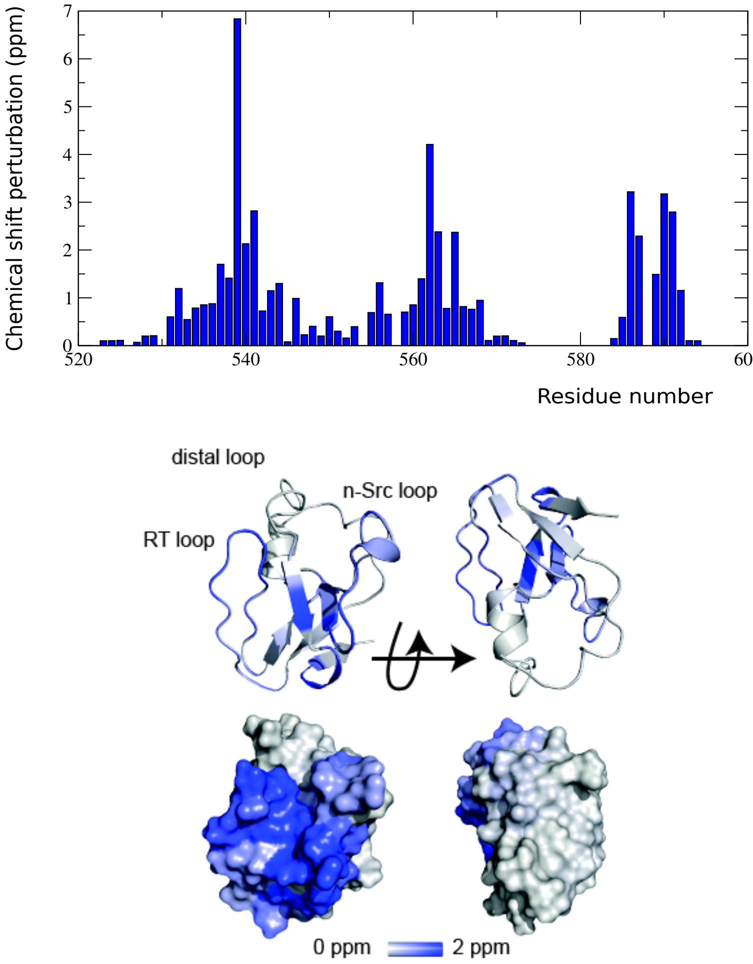


Figure 9.4: Chemical shift perturbations of Bin1 SH3 domain when titrated with NS5A D2D3 (at 1:1 ratio) (*top*) summed as $((10\delta(^{1}\mathrm{H}))^2 + (\delta(^{15}\mathrm{N}))^2)^{1/2}$ (*top*). The chemical shift perturbations were mapped onto the structure of Bin1 SH3, with the blue color deepening with the magnitude of the chemical shift perturbation observed for the amino acid (*bottom*).

the PxxP region (Fig. 8.1). It is possible, that there is dynamics in the bound form in the intermediate exchange regime, causing line broadening, because of exchange between different possible binding modes.

9.3.2 Characterization of the high affinity complex of NS5A D2D3 and Bin1 SH3

To characterize in more detail the effect of Bin1 SH3 binding on the structure of NS5A D2D3, we conducted further studies. We assumed that the low affinity binding is negligible as long as the high affinity binding site is not saturated. This assumption is justified by the fact, that chemical shifts of the low affinity binding sites are only perturbed significantly at an excess of Bin1 SH3 (higher than 1:1 Bin1 SH3:NS5A ratio).

As mentioned before, we observed small chemical shift perturbations for nearly all residues along domain 3 at sub-stoichiometric Bin1 SH3:NS5A ratios. Peaks of the bound state and the free state are present at the same time at sub-stoichiometric Bin1 SH3:NS5A ratios, indicating that the exchange regime is slow and that these changes are a direct consequence of the high affinity binding. In addition, some residues in D2 also showed small chemical shift changes that most likely also result from the binding of Bin1 SH3 to the PxxP region. Examples of these chemical shift changes for several residues from D3 and a few from D2 are shown in Fig. 9.5 a. Chemical shift perturbations measured at Bin1 SH3:NS5A 0.7:1 ratio show, that the entire D3 is perturbed by the binding of the PxxP region to Bin1 SH3 domain (Fig. 9.5 b), in addition to a few perturbed regions in D2. The largest effects in D2 are observed for residues 262-266, 301-306, 317-318 and 329-334.

We measured C^α chemical shifts on the 1:1 complex, in order to see if there are changes in secondary structural propensities for the regions, which were identified from chemical shift changes to be affected by the binding. We calculated secondary C^α chemical shifts from these data and compared it to those of free NS5A (Fig. 9.5). C^α chemical shift changes are very small, showing that there are *no significant changes in transient secondary structure* resulting from the high affinity binding event, associated with the chemical shift perturbations in the slow exchange regime observed in remote regions.

^{15}N relaxation data was also recorded to characterize the impact of the binding of one equivalent of Bin1 SH3 domain on the dynamics of NS5A D2D3. The measured relaxation time constants are plotted in Fig. 9.6 The figure shows, that T_2 relaxation time constants are slightly smaller in the complex along than in the free state for D3 residues. Heteronuclear NOEs also indicate that D3 in the bound state is more flexible than in the free state. In order to analyze in more detail the relaxation data, spectral densities were mapped. The data are shown in Figure 9.7.

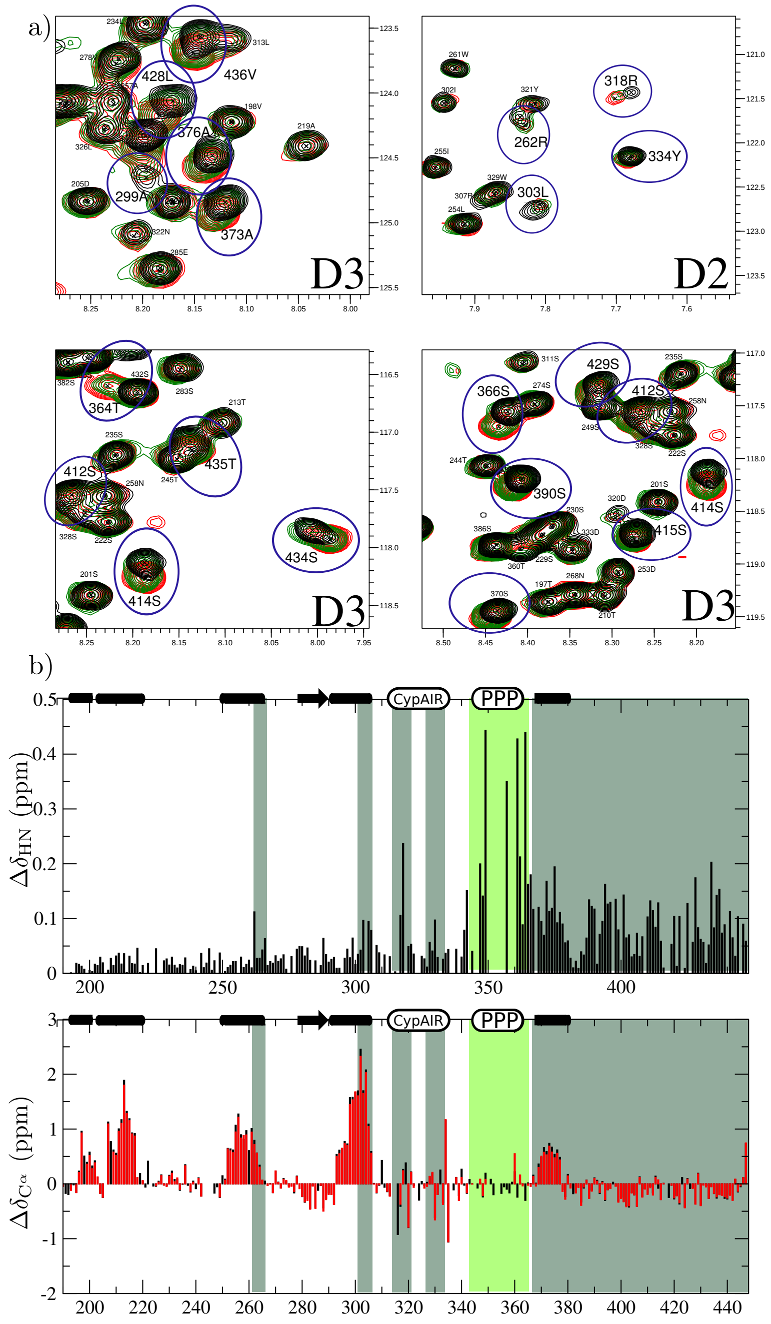


Figure 9.5: a) Selected regions of the BEST-TROSY spectra upon titration of NS5A D2D3 with Bin1 SH3. The spectra were recorded at 800 MHz field at 5 °C at NS5A:Bin1 SH3 ratios 1:0 (*black*), 1:0.4 (*green*) and 1:0.7 (*red*). The circle highlights shifting residues showing slow exchange. b) ^{1}H-^{15}N chemical shift perturbations of NS5A D2D3 are shown on *top*, measured by recording BEST-TROSY spectra at 1:0 and at 1:1 ratios of NS5A D2D3:Bin1 SH3 at 5 °C and 800 MHz. They were summed according to $((10\delta(^{1}\mathrm{H}))^{2} + (\delta(^{15}\mathrm{N}))^{2})^{1/2}$. The PxxP binding region is highlighted in *green*, remote regions affected by this binding event are highlighted in *blue*. Secondary C$^{\alpha}$ chemical shifts for unbound NS5A D2D3 (*black*) and 1:1 NS5A D2D3:Bin1 SH3 (*red*) are shown, measured from BEST-TROSY HNCA spectra recorded at 25 °C at 800 MHz field.

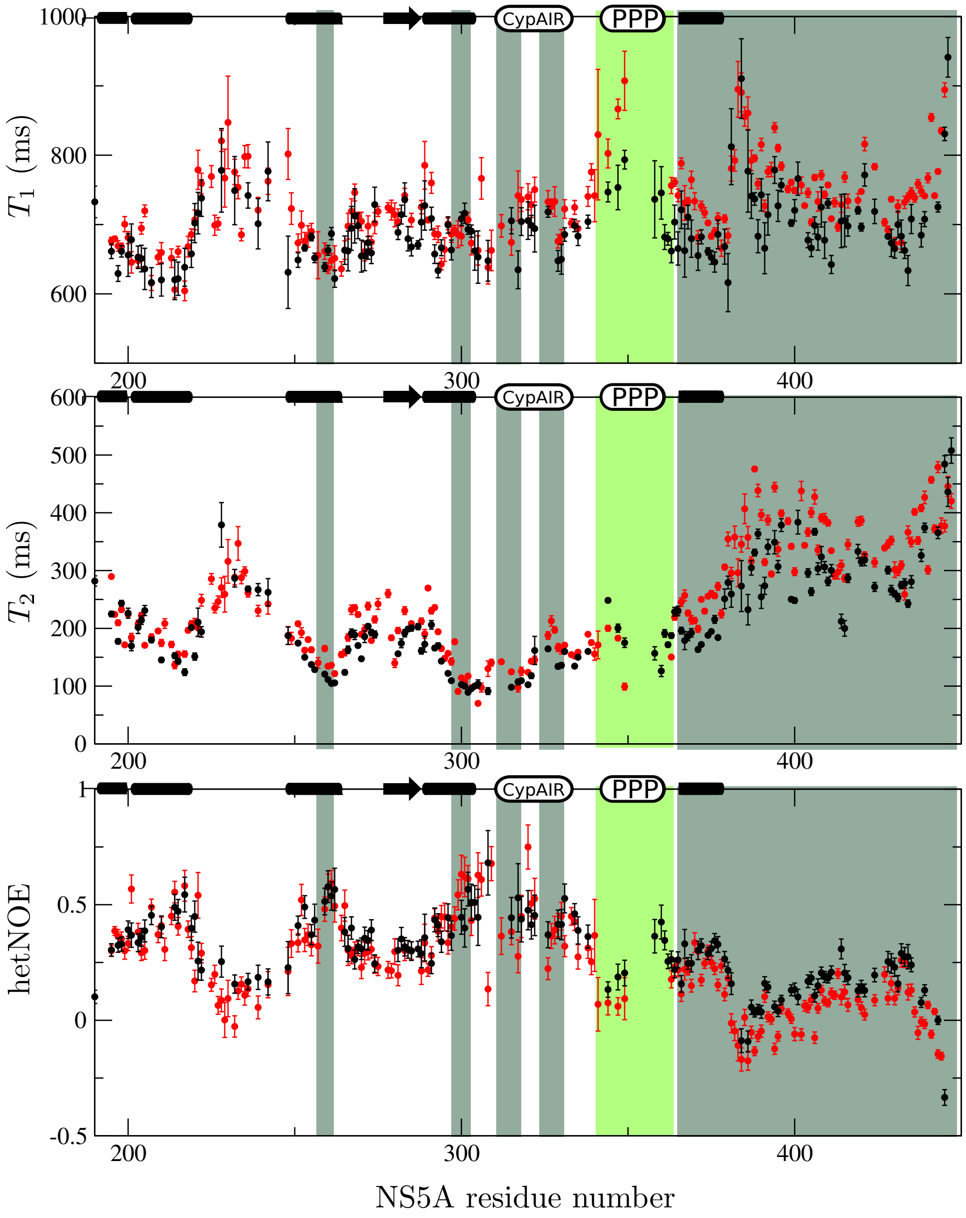


Figure 9.6: ^{15}N relaxation data recorded at 800 MHz of NS5A D2D3 (*black*) and its complex (*red*) formed by one equivalent of Bin1 SH3 domain. The regions affected by appearance of two sets of peaks in slow exchange are highlighted.

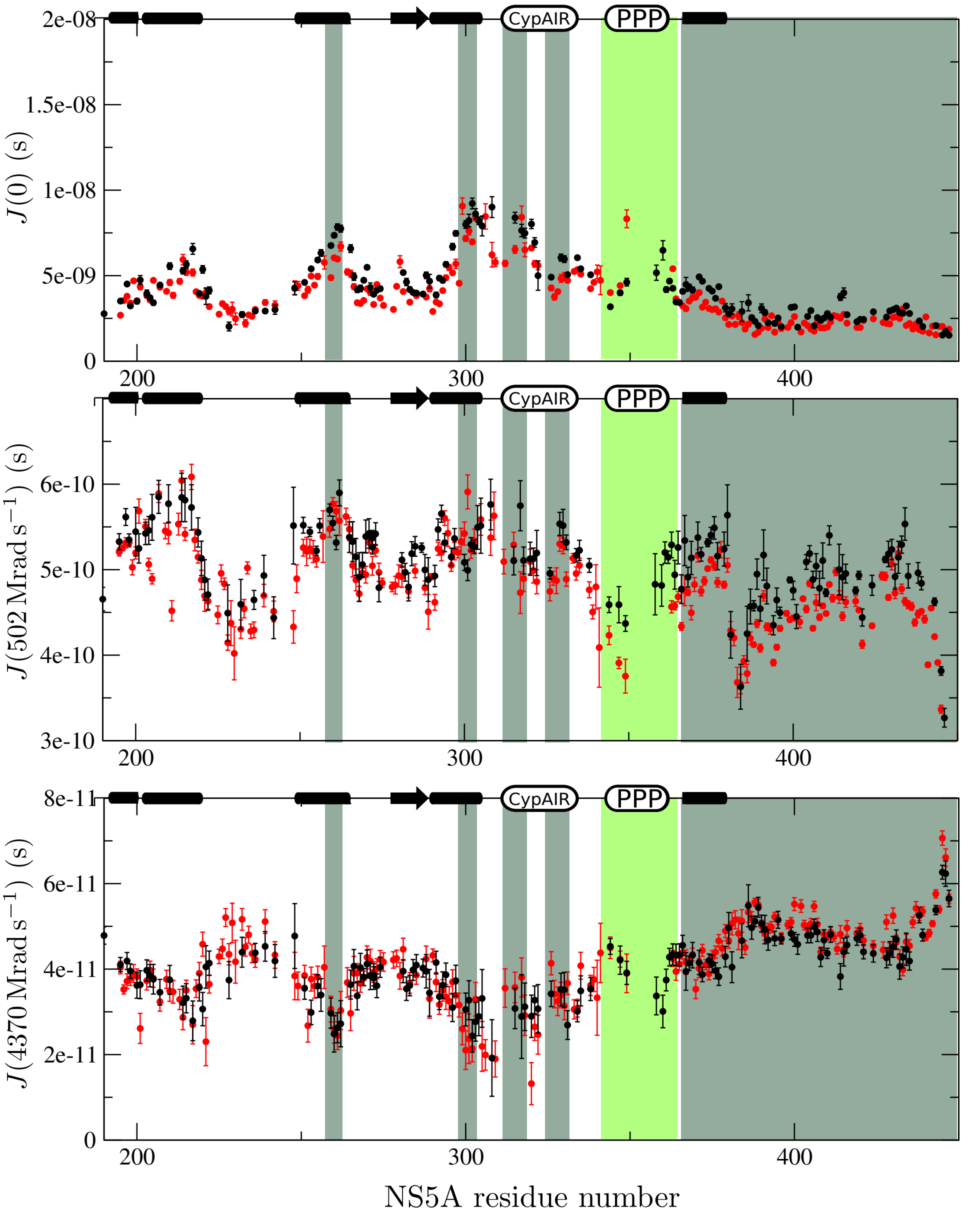


Figure 9.7: Power spectral densities of NS5A D2D3 (*black*) and its complex with Bin1 SH3 domain (*red*) extracted by reduced spectral density mapping approach from the ^{15}N relaxation data recorded at 800 MHz at 25 °C.

The spectral densities at zero frequency $J(0)$ are proportional to the rotational tumbling correlation time τ_C for a rigid spherical molecule. In case of an IDP the situation is more complex, but in general higher $J(0)$ is likely to mean slower tumbling. The data shows that the entire domain 3 is tumbling faster in the complex than in the free protein. In addition some effect can be observed as well on the broad peaks of the region of residues 316-320, but the large error bars do not allow to draw major conclusions about this region from the data. This indicates, that D3 is more flexible in the complex than in the unbound state.

We also performed ITC titrations in order to characterize the thermodynamics of the binding, the titration curves are shown in Figure 9.8. The titration was performed by Melanie Schwarten at the biophysical platform of the PSB site. The thermodynamics of binding of Bin1 SH3 to the D2 and D2D3 constructs has been compared as well. The resulting K_D was found to be higher when binding to D2D3 than when binding to D2, this means binding affinity is slightly reduced by the presence of domain 3. This could be a result of the interaction becoming energetically less favored as a long range, energetically favorable interaction (probably governed by electrostatics) is disrupted. This result also supports the hypothesis of Bin1 SH3 binding to the PxxP motif impairing the transient long range interaction of NS5A domain 3 with domain 2.

Gel filtration has also been performed on the free protein and on the complex, the data are shown in Fig. 9.9. The complex is slightly larger as seen by gel filtration than the free state, this is in agreement with the hypothesis of hindrance of a transient long range interaction by binding, though the effect is minor. Certainly, there is no compaction as would be expected by D3 establishing direct contacts with Bin1 SH3.

MALLS data have been measured on the high affinity complex, as well as on the free proteins. The data is shown in Fig. 9.10. As mentioned before, MALLS allows measurements of molecular weights. The measured molecular weights are shown in Table 9.2. The measured molecular weight of NS5A D2D3 is higher (35 kDa) than the expected 26 kDa for the monomeric state. 35 kDa is not an integer multiple of 26 kDa (the monomeric MW). Therefore, SEC cannot resolve the different oligomeric states. The different oligomers are supposed to exchange on a faster timescale than what allows separation by SEC. A minor peak could also be observed with 26 kDa molecular weight. However, it is unlikely that this would correspond to monomeric NS5A D2D3, as mentioned before, the timescale of exchange between monomers and oligomers is faster than what can be separated by SEC. Surprisingly, Bin1 SH3 showed a tendency to oligomerize in the low ionic strength buffer used for the NMR experiments. This explains the measured 27 kDa molecular weight (the experiment was repeated with with high ionic strength as well and Bin1 SH3 was found to be monomeric). This oligomerization of Bin1, however, did not perturb its binding to NS5A D2D3. The

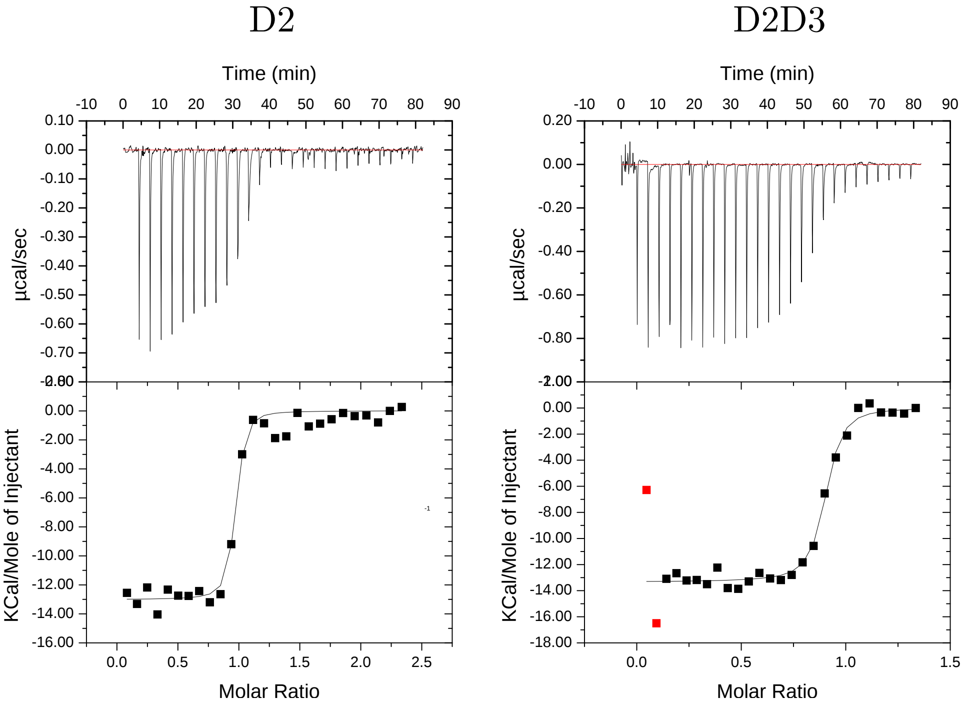


Figure 9.8: Isothermal titration calorimetry data of NS5A D2 (*left*) and D2D3(*right*). The heat response to step-wise injection of Bin1 SH3 domain is shown in top as function of the molar ratio. The reaction heat is shown in bottom assuming a stoichiometric binding and the fit to the data as continuous line.

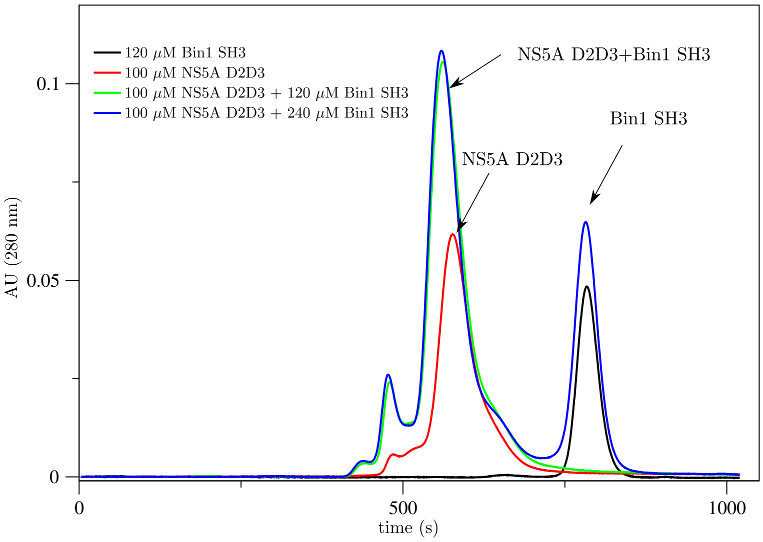


Figure 9.9: Size exclusion chromatography data on NS5A D2D3 and its 1:1 complex formed with Bin1 SH3.

	MW (kDa)	mass fraction (%)
NS5A D2D3		
peak 1	35.1 ± 2%	85.3
peak 2	26.2 ± 11%	14.7
NS5A D2D3+Bin1 SH3		
peak 2	44.6 ± 1.3%	68.3
peak 4	36.2 ± 7.8%	17.8
Bin1 SH3	27.1 ± 5.7%	76.2

Table 9.2: MALLS measurement results showing the measured molecular weights for the peaks eluting after separation by SEC, according to the numbering in Fig. 9.10

data confirmed, that one equivalent of Bin1 SH3 domain binds NS5A, as mass increase for both of the NS5A containing peaks was about 10 kDa, corresponding to the molecular weight of Bin1 SH3 domain. The slight expansion upon complex formation can also be seen here on the elution times.

Overall, the available data suggests that upon binding of Bin1 SH3 domain to NS5A D2D3 a long range interaction between the two domains is hindered. Based on the model of NS5A D2D3 shown at the end of the previous chapter, we suggest a model illustrating how transient long range interactions are altered by binding of the SH3 domain (Fig.9.11). A proof of this would be showing the interaction between the domains by PREs. As mentioned before, the main obstacle is the tendency of NS5A D2D3 to oligomerize, as seen by SEC, MALLS and SAXS. As intra- and intermolecular PREs would be indistinguishable, data analysis would be very difficult or even impossible.

9.4 Interaction of Bin1 SH3 with the low affinity NS5A binding sites

Having discussed the results on the characterization of the high affinity complex we will now discuss the impact on the structure of NS5A of its low affinity interactions with Bin1 SH3. To study the low affinity interactions only, a smaller construct was expressed and purified (NS5A 191-342). This construct is NS5A ΔPxxP lacking the PxxP motif located in LCSII (Fig.8.1)

As has been seen from chemical shift perturbations of the D2D3 construct upon titration with Bin1 SH3, two of the transiently structured α-helical

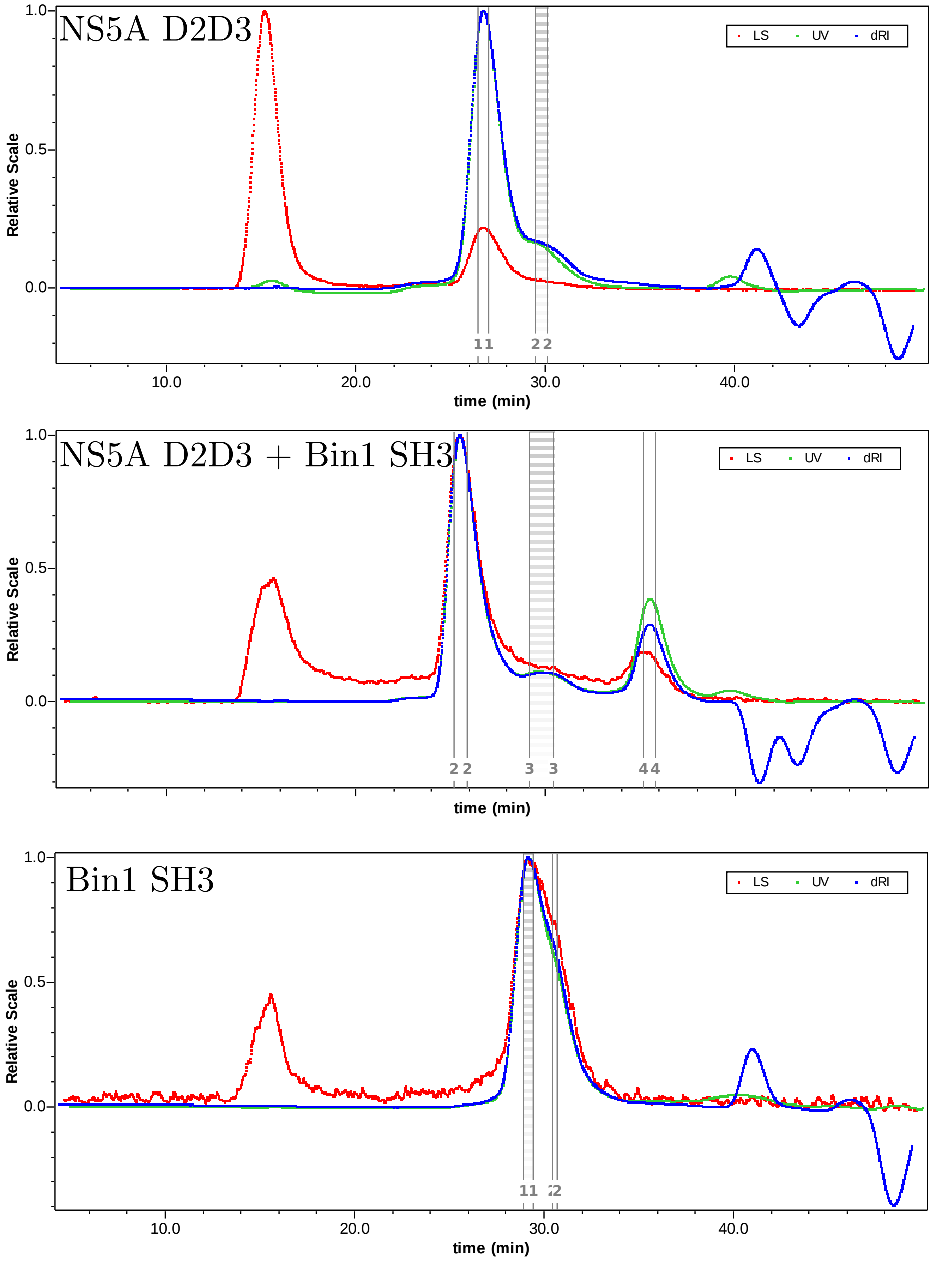


Figure 9.10: MALLS data of NS5A D2D3 (*top*), the 1:1 complex on Bin1 SH3 (*middle*) and NS5A D2D3 and Bin1 SH3 domain alone (*bottom*).

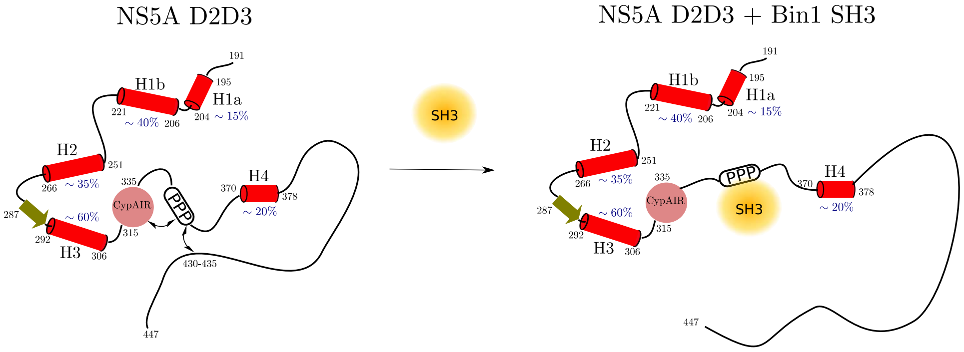
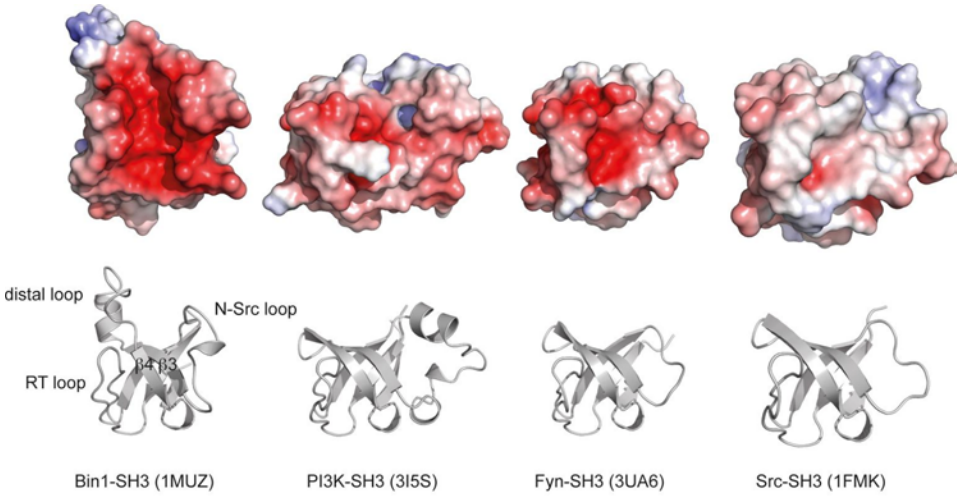


Figure 9.11: Model of the altered transient long range interactions within the structural ensemble of NS5A D2D3 upon binding of an SH3 domain to its PxxP domain.

regions, H1 and H3 are affected by low affinity binding of Bin1 SH3 to NS5A D2D3 (denoted as binding regions B1 and B2), and also minor chemical shift perturbations could be observed in the region corresponding to the Cyclophilin A interaction region of NS5A (residues 316-320). Furthermore, broadening of the amide resonances could be observed in the center of these binding sites, presumably due to exchange dynamics on the μs-ms timescale and/or increased rotational tumbling correlation time (τ_C).

Our first observation upon inspection of the sequence was that the residues corresponding to the broadened peaks are mostly positively charged. In PxxP motifs at least one positively charged residue has to be present adjacent to the prolines (PxxP). This is also true for the PxxP region of NS5A which interacts with Bin1 SH3 with high affinity, the stretch of proline residues are followed by several arginines and lysines. Thus we mapped electrostatics on the structures of the SH3 domains (Figure 9.12). It is apparent from the figure, that the canonical PxxP binding groove of Bin1 SH3 is the most negatively charged among all the studied SH3 domains. This observation hints the role of electrostatics in these low affinity, non-canonical interactions. Examining the charge distribution of the amino acids in the transiently helical regions of NS5A (Fig. 8.10) shows that indeed, H1 and H3 contain several positively charged residues, as opposed to H2 that is mainly negatively charged.

The titrations performed prior to this thesis in the lab (Figure 9.1) show that binding affinity of the low affinity B1 and B2 binding sites of NS5A to the various SH3 domains, as suspected, correlate with the amount of negatively charged residues in the binding pocket of the SH3 domains. The data shows, that the chemical shifts of low affinity non-canonical binding sites on NS5A are most perturbed by binding to Bin1 SH3, which is the most negatively charged, while in the case of Src SH3 practically no interaction was seen at binding site B1 at the transiently helical region H1.



Bin1-SH3 (1MUZ) PI3K-SH3 (3I5S) Fyn-SH3 (3UA6) Src-SH3 (1FMK)

Figure 9.12: Electrostatic mapping of the structures of SH3 domains, including Bin1 SH3. Negatively charged residues are mapped in *red*, positively charged residues are mapped in *blue*.

In order to characterize changes in structure and dynamics of NS5A upon binding to the binding regions B1 and B2, we performed titrations on the ΔPxxP construct and tried to saturate the binding sites. C^{α} and C' chemical shift changes were measured on the complex and compared to chemical shifts of the free state. ^{15}N relaxation data were also recorded on the bound form and compared to relaxation data of the free form (Figure 9.13).

Our data shows that secondary $^{13}C'$ and $^{13}C^{\alpha}$ chemical shifts in B1 and B2 decrease in the complex compared to the bound form. In addition, heteronuclear NOEs alo decrease upon binding, as shown by the data in Fig. 9.13. These observations indicate that the tendency to form transiently structured α helices is lost in these region upon binding to Bin1 SH3. A possible explanation for these data can be given considering the fact that SH3 domains are known to bind left handed PPII conformation. $^{13}C^{\alpha}$ chemical shifts of PPII conformations are very close to random coil chemical shifts, they cannot be distinguished from them. Thus our data suggests that PPII conformations are selected and stabilized by binding to the SH3 domain, or the structure of NS5A in the bound form is random coil-like. Both interpretations would also explain the observed decrease in heteronuclear NOEs. Non-proline residues, especially glutamines and alanines can also adopt PPII conformations [162]. Actually, a pure alanine polypeptide adopts PPII conformation in solution as shown by CD spectroscopy [163]. Smaller heteronuclear NOE ratios mean faster ps-ns timescale motions of the amide NH bond. In a PPII helix, in contrast to α-helical and β-strand

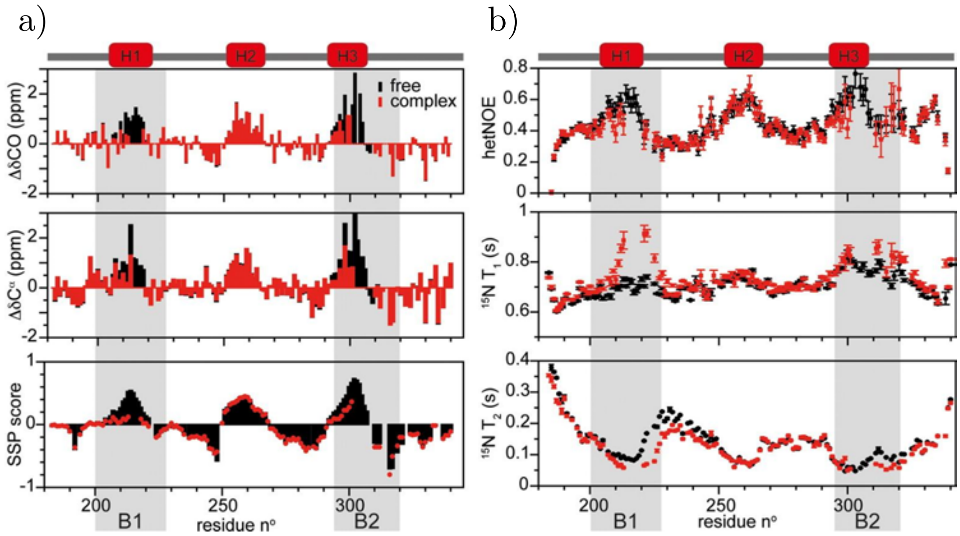
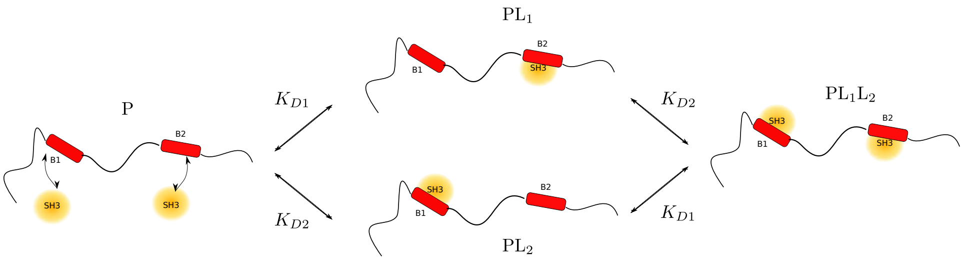


Figure 9.13: Characterization of changes in secondary structure (a) and dynamics (b) upon non-canonical binding of NS5A ΔPxxP to Bin1 SH3 domain. The data for the free protein are shon in *black* and for the 1:10 complex shown in *red*.

conformations, amide protons are not involved in hydrogen bonding. Thus the lack of hydrogen bonding may result in larger freedom for fast motions of the N–H bond. In this scenario conformational selection occurs, but rather than the preformed α-helix, PPII conformations are selected. As shown by Ozenne *et al.* [164], it is not possible to determine PPII sampling, based on ^{13}C chemical shifts alone. They showed that ^{1}H and ^{15}N chemical shifts are more sensitive to it, but these chemical shifts are strongly influenced by the pH and temperature. A set of RDCs combined with chemical shifts makes it possible to detect sampling of PPII conformation, but as shown in Section 8.9, RDCs could not be obtained for our NS5A constructs due to interactions with the alignment media.

Another possible explanation is the formation of a "fuzzy" complex, a term that has appeared in the literature recently for describing complex formation by transient interactions where disorder prevails in the bound state. Some of the examples of such an interaction seem to be governed by electrostatic contacts [27, 165] In a "fuzzy" complex at least one binding partner remains disordered, in other words interchanges between different conformations in the bound state. This binding mode is opposed to the conformational selection or folding upon binding mechanisms, which explain interactions of IDPs where the IDP adopts a well-defined folded conformation in the bound state. Due to increased tumbling correlation time of the complex and/or conformational dynamics in the bound state the NMR signals are broadened beyond detection in the center of the binding site, thus it is not possible to decide on this issue.



$$[\mathrm{P}]+[\mathrm{L}] \xleftrightarrow{K_{D1}} [\mathrm{PL_1}] \qquad [\mathrm{P}]+[\mathrm{L}] \xleftrightarrow{K_{D2}} [\mathrm{PL_2}]$$

$$[\mathrm{PL_1}]+[\mathrm{L}] \xleftrightarrow{K_{D2}} [\mathrm{PL_1L_2}] \qquad [\mathrm{PL_2}]+[\mathrm{L}] \xleftrightarrow{K_{D1}} [\mathrm{PL_1L_2}]$$

Figure 9.14: The model of the independent two-site binding event with comparable affinities, assumed for fitting for K_Ds.

Measurement of chemical shift perturbations upon titration allows determination of dissociation constants. In the case of the two low affinity binding events we fitted the data to a model where two binding sites are competing for the ligand with comparable affinity. Furthermore, binding to one of the sites is independent of whether a ligand is already bound to the other site, so the corresponding K_Ds are assumed to be the same (Fig. 9.14). We fitted for K_Ds with according to this model. The chemical shift perturbations observable at a given point of the titration depend on the populations of the different states P, PL_1, PL_2 and PL_1L_2

$$\begin{aligned} \Delta\delta_{B1} &= \Delta\delta_{max}([\mathrm{PL_1}] + [\mathrm{PL_1L_2}])/\mathrm{P_0} \\ \Delta\delta_{B2} &= \Delta\delta_{max}([\mathrm{PL_2}] + [\mathrm{PL_1L_2}])/\mathrm{P_0} \end{aligned} \tag{9.1}$$

where $\Delta\delta_{B1}$ and $\Delta\delta_{B2}$ are the observed chemical shift perturbations for a residue in B1 or in B2, respectively. We made the assumption here, that chemical shift perturbation at the site B results from a ligand binding to that given site. The concentrations of the different states in equilibrium can be calculated according to the equations shown in Figure 9.14. With $[\mathrm{P_0}]$ denoting the initial concentration of NS5A and $[\mathrm{L_0}]$ that of Bin1 SH3 domain. The concentrations of the other forms are denoted according to the illustration in the figure. After some algebra we get the following equations for the concentrations of the different forms:

$$\begin{aligned} [\mathrm{L}] &= -a/3 + [2(a^2 - 3b)^{1/2} \cos(\phi/3)]/3 \\ [\mathrm{P}] &= ([\mathrm{P_0}] K_{D1} K_{D2})/d \\ [\mathrm{PL_1}] &= ([\mathrm{P_0}][\mathrm{L_0}] K_{D2})/d \\ [\mathrm{PL_2}] &= ([\mathrm{P_0}][\mathrm{L_0}] K_{D1})/d \\ [\mathrm{PL_1L_2}] &= ([\mathrm{P_0}][\mathrm{L_0}^2])/d \end{aligned} \tag{9.2}$$

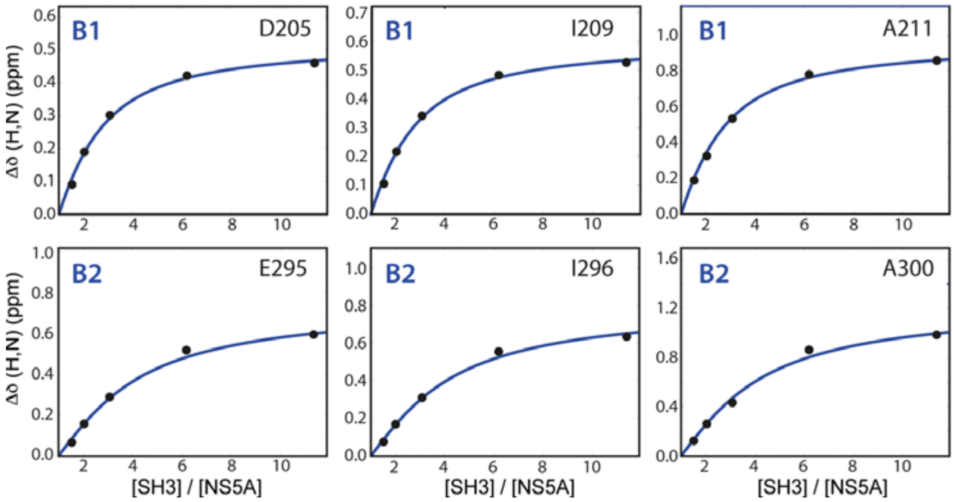


Figure 9.15: Fits for selected residues obtained by global fitting of the experimentally observed chemical shift perturbations to Eqs. 9.2 and 9.3.

where

$$
\begin{aligned}
a &= K_{D1} + K_{D2} - [\mathrm{L}_0] + 2[\mathrm{P}_0] \\
b &= K_{D1}K_{D2} - [\mathrm{L}_0]K_{D1} - [\mathrm{L}_0]K_{D2} + [\mathrm{P}_0](K_{D1} + K_{D2}) \\
c &= -[\mathrm{L}_0]K_{D1}K_{D2} \\
d &= K_{D1}K_{D2} + [\mathrm{L}_0](K_{D1} + K_{D2}) + [\mathrm{L}_0]^2 \\
\phi &= \arccos\,(-2a^3 + 9ab - 27c)/[2\sqrt{(a^2 - 3b)^3}]
\end{aligned}
\tag{9.3}
$$

Chemical shift perturbations were fitted globally to the equations above with a script written by Enrico Rennella, with assuming identical K_Ds for each residue within a binding site. The resulting K_Ds were $K_{D1} = 100 \pm 50\,\mu\mathrm{M}$ and $K_{D2} = 240 \pm 50\,\mu\mathrm{M}$. Representative fits are shown in Figure 9.15.

These results show that the binding affinity is 2-3 times higher for B1 than for B2. As mentioned before, $^{13}C^{\alpha}$ and $^{13}C'$ chemical shift data show that the increased α-helical sampling of region H1 and H3 get reduced upon binding the SH3 domain. The lower binding affinity of B2 compared to B1 may be explained by the higher propensity of NS5A to sample α-helical conformations in the H3 region that in the H1 region.

These binding affinities are comparably low. However, Bin1 has been shown to form dimers via its BAR domain [166]. This could result in a very stable complex formed *in vivo* because of the spatial proximity of the second SH3 domain to the low affinity binding sites after binding of the first one to the high affinity binding site, B3 (the PxxP motif). This binding could play a role for example in inhibiting interaction of transient helices H1 and H3 with some, so far unknown binding partner. Interaction with Bin1 has been reported to reduce phosphorylation of NS5A [160]. It is possible, that one of the low affinity binding sites contains a phosphorylation site. Bin1 SH3 binding could prevent interaction of the responsible kinase with the phosphorylation site. However, all these possibilities remain speculations at present.

Chapter 10

Study of phosphorylation of NS5A

Post-translational modifications (PTMs), among them phosphorylation are chemical events, capable of changing the properties of a protein. Phosphorylation of an amino acid introduces a negatively charged phosphate group at the place of otherwise neutral or positively charged amino acids. This may also result in a change of the overall conformation of the protein and therefore result in a change of function. In case the overall conformation is not changed, phosphorylation may fulfill its role of changing the properties of the protein, by changing the amino acid (and charge) compositions of a short linear motif (SLiM) and therefore changing the ability of interaction.

NS5A protein of HCV is a phosphoprotein that exists in at least two distinct phosphoforms as seen by SDS-PAGE. In this chapter our NMR studies on the phosphorylation of the D2D3 construct of NS5A are described.

10.1 Post-translational modifications

One of the multiple roles of proteins in living organisms is to exert regulatory functions. Post-translational modifications often play roles in regulatory processes. PTMs can either occur directly after translation in order to produce the mature protein, or later, depending on the stage of the cell cycle, as a result of a regulatory mechanism in order to tune the properties of the molecule according to a specific requirement. An example for the first scenario is the formation of disulfide bridges. Phosphorylation, on the contrary, usually belongs to the latter group of PTMs. Post-translational modifications, including phosphorylation, may be able to turn on or off enzymatic functions. An example of the latter is the case of the Src kinase, which when phosphorylated, is incapable of phosphorylation of its substrate [167]. Phosphorylation, such as other PTMs, is a reversible process and is a result of an enzymatic reaction. Transfer of one phosphate group to the protein requires

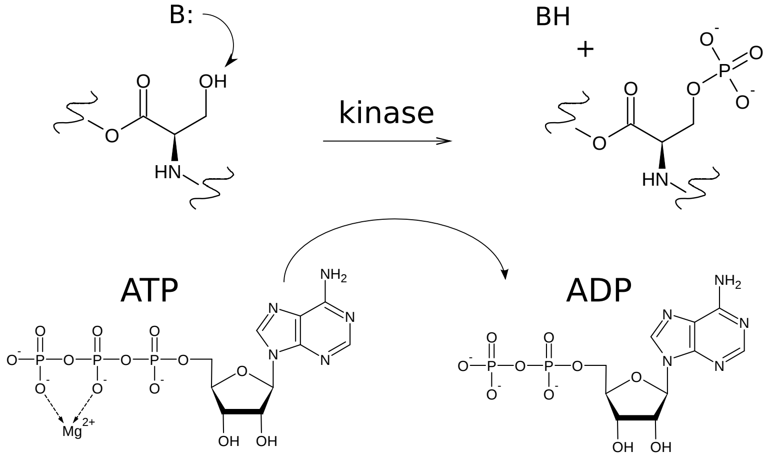


Figure 10.1: Mechanism of phosphorylation.

energy, that is provided by hydrolysis of an ATP molecule as shown in Fig. 10.1. The enzymes responsible for phosphorylation are termed kinases, while the enzymes responsible for dephosphorylation are called phosphatases. The most frequently phosphorylated residues are serines and threonines. Tyrosine phosphorylation is also common and perturbations of tyrosine kinase signaling is associated with development of several diseases such as cancer and diabetes. Phosphorylation of positively charged residues also occurs, and produces phosphoramidite bonds that are less stable compared to phosphoesters. Phosphorylation of histidine residues is relatively common (more frequent than tyrosine phosphorylation) and there are also examples of lysine and arginine phosphorylation.

10.2 NMR as a tool to study phosphorylation

Mass spectrometry (MS) is the most commonly used method for identification of phosphorylation sites. However, MS studies require purification of sufficient amount of protein from cell lysates and their subsequent digestion, in order to obtain fragments that can be ionized. Difficulties in identification of the exact location of phosphorylation sites by MS can be encountered if there are several potential phosphorylation sites in proximity, for example a cluster of serines, or if the sequence is repetitive. Another possible method is immunodetection, but this assay requires specific antibodies to be available for the phosphopeptides corresponding to the potential phosphorylation sites.

Once the role of a kinase in a phosphorylation event is established, phosphorylation can also be studied *in vitro* by NMR spectroscopy, as shown in the pioneer study of Landrieu *et al.* [168]. They showed that phosphorylation can be followed in a time-resolved manner by recording a series of HSQC spectra exploiting the fact that phosphorylation results in a large downfield shift for protons of the amide resonances. Furthermore, if ^{15}N-labeled protein can be delivered into the cell in sufficient quantities, phosphorylation can be followed *in vivo* and analyzed by NMR, provided that there is an assignment of the fingerprint spectrum of the protein available and the fingerprint spectrum in-cell is well enough resolved to detect site-specific changes in a time-resolved manner. This was shown for the first time by Selenko *et al.* [169].

Since these first studies there have been numerous NMR studies on phosphorylation of different proteins *in vitro* as well as *in vivo*. Recently, in-cell NMR phosphorylation studies have been performed to follow the response of the cell for certain stimuli, such as UV irradiation, by following changes in the phosphorylation pattern [87].

IDPs are especially suited for such studies, as usually there are no amide signals in the unphosphorylated form of the protein corresponding to the proton frequencies of phosphoserines. However, the size of the proteins that can be studied is limited because of signal overlap it is difficult to follow phosphorylation-induced changes. In case of lack of well-resolved resonances quantification of the kinetics is not possible. Furthermore, the elevated pH and temperature required for kinase activity imposes a further challenge for NMR spectroscopy because of line broadening of the amide resonances due to chemical exchange and increase of the proton T_1 of water, as discussed in detail in Chapters 5 and 6. Thus resolution- and sensitivity-optimized methods are required to conduct such studies especially for the crowded spectra of IDPs.

10.3 Phosphorylation of NS5A

Post-translational modifications of all the HCV proteins have been recently reviewed [170]. Phosphorylation of NS5A has been subject to intense research. NS5A is known to play roles in RNA replication and in viral particle assembly and seems to regulate these processes. It has been hypothesized that phosphorylation induces a switch between these crucial functions of the protein but there is no direct evidence yet, and the molecular mechanism remains unclear.

Early studies of NS5A have already shown that there are two forms of NS5A migrating on a gel, termed hyperphosphorylated with apparent MW of 58 kDa and basal-phosphorylated, with an apparent MW of 56 kDa. Phosphoamino acid analysis has shown that NS5A is predominantly phosphorylated on serine residues with a minor fraction of phosphorylated threonines present [171]. In addition, it has been shown that presence of NS3 is required for phosphorylation of NS5A. Several efforts followed in order to find the kinases involved in phosphorylation of NS5A and to identify the phosphorylation sites. By kinase inhibitor assays, mutagenesis studies and *in vitro* phosphorylation assays casein kinase 1 (CK1), casein kinase 2 (CK2), Polo-like kinase 1 (Plk1), glycogen synthase kinase 3 (GSK3) and p38 mitogen-activated protein kinase (p38MAPK) have been suggested to be involved in phosphorylation of NS5A.

Furthermore, it has been shown, that alanine mutations in the serine cluster of domain 3 result in impairment of the interaction of NS5A and the HCV core protein, an interaction that has been shown to be crucial for viral particle assembly [128]. CK2 has been shown in this study to play a role in phosphorylation of these serines. Its positive regulatory role on viral particle production has been shown by RNA silencing and use of a CK2 inhibitor.

Mutations introduced in the region of NS5A, where the suggested hyperphosphorylation sites are, besides reduction of hyperphosphorylation of NS5A, resulted in increased replication. In addition, replication-incapable mutants of NS5A could be rescued by introduction of mutations in that region. Thus it was shown, that phosphorylation regulates the role of NS5A in replication.

The role of CK1α in phosphorylation of NS5A has also been rather well established, by design of kinase inhibitors inhibiting production of hyperphosphorylated NS5A and subsequent kinase inhibitor affinity chromatography [172]. By *in vitro* phosphorylation with CK1δ, it was found that prephosphorylation of S229 is required for efficient phosphorylation of S332 by CK1δ *in vitro*. Recently, it has been shown by mass spectrometric analysis of NS5A purified from HCV replicating cells that several serines in the cluster of residues starting from S222 are phosphorylated in hyperphosphorylated NS5A [173]. Phosphorylation in this region has been shown to regulate the interaction of NS5A with hVAP-A [174]. Interestingly, Gupta *et al.* found that residues in domain 3 are responsible for binding hVAP-B (assuming similar binding to hVAP-A and hVAP-B) and binding seemed to be stronger with D3 construct alone as compared to the full-length [165]. It is worth noting, however, that upon dissecting NS5A they did not include the LCS I region including the putative hyperphosphorylation sites in any of their constructs.

A very recent comprehensive study by Masaki *et al.* has adopted a kinome-wide screening approach for identification of kinases that bind NS5A and analyzed their in vitro phosphorylation abilities for phosphorylation of NS5A [175]. They have also identified CK2α, CK1α, Plk1 and PKAC and some other CK1 and CK2 isoforms to bind and phosphorylate NS5A in vitro. They performed a wide range of *in vivo* tests on effects of silencing of these kinases on virion production and replication. The largest effect on the viral particle production was found by silencing of CK1α. By phosphomimetic and phosphoablatant mutations as well as mass spectroscopy they found that residues S225 and S232 played major roles in viral particle assembly and are phosphorylated by CK1α. To summarize, it is now well established, that hyperphosphorylation occurs (at least partially) in the cluster of serine residues starting from S222 in LCS I, and CK1α plays a role in the phosphorylation of these residues.

Despite the numerous studies outlined above, the question of sequentiality of phosphorylation events remained unanswered, and the direct connection between function and phosphorylation sites with a detailed mechanism of action has yet to be estabished. Masaki *et al.* speculate that there is a host factor interacting differentially with hyperphosphorylated and basal phosphorylated NS5A and transfers it to the lipid droplets where viral particle assembly takes place. Further studies will be needed to elucidate the exact mechanism, but the hVAP-A and DGAT1 proteins have been suggested to be potential host factors that could mediate this process.

Before the comprehensive study of Masaki *et al.* took place, in the course of this thesis we decided to conduct a study on in vitro phosphorylation of the NS5A D2D3 construct. We tested *in vitro* phosphorylation by some kinases that have been suggested in the literature to play a role in phosphorylation of NS5A. Our major candidates were CK2, CK1α and Plk1, and in addition we also tested PKA. The results will be presented in the following sections.

10.4 Phosphorylation site predictions

At the beginning of our studies we performed predictions of the phosphorylation sites on NS5A. We used the Netphos and NetphosK webservers [176] for prediction of phosphorylation sites for NS5A D2D3. The results are summarized in Fig. 10.2. On panel a) it is shown that many sites, mainly serines but also threonines and a few tyrosines are predicted to be phosphorylated (having phosphorylation potential over the threshold), with the highest probability for S222. Several sites in domain 3 are also predicted to become phosphorylated. Panel b) shows kinase specific predictions based on the consensus sequence of the substrate of the kinase. CK2 is predicted to phosphorylate several residues in domain 3 near the C terminus, a prediction consistent with the findings of Tellinghuisen *et al.* [128]. The cluster of serines in LCSI reported in the context of hyperphosphorylation are predicted to become phosphorylated, but the predicted kinases (PKC and DNAPK) have not been reported in the literature to be involved in phosphorylation of NS5A. The interesting exception is S222, which is predicted to get phosphorylated by three kinases, and all three of them have been reported in the context of NS5A phosphorylation. The three kinases are p38MAPK, PKA and GSK3. Interestingly, S222 is the site that is required to be prephosphorylated, so that CK1 can phosphorylate the other serines within the cluster. In the literature Plk1 has been suggested to be responsible for this prephosphorylation. however, the NetPhosK server does not predict for Plk1. Another useful site for prediction of phosphorylation is the ELM resource [19] that gives information on the presence of linear motifs that are the recognition sites of different kinases based on their recognition motifs (Fig. 10.3). These predictions and the data available in the literature was our basis for the selection of the kinases listed at the end of the previous section, for pursuing with our *in vitro* phosphorylation studies.

10.5 NMR studies of NS5A phosphorylation

Our strategy was to use as little quantity of protein as possible for an initial test. In these tests we recorded a reference spectrum, dialysed the protein to the kinase buffer, added the kinase and incubated the system overnight. Then we dialysed the protein back to the NMR buffer and recorded again the spectrum in identical experimental conditions as the reference. In this way we tested CK2, CK1α, Plk1 and PKA kinases.

10.5.1 Phosphorylation by PKA

A phosphorylation test was performed with PKA. Although it phosphorylates NS5A *in vitro*, it has been reported probably not to be biologically relevant for NS5A phosphorylation [171]. The reason why we tested this kinase, in spite of this, is that it is predicted to phosphorylate serine 222. Phosphorylation of S222 has been suggested to be the required pre-phosphorylation for phosphorylation by CK1 of residues in the serine cluster in LCSI.

The phosphorylation was not followed in real-time. We conducted a test in order to see which residues become phosphorylated by PKA, and in case S222 is phosphorylated, to pursue with phosphorylation by CK1. The protein was incubated with PKA kinase at room temperature overnight. The reference spectrum and the spectrum recorded after phosphorylation, and after dyalisis to NMR buffer is shown in Figure 10.4. An amide resonance appears at 9.35 ppm proton chemical shift, and a second, less intense amide resonance appears in the region characteristic for phosphoserines.

Line broadening beyond detection was observed for residues 358-363 in the spectrum and shifts for several residues especially in the transiently helical region H1. Inspection of the prediction results shown on Fig. 10.2 shows that T360 is predicted to be phosphorylated by PKA, probably the amide of this phosphorylated threonine resonates at 9.35 ppm proton chemical shift. The correlation peaks of residues near T360 should be shifting because of the changing chemical environment. However, no additional correlation peaks could be observed, only near residue K358, but there is also a His residue nearby that seemed to have shifted due to a slight change in pH. Otherwise a few new peaks can be observed in the spectrum with low intensity, that may result from the minor phosphorylation site, or belong to residues 358-363. The reason for broadening of residues near T360 is yet unclear. It may be,

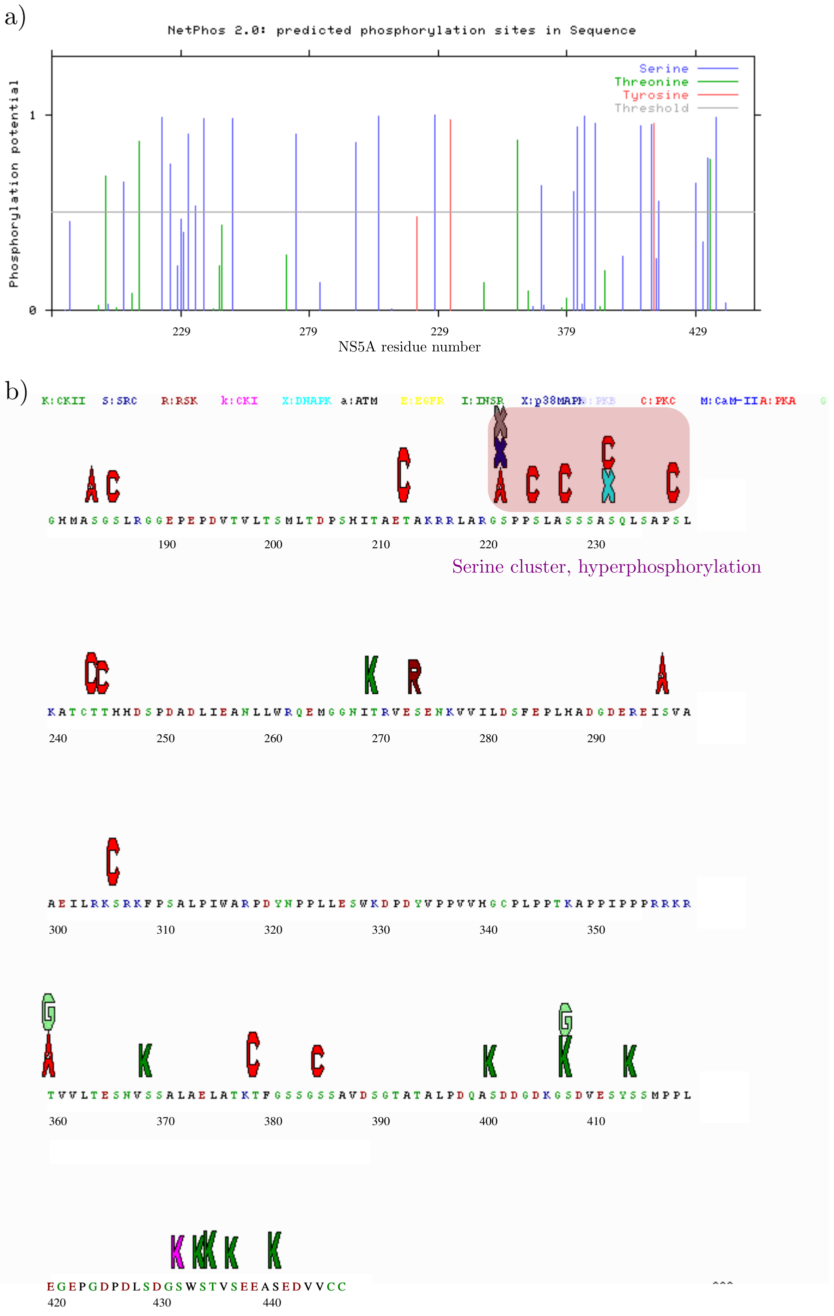


Figure 10.2: Prediction output of the NetPhos software. a) The predicted phosphorylation sites are shown, the column height is proportional to the probability of the given site to become phosphorylated. Serines are shown in *blue*, threonines in *green* and tyrosines in *red*. b) Kinase specific predictions by NetPhosK. The different kinases are denoted with letters in different colors according to the label. The size of the letter is proportional to the likelihood of phosphorylation by that kinase.

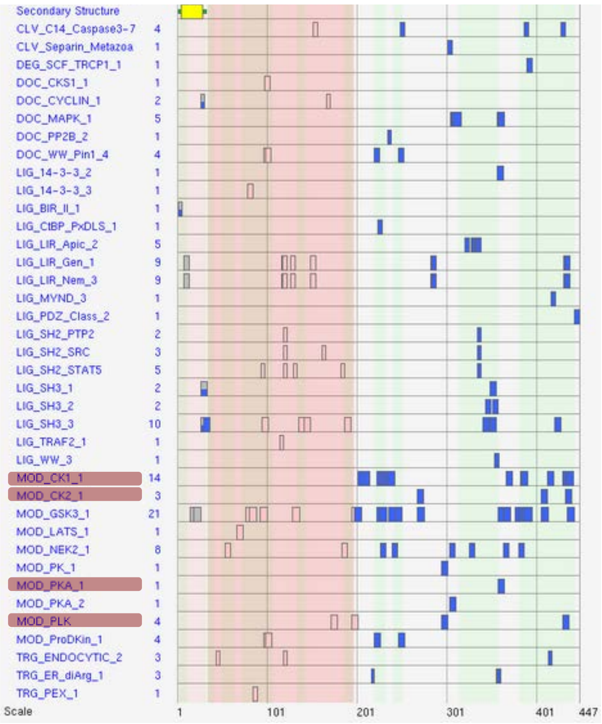


Figure 10.3: Linear motifs identified by a search for the full-legth NS5A sequence on the ELM webserver. The kinase recognition sites relevant for our study are highlighted in red. The region of the protein predicted to be disordered by IUPRED is shown with green background while the region predicted to be folded is shown with red background. Blue filled rectangles show the sequence position of the identified, presumably exposed linear motifs, while linear motifs in the folded region are shown by transparent rectangles.

that they are resonating at different frequencies, but overlapping with other peaks, or they are broadened because of structural alterations.

The other amide resonance, appearing at chemical shifts characteristic for phosphoserines could not be identified. S222 showed only slight intensity change in the spectra recorded before and after phosphorylation (highlighted in Fig. 10.4 a). Such intensity changes, however, could be observed for most of the amino acids, as the major phosphorylation site seemed to cause alterations of the structure in remote regions.

We identified one major phosphorylation site being phosphorylated by PKA (supposedly T360), and a minor site, probably a serine, phosphorylated significantly slower. The identity of this second site is yet unclear.

However, as biological relevance of this kinase for NS5A phosphorylation *in vivo* was not established at the time of our investigation, and S222 was not found to be phosphorylated by PKA with fast kinetics, we did not pursue this line of investigation.

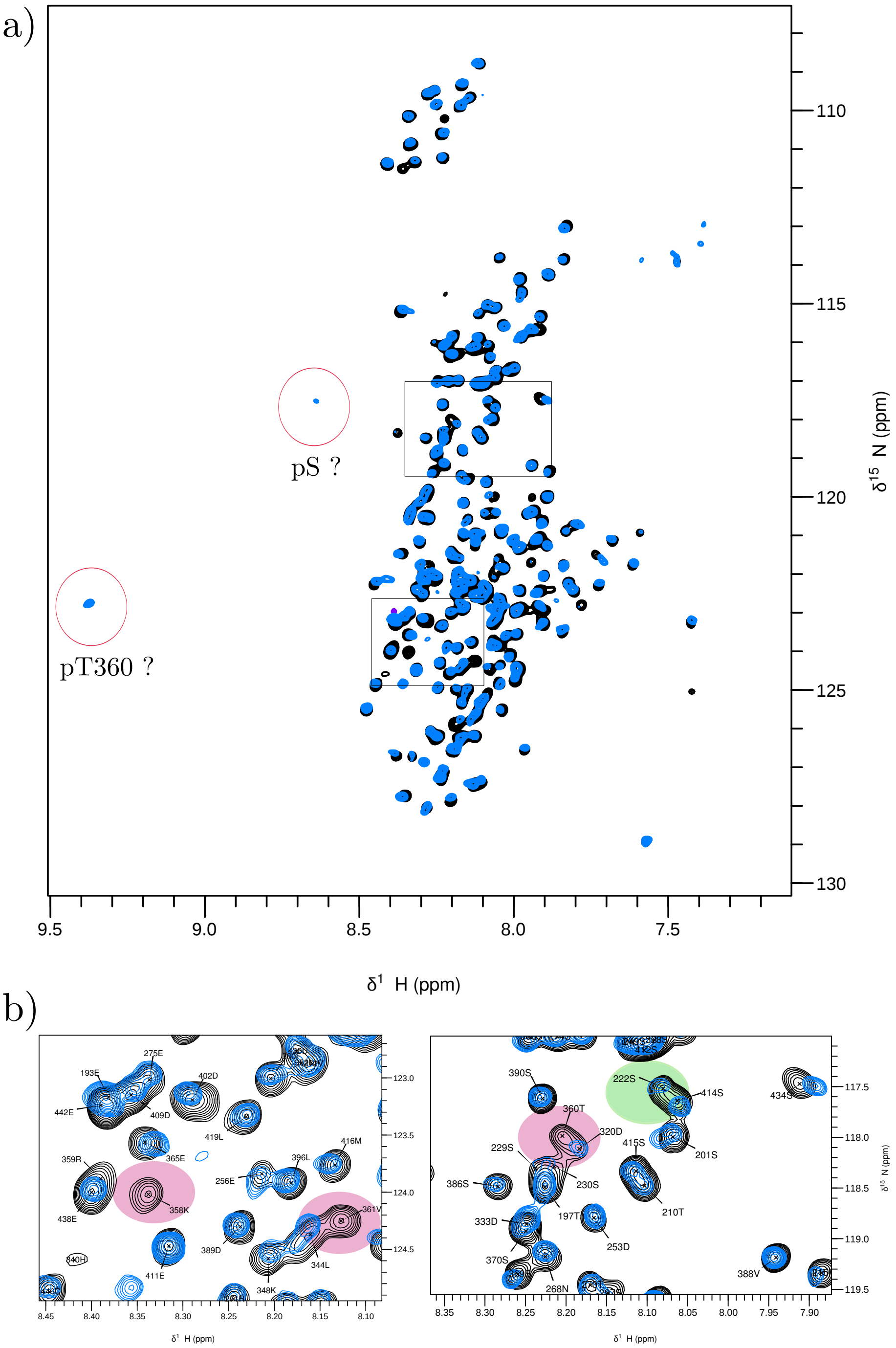


Figure 10.4: a) NS5A BEST-TROSY spectra recorded before *(black)* and after *(blue)* phosphorylation with PKA in the NMR buffer. Recorded at 800 MHz field, at 5 °C. b) Zoom of selected regions of the spectra shown in a), showing the observed line broadening for residues 361 and 358 (left) and 360 (right). S222, showing only slight intensity change is highlighted in green.

10.5.2 Phosphorylation by Plk1

We attempted phosphorylation of NS5A D2D3 with Plk1 kinase that has recently been identified as a kinase relevant for NS5A phosphorylation. It was reported to co-localize with NS5A and to phosphorylate it *in vitro* [177]. The consensus recognition motif of Plk1 is [E/D]X[S/T][I/L/V/M]X[E]. S297 and S432 correspond to this recognition motif, therefore they are expected to become phosphorylated by Plk1. In addition, phosphorylation of serines or threonines by other kinases might result in further Plk1 sites, because of introduction of negatively charged residues, necessary for recognition.

In spite of repeated attempts we were unable to detect phosphorylation of the sample. The reason could be that proteolytic degradation of NS5A is competing with phosphorylation, and the sample degrades before we could detect phosphorylation-induced changes. Another possible reason is that the presence of D1 is required for the phosphorylation by Plk1.

10.5.3 Phosphorylation by CK2

CK2, besides CK1, is the kinase most mentioned and studied in the context of NS5A. In this study we focused on a more detailed characterization of the phosphorylation by this kinase, as its biological relevance has already been established by others, especially by the study of Tellinghuisen *et al.* [128]. In their study, they showed that silencing of CK2 disrupts virus production, and DMAT, the selective CK2 inhibitor, has the same effect. The major phosphorylation site identified by their study is S457, based on mutagenesis of several serine residues. They conducted their studies on NS5A from HCV 2a genotype, where S457 is located in the context of a SEED sequence, that is an ideal recognition site for CK2. However, sequence alignment of this region between different genotypes shows variability of the negatively charged residue responsible for recognition by CK2. For example in our construct (1b genotype) the aspartate is replaced by an alanine. Sequence alignment also shows the presence of highly-conserved serines in this region of the protein. It is worth noting, that the authors also observed phosphorylation when S457 was mutated to an alanine. Furthermore, phosphorylation by CK2 was found important for interaction of NS5A with Core.

The spectra obtained from the test phosphorylation are shown in Figure 10.6. Four resonances appeared in the phosphoserine region of the spectrum (to of them partially overlapping), and a fifth one with low intensity. From this we deduced that four serines were phosphorylated during the overnight incubation with CK2. From comparison of the reference and the phospho-NS5A spectra, with the assignment of the disappearing residues and the shifting neighboring residues we suggest that the residues that became phosphorylated are: S408, S429, S432, S434. The general substrate of CK2 is SXXE/D, but there are also examples of SXE/D and SE/D sub-

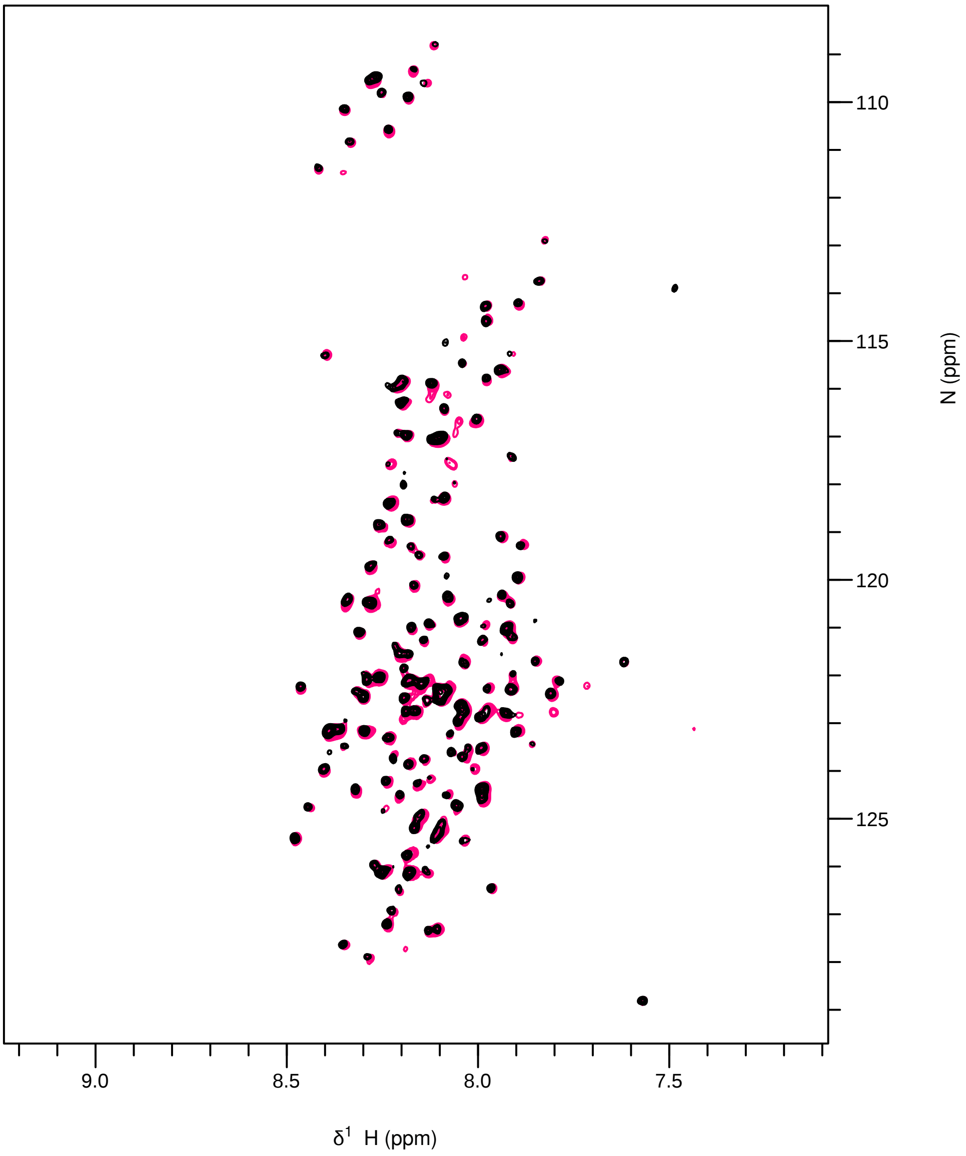


Figure 10.5: NS5A BEST-TROSY spectra recorded before (*black*) and after (*pink*) phosphorylation with Plk1 in the kinase buffer. Recorded at 800 MHz field, at 5 °C.

strates. For residue S408 there is a D in position $n+1$, and in position $n-3$. S429 also has a D neighbor in position $n+1$, furthermore when S432 and S434 become phosphorylated additional negatively charged residues will be introduced in the surroundings. S437 is predicted to be a good substrate of CK2, but interestingly, instead, S434 became phosphorylated that is not predicted to be as good substrate of CK2 (Fig. 10.2). Interestingly, we also observed shifting residues in the region far from the phosphorylated sites; this could be a result of modification of the long range interactions upon phosphorylation. These shifts are the largest for residues in regions H1 and H3, furthermore near the PPII motif, around residue 360. These regions are characterized by stretches of positively charged residues (and coincide with the Bin1 SH3 high affinity and low affinity binding sites). This observation can probably be explained by perturbation of the transient, electrostatically mediated, long range interactions, by introduction of the charged phosphoserine residues. The experiment was repeated in a different setup with a larger amount of protein in order to gain sensitivity and to be able to follow the kinetics of the reaction. The results of the study of kinetics will be discussed in Section 10.6.

The structural changes were also further analyzed. In spite of repeated efforts we were not able to record relaxation data on the phosphorylated sample to characterize its dynamics, because of short lifetime of the sample due to proteolytic degradation (commercial kinases always seem to be contaminated with proteases, they are typically only of 80% purity)

10.5.4 Phosphorylation by CK1δ

According to the study of Quintavalle *et al.* [172], CK1α is likely to be the biologically relevant isoform of CK1 for NS5A phosphorylation. However, in their study the commercially available CK1δ was used to phosphorylate NS5A in vitro. The reason for using the δ isoform is that it is commercially available, unlike CK1α. The *in vitro* phosphorylation produced two distinctly phosphorylated species migrating on a gel a similar way to *in vivo* isolates of basal and hyperphosphorylated NS5A. We attempted to reproduce these results and to follow the kinetics by NMR. The further objective was to assign the two species that were observed and suspected to be basal- and hyperphosphorylated NS5A.

We found in our first NMR experiment that phosphorylation by CK1δ of NS5A D2D3 was slow compared to phosphorylation by the other tested kinases. However, phosphorylation was observed after several days, at eight sites, but none of them was complete. We found the major sites phosphorylated by CK1δ *in vitro* to be identical with the ones phosphorylated by CK2 as seen by overlay of the spectra. The minor sites could not be identified. We decided to follow the kinetics of these phosphorylation events and to use this system to test the applicability of the $\mathrm{H}^{\mathrm{N-BEST}}\mathrm{CON}$ experiment,

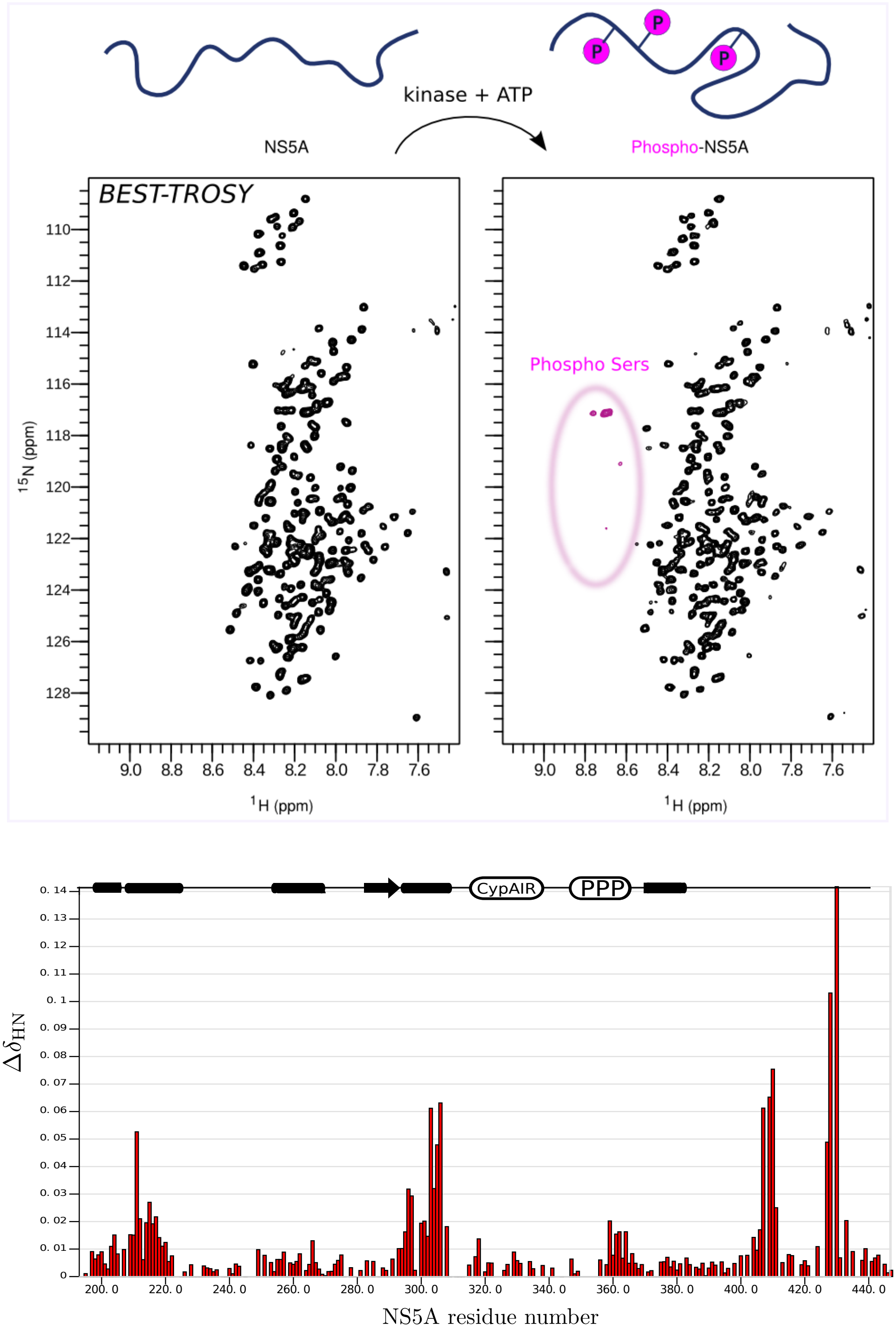


Figure 10.6: BEST-TROSY spectra of NS5A D2D3 before and after incubation with CK2 (*top*). The peaks belonging to the phosphoserines are colored in *pink*. Chemical shift perturbations plotted against the residue number are shown (*bottom*), calculated by $((\delta(^{1}\mathrm{H}))^{2} + (0.1\delta(^{15}\mathrm{N}))^{2})^{1/2}$.

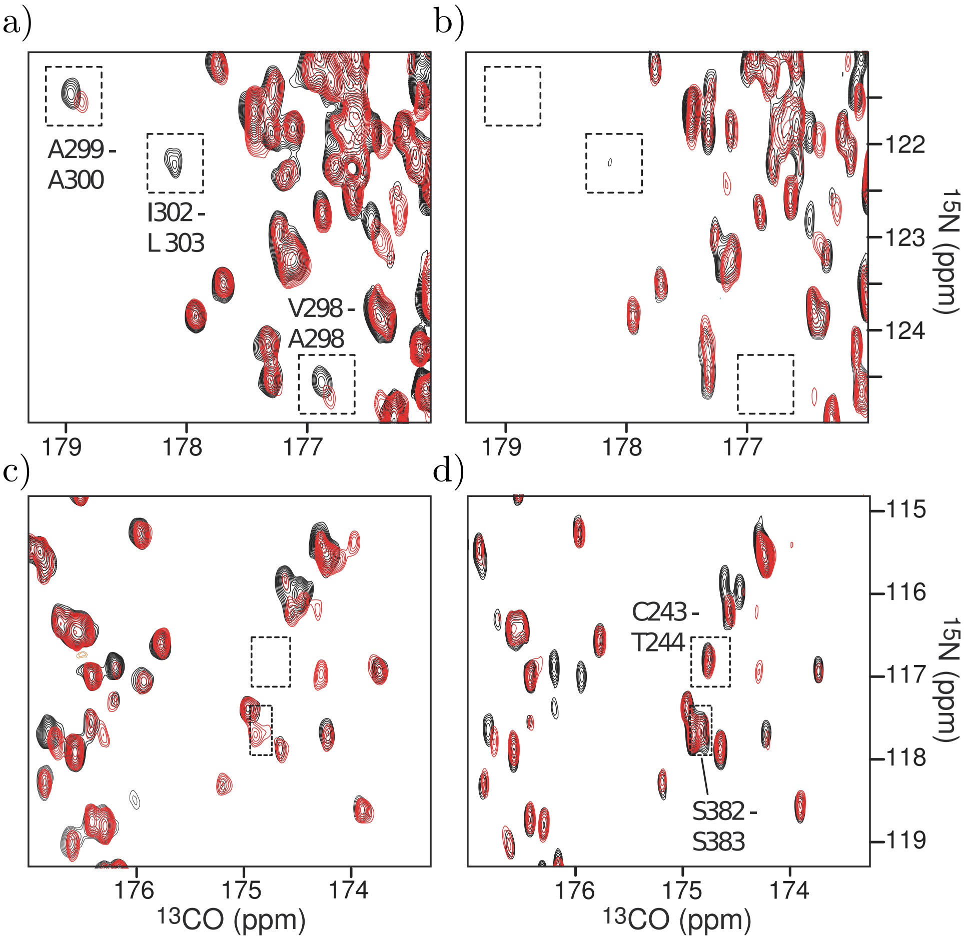


Figure 10.7: Following phosphorylation by the $H^{N-BEST}CON$ experiment (a and c) compared to the CON experiment (b and d). The reference spectra are shown in *black*, the spectra after phosphorylation by CK1δ in *red*. In a) and b) a region of the spectrum is selected showing correlation peaks of residues in transiently structured regions. c) and d) illustrate that for residues involved in very fast exchange with the solvent,

presented in Section 5.8. In order to compare it with the conventional CON experiment, we followed phosphorylation by recording a series of CON and $H^{N-BEST}CON$ spectra in an alternating manner with 4h measurement time each. The experiments were run for a total of four days. As shown in Figure 10.7 a and b, signals from residues in transiently structured regions, such as H3, are observable with much higher intensity in the $H^{N-BEST}CON$ than in the CON experiment. Transiently structured regions profit from longitudinal relaxation enhancement by selective pulsing, therefore a large gain in sensitivity can be achieved for these sites. In the CON spectra they cannot be observed, because of their smaller T_2 relaxation time constants compared to flexible regions. On the contrary, a few sites are not observable in the $\mathrm{H}^{\mathrm{N-BEST}}$CON spectra (Fig. 10.7 c) due to broadening resulting from very fast solvent exchange. These regions are very flexible therefore they are observable in the CON spectrum (Fig. 10.7 d).

10.6 Kinetics of phosphorylation by CK2

Several experimental setups were tested to follow the kinetics of the reaction. The limitation is solvent exchange, the higher the pH and the temperature, the faster the exchange. Serines, because of their chemical properties exchange the fastest among the amino acids. On the other hand, the kinase is only active at elevated temperature and pH. The pH was set to 7.5 as in the specifications, and the temperature was set to 25 °C, to reach a reasonable compromise between kinase activity, solvent exchange broadening of resonances, and protease activity. This latter was the main reason for the repeated tests for optimal conditions, as NS5A is extremely prone to degradation and commercial kinases, as mentioned before, are impure. Protease inhibitor cocktails were tested, but they were found not to be effective against the proteases responsible for degradation of NS5A. Addition of glycerol was also tested, in order to improve spectral quality by slowing down solvent exchange as suggested in Ref. [87]. However, it seemed that glycerol drastically reduced activity of the kinase, while the protease activity was reduced less. This resulted in degradation of the protein becoming faster than its phosphorylation. Therefore, glycerol was not used in further experiments. As outlined in Section 5.6 we found that elevated pH and temperature results in effective proton T_1 relaxation time constants as short as 60 ms. Thus the optimal recycle delay for recovering the signal of exchanging protons was tested and was found to be only 50 ms, following the 70 ms acquisition period. With these settings, each 2D BEST-TROSY spectrum took 10 min time to record.

In this experiment, recorded in a way to allow analysis of the kinetics, the phosphorylation progressed further, more residues became phosphorylated than in the test phosphorylation. Preliminary analysis of the data shows, that S408, S434, S432 and S429 are phosphorylated fastest. However, as there was still too much solvent exchange broadening, the serine resonances could not be used for fits, only the resonances of the neighboring residues. Therefore, it was not possible to decide which of these is phosphorylated first, it is possible that the bound kinase slides and phosphorylates the other sites within one binding event. S414 and S437 are phosphorylated only partially as for the neighboring resonances two sets of peaks are present. The analysis of this data is still in progress. For nearly all peaks, which are not near the phosphorylation sites, there is a slight increase of peak intensities as phosphorylation progresses, such as for residue 363 (shown in

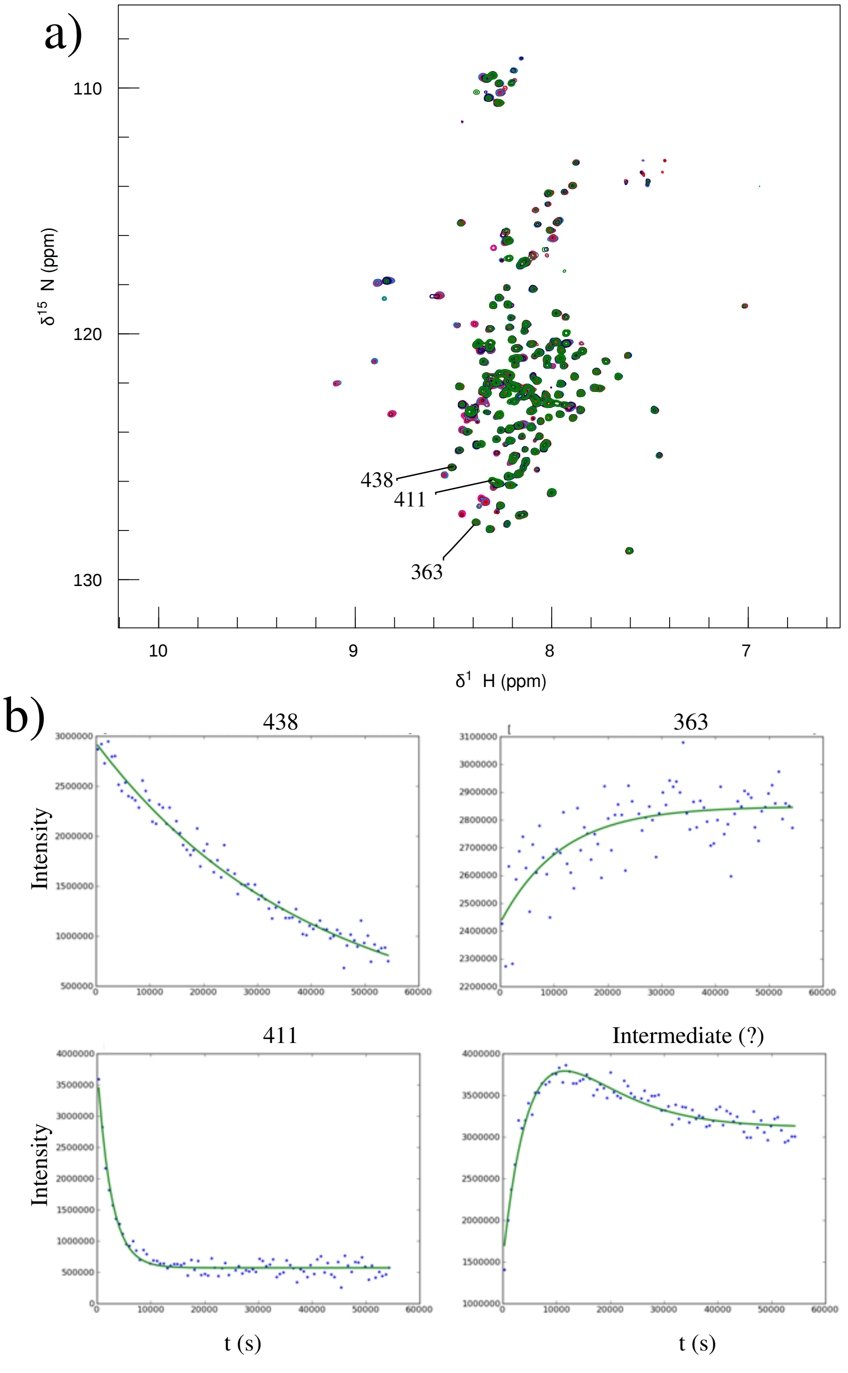


Figure 10.8: Series of BEST-TROSY spectra to follow phosphorylation. The first, fifth, 10th, 20th, 40th, 60th and 40th time points are shown superposed.

residue number	k (s^{-1})
363	$7\times10^{-5}\pm2\times10^{-5}$
401	$2.8\times10^{-5}\pm2.5\times10^{-6}$
407	$3.2\times10^{-4}\pm3.5\times10^{-5}$
433	$2.0\times10^{-4}\pm1.6\times10^{-5}$
438	$2.2\times10^{-5}\pm2.5\times10^{-6}$

Table 10.1: Reaction rate constants of phosphorylation of NS5A by CK2 for selected residues obtained by fitting a monoexponential function $I = I_0 \exp(-kt)$ to the BEST-TROSY peak intensities measured as function of time.

Fig. 10.8 b. Fitted reaction rate constants for selected residues close to the phosphorylation sites, as well as for residue 363, that has no nearby phosphorylation site, are listed in Table 10.1. Fits were obtained by fitting a monoexponential function to the peak intensities. Fitting biexponential functions for the residues under the impact of several phosphorylation sites is in progress. However, a more detailed analysis of these data will require further experiments, as the protein was assigned at pH 6.5, and this experiment was performed at pH 7.5. For the more resolved regions, the NMR assignment could be transferred, but in many cases the assignment is still ambiguous.

The resulting phosphoprotein, after phosphorylation by CK2 and measurement of the kinetics data, has been submitted to mass spectrometric analysis to confirm the number of phosphorylation sites and to obtain information on the ratio of differently phosphorylated states in the sample. Furthermore, we wanted to exclude the presence of proteolytic degradation, that would compromise the kinetics data.

The data shows that the sample became multiply phosphorylated and no significant degradation took place. The spectra show, that the phospho-NS5A sample is heterogeneous, most of the protein became phosphorylated at 6 sites but there is also 5-fold and 7-fold phosphorylated NS5A present in the sample. Their ratios can also be estimated from the data. 0.6:1:0.6 are the relative intensities of the peaks belonging to the 5-fold:6-fold:7-fold phosphorylated NS5A species.

10.7 Conclusions and perspectives

The phosphorylation of NS5A is a long-studied subject. We conducted *in vitro* phosphorylation studies with several kinases, and used NMR spectroscopy as the method for identification of the phosphorylation sites.

We chose kinases, whose role in phosphorylation of NS5A was rather well established. Therefore we used CK2, CK1δ and Plk1 for our phos-

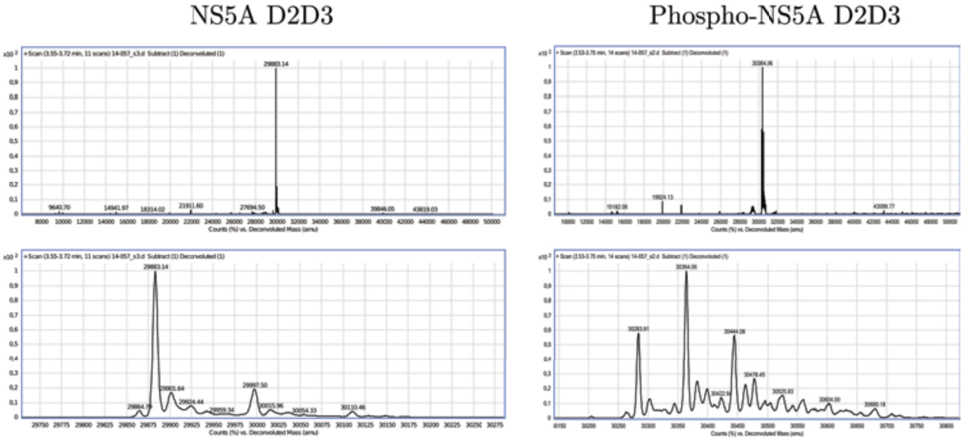


Figure 10.9: Mass spectrometric data of NS5A D2D3 (*left*) and CK2 phospho-NS5A D2D3 (*right*).

phorylation studies and PKA, that was reported to phosphorylate NS5A *in vitro* in an early study [178], but its biological relevance was not yet established at the time of our study. The reason for our investigation was that it was predicted to phosphorylate S222, yielding the phosphoserine required for phosphorylation by CK1. We identified phosphorylation at T360, but not, or only partially at S222, so we did not pursue our experiments with this kinase. Shortly before finishing writing this thesis, an article was published [179] that showed the biological relevance of this kinase in RNA replication of the virus and reported the same phosphorylation site as we found in our study. However, our observation of NMR line-broadening of some resonances around the phosphorylation site is not in agreement with the intensity increase that was reported in their study. This discrepancy may require further investigation. In addition, we observed that this phosphorylation had an impact on the structural ensemble of the protein as sen by intensity changes and small chemical shift changes also in regions far from the phosphorylation site. Now, that biological relevance of this kinase has been established, these structural changes may also deserve further investigation.

CK2 was found to phosphorylate residues in D3, though different ones than S457 (corresponding to S437 in our genotype), reported to be phosphorylated by Tellinghuisen *et al.* on NS5A of HCV genotype 2a [128]. Actually, we found the kinase to phosphorylate a cluster of three serines (S429, S432, S434) with the highest preference. In addition to these sites, S408 and S414 were identified to be phosphorylated, with lower preference. This finding is surprising, as S429, S432 and S434 are not in the context of the most typical recognition motif of CK2, with a negatively charged residue at position $n + 3$. For S434, however, an E can be found at sequence position $n + 4$.

Preliminary data on the kinetics suggest that these sites are phosphorylated with similar rates, therefore it is possible, that the kinase is not releasing the substrate, before phosphorylation of these sites. To confirm this hypothesis, further experiments will be required, including assignment at pH 7.5

We conducted phosphorylation by CK1δ. We found that the kinetics was much slower with this kinase than with PKA or CK2. Substrate recognition by CK1 requires a phosphoserine. Predictions show, that provided that S222 became phosphorylated, CK1 could phosphorylate several serines among the serines of the hyperphosphorylation serine cluster.

Summarizing the above stated, we identified several new CK2 phosphorylation sites in D3 of NS5A. In addition, we have shown that phosphorylation in D3 results in slight chemical shift perturbations in positively charged regions of D2, at H1, H3 and near residue 360. The SWS motif (residues 432-434) is highly conserved among the HCV genotypes. Phosphorylation of these residues may be required for an interaction, possibly for its interaction with the highly positively charged core protein of HCV. Actually, the effect of the CK2 inhibitor, DMAT, was reduction of viral partial assembly via modulation of the interaction of core and NS5A according to Tellinghuisen *et al.* [128]. A short-term perspective of this work is to conduct interaction studies between CK2-phospho-NS5A and core protein of HCV, compared to interaction of unhosphorylated NS5A with core of HCV.

Appendices

List of publications

- Transient structure and SH3 interaction sites in an intrinsically disordered fragment of the hepatitis C virus protein NS5A. Feuerstein S, Solyom Z, Aladag A, *et al.*, *J. Mol. Biol.*, 2012.
- BEST-TROSY experiments for time-efficient sequential resonance assignment of large disordered proteins. Solyom Z, Schwarten M, Geist L *et al.*, *J. Biomol. NMR*, 2013.
- Interaction of nonstructural protein 5A of the hepatitis C virus with Src homology 3 domains using noncanonical binding sites. Schwarten M, Solyom Z, Feuerstein S, *et al.*, *Biochemistry*, 2013
- NMR spectroscopic studies of intrinsically disordered proteins at near-physiological conditions. Gil S, Hosek T, Solyom Z, *et al.* *Angew. Chem. Int. Ed. Engl.*, 2013.
- Measuring hydrogen exchange in proteins by selective water saturation in ^{1}H-^{15}N SOFAST/BEST-type experiments: advantages and limitations. Rennella E, Solyom Z and Brutscher B, *J. Biomol. NMR*, in press

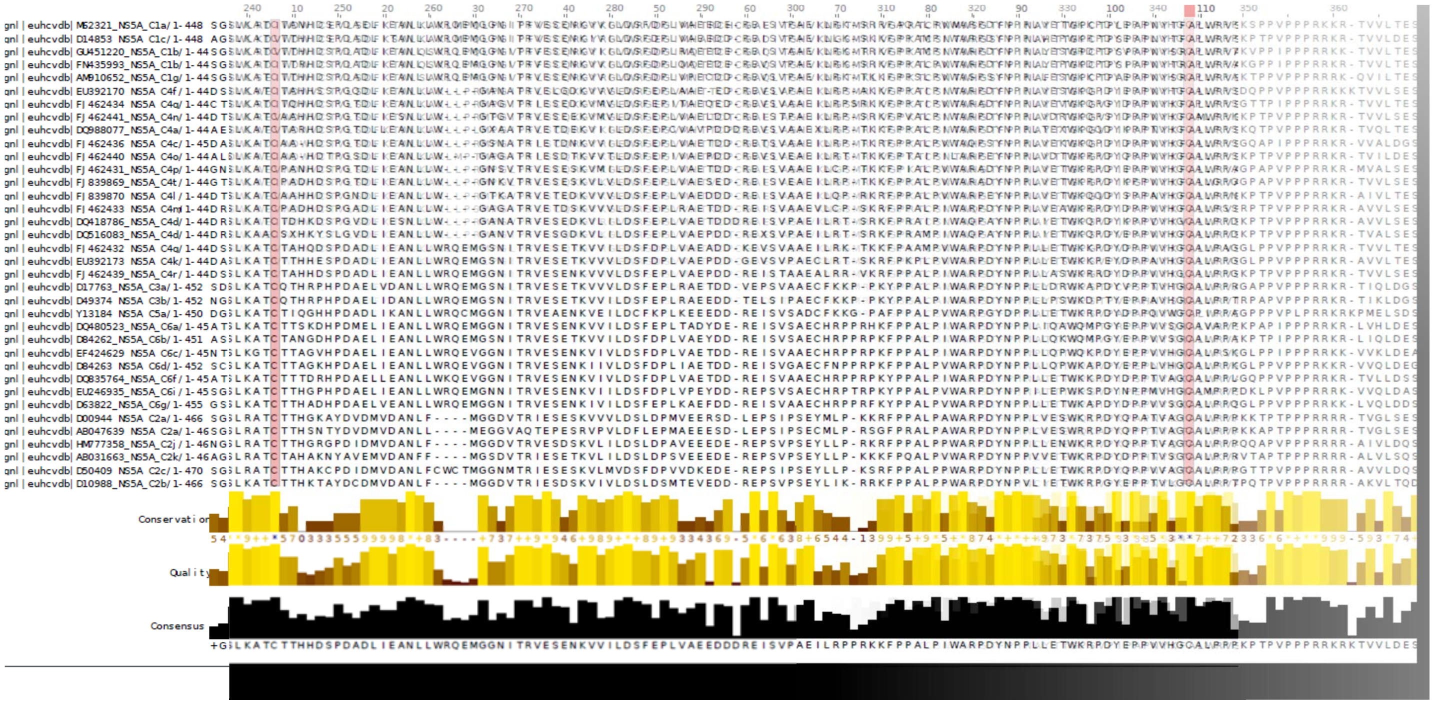


Figure 10: Multiple sequence alignment of NS5A of example sequences of all the NS5A genotypes found in the European HCV database. The alignment has been made by the ClustalW software. The figure illustrates conservation of 243C and 342C (highlighted in *red*).

Expression and purification of His-CTev_NS5A_D2D3

Day 1: Transformation of pET-41c-His-D2D3 in *E. Coli* Bl21 DE3
incubate over night at 37 ºC on LB-Kanamycin plates

Day 2: 1st pre-culture in 5 ml LB with Kanamycin
in the evening centrifuge pre-culture
resuspend cell pellet in 50 ml M9-Km
over night at 37 ºC

Day 3: centrifuge over-night culture
resuspend cell pellet in 1000 ml M9-Km
until OD at 0.6-0.8 at 37 ºC
induce expression with 1 mM IPTG
4 h at 25 ºC
haverst cells at 4000 xg for 20 min
resuspend cell pellet in 1x PBS
centrifuge again
store cell pellet at -80 ºC

Purification

Buffers:

Lysis-Buffer: 100 mM NaH_2PO_4 pH 8.0 adjusted with NaOH
10 mM Tris-HCl
8 M urea
20 mM imidazole
5 mM DTT

Wash-Buffers: Lysis-Buffer with increasing concentrations of imidazole
Wash 1 20 mM imidazole
Wash 2 50 mM imidazole
Wash 3 100 mM imidazole
Wash 4 250 mM imidazole
Wash 5 500 mM imidazole

Cleavage buffer: 50mM KH2PO4
200 mM NaCl
2 mM 2-Mercaptoethanol
pH=7.0

- resuspend the cell pellet in 5 ml per g Lysis-Buffer
- incubate for 30min RT on shaker
- sonificate for 2 min on ice with 50 % amplitude 2 sec. on 8 sec. off
- centrifuge for 45 min. at 45000 xg
- incubate for 30 min. the supernatant with 3.5 ml Ni-NTA (equilibrated with Lysis-Buffer) -->

use more the next time!

- wash with 2 CV of each wash buffer
- pool His C NS5A-D2D3 containing fractions (W2, W3, W4)

Cleavage

- dialyse against cleavage buffer
- cleavage for 2 h in RT, with 1mgTEV/10-20mg protein, pH=7
- (TEV partially precipitates during cleavage)

Ni-NTA

- load it on column, incubate at RT for 30 min
- wash with 3 CV of cleavage buffer
- elute with 2 CV of 20 mM, 50mM, 100mM, 500mM imidazole

At this point based on the SEC I did, the protein is pure. But if you want to do a SEC step for safety:

Dialysis : 50 mM KH_2PO_4 pH8
200 mM NaCl
20 mM DTT

- dialysis for 2x 2 h and once overnight against 3 times 1l Dialysis-Buffer
- centrifuge for 20 min at 4000 xg
- concentrate supernatant to 3 ml --> attention... maybe less next time, as protein precipitates!

SEC-Buffer: 50 mM KH_2PO_4 pH8
200 mM NaCl
2 mM 2-Mercaptoethanol

- SEC: High load 16/60 Superdex 75 pg flowrate 1 ml/min
- pool pure NS5A-D2D3 fractions
- concentrate

NMR-Buffer: 50 mM KP_i pH 6.5
20 mM NaCl
2 mM TCEP

- dialyse against NMR-buffer 3 times 2 hours

Optional for sample stabilization:
heating 20 minutes at 90 °C
then add TCEP to a final concentration of 5mM
(Attention, the protein only resists heating in the NMR buffer. At pH 7.5 precipitation was observed!)

J Biomol NMR (2013) 55:311–321
DOI 10.1007/s10858-013-9715-0

ARTICLE

BEST-TROSY experiments for time-efficient sequential resonance assignment of large disordered proteins

**Zsofia Solyom · Melanie Schwarten ·
Leonhard Geist · Robert Konrat · Dieter Willbold ·
Bernhard Brutscher**

Received: 17 January 2013 / Accepted: 15 February 2013 / Published online: 24 February 2013

Abstract The characterization of the conformational properties of intrinsically disordered proteins (IDPs), and their interaction modes with physiological partners has recently become a major research topic for understanding biological function on the molecular level. Although multidimensional NMR spectroscopy is the technique of choice for the study of IDPs at atomic resolution, the intrinsically low resolution, and the large peak intensity variations often observed in NMR spectra of IDPs call for resolution- and sensitivity-optimized pulse schemes. We present here a set of amide proton-detected 3D BEST-TROSY correlation experiments that yield the required sensitivity and spectral resolution for time-efficient sequential resonance assignment of large IDPs. In addition, we introduce two proline-edited 2D experiments that allow unambiguous identification of residues adjacent to proline that is one of the most abundant amino acids in IDPs. The performance of these experiments, and the advantages of BEST-TROSY pulse schemes are discussed and illustrated for two IDPs of similar length (~270 residues) but with different conformational sampling properties.

Electronic supplementary material The online version of this article (doi:10.1007/s10858-013-9715-0) contains supplementary material, which is available to authorized users.

Z. Solyom · M. Schwarten · B. Brutscher (✉)
Institut de Biologie Structurale, Université Grenoble 1, 41 Rue Jules Horowitz, 38027 Grenoble Cedex 1, France
e-mail: bernhard.brutscher@ibs.fr

Z. Solyom · M. Schwarten · D. Willbold · B. Brutscher
Commissariat à l'Energie Atomique et aux Energies Alternatives (CEA), Grenoble, France

Z. Solyom · M. Schwarten · B. Brutscher
Centre National de Recherche Scientifique (CNRS), Grenoble, France

L. Geist · R. Konrat
Department of Computational and Structural Biology, Max F. Perutz Laboratories, Campus Vienna Biocenter 5, A-1030 Vienna, Austria

D. Willbold
Institute of Complex Systems (ICS-6) Structural Biochemistry, Forschungszentrum Jülich, 52425 Jülich, Germany

D. Willbold
Institut für Physikalische Biologie, Heinrich-Heine-Universität, 40225 Düsseldorf, Germany

Keywords BEST · TROSY · IDP · Viral protein · Amino-acid-type editing · Longitudinal-relaxation enhancement

Introduction

The structure–function paradigm stating that a well-defined structure is required for a protein's function has been challenged by the discovery of a large number of highly flexible proteins or protein segments that exist as ensembles of partly collapsed or extended structural conformers in the cell and that are functional as such (Dunker et al. 2001; Tompa 2002; Wright and Dyson 1999; Uversky and Dunker 2010). Intrinsically disordered proteins (IDPs) or protein regions (IDRs) have been shown to play important roles in regulatory and signaling processes where the structural flexibility allows the protein to adapt to and interact with a large number of distinct molecular partners (Tompa 2012). Similarly, structural disorder is also abundant in viral proteins (Davey et al. 2011; Xue et al. 2010). Viruses are often characterized by a small genome, only coding for a few proteins. High mutation rates in these genomes allow to adapt to changing environments and to escape the defense mechanisms of the host cell. Again, structural flexibility presents a functional advantage in terms of binding promiscuity, as well as a high tolerance to mutations.

During recent years, NMR spectroscopy has become the technique of choice to obtain atomic-resolution information for IDPs, and to extract useful information on the structural ensemble that the IDP forms in solution. In particular, NMR allows the identification of peptide regions with increased propensity to form α-helical or extended (β-strand) structures that often play a role in molecular recognition events, or the characterization of transient long-range interactions. Furthermore, NMR is a powerful technique to characterize binding events in terms of interaction surfaces, and to study eventual conformational transitions of the IDP upon binding to its partner(s). The rapid interconversion between the large number of conformations sampled by the ensemble typically results in a single set of sharp NMR signals. However, as a consequence of the lack of a stable structure, NMR spectra of IDPs are characterized by low chemical shift dispersion as compared to well-structured globular proteins, which makes NMR studies of large IDPs a challenging task. In addition, fast solvent exchange of the solvent-exposed labile protons with water protons may result in extensive line broadening of amide proton resonances. Therefore, amide ^{1}H-detected NMR experiments often need to be performed at lower than ambient sample temperature to slow down hydrogen exchange. Alternatively, H^{α}-detected (Mantylahti et al. 2010) or ^{13}C-detected experiments (Felli and Brutscher 2009; Csizmok et al. 2008; Bermel et al. 2012) can be used to circumvent the problem of unobservable amide proton resonances. However, H^{α}-detected experiments require protein samples in deuterated solvents, while ^{13}C-detected experiments suffer from reduced sensitivity due to the lower gyromagnetic ratio of ^{13}C with respect to ^{1}H. Other concerns for NMR studies of IDPs are often limited sample stability, low sample concentrations to avoid protein aggregation, and substantial peak intensity heterogeneities in the NMR spectra. Therefore, sensitive NMR pulse schemes are required to detect correlation peaks also for the sites with the lowest signal intensity; high dimensional (≥3D) NMR experiments with long acquisition times in all dimensions are needed in order to resolve overlapping correlation peaks; and last but not least fast acquisition techniques are mandatory to enable multidimensional data acquisition in a reasonable amount of time.

Herein, we present BEST-TROSY pulse sequences that yield the required sensitivity and spectral resolution for time-efficient sequential resonance assignment of large IDPs from a set of typically four 3D correlation spectra. These experiments are particularly attractive for NMR studies performed at high magnetic field strengths. Compared to other ^{1}H-detected pulse schemes, BEST-TROSY experiments provide significant advantages in terms of experimental sensitivity and spectral resolution as demonstrated for two IDPs of about 270 residues in length, the C-terminal part of the NS5A protein from hepatitis C virus (268 residues including tag) that acts as a multifunctional regulator of cellular pathways, and the chicken BASP1 protein (270 residues including tag) involved in transcription regulation.

Materials and methods

NMR measurements were performed on an Agilent VNMRS 800 MHz spectrometer at 278 K equipped with a cryogenically cooled triple-resonance (HCN) probe, and pulsed z-field gradients.

The presented pulse sequences were tested on two NMR samples: the first contains 0.12 mM [U–^{13}C, U–^{15}N] enriched NS5A (Non-Structural Protein 5A, residues 191–447 plus an 11-residues tag remaining after cleavage of a His tag) from hepatitis C virus in 50 mM potassium phosphate buffer (pH 6.5), 20 mM NaCl, 2 mM β-mercaptoethanol and 5 % (v/v) D_2O, the second sample contains 0.6 mM [U–^{13}C, U–^{15}N] chicken BASP1 (Brain Acid-Soluble Protein 1, 244 residues) in 20 mM citrate buffer (pH 2.0), 0.2 % NaN_3 and 10 % (v/v) D_2O. The BASP1 construct includes a 26-residue cleavable His tag.

To characterize the conformational dynamics of the two proteins, ^{15}N relaxation experiments (T_1, T_2, HETNOE) were performed using standard pulse sequences (Farrow et al. 1994). The relaxation curves were fitted onto the extracted peak intensities from the subspectra measured with 10 relaxation delays, ranging from 0.01 to 1.6 s in the case of T_1 and from 0.01 to 0.25 s for T_2.

Proton T_1 relaxation time constants were measured by inversion recovery experiments, with amide selective inversion, water-flip-back (WFB) and non-selective inversion using the pulse sequence elements of Figure S1. The spectra were recorded with 10 relaxation delays ranging from 0 to 1.4 s in the non-selective and WFB case and 0–0.9 s in the amide selective case. All relaxation curves were fitted to a mono-exponential function using a Python program.

3D HNcoCACB, iHNCACB, hNcocaNH, and hnCOcaNH correlation spectra (Fig. 3) were recorded for NMR assignment of NS5A with the recycle delay set to $T_{rec} = 0.15$ s, the acquisition time to 70 ms with 4 scans per (t_1, t_2) increment. In the ^{15}N dimension, 150 complex points were acquired for a spectral width of 2,000 Hz resulting in a maximal evolution time of $t_2^{max} = 75$ ms. The number of recorded data points (spectral widths) for the other indirect dimensions are: 110 (10,000 Hz) for CACB, 100 (2,000 Hz) for N, and 90 (2,000 Hz) for CO, resulting in total data acquisition times of about 20 h per experiment.

Results and discussion

Amide ^{1}H-detected BEST-TROSY pulse schemes provide increased sensitivity and spectral resolution, while allowing for short overall experimental times, by exploiting three complementary effects: (i) The selective manipulation of amide ^{1}H by means of band-selective radio-frequency (RF) pulses enhances longitudinal ^{1}H relaxation (Pervushin et al. 2002; Schanda 2009), and thus provides higher sensitivity due to an increased ^{1}H steady-state polarization (BEST principle) (Schanda et al. 2006). (ii) Single-transition spin-state selection results in favorable relaxation properties for ^{1}H and ^{15}N due to CSA-dipolar cross-correlation (TROSY principle) (Pervushin et al. 1997). (iii) BEST-TROSY contains a build-in module for the conversion of (undetected) ^{1}H polarization that builds up during the pulse sequence due to spin relaxation, into enhanced ^{15}N polarization that contributes to the detected signal in the subsequent scan. The efficiency of this ^{15}N polarization enhancement mechanism depends on the longitudinal relaxation times (T_1) of both ^{1}H and ^{15}N, with the highest gains obtained for short ^{1}H T_1 and long ^{15}N T_1 (Favier and Brutscher 2011).

3D BEST-TROSY HNC correlation experiments

So far, BEST-TROSY (BT) techniques have been successfully applied to the study of globular proteins (Favier and Brutscher 2011) and nucleic acids (Farjon et al. 2009), but not yet to long, highly disordered polypeptides which is the subject of the present report. In order to illustrate the performance of BEST-TROSY for large IDPs we have chosen two proteins (or protein fragments), NS5A and BASP1 that are similar in size (~270 residues), but differ in their amino-acid-type composition and their structural compactness, resulting in different NMR properties (chemical shift dispersion, relaxation times, …) as shown in Figs. 1 and 5c. In particular, the structural dynamics in the BASP1 protein are quite uniform along the polypeptide chain, while NS5A shows a more heterogeneous behavior with several peptide regions that are characterized by increased local tumbling correlation times (τ_c) and higher local order (larger HET-NOE values). The 2D ^{1}H–^{15}N BEST-TROSY correlation spectra shown in Fig. 1 for NS5A (left) and BASP1 (right) show a large number of resolved resonances (especially for NS5A) despite the intrinsically low frequency dispersion. This high resolution can be maintained in 3D H–N–C correlation spectra, required for NMR assignment, by using semi-constant-time (semi-CT) editing (Grzesiek and Bax 1993) in the ^{15}N dimension allowing for long maximal evolution times. A series of such BT-optimized triple-resonance pulse sequences is displayed in Fig. 2. We used these experiments for the assignment of NS5A, the BASP1 protein had been assigned previously by other means. These experiments correlate the backbone amide ^{1}H and ^{15}N either with the ^{13}CA, ^{13}CB, ^{13}CO, or ^{15}N of the same (intra-residue correlation) or a sequentially adjacent residue (sequential correlation). A 2-step phase cycle is sufficient to obtain artifact-free correlation spectra. As discussed in more detail below, BT-optimization results in accelerated longitudinal relaxation of amide protons, yielding significantly reduced inter-scan delays required for optimal sensitivity, thus enhancing the overall sensitivity and reducing the total experimental time. As a consequence, even for large IDPs such as NS5A or BASP1 high-resolution 3D data sets can be recorded in only a few hours using a regular uniform sampling grid and Fourier-transform (FT) processing. Example strip extracted from NS5A spectra are shown in Fig. 3.

Particularly useful for sequential resonance assignment of large IDPs are the 3D hNcocaNH (Fig. 3a) and 3D hnCOcaNH (Fig. 3b) spectra that allow building sequential connectivities on the basis of backbone ^{15}N and ^{13}CO chemical shifts, characterized by a higher chemical shift dispersion than ^{13}CA or ^{13}CB in highly flexible proteins (Panchal et al. 2001, Kumar et al. 2010). Note that in these experiments the signal originating from ^{15}N polarization only enhances the "out-and-back" coherence transfer pathways resulting in the diagonal peaks.

BEST versus conventional techniques

In the following we will discuss the advantages of BEST-TROSY techniques (Fig. 2) with respect to alternative conventional pulse schemes. Longitudinal ^{1}H relaxation

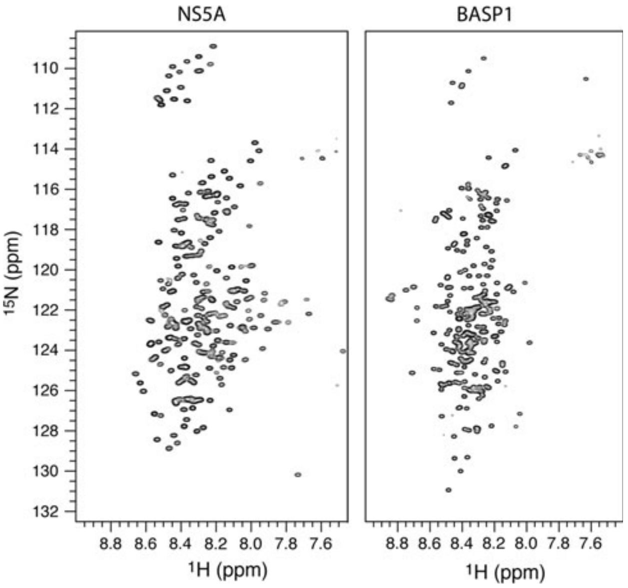
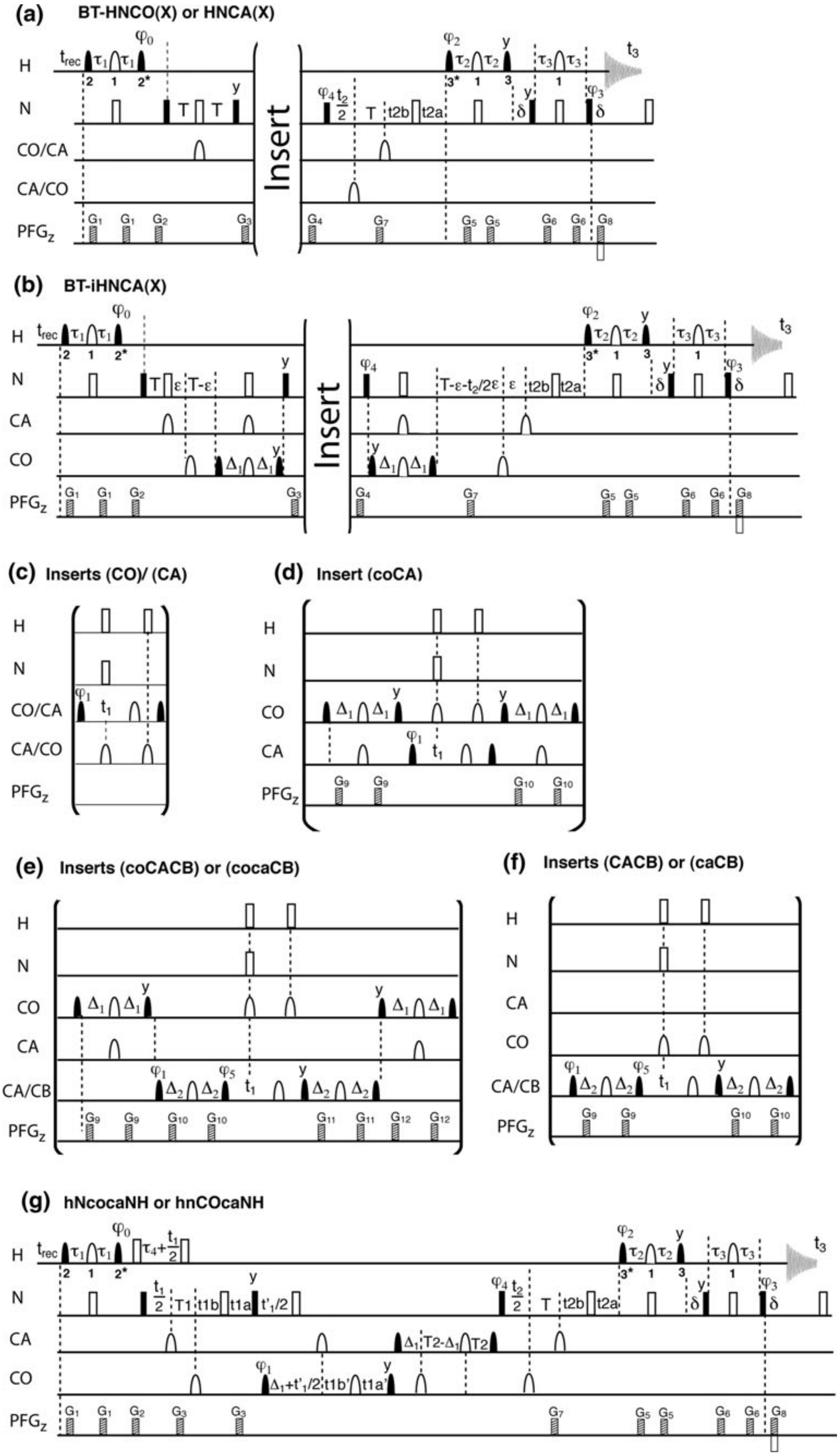


Fig. 1 2D ^{1}H–^{15}N BEST-TROSY correlation spectra of NS5A(191–447) (*left*) and BASP1 (*right*)


(a) BT-HNCO(X) or HNCA(X)
(b) BT-iHNCA(X)
(c) Inserts (CO)/ (CA)
(d) Insert (coCA)
(e) Inserts (coCACB) or (cocaCB)
(f) Inserts (CACB) or (caCB)
(g) hNcocaNH or hnCOcaNH
Insert
H
N
CO/CA
CA/CO
CA
CO
CA/CB
PFGz
trec
τ1
τ2
τ3
τ4
φ0
φ1
φ2
φ3
φ4
φ5
y
2
1
2*
3*
3
T
T-ε
ε
T-ε-t2/2ε
t2/2
t2b
t2a
t1
t1/2
T1
t1b
t1a
t'1/2
Δ1+t'1/2
t1b'
t1a'
Δ1
Δ2
T2-Δ1
T2
δ
t3
G1
G2
G3
G4
G5
G6
G7
G8
G9
G10
G11
G12

◀**Fig. 2** BEST-TROSY sequences: (a+c) HNCO or HNCA, (a+d) HNcoCA, (a+e) HNcoCACB or HNcocaCB, (a+f) HNCACB or HNcaCB, (b+c) iHNCA, (b+f) iHNCACB, (g) hNcocaNH and hnCOcaNH. Filled and open pulse symbols indicate 90° and 180° rf pulses. Unless indicated, all pulses are applied with phase x. All selective ^{1}H pulses are centered at 8.5 ppm, covering a bandwidth of 3.5 ppm, with the following shapes: [1] REBURP (Geen and Freeman 1991), [2] PC9 (Kupce and Freeman 1994) and [3] E-BURP2 (Geen and Freeman 1991). A star indicates a flip-back pulse obtained by time inversion of the excitation pulse shape. Open squares on ^{1}H indicate BIP-720-50-20 broadband inversion pulses (Smith et al. 2001). CO pulses have the shape of the center lobe of a sinx/x function, whereas CA and CA/CB pulses are applied with a rectangular shape and zero excitation at the CO frequency. The transfer delays common to all BEST-TROSY sequences are adjusted to $\tau_1 = 1\backslash(4J_{NH}) - 0.5\delta_1 - 0.5\delta_2$, $\tau_2 = 1\backslash(4J_{NH}) - 0.5\delta_1 - k\delta_3$, $\tau_3 = 1\backslash(4J_{NH})$ with $1\backslash(4J_{NH}) \approx 2.7$ ms, $\tau_4 = 1\backslash(2J_{NH})$, t2a = T-t$_2$/2 + t2b, and t2b = 0. The delays δ_1, δ_2, and δ_3 correspond to the lengths of the REBURP, PC9, and E-BURP2 pulses, respectively, and the parameter $k \approx 0.7$ can be fine-tuned to equilibrate the transfer amplitudes of the different coherence transfer pathways (Clean-TROSY) for optimal suppression of the unwanted quadruplet components in the spectrum (Schulte-Herbruggen and Sorensen 2000). For semi-CT ^{15}N editing, the delay t2b is incremented together with t$_2$ using the following time increment: Δt2b = (t$_2^{max}$/2-T)/N$_2$ with N$_2$ the number of total increments in the t$_2$ dimension. The additional delays are set to **a** T = 15 ms; **b** T = 17.5 ms; $\varepsilon = 0.5$ ms; $\Delta_1 = 4.5$ ms; **d** $\Delta_1 = 4.5$ ms; **e** and **f** $\Delta_1 = 4.5$ ms, $\Delta_2 = 3.5$ ms (for HNcoCACB) and $\Delta_2 = 7$ ms (for HNcocaCB); **g** T1 = 15 ms; T2 = 13.5 ms; $\Delta_1 = 4.5$ ms; t1a = T-t$_1$/2 + t1b; t1b = 0; t1a' = Δ_1-t'$_1$/2 + t1b'; t1b' = 0. For semi-CT ^{15}N (^{13}CO) editing, the delay t1b (t1b') is incremented together with t$_1$ (t$_1$') using the following time increment: Δt1b = (t$_1^{max}$/2-T)/N$_1$ (Δt1b = (t$_1^{max}$/2-Δ_1)/N$_1$) with N$_1$ the number of total increments in the t$_1$ dimension. Pulsed field gradients G$_1$–G$_8$ are applied along the z-axis (PFG$_z$) with durations of 200 μs to 2 ms and field strengths ranging from 5 to 40 G/cm. The 2-step phase cycling is: $\varphi_1 = -$x, x; $\varphi_2 = -$y $\varphi_3 = -$x; φ_4 = x, x; φ_5 = y; $\varphi_{rec} = -$x, x. The relative durations of G$_7$ and G$_8$ are given by the gyromagnetic ratios G$_7$/G$_8$ = γ_H/γ_N. Quadrature detection in t$_1$ is obtained by time-proportional phase incrementation of φ_1 (and φ_5) according to TPPI-States. For quadrature detection in t$_2$, echo-antiecho data are recorded by inverting the sign of gradient G$_8$ together with phases φ_2 and φ_3. All pulse sequences (in Agilent pulse program language) are available from the authors upon request

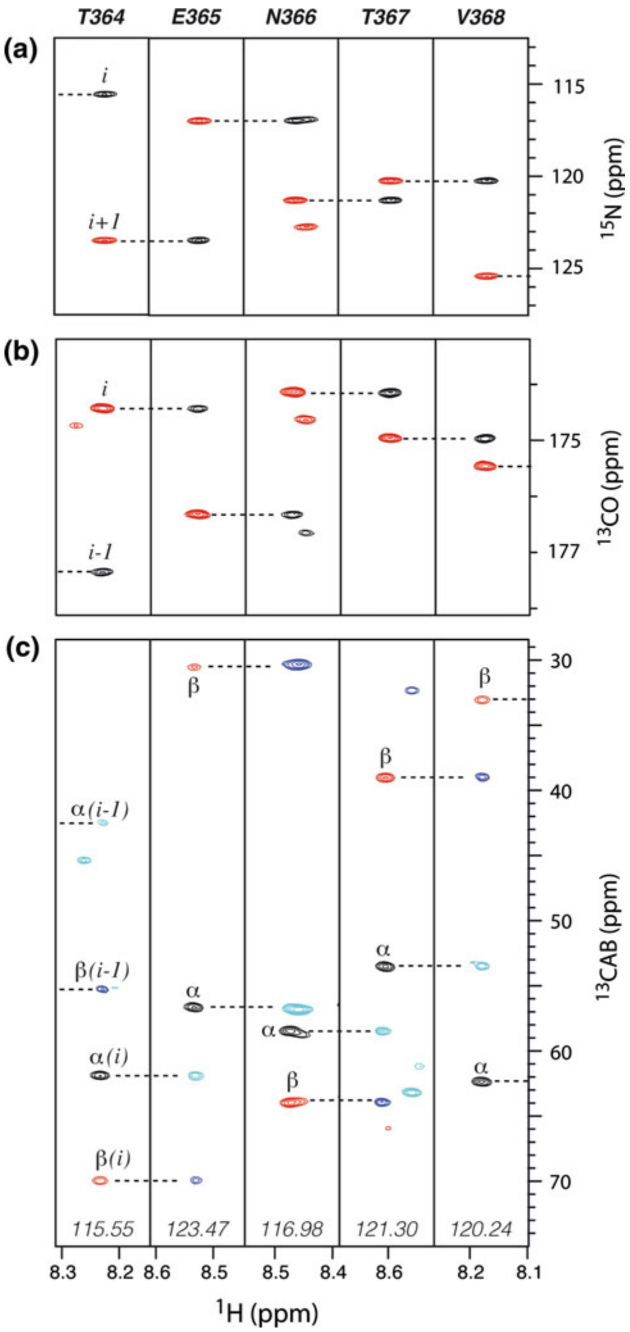


Fig. 3 Sequential resonance assignment of the NS5A segment 364-368 based on **a** 3D hNcocaNH, **b** 3D hnCOcaNH, and **c** 3D HNcoCACB and iHNCACB spectra. In **a** and **b** 2 correlation *peaks* of opposite sign (except if Gly residues are involved) are detected per amide group, one corresponding to the intra-residue, one to a sequential correlation. In **c** two spectra, each containing only intra-residue or sequential correlations, are superposed on the same graph

enhancement as achieved by BEST-type sequences exploits the fact that unperturbed proton spins that are coupled to the excited amide protons either by dipolar interactions with other aliphatic or aromatic protons, or chemical exchange with water protons take up some of the energy put into the spin system, thus accelerating spin–lattice (longitudinal) relaxation. While NS5A was studied at pH 6.5, for the BASP1 sample the pH was adjusted to two in order to investigate the relative contributions of ^{1}H–^{1}H dipolar interactions and amide-solvent ^{1}H exchange (rendered inefficient at low pH) on longitudinal relaxation enhancement. In order to quantify these effects for the two IDPs studied here, we have performed inversion-recovery experiments using three different inversion pulse schemes to simulate the proton relaxation behavior encountered in different types of experiments: (A) selective inversion of amide protons (BEST-type experiments), (B) inversion of all but water protons (water-flip-back (WFB) experiments), (C) inversion of the entire proton spectrum (conventional non-selective experiments). The inversion sequence is followed by a variable relaxation delay and a ^{1}H-^{15}N BEST-HSQC sequence for readout of the ^{1}H polarization (see Figure S1). The measured ^{1}H T_1 values are displayed in Fig. 4a as a function of the protein sequences of NS5A and BASP1, respectively. The average ^{1}H T_1 values measured for the two proteins are 0.92 ± 0.11 s (NS5A) and 0.91 ± 0.08 s (BASP1) for the non-selective inversion (black bars), 0.70 ± 0.15 s (NS5A) and 0.89 ± 0.07 s (BASP1) for the WFB situation (green bars), and 0.21 ± 0.06 s (NS5A) and 0.31 ± 0.06 s (BASP1) for the amide-selective (BEST-type) experiment (red bars). The measured relaxation times are quite uniform along the peptide sequence for the different experimental scenarios, with local differences mainly reflecting the variation in local structure (water accessibility and effective tumbling correlation time). The first interesting conclusion from these measurements is that for both proteins the non-selective ^{1}H T_1 of about 900 ms is reduced to 200–300 ms in BEST-type experiments. A second observation is that for both proteins the major relaxation enhancement mechanisms are dipolar interactions of the amide proton with surrounding protons, rather than chemical exchange with water protons. This conclusion, however, may be different for samples studied at higher pH and higher temperature. Finally, a third conclusion from these data is that longitudinal relaxation enhancement is more efficient for NS5A than for BASP1. This is mainly explained by the more compact conformers in the structural ensemble of NS5A resulting in more efficient proton–proton spin diffusion within the molecule, and some additional contribution from hydrogen-exchange mediated polarization transfer in NS5A.

The average ^{1}H T_1 values can be used to compute the expected signal-to-noise ratio per unit time (SNR_{UT}) as a function of the recycle delay (T_{rec}) between scans according to the analytical relation.

$$SNR_{UT} = (1 - \exp(-T_{rec}/T_1))/\sqrt{T_{scan}}. \quad (1)$$

Note that T_{rec} includes the data acquisition time, and that $T_{scan} = T_{rec} + T_{seq}$ with the sequence length arbitrarily set to $T_{seq} = 100$ ms. The resulting sensitivity curves for NS5A and BASP1, shown in Fig. 4b, indicate sensitivity gains for BEST-type optimized sequences of about a factor of 2 with the maximal sensitivity obtained for a recycle delay $T_{rec} \approx 1.25\ T_1$.

BEST-TROSY versus BEST-HSQC

Having shown in the previous paragraph that BEST sequences provide a sensitivity (and time) advantage with respect to non-selective or WFB ^{1}H-^{15}N correlation experiments, we now want to address the question how

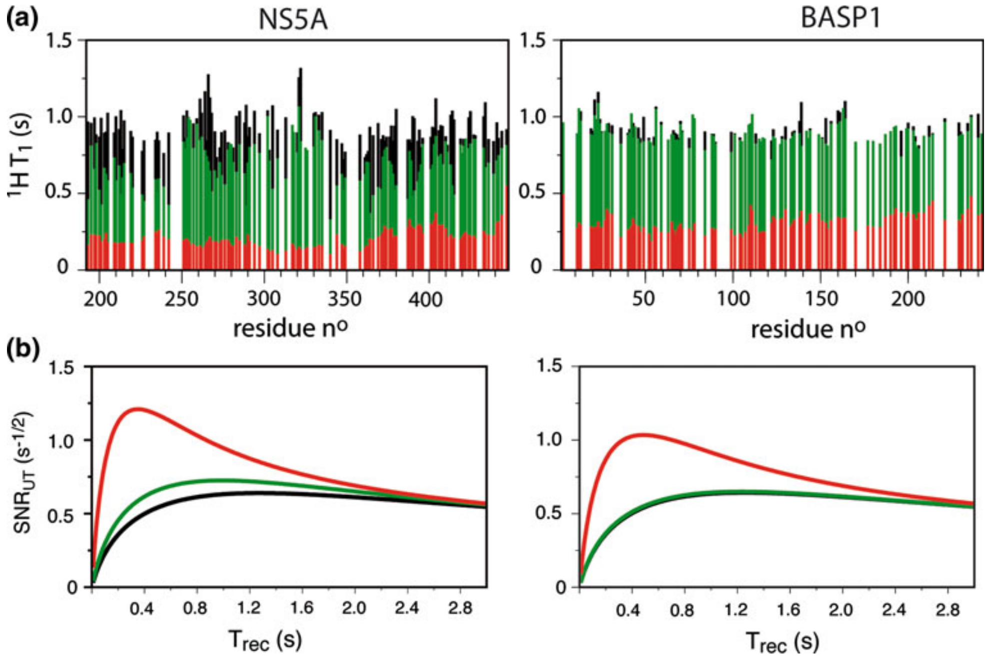


Fig. 4 Backbone amide ^{1}H T_1 relaxation time constants measured for NS5A (*left*) and BASP1 (*right*) are shown in **a** as function of the peptide sequence. The residue numbering corresponds to the real protein sequence, and the values measured for the tag-residues are not shown. Relaxation rates were measured by inversion-recovery using different inversion sequences (see Figure S1): non-selective ^{1}H inversion (*black bars*), water-flip-back ^{1}H inversion (*green bars*), and selective amide ^{1}H inversion (*red bars*). In **b** the expected average signal-to noise-ratio (SNR) per unit time (UT) has been computed from equation [1] as a function of the inter-scan delay (T_{rec}) for the three experimental scenarios: non-selective (*black*), water-flip-back (*green*), and amide-selective (*red*) BEST-type pulse sequences. The residue numbering corresponds to the real protein sequence; values measured for the tag-residues are not shown

BEST-TROSY (BT) compares to BEST-HSQC (BH) sequences. In order to do so, we have recorded 2D HNco planes of both proteins using either BEST-TROSY (Fig. 2a, c) or BEST-HSQC (Lescop et al. 2007) versions of the experiment. Except for the pulse sequence, all other acquisition and processing parameters were chosen identical. As can be appreciated from the spectra shown in Fig. 5a, the BEST-TROSY-optimized sequence yields higher spectral resolution due to the narrower single-transition line widths, and the signal contributed from ^{15}N polarization. Actually, as the ^{15}N labeling period t_2 becomes longer, more ^{1}H polarization builds up as a consequence of longitudinal spin relaxation during t_2, that is converted into enhanced ^{15}N polarization by the final coherence transfer step (Favier and Brutscher, 2011). Furthermore, also the peak intensities are increased in the BEST-TROSY spectra, on average by 80 % for NS5A and by 20 % for BASP1 despite an intrinsic factor-of-2 signal loss due to the single-transition selection. This signal enhancement reaches a factor of 3 (NS5A) and 1.6 (BASP1) for individual residues (Fig. 5b). Note however, that despite this overall increase in signal intensity, some residues in BASP1 have a reduced intensity in BEST-TROSY with respect to BEST-HSQC spectra. The differences observed for the two proteins are most likely ascribed to the shorter ^{1}H T_1 and longer ^{15}N T_1 values in NS5A, resulting in increased ^{15}N polarization enhancement, as well as more efficient CSA-dipolar cross-correlation (TROSY line narrowing effect) due to a more compact average structure in the N-terminal half of the NS5A protein (Feuerstein et al. 2012b). The highest intensity gains observed correlate well with regions of increased structural rigidity, as revealed by the larger τ_c and HETNOE values (Fig. 5c). Finally, BEST-TROSY also leads to a more uniform intensity distribution in the NMR spectra of these IDPs, as the weakest NMR signals become enhanced most (Fig. 5d).

Proline-selective ^{1}H–^{15}N BEST-TROSY correlation experiments

A specific feature of IDPs is that they often have repetitive sequence parts, and that they are rich in proline (Pro) residues (Tompa 2002). Because Pro residues are break points in the sequential assignment walk, this presents an additional complication for sequence-specific resonance assignment of IDPs. Therefore, NMR tools that allow identifying Pro-neighboring residues are of particular interest for IDP resonance assignment. Such information can be either obtained by 3D HNCAN-type pulse sequences (Lohr et al. 2000) that allow coherence transfer across Pro residues, or by Pro-edited 2D ^{1}H–^{15}N correlation experiments (Schubert et al. 2000) that exploit the particular spin-coupling topology of the Pro-side-chain to selectively detect coherence-transfer pathways from residues preceding or following Pro.

Here we present new pulse sequences for recording Pro-edited 2D ^{1}H–^{15}N correlation spectra that are of the BEST-TROSY type and therefore benefit from all the advantages discussed in the previous paragraphs. The Pro-HNcocan experiment selectively detects ^{1}H–^{15}N correlations of residues following a proline in the peptide sequence, while the Pro-iHNcan is selective to residues preceding a proline. The pulse sequences are obtained by inserting the pulse sequence element depicted in Fig. 6a into either the BT-HNCO(X) building block of Fig. 2a (Pro-HNcocan) or the BT-iHNCA(X) building block of Fig. 2b (Pro-iHNcan)—see also Figure S2 of the Supporting Information. The experiments exploit the fact that the ^{15}N chemical shift range of Pro in IDPs is well separated from all other amino-acid types (Fig. 6b). This makes it possible to selectively dephase C^α coherence during a constant-time delay T with respect to either the one-bond (Pro-HNcocan) or two-bond (Pro-iHNcan) J_{CAN} coupling, similar to a technique presented recently for the selective detection of phsophoserine and phosphothreonine residues (McIntosh et al. 2009). Two experiments need to be performed either with (*transfer experiment*) or without (*reference experiment*) application of the Pro-selective ^{15}N shaped pulse. In the transfer experiment the ^{13}C–^{15}N coupling evolution is active during the constant time delay T while in the reference experiment, coupling evolution is refocused. Subtraction of the 2 data sets (*difference spectrum*) then yields the desired Pro-selective ^{1}H–^{15}N correlation maps. The relative sensitivity of these experiments with respect to either HNcoca or iHNca experiments can be estimated by computing the transfer amplitude.

$$\mathrm{TA}_{pro} = 0.5(1 - \cos(\pi J_{NCA}\mathrm{T}))\cos(\pi J_{CC}\mathrm{T})\exp(-T/T_2), \tag{2}$$

with T_2 the transverse relaxation time constant of the C^α coherence present during the time T, $J_{cc} = \approx 35$ Hz the one-bond C^α–C^β coupling constant, and the factor 0.5 accounting for the sensitivity loss induced by difference spectroscopy. In Fig. 6c, the computed transfer amplitudes are plotted as a function of the delay T for the two experiments assuming relaxation time constants T_2 in the range from 30 to 100 ms as typical for highly flexible proteins. For both experiments, highest sensitivity (largest transfer amplitude) is obtained for T values of about 58 ms, with the overall sensitivity about a factor of 2 higher in the Pro-HNcocan with respect to Pro-iHNcan experiment which is explained by the larger one-bond ^{13}C-^{15}N coupling 12 Hz active in Pro-HNcocan with respect to the smaller two-bond coupling J_{NCA} ≈ 7 Hz that is active in Pro-iHNcan. Pro-HNcocan and Pro-iHNcan spectra of BASP1, recorded in 15 min and 30 min, respectively, are

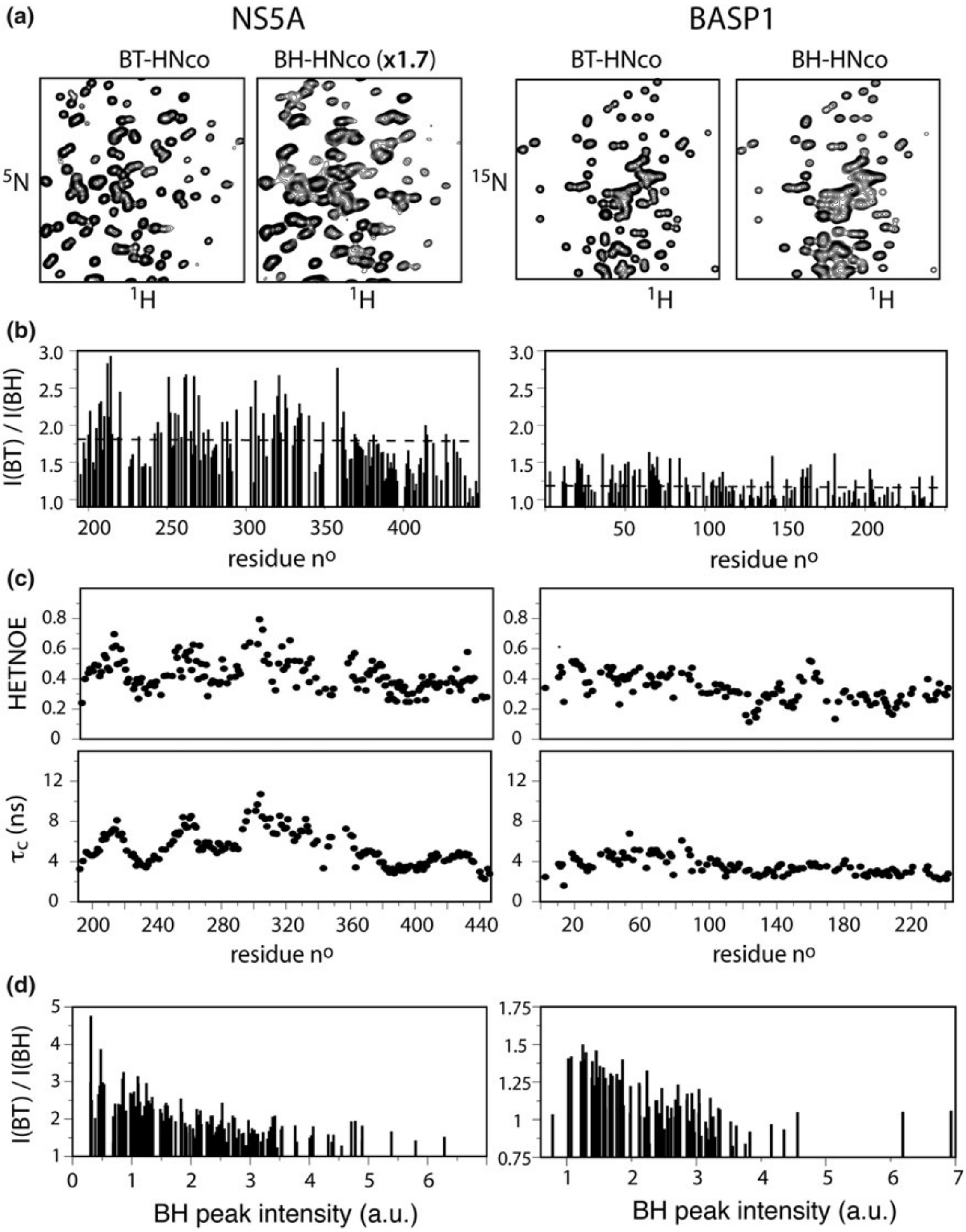


Fig. 5 Experimental comparison of the performance of BEST-HSQC (BH) and BEST-TROSY (BT)—type HNC correlation experiments for the two IDPs NS5A (*left*) and BASP1 (*right*). Extracts from 2D BT-HNco and BH-HNco spectra recorded with identical experimental parameters (except for the pulse sequence) are shown in **a**. For NS5A the BH-HNco spectrum was scaled by a factor 1.7 to obtain similar *peak* intensities for a better appreciation of the differences in spectral resolution in the two spectra. The *peak* intensity ratios (BT over BH) measured for the two proteins are plotted in **b** as a function of the peptide sequence. **c** ^{15}N relaxation data of NS5A(191–447) (*left*) and BASP1 (*right*). Local tumbling correlation times (τ_c) were obtained from the measured ^{15}N T_1 and T_2 relaxation time constants by computing the following expression $\tau_c \cong \sqrt{(6T_1/T_2 - 7)}\big/(4\pi\nu_N)$. **d** Same data as shown in **b**, but here the intensity gain (BT/BH) is plotted as a function of the peak intensity measured in the BEST-HSQC spectrum, in order to illustrate that highest signal enhancement is obtained for the weakest peaks leading to a more uniform spectral intensity distribution

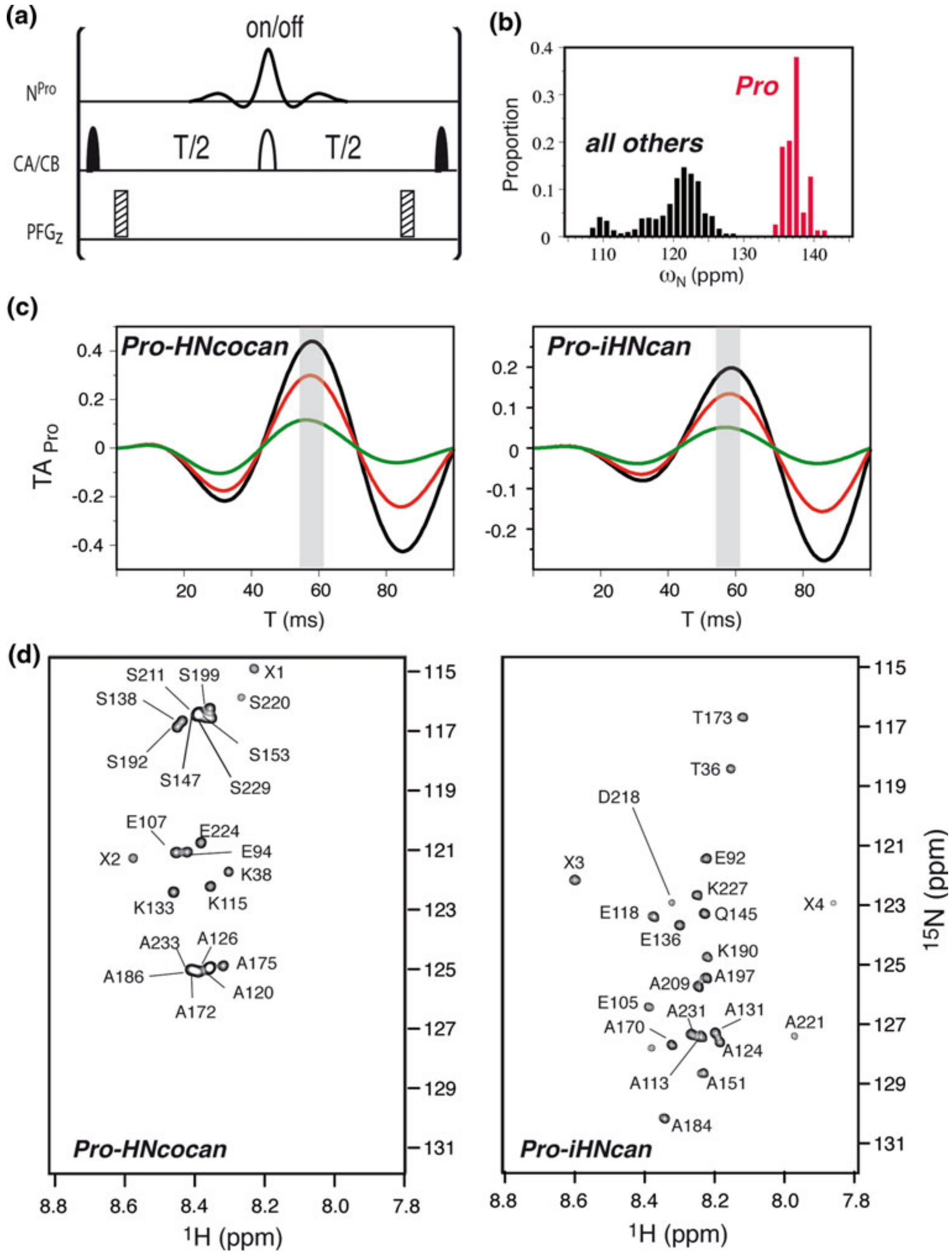


Fig. 6 Pro-selective ^{1}H–^{15}N spin-echo difference experiments. The pulse sequence element shown in **a** can be inserted into the BEST-TROSY sequence of Fig. 2a or 2b to obtain the pulse sequence of Pro-HNcocan or Pro-iHNcan, respectively. Pro-selective correlation spectra are obtained by subtracting two data sets recorded with and without the Pro ^{15}N refocusing pulse, centered at 138 ppm and covering a band width of 7 ppm (corresponding to a pulse length of 8.6 ms at 80 MHz). Use of a REBURP pulse shape (Geen and Freeman 1991) for Pro ^{15}N refocusing, makes sure that C$^{\alpha}$–N coupling evolution remains active during the entire delay T (Lescop et al. 2010), thus ensuring optimal sensitivity. **b** ^{15}N chemical shift statistics of IDPs obtained from 14 data sets (1,562 chemical shifts) deposited with the BMRB for non-Pro residues (*black*), and 6 data sets (79 chemical shifts) for Pro residues (*red*), including NS5A and BASP1. **c** Transfer amplitudes (TA_{Pro}) computed as a function of the filter delay (T) using equation [2] assuming different relaxation time constants: $T_2 = 100$ ms (*black*), 60 ms (*red*) and 30 ms (*green*). Highest transfer efficiency is obtained for a filter delay $T \cong 58$ ms. (d) Pro-HNcocan (*left*) and Pro-iHNcan (*right*) spectra recorded for BASP1 at 25 °C in 15 min and 30 min, respectively. All residues adjacent to prolines can be identified unambiguously from these spectra. Peaks are annotated by the residue number (and amino-acid type) of the detected amide ^{1}H–^{15}N correlations. An annotation 'X' refers to residues within the N-terminal His-tag extension of the BASP1 protein

displayed in Fig. 6d. All residues preceding or following Pro can be unambiguously identified from these spectra providing valuable starting points for the assignment of this 270-residue IDP. The corresponding spectra recorded for NS5A are shown in Figure S2 of the Supporting Information.

Recently we have introduced HADAMAC (Lescop et al. 2008), a sensitive NMR technique that allows distinguishing between seven different classes of amino-acid types from a set of amino-acid-type-edited 2D ^{1}H–^{15}N correlation spectra. HADAMAC is particularly useful for sequential resonance assignment of IDPs (Feuerstein et al. 2012a), but Pro residues cannot be distinguished from Arg, Glu, Lys, Gln, Met, and Leu residues, because they all fall within the so-called 'rest' class, and are detected in the same 2D spectrum. Thus, the Pro-edited ^{1}H–^{15}N BEST-TROSY correlation experiments introduced here, nicely complement HADAMAC for efficient amino-acid-type discrimination in IDPs.

In summary, we have proposed a set of sensitivity- and resolution-enhanced correlation experiments for sequential resonance assignment of highly flexible protein systems such as IDPs. We have demonstrated the performance of these BEST-TROSY experiments for two IDPs of ~270 residues in length. Short inter-scan delays of about 200 ms provide high sensitivity and allow to perform 3D data acquisition with long maximal evolution times in indirect dimensions (especially ^{15}N) in a total experimental time of only a few hours. If the resolution in 3D spectral space is not sufficient, the proposed experiments can be easily extended to 4D versions by additionally editing the chemical shifts of ^{13}CO that is involved in the coherence transfer pathways of all sequential and intra-residue correlation experiments. In addition, Pro-edited 2D ^{1}H-^{15}N correlation experiments have been introduced that provide complementary amino-acid-type information for resonance assignment of IDPs.

Acknowledgments We are grateful to Isabel Ayala and Adrien Favier for help in protein production and technical support. This work has been supported by grants from the European Commission (FP7-ITN IDPbyNMR contract No. 264257 and FP7-I3 BIO-NMR contract No. 261863), from the DFG (SFB974, A11), and from the Austrian Science Foundation FWF (W1221-B03 and P 20549-N19).

References

Bermel W, Bertini I, Felli IC, Gonnelli L, Kozminski W, Piai A, Pierattelli R, Stanek J (2012) Speeding up sequence specific assignment of IDPs. J Biomol NMR 53:293–301

Csizmok V, Felli IC, Tompa P, Banci L, Bertini I (2008) Structural and dynamic characterization of intrinsically disordered human securin by NMR spectroscopy. J Am Chem Soc 130:16873–16879

Davey NE, Trave G, Gibson TJ (2011) How viruses hijack cell regulation. Trends Biochem Sci 36:159–169

Dunker AK, Lawson JD, Brown CJ, Williams RM, Romero P, Oh JS, Oldfield CJ, Campen AM, Ratliff CR, Hipps KW, Ausio J, Nissen MS, Reeves R, Kang CH, Kissinger CR, Bailey RW, Griswold MD, Chiu M, Garner EC, Obradovic Z (2001) Intrinsically disordered protein. J Mol Graph Model 19:26–59

Farjon J, Boisbouvier J, Schanda P, Pardi A, Simorre JP, Brutscher B (2009) Longitudinal relaxation enhanced NMR experiments for the study of nucleic acids in solution. J Am Chem Soc 131:8571–8577

Farrow NA, Muhandiram R, Singer AU, Pascal SM, Kay CM, Gish G, Shoelson SE, Pawson T, Formankay JD, Kay LE (1994) Backbone dynamics of a free and a phosphopeptide-complexed Src homology-2 domain studied by 15 N NMR relaxation. Biochemistry 33:5984–6003

Favier A, Brutscher B (2011) Recovering lost magnetization: polarization enhancement in biomolecular NMR. J Biomol NMR 49:9–15

Felli IC, Brutscher B (2009) Recent advances in solution NMR: fast methods and heteronuclear direct detection. Chem Phys Chem 10:1356–1368

Feuerstein S, Plevin MJ, Willbold D, Brutscher B (2012a) iHADAMAC: a complementary tool for sequential resonance assignment of globular and highly disordered proteins. J Magn Reson 214:329–334

Feuerstein S, Solyom Z, Aladag A, Favier A, Schwarten M, Hoffmann S, Willbold D, Brutscher B (2012b) Transient structure and SH3 interaction sites in an intrinsically disordered fragment of the hepatitis C virus protein NS5A. J Mol Biol 420:310–323

Geen H, Freeman R (1991) Band-selective radiofrequency pulses. J Magn Reson 93:93–141

Grzesiek S, Bax A (1993) Amino-acid type determination in the sequential assignment procedure of uniformly 13C/15 N-enriched porteins. J Biomol NMR 3:185–204

Kumar D, Paul S, Hosur RV (2010) BEST-HNN and 2D-(HN)NH experiments for rapid backbone assignment in proteins. J Magn Reson 204:111–117

Kupce E, Freeman R (1994) Wide-band excitation with polychromatic pulses. J Magn Reson A 108:268–273

Lescop E, Schanda P, Brutscher B (2007) A set of BEST triple-resonance experiments for time-optimized protein resonance assignment. J Magn Reson 187:163–169

Lescop E, Rasia R, Brutscher B (2008) Hadamard amino-acid-type edited NMR experiment for fast protein resonance assignment. J Am Chem Soc 130:5014–5015

Lescop E, Kern T, Brutscher B (2010) Guidelines for the use of band-selective radiofrequency pulses in hetero-nuclear NMR: example of longitudinal-relaxation-enhanced BEST-type H-1-N-15 correlation experiments. J Magn Reson 203:190–198

Lohr F, Pfeiffer S, Lin YJ, Hartleib J, Klimmek O, Ruterjans H (2000) HNCAN pulse sequences for sequential backbone resonance assignment across proline residues in perdeuterated proteins. J Biomol NMR 18:337–346

Mantylahti S, Aitio O, Hellman M, Permi P (2010) HA-detected experiments for the backbone assignment of intrinsically disordered proteins. J Biomol NMR 47:171–181

Mcintosh LP, Kang HS, Okon M, Nelson ML, Graves BJ, Brutscher B (2009) Detection and assignment of phosphoserine and phosphothreonine residues by (13)C-(31)P spin-echo difference NMR spectroscopy. J Biomol NMR 43:31–37

Panchal SC, Bhavesh NS, Hosur RV (2001) Improved 3D triple resonance experiments, HNN and HN(C)N, for H-N and N-15 sequential correlations in (C-13, N-15) labeled proteins: application to unfolded proteins. J Biomol NMR 20:135–147

Pervushin K, Riek R, Wider G, Wüthrich K (1997) Attenuated T-2 relaxation by mutual cancellation of dipole–dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. Proc Natl Acad Sci USA 94:12366–12371

Pervushin K, Vögeli B, Eletsky A (2002) Longitudinal H-1 relaxation optimization in TROSY NMR spectroscopy. J Am Chem Soc 124:12898–12902

Schanda P (2009) Fast-pulsing longitudinal relaxation optimized techniques: enriching the toolbox of fast biomolecular NMR spectroscopy. Prog NMR Scpectrosc 55:238–265

Schanda P, Van Melckebeke H, Brutscher B (2006) Speeding up three-dimensional protein NMR experiments to a few minutes. J Am Chem Soc 128:9042–9043

Schubert M, Ball LJ, Oschkinat H, Schmieder P (2000) Bridging the gap: a set of selective H-1-N-15-correlations to link sequential neighbors of prolines. J Biomol NMR 17:331–335

Schulte-Herbruggen T, Sorensen OW (2000) Clean TROSY: compensation for relaxation-induced artifacts. J Magn Reson 144:123–128

Smith MA, Hu H, Shaka AJ (2001) Improved broadband inversion performance for NMR in liquids. J Magn Reson 151:269–283

Tompa P (2002) Intrinsically unstructured proteins. Trends Biochem Sci 27:527–533

Tompa P (2012) Intrinsically disordered proteins: a 10-year recap. Trends Biochem Sci 37:509–516

Uversky VN, Dunker AK (2010) Understanding protein non-folding. BBA-proteins proteom 1804:1231–1264

Wright PE, Dyson HJ (1999) Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. J Mol Biol 293:321–331

Xue B, Williams RW, Oldfield CJ, Goh GKM, Dunker AK, Uversky VN (2010) Viral disorder or disordered viruses: do viral proteins possess unique features? Protein Peptide Lett 17:932–951

NMR Spectroscopy of Proteins

DOI: 10.1002/anie.201304272

NMR Spectroscopic Studies of Intrinsically Disordered Proteins at Near-Physiological Conditions[**]

*Sergio Gil, Tomáš Hošek, Zsofia Solyom, Rainer Kümmerle, Bernhard Brutscher, Roberta Pierattelli, and Isabella C. Felli**

Intrinsically disordered proteins (IDPs) have recently attracted the attention of the scientific community because of their peculiar features that expand our view of how protein function is determined by the conformational properties of a polypeptide chain. The discovery of numerous physiological functions performed by IDPs has challenged the traditional structure–function paradigm.[1–4] The lack of a unique stable 3D structure and the high extent of local mobility provide functional advantages to IDPs in terms of structural plasticity and binding promiscuity.

Among the experimental techniques available to obtain atomic-resolution information on IDPs, solution-state NMR spectroscopy plays a predominant role. As a spectroscopic method, NMR is equally well applicable to structured as well as highly disordered proteins. However, the absence of a stable 3D structure and the highly dynamic nature of IDPs causes averaging of NMR chemical shifts, resulting in extensive cross-peak overlap in the NMR spectra.[5–9] Furthermore, fast hydrogen exchange of the solvent-exposed amide and other labile protons with water protons causes extensive line broadening or even complete disappearance of the corresponding resonances in the NMR spectra. Therefore, new NMR methods need to be derived or conventional techniques developed for folded proteins need to be adapted to account for the particular spectroscopic properties of IDPs.

To reduce the spectral overlap problem, uniform isotopic enrichment of the IDP with stable isotopes (^{13}C, ^{15}N) is mandatory. Indeed, ^{13}C and ^{15}N nuclei are characterized by an increased chemical shift dispersion with respect to protons. Therefore, proton-detected triple-resonance experiments[10,11] provide the spectral resolution required for site-resolved investigation of IDP structure, dynamics, and interaction modes.[12–14] Recently, the advent of more sensitive NMR instrumentation has brought carbon detection in the suitable sensitivity range for biomolecular NMR applications.[15–17] ^{13}C-Detected exclusively heteronuclear NMR experiments have been developed and offer valuable ways for the characterization of IDPs.[18–21]

The impact of hydrogen protein–solvent exchange processes on the NMR spectra of IDPs has never been discussed in detail, despite its importance, as most NMR experiments used to characterize IDPs are based on amide proton detection. Hydrogen exchange rates are highly dependent on the sample pH and temperature,[22] and changes in these two parameters can thus strongly influence the quality of the spectra. Inspecting the IDP literature reveals that most NMR studies have been performed at low temperatures and (slightly) acidic pH to reduce exchange-induced line broadening in the ^{1}H–^{15}N spectra, and in triple-resonance experiments that are based on amide ^{1}H detection.

The ensemble of conformers that an IDP adopts in solution and its ability to bind to molecular partners, may be strongly dependent on the environmental conditions. Therefore, it is important to have NMR methods to access atomic-resolution information under conditions that more closely approach physiological conditions (neutral pH, body temperature). In this context, ^{13}C detection provides a valuable method of investigation as ^{13}C nuclei are not only characterized by a good chemical shift dispersion even in absence of a stable 3D structure but they are also insensitive to hydrogen-exchange-induced line broadening.[23–27] With increasing temperature and pH, the quality of 2D ^{1}H–^{15}N correlation spectra deteriorates owing to the more efficient hydrogen-exchange mechanism, resulting in increasing number of peaks that are extensively line broadened, while at the same time the quality of the ^{13}C-detected ^{13}CO–^{15}N spectra is maintained, if not improved. This is illustrated in Figure 1, which shows ^{1}H–^{15}N and ^{13}CO–^{15}N spectra recorded in the temperature range of 285.7 K to 304.8 K (pH 7.4) for the IDP α-synuclein, a 140 residue human protein implicated in neurodegenerative diseases.[28] This comparison, which is a general behavior observed for IDPs, demonstrates that ^{13}C-start ^{13}C-detected experiments enable us to recover atomic resolution information that is not available from amide ^{1}H-detected experiments, thus opening new possibilities for the characterization of IDPs under close to physiological sample conditions.

---

[*] T. Hošek, Prof. R. Pierattelli, Prof. I. C. Felli
CERM and Department of Chemistry "Ugo Schiff"
University of Florence
Via Luigi Sacconi 6, 50019 Sesto Fiorentino, Florence (Italy)
E-mail: felli@cerm.unifi.it

Dr. S. Gil, Dr. R. Kümmerle
Bruker BioSpin AG
Industriestrasse 26, 8117 Fällanden (Switzerland)

Z. Solyom, Dr. B. Brutscher
Institut de Biologie Structurale, Université Grenoble 1, CNRS, CEA
Rue Jules Horowitz 41, 38027 Grenoble Cedex 1 (France)

[**] We thank Drs. Melanie Schwarten and Dieter Willbold (FZ Jülich (Germany)) for their contributions to the NS5A phosphorylation study, Leonardo Gonnelli (CERM) for the α-synuclein samples, and Dr. Klaus-Peter Neidig and Dr. Wolfgang Bermel (Bruker BioSpin GmbH) for stimulating discussions. This work has been supported in part by the EC Project BioNMR (Contract no 261863). S.G., T.H., and Z.S. are fellows of the IDPbyNMR Marie Curie action of the EC (Contract no 264257).

Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/anie.201304272.

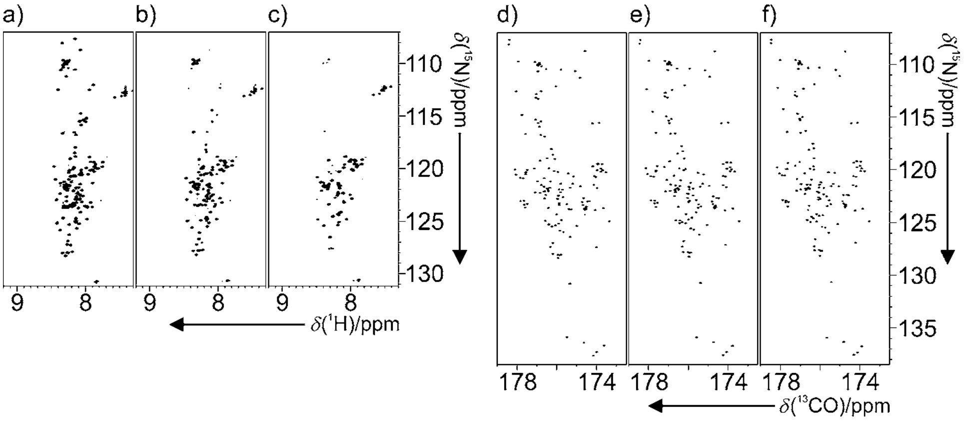


Figure 1. 2D spectra correlating the backbone amide nitrogen either with the directly bound amide proton or with the directly bound carbonyl. a)–c) ^{1}H–^{15}N HSQC and d)–f) ^{13}C–^{15}N CON acquired on α-synuclein at pH 7.4 are shown as a function of increasing temperature: a), d) 285.7 K; b), e) 295.5 K; c), f) 304.8 K. Each spectrum was acquired with one scan per increment, and the same spectral resolution (in Hz) was chosen for the two experiments. Similar behaviour is observed employing different variants of 2D ^{1}H–^{15}N correlation experiments (Supporting Information, Figure S1).

A major drawback of ^{13}C-detected experiments remains their low intrinsic sensitivity owing the approximately four times lower gyromagnetic ratio of ^{13}C with respect to ^{1}H ($\gamma_C \approx \gamma_H/4$). Therefore, it has been proposed to use the larger ^{1}H (instead of ^{13}C) polarization as a starting point of the coherence transfer pathways to increase the sensitivity of ^{13}C-detected experiments.[15,18] The sensitivity of these ^{1}H-start, ^{13}C-detected experiments can be further improved by longitudinal ^{1}H relaxation enhancement techniques.[29–34] These methods are based on the selective manipulation of a subset of proton spins that then relax back to thermodynamic equilibrium much faster owing to an energy transfer mechanism from the excited protons to the protons that remained unperturbed by the pulse sequence. The extent of longitudinal relaxation enhancement obtained by selective manipulation of amide ^{1}H spins in α-synuclein is shown in Figure 2. ^{1}H polarization inversion recovery has been measured for individual amide sites by inverting different pools of ^{1}H spins, followed by a relaxation delay, and a 2D $H^{N\text{-flip}}$CON readout sequence (Supporting Information, Figure S2) at pH 7.4 and 295.5 K (additional profiles at pH 6.4 are shown in the Supporting Information, Figure S3). Selective inversion of amide protons results in very fast ^{1}H recovery with effective T_1 time constants of about 60 ms (1.4 times the zero-crossing time point), instead of $T_1 \approx 2$ s measured for non-selective ^{1}H inversion. This circa 30-fold decrease in ^{1}H T_1 is impressive, and much higher than what is typically observed for folded proteins with reported selective amide ^{1}H T_1 of 200–400 ms compared to non-selective ^{1}H T_1 of 1.0 to 1.5 s.[35,36] Our inversion recovery data also demonstrate that the major source of ^{1}H relaxation enhancement under these experimental conditions can be attributed to hydrogen-exchange processes rather than ^{1}H–^{1}H dipolar interactions. The latter are responsible for the observed relaxation-enhancement effects in globular proteins,[30–32,34] as well as in IDPs studied at low temperature and acidic pH values.[14]

Inversion recovery curves also show that perturbation of the water resonance dramatically slows down recovery of amide protons to equilibrium. This has a strong impact on the performance of NMR spectra of IDPs, in particular when approaching physiological conditions. Therefore avoiding perturbation of water protons becomes a key general aspect to consider in the design of NMR spectra optimized for the study of IDPs.

The extremely short recovery times observed for amide ^{1}H in α-synuclein at close to physiological conditions prompted us to design new $^{1}H^{N}$-start ^{13}C-detected experiments with minimal perturbation of the water ^{1}H spins that are expected to yield significantly improved sensitivity. A particularly useful example is the 2D $H^{N\text{-BEST}}$CON experiment shown in Figure 3a that yields "fingerprint" spectra of the protein backbone that can be used for chemical shift mapping purposes. The acronym BEST[31] refers to the use of band-selective amide ^{1}H pulses to achieve longitudinal relaxation enhancement. The BEST technique was preferred over alternative solutions, for example, flip-back approaches, as it performs best with respect to the remaining amount of water ^{1}H polarization after one scan. Because of the long ^{1}H T_1 of water (ca. 3 s at

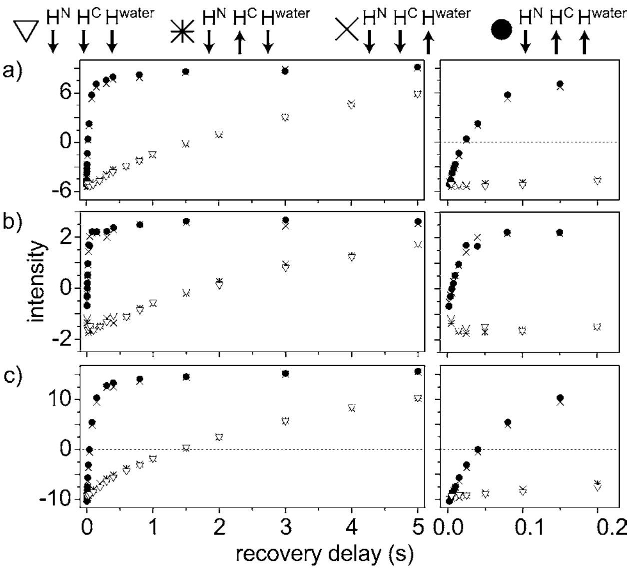


Figure 2. Inversion recovery profiles of amide protons acquired with the variant of the $H^{N\text{-flip}}$CON (described in the Supporting Information, Figure S2) for selected amino acids of α-synuclein: a) Val 40; b) His 50; c) Leu 100. Different initial conditions are indicated as follows: ▽ non-selective, ⋆ H^N–H^{water}-selective, × H^N–H^C-selective, ● H^N-selective.

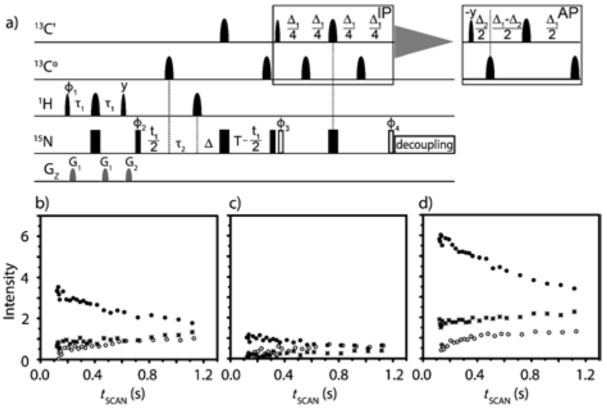


Figure 3. a) H$^{\text{N-BEST}}$CON pulse sequence. The delays are: $\tau_1 = 1/(4J_{\text{NH}}) - 0.5\delta_1 - 0.5\delta_2$, $\tau_2 = 1/(4J_{\text{NH}}) - 0.5\delta_2$, $\Delta_1 = 1/(2J_{\text{CON}})$, 33.2 ms, $\Delta_2 = 1/(2J_{\text{COC}\alpha})$, 9 ms, $\Delta = 1/(4J_{\text{CON}}) - 1/(4J_{\text{NH}}) - t_1/2$, $T = 1/(4J_{\text{CON}})$. The delays δ_1 and δ_2 correspond to the lengths of the PC9 and REBURP pulses that are used for 90° and 180° pulses, respectively. The phase cycle is $\phi_1 = x, -x$; $\phi_2 = 2(x), 2(-x)$; $\phi_3 = 4(x), 4(-x)$; $\phi_4 = 4(y), 4(-y)$; $\phi_{\text{rec}} = x, -x, x, -x$. Quadrature detection was obtained by incrementing phase ϕ_2 (t_1) in States-TPPI manner. For the ^{15}N virtual decoupling version, two 90° ^{15}N radio-frequency pulses are applied (white rectangles) instead of the one 90° ^{15}N rf pulse preceding IPAP block and ^{15}N decoupling during acquisition. b)–d) Intensity normalized accordingly to T_{scan} time, for CON (○), H$^{\text{N-start}}$CON (⋆), and H$^{\text{N-BEST}}$CON (●) for selected amino acids of α-synuclein, b) Val 40, c) His 50, and d) Leu 100 at pH 7.4 and 295.5 K.

room temperature),[37,38] even minimal perturbation of the water ^{1}H results in significantly reduced steady-state water ^{1}H polarization under fast-pulsing conditions, and thus in a reduction of the longitudinal enhancement effect required for optimal sensitivity.

In fact, ^{1}H recovery starts right after the last ^{1}H pulse, well before the end of the pulse sequence, in contrast to ^{1}H-detected experiments that typically end with ^{1}H pulses. Therefore, at the end of the H$^{\text{N-BEST}}$CON pulse sequence (including signal acquisition), sufficient ^{1}H polarization may already have built up to immediately start the next scan without any additional delay. To allow for long ^{13}C acquisition times to take benefit of the favorable transverse relaxation properties of IDPs, and avoid problems related to probe heating owing to ^{15}N decoupling during acquisition, we have implemented virtual ^{15}N decoupling in the pulse sequence.[18,39]

The performance of the new experiment in terms of sensitivity was evaluated by recording a series of 2D ^{13}CO–^{15}N correlation spectra of α-synuclein with different inter-scan delays, ranging from 0 to 1 s, using either H$^{\text{N-BEST}}$CON, the analogous experiment employing non-selective ^{1}H pulses, H$^{\text{N-start}}$CON, or the ^{13}C-start CON pulse schemes. The results obtained for selected residues of α-synuclein are shown in Figure 3b–d. The first observation from these data is that starting with ^{1}H as a polarization source in the H$^{\text{N-start}}$CON experiment does not result in a significant sensitivity increase with respect to the ^{13}C-start CON version under these experimental conditions. This is mainly explained by the long non-selective ^{1}H T_1 of about 2 s compared to the ^{13}CO T_1 of about 1 s (estimated value), and by additional ^{1}H relaxation occurring during the ^{1}H-start pulse sequence. It is interesting to note how under these conditions, amide ^{1}H polarization recovery for IDPs may become slower than ^{13}C recovery of carbonyl nuclei. Much larger signal gains are obtained for the H$^{\text{N-BEST}}$CON experiment that benefits from the extremely short selective recovery times of amide ^{1}H that are used as starting polarization. The observed sensitivity gains (H$^{\text{N-BEST}}$CON over ^{13}C-start CON), determined for short recycle times of 150 ms, vary between different sites from a factor of 4 to a factor of 12 (Supporting Information, Figure S4). The example curves shown in Figure 3b–d have been chosen to be representative of residues characterized by solvent exchange rates of different magnitude. For residues characterized by (relatively) slow exchange under these conditions (for example, Leu 100), a higher sensitivity increase is observed than for residues with faster exchanging amide protons (for example, His 50). This demonstrates that starting from amide ^{1}H introduces again a dependency on the solvent exchange rates, and that there is an optimal exchange regime for the use of such BEST-type ^{1}H-start ^{13}C-detected experiments. Maximal sensitivity is achieved if the exchange rate is large with respect to the longitudinal relaxation rate R_1, but still small compared to the transverse relaxation rate R_2 ($1\,\text{s}^{-1} < k_{\text{ex}} < 10^2\,\text{s}^{-1}$). In the presence of even faster hydrogen exchange, experimental variants exploiting ^{13}C,[23,25] or non-exchangeable ^{1}H,[18] as the starting polarization source are expected to yield higher sensitivity.

The H$^{\text{N-BEST}}$CON experiment is particularly useful to study the behavior of IDPs inside living cells (in-cell NMR) or to monitor post-translational modifications occurring either in vitro or in vivo, because to be biologically relevant, all of these studies are preferably performed under sample conditions that are close to physiological pH and temperature.

In-cell NMR spectroscopy provides a unique spectroscopic method to investigate the fate of a protein in a cellular context. In particular, it allows changes induced by different cellular stimuli to be monitored or protein spectra obtained for different cell types to be compared.[40–44] A number of techniques, for example bacterial over-expression, injection, electroporation, or direct expression of proteins in human cells, have been developed during recent years that allow to introduce isotopic labeled protein into various cell types.[40–44] The extensive line broadening that is typical for in-cell spectra, in combination with the drastically reduced chemical shift dispersion and high solvent exchange rates of IDPs, has a strong negative impact on the amount of residue-specific information that can be obtained through 2D ^{1}H–^{15}N correlation experiments. ^{13}C-detected experiments offer a valuable solution to this problem thanks to the higher chemical shift dispersion and reduced sensitivity to exchange broadening.[25,45] However, high sensitivity is required to ensure short experimental times to cope with the limited lifetime of meaningful in-cell NMR samples. The use of ^{1}H as a starting polarization source as well as longitudinal relaxation enhancement are important features for the study of IDPs in cells through ^{13}C-detected exclusively heteronuclear NMR experiments.[46] An example of an in-cell H$^{\text{N-BEST}}$CON spectrum of α-synuclein over-expressed in *E. coli* cells is shown in Figure 4. The inset clearly shows how the H$^{\text{N-BEST}}$CON experiment enables to detect correlations that were lost

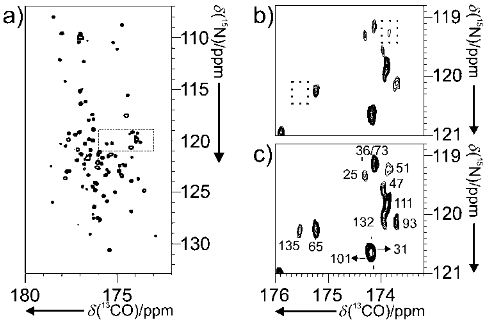


Figure 4. a) 2D ^{13}C–^{15}N H$^{N\text{-BEST}}$CON spectrum of α-synuclein in *E. coli* cells acquired in 20 min. The region of the spectrum enclosed by a rectangular box is enlarged to highlight how c) H$^{N\text{–BEST}}$CON enables a better sensitivity to be obtained compared to b) H$^{N\text{–flip}}$CON and to detect correlations that otherwise would not have been detected in the same experimental time.

under the spectral noise in the corresponding H$^{N\text{–flip}}$ implementation of the experiment.

Another interesting application of NMR spectroscopy at nearly physiological conditions consists in following post-translational modifications modulating protein function. Among them, recent studies have focused on monitoring phosphorylation events involving intrinsically disordered proteins.[47–49] Also in this case, ^{13}C-detected NMR experiments complement well the ^{1}H-detected experiments for obtaining site-resolved information on phosoporylation events in intrinsically disordered proteins (for an example, see the Supporting Information, Figure S5).

In conclusion, we have shown here that ^{13}C-start ^{13}C-detected experiments do not suffer from fast hydrogen exchange between amide and solvent protons in IDP samples studied at close to physiological conditions, thus enabling us to recover information that would be difficult or even impossible to obtain through amide ^{1}H-detected experiments. Furthermore, in favourable cases the fast hydrogen exchange rates can even be turned into a spectroscopic advantage. By combining longitudinal ^{1}H relaxation optimized BEST-type techniques with ^{13}C-direct detection pulse schemes, important sensitivity improvements can be achieved, and experimental times can be significantly reduced. This opens up new applications for monitoring chemical shift changes in IDPs upon interaction to a binding partner, chemical modification, or by changing the environment, under sample conditions that were inaccessible by conventional techniques. This concept can be easily extended to other ^{1}H^{N}-start ^{13}C-detected experiments. Furthermore, ^{13}C-detected NMR experiments can be used to obtain semi-quantitative information about hydrogen-exchange rates without directly detecting the exchangeable amide ^{1}H by quantifying the signal enhancement observed in H$^{N\text{–BEST}}$CON with respect to H$^{N\text{-start}}$CON spectra.[18,33] This may become useful to monitor changes in the exchange rates after for example, some binding event, reflecting a change in the solvent accessibility at the binding site of the IDP.

Received: May 17, 2013
Revised: July 30, 2013

Keywords: intrinsically disordered proteins · longitudinal relaxation enhancement · NMR spectroscopy · α-synuclein

---

[1] P. E. Wright, H. J. Dyson, *J. Mol. Biol.* **1999**, *293*, 321 – 331.
[2] P. Tompa, M. Fuxreiter, *Trends Biochem. Sci.* **2008**, *33*, 2 – 8.
[3] P. Tompa, *Structure and function of intrinsically disordered proteins*, Taylor and Francis Group, Boca Raton, FL, **2009**.
[4] V. N. Uversky, A. K. Dunker, *Biochim. Biophys. Acta Proteins Proteomics* **2010**, *1804*, 1231 – 1264.
[5] D. Neri, G. Wider, K. Wüthrich, *Proc. Natl. Acad. Sci. USA* **1989**, *89*, 4397 – 4401.
[6] H. J. Dyson, P. E. Wright, *Methods Enzymol.* **2001**, *339*, 258 – 271.
[7] H. Zhang, S. Neal, D. S. Wishart, *J. Biomol. NMR* **2003**, *25*, 173 – 195.
[8] D. Eliezer, *Curr. Opin. Struct. Biol.* **2009**, *19*, 23 – 30.
[9] J. Marsh, S. A. Teichmann, J. D. Forman-Kay, *Curr. Opin. Struct. Biol.* **2012**, *22*, 643 – 650.
[10] M. Sattler, J. Schleucher, C. Griesinger, *Prog. Nucl. Magn. Reson. Spectrosc.* **1999**, *34*, 93 – 158.
[11] A. Bax, S. Grzesiek, *Acc. Chem. Res.* **1993**, *26*, 131 – 138.
[12] M. D. Mukrasch, S. Bibow, J. Korukottu, S. Jeganathan, J. Biernat, C. Griesinger, E. Mendelkow, M. Zweckstetter, *PLoS Biol.* **2009**, *7*, e34.
[13] V. Motáčková, J. Nováček, A. Zawadzka-Kazimierczuk, K. Kazimierczuk, L. Žídek, H. Šanderová, L. Krásný, W. Koźmiński, V. Sklenář, *J. Biomol. NMR* **2010**, *48*, 169 – 177.
[14] Z. Solyom, M. Schwarten, L. Geist, R. Konrat, D. Willbold, B. Brutscher, *J. Biomol. NMR* **2013**, *55*, 311 – 321.
[15] Z. Serber, C. Richter, D. Moskau, J.-M. Boehlen, T. Gerfin, D. Marek, M. Haeberli, L. Baselgia, F. Laukien, A. S. Stern, J. C. Hoch, V. Dötsch, *J. Am. Chem. Soc.* **2000**, *122*, 3554 – 3555.
[16] I. Bertini, L. Duma, I. C. Felli, M. Fey, C. Luchinat, R. Pierattelli, P. R. Vasos, *Angew. Chem.* **2004**, *116*, 2307 – 2309; *Angew. Chem. Int. Ed.* **2004**, *43*, 2257 – 2259.
[17] H. Kovacs, D. Moskau, M. Spraul, *Prog. Nucl. Magn. Reson. Spectrosc.* **2005**, *46*, 131 – 155.
[18] W. Bermel, I. Bertini, V. Csizmok, I. C. Felli, R. Pierattelli, P. Tompa, *J. Magn. Reson.* **2009**, *198*, 275 – 281.
[19] W. Bermel, I. Bertini, I. C. Felli, R. Peruzzini, R. Pierattelli, *ChemPhysChem* **2010**, *11*, 689 – 695.
[20] J. Nováček, A. Zawadzka-Kazimierczuk, V. Papoušková, L. Žídek, H. Šanderová, L. Krásny, W. Koźmiński, V. Sklenář, *J. Biomol. NMR* **2011**, *50*, 1 – 11.
[21] W. Bermel, I. Bertini, L. Gonnelli, I. C. Felli, W. Kozminski, A. Piai, R. Pierattelli, J. Stanek, *J. Biomol. NMR* **2012**, *53*, 293 – 301.
[22] Y. W. Bai, J. S. Milne, L. Mayne, S. W. Englander, *Proteins Struct. Funct. Genet.* **1993**, *17*, 75 – 86.
[23] W. Bermel, I. Bertini, I. C. Felli, R. Kümmerle, R. Pierattelli, *J. Magn. Reson.* **2006**, *178*, 56 – 64.
[24] I. Bertini, I. C. Felli, L. Gonnelli, R. Pierattelli, Z. Spyranti, G. A. Spyroulias, *J. Biomol. NMR* **2006**, *36*, 111 – 122.
[25] W. Bermel, I. Bertini, I. C. Felli, Y.-M. Lee, C. Luchinat, R. Pierattelli, *J. Am. Chem. Soc.* **2006**, *128*, 3918 – 3919.
[26] S. T. Hsu, C. W. Bertoncini, C. M. Dobson, *J. Am. Chem. Soc.* **2009**, *131*, 7222 – 7223.
[27] L. Skora, S. Becker, M. Zweckstetter, *J. Am. Chem. Soc.* **2010**, *132*, 9223 – 9225.
[28] P. H. Weinreb, W. G. Zhen, A. W. Poon, K. A. Conway, P. T. Lansbury, Jr., *Biochemistry* **1996**, *35*, 13709 – 13715.
[29] K. Pervushin, B. Vogeli, A. Eletsky, *J. Am. Chem. Soc.* **2002**, *124*, 12898 – 12902.
[30] P. Schanda, B. Brutscher, *J. Am. Chem. Soc.* **2005**, *127*, 8014 – 8015.

[31] P. Schanda, H. Van Melckebeke, B. Brutscher, *J. Am. Chem. Soc.* **2006**, *128*, 9042–9043.
[32] E. Lescop, P. Schanda, B. Brutscher, *J. Magn. Reson.* **2007**, *187*, 163–169.
[33] W. Bermel, I. Bertini, I. C. Felli, R. Pierattelli, *J. Am. Chem. Soc.* **2009**, *131*, 15339–15345.
[34] A. Favier, B. Brutscher, *J. Biomol. NMR* **2011**, *49*, 9–15.
[35] B. Brutscher, P. Schanda in *Encyclopedia of NMR* (Eds.: M. Grant, R. K. Harris), J. Wiley and Sons, Chichester, **2009**.
[36] P. Schanda, *Prog. Nucl. Magn. Reson. Spectrosc.* **2009**, *55*, 238–265.
[37] H. Y. Carr, E. M. Purcell, *Phys. Rev.* **1954**, *94*, 630–638.
[38] H. J. Simpson, H. Y. Carr, *Phys. Rev.* **1958**, *111*, 1201–1202.
[39] T. Kern, P. Schanda, B. Brutscher, *J. Magn. Reson.* **2008**, *190*, 333–338.
[40] S. Reckel, F. Lohr, V. Dotsch, *ChemBioChem* **2005**, *6*, 1601–1606.
[41] K. Inomata, A. Ohno, H. Tochio, S. Isogai, T. Tenno, I. Nakase, T. Takeuchi, S. Futaki, Y. Ito, H. Hirokai, M. Shirakawa, *Nature* **2009**, *458*, 106–109.
[42] S. Liokatis, A. Stuetzer, S. J. Elsaesser, F. X. Theillet, R. Klingberg, B. van Rossum, D. Schwarzer, C. D. Allis, W. Fischle, P. Selenko, *Nat. Struct. Mol. Biol.* **2012**, *19*, 819–823.
[43] F. X. Theillet, C. Smet-Nocca, S. Liokatis, R. Thongwichian, J. Kosten, M. K. Yoon, R. W. Kriwacki, I. Landrieu, G. Lippens, P. Selenko, *J. Biomol. NMR* **2012**, *54*, 217–236.
[44] L. Banci, L. Barbieri, I. Bertini, E. Luchinat, E. Secci, Y. Zhao, A. R. Aricescu, *Nat. Chem. Biol.* **2013**, *9*, 297–299.
[45] A. Binolfi, F. X. Theillet, P. Selenko, *Biochem. Soc. Trans.* **2012**, *40*, 950–954.
[46] I. Bertini, I. C. Felli, L. Gonnelli, V. M. V. Kumar, R. Pierattelli, *Angew. Chem.* **2011**, *123*, 2387–2389; *Angew. Chem. Int. Ed.* **2011**, *50*, 2339–2341.
[47] G. Lippens, L. Amniai, M. R. Wieckowski, A. Sillen, A. Leroy, I. Landrieu, *Biochem. Soc. Trans.* **2012**, *40*, 698–703.
[48] F. X. Theillet, H. M. Rose, S. Liokatis, A. Binolfi, R. Thongwichian, M. Stuiver, P. Selenko, *Nat. Protoc.* **2013**, *8*, 1416–1432.
[49] I. Amata, M. Maffei, A. Igea, M. Gay, M. Vilaseca, A. R. Nebreda, M. Pons, *ChemBioChem* **2013**, DOI: 10.1002/cbic.201300139.

J Biomol NMR
DOI 10.1007/s10858-014-9857-8

ARTICLE

Measuring hydrogen exchange in proteins by selective water saturation in ^{1}H–^{15}N SOFAST/BEST-type experiments: advantages and limitations

Enrico Rennella · Zsofia Solyom · Bernhard Brutscher

Received: 25 June 2014 / Accepted: 22 August 2014

Abstract HET^{ex}-SOFAST NMR (Schanda et al. in J Biomol NMR 33:199–211, 2006) has been proposed some years ago as a fast and sensitive method for semi-quantitative measurement of site-specific amide-water hydrogen exchange effects along the backbone of proteins. Here we extend this concept to BEST readout sequences that provide a better resolution at the expense of some loss in sensitivity. We discuss the theoretical background and implementation of the experiment, and demonstrate its performance for an intrinsically disordered protein, 2 well folded globular proteins, and a transiently populated folding intermediate state. We also provide a critical evaluation of the level of accuracy that can be obtained when extracting quantitative exchange rates from HET^{ex} NMR measurements.

Keywords BEST · SOFAST · Fast NMR · Hydrogen exchange · Protein · Folding · Unfolding · Compactness · α-synuclein

---

Electronic supplementary material The online version of this article (doi:10.1007/s10858-014-9857-8) contains supplementary material, which is available to authorized users.

E. Rennella · Z. Solyom · B. Brutscher (✉)
Institut de Biologie Structurale, Université Grenoble 1,
71 avenue des Martyrs, 38044 Grenoble Cedex 9, France
e-mail: bernhard.brutscher@ibs.fr

E. Rennella · Z. Solyom · B. Brutscher
Commissariat à l'Energie Atomique et aux Energies Alternatives (CEA), Grenoble, France

E. Rennella · Z. Solyom · B. Brutscher
Centre National de Recherche Scientifique (CNRS), Grenoble, France

Introduction

Hydrogen exchange measurements of amide protons with solvent protons provide useful information on local structural and dynamic properties of proteins, e.g. solvent accessibility, hydrogen bond strength, local and global unfolding events. Such measurements also allow mapping protein–ligand and protein–protein interactions via a change in the apparent solvent accessibility of amide protons at the protein surface. NMR spectroscopy is a well-established technique to measure hydrogen exchange rates in proteins in a site-resolved manner (Dempsey 2001; Englander and Kallenbach 1983). In the most widely used experimental setup, the protonated protein is quickly dissolved in D_2O, and the disappearance of amide proton NMR signals as a consequence of H/D exchange processes is followed in real-time by recording a series of ^{1}H–^{15}N correlation spectra (Wagner and Wüthrich 1982). Using fast acquisition techniques, such H/D exchange methods are suitable for measuring exchange rates up to $k_{ex} < 0.1\ s^{-1}$ (Gal et al. 2007; Schanda et al. 2007). To detect and quantify faster exchange processes, a variety of conceptually similar magnetization transfer techniques have been proposed (Fan et al. 2011; Gemmecker et al. 1993; Grzesiek and Bax 1993; Hwang et al. 1998; Mori et al. 1994; Wider et al. 1996). The basic idea of these methods is to selectively excite (or frequency-label) the water proton spins, and then observe the transfer of magnetization to amide protons, by repeating the experiment for increasing transfer times. Magnetization transfer techniques provide access to exchange rates in the range $1\ s^{-1} < k_{ex} < 100\ s^{-1}$. The upper limit of quantification of exchange rates is determined by the signal-to-noise ratio of the corresponding exchange line-broadened NMR signals, while the lower limit is imposed by longitudinal relaxation.

Even faster exchange, resulting in unobservable proton resonances, can be quantified by measuring the exchange-induced modulation of ^{15}N-^{1}H scalar couplings (Segawa et al. 2008). The CLEANEX-PM transfer sequence (Hwang et al. 1998) is arguably the current gold standard for measuring fast exchange rates of NMR-visible amide protons in globular proteins, as it suppresses to a large extent direct and exchange-relayed NOE contributions that otherwise introduce systematic errors in the measured exchange rates. A different method of chemical exchange spectroscopy is SOLEXSY (Chevelkov et al. 2010) where magnetization transfer from a protonated to deuterated ^{15}N, and vice versa is measured. While SOLEXSY is not affected by additional magnetization transfer pathways, this method suffers from relatively low intrinsic experimental sensitivity. A few years ago, we introduced yet another technique, called HET$^{\mathrm{ex}}$-SOFAST NMR (Schanda et al. 2006), that provides a semi-quantitative measure of fast hydrogen exchange rates along the polypeptide chain of proteins. In this experiment, the effect of saturating the water protons on the longitudinal relaxation rates (R_1) of amide protons during the recycle (inter-scan) delay is exploited as a measure of the hydrogen exchange efficiency. The main advantages of HET$^{\mathrm{ex}}$-SOFAST with respect to other magnetization transfer techniques are high experimental sensitivity and short overall acquisition times. In the present manuscript, we discuss the experimental and theoretical background of HET$^{\mathrm{ex}}$-SOFAST, and demonstrate that the same idea can be easily extended to BEST-type experiments (HET$^{\mathrm{ex}}$-BEST). In addition, we evaluate experimentally the range of exchange rates accessible to this technique, and the level of accuracy that can be obtained from their quantification.

Theoretical background

If we assume for simplicity a spin system consisting of a labile (exchange-prone) amide proton (A) from the protein under investigation, and a second proton (W) from the bulk water, the time evolution of spin polarization is described by the following set of 1st order differential Bloch-McConnell equations:

$$\frac{d}{dt}\begin{pmatrix} W_z(t) \\ A_z(t) \end{pmatrix} = -\begin{pmatrix} R_1^W & 0 \\ 0 & R_1^A \end{pmatrix}\begin{pmatrix} W_z(t) - W_z^{eq} \\ A_z(t) - A_z^{eq} \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ k_{ex} & -k_{ex} \end{pmatrix}\begin{pmatrix} W_z(t) \\ A_z(t) \end{pmatrix} \tag{1}$$

with R_1^W and R_1^A the longitudinal relaxation rates of water and amide protons, and k_{ex} the hydrogen exchange rate. Equation (1) assumes that the bulk water polarization is not significantly changed by hydrogen exchange with the protein. We now consider two different scenarios that are relevant for the HET$^{\mathrm{ex}}$-SOFAST/BEST experiment described below: (1) in the *reference experiment*, the amide polarization at time $t = 0$ equals $A_z^{ref}(0) = 0$, and the water polarization is close to equilibrium ($W_z^{ref}(t) \approx W_z^{eq}$) throughout the experiment. (2) In the saturation experiment, the amide polarization at time $t = 0$ also equals $A_z^{sat}(0) = 0$, and the water polarization $W_z^{sat}(t) = 0$ for all relaxation times. Under these conditions, solving Eq. (1) yields the following solutions for the evolution of amide proton polarization A_z^{ref} and A_z^{sat} as a function of a relaxation delay d_{relax}:

$$A_z^{ref}(d_{relax}) = A_z^{eq}\frac{R_1^A + k_{ex}W_z^{ref}(d_{relax})}{R_1^A + k_{ex}}\left[1 - \exp\{-(R_1^A + k_{ex})d_{relax}\}\right] \tag{2}$$

$$A_z^{sat}(d_{relax}) = A_z^{eq}\frac{R_1^A}{R_1^A + k_{ex}}\left[1 - \exp\{-(R_1^A + k_{ex})d_{relax}\}\right] \tag{3}$$

In order to determine reliable exchange rates k_{ex} (as well as R_1^A and A_z^{eq}) from Eqs. (2) and (3), measurements of $A_z^{ref}(d_{relax})$, $A_z^{ref}(d_{relax})$, and $W^{ref}(d_{relax})$ need to be performed at a minimum of 2 different relaxation times d_{relax}. Alternatively, if such data are only available for a single relaxation time, we still can compute the following expression, yielding a measure of k_{ex}/R_1^A:

$$HET^{ex}ratio = \frac{\left(A_z^{ref}(d_{relax})/A_z^{sat}(d_{relax}) - 1\right)}{W_z^{ref}(d_{relax})} = \frac{k_{ex}}{R_1^A} \tag{4}$$

In a real protein, dipolar interactions with other protons close in space that exchange (or not) with water protons will offer additional relaxation pathways via NOE and exchange-relayed NOE effects that result in more complex spin dynamics that are not accounted for by a simple fit of the measured data to Eqs. (2) and (3).

The HET$^{\mathrm{ex}}$-SOFAST/BEST experiment

In order to evaluate the effect of hydrogen exchange on the apparent longitudinal relaxation rates of amide protons along the protein backbone, 2 complementary 2D ^{1}H–^{15}N correlation spectra need to be recorded: (1) a reference spectrum, and (2) a water-saturated spectrum. For the reference spectrum, an experiment is required that yields minimal perturbation of water proton polarization, in order to allow for fast repetition of the pulse sequence. This is achieved by SOFAST-HMQC (Schanda and Brutscher 2005; Schanda et al. 2005), BEST-HSQC (Lescop et al. 2007) or BEST-TROSY (Favier and Brutscher 2011) pulse schemes that affect only little the water ^{1}H polarization,

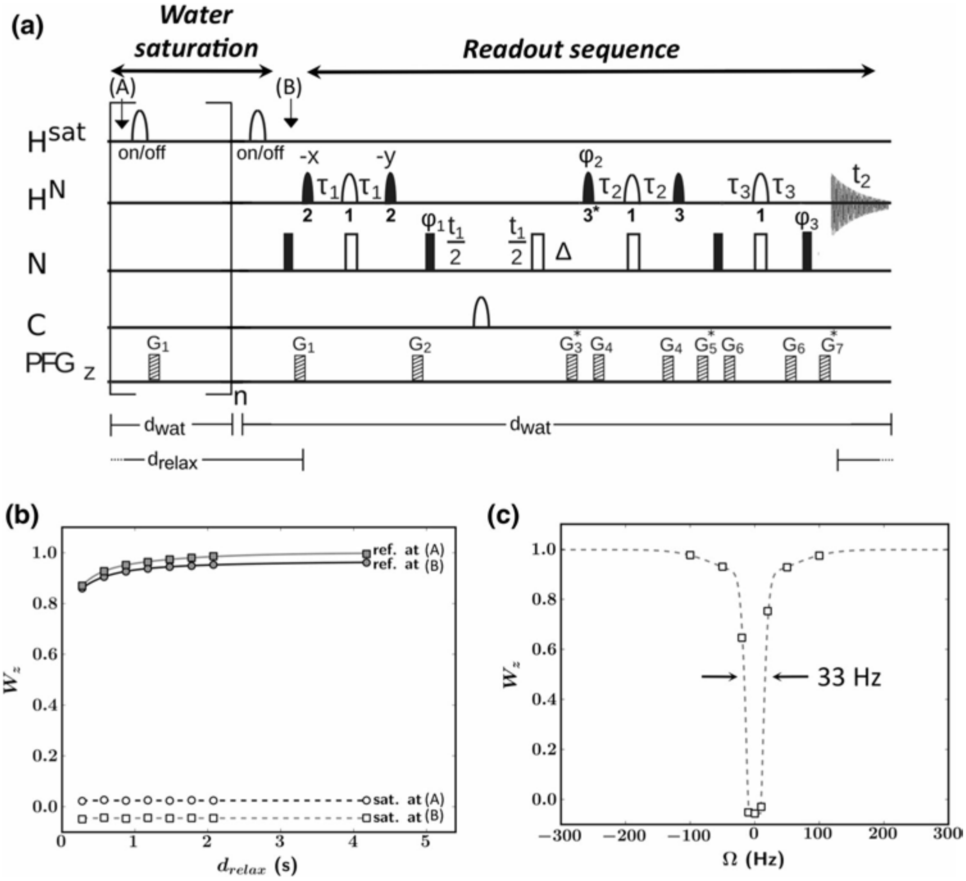


Fig. 1 **a** Pulse sequence of HET^{ex}-BEST-TROSY experiment (the Bruker code is provided in the Supporting Information). The pulse sequence is basically a BEST-TROSY experiment (Favier and Brutscher 2011) to read-out the effect of water saturation on the amide ^{1}H relaxation. Note that only the ^{1}H pathway is detected, while the ^{15}N pathway is removed (by a 90° ^{15}N pulse followed by a pulsed field gradient). For hydrogen exchange measurements, 2D ^{1}H–^{15}N spectra are recorded with (saturation spectrum) and without (reference spectrum) the water saturation scheme applied during the relaxation delay d_{relax}. Water saturation is achieved by a series of on-resonance ISNOB-5 pulses (pulse duration of 150 ms at 600 MHz) followed by a pulsed field gradient to suppress radiation damping. The repetition time d_{wat} is typically set to the duration of the BEST-TROSY readout sequence (including the acquisition period and the length of a single ISNOB-5 pulse). Filled and open pulse symbols indicate 90° and 180° pulses, respectively. Unless indicated, all pulses are applied with phase x. The following pulse shapes are used for H^{N}: [1] REBURP (Geen and Freeman 1991), [2] PC9 (Kupce and Freeman 1994), and [3] E-BURP2 (Geen and Freeman 1991). A *star* indicates a flip-back pulse obtained by time inversion of the excitation pulse shape. H^{N} pulses are typically centred at 8.5 ppm, covering a bandwidth of 4.0 ppm. The transfer delays are adjusted to $\tau_1 = 1/(4J_{NH}) - 0.5\delta_1 - 0.5\ \delta_2$, $\tau_2 = 0.8/(4J_{NH}) - 0.5\ \delta_3 - 0.5\ \delta_1$, $\tau_3 = 1/(4J_{NH}) - 0.5\ \delta_1$, with $1/(4J_{NH}) \approx 2.7$ ms. The delays δ_1, δ_2 and δ_3 correspond to the ^{1}H pulse lengths of the REBURP, PC9, and E-BURP2, respectively. A 2-step phase cycle is used: $\varphi_1 = x, -x$; $\varphi_2 = y$; $\varphi_3 = y$ For echo/antiecho-type quadrature detection in t_1, 2 data sets are recorded with the relative gradient strengths of G3, G5 and G7 set to (−8: 2: 3.013) and (−7: 3: 1.987), respectively, together with a 180° phase shift of φ_2 and φ_3. **b** Normalized water polarization measured in the reference and saturation experiments as a function of the relaxation delay d_{relax}. For these measurements a small-angle (∼1°) read-out pulse, preceded by a strong pulsed field gradient pulse, is inserted at time points (**a**) or (**b**) in the pulse sequence. **c** Normalized water polarization measured at time point (**b**) of the pulse sequence as a function of the offset of the H^{sat} pulses from the water resonance frequency

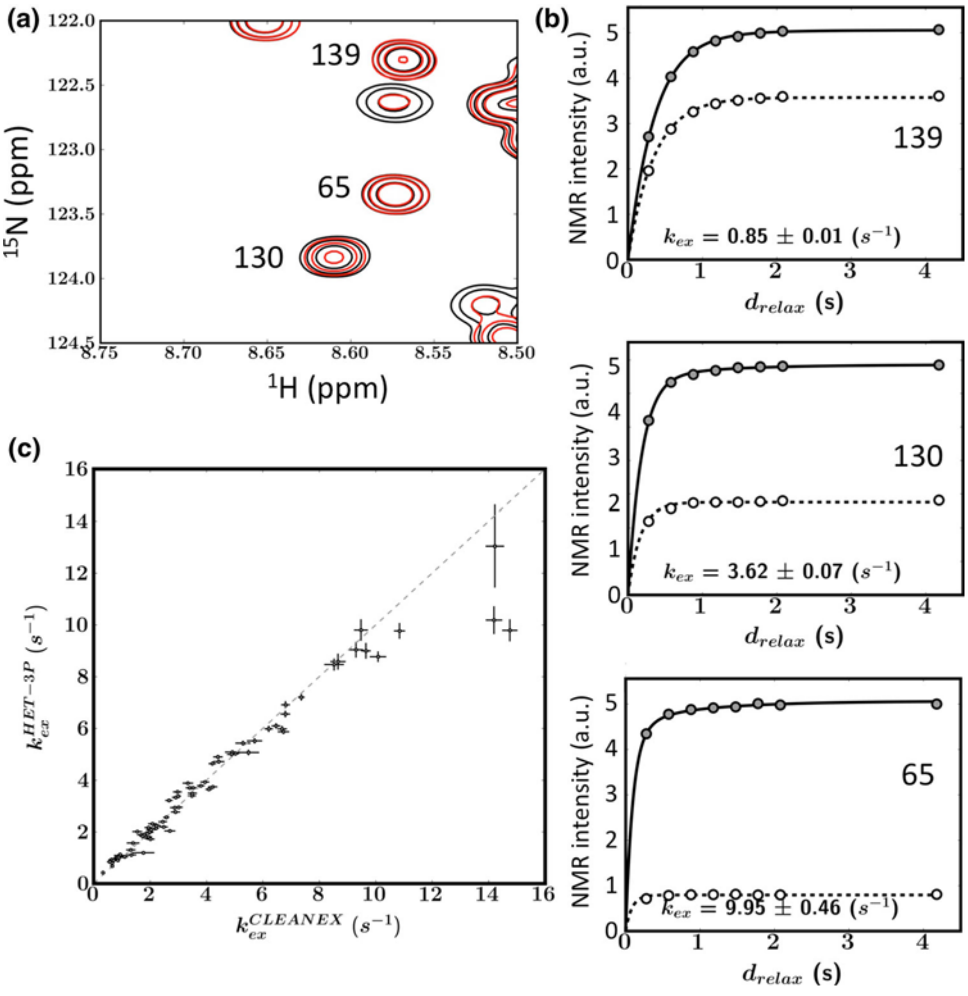
even under fast-pulsing conditions. Note that even a slight saturation (5–10 %) of the water polarization by the pulse sequence has a dramatic affect on the steady-state water polarization due to the long water T_1 of several seconds ($R_1^w = 0.3\ s^{-1}$ at 25 °C). For the water-saturation experiment, the same ^{1}H–^{15}N spectrum is recorded but with an additional train of regularly spaced selective ^{1}H 180° pulses applied at the water ^{1}H resonance during the recycle delay between subsequent scans. The pulse sequence for HET^{ex}-BEST-TROSY is shown in Fig. 1a. The BEST-TROSY sequence has been chosen here, because it yields highest spectral resolution, especially at high magnetic fields, at the expense of somewhat reduced sensitivity with respect to SOFAST-HMQC (Favier and Brutscher 2011). In order to measure the effect of the pulse sequence on the water polarization, we have added a small-angle (∼1°) read-out pulse at positions (a) or (b) in the pulse sequence. The result of such measurements, shown in Fig. 1b, indicates that the water ^{1}H polarization stays at more than 80 % of its equilibrium value, even for relaxation delays as short as 200 ms.

A critical point for this experiment is the choice of the selective ^{1}H pulses to achieve good water saturation over a narrow bandwidth, independent of the relaxation delay. In addition, if a high-Q cryogenically cooled probe is used for the experiment, the effective B_1 field needs to be stronger than the B_1 field induced by radiation damping (Krishnan and Murali 2013; Shishmarev and Otting 2011) in order to achieve proper inversion of the water ^{1}H polarization by the pulse. As an example, a Gaussian pulse shape of 100 ms length at 600 MHz ^{1}H frequency, as used in our original HET^{ex}-SOFAST experiment (Schanda et al. 2006) does not result in more than 50 % saturation of the water on a high-Q probe under steady-state conditions. Significantly shorter pulse lengths (<40 ms) with higher B_1 fields are required to achieve efficient water saturation, that will have a strong negative impact on the selectivity of the saturation scheme. We therefore have tested different pulse shapes and pulse lengths, and we obtained good results with an ISNOB-5 pulse (Kupce et al. 1995) of about 150 ms duration. The saturation scheme consists in a repetitive series of a water-selective 180° pulse followed by a relaxation delay (equal to the duration of the readout pulse sequence). The shortest possible relaxation delay therefore corresponds to the length of the selective pulse. After about 40 repetitions of this saturation scheme (dummy scans), steady-state conditions are reached with a residual water ^{1}H polarization close to zero (Fig. 1b). The experimentally determined offset-dependence of water saturation is shown in Fig. 1c, indicating that only protons resonating in close proximity of the water ^{1}H (± 20 Hz) are significantly affected by this pulse scheme.

Application to an intrinsically disordered protein

A first application of HET^{ex}-BEST-TROSY to the intrinsically disordered 140-residue protein (IDP) α-synuclein (pH 7.4, 5 °C) is shown in Fig. 2. For IDPs, hydrogen-exchange measurements may be useful probes to identify

Fig. 2 Hydrogen exchange measurements for amide protons in α-synuclein. **a** Small spectral region of the ^{1}H–^{15}N HET^{ex}-BEST-TROSY reference (*black*) and saturation (*red*) spectra recorded for $d_{relax} = 270$ ms. **b** Intensity *buildup curves* measured for different residues, annotated in (**a**), in a series of reference spectra (*straight lines*) and saturation spectra (*dashed lines*), recorded for different relaxation delays. **c** Correlation plot of exchange rates k_{ex} obtained from a 3-point data fit to Eqs. (2) and (3) and from CLEANEX-PM measurements. Data are shown for all α-synuclein residues for which a resolved ^{1}H–^{15}N correlation peak was detected in the NMR spectra. *Error bars* were obtained from Monte Carlo simulations based on the spectral noise level



partially structured peptide segments, and to evaluate changes in solvent accessibility under different sample conditions, or upon binding to a molecular partner (Croke et al. 2008; Crowhurst et al. 2002; Csizmok et al. 2008). HET^{ex}-BEST-TROSY (reference and saturation) spectra have been measured for relaxation delays d_{relax} in the range 200 ms–4.2 s. The relaxation curves for 3 representative residues (Fig. 2a) are plotted in Fig. 2b. These data nicely illustrate the differences observed in the relaxation behavior of amide protons along the backbone of α-synuclein under water-saturating conditions. The fast amide ^{1}H relaxation observed for fast exchanging amide sites can also be exploited to enhance the signal intensity of the corresponding correlation peaks in SOFAST- and BEST-type experiments (Gil et al. 2013; Yao et al. 2011). In order to extract quantitative exchange rates, these relaxation curves were fitted to Eqs. (2) and (3), using either all measured time points, or a minimal data set (3 points) consisting in 2 reference spectra "together with measurement of the water polarization $W^{ref}(d_{relax})$", measured at 2 relaxation times (d_{relax} = 200 and 500 ms), and one saturation spectrum obtained for a single relaxation delay of d_{relax} = 200 ms. As both fits yield very similar results (see figure S1 of the Supporting Information), only exchange rates obtained from the 3-point fits will be discussed in the following.

In order to evaluate the accuracy of the exchange rates extracted from the HET^{ex}-BEST-TROSY data, we also have performed CLEANEX-PM measurements (Hwang et al. 1998) on the same sample. Details on the experimental setup are provided in figure S2 of the Supporting Information. As shown in Fig. 2c, the CLEANEX-PM "reference" exchange rates correlate well with the rates extracted from HET^{ex}-BEST-TROSY in the range $0.1\ s^{-1} < k_{ex} < 10\ s^{-1}$. Higher exchange rates cannot be quantified accurately from HET^{ex}-BEST-TROSY data, as the relaxation curves under water-saturation conditions become flat, in other words the plateau peak intensity has been reached within the shortest relaxation time (d_{relax} = 200 ms). However, it is worth noting that the CLEANEX-PM sequence has also a limited accuracy in this regime, as for peptide segments that undergo fast time-scale local motions the conditions underlying the suppression scheme of NOE and ROE effects are no longer satisfied.

In conclusion of this part, HET^{ex}-BEST-TROSY yields accurate exchange rate measurements for highly dynamic proteins or protein segments over a range of about 2 orders of magnitude ($0.1\ s^{-1} < k_{ex} < 10\ s^{-1}$), by measuring a minimal data set of 3 ^{1}H–^{15}N correlation spectra, and fitting the measured data to the analytical solutions of the Bloch-McConnell equations for an isolated 2-spin system as given by Eqs. (2) and (3). The main advantage of this approach is that the overall data acquisition time for HET^{ex}-BEST-TROSY was only 40 min, while CLEANEX-PM required 24 h of data collection on the same NMR spectrometer.

Application to globular proteins

The same strategy for evaluating the performance of HET^{ex}-BEST-TROSY, as outlined above for α-synuclein, has been applied to 2 small well-folded proteins: (1) the 76-residue *human* ubiquitin (pH 6.4, 20 °C), and (2) the 70-residue *E. coli* heat-shock protein CspA (pH 7.7, 15 °C). Correlation plots of exchange rates extracted from HET^{ex}-BEST-TROSY and CLEANEX-PM data are shown in Fig. 3a (ubiquitin), d (CspA). The majority of amide protons in these globular proteins are involved in hydrogen-bonded secondary structural elements, and thus exhibits relatively slow solvent exchange with k_{ex} rates of $< 2\ s^{-1}$. Comparison with the rates obtained from CLEANEX-PM shows that there is almost no correlation for amides in the structured parts of the proteins, while the correlation is quite good for the few residues in mobile loop regions that exhibit faster exchange, $k_{ex} > 2\ s^{-1}$. The lack of correlation in the structured protein regions is explained by the fact that dipolar (NOE) interactions between the amide proton and nearby labile (O–H, N–H, N–H$_2$) protons, or H$^\alpha$ protons resonating close to the water frequency, contribute up to 2 s^{-1} to the apparent exchange rates in HET^{ex}-BEST-TROSY, while they are suppressed to a large extent in CLEANEX-PM. A particularly striking example is residue S27 of CspA that shows the largest discrepancy between the HET^{ex}-BEST-TROSY and CLEANEX-PM data. A zoom into the structural region surrounding S27 (see figure S3 of the Supporting Information) shows that the amide proton is only at a distance of 2.2 Å from its own side-chain O–H. The apparent $k_{ex} = 1.7\ s^{-1}$ extracted from HET^{ex}-BEST-TROSY is thus due to an exchange-relayed NOE effect with this side chain proton, and not to a direct hydrogen exchange of the amide proton with water. If the exchange process is in the so-called EX2 regime, where the folding (closing) rate is much faster than the intrinsic chemical exchange rate, the accuracy of the measured exchange rates can be increased by repeating the HET^{ex}-BEST-TROSY experiment for different pH values, as proposed by Bax and coworkers (Fitzkee et al. 2011). However, due to the limited range of exchange rates accessible by this technique, only small pH variations should be used in practice.

As an alternative to the extraction of quantitative exchange rates from HET^{ex}-BEST-TROSY data measured for multiple relaxation delays, we can compute the HET^{ex}ratio, introduced in Eq. 4, from a pair of HET^{ex}-BEST-TROSY spectra recorded for a single relaxation

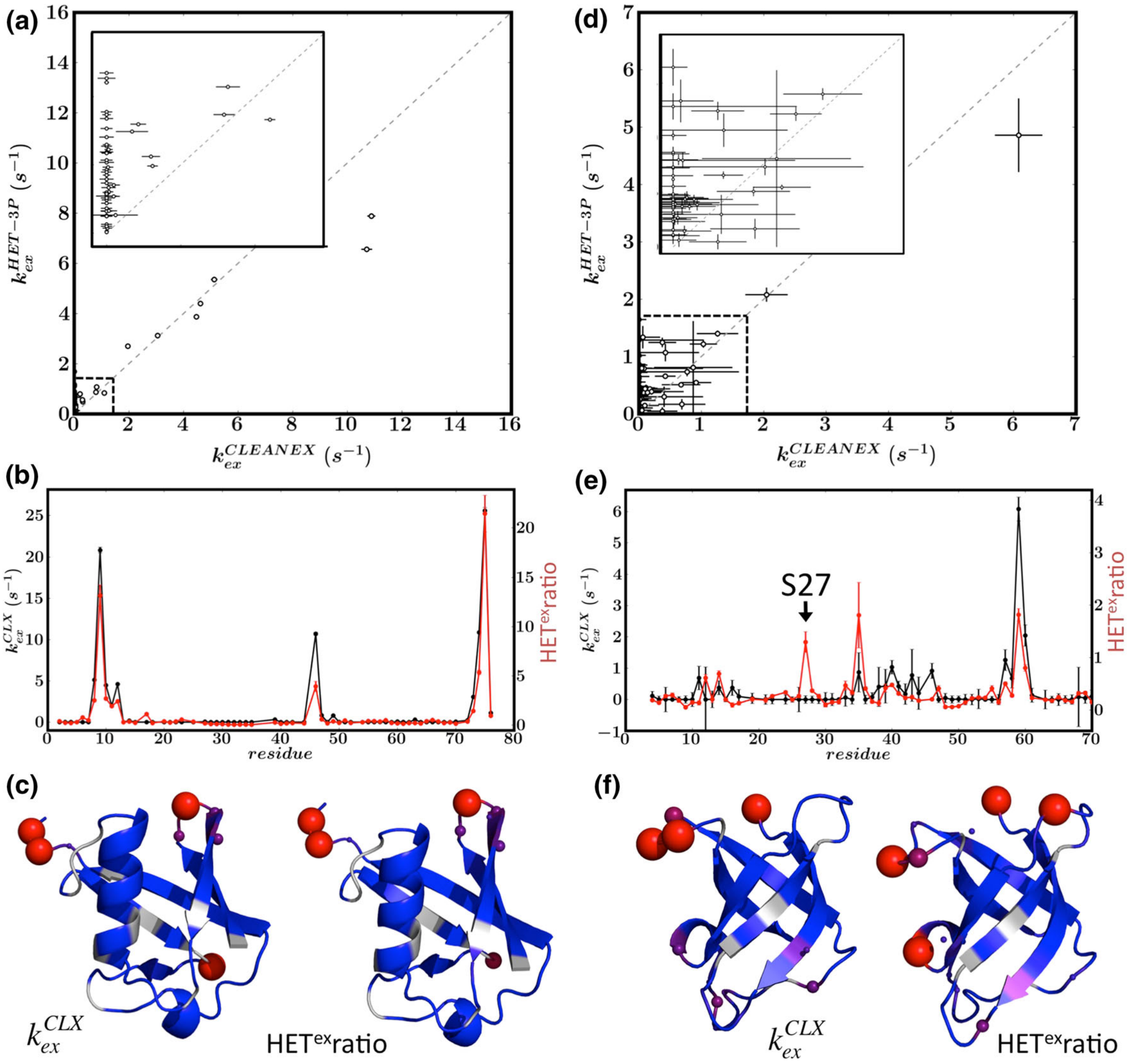


Fig. 3 Hydrogen exchange measurements for ubiquitin (*left panel*) and CspA (*right panel*). **a** and **d** Correlation of exchange rates k_{ex} extracted from HET^{ex}-BEST-TROSY and CLEANEX-PM data. CLEANEX-PM k_{ex} rates (*black lines*) and *HET^{ex}ratios* (*red lines*) are plotted as a function of peptide sequence in (**b**) and (**e**), and color-coded on the crystal structures of ubiquitin (PDB 1UBQ) in (**c**) and CspA (PDB 1MJC) in (**f**). *A larger ball size* indicates faster exchange; while *gray* regions indicate the absence of NMR data, either because of spectral overlap or due to missing assignment. *Error bars* for k_{ex} rates were obtained from Monte Carlo simulations based on the measured spectral noise level, while *error bars* for *HET^{ex}ratios* were calculated by error propagation of the spectral noise

delay. If we assume that the longitudinal ^{1}H relaxation rate R_1^A is changing only little from one site to another, a hypothesis that is typically satisfied for well-structured protein regions, HET^{ex}ratios provide a measure of the change in solvent accessibility along the protein backbone. This is shown in Fig. 3b, e for ubiquitin and CspA, respectively. The measured HET^{ex}ratios provide *solvent-exchange profiles* that closely match the results obtained from quantification of exchange rates using CLEANEX-PM. This is especially true if the measured exchange rates fall within the range 0.1 s^{-1} $< k_{ex} <$ 10 s^{-1} (example of ubiquitin), where the HET^{ex}-BEST-TROSY approach is most sensitive. For both proteins, the measured HET^{ex}-ratios allow to differentiate between solvent-exposed and hydrogen-bonded amide sites as illustrated by the structural drawings in Fig. 3c, f.

We can conclude that HET^{ex}-BEST-TROSY, if performed for a single pH value, does not provide a valuable tool for accurate quantification of exchange rates in well-structured hydrogen-bonded parts of proteins, but it provides a fast and sensitive technique for the characterization of local structural compactness and water accessibility in globular proteins, as already outlined in our previous work (Schanda et al. 2006).

Application to a transiently populated protein folding intermediate

The HET^{ex} SOFAST/BEST technique is particularly interesting for applications to short-lived or transiently populated protein states, for which fast and sensitive NMR methods are mandatory. This is demonstrated here for the example of the amyloidogenic protein β2-microglobulin (B2M). It has been shown previously that during the refolding process from a highly unstructured to the well-structured native state, a folding intermediate (I-state) accumulates that is characterized by a non-native *trans* peptide bond between residues His^{31} and Pro^{32} (Kameda et al. 2005). This I-state is thought to be involved in the

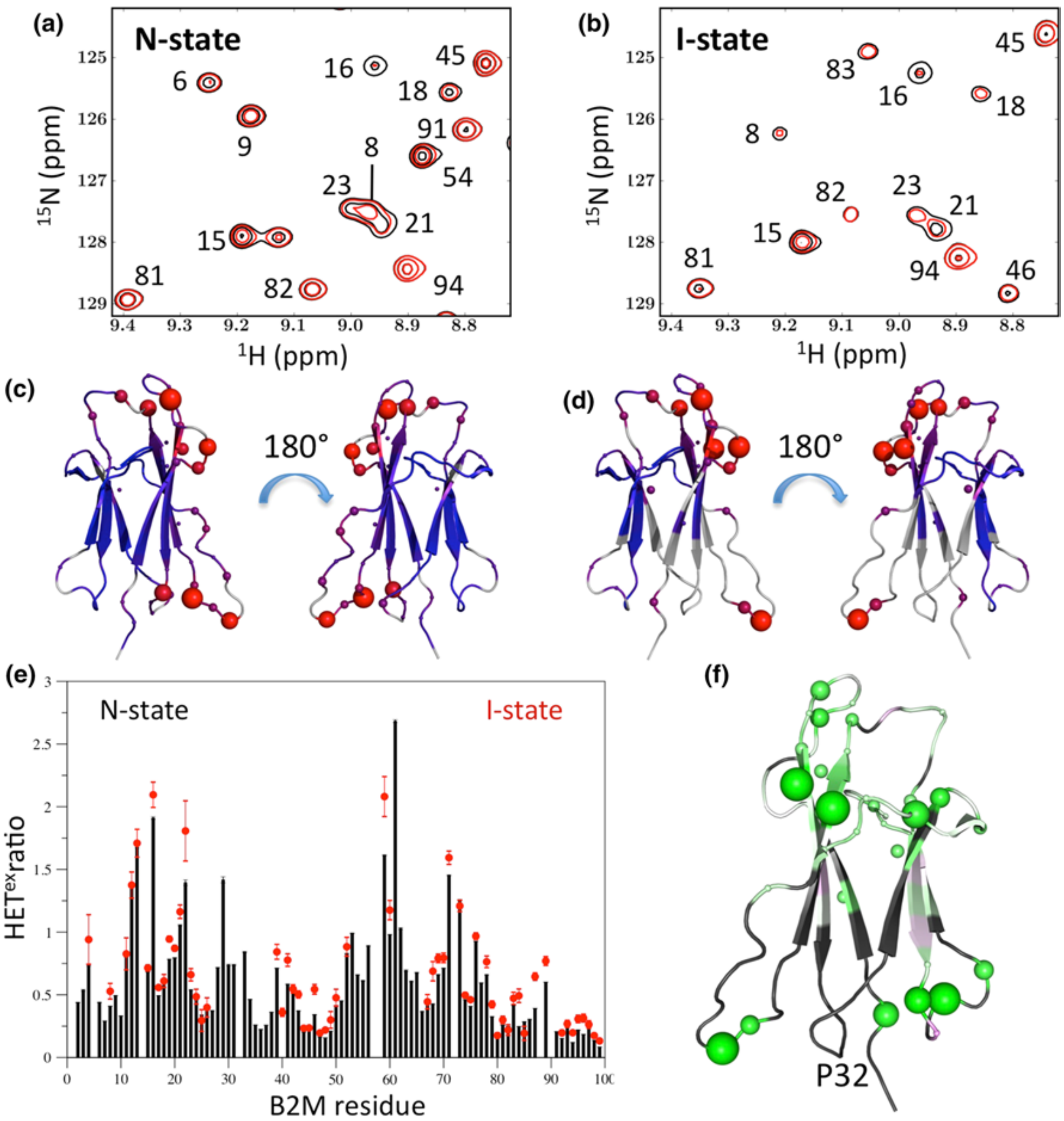


Fig. 4 Real-time hydrogen exchange measurements during the refolding of the amyloidogenic protein β2-microglobulin (B2M). Spectral regions of the ^{1}H–^{15}N HET^{ex}-BEST-TROSY reference (*black*) and saturation (*red*) spectra recorded for the N-state (**a**) and I-state (**b**) of B2M. CLEANEX-PM k_{ex} rates and *HET*^{ex}*ratios* measured for the N-state (**c**) and I-state (**d**) color-coded on the solution structure of B2M (PDB 2XKS). In **e** the same data are plotted as a function of the peptide sequence for the N-state (*black bars*) and I-state (*red circles*). *Error bars* for *HET*^{ex}*ratios* were calculated by error propagation of the spectral noise. **f** Differences between *HET*^{ex}*ratios* measured for the N- and I-states color-coded on the crystal structure of B2M. *A larger ball size* indicates faster exchange; while *gray* regions indicate the absence of NMR data, either because of spectral overlap or due to missing assignment

onset of amyloid fibril formation. Therefore obtaining atomic-resolution information on the I-state is of fundamental importance for a better understanding of the process of amyloidogenesis. Recently, using real-time multidimensional NMR, we have obtained sequential resonance assignments, as well as some information on the conformational dynamics in the I-state (Rennella et al. 2012). We could also demonstrate that the I-state binds to the hydrophobic dye ANS, while no such interaction is observed for the protein in its native state (Rennella et al. 2013). Under the chosen temperature of 15 °C, the I-state has a half-life time of $\sim$20 min, meaning that the maximum available experimental time for an NMR experiment of the I-state is about 40 min. Figure 4a, b show a small region of HET^{ex}-BEST-TROSY spectra recorded for the N-state, and the I-state of B2M, reconstructed from real-time NMR data. The spectral quality is sufficient to calculate HET^{ex}ratios for a large number of amide protons in the I-state. The measured HET^{ex}ratios are color-coded on the B2M structure in Fig. 4c (N-state), d (I-state), and plotted as a function of the protein sequence in Fig. 4e. A first conclusion from these data is that HET^{ex}-BEST-TROSY yields meaningful information on the solvent accessibility along the peptide chain of short-lived protein folding intermediates. Overall, the solvent-exchange profiles measured for the N- and I-states are very similar. Interestingly, however, a closer inspection of the rate differences between I- and N-states indicates a higher apparent solvent accessibility in the I-state loop regions at both sides of the protein (see Fig. 4f), while the hydrogen-bonding in the β-sheet region seems to be largely identical in the 2 states, in agreement with ^{13}C secondary chemical shift data reported previously (Rennella et al. 2012). An alternative interpretation of the observed differences between the N and I states would be a more compact overall conformation of the folding intermediate that may result in an increase of the NOE contribution to the measured HET^{ex}ratios.

Conclusions

We have demonstrated here that the measurement of amide ^{1}H polarization recovery rates under water saturating and non-saturating conditions provides a fast measure of hydrogen exchange processes along the backbone of ^{15}N-labeled proteins. SOFAST- and BEST-type readout sequences are ideally suited for HET^{ex} NMR as they minimally perturb the water ^{1}H polarization, thus allowing for high repetition rates (fast-pulsing), short overall acquisition times, and high experimental sensitivity. Typically, in case of abundant sensitivity, a full HET^{ex}–BEST data set can be recorded in a few minutes (and even less in case of HET^{ex}–SOFAST). We have presented two different approaches of data analysis: (1) Residue-specific HET^{ex} ratios can be computed from a pair of 2D NMR spectra recorded for a short relaxation delay, providing a semi-quantitative measure of the relative exchange efficiency along the peptide chain. This approach is most suitable for globular proteins or heterogeneous proteins with structured and unstructured parts, and it provides an efficient and fast method for distinguishing residues in structured and highly mobile protein regions. (2) If a second reference spectrum recorded for a longer relaxation delay is available, this allows extracting quantitative exchange rate constants, as long as Eq. (1) is valid. We have shown that for IDPs or highly disordered protein fragments, this approach yields accurate exchange rates in the range $0.1\ \mathrm{s}^{-1} < k_{ex} < 10\ \mathrm{s}^{-1}$, and thus provides an attractive alternative to CLEANEX-PM or other magnetization-transfer techniques, yielding similar results in much shorter experimental time. Finally, the potentially most interesting application of HET^{ex} NMR is the investigation of fast hydrogen exchange processes in short-lived protein states, as illustrated here for a folding intermediate of the amyloidogenic human protein β2-microglobulin that can only be studied by real-time NMR methods, requiring fast multidimensional data acquisition. HET^{ex} NMR adds a new probe of structural information for such transiently populated protein states. Similarly, HET^{ex} NMR may prove useful for investigating protein stability inside living cells or in cell extracts (Smith et al. 2013) where protein leakage and protein degradation limits the available measurement time for a single sample.

Acknowledgments We thank Isabel Ayala and Karine Giandoreggio for expert protein sample preparation, and we acknowledge support from ANR grant Blanc-InterII-SIMI7-2011 (RNAfolding). This work used the NMR and isotope labeling platforms of the Grenoble Instruct centre (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

Chevelkov V, Xue Y, Rao DK, Forman-Kay JD, Skrynnikov NR (2010) 15 N H/D-SOLEXSY experiment for accurate measurement of amide solvent exchange rates: application to denatured drkN SH3. J Biomol NMR 46:227–244

Croke RL, Sallum CO, Watson E, Watt ED, Alexandrescu AT (2008) Hydrogen exchange of monomeric alpha-synuclein shows unfolded structure persists at physiological temperature and is independent of molecular crowding in Escherichia coli. Protein Sci 17:1434–1445

Crowhurst KA, Tollinger M, Forman-Kay JD (2002) Cooperative Interactions and a Non-native Buried Trp in the Unfolded State of an SH3 Domain. J Mol Biol 322:163–178

Csizmok V, Felli IC, Tompa P, Banci L, Bertini I (2008) Structural and dynamic characterization of intrinsically disordered human securin by NMR spectroscopy. J Am Chem Soc 130:16873–16879

Dempsey CE (2001) Hydrogen exchange in peptides and proteins using NMR spectroscopy. Prog in NMR Spectrosc 39:135–170

Englander SW, Kallenbach NR (1983) Hydrogen exchange and structural dynamics of proteins and nucleic acids. Q Rev Biophys 16:521–655

Fan J-S, Lim J, Yu B, Yang D (2011) Measurement of amide hydrogen exchange rates with the use of radiation damping. J Biomol NMR 51:151–162

Favier A, Brutscher B (2011) Recovering lost magnetization: polarization enhancement in biomolecular NMR. J Biomol NMR 49:9–15

Fitzkee NC, Da Torchia, Bax A (2011) Measuring rapid hydrogen exchange in the homodimeric 36 kDa HIV-1 integrase catalytic core domain. Protein Sci 20:500–512

Gal M, Schanda P, Brutscher B, Frydman L (2007) UltraSOFAST HMQC NMR and the repetitive acquisition of 2D protein spectra at Hz rates. J Am Chem Soc 129:1372–1377

Geen H, Freeman R (1991) Band-selective radiofrequency pulses. J Magn Reson 93:93–141

Gemmecker G, Jahnke W, Kessler H (1993) Measurement of fast proton exchange rates in isotopically labeled compounds. J Am Chem Soc 115:11620–11621

Gil S, Hošek T, Solyom Z, Kümmerle R, Brutscher B, Pierattelli R, Felli IC (2013) NMR Spectroscopic Studies of Intrinsically Disordered Proteins at Near-Physiological Conditions. Angew Chem Int Ed Engl 52:11808-11812

Grzesiek S, Bax A (1993) Measurement of amide proton-exchange rates and Noes with Water in C-13/N-15-Enriched Calcineurin-B. J Biomol NMR 3:627–638

Hwang TL, van Zijl PC, Mori S (1998) Accurate quantitation of water-amide proton exchange rates using the phase-modulated CLEAN chemical EXchange (CLEANEX-PM) approach with a Fast-HSQC (FHSQC) detection scheme. J Biomol NMR 11:221–226

Kameda A, Hoshino M, Higurashi T, Takahashi S, Naiki H, Goto Y (2005) Nuclear magnetic resonance characterization of the refolding intermediate of beta(2)-microglobulin trapped by non-native prolyl peptide bond. J Mol Biol 348:383–397

Krishnan VV, Murali N (2013) Radiation damping in modern NMR experiments: progress and challenges. Prog in NMR Spectrosc 68:41–57

Kupce E, Freeman R (1994) Wide-Band Excitation with Polychromatic Pulses. J Magn Reson A 108:268–273

Kupce E, Boyd J, Campbell ID (1995) Short selective pulses for biochemical applications. J Magn Reson B 106:300–303

Lescop E, Schanda P, Brutscher B (2007) A set of BEST triple-resonance experiments for time-optimized protein resonance assignment. J Magn Reson 187:163–169

Mori S, O'Neil Johnson M, Berg JM, Van Zijl PCM (1994) Water exchange filter (WEX Filter) for nuclear magnetic resonance studies of macromolecules. J Am Chem Soc 116:11982–11984

Rennella E, Cutuil T, Schanda P, Ayala I, Forge V, Brutscher B (2012) Real-time NMR characterization of structure and dynamics in a transiently populated protein folding intermediate. J Am Chem Soc 134:8066–8069

Rennella E et al (2013) Oligomeric states along the folding pathways of β2-microglobulin: kinetics, thermodynamics, and structure. J Mol Biol 425:2722–2736

Schanda P, Brutscher B (2005) Very fast two-dimensional NMR spectroscopy for real-time investigation of dynamic events in proteins on the time scale of seconds. J Am Chem Soc 127:8014–8015

Schanda P, Kupce E, Brutscher B (2005) SOFAST-HMQC experiments for recording two-dimensional heteronuclear correlation spectra of proteins within a few seconds. J Biomol NMR 33:199–211

Schanda P, Forge V, Brutscher B (2006) HET-SOFAST NMR for fast detection of structural compactness and heterogeneity along polypeptide chains. Magn Reson Chem 44:S177–S184

Schanda P, Forge V, Brutscher B (2007) Protein folding and unfolding studied at atomic resolution by fast two-dimensional NMR spectroscopy. Proc Natl Acad Sci USA 104:11257–11262

Segawa T, Kateb F, Duma L, Bodenhausen G, Pelupessy P (2008) Exchange rate constants of invisible protons in proteins determined by NMR spectroscopy. ChemBioChem 9:537–542

Shishmarev D, Otting G (2011) Radiation damping on cryoprobes. J Magn Reson 213:76–81

Smith AE, Sarkar M, Young GB, Pielak GJ (2013) Amide proton exchange of a dynamic loop in cell extracts. Protein Sci 22:1313–1319

Wagner G, Wüthrich K (1982) Amide Proton Exchange and Surface Conformation Basic Pancreatic Trypsin Inhibitor in Solution. J Mol Biol 160:343–361

Wider G, Riek R, Wüthrich K (1996) Diffusion filters for separation of solvent—protein and protein–protein nuclear overhauser effects (HYDRA). J Am Chem Soc 7863:11629–11634

Yao S, Hinds MG, Murphy JM, Norton RS (2011) Exchange enhanced sensitivity gain for solvent-exchangeable protons in 2D 1H-15 N heteronuclear correlation spectra acquired with band-selective pulses. J Magn Reson 211:243–247

doi:10.1016/j.jmb.2012.04.023 J. Mol. Biol. (2012) **420**, 310–323

Contents lists available at www.sciencedirect.com

Journal of Molecular Biology

journal homepage: http://ees.elsevier.com.jmb

Transient Structure and SH3 Interaction Sites in an Intrinsically Disordered Fragment of the Hepatitis C Virus Protein NS5A

Sophie Feuerstein[1,2,3], Zsofia Solyom[1,2,3], Amine Aladag[4,5], Adrien Favier[1,2,3], Melanie Schwarten[1,2,3], Silke Hoffmann[4], Dieter Willbold[2,4,5] and Bernhard Brutscher[1,2,3]*

[1]*Institut de Biologie Structurale, Université Grenoble 1, 41 rue Jules Horowitz, 38027 Grenoble Cedex 1, France*
[2]*Commissariat à l'Energie Atomique et aux Energies Alternatives (CEA), Grenoble, France*
[3]*Centre National de Recherche Scientifique (CNRS), Grenoble, France*
[4]*Institute of Complex Systems (ICS-6) Structural Biochemistry, Forschungszentrum Jülich, 52425 Jülich, Germany*
[5]*Institut für Physikalische Biologie, Heinrich-Heine-Universität, 40225 Düsseldorf, Germany*

Received 12 April 2012;
accepted 21 April 2012
Available online
26 April 2012

Edited by P. Wright

Keywords:
Bin1;
hepatitis C virus;
intrinsic disorder;
NMR;
NS5A

Understanding the molecular mechanisms involved in virus replication and particle assembly is of primary fundamental and biomedical importance. Intrinsic conformational disorder plays a prominent role in viral proteins and their interaction with other viral and host cell proteins via transiently populated structural elements. Here, we report on the results of an investigation of an intrinsically disordered 188-residue fragment of the hepatitis C virus non-structural protein 5A (NS5A), which contains a classical poly-proline Src homology 3 (SH3) binding motif, using sensitivity- and resolution-optimized multidimensional NMR methods, complemented by small-angle X-ray scattering data. Our study provides detailed atomic-resolution information on transient local and long-range structure, as well as fast time scale dynamics in this NS5A fragment. In addition, we could characterize two distinct interaction modes with the SH3 domain of Bin1 (bridging integrator protein 1), a pro-apoptotic tumor suppressor. Despite being largely disordered, the protein contains three regions that transiently adopt α-helical structures, partly stabilized by long-range tertiary interactions. Two of these transient α-helices form a noncanonical SH3-binding motif, which allows low-affinity SH3 binding. Our results contribute to a better understanding of the role of the NS5A protein during hepatitis C virus infection. The present work also highlights the power of NMR spectroscopy to characterize multiple binding events including short-lived transient interactions between globular and highly disordered proteins.

---

*Corresponding author. Biomolecular NMR Spectroscopy Group, Institut de Biologie Structurale—Jean Pierre Ebel 41, rue Jules Horowitz, 38027 Grenoble Cedex 1, France. E-mail address: bernhard.brutscher@ibs.fr.

Abbreviations used: BEST, band-selective excitation short-transient; Bin1, bridging integrator protein 1; EOM, ensemble optimization method; HADAMAC, Hadamard-encoded amino-acid-type editing; HCV, hepatitis C virus; HETNOE, heteronuclear nuclear Overhauser enhancement; IDP, intrinsically disordered protein; LCS, low-complexity sequence; MTSL, (1-oxyl-2,2,5,5-tetramethyl-Δ^3-pyrroline-3-methyl)methanethiosulfonate; NOE, nuclear Overhauser enhancement; NS5A, non-structural protein 5A; PRE, paramagnetic relaxation enhancement; SAXS, small-angle X-ray scattering; SH3, Src homology 3; TROSY, transverse relaxation optimized spectroscopy; 2D, two-dimensional; 3D, three-dimensional.

Introduction

Hepatitis C virus (HCV), a member of the *Flaviviridae* family, is a major human pathogen that causes diseases such as chronic hepatitis, liver cirrhosis, and hepatocellular carcinoma.[1] As there is still no protective vaccine available and current drug treatments show limited efficiency, HCV infections present a major health problem worldwide. A better understanding of the molecular mechanisms involved in the HCV life cycle is required to develop new antiviral therapies. The HCV genome is translated into a single polyprotein of ~3000 amino acid residues that is then further processed by both viral and human proteases to a total of 10 different proteins.[2] These HCV proteins are classified into structural proteins that form the virus envelope, and non-structural proteins that are involved in a variety of vital functions for the virus. In particular, the non-structural protein 5A (NS5A), which exists as a 56-kDa basal or 58-kDa hyperphosphorylated form,[3] is essential for both HCV replication and particle assembly. Although its exact cellular function remains to be established, it has been reported that NS5A acts as a multifunctional regulator of cellular pathways via interactions with viral RNA,[4] as well as a variety of viral and human proteins like recently described for

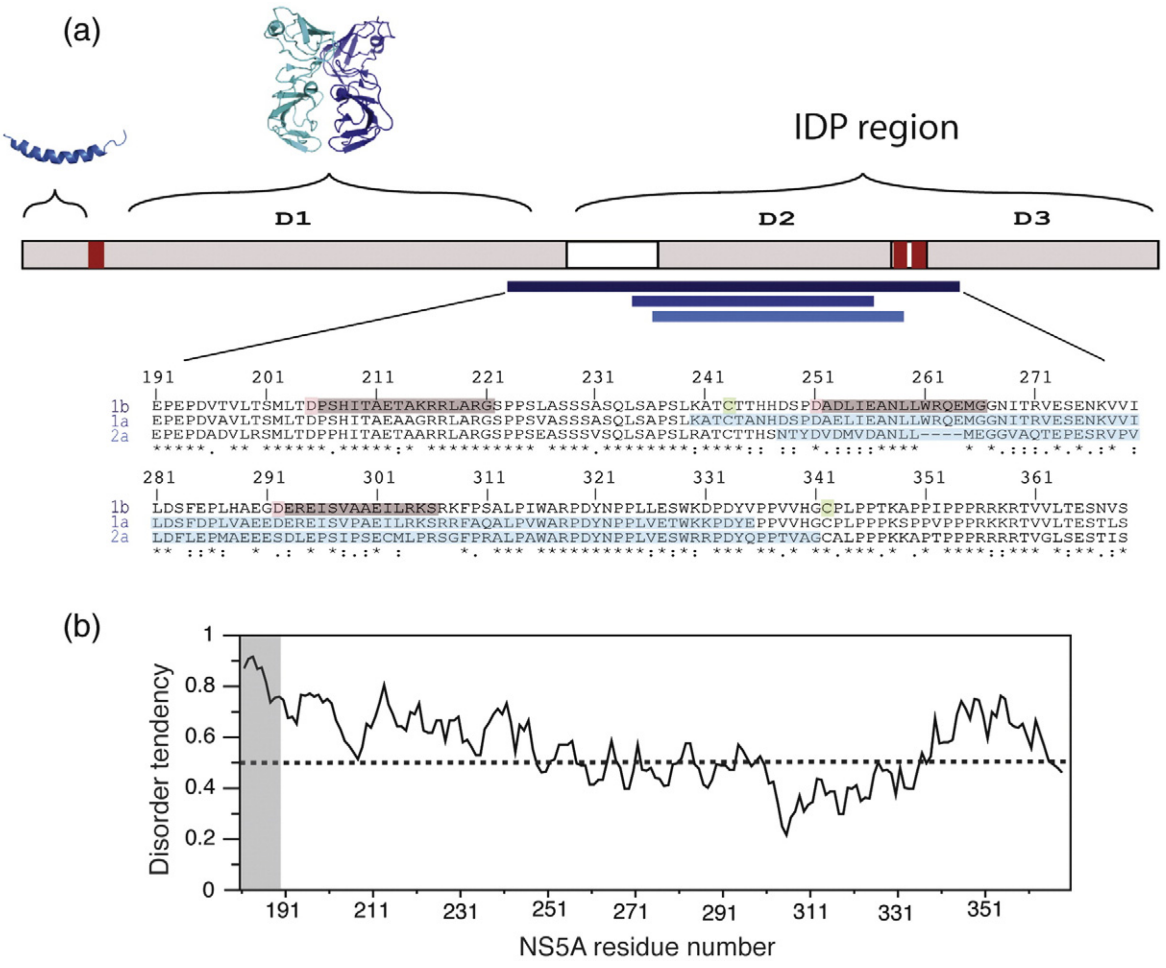


Fig. 1. (a) Domain organization and sequence alignment of NS5A from different genotypes for which NMR data are available. The gray bar describes the domain organization of NS5A including LCS linkers (white) and PxxP motifs (red). The corresponding ribbon diagrams of the N-terminal amphipathic helix[7] and the crystal structure[8] of domain 1 are shown on top. The blue bars beneath show the position of NS5A(191–369) (dark blue), as well as previously studied constructs of genotypes 1a (blue)[9] and 2a (light blue).[10] Highlighted in the sequence alignment are the regions of the three helices (gray) transiently formed in 1b, the N-capping aspartic acid residues (pink), and the cysteines (green) used for paramagnetic labeling, as well as the NS5A fragments studied previously for other genotypes (light blue). Sequence numbering is according to Kuiken *et al.*[11] (b) Prediction of conformational disorder along the peptide chain of the NS5A(191–369) fragment using the IUPred predictor.[12] Scores above a threshold value of 0.5 are considered to adopt highly flexible conformations.

c-Src[5] or lipid kinase PI4KIIIα.[6] NS5A is attached to the endoplasmic reticulum membrane via an N-terminal amphipathic α-helix. The remaining cytoplasmic part of NS5A is organized into three domains (Fig. 1), as defined by bioinformatic tools and limited proteolysis, connected by linker regions of low-complexity sequences (LCSs).[13] Among different HCV genotypes, domain 1 (D1) of NS5A is highly conserved with a well-defined globular fold[8,14] comprising a zinc-binding domain with RNA binding activity. The less conserved domains 2 and 3 (D2 and D3, respectively) have recently been shown[9,10,15] to belong to the class of intrinsically disordered proteins (IDPs) characterized by a high degree of conformational flexibility. Despite the lack of a persistent structure, IDPs can carry out important cellular functions such as signal transduction and transcription regulation by binding to various molecular partners, typically with high specificity but rather low affinity. Often, the molecular recognition sites of IDPs are transiently structured, and binding occurs via a combination of *induced fit* or *folding upon binding*[16] and *conformational selection* mechanisms.[17] Typically, the first binding event with an IDP occurs via a nonspecific "fly-casting" interaction, where disorder is favorable. After this nonspecific interaction has been established and the IDP is close to its binding partner, conformational selection plays a role in establishing specific interaction between the preformed structural element of the IDP and the target. Finally, the structure of the complex may be adjusted by binding induced folding.[18] IDPs are not amenable for structure analysis by crystallographic methods. Thus, multidimensional NMR spectroscopy is the only available technique providing atomic-resolution information on structural propensities, conformational ensembles, and local dynamics of IDPs in the absence or presence of a binding partner.

Our study focuses on NS5A(191–369) of genotype 1b comprising domain 2, as well as elongations in N- and C-terminal directions that contain both LCS linkers (Fig. 1). This NS5A fragment includes crucial binding motifs necessary for interactions with other viral and host proteins,[19] and in particular two class II PxxP motifs. These proline-rich sequences form poly-L-proline helices of type II (PPII) and are canonical binding motifs for Src homology 3 (SH3) domains. Several interactions between the C-terminal class II PxxP motif of NS5A and host cell SH3-containing proteins such as Grb2, Bin1 (bridging integrator protein 1), PI3K, and members of the Src-family of tyrosine kinases (Fyn, Lck, Hck, Lyn) have been reported.[20–25] These interactions play a potential role in the perturbation of multiple biological regulation processes such as cellular signaling and apoptosis. The atomistic view provided by our NMR study of structural propensities, conformational dynamics, and SH3-binding modes of this NS5A fragment contributes to a better understanding of NS5A promiscuity and function.

Results

The reduced frequency dispersion observed in the NMR spectra, especially in the ^{1}H dimension (Fig. S1), and the prediction of structural disorder by IUPred[12] (Fig. 1b) or various other algorithms indicate that NS5A(191–369) shows a high degree of conformational flexibility characteristic of an IDP. This observation is in agreement with results obtained previously[9,10] on shorter NS5A fragments of different HCV genotypes. However, AGADIR predictions of helical propensities (see Fig. 2f) revealed the presence of two significantly populated helical segments and a third one with a small helical population. Therefore, we have undertaken an NMR study providing site-resolved information on structural and dynamic features along the peptide chain of this NS5A fragment.

Structural and dynamic features of NS5A(191–369)

For NMR studies of globular proteins, structural information is mainly obtained from long-range ^{1}H–^{1}H nuclear Overhauser enhancements (NOEs). For NS5A(191–369), and IDPs in general, the observation of NOE cross peaks is generally limited to correlations within a single residue and between sequentially neighboring residues characterized by short ^{1}H–^{1}H distances independent of the protein's conformation. Therefore, only very limited structural information is obtained from NOE data of IDPs. Fortunately, a variety of other NMR observables that can be accurately quantified are available, thus providing useful information on the structural propensities and conformational dynamics of the IDP under investigation. Here, we used NMR chemical shifts, ^{15}N relaxation rates, and paramagnetic relaxation enhancement (PRE) effects to characterize the NS5A(191–369) fragment in terms of structure and dynamics at single-residue resolution.

Sequential resonance assignment of this 188-residue IDP was obtained using a set of band-selective excitation short-transient (BEST)-transverse relaxation optimized spectroscopy (TROSY)[31–35] three-dimensional (3D) correlation experiments and an additional amino-acid-type edited HADAMAC (Hadamard-encoded amino-acid-type editing) experiment.[36] BEST-TROSY experiments yield improved sensitivity and resolution, while allowing for shorter overall acquisition times. These experiments are therefore especially attractive for large, highly flexible proteins, characterized by a high level of

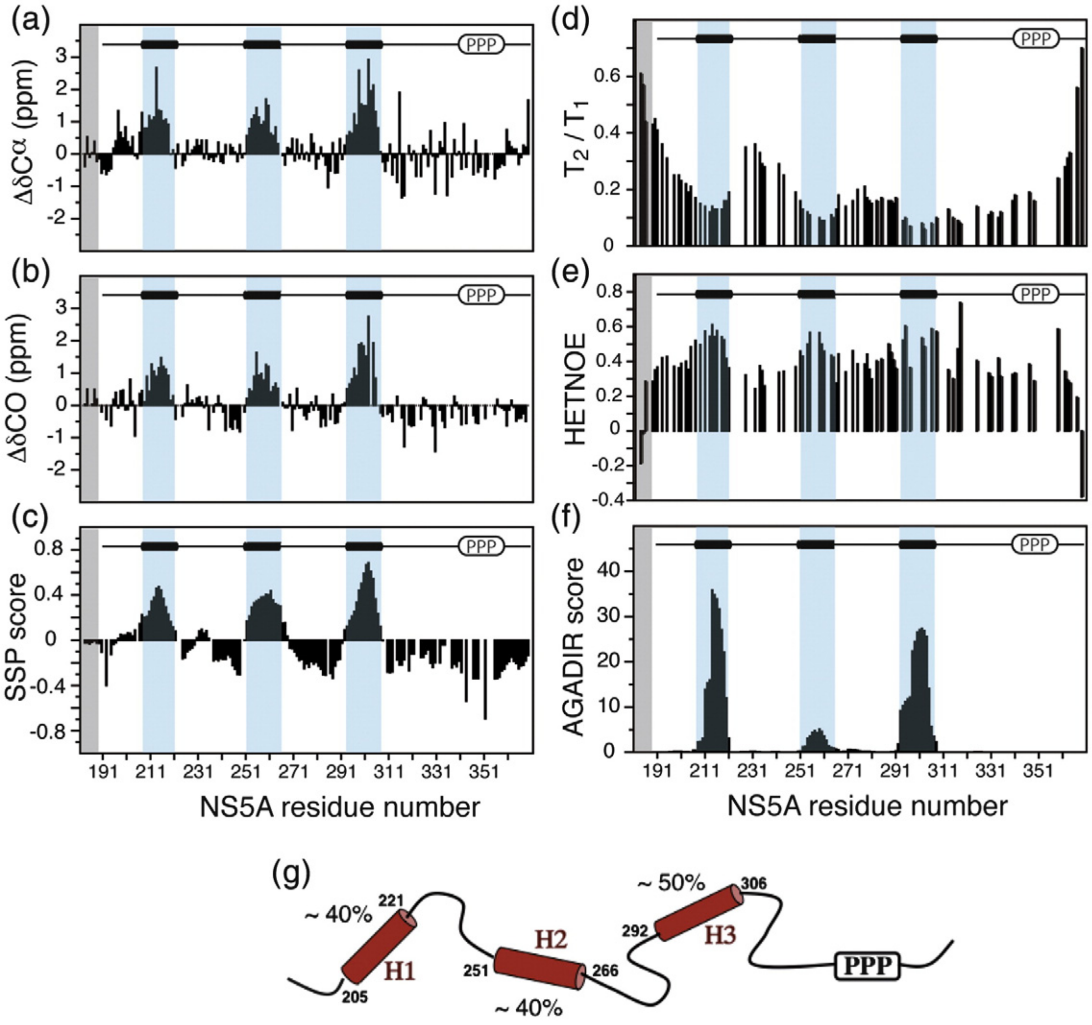


Fig. 2. Structural and dynamic characterization of NS5A(191–369) at 5 °C from NMR chemical shift, and ^{15}N relaxation data. (a) C^{α} and (b) CO secondary chemical shifts calculated as the difference between measured and random-coil chemical shifts taken from Zhang *et al.*[26] and corrected for next-neighbor effects.[27] (c) Structural propensities calculated with the SSP program.[28] All available chemical shifts (H^N, N, C^{α}, C^{β}, and CO) have been used as input data for the SSP calculation. Positive score values indicate the amount of α-helical conformation present along the sequence, while negative values are indicative of extended or β-strand conformations. (d) Ratios of measured transverse and longitudinal ^{15}N relaxation times (T_2/T_1) providing an estimate of the local rotational correlation times τ_c along the peptide chain. (e) $\{^1H\}^{15}N$ NOE (HETNOE) values indicative of the amplitude (order parameter) of local subnanosecond motions (high values: restricted motion, low values: high-amplitude motions). (f) Prediction of helical propensities based on the physical chemistry of the polypeptide chain using the program AGADIR.[29,30] (g) Structural model derived from the recorded NMR data showing three α-helical segments of similar length that are formed with propensities of ~40% for H1 and H2 and ~50% for H3, connected by flexible linker regions. The corresponding data at 20 °C are shown in Fig. S2.

NMR frequency degeneracy, as well as low sample concentration (to avoid aggregation) and limited sample stability. An experimental comparison of the overall performance of BEST-TROSY with respect to conventional implementations for the 188-residue NS5A protein fragment is shown in Fig. S3a and b. A specific feature of IDPs is that they often have repetitive sequence parts and that they are rich in proline residues. The NS5A(191–369) fragment has a total of 24 proline residues including seven Pro–Pro pairs. This presents an additional complication for the sequence-specific resonance assignment of IDPs because proline residues are break points in the sequential assignment walk. Therefore, additional amino-acid-type information for individual residues is mandatory. Here, we have used a HADAMAC experiment[36] that allows distinguishing between seven different classes of amino acids (Fig. S3d). The combination of a set of high-resolution 3D BEST-TROSY correlation experiments and HADAMAC data enabled us to obtain a high level of sequence-specific resonance assignment at 5 °C (20 °C): H^N: 98% (88%), ^{15}N: 85% (77%), $^{13}C'$: 85% (77%), $^{13}C^{\alpha}$: 96% (92%), and $^{13}C^{\beta}$: 96% (76%). The NMR assignment at 5 °C has been reported previously[37] and deposited with the BioMagResBank under accession number 17468.

Chemical shifts are sensitive reporters on the local chemical environment. In particular, secondary chemical shifts, the deviation of the measured resonance frequencies from random-coil values of backbone nuclei, depend on backbone dihedral angles and thus can be used to estimate the amount of regular secondary structure present along the peptide chain. $^{13}C^{\alpha}$ and $^{13}C'$ secondary chemical shifts measured for NS5A(191–369) are shown in Fig. 2a and b, respectively. For most residues, the absolute values of these secondary chemical shifts are below 1.0 to 1.5 ppm, indicating the absence of any persistent secondary structure in this protein fragment consistent with an IDP. Nevertheless, several continuous stretches of positive $^{13}C^{\alpha}$ and $^{13}C'$ secondary chemical shifts, each comprising about 15 residues, can be clearly identified, indicating peptide segments with a significant propensity to form α-helical structure. Using the SSP score[28] shown in Fig. 2c, we can estimate α-helical propensities of up to ~38% (SSP score averaged over 10 consecutive residues) for the segments comprising residues 205–221 (**H1**) and 251–266 (**H2**) and ~51% for the segment 292–306 (**H3**). Interestingly, aspartic acid residues D205, D251, and D292 (replaced by a Ser in genotype 2a) that are well conserved among the different genotypes (see Fig. 1) are found at the beginning of each of these transiently populated helices. Aspartic acid is the most prevalently found helix-capping residue in high-resolution structures of globular proteins.[38]

The structural model emerging from the NMR chemical shift data is further supported by ^{15}N relaxation measurements [T_1, T_2, and heteronuclear NOE (HETNOE)]. The T_2/T_1 ratio (Fig. 2d) provides an estimate of the effective local tumbling correlation time along the peptide chain, while variations in the HETNOE values (Fig. 2e) are indicative of differences in local order. Interestingly, the peptide segments (H1, H2, and H3) that have been identified by the chemical shift data to have a high propensity for α-helical conformation also show increased tumbling correlation times τ_c (low T_2/T_1 ratios) and increased local order (HETNOE>0.5). The ^{15}N relaxation data are thus in agreement with a model of NS5A(191–369) consisting of three helical segments that mainly experience collective segmental motion separated by more flexible linker regions. All three helices show some fraying effects in the sense that the effective τ_c values (T_2/T_1 ratios) decrease (increase) toward the helix ends. This behavior is generally not observed for helices in globular proteins but is a typical feature of highly flexible, partially unfolded proteins with transiently populated helices that are mainly stabilized by intrasegmental, rather than by long-lived tertiary interactions.[39]

To investigate the importance of long-range interactions in NS5A(191–369), we first have compared the experimental NMR-derived results with AGADIR predictions.[29] The AGADIR algorithm uses helix/coil transition theory to estimate α-helical propensities (in percentage) along peptide chains with a confidence level of about ±10%†. Significant deviations from the AGADIR predictions are therefore indicative of the presence of additional long-range interactions between different segments of the peptide chain. The results obtained after submitting the NS5A(191–369) sequence to the AGADIR program are shown in Fig. 2f. Interestingly, the same three segments as identified by our NMR data are predicted by AGADIR with α-helical propensities ranging from ~5% for H2, to ~25% for H3, to ~35% for H1. The values predicted for helices H2 and H3 are significantly lower than the experimental estimates based on NMR chemical shifts, yielding a first indication of the presence of transient long-range interactions that contribute to the stabilization of these helical structures. An experimental indication of the presence of tertiary interactions is provided by the stability of these transient helices with respect to a temperature increase. The NMR results (secondary chemical shifts and ^{15}N relaxation data) obtained for a sample temperature of 20 °C are shown in Fig. S2. While AGADIR predicts helical propensities reduced by about 30% when increasing the temperature from 5 to 20 °C (Fig. S2f), the experimentally derived helix populations are only little affected by this temperature change (Fig. S2a–c). Furthermore, similar HETNOE values are measured within these transiently formed structural elements (Fig. S2e), indicating that the local dynamics is also only little affected by this temperature change. The major differences are observed in the T_2/T_1 ratios (Fig. S2d), in agreement with faster rotational tumbling of the helical segments as expected from the reduced viscosity of the solvent at higher temperature.

In order to obtain some direct evidence of long-range interactions in this NS5A fragment, we have performed PRE measurements. The native NS5A(191–369) sequence contains two cysteine residues (C243 and C342), both located outside the helical segments, that provide convenient attachment points for a nitroxide spin label [(1-oxyl-2,2,5,5-tetramethyl-Δ^3-pyrroline-3-methyl)methanethiosulfonate (MTSL)]. PRE data were measured for two single-cysteine mutants (NS5A_C243S and NS5A_C342S, respectively) where one of the two wild-type

---

† http://agadir.crg.es/agadir.jsp

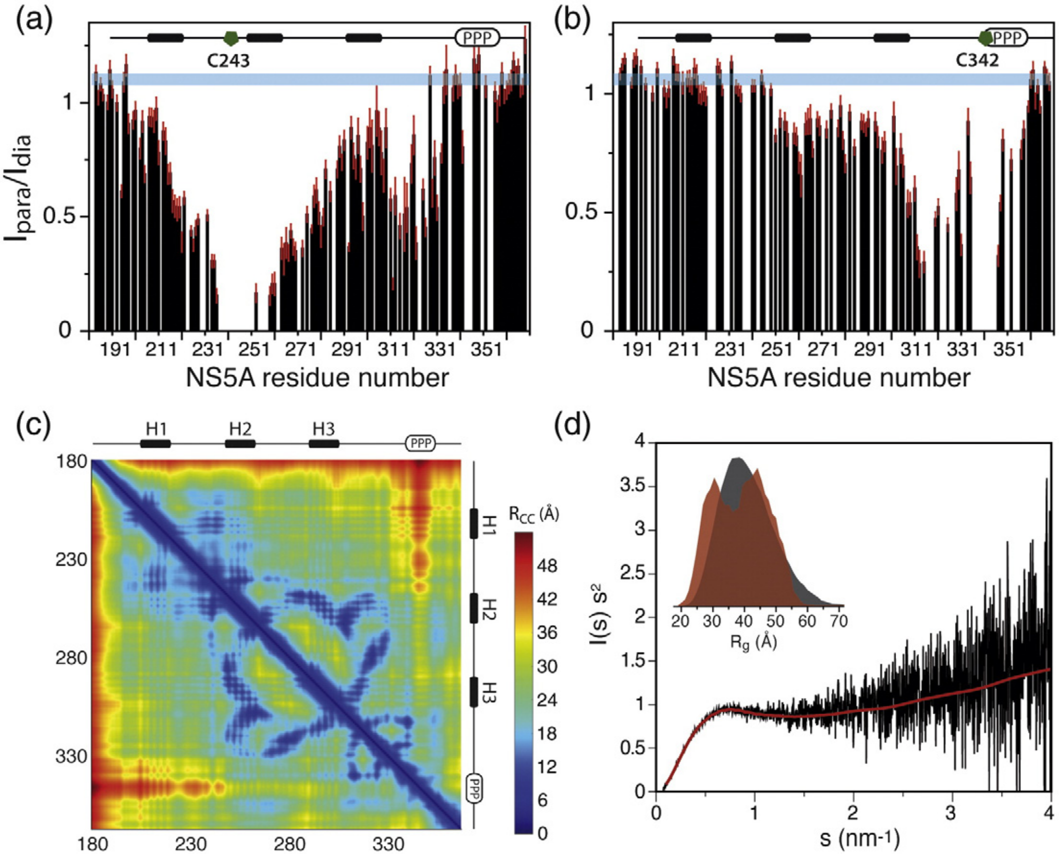


Fig. 3. PRE effects measured for MTSL-labeled NS5A mutants. Plotted are the intensity ratios $I_{\mathrm{para}}/I_{\mathrm{dia}}$ *versus* the primary sequence of (a) NS5A_C342S and (b) NS5A_C243S. Because of longitudinal relaxation enhancement effects in the presence of paramagnetic ions in the NMR sample, the threshold value in our experimental setup is $I_{\mathrm{para}}/I_{\mathrm{dia}} \approx 1.1$ (horizontal blue bar). The position of the nitroxide spin label (in green), the three transient helices, and the proline-rich region are indicated on top of the histogram. (c) Results of single-molecule structure calculations using PRE-derived distance restraints and fixed α-helical geometry for the peptide regions 205–221, 251–266, and 292–306. The average C^{α}–C^{α} contact map, computed for the 20 lowest-energy structures, highlights the transient close contacts in the NS5A structural ensemble. (d) Kratky plot representation of SAXS data recorded on a 90-μM sample of NS5A(191–369) at 20 °C. The red line represents the theoretical scattering curve from the EOM-optimized structural ensemble. The inset shows the radius of gyration (R_g) distribution of the initial pool of conformers (black) and of the conformers that fit the SAXS data (red).

cysteines was replaced by a serine, and the MTSL spin label was chemically attached to the remaining cysteine residue. Figure 3 shows the intensity ratios of the paramagnetic and diamagnetic states of the nitroxide radical ($I_{\mathrm{para}}/I_{\mathrm{dia}}$) observed for individual amide protons in both MTSL-labeled-NS5A(191–369) mutants. PRE effects ($I_{\mathrm{para}}/I_{\mathrm{dia}} <$ threshold value) report on the average spatial proximity of a given amide proton to the paramagnetic spin label. In the case of an extended polypeptide chain, the PRE effect is expected to be limited to neighboring residues. In clear contrast, we observe PRE effects for many residues far away in sequence from the site of spin labeling. In particular, placing a spin label at position 342 (close to the poly-proline region or PxxP motifs) has an effect ($I_{\mathrm{para}}/I_{\mathrm{dia}} < 0.8$) on all preceding residues up to about residue 250 including helices H2 and H3. Similarly, a spin label at position 243 affects all residues between positions 200 and 330 with a particularly strong effect in the region comprising residues 311 to 326.

NS5A(191–369) adopts a multitude of distinct conformations that interconvert on a subnanosecond time scale. The measured PREs are thus averages over all possible conformations in this structural ensemble. Furthermore, because of the r^{-6} dependence of PREs from the electron–nuclear distance r, more compact structures with short distances r have a much stronger contribution to the measured PRE effect than more extended structures. Although the quantity and quality of the available PRE-derived distance restraints do not allow the determination of a realistic structural ensemble, we have undertaken calculations where all PRE-derived distance restraints are simultaneously enforced to the same structure, thus providing some information about compact conformations that NS5A may adopt in solution. In addition, the backbone dihedral angles in the transiently populated α-helical peptide regions were fixed to canonical α-helical geometry. The 20 lowest-energy structures obtained from such a calculation (Fig. S5) violate the PRE-derived distance restraints by less than 3 Å, a value that is well below a reasonable estimate of the experimental uncertainty. The calculated structures provide valuable information on long-range interactions in the structural ensemble that NS5A forms in solution, although they do not have to occur at the same time. The characteristic structural features that are conserved among these 20 lowest-energy structures are best appreciated from the contact map shown in Fig. 3c, where dark blue spots represent average C^{α}–C^{α} distances of less than 12 Å. Interestingly, the transiently populated helices H2 and H3 are in close proximity in all of the 20 structures, and their relative position is identical within an RMSD of 1.3 Å (Fig. S5c). These observations further support the idea of a cooperative folding process of helices H2 and H3 with long-range interactions that stabilize these transient helix folds, thus explaining the deviations of the experimentally derived helical propensities from the AGADIR predictions.

Additional information on the overall dimension of the structural ensemble that NS5A(191–369) forms in solution was obtained from small-angle X-ray scattering (SAXS) experiments. SAXS data recorded for two different protein concentrations show no significant difference (Fig. S4), indicating that intermolecular interactions can be neglected. The Kratky representation of these data, shown in Fig. 3d, is characteristic of a highly disordered protein. Nevertheless, a small bump observed in this scattering curve (at $s \approx 0.7$) is indicative of a certain degree of compactness.[40] To get a more quantitative description of the size distribution in the conformational ensemble, we have used the genetic algorithm EOM (ensemble optimization method)[41] that fits the SAXS data by selecting a set of structures out of a large pool representing the conformational space accessible to a highly flexible protein. The EOM-optimized ensemble of conformers (Fig. 3d) shows a bimodal R_g distribution containing both more compact and more extended structures. This observation is in qualitative agreement with the NMR-derived picture of NS5A undergoing transient compaction in conjunction with cooperative α-helix formation.

Interaction of NS5A(191–369) with a human SH3 domain

In addition to the characterization of conformation and dynamics of NS5A(191–369), we investigated its interaction with Bin1–SH3 as an example of a human SH3 domain. The changes in peak positions and line widths observed in the NMR spectra between the bound and the free protein provide information on binding sites and affinities of the studied protein–protein interaction. NS5A has been shown to bind various SH3 domains, thereby influencing host cell signaling pathways. So far, these interactions were mostly analyzed by co-immunoprecipitation, yeast two-hybrid systems, and co-localization,[21,23,25,42,43] but to our knowledge, no atomic-resolution information is yet available for these NS5A–SH3 interactions. The binding of NS5A to the tumor suppressor Bin1 (also known as *amphiphysin II*) was first reported by Zech *et al.*[44] Consistent with the membrane-tubulating potential of Bin1 (mediated by its BAR domain), a role of the NS5A–Bin1 interaction during membranous web formation was speculated. Another possible function of this NS5A–Bin1–SH3 interaction is the inhibition of the tumor suppressor activity of Bin1 by competing with c-Myc for SH3 binding. Furthermore, it has been reported that Bin1 prevents NS5A from hyperphosphorylation, thus ensuring HCV RNA replication.[22]

A detailed view of the NS5A binding sites for Bin1–SH3 was obtained from NMR titration experiments. Two-dimensional (2D) ^{1}H–^{15}N BEST-TROSY spectra of [U-^{13}C, U-^{15}N]-NS5A(191–369) were recorded after adding increasing amounts of unlabeled Bin1–SH3 up to a final molar ratio of 1:10 (NS5A:Bin1–SH3). Representative parts of the titration spectra are shown in Fig. 4a and b, while the full spectra are provided as Fig. S8. Adding a small amount of Bin1–SH3 (molar ratio of 5:1) only affects the line widths and chemical shifts of amide resonances in the region 342 to 366 (Fig. 4a) comprising three canonical SH3 binding sites (PPII helix motifs): one of class I [PP1.2 (348–354)] and two of class II [PP2.1 (343–348) and PP2.2 (351–356)] (Fig. 4c). Separate correlation peaks are detected for the free and SH3-bound NS5A fragment indicating slow exchange kinetics on the NMR chemical shift time scale. This refers to a macroscopic system with a high binding affinity of the two proteins, consistent with a K_d of 0.1 μM measured by surface plasmon resonance for a short NS5A(333–369) fragment that comprises the entire poly-proline rich region (Fig. S6). Consequently, at an equimolar mixture of the two proteins at 100 μM concentration, this high-affinity NS5A(191–369) binding site is saturated, and only peaks from the bound state were observable in the NMR spectra. Furthermore, for the central residues I352 and R356 to T360, the bound-state peak is broadened beyond the NMR detection limit (Fig. 4a) and other peaks, for example, A349, are very weak in the bound state. A first explanation of the observed line broadening is the presence of conformational dynamics and/or heterogeneity in

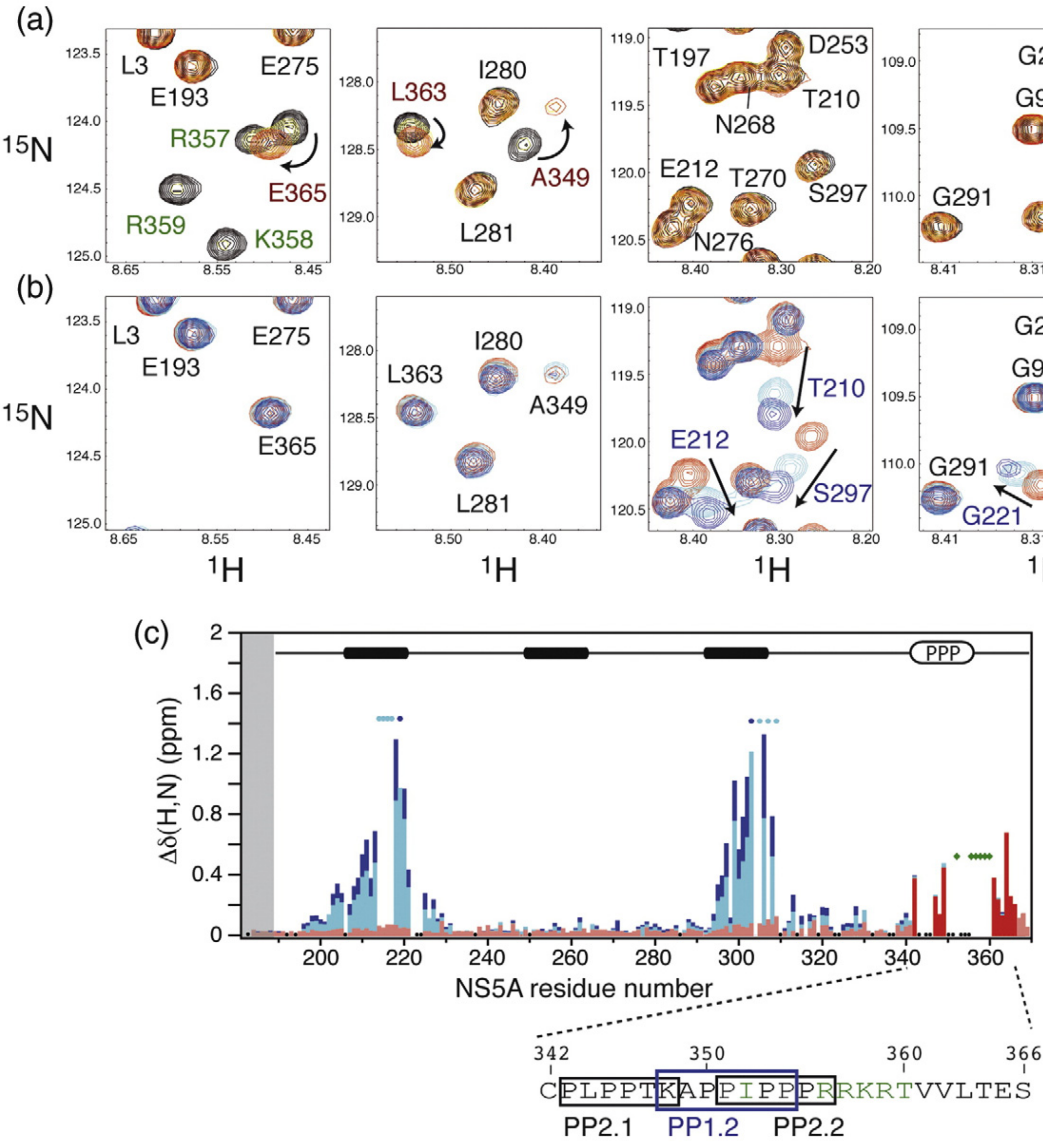


Fig. 4. Interaction of NS5A(191–369) with human Bin1–SH3. Sections of overlaid 2D ^{1}H–^{15}N BEST-TROSY spectra (including peak assignments) of [U-^{13}C, U-^{15}N]-NS5A(191–369) in the presence of increasing amounts of unlabeled Bin1–SH3. (a) High-affinity binding event represented by a reference spectrum (black), and spectra recorded at molar NS5A(191–369):Bin1–SH3 ratios of 2:1 (yellow) and 1:1 (red). (b) Low-affinity binding site. Spectra are shown for NS5A(191–369):Bin1–SH3 mixtures of 1:1 (red), 1:5 (cyan), and 1:10 (blue). The spectral regions comprise residues from both helices H1 and H3 that are part of the second binding site. (c) Chemical shift changes $\Delta\delta(\mathrm{H,N}) = ((10*\Delta\delta\mathrm{H}^{\mathrm{N}})^2 + (\Delta\delta\mathrm{N})^2)^{1/2}$ plotted as a function of the primary sequence of NS5A(191–369). Observed chemical shift changes for residues undergoing slow exchange are indicated by red bars (residues 342 to 366), while measured peak shifts are shown for molar ratios of 1:1 (light red bars), 1:5 (cyan bars), and 1:10 (blue bars). Missing peaks due to line broadening in the bound states are indicated by green diamonds for the 1:1 spectra and by cyan circles for the 1:5 spectra. The chemical shift changes of amino acids A219 and L303 at a molar ratio of 1:10 (indicated by a blue circle) could not be analyzed because the corresponding peaks were not distinguishable from signals of residues A252 and D331. Missing bars for residues R304 and W316 are due to lack of assignment. The positions of proline residues are displayed by black circles, and the transiently formed α-helices are shown on top of the histogram. In addition, the exact amino acid sequence and the location of the canonical SH3 class I and II binding motifs within the high-affinity binding site are given below.

the bound state, e.g., binding to different SH3 binding motifs within the peptide region affected by SH3 binding (Fig. 4c). A second explanation could be a change in the effective local tumbling correlation time upon SH3 binding. At the low temperature (5 °C) at which these titration experiments were performed, this may lead to significantly enhanced transverse spin relaxation possibly explaining the reduced signal intensity observed for tightly bound NS5A residues.

Interestingly, we also could map a second, low-affinity binding site that consists of two peptide regions, (1) residue 198 to 231 and (2) residues 294 to 330, comprising the transiently populated helices H1 and H3, respectively. While no chemical shift changes were observed for residues located in these regions up to an equimolar mixture of the two proteins, the corresponding ^{1}H–^{15}N correlation peaks start to move linearly towards the chemical shifts of the bound state (Fig. 4b). Under conditions of a fivefold excess of Bin1–SH3, NMR signals from residues A214, K215, R216, and R217 and from residues K305, R307, and F309 in the central regions of H1 and H3, respectively, were undetectable due to extensive line broadening. Similar to what we discussed for the high-affinity binding site, this may be explained by a broadening of peaks in the bound state because of either reduced local molecular tumbling or conformational heterogeneity. It also could be that the exchange between free and SH3-bound NS5A is in the intermediate time regime ($k_{ex} \approx 2\pi\Delta\nu$). In the latter case, we can estimate from the maximal chemical shift change $\Delta\nu_{max}$ observed in the central parts of the NS5A binding site an exchange rate constant on the order of $k_{ex} \approx 10^{3}$ s^{-1}. The measured chemical shift changes $\Delta\delta(H^{N},N)$ corresponding to molar ratios NS5A:Bin1–SH3 of 1:1, 1:5, and 1:10 are shown in Fig. 4c as a function of the NS5A sequence. No changes in chemical shift were observed for residues 252 to 266 corresponding to H2, indicating that this transiently formed helix is not affected by Bin1–SH3 binding. An estimate of the dissociation constant for this second binding mode was obtained from the chemical shift changes observed in the NMR titration experiment assuming a simple single-site binding model (Fig. S9). The K_d values obtained by fitting the chemical shift data for different residues within the binding site of NS5A varied between 200 μM and 900 μM. The large scattering in the observed K_d values is not surprising, because even at the final NS5A:Bin1–SH3 molar ratio of 1:10, this second binding site was still far from being saturated. Therefore, our chemical shift data only allow to estimate a lower limit for the dissociation constant $K_d > 200$ μM, while a more precise estimation would require further titration points to reach saturation of this binding site. The NS5A regions affected by this low-affinity SH3 binding mode significantly extend outside the boundaries of the transiently formed helices as identified from ^{13}C secondary chemical shifts (Fig. 2). This observation suggests localized changes in the structure and conformational dynamics of NS5A when bound to Bin1–SH3. However, we cannot conclude from our data whether this corresponds to a more structured conformation of NS5A (folding upon binding) or whether it reflects additional contacts with the binding partner, for example, with the proline-rich region 310–330. Additional data will be required to get more detailed information on the NS5A bound-state structure.

Discussion

Structural features in NS5A(191–369): Transient order in an IDP

NMR chemical shift and ^{15}N relaxation data revealed the presence of three segments in NS5A(191–369) with a significant propensity (40–50%) to adopt α-helical structures. While the positions of these helical regions are in good agreement with bioinformatic predictions, the experimentally derived helical propensities for two of the three helical segments are much higher than predicted, and the helix populations remain almost unchanged upon temperature increase from 5 to 20 °C. This discrepancy is explained by short-lived tertiary interactions that possibly stabilize the helical folds of H2 and H3. From all the NMR data reported herein, a structural picture emerges for NS5A(191–369) that significantly differs from a simple random-coil chain. Although the low chemical shift dispersion observed in the NMR spectra and the ^{15}N relaxation data indicate that the protein explores a wide range of conformations interconverting on a fast (less than a nanosecond) time scale, there is also clear evidence for the presence of more ordered conformations. Most likely, the formation of transient secondary and tertiary structures plays a role in the recognition of various binding partners, both viral and host cell proteins, that confers a functional advantage to the NS5A protein.

Our results differ from other NMR studies reported in the literature on similar NS5A fragments coded by HCV genotypes 1a[9] and 2a.[10] In both cases, the studied constructs are much shorter (Fig. 1a) and lack the poly-proline motifs at the C-terminus, as well as the N-terminal extension (comprising helix H1 in genotype 1b). Both studies not only conclude on the absence of any significantly populated secondary structural elements but also observe a tendency of α-helical structure formation upon the addition of TFE as revealed by circular dichroism. AGADIR predictions for the NS5A protein coded by the HCV genotypes 1a, 1b, and 2a (Fig. S7) show a conserved propensity of helix formation in the H1 region, although with strong variations in the absolute value, indicating that this structural motif is well conserved among different genotypes. While the sequences of sub-genotypes 1a and 1b are also similar in helical regions H2 and H3, this part of NS5A seems to be less well conserved between different genotypes. The sequence identity between 1b and 2a is only 50%

for H2 and 26% for H3. As a consequence, almost no helical structure is predicted for these NS5A regions of genotype 2a, but no experimental data that confirm (or disconfirm) these predictions are available so far. It would also be interesting to study a longer NS5A construct comprising, in addition to the fragment studied here, the entire D3. This domain also contains a segment with significant helical propensity[45] that is well conserved among these three genotypes (Fig. S7).

NS5A: A molecular interaction platform

NS5A is believed to perform a variety of viral functions in the infected cell that are vital for HCV replication and particle formation. A main role of NS5A is to transiently interact with human proteins in order to recruit them to the replication complex or to interfere with cellular processes. Many identified interaction sites of NS5A with host cell proteins are localized in D2 and the adjacent LCS regions. For a better understanding of NS5A binding modes, we studied here the interaction of NS5A(191–369) with the human Bin1–SH3 domain. SH3 domains meditate protein–protein interactions and are involved in many signaling processes in the cytoplasm. Besides their well-described binding to PPII helices,[46,47] recent studies also showed the possibility of different binding mechanisms, mainly involving helical motifs.[48–51] In addition, based on a combined approach of comparative genomics and disorder prediction, it was suggested that SH3 domains may bind to disordered protein regions.[52]

The NMR results reported here show that SH3 domain binding to NS5A is a more complex process that is not entirely described by PxxP motif recognition but may also use transient interactions with α-helical peptide regions that function as molecular recognition elements. Bin1–SH3 binding to NS5A does not seem to be a cooperative process, suggesting a model where two different SH3 domains bind to distinct NS5A binding sites, either the PxxP motif or the two α-helices. Such multivalent binding mechanisms of SH3 domains were described previously among others for p67phox–p47phos[50] and Pex13p–Pex14p–Pex5p.[53] The exact role and functional importance of this noncanonical, low-affinity binding site still requires further investigation. So far, we can only speculate about its biological relevance. One role of this divalent binding motif could be preferential binding of proteins that either contain two SH3 domains or exist as dimers. Actually, Bin1 has shown to be able to form dimers via its BAR domain.[54] The SH3 binding activity of NS5A may well be disrupting SH3-domain-mediated signaling pathways, especially if the SH3-domain-containing proteins are membrane anchored, regardless whether this is of additional advantage for HCV replication or a collateral damage. Another potential function of the NS5A–SH3 interaction may be interference with other binding events that are relevant for a variety of NS5A activities and that require access to protein regions within or close to the identified Bin1–SH3 binding sites. In addition to SH3 domains of various human proteins[21,25,44] competing for the same binding sites, other examples of interfering interaction sites are the serine residues near position 230 that are described to be phosphorylated *in vivo*,[3] the interferon sensitivity-determining region,[55] the nuclear localization signal,[56] the binding site for the cellular hVAP-A protein,[57] and the viral core protein,[58] as well as viral RNA.[59] Furthermore, Bin1 binding may also affect proline 319, which is a substrate for isomerization by the peptidyl prolyl isomerase cyclophilin A.[60] Thus, NS5A may act as a complex molecular interaction platform, and the potential interferences between numerous NS5A activities still remain to be elucidated.

In summary, we have reported a detailed analysis of the structural and dynamic features of NS5A(191–369), as well as its modes of interaction with the SH3 domain of Bin1. Our NMR data indicate three regions that transiently adopt α-helical structures, which are partly stabilized by long-range tertiary interactions. Two of the three transient helices are involved in noncanonical SH3 binding, forming a second binding motif, in addition to the classical binding of PxxP motifs. Our study also highlights the power of NMR spectroscopy for the characterization of multiple binding events, including short-lived transient interactions between globular and highly disordered proteins, a challenging task for most alternative biochemical or spectroscopic techniques. The present work contributes to a better understanding of the molecular basis of the role of NS5A in HCV diversity and pathogenicity.

Materials and Methods

Expression and purification of [*U*-^{13}C, *U*-^{15}N]-NS5A(191–369)

Expression and purification of the uniformly ^{13}C- and ^{15}N-labeled NS5A fragment comprising residues 191 to 369 [NS5A(191–369)] of genotype 1b (strain HC-J4) were done as described recently.[37] [*U*-^{13}C, *U*-^{15}N]-NS5A(191–369)-containing samples were stored in 50 mM potassium phosphate buffer (pH 6.5), 20 mM NaCl, and 2 mM β-mercaptoethanol, at −20 °C.

Expression and purification of Bin1–SH3

The gene region coding for Bin1–SH3 domain comprising amino acid residues 517 to 593 was subcloned via PCR from the vector IRAUp969F0624D (Image3688734; RZPD, Berlin, Germany) into expression vector pGEX-6P-2 via

BamHI and XhoI restriction sites and verified by DNA sequencing. Protein expression, purification, and cleavage of the glutathione *S*-transferase–Bin1–SH3 fusion protein followed a protocol described previously for other SH3 domains.[61]

NMR spectroscopy

NMR measurements were performed on Agilent VNMRS 600-MHz (^{15}N relaxation experiments) or 800-MHz (NMR assignment, PREs, and SH3 titration experiments) systems equipped with cryogenically cooled triple-resonance probes and pulsed z-field gradients. For all 2D H–N and 3D H–N–C correlation experiments, BEST-TROSY optimized pulse sequences[35] were used. First, a series of ^{1}H–^{15}N correlation spectra of NS5A(191–369) were recorded (Fig. S1) as a function of sample temperature (5–25 °C). For the subsequent NMR assignment, ^{15}N relaxation, PRE measurements, and SH3 titration experiments, the sample temperature was set to 5 °C, as the highest spectral quality (number of resolved peaks) was obtained at low temperature. To confirm that the results obtained at low temperature are also relevant for higher temperatures, closer to physiological conditions, some of the NMR experiments were repeated at 20 °C.

NMR data collection for sequential resonance assignment and ^{15}N relaxation experiments were performed on an NMR sample containing 130 μM [*U*-^{13}C, *U*-^{15}N]-enriched NS5A(191–369) in 50 mM potassium phosphate buffer (pH 6.5), 20 mM NaCl, 2 mM β-mercaptoethanol, and 5% (v/v) D_2O. The sample was put into a Shigemi tube and stored at 4 °C between the NMR data recordings. A few weeks after sample preparation, additional peaks appeared in the spectrum, and a fresh sample was prepared for further NMR experiments.

To achieve sequential resonance assignment, we recorded a data set consisting of seven triple-resonance 3D correlation experiments: 3D HNCO, 3D HNcoCA, 3D iHNCA, 3D HNcoCACB, 3D iHNCACB, 3D hNcocaNH, and 3D hnCOcaNH. The BEST-TROSY version of these experiments combined with semi-CT frequency editing in the ^{15}N dimension allowed for high spectral resolution and high sensitivity in a short overall experimental time ranging from a few hours for the most sensitive experiments to ~2 days for the less sensitive ones. In addition, a HADAMAC spectrum was recorded for efficient amino-acid-type discrimination purposes.[36] More experimental details can be found in Table S1.

To further characterize the conformational dynamics of NS5A(191–369), we performed ^{15}N relaxation experiments (T_1, T_2, HETNOE) on the ^{13}C-, ^{15}N-labeled sample using standard pulse sequences.[62] Experimental details are provided in the Table S2. Relaxation rates for a total of 93 and 83 backbone amide sites could be obtained at 5 °C and 20 °C, respectively, from cross peaks that were well resolved in the 2D ^{1}H–^{15}N spectra.

All NMR data were processed using NMRPipe[63] and evaluated with the NMRView software.[64]

Paramagnetic relaxation enhancement

For PRE experiments, two single-cysteine mutants (NS5A_C243S and NS5A_C342S) of WT NS5A(191–369) were produced using QuikChange Site-Directed Mutagenesis Kit (Stratagene) and verified by DNA sequencing. Expression and purification of both mutants were performed as described for WT NS5A(191–369). The paramagnetic nitroxide spin label MTSL (Toronto Research Chemicals, Toronto) was attached to the thiol group of the remaining cysteine by the following procedure. A 10-fold excess of DTT was added to the protein sample and incubated for 3 h at room temperature to ensure that cysteines were fully reduced. DTT was then removed by buffer exchange with a NAP-10 column (GE Healthcare) to 50 mM potassium phosphate (pH 6.5) and 20 mM NaCl. A fivefold excess of MTSL dissolved in methanol was added and the mixture was incubated overnight at 4 °C. Removing free MTSL and sample concentration was performed by buffer exchange with 1.5-ml concentrators (Vivaspin500, Sartorius).

To avoid any effect from intermolecular paramagnetic relaxation, we detected PRE effects on diluted samples of 35 μM MTSL-labeled NS5A_C243S and NS5A_C342S, respectively, in 50 mM potassium phosphate buffer (pH 6.5), 20 mM NaCl, and 5% (v/v) D_2O. Diamagnetic references were obtained by adding a 10-fold excess of ascorbic acid to the MTSL-labeled sample and subsequent incubation for 5 h at 4 °C. 2D ^{1}H–^{15}N BEST-TROSY experiments of both mutants were recorded with MTSL in both paramagnetic and diamagnetic forms. The effect of PRE was quantitatively analyzed by the ratio of corresponding peak intensities $I_{\mathrm{para}}/I_{\mathrm{dia}}$.

PRE-based structure calculations

The two sets of measured PREs were normalized and converted into intraresidue distance restraints between the sulfur atom of the cysteine residue with the attached MTSL label and the amide proton of the residue for which a PRE was measured. In order to do so, we exploited the r^{-6} dependence of the measured intensity ratios assuming that a single conformation accounts for the observed effects. For the conversion, we chose somehow arbitrarily an effective correlation time of 4 ns for the relaxation-active electron–nuclear interaction and neglected differences in diamagnetic relaxation throughout the peptide chain. This resulted in internuclear distance restraints ranging from 10 Å (PRE ~0.1) to 25 Å (PRE ~0.9). For PREs larger than 0.9, the minimal distance was set to 25 Å, and no upper distance restraint was used, while for PREs smaller than 0.2, the upper distance limit was set to 10 Å. For all others, the lower and upper distance restraints were determined from the measured intensity ratios and error margins.

Structure calculations were performed using CYANA version 2.1,[65] following a protocol adapted from a recent work of Mukrasch *et al.* on protein tau.[66] A total of 1000 random structures were generated and minimized restraining the Ψ and Φ dihedral angles of the residues located in the three helical regions of the protein to an α-helical conformation and applying a lower-bound distance restraint of 500 Å between the C^{α} atoms of the N-terminal and C-terminal residues in order to get extended conformations. A total of 390 PRE-derived upper- and lower-bound distance restraints were then applied during a torsion-angle dynamics simulated-annealing protocol of 10,000 steps during which the Ψ and Φ dihedral angles in the helical regions were maintained fixed.

Small-angle X-ray scattering

SAXS data were measured at the DESY X33 beamline at the EMBL Hamburg. Scattering data were recorded at 20 °C for two samples of NS5A(191–369) with concentrations of 90 μM and 50 μM and identical sample conditions to the NMR experiments (buffer) using an overall exposure time of 2 min. After normalization to the intensity of the transmitted beam, the scattering of the buffer was subtracted, and the data were scaled according to the measured concentrations using the program PRIMUS. The EOM analysis[41] was performed using the web interface of the EMBL Hamburg. EOM uses a genetic algorithm that selects an ensemble of 20 structures from a pool of 10,000 random conformations generated by the program RANCH that reproduces the scattering data. The calculation was repeated 100 times, and a distribution of the radii of gyration (R_g) of all structures in the selected ensembles was created as the final result of the analysis.

Mapping of chemical shift perturbations by NMR titration experiments

The interaction of NS5A(191–369) and Bin1–SH3 was studied by mapping the ^{1}H and ^{15}N chemical shift changes upon adding increasing amounts of unlabeled ligand (Bin1–SH3). To ensure identical buffer conditions, we dialyzed solutions of *U*-^{13}C, *U*-^{15}N-labeled NS5A(191–369) and unlabeled Bin1–SH3 overnight against 50 mM potassium phosphate buffer (pH 6.5), 20 mM NaCl, and 2 mM β-mercaptoethanol at 4 °C. One hundred twenty microliters of 100 μM *U*-^{13}C, *U*-^{15}N-labeled NS5A(191–369) was filled into a 2.5-mm micro-NMR tube (New Era) and then put into a standard 5-mm NMR tube filled with D_2O. Unlabeled Bin1–SH3 from a 1-mM stock solution was added to the NMR sample with increasing molar ratios and mixed by using an eVol pipetting device (SGE Analytical Science). 2D ^{1}H–^{15}N BEST-TROSY spectra were recorded at 5 °C at molar ratios [Bin1–SH3]/[NS5A(191–369)] of 0, 0.2, 0.5, 1, 1.5, 2, 5, and 10. The total chemical shift changes $\Delta_{\text{total}}\delta(H^N,N)$ were calculated according to the following equation: $\Delta_{\text{total}}\delta(H^N,N) = ((10*\Delta\delta H^N)^2 + (\Delta\delta N)^2)^{1/2}$ in ppm, where $\Delta\delta H^N$ and $\Delta\delta N$ are the chemical shift changes in the proton and nitrogen dimensions, respectively.

Acknowledgements

The authors are grateful to Isabel Ayala, Ombeline Pessey, and Lionel Imbert for help in protein production. We also acknowledge access to the EMBL X33 beamline at the DORIS storage ring, Deutsches Elektronen-Synchrotron, Hamburg, and we thank Clement Blanchet, Giancarlo Tria, and Dmitri Svergun for technical support and assistance with data analysis. This work was supported by the Commisariat à lEnergie Atomique et aux Energies Alternatives, the Centre National de la Recherche Scientifique, the University Grenoble1, and the Deutsche Forschungsgemeinschaft (SFB575), and by a European Marie-Curie grant (FP7-ITN-IDP-byNMR contract No. 264257).

Supplementary Data

Supplementary data to this article can be found online at doi:10.1016/j.jmb.2012.04.023

References

1. Poynard, T., Yuen, M. F., Ratziu, V. & Lai, C. L. (2003). Viral hepatitis C. *Lancet*, **362**, 2095–2100.
2. Moradpour, D., Penin, F. & Rice, C. M. (2007). Replication of hepatitis C virus. *Nat. Rev., Microbiol.* **5**, 453–463.
3. Tanji, Y., Kaneko, T., Satoh, S. & Shimotohno, K. (1995). Phosphorylation of hepatitis C virus-encoded nonstructural protein NS5A. *J. Virol.* **69**, 3980–3986.
4. Foster, T. L., Belyaeva, T., Stonehouse, N. J., Pearson, A. R. & Harris, M. (2010). All three domains of the hepatitis C virus nonstructural NS5A protein contribute to RNA binding. *J. Virol.* **84**, 9267–9277.
5. Pfannkuche, A., Böther, K., Karthe, J., Poenisch, M., Bartenschlager, R., Trilling, M. *et al.* (2011). c-Src is required for complex formation between the hepatitis C virus-encoded proteins NS5A and NS5B: a prerequisite for replication. *Hepatology*, **53**, 1127–1136.
6. Reiss, S., Rebhan, I., Backes, P., Romero-Brey, I., Erfle, H., Matula, P. *et al.* (2011). Recruitment and activation of a lipid kinase by hepatitis C virus NS5A is essential for integrity of the membranous replication compartment. *Cell Host Microbe*, **9**, 32–45.
7. Penin, F., Brass, V., Appel, N., Ramboarina, S., Montserret, R., Ficheux, D. *et al.* (2004). Structure and function of the membrane anchor domain of hepatitis C virus nonstructural protein 5A. *J. Biol. Chem.* **279**, 40835–40843.
8. Love, R., Brodsky, O., Hickey, M. J., Wells, P. A. & Cronin, C. N. (2009). Crystal structure of a novel dimeric form of NS5A domain I protein from hepatitis C virus. *J. Virol.* **83**, 4395–4403.
9. Liang, Y., Ye, H., Kang, C. B. & Yoon, H. S. (2007). Domain 2 of nonstructural protein 5A (NS5A) of hepatitis C virus is natively unfolded. *Biochemistry*, **46**, 11550–11558.
10. Hanoulle, X., Badillo, A., Verdegem, D., Penin, F. & Lippens, G. (2010). The domain 2 of the HCV NS5A protein is intrinsically unstructured. *Protein Pept. Lett.* **17**, 1012–1018.
11. Kuiken, C., Combet, C., Bukh, J., Shin-I, T., Deleage, G., Mizokami, M. *et al.* (2006). A comprehensive system for consistent numbering of HCV sequences, proteins and epitopes. *Hepatology*, **44**, 1355–1361.
12. Dosztanyi, Z., Csizmok, V., Tompa, P. & Simon, I. (2005). IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. *Bioinformatics*, **21**, 3433–3434.
13. Tellinghuisen, T. L., Marcotrigiano, J., Gorbalenya, A. E. & Rice, C. M. (2004). The NS5A protein of hepatitis C virus is a zinc metalloprotein. *J. Biol. Chem.* **279**, 48576–48587.

14. Tellinghuisen, T. L., Marcotrigiano, J. & Rice, C. M. (2005). Structure of the zinc-binding domain of an essential component of the hepatitis C virus replicase. *Nature*, **435**, 374–379.
15. Hanoulle, X., Verdegem, D., Badillo, A., Wieruszeski, J. M., Penin, F. & Lippens, G. (2009). Domain 3 of non-structural protein 5A from hepatitis C virus is natively unfolded. *Biochem. Biophys. Res. Commun.* **381**, 634–638.
16. Wright, P. E. & Dyson, H. J. (2009). Linking folding and binding. *Curr. Opin. Struct. Biol.* **19**, 31–38.
17. Tsai, C. J., Ma, B. Y., Sham, Y. Y., Kumar, S. & Nussinov, R. (2001). Structured disorder and conformational selection. *Proteins: Struct. Funct. Genet.* **44**, 418–427.
18. Espinoza-Fonseca, L. M. (2009). Reconciling binding mechanisms of intrinsically disordered proteins. *Biochem. Biophys. Res. Commun.* **382**, 479–482.
19. de Chassey, B., Navratil, V., Tafforeau, L., Hiet, M. S., Aublin-Gex, A., Agaugue, S. *et al.* (2008). Hepatitis C virus infection protein network. *Mol. Syst. Biol.* **4**, 230.
20. He, Y., Nakao, H., Tan, S. L., Polyak, S. J., Neddermann, P., Vijaysri, S. *et al.* (2002). Subversion of cell signaling pathways by hepatitis C virus nonstructural 5A protein via interaction with Grb2 and P85 phosphatidylinositol 3-kinase. *J. Virol.* **76**, 9207–9217.
21. Macdonald, A., Crowder, K., Street, A., McCormick, C. & Harris, M. (2004). The hepatitis C virus NS5A protein binds to members of the Src family of tyrosine kinases and regulates kinase activity. *J. Gen. Virol.* **85**, 721–729.
22. Masumi, A., Aizaki, H., Suzuki, T., DuHadaway, J. B., Prendergast, G. C., Komuro, K. & Fukazawa, H. (2005). Reduction of hepatitis C virus NS5A phosphorylation through its interaction with amphiphysin II. *Biochem. Biophys. Res. Commun.* **336**, 572–578.
23. Nanda, S. K., Herion, D. & Liang, T. J. (2006). The SH3 binding motif of HCV NS5A protein interacts with Bin1 and is important for apoptosis and infectivity. *Gastroenterology*, **130**, 794–809.
24. Street, A., Macdonald, A., Crowder, K. & Harris, M. (2004). The Hepatitis C virus NS5A protein activates a phosphoinositide 3-kinase-dependent survival signaling cascade. *J. Biol. Chem.* **279**, 12232–12241.
25. Tan, S. L., Nakao, H., He, Y., Vijaysri, S., Neddermann, P., Jacobs, B. L. *et al.* (1999). NS5A, a nonstructural protein of hepatitis C virus, binds growth factor receptor-bound protein 2 adaptor protein in a Src homology 3 domain/ligand-dependent manner and perturbs mitogenic signaling. *Proc. Natl Acad. Sci. USA*, **96**, 5533–5538.
26. Zhang, H. Y., Neal, S. & Wishart, D. S. (2003). RefDB: a database of uniformly referenced protein chemical shifts. *J. Biomol. NMR*, **25**, 173–195.
27. Schwarzinger, S., Kroon, G. J., Foss, T. R., Chung, J., Wright, P. E. & Dyson, H. J. (2001). Sequence-dependent correction of random coil NMR chemical shifts. *J. Am. Chem. Soc.* **123**, 2970–2978.
28. Marsh, J. A., Singh, V. K., Jia, Z. C. & Forman-Kay, J. D. (2006). Sensitivity of secondary structure propensities to sequence differences between alpha- and gamma-synuclein: implications for fibrillation. *Protein Sci.* **15**, 2795–2804.
29. Munoz, V. & Serrano, L. (1994). Elucidating the folding problem of helical peptides using empirical parameters. *Nat. Struct. Biol.* **1**, 399–409.
30. Lacroix, E., Viguera, A. R. & Serrano, L. (1998). Elucidating the folding problem of alpha-helices: local motifs, long-range electrostatics, ionic-strength dependence and prediction of NMR parameters. *J. Mol. Biol.* **284**, 173–191.
31. Schanda, P., Van Melckebeke, H. & Brutscher, B. (2006). Speeding up three-dimensional protein NMR experiments to a few minutes. *J. Am. Chem. Soc.* **128**, 9042–9043.
32. Lescop, E., Schanda, P. & Brutscher, B. (2007). A set of BEST triple-resonance experiments for time-optimized protein resonance assignment. *J. Magn. Reson.* **187**, 163–169.
33. Schanda, P. (2009). Fast-pulsing longitudinal relaxation optimized techniques: enriching the toolbox of fast biomolecular NMR spectroscopy. *Prog. NMR Spectrosc.* **55**, 238–265.
34. Farjon, J., Boisbouvier, J., Schanda, P., Pardi, A., Simorre, J. P. & Brutscher, B. (2009). Longitudinal relaxation enhanced NMR experiments for the study of nucleic acids in solution. *J. Am. Chem. Soc.* **131**, 8571–8577.
35. Favier, A. & Brutscher, B. (2011). Recovering lost magnetization: polarization enhancement in biomolecular NMR. *J. Biomol. NMR*, **49**, 9–15.
36. Lescop, E., Rasia, R. & Brutscher, B. (2008). Hadamard amino-acid-type edited NMR experiment for fast protein resonance assignment. *J. Am. Chem. Soc.* **130**, 5014–5015.
37. Feuerstein, S., Solyom, Z., Aladag, A., Hoffmann, S., Willbold, D. & Brutscher, B. (2011). ^{1}H, ^{13}C, and ^{15}N resonance assignment of a 179 residue fragment of hepatitis C virus non-structural protein 5A. *Biomol. NMR Assign.* **5**, 241–243.
38. Parker, M. H. & Hefford, M. A. (1997). A consensus residue analysis of loop and helix-capping residues in four-alpha-helical-bundle proteins. *Protein Eng.* **10**, 487–496.
39. Brutscher, B., Bruschweiler, R. & Ernst, R. R. (1997). Backbone dynamics and structural characterization of the partially folded A state of ubiquitin by H-1, C-13, and N-15 nuclear magnetic resonance spectroscopy. *Biochemistry*, **36**, 13043–13053.
40. Bernado, P. & Svergun, D. I. (2012). Structural analysis of intrinsically disordered proteins by small-angle X-ray scattering. *Mol. Biosyst.* **8**, 151–167.
41. Bernado, P., Mylonas, E., Petoukhov, M. V., Blackledge, M. & Svergun, D. I. (2007). Structural characterization of flexible proteins using small-angle X-ray scattering. *J. Am. Chem. Soc.* **129**, 5656–5664.
42. Macdonald, A., Mazaleyrat, S., McCormick, C., Street, A., Burgoyne, N. J., Jackson, R. M. *et al.* (2005). Further studies on hepatitis C virus NS5A–SH3 domain interactions: identification of residues critical for binding and implications for viral RNA replication and modulation of cell signalling. *J. Gen. Virol.* **86**, 1035–1044.
43. Shelton, H. & Harris, M. (2008). Hepatitis C virus NS5A protein binds the SH3 domain of the Fyn tyrosine kinase with high affinity: mutagenic analysis of residues within the SH3 domain that contribute to the interaction. *Virol. J.* **5**, 24.
44. Zech, B., Kurtenbach, A., Krieger, N., Strand, D., Blencke, S., Morbitzer, M. *et al.* (2003). Identification

and characterization of amphiphysin II as a novel cellular interaction partner of the hepatitis C virus NS5A protein. *J. Gen. Virol.* **84**, 555–560.

45. Verdegem, D., Badillo, A., Wieruszeski, J. M., Landrieu, I., Leroy, A., Bartenschlager, R. *et al.* (2011). Domain 3 of NS5A protein from the hepatitis C virus has intrinsic alpha-helical propensity and is a substrate of cyclophilin A. *J. Biol. Chem.* **286**, 20441–20454.
46. Ren, R. B., Mayer, B. J., Cicchetti, P. & Baltimore, D. (1993). Identification of a 10-amino acid proline-rich SH3 binding site. *Science*, **259**, 1157–1161.
47. Musacchio, A. (2002). How SH3 domains recognize proline. *Adv. Protein Chem.* **61**, 211–268.
48. Agrawal, V. & Kishan, K. V. R. (2002). Promiscuous binding nature of SH3 domains to their target proteins. *Protein Pept. Lett.* **9**, 185–193.
49. Nobuhisa, I., Takeya, R., Ogura, K., Ueno, N., Kohda, D., Inagaki, F. & Sumimoto, H. (2006). Activation of the superoxide-producing phagocyte NADPH oxidase requires co-operation between the tandem SH3 domains of p47(phox) in recognition of a polyproline type II helix and an adjacent alpha-helix of p22(phox). *Biochem. J.* **396**, 183–192.
50. Kami, K., Takeya, R., Sumimoto, H. & Kohda, D. (2002). Diverse recognition of non-PxxP peptide ligands by the SH3 domains from p67(phox), Grb2 and Pex13p. *EMBO J.* **21**, 4268–4276.
51. Douangamath, A., Filipp, F. V., Klein, A. T. J., Barnett, P., Zou, P. J., Voorn-Brouwer, T. *et al.* (2002). Topography for independent binding of alpha-helical and PPII-helical ligands to a peroxisomal SH3 domain.. *Mol. Cell*, **10**, 1007–1017.
52. Beltrao, P. & Serrano, L. (2005). Comparative genomics and disorder prediction identify biologically relevant SH3 protein interactions. *PLoS Comput. Biol.* **1**, 202–211.
53. Barnett, P., Bottger, G., Klein, A. T. J., Tabak, H. F. & Distel, B. (2000). The peroxisomal membrane protein Pex13p shows a novel mode of SH3 interaction. *EMBO J.* **19**, 6382–6391.
54. Casal, E., Federici, L., Zhang, W., Fernandez-Recio, J., Priego, E. M., Miguel, R. N. *et al.* (2006). The crystal structure of the BAR domain from human Bin1/Amphiphysin II and its implications for molecular recognition. *Biochemistry*, **45**, 12917–12928.
55. Tan, S. L. & Katze, M. G. (2001). How hepatitis C virus counteracts the interferon response: the jury is still out on NS5A. *Virology*, **284**, 1–12.
56. Ide, Y., Zhang, L. W., Chen, M., Inchauspe, G., Bahl, C., Sasaguri, Y. & Padmanabhan, R. (1996). Characterization of the nuclear localization signal and subcellular distribution of hepatitis C virus nonstructural protein NS5A. *Gene*, **182**, 203–211.
57. Evans, M. J., Rice, C. M. & Goff, S. P. (2004). Phosphorylation of hepatitis C virus nonstructural protein 5A modulates its protein interactions and viral RNA replication. *Proc. Natl Acad. Sci. USA*, **101**, 13038–13043.
58. Goh, P. Y., Tan, Y. J., Lim, S. P., Lim, S. G., Tan, Y. H. & Hong, W. J. (2001). The hepatitis C virus core protein interacts with NS5A and activates its caspase-mediated proteolytic cleavage. *Virology*, **290**, 224–236.
59. Hwang, J., Huang, L. Y., Cordek, D. G., Vaughan, R., Reynolds, S. L., Kihara, G. *et al.* (2010). Hepatitis C virus nonstructural protein 5A: biochemical characterization of a novel structural class of RNA-binding proteins. *J. Virol.* **84**, 12480–12491.
60. Foster, T. L., Gallay, P., Stonehouse, N. J. & Harris, M. (2011). Cyclophilin A interacts with domain II of hepatitis C virus NS5A and stimulates RNA binding in an isomerase-dependent manner. *J. Virol.* **85**, 7460–7464.
61. Schmidt, H., Hoffmann, S., Tran, T., Stoldt, M., Stangler, T., Wiesehan, K. & Willbold, D. (2007). Solution structure of a Hck SH3 domain ligand complex reveals novel interaction modes. *J. Mol. Biol.* **365**, 1517–1532.
62. Farrow, N. A., Muhandiram, R., Singer, A. U., Pascal, S. M., Kay, C. M., Gish, G. *et al.* (1994). Backbone dynamics of a free and a phoshpopeptide-complexed Src homology-2 domain studied by N15 NMR relaxation. *Biochemistry*, **33**, 5984–6003.
63. Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J. & Bax, A. (1995). NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR*, **6**, 277–293.
64. Johnson, B. A. (2004). Using NMRView to visualize and analyze the NMR spectra of macromolecules. *Methods Mol. Biol.* **278**, 313–352.
65. Güntert, P., Mumenthaler, C. & Wuthrich, K. (1997). Torsion angle dynamics for NMR structure calculation with the new program DYANA. *J. Mol. Biol.* **273**, 283–298.
66. Mukrasch, M. D., Bibow, S., Korukottu, J., Jeganathan, S., Biernat, J., Griesinger, C. *et al.* (2009). Structural polymorphism of 441-residue tau at single residue resolution. *PLoS Biol.* **7**, 399–414.

BIOCHEMISTRY
including biophysical chemistry & molecular biology

Article

pubs.acs.org/biochemistry

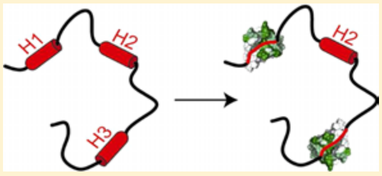
Interaction of Nonstructural Protein 5A of the Hepatitis C Virus with Src Homology 3 Domains Using Noncanonical Binding Sites

Melanie Schwarten,[†,‡,§,∥] Zsófia Sólyom,[†,‡,§] Sophie Feuerstein,[†,‡,§] Amine Aladağ,[∥,⊥] Silke Hoffmann,[∥] Dieter Willbold,[‡,∥,⊥] and Bernhard Brutscher*[,†,‡,§]

[†]Institut de Biologie Structurale, Université Grenoble 1, 41 rue Jules Horowitz, 38027 Grenoble Cedex 1, France
[‡]Commissariat à l'Energie Atomique et aux Energies Alternatives (CEA), Grenoble, France
[§]Centre National de Recherche Scientifique (CNRS), Grenoble, France
[∥]Institute of Complex Systems (ICS-6) Structural Biochemistry, Forschungszentrum Jülich, 52425 Jülich, Germany
[⊥]Institut für Physikalische Biologie, Heinrich-Heine-Universität, 40225 Düsseldorf, Germany

Supporting Information

ABSTRACT: Src homology 3 (SH3) domains are widely known for their ability to interact with other proteins using the canonical PxxP binding motif. Besides those well-characterized interaction modes, there is an increasing number of SH3 domain-containing complexes that lack this motif. Here we characterize the interaction of SH3 domains, in particular the Bin1-SH3 domain, with the intrinsically disordered part of nonstructural protein 5A of the hepatitis C virus using noncanonical binding sites in addition to its PxxP motif. These binding regions partially overlap with regions that have previously been identified as having an increased propensity to form α-helices. Remarkably, upon interaction with the Bin1-SH3 domain, the α-helical propensity decreases and a fuzzy complex is formed.

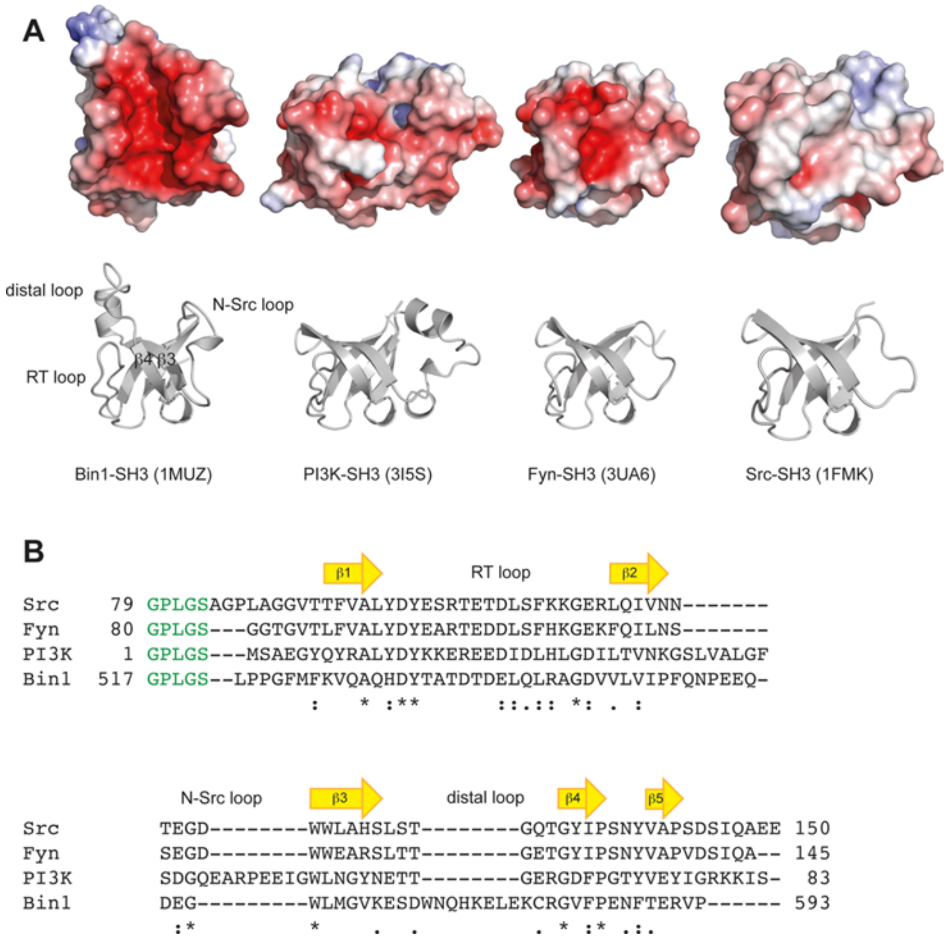


Src homology 3 domains (SH3 domains) are small, 50−80-residue protein domains that mediate protein−protein interactions. They can be found in a variety of proteins regulating dynamic cellular processes like signal transduction. Their structure is characterized by a five-stranded β-barrel with a hydrophobic cleft on the surface (Figure 1A), which is mainly formed by aromatic residues. The hydrophobic pocket recognizes left-handed polyproline type II (PPII) helices. This proline-rich sequence motif, called the PxxP motif, can interact in two orientations with the SH3 domain and can thus be classified into two distinct groups. Class I ligands have the consensus sequence +pxPpxP in common (P, conserved proline residue; p, often proline residue; +, positively charged amino acid residue; x, any amino acid residue), whereas class II ligands consist of a PpxPp+ sequence. The basic amino acid residue interacts with a negatively charged residue in the RT loop of the SH3 domain and thus defines the orientation of the ligand (for a review, see ref 1).

In addition to this canonical PxxP binding mode, SH3 domains also interact with protein partners via other, noncanonical binding modes. An increasing number of complexes between SH3 domains and peptides that lack the canonical PxxP motif have been identified and structurally characterized.[2] These nonconsensus SH3 ligands can be very diverse, ranging from the SAMP motif that interacts with the SH3 domain of DDEF1 (development- and differentiation-enhancing factor 1)[3] to the PxxDY motif that interacts with Eps8L1[4] and a (R/K)xx(R/K) motif that binds to the Gads-SH3 domain.[5,6] Recently, Perez et al.[7] reported an allosteric mechanism for the SH3 domain of c-Src that interacts via a noncanonical binding mode with its Unique domain. This interaction is prevented when a polyproline ligand is bound to the canonical PxxP binding site at the opposite surface of the SH3 domain.[7] Another motif containing mainly positively charged amino acid residues was reported to bind to the SH3 domain of bridging integrator protein 1 (Bin1),[8] a pro-apoptotic tumor suppressor, which is the major target of this study. Via its SH3 domain, Bin1 interacts with a variety of other proteins, including c-Myc and dynamin.[9,10] Furthermore, interactions of the Bin1-SH3 domain with viral proteins, including nonstructural protein 3 of alphaviruses[11] and nonstructural protein 5A (NS5A) of the hepatitis C virus, have been reported.[12−14]

Nonstructural protein 5A (NS5A) of the hepatitis C virus (HCV) is involved in a variety of viral and cellular processes. NS5A is indispensable for viral replication and particle assembly, although no direct enzymatic activity has been attributed to it. The domain structure of NS5A is shown in Figure 2A. Its N-terminal region contains an amphipathic helix, anchoring NS5A to the membranes of the endoplasmic reticulum, followed by a well-folded zinc-binding domain (D1).[15,16] The C-terminal part of NS5A, comprising domains D2 and D3, is intrinsically disordered.[17−19] The three domains are linked by so-called low-complexity sequences LCS-1 and LCS-2. LCS-2 contains a proline-rich region comprising two class II PxxP motifs (PP2.1 and PP2.2) as well as a class I motif

Received: March 21, 2013
Revised: July 12, 2013
Published: August 15, 2013



```
Src    79 GPLGSAGPLAGGVTTFVALYDYESRTETDLSFKKGERLQIVNN-------
Fyn    80 GPLGS---GGTGVTLFVALYDYEARTEDDLSFHKGEKFQILNS-------
PI3K    1 GPLGS---MSAEGYQYRALYDYKKEREEDIDLHLGDILTVNKGSLVALGF
Bin1  517 GPLGS--LPPGFMFKVQAQHDYTATDTDELQLRAGDVVLVIPFQNPEEQ-
                       :   * :**      ::.:: *: . :

Src        TEGD--------WWLAHSLST--------GQTGYIPSNYVAPSDSIQAEE 150
Fyn        SEGD--------WWEARSLTT--------GETGYIPSNYVAPVDSIQA-- 145
PI3K       SDGQEARPEEIGWLNGYNETT--------GERGDFPGTYVEYIGRKKIS-  83
Bin1       DEG---------WLMGVKESDWNQHKELEKCRGVFPENFTERVP------ 593
            :*         *    .  .         . * :* .:.
```

Figure 1. (A) Structures and electrostatic surface potentials of SH3 domains of Src, Fyn, PI3K, and Bin1 used in this study. Surface electrostatic potentials of the Bin1-SH3 [Protein Data Bank (PDB) entry 1MUZ], PI3K-SH3 (PDB entry 3I5S), Fyn-SH3 (PDB entry 3UA6), and Src-SH3 (PDB entry 1FMK) domains were calculated using the APBS tools[47] implemented in PyMOL.[48] (B) Sequence alignment of the different SH3 domains. Positions of the *β*-strands and the loop regions are indicated at the top. The five N-terminal amino acid residues that do not belong to the SH3 domain but result from cloning and PreScission cleavage are colored green.

(PP1.2). It has been shown that the PP2.2 motif is highly conserved among different HCV genotypes and can interact with a variety of SH3 domains of the Src kinase family, including Fyn, Lyn, Lck, and Hck, as well as the SH3 domains of the adaptor proteins Grb2 and Bin1.[20,21] Here we report on the results of a nuclear magnetic resonance (NMR) investigation of the interaction of NS5A with the SH3 domains of Bin1, PI3K, Fyn, and Src (Figure 1B). Our results reveal that in addition to the canonical SH3 binding to the PxxP motif, NS5A has two additional low-affinity binding sites for noncanonical SH3 binding. These noncanonical modes of binding between NS5A and the Bin1-SH3 domain have been further investigated in terms of structure and conformational dynamics.

■ MATERIALS AND METHODS

Protein Production and Purification. The NS5A fragment containing residues 191–369 [NS5A(191–369)] was expressed and purified as previously described.[22] A second NS5A fragment, NS5A(191–340), is a shorter variant of NS5A(191–369) featuring a deletion of the second LCS that contains the canonical SH3 binding motifs (see Figure 1B). It was generated by polymerase chain reaction (PCR) using NS5A(191–369) as a template with 5′-GGAGGAGGATCCCTCCGCGGCGGTGAACCGGAACC-3′ and 5′-GGAGGAGCGGCCGCTTAATGCACCACCGGCGGCACATAATC-3′ as primers. The PCR product was subcloned into pGex-6P-2 using BamHI and NotI restriction sites. Expression and prepurification of NS5A(191–340) were conducted as described for NS5A(191–369), whereas the PreScission cleavage products were subjected to a second gluthatione-Sepharose 4B affinity chromatography (GE Healthcare) step, followed by size exclusion chromatography [HighLoad SD75 16/60 (GE Healthcare)]. SH3 domains were expressed as GST fusion proteins and purified as previously described.[17,23–25] A sequence alignment of these SH3 domains is shown in Figure 1B.

NMR Spectroscopy. NMR experiments were performed on Agilent VNMRS 600 and 800 MHz spectrometers equipped with cryogenically cooled triple-resonance (HCN) probes with pulsed z-field gradients. All NMR data were processed using NMRPipe[26] and evaluated using CcpNMR.[27]

The interactions of NS5A(191–369) with the Src-, Fyn-, PI3K-, and Bin1-SH3 domains were studied as reported previously.[17] Briefly, increasing amounts of an unlabeled SH3 domain (1.0 mM stock solution) were added to [U-^{13}C;

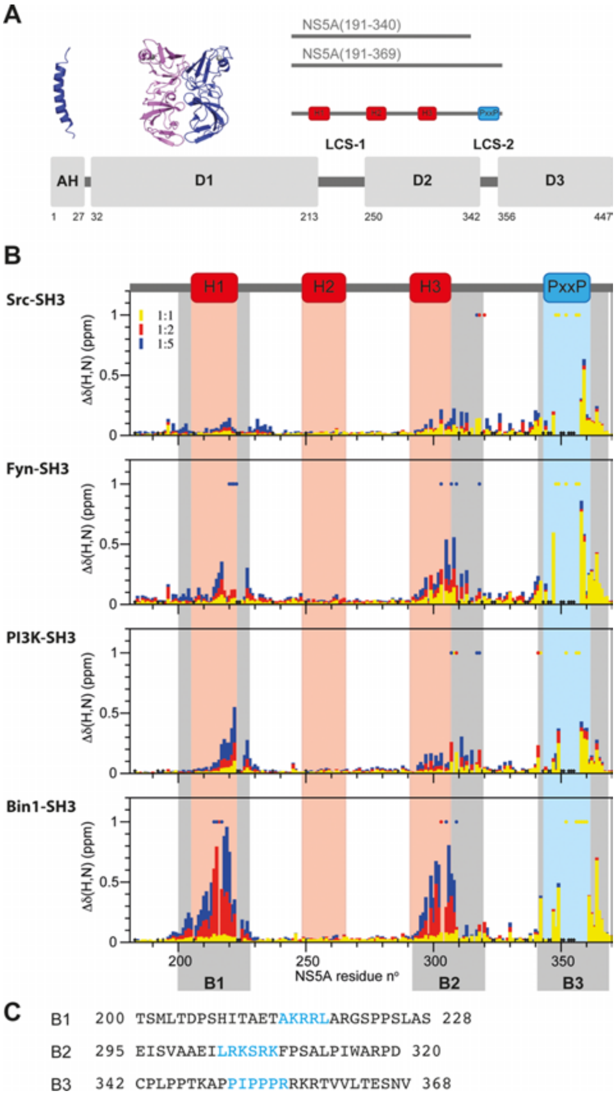


Figure 2. (A) Schematic representation of the domain organization of NS5A. Ribbon representations of the N-terminal amphipathic helix (AH, blue, PDB entry 1R7E) and dimeric globular domain 1 (D1, pink and blue, PDB entry 3FQM) are shown. Red and blue bars indicate the location of the transiently populated α-helices and the PxxP motifs, respectively. (B) Chemical shift perturbations in NS5A(191−369) upon titration with the SH3 domains of Src, Fyn, PI3K, and Bin1 are mapped on the primary sequence. On top of the histograms, the locations of the transiently formed α-helices and the PxxP motifs are shown. For a NS5A:SH3 ratio of 1:1 (yellow), chemical shift changes can mainly be detected close to the PxxP motifs. Chemical shift perturbations at ratios of 1:2 and 1:5 are colored red and blue, respectively. Residues for which no signal could be detected due to line broadening upon interaction are marked with yellow circles for the 1:1 ratio and red and blue circles for the 1:2 and 1:5 ratios, respectively. Proline residues are represented by black diamonds. (C) Amino acid sequences of the binding regions.

U-^{15}N]-NS5A(191−369) (starting concentration of 0.1 mM), and two-dimensional (2D) ^{1}H−^{15}N BEST-TROSY[28] spectra were recorded at [NS5A(191−369)]:[Bin1-SH3] molar ratios of 1:0, 1:0.5, 1:1, 1:2, and 1:5 at 5 °C. To identify the binding interfaces of NS5A(191−340) and Bin1-SH3, chemical shift changes upon addition of Bin1-SH3 to NS5A(191−340) were mapped. Experiments were conducted in a single titration series by adding increasing amounts of U-^{15}N-labeled Bin1-SH3 (1.54 mM stock solution) to [U-^{13}C; U-^{15}N]-NS5A(191−340) (starting concentration of 0.16 mM). 2D BEST-TROSY-HNco experiments,[29] which allow the separation of signals belonging to [U-^{13}C; U-^{15}N]-NS5A(191−340) from those belonging to [U-^{15}N]-NS5A Bin1-SH3, were conducted at 5 °C at [NS5A(191−340)]:[Bin1-SH3] molar ratios of 1:0, 1:0.5, 1:1, 1:2, 1:5, 1:10, and 0:1. The chosen NS5A concentration is close to the solubility limit of NS5A. Therefore, we did not attempt to record Bin1-SH3 spectra in the presence of a large excess of NS5A. The weighted chemical shift changes, $\Delta\delta$(H,N), were calculated using the following equation: $\Delta\delta(\mathrm{H,N}) = \{[10\Delta\delta(\mathrm{H})]^2 + \Delta\delta(\mathrm{N})^2\}^{1/2}$. The dissociation constants K_d^1 (for binding of SH3 to B1) and K_d^2 (for binding of SH3 B2) were then estimated by fitting the data to the equations $\Delta\delta_{\mathrm{B1}} = \Delta\delta_{\max}([\mathrm{PL_1}] + [\mathrm{PL_1L_2}])/[\mathrm{P_0}]$ and $\Delta\delta_{\mathrm{B2}} = \Delta\delta_{\max}([\mathrm{PL_2}] + [\mathrm{PL_1L_2}])/[\mathrm{P_0}]$, with the concentrations of free SH3 (L), free NS5A (P), SH3-bound B1 (PL$_1$), SH3-bound B2 (PL$_2$), and SH3-bound B1 and B2 (PL$_1$L$_2$) given by the equations $[\mathrm{L}] = -a/3 + [2(a^2 - 3b)^{1/2}\cos(\varphi/3)]/3$, $[\mathrm{P}] = ([\mathrm{P_0}]K_\mathrm{d}^1K_\mathrm{d}^2)/d$, $[\mathrm{PL_1}] = ([\mathrm{P_0}][\mathrm{L}]K_\mathrm{d}^2)/d$, $[\mathrm{PL_2}] = ([\mathrm{P_0}][\mathrm{L}]K_\mathrm{d}^1)/d$, $[\mathrm{PL_1L_2}] = ([\mathrm{P_0}][\mathrm{L}]^2)/d$, respectively. [P$_0$] and [L$_0$] are the total NS5A and SH3 concentrations, respectively, and the parameters a, b, c, d, and φ are given by

$$a = K_\mathrm{d}^1 + K_\mathrm{d}^2 - [\mathrm{L_0}] + 2[\mathrm{P_0}]$$

$$b = K_\mathrm{d}^1K_\mathrm{d}^2 - [\mathrm{L_0}]K_\mathrm{d}^1 - [\mathrm{L_0}]K_\mathrm{d}^2 + [\mathrm{P_0}]K_\mathrm{d}^1 + [\mathrm{P_0}]K_\mathrm{d}^2$$

$$c = -[\mathrm{L_0}]K_\mathrm{d}^1K_\mathrm{d}^2$$

$$d = K_\mathrm{d}^1K_\mathrm{d}^2 + [\mathrm{L}](K_\mathrm{d}^1 + K_\mathrm{d}^2) + [\mathrm{L}]^2$$

$$\varphi = \arccos(-2a^3 + 9ab - 27c)/\left[2\sqrt{(a^2 - 3b)^3}\right]$$

To identify Bin1-SH3 residues affected by either the B1 or B2 binding sites separately, we used the synthetic NS5A peptides NS5A(200−228) and NS5A(295−320), which were purchased as C18 reversed-phase high-performance liquid chromatography-purified products (JPT Peptide Technologies) containing an acetylated N-terminus and an amidated C-terminus. For chemical shift mapping purposes, 2D ^{1}H−^{15}N BEST-TROSY spectra of the U-^{15}N-labeled Bin1-SH3 domain in isolation and in the presence of an approximately 10-fold excess of the individual peptides were recorded.

Carbonyl (CO) and C$^\alpha$ chemical shifts of NS5A(191−340) in the free and complexed state (10-fold excess of Bin1-SH3) were determined using three-dimensional BEST-TROSY HNCO and HNCA experiments.[29] Secondary chemical shifts were calculated on the basis of random-coil chemical shifts and corrected for next neighbor effects.[30,31] Secondary structure propensities were determined with the SSP program using H$^\mathrm{N}$, N, Cα, and CO chemical shifts as input data.[32]

^{15}N relaxation experiments for measuring T_1, T_2, and heteronuclear {^{1}H}−^{15}N NOE (hetNOE) values were performed at a ^{1}H frequency of 800 MHz for the free and SH3-bound NS5A(191−340) using standard pulse sequences.[33]

■ RESULTS

NS5A Binds to SH3 Domains Using Different Interaction Modes. The interaction of the intrinsically disordered central part of NS5A (residues 191−369) with the

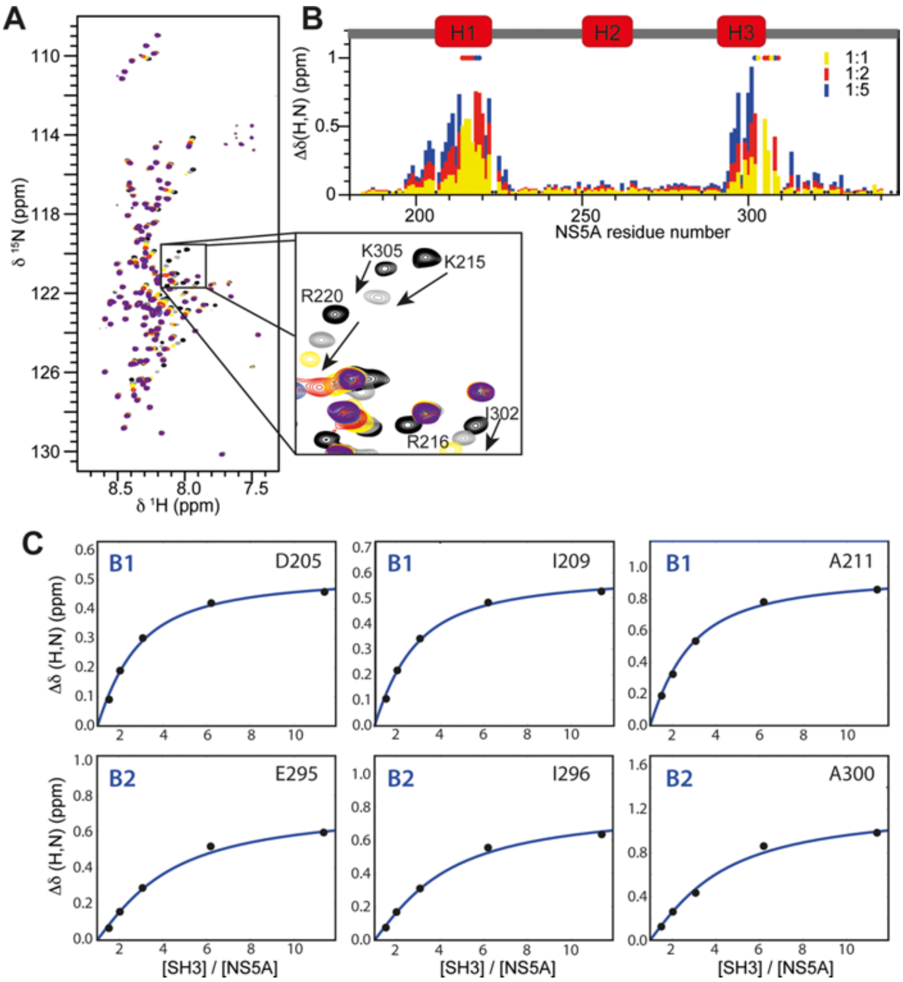


Figure 3. Interaction of NS5A(191–340) with Bin1-SH3. (A) Superposed ^{1}H–^{15}N correlation spectra of NS5A(191–340) upon titration with increasing amounts of Bin1-SH3. The different color-coded spectra correspond to [NS5A(191–369)]:[Bin1-SH3] molar ratios of 1:0 (black), 1:0.5 (gray), 1:1 (yellow), 1:2 (red), 1:5 (blue), and 1:10 (purple). (B) Chemical shift perturbations measured at NS5A(191–340):Bin1-SH3 ratios of 1:1, 1:2, and 1:5 are colored yellow, red, and blue, respectively. (C) Binding curves obtained for representative residues in the two binding regions, B1 and B2. A global fit yielded a K_d^1 of 100 ± 50 μM for the B1 region and a K_d^2 of 240 ± 50 μM for the B2 region.

SH3 domains of Src-, Fyn-, and PI3-kinases as well as the Bin1-SH3 domain was mapped using NMR titration experiments in which increasing amounts of SH3 were added to an NS5A sample. All four SH3 domains induced changes in the NMR spectra of NS5A (Figure 2B), thus indicating that NS5A interacts with each of these SH3 domains. First, addition of small amounts of SH3 led to line broadening of several NS5A residues within the PP2.2 motif [namely, I352, R356, and R357 (see Figure 2A)]. In addition, NMR signals from amino acid residues adjacent to the PP2.2 motif showed chemical shift changes during titration. Titration of NS5A with either Src-SH3, Fyn-SH3, or PI3K-SH3 at substoichiometric ratios resulted in a continuous change in the chemical shifts, indicating fast exchange kinetics on the NMR time scale, which is indicative of a comparatively low binding affinity. In contrast, as already reported previously,[17] upon titration with Bin1-SH3, separate NMR signals are detected for the SH3-bound and free states of NS5A, indicating a slow exchange process and comparatively high binding affinity.

Besides the interaction with SH3 domains mediated by a canonical PxxP motif located in the LCS-2 region of NS5A (binding site B3), we were able to identify two additional peptide regions in NS5A that interact with the four SH3 domains mentioned above (Figure 2B). Upon addition of SH3 at concentrations higher than the equimolar concentration, chemical shift changes were observed for NS5A residues 200–228 (binding site B1) and 295–320 (binding site B2). Interestingly, these binding sites comprise peptide regions that have been shown to transiently adopt α-helical structure.[17] While for the Src-SH3 domain only marginal chemical shift changes in the B1 and B2 regions were observed, possibly indicating a very low binding affinity, large chemical shift changes and broadening of some NMR signals are characteristic of the titration with the Bin1-SH3 domain. For the Fyn-SH3 and PI3K-SH3 domains, an intermediate situation was observed. Upon addition of Fyn-SH3 to NS5A, residues 214–230 showed chemical shift changes and central residues 218–221 experienced line broadening beyond the NMR detection limit at a 1:5 NS5A:SH3 ratio. Similarly, in the second binding region, chemical shift changes are observed for residues 295–320 with several signals disappearing at a 1:5 ratio. For the PI3K-SH3 domain, no severe line broadening was detected in the B1 region, but several signals in the B2 region disappeared upon titration.

Noncanonical SH3 Binding to NS5A. To further investigate the interaction of the SH3 domain with low-affinity NS5A binding sites B1 and B2, we designed a shorter NS5A construct, NS5A(191−340), that lacks the high-affinity SH3 binding site (PxxP motif). The ^{1}H−^{15}N correlation spectra of NS5A(191−340) and NS5A(191−369) almost perfectly overlap (Figure S1 of the Supporting Information), indicating that the structural and dynamic characteristics of NS5A are very similar within the two constructs. Importantly, the transient helical structures that have been identified in the longer NS5A(191−369) fragment[17] remain unaltered in NS5A(191−340), making this a suitable protein construct for studying the noncanonical binding of NS5A to SH3 domains. In the following, we will focus on the interaction of NS5A(191−340) with the Bin1-SH3 domain that induces the largest chemical shift changes in the NS5A spectrum. A very similar pattern of chemical shift changes and line broadening was observed in NMR titration experiments (Figure 3B) as shown in Figure 2B for the longer NS5A construct, indicating that the observed perturbation is due to direct SH3 binding events, and not to conformational changes induced by the high-affinity interaction of the SH3 domain with the PxxP motif of NS5A. In Figure 3C, the chemical shift changes of representative residues within the two binding regions that do not show significant line broadening are plotted as a function of the SH3:NS5A concentration ratio for the NS5A(191−340) construct. A global fit of the titration curves assuming a kinetic model with two independent binding sites, B1 and B2, yields a K_d^1 of 100 ± 50 μM for the B1 region and a K_d^2 of 240 ± 50 μM for the B2 region. As shown previously for the long NS5A(191−369) construct, the B1 site has a 2−3-fold higher binding affinity for the Bin1-SH3 domain than the B2 site. The absolute K_d values should be taken with care as the NS5A constructs have a tendency to aggregate, and it is therefore difficult to accurately estimate the binding-competent NS5A concentration.

Our titration data also provided information about the Bin1-SH3 residues that are involved in the interaction with the noncanonical low-affinity binding sites of NS5A. The measured ^{1}H and ^{15}N chemical shift changes are plotted in Figure 4A as a function of the Bin1-SH3 sequence, and in Figure 4B, they are color-coded on the structure of Bin1-SH3. Binding of SH3 to NS5A segments B1 and B2 affects mainly the residues located in the canonical binding pocket of SH3 domains for polyproline motifs, which is formed by the RT and N-Src loop as well as β-strands 3 and 4. Low- and high-affinity NS5A binding thus affects the same binding pocket on the surface of the SH3 domain. Possible differences in the modes of binding of B1 and B2 to Bin1-SH3 were further investigated by NMR titration experiments of U-^{15}N-labeled Bin1-SH3 with the unlabeled synthetic peptides B1(200−228) and B2(295−320). For both peptides, very similar chemical shift changes were obtained as in the case of the NS5A(191−340) construct containing both binding sites (Figure S2 of the Supporting Information).

Our results confirm that each SH3 domain can bind to only a single NS5A interaction site, and that the different binding modes are mutually exclusive as they all compete for the same binding pocket on the SH3 domain.

Changes in the NS5A Structural Ensemble Induced by SH3 Binding. As the two low-affinity binding regions B1 and B2 partly overlap with peptide segments that were previously identified as having increased α-helical propensities, we next investigated how SH3 binding influences the conformational sampling properties of NS5A. ^{13}C secondary chemical shifts are sensitive reporters of the local backbone geometry and are routinely used in NMR studies of proteins as a measure of secondary structural propensities along the polypeptide chain. Figure 5A shows C$^\alpha$ and CO secondary chemical shifts measured in the free and Bin1-SH3-bound forms of NS5A(191−340). In addition, the SSP score[32] that allows combination of all measured chemical shifts into a single value is plotted. A positive SSP score provides a quantitative measure of the percentage of conformers in the structural ensemble with α-helical geometry, while a negative SSP score indicates the percentage of extended local geometry. The SSP scores computed for free NS5A(191−340) are in good agreement with the previously reported results obtained for the longer NS5A(191−369) construct.[17] From the SSP scores, we can identify three NS5A segments with α-helical propensities of approximately 40% for H1 (residues 205−221) and H2 (residues 251−266) and approximately 50% for H3 (residues 292−306). As expected, only the SSP scores in the two SH3 interaction regions, B1 and B2, comprising helical segments H1 and H3, respectively, were affected by Bin1-SH3 binding. Somewhat surprisingly, the population of α-helical structure decreases upon binding: the SSP scores in segment B1 are reduced to almost zero, while for B2, they are significantly

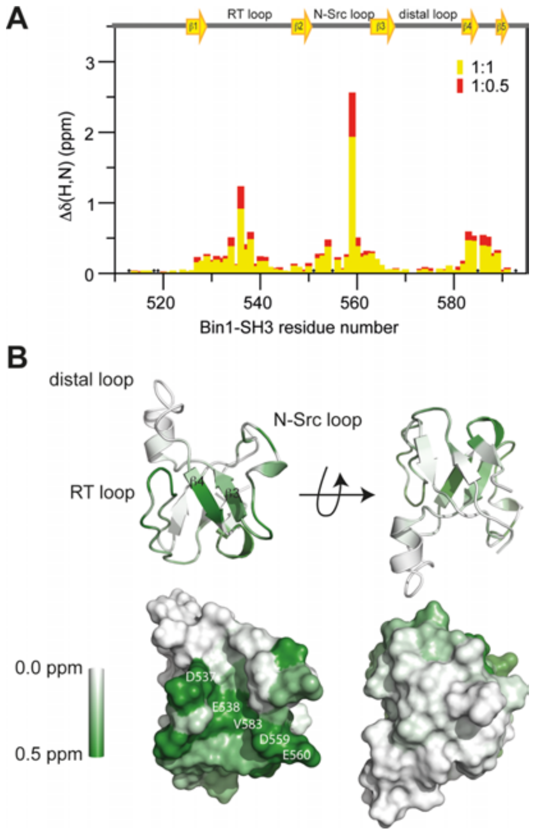


Figure 4. Chemical shift changes observed upon interaction of NS5A(191−340) with Bin1-SH3 plotted as a function of the Bin1-SH3 sequence (A) and mapped onto the Bin1-SH3 structure (B). (A) On top of the histogram, the locations of the β-strands and loop regions are shown. Chemical shift perturbations at NS5A(191−340):Bin1-SH3 ratios of 1:1 and 1:0.5 are colored yellow and red, respectively. Proline residues are represented by black diamonds. (B) Regions that interact are colored green, whereas noninteracting regions are colored white.

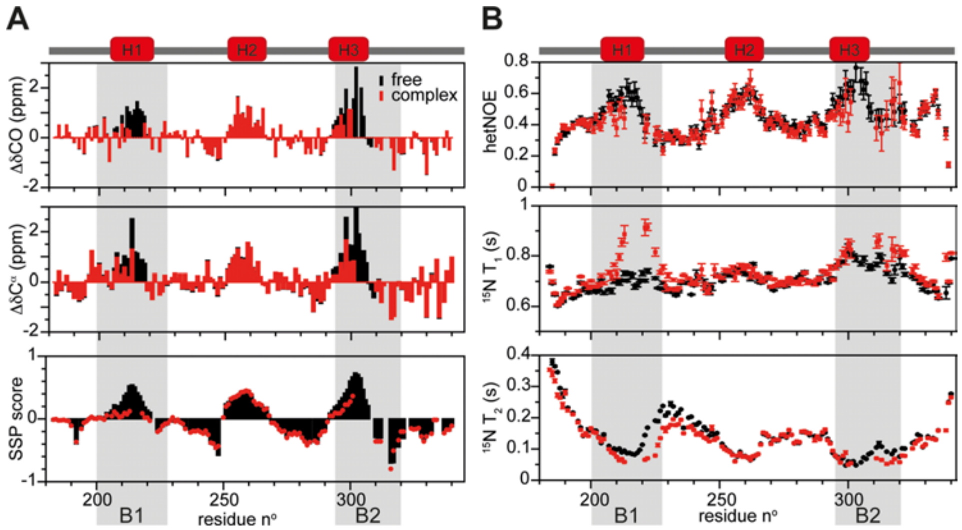


Figure 5. Structural and dynamical properties of NS5A(191−340) in its free (black) and Bin1-SH3-bound (red) states as derived from (A) CO and C^{α} secondary chemical shifts and secondary structure propensities (SSP score) and (B) ^{15}N relaxation (T_1, T_2, and hetNOE). The regions that form transient α-helices H1−H3 are indicated at the top; the Bin1-SH3 binding regions are highlighted with gray bars.

reduced (although no ^{13}C chemical shifts could be measured for residues 214−217 and 302−309 because of extensive line broadening). The observed NMR parameters, i.e., chemical shifts, are population-weighted averages. Therefore, the differences observed between B1 and B2 might be due to the lower binding affinity of B2 resulting in a reduced population of NS5A molecules in complex with SH3 at the highest NS5A:SH3 molar ratio (1:10) that was used. For B1, more than 80% of the NS5A molecules are bound to SH3, while for B2, this number is reduced to 65%.

Complementary information about the conformational dynamics of NS5A in its free state and in complex with Bin1-SH3 was obtained from ^{15}N relaxation data (T_1, T_2, and hetNOE), plotted in Figure 5B as a function of NS5A sequence. Transiently populated helical segments H1, H2 and H3 are characterized by hetNOE values higher than those of the rest of the peptide chain, indicative of an increased level of local order, as well as a longer T_1 and a shorter T_2, in agreement with an increase in the effective local tumbling correlation time, τ_C. As expected from the higher molecular weight of the complex, when SH3 binds tumbling correlation time τ_C is increased (longer T_1 and shorter T_2) in peptide regions B1 and B2 that bind to the SH3 domain. This observation most likely also explains the weak (or absent) NMR signals for the central residues in the binding site. More surprisingly, the hetNOE values of NS5A residues in the binding regions decrease upon interaction with SH3, in agreement with a more flexible structural ensemble. The interaction of NS5A with Bin1-SH3 via low-affinity binding sites B1 and B2 thus adds another example of a "fuzzy" complex[34−36] formed between a globular protein domain and an intrinsically disordered protein (IDP).

■ DISCUSSION

Viruses extensively use the molecular machinery of the host cell for either their own replication or to interfere with the cellular defense mechanisms.[37] Nonstructural protein 5A of the hepatitis C virus is a prominent example for such a protein that has multiple functions during the viral life cycle and therefore needs to interact with a variety of viral and cellular proteins.[38] NS5A belongs to the group of IDPs for which structural flexibility of the peptide chain presents a functional advantage in terms of binding promiscuity, as well as a high tolerance to mutations in the viral genome. Interactions are often meditated by short linear motifs that mimic binding sites of host proteins. A prominent example is the SH3 domain binding via PxxP motifs. Three such PxxP motifs are found in the low-complexity sequence LCS-2 of NS5A connecting domains 2 and 3.[20] Especially the second motif, PP2.2, was shown to interact with various SH3 domains.[21] In this work, we used NMR spectroscopy to characterize these interactions at an atomic level.

Canonical SH3 Binding of NS5A Mediated by its PxxP Motifs (B3). The SH3 domains of the four selected proteins, Src, Fyn, PI3K, and Bin1, all interact with NS5A via the PP2.2 motif located in the LCS-2 region, although with different binding affinities. The strongest binder is the SH3 domain of Bin1 that has the most negatively charged binding groove (see Figure 1A). As the PP2.2 motif of NS5A is followed by three positively charged residues (R357, K358, and R359), the high affinity observed for Bin1-SH3 may be explained by additional electrostatic interactions in the binding pocket that are less efficient for the SH3 domains of kinases Src, Fyn, and PI3K.

Interestingly, some of our experimental results contradict previous reports in the literature. First, we observe binding of NS5A to Src-SH3, whereas no such interaction was detected in pull-down experiments.[21] This discrepancy may be explained by the rather low affinity of NS5A for Src-SH3, and a large number of other SH3 domains present in the cell that compete for binding to NS5A. Second, the interaction of PI3K-SH3 with NS5A was reported previously to be independent of the PxxP motif in the LCS-2 region.[39,40] However, while He et al.,[39] according to their pull-down experiments, suggest that the 110 N-terminal residues are necessary for interaction, Street et al.[40] conclude from their work on deletion mutants that the NS5A region of residues 270−300 is indispensable for SH3 binding. Our NMR results contradict both of these earlier findings, clearly demonstrating that under *in vitro* conditions PI3K-SH3 behaves very much like the other SH3 domains studied here, although we cannot completely rule out possible cross-talk between binding regions B1 and B2 that is eliminated by the deletion of the intermediate fragment. In addition, under *in vivo* conditions where NS5A is attached to membranes and present in a dimeric state, cooperativity or cross-talk between the different binding sites might occur, which will influence the interaction modes. Finally, ITC measurements with Fyn-SH3 in complex with a short NS5A peptide containing the PP2.2 motif[41] as well as SPR data for the complex of Fyn-SH3 with a NS5A D2-D3 construct[42] are in agreement with a K_d in the submicromolar range, while our NMR observation of a fast exchange process between the free and bound form points toward a much lower binding affinity for this complex.

Noncanonical NS5A Binding to SH3 Domains Induces an Order-to-Disorder Transition. Here we have shown that NS5A interacts with SH3 domains via three distinct binding regions, with one (B3) corresponding to the canonical PxxP motif and the other two (B1 and B2) not containing such a canonical SH3 recognition element. Both of these additional binding sites are characterized by the presence of positively charged residues, similar to the positively charged segment in the C-terminal region of B3. These binding motifs, AKRRL for B1 and LRKSRK for B2 (Figure 2C), are similar to the (R/K)xx(R/K) motif that has been reported to bind Gads-SH3[5,6] and the (K/R)xxxxKx(K/R)(K/R) motif that also binds to the Bin1-SH3 domain.[8] In addition, as reported by Kojima et al.,[8] these positively charged peptide motifs interact with Bin1-SH3 through a surface region very similar to what is used for binding to PxxP motifs.

As noncanonical binding regions B1 and B2 correspond to a large extent to NS5A chain segments with a high propensity (40−50%) to exist in an α-helical conformation, we were also interested in structural changes induced by SH3 binding. Often, IDPs interact with their molecular targets via transiently populated secondary structural elements (α-helices or β-strands) that act as molecular recognition elements. Binding occurs via a conformational selection process in which the preformed structured conformation is stabilized in the complex, and the IDP undergoes an apparent disorder-to-order transition.[43,44] However, it has also been reported that in some cases the IDP stays disordered even when it is bound to its target protein, forming a so-called fuzzy complex.[34,35,45] One of these examples concerns NS5A, for which it has been reported [36] that domain 3 remains unstructured upon interaction with the major sperm protein (MSP) domain of VAPB forming a fuzzy complex. Here, in the case of the interaction of Bin1-SH3 with the noncanonical NS5A binding sites, we even observe an order-to-disorder transition, as the α-helical propensity and local order parameters (hetNOEs) are reduced upon complex formation. The fuzzy nature of the complex adds an entropic contribution to the binding free energy that otherwise seems to be governed by electrostatic interactions between the positively charged residues of NS5A in the binding region and the negatively charged interaction surface of Bin1-SH3. The fact that the transiently formed α-helices are not the recognition motifs for SH3 binding does not exclude the possibility that these preformed structural elements are of importance for interactions with other proteins. In that case, we may speculate that binding of NS5A to SH3 domains via these noncanonical binding sites inhibits binding to some other (still unknown) protein(s), thus contributing to the regulation of cellular processes by the virus.

At present, we can only speculate about the biological relevance of the different NS5A binding sites and modes of binding to SH3 domains. Bin1 is known to form dimers *in vivo* through its BAR domain.[46] One may thus argue that such Bin1 dimers can form a very stable complex with a single NS5A molecule by simultaneously binding to the conventional PxxP motif and one of the noncanonical binding regions. This may allow translocation of Bin1 even at low cellular NS5A concentrations. It has also been shown that the interaction of Bin1 with NS5A inhibits its phosphorylation, although the exact molecular mechanisms remain unknown.[12] A possible role of binding of Bin1-SH3 to the noncanonical binding regions in NS5A is therefore the inhibition of phosphorylation by preventing kinases from accessing the phosphorylation sites.

In conclusion, we have characterized the interaction of viral protein NS5A with SH3 domains of several kinases and tumor suppressor Bin1. Besides the well-know PxxP motif interacting with the SH3 domain, we were able to identify two additional binding regions in NS5A that bind at an SH3 surface location similar to that of the PxxP motif, mainly through electrostatic interactions. Surprisingly, the transiently formed α-helices that are present in the binding regions of free NS5A are destabilized in the NS5A−SH3 complex. These additional SH3 binding modes may be important for the regulation of molecular interaction and phosphorylation events that NS5A undergoes during the viral life cycle.

■ ASSOCIATED CONTENT

Ⓢ Supporting Information

An overlay of 2D BEST-TROSY spectra of NS5A(191−340) and NS5A(191−369) (Figure S1) and measured chemical shift changes for Bin1-SH3 in the presence of either the B1(200−228) or the B2(295−320) peptide (Figure S2). This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*Biomolecular NMR Spectroscopy group, Institut de Biologie Structurale, Jean Pierre Ebel 41, rue Jules Horowitz, 38027 Grenoble Cedex 1, France. E-mail: bernhard.brutscher@ibs.fr. Phone: +33 4 38 78 95 62. Fax: +33 4 38 78 54 94.

Funding

This work has been supported by grants from the European Commission (FP7-ITN IDPbyNMR Contract 264257) and the Deutsche Forschungsgemeinschaft (SFB974, A11).

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank Adrian Favier and Isabel Ayala for technical support and help with protein expression and Enrico Rennella for the fitting of the titration curves.

■ ABBREVIATIONS

BEST, band-selective excitation short transient; Bin1, bridging integrator protein 1; DDEF1, development- and differentiation-enhancing factor 1; HCV, hepatitis C virus; hetNOE, heteronuclear $\{^1H\}-^{15}N$ nuclear Overhauser enhancement; IDP, intrinsically disordered protein; LCS, low-complexity sequence; NS5A, nonstructural protein 5A; PDB, Protein Data Bank; PPII, polyproline type II; SH3, Src homology 3; TROSY, transverse relaxation-optimized spectroscopy.

REFERENCES

(1) Musacchio, A. (2002) How SH3 domains recognize proline. *Adv. Protein Chem. 61*, 211−268.

(2) Saksela, K., and Permi, P. (2012) SH3 domain ligand binding: What's the consensus and where's the specificity? *FEBS Lett. 586*, 2609−2614.

(3) Kaieda, S., Matsui, C., Mimori-Kiyosue, Y., and Ikegami, T. (2010) Structural basis of the recognition of the SAMP motif of adenomatous polyposis coli by the Src-homology 3 domain. *Biochemistry 49*, 5143−5153.

(4) Aitio, O., Hellman, M., Kesti, T., Kleino, I., Samuilova, O., Paakkonen, K., Tossavainen, H., Saksela, K., and Permi, P. (2008) Structural basis of PxxDY motif recognition in SH3 binding. *J. Mol. Biol. 382*, 167−178.

(5) Berry, D. M., Nash, P., Liu, S. K., Pawson, T., and McGlade, C. J. (2002) A high-affinity Arg-X-X-Lys SH3 binding motif confers specificity for the interaction between Gads and SLP-76 in T cell signaling. *Curr. Biol. 12*, 1336−1341.

(6) Liu, Q., Berry, D., Nash, P., Pawson, T., McGlade, C. J., and Li, S. S. (2003) Structural basis for specific binding of the Gads SH3 domain to an RxxK motif-containing SLP-76 peptide: A novel mode of peptide recognition. *Mol. Cell 11*, 471−481.

(7) Perez, Y., Maffei, M., Igea, A., Amata, I., Gairi, M., Nebreda, A. R., Bernado, P., and Pons, M. (2013) Lipid binding by the Unique and SH3 domains of c-Src suggests a new regulatory mechanism. *Sci. Rep. 3*, 1295.

(8) Kojima, C., Hashimoto, A., Yabuta, I., Hirose, M., Hashimoto, S., Kanaho, Y., Sumimoto, H., Ikegami, T., and Sabe, H. (2004) Regulation of Bin1 SH3 domain binding by phosphoinositides. *EMBO J. 23*, 4413−4422.

(9) Nicot, A. S., Toussaint, A., Tosch, V., Kretz, C., Wallgren-Pettersson, C., Iwarsson, E., Kingston, H., Garnier, J. M., Biancalana, V., Oldfors, A., Mandel, J. L., and Laporte, J. (2007) Mutations in amphiphysin 2 (BIN1) disrupt interaction with dynamin 2 and cause autosomal recessive centronuclear myopathy. *Nat. Genet. 39*, 1134−1139.

(10) Sakamuro, D., Elliott, K. J., Wechsler-Reya, R., and Prendergast, G. C. (1996) BIN1 is a novel MYC-interacting protein with features of a tumour suppressor. *Nat. Genet. 14*, 69−77.

(11) Neuvonen, M., Kazlauskas, A., Martikainen, M., Hinkkanen, A., Ahola, T., and Saksela, K. (2011) SH3 domain-mediated recruitment of host cell amphiphysins by alphavirus nsP3 promotes viral RNA replication. *PLoS Pathog. 7*, e1002383.

(12) Masumi, A., Aizaki, H., Suzuki, T., DuHadaway, J. B., Prendergast, G. C., Komuro, K., and Fukazawa, H. (2005) Reduction of hepatitis C virus NS5A phosphorylation through its interaction with amphiphysin II. *Biochem. Biophys. Res. Commun. 336*, 572−578.

(13) Nanda, S. K., Herion, D., and Liang, T. J. (2006) The SH3 binding motif of HCV NS5A protein interacts with Bin1 and is important for apoptosis and infectivity. *Gastroenterology 130*, 794−809.

(14) Zech, B., Kurtenbach, A., Krieger, N., Strand, D., Blencke, S., Morbitzer, M., Salassidis, K., Cotten, M., Wissing, J., Obert, S., Bartenschlager, R., Herget, T., and Daub, H. (2003) Identification and characterization of amphiphysin II as a novel cellular interaction partner of the hepatitis C virus NS5A protein. *J. Gen. Virol. 84*, 555−560.

(15) Tellinghuisen, T. L., Marcotrigiano, J., Gorbalenya, A. E., and Rice, C. M. (2004) The NS5A protein of hepatitis C virus is a zinc metalloprotein. *J. Biol. Chem. 279*, 48576−48587.

(16) Brass, V., Bieck, E., Montserret, R., Wolk, B., Hellings, J. A., Blum, H. E., Penin, F., and Moradpour, D. (2002) An amino-terminal amphipathic α-helix mediates membrane association of the hepatitis C virus nonstructural protein 5A. *J. Biol. Chem. 277*, 8130−8139.

(17) Feuerstein, S., Solyom, Z., Aladag, A., Favier, A., Schwarten, M., Hoffmann, S., Willbold, D., and Brutscher, B. (2012) Transient structure and SH3 interaction sites in an intrinsically disordered fragment of the hepatitis C virus protein NS5A. *J. Mol. Biol. 420*, 310−323.

(18) Hanoulle, X., Badillo, A., Verdegem, D., Penin, F., and Lippens, G. (2010) The domain 2 of the HCV NS5A protein is intrinsically unstructured. *Protein Pept. Lett. 17*, 1012−1018.

(19) Hanoulle, X., Verdegem, D., Badillo, A., Wieruszeski, J. M., Penin, F., and Lippens, G. (2009) Domain 3 of non-structural protein 5A from hepatitis C virus is natively unfolded. *Biochem. Biophys. Res. Commun. 381*, 634−638.

(20) Tan, S. L., Nakao, H., He, Y., Vijaysri, S., Neddermann, P., Jacobs, B. L., Mayer, B. J., and Katze, M. G. (1999) NS5A, a nonstructural protein of hepatitis C virus, binds growth factor receptor-bound protein 2 adaptor protein in a Src homology 3 domain/ligand-dependent manner and perturbs mitogenic signaling. *Proc. Natl. Acad. Sci. U.S.A. 96*, 5533−5538.

(21) Macdonald, A., Crowder, K., Street, A., McCormick, C., and Harris, M. (2004) The hepatitis C virus NS5A protein binds to members of the Src family of tyrosine kinases and regulates kinase activity. *J. Gen. Virol. 85*, 721−729.

(22) Feuerstein, S., Solyom, Z., Aladag, A., Hoffmann, S., Willbold, D., and Brutscher, B. (2011) ^{1}H, ^{13}C, and ^{15}N resonance assignment of a 179 residue fragment of hepatitis C virus non-structural protein 5A. *Biomol. NMR Assignments 5*, 241−243.

(23) Batra-Safferling, R., Granzin, J., Modder, S., Hoffmann, S., and Willbold, D. (2010) Structural studies of the phosphatidylinositol 3-kinase (PI3K) SH3 domain in complex with a peptide ligand: Role of the anchor residue in ligand binding. *Biol. Chem. 391*, 33−42.

(24) Schmidt, H., Hoffmann, S., Tran, T., Stoldt, M., Stangler, T., Wiesehan, K., and Willbold, D. (2007) Solution structure of a Hck SH3 domain ligand complex reveals novel interaction modes. *J. Mol. Biol. 365*, 1517−1532.

(25) Tran, T., Hoffmann, S., Wiesehan, K., Jonas, E., Luge, C., Aladag, A., and Willbold, D. (2005) Insights into human Lck SH3 domain binding specificity: Different binding modes of artificial and native ligands. *Biochemistry 44*, 15042−15052.

(26) Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., and Bax, A. (1995) NMRPipe: A multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR 6*, 277−293.

(27) Vranken, W. F., Boucher, W., Stevens, T. J., Fogh, R. H., Pajon, A., Llinas, M., Ulrich, E. L., Markley, J. L., Ionides, J., and Laue, E. D. (2005) The CCPN data model for NMR spectroscopy: Development of a software pipeline. *Proteins 59*, 687−696.

(28) Favier, A., and Brutscher, B. (2011) Recovering lost magnetization: Polarization enhancement in biomolecular NMR. *J. Biomol. NMR 49*, 9−15.

(29) Solyom, Z., Schwarten, M., Geist, L., Konrat, R., Willbold, D., and Brutscher, B. (2013) BEST-TROSY experiments for time-efficient sequential resonance assignment of large disordered proteins. *J. Biomol. NMR 55*, 311−321.

(30) Schwarzinger, S., Kroon, G. J., Foss, T. R., Chung, J., Wright, P. E., and Dyson, H. J. (2001) Sequence-dependent correction of random coil NMR chemical shifts. *J. Am. Chem. Soc. 123*, 2970−2978.

(31) Zhang, H., Neal, S., and Wishart, D. S. (2003) RefDB: A database of uniformly referenced protein chemical shifts. *J. Biomol. NMR 25*, 173−195.

(32) Marsh, J. A., Singh, V. K., Jia, Z., and Forman-Kay, J. D. (2006) Sensitivity of secondary structure propensities to sequence differences between α- and γ-synuclein: Implications for fibrillation. *Protein Sci. 15*, 2795−2804.

(33) Farrow, N. A., Muhandiram, R., Singer, A. U., Pascal, S. M., Kay, C. M., Gish, G., Shoelson, S. E., Pawson, T., Forman-Kay, J. D., and Kay, L. E. (1994) Backbone dynamics of a free and phosphopeptide-complexed Src homology 2 domain studied by ^{15}N NMR relaxation. *Biochemistry 33*, 5984−6003.

(34) Fuxreiter, M. (2012) Fuzziness: Linking regulation to protein dynamics. *Mol. BioSyst. 8*, 168−177.

(35) Fuxreiter, M., and Tompa, P. (2012) Fuzzy complexes: A more stochastic view of protein function. *Adv. Exp. Med. Biol. 725*, 1−14.

(36) Gupta, G., Qin, H., and Song, J. (2012) Intrinsically unstructured domain 3 of hepatitis C Virus NS5A forms a "fuzzy

complex" with VAPB-MSP domain which carries ALS-causing mutations. *PLoS One* *7*, e39261.

(37) Davey, N. E., Trave, G., and Gibson, T. J. (2011) How viruses hijack cell regulation. *Trends Biochem. Sci.* *36*, 159−169.

(38) Bode, J. G., Brenndorfer, E. D., Karthe, J., and Haussinger, D. (2009) Interplay between host cell and hepatitis C virus in regulating viral replication. *Biol. Chem.* *390*, 1013−1032.

(39) He, Y., Nakao, H., Tan, S. L., Polyak, S. J., Neddermann, P., Vijaysri, S., Jacobs, B. L., and Katze, M. G. (2002) Subversion of cell signaling pathways by hepatitis C virus nonstructural 5A protein via interaction with Grb2 and P85 phosphatidylinositol 3-kinase. *J. Virol.* *76*, 9207−9217.

(40) Street, A., Macdonald, A., Crowder, K., and Harris, M. (2004) The hepatitis C virus NS5A protein activates a phosphoinositide 3-kinase-dependent survival signaling cascade. *J. Biol. Chem.* *279*, 12232−12241.

(41) Martin-Garcia, J. M., Luque, I., Ruiz-Sanz, J., and Camara-Artigas, A. (2012) The promiscuous binding of the Fyn SH3 domain to a peptide from the NS5A protein. *Acta Crystallogr.* *D68*, 1030−1040.

(42) Shelton, H., and Harris, M. (2008) Hepatitis C virus NS5A protein binds the SH3 domain of the Fyn tyrosine kinase with high affinity: Mutagenic analysis of residues within the SH3 domain that contribute to the interaction. *Virol. J.* *5*, 24.

(43) Sugase, K., Dyson, H. J., and Wright, P. E. (2007) Mechanism of coupled folding and binding of an intrinsically disordered protein. *Nature* *447*, 1021−1025.

(44) Wright, P. E., and Dyson, H. J. (2009) Linking folding and binding. *Curr. Opin. Struct. Biol.* *19*, 31−38.

(45) Mittag, T., Kay, L. E., and Forman-Kay, J. D. (2010) Protein dynamics and conformational disorder in molecular recognition. *J. Mol. Recognit.* *23*, 105−116.

(46) Casal, E., Federici, L., Zhang, W., Fernandez-Recio, J., Priego, E. M., Miguel, R. N., DuHadaway, J. B., Prendergast, G. C., Luisi, B. F., and Laue, E. D. (2006) The crystal structure of the BAR domain from human Bin1/amphiphysin II and its implications for molecular recognition. *Biochemistry* *45*, 12917−12928.

(47) Baker, N. A., Sept, D., Joseph, S., Holst, M. J., and McCammon, J. A. (2001) Electrostatics of nanosystems: Application to microtubules and the ribosome. *Proc. Natl. Acad. Sci. U.S.A.* *98*, 10037−10041.

(48) *The PyMOL Molecular Graphics System* (2010) Schrodinger, LLC.

Résumé en français

Le but de cette thèse était double: il comprend des développements méthodologiques RMN pour les protéines intrinsèquement désordonnées et l'application de ces méthodes pour une système intrinsèquement désordonnée biologiquement important, la protein NS5A du virus hépatite C.

La première partie de cette thèse est une introduction sur les protéines désordonnées, avec une partie spécifiquement consacrée aux protéines virales. Pour les protéines virales le désordre donne un avantage particulier, résultant d'une tolérance aux mutations (permit par l'absence d'une structure stable), une évolution rapide et une adaptabilité à l'environnement plus élevé. Les protéines désordonnées sont abondantes parmi les eukaryotes. Leur abondance dans les domaines différent peut être évalué en utilisent des prédicteurs de désordre. Deux prédicteurs de désordre ont été utilisées pendant cette thèse, PONDR et IUPRED. A partir de la séquence des acides aminés ces deux prédicteurs permettent d'identifier les régions ayant une pré-disposition à être désordonné. Cependant, pour être capable d'étudier la protéine avec des méthodes biophysiques on doit produire un échantillon. Les protéines désordonnées sont habituellement produites par expression en E. Coli. Leur marquage isotopique pour les études par spectroscopie RMN est également possible aussi. Des problèmes de solubilité peuvent apparaître lors de leur expression, ces problèmes proviennent de leur tendances à agréger. Une des autres caractéristiques des protéines désordonnées est leur susceptibilité à la dégradation protéolytique, qui est causée par la nature très exposée de la chaine des acides aminés.

Différentes méthodes biophysiques peuvent être utilisées pour caractériser les protéines désordonnées. La méthode la plus souvent utilisée au début de la recherche sur les protéines désordonnées était la spectroscopie CD. Le principe de cette méthode spectroscopique est de mesurer la différence entre l'absorption d'une échantillon optiquement actif et la lumière circulairement polarisé dans les sens gauche et droite. Cette méthode permet d'identifier la proportion de structure secondaire présente dans la protéine. Une autre méthode très utile pour la caractérisation des protéines désordonnées est la diffusion de lumière statique (SLS) et dynamique (DLS). La SLS permet de déterminer la masse moléculaire des particules présentes dans la solution. La DLS permet l'identification de la différence dans la mobilité des particules et donc permet d'évaluer la polydispersité éventuelle. La diffusion de lumière rayons-X de petits angles (SAXS) permet la mesure de rayon de gyration de la molécules étant donné une monodispersité de la solution. La calorimetrie de dosage isothermale (ITC) est une méthode utile pour la caractérisation thermodynamique des interactions. Elle permet la mesure des constantes de dissociation. La méthode RMN qui donne l'information la plus détaillé sur les protéines désordonnées est introduit dans la deuxième partie de cette thèse.

La spectroscopie RMN est une méthode très puissante pour la caractérisation des biomolécules avec une résolution atomique. Cependant, elle a une sensibilité assez faible résultant des temps de mesures et/ou de la concentration d'échantillon requise élevés. Il y a quelques ans une méthode pour augmenter la sensibilité des méthodes RMN a été mise au point à l'IBS. Les séquences d'impulsion BEST et SOFAST celles-ci exploitent les fréquences de résonances particulière des spins HN qui sont généralement observés en RMN des protéines, en applicant des impulsions qui excitent selectivement ces spins. Cette méthode permet d'augmenter la vitesse de relaxation T1 apparente. Comme la relaxation est plus rapide, un temps d'attente plus court est suffisant pour le retour de la polarisation a l'équilibre. Pendant cette thèse des expériences 3D BEST-TROSY ont été dévéloppés et caracterisés pour l'attribution rapide et efficace des protéines intrinsèquement désordonnées. On a montré que les expériences BEST-TROSY permettent d'obtenir une résolution et une sensibilité plus élevé pour les protéines désordonnées que les expériences traditionnelles. En plus, pour faciliter l'attribution, une méthode, BEST-TROSY-Pro-HNcoca(N) a été mis en place. Celle-ci permet la détection sélective des pics de corrélation correspondants aux acides aminés adjacent aux prolines. Cette expérience est basée sur les déplacements chimiques distincts des azote amide des prolines par rapport à tous les autres acides aminés. L'expérience inclut un transfert sélectif d'aimantation aux noyeaux azotes des prolines, pendant que l'évolution de l'aimantation est refocusé. En calculant la différence entre deux spectres, le réference (ou l'impulsion sélective n'est pas appliqué) et le transfert (ou l'impulsion sélective est appliqué) on observe uniquement les pics de corrélation amides des acides aminés adjacent des prolines. De plus, une experience a été dévéloppé, HET$^{\text{ex}}$-BEST-TROSY, qui permet une détermination rapide des taux d'échanges des protons des amides avec le solvant pour les protéines désordonnées.

La troisième partie de cette thèse comprend les resultats qu'on a obtenu utilisent les méthodes précédemment dévéloppées pour la caractérisation d'une fragment désordonné de la protéine NS5A du virus de l'hépatite C. Cette protéin joue un rôle important dans la cycle de vie du virus, elle participe au processus de réplication et à l'assemblage de la particule virale. Cependant, les mécanismes exactes de son fonctionnement ne sont pas encore connus. Pendant cette thèse on a caractérisé la structure et dynamique des trois fragments de la protéine, dont le plus longue comprend les domaines 2 et 3 qui correspondent à la partie intrinsèquement désordonnée intégrale de cette protéine. Les predicteurs de désordre IUPRED et PONDR ont donné des scores de désordre très élevés pour le domaine 3, tandis que pour la domaine 2 les scores était a la limite de désordre et ordre. La protéine a été attribué avec des expériences de corrélation 3D BEST-TROSY qui ont été introduits dans la deuxième partie de cette thèse. La structure secondaire a été caractérisé avec un calcul des déplacement chimiques secondaires pour les noyeaux C^{α}, C', C^{β}, H^{α}. Avec le comparaison des déplacements chimiques on a identifié 4 régions avec des propensités significatifs de former des hélices alphas et une region avec des déplacements chimiques secondaires particuliers et quelques régions avec probablement des tendances faibles pour former des conformations étendues, comme des feuillet bêta. Les hélices alphas transitoires H1b et H3 comprennent des résidus positivement chargés, tandis que dans le domaine 3 la majorité des acides amines sont négativement chargés. La dynamique a été caracterisé à deux températures, qui a montré la présence d'hétérogénéité dans le domaine 3, où on aucune structure secondaire n'a pu être identifié a partir des déplacement chimiques secondaires. De plus, la comparaison des données de relaxation et déplacement chimique permet d'identifier des regions qui sont différentes, ce phénomène peut être expliqué par la présence d'interaction àlongue distance. En observant la séquence une correlation entre les charges et les interactions a pu être mis en évidence. L'ensemble de ces informations permet de montrer que ces interactions sont gouvernées par des forces électrostatiques.

L'interaction de NS5A avec le domaine SH3 de la protéine Bin1 a été étudié. Le domaine SH3 reconnaît les motif polyproline, NS5A possède 3 motifs polyproline. Les régions d'interaction entre NS5A et Bin1 SH3 ont été déterminé en réalisent une titration par RMN. De manière surprenante, des régions d'interaction sans motif polyproline ont pu être observé, mais ces régions comprennent des groupes des acides aminés chargés positivement. En résumé, jusqu'à un rapport équimolaire de NS5A et Bin1 on observe des perturbations des déplacement chimiques pour les résidues autour des motifs polyprolines avec un échange lent. De plus, on pouvait observer aussi des petit perturbations des déplacements chimiques pour presque toutes les résidues dans le domaine 3. Cependant, quand il y a un excès de Bin1 SH3, on observe des perturbations des déplacements chimiques avec un échange rapide autour des résidues positivement chargés dans les régions partiellement structurées en hélice. Pour trouver une explication pour les petits perturbations observés en domaine 3, la dynamique du complex 1:1 a été comparé avec l'état libre de NS5A. Les mesures de relaxation ont montré que le domaine 3 devient plus flexible dans le complexe que dans l'état libre. Cette observation permet d'expliquer qu'une interaction à longue distance peut être perturbé par l'interaction entre NS5A et Bin1 SH3.

NS5A étant une phosphoprotéine on a étudié sa a été étudié par spectroscopie RMN *in vitro*. Trois kinases ont été testé, CK2, CK1 et PKA. Avec le kinase CK2 on a observé 7 phosphorylations, dont 4 phosphorylation plus rapides qui ont pu être mis en évidence avec un suivi en temps réel de la phosphorylation par des éxperiences BEST-TROSY. PKA permet la phosphorylation d'une thréonine proche d'un motif polyproline. Le résultat des phosphorylations par CK2 est une perturbation des déplacement chimiques dans les régions comprenant de nombreaux résidus chargés positivement. Ce phénomène étant probablement causé par une perturbation des interactions électrostatiques.

Abstract in French

Les protéines intrinsèquement désordonnées sont caractérisées par un manque de structure 3D stable et sont biologiquements actives dans cet état. La spectroscopie RMN est la méthode de choix pour leurs études à une résolution atomiques, car la cristallographie aux rayons X ne permet pas leur étude en raison de leur caractère hautement dynamique. Cependant, l'étude par spectroscopie RMN de ces protéines est difficiles à cause du grand nombre de recouvrement entre les signaux dans le spectre résultant de l'absence d'un réseau de liaison hydrogène qui pourrait stabiliser la structure et permettre d'obtenir une dispersion des signaux plus élevé. Un autre problème est la sensibilité expérimentale car souvent le temps de mesure est limité en raison de leur prédisposition à la dégradation protéolytique. Dans la première partie de cette thèse les protéines intrinsèquement désordonnées sont introduites. La deuxième partie porte sur la spectroscopie RMN des protéines intrinsèquement désordonnées, des expériences RMN de type BEST-TROSY sont présentées et sont montrées comme étant bien adapté pour l'étude de protéines intrinsèquement désordonnées, en particulier pour celle avec une grande étendue de structure résiduelle. Des expériences 3D BEST-TROSY sont présentées pour leur attribution, une version proline-éditée permet d'aider à l'identification de ce type d'acide aminé et enfin l'expérience HETex-BEST-TROSY qui permet une mesure rapide des taux de change de solvants. Dans la troisième partie de cette thèse ces expériences RMN sont appliquées pour l'étude de la région intrinsèquement désordonnés (domaines 2 et 3) de la protéine NS5A du virus de l'hépatite C (VHC). La structure secondaire résiduel présente dans le fragment de la protéine est analysée. La comparaison des données RMN sur trois constructions de la protéine de différentes longueurs ainsi que les données de SAXS permettent l'identification des interactions transitoires à longue portée entre les différentes régions de cette protéine. En outre, les modes de liaison de ce fragment de protéine à Bin1 domaine SH3 sont analysés. Enfin, les résultats préliminaires obtenus sur l'étude de la phosphorylation de NS5A du VHC par certaines kinases, qui ont été montrées comme biologiquement pertinents, sont présentés.

Abstract in English

Intrinsically disordered proteins are characterized by a lack of a stable, 3D structure and fulfill their biological role as such. NMR spectroscopy is the method of choice for their atomic resolution studies, as X-ray crystallography is not amenable to them due to their highly dynamic character. However, NMR spectroscopic studies of these proteins are challenging, because of the high extent of signal overlap in the spectra, resulting from the absence of a hydrogen-bonding network that would lead to structuring and higher signal dispersion. A further problem is experimental sensitivity as often measurement time is limited due to their predisposition for proteolytic degradation. In the fist part of this thesis intrinsically disordered proteins are introduced. The second part focuses on NMR spectroscopy of IDPs, BEST-TROSY-type NMR methods are presented and are shown to be well suited for large IDPs, especially for those with high extent of residual structure. 3D BEST-TROSY experiments are presented for assignment, a proline-edited version for aiding amino acid-type identification, and the HETex-BEST-TROSY experiment that allows rapid measurement of solvent exchange rates. In the third part of this thesis NMR methods are applied for study of the entire intrinsically disordered region (domains 2 and 3) of NS5A protein of hepatitis C virus. The residual secondary structure in this protein fragment is analyzed. Comparison of NMR data on three protein constructs of different lengths together with SAXS data allows identification of transient long range interactions between different regions of this protein. Furthermore, the binding modes of this protein fragment to Bin1 SH3 domain are analyzed. Finally, the preliminary results obtained on investigation of phosphorylation of NS5A of HCV by certain kinases, reported to be biologically relevant, are presented.

Bibliography

[1] P J Mitchell and R Tjian. Transcriptional regulation in mammalian cells by sequence-specific DNA binding proteins. *Science (New York, N.Y.)*, 245(4916):371–8, July 1989.

[2] A.Keith Dunker, J.David Lawson, Celeste J Brown, Ryan M Williams, Pedro Romero, Jeong S Oh, Christopher J Oldfield, Andrew M Campen, Catherine M Ratliff, Kerry W Hipps, Juan Ausio, Mark S Nissen, Raymond Reeves, ChulHee Kang, Charles R Kissinger, Robert W Bailey, Michael D Griswold, Wah Chiu, Ethan C Garner, and Zoran Obradovic. Intrinsically disordered protein. *Journal of Molecular Graphics and Modelling*, 19(1):26–59, February 2001.

[3] John C. Wootton. Non-globular domains in protein sequences: Automated segmentation using complexity measures. *Computers & Chemistry*, 18(3):269–285, September 1994.

[4] P Romero, Z Obradovic, X Li, E C Garner, C J Brown, and A K Dunker. Sequence complexity of disordered protein. *Proteins*, 42(1):38–48, January 2001.

[5] P E Wright and H J Dyson. Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. *Journal of molecular biology*, 293(2):321–31, October 1999.

[6] Peter Tompa. Intrinsically unstructured proteins. *Trends in Biochemical Sciences*, 27(10):527–533, October 2002.

[7] Megan Sickmeier, Justin A Hamilton, Tanguy LeGall, Vladimir Vacic, Marc S Cortese, Agnes Tantos, Beata Szabo, Peter Tompa, Jake Chen, Vladimir N Uversky, Zoran Obradovic, and A Keith Dunker. DisProt: the Database of Disordered Proteins. *Nucleic acids research*, 35(Database issue):D786–93, January 2007.

[8] Mihaly Varadi, Simone Kosol, Pierre Lebrun, Erica Valentini, Martin Blackledge, A Keith Dunker, Isabella C Felli, Julie D Forman-Kay,

Richard W Kriwacki, Roberta Pierattelli, Joel Sussman, Dmitri I Svergun, Vladimir N Uversky, Michele Vendruscolo, David Wishart, Peter E Wright, and Peter Tompa. pE-DB: a database of structural ensembles of intrinsically disordered and of unfolded proteins. *Nucleic acids research*, 42(Database issue):D326–35, January 2014.

[9] P Romero, Z Obradovic, C Kissinger, J E Villafranca, and A K Dunker. Identifying Disordered Regions in Proteins from Amino Acid Sequence. *Proceedings of International Conference on Neural Networks (ICNN'97)*, 1, 1997.

[10] Vladimir N Uversky. Unusual biophysics of intrinsically disordered proteins. *Biochimica et biophysica acta*, 1834(5):932–51, May 2013.

[11] Zsuzsanna Dosztányi, Veronika Csizmok, Peter Tompa, and István Simon. IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. *Bioinformatics (Oxford, England)*, 21(16):3433–4, August 2005.

[12] Zsuzsanna Dosztányi, Veronika Csizmók, Péter Tompa, and István Simon. The pairwise energy content estimated from amino acid composition discriminates between folded and intrinsically unstructured proteins. *Journal of molecular biology*, 347(4):827–39, May 2005.

[13] Bo He, Kejun Wang, Yunlong Liu, Bin Xue, Vladimir N Uversky, and A Keith Dunker. Predicting intrinsic disorder in proteins: an overview. *Cell research*, 19(8):929–49, August 2009.

[14] A Keith Dunker, Marc S Cortese, Pedro Romero, Lilia M Iakoucheva, and Vladimir N Uversky. Flexible nets. The roles of intrinsic disorder in protein interaction networks. *The FEBS journal*, 272(20):5129–48, October 2005.

[15] Lilia M Iakoucheva, Predrag Radivojac, Celeste J Brown, Timothy R O'Connor, Jason G Sikes, Zoran Obradovic, and A Keith Dunker. The importance of intrinsic disorder for protein phosphorylation. *Nucleic acids research*, 32(3):1037–49, January 2004.

[16] Cathleen J Cox, Kaushik Dutta, Edward T Petri, William C Hwang, Yaqiong Lin, Steven M Pascal, and Ravi Basavappa. The regions of securin and cyclin B proteins recognized by the ubiquitination machinery are natively unfolded. *FEBS letters*, 527(1-3):303–8, September 2002.

[17] Peter Tompa and Peter Csermely. The role of structural disorder in the function of RNA and protein chaperones. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology*, 18(11):1169–75, August 2004.

[18] Peter Tompa. The interplay between structure and function in intrinsically unstructured proteins. *FEBS letters*, 579(15):3346–54, June 2005.

[19] Holger Dinkel, Sushama Michael, Robert J Weatheritt, Norman E Davey, Kim Van Roey, Brigitte Altenberg, Grischa Toedt, Bora Uyar, Markus Seiler, Aidan Budd, Lisa Jödicke, Marcel A Dammert, Christian Schroeter, Maria Hammer, Tobias Schmidt, Peter Jehl, Caroline McGuigan, Magdalena Dymecka, Claudia Chica, Katja Luck, Allegra Via, Andrew Chatr-Aryamontri, Niall Haslam, Gleb Grebnev, Richard J Edwards, Michel O Steinmetz, Heike Meiselbach, Francesca Diella, and Toby J Gibson. ELM–the database of eukaryotic linear motifs. *Nucleic acids research*, 40(Database issue):D242–51, January 2012.

[20] Norman E Davey, Gilles Travé, and Toby J Gibson. How viruses hijack cell regulation. *Trends in biochemical sciences*, 36(3):159–69, March 2011.

[21] Patrick S Moore and Yuan Chang. Why do viruses cause cancer? Highlights of the first century of human tumour virology. *Nature reviews. Cancer*, 10(12):878–89, December 2010.

[22] W-L Tsai and R T Chung. Viral hepatocarcinogenesis. *Oncogene*, 29(16):2309–24, April 2010.

[23] Peter E Wright and H Jane Dyson. Linking folding and binding. *Current opinion in structural biology*, 19(1):31–8, February 2009.

[24] Elio A Cino, Ryan C Killoran, Mikko Karttunen, and Wing-Yiu Choy. Binding of disordered proteins to a protein hub. *Scientific reports*, 3:2305, January 2013.

[25] Peter Tompa and Monika Fuxreiter. Fuzzy complexes: polymorphism and structural disorder in protein-protein interactions. *Trends in biochemical sciences*, 33(1):2–8, January 2008.

[26] Monika Fuxreiter. Fuzziness: linking regulation to protein dynamics. *Molecular bioSystems*, 8(1):168–77, January 2012.

[27] Tanja Mittag, Stephen Orlicky, Wing-Yiu Choy, Xiaojing Tang, Hong Lin, Frank Sicheri, Lewis E Kay, Mike Tyers, and Julie D Forman-Kay. Dynamic equilibrium engagement of a polyvalent ligand with a single-site receptor. *Proceedings of the National Academy of Sciences of the United States of America*, 105(46):17772–7, November 2008.

[28] Aaron A Santner, Carrie H Croy, Farha H Vasanwala, Vladimir N Uversky, Ya-Yue J Van, and A Keith Dunker. Sweeping away protein aggregation with entropic bristles: intrinsically disordered protein fusions enhance soluble expression. *Biochemistry*, 51(37):7250–62, October 2012.

[29] Albert H Mao, Scott L Crick, Andreas Vitalis, Caitlin L Chicoine, and Rohit V Pappu. Net charge per residue modulates conformational ensembles of intrinsically disordered proteins. *Proceedings of the National Academy of Sciences of the United States of America*, 107(18):8183–8, May 2010.

[30] Leonhard Geist, Morkos A Henen, Sandra Haiderer, Thomas C Schwarz, Dennis Kurzbach, Anna Zawadzka-Kazimierczuk, Saurabh Saxena, Szymon Zerko, Wiktor Koźmiski, Dariush Hinderberger, and Robert Konrat. Protonation-dependent conformational variability of intrinsically disordered proteins. *Protein science : a publication of the Protein Society*, 22(9):1196–205, September 2013.

[31] Magnus Kjaergaard, Ann-Beth Nø rholm, Ruth Hendus-Altenburger, Stine F Pedersen, Flemming M Poulsen, and Birthe B Kragelund. Temperature-dependent structural changes in intrinsically disordered proteins: formation of alpha-helices or loss of polyproline II? *Protein science : a publication of the Protein Society*, 19(8):1555–64, August 2010.

[32] Mark Hochstrasser. Protein Degradation or Regulation: Ub the Judge. *Cell*, 84(6):813–815, March 1996.

[33] Marcin J Suskiewicz, Joel L Sussman, Israel Silman, and Yosef Shaul. Context-dependent resistance to proteolysis of intrinsically disordered proteins. *Protein science : a publication of the Protein Society*, 20(8):1285–97, August 2011.

[34] Peter Tsvetkov, Nina Reuven, and Yosef Shaul. The nanny model for IDPs. *Nature chemical biology*, 5(11):778–81, November 2009.

[35] Muhamad Ali, Hirotatsu Suzuki, Takako Fukuba, Xiuping Jiang, Hideo Nakano, and Tsuneo Yamane. Improvements in the cell-free production of functional antibodies using cell extract from protease-deficient Escherichia coli mutant. *Journal of bioscience and bioengineering*, 99(2):181–6, March 2005.

[36] V N Uversky, J Li, and A L Fink. Evidence for a partially folded intermediate in alpha-synuclein fibril formation. *The Journal of biological chemistry*, 276(14):10737–44, April 2001.

[37] Vladimir Uversky and Sonia Longhi. *Instrumental analysis of intrinsically disordered proteins: assessing structure and conformation*, volume 3. John Wiley & Sons, 2011.

[38] Pau Bernadó and Dmitri I Svergun. Structural analysis of intrinsically disordered proteins by small-angle X-ray scattering. *Molecular bioSystems*, 8(1):151–67, January 2012.

[39] Pau Bernadó, Efstratios Mylonas, Maxim V Petoukhov, Martin Blackledge, and Dmitri I Svergun. Structural characterization of flexible proteins using small-angle X-ray scattering. *Journal of the American Chemical Society*, 129(17):5656–64, May 2007.

[40] Matthew W Freyer and Edwin A Lewis. Isothermal titration calorimetry: experimental design, data analysis, and probing macromolecule/ligand binding and kinetic interactions. *Methods in cell biology*, 84:79–113, January 2008.

[41] E M Purcell, H C Torrey, and R V Pound. Resonance Absorption by Nuclear Magnetic Moments in a Solid. *Phys. Rev.*, 69(1-2):37–38, January 1946.

[42] F Bloch. Nuclear Induction. *Phys. Rev.*, 70(7-8):460–474, October 1946.

[43] R Ishima and D A Torchia. Protein dynamics from NMR. *Nature structural biology*, 7(9):740–3, September 2000.

[44] Modesto Orozco. A theoretical view of protein dynamics. *Chemical Society Reviews*, 2014.

[45] R R Ernst, G Bodenhausen, and A Wokaun. *Principles of Nuclear Magnetic Resonance in One and Two Dimensions.* International series of monographs on chemistry. Clarendon Press, 1990.

[46] M H Levitt. *Spin Dynamics: Basics of Nuclear Magnetic Resonance.* Wiley, 2013.

[47] J Cavanagh, W J Fairbrother, A G Palmer, N J Skelton, and M Rance. *Protein NMR Spectroscopy: Principles and Practice.* Elsevier Science, 2010.

[48] Jung Ho Lee, Yusuke Okuno, and Silvia Cavagnero. Sensitivity enhancement in solution NMR: emerging ideas and new frontiers. *Journal of magnetic resonance (San Diego, Calif. : 1997)*, 241:18–31, April 2014.

[49] Aaron J Rossini, Alexandre Zagdoun, Moreno Lelli, Anne Lesage, Christophe Copéret, and Lyndon Emsley. Dynamic nuclear polarization surface enhanced NMR spectroscopy. *Accounts of chemical research*, 46(9):1942–51, September 2013.

[50] Ivano Bertini, Luminita Duma, Isabella C. Felli, Michael Fey, Claudio Luchinat, Roberta Pierattelli, and Paul R. Vasos. A Heteronuclear Direct-Detection NMR Spectroscopy Experiment for Protein-Backbone Assignment. *Angewandte Chemie International Edition*, 43(17):2257–2259, April 2004.

[51] Wolfgang Bermel, Ivano Bertini, Isabella C Felli, Yong-Min Lee, Claudio Luchinat, and Roberta Pierattelli. Protonless NMR experiments for sequence-specific assignment of backbone nuclei in unfolded proteins. *Journal of the American Chemical Society*, 128(12):3918–9, March 2006.

[52] Andres Binolfi, Francois-Xavier Theillet, and Philipp Selenko. Bacterial in-cell NMR of human α-synuclein: a disordered monomer by nature? *Biochemical Society transactions*, 40(5):950–4, October 2012.

[53] Andrew Campen, Ryan M Williams, Celeste J Brown, Jingwei Meng, Vladimir N Uversky, and A Keith Dunker. TOP-IDP-scale: a new amino acid scale measuring propensity for intrinsic disorder. *Protein and peptide letters*, 15(9):956–63, January 2008.

[54] N J Baxter and M P Williamson. Temperature dependence of 1H chemical shifts in proteins. *Journal of biomolecular NMR*, 9(4):359–69, June 1997.

[55] Andrea Cavalli, Xavier Salvatella, Christopher M Dobson, and Michele Vendruscolo. Protein structure determination from NMR chemical shifts. *Proceedings of the National Academy of Sciences of the United States of America*, 104(23):9615–20, June 2007.

[56] Lorna J Smith, Klaus M Fiebig, Harald Schwalbe, and Christopher M Dobson. The concept of a random coil Residual structure in peptides and denatured proteins. *Folding and Design*, 1(5):95–106, 1996.

[57] Paul J. Flory. Effect of Volume Exclusion on the Dimensions of Polymer Chains. *The Journal of Chemical Physics*, 44(6):2243, May 1966.

[58] Paul J Flory. *Principles of Polymer Chemistry.* Baker lectures 1948. Cornell University Press, 1953.

[59] Rohit V Pappu, Rajgopal Srinivasan, and George D Rose. The Flory isolated-pair hypothesis is not valid for polypeptide chains : Impli-

cations for protein folding. *Proceedings of the National Academy of Sciences of the United States of America*, 97(23):12565–12570, 2000.

[60] Haiyan Zhang, Stephen Neal, and DS Wishart. RefDB: a database of uniformly referenced protein chemical shifts. *Journal of biomolecular NMR*, 25:173–195, 2003.

[61] D S Wishart, C G Bigam, a Holm, R S Hodges, and B D Sykes. (1)H, (13)C and (15)N random coil NMR chemical shifts of the common amino acids. I. Investigations of nearest-neighbor effects. *Journal of biomolecular NMR*, 5(3):332, April 1995.

[62] Magnus Kjaergaard, Sø ren Brander, and Flemming M Poulsen. Random coil chemical shift for intrinsically disordered proteins: effects of temperature and pH. *Journal of biomolecular NMR*, 49(2):139–49, February 2011.

[63] Kamil Tamiola, Burçin Acar, and Frans A A Mulder. Sequence-specific random coil chemical shifts of intrinsically disordered proteins. *Journal of the American Chemical Society*, 132(51):18000–3, December 2010.

[64] Stephan Schwarzinger, Gerard J. A. Kroon, Ted R. Foss, John Chung, Peter E. Wright, and H. Jane Dyson. Sequence-Dependent Correction of Random Coil NMR Chemical Shifts. *Journal of the American Chemical Society*, 123(13):2970–2978, April 2001.

[65] Magnus Kjaergaard and Flemming M Poulsen. Sequence correction of random coil chemical shifts: correlation between neighbor correction factors and changes in the Ramachandran distribution. *Journal of biomolecular NMR*, 50(2):157–65, June 2011.

[66] Pau Bernadó, Laurence Blanchard, Peter Timmins, Dominique Marion, Rob W H Ruigrok, and Martin Blackledge. A structural model for unfolded proteins from residual dipolar couplings and small-angle x-ray scattering. *Proceedings of the National Academy of Sciences of the United States of America*, 102(47):17002–7, November 2005.

[67] Valéry Ozenne, Frédéric Bauer, Loïc Salmon, Jie-Rong Huang, Malene Ringkjø bing Jensen, Stéphane Segard, Pau Bernadó, Céline Charavay, and Martin Blackledge. Flexible-meccano: a tool for the generation of explicit ensemble descriptions of intrinsically disordered proteins and their associated experimental observables. *Bioinformatics (Oxford, England)*, 28(11):1463–70, June 2012.

[68] Giovanni Lipari and Attila Szabo. Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 1. Theory and range of validity. *Journal of the American Chemical Society*, 104(17):4546–4559, August 1982.

[69] L E Kay, D A Torchia, and A Bax. Backbone dynamics of proteins as studied by 15N inverse detected heteronuclear NMR spectroscopy: application to staphylococcal nuclease. *Biochemistry*, 28(23):8972–9, November 1989.

[70] N a Farrow, O Zhang, a Szabo, D a Torchia, and L E Kay. Spectral density function mapping using 15N relaxation data exclusively. *Journal of biomolecular NMR*, 6(2):153–62, September 1995.

[71] Pavel Kadeávek, Vojtěch Zapletal, Alžbeta Rabatinová, Libor Krásný, Vladimír Sklená, and Lukáš Žídek. Spectral density mapping protocols for analysis of molecular motions in disordered proteins. *Journal of biomolecular NMR*, 58(3):193–207, March 2014.

[72] Jan Stanek, Saurabh Saxena, Leonhard Geist, Robert Konrat, and Wiktor Koźmiski. Probing local backbone geometries in intrinsically disordered proteins by cross-correlated NMR relaxation. *Angewandte Chemie (International ed. in English)*, 52(17):4604–6, April 2013.

[73] G Marius Clore and Junji Iwahara. Theory, practice, and applications of paramagnetic relaxation enhancement for the characterization of transient low-population states of biological macromolecules and their complexes. *Chemical reviews*, 109(9):4108–39, September 2009.

[74] Gabrielle Nodet, Loïc Salmon, Valéry Ozenne, Sebastian Meier, Malene Ringkjø bing Jensen, and Martin Blackledge. Quantitative description of backbone conformational sampling of unfolded proteins at amino acid resolution from NMR residual dipolar couplings. *Journal of the American Chemical Society*, 131(49):17908–18, December 2009.

[75] Joseph A Marsh and Julie D Forman-Kay. Ensemble modeling of protein disordered states: Experimental restraint contributions and validation. *Proteins*, October 2011.

[76] Mickaël Krzeminski, Joseph a Marsh, Chris Neale, Wing-Yiu Choy, and Julie D Forman-Kay. Characterization of disordered proteins with ENSEMBLE. *Bioinformatics (Oxford, England)*, 29(3):398–9, February 2013.

[77] Paul Schanda. Fast-pulsing longitudinal relaxation optimized techniques: Enriching the toolbox of fast biomolecular NMR spectroscopy. *Progress in Nuclear Magnetic Resonance Spectroscopy*, 55(3):238–265, October 2009.

[78] R R Ernst and W A Anderson. Application of Fourier transform spectroscopy to magnetic resonance. *Review of Scientific Instruments*, 37:93, 1966.

[79] A Ross, M Salzmann, and H Senn. Fast-HMQC using Ernst angle pulses: An efficient tool for screening of ligand binding to target proteins. *Journal of biomolecular NMR*, 10(4):389–96, December 1997.

[80] Paul Schanda, Eriks Kupce, and Bernhard Brutscher. SOFAST-HMQC experiments for recording two-dimensional heteronuclear correlation spectra of proteins within a few seconds. *Journal of biomolecular NMR*, 33(4):199–211, December 2005.

[81] Konstantin Pervushin, Beat Vögeli, and Alexander Eletsky. Longitudinal (1)H relaxation optimization in TROSY NMR spectroscopy. *Journal of the American Chemical Society*, 124(43):12898–902, October 2002.

[82] G Otting, E Liepinsh, and K Wüthrich. Protein hydration in aqueous solution. *Science (New York, N.Y.)*, 254(5034):974–80, November 1991.

[83] Tammo Diercks, Mark Daniels, and Robert Kaptein. Extended flip-back schemes for sensitivity enhancement in multidimensional HSQC-type out-and-back experiments. *Journal of biomolecular NMR*, 33(4):243–59, December 2005.

[84] Wolfgang Bermel, Ivano Bertini, Veronika Csizmok, Isabella C Felli, Roberta Pierattelli, and Peter Tompa. H-start for exclusively heteronuclear NMR spectroscopy: the case of intrinsically disordered proteins. *Journal of magnetic resonance (San Diego, Calif. : 1997)*, 198(2):275–81, June 2009.

[85] Ewen Lescop, Thomas Kern, and Bernhard Brutscher. Guidelines for the use of band-selective radiofrequency pulses in hetero-nuclear NMR: example of longitudinal-relaxation-enhanced BEST-type 1H-15N correlation experiments. *Journal of magnetic resonance (San Diego, Calif. : 1997)*, 203(1):190–8, March 2010.

[86] Paul Schanda and Bernhard Brutscher. Very fast two-dimensional NMR spectroscopy for real-time investigation of dynamic events in proteins on the time scale of seconds. *Journal of the American Chemical Society*, 127(22):8014–5, June 2005.

[87] Francois-Xavier Theillet, Honor May Rose, Stamatios Liokatis, Andres Binolfi, Rossukon Thongwichian, Marchel Stuiver, and Philipp Selenko. Site-specific NMR mapping and time-resolved monitoring of serine and threonine phosphorylation in reconstituted kinase reactions and mammalian cell extracts. *Nature protocols*, 8(7):1416–32, June 2013.

[88] Carlos Amero, Paul Schanda, M Asunción Durá, Isabel Ayala, Dominique Marion, Bruno Franzetti, Bernhard Brutscher, and Jérôme Boisbouvier. Fast two-dimensional NMR spectroscopy of high molecular weight protein assemblies. *Journal of the American Chemical Society*, 131(10):3448–9, March 2009.

[89] Enrico Rennella, Thomas Cutuil, Paul Schanda, Isabel Ayala, Vincent Forge, and Bernhard Brutscher. Real-time NMR characterization of structure and dynamics in a transiently populated protein folding intermediate. *Journal of the American Chemical Society*, 134(19):8066–9, May 2012.

[90] Paul Schanda, Hélène Van Melckebeke, and Bernhard Brutscher. Speeding up three-dimensional protein NMR experiments to a few minutes. *Journal of the American Chemical Society*, 128(28):9042–3, July 2006.

[91] Ewen Lescop, Paul Schanda, and Bernhard Brutscher. A set of BEST triple-resonance experiments for time-optimized protein resonance assignment. *Journal of magnetic resonance (San Diego, Calif. : 1997)*, 187(1):163–9, July 2007.

[92] Paul Schanda, Ewen Lescop, Mirjam Falge, Rémy Sounier, Jérôme Boisbouvier, and Bernhard Brutscher. Sensitivity-optimized experiment for the measurement of residual dipolar couplings between amide protons. *Journal of biomolecular NMR*, 38(1):47–55, May 2007.

[93] Dominique Marion. Combining methods for speeding up multidimensional acquisition. Sparse sampling and fast pulsing methods for unfolded proteins. *Journal of magnetic resonance (San Diego, Calif. : 1997)*, 206(1):81–7, September 2010.

[94] Y Z Zhang. *Protein and Peptide Structure and Interactions Studied by Hydrogen Exchange and NMR.* University of Pennsylvania, 1995.

[95] J H Simpson and H Y Carr. Diffusion and Nuclear Spin Relaxation in Water. *Phys. Rev.*, 111(5):1201–1202, September 1958.

[96] Sheng Cai, Candace Seu, Zoltan Kovacs, A Dean Sherry, and Yuan Chen. Sensitivity Enhancemen of Multidimensional NMR Experiments by Paramagnetic Relaxation Effects. *Journal of the American Chemical Society*, 128(5):13474–13478, 2006.

[97] Alexander Eletsky, Osvaldo Moreira, Helena Kovacs, and Konstantin Pervushin. A novel strategy for the assignment of side-chain resonances in completely deuterated large proteins using 13C spectroscopy. *Journal of biomolecular NMR*, 26(2):167–79, June 2003.

[98] Sebastian Hiller, Gerhard Wider, Touraj Etezady-Esfarjani, Reto Horst, and Kurt Wüthrich. Managing the solvent water polarization to obtain improved NMR spectra of large molecular structures. *Journal of biomolecular NMR*, 32(1):61–70, May 2005.

[99] François-Xavier Theillet, Andres Binolfi, Stamatis Liokatis, Silvia Verzini, and Philipp Selenko. Paramagnetic relaxation enhancement to improve sensitivity of fast NMR methods: application to intrinsically disordered proteins. *Journal of biomolecular NMR*, 51(4):487–95, December 2011.

[100] Nathalie Sibille, Gaëtan Bellot, Jing Wang, and Hélène Déméné. Low concentration of a Gd-chelate increases the signal-to-noise ratio in fast pulsing BEST experiments. *Journal of magnetic resonance (San Diego, Calif. : 1997)*, 224:32–7, November 2012.

[101] K Pervushin, R Riek, G Wider, and K Wüthrich. Attenuated T2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. *Proceedings of the National Academy of Sciences of the United States of America*, 94(23):12366–71, November 1997.

[102] Jonathan Farjon, Jérôme Boisbouvier, Paul Schanda, Arthur Pardi, Jean-Pierre Simorre, and Bernhard Brutscher. Longitudinal-relaxation-enhanced NMR experiments for the study of nucleic acids in solution. *Journal of the American Chemical Society*, 131(24):8571–7, June 2009.

[103] Adrien Favier and Bernhard Brutscher. Recovering lost magnetization: polarization enhancement in biomolecular NMR. *Journal of biomolecular NMR*, 49(1):9–15, January 2011.

[104] Helen Geen and Ray Freeman. Band-selective radiofrequency pulses. *Journal of Magnetic Resonance (1969)*, 93(1):93–141, June 1991.

[105] E Kupce and R Freeman. Wideband excitation with polychromatic pulses. *Journal of Magnetic Resonance, Series A*, 108:268–273, 1994.

[106] Alexander S Maltsev, Jinfa Ying, and Ad Bax. Deuterium isotope shifts for backbone H, N and C nuclei in intrinsically disordered protein α-synuclein. *Journal of biomolecular NMR*, 54(2):181–91, October 2012.

[107] Isabella C Felli and Bernhard Brutscher. Recent advances in solution NMR: fast methods and heteronuclear direct detection. *Chemphyschem : a European journal of chemical physics and physical chemistry*, 10(9-10):1356–68, July 2009.

[108] Mari a Smith, Haitao Hu, and a.J Shaka. Improved Broadband Inversion Performance for NMR in Liquids. *Journal of Magnetic Resonance*, 151(2):269–283, August 2001.

[109] Zsofia Solyom, Melanie Schwarten, Leonhard Geist, Robert Konrat, Dieter Willbold, and Bernhard Brutscher. BEST-TROSY experiments for time-efficient sequential resonance assignment of large disordered proteins. *Journal of biomolecular NMR*, 55(4):311–21, April 2013.

[110] Peter Tompa. Intrinsically disordered proteins: a 10-year recap. *Trends in biochemical sciences*, 37(12):509–16, December 2012.

[111] Ewen Lescop, Rodolfo Rasia, and Bernhard Brutscher. Hadamard amino-acid-type edited NMR experiment for fast protein resonance assignment. *Journal of the American Chemical Society*, 130(15):5014–5, April 2008.

[112] Sophie Feuerstein, Zsofia Solyom, Amine Aladag, Adrien Favier, Melanie Schwarten, Silke Hoffmann, Dieter Willbold, and Bernhard Brutscher. Transient structure and SH3 interaction sites in an intrinsically disordered fragment of the hepatitis C virus protein NS5A. *Journal of molecular biology*, 420(4-5):310–23, July 2012.

[113] G Wagner and K Wüthrich. Amide protein exchange and surface conformation of the basic pancreatic trypsin inhibitor in solution. Studies with two-dimensional nuclear magnetic resonance. *Journal of molecular biology*, 160(2):343–61, September 1982.

[114] Maayan Gal, Paul Schanda, Bernhard Brutscher, and Lucio Frydman. UltraSOFAST HMQC NMR and the repetitive acquisition of 2D protein spectra at Hz rates. *Journal of the American Chemical Society*, 129(5):1372–7, February 2007.

[115] Jing-Song Fan, Jackwee Lim, Binhan Yu, and Daiwen Yang. Measurement of amide hydrogen exchange rates with the use of radiation damping. *Journal of biomolecular NMR*, 51(1-2):151–62, September 2011.

[116] S Grzesiek and A Bax. Measurement of amide proton exchange rates and NOEs with water in 13C/15N-enriched calcineurin B. *Journal of biomolecular NMR*, 3(6):627–38, November 1993.

[117] T L Hwang, P C van Zijl, and S Mori. Accurate quantitation of water-amide proton exchange rates using the phase-modulated CLEAN chemical EXchange (CLEANEX-PM) approach with a Fast-HSQC (FHSQC) detection scheme. *Journal of biomolecular NMR*, 11(2):221–6, February 1998.

[118] S Mori, C Abeygunawardana, P C van Zijl, and J M Berg. Water exchange filter with improved sensitivity (WEX II) to study solvent-exchangeable protons. Application to the consensus zinc finger peptide CP-1. *Journal of magnetic resonance. Series B*, 110(1):96–101, January 1996.

[119] Gerhard Wider, Roland Riek, and Kurt Wüthrich. Diffusion Filters for Separation of SolventProtein and ProteinProtein Nuclear Overhauser Effects (HYDRA). *Journal of the American Chemical Society*, 118(46):11629–11634, January 1996.

[120] Veniamin Chevelkov, Yi Xue, D Krishna Rao, Julie D Forman-Kay, and Nikolai R Skrynnikov. 15N H/D-SOLEXSY experiment for accurate measurement of amide solvent exchange rates: application to denatured drkN SH3. *Journal of biomolecular NMR*, 46(3):227–44, March 2010.

[121] Paul Schanda, Vincent Forge, and Bernhard Brutscher. HET-SOFAST NMR for fast detection of structural compactness and heterogeneity along polypeptide chains. *Magnetic resonance in chemistry : MRC*, 44 Spec No:S177–84, July 2006.

[122] Markus H Heim. 25 years of interferon-based treatment of chronic hepatitis C: an epoch coming to an end. *Nature reviews. Immunology*, 13(7):535–42, July 2013.

[123] Q L Choo, G Kuo, A J Weiner, L R Overby, D W Bradley, and M Houghton. Isolation of a cDNA clone derived from a blood-borne non-A, non-B viral hepatitis genome. *Science (New York, N.Y.)*, 244(4902):359–62, April 1989.

[124] Darius Moradpour, François Penin, and Charles M Rice. Replication of hepatitis C virus. *Nature reviews. Microbiology*, 5(6):453–63, June 2007.

[125] Gaël Cristofari, Roland Ivanyi-Nagy, Caroline Gabus, Steeve Boulant, Jean-Pierre Lavergne, François Penin, and Jean-Luc Darlix. The hepatitis C virus Core protein is a potent nucleic acid chaperone that directs dimerization of the viral (+) strand RNA in vitro. *Nucleic acids research*, 32(8):2623–31, January 2004.

[126] Roland Ivanyi-Nagy, Jean-Pierre Lavergne, Caroline Gabus, Damien Ficheux, and Jean-Luc Darlix. RNA chaperoning and intrinsic disorder in the core proteins of Flaviviridae. *Nucleic acids research*, 36(3):712–25, February 2008.

[127] B de Chassey, V Navratil, L Tafforeau, M S Hiet, a Aublin-Gex, S Agaugué, G Meiffren, F Pradezynski, B F Faria, T Chantier, M Le Breton, J Pellet, N Davoust, P E Mangeot, a Chaboud, F Penin, Y Jacob, P O Vidalain, M Vidal, P André, C Rabourdin-Combe, and V Lotteau. Hepatitis C virus infection protein network. *Molecular systems biology*, 4(230):230, January 2008.

[128] Timothy L Tellinghuisen, Katie L Foss, and Jason Treadaway. Regulation of hepatitis C virion production via phosphorylation of the NS5A protein. *PLoS pathogens*, 4(3):e1000032, March 2008.

[129] Bo OuYang, Shiqi Xie, Marcelo J Berardi, Xinhao Zhao, Jyoti Dev, Wenjing Yu, Bing Sun, and James J Chou. Unusual architecture of the p7 channel from hepatitis C virus. *Nature*, 498(7455):521–5, June 2013.

[130] Nicole Appel, Thomas Pietschmann, and Ralf Bartenschlager. Mutational analysis of hepatitis C virus nonstructural protein 5A: potential role of differential phosphorylation in RNA replication and identification of a genetically flexible domain. *Journal of virology*, 79(5):3187–94, March 2005.

[131] Xavier Hanoulle, Aurélie Badillo, Jean-Michel Wieruszeski, Dries Verdegem, Isabelle Landrieu, Ralf Bartenschlager, François Penin, and Guy Lippens. Hepatitis C virus NS5A protein is a substrate for the peptidyl-prolyl cis/trans isomerase activity of cyclophilins A and B. *The Journal of biological chemistry*, 284:13589–13601, 2009.

[132] Timothy L Tellinghuisen, Joseph Marcotrigiano, Alexander E Gorbalenya, and Charles M Rice. The NS5A protein of hepatitis C virus is a zinc metalloprotein. *The Journal of biological chemistry*, 279(47):48576–87, November 2004.

[133] Timothy L Tellinghuisen, Joseph Marcotrigiano, and Charles M Rice. Structure of the zinc-binding domain of an essential component of the hepatitis C virus replicase. *Nature*, 435(7040):374–9, May 2005.

[134] Jungwook Hwang, Luyun Huang, Daniel G Cordek, Robert Vaughan, Shelley L Reynolds, George Kihara, Kevin D Raney, C Cheng Kao, and Craig E Cameron. Hepatitis C virus nonstructural protein 5A: biochemical characterization of a novel structural class of RNA-binding proteins. *Journal of virology*, 84(24):12480–91, December 2010.

[135] Darius Moradpour, Volker Brass, and Francois Penin. Function follows form: the structure of the N-terminal domain of HCV NS5A. *Hepatology (Baltimore, Md.)*, 42(3):732–5, September 2005.

[136] Gregory Camus, Eva Herker, Ankit A Modi, Joel T Haas, Holly R Ramage, Robert V Farese, and Melanie Ott. Diacylglycerol acyltransferase-1 localizes hepatitis C virus NS5A protein to lipid droplets and enhances NS5A interaction with the viral capsid core. *The Journal of biological chemistry*, 288(14):9915–23, April 2013.

[137] Robert A Love, Oleg Brodsky, Michael J Hickey, Peter A Wells, and Ciarán N Cronin. Crystal structure of a novel dimeric form of NS5A domain I protein from hepatitis C virus. *Journal of virology*, 83(9):4395–403, May 2009.

[138] Dries Verdegem, Aurélie Badillo, Jean-Michel Wieruszeski, Isabelle Landrieu, Arnaud Leroy, Ralf Bartenschlager, François Penin, Guy Lippens, and Xavier Hanoulle. Domain 3 of NS5A protein from the hepatitis C virus has intrinsic alpha-helical propensity and is a substrate of cyclophilin A. *The Journal of biological chemistry*, 286(23):20441–54, June 2011.

[139] Xavier Hanoulle, Aurelie Badillo, Dries Verdegem, Francois Penin, and Guy Lippens. The domain 2 of the HCV NS5A protein is intrinsically unstructured. *Protein and peptide letters*, 17(8):1012–1018, 2010.

[140] Yu Liang, Hong Ye, Cong Bao Kang, and Ho Sup Yoon. Domain 2 of nonstructural protein 5A (NS5A) of hepatitis C virus is natively unfolded. *Biochemistry*, 46(41):11550–8, October 2007.

[141] Melanie Schwarten, Zsófia Sólyom, Sophie Feuerstein, Amine Aladag, Silke Hoffmann, Dieter Willbold, and Bernhard Brutscher. Interaction of nonstructural protein 5A of the hepatitis C virus with Src homology 3 domains using noncanonical binding sites. *Biochemistry*, 52(36):6160–8, September 2013.

[142] Alexandra M Livernois, Daniel J Hnatchuk, Emma E Findlater, and Steffen P Graether. Obtaining highly purified intrinsically disordered protein by boiling lysis and single step ion exchange. *Analytical biochemistry*, 392(1):70–6, September 2009.

[143] Sophie Feuerstein, Michael J Plevin, Dieter Willbold, and Bernhard Brutscher. iHADAMAC: a complementary tool for sequential resonance assignment of globular and highly disordered proteins. *Journal of magnetic resonance (San Diego, Calif. : 1997)*, 214(1):329–34, January 2012.

[144] Sophie Feuerstein, Zsófia Solyom, Amine Aladag, Silke Hoffmann, Dieter Willbold, and Bernhard Brutscher. 1H, 13C, and 15N resonance assignment of a 179 residue fragment of hepatitis C virus non-

structural protein 5A. *Biomolecular NMR assignments*, 5(2):241–3, October 2011.

[145] V Muñoz and L Serrano. Elucidating the folding problem of helical peptides using empirical parameters. *Nature structural biology*, 1(6):399–409, June 1994.

[146] Joseph A Marsh, Vinay K Singh, Zongchao Jia, and Julie D Forman-Kay. Sensitivity of secondary structure propensities to sequence differences between alpha- and gamma-synuclein: implications for fibrillation. *Protein science : a publication of the Protein Society*, 15(12):2795–804, December 2006.

[147] AngelaM. Gronenborn and G. Marius Clore. Identification of N-terminal helix capping boxes by means of 13C chemical shifts. *Journal of Biomolecular NMR*, 4(3), May 1994.

[148] Claire Rosnoblet, Bernd Fritzinger, Dominique Legrand, Hélène Launay, Jean-Michel Wieruszeski, Guy Lippens, and Xavier Hanoulle. Hepatitis C virus NS5B and host cyclophilin A share a common binding site on NS5A. *The Journal of biological chemistry*, 287(53):44249–60, December 2012.

[149] D. Svergun, C. Barberato, and M.H.J. Koch. CRYSOL - a Program to Evaluate X-ray Solution Scattering of Biological Macromolecules from Atomic Coordinates, December 1995.

[150] J J Chou, S Gaemers, B Howder, J M Louis, and a Bax. A simple apparatus for generating stretched polyacrylamide gels, yielding uniform alignment of proteins and detergent micelles. *Journal of biomolecular NMR*, 21(4):377–82, December 2001.

[151] C H Lee, B Leung, M A Lemmon, J Zheng, D Cowburn, J Kuriyan, and K Saksela. A single amino acid in the SH3 domain of Hck determines its high affinity and specificity in binding to HIV-1 Nef protein. *The EMBO journal*, 14(20):5006–15, October 1995.

[152] BJ Mayer. SH3 domains: complexity in moderation. *J. Cell Sci.*, 114(7):1253–1263, April 2001.

[153] Maria Harkiolaki, Marc Lewitzky, Robert J C Gilbert, E Yvonne Jones, Roland P Bourette, Guy Mouchiroud, Holger Sondermann, Ismail Moarefi, and Stephan M Feller. Structural basis for SH3 domain-mediated high-affinity binding between Mona/Gads and SLP-76. *The EMBO journal*, 22(11):2571–82, June 2003.

[154] Maryna Gorelik and Alan R Davidson. Distinct peptide binding specificities of Src homology 3 (SH3) protein domains can be determined by modulation of local energetics across the binding interface. *The Journal of biological chemistry*, 287(12):9168–77, March 2012.

[155] Mario Angelo Pagano, Elena Tibaldi, Giorgio Palù, and Anna Maria Brunati. Viral proteins and Src family kinases: Mechanisms of pathogenicity from a "liaison dangereuse". *World journal of virology*, 2(2):71–78, May 2013.

[156] S L Tan, H Nakao, Y He, S Vijaysri, P Neddermann, B L Jacobs, B J Mayer, and M G Katze. NS5A, a nonstructural protein of hepatitis C virus, binds growth factor receptor-bound protein 2 adaptor protein in a Src homology 3 domain/ligand-dependent manner and perturbs mitogenic signaling. *Proceedings of the National Academy of Sciences of the United States of America*, 96(10):5533–8, May 1999.

[157] Andrew Macdonald, Katherine Crowder, Andrew Street, Christopher McCormick, and Mark Harris. The hepatitis C virus NS5A protein binds to members of the Src family of tyrosine kinases and regulates kinase activity. *The Journal of general virology*, 85(Pt 3):721–9, March 2004.

[158] Holly Shelton and Mark Harris. Hepatitis C virus NS5A protein binds the SH3 domain of the Fyn tyrosine kinase with high affinity: mutagenic analysis of residues within the SH3 domain that contribute to the interaction. *Virology journal*, 5:24, January 2008.

[159] Santosh K Nanda, David Herion, and T Jake Liang. The SH3 binding motif of HCV [corrected] NS5A protein interacts with Bin1 and is important for apoptosis and infectivity. *Gastroenterology*, 130(3):794–809, March 2006.

[160] Atsuko Masumi, Hideki Aizaki, Tetsuro Suzuki, James B DuHadaway, George C Prendergast, Katsutoshi Komuro, and Hidesuke Fukazawa. Reduction of hepatitis C virus NS5A phosphorylation through its interaction with amphiphysin II. *Biochemical and biophysical research communications*, 336(2):572–8, October 2005.

[161] Cecilia Andresen, Sara Helander, Alexander Lemak, Christophe Farès, Veronika Csizmok, Jonas Carlsson, Linda Z Penn, Julie D Forman-Kay, Cheryl H Arrowsmith, Patrik Lundström, and Maria Sunnerhagen. Transient structure and dynamics in the disordered c-Myc transactivation domain affect Bin1 binding. *Nucleic acids research*, 40(13):6353–66, July 2012.

[162] Brian W Chellgren and Trevor P Creamer. Short sequences of non-proline residues can adopt the polyproline II helical conformation. *Biochemistry*, 43(19):5864–9, May 2004.

[163] Reinhard Schweitzer-Stenner, Fatma Eker, Kai Griebenow, Xiaolin Cao, and Laurence A Nafie. The conformation of tetraalanine in water determined by polarized Raman, FT-IR, and VCD spectroscopy. *Journal of the American Chemical Society*, 126(9):2768–76, March 2004.

[164] Valéry Ozenne, Robert Schneider, Mingxi Yao, Jie-rong Huang, Loïc Salmon, Markus Zweckstetter, Malene Ringkjø bing Jensen, and Martin Blackledge. Mapping the potential energy landscape of intrinsically disordered proteins at amino acid resolution. *Journal of the American Chemical Society*, 134(36):15138–48, September 2012.

[165] Garvita Gupta, Haina Qin, and Jianxing Song. Intrinsically unstructured domain 3 of hepatitis C Virus NS5A forms a "fuzzy complex" with VAPB-MSP domain which carries ALS-causing mutations. *PloS one*, 7(6):e39261, January 2012.

[166] Eva Casal, Luca Federici, Wei Zhang, Juan Fernandez-Recio, Eva-Maria Priego, Ricardo Nuñez Miguel, James B DuHadaway, George C Prendergast, Ben F Luisi, and Ernest D Laue. The crystal structure of the BAR domain from human Bin1/amphiphysin II and its implications for molecular recognition. *Biochemistry*, 45(43):12917–28, October 2006.

[167] Robert Roskoski. Src kinase regulation by phosphorylation and dephosphorylation. *Biochemical and biophysical research communications*, 331(1):1–14, May 2005.

[168] Isabelle Landrieu, Ludovic Lacosse, Arnaud Leroy, Jean-Michel Wieruszeski, Xavier Trivelli, Alain Sillen, Nathalie Sibille, Harald Schwalbe, Krishna Saxena, Thomas Langer, and Guy Lippens. NMR analysis of a Tau phosphorylation pattern. *Journal of the American Chemical Society*, 128(11):3575–83, March 2006.

[169] Philipp Selenko, Dominique P Frueh, Simon J Elsaesser, Wilhelm Haas, Steven P Gygi, and Gerhard Wagner. In situ observation of protein phosphorylation by high-resolution NMR spectroscopy. *Nature structural & molecular biology*, 15(3):321–9, March 2008.

[170] Jana Hundt, Zhubing Li, and Qiang Liu. Post-translational modifications of hepatitis C viral proteins and their biological significance. *World journal of gastroenterology : WJG*, 19(47):8929–39, December 2013.

[171] Karen E Reed, Jian Xu, and Charles M Rice. Phosphorylation of the hepatitis C virus NS5A protein in vitro and in vivo: properties of the NS5A-associated kinase. *Journal of Virology*, 71(10):7187–7197, 1997.

[172] Manuela Quintavalle, Sonia Sambucini, Vincenzo Summa, Laura Orsatti, Fabio Talamo, Raffaele De Francesco, and Petra Neddermann. Hepatitis C virus NS5A is a direct substrate of casein kinase I-alpha, a cellular kinase identified by inhibitor affinity chromatography using specific NS5A hyperphosphorylation inhibitors. *The Journal of biological chemistry*, 282(8):5536–44, February 2007.

[173] Douglas Ross-Thriepland and Mark Harris. Insights into the complexity and functionality of hepatitis C virus NS5A phosphorylation. *Journal of virology*, 88(3):1421–32, February 2014.

[174] Matthew J Evans, Charles M Rice, and Stephen P Goff. Phosphorylation of hepatitis C virus nonstructural protein 5A modulates its protein interactions and viral RNA replication. *Proceedings of the National Academy of Sciences of the United States of America*, 101(35):13038–43, August 2004.

[175] Takahiro Masaki, Satoko Matsunaga, Hirotaka Takahashi, Kenji Nakashima, Yayoi Kimura, Masahiko Ito, Mami Matsuda, Asako Murayama, Takanobu Kato, Hisashi Hirano, Yaeta Endo, Stanley M Lemon, Takaji Wakita, Tatsuya Sawasaki, and Tetsuro Suzuki. Involvement of Hepatitis C Virus NS5A Hyperphosphorylation Mediated by Casein Kinase I-α in Infectious Virus Production. *Journal of virology*, 88(13):7541–7555, July 2014.

[176] N Blom, S Gammeltoft, and S Brunak. Sequence and structure-based prediction of eukaryotic protein phosphorylation sites. *Journal of molecular biology*, 294(5):1351–62, December 1999.

[177] Yung-Chia Chen, Wen-Chi Su, Jing-Ying Huang, Ti-Chun Chao, King-Song Jeng, Keigo Machida, and Michael M C Lai. Polo-like kinase 1 is involved in hepatitis C virus replication by hyperphosphorylating NS5A. *Journal of virology*, 84(16):7983–93, August 2010.

[178] K E Reed and C M Rice. Identification of the major phosphorylation site of the hepatitis C virus H strain NS5A protein as serine 2321. *The Journal of biological chemistry*, 274(39):28011–8, September 1999.

[179] Daniel G Cordek, Tayler J Croom-Perez, Jungwook Hwang, Michele R S Hargittai, Chennareddy V Subba-Reddy, Qingxia Han, Maria Fernanda Lodeiro, Gang Ning, Thomas S McCrory, Jamie J Arnold, Hasan Koc, Brett D Lindenbach, Scott a Showalter, and Craig E

Cameron. Expanding the proteome of an RNA virus by phosphorylation of an intrinsically disordered viral protein. *The Journal of biological chemistry*, July 2014.