UNIVERSITÀ DEGLI STUDI FIRENZE

DOTTORATO DI RICERCA
INTERNATIONAL DOCTORATE IN STRUCTURAL BIOLOGY

CYCLE XXVIII

COORDINATOR Prof. Claudio Luchinat

Characterizing structural disorder through NMR: new methods and applications

Settore Scientifico Disciplinare CHIM/03

Candidate
Alessandro Piai

_____

Tutors
Prof. Isabella C. Felli

_____

Prof. Roberta Pierattelli

_____

Coordinator
Prof. Claudio Luchinat

_____

November 2012 - 2015

This thesis has been approved by the University of Florence, the University of Frankfurt and the Utrecht University

Universiteit Utrecht

CERM Firenze

FRANKFURT AM MAIN

In memory of
Domenico Piai

Acknowledgments

I would like to express my deepest gratitude to my supervisors and mentors, Prof. *Isabella C. Felli* and Prof. *Roberta Pierattelli*, for their constant guidance, trust and support throughout my doctoral studies. Through their patient teaching and insightful advices I have learnt how to carry out scientific research at high level and how to set and achieve goals. I truly appreciate the working environment they have created in order to let me develop my own ideas while always directing me with their outstanding scientific knowledge. Working with them has enriched me as a scientist and as a person.

I am particularly in debt to Prof. *Wiktor Koźmiński*, Dr. *Wolfgang Bermel* and Prof. *Vladimir N. Uversky* for having believed in my talent since I was a Master student. Beyond the true admiration I feel towards their invaluable scientific skills, more precious is the friendship we have established.

My sincere thanks go to Prof. *Peter Tompa* and Prof. *Xavier Salvatella* for having contributed to my professional grow through inspiring discussions in occasion of our exciting collaborations.

I wish to extend my gratitude to the bright young researchers who has worked with me in my projects: *Bahareh Eftekharzadeh*, *Jan Stanek*, *Sara Contreras Martos*, *Anna Zawadzka-Kazimierczuk* and *Simone Kosol*. With them I built links of trust and friendship that go well beyond the professional relationship.

I am very grateful to the *IDPbyNMR* project for having given me the opportunity to meet and exchange ideas with the most insightful scientists working in the field of biomolecular NMR and intrinsically disordered proteins.

I would like to thank all the people and friends in CERM who have become much more than colleagues to me: *Linda Cerofolini*, *Angelo Gallo*, *Tomáš Hošek*, *Soraya Serrano*, *Vasantha Kumar M.V.*, *Daniela Lalli*, *Tommaso Martelli*, *Mauro Rinaldelli*, *Eduardo O. Calçada* and *Magdalena Korsak*.

Special thanks go to *Fabio Calogiuri*, *Massimo Lucci*, *Enrico Morelli*, *Rebecca Del Conte*, *Leonardo Gonnelli*, *Marco Allegrozzi* and *Roberto Becattini* for prompt technical assistance every time I was in need.

I wish to thank the clever Master students *Thomas Tarenzi*, *Alessandro Del Grande* and *Daniele Mungianu* who helped me a lot in my research with their remarkable learning skills and motivation.

I wish to express warm thanks to my great friends *Jeffrey T. Rubino*, *Gianluca D'Orsi*, *Mauro Perfetti*, *Lorenzo Sernissi* and *Bianca Cecconi*, who have been always at my side and shared with me so many important moments of my life.

Very sweet thanks go to *Yana Valasatava* for her immense kindness, support and care, which have been source of constant joy.

Finally, I wish to deeply thank *Paola Fabrizio*, *Pierluigi Piai*, *Alberto Piai* and all the other members of my family for their never-ending support, understanding and continuous encouragement throughout my entire doctorate.

“Science is a collaborative effort. The combined results of several people working together is often much more effective than could be that of an individual scientist working alone.”

John Bardeen

I wish to thank all the people who made this thesis possible

Contents

List of abbreviations

AP – Anti-Phase
APSY – Automated Projection Spectroscopy
AR – Androgen Receptor
ARE – Androgen Response Element
BEST – Band-Selective Excitation Short-Transient
BMRB – Biological Magnetic Resonance Bank
BROMO – Bromodomain
BT – BEST TROSY
CBP – CREB Binding Protein
CPMG – Car-Purcell-Meiboom-Gill
CREB – Cyclic-AMP-Response-Element-Binding
CS – Compressed Sensing
CSA – Chemical Shift Anisotropy
CSI – Chemical Shift Index
DBD – DNA-Binding Domain
DD – Dipolar Coupling
EDTA – Ethylenediaminetetraacetic acid
FT – Fourier Transform
HAT – Histone Acetyl-Transferase
HEPES – 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HSQC – Heteronuclear Single Quantum Correlation
IDP – Intrinsically Disordered Protein
IDR – Intrinsically Disordered Region
IP – In Phase
IST – Iterative Soft Thresholding
LBD – Ligand-Binding Domain
LCR – Low Complexity Region

LRE – Longitudinal Relaxation Enhancement

MATCH – Memetic Algorithm and Combinatorial Optimization Heuristics

MDD – MultiDimensional Decomposition

ME – Maximum Entropy

MFT – Multidimensional Fourier Transform

MoRF – Molecular Recognition Feature

NCBD – Nuclear-receptor Coactivator-Binding Domain

ncSP – neighbor corrected Structural Propensity

ncSPC – neighbor corrected Structural Propensity Calculator

NMR – Nuclear Magnetic Resonance

NOE – Nuclear Overhauser Effect

NTD – N-Terminal Domain

NUS – Non-Uniform Sampling

PDB – Protein Data Bank

PED – Protein Ensemble Database

PHD – Plant HomeoDomain

PRE – Paramagnetic Relaxation Enhancement

PreSMo – Pre-Structured Motif

PTM – Post-Translational Modification

RDC – Residual Dipolar Coupling

SBMA – Spinal Bulbar Muscular Atrophy

SIFT – Spectroscopy by Integration of Frequency and Time-domain information

SLiM – Short Linear Motif

SMFT – Sparse Multidimensional Fourier Transform

SSA – Signal Separation Algorithm

SSP – Secondary Structure Propensity

TAZ1 – Transcriptional-Adaptor Zinc-finger-1

TAZ2 – Transcriptional-Adaptor Zinc-finger-2

TCEP – Tris(2-carboxyethyl)phosphine

TSAR – Tool for SMFT-based Assignment of Resonances

TROSY – Transverse Relaxation-Optimized Spectroscopy

ZFP – Zinc Finger Protein

Abstract

Intrinsically Disordered Proteins (IDPs) are flexible proteins that challenge structural biology due to the fact that they cannot be studied with the standard methods developed to characterize well-folded proteins. In the last decades, the discovery of their widespread presence and involvement in many biological functions, despite their deviation from the structure-function paradigm, has pushed the scientific community towards a growing acceptance of the importance of IDPs and to the development of new tools for studying their structure, dynamics and functions. In this context, Nuclear Magnetic Resonance (NMR) spectroscopy plays the leading role of most suitable technique to characterize IDPs.

In this doctoral thesis, my contribution to the advancement of NMR spectroscopy, achieved by developing new experiments to study IDPs, is described. The new methods enable the characterization of structural disorder and allow to address topics of general interest such as the study of protein linkers and low-complexity regions, two areas of high biological relevance for which only very limited atomic resolution information is available so far.

The methodological progress in NMR and the findings on the studied IDPs reported here give a contribution to the discovery of new roles for structural disorder and prompt towards an unified understanding of structure-dynamics-disorder/function relationships.

Keywords

Intrinsically Disordered Proteins · Nuclear Magnetic Resonance · ^{13}C detection · Non-Uniform Sampling · Longitudinal Relaxation Enhancement · α-synuclein · CREB Binding Protein · Androgen Receptor

Aim of the thesis and summary of the project

The recent birth of the Protein Ensemble Database (PED)[1] (Varadi et al., 2014), in which experimental data and calculated structure ensembles can be deposited, shows the increasing awareness of the importance of focusing on the high-resolution investigation of Intrinsically Disordered Proteins (IDPs). Indeed, with respect to the large variety of studies on globular and well-folded proteins, the investigation and characterization of structural disorder are still at their early stages. Only a limited amount of high-resolution structural and dynamic information is available on highly flexible disordered systems. The molecular basis of the mechanisms through which IDPs perform their biological functions are still largely unknown.

The aim of this thesis has been to contribute to this exciting area of research by developing a suite of new NMR experiments to allow fast and efficient study of highly flexible IDPs. The new methods, besides making the experimental characterization of IDPs easier and more accurate, aim to open new investigation avenues for researches that would not have been possible otherwise. They enable the structural and dynamic investigations at high-resolution of large and heterogeneous IDPs and allow focusing on general themes of high biological importance that have not been addressed so far.

Functional studies of IDPs have illustrated the broad impact of disorder on many areas of cellular biology, ranging from signaling processes to transcriptional and regulation mechanisms. However, malfunctioning of IDPs is also known to play key roles in many neurodegenerative diseases, thus giving biomedical relevance to this field of study.

[1] http://pedb.vib.be/

Among the many open questions in the field of IDPs, I have specifically addressed through the new methods the study of the roles of flexible protein linkers and of the so-called “low-complexity regions” in large proteins. These two general themes have been studied by focusing on two selected examples of high biomedical relevance, i.e. the disordered regions of the CREB Binding Protein and the N-terminal domain of the Androgen Receptor. The results highlight new roles for protein function, strongly related to the highly heterogeneous structural and dynamic properties of the proteins.

After an introductory chapter on IDPs and NMR, the results obtained during my doctoral studies along these two specific lines of research are described in the context of the scientific literature. In chapter 2, after reviewing the available methodologies and carefully investigating how to optimize the NMR parameters to study IDPs, the appropriate approaches to improve the NMR methodology for the study of IDPs are described; the experimental design and the assignment strategy tailored to characterize highly disordered systems are illustrated. In chapter 3, the thorough characterization of the studied linkers and low-complexity regions, achieved thanks to the developed methods, is presented; the findings on the two topics of general interest that have been addressed open new perspectives for the general understanding of how the highly heterogeneous structural and dynamic properties of proteins influence their function, in a much broader and more sophisticated way with respect to what is known so far. The scientific publications and review articles, presenting the results obtained in a more extended manner, are provided in the Appendix.

Chapter 1

Introduction

Structural biology is the study of the structure and function of biomolecules in order to understand their relationship. At its basis there is the structure-function paradigm, which explains that the three-dimensional structure of a molecule determines and influences its function. For example, the folding of a polypeptide allows specific residues to be spatially close as to provide local and specific properties to the molecule. This concept was derived from early studies examining the specificity of enzymes and the structures of folded proteins determined primarily via crystal X-ray diffraction. However, this picture is not complete, as structural disorder has been identified to play a role of equal importance in determining the properties of a molecule.

With the recent discovery of structural disorder, structural biology is now asked new questions whose answers go beyond the structure-function paradigm (Figure 1.1). It was discovered that molten globules and disordered protein states can exploit their high flexibility to perform specific biological functions which can have also peculiar origins, like disorder-to-order transitions or form fuzzy complexes, or to enable interactions with multiple partners when highly specific and low-affinity interactions are required for reversible binding (Uversky et al., 2000; Demchenko, 2001; Uversky, 2002a; Dyson et al., 2002; Dyson et al., 2005; Habchi et al., 2014; van der Lee et al., 2014; Tompa et al., 2014).

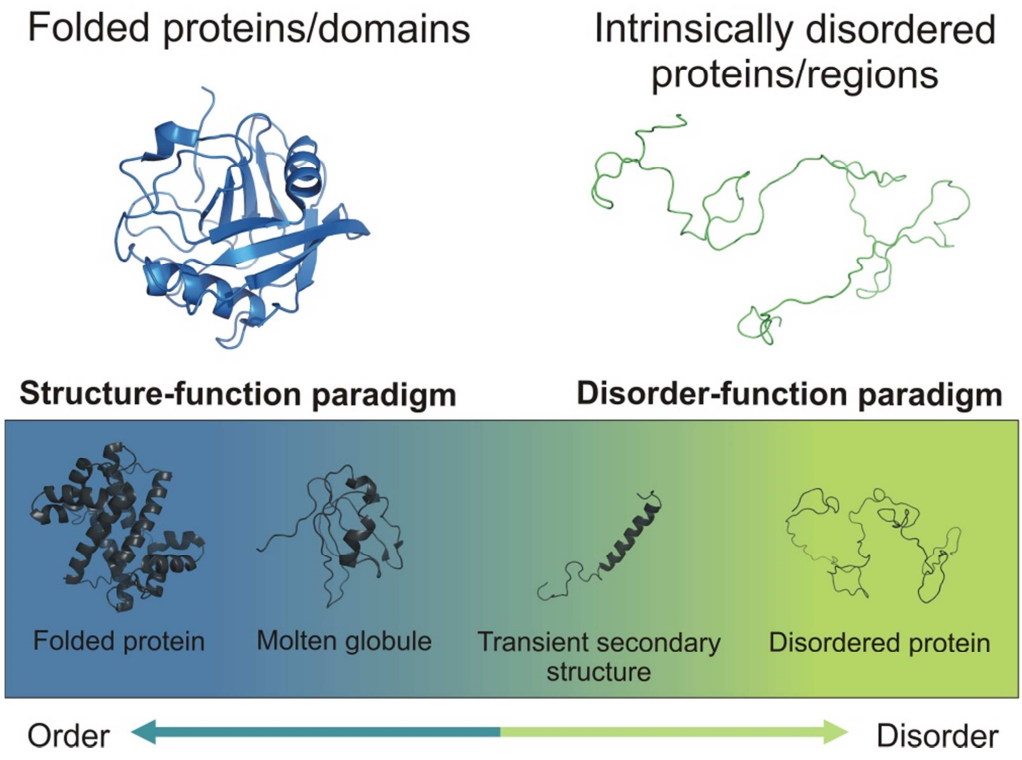


Figure 1.1 Folded proteins/domains and intrinsically disordered proteins/regions represent the two opposite extreme cases of three-dimensional organization of proteins. The interplay between structure and disorder increases the functional versatility of proteins through the so-called continuum model of protein structure, schematically shown in the bottom of the figure. The color gradient represents a continuum of conformational states ranging from compact, globular, folded states (blue) to dynamic, flexible and disordered conformational ensembles (green). As highlighted, the established structure-function paradigm depicts only a portion of the possible protein functions. Therefore, understanding how nature exploits structure and disorder will be crucial for uncovering the full extent of protein functions.

Classical biochemical methods to isolate and purify proteins are strongly biased towards the production and characterization of folded and active proteins. Indeed, standard preparation methods calls for the isolation and homogenization of plant or animal tissues, which are then assayed for the activity of interest before the active protein is purified. This methodology highly favors isolation of folded proteins, as during homogenate formation release of proteases results in the degradation of disordered proteins, which are much more sensitive than folded proteins in these conditions. Thanks to recent advances in bioinformatics, scientists have now access to a vast library of gene sequences, which can be

exploited using recombinant technology to express and purify disordered proteins. The study of these particular proteins, the so-called Intrinsically Disordered Proteins (IDPs), is the subject of this doctoral thesis.

1.1 Intrinsically Disordered Proteins

IDPs are a peculiar class of proteins characterized by the absence of a single and stable tridimensional structure in the physiological state. In solution these proteins possess multiple conformations separated by low energy barriers, therefore their structure constantly interconverts between different states (Tompa, 2005; Uversky et al., 2010). Their discovery was a breakthrough in the structural biology field as it opened new avenues to the comprehension of their physiological relevance in mechanisms at the basis of fundamental biological processes (Xie et al., 2007; Vucetic et al., 2007; Uversky, 2011a; Uversky, 2013; Wright et al., 2015). In fact, the high flexibility of IDPs allows them to perform a multitude of functions in the cell, which range from gene regulation to signaling processes (Wright et al., 1999; Iakoucheva et al., 2002; Tompa, 2002; Iakoucheva et al., 2004); structural disorder provides a unique tool to modulate and fine-tune highly-specific interactions. The lack of a stable tertiary structure exposes the protein backbone so that Short Linear Motifs (SLiMs) are allowed to interact with different partners (Fuxreiter et al., 2007; Van Roey et al., 2014) and Post-Translational Modifications (PTMs) sites become accessible for reaction (Lu et al., 2002; Theillet et al., 2012); transient local structural elements and pre-structured motifs (PreSMos) act as specificity determinants for recognition by target proteins (Fuxreiter et al., 2004; Lee et al., 2012; Lee et al., 2014); highly-specific and interacting-prone disordered regions called Molecular Recognition Features (MoRFs) (Vacic et al., 2007) fold upon binding and exploit disorder-to-order transitions to provide new functions to the protein. Although IDPs are involved in many physiological pathways, they are also found to have direct implications in several diseases and pathologies involving amyloid formation, like Alzheimer, or

different types of cancers (Masters et al., 1985; Bellotti et al., 1999; Iakoucheva et al., 2002; Dev et al., 2003; Uversky et al., 2008; Uversky, 2011b). This makes this class of proteins absolutely worth studying.

IDPs are quite abundant in living organisms. Bioinformatics studies predict that IDPs and Intrinsically Disordered Regions (IDRs) longer than 30 residues constitute about 30% of the eukaryotic genomes and that this number increases with the complexity of an organism (Dunker et al., 2000; Dunker et al., 2002; Uversky, 2002b; Ward et al., 2004). Such studies identify IDPs by targeting peculiar properties in their amino acid composition: bias towards the so-called disorder-promoting residues (Dunker et al., 2001; Vucetic et al., 2003; Hansen et al., 2006), high abundance of glycine and proline residues (Theillet et al., 2013), preference for charged and hydrophilic residues rather than hydrophobic ones (Romero et al., 2001). Interestingly, amino acidic sequences of IDPs and IDRs correspond to those for which no electron density is available in the Protein Data Bank (PDB)[2] (Berman et al., 2000) and, consequently, no structural information is known. On these basis, several structural-disorder predictors have been developed to recognize IDPs from their amino acid sequence, like PONDR-FIT[3] (Xue et al., 2010), IUPred[4] (Dosztányi et al., 2005), FoldIndex[5] (Prilusky et al., 2005), DisEMBL 1.5[6] (Linding et al., 2003b), GlobPlot 2.3[7] (Linding et al., 2003a) and DISOPRED3[8] (Jones et al., 2015). These tools have been used to drive the design of laboratory experiments to target structural disorder so that, at the present time, several hundreds of proteins have been shown experimentally to be either completely or partially disordered. According to the latest release of DisProt[9] (May 2013) (Sickmeier et al., 2007), an established database of protein

[2] http://www.rcsb.org/pdb/home/home.do
[3] http://www.disprot.org/pondr-fit.php
[4] http://iupred.enzim.hu/
[5] http://bip.weizmann.ac.il/fldbin/findex
[6] http://dis.embl.de/
[7] http://globplot.embl.de/
[8] http://bioinf.cs.ucl.ac.uk/psipred/?disopred=1
[9] http://www.disprot.org/

disorder which collects experimental data on IDPs/IDRs, 1539 disordered regions have been experimentally identified in 694 proteins.

1.2 NMR: a strategic tool to study IDPs

Nuclear Magnetic Resonance (NMR) spectroscopy is the most powerful technique to study IDPs. NMR enables the characterization of IDPs and the investigation of their biological functions thanks to its unique ability to provide structural and dynamic information at atomic resolution (Ishima et al., 2000; Kay, 2005; Bonvin et al., 2005; Mittag et al., 2007; Kleckner et al., 2011; Jensen et al., 2014; Parigi et al., 2014). Contrary to X-ray crystallography, NMR does not require the protein of interest to be crystallized, but allows for study while in solution. In fact, the high flexibility of IDPs usually prevents crystallization, although they may be suitable candidates for X-ray crystallography when/if tightly bound to an interacting partner. Therefore, NMR provides the great advantage of examining IDPs in a more physiologically relevant context, in which they are free to adopt different conformations.

For most applications, NMR requires the biomolecule to be isotopically enriched (commonly in ^{13}C and ^{15}N, but also in ^{2}H in particular conditions). The replacement of NMR-silent nuclei with isotopes with an experimentally-suitable nuclear magnetic moment allows the collection of chemical shifts, the frequencies at which nuclei resonate. Chemical shifts values provide structural information, including how atoms are bound and their local chemical environment. Once chemical shifts are associated to the originating nuclei through the so-called sequence-specific resonance assignment, they can be used to derive structural information or monitor events like protein interactions at atomic resolution (Bonvin et al., 2005; Cavalli et al., 2007; Shen et al., 2008). Particularly valuable for IDPs is the collection of backbone C^{α} and C^{β} chemical shifts, as the difference between experimental and theoretical random-coil values directly correlates with the presence of protein secondary structures

(Spera et al., 1991). This analysis, termed Chemical Shift Index (CSI) (Wishart et al., 1994; Wishart et al., 1995), can be used to identify and determine the location of residual structural elements in IDPs (Kragelj et al., 2013). CSIs from different nuclei can also be combined to obtain an overall Secondary Structure Propensity (SSP) score (Marsh et al., 2006; Tamiola et al., 2012), which describes the conformational sampling of the protein. Recently, the establishment of a sequence-specific random-coil chemical shift library of IDPs has improved the accuracy of this analysis (Tamiola et al., 2010).

In addition to chemical shifts, there are many other sources of information provided by NMR. Intensities and volumes of NMR signals can encode observables which complement the information provided by chemical shifts and allow a full characterization of the protein under investigation (Bertini et al., 2012; Brutscher et al., 2015). Examples include, ^{15}N relaxation parameters (^{15}N T_1, ^{15}N T_2, ^{15}N{^{1}H} NOEs) (Kay et al., 1989; Barbato et al., 1992; Farrow et al., 1994; Peng et al., 1994; Farrow et al., 1995) allow the study the dynamic properties of the protein, as they are sensitive to motions occurring from the pico- to nanosecond timescale. Being sensitive to fast local motions, they allow to identify protein regions characterized by different extent of mobility and thus can be used to validate structural data. Complementary dynamic information on the millisecond timescale can be retrieved by performing Car-Purcell-Meiboom-Gill (CPMG) relaxation dispersion measurements (Dittmer et al., 2004). ^{1}H-^{1}H Nuclear Overhauser (NOE) distance constraints can be collected to identify and calculate the structure of secondary structure elements (Wüthrich, 1986). Determination of ^3J scalar couplings, like $^3J_{\text{HN-H}\alpha}$, provides information about the dihedral torsion angles of the protein backbone through Karplus equations (Vuister et al., 1993; Bax et al., 1995; Wang et al., 1996; Griesinger et al., 1999; Vögeli et al., 2007; Li et al., 2015). Quantification of chemical exchange at amide sites reports about compactness and solvent exposure of the protein (Hwang et al., 1998; Melacini et al., 1999; Schanda et al., 2006a; Bertini et al., 2011;

Rennella et al., 2014). Measurement of Residual Dipolar Couplings (RDCs) yields orientational restraints, which describe the spatial arrangement of different domains (Louhivouri M. et al., 2003; Bernadó et al., 2005; Obolensky et al., 2007; Salmon et al., 2012). Paramagnetic Relaxation Enhancement (PRE) provides long-range distance information that can complement NOE restraints by deriving distances between the spin label and the NMR active nuclei from increased R_2 relaxation rates (Bentrop et al., 1997; Bertini et al., 2001a; Bertini et al., 2001b; Teilum et al., 2002; Dedmon et al., 2005; Bertoncini et al., 2005; Allison et al., 2009). They are particularly suited also to monitor weak protein-protein interactions, provided that the paramagnetic tag engineered in the protein does not alter its properties (Wu et al., 2010).

These and other observables constitute the wealth of high-resolution information provided by NMR. The possibility to detect local residual secondary structures and obtain a full description of the dynamic properties of the protein under investigation renders this technique extremely useful for the study of highly dynamic systems such as IDPs.

1.3 Challenges of IDPs to NMR

NMR is the most appropriate technique to study IDPs at atomic resolution. However, the peculiar properties of IDPs often complicate, even dramatically, their spectroscopic characterization. The high flexibility and the lack of a stable tridimensional structure have a strong impact on the resulting NMR spectra.

The possibility to explore a vast conformational space leads to an averaging of most of the contributions to the chemical shifts. As a consequence, chemical shifts of the same amino acid type result poorly differentiated. NMR spectra of IDPs are thus characterized by a smaller chemical shift dispersion with respect to analogues of folded proteins, especially in the proton dimension (Figure 1.2). The

great extent of cross-peak overlaps allows standard triple-resonance 3D experiments to resolve just a small fraction of the NMR signals.

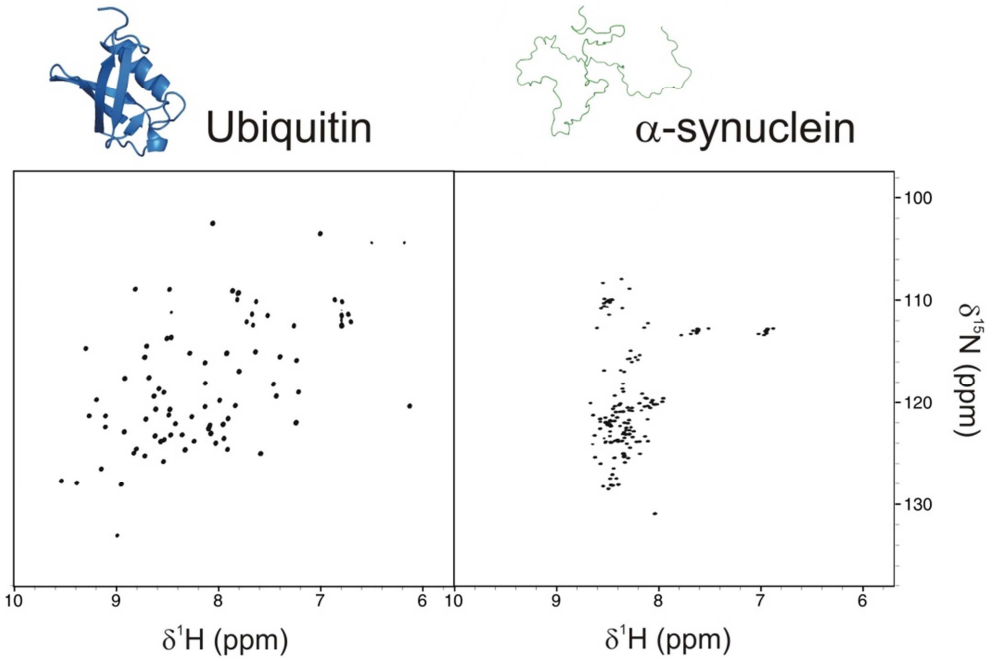


Figure 1.2 Spectroscopic properties of folded and disordered proteins. The different chemical shift dispersion of folded proteins and IDPs is highlighted by comparing the 2D ^{1}H-^{15}N Heteronuclear Single Quantum Correlation (HSQC) spectra of human ubiquitin (left) and α-synuclein (right). The spectra, reported with identical scales in Hz, were acquired on a ^{13}C, ^{15}N-labeled sample of human Ubiquitin (0.5 mM Ubiquitin, 20.0 mM HEPES, pH 7.0, 298.0 K) and on a ^{13}C, ^{15}N-labeled sample of human α-synuclein (1.0 mM α-synuclein, 20 mM Pi, 200 mM NaCl, 0.5 mM EDTA, pH 6.5, 285.5 K) with a Bruker Avance III 950 MHz spectrometer equipped with a cryogenically-cooled triple-resonance probe. Despite their different size, ubiquitin and α-synuclein were chosen in the example since they are well-established standard samples routinely used to setup NMR experiments.

The solvent-exposed backbone hampers the formation of intra-molecular hydrogen bonds, resulting in $^{1}H^{N}$ proton nuclei pronounced chemical exchange at amide sites, especially when physiological conditions are approached (high temperature and pH). This reduces the sensitivity of the NMR experiments and causes extensive line broadening, further decreasing the spectral resolution. For fast proton exchange rates, cross-peaks are often broadened even beyond detection. As a consequence, routine NMR experiments based on $^{1}H^{N}$ detection fail to provide complete information (Figure 1.3).

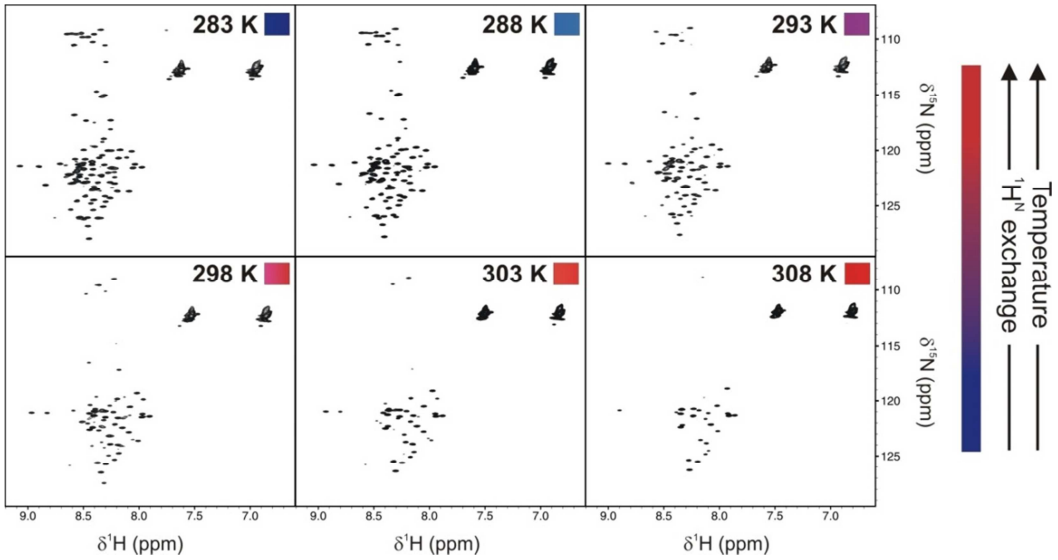


Figure 1.3 Increasing chemical exchange at amide sites of IDPs with consequent line broadening, even beyond detection, when approaching physiological conditions. The 2D ^{1}H-^{15}N HSQC spectra were acquired at different temperatures on a ^{13}C, ^{15}N-labeled sample of the N-terminal domain of human Androgen Receptor "4Q" (0.4 mM 4Q, 20.0 mM Pi, 1.0 mM TCEP, pH 7.4) with a Bruker Avance 700 MHz spectrometer equipped with a cryogenically-cooled triple-resonance probe optimized for ^{13}C direct detection. All the spectra are plotted with the same contour levels.

The biased amino acid composition causes amino acid sequences of IDPs to be highly repetitive. This complicates the sequence-specific resonance assignment, as many ambiguities arise due to the redundancy of amino acid patterns. Proline residues, very abundant in IDPs as their peculiar structure is not compatible with secondary structure elements, further hinder the chemical shift assignment. Indeed, this residue is not observable in experiments employing amide proton detection, breaking chains of sequential connectivities.

Combination of all the above factors makes the characterization of IDPs by NMR a challenging task. Therefore, in recent years a variety of NMR experiments tailored for IDPs has been developed (Panchal et al., 2001; Hiller et al., 2007; Motáčkova et al., 2010; Zawadzka-Kazimierczuk et al., 2010; Mäntylahti et al., 2010; Narayanan et al., 2010; Nováček et al., 2011; Mäntylahti et al., 2011; Bermel et al., 2012b; Zawadzka-Kazimierczuk et al., 2012b; Bermel et al., 2013b; Kazimierczuk et al., 2013; Sólyom et al., 2013; Pantoja-Uceda et al., 2013a; Piai et al., 2014; Pantoja-Uceda et al., 2014; Dziekański et al., 2015; Żerko et al., 2015).

Chapter 2

Methodological advancement: improving NMR experiments to characterize IDPs

Implementation of proper measures while designing the experiments may mitigate the detrimental effects impacting the quality of NMR spectra of IDPs which derive from the high mobility of disordered polypeptides. Indeed, the exploitation of the methods described in the next paragraphs makes a difference in the feasibility of the NMR characterization of IDPs, especially when complex and large proteins are studied (Dyson et al., 2001; Dyson et al., 2004; Felli et al., 2014; Konrat, 2014; Nováček et al., 2014; Brutscher et al., 2015).

In recent years, several methods have been successfully developed to study IDPs. These methods should not be regarded in a competitive context, but rather as a great resource that allows for more possible studies of IDPs. Each method has its particular advantages and disadvantages; some approaches may be more suitable than others depending sample conditions. In fact, use of more than one of the methods would be preferable as together they could provide complementary data, and ultimately a more thorough understanding of IDPs.

2.1 Theoretical background and technical aspects

As discussed in section 1.2, NMR represents a unique spectroscopic tool for the high-resolution investigation of IDPs. Before starting any methodological development, it is important to study and review the state of the art of the available techniques. Indeed, the most appropriate already-existing approaches

can be employed and/or optimized to achieve the desired goals; new methods can be developed based on the knowledge of the already established ones. With the aim to develop new NMR experiments to study disordered systems, the most recent NMR advancements which were proven useful to deal with the spectroscopic features of IDPs have been reviewed in two book chapters.

In the first one, the main focus is on the recently developed NMR methods based on ^{13}C direct detection, which were developed to provide additional complementary information to that available through ^{1}H detected NMR experiments. While applicable as a general tool for biomolecular applications, the suite of ^{13}C detected experiments can provide unique additional information on proteins when ^{1}H detection finds limitations, such as in paramagnetic proteins, very large macromolecular assemblies or proteins characterized by a high extent of disorder, flexibility and solvent exchange, such as IDPs. This first review comprises a broad range of applications of the suite of ^{13}C detected experiments and provides an overall description of the experiments including key technical aspects (Appendix – Article 1).

The second review focuses on the optimal methods for the investigation of IDPs, comprising both ^{13}C and ^{1}H detected experiments. Different aspects ranging from the basic principles of NMR spectroscopy, including hardware requirements, to the design and application of complex NMR experiments, are discussed. Practical guidelines to the collection and interpretation of NMR data are given, describing the currently best strategies and approaches to obtain structural and dynamic properties of highly disordered systems. In particular, benefits provided both by ^{1}H and ^{13}C detection are highlighted considering the intrinsically different properties of these two different nuclei, which provide the rationale for a great variety of developed experiments. The technical aspects that need to be solved, as well as the many experimental variants available which can be exploited to study IDPs, are presented (Appendix – Article 2).

2.2 Increasing the chemical shift dispersion

The small chemical shift dispersion observed in the NMR spectra of IDPs is one of the major drawbacks for their analysis. The ability to resolve resonances is very important to extract information at atomic resolution. Therefore, approaches able to increase spectral resolution and/or reduce spectral overlaps are recommended to be used.

Studies of unfolded protein states have shown that heteronuclei exhibit greater chemical shift dispersion with respect to proton nuclei ($^{1}H^{N}$ < $^{13}C'$ < ^{15}N) (Yao et al., 2001; Wirmer et al., 2006; Gerum et al., 2009). Indeed, the proton dimension in NMR spectra of IDPs is the most collapsed (Figure 2.1). Consequently, the superior chemical shift dispersion of heteronuclei should be exploited by preferring to acquire ^{13}C and ^{15}N dimensions. In ^{1}H detected experiments, this is done by labeling chemical shifts of heteronuclei in the indirect dimensions. Alternatively, ^{13}C detection represents a valuable option, as it provides high-resolution also in the direct dimension. Therefore, this approach is recommended in order to diminish accidental cross-peak overlaps in the spectrum. However, this effect is particularly pronounced in the case of IDPs. In fact, chemical shift dispersion of amide proton nuclei and carbonyl carbon nuclei of folded proteins is comparable, but the acquisition of the former provides more experimental sensitivity.

Spectral resolution can also be improved by designing experiments in which detected correlations involve neighboring residues. Indeed, due to the lack of a stable structure, contributions to the chemical shifts for a given amino acid type are almost identical with the consequence that chemical shifts result very similar. Nonetheless, if contributions deriving from neighboring residues are exploited by transferring magnetization from/to them, NMR signals result more disperse in the spectrum and can be better discriminated (Bermel et al., 2013a; Brutscher et al., 2015).

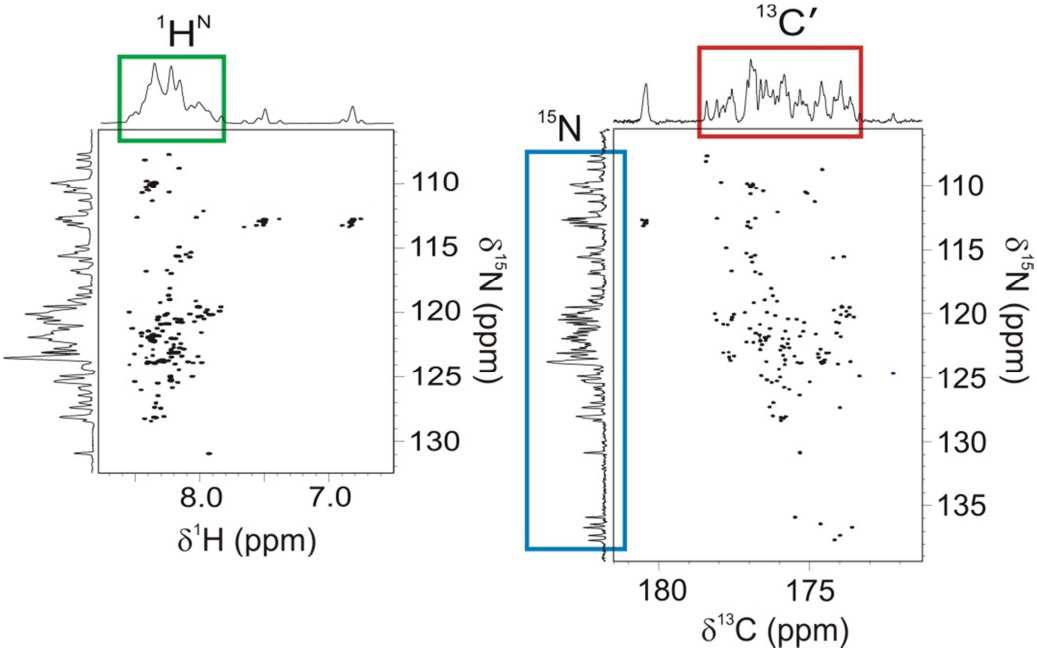


Figure 2.1 Chemical shift dispersion in IDPs. The 2D ^{1}H-^{15}N HSQC and 2D CON spectra of human α-synuclein are shown. Both spectra were acquired on a ^{13}C, ^{15}N-labeled sample of α-synuclein (1.0 mM α-synuclein, 20 mM Pi, 200 mM NaCl, 0.5 mM EDTA, pH 6.5, 285.5 K) with a Bruker Avance 700 MHz spectrometer equipped with a cryogenically-cooled triple-resonance probe optimized for ^{13}C direct detection. The different chemical shift dispersions of the most exploited backbone nuclei, namely $^{1}H^{N}$, $^{13}C'$ and ^{15}N, are highlighted in the 1D projections of the spectra (green, red and blue boxes, respectively). The spectra are reported with identical scales in Hz.

Finally, spectral resolution can be increased by acquiring multidimensional (≥4D) NMR experiments in which cross-peaks are separated in multiple indirect dimensions (Narayanan et al., 2010; Motáčkova et al., 2010; Zawadzka-Kazimierczuk et al., 2010; Nováček et al., 2011; Bermel et al., 2012b; Zawadzka-Kazimierczuk et al., 2012b; Kazimierczuk et al., 2013; Bermel et al., 2013b; Piai et al., 2014; Nowakowski et al., 2015; Dziekański et al., 2015; Żerko et al., 2015). Alternatively, spectral crowding can be simplified by performing amino acid-selective experiments, providing NMR spectra in which only signals belonging to particular amino acids, selected on the basis of their topology, are detected (Dötsch et al., 1996b; Dötsch et al., 1996c; Dötsch et al., 1996a; Dötsch et al., 1996d; Grzesiek et al., 1992; Schubert et al., 2001a; Schubert et al., 2001b; Schubert et al., 2001c; Schubert et al., 2005; Lescop et al., 2008; Pantoja-Uceda

et al., 2008; Bermel et al., 2012a; Pantoja-Uceda et al., 2012; Pantoja-Uceda et al., 2013b).

2.3 ^{13}C detection

^{13}C detection is a widely-established technique commonly used in the study of small molecules. Nowadays, it is usefully applied also in several fields of biomolecular NMR, like in the characterization of large and/or paramagnetic proteins and for solid-state applications. In recent years, ^{13}C detection has become a key tool in the research field of IDPs (Bermel et al., 2006; Csizmok et al., 2008; Nováček et al., 2011; Felli et al., 2013; Felli et al., 2014; Pantoja-Uceda et al., 2014; Dziekański et al., 2015).

The greatest advantage provided by ^{13}C detection relies on the superior chemical shift dispersion of ^{13}C with respect to ^{1}H when studying flexible systems such as IDPs (Figure 2.2). This contributes considerably to the enhancement the spectral resolution by reducing accidental cross-peaks overlap. However, this comes at the price of lower sensitivity, due to smaller gyromagnetic ratio of ^{13}C with respect to ^{1}H. Nowadays, thanks to major improvements in instrumental technologies, ^{13}C detection guarantees remarkable results if sample concentration exceeds 200-300 µM, as in many situations of practical interest (Ardenkjaer-Larsen et al., 2015). Therefore, it represents the ideal tool for investigation of IDPs due to a good compromise between resolution and sensitivity, in particular when physiological conditions are approached (Gil et al., 2013). Indeed, carbon nuclei are insensitive to chemical exchange processes, contrary to amide proton nuclei which experience severe line broadening at high temperature and pH. Moreover, carbon nuclei are less-sensitive to detrimental effects of high salt concentrations, often required to prevent sample aggregation (Serber et al., 2001). Finally, ^{13}C detection allows to directly observe proline residues, very abundant in IDPs and therefore precious sources of information towards a complete characterization of the protein under investigation.

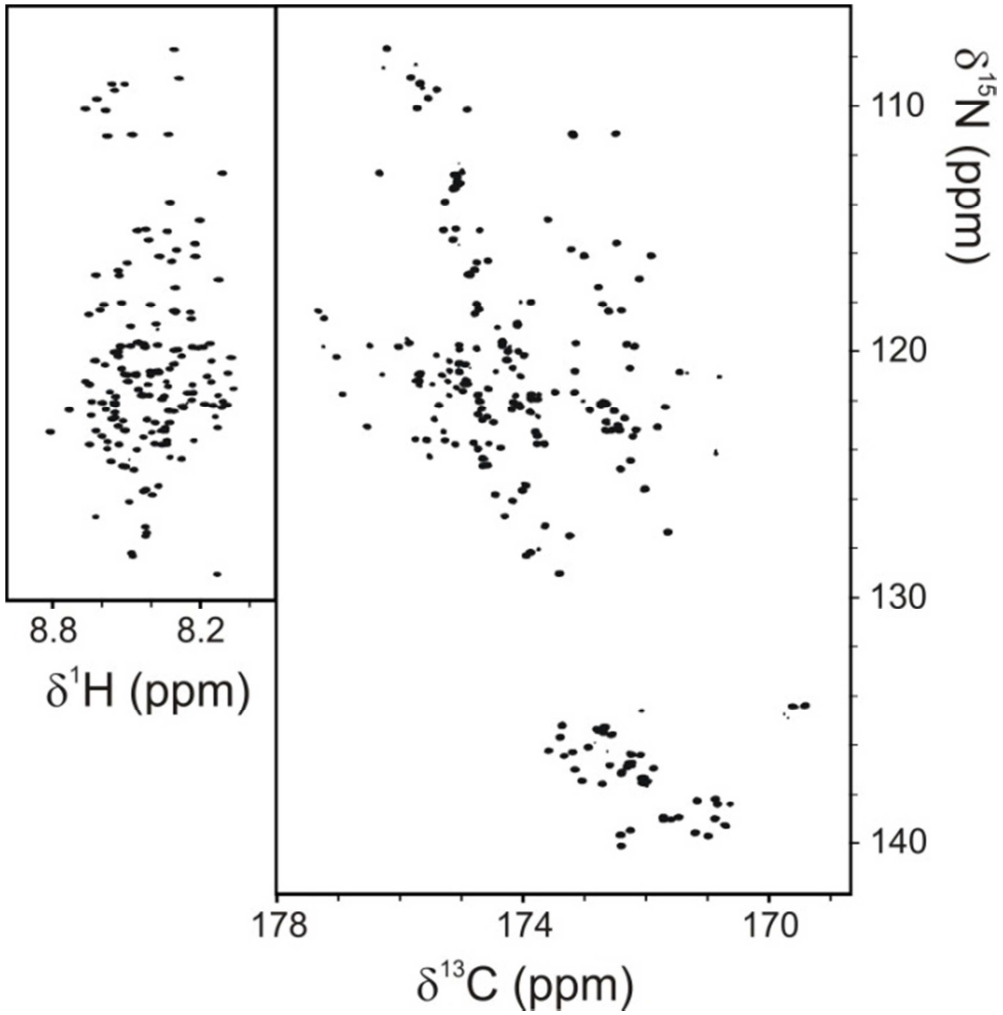


Figure 2.2 The 2D ^{1}H-^{15}N HSQC (left) and 2D CON (right) spectra of the fourth disordered region of human CREB Binding Protein "ID4" are shown. The spectra, reported with identical scales in Hz, emphasize the superior chemical shift dispersion provided by ^{13}C detection with respect to ^{1}H detection. At 16.4 T, the resolution in the direct dimension increases from about 500 Hz to 1200 Hz. Noteworthy, ^{13}C detection allows to observe proline residues. The experiments were performed on a ^{13}C, ^{15}N-labeled sample of ID4 (0.6 mM ID4, 20 mM Pi, 100 mM KCl, pH 6.5, 283.0 K) with a Bruker Avance 700 MHz spectrometer equipped with a cryogenically-cooled triple-resonance probe optimized for ^{13}C direct detection.

Differently from proton nuclei, homonuclear couplings between carbon nuclei are very large. This results in the presence of significant line splittings in the forms of doublets (^{13}C′ detection) or doublets of doublets ($^{13}C^{\alpha}$ detection) in the direct dimension of the spectra. For this reason, clever strategies to achieve homonuclear decoupling have been developed (Shimba et al., 2003; Bermel et al., 2008; Ying et al., 2014). Among them, virtual decoupling methods such as

IPAP, DIPAP and S^3E have become the most widely used due to their efficient and robust performance (Figure 2.3) (Bermel et al., 2005; Bermel et al., 2008; Felli et al., 2015). Of course, these methods rely on the assumption that homonuclear couplings have almost identical magnitude through all the protein, which is the case of the large one-bond couplings between α and carbonyl carbon nuclei ($^1J_{C'C\alpha} \approx 52$ Hz) and α and β carbon nuclei ($^1J_{C\alpha C\beta} \approx 35$ Hz).

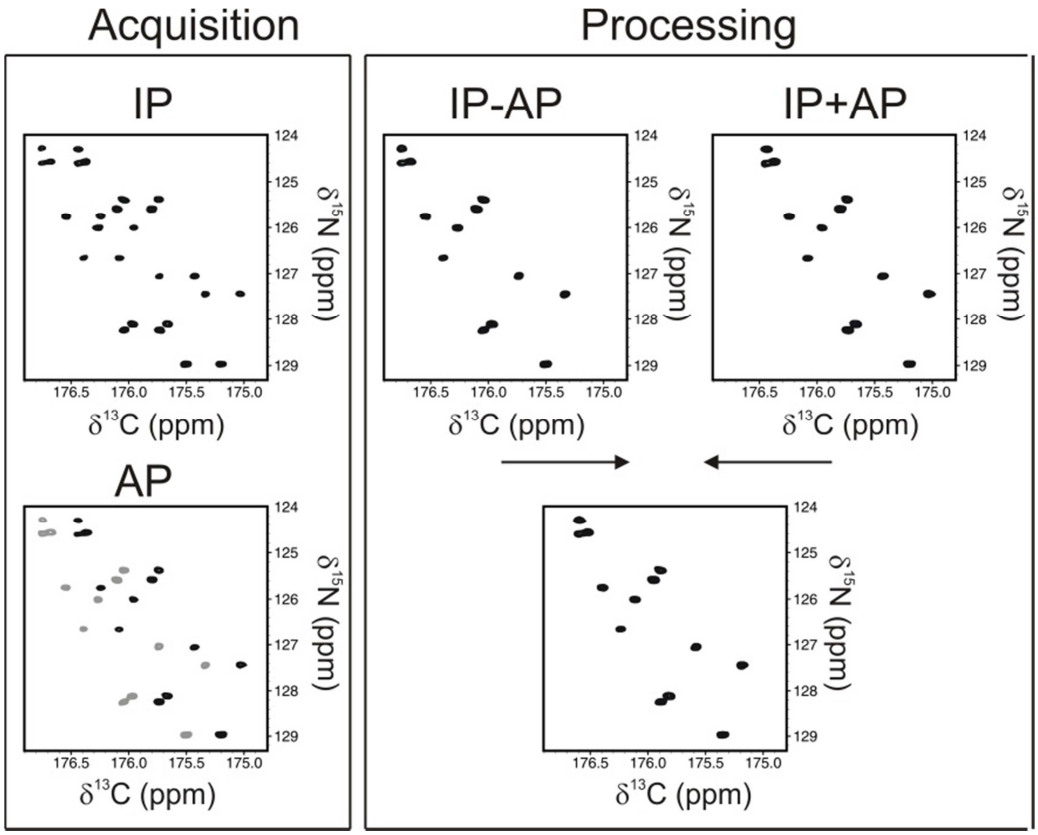


Figure 2.3 C' virtual decoupling IPAP scheme for the removal of $^1J_{C'C\alpha}$ homonuclear scalar couplings. A portion of the 2D CON spectrum of human ID4, acquired in the experimental conditions reported in the caption of Figure 2.2, is shown to exemplify the use of the IPAP scheme. First, the In Phase (IP) and Anti-Phase (AP) components of the signal, reported on the left, are acquired (negative cross-peaks are shown in grey). Then, as shown on the right, linear combinations are performed to separate the two multiplet components, which are shifted to the center of the original multiplet by $^1J_{C'C\alpha}/2$ and finally summed together, thus removing the splitting due to the one-bond homonuclear scalar coupling (bottom right).

Experimental sensitivity of ^{13}C detected experiments can be enhanced by exploiting proton nuclei as starting source of polarization. In this way, it is

possible to take advantage of the higher gyromagnetic ratio of proton nuclei while still retaining the experiment "*exclusively heteronuclear*" (Bermel et al., 2009b). The most convenient pool of proton nuclei ($^{1}H^{N}$, $^{1}H^{\alpha}$ or $^{1}H^{ali}$) can be chosen according to the particular design of the NMR experiment (Bermel et al., 2009b; Hošek et al., 2015). However, selection of amide proton nuclei will result in a loss of correlations originating from proline residues.

An additional motivation for the exploitation of aliphatic proton nuclei as starting source of polarization comes from the fact that recycle delays of the experiments can be shortened, as proton nuclei exhibit faster longitudinal recoveries with respect to carbon nuclei. Further improvements can be achieved by implementing techniques for Longitudinal Relaxation Enhancement (LRE) (Schanda et al., 2005; Schanda et al., 2006b; Lescop et al., 2007; Bermel et al., 2009a; Hošek et al., 2015). For example, recovery rates of amide proton nuclei can be enhanced by using either the H^{N}-flip (Bermel et al., 2009a; Bertini et al., 2011) or the Band-Selective Excitation Short-Transient (BEST) approach (Schanda et al., 2006b; Lescop et al., 2007).

2.4 ^{1}H detection

A great number of NMR experiments for structural biology applications is based on ^{1}H detection. Despite the fact that it provides less resolution with respect to ^{13}C detection, it can be still a valuable option to study IDPs, especially when dealing with low concentrated or short-living samples. Indeed, ^{1}H detection provides very good experimental sensitivity thanks to the higher gyromagnetic ratio of ^{1}H. $^{1}H^{N}$ detection, the most established, can be applied to IDPs in conditions able to prevent or diminish chemical exchange rates of amide proton nuclei, i.e. at low temperature and pH. The great extent of cross-peaks overlap, due to the narrow chemical shift dispersion of $^{1}H^{N}$, can be partially mitigated by using Transverse Relaxation-Optimized Spectroscopy (TROSY) (Pervushin et al., 1997). This approach is based on the interference between two spin relaxation

mechanisms, the Dipolar Coupling (DD) and the Chemical Shift Anisotropy (CSA), in scalar coupled $^{1}H^{N}$ and N backbone nuclei. For each component of the multiplet, this interference can be either constructive or destructive, resulting in a faster or slower apparent relaxation rate, respectively. TROSY enhances spectral resolution by retaining only the slowest relaxing component of each cross-peak, thus providing sharper lines. TROSY-based NMR experiments can be combined with BEST, providing a twofold advantage: i) LRE is achieved and faster experiments can be performed; ii) the loss in sensitivity deriving by the application of TROSY can be regained by acquiring a higher number of transients in a given amount of time (Sólyom et al., 2013). Additionally, the simultaneous application of both approaches allows to further increase the experimental sensitivity by enhancing ^{15}N polarization through the recovery of undetectable ^{1}H polarization which originates from longitudinal relaxation during the pulse sequence (Favier et al., 2011).

Alternatively, $^{1}H^{\alpha}$ detection can be employed. Even if less frequently used than $^{1}H^{N}$ detection, examples of its application for the study of IDPs are present in literature (Mäntylahti et al., 2010; Mäntylahti et al., 2011; Piai et al., 2014). $^{1}H^{\alpha}$ detection provides good sensitivity, allows to observe proline residues and enables to work at physiological conditions, since $^{1}H^{\alpha}$ are not affected by chemical exchange. On the other hand, chemical shift dispersion of α proton nuclei is even smaller than that of amide proton nuclei, and $^{1}H^{\alpha}$ resonate nearby the water signal, therefore an excellent experimental setup is required to achieve optimal water suppression. As an alternative, fully deuterated solvent can be used in order to avoid solvent suppression, provided that $^{1}H^{N}$ are not exploited in the pulse sequence and measured chemical shifts are corrected for the isotopic shift.

2.5 Non-Uniform Sampling

A possible solution to overcome the resolution problem of IDPs is to acquire experiments containing a high number of dimensions. The possibility to correlate several chemical shift evolutions enhances spectral resolution by spreading overlapping resonances into a larger frequency space. In practice, the maximum number of dimensions of a NMR experiment is limited by the fact that each indirect dimension costs in terms of experimental time and sensitivity. The latter, resulting from magnetization losses due to increasing length of the pulse sequence with the number of dimensions, has a minor impact on highly flexible IDPs with respect to what is found for folded proteins, as high flexibility translates into favorable relaxation properties which allow coherence to sustain long transfers and delays. On the contrary, the problem of excessive experimental time is independent from the properties of the protein under investigation, but is intrinsically related to the way multidimensional NMR experiments are built-up. Indeed, each indirect dimension causes an increase of the overall acquisition time proportionally to the desired resolution and according to the requirements of phase-sensitive quadrature detection. As a result, high-resolution multidimensional experiments cannot be performed without the use of proper strategies to reduce their experimental duration.

In recent years, many different approaches have been proposed to overcome the sampling limitation problem: i) LRE, which allows for faster recovery of equilibrium magnetization and thus higher experimental repetition rate; ii) spatial encoding of spectral frequencies, employed for ultrafast measurement of multidimensional spectra in a single scan (Frydman et al., 2002; Mishkovsky et al., 2009); iii) Non-Uniform Sampling (NUS) (Kazimierczuk et al., 2006; Mobli et al., 2006; Kazimierczuk et al., 2007; Coggins et al., 2007; Kazimierczuk et al., 2010; Kazimierczuk et al., 2010; Kazimierczuk et al., 2012), in which fewer data points than those required by the Nyquist theorem are acquired (Figure 2.4). In particular, the latter approach allows to obtain excellent results in terms of

maximum spectral resolution and number of dimensions, and therefore is particularly advantageous for the study of IDPs.

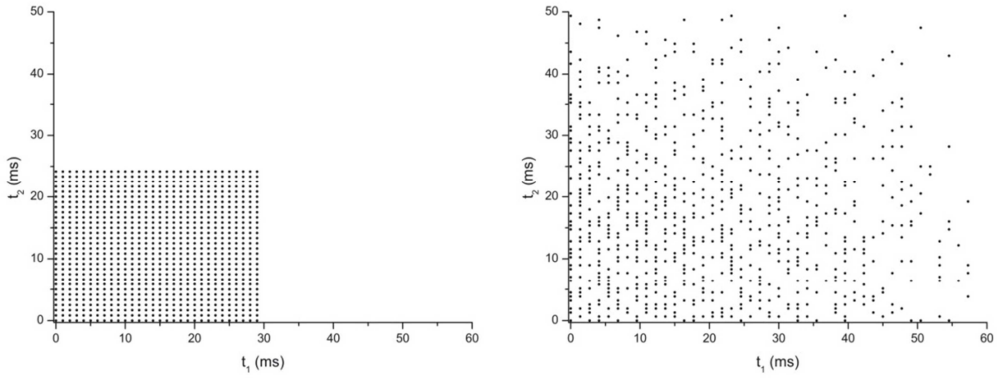


Figure 2.4 Conventional and non-uniform sampling. On the left, 900 sampling points regularly spaced on the time grid, as required to build up the indirect dimensions in conventional NMR experiments (3D in the example), are shown. On the right, the same number of points (which translates in the same experimental time) are distributed over the time domain space with a sampling density of 25%, thus providing a spectrum with twice higher resolution in both indirect dimensions. The NUS sampling points were generated using the Poisson disk algorithm (Kazimierczuk et al., 2008) and the distribution was relaxation-optimized according to the Gaussian distribution $\exp(-t^2/\sigma^2)$, with σ=0.5. Contrary to conventional sampling, NUS data sets cannot be processed using the discrete Fourier transform.

Requirements for a proper performance of NUS are: i) sufficient experimental sensitivity; ii) optimization of the sampling scheme and number of points; iii) choice of the most appropriate method to process the data. The first issue comes from the fact that in NUS experiments signal to noise is traded for shorter acquisition time and higher resolution. In practice, this is achieved by selecting sampled points at longer acquisition times, which provides higher resolution but also greater relaxation losses. Therefore, experimental sensitivity should be maximized, for example by matching the sampling list with the relaxation decay of the signal. The second point arises from the fact that under-sampling translates into the presence of spectral artefacts, resulting from missing information with respect to conventional sampling. Consequently, artefacts must be minimized by designing proper sampling schemes, taking into account that the arrangement and the number of sampled points determine the spectral

distribution of artefacts and the signal-to-artefact ratio, respectively (Kazimierczuk et al., 2007). NUS, in the great majority of cases, precludes the use of the discrete Fourier transform (FT) for computing spectra (Ernst et al., 1966). Therefore, alternative methods should be employed based on their reconstruction efficiency and computational cost. Currently, the most established methods range from the general Multidimensional Fourier Transform (MFT) (Kazimierczuk et al., 2007) to more sophisticated methods such as Maximum Entropy (ME) (Hoch et al., 1996; Mobli et al., 2007; Mobli et al., 2008), MultiDimensional Decomposition (MDD) (Luan et al., 2005; Jaravine et al., 2006; Jaravine et al., 2008; Orekhov et al., 2011a), Compressed Sensing (CS) (Orekhov et al., 2011b; Holland et al., 2011) and Spectroscopy by Integration of Frequency and Time-domain information (SIFT) (Matsuki et al., 2009). Additional post-processing cleaning procedures like Iterative Soft Thresholding (IST) (Hyberts et al., 2012), CLEAN (Coggins et al., 2008) and Signal Separation Algorithm (SSA) (Stanek et al., 2012), can be exploited to further reduce residual NUS artefacts.

4-5D experiments are extremely useful for the study of IDPs because of their higher resolving power. A very convenient way to process such high-dimensional spectra is by employing MFT. Indeed, one of the advantages of this method is the possibility to compute only a subspace of the full multidimensional spectrum at arbitrary frequency coordinates. This approach, known as Sparse Multidimensional Fourier Transform (SMFT) (Kazimierczuk et al., 2009), allows to simplify the analysis of multidimensional spectra by displaying only 2D cross-sections computed at some pre-known frequencies collected in a lower dimensionality “basis spectrum” (2-3D), which shares the same dimensions with the higher dimensionality experiment (4-5D) (Figure 2.5). In this way, the interpretation of 4-5D spectra becomes very manageable and the collection of cross-peaks results facilitated. Importantly, SMFT provides additional functionality over automated protocols, as it allows to manually inspect and cross-check results of crowded spectral regions with severe peak overlaps.

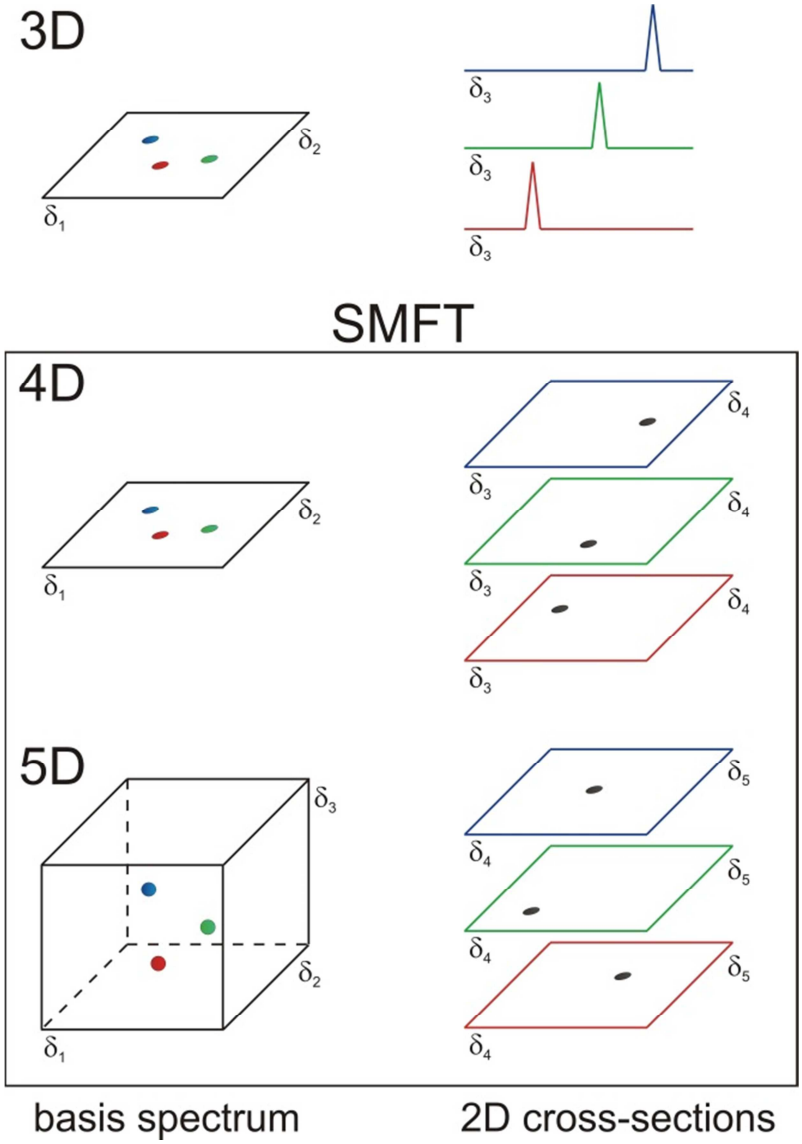


Figure 2.5 Schematic illustration of a possible method to process and visualize multidimensional NMR spectra. A 3D spectrum can be thought as a bidimensional spectrum in which each cross-peak is related to another frequency in the third dimension, as shown with colors. Analogously, a 4D spectrum can be imagined as a 2D basis spectrum, which shares two dimensions with the 4D spectrum (δ_1 and δ_2), in which each peak is associated to another 2D spectrum (cross-sections) containing the two further dimensions (δ_3 and δ_4); a 5D spectrum can be analyzed as a series of two-dimensional cross-sections (δ_4 and δ_5), each one correlated to a given peak of the related 3D basis spectrum (δ_1, δ_2 and δ_3). SMFT does not compute the full spectrum, but just the series of highly resolved 2D cross-sections, which provide the added information content with respect to the reference spectrum.

Another approach which can be used to reduce the number of dimensions of a high multidimensional spectrum is projection spectroscopy, in which a series of

projections are acquired rather than the full spectrum (Szyperski et al., 1993; Malmodin et al., 2005; Coggins et al., 2010). In this way, the analysis of the spectrum results facilitated as it consists in that of a collection of a series of 2D projections. Automation of such analysis, called Automated Projection Spectroscopy (APSY) (Hiller et al., 2005), yields a peak list of the full-dimensionality spectrum, without reconstructing it. Therefore, this approach results particularly convenient to analyze spectra which could not be directly inspected, for example due to excessive data storage size, like 7D spectra (Hiller et al., 2007; Narayanan et al., 2010).

2.6 Strategies for sequence-specific resonance assignment of IDPs

The first step towards a full characterization of an IDP by NMR is the sequence-specific resonance assignment. The assignment procedures rely on successful identification of connectivities between neighboring residues by using one or more chemical shifts to establish unambiguous sequential correlations. Due to the very poor chemical shift dispersion of IDPs, chances to obtain unique connections are reduced, with the consequence that the complexity of the process grows with the size of the protein under investigation. Therefore, the success of an assignment greatly depends on the choice of nuclei to correlate, since best connections are those providing the lowest possible number of ambiguities.

Contrary to folded proteins, C^{α} and C^{β} chemical shifts of IDPs are not the most dispersed, although they still remain characteristic for different amino acid types. Therefore, they are primarily exploited to identify residue type of spin systems rather than to establish sequential correlations. Given the lower dispersion of H^{N} chemical shifts, N and C' result the most appropriate nuclei to identify connectivities between neighboring residues (Bermel et al., 2013b; Pantoja-Uceda et al., 2013a; Piai et al., 2014; Nováček et al., 2014). The fact that C' and N

evolution times can be set to tens of milliseconds satisfies the need for utmost resolution necessary to obtain the sequence-specific assignment of an IDP.

In recent years, several NMR experiments dedicated to backbone and side-chain assignment of IDPs have been developed. Different techniques have been exploited to improve and speed-up the resonance assignment procedure. Among all the different approaches, NUS (par. 2.5) has been proven to provide greatest benefits. Indeed, the possibility to introduce a high number of dimensions with high-resolution increases the resolving power of the experiment by correlating more frequencies at the same time and by establishing more robust connectivities, thus reducing the number of ambiguous correlations. Combined use of NUS with ^{13}C detection and techniques for LRE guarantees further improvements in terms of chemical shift dispersion, experimental time and sensitivity.

For these reasons, these approaches have been profitably merged together is the so-called "*CON-CON*" assignment strategy (Piai et al., 2014): 5D ($H^{N-flip}N$)CONCACON, 5D (HCA)CONCACON and 5D (H)CACON(CA)CON experiments (Bermel et al., 2013b) have been designed to correlate frequencies of neighboring residues through the simultaneous exploitation of C' and N dimensions (Appendix – Article 3). In this way, sequential correlations between cross-peaks of a 2D CON spectrum are established (Figure 2.6). Building up such connections throughout larger numbers of subsequent residues allows to map them onto the protein sequence, using complementary information provided by 4D HCBCACON or 4D HCCCON experiments to recognize different amino acid types (Bermel et al., 2012b). Despite the intrinsic lower sensitivity of ^{13}C detection, these experiments can be acquired in a short time at very high-resolution in all the indirect dimensions thanks to the use of NUS, ^{1}H-excitation and techniques to achieve LRE, e.g the H^{N-flip}.

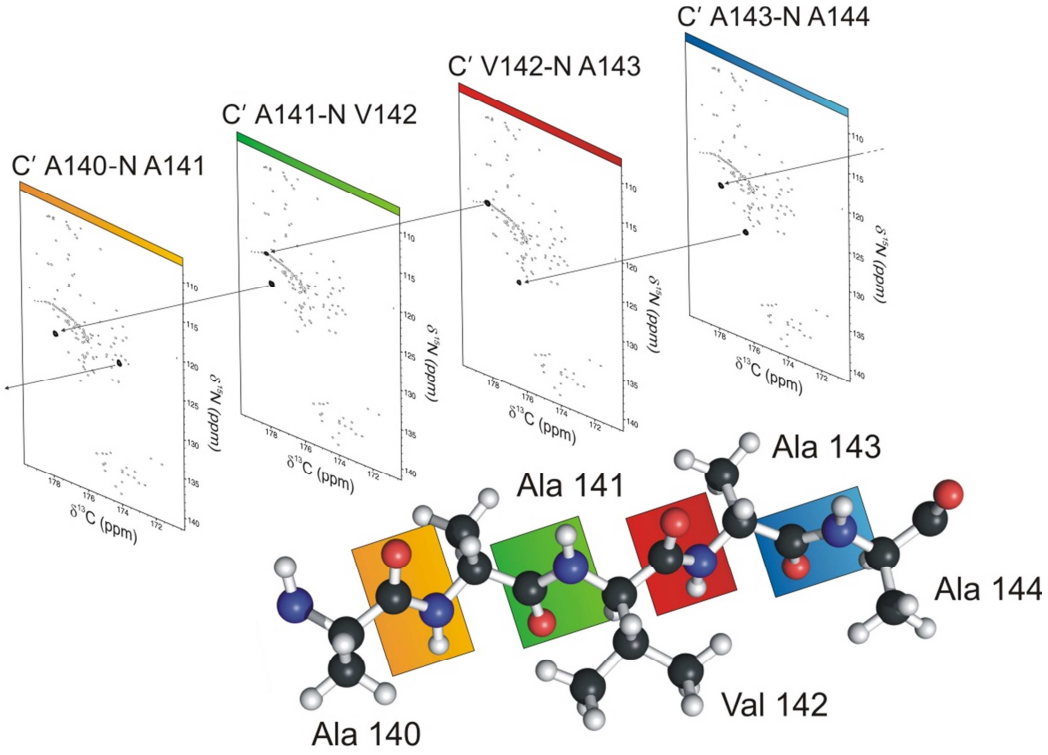


Figure 2.6 "*CON-CON*" assignment strategy exemplified by showing a series of 2D cross-sections of a 5D (HACA)CONCACON spectrum acquired on a ^{13}C, ^{15}N-labeled sample of the N-terminal domain of human Androgen Receptor "25Q" (0.4 mM 25Q, 20.0 mM Pi, 1.0 mM TCEP, pH 7.4, 278.0 K) with a Bruker Avance 700 MHz spectrometer equipped with a cryogenically-cooled triple-resonance probe optimized for ^{13}C direct detection. In each cross-sections, two pairs of sequential $^{13}C'$ and ^{15}N frequencies are present (filled black cross-peaks; empty cross-peaks correspond to the superimposed 2D CON spectrum). The sequential specific assignment is performed by connecting the cross-peaks of consecutive cross-sections, exploiting simultaneously $^{13}C'$ and ^{15}N frequencies to minimize the occurrence of possible overlaps. As shown by colors, the order of the 2D cross-sections reflects that of $^{13}C'$ and ^{15}N pairs along the protein backbone.

With the goal to form a complete set of experiments for sequence-specific assignment of highly flexible IDPs, three analogous 5D experiments exploiting a conceptually identical approach, but based on ^{1}H detection, have been developed (Appendix – Article 4). The experiments, i.e. 5D (HACA)CON(CACO)NCO(CA)HA, 5D BT-(H)NCO(CAN)CONNH and 5D BT-HN(COCAN)CONNH (Piai et al., 2014), taking advantage either of the properties of α or amide proton nuclei, overcome the low dispersion of proton chemical shifts by exploiting the best resolved correlations provided by C' and N nuclei in the indirect dimensions. Also for this set of experiments, NUS has been

employed to achieve excellent resolution in all spectral dimensions; in particular, in the case of $^{1}H^{N}$ detection, it allows a full exploitation of the resolution enhancement provided by TROSY, while BEST further speeds up the experiments. Noteworthy, the information derived from the two BEST TROSY (BT) experiments can be combined so to extend the assignment strategy to "*HNCO-HNCO*", as three frequencies, those collected in a 3D HNCO spectrum, are simultaneously employed to establish almost unambiguous sequential correlations (Figure 2.7).

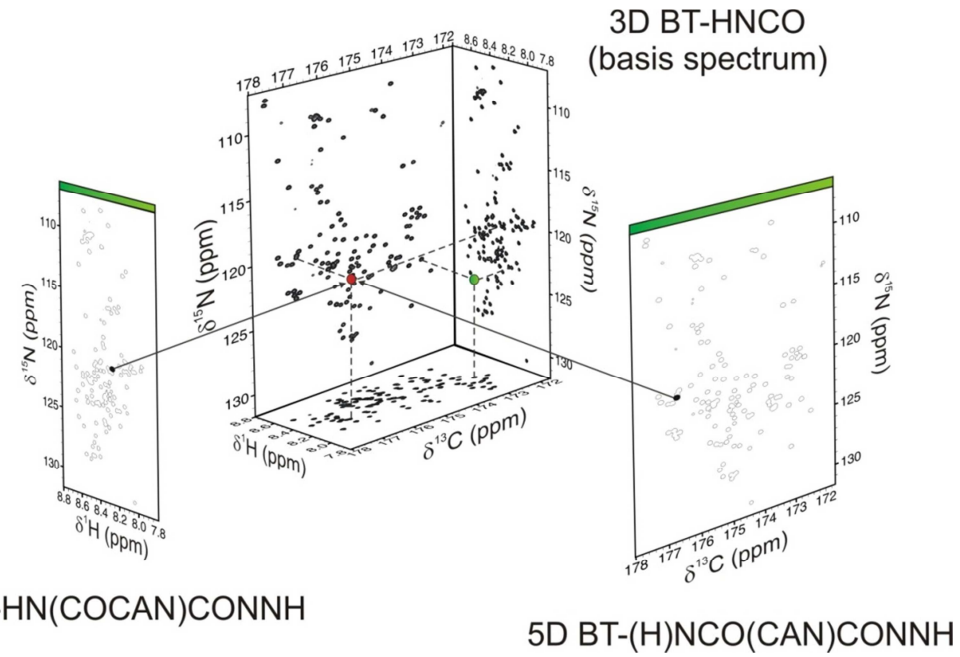


Figure 2.7 "*HNCO-HNCO*" assignment strategy exploiting 5D BT-(H)NCO(CAN)CONNH and 5D BT-HN(COCAN)CONNH experiments. All the spectra were acquired on a ^{13}C, ^{15}N-labeled sample of human α-synuclein (1.0 mM α-synuclein, 20 mM Pi, 200 mM NaCl, 0.5 mM EDTA, pH 6.5, 285.5 K) with a Bruker Avance III 950 MHz spectrometer equipped with a cryogenically-cooled triple-resonance probe. 2D cross-sections of 5D BT-(H)NCO(CAN)CONNH and 5D BT-HN(COCAN)CONNH spectra are reported on the left and on the right, respectively; the 3D BT-HNCO basis spectrum, required to process the 5D experiments with the SMFT method, is shown in the middle (only its 2D projections are displayed). The sequence-specific assignment is performed through simultaneous exploitation of $^{13}C'$, ^{15}N and $^{1}H^{N}$ frequencies. First, a cross-peak (e.g. the green one, Gly68$^{13}C'$-Ala69^{15}N-Ala69$^{1}H^{N}$) of the 3D basis spectrum is chosen. Then, the related 2D cross-sections of the two 5D spectra are inspected in order to retrieve the $^{13}C'$-^{15}N and $^{1}H^{N}$-^{15}N frequencies of the residue immediately following the one selected in the basis spectrum (only the filled black cross-peaks are observed in the cross-sections; empty cross-peaks represent the superimposed 2D projections). Since the cross-sections of the two 5D spectra share the nitrogen dimension, the provided information can be merged identifying simultaneously the $^{13}C'$, ^{15}N and $^{1}H^{N}$ frequencies of the following residue. Once such cross-peak (the red one in the example, Ala69$^{13}C'$-Val70^{15}N-Val70$^{1}H^{N}$) is found in the 3D basis spectrum, it becomes sequence-specifically linked to the previous. Analogously, the assignment procedure continues forming chains of sequential $^{13}C'_{i}$-$^{15}N_{i+1}$-$^{1}H^{N}_{i+1}$ cross-peaks.

Considered together, these experiments allow to obtain a highly efficient backbone resonance assignment of flexible proteins in a great variety of experimental conditions. Indeed, they present different advantages and disadvantages depending on the particular sample conditions (concentration, lifetime, pH, temperature, etc.). For example, while the ^{13}C or $^{1}H^{\alpha}$ detected experiments are especially useful to study IDPs near physiological conditions and/or characterized by proline-rich repetitive amino acid sequences, the $^{1}H^{N}$ detected ones are expected to be favorable for low concentrated or short-lived IDPs, provided they can be studied under conditions where solvent exchange is sufficiently slowed down. Needless to say, the availability of such a variety of complementary experiments is extremely important for accurate determination of resonance frequencies in the case of complex IDPs.

Highly-resolved correlations provided by multidimensional experiments result particularly suitable for automation in the sequence-specific resonance assignment process. For this reason, several algorithms for automatic assignment like MARS (Jung et al., 2004), FLYA (Schmidt et al., 2012), Memetic Algorithm and Combinatorial Optimization Heuristics (MATCH) (Volk et al., 2008) and Tool for SMFT-based Assignment of Resonances (TSAR) (Zawadzka-Kazimierczuk et al., 2012a) have been developed and successfully employed to assign IDPs.

In particular, the automatic assignment program TSAR, tailored to analyze 2D cross-sections of high dimensional spectra processed with SMFT, has been optimized to exploit all the information provided by the experiments described above and allows to speed up the assignment of IDPs when employing the "*CON-CON*" strategy (Bermel et al., 2013b). In addition, a set of ^{13}C detected experiments for spin-system identification has been improved so to provide complementary information on amino acid type to further improve TSAR performances. The acquisition time necessary to acquire a series of amino acid selective 2D (HCA)CON and (HCA)NCO experiments has been reduced, at no cost for spectral resolution, by exploiting NUS; a novel 4D γ-selective-HCBCACON

experiment in which cross-peak signs are modulated according to the number and type of γ carbon nuclei present in each amino acid side-chain, has been designed in order to diminish the number of possible mistakes in automatic amino acid recognition (Appendix – Article 5). With this method, a fast and reliable automatic assignment of chemical shifts of α-synuclein has been accomplished.

The developed strategy, from the acquisition of the data to the assignment of the frequencies, enables the investigation of IDPs of increasing size and complexity and provides a solid and robust method to characterize structural disorder, as shown in chapter 3.

2.7 α-synuclein: a suitable standard for the development of new NMR experiments and a reporter of the effects of cellular crowding

α-synuclein is a presynaptic neuronal protein constituted by 140 residues typically found in the brain, but in smaller amounts also in the heart and muscles. In the brain, α-synuclein is found mainly in the presynaptic terminals of neurons, where it interacts with membrane lipids and other proteins (Jao et al., 2008). Although its physiological function is yet not well understood, studies suggest that it plays a crucial role in maintaining a supply of synaptic vesicles for the release of neurotransmitters, e.g. dopamine (Alves da Costa, 2003; Burré et al., 2010). It is also known that α-synuclein can form aberrant oligomers responsible for Parkinson's disease (Dev et al., 2003; Spinelli et al., 2014), a progressive pathology which causes resting tremor, rigidity, postural instability and bradykinesia. Indeed, these toxic species are the major constituents of Lewy bodies, protein clumps that are at the basis of this neurodegenerative pathology (McLean et al., 2000; Cookson, 2009). Recently, it has been reported that α-synuclein may have also an involvement in Alzheimer's disease, even if this statement is still controversial (Larson et al., 2012).

Because of its relevant biological importance, α-synuclein has been widely-studied. The efficient expression and purification protocol, the high solubility and moderate size and the acceptable sample stability make α-synuclein an excellent choice as a standard for the development of NMR experiments tailored for the study of IDPs. The sequence-specific assignment previously reported (Bermel et al., 2006) in the Biological Magnetic Resonance Bank (BMRB)[10] (Ulrich et al., 2008) has been initially used and then extended to incorporate chemical shifts of a larger set of nuclear spins as well as to report chemical shift changes upon variations of the experimental conditions in order to approach physiological ones (Appendix – Table 2). Then, the protein has been used as a benchmark for the development of new experiments.

The obtained full chemical shift assignment of α-synuclein has been also used to address one of the many open questions in the field of IDPs, i.e. the effects of molecular crowding on the properties of a protein. The cellular environment contains a variety of macromolecules occupying up to 30% of the available volume which can have significant effects on a variety of processes occurring in the cell, including protein aggregation. Different crowding agents have been used to investigate the effects on the kinetics and mechanism of aggregation of several proteins with a variable degree of intrinsic disorder. The results indicate that factors other than the excluded volume effect, such as non-specific chemical interactions of proteins with polymers, are likely to be involved in the phenomenon of macromolecular crowding and that the effects of neutral polysaccharides on protein aggregation are strongly dependent on the structure of both proteins and polysaccharides. In general, crowding has been found to promote the formation of the most stable and/or compact protein conformation and inhibited aggregation of proteins forming stable oligomers or folded monomers; on the contrary, compact conformations stabilized by crowding of

[10] http://www.bmrb.wisc.edu/

partially or completely disordered proteins are likely to be on the pathway to amyloid formation (Appendix – Article 6).

Chapter 3

Structural disorder, flexibility and new functional roles

As discussed in chapter 1, functional studies of IDPs have illustrated the broad impact of disorder on many areas of cellular biology, ranging from signaling processes to transcriptional regulation mechanisms. Such a wide range of functions is reflected by a great variety of disorder degrees, which make each IDP specifically tuned to perform its biological role. Out of the many possible categories of structural disorder, two important classes are represented by protein linkers, connecting globular domains of complex proteins, and low-complexity regions (LCRs).

Many eukaryotic proteins contain flexible linker regions, separating globular domains, about which no structural and dynamic information is available at high-resolution. Linker sequences can be very different for length and composition; however, many of them are often rich in polar and uncharged amino acids, small residues (such as alanine and glycine residues) and proline residues. Most of these residues have a propensity to populate the polyproline-II region of the Ramachandran plot (Shi et al., 2002; Rucker et al., 2002; Rucker et al., 2003), which means that they prefer to adopt a highly extended conformation, even if remaining overall flexible. Therefore, in the absence of their targets, the primary function of a linker is to guarantee a relatively unhindered spatial search by the attached domains. However, binding can induce structural changes in protein linkers which can have significant functional consequences, like a proper orientation of the neighboring domains, the formation of weak but highly specific interactions, the fine-tuning of biological processes. In this regard, the

new NMR experiments I developed extend the number of available tools to further understand the array of functions of protein linkers.

To date, atomic resolution information on low-complexity regions (LCRs) has not been accessible in their native contexts. Indeed, despite their abundance in protein sequence, LCRs of disordered protein regions are largely under-represented in the PDB. LCRs are characterized by little variety in amino acid composition, which can range from regions comprising few different amino acids to those comprising just one residue type, such as polyglutamine (polyQ) tracts (Wootton et al., 1993; Kumari et al., 2015). Frequency degeneracies and signal overlaps often prevent to derive site-specific structural properties by NMR, which are fundamental towards a full understanding of the wide range of biological function performed by such disordered elements, most of which remain still to be elucidated. LCRs promise to be a very important subject of study as they are found in a great number of proteins (Wootton, 1994): they are estimated to represent about 12% of the protein content classified in the current release of UniProt (The UniProt Consortium, 2015). Recently, it has been discovered that LCRs are preferentially located towards protein termini and that their positions within the sequence have an influence in determining their binding properties and biological roles (Coletta et al., 2010; Martins et al., 2015). Noteworthy, mutations or alterations of such regions often lead to the development of severe neurodegenerative diseases characterized by the occurrence of protein aggregation (Das et al., 2014). Therefore, a deeper understanding of their biological roles will result extremely useful in order to design new medical applications. The NMR methods developed during my doctoral research provide new strategies to tackle such systems.

3.1 The role of flexible linkers: characterization of three IDRs of CREB Binding Protein

Long extended protein linkers are probably the most widespread disordered regions within proteins. Many proteins, especially large ones, exploit such modules to provide conformational flexibility to the attached globular domains. However, linkers may also perform additional functions and be involved in recognition and regulation processes, e.g. the fine-tuning of highly specific interactions. With the goal to get new insights into the structural and functional roles of these protein elements, ID3, ID4 and ID5 linkers of CREB Binding Protein (corresponding to residues 674-1079, 1851-2058 and 2109-2442 of the full-length protein) have been characterized. The choice of CREB Binding Protein as a model system to address the theme of protein linkers was due to the fact that such protein, well-studied for its high biological relevance, is constituted by several disordered domains and it is known to have a multitude of interacting partners (Dyson et al., 2005).

Cyclic-AMP-Response-Element-Binding protein (CREB)-Binding Protein (CBP), like its paralog p300, is a transcription factor which integrates signals from numerous signal transduction pathways and plays critical roles in basic cellular processes ranging from embryonic development to DNA repair (Goodman et al., 2000; Smith et al., 2004; Dyson et al., 2005). Its biological function is related to its ability to interact with a large number of proteins, functioning as a scaffold for the recruitment and assembly of the transcriptional machinery, as well as to its acetyl-transferase activity, through which it modifies both chromatin and transcription factors (Giordano et al., 1999; Goodman et al., 2000; Blobel, 2000). CBP inactivation or malfunctioning is associated with the occurrence of cancer.

CBP is a modular protein constituted by seven folded domains performing different activities, whose 3D structures have been determined in recent years either by X-ray crystallography or NMR (Legge et al., 2004; De Guzman et al.,

2005; Liu et al., 2008; Miller et al., 2009; Kjaergaard et al., 2010; Lee et al., 2010; Wang et al., 2012; Filippakopoulos et al., 2012; Plotnikov et al., 2014). Four of them require zinc binding to adopt a stable fold: the Transcriptional-Adaptor Zinc-finger-1 domain (TAZ1, residues 347-439), the Plant HomeoDomain (PHD, residues 1224-1317), the zinc-binding domain near the dystrophin WW domain (ZZ, residues 1694-1753) and the Transcriptional-Adaptor Zinc-finger-2 domain (TAZ2, residues 1762-1855). The other folded domains are the KID-binding domain (KIX, residues 586-673), the Bromodomain (Bromo, residues 1081-1197) and the Histone Acetyl-Transferase domain (HAT, residues 1318-1693); finally, a domain named Nuclear-receptor Coactivator-Binding Domain (NCBD, residues 2057-2116) is intrinsically disordered, but folds upon binding to its partner(s).

More than half of the 2,442 residues of the protein are predicted to be intrinsically disordered and have not been characterized yet at atomic resolution (Dyson et al., 2005). They form five linker regions named ID1 (N-terminus), ID2 (connecting TAZ1 and KIX domains), ID3 (connecting KIX and Bromo domains), ID4 (connecting TAZ2 domain and NCBD) and ID5 (C-terminus) (Figure 3.1). A peculiar feature of such linkers is that they are evolutionary conserved in amino acid composition (≈ 70% are polar residues), even if not in sequence and length. Although they exhibit the amino acidic compositional bias typical of IDPs, they also include several conserved segments containing relatively high proportions of charged and hydrophobic residues, which may represent yet unidentified interaction motifs.

Even if the main role of CBP linkers seems to be to confer enough flexibility for the assembly of the transcriptional machinery (Gunasekaran et al., 2003), they may also provide binding sites for transcriptional regulatory proteins or recruit protein factors through the so-called “fly-casting” mechanism (Shoemaker et al., 2000; Huang et al., 2009). I have thus undertaken the characterization of these IDRs.

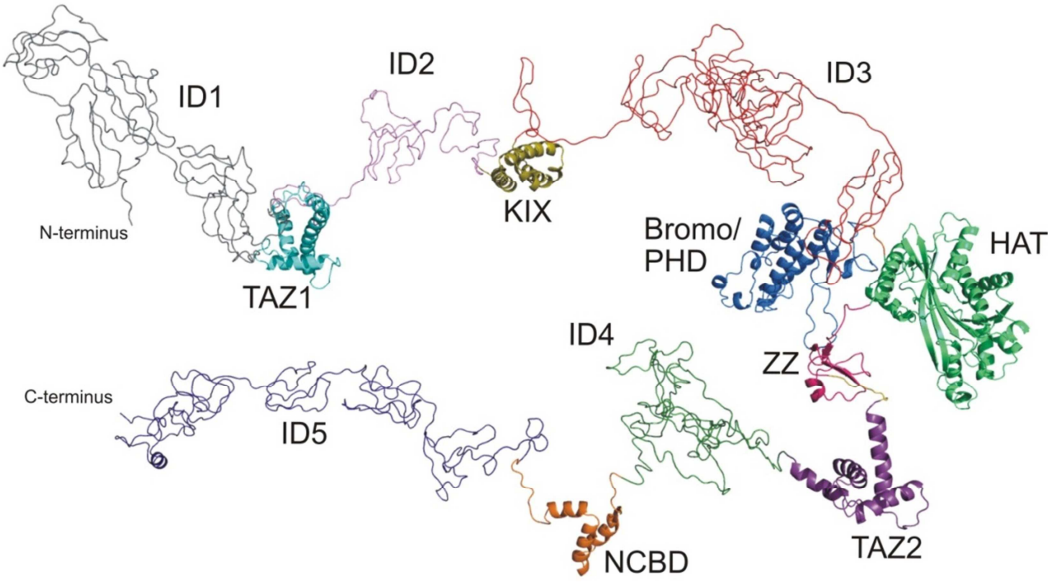


Figure 3.1 Structural organization of CBP: ID1 (gray), TAZ1 (cyan, PDB code: 1U2N), ID2 (violet), KIX (gold, PDB code: 2LQI), ID3 (red), Bromo and PHD (light blue, PDB code: 4N4F), HAT (light green, PDB code: 3BIY), ZZ (pink, PDB code: 1TOT), TAZ2 (purple, PDB code: 2KJE), ID4 (green), NCBD (orange, PDB code: 2KKJ) and ID5 (blue). Disordered regions have been generated using the IntFOLD[11] web resource (Roche et al., 2011).

The three which were assigned so far resulted structurally very heterogeneous (Figure 3.2). The longest linker, ID3, even if overall disordered, is characterized by several regions exhibiting small but significant structural propensities, which render its behavior not that of a fully random-coil protein (manuscript in preparation). ID4, the linker containing the highest percentage of proline residues (22%), appears to be highly flexible except the regions encompassing residues 2-25 and 101-128 (1852-1875 and 1951-1978 in the full-length CBP) which exhibit a high degree of α-helical sampling. Interestingly, proline residues have been found to be uniformly distributed along the linker except for these two structured regions, indicating that they play an active role in modulating the structural features of this CBP fragment (Appendix – Article 7). ID5, the C-terminal disordered region of CBP, is the one containing the higher structural content among the all. Indeed, even if the sequence-specific resonance assignment of this IDR is still in progress, preliminary results indicate that the

[11] http://www.reading.ac.uk/bioinf/IntFOLD/

regions encompassing the 18 residues long polyQ tract (residues 66-88) and residues 164-174 (2189-2211 and 2287-2297 in full-length CBP, respectively) adopt an α-helical conformation, while the polypeptide is completely disordered elsewhere (manuscript in preparation). In all cases, the experiments described in chapter 2 resulted fundamental to achieve the full characterization of the linkers at atomic resolution.

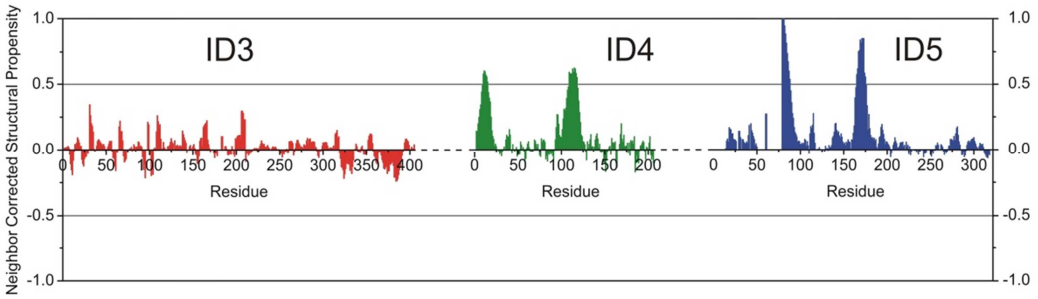


Figure 3.2 Neighbor corrected Structural Propensity (ncSP) plots of ID3 (red), ID4 (green) and ID5 (blue) linkers of CBP. In the figure, the numbering of each linker has been considered out of the context of the full-length protein. The SSP scores were calculated from experimentally measured N, C′, C^{α} and C^{β} chemical shifts by using the neighbor corrected Structural Propensity Calculator (ncSPC) tool[12] (Tamiola et al., 2012). The Tamiola, Acar and Mulder random coil chemical shift library was chosen for the analysis (Tamiola et al., 2010). Positive and negative values correspond to α-helical and β-sheet propensities, respectively.

The atomic resolution investigations of the structural and dynamic properties of these IDRs provide a striking example of how the concept of protein linkers as mere connecting elements between structured domains is far from the truth. Indeed, PreSMos as those identified in ID4 and ID5 may represent so far uncharacterized functional modules which can exploit their plasticity to modulate the relative orientations of neighboring folded domains, while at the same time fine-tuning interactions between CBP and its partners.

My colleagues and I are currently working in getting more insights into their biological functions by studying possible interactions of the linkers with proteins of high biomedical relevance. In particular, we have started collecting the first

[12] http://nmr.chem.rug.nl

evidences of the interaction between ID3 and a disordered fragment of the Zinc Finger Protein 106 (ZFP106) (Grasberger et al., 2005) and between ID5 and a fragment of the Ras GTPase-activating-like protein IQGAP1, a scaffold protein known to be overexpressed in many types of cancer (Jadeski et al., 2008;Abel et al., 2015). These evidences, if confirmed by additional experiments, will provide two strong proofs that protein linkers do actually contribute to the interaction and/or recruitment of different proteins, thus actively participating in the fine regulation of several cellular processes. In this regard, the sequence-specific assignments of ID3 and ID5 will result fundamental to study these interactions by NMR at high-resolution.

3.2 The molecular basis of a polyQ disease: the case study of the N-terminal domain of Androgen Receptor

Polyglutamine repeats are renowned exponents of the category of low-complexity regions. Despite their abundance in proteins, they are mostly known for their implication in several neurodegenerative disorders rather than for their physiological functions, which are still unclear. To get more insights into the development of one of these pathologies and characterize the structural properties of a polyQ tract in its native protein context, a N-terminal fragment of Androgen Receptor was studied.

Androgen Receptor (AR) is a steroid-hormone activated transcription factor which regulates the development and maintenance of the male phenotype (Gelmann, 2002). The protein, constituted by 919 residues, has three major functional domains: the intrinsically disordered N-Terminal Domain (NTD, residues 1-558) (McEwan et al., 2007), the zinc-finger DNA-Binding Domain (DBD, residues 559-631) (Shaffer et al., 2004) and the Ligand-Binding Domain (LBD, residues 670-919) (Pereira de Jésus-Tran et al., 2006). Between DBD and LBD, a hinge region that contains a nuclear localization sequence is present (Figure 3.3). In the resting state, AR is a cytosolic protein in complex with molecular

chaperones bound to the NTD. Upon binding the hormone ligand to the LBD, the receptor dissociates from accessory proteins and translocates into the nucleus, where it dimerizes and then stimulates transcription of androgen responsive genes called Androgen Response Elements (AREs) (Shaffer et al., 2004). Once the DBD is bound to the DNA, the NTD recruits the general transcription factors and co-regulators needed to assemble the transcription machinery (Wang et al., 2007). The NTD of the protein contains a polyQ tract which is typically constituted by 9 up to 36 glutamine residues (Kumar et al., 2011; Ryan et al., 2011). Its variable length results from an expansion of the CAG and GTC trinucleotide repeats in the AR gene due to their propensity to form non-B-DNA-structures, which cause slippage during DNA replication (Wells et al., 2005; Mirkin, 2007). When the length of the polyQ tract exceeds 40 glutamine residues, a neuromuscular disease called Spinal Bulbar Muscular Atrophy (SBMA) is developed (La Spada et al., 1991; Zoghbi et al., 2000).

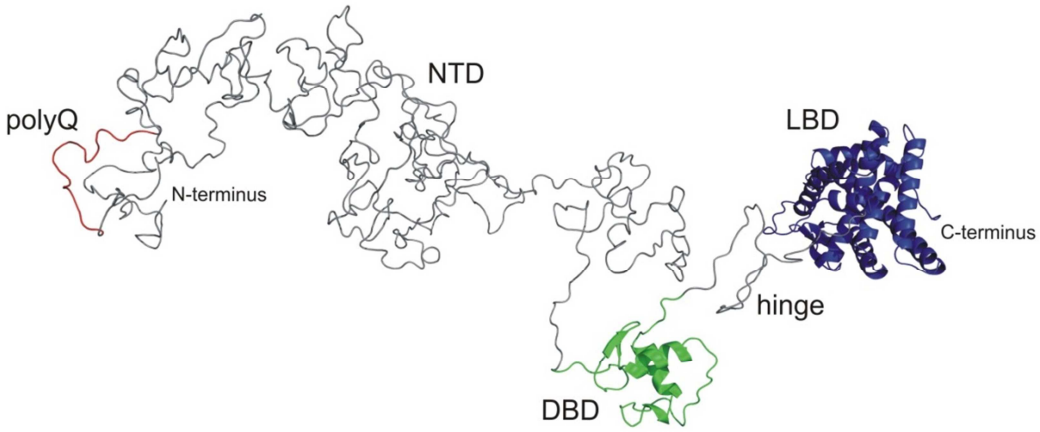


Figure 3.3 Structural organization of AR: NTD (gray), DBD (green, PDB code: 1R4I) and LBD (blue, PDB code: 2AM9). The polyQ tract is highlighted in red. Disordered regions have been generated using the IntFOLD web resource.

SBMA, also known as Kennedy disease (Kennedy et al., 1968), is a rare disorder which damages motor neurons, leading to muscles' weakness and atrophy (Katsuno et al., 2006; Thomas et al., 2006; Rhodes et al., 2009). Although the

molecular basis of the disease are not yet fully understood, it has been reported that caspase cleavage of AR is a crucial event in cytotoxicity (Merry et al., 1998). Indeed, protein fragments containing the expanded polyQ tract accumulate within the nerve cells and interfere with normal cell functions by formation of toxic aggregates (Li et al., 1998; Chen et al., 2002; Adachi et al., 2005). However, the molecular mechanism by which the polyQ tract is involved in the formation of the nuclear inclusions associated with SBMA has not been described so far.

To get more insights into the pathological mechanisms of SBMA and characterize the structural and dynamic properties of a polyQ tract within its native protein context, a N-terminal fragment of AR, containing the polyQ tract responsible for the onset of the disease, has been studied. This fragment, constituted by the first 156 residues of the NTD, is the product of proteolytic cleavage by caspase 3 (Merry et al., 1998; Kobayashi et al., 1998; Ellerby et al., 1999).

Two different constructs, featuring a polyQ tract spanning either 4 or 25 glutamine residues (4Q and 25Q), have been characterized at high-resolution employing the NMR experiments and the NMR approaches described in chapter 2 (Appendix – Article 8). Analysis of chemical shifts and ^{15}N relaxation measurements revealed that there is a correlation between the structural properties and the length of the polyQ tract, as the latter has been found to adopt an α-helical conformation when the number of glutamine residues is 25 (Figure 3.4a). The study of the aggregation properties of a third construct (25QΔL), in which the 4 leucine residues immediately preceding the polyQ tract have been removed, allowed to understand that this short protein sequence flanking the polyQ tract is responsible for the induction of α-helical propensity in the polyQ tract, which in turn reduces fibril formation by adding a helix unfolding step in the mechanism of the transition. Although this can act as a mechanism adopted by nature to protect against aggregation, the influence of the flanking region on the secondary structure of the polyQ tract spans a limited range of

residues, thus ceasing to be effective when the polyQ tract is too extended as found in SBMA.

To further characterize the mechanism of this process, the early stages of oligomerization of 25Q have also been studied (Appendix – Article 9). Time-resolved NMR experiments have been used to achieve a direct and unprecedented characterization of the mechanism of self-assembly and to identify precisely which residues undergo any structural changes during oligomerization. The hydrophobic region encompassing residues 22-36, distant from the polyQ tract, has been discovered to be responsible for the inter-molecular interactions that trigger AR oligomerization (Figure 3.4b); the latter then proceeds only if the length of the polyQ tract is sufficiently long.

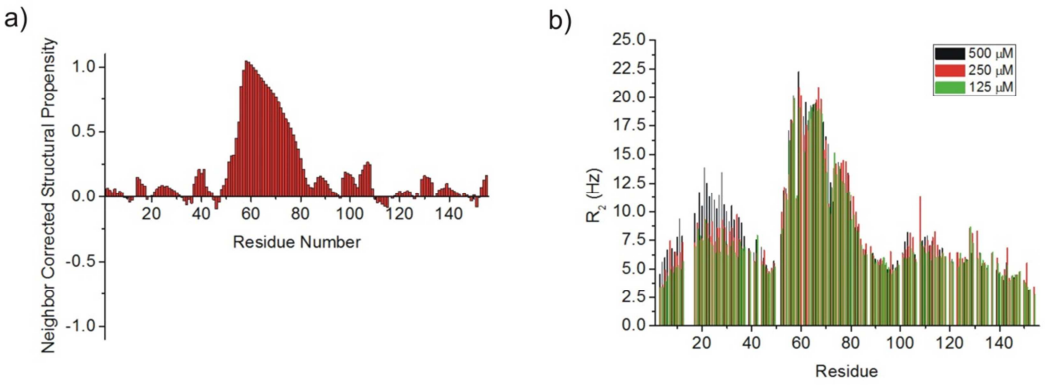


Figure 3.4 Structural and dynamic properties of 25Q. a) Neighbor corrected Structural Propensity (ncSP) of 25Q. The SSP score was calculated from experimentally measured N, C′, C^{α} and C^{β} chemical shifts by using the neighbor corrected Structural Propensity Calculator (ncSPC) tool. The Tamiola, Acar and Mulder random coil chemical shift library was chosen for the analysis. Positive and negative values correspond to α-helical and β-sheet propensities, respectively. b) Concentration-dependent backbone dynamics in 25Q. ^{15}N R_2 relaxation rates are reported as a function of residue number at three different protein concentrations: 125 (green), 250 (red) and 500 µM (black). The error bars have been removed from the figure to facilitate the comparison.

Taken together, these findings provide new insights into the structural properties of polyQ tracts and offer a compelling example of how flanking regions can affect the rate and extent of formation of toxic fibrillar species by proteins harboring this peculiar sequences. The increased understanding of how sequence context

influences the properties of polyQ tracts opens up new avenues for the development of therapeutic strategies for polyQ diseases based on targeting the flanking regions rather than the polyQ tracts themselves. In particular, further insights into the mechanism of AR oligomerization in SBMA have been achieved, leading to a simple explanation for the relationship between cleavage of AR by caspase-3 and the formation of nuclear inclusions.

Chapter 4

Conclusions and perspectives

In recent years, IDPs have turned out to be an important category of proteins. They have been shown to be involved in many cellular processes, able to exploit their flexibility and plasticity to perform highly specific biological functions. Therefore, the study of structural disorder is essential to increase our understanding of key cellular processes, whose malfunction can cause severe diseases.

Among the biophysical techniques available to study IDPs, solution NMR spectroscopy occupies a unique position thanks to its ability to provide high-resolution site-specific information and address protein dynamics. The possibility to obtain simultaneously structural and dynamic information can provide a huge contribution to the understanding of how IDPs function. Nevertheless, so far the application of NMR has been mostly targeted to the study of well-folded proteins. Indeed, the peculiar properties of IDPs decrease dramatically the performances of established NMR experiments and render them inadequate in terms of sensitivity and resolution. For these reasons, my doctoral project has focused on improving the available NMR methods to study IDPs.

Continuous increase in magnetic field strengths, improvements in instrumentation and renovation of acquisition and processing software, make feasible long and complex NMR experiments. Such a technological progress, together with the fast molecular tumbling of highly flexible IDPs which results in slowly relaxing spin coherences, allows the design of advanced pulse sequences correlating nuclei in a larger number of dimensions beside the standard three. 4-5D NMR experiments provide enough resolution to simplify crowded and overlapping spectra such as those of large disordered systems. This thesis

gathers a suite of novel 5D NMR pulse sequences tailored to achieve sequence-specific assignment of IDPs, forming a robust protein backbone assignment protocol ("*CON-CON*" strategy). The experiments have been designed in such a way to reduce signal overlaps, to detect proline residues and to provide sufficient experimental sensitivity for practical use. Approaches for chemical shift labeling of backbone carbonyl carbons and nitrogen nuclei to exploit their larger chemical shift dispersion have been included in all the experiments, at the same time taking full advantage of LRE and NUS to enhance experiments' performances.

Each pulse sequence has been optimized on α-synuclein first and tested to reproduce the assignment of such well-known IDP; then the experiments have been employed to study so far uncharacterized disordered systems to gain knowledge on their structural and dynamic properties and mechanisms of action. Finally, all the NMR experiments have been assessed for automation in their analysis by customizing the automatic assignment program TSAR to fully exploit their information content. Tests performed on different combinations of spectra acquired on α-synuclein led to the conclusion that, for disordered proteins of moderate size, about 80-90% of cross-peaks can be correctly assigned with such tool, speeding up the sequence-specific assignment process.

The developed NMR experiments resulted particularly useful to characterize several IDPs/IDRs that could have not been studied otherwise.

Traditionally, IDRs have been considered to be simply passive segments in protein sequences that link structured domains. However, the NMR characterizations of the ID3, ID4 and ID5 linkers of CBP provide a striking example of how IDRs can be heterogeneous in terms of structural and dynamic properties. These features may increase the functional states in which a protein exist in the cell by enable the interaction and/or the recruitment of different proteins, thus contributing to the fine regulation of several cellular processes.

The study of the N-terminal domain of Androgen Receptor allowed an unprecedented NMR characterization of a 25 residues-long polyQ tract in its native sequence context. Flanking regions of the polyQ tract have been discovered to induce secondary structure elements in the polyQ tract itself, modulating its aggregation properties. On this basis, a model describing the events occurring during the initial stage of AR oligomerization in SBMA has been proposed.

In conclusion, the presented NMR experiments expand the number of IDPs that can be studied at atomic resolution through NMR, particularly near physiological conditions. The exploitation of a high number of dimensions in the experiments results essential to extend the size and complexity of IDPs that can be characterized. The proposed methods allow to overcome potential limitations deriving from the peculiar properties of IDPs (high proline content, extensive spectral overlap, fast amide proton nuclei exchange) and open new possibilities for the characterization at atomic level of PTMs and intermolecular interactions; they contribute to the methodological advancement of NMR and increase the number of available applications to study IDPs. Furthermore, topics of general interest such as the characterization of protein linkers and low-complexity regions, chosen as key examples in the broad area of IDPs, were addressed; the thorough NMR characterization of several disordered systems such as the N-terminus of AR and the ID3, ID4 and ID5 linkers of CBP provided important insights into their structural properties and functions.

Clearly, much work still needs to be done. It is very probable that the study of IDPs will change the concept of functional protein, which will evolve from a static picture to that of a highly dynamic ensemble of conformations. The advent of more powerful computational methods to screen protein sequences for intrinsic disorder will certainly reveal many more proteins that belong to this class: hence the importance of further improving NMR experiments. Experimental characterization of such proteins is expected to provide a large amount of data

contributing to increase our understanding of the molecular basis responsible for their function and to reveal a much larger number of ways in which proteins communicate in the cell, filling a gap of knowledge of about 50 years with respect to the study of well-folded proteins.

The NMR methods presented in this thesis are a key contribution to make possible such kind of studies.

Bibliography

Abel,A.M., Schuldt,K.M., Rajasekaran,K., Hwang,D., Riese,M.J., Rao,S., Thakar,M.S. and Malarkannan,S. (2015). IQGAP1: insights into the function of a molecular puppeteer. Mol. Immunol. 65, 336-349.

Adachi,H., Katsuno,M., Minamiyama,M., Waza,M., Sang,C., Nakagomi,Y., Kobayashi,Y., Tanaka,F., Doyu,M., Inukai,A., Yoshida,M., Hashizume,Y. and Sobue,G. (2005). Widespread nuclear and cytoplasmic accumulation of mutant androgen receptor in SBMA patients. Brain 128, 569-670.

Allison,J.R., Varnai,P., Dobson,C.M. and Vendruscolo,M. (2009). Determination of the Free Energy Landscape of α-Synuclein Using Spin Label Nuclear Magnetic Resonance Measurements. J. Am. Chem. Soc. 131, 18314-18326.

Alves da Costa,C. (2003). Recent advances on alpha-synuclein cell biology: functions and dysfunctions. Curr. Mol. Med. 3, 17-24.

Ardenkjaer-Larsen,J.-H., Boebinger,G.S., Comment,A., Duckett,S.B., Edison,A.S., Engelke,F., Griesinger,C., Griffin,R.G., Hilty,C., Maeda,H., Parigi,G., Prisner,T.F., Ravera,E., van Bentum,G.J.M., Vega,S., Webb,A., Luchinat,C., Schwalbe,H. and Frydman,L. (2015). Facing and overcoming biomolecular NMR's sensitivity challanges. Angew. Chem. Int. Ed. 54, 9162-9185.

Barbato,G., Ikura,M., Kay,L.E., Pastor,R.W. and Bax,A. (1992). Backbone dynamics of calmodulin studied by ^{15}N relaxation using inverse detected two-dimensional NMR spectroscopy: the central helix is flexible. Biochemistry 31, 5269-5278.

Bax,A. and Wang,A.C. (1995). Reparametrization of the Karplus Relation for $^{3}J(H^{\alpha}\text{-}N)$ and $^{3}J(H^{N}\text{-}C')$ in Peptides from Uniformly $^{13}C/^{15}N$-Enriched Human Ubiquitin. J. Am. Chem. Soc. 117, 1810-1813.

Bellotti,V., Mangione,P. and Stoppini,M. (1999). Biological activity and pathological implications of misfolded proteins. Cell. Mol. Life Sci. 55, 977-991.

Bentrop,D., Bertini,I., Cremonini,M.A., Forsén,S., Luchinat,C. and Malmendal,A. (1997). The solution structure of the paramagnetic complex of the N-terminal domain of calmodulin with two Ce^{3+} ions by ^{1}H NMR. Biochemistry 36, 11605-11618.

Berman,H.M., Westbrook,J., Feng,Z., Gilliland,G., Bhat,T.N., Weissig,H., Shindyalov,I.N. and Bourne,P.E. (2000). The Protein Data Bank. Nucleic Acids Res. 28, 235-242.

Bermel,W., Bertini,I., Duma,L., Emsley,L., Felli,I.C., Pierattelli,R. and Vasos,P.R. (2005). Complete assignment of heteronuclear protein resonances by protonless NMR spectroscopy. Angew. Chem. Int. Ed. 44, 3089-3092.

Bermel,W., Bertini,I., Felli,I.C., Lee,Y.-M., Luchinat,C. and Pierattelli,R. (2006). Protonless NMR experiments for sequence-specific assignment of backbone nuclei in unfolded proteins. J. Am. Chem. Soc. 128, 3918-3919.

Bermel,W., Felli,I.C., Kümmerle,R. and Pierattelli,R. (2008). ^{13}C direct-detection biomolecular NMR. Concepts Magn. Reson. 32A, 183-200.

Bermel,W., Bertini,I., Felli,I.C. and Pierattelli,R. (2009a). Speeding up ^{13}C direct detection Biomolecular NMR experiments. J. Am. Chem. Soc. 131, 15339-15345.

Bermel,W., Bertini,I., Csizmok,V., Felli,I.C., Pierattelli,R. and Tompa,P. (2009b). H-start for exclusively heteronuclear NMR spectroscopy: the case of intrinsically disordered proteins. J. Magn. Reson. 198, 275-281.

Bermel,W., Bertini,I., Chill,J.H., Felli,I.C., Haba,N., Kumar,V.M.V. and Pierattelli,R. (2012a). Exclusively heteronuclear ^{13}C-detected amino-acid-selective NMR experiments for the study of instrinsically disordered proteins (IDPs). ChemBioChem 13, 2425-2432.

Bermel,W., Bertini,I., Gonnelli,L., Felli,I.C., Koźmiński,W., Piai,A., Pierattelli,R. and Stanek,J. (2012b). Speeding up sequence specific assignment of IDPs. J. Biomol. NMR 53, 293-301.

Bermel,W., Bruix,M., Felli,I.C., Kumar,V.M.V., Pierattelli,R. and Serrano,S. (2013a). Improving the chemical shift dispersion of multidimensional NMR spectra of intrinsically disordered proteins. J. Biomol. NMR 55, 231-237.

Bermel,W., Felli,I.C., Gonnelli,L., Koźmiński,W., Piai,A., Pierattelli,R. and Zawadzka-Kazimierczuk,A. (2013b). High-dimensionality ^{13}C direct-detected NMR experiments for the automatic assignment of intrinsically disordered proteins. J. Biomol. NMR 57, 353-361.

Bernadó,P., Bertoncini,C.W., Griesinger,C., Zweckstetter,M. and Blackledge,M. (2005). Defining Long-Range Order and Local Disorder in Native α-Synuclein Using Residual Dipolar Couplings. J. Am. Chem. Soc. 127, 17968-17969.

Bertini,I., Luchinat,C. and Parigi,G. (2001a). Solution NMR of paramagnetic molecules. Elsevier, Amsterdam.

Bertini,I., Donaire,A., Jiménez,B., Luchinat,C., Parigi,G., Piccioli,M. and Poggi,L. (2001b). Paramagnetism-based Versus Classical Constraints: An Analysis of the Solution Structure of Ca Ln Calbindin D9k. J. Biomol. NMR 21, 85-98.

Bertini,I., Felli,I.C., Gonnelli,L., Kumar,V.M.V. and Pierattelli,R. (2011). High-resolution characterization of intrinsic disorder in proteins: expanding the suite of ^{13}C detected NMR experiments to determine key observables. ChemBioChem 12, 2347-2352.

Bertini,I., Mc Greevy,K. and Paigi,G. eds. (2012). NMR of biomolecules. Wiley-Blackwell, Weinheim, Germany.

Bertoncini,C.W., Jung,Y.-S., Fernández,C.O., Hoyer,W., Griesinger,C., Jovin,T.M. and Zweckstetter,M. (2005). Release of long-range tertiary interactions potentiates aggregation of natively unstructured a-synuclein. Proc. Natl. Acad. Sci. USA 102, 1430-1435.

Blobel,G.A. (2000). CREB-binding protein and p300: molecular integrators of hematopoietic transcription. Blood 95, 745-755.

Bonvin,A.M., Boelens,R. and Kaptein,R. (2005). NMR analysis of protein interactions. Curr. Opin. Chem. Biol. 9, 501-508.

Brutscher,B., Felli,I.C., Gil-Caballero,S., Hošek,T., Kümmerle,R., Piai,A., Pierattelli,R. and Sólyom,Z. (2015). NMR Methods for the Study of Instrinsically Disordered Proteins Structure, Dynamics, and Interactions: General Overview and Practical Guidelines. Adv. Exp. Med. Biol. 870, 122.

Burré,J., Sharma,M., Tsetsenis,T., Buchman,V., Etherton,M.R., and Südhof,T.C. (2010). Alpha-synuclein promotes SNARE-complex assembly in vivo and in vitro. Science 329, 1663-1667.

Cavalli,A., Salvatella,X., Dobson,C.M. and Vendruscolo,M. (2007). Protein structure determination from NMR chemical shifts. Proc. Natl. Acad. Sci. USA 104, 9615-9620.

Chen,S., Berthelier,V., Hamilton,J.B., O'Nuallain,B. and Wetzel,R. (2002). Amyloid-like features of polyglutamine aggregates and their assembly kinetics. Biochemistry 41, 7391-7399.

Coggins,B.E. and Zhou,P. (2007). Sampling of the NMR time domain along concentric rings. J. Magn Reson. 184, 207-221.

Coggins,B.E. and Zhou,P. (2008). High resolution 4-D spectroscopy with sparse concentric shell sampling and FFT-CLEAN. J. Biomol. NMR 42, 225-239.

Coggins,B.E., Venters,R.A. and Zhou,P. (2010). Radial sampling for fast NMR: Concepts and practices over three decades. Prog. Nucl. Magn. Reson. Spectrosc. 57, 381-419.

Coletta,A., Pinney,J.W., Solís,D.Y., Marsh,J., Pettifer,S.R. and Attwood,T.K. (2010). Low-complexity regions within protein sequences have position-dependent roles. BMC Syst. Biol. 4.

Cookson,M.R. (2009). alpha-Synuclein and neuronal cell death. Mol. Neurodegener. 4:9.

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

Das,S., Pal,U., Das,S., Bagga,K., Roy,A., Mrigwani,A. and Maiti,N.C. (2014). Sequence complexity of amyloidogenic regions in intrinsically disordered human proteins. Plos ONE 9, e89781.

De Guzman,R.N., Wojciak,J.M., Martinez-Yamout,M.A., Dyson,H.J. and Wright,P.E. (2005). CBP/p300 TAZ1 domain forms a structured scaffold for ligand binding. Biochemistry 44, 490-497.

Dedmon,M.M., Lindorff-Larsen,K., Christodoulou,J., Vendruscolo,M. and Dobson,C.M. (2005). Mapping long-range interactions in alpha-synuclein using spin-label NMR and ensemble molecular dynamics simulations. J. Am. Chem. Soc. 127, 476-477.

Demchenko,A.P. (2001). Recognition between flexible protein molecules: induced and assisted folding. J. Mol. Recognit. 14, 42-61.

Dev,K.K., Hofele,K., Barbieri,S., Buchman,V.L., and van der Putten,H. (2003). Part II: alpha-synuclein and its molecular pathophysiological role in neurodegenerative disease. Neuropharm. 45, 14-44.

Dittmer,J. and Bodenhausen,G. (2004). Evidence for slow motion in proteins by multiple refocusing of heteronuclear nitrogen/proton multiple quantum coherences in NMR. J. Am. Chem. Soc. 126, 1314-1315.

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

Dötsch,V. and Wagner,G. (1996a). Editing for amino-acid type in CBCACONH experiments based on the $^{13}C^{\beta}$-$^{13}C^{\gamma}$ coupling. J. Magn Reson. B 111, 310-313.

Dötsch,V., Oswald,R.E., and Wagner,G. (1996b). Amino-acid type-selective triple-resonance experiments. J. Magn Reson. B 110, 107-111.

Dötsch,V., Oswald,R.E. and Wagner,G. (1996c). Selective identification of threonine, valine and isoleucine sequential connectivities with a TVI-CBCACONH experiment. J. Magn Reson. B 110, 304-308.

Dötsch,V., Matsuo,H. and Wagner,G. (1996d). Amino-acid type identification for deuterated protiens with a β-carbon-edited HNCOCACB experiment. J. Magn Reson. B 112, 95-100.

Dunker,A.K., Obradović,Z., Romero,P., Garner,E.C. and Brown,C.J. (2000). Intrinsic Protein Disorder in Complete Genomes. Genome Informatics 11, 161-171.

Dunker,A.K., Lawson,J.D., Brown,C.J., Williams,R.M., Romero,P., Oh,J.S., Ratliff,C.M., Hipps,K.W., Ausio,J., Nissen,M.S., Reeves,R., Kang,C., Kissinger,C.R., Bailey,R.W., Griswold,M.D., Chiu,W., and Garner,E.C. and Obradović,Z. (2001). Intrinsically disordered protein. J. Mol. Graph. Model 19, 26-59.

Dunker,A.K., Brown,C.J., Lawson,J.D., Iakouchova,L.M. and Obradović,Z. (2002). Intrinsic disorder and protein function. Biochemistry 41, 6573-6582.

Dyson,H.J. and Wright,P.E. (2001). Nuclear magnetic resonance methods for the elucidation of structure and dynamics in disordered states. Methods In Enzymology 339, 258-271.

Dyson,H.J. and Wright,P.E. (2002). Coupling of folding and binding for unstructured proteins. Curr Opin Struct Biol 12, 54-60.

Dyson,H.J. and Wright,P.E. (2004). Unfolded proteins and protein folding studied by NMR. Chem. Rev. 104, 3607-3622.

Dyson,H.J. and Wright,P.E. (2005). Intrinsically unstructured proteins and their functions. Nat. Rev. Mol. Cell Biol. 6, 197-208.

Dziekański,P., Grudziaż,K., Jarvoll,P., Koźmiński,W. and Zawadzka-Kazimierczuk,A. (2015). (13)C-detected NMR experiments for automatic resonance assignment of IDPs and multiple-fixing SMFT processing. J. Biomol. NMR 62, 179-190.

Ellerby,L.M., Hackam,A.S., Propp,S.S., Ellerby,H.M., Rabizadeh,S., Cashman,N.R., Trifiro,M.A., Pinsky,L., Wellington,C.L., Salvesen,G.S., Hayden,M.R. and Bredesen,D.E. (1999). Kennedy's disease: caspase cleavage of the androgen receptor is a crucial event in cytotoxicity. J. Neurochem. 72, 185-195.

Ernst,R.R. and Anderson,W.A. (1966). Application of Fourier transform spectroscopy to magnetic resonance. Rev. Sci. Instr. 37, 93.

Farrow,N.A., Zhang,O., Forman-Kay,J. and Kay,L.E. (1994). A heteronuclear correlation experiment for simultaneous determination of ^{15}N longitudinal decay and chemical exchange rates of systems in slow equilibrium. J. Biomol. NMR 4, 727-734.

Farrow,N.A., Zhang,O., Szabo,A., Torchia,D.A. and Kay,L.E. (1995). Spectral density function mapping using ^{15}N relaxation data exclusively. J. Biomol. NMR 6, 153-162.

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

Felli,I.C., Piai,A. and Pierattelli,R. (2013). Recent advances in solution NMR studies: ^{13}C direct detection for biomolecular NMR applications. Ann. Rep. NMR Spectroscop. 359-418.

Felli,I.C. and Pierattelli,R. (2014). Novel methods based on ^{13}C detection to study intrinsically disordered proteins. J. Magn. Reson. 241, 115-125.

Felli,I.C. and Pierattelli,R. (2015). Spin-state-selctive methods in solution- and solid-state biomolecular ^{13}C NMR. Prog. NMR Spectrosc. 84, 1-13.

Filippakopoulos,P., Picaud,S., Mangos,M., Keates,T., Lambert,J.P., Barsyte-Lovejoy,D., Felletar,I., Volkmer,R., Müller,S., Pawson,T., Gingras,A.C., Arrowsmith,C.H. and Knapp,S. (2012). Histone recognition and large-scale structural analysis of the human bromodomain family. Cell 149, 214-231.

Frydman,L., Scherf,T. and Lupulescu,A. (2002). The acquisition of multidimensional NMR spectra within a single scan. Proc. Natl. Acad. Sci. USA 99, 15858-15862.

Fuxreiter,M., Simon,I., Friedrich,P. and Tompa,P. (2004). Preformed structural elements feature in partner recognition by intrinsically unstructured proteins. J. Mol. Biol. 338, 1015-1026.

Fuxreiter,M., Tompa,P. and Simon,I. (2007). Local structural disorder imparts plasticity on linear motifs. Bioinformatics 23, 950-956.

Gelmann,E.P. (2002). Molecular biology of the androgen receptor. J. Clin. Oncol. 20, 3001-3015.

Gerum,C., Silver,R., Wirmer-Bartoschek,J.,ì and Schwalbe,H. (2009). Unfolded-state structure and dynamics influence the fibril formation of human prion protein. Angew. Chem. Int. Ed. 48, 9452-9456.

Gil,S., Hošek,T., Sólyom,Z., Kümmerle,R., Brutscher,B., Pierattelli,R. and Felli,I.C. (2013). NMR studies of intrinsically disordered proteins near physiological conditions. Angew. Chem. Int. Ed. 52, 11808-11812.

Giordano,A. and Avantaggiati,M.L. (1999). p300 and CBP: partners for life and death. J. Cell Physiol. 181, 218-230.

Goodman,R.H. and Smolik,S. (2000). CBP/p300 in cell growth, transformation, and development. Genes Dev. 14, 1553-1577.

Grasberger,H. and Bell,G.I. (2005). Subcellular recruitment by TSG118 and TSPYL implicates a role for zinc finger protein 106 in a novel developmental pathway. Int. J. Biochem. Cell Biol. 37, 1421-1437.

Griesinger,C., Hennig,M., Marino,J.P., Reif,B., Richter,C. and Schwalbe,H. (1999). Modern Techniques in Protein NMR: Methods for the determination of Torsion Angle Restraints in Biomacromolecules. Biol. Magn. Reson. 16, 259-367.

Grzesiek,S. and Bax,A. (1992). Correlating backbone amide and side chain resonances in larger proteins by multiple relayed triple resonance NMR. J. Am. Chem. Soc. 114, 6291-6293.

Gunasekaran,K., Tsai,C.J., Kumar,S., Zanuy,D., and Nussinov,R. (2003). Extended disordered proteins: targeting function with less scaffold. Trends Biochem. Sci. 28, 81-85.

Habchi,J., Tompa,P., Longhi,S. and Uversky,V.N. (2014). Introducing protein intrinsic disorder. Chem. Rev. 114, 6561-6588.

Hansen,J.C., Lu,X., Ross,E.D. and Woody,R.W. (2006). Intrinsic protein disorder, amino acid composition, and histone terminal domains. J. Biol. Chem. 281, 1853-1856.

Hiller,S., Fiorito,F., Wüthrich,K. and Wider,G. (2005). Automated projection spectroscopy (APSY). Proc. Natl. Acad. Sci. USA 102, 10876-10881.

Hiller,S., Wasmer,C., Wider,G. and Wüthrich,K. (2007). Sequence-specific resonance assignment of soluble nonglobular proteins by 7D APSY-NMR spectroscopy. J. Am. Chem. Soc. 129, 10823-10828.

Hoch,J.C. and Stern,A.S. (1996). NMR data Processing. Wiley-Interscience, New York.

Holland,D.J., Bostock,M.J., Gladden,L.F. and Nietlispach,D. (2011). Fast multidimensional NMR spectroscopy using compressed sensing. Angew. Chem. Int. Ed. 50, 6548-6551.

Hošek,T., Gil-Caballero,S., Pierattelli,R., Brutscher,B. and Felli,I.C. (2015). Longitudinal relaxation properties of $^{1}H^{N}$ and $^{1}H^{\alpha}$ determined by direct-detected ^{13}C NMR experiments to study intrinsically disordered. J. Magn. Reson. 254, 19-26.

Huang,Y. and Liu,Z. (2009). Kinetic advantage of intrinsically disordered proteins in coupled folding-binding process: a critical assessment of the "fly-casting" mechanism. J. Mol. Biol. 393, 1143-1159.

Hwang,T.L., Van Zijl,P.C.M. and Mori,S. (1998). Accurate quantification 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.

Hyberts,S.G., Milbradt,A.G., Wagner,A.B., Arthanari,H. and Wagner,G. (2012). Application of iterative soft thresholding for fast reconstruction of NMR data non-uniformly sampled with multidimensional Poisson Gap scheduling. J. Biomol. NMR 52, 315-327.

Iakoucheva,L.M., Brown,C.J., Lawson,J.D., Obradović,Z., and Dunker,A.K. (2002). Intrinsic disorder in cell-signaling and cancer-associated proteins. J Mol Biol 323, 573-584.

Iakoucheva,L.M., Radivojac,P., Brown,C.J., O'Connor,T.R., Sikes,J.G., Obradović,Z. and Dunker,A.K. (2004). The importance of intrinsic disorder for protein phosphorylation. Nucleic Acids Res. 32, 1037-1049.

Ishima,R. and Torchia,D.A. (2000). Protein dynamics from NMR. Nature Struct. Biol. 7, 740-743.

Jadeski,L., Mataraza,J.M., Jeong,H.W., Li,Z. and Sacks,D.B. (2008). IQGAP1 stimulates proliferation and enhances tumorigenesis of human breast epithelial cells. J. Biol. Chem. 283, 1008-1017.

Jao,C.C., Hedge,B.G., Chen,J., Haworth,I.S., and Langen,R. (2008). Structure of membrane-bound alpha-synuclein from site-directed spin labeling and computational refinement. Proc. Natl. Acad. Sci. USA 105, 19666-19671.

Jaravine,V., Ibraghimov,I. and Orekhov,V.Y. (2006). Removal of a time barrier for high-resolution multidimensional NMR spectroscopy. Nat. Methods 3, 605-607.

Jaravine,V.A., Zhuravleva,A.V., Permi,P., Ibraghimov,I. and Orekhov,V.Y. (2008). Hyperdimensional NMR spectroscopy with nonlinear sampling. J. Am. Chem. Soc. 130, 3927-3936.

Jensen,M.R., Zweckstetter,M., Huang,J.R. and Blackledge,M. (2014). Exploring free-energy landscapes of intrinsically disordered proteins at atomic resolution using NMR spectroscopy. Chem. Rev. 114, 6632-6660.

Jones,D.T. and Cozzetto,D. (2015). DISOPRED3: precise disordered region predictions with annotated protein-binding activity. Bioinformatics 31, 857-863.

Jung,Y.S. and Zweckstetter,M. (2004). Mars: robust automatic backbone assignment of proteins. J. Biomol. NMR 30, 11-23.

Katsuno,M., Adachi,H., Waza,M., Banno,H., Suzuki,K., Tanaka,F., Doyu,M. and Sobue,G. (2006). Pathogenesis, animal models and therapeutics in spinal and bulbar muscular atrophy (SBMA). Exp. Neurol. 200, 8-18.

Kay,L.E., Torchia,D.A. and Bax,A. (1989). Backbone dynamics of proteins as studied by ^{15}N inverse detected heteronuclear NMR spectroscopy: application to staphylococcal nuclease. Biochemistry 28, 8972-8979.

Kay,L.E. (2005). NMR studies of protein structure and dynamics. J. Magn. Reson. 173, 193-207.

Kazimierczuk,K., Zawadzka,A., Koźmiński,W. and Zhukov,I. (2006). Random sampling of evolution time space and Fourier transform processing. J. Biomol. NMR 36, 157-168.

Kazimierczuk,K., Zawadzka,A., Koźmiński,W. and Zhukov,I. (2007). Lineshapes and artifacts in Multidimensional Fourier Transform of arbitrary sampled NMR data sets. J. Magn. Reson. 188, 344-356.

Kazimierczuk,K., Zawadzka,A. and Koźmiński,W. (2008). Optimization of random time domain sampling in multidimensional NMR. J. Magn. Reson. 192, 123-130.

Kazimierczuk,K., Zawadzka,A. and Koźmiński,W. (2009). Narrow peaks and high dimensionalities: Exploiting the advantages of random sampling. J. Magn. Reson. 197, 219-228.

Kazimierczuk,K., Stanek,J., Zawadzka-Kazimierczuk,A. and Koźmiński,W. (2010). Random sampling in multidimensional NMR spectroscopy. Prog. NMR Spectrosc. 57, 420-434.

Kazimierczuk,K., Misiak,M., Stanek,J., Zawadzka-Kazimierczuk,A. and Koźmiński,W. (2012). Generalized Fourier transform for non-uniform sampled data. Top. Curr. Chem. 316, 79-124.

Kazimierczuk,K., Stanek,J., Zawadzka-Kazimierczuk,A. and Koźmiński,W. (2013). High-dimensional NMR spectra for structural studies of biomolecules. ChemPhysChem 14, 3015-3025.

Kennedy,W.R. and Alter,M.S. (1968). Progressive proximal spinal and bulbar muscular atrophy of late onset. A sex-linked recessive trait. Neurology 18, 671-680.

Kjaergaard,M., Teilum,K. and Poulsen,F.M. (2010). Conformational selection in the molten globule state of the nuclear coactivator binding domain of CBP. Proc. Natl. Acad. Sci. USA 107, 12535-12540.

Kleckner,I.R. and Foster,M.P. (2011). An introductio to NMR-based approaches for measuirng protein dynamics. Biochim. Biophys. Acta 1814, 942-968.

Kobayashi,Y., Miwa,S., Merry,D.E., Kume,A., Mei,L., Doyu,M. and Sobue,G. (1998). Caspase-3 cleaves the expanded androgen receptor protein of spinal and bulbar muscular atrophy in a polyglutamine repeat length-dependent manner. Biochem. Biophys. Res. Commun. 252, 145-150.

Konrat,R. (2014). NMR contributions to structural dynamics studies of intrinsically disordered proteins. J. Magn. Reson. 241, 74-85.

Kragelj,J., Ozenne,V., Blackledge,M. and Jensen,M.R. (2013). Conformational propensities of intrinsically disordered proteins from NMR chemical shifts. Chemphyschem. 14, 3034-3045.

Kumar,R., Atamna,H., Zakharov,M.N., Bashin,S., Khan,S.H. and Jasuja,R. (2011). Role of the androgen receptor CAG repeat polymorphism in prostate cancer, and spinal and bulbar muscular atrophy. Life Sci. 88, 565-571.

Kumari,B., Kumar,R. and Kumar,M. (2015). Low complexity and disordered regions of proteins have different structural and amino acid preferences. Mol. Biosyst. 11, 585-594.

La Spada,A.R., Wilson,E.M., Lubahn,D.B., Harding,A.E. and Fischbeck,K.H. (1991). Androgen receptor gene mutations in X-linked spinal and bulbar muscular atrophy. Nature 352, 77-79.

Larson,M.E., Sherman,M.A., Greimel,S., Kuskowski,M., Schneider,J.A., Bennett,D.A. and Lesné,S.E. (2012). Soluble a-synuclein is a novel modulator of Alzheimer's disease pathophysiology. J. Neurosci. 32, 10253-10266.

Lee,C., Kalmar,L., Xue,B., Tompa,P., Daughdrill,G.W., Uversky,V.N. and Han,K.H. (2014). Contribution of proline to the pre-structuring tendency of transient helical secondary structure elements in intrinsically disordered proteins. Biochim. Biophys. Acta. 1840, 993-1003.

Lee,C.W., Martinez-Yamout,M.A., Dyson,H.J. and Wright,P.E. (2010). Structure of the p53 transactivation domain in complex with the nuclear receptor

coactivator binding domain of CREB binding protein. Biochemistry 49, 9964-9971.

Lee,S.H., Kim,D.H., Han,J.J., Cha,E.J., Lim,J.E., Cho,Y.J., Lee,C. and Han,K.H. (2012). Understanding pre-structured motifs (PreSMos) in intrinsically unfolded proteins. Curr. Prot. Pept. Sci. 13, 35-54.

Legge,G.B., Martinez-Yamout,M.A., Hambly,D.M., Trinh,T., Lee,B.M., Dyson,H.J. and Wright,P.E. (2004). ZZ domain of CBP: an unusual zinc finger fold in a protein interaction module. J. Mol. Biol. 343, 1081-1093.

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

Li,F., Lee,J.H., Grishaev,A., Ying,J. and Bax,A. (2015). High accuracy of Karplus equations for relating three-bond J couplings to protein backbone torsion angles. ChemPhysChem 16, 572-578.

Li,M., Miwa,S., Kobayashi,Y., Merry,D.E., Yamamoto,M., Tanaka,F., Doyu,M., Hashizume,Y., Fischbeck,K.H. and Sobue,G. (1998). Nuclear inclusions of the androgen receptor protein in spinal and bulbar muscular atrophy. Ann. Neurol. 44, 249-254.

Linding,R., Russel,R.B., Neduva,V. and Gibson,T.J. (2003a). GlobPlot: exploring protein sequences for globularity and disorder. Nucleic Acids Res. 31, 3701-3708.

Linding,R., Jensen,L.J., Diella,F., Bork,P., Gibson,T.J. and Russell,R.B. (2003b). Protein disorder prediction: Implications for structural proteomics. Structure 11, 1453-1459.

Liu,X., Wang,L., Zhao,K., Thompson,P.R., Hwang,Y., Marmorstein,R. and Cole,P.A. (2008). The structural basis of protein acetylation by the p300/CBP transcriptional coactivator. Nature 451, 846-850.

Louhivouri M., Pääkkönen,K., Fredriksson K., Permi,P., Lounila J. and Annila,A. (2003). On the origin of residual dipolar couplings from denatured proteins. J. Am. Chem. Soc. 125, 15647-15650.

Lu,K.P., Liou,Y.C. and Zhou,X.Z. (2002). Pinning down proline-directed phosphorylation signaling. Trends Cell Biol. 12, 164-172.

Luan,T., Jaravine,V., Yee,A., Arrowsmith,C. and Orekhov,V.Y. (2005). Optimization of resolution and sensitivity of 4D NOESY using multi-dimensional decomposition. J. Biomol. NMR 33, 1-14.

Malmodin,D. and Billeter,M. (2005). Multiway Decomposition of NMR Spectra with Coupled Evolution Periods. J. Am. Chem. Soc. 127, 13486-13487.

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

Mäntylahti,S., Hellman,M. and Permi,P. (2011). Extension of the HA-detection based approach: (HCA)CON(CA)H and (HCA)NCO(CA)H experiments for the main-chain assignment of intrinsically disordered proteins. J. Biomol. NMR 49, 99-109.

Marsh,J.A., Singh,V.K., Jia,Z. and Forman-Kay,J.D. (2006). Sensitivity of secondary structural propensities to sequence differences between α– and γ–synuclein: implications for fibrillation. Protein Sci. 15, 2795-2804.

Martins,F., Gonçalves,R., Oliveira,J., Cruz-Monteaqudo,M., Nieto-Villar,J.M., Paz-Y-Miño,C., Rebelo,I. and Tejera,E. (2015). Unravelling the relationship between protein sequence and low-complexity regions entropies: Interactome implications. J. Theor. Biol. 382, 320-327.

Masters,C.L., Multhaup,G., Simms,G., Pottgiesser,J., Martins,R.N. and Beyreuther,K. (1985). Neuronal Origin of A Cerebral Amyloid - Neurofibrillary Tangles of Alzheimers-Disease Contain the Same Protein As the Amyloid of Plaque Cores and Blood-Vessels. Embo Journal 4, 2757-2763.

Matsuki,Y., Eddy,M.T. and Herzfeld,J. (2009). Spectroscopy by integration of frequency and time domain information for fast acquisition of high-resolution dark spectra. J. Am. Chem. Soc. 131, 4648-4656.

McEwan,I.J., Lavery,D., Fischer,K. and Watt,K. (2007). Natural disordered sequences in the amino terminal domain of nuclear receptors: lessons from the androgen and glucocorticoid receptors. Nucl. Recept. Signal. 5, e001.

McLean,P.J., Kawamata,H., Ribich,S. and Hyman,B.T. (2000). Membrane association and protein conformation of alpha-synuclein in intact neurons. Effect of Parkinson's disease-linked mutations. J. Biol. Chem. 275, 8812-8816.

Melacini,G., Kaptein,R. and Boelens,R. (1999). Editing of Chemical Exchange-Relayed NOEs in NMR Experiments for the observation of protein-Water interactions. J. Magn. Reson. 136, 214-218.

Merry,D.E., Kobayashi,Y., Bailey,C.K., Taye,A.A. and Fischbeck,K.H. (1998). Cleavage, aggregation and toxicity of the expanded androgen receptor in spinal and bulbar muscular atrophy. Hum. Mol. Genet. 7, 693-701.

Miller,M., Dauter,Z., Cherry,S., Tropea,J.E. and Wlodawer,A. (2009). Structure of the Taz2 domain of p300: insights into ligand binding. Acta Crystallogr. D Biol. Crystallogr. 65, 1301-1308.

Mirkin,S.M. (2007). Expandable DNA repeats and human disease. Nature 447, 932-940.

Mishkovsky,M. and Frydman,L. (2009). Principles and progress in ultrafast multidimensional nuclear magnetic resonance. Annu. Rev. Phys. Chem. 60, 429-448.

Mittag,T. and Forman-Kay,J. (2007). Atomic-level characterization of disordered protein ensembles. Curr. Opin. Struct. Biol. 17, 3-14.

Mobli,M., Stern,A.S. and Hoch,J.C. (2006). Spectral reconstruction methods in fast NMR: Reduced dimensionality, random sampling and maximum entropy. J. Magn. Reson. 182, 96-105.

Mobli,M., Maciejewski,M.W., Gryk,M.R. and Hoch,J.C. (2007). Automatic maximum entropy spectral reconstruction in NMR. J. Biomol. NMR 39, 133-139.

Mobli,M. and Hoch,J.C. (2008). Maximum Entropy Spectral Reconstruction of Non-Uniformly Sampled Data. Concepts Magn. Reson. A 32A, 436-448.

Motáčkova,V., Nováček,J., Zawadzka-Kazimierczuk,A., Kazimierczuk,K., Židek,L., Sanderová,H., Krásný,L., Koźmiński,W. and Sklenář,V. (2010). Strategy for complete NMR assignment of disordered proteins with highly repetitive sequences based on resolution-enhanced 5D experiments. J. Biomol. NMR 48, 169-177.

Narayanan,R.L., Dürr,H.N., Bilbow,S., Biernat,J., Mendelkow,E. and Zweckstetter,M. (2010). Automatic Assignment of the Intrinsically Disordered Protein Tau with 441-Residues. J. Am. Chem. Soc. 132, 11906-11907.

Nováček,J., Zawadzka-Kazimierczuk,A., Papoušková,V., Židek,L., Sanderová,H., Krásný,L., Koźmiński,W. and Sklenář,V. (2011). 5D ^{13}C-detected experiments for backbone assignment of unstructured proteins with a very low signal dispersion. J. Biomol. NMR 50, 1-11.

Nováček,J., Žídek,L. and Sklenář,V. (2014). Toward optimal-resolution NMR of intrinsically disordered proteins. J. Magn. Reson. 241, 41-52.

Nowakowski,M., Saxena,S., Stanek,J., Żerko,S. and Koźmiński,W. (2015). Applications of high dimensionality experiments to biomolecular NMR. Progr. NMR Spectrosc. 90-91, 49-73.

Obolensky,OI., Schlepckow,K., Schwabe,H. and Solov'yov,A. (2007). Theoretical framework for NMR residual dipolar couplings in unfolded proteins. J Biomol NMR 39, 1-16.

Orekhov,V.Y. and Jaravine,V.A. (2011a). Analysis of non-uniformly sampled spectra with multi-dimensional decomposition. Progress in NMR Spectroscopy 59, 271-292.

Orekhov,V.Y. and Kazimierczuk,K. (2011b). Accelerated NMR spectroscopy by using compressed sensing. Angew. Chem Int. Ed. 50, 5556-5559.

Panchal,S.C., Bhavesh,N.S. and Hosur,R.V. (2001). Improved 3D triple resonance experiments, HNN and HN(C)N, for H^N and ^{15}N sequential correlations (^{13}C, ^{15}N) labeled proteins: application to unfolded proteins. J. Biomol. NMR 20, 135-147.

Pantoja-Uceda,D. and Santoro,M.M. (2008). Amino acid type identification in NMR spectra of proteins via β- and γ-carbon edited experiments. J. Magn Reson. 195, 187-195.

Pantoja-Uceda,D. and Santoro,J. (2012). New amino acid residue type identification experiments valid for protonated and deuterated proteins. J. Biomol. NMR 54, 145-153.

Pantoja-Uceda,D. and Santoro,J. (2013a). Direct correlation of consecutive C'-N groups in proteins: a method for the assignment of intrinsically disordered proteins. J. Biomol. NMR 57, 57-63.

Pantoja-Uceda,D. and Santoro,J. (2013b). A suite of amino acid residue type classification pulse sequences for ^{13}C-detected NMR of proteins. J. Magn. Reson. 234, 190-196.

Pantoja-Uceda,D. and Santoro,J. (2014). New ^{13}C-detected experiments for the assignment of intrinsically disordered proteins. J. Biomol. NMR 59, 43-50.

Parigi,G., Rezaei-Ghaleh,N., Giachetti,A., Becker,S., Fernandez,C., Blackledge,M., Griesinger,C., Zweckstetter,M. and Luchinat,C. (2014). Long-Range Correlated Dynamics in Intrinsically Disordered Proteins. J. Am. Chem. Soc. 136, 16201-16209.

Peng,J.W. and Wagner,G. (1994). Investigation of protein motions via relaxation measurements. Methods Enzymol. 239, 563-596.

Pereira de Jésus-Tran,K., Côté,P.L., Cantin,L., Blanchet,J., Labrie,F. and Breton,R. (2006). Comparison of crystal structures of human androgen receptor ligand-binding domain complexed with various agonists reveals molecular determinants responsible for binding affinity. Protein Sci. 15, 987-999.

Pervushin,K., Riek,R., Wider,G. and 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.

Piai,A., Hošek,T., Gonnelli,L., Zawadzka-Kazimierczuk,A., Koźmiński,W., Brutscher,B., Bermel,W., Pierattelli,R. and Felli,I.C. (2014). "CON-CON" assignment strategy for highly flexible intrinsically disordered proteins. J. Biomol. NMR 60, 209-218.

Plotnikov,A.N., Yang,S., Zhou,T.J., Rusinova,E., Frasca,A. and Zhou,M.M. (2014). Structural insights into acetylated-histone H4 recognition by the bromodomain-PHD finger module of human transcriptional coactivator CBP. Structure 22, 353-360.

Prilusky,J., Felder,C.E., Zeev-Ben-Mordehai,T., Rydberg,E.H., Man,O., Beckmann,J.S., Silman,I. and Sussman,J.L. (2005). FoldIndex: a simple tool to predict whether a given protein sequence is intrinsically unfolded. Bioinformatics 21, 3435.

Rennella,E., Sólyom,Z. and Brutscher,B. (2014). Measuring hydrogen exchange in proteins by selective water saturation in (1)H- (15)N SOFAST/BEST-type experiments: advantages and limitations. J. Biomol. NMR 60, 99-107.

Rhodes,L.E., Freeman,B.K., Aau,S., Kokkinis,A.D., La Pean,A., Chen,C., Lehky,T.J., Shrader,J.A., Levy,E.W., Harris-Love,M., Di Prospero,N.A. and Fischbeck,K.H. (2009). Clinical features of spinal and bulbar muscular atrophy. Brain 132, 3242-3251.

Roche,D.B., Buenavista,M.T., Tetchner,S.J. and McGuffin,L.J. (2011). The IntFOLD server: an integrated web resource for protein fold recognition, 3D model quality assessment, intrinsic disorder prediction, domain prediction and ligand binding site prediction. Nucleic Acids Res. 39, W171-W176.

Romero,P., Obradović,Z., Li,X.H., Garner,E.C., Brown,C.J. and Dunker,A.K. (2001). Sequence complexity of disordered protein. Proteins-Structure Function and Genetics 42, 38-48.

Rucker,A.L. and Creamer,T.P. (2002). Polyproline II helical structure in protein unfolded states: lysine peptides revisited. Protein Sci. 11, 980-985.

Rucker,A.L., Pager,C.T., Campbell,M.N., Qualls,J.E. and Creamer,T.P. (2003). Host-guest scale of left-handed polyproline II helix formation. Proteins 53, 68-75.

Ryan,C.P. and Crespi,B.J. (2011). Androgen receptor polyglutamine repeat number: models of selection and disease susceptibility. Evol. Appl. 6, 180-196.

Salmon,L., Jensen,M.R., Bernadó,P. and Blackledge,M. (2012). Measurement and analysis of NMR residual dipolar couplings for the study of intrinsically disordered proteins. Methods Mol. Biol. 895, 115-25.

Schanda,P., Kupce,E. and 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. and Brutscher,B. (2006a). HET-SOFAST NMR for fast detection of structural compactness and heterogeneity along polypeptide chains. Magn. Reson. Chem. 44, S177-S184.

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

Schmidt,E. and Güntert,P. (2012). A new algorithm for reliable and general NMR resonance assignment. J. Am. Chem. Soc. 134, 12817-12829.

Schubert,M., Oschkinat,H. and Schmieder,P. (2001a). Amino acid type-selective backbone ^{1}H-^{15}N correlations for Arg and Lys. J. Biomol. NMR 20, 379-384.

Schubert,M., Oschkinat,H. and Schmieder,P. (2001b). MUSIC, selective pulses, and tuned delays: amino acid-type selective ^{1}H-^{15}N correlations, II. J. Magn. Reson. 148, 61-72.

Schubert,M., Oschkinat,H. and Schmieder,P. (2001c). MUSIC and aromatic residues: amino acid type-selective ^{1}H-^{15}N correlations, III. J. Magn. Reson. 153, 186-192.

Schubert,M., Labudde,D., Leitner,D., Oschkinat,H. and Schmieder,P. (2005). A modified strategy for sequence specific assignment of protein NMR spectra based on amino acid type selective experiments. J. Biomol. NMR 31, 115-128.

Serber,Z., Richter,C. and Dötsch,V. (2001). Carbon-detected NMR experiments to investigate structure and dynamics of biological macromolecules. ChemBioChem 2, 247-251.

Shaffer,P.L., Jivan,A., Dollins,D.E., Claessens,F. and Gewirth,D.T. (2004). Structural basis of androgen receptor binding to selective androgen response elements. Proc. Natl. Acad. Sci. USA 101, 4758-4763.

Shen,Y., Lange,O., Delaglio,F., Rossi,P., Aramini,J.M., Liu,G., Eletsky,A., Wu,Y., Singarapu,K.K., Lemak,A., Ignatchenko,A., Arrowsmith,C.H., Szyperski,T., Montelione,G.T., Baker,D. and Bax,A. (2008). Consistent blind protein structure generation from NMR chemical shift data. Proc. Natl. Acad. Sci. USA 105, 4685-4690.

Shi,Z., Olson,C.A., Rose,G.D., Baldwin,R.L. and Kallenbach,N.R. (2002). Polyproline II structure in a sequence of seven alanine residues. Proc. Natl. Acad. Sci. USA 99, 9190-9195.

Shimba,N., Stern,A.S., Craik,C.S., Hoch,J.C. and Dötsch,V. (2003). Elimination of $^{13}C^{\alpha}$ splitting in protein NMR spectra by deconvolution with maximum entropy reconstruction. J. Am. Chem. Soc. 125, 2382-2383.

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

Sickmeier,M., Hamilton,J.A., LeGall,T., Vacic,V., Cortese,M.S., Tantos,A., Szabo,B., Tompa,P., Chen,J., Uversky,V.N., Obradović,Z. and Dunker,A.K. (2007). DisProt: the database of disordered proteins. Nucleic Acids Res. 35, D786-D793.

Smith,J.L., Freebern,W.J., Collins,I., De Siervi,A., Montano,I., Haggerty,C.M., McNutt,M.C., Butscher,W.G., Dzekunova,I., Petersen,D.W., Kawasaki,E., Merchant,J.L. and Gardner,K. (2004). Kinetic profiles of p300 occupancy in vivo predict common features of promoter structure and coactivator recruitment. Proc. Natl. Acad. Sci. USA 101, 11554-11559.

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.

Spera,S. and Bax,A. (1991). Empirical Correlation between Protein Backbone Conformation and C^{α} and C^{β} ^{13}C Nuclear Magnetic Resonance Chemical Shifts. J. Am. Chem. Soc. 113, 5490-5492.

Spinelli,K.J., Taylor,J.K., Osterberg,V.R., Churchill,M.J., Pollock,E., Moore,C., Meshul,C.K. and Unni,V.K. (2014). Presynaptic alpha-synuclein aggregation in a mouse model of Parkinson's disease. J. Neurosci. 34, 2037-2050.

Stanek,J., Augustyniak,R. and Koźmiński,W. (2012). Suppression of sampling artefacts in high-resolution four-dimensional NMR spectra using signal separation algorithm. J. Magn. Reson. 214, 91-102.

Szyperski,T., Wider,G., Bushweller,J.H. and Wüthrich,K. (1993). Reduced dimensionality in triple resonance experiments. J. Am. Chem. Soc. 115, 9307-9308.

Tamiola,K., Acar,B. and Mulder,F.A. (2010). Sequence-Specific Random Coil Chemical Shifts of Intrinsically Disordered Proteins. J. Am. Chem. Soc. 132, 18000-18003.

Tamiola,K. and Mulder,F.A. (2012). Using NMR chemical shifts to calculate the propensity for structural order and disorder in proteins. Biochem. Soc. Trans. 40, 1014-1020.

Teilum,K., Kragelund,B.B. and Poulsen,F.M. (2002). Transient structure formation in unfolded acyl-coenzyme A-binding protien observed by site-directed spin labeling. J. Mol. Biol. 324, 349-357.

The UniProt Consortium (2015). UniProt: a hub for protein information. Nucleic Acids Res. 43, D204-D212.

Theillet,F.X., Smet-Nocca,C., Liokatis,S., Thongwichian,R., Kosten,J., Yoon,M.K., Kriwacki,R.W., Landrieu,I., Lippens,G. and Selenko,P. (2012). Cell signaling, post-translational protein modifications and NMR spectroscopy. J. Biomol. NMR 54, 217-236.

Theillet,F.X., Kalmar,L., Tompa,P., Han,K.H., Selenko,P., Dunker,A.K., Daughdrill,G.W. and Uversky,V.N. (2013). The alphabet of intrinsic disorder: I. Act like a Pro: On the abundance and roles of proline residues in intrinsically disordered proteins. Intr. Dis. Prot. 1, e24360.

Thomas,P.S.Jr., Fraley,G.S., Damian,V., Woodke,L.B., Zapata,F., Sopher,B.L., Plymate,S.R. and La Spada,A.R. (2006). Loss of endogenous androgen receptor protein accelerates motor neuron degeneration and accentuates androgen insensitivity in a mouse model of X-linked spinal and bulbar muscular atrophy. Hum. Mol. Genet. 15, 2225-2238.

Tompa,P. (2002). Intrinsically unstructured proteins. Trends Biochem. Sci. 27, 527-533.

Tompa,P. (2005). The interplay between structure and function in intrinsically unstructured proteins. FEBS Lett. 579, 3346-3354.

Tompa,P. and Varadi,M. (2014). Predicting the predictive power of IDP ensembles. Structure 22, 177-178.

Ulrich,E.L., Akutsu,H., Doreleijers,J.F., Harano,Y., Ioannidis,Y.E., Lin,J., Livny,M., Mading,S., Maziuk,D., Miller,Z., Nakatani,E., Schulte,C.F., Tolmie,D.E.,

Wenger,R.K., Yao,H. and Markley,J.L. (2008). BioMagResBank. Nucleic Acids Res. 36, D402-D408.

Uversky,V.N., Gillespie,J.R. and Fink,A.L. (2000). Why are "natively unfolded" proteins unstructured under physiologic conditions? Proteins Struct. Funct. Genet. 41, 415-427.

Uversky,V.N. (2002a). What does it mean to be natively unfolded? Eur J. Biochem. 269, 2-12.

Uversky,V.N. (2002b). Natively unfolded proteins: a point where biology waits for physics. Protein Sci. 11, 739-756.

Uversky,V., Oldfield,C.J. and Dunker,A.K. (2008). Intrinsically disordered proteins in human diseases: introducing the D2 concept. Annu. Rev. Biophys. 37, 215-246.

Uversky,V.N., and Dunker,A.K. (2010). Understanding protein non-folding. Biochim. Biophys. Acta. 1804, 1231-1264.

Uversky,V.N. (2011a). Intrinsically disordered proteins from A to Z. Biochem. Cell Biol. 43, 1090-1103.

Uversky,V.N. (2011b). Multitude of binding modes attainable by intrinsically disordered proteins: a portrait gallery of disorder-based complexes. Chem Soc Rev 40, 1623-1634.

Uversky,V.N. (2013). A decade and a half of protein intrinsic disorder: biology still waits for physics. Protein Sci. 22, 693-724.

Vacic,V., Oldfield,C.J., Mohan,A., Radivojac,P., Cortese,M.S., Uversky,V. and Dunker,A.K. (2007). Characterization of molecular recognition features, MoRFs, and their binding partners. Proteomics 6, 2351-2366.

van der Lee,R., Buljan,M., Lang,B., Weatheritt,R.J., Daughdrill,G.W., Dunker,A.K., Fuxreiter,M., Gough,J., Gsponer,J., Jones,D.T., Kim,P.M., Kriwacki,R.W., Oldfield,C.J., Pappu,R.V., Tompa,P., Uversky,V.N., Wright,P.E. and Babu,M.M. (2014). Classification of intrinsically disordered regions and proteins. Chem. Rev. 114, 6589-6631.

Van Roey,K., Uyar,B., Weatheritt,R.J., Dinkel,H., Seiler,M., Budd,A., Gibson,T.J. and Davey,N.E. (2014). Short linear motifs: ubiquitous and functionally diverse protein interaction modules directing cell regulation. Chem. Rev. 114, 6733-6778.

Varadi,M., Kosol,S., Lebrun,P., Valentini,E., Blackledge,M., Dunker,A.K., Felli,I.C., Forman-Kay,J.D., Kriwacki,R.W., Pierattelli,R., Sussman,J.L., Svergun,D.I.,

Uversky,V.N., Vendruscolo,M., Wishart,D.S., Wright,P.E. and Tompa,P. (2014). pE-DB: a database of structural ensemble of intrinsically disordered and of unfolded proteins. Nucleic Acids Res. 42, D326-D335.

Vögeli,B., Ying,J., Grishaev,A., and Bax,A. (2007). Limits on variations in protein backbone dynamics from precise measurements of scalar couplings. J. Am. Chem. Soc. 129, 9377-9385.

Volk,J., Herrmann,T. and Wüthrich,K. (2008). Automated sequence-specific protein NMR assignment using the memetic algorithm MATCH. J. Biomol. NMR 41, 127-138.

Vucetic,S., Brown,C.J., Dunker,A.K. and Obradovic,Z. (2003). Flavors of protein disorder. Proteins 52, 573-584.

Vucetic,S., Xie,H., Iakoucheva,L.M., Oldfield,C.J., Dunker,A.K., Obradovic,Z. and Uversky,V.N. (2007). Functional anthology of intrinsic disorder.2.Cellular components, domains, technical terms, developmental processes and coding sequence diversities correlated with long disordered regions. J. Proteome. Res. 6, 1899-1916.

Vuister,G.W. and Bax,A. (1993). Quantitative J correlation: a new approach for measuring homonuclear three-bond J($H^N H^\alpha$) coupling constants in ^{15}N-enriched proteins. J. Am. Chem. Soc. 115, 7772-7777.

Wang,A.C. and Bax,A. (1996). Determination of the backbone dihedral angles ϕ in human ubiquitin from reparametrized empirical Karplus equations. J. Am. Chem. Soc. 118, 2483-2494.

Wang,F., Marshall,C.B., Yamamoto,K., Li,G.Y., Gasmi-Seabrook,G.M., Okada,H., Mark,T.W. and Ikura,M. (2012). Structures of KIX domain of CBP in complex with two FOXO3a transactivation domains reveal promiscuity and plasticity in coactivator recruitment. Proc. Natl. Acad. Sci. USA 109, 6078-6083.

Wang,Q., Li,W., Liu,X.S., Carroll,J.S., Jänne,O.A., Keeton,E.K., Chinnaiyan,A.M., Pienta,K.J. and Brown,M. (2007). A hierarchical network of transcription factors governs androgen receptor-dependent prostate cancer growth. Mol. Cell. 27, 380-392.

Ward,J.J., Sodhi,J.S., McGuffin,L.J., Buxton,B.F. and Jones,D.T. (2004). Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. Journal of Molecular Biology 337, 635-645.

Wells,R.D., Dere,R., Hebert,M.L., Napierala,M. and Son,L.S. (2005). Advances in mechanisms of genetic instability related to hereditary neurological diseases. Nucleic Acids Res. 33, 3785-3798.

Wirmer,J., Berk,H., Ugolini,R., Redfield,C. and Schwalbe,H. (2006). Characterization of the unfolded state of bovine α-lactalbumin and comparison with unfolded states of homologous proteins. Protein Sci. 15, 1397-1407.

Wishart,D.S. and Sykes,B.D. (1994). The ^{13}C chemical shift index: a simple method for the identification of protein secondary structure using ^{13}C chemical shift data. J. Biomol. NMR 4, 171-180.

Wishart,D.S., Bigam,C.G., Holm,A., Hodges,R.S. and Sykes,B.D. (1995). ^{1}H, ^{13}C and ^{15}N random coil NMR chemical shifts of the common amino acids. I. Investigations of nearest-neighbor effects. J. Biomol. NMR 5, 67-81.

Wootton,J.C. (1994). Sequences with 'unusual' amino acid compositions. Curr. Opin. Struct. Biol. 4, 413-421.

Wootton,J.C. and Federhen,S. (1993). Statistics of local complexity in amino acid sequences and sequence databases. Computers&Chemistry 17, 149-163.

Wright,P.E. and Dyson,H.J. (1999). Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. J. Mol. Biol. 293, 321-331.

Wright,P.E. and Dyson,H.J. (2015). Intrinsically disordered proteins in cellular signalling and regulation. Nat. Rev. Mol. Cell Biol. 16, 18-29.

Wu,K.P. and Baum,J. (2010). Detection of transient interchain interactions in the intrinsically disordered protein alpha-synuclein by NMR paramagnetic relaxation enhancement. J. Am. Chem. Soc. 132, 5546-47.

Wüthrich,K. (1986). NMR of Proteins and Nucleic Acids. Wiley, New York.

Xie,H., Vucetic,S., Iakoucheva,L.M., Oldfield,C.J., Dunker,A.K., Uversky,V.N. and Obradovic,Z. (2007). Functional anthology of intrinsic disorder. 1. Biological processes and functions of proteins with long disordered regions. J. Proteome. Res. 6, 1882-1898.

Xue,B., Dunbrack,R.L.Jr., Williams,R.W., Dunker,A.K. and Uversky,V. (2010). PONDR-FIT: A meta-predictor of intrinsically disordered amino acids. Biochim Biophys Acta 1804, 996-1010.

Yao,J., Chung,J., Eliezer,D., Wright,P.E. and Dyson,H.J. (2001). NMR structural and dynamic characterization of the acid-unfolded state of apomyoglobin provides insights into the early events in protein folding. Biochemistry 40, 3561-3571.

Ying,J., Li,F., Lee,J.H. and Bax,A. (2014). $^{13}C^{\alpha}$ decoupling during direct observation of carbonyl resonances in solution NMR of isotopically enriched proteins. J. Biomol. NMR 60, 15-21.

Zawadzka-Kazimierczuk,A., Kazimierczuk,K. and Koźmiński,W. (2010). A set of 4D NMR experiments of enhanced resolution for easy resonance assignment in proteins. J. Magn. Reson. 202, 109-116.

Zawadzka-Kazimierczuk,A., Koźmiński,W. and Billeter,M. (2012a). TSAR: a program for automatic resonance assignment using 2D cross-sections of high dimensionality, high-resolution spectra. J. Biomol. NMR 54, 81-95.

Zawadzka-Kazimierczuk,A., Koźmiński,W., Sanderová,H. and Krásný,L. (2012b). High dimensional and high resolution pulse seqeunces for backbone resonance assignment of intrinsically disordered proteins. J. Biomol. NMR 52, 329-337.

Żerko,S. and Koźmiński,W. (2015). Six- and seven-dimensional experiments by combination of sparse random sampling and projection spectroscopy dedicated for backbone resonance assignment of intrinsically disordered proteins. J. Biomol. NMR. 63, 283-290.

Zoghbi,H.Y. and Orr,H.T. (2000). Glutamine repeats and neurodegeneration. Annu. Rev. Neurosci. 23, 217-247.

Appendix

Developed pulse sequences

The new NMR experiments are listed in the table below.

Experiment	Provided correlations
5D (HCA)CONCACON	$^{13}C'_{i-1}$-$^{15}N_i$-$^{13}C^{\alpha}_i$-$^{15}N_{i+1}$-$^{13}C'_i$ $^{13}C'_i$-$^{15}N_{i+1}$-$^{13}C^{\alpha}_i$-$^{15}N_{i+1}$-$^{13}C'_i$
5D ($H^{N\text{-}flip}N$)CONCACON	$^{13}C'_{i-1}$-$^{15}N_i$-$^{13}C^{\alpha}_i$-$^{15}N_{i+1}$-$^{13}C'_i$ $^{13}C'_i$-$^{15}N_{i+1}$-$^{13}C^{\alpha}_i$-$^{15}N_{i+1}$-$^{13}C'_i$
5D (H)CACON(CA)CON	$^{13}C^{\alpha}_i$-$^{13}C'_i$-$^{15}N_{i+1}$-$^{13}C'_{i+1}$-$^{15}N_{i+2}$ $^{13}C^{\alpha}_i$-$^{13}C'_i$-$^{15}N_{i+1}$-$^{13}C'_i$-$^{15}N_{i+1}$
5D (HACA)CON(CACO)NCO(CA)HA	$^{13}C'_{i-1}$-$^{15}N_i$-$^{15}N_{i+1}$-$^{13}C'_i$-$^{1}H^{\alpha}_i$
5D BT-(H)NCO(CAN)CONNH	$^{15}N_{i+1}$-$^{13}C'_i$-$^{13}C'_{i-1}$-$^{15}N_i$-$^{1}H^{N}_i$
5D BT-HN(COCAN)CONNH	$^{1}H^{N}_{i+1}$-$^{15}N_{i+1}$-$^{13}C'_{i-1}$-$^{15}N_i$-$^{1}H^{N}_i$
4D γ-selective-HCBCACON	$^{1}H^{\alpha/\beta}_i$-$^{13}C^{\alpha/\beta}_i$-$^{15}N_{i+1}$-$^{13}C'_i$

Chemical shift assignments

The following NMR chemical shifts assignments have been deposited in the BMRB.

BMRB entry	Protein	Experimental conditions
6968	α-synuclein	20.0 mM Pi, 200.0 mM NaCl, 0.5 mM EDTA, pH 6.5, 285.5 K
26671	α-synuclein	20.0 mM Pi, 200.0 mM NaCl, 0.5 mM EDTA, pH 7.4, 308.0 K
26616	ID4	20.0 mM Pi, 100.0 mM KCl, pH 6.5, 308.0 K
26639	ID4	20.0 mM Pi, 100.0 mM KCl, pH 6.5, 308.0 K
25606	4Q	20.0 mM Pi, 1.0 mM TCEP, pH 7.4, 278.0 K
25607	4Q	20.0 mM Pi, 1.0 mM TCEP, pH 7.4, 308.0 K
25608	25Q	20.0 mM Pi, 1.0 mM TCEP, pH 7.4, 278.0 K

Article 1. Recent Advances in Solution NMR Studies: ^{13}C Direct Detection for Biomolecular NMR Applications

Isabella C. Felli, Alessandro Piai, Roberta Pierattelli

Magnetic Resonance Center and Department of Chemistry “Ugo Schiff”, University of Florence, Via L. Sacconi 6, Sesto Fiorentino, Italy.

Published in 2013 in Annu.Rep.NMR Spectrosc., 80, 359-418

CHAPTER SIX

Recent Advances in Solution NMR Studies: ^{13}C Direct Detection for Biomolecular NMR Applications

Isabella C. Felli, Alessandro Piai, Roberta Pierattelli

Magnetic Resonance Center and Department of Chemistry "Ugo Schiff", University of Florence, Via L. Sacconi 6, Sesto Fiorentino, Italy

Contents

Abstract

Carbon-13 direct detection NMR provides a complementary tool for biomolecular applications, thanks to the development of a large variety of experimental variants and to recent improvements in instrumental sensitivity. It can be used as a general tool for any protein study and it can provide unique information very relevant for the characterization of a variety of different systems such as paramagnetic proteins and large multimeric protein assemblies, as well as intrinsically disordered proteins.

The different properties of ^{13}C respect to ^{1}H, which provide the rational for the experiments developed, are discussed. The technical aspects that needed to be solved, as well as the many experimental variants developed to address different cases, are presented. Application areas where these experiment result particularly useful are also described.

Annual Reports on NMR Spectroscopy, Volume 80
ISSN 0066-4103
http://dx.doi.org/10.1016/B978-0-12-408097-3.00006-8

Key Words: ^{13}C direct detection, Virtual decoupling, Protonless NMR, Exclusively heteronuclear NMR, Multidimensional NMR, Longitudinal relaxation enhancement (LRE), Non-uniform sampling (NUS), Paramagnetic proteins, Very large proteins, Intrinsically disordered proteins (IDPs)

1. INTRODUCTION

Carbon-13 direct-detected NMR experiments are widely used for the study of small molecules, thanks to the great amount of information that can be obtained from the observed chemical shifts and coupling topologies. However, after the pioneering studies by the group of John Markley, ^{13}C direct detection for biomolecular NMR applications in solution has been abandoned. A wide variety of experiments based on ^{1}H detection, which exploit heteronuclei in the indirect dimensions, have been developed, on the grounds of the much higher intrinsic sensitivity of ^{1}H with respect to that of ^{13}C, and are routinely used. Therefore, why worry about ^{13}C direct detection for biomolecular applications in solution?

The reason is probably twofold. On the one hand, the great improvement in instrumental sensitivity we have achieved in the past decade, initially exploited to study systems with limited solubility, can also be exploited to investigate nuclear spins characterized by a smaller gyromagnetic ratio and thus lower sensitivity. On the other, the great progress in the hardware, the vast array of experimental methods that have been developed, as well as progress in sample preparation and isotopic labelling strategies, have opened the way to the characterization through NMR spectroscopy of systems of increasing complexity. Several areas in which ^{1}H-detected experiments find limitations (despite the high sensitivity) have thus emerged, stimulating the design of alternative methods able to exploit the intrinsically different properties of the various nuclear spins. The fast progress in the implementation of hyperpolarization techniques might completely change our way of thinking in the future, with the intrinsic sensitivity of a specific nucleus playing a minor role in the choice of the specific approach!

Most of the high-field high-resolution NMR instruments are nowadays equipped with probe-heads characterized by an excellent sensitivity, not only for ^{1}H but also for ^{13}C. A wide variety of multidimensional NMR experiments based on ^{13}C direct detection have been developed for biomolecular NMR applications. It is thus time to evaluate the progress and the perspectives in this field. In this review, we focus on the peculiar properties of the heteronuclei and of ^{13}C in particular, to identify the key features of

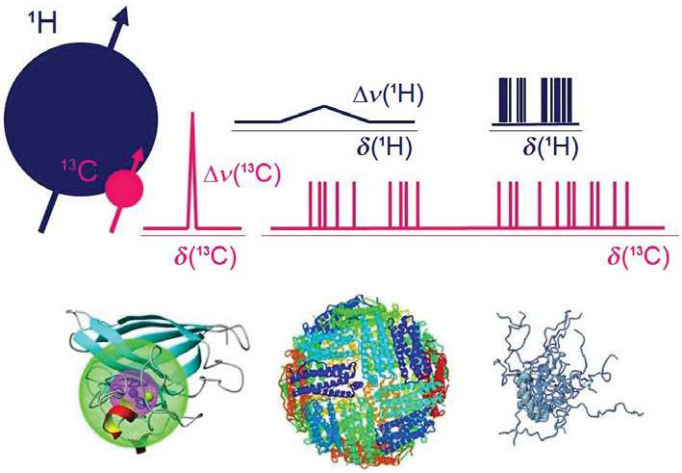


Figure 6.1 A schematic picture highlighting the intrinsically different properties of ^{1}H and ^{13}C spins, as well as different classes of proteins (paramagnetic, very large, intrinsically disordered) for which ^{13}C direct-detected *exclusively heteronuclear* NMR experiments can provide additional information respect to that available through ^{1}H-detected experiments. The proteins shown at the bottom are monomeric Cu(II)/Zn(II) superoxide dismutase, ferritin and oxidized cox17. *Adapted with permission by Wiley.*

heteronuclei that may be useful and complementary to the ones of protons (Fig. 6.1). We then summarize the many approaches proposed to overcome the problem of homonuclear couplings in the direct acquisition dimension that could potentially complicate the spectra and reduce the sensitivity. A survey of experiments that have been developed for a variety of different cases will then be presented, highlighting the many variants proposed. Finally, as biomolecular ^{13}C direct detection in many cases is not limited any longer by sensitivity, we show how all the approaches to reduce experimental time that have recently been implemented find application in the various ^{13}C-detected experiments. The suite of experiments developed can thus be used as a general complementary tool for biomolecular applications and can provide additional unique information for different kinds of proteins, such as paramagnetic proteins, large multimeric assemblies and intrinsically disordered proteins (IDPs) (Fig. 6.1).

2. INTRINSIC PROPERTIES OF HETERONUCLEI (^{13}C, ^{15}N)

Despite sharing the same nuclear spin $S=1/2$, the isotopes ^{1}H, ^{13}C and ^{15}N on which protein biomolecular NMR is based[1,2] do have

significantly different properties. The possibility to exploit three different nuclei that cover nearly all sites in a protein, and to correlate them one to the other through different interactions, accounts for the great success of NMR spectroscopy to study biological macromolecules with a wide variety of NMR experiments under continuous development. The three different nuclei differ in their gyromagnetic ratio, the contributions to their chemical shifts, the spin–spin coupling constants and, in general, the coupling topologies as well as the nuclear relaxation properties. The final result of a specific experiment depends on all the above and in several instances it may be useful to perform heteronuclear-detected experiments even if it is well known that heteronuclei are characterized by a lower gyromagnetic ratio and thus by a lower intrinsic sensitivity compared to proton.

The intrinsic instrumental sensitivity depends on a variety of factors, including the static magnetic field and construction details of a specific probe-head, but assuming them constant, it depends on the gyromagnetic ratio of the excited (γ_e) and observed nuclei (γ_o), according to $\gamma_e\gamma_o^{3/2}$. This explains the great success of inverse detection of heteronuclei through NMR experiments in which protons are used as a starting polarization source, heteronuclei are frequency labelled in the indirect dimensions, and then coherence is transferred back to protons and detected.[3,4] For this reason, throughout the years, probes were generally optimized for ^{1}H detection at the expense of ^{13}C sensitivity. However, thanks to initial successful case studies,[5–10] the design of room temperature probes optimized with the internal, more sensitive coil dedicated to ^{13}C,[11,12] and the leap in sensitivity deriving from the introduction of cryogenic technology in the hardware design,[13,14] ^{13}C sensitivity has experienced an increase of more than an order of magnitude in the past 10 years, making experiments possible nowadays that we could only think of until a few years ago. This instrumental improvement in the intrinsic sensitivity partly compensates for the unavoidable loss experienced when shifting from ^{1}H to ^{13}C detection, allowing a large number of applications. The variety of successful examples in which ^{13}C direct detection does contribute to recover information that would be lost using ^{1}H detection is described at the end of this chapter.

The gyromagnetic ratio, related by definition to the magnitude of the magnetic moment associated to the nuclear spin, and for this reason responsible for the intrinsic sensitivity of an NMR experiment, is also responsible for the magnitude of the dipole–dipole interactions between nuclear spins. These are averaged out by fast molecular tumbling in

solution and thus do not contribute to the signal frequencies, unless a partial orientation of the molecule is induced, but they do provide the major contributions to nuclear relaxation. The dipolar contributions to relaxation depend, among other factors, on the product of the square of the gyromagnetic ratio of the two spins involved in the dipole–dipole interaction as well as on the effective correlation time, through the values of the spectral density function at the relevant frequencies. It is well known that one of the major bottlenecks for NMR applications arises when contributions to transverse relaxation shorten coherence lifetimes, broadening lines beyond detection. In this frame, protons, through their high gyromagnetic ratio, are one of the major sources of dipolar contributions to relaxation. Therefore, focusing on heteronuclei constitutes one of the approaches to reduce contributions to transverse relaxation and thus provides valuable alternatives when ^{1}H spins are not detectable any longer or when the dense network of ^{1}H spins does constitute the limiting factor for a specific application. Two well-known examples that are described in detail in the last part of this chapter are the presence of a paramagnetic centre and the increase in molecular mass. In the first case, the presence of one or more unpaired electrons, through the much higher magnetic moment in respect to nuclear ones, causes a wealth of additional contributions to nuclear relaxation that often broaden lines beyond detection.[15,16] As dipolar contributions to relaxation deriving from a paramagnetic centre depend on the square of the gyromagnetic ratio of the observed nuclear spin, the shift from ^{1}H detection to ^{13}C detection, with other parameters remaining the same, causes a reduction of a factor of $(\gamma_H/\gamma_C)^2$, which may make the difference between being able or not to detect signals.[10,16] The other major source of transverse relaxation is caused by the increase in molecular mass that causes a growth of the rotational correlation time τ_r, and thus of all the contributions to transverse relaxation, with the dipolar ones involving protons playing a major role. Therefore, one of the strategies to extend the molecular mass limit of proteins that can be studied through NMR is the isotopic enrichment with ^{2}H, characterized by a smaller gyromagnetic ratio in respect to protons, in order to reduce the magnitude of the dipolar contributions to transverse relaxation deriving from the dense ^{1}H network of dipole–dipole interactions.[17] When all ^{1}H nuclear spins are substituted by ^{2}H spins, the most favourable spin to detect is thus ^{13}C. This shows how the high gyromagnetic ratio of a nucleus is not necessarily an advantage and that it may actually constitute a limitation.

Other possible sources of fast transverse relaxation potentially able to broaden lines beyond detectable limits are the contributions originating from exchange processes. Indeed, it is well known that local fluctuations between different conformations may provide fluctuating local magnetic fields for efficient transverse relaxation. Of course, the conformational changes experienced will differently affect the chemical shifts depending on the kind of nuclear spins considered. This means that not all kinds of spins are equally affected by conformational exchange processes and therefore the possibility to look not only at protons but also directly at the heteronuclei provides an additional tool to detect resonances, also in cases of conformational exchange processes.[18] The most useful cases in which heteronuclei may provide alternative solutions are perhaps those in which exchange processes with the solvent are responsible for ^{1}H broadening.[19,20] This often occurs either for solvent exposed loops or for specific conditions of pH that are of physiological relevance.[21] Obviously, 2D ^{1}H–^{15}N correlation experiments, as well as all the amide proton-detected 3D and multidimensional NMR experiments, which play a central role in biomolecular NMR applications, suffer from exchange broadening of amide protons, whereas non-exchangeable nuclei do not. So, it may be useful to develop experiments that exploit non-exchangeable nuclei, such as for example carbonyls and C^{α}s.

Another nice property of heteronuclei consists in the large chemical shift dispersion that has contributed to one of the major steps forward in biomolecular NMR applications. Indeed, the possibility to sample not only proton sites but also to directly detect heteronuclei, provides a tool to monitor the vast majority of nuclei in a protein. In addition the heteronuclei, thanks to their improved chemical shift dispersion compared to protons, contribute to increase the resolution. For example, the 10 ppm ^{1}H dispersion of a typical protein spectrum becomes 200 ppm for ^{13}C spins. Considering instead specific types of spins in proteins, H^{α}s resonate in a 2.5 ppm region and C^{α}s in a 20 ppm region, showing the improved dispersion. Finally, different types of ^{13}C spins (carbonyls, aliphatic, aromatic, etc.) can be manipulated independently through band-selective pulses and this feature has stimulated the design of a variety of experiments.

Therefore, the possibility to perform experiments based on ^{13}C direct detection in principle provides a tool to contribute to enhance the chemical shift dispersion of the observed correlations. In general, this constitutes an added value but it becomes particularly relevant in all the cases in which the major limitation to NMR applications derives from resonance overlap

rather than from fast transverse relaxation, as in the case of systems that lack a stable 3D structure, such as unfolded proteins or IDPs.

While the large dispersion of isotropic chemical shifts does provide unquestionable advantages, the anisotropic part of the chemical shift interaction should also be considered, as it can provide sizeable contributions to nuclear relaxation. While for aliphatic ^{13}C spins the chemical shift anisotropy is very small and does not have a large impact on nuclear relaxation, it may become sizeable for aromatic carbons and carbonyls, having a significant impact on transverse nuclear relaxation, in particular when molecular mass increases. This aspect should thus be carefully taken into account in the experimental design for specific kinds of applications.

Isotopic enrichment of proteins with ^{13}C and ^{15}N has stimulated the development of a large array of experiments, thanks also to the large number of scalar coupling constants that can be exploited for coherence transfer. Among them are the large one-bond ^{13}C–^{13}C scalar couplings (a specific property of ^{13}C spins with respect to ^{1}H spins) that mediate very efficient coherence transfers requiring only a short time, thus minimizing relaxation losses, as well as the ^{13}C–^{15}N scalar couplings that can be used in a very selective and complementary way to correlate different spins and to achieve information for sequential assignment. The ^{3}J ^{13}C–^{13}C scalar couplings (and possibly the ^{3}J ^{15}N–^{15}N), although much smaller than the ^{3}J ^{1}H–^{1}H analogues, can also be used to achieve coherence transfers when designing NMR experiments or to access information on dihedral angles. The ^{1}H–^{13}C and ^{1}H–^{15}N couplings can be exploited not only to achieve coherence transfers and/or determine residual dipolar couplings but also to exploit longitudinal relaxation enhancement (LRE) as described later.

Finally, it is worth mentioning more technical features of ^{13}C direct detection that may prove useful. Indeed, for protein NMR the most widely used solvent is H_2O, which contributes to a very intense signal (in general, about five orders of magnitude larger than the solute's signals). Exchange of the water with D_2O enables to avoid the problem of solvent suppression at the expense of all exchangeable protons, which are lost in the exchange. Many clever solutions have been proposed to suppress the solvent signal,[22] but in some cases, it may still be useful to acquire spectra of proteins in water without the need for suppressing the strong solvent signal. The other advantage of ^{13}C detection in respect to proton detection is when dealing with high ionic strength samples. In fact the influence of high salty samples is much less pronounced when dealing with ^{13}C direct detection, again

because of the lower frequency of detection.[13,23] Thus ^{13}C direct detection can be exploited in the study of very salty samples.[13,23]

2.1. The homonuclear coupling

The presence of homonuclear couplings, which complicate the spectra based on ^{13}C direct detection, is readily evident from the example shown in Fig. 6.2, reporting the 2D $^{13}C'$–^{15}N correlation spectrum (CON) acquired on human securin, an IDP of 202 amino acids.[24] As shown in the figure, the nice chemical shift dispersion of the correlations observed would not be very useful without performing homonuclear decoupling in the direct acquisition dimension. This problem is less dramatic when dealing with spectra based on ^{1}H detection since, in this case, homonuclear couplings are roughly an order of magnitude smaller in respect to the case of ^{13}C and often of comparable magnitude to the linewidths. On the other hand, if we manage to suppress the large one-bond ^{13}C–^{13}C couplings, the remaining homonuclear ones involving ^{13}C are smaller than in the case of ^{1}H, providing little contributions to the observed signals' linewidth.

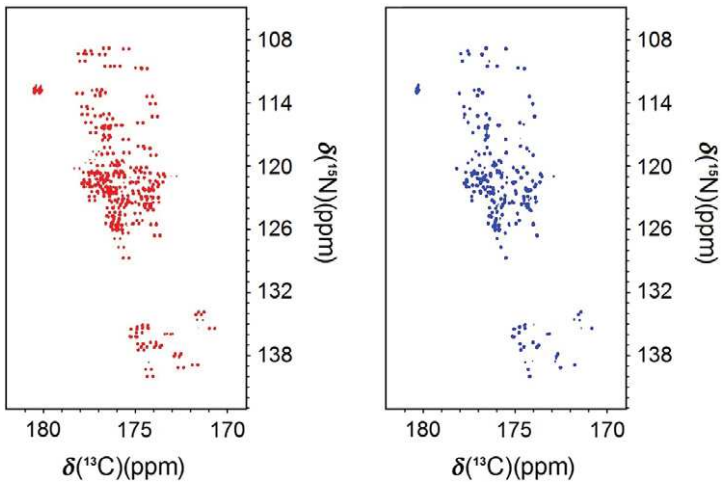


Figure 6.2 The problem of the ^{13}C–^{13}C homonuclear couplings in the direct acquisition dimension. The 2D CON spectra of human securin are reported here, without (left) and with (right) ^{13}C–^{13}C homonuclear decoupling. The huge improvement in resolution provided by the application of homonuclear decoupling is clearly visible. Spectra were acquired at 700 MHz, 283 K on a ^{13}C, ^{15}N-labelled sample of 0.7 mM human securin (25 mM Na_2HPO_4, 150 mM KCl and 10 mM 2-mercaptoethanol, pH 7.2). The IPAP approach was used for homonuclear decoupling of carbonyls to suppress the one-bond C'–C^{α} coupling in the direct acquisition dimension.

Heteronuclear decoupling is easier to perform both in the direct as well as in the indirect dimensions, simply by applying one of the many decoupling sequences or using 180° pulses in indirect dimensions. For the homonuclear case, when performing decoupling in indirect dimensions of NMR experiments, it is quite straightforward to decouple C' from C^{α} and vice versa by using band-selective pulses.[1,2] Alternatively, if transverse relaxation allows, constant time versions of the experiments can be used to refocus the evolution of homonuclear couplings.

The additional complications related to performing homonuclear decoupling in the direct acquisition dimension derive from the fact that we would need to give radio frequency pulses at a specific frequency very similar to the one that we would like to sample, while keeping the receiver open for acquisition. Therefore, it is not trivial to filter one from the other. A solution is conceived in sharing the acquisition time between the "acquisition-mode" and the "decoupling-mode". This switch should of course be very quick, faster than the sampling speed that we want to achieve, in order to allow some time for decoupling between each acquired data point. A similar approach has been proposed also for ^{1}H detection in order to identify coupling partners by irradiating at the frequency of a specific signal and monitoring which multiplets are then perturbed.[25]

In case of ^{13}C detection, if the frequency ranges of different types of spins are well separated and far from each other, such as C' and C^{α} spins, it is easy to perform band-selective irradiation of C^{α} while acquiring carbonyls (and vice versa), removing in this way the large ^{13}C–^{13}C coupling affecting carbonyls (for all carbonyls at once).[10,26] However, one of the drawbacks of this approach consists in the appearance of decoupling side bands, and not for all NMR signals the expected gain in intensity is achieved. This approach works fairly well to decouple carbonyls from C^{α}s and vice versa, whereas it is more difficult to implement complete decoupling of C^{α}s from C^{β}s, which also share a large one-bond coupling, due to the very close frequency ranges between these two kinds of nuclear spins.[26] Moreover, from a more technical point of view, this set-up may require the use of special filters.

Another large class of methods that has been proposed relies on maintaining the coupling active and on letting it selectively evolve in alternate experiments, in order to sample the in-phase (IP) and the anti-phase (AP) components of the signal or their different linear combinations.[12,27–30] These approaches are very elegant and do not require an additional radio frequency channel, since the effect can be achieved by using band-selective pulses.

This idea is very simple and is illustrated schematically in Fig. 6.3 (top panels). Once the carbonyls are transverse, it is possible to refocus or to allow

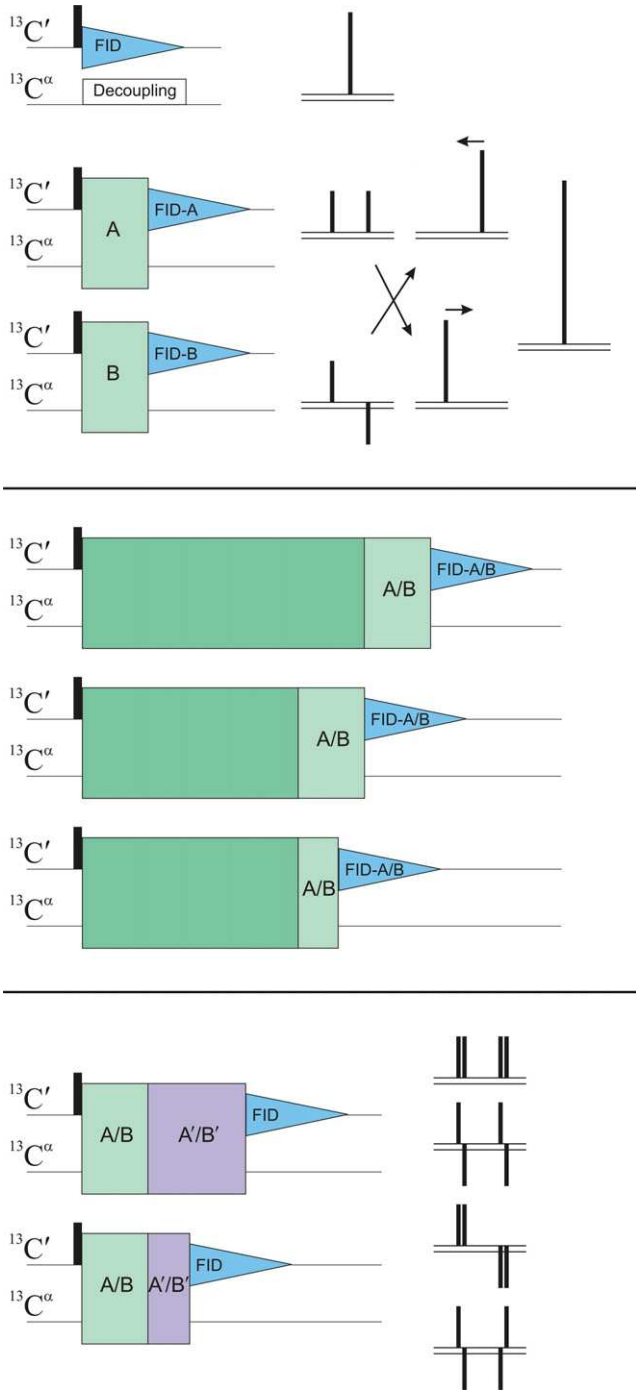


Figure 6.3 Schematic representation of the methods for homonuclear decoupling (top), of how they can be implemented in different kinds of experiments (middle)

the evolution of the C'–C^{α} couplings for a selected overall time ($\Delta = 1/2J$) to obtain either the IP or AP components of the signal, of course using the same number of pulses and delays to make sure that potential pulse imperfections or relaxation losses affect both components in the same way. The figure also shows the linear combinations that need to be performed to separate the two multiplet components, which are then shifted to the centre of the original multiplet and summed together, removing in this way the splitting due to the one-bond coupling. The implementation of this approach (IPAP)[31] to the simplest case of carbonyl decoupling in the 1D mode is shown in Fig. 6.4A.[12,29] An alternative approach (S^3E) is based on the reduction of the overall time in which the coupling is allowed to evolve to $\Delta/2 = 1/4J$, so that both components of the signal (IP and AP) are simultaneously present at the time of acquisition (Fig. 6.4B).[30,32] In this case, separation of the two multiplet components is achieved by acquiring two different linear combinations of IP and AP components by changing the relative sign of the two. For example, a 180° band-selective ^{13}C pulse right before or right after the J evolution period can be used to selectively change the sign of only one of the two components in alternate experiments.[32] Indeed, while at the beginning of the block only IP magnetization is present and so the sign change affects both components deriving from the initial IP signal, at the end of the block the two components result orthogonal one to the other, and so the 180° pulse changes only the sign of one of them while preserving the one of the other. Analogously, the two independent

and how to extend the approach to more than one large coupling (bottom). Top: the three panels indicate schematically the various options to achieve homonuclear decoupling which include band-selective homonuclear decoupling as well as the design of different variants to acquire the in-phase (IP) and anti phase (AP) components of the signals, in order to have the necessary data for subsequent linear combinations to suppress the coupling in the direct acquisition dimension. Middle: the three panels schematically indicate how the different spin-state-selective variants can be implemented in different kinds of experiments, either by adding an extra block at the end of the experiment or by introducing specific variants in the last coherence transfer step or finally, even shortening the pulse sequence by implementing the S^3E variant when the last coherence transfer step consists in refocusing of the coupling that we need to suppress. Bottom: the two panels schematically indicate how one can design experimental variants to suppress more than one coupling, such as adding one block after the other or merging them to reduce the overall duration of the block and thus the relative relaxation losses.

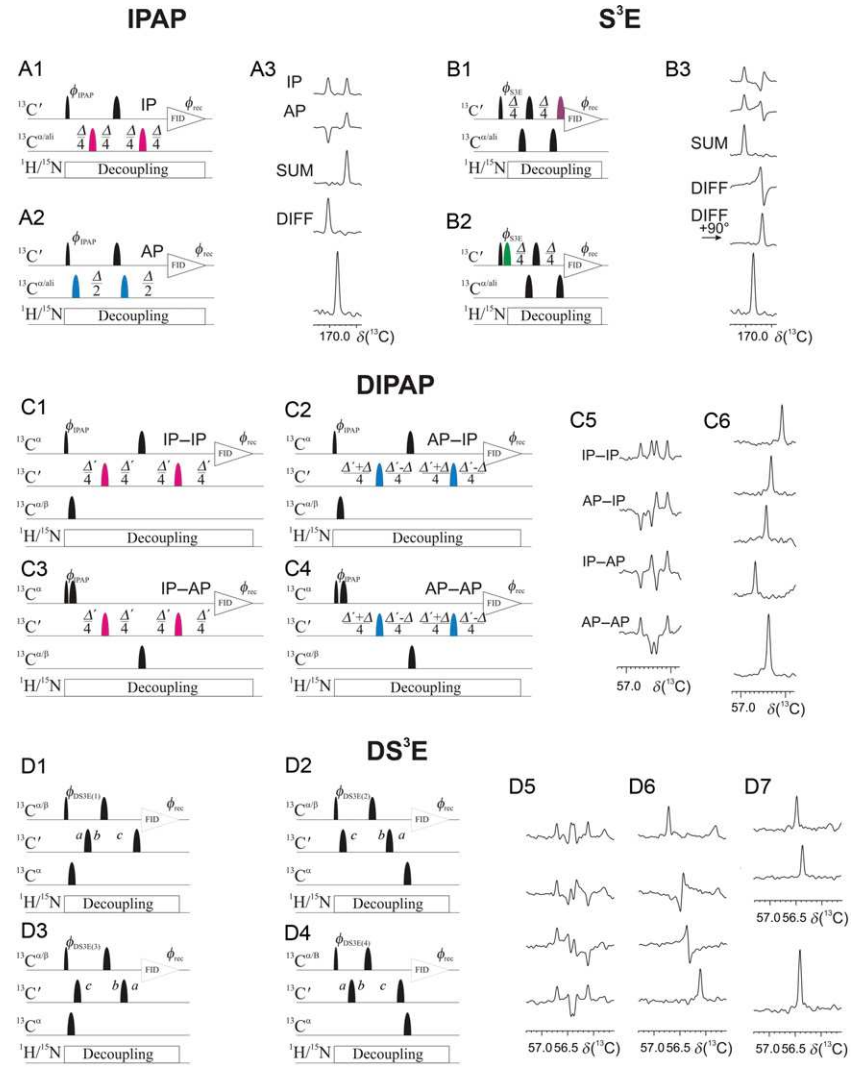


Figure 6.4 Implementation of C′ and C^{α} virtual decoupling in the 1D mode. The IPAP, S^3E, DIPAP and DS^3E building blocks for virtual decoupling are reported here. The first two can be used in C′ direct-detected experiments, while the second two in C^{α} direct-detected experiments. Band-selective ^{13}C pulses are denoted by shapes (narrow and wide ones represent 90° and 180° pulses, respectively). The pulses are applied along the *x*-axis unless noted differently. The lengths of the delays and the phase cycles are reported case by case. For each different pulse scheme, the resulting 1D spectrum is shown as an example. All the spectra have been acquired on a ^{13}C, ^{15}N-labelled cyclic octapeptide hymenistatin. The IPAP building block is reported in panel (A). Panels (A1) and (A2) show the two variants of IPAP pulse scheme to acquire the in-phase and the anti-phase components, respectively. The results of both experiments are reported in panel (A3) and indicated with IP and AP, respectively. Panel (A3) also shows the

components are then summed and subtracted in order to separate the two multiplet components. After applying a 90° phase correction to one of the two, they are again shifted to the centre of the original multiplet and summed to obtain virtual decoupling (Fig. 6.4B). Slightly different variants of this approach have also been proposed.[33,34]

These methods perform very well for C′ homonuclear decoupling due to the fact that the couplings are very uniform throughout the sequence, so the block can be optimized at once for all the signals. Good separation between the two components can be achieved with an improvement of the signal of a factor of two, as expected. However, the gain in S/N is also affected by the unwanted increase in the noise, which arises from the linear combination of the different FIDs. Taking into consideration all these effects, the final gain is

mathematical approach employed to treat the data. The delay Δ is $1/(2J_{C\alpha C'})=9$ ms, where $2J_{C\alpha C'}$ is the coupling to suppress. The phase cycles are: (A1) $\phi_{IPAP}=x, -x$ and $\phi_{rec}=x, -x$; (A2) $\phi_{IPAP}=-y, y$ and $\phi_{rec}=x, -x$. The S^3E building block is reported in panel (B). Panels (B1) and (B2) show the two variants of S^3E pulse scheme to acquire the two independent FIDs needed to separate the two C′ multiplet components (with respect to C^α). The results of both experiments as well as schematic representation of the manipulations necessary to separate the two multiplet components and shift them to the centre of the original multiplet to obtain virtual decoupling are reported in panel (B3). The delay Δ is $1/(2J_{C\alpha C'})=9$ ms, where $2J_{C\alpha C'}$ is the coupling to suppress. The phase cycles are: (B1) $\phi_{S^3E}=x, -x$ and $\phi_{rec}=x, -x$; (B2) $\phi_{S^3E}=y, -y$ and $\phi_{rec}=x, -x$. The DIPAP method for C^α direct detection to remove the two large C^α–C′ and C^α–C^β splittings in the direct acquisition dimension is shown in panel (C). Panels (C1)–(C4) report the four variants of the pulse scheme to acquire and store separately the four components indicated with IP–IP, AP–IP, IP–AP and AP–AP, respectively. The results of the four experiments are shown in panel (C5). The four different linear combinations as well as the final results are shown in panel (C6). The latter is scaled relative to the other traces in order to have the same noise intensity. The delay Δ is $1/(2J_{C\alpha C'})=9$ ms, while Δ' is set to $1/(2J_{C\alpha C'})=14$ ms. The phase cycles are: ϕ_{DIPAP} (C1)$=x, -x$; ϕ_{DIPAP} (C2)$=y, -y$; ϕ_{DIPAP} (C3)$=-y, y$; ϕ_{DIPAP} (C4)$=x, -x$; $\phi_{rec}=x, -x$. The DS^3E building block for C^α direct detection to remove the two large C^α–C′ and C^α–C^β splittings is reported in panel (D). The four variants of the 1D experiment that are necessary to separate, through linear combinations, the four components of the C^α multiplet are reported in panels (D1)–(D4). The four different 1D experiments are reported in panel (D5); the four different linear combinations are reported in panel (D6); the two dispersive components, after applying a first order 90° phase correction, are shown in panel (D7) together with the final result. The delays a, b and c are set to allow for the evolution of the one-bond C^α–C′ and C^α–C^β couplings for a total evolution time of $1/4J$ each ($a+b+c=1/4J_{C\alpha C\beta}=7.1$ ms and $a-b+c=1/4J_{C\alpha C'}=4.5$ ms, with $a+b=c$). The phase cycles are: ϕ_{DS^3E} (1)$=x, -x$; ϕ_{DS^3E} (2)$=-y, y$; ϕ_{DS^3E} (3)$=y, -y$; ϕ_{DS^3E} (4)$=x, -x$; $\phi_{rec}=x, -x$.

thus reduced to the square root of two, ignoring losses due to relaxation and/or pulse imperfections.[35]

For simplicity, the principles on which these approaches are based have been described referring to transverse carbonyl coherence. Therefore, this block can be added at the end of any pulse sequence in which carbonyls are acquired (Fig. 6.3, middle panels). The same results can be obtained also starting from an AP coherence, just inverting the two blocks (the one that prevents the evolution of the coupling yields the AP, whereas the other the IP component). The IPAP approach can thus be implemented without the inclusion of additional delays in all the experiments in which the last coherence transfer step is used to refocus a carbonyl coherence. Band-selective pulses on C^{α} spins should be introduced to modulate the extent of the evolution of the scalar coupling to obtain the two independent components. Indeed, the IPAP approach is often used in 2D CON experiments and more generally in all the experiments that exploit the N to C′ transfer step in the last part, since this final transfer step is equal to $1/2J_{NC'}$, which is longer than $1/2J_{C\alpha C'}$. If instead the experiment ends with a C^{α} to C′ transfer step, in substitution to the IPAP approach, it is possible to implement the S^3E scheme, which provides a reduction of the duration of the pulse sequence (Fig. 6.4B).[32,35] This strategy may be particularly useful when transverse relaxation is an issue.[32]

Another approach that has been proposed for 2D CACO experiments, in which C^{α}_i–C'_i correlations are retrieved, is referred to as "COCAINE".[36] In this experiment, the orthogonal component of the signal that comes from the chemical shift evolution period is recovered through a very simple approach and used to perform homonuclear decoupling.

Finally, all these spin-state-selective approaches can also be used for heteronuclear decoupling. For example, the one described above has been implemented in the 2D CON experiment for ^{15}N heteronuclear decoupling[37] (in a way similar to what was proposed also for ^{1}H–^{15}N decoupling in 2D HSQC experiments[38]), allowing to acquire the carbonyl FID as long as necessary without being limited by the maximal duration imposed by ^{15}N decoupling.

The extension of these ideas to the situation of a spin sharing two scalar couplings is conceptually straightforward (Fig. 6.3, bottom panels). In principle, two blocks, each one optimized for the specific coupling to suppress, can be designed and implemented one after the other as initially proposed in solid-state applications.[27,28] They can also be merged reducing the total duration of the block to that of the longer one of the two.[12,32] Both

strategies, named DIPAP[12] and DS^{3}E,[32] can be implemented, as shown in the case of C^α direct detection. C^α spins are indeed characterized by two coupling partners, carbonyl and C^β (except for glycines). A technical complication derives from the fact that, due to the proximity of C^α and C^β chemical shifts, it is very difficult to selectively invert C^βs with respect to C^αs. A possible solution to this problem consists in the employment of ^{13}C 180° band-selective pulses only on C^αs or, alternatively, on both C^αs and C^βs, providing a tool to achieve the discrimination needed. The variants developed (DIPAP and DS^{3}E, where D stands for "double", referring to two couplings involved) are shown in Fig. 6.4(C and D). As it can be observed, the separation of the four different lines in the multiplet is very efficient as well as the final result of simplification of the original multiplet, with the removal of the two large couplings. Different manipulations of the acquired data also yield properly decoupled C^α resonances for glycines, which lack the C^β, without the need to acquire a second set of data. In terms of sensitivity, also in this case, the manipulations of the different FIDs cause a reduction of the expected increase in S/N ratio, since not only signals but also noise is summed. The two approaches (DIPAP and DS^{3}E) can be easily implemented in experiments ending with IP C^α magnetization, as for example in ^{13}C–^{13}C NOESY experiments or in experiments ending with a C′–C^α or N–C^α block.[35]

It is worth noting that the approaches for "virtual" homonuclear ^{13}C decoupling described above, initially proposed for solid-state NMR spectroscopy of biological systems,[27,28] where ^{13}C has been, at least until recently, the nucleus of choice for direct detection, have then been extended and further developed for solution NMR applications,[12,35] and then introduced back in solid-state NMR.[39,40] This provides an interesting example of how effective cross-fertilization between these two fields has stimulated a faster progress, showing how common and also complementary aspects between solution and solid-state NMR provide key tools to extend the range of systems that can be investigated through NMR.

One interesting property of the "virtual" decoupling methods that emerged while implementing these techniques in solid-state NMR experiments, where many of the commonly used experiments are based on the acquisition of both C′ and C^α, is that the coupling interaction is preserved, allowing thus to mutually decouple the two spins involved in the coupling, without losing any of the information on the two. This is a subtle but interesting difference in respect to the methods in which radio frequency irradiation decoupling schemes are applied to one of the two spins involved

in the coupling (and thus the information on one of the two spins is lost). This property has been exploited in the solid state[39,41] and might also prove useful for solution state NMR.

Finally, in several cases, it may not be trivial to implement homonuclear decoupling, or it may even be preferable to avoid it. For example, when acquiring all aliphatic signals that are characterized by different coupling topologies, a unified approach is not possible and thus couplings are retained.[42] Dedicated versions for specific applications (methyl groups, aromatic rings) can be designed.[32,43] In other cases in which fast transverse relaxation might be a limiting factor (e.g. very large or paramagnetic proteins), it has been proposed to acquire only the AP component of the signal in order to shorten, and thus simplify, the pulse sequences.[7,44]

2.2. The suite of NMR experiments

Once direct ^{13}C detection on uniformly ^{13}C, ^{15}N-labelled proteins can be performed with ^{13}C homonuclear decoupling, a wide variety of different experiments can be developed. In addition, dedicated isotopic enrichment schemes can be designed to improve the performance of key experiments that may prove useful in particular situations. In analogy to the suite of ^{1}H-detected experiments, ^{13}C-detected experiments can be differentiated on the basis of the active interactions exploited in the coherence transfer steps (scalar couplings, cross-relaxation rates), in the starting polarization source (^{13}C or ^{1}H), in the kind of detected nuclei (C', C^{α}, C^{ali} or others) and in the kind of homonuclear decoupling method used (IPAP, S^3E, DIPAP, DS^3E).

After pioneering papers proposing ^{13}C direct detection to study proteins by the group of Markley stimulated by the advent of uniform isotopic labelling of proteins with ^{13}C and ^{15}N,[45–47] inverse detection of heteronuclei[3,4] promoted the development of the suite of ^{1}H-detected triple resonance experiments that is being used for many years to study proteins through NMR.[1,2] However, recent technical improvements in instrumental sensitivity, as well as limits encountered in the application of ^{1}H-detected experiments on various kinds of proteins (paramagnetic proteins, very large multimeric assemblies, high salt conditions, etc.), attracted interest again in the direct detection of heteronuclei as a tool to overcome limitations imposed by the properties of protons and proton detection in general. Indeed, fast proton transverse relaxation in the presence of a paramagnetic centre as well as with increasing molecular mass constitutes a severe

limitation and may cause broadening of signals beyond detection. Therefore, exploitation of protons in any of the magnetization transfer steps of an experiment has a detrimental effect. This idea stimulated the design of 2D experiments based on ^{13}C direct detection in which only the heteronuclei were exploited in the magnetization transfer steps to be used as a complementary tool to ^{1}H-detected experiments.[7,8,48,49] The experiments that were initially proposed were based on ^{13}C-start, ^{13}C-detection and completely avoided protons in any of the magnetization transfer steps. For this reason, they were referred to as *protonless* NMR experiments.[50] The initial set of protonless NMR experiments included various kinds of 2D experiments correlating carbon spins, such as the ^{13}C–^{13}C COSY and the ^{13}C–^{13}C TOCSY, as well as the CACO or COCA experiments, correlating C^{α} and C' in different ways. Indeed, the large one-bond ^{13}C–^{13}C scalar couplings provide a very efficient coherence transfer mechanism. In addition, the 2D experiments correlating C^{α} or C' to nitrogen (CAN and CON) were also proposed.[8–10] Furthermore, it was suggested to exploit the ^{13}C–^{13}C NOESY experiment as an additional tool to be used in cases in which fast transverse relaxation constitutes the major limitation.[29,51,52] In fact, when fast transverse relaxation still constitutes a limiting factor, it might still be possible to detect additional correlations in ^{13}C–^{13}C NOESY experiments, in which cross-peaks originate from relaxation processes occurring, while magnetization is longitudinal (longitudinal auto-relaxation and cross-relaxation rates). Despite the ^{13}C–^{13}C cross-relaxation rates are expected to be much smaller than ^{1}H–^{1}H ones, the ^{13}C–^{13}C NOE effects between directly bound carbons are easily detectable and it was soon realized that spin diffusion through directly bound carbons becomes a very efficient process providing in this way 2D spectra similar to either ^{13}C–^{13}C TOCSY spectra in solution or one of the many ^{13}C–^{13}C dipolar correlation experiments in the solid state.[29] In parallel, other experiments based on ^{13}C direct detection that exploit ^{1}H as a starting polarization source were proposed. These proved useful in several cases, such as long out-and-back experiments,[5,23,53–57] affected by losses due to long coherence transfer pathways or for very salty samples that cause dramatic reduction in ^{1}H sensitivity while ^{13}C is less affected,[23] or because there is no need to implement solvent suppression.

Thanks to these examples demonstrating the usefulness of ^{13}C-detected experiments in several situations of practical interest, in the following years a leap in ^{13}C sensitivity was achieved. This was initially obtained through the design of dedicated room temperature probes with optimized ^{13}C sensitivity, followed by the implementation of the "cryo-technology" also to ^{13}C

direct detection.[11,13,14] This opened the way to the implementation of more challenging experiments. It was thus shown that 3D protonless NMR experiments on proteins could be conducted in a few days;[11,30] in parallel, all the methods to achieve homonuclear decoupling were tested and implemented in 2D and 3D experiments. This enabled the design of the complete suite of 2D/3D experiments based on ^{13}C direct detection.[30,37] The correlations that can be detected in these experiments and the relevant scalar couplings actively used are shown schematically in Fig. 6.5. Carbonyl detection was initially selected due to the presence of only one large coupling to suppress the direct acquisition dimension and to the efficiency of homonuclear decoupling methods.[35] Carbonyls can thus be correlated to the neighbouring carbon (C^{α}) through a ^{13}C–^{13}C transfer step. In addition, C^{α} can be further correlated to the C^{β} through a ^{13}C–^{13}C COSY step and to all the aliphatic side chain through a ^{13}C–^{13}C TOCSY step to obtain the 2D/3D CBCACO[12] and CCCO[30] experiments, respectively. By introducing a further transfer step, it is also possible to include an additional dimension in which the attached nitrogen (in respect to the carbonyl) is frequency labelled. Therefore, the 3D experiments CACON, CBCACON and CCCON were designed.[30] These enable the identification of all the spin systems in a protein, including those involving prolines. Several additional experiments were then proposed in which different couplings are exploited to detect the complementary correlations to link amino acid spin systems in a sequence-specific manner. Experiments featuring $^{1}J_{CAN}/^{2}J_{CAN}$ (CANCO)[11,58] coherence transfers, as well as the variant including also the information about the C^{β} chemical shift (CBCANCO),[24] provide the complementary information to that available in the CACON/CBCACON experiments for sequence-specific assignment. In addition, other experiments providing correlations which involve carbonyl carbons or nitrogen atom nuclei of the previous and of the following amino acids were proposed as a further tool to reduce potential ambiguities. These include the COCON[59] experiment, exploiting a ^{13}C–^{13}C transfer step mediated by a $^{3}J_{COCO}$, which gives best results when using the MOCCA scheme for ^{13}C–^{13}C transfer,[60] as well as the (H)NCANCO[37] experiment in which the N–C^{α}–N step is tuned to correlate the ^{15}N of one amino acid with two neighbouring ones.

Another backbone ^{13}C nuclear spin that can be tackled in a general way and used to design a whole series of experiments is C^{α}.[12,61–63] It is even characterized by a larger chemical shift dispersion with respect to C′, so in principle, it can provide even larger resolution than C′. However, as stated

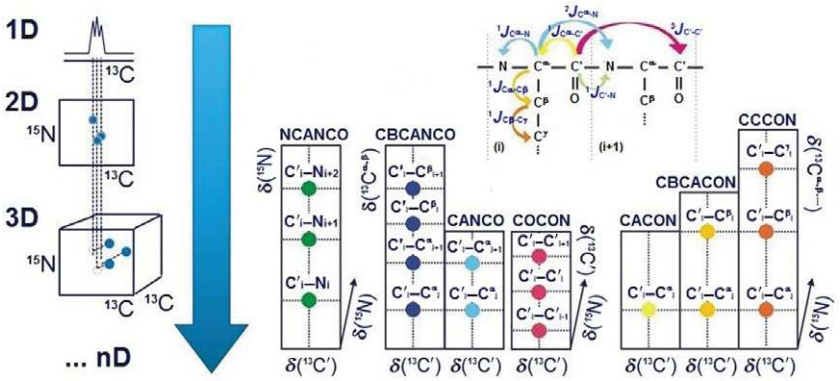


Figure 6.5 The suite of 2/3D $^{13}C'$ direct-detected NMR experiments. The schematic illustration shows a series of experiments which correlate the direct dimension, where $^{13}C'$ is frequency labelled, to the other heteronuclei. As reported on the left, all these experiments can be considered an expansion of the 2D CON experiment, with the possibility to add in the third dimension correlations useful to identify the spin systems or for sequential assignment. That is why only these planes are reported in the picture, whereas the third dimension in which nitrogen evolves has just been represented as an arrow. These experiments can also be conducted in the 2D mode: in this case, their final appearance would be as drawn. All the cross-peaks have been coloured according to the particular scalar coupling/s exploited in each experiment, as shown in the top panel. The resulting spectra are schematically illustrated in the lower part of the illustration. The NCANCO experiment can be used to retrieve sequential connectivities towards both the N- and C-terminal part of the protein, since it correlates simultaneously three neighbouring nitrogens. Similarly, the COCON experiment can be used for the same purpose, since it correlates three adjacent carbonyls. The CANCO and its expansion CBCANCO experiments can instead be exploited to establish sequential correlations along the C^{α} and C^{α}/C^{β} dimensions. Finally, the CACON, CBCACON and CCCON experiments exploit the side chain carbon nuclei in order to unambiguously identify the spin systems. All experiments can be conducted either starting with ^{13}C polarization (^{13}C-start) or, with small modifications, with ^{1}H polarization in the non-selective (^{1}H-start) or selective (^{1}H-flip) modes. The NCANCO, which in principle can be acquired starting from ^{15}N polarization, is generally acquired in either the ^{1}H-start or ^{1}H-flip modes to increase the sensitivity. *Printed with permission by Wiley.*

above, it is also coupled to the C^{β} (except for glycines) through a large one-bond coupling which makes homonuclear decoupling a bit more demanding, even if feasible. Indeed, band-selective homonuclear decoupling is not easy to implement and virtual decoupling approaches, which instead perform very nicely, do require that we pay a price in terms of reduced sensitivity. For these reasons, when possible, carbonyls are usually preferred to

C^{α}s and thus the experiments proposed exploiting C^{α} detection were developed a bit later and in a smaller number. Nevertheless, the CAN experiment was proposed to detect the two correlations between C^{α} and the intra and sequential nitrogen, as well as a variant to highlight the sequential correlation and thus discriminate it from the intra-residue one.[61,62] In addition, the C^{α}-detected versions of the COCA, ^{13}C–^{13}C TOCSY and ^{13}C–^{13}C NOESY experiments, together with different variants of virtual decoupling (DIPAP and DS^3E, Fig. 6.4), are available.[50]

Thanks to the potential of experiments based on C^{α} direct detection for the study of higher molecular mass proteins, it was proposed to exploit the isotopic labelling strategy that avoids neighbouring ^{13}C spins (and thus also the need of homonuclear decoupling in the direct ^{13}C dimension), designed for solid-state applications.[64] Several experiments based on C^{α} direct detection that exploit this isotopic labelling scheme were thus proposed, including the CANCA and the CACA TOCSY experiments.[61,62] Of course, these experiments can also be implemented on fully labelled samples using one of the homonuclear decoupling tricks. More recently, it was realized that the experiments based on C^{α} direct detection might also prove useful for the study of IDPs.[63]

For all the systems in which fast transverse ^{1}H relaxation is not an issue but poor chemical shift dispersion is the limiting factor, such as in disordered or unfolded proteins, it is possible to exploit ^{1}H as a starting polarization source and design a set of experiments that provide the same correlations as the ones described above (Fig. 6.5), but with a higher sensitivity.[37] In order to maximize the dispersion of the correlations, only the heteronuclei are frequency labelled in the detected dimensions. For this reason, this set of experiments based on ^{13}C detection is indicated with the more general term of "exclusively heteronuclear NMR experiments" in respect to the initial one ("*protonless* NMR experiments"). However, analogous versions in which ^{1}H is frequency labelled in the indirect dimensions have also been proposed.[56,57]

2.3. Speeding up ^{13}C direct-detected experiments

Since ^{13}C detection is not limited any longer by the sensitivity, at least in many cases of practical interest, it is possible to implement all the variants that have been recently developed to reduce the duration of NMR experiments or to improve the resolution in a given amount of time (Fig. 6.6).[65] These include longitudinal relaxation enhancement (LRE) strategies to

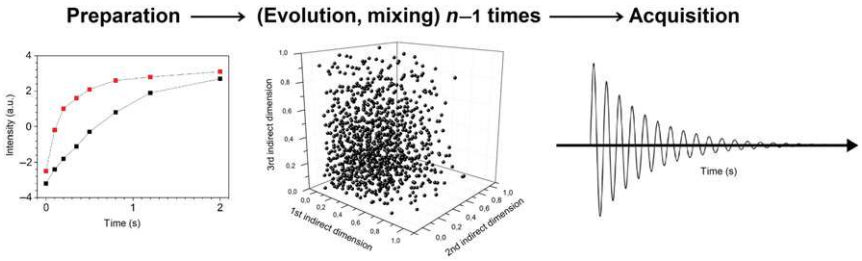


Figure 6.6 Speeding up the acquisition of multidimensional NMR spectra. An *n*-dimensional NMR experiment divided into four basics elements: preparation, evolution, mixing and acquisition periods. The evolution and mixing periods are repeated $n-1$ times, where n is the dimensionality of the experiment. In the illustration, a 4D experiment is shown as example ($n=4$). The experiments can be shortened in two ways: by reducing the inter-scan delay, exploiting longitudinal relaxation enhancement (LRE) in order enhance the recovery of the magnetization (left panel, from black to red curve), and by sampling only a subset of the time points needed to build the indirect dimensions of the experiment, using non-uniform sampling (NUS) schemes.

reduce the time between the acquisition of two subsequent FIDs, as well as non-uniform sampling (NUS) strategies to reduce the number of points necessary to build multidimensional NMR experiments. These approaches are briefly explained hereafter and their implementation in ^{13}C-detected experiments are described.

2.3.1 Longitudinal relaxation enhancement

It is well known that recovery of magnetization to equilibrium does not only depend on the intrinsic relaxation rates (auto- and cross-relaxation rates) but also on the initial conditions. Indeed, it has been shown that in the macromolecular limit (for folded proteins), selective perturbation of a subset of protons, for example, amide protons, while leaving all other spins unperturbed, greatly enhances the recovery of magnetization to the equilibrium position. This effect, which can be demonstrated through the relaxation matrix approach, has been implemented in proton-detected experiments in order to reduce the inter-scan delay and thus shorten the duration of a given experiment. This effect, first introduced for the acquisition of 2D HN experiments,[66–71] was then extended to the whole set of triple resonance experiments,[70] contributing to a drastic reduction of the time necessary to collect the multidimensional experiments necessary for the complete assignment of a protein.

Along the same lines, the idea to selectively manipulate a subset of spins, while leaving all others unperturbed with the aim of speeding up their recovery, can be implemented in an even easier way in ^{1}H-start-^{13}C-detected experiments.[72,73] Indeed, since in this case ^{1}H spins are only used as a starting polarization source, after the initial polarization transfer step there is no need to perturb them again. Since one of the nice features of the suite is that one can access information about all amino acids, including prolines, the first implementation of the approach was to $^{1}H^{\alpha}$ start experiments.[72] Therefore, in place of band-selective pulses proposed to focus on amide protons, it was suggested to exploit the evolution of the ^{1}H–^{13}C scalar coupling in order to only perturb H^{α} spins and return all other spins to the $+z$-axis. If necessary, during chemical shift evolution periods, the evolution of heteronuclear-proton scalar couplings can be refocused through the use of 180° ^{1}H pulses, which can be applied in pairs one close to the other to flip ^{1}H

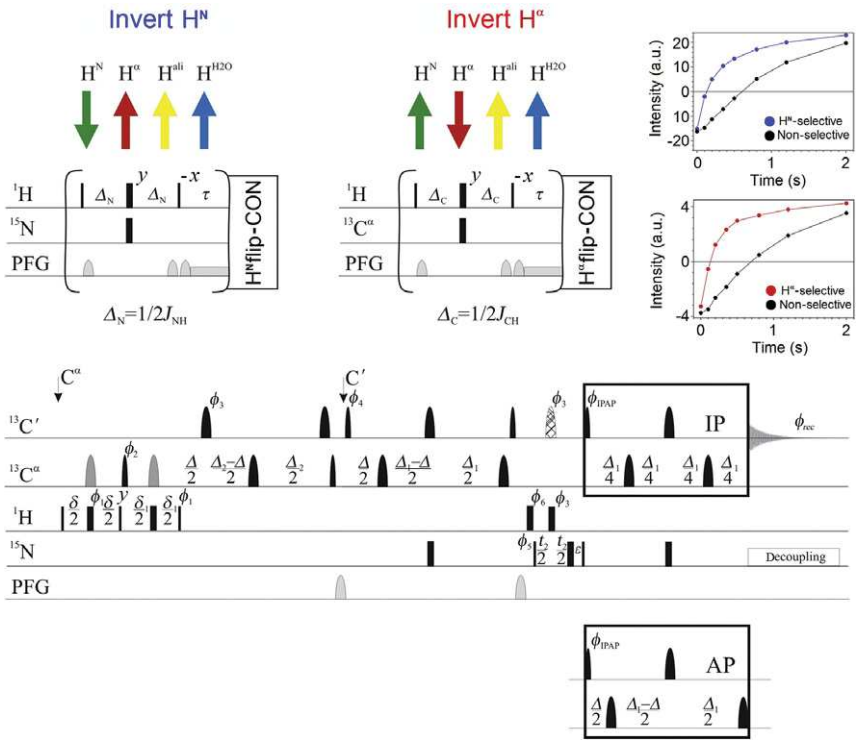


Figure 6.7 The H-flip approach for longitudinal relaxation enhancement in ^{13}C detected exclusively heteronuclear NMR experiments. The blocks needed to acquire the inversion recovery profiles of H^{N} and H^{α} protons in the selective and non-selective (omitting the 180° pulse on the heteronucleus and changing the phase of the last

magnetization back to $+z$-axis. This approach, referred to as $H^{\alpha\text{-flip}}$, is schematically illustrated in Fig. 6.7. The performance of it on a well-known protein sample (ubiquitin) is also shown in the figure by the inversion recovery profiles acquired in the selective and in the non-selective modes. The modifications necessary to implement the $H^{\alpha\text{-flip}}$ in the CON pulse sequence are also shown (Fig. 6.7). The same approach was then extended also to the H^N start versions of the experiments, providing also in this case a significant LRE.[73,74]

The next obvious question is whether similar ideas could also be exploited to contribute to the longitudinal enhancement of ^{13}C magnetization. In this case, the magnitude of homonuclear cross-relaxation rates becomes comparable with the one of longitudinal relaxation rates only when dealing with extensively ^2H-enriched proteins of large molecular mass, as indicated by the inversion recovery profiles acquired either inverting all spins or only a subset of them for different proteins of different

proton pulse by 180°) mode are reported in the top of the illustration. On the right, the H^N and H^α inversion recovery profiles of the residue 56 of human ubiquitin are shown. The benefit that can be gained perturbing selectively amide or aliphatic protons is clearly visible. In the bottom part, the pulse sequence of the ($H^{\alpha\text{-flip}}$)CON experiment is reported. The H-flip version of the experiment can be easily obtained by introducing minor modifications in the first part of the pulse sequence to restore on the *z*-axis the magnetization of proton spins not involved in the first magnetization transfer step, promoting longitudinal relaxation enhancement. All pulses are given with phase *x*, unless otherwise specified. Band-selective ^{13}C pulses are denoted by shapes and rectangular pulses represent hard pulses. The wide and narrow pulses correspond to π and $\pi/2$ flip angles. Pulse field gradients (PFG line) are also indicated by shapes. The delays are: $\varepsilon = t_1(0)$, $\delta = 3.6$ ms, $\delta_1 = 2.2$ ms, $\Delta = 9$ ms, $\Delta_1 = 25$ ms, $\Delta_2 = 27$ ms. The phase cycles are: $\phi_1 = -x$; $\phi_2 = 4(x)$, $4(-x)$; $\phi_3 = 8(x)$, $8(-x)$; $\phi_4 = x$, $(-x)$; $\phi_5 = 2(x)$, $2(-x)$; $\phi_6 = 8(-x)$, $8(x)$; $\phi_{IPAP}(IP) = x$; $\phi_{IPAP}(AP) = -y$; $\phi_{rec} = x$, $(-x)$, $(-x)$, x, $(-x)$, x, x, $(-x)$. Quadrature detection was obtained by incrementing phase ϕ_5 (t_1) in a States-TPPI manner. For ^{13}C band-selective $\pi/2$ and π flip angle pulses Q5 (or time reversed Q5) and Q3 shapes were used with durations of 274 and 220 μs, respectively, except for the π pulse indicated in grey (band-selective inversion of C^α spins respect to C^β, Q3, 860 ms) and for the π pulse indicated in crossed stripes (adiabatic inversion pulse over the C′ and C^α regions, Chirp, 500 μs, 25% smoothing, 80 kHz sweep, 11.3 kHz strength). The ^1H and ^{15}N carriers are placed at 4.7 and 125 ppm, respectively. The change in the position of the ^{13}C carrier (39 ppm for C^{ali}, 55 ppm for C^α and 173 ppm for C′, 100 ppm for the adiabatic pulse) is indicated by vertical arrows. Decoupling of the ^1H and ^{15}N was achieved with waltz-16 (1.7 kHz) and garp-4 (1.0 kHz), respectively. The experiment employs the IPAP approach to suppress the C^α–C′ coupling and the in-phase (IP) and anti-phase (AP) components are acquired and stored separately by using the pulse schemes illustrated, doubling the number of FIDs recorded in one of the indirect dimensions.

sizes. As it can be observed, for ^{13}C recovery the effect between selective and non-selective inversion of all spins is minor for proteins of molecular size as ubiquitin, while it becomes sizeable going to proteins with increasing molecular mass, as indicated in Fig. 6.8 (top panels). Therefore, the effect is of potential interest and could be exploited. For proteins of

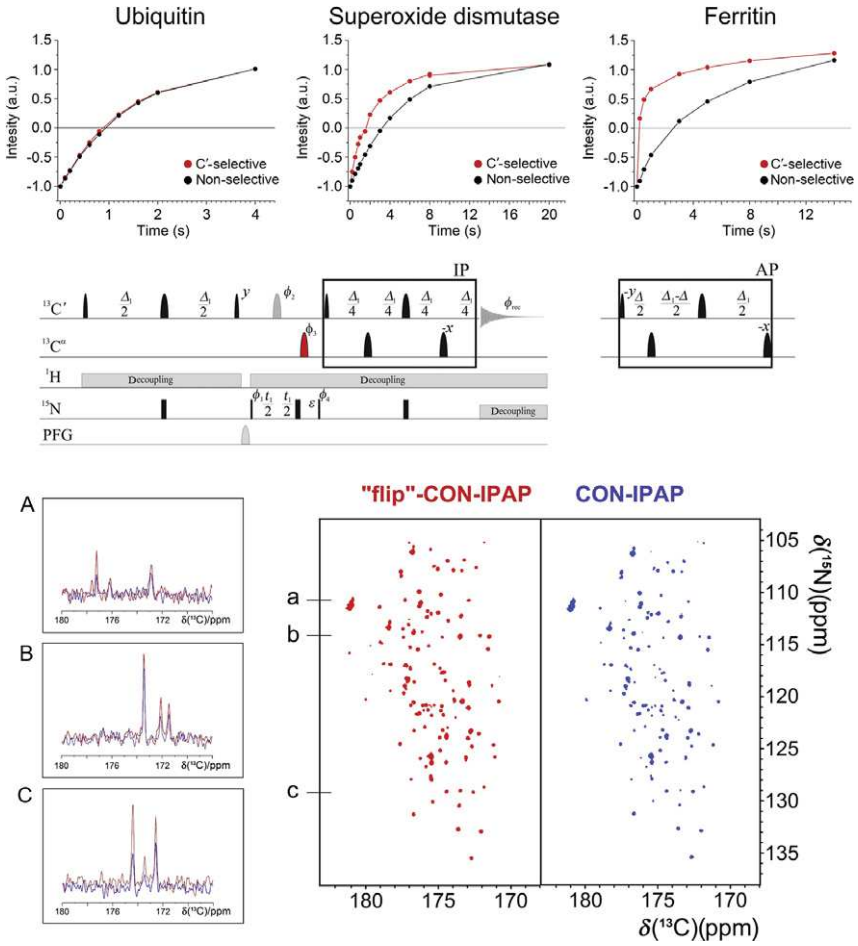


Figure 6.8 How about ^{13}C longitudinal relaxation enhancement? Three inversion recovery profiles obtained perturbing selectively the carbonyl spins or all the carbons spins are reported in the top of the figure for ubiquitin (8.5 kDa), Cu(I)/Zn(II) superoxide dismutase (32 kDa) and ferritin (480 kDa). From their comparison, it is readily visible how the gain obtained perturbing only the carbonyl spins increases together with the molecular mass of the protein. In the middle of the figure, it is reported the flip-CON-IPAP pulse sequence, which allows to exploit this effect in order to reduce the duration of the inter-scan delay. The only difference between this pulse sequence and its analogue

intermediate size, such as superoxide dismutase (32 kDa), it was shown that by taking care of flipping back to $+z$-axis all "unused" ^{13}C spins in CON-type and CAN-type experiments, the overall sensitivity can be improved (Fig. 6.8, bottom panels). Therefore, on proteins of bigger size such effect is expected to be even larger. However, it should also be considered that most of the scalar-type experiments will also suffer from reduced sensitivity.

2.3.2 Non-uniform sampling

A complementary approach to speed up the acquisition time reducing the inter-scan delay between each repetition of a given experiment consists in reducing the overall number of repetitions of the experiment necessary to build indirect dimensions of NMR experiments.

Conventionally, the acquisition of multidimensional NMR experiments is in fact performed sampling, in each indirect dimension, time points equally spaced on the Cartesian grid, where by time point we refer to each repetition of the experiment with different delay/s for chemical shift evolution/s. The grid spacing is dictated by the Nyquist theorem, in the same way the sampling interval is determined in the direct dimension. Such a theorem requires the sampling rate to be faster than the highest frequency expected. In other words, the delay between data points placed along a particular dimension of the Cartesian grid is calculated as the reciprocal of the desired spectral width. Unfortunately, the observation of such rule results in an enormous growth of experimental time as the

CON-IPAP is the presence of a 180° shaped pulse (shown in red) selectively applied on aliphatic ^{13}C spins to flip back ^{13}C spins inverted by the previous pulse. The effect of the addition of this pulse are reported in the lower part of the figure, where the spectra of these two versions of the 2D CON-IPAP experiment acquired on Cu(I)/Zn(II) superoxide dismutase are reported. Fixing the experimental time, the gain in signal intensities is evident, as also shown in the bottom left for three sections of these spectra. All the pulses of the flip-CON-IPAP pulse sequence are given with phase x, unless otherwise specified. Band-selective ^{13}C pulses are denoted by shapes and rectangular pulses represent hard pulses. The wide and narrow pulses correspond to π and $\pi/2$ flip angles. Pulse field gradients (PFG line) are also indicated by shapes. The delays are: $\Delta = 9$ ms, $\Delta_1 = 25$ ms, $\varepsilon = t_1(0) + p_C - p_{Cali}$, where p_C is the duration of the adiabatic 180° pulse (shown in grey) in the middle of the ^{15}N chemical shift evolution time and p_{Cali} is 180° is the duration of the shaped pulse (shown in red) selectively applied on aliphatic ^{13}C spins. The phase cycles are: $\phi_1 = x, -x$; $\phi_2 = 2(x), 2(-x)$; $\phi_3 = 2(-x), 4(x)$; $\phi_4 = 4(x), 4(-x)$; $\phi_{rec} = x, (-x), x, 2(-x), x, -x, -x$. All the other parameters are the same as reported in Fig. 6.7.

dimensionality of the experiment is increased, especially when a consistent number of increments are required to achieve optimal resolution in the indirect dimensions.

In the past two decades, an alternative sampling approach known as NUS has been introduced in NMR to reduce experimental time, or achieve higher resolution in the same amount of time or be able to acquire spectra of high dimensionality with appropriate resolution. NUS eludes the limitation of conventional sampling by sampling only a subset of the time points of the Cartesian grid, according to some predetermined sampling scheme. However, sparsely sampled data cannot be processed with the well-known fast Fourier transform, for which signals would be corrupted due to missing data, but need different strategies for a proper treatment. Clearly, the quality of the resulting spectra is therefore determined by the sampling schedule and spectral reconstruction or processing method adopted, which have thus become growing fields of research. Currently, several non-linear algorithms for data reconstruction or processing have been proposed and optimized, such as maximum entropy,[75,76] forward maximum entropy,[77,78] multidimensional decomposition (MDD),[79] compressed sensing (CS),[80] spectroscopy by integration of frequency and time domain information (SIFT)[81] and multidimensional Fourier transform (MFT).[82] Principally, these methods differ in the assumptions that have to be introduced to reconstruct the missing data. For example, ME methods rely on the optimization of a penalty function in order to find the spectrum with highest entropy, while CS performs an l_1-norm minimization to recover the sparsest spectrum; MDD compensates the lack of sampled points simplifying the description of the spectrum as a direct product of Lorentzian shapes, whereas SIFT replaces the time data omitted with some spectral points assumed to be equal to zero; finally, MFT simply Fourier transforms the NUS data with a multidimensional procedure. Of course, since different assumptions may be more appropriate for different NUS spectra, the overall efficiency of each method generally depends from case to case. Apart from that, other relevant differences between these methods are the computational efficiency and the kind of input and output, as well as the degree of automation.

However, a common drawback of all these methods is the appearance of spectral artefacts in the final spectrum, as an effect of the transformation of the so-called "sampling noise", namely the noise that can be attributed to the irregular data points distribution rather than to sensitivity. The artefacts can manifest in many forms, such as ridges, phase distortions or peaks of various

intensity: in fact, artefact final appearance and distribution largely depend on the type of sampling schedule and algorithm used to reconstruct the data. For most common applications however, the artefact level is generally very low. In particular, this is true for spectra in which peaks have similar amplitudes, since artefact intensities are proportional to the ones of the real peaks, and in such case, there is almost no risk to confuse "true" and "false" signals one with the other. On the contrary, when peak intensities are very different (as e.g. in NOESY-type spectra), smaller peaks can be hidden under the sampling noise. For such cases, ME and MDD are recommended, since these methods are able to efficiently reduce the artificial noise, even if at elevate computational expense. For other methods which do not directly treat the sampling noise, like MFT, very effective post-processing cleaning procedure, like the signal separation algorithm[83,84] and CLEAN,[85] can be used to achieve similar results.

In parallel, many efforts have been put into the optimization of the sampling schemes, since it has been proved that the artefact patterns are strictly related to the distribution of the time points.[86] Several sampling schemes have been evaluated: possibilities range from more ordered distribution of time points, like spiral and radial sampling, to completely disordered distributions, like random sampling. Evidences have shown that the higher is the degree of randomness in the time-point schedules, the lower are the overall artefact levels. In fact, random sampling schemes do not cluster the artefacts in some specific spectral regions, but spread the artefacts over the entire frequency domain. However, it has been demonstrated that a certain degree of "order" can be beneficial too, in term of artefact control. Indeed, the use of explicit restraints for the minimum relative distances between the time points, which therefore leads to more even (but still random) schemes, has been shown to shift the artefacts towards empty spectral regions and to provide relatively clean areas in the proximity of the signals.[86] Among several algorithms which generate sampling schedules imposing distance restraints between the time points, Poisson-disk sampling scheme is one of the most effective.

In conclusion, all the methodologies and improvements described in this paragraph allow thus a fast and accurate acquisition of several typologies of NMR experiments. In particular, NUS and LRE can be effectively exploited also for ^{13}C-detected applications, and combined with all the approaches described before, such as the ones for virtual homodecoupling. As an example, the acquisition of 3D ^{13}C-detected experiments could be significantly reduced as shown in Fig. 6.9.

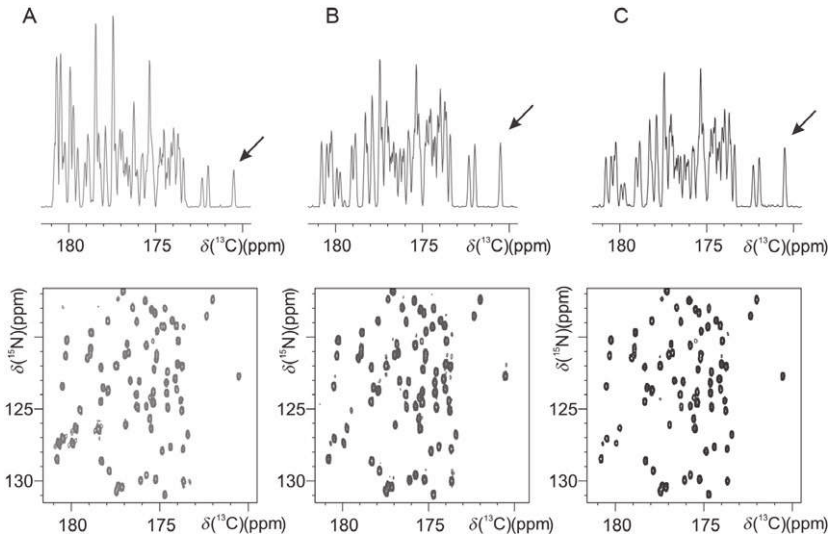


Figure 6.9 Implementation of longitudinal relaxation enhancement (LRE) and non-uniform sampling (NUS) in one of the 3D ^{13}C-detected *exclusively heteronuclear* NMR experiments, the (H)CANCO. The figure shows the C′–N projections of the 3D CANCO spectra acquired with the ^{1}H-start variant (A), with the ^{1}H-flip variant (B) and with the ^{1}H-flip variant also reducing the sampling in indirect dimensions to 40% of the data (C). The overall time necessary to conduct the experiment is significantly reduced (about a factor of 10), still showing all the expected cross-peaks without causing a reduction in the signal-to-noise. This fact is clearly evident looking for example at the peak highlighted with an arrow in the 1D projections. *Adapted with permission by ACS.*

2.4. ^{13}C direct-detected NMR experiments of high dimensionality

Thanks to the methodological improvements described in the previous paragraph, experiments of high dimensionality can be conducted with high resolution in a reasonable amount of time, with the only limitation being the relaxation properties of the molecule under investigation.

Multidimensional experiments which involve four to five chemical shift evolutions are feasible only with the use of LRE and NUS: many improvements have thus been done to facilitate the generation of the sampling schedules and to simplify the processing of the acquired data. The setup of the sampling lists has become very user-friendly, since in most of the programmes only a few steps are generally required to define the type of sampling, the total number of time points and the maximum evolution times. Also, the programmes for data reconstruction have been significantly improved to automatically read the experimental parameters required for

processing, and some of them have already been implemented in the acquisition softwares of the spectrometers to be routinely used.

However, the main barrier which initially limited the reception and diffusion of high multidimensionality experiments was the prejudiced idea that more than three dimensions could be practically very difficult to handle. For a long time, the common feeling has been that the abundance of information encoded in such spectra cannot be totally retrieved, since no proper strategies to visualize and analyse the spectra were available at the moment. Recently, many efforts have thus been made in this direction, with the aim of familiarising the user in his confrontation with these kinds of experiments. Among all the proposals, a new approach known as sparse multidimensional Fourier transform (SMFT)[82] has turned out to be one of the most interesting ones, for its simplicity and effectiveness. In fact, this method greatly simplifies the visualization of the *n*-dimensional space of any nD experiment (like a 4 or 5D) decomposing it as a series of lower dimensionality spaces (like 2 or 3D), which of course are familiar to almost all the NMR users.

The acronym SMFT, which just means the "sparse" or "local" application of the MFT procedure, already explains how the decomposition of the *n*-dimensional space is performed: the features of MFT (mentioned in the previous paragraph) are exploited to focus on spectral regions of multidimensional spectra only at arbitrary chosen frequencies. In other words, instead of processing the full spectrum, only its slices which contain peaks are computed, thus greatly simplifying the picking of the signals. As an example, if we consider a 4D spectrum in which δ_i, δ_j, δ_k and δ_l chemical shift evolutions are labelled, the spectrum is processed as a series of 2D cross-sections (e.g. δ_i and δ_j) obtained fixing, for each of them, a particular pair of the other two frequencies (δ_k and δ_l). The frequencies to be fixed can be selected based on the information retrieved from a spectrum of lower dimensionality previously recorded, which shares these two dimensions with the 4D experiment (e.g. a 2D spectrum containing δ_k and δ_l dimensions). Thus, a 4D spectrum can be thought of as a 2D spectrum (reference spectrum) in which each peak is associated to another 2D spectrum (cross-section). Analogously, any 5D experiment can be processed splitting the full 5D spectrum in a 3D reference spectrum in which each peak is coupled to a further 2D spectrum. This approach is schematically illustrated for all possible cases in Fig. 6.10.

Therefore, SMFT greatly facilitates the analysis of high-dimensional spectra, since it provides a way to collect the resonances without the need

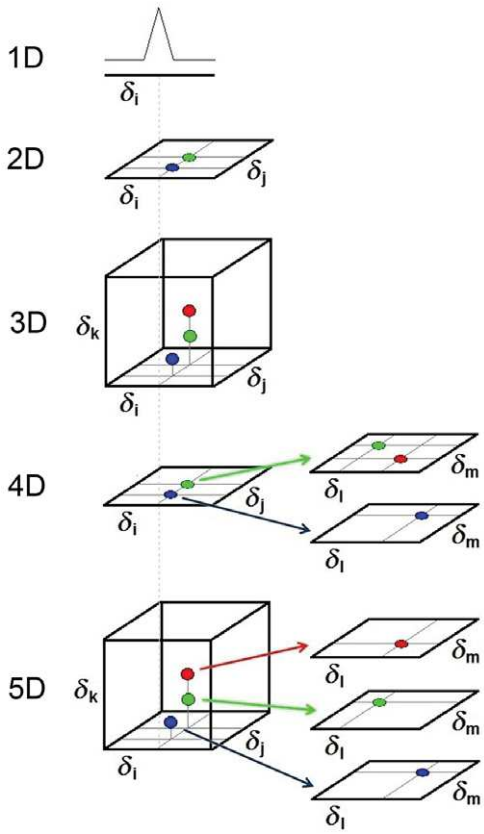


Figure 6.10 A strategy for the visualization of multidimensional NMR spectra. The progressive steps towards experiments of higher dimensionality are schematically shown here from the 1D to the 5D case, together with a simple approach which can be used to easily retrieve the information content of a spectrum of any dimensionality. It is also shown how the resolving power increases along with the addition of further dimensions. The transition from 1D to 2D, and similarly from 2D to 3D, is well known by all NMR spectroscopists, and is reported here only for sake of completeness. The figure shows how such spectra are usually visualized and the benefits gained in term of resolution due to the addition of a further dimension. The following step from 3D to 4D, which could seem more awkward, is instead very easy. A 4D spectrum can be visualized associating each peak of a 2D reference spectrum, which shares two dimensions with the 4D spectrum (δ_i and δ_j), to an another 2D spectrum containing the other two dimensions (δ_l and δ_m). In this way, instead of the full 4D, only the slices containing peaks, numerically equal to the number of detected peaks in the 2D reference spectrum, are computed. Analogously, the same way of proceeding can be used for the final step from 4D to 5D. The only difference in respect to the previous case is in the choice of the reference spectrum, which in this case should be a 3D. As a consequence, the five-dimensional space is split in the sum of a three-dimensional (δ_i, δ_j and δ_k) and a two-dimensional space (δ_l and δ_m).

to explore all the huge *n*-dimensional space of the full spectrum. By the way, this approach provides also other important advantages. For example, it allows the use of automatic peak-picking tools, which can perform their task very reliably and efficiently because the signals are generally well resolved due to the several dimensions employed in such experiments. This typology of processing allows also to save a huge amount of disk space, since the spectra of high dimensionality are not fully processed, but only a limited number of bidimensional sub-spectra are computed: in this way, it is possible to maintain a high digital and spectral resolution in all the detected dimensions without any dramatic increase of file sizes.

Of course, the discussed benefits provided by the exploitation of the heteronuclei and ^{13}C-direct detection can be easily combined together with the use of several dimensions to achieve results as good as possible. Recently, several 4/5D heteronuclear experiments which include all these approaches have been established and efficiently used for the investigation of specific and complex biological systems, for example the sequence-specific assignment of several large-sized IDPs (discussed later). Here, we would like to give to the reader just an overview of the most recent experiments, as an example of the great results that can be obtained when all these approaches are combined together. These experiments are reported in Fig. 6.11.

The suite of the 4D experiments has been built on the basis of the most simple heteronuclear 2D CON experiment, chosen as reference spectrum, in which C'_i–N_{i+1} correlations are present.[87,88] All the 4D experiments can thus be thought of as the association of each cross-peak of the 2D CON spectrum with two extra dimensions, in which two particular frequencies are chemical shift labelled depending on the specificity of each experiment. In this way, once the signals present in the 2D reference spectrum are collected, the analysis of the complete set of 4D spectra simply consists in the inspection of the 2D planes associated to each cross-peak retrieved in the CON spectrum. The 4D experiments reported in Fig. 6.11, tailored both for spin-system identification and sequence-specific assignment, have been recently proposed.[87] The correlations present in each experiment, as well as the magnetization transfer pathways involved, are also reported in Fig. 6.11. Briefly, the 4D HCBCACON and HCCCON experiments allow to identify the type of amino acid to which each CON peak belongs, whereas the 4D HCBCANCO, (H^N-flip)NCACON and (H^N-flip) NCANCO experiments are specifically designed for the sequential assignment of the resonances. An alternative version of the 4D HCBCANCO experiment, in which a ^1J(NCα) selective transfer is used, has also been proposed.[88]

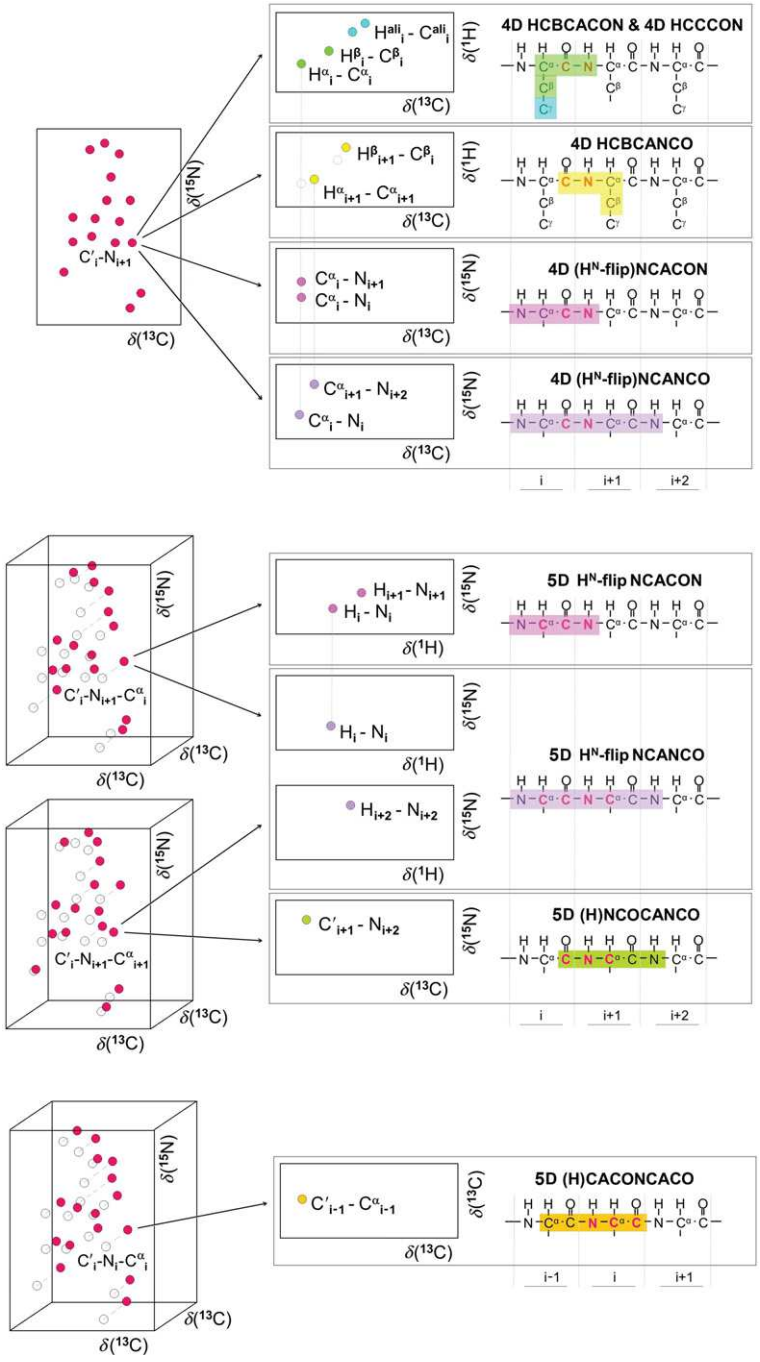


Figure 6.11 An overview of the most recent 4/5D ^{13}C direct-detected NMR experiments. A series of ^{13}C direct-detected experiments of high dimensionality tailored both

Extending the same concepts with the addition of a further dimension, several 5D experiments have been established in order to provide extremely resolved signals, with the aim of making the specific sequential assignment even easier.[87,89] However, contrary to the 4D case previously described, these experiments do not correlate a common set of three spin resonances, and thus a unique 3D spectrum cannot be chosen to process all the 5D spectra exploiting the features of the SMFT algorithm. Therefore, different 3D reference spectra have to be recorded depending on the correlations expected in the 5D experiments. Of course, also in this case, the analysis of each 5D is then reduced to the inspection of the 2D cross-sections associated to each peak of the reference spectrum. The experiments reported in Fig. 6.11 are the 5D H^N-flip NCACON and H^N-flip NCANCO,[87] which add a proton dimension to their 4D analogues, and the 5D (H)NCOCANCO and (H^N-flip)CACONCACO,[89] which instead exploit C′ and N frequencies for the sequential assignment.

2.5. ^{13}C direct-detected amino acid-type-selective NMR experiments

Another possibility to simplify the spectra and reduce signal overlap consists in exploiting the methods proposed in the literature to select specific types of amino acids or amino acid groups in order to drastically reduce the amount of cross-peaks that are detected in the spectra. There are a variety of different

for spin-system identification and sequential assignment is schematically reported here. The 4D experiments are presented in the upper part of the figure, while the 5D experiments are described below. The experiments are shown as processed using the SMFT algorithm, retaining thus a simple intuition for the analysis of the data. For each experiment, its reference spectrum is reported on the left side of the picture, whereas an example of one of the additional 2D planes containing the specific information depending on the design of each particular experiment is shown on the right side. Since all the 4D experiments share the chemical shift evolution of the carbonyls and the attached nitrogen, the 2D CON experiment is chosen as the reference spectrum. The additional information content of the various 4D experiments can be retrieved in the associated 2D cross-sections; the expected correlations are schematically illustrated, as well as the nuclei involved in the coherence transfer pathways. The 5D experiments are reported in the middle and lower part of the panel. In this case, different 3D reference spectra are needed depending on the magnetization transfer pathways exploited. In particular, the 5D (H^N-flip)NCANCO experiment requires two different 3D reference spectra to retrieve all the information.

approaches to select a specific amino acid or a group of amino acids with common spectroscopic characteristics or on the basis of their chemical structure.[90–108] These experiments are based on a variety of selection criteria such as multiple quantum filters to select XH_n groups (with X being ^{13}C or ^{15}N), band-selective pulses on heteronuclei (^{13}C, ^{15}N) to irradiate specific frequency ranges, tuning of delays to pick up specific coherence transfer pathways, and matching of increasing numbers of coherence transfer steps to a particular spin topology.

In principle, all amino acids can be distinguished using different permutations of such selection criteria. Importantly, the overall selection block which generally exploits specific features of amino acid side chains can be implemented in different kinds of pulse sequences. These experiments are particularly well suited to simplify crowded spectra, such as those of IDPs. It is precisely the problematic spectral resolution of IDPs which renders them excellent candidates for such an approach, since amino acid side chain resonances are expected to cluster in very narrow frequency bands, correlating far better with amino acid type than in structured proteins.[109] In addition, the succession of filter and transfer steps, which are the essence of such experiments, greatly benefits from the flexibility of IDPs which mitigates loss of signal due to nuclear relaxation. This suggests the advantage of realizing the ^{13}C direct detection and amino acid-selective NMR (CAS-NMR) experiments for IDPs while utilizing the exclusively heteronuclear approach.[110]

The (H)CBCACON pulse sequence represents the most suitable template to be used as a starting variant to introduce the modifications to select correlations deriving only from specific types of amino acids (Fig. 6.12). The pulse sequence starts with ^{1}H polarization and thus enables to use the editing tricks based on the multiplicity of the XH_n group. The MUSIC variant was implemented following the approach proposed by the group of Oschkinat.[104–105] This enables to easily distinguish between CH_3, CH_2 groups and of course also to select NH_2 side chain groups of asparagine and glutamine amino acids. Another very useful approach to discriminate between the different amino acids is to combine this selection criterion with the number of carbon–carbon (–C–C–) transfer steps in the pulse sequence. For example, the simple use of these two criteria readily suggests very straightforward modifications of the experiment to select correlations deriving from glycine residues (a CH_2 and only C) as well as of alanine residues (a CH_3 and only one C transfer). The other very useful way to discriminate between the different amino acids types consists in the use of band-selective ^{13}C pulses.

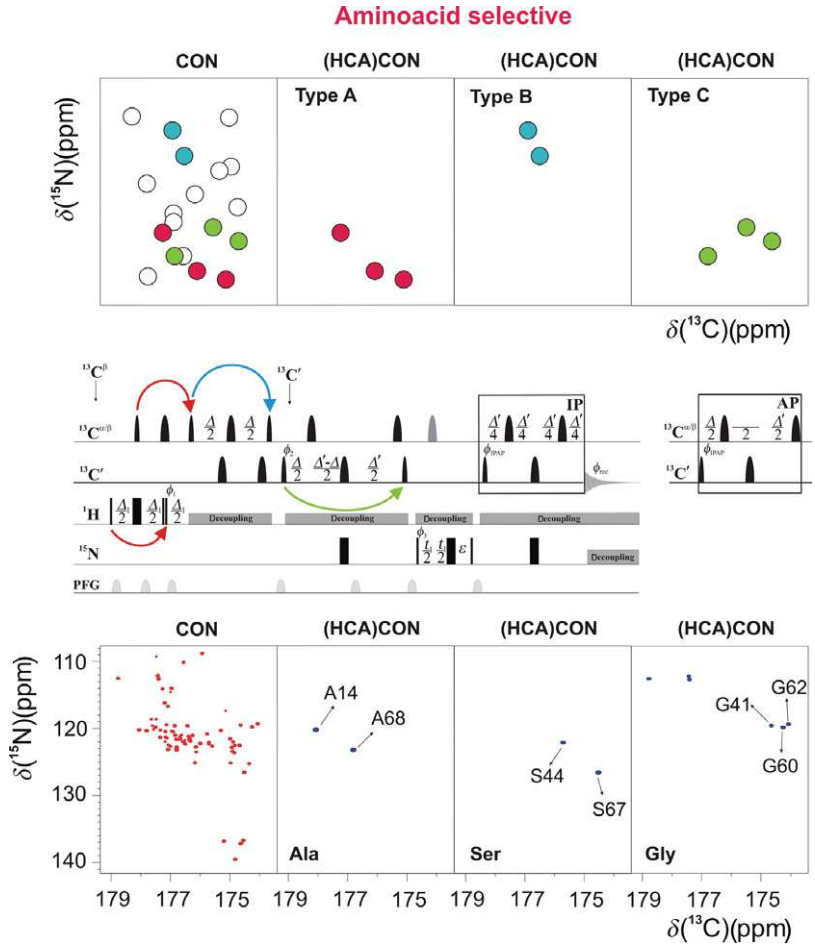


Figure 6.12 Amino acid selection to simplify crowded spectra. In top of the figure, a 2D CON spectrum and three 2D (HCA)CON spectra in which three different amino acids (types A, B and C) are selected are schematically represented. In the bottom of the figure, the same spectra acquired on reduced cox17 are reported to show how the CON spectrum can greatly be simplified. In the three examples, alanine, serine and glycine are specifically selected. The 2D (HCA)CON experiment used to acquire the spectra is reported in the middle of the figure. The arrows indicate the key coherence transfer steps: $H \rightarrow C$ (with triple quantum selection) and $C^{\beta} \rightarrow C^{\alpha}$, $C^{\alpha} \rightarrow C'$; other amino acids that have a CH_3 group do not contribute to the spectra, as more C–C transfer steps are necessary. Band-selective ^{13}C pulses are denoted by shapes and rectangular pulses represent hard pulses. The wide and narrow pulses correspond to π and $\pi/2$ flip angles. Pulse field gradients (PFG line) are also indicated by shapes. The delays are: $\Delta = 9$ ms, $\Delta_1 = 8$ ms, $\Delta' = 25$ ms, $\varepsilon = t_1(0) + p_C$, where p_C is the duration of the adiabatic 180° pulse (shown in grey) in the middle of the ^{15}N chemical shift evolution period. The phase cycles are: $\phi_1 = 2(30), 2(90), 2(150), 2(210), 2(270), 2(330)$; $\phi_2 = 12(x), 12(-x)$; $\phi_3 = x, -x$; ϕ_{IPAP} (IP) $= x$; ϕ_{IPAP} (AP) $= -y$; $\phi_{rec} = x, 2(-x), 2(x), 2(-x), 2(x), 2(-x), x, -x, 2(x), 2(-x), 2(x), 2(-x), 2(x), -x$. Quadrature detection in the F1 dimension was obtained by incrementing ϕ_3 in a States-TPPI manner.

As already noted several times, the well dispersed ^{13}C resonances render easier the design of band-selective ^{13}C pulses to excite or invert ^{13}C spins belonging only to the selected amino acids. Indeed, it is sufficient to use π band-selective pulses to refocus interactions only for selected amino acids. A very pertinent example in this respect is the selection of serine residues that result by a combination of a CH_2 filter, only one carbon–carbon (–C–C–) transfer, using a π pulse that only covers the C–C region of serine residues. The pulse sequences may become more complicated for different kinds of amino acids or some variants may enable selection of groups of amino acids. However, with combinations of these approaches, it is possible to simplify the spectra and through different combinations to identify resonances belonging to virtually all different amino acids (with just a couple of amino acids that are more difficult to be discriminated so far).

These selection criteria, since they basically exploit differences in the spectroscopic properties of amino acid side chains can also be implemented using just the (H)CBCACO experiment as a starting sequence or implemented in (H)CBCANCO-type pulse sequences, providing in this way the complementary correlations to be used for sequence-specific assignment and/or just a different variant.[110]

2.6. NMR observables for structural and dynamics characterization of proteins

The chemical shifts of the heteronuclei are the first valuable source of structural information deriving from ^{13}C detected exclusively heteronuclear NMR experiments.[111–113] Inter-nuclear distance information from ^{13}C–^{13}C nuclear Overhauser effects is instead not trivial to achieve. Indeed, the long-range distances between carbon spins are in general larger than the analogous ones between their directly bound protons responsible for the reduced ^{13}C–^{13}C long-range NOEs in respect to ^{1}H–^{1}H ones. Therefore, ^{13}C–^{13}C NOESY spectra result very useful as a complementary tool for spin-system identification, but not to achieve inter-nuclear distance information in an analogous way to what routinely done exploiting ^{1}H–^{1}H NOESY spectra. Therefore, the structural and dynamic information that can be achieved from ^{13}C-detected experiments, in addition to that naturally encoded in chemical shifts, resides in all the effects that can be determined by modifications of a basic 2D experiments, such as the CACO or the CON experiment, such as for example couplings (in particular residual dipolar couplings), relaxation rates and exchange effects.[73,114–116]

A series of experiments to determine one-bond couplings, which can thus be used for the determination of residual dipolar couplings, has been recently proposed.[114] These experiments exploit the 2D CON as a reference, since it is used for many applications. To enhance the sensitivity of the experiments without actively labelling proton chemical shifts in any of the dimensions of the NMR experiment, ^{1}H is used only as a starting polarization source, even if it is not a necessary requirement. H^{α} protons, which provide a starting pool for all residues in the protein and are generally less prone to exchange processes that may broaden the NMR lines, are used as a starting polarization source, through an extra coherence transfer step involving C^{α} to obtain in the end a C′-N correlation spectrum, which however also allows to determine observables involving C^{α}. One peak per amino acid is detected permitting to have complete information on all the residues in the protein, still keeping the overall number of peaks as small as possible.

Any experiment can be modified to determine scalar couplings as long as transverse coherence is created involving one of the two nuclei sharing the scalar coupling that we would like to determine. Several different experimental approaches to determine couplings have been proposed in the literature (E. COSY,[117–120] quantitative J correlation,[121–124] fitting modulations,[125] IPAP[31]). Among them, the IPAP approach was selected for its simplicity and robustness.[31] Simple modifications of the pulse sequence enable the determination of the C′–N coupling, and of the $N–H^{N}$ coupling which can in this way be determined also in cases in which the amide proton is broadened beyond detection. Separation of the two multiplet components into two sub-spectra is always useful to minimize the occurrence of peak overlap. The remaining one-bond couplings that are of interest ($C'–C^{\alpha}$, $C^{\alpha}–H^{\alpha}$) involve nuclei other than ^{15}N, which is frequency labelled in the indirect dimension of the experiment. In these cases, it is possible to combine the evolution of the desired coupling with the ^{15}N chemical shift evolution by a joint, synchronous increment of another delay in the pulse sequence, in addition to the one incremented to achieve ^{15}N chemical shift evolution. Thus, the evolution of the desired coupling induces a modulation in the detected signal that will cause a splitting in the indirect dimension of the experiment, proportional to the coupling itself. For the scalar coupling determination, this approach allows to encode the information about the desired scalar coupling in the most appropriate indirect dimension.[126] Therefore, all the desired couplings can be determined maintaining the same basic pulse sequence and type of experiment, chosen on the basis of the favourable chemical shift dispersion

properties of the two nuclear spins that are frequency labelled in the two dimensions of the experiment. This is a very important aspect in general, but it becomes one of the key features for the study of systems that lack the signal dispersion typical of well-structured molecules, such as in the case of unfolded proteins or IDPs. Several variants of the experiment can be designed to determine the C'–C^{α} coupling, since transverse coherence of each of the two nuclear spins is created, in different time intervals in the pulse sequence. The best results were obtained by exploiting C' transverse coherence rather than C^{α}.

A comment is due on the possibility to determine both C'–N and C'–C^{α} couplings in the direct acquisition dimension. Indeed, an elegant experiment based on C' direct detection has recently been proposed to determine the C'–C^{α} coupling.[127] Moreover, the spin-state-selective block used for C' homonuclear decoupling,[35,50] recently extended also to C'–N decoupling,[37] provides the coupling information "for free". However, the presence of additional ^{13}C–^{13}C couplings in the direct acquisition dimension, even if much smaller than the one-bond ones, may cause small distortions in the signals that are acceptable for the purpose of virtual decoupling but are inadequate for the precise determination of the desired scalar coupling. Therefore, as long as a sufficient resolution can be obtained in the indirect dimension, the determination of the coupling in this dimension has to be preferred.

Also, relaxation rates are widely used to characterize the structural and dynamic properties of a protein and can conveniently be measured with ^{13}C direct detection experiments, in order to obtain complementary information of those obtained with ^{1}H-detected experiments.[114,115] In particular, the basic pulse sequences can be modified for the determination of longitudinal and transverse carbonyl and nitrogen relaxation rates. The determination of transverse ^{13}C relaxation for a ^{13}C-enriched sample is complicated by the presence of the homonuclear scalar couplings that induce oscillatory behaviour in the transverse decays.[128,129] This effect can be explicitly included in the fitting function. However, inclusion of more fitting parameters may render the determination of the rates less accurate. Therefore, it is better to prevent the evolution of these couplings, as also previously proposed for the determination of carbonyl relaxation dispersion profiles.[129] In this way, the oscillations due to the large one-bond C^{α}–C' scalar coupling are avoided (small oscillatory behaviours may remain for residues such as aspartic acid, asparagine, glutamic acid and glutamine, due to the presence of C'–C^{γ} and C'–C^{δ} couplings which cannot be refocused).

Finally, exchange processes with the solvent provide indirect structural information which can however prove useful to discriminate between amide protons buried in hydrophobic cores or involved in strong hydrogen bonds and the ones located insolvent exposed loops. The ^{13}C-detected experiments offer in this frame a valuable tool to access this information indirectly, enabling minimal perturbation of the water resonance.

3. EXAMPLES OF APPLICATIONS

The great improvement in instrumental sensitivity that we have assisted to in recent years, in parallel with the experimental methods that have been developed to exploit heteronuclear properties, in particular in all the cases in which proton detection approaches find their limitations, have opened a new chapter in biomolecular NMR providing a new tool of investigation to be exploited. The suite of experiments based on ^{13}C direct detection can thus be used in parallel to the one based on ^{1}H direct detection to access complementary information in any protein study. In addition, for a variety of different kinds of applications, these experiments provide unique additional information. In this sense, ^{13}C detection can be considered paradoxically more sensitive than ^{1}H detection when it helps to recover information that would be lost using only ^{1}H detection. These applications are described hereafter.

3.1. Paramagnetic molecules

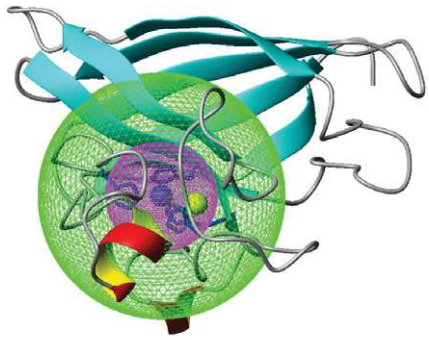
The initial motivation to focus on heteronuclear direct detection was provided by the need for improved methods for the characterization of paramagnetic proteins. Paramagnetic molecules are characterized by the presence of additional contributions to chemical shifts and nuclear relaxation arising from the so-called hyperfine interaction, that is the interaction between the nuclear spin I and the electron spin S. This interaction occurs via two different mechanisms, a through-bond interaction which depends on the amount of unpaired electron spin-density delocalized onto the investigated nuclear spin (contact interaction)[130,131] and a through-space interaction[132,133] which depends on the dipole–dipole interaction between nuclear and electron spins. Both effects contribute to chemical shifts and nuclear relaxation and provide a wealth of structural and dynamic information that cannot be accessed otherwise. The specific effects sensed by nuclear spins (chemical shifts, relaxation, partial alignment of molecules in high magnetic fields) depend on the properties of the paramagnetic centre present in the

molecule and thus, thanks to their peculiar nature, can be exploited for biomolecular NMR.

Due to the paramagnetism-induced relaxation, the NMR signals belonging to nuclei located in the surrounding of the paramagnetic centre can be broadened beyond detection. Independently from the mechanism by which it originates, paramagnetic relaxation depends on the square of the gyromagnetic ratio γ of the observed nucleus. This means that, assuming that all the other conditions remain the same (type of paramagnetic centre, distance from it), shifting the NMR detector form ^{1}H to ^{13}C enables to gain a factor of $(\gamma_H/\gamma_C)^2$ in terms of reduced contributions to relaxation. This is not the case for paramagnetism-induced chemical shifts. Therefore, ^{13}C direct-detected experiments provide in principle the same content of information as ^{1}H ones, but the loss of information due to paramagnetism-induced line broadening will be much less effective. The keynote finding is that hyperfine relaxation provide metal-nucleus distance restraints[16] and this observable, based on ^{13}C nuclei, may actually provide structural information in regions of a protein where ^{1}H resonances are broadened beyond detectable limits, as long as ^{1}H are never exploited in any of the magnetization transfer steps of the experiment ("*protonless* NMR"). In other words, the use of ^{13}C-detected *protonless* NMR experiments enables to recover information that would not be available through ^{1}H detection and improve the characterization of the system (Fig. 6.13).

Depending on the specific kind of paramagnetic centre, it may be more appropriate to quantify paramagnetic contributions to relaxation measuring either longitudinal or transverse relaxation rates. This effect is often used to access structural information by determining rates on a diamagnetic state and on a paramagnetic one, either exploiting paramagnetic centres naturally present in a protein, or recurring to metal substitution to exploit these effects or finally by engineering on purpose paramagnetic tags in a protein. In all these cases, the use of ^{13}C-detected NMR experiments enables to recover valuable information, in a complementary range to that accessible through ^{1}H detection, to improve the characterization of the system.

The experiments established on ^{13}C-based approach can be further optimized for paramagnetic molecules by selecting the most efficient coherence transfer pathways and identifying those that are least affected by fast relaxation. In principle, the most sensitive experiments are those based on coherence transfer mechanisms mediated by large scalar couplings in order to minimize the duration of delays present in the pulse sequence. The COSY-based experiments were the first to be used for paramagnetic



$$R_1^I, R_2^I \propto \gamma_I^2 \cdot \gamma_S^2 \cdot (1/r_I^6) \cdot f(\tau)$$

Figure 6.13 The paramagnetic impact on nuclear relaxation illustrated for monomeric Cu(II)/Zn(II) superoxide dismutase. In paramagnetic proteins, the longitudinal and transverse nuclear relaxation rates R_1 and R_2 are proportional to the product of terms listed in the figure, namely the square of the gyromagnetic ratio γ_I of the nuclear spin, the square of the gyromagnetic ratio γ_S of the unpaired electron spin, the reciprocal of their distance to the power of six and a function of the correlation time $f(\tau)$. As a consequence, the closer to the paramagnetic centre the atoms, the higher the relaxation rates are. In the case of Cu(II)/Zn(II) superoxide dismutase, the protons located inside the green sphere (11 Å radius) cannot be detected since are broadened beyond detection. ^{13}C direct detection allows instead to detect the carbon spins of the residues situated in proximity of the paramagnetic centre, except the ones belonging to the ligands (inside the purple sphere, 5 Å radius), thanks to their lower gyromagnetic ratio in respect to the one of protons. *Adapted with permission by ACS.*

proteins.[6,7,9,48,49] As an alternative, several other experiments based on coherence transfer via the $C'–C^{\alpha}$ scalar couplings, and also via the smaller $N–C'$ scalar coupling, exploiting either the single quantum or multiple quantum coherence transfer, can be used.[8–10] Signal losses due to fast relaxation can be reduced by shortening the coherence transfer delays[10] and even further by completely removing the building block in which the AP $C'–C^{\alpha}$ coherence is refocused and detecting directly the AP component.[134,135] However, when transverse relaxation is much faster than longitudinal relaxation, experiments based on the $^{13}C–^{13}C$ nuclear Overhauser effect can be usefully exploited, especially at higher fields.[136]

The use of *protonless* NMR experiments based on ^{13}C detection for structural purposes was initially applied to the study of copper binding

proteins, in particular to those containing a type-II Cu^{2+} centre that is known to cause broadening of ^{1}H signals in a large sphere of about 11 Å around the copper ion.[9,10,137] The *protonless* experiments enabled to detect all heteronuclear signals in the blind sphere, except the metal ligands, and ^{13}C-detected saturation recovery experiments were proposed as a tool to quantify LREs. These were used in conjunction with dihedral angle constraints deriving from chemical shifts, to determine a 3D structural model of the protein, despite the large "blind sphere" for ^{1}H detection.[137]

This approach was also applied to the study of several other metal binding proteins, such as other copper binding proteins,[138,139] iron binding proteins[140–143] as well as nichel containing ones.[8,144]. In addition, metal substitution of a diamagnetic metal ion with a paramagnetic one can be used to determine additional observables which involve the paramagnetic centre and thus contain structural and dynamic information. The most well-known example consists in calcium binding proteins, such as calbindin and oncomodulin, where the calcium ion was substituted by a lanthanide metal ion.[7,43,136,145–148] In these studies, thanks to the large anisotropy characteristic of the metal ions, it was possible to determine pseudocontact shifts and use them to achieve structural information. The large anisotropy of the metal ion also provides the driving force for partial orientation of the proteins in high magnetic fields so that this effect can be exploited to determine residual dipolar couplings. ^{13}C-detected *protonless* experiments provide a particularly valuable tool to this end at they enable to determine such effects in regions where protons are broadened beyond detection and also residual dipolar couplings involving protons, if proton longitudinal relaxation is not too large.[116]

Another area in which the *protonless* strategy finds application is that of paramagnetic tags, engineered on purpose on protein surfaces to access long-range structural information either within the protein itself or on its interactions with partners.[149] In this case, in addition to metal ions, also organic radicals are used to introduce paramagnetic effects. These have markedly different relaxation properties in respect to metal centres and thus in this case it is more convenient to quantify contributions to transverse relaxation rather than to longitudinal relaxation on the grounds that the latter are predicted to be much smaller and thus difficult to quantify accurately. Indeed, the same experiment is first conducted with the tag in a diamagnetic state, and then after the tag has been brought to a paramagnetic state, comparing the two spectra. A nice example in this context is provided by the case of the RNA recognition motif, a protein composed by two globular

domains connected by a flexible linker, in which the paramagnetic tags were used to define the relative orientations of two domains of the protein, important for its function.[149] It was shown that the complementary information derived from the combined use of ^{1}H- detected and ^{13}C-detected protonless experiments on the protein with one spin label was as useful as the combination of the ^{1}H-detected spectra acquired on a much larger number of tags.

Finally, highly relaxing agents, such as gadolinium complexes, can be added to a protein solution to identify protein surfaces.[150] Again the complementary use of ^{1}H-detected and ^{13}C-detected experiments to monitor paramagnetic relaxation enhancements were shown to provide useful information to characterize protein surfaces, to contribute to solution structure refinement and to monitor complex formation.

3.2. Large proteins

The problems deriving from too fast transverse relaxation present in paramagnetic molecules are also common to the field of proteins of increasing molecular mass. *Protonless* experiments were thus applied to proteins of different molecular size in order to evaluate their potential in the investigation of large proteins.[32,52,151] Results showed that they can give a crucial contribution in the study of systems that are so large that only a limited amount of information is available using solution state NMR, such as, for example, very large multimeric, highly symmetric proteins, or labelled proteins in otherwise unlabelled large macromolecular complexes, where the increase in molecular mass is not reflected by an analogous increase in the number of expected signals.[52,61,62,139,150,152]

It is well known that for folded proteins the increase in molecular mass causes an increase in the rotational correlation time. This has an impact on transverse relaxation rates, which contain contributions that are linearly proportional to the rotational correlation time. The first limitation that it is encountered is due to the large dipole–dipole interactions involving protons. As a consequence, it has been proposed to isotopically label proteins with ^{2}H.[17] Deuterium, despite the fact that it is characterized by $S=1$ and thus is more difficult to be directly exploited, is characterized by a lower gyromagnetic ratio compared to ^{1}H, and thus reduces the large contributions to relaxation which would have derived from the dense ^{1}H network present in proteins, even if at expense of the number of nuclear spins that can be directly sampled. Very clever experiments, designed to recover as much

information as possible through the constructive use of cross-correlated relaxation phenomena, have been proposed.[153] Specifically, the information is recovered selectively reintroducing amide protons (exploiting the $^{1}H–^{15}N/^{15}N$ dipole-chemical shift anisotropy cross-correlation rate)[153–155] or methyl groups (exploiting the $^{1}H–^{13}C$ dipole–dipole interactions).[156,157] Also, other effects have been proved to be potentially useful (CH_2 groups, CH aromatic groups),[154] but they have not become widely used.

Another important consideration when focusing on large macromolecules consists in evaluating the most promising kinds of ^{13}C spins for their investigation. Indeed, it is well known that carbonyls, characterized by a fairly large chemical shift anisotropy, are strongly influenced by the chemical shift anisotropy relaxation mechanism which quadratically grows with the increase in magnetic field strength, generally necessary when studying large (and thus complex) systems.[158] Therefore, the most suitable nuclear spins for the investigation of large macromolecules in solution are aliphatic ^{13}C spins in ^{2}H-labelled molecules, thanks to the favourable relaxation properties also upon increase in rotational correlation time at high magnetic fields. Other relaxation properties that benefit by an increase in rotational correlation time are the homonuclear cross-relaxation rates. These are responsible for the observation of cross-peaks in NOESY experiments. Although $^{13}C–^{13}C$ cross-relaxation rates are smaller than $^{1}H–^{1}H$ ones, because of the smaller gyromagnetic ^{13}C ratio they can be allowed to evolve for longer mixing times thanks to the slower longitudinal relaxation rates, conditions that are very favourable for the detection of cross-peaks in $^{13}C–^{13}C$ NOESY experiments. Moreover, in ^{13}C uniformly labelled proteins, the short $^{13}C–^{13}C$ distance between directly bound carbon spins partly compensates the smaller gyromagnetic ratios, enabling an easy detection of the correlations between neighbouring carbons. Spin diffusion within the network of directly bound ^{13}C chains becomes a very efficient phenomenon (Fig. 6.14). Therefore, $^{13}C–^{13}C$ NOESY-type experiments are promising for the study of very large macromolecules. On the other hand, experiments that feature long delays in which carbonyls are in the transverse plane, such as for example CON-type experiments (and more in general the CON-based suite), suffer by pronounced relaxation losses due to the large carbonyl transverse relaxation rates.

The performance of $^{13}C–^{13}C$ NOESY experiments was first tested on superoxide dismutase, a 153 amino acid protein that can be obtained in the monomeric and dimeric forms allowing the assessment of the effect of increasing molecular mass on the various kinds of experiments.[29,51] It was

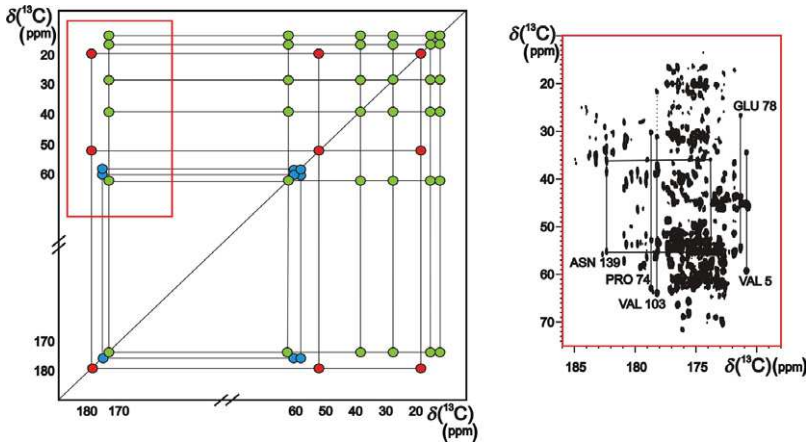


Figure 6.14 The left panel schematically illustrates the information content of ^{13}C–^{13}C NOESY spectra in solution, as exemplified for three different amino acids: alanine (red), serine (blue) and isoleucine (green). As an example, a portion of the 2D ^{13}C–^{13}C NOESY spectrum acquired on dimeric Cu(II)/Zn(II) superoxide dismutase is shown in the right panel.

shown that one-bond correlations can be easily detected and that the effect of spin diffusion becomes operative increasing the molecular mass.[29,51] ^{13}C–^{13}C NOESY experiments were then acquired on ferritin, a 480 kDa highly symmetric multimeric protein.[52] In this case, the amount of information that could be detected through ^{1}H, relaxation optimized experiments was very poor, whereas a variety of spin systems are detectable in the 2D spectrum. Several spin systems of amino acids characterized by well isolated chemical shifts can be detected[52] and resolution in crowded regions can be improved using appropriate acquisition schemes.[32] So, in principle, and as it has been shown, these ^{13}C–^{13}C NOESY experiments can be used to follow changes upon changes in experimental conditions, just as ^{1}H–^{15}N HSQC are generally used in all cases amenable to ^{1}H detection.[161] This means that information that would not be available otherwise can be recovered.

However, despite the correlations detected, one is left with the problem of sequence-specific assignment. Indeed, all the methods that one can think of, based on exploiting scalar couplings in solution, end up with either too small couplings or too large transverse relaxation of at least one of the two spins involved in the coupling. Another option would be to detect longer range correlations in ^{13}C–^{13}C NOESY experiments. Indeed, if sensitivity

would allow to access long-range ^{13}C–^{13}C distance information, this would open new avenues to access structural information never seen before in solution NMR. However, these effects are expected to be very small, and on completely isotopically enriched samples long-range correlations would be at least two orders of magnitude smaller in respect to the one-bond correlations (and spin diffusion ones). Despite the evidence that these correlations are in principle detectable, they are really at the limits of instrumental sensitivity; therefore, another significant leap in sensitivity would be necessary, either in the instrument technology or perhaps through one of the methods for hyperpolarization.

One solution proposed to circumvent the problem of the assignment in case of very large proteins is the combined use of solution and solid-state NMR spectroscopy.[159–161] Indeed, the slow re-orientation of large molecules in solution, which is responsible for line broadening, does not constitute a limitation in the solid state. Therefore, backbone carbonyl carbon and nitrogen atoms of large molecules become detectable in MAS solid-state experiments and a variety of experiments can be used for sequence-specific assignment. An increasing number of MAS solid-state NMR experiments are now being developed enabling the study of more complex molecules.[162,163] Moreover, it has recently been shown that highly concentrated solutions of high molecular mass proteins tend to sediment upon magic angle spinning inside rotors, providing a complementary tool for solution state studies of large systems.[164] As this area is experiencing very fast development, combination of solution and solid-state MAS NMR experiments is likely to become a very valuable tool in the near future.

3.3. Intrinsically disordered proteins (IDPs)

IDPs are a class of proteins characterized by the lack of a well-structured three-dimensional fold. In fact, in physiological conditions, IDPs do not undergo the folding process,[165–167] but maintain highly flexible conformations during their resting state.

Despite the absence of a rigid secondary or tertiary structure, IDPs fulfil a number of biological functions.[168–170] In fact, IDPs carry out many important activities in the cell, which range from regulation of transcription and translation, storage of small molecules and regulation of the self-assembly of large multi-protein complexes. Moreover, some IDPs are involved in binding to several partners, playing in this way an importantrole in signal transduction and various regulatory processes.[171]

Bioinformatics indeed shows that a large portion of the genome encodes for proteins that remain fully or partially disordered in the cell: more or less one third of the eukaryotic proteins (25–41%) contain in fact unstructured regions longer than 30 consecutive residues that do not possess any regular secondary or tertiary structure.[172] This is the main reason why, in the past decade, unstructured polypeptide sequences have been attracting an increased attention of biochemists and structural biologists.[172–177] It is thus becoming accepted that the peculiar properties of IDPs do confer them functional advantages that are highly complementary to those deriving from the presence of a well-defined 3D structure, demanding the extension of the well-accepted structure–function paradigm in a new dimension of high flexibility and disorder.

The peculiar features of IDPs prevent them from being suitable targets for crystallographic studies. In fact, the lack of a well-defined structure and the consequent high flexibility allow these proteins to adopt multiple and variable conformations, making impossible the preparation of ordered crystals and thus leading to failure in crystallographic structural analysis. As a consequence, NMR spectroscopy plays a crucial role in the investigation of disordered proteins, being the only method that allows a high resolution description of IDPs' conformations and dynamics in solution.[167,175,178–180]

However, the biophysical characterization of IDPs by NMR presents several drawbacks too. In fact, the peculiar properties of IDPs, namely a high degree of structural and dynamic heterogeneity, as well as the presence of highly solvent exposed backbones do have a significant impact on the NMR observables that strongly influences the quality of NMR spectra, the information that can be retrieved as well as the subsequent interpretation of the results. First, the lack of a well-defined 3D structure and the dynamic inter-conversion between different conformers average out the majority of the contributions to chemical shifts arising from the 3D structure (neighbouring groups, dihedral angles, etc.). This means that the chemical shift ranges are drastically reduced causing a great degeneracy of NMR signals. As a consequence, the overlap in the spectra is often so severe to prevent standard NMR experiments to provide sufficient resolving power for the unambiguous assignment of the resonances, especially in the case of amino acid sequences with highly repetitive motifs. Second, under physiological conditions, labile amide protons, which are much exposed to the solvent due to the lack of structured elements exchange rapidly with the solvent, broadening $^{1}H^{N}$ resonances even beyond detection, with consequent

decrease in sensitivity and resolution of the set of experiments based on amide proton detection. In addition, IDPs often contain multiple proline-rich polypeptide segments which, lacking of the amide protons, cause difficulties in sequential assignment of residues. All these peculiar properties of IDPs should thus be taken into account in the design of NMR methods to be able to increase the size and complexity of IDPs that can be studied at atomic resolution through NMR.[180]

In the past years, great improvements have been made in the optimization of several NMR experiments for the investigation of IDPs.[37,87,89,181–184] In particular, the most promising ones exploit the chemical shift labelling of several heteronuclei and use ^{13}C-direct detection. There are many reasons behind this choice.

To begin with, ^{13}C-direct detection has the great advantage to be not influenced by the rate of $^{1}H^{N}$ exchange, and therefore is very useful to recover information that might be lost with $^{1}H^{N}$ detection.[19–21] Indeed, when looking carefully in the literature, it can be seen that many of the studies performed on IDPs have been conducted at low pH and/or low temperature, in order to minimize the contributions to the $^{1}H^{N}$ exchange broadening and improve the quality of the ^{1}H–^{15}N HSQC spectra and related suites of triple resonance experiments. However, it is much more interesting to study IDPs in conditions closer to physiological ones, and in this frame, the exclusively heteronuclear NMR experiments based on carbon detection offer a valuable tool of investigation. Moreover, since the reduction in the chemical shift dispersion upon loss of a stable 3D structure is more pronounced for ^{1}H than for ^{13}C (Fig. 6.15), carbon and the other heteronuclei result better candidates to be exploited for the study of IDPs thanks to the higher chemical shift dispersion in respect to protons.[59,109,185] These considerations can be used to decide the most appropriate experiments to achieve the sequence-specific assignment of the polypeptide. As an example, assuming a completely random coil behaviour, which represents the minimal chemical shift dispersion situation, one can predict the dispersion of cross-peaks in the most sensitive 2D experiments correlating the backbone nitrogen with the directly bound proton or carbon (amide proton and carbonyl carbon, Fig. 6.16). Indeed, chemical shifts can be calculated for any protein by simply taking the primary sequence and using random coil values typical of each amino acid,[111] properly corrected using the contributions from neighbouring amino acids.[186] This shows that C′–N correlations display the most favourable chemical shift dispersion and that proline residues, often very abundant in IDPs, can be easily detected by using ^{13}C direct

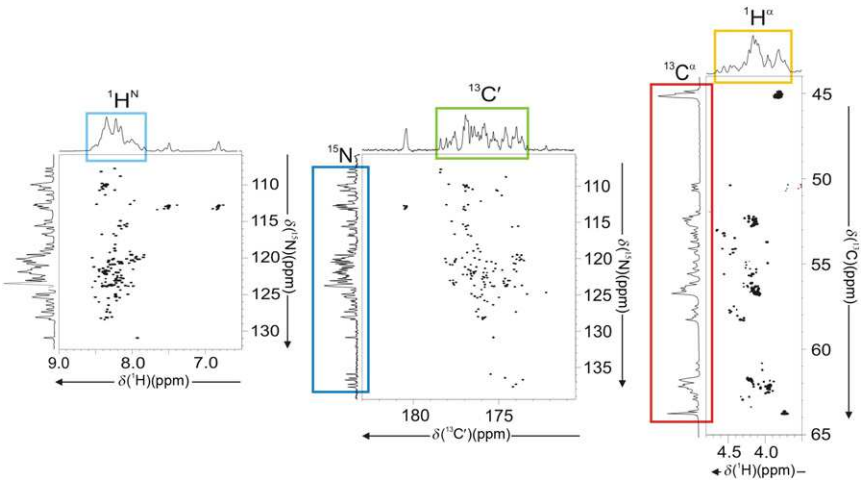


Figure 6.15 The chemical shift dispersion in IDPs. The 2D ^{1}H–^{15}N HSQC, ^{13}C–^{15}N CON-IPAP and ^{1}H–^{13}C HSQC spectra of ^{13}C, ^{15}N human α-synuclein are shown. The different chemical shift dispersion of the nuclei belonging to the backbone, namely $^{1}H^{\alpha}$, $^{1}H^{N}$, $^{13}C^{\alpha}$, $^{13}C'$ and ^{15}N, is highlighted in the 1D projections of the spectra. The three spectra, recorded at 16.4 T, are reported with the same Hz/cm ratio.

detection NMR experiments (Fig. 6.16). The possibility to perform ^{13}C direct detection has of course opened new avenues for the study of IDPs, enabling to take maximum advantage of the large heteronuclear dispersion, in all detected dimension of multidimensional NMR experiments.

Experiments to perform sequence-specific assignment of IDPs can be designed to detect correlations between backbone heteronuclei, in particular carbonyls and nitrogens, exploiting the small $^{3}J_{C'C'}$ as well as multiple magnetization transfer steps mediated by heteronuclear scalar couplings such as $^{1}J_{C'N}$, $^{1}J_{CN}$, $^{2}J_{CN}$.[37,59,60,87,89] Indeed the conventional strategy used for sequence-specific assignment of folded proteins, based on C^{α} and C^{β} chemical shifts, does not really work well for IDPs since the nuclei of amino acid side chains tend to be affected to a minor extent by the contributions of the neighbouring amino acids in the primary sequence and, in the absence of a partially structured conformation, tend to average to the random coil values typical of each amino acid (Fig. 6.15). However, this aspect, which on one hand causes extensive spectral overlap, can be profitably used as a criterion to identify amino acid types, provided other nuclear spins are used to resolve the resonances coming from different amino acids. To this end, the most suitable nuclear spins are backbone

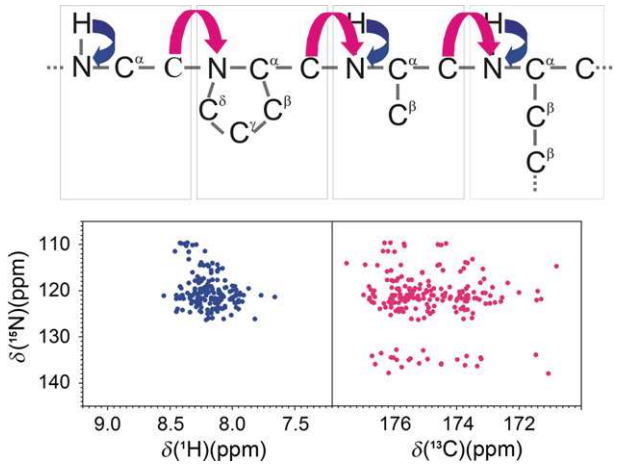


Figure 6.16 The predicted chemical shift dispersion of cross-peaks in 2D $^{1}H–^{15}N$ HSQC and $^{13}C–^{15}N$ CON-IPAP spectra. The two plots show how it is possible to improve the chemical shift dispersion correlating nitrogens to attached carbonyls rather than to protons. Random coil chemical shifts, properly corrected for the contributions of neighbouring aminacids in the primary sequence, were used for the plots. The primary sequence of human securin was chosen as an example. *Printed with permission by Wiley.*

ones, which are more influenced by the contributions of neighbouring amino acids, in particular those involved in the peptide bond. Moreover, since in disordered proteins chemical shifts tend to cluster to those typical of each amino acid type, correlating nuclear spins deriving from different amino acids (a natural feature contributing to the nice resolution of the CON experiment) provides a way to reduce the problem of extensive overlap.[63] Therefore, multidimensional experiments devoted to spin-system identification can be merged with information retrieved from experiments designed to obtain sequential correlations in a very efficient way. Fortunately, the almost unrestricted motional freedom makes the relaxation properties of IDPs ideal for high multidimensional NMR studies, thanks to the relatively long transverse relaxation times even on relatively long polypeptide chains. Exploitation of a higher number of dimensions appears as the only way to overcome the issue. In fact, further chemical shifts evolution periods can spread the overlapping resonances and thus lead to a more reliable and less ambiguous assignment. The more the frequency-labelled nuclei are chosen among the ones which exhibit the

greater chemical shift dispersions and belong to two or even three different residues, the less are the chances to still not resolve overlapping peaks. Therefore, appropriate magnetization transfer pathways can be selected in order to identify less ambiguous sequential linkages, in respect to the connectivities provided by 3D experiments, making this approach extremely useful to sequentially join consecutives residues especially for proteins characterized by a very repetitive sequence. Therefore, experiments of higher dimensionality are feasible on IDPs and can be extremely helpful where the suites of well established and routinely used 3D NMR experiments fail. Thanks to the great improvement in NUS and to the available programs for data processing, experiments which involve several chemical shifts evolution periods can be efficiently used for IDPs. These methods allow to obtain spectra in which high resolution is provided in all indirect dimensions, making the resolving power of such experiments adequate for IDPs. The high number of dimensions, together with the high resolution, is the key to minimize as much as possible the occurrence of the cross-peak overlaps.

Additional experiments based on ^{13}C detection that exploit the nice dispersion of the 2D CON as a read-out scheme, have been developed to achieve additional information for the structural and dynamic characterization of IDPs. These include, as described above, experiments to determine relaxation rates, one-bond couplings to access residual dipolar couplings, as well as experiments to determine exchange processes with the solvent as well as heteronuclear NOEs. These experiments are expected to contribute to extending the size and complexity of IDPs that can be characterized at atomic resolution and have already been used in a number of cases.[19,20,24,37,57,59,74,88,89,187–189]

REFERENCES

1. Sattler M, Schleucher J, Griesinger C. Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. *Prog NMR Spectrosc* 1999;**34**:93.
2. Bax A, Grzesiek S. Methodological advances in protein NMR. *Acc Chem Res* 1993;**26**:131.
3. Bodenhausen G, Ruben DJ. Natural abundance nitrogen-15 NMR by enhanced heteronuclear spectroscopy. *Chem Phys Lett* 1980;**69**:185.
4. Bax A, Griffey RH, Hawkins BL. Sensitivity-enhanced correlation of ^{15}N and ^{1}H chemical shifts in natural-abundance samples via multiple quantum coherence. *J Am Chem Soc* 1983;**105**:7188.
5. Kolczak U, Salgado J, Siegal G, Saraste M, Canters GW. Paramagnetic NMR studies of blue and purple copper proteins. *Biospectroscopy* 1999;**5**:S19.

6. Serber Z, Richter C, Dötsch V. Carbon-detected NMR experiments to investigate structure and dynamics of biological macromolecules. *Chembiochem* 2001;**2**:247.
7. Bertini I, Lee Y-M, Luchinat C, Piccioli M, Poggi L. Locating the metal ion in calcium-binding proteins by using cerium(III) as a probe. *Chembiochem* 2001;**2**:550.
8. Kostic M, Pochapsky SS, Pochapsky TC. Rapid recycle ^{13}C, ^{15}N and ^{13}C,^{13}C heteronuclear and homonuclear multiple quantum coherence detection for resonance assignments in paramagnetic proteins: example of Ni^{2+}-containing acireductone dioxygenase. *J Am Chem Soc* 2002;**124**:9054.
9. Arnesano F, Banci L, Bertini I, Felli IC, Luchinat C, Thompsett AR. A strategy for the NMR characterization of type II copper(II) proteins: the case of the copper trafficking protein CopC from *Pseudomonas syringae*. *J Am Chem Soc* 2003;**125**:7200.
10. Bermel W, Bertini I, Felli IC, Kümmerle R, Pierattelli R. ^{13}C direct detection experiments on the paramagnetic oxidized monomeric copper, zinc superoxide dismutase. *J Am Chem Soc* 2003;**125**:16423.
11. Bertini I, Duma L, Felli IC, Fey M, Luchinat C, Pierattelli R, et al. A heteronuclear direct detection NMR experiment for protein backbone assignment. *Angew Chem Int Ed* 2004;**43**:2257.
12. Bermel W, Bertini I, Duma L, Emsley L, Felli IC, Pierattelli R, et al. Complete assignment of heteronuclear protein resonances by *protonless* NMR spectroscopy. *Angew Chem Int Ed* 2005;**44**:3089.
13. Serber Z, Richter C, Moskau D, Boehlen J-M, Gerfin T, Marek D, et al. New carbon-detected protein NMR experiments using cryoprobes. *J Am Chem Soc* 2000;**122**:3554.
14. Kovacs H, Moskau D, Spraul M. Cryogenically cooled probes—a leap in NMR technology. *Prog NMR Spectrosc* 2005;**46**:131.
15. Bertini I, Luchinat C, Parigi G. *Solution NMR of paramagnetic molecules*. Amsterdam: Elsevier; 2001.
16. Bertini I, Luchinat C, Parigi G, Pierattelli R. NMR of paramagnetic metalloproteins. *Chembiochem* 2005;**6**:1536.
17. LeMaster DM. Isotope labeling in solution protein assignment and structural analysis. *Prog NMR Spectrosc* 1994;**26**:371.
18. Bertini I, Felli IC, Gonnelli L, Pierattelli R, Spyranti Z, Spyroulias GA. Mapping protein-protein interaction by $^{13}C'$-detected heteronuclear NMR spectroscopy. *J Biomol NMR* 2006;**36**:111.
19. Hsu ST, Bertoncini CW, Dobson CM. Use of *protonless* NMR spectroscopy to alleviate the loss of information resulting from exchange-broadening. *J Am Chem Soc* 2009;**131**:7222.
20. Skora L, Becker S, Zweckstetter M. Molten globule precursor states are conformationally correlated to amyloid fibrils of human beta-2-microglobulin. *J Am Chem Soc* 2010;**132**:9223.
21. Gil S, Hosek T, Solyom Z, Kümmerle R, Brutscher B, Pierattelli R, et al. NMR studies of intrinsically disordered proteins near physiological conditions. 2013.
22. Zheng G, Price WS. Solvent suppression in NMR. *Prog NMR Spectrosc* 2010;**56**:267.
23. Shimba N, Kovacs H, Stern AS, Nomura AM, Shimada I, Hoch JC, et al. Optimization of ^{13}C direct detection NMR methods. *J Biomol NMR* 2004;**30**:175.
24. Csizmok V, Felli IC, Tompa P, Banci L, Bertini I. Structural and dynamic characterization of intrinsically disordered human securin by NMR. *J Am Chem Soc* 2008;**130**:16873.
25. Hammarström A, Otting G. Improved spectral resolution in ^{1}H NMR spectroscopy by homonuclear semiselective shaped pulse decoupling during acquisition. *J Am Chem Soc* 1994;**116**:8847.

26. Vögeli B, Kovacs H, Pervushin K. Simultaneous ^{1}H- or ^{2}H-, ^{15}N- and multiple-band-selective ^{13}C-decoupling during acquisition in ^{13}C-detected experiments with proteins and oligonucleotides. *J Biomol NMR* 2005;**31**:1.
27. Duma L, Hediger S, Brutscher B, Böckmann A, Emsley L. Resolution enhancement in multidimensional solid-state NMR spectroscopy of proteins using spin-state selection. *J Am Chem Soc* 2003;**125**:11816.
28. Duma L, Hediger S, Lesage A, Emsley L. Spin-state selection in solid-state NMR. *J Magn Reson* 2003;**164**:187.
29. Bertini I, Felli IC, Kümmerle R, Luchinat C, Pierattelli R. ^{13}C-^{13}C NOESY: a constructive use of ^{13}C-^{13}C spin-diffusion. *J Biomol NMR* 2004;**30**:245.
30. Bermel W, Bertini I, Felli IC, Kümmerle R, Pierattelli R. Novel ^{13}C direct detection experiments, including extension to the third dimension, to perform the complete assignment of proteins. *J Magn Reson* 2006;**178**:56.
31. Ottiger M, Delaglio F, Bax A. Measurement of J and dipolar couplings from simplified two-dimensional NMR spectra. *J Magn Reson* 1998;**131**:373.
32. Bermel W, Felli IC, Matzapetakis M, Pierattelli R, Theil EC, Turano P. A method for C^{α} direct-detection in protonless NMR. *J Magn Reson* 2007;**188**:301.
33. Lerche MH, Meissner A, Poulsen FM, Sørensen OW. Pulse sequences for measurement of one-bond (15)N-(1)H coupling constants in the protein backbone. *J Magn Reson* 1999;**140**:259.
34. Verel R, Manolikas T, Siemer AB, Meier BH. Improved resolution in ^{13}C solid-state spectra through spin-state-selection. *J Magn Reson* 2007;**184**:322.
35. Bermel W, Felli IC, Kümmerle R, Pierattelli R. ^{13}C direct-detection biomolecular NMR. *Concepts Magn Reson* 2008;**32A**:183.
36. Lee D, Vögeli B, Pervushin K. Detection of C′, Ca correlations in proteins using a new time- and sensitivity-optimal experiment. *J Biomol NMR* 2005;**31**:273.
37. Bermel W, Bertini I, Csizmok V, Felli IC, Pierattelli R, Tompa P. H-start for exclusively heteronuclear NMR spectroscopy: the case of intrinsically disordered proteins. *J Magn Reson* 2009;**198**:275.
38. Kern T, Schanda P, Brutscher B. Sensitivity-enhanced IPAP-SOFAST-HMQC for fast-pulsing 2D NMR with reduced radiofrequency load. *J Magn Reson* 2008;**190**:333.
39. Laage S, Lesage A, Emsley L, Bertini I, Felli IC, Pierattelli R, et al. Transverse-dephasing optimized homonuclear J-decoupling in solid-state NMR spectroscopy of uniformly 13^{C}-labeled proteins. *J Am Chem Soc* 2009;**131**:10816.
40. Bertini I, Emsley L, Felli IC, Laage S, Lesage A, Lewandoski DA, et al. High-resolution and sensitivity through-bond correlations in ultra-fast MAS solid-state NMR. *Chem Sci* 2011;**2**:345.
41. Barbet-Massin E, Pell AJ, Knight MJ, Webber AL, Felli IC, Pierattelli R, et al. ^{13}C-detected through-bond correlation experiments for protein resonance assignment by ultra-fast MAS NMR. *Chemphyschem* 2013; Epub ahead of print.
42. Vögeli B, Kovacs H, Pervushin K. Measurements of side chain ^{13}C-^{13}C residual dipolar coupling in uniformly deuterated proteins. *J Am Chem Soc* 2004;**126**:2414.
43. Babini E, Felli IC, Lelli M, Luchinat C, Pierattelli R. Backbone and side-chains ^{1}H, ^{13}C and ^{15}N NMR assignment of human β-parvalbumin. *J Biomol NMR* 2005;**33**:137.
44. Hu K, Eletsky A, Pervushin K. Backbone resonance assignment in large protonated proteins using a combination of new 3D TROSY-HN(CA)HA, 4D TROSY-HACANH and 13C detected HACACO experiments. *J Biomol NMR* 2003;**26**:69.
45. Oh B-H, Westler WM, Darba P, Markley JL. Protein carbon-13 spin systems by a single two-dimensional nuclear magnetic resonance experiment. *Science* 1988;**240**:908.
46. Westler WM, Kainosho M, Nagao H, Tomonaga M, Markley JL. Two-dimensional NMR strategies for carbon-carbon correlations and sequence-specific assignments in carbon-13 labeled proteins. *J Am Chem Soc* 1988;**110**:4093.

47. Oh B-H, Westler WM, Markley JL. Carbon-13 spin system directed strategy for assigning cross peaks in the COSY fingerprint region of a protein. *J Am Chem Soc* 1989;**111**:3083.
48. Machonkin TE, Westler WM, Markley JL. ^{13}C-^{13}C 2D NMR: a novel strategy for the study of paramagnetic proteins with slow electronic relaxation times. *J Am Chem Soc* 2002;**124**:3204.
49. Machonkin TE, Westler WM, Markley JL. Strategy for the study of paramagnetic proteins with slow electronic relaxation rates by NMR spectroscopy application to oxidized human [2Fe-2S] ferredoxin. *J Am Chem Soc* 2004;**126**:5413.
50. Bermel W, Bertini I, Felli IC, Piccioli M, Pierattelli R. ^{13}C-detected *protonless* NMR spectroscopy of proteins in solution. *Prog NMR Spectrosc* 2006;**48**:25.
51. Bertini I, Felli IC, Kümmerle R, Moskau D, Pierattelli R. ^{13}C-^{13}C NOESY: an attractive alternative for studying large macromolecules. *J Am Chem Soc* 2004;**126**:464.
52. Matzapetakis M, Turano P, Theil EC, Bertini I. ^{13}C-^{13}C NOESY spectra of a 480 kDa protein: solution NMR of ferritin. *J Biomol NMR* 2007;**38**:237.
53. Eletsky A, Moreira O, Kovacs H, Pervushin K. A novel strategy for the assignment of side-chain resonances in completely deuterated large proteins using ^{13}C spectroscopy. *J Biomol NMR* 2003;**26**:167.
54. Pervushin K, Eletsky A. A new strategy for backbone resonance assignment in large proteins using a MQ-HACACO experiment. *J Biomol NMR* 2003;**25**:147.
55. Hu K, Vögeli B, Clore GM. ^{13}C detected HN(CA)C and HMCMC experiments using a single methyl reprotonated sample for unambiguous methyl-resonance assignment. *J Biomol NMR* 2006;**36**:259.
56. Jordan JB, Kovacs H, Wang Y, Mobli M, Luo R, Anklin C, et al. Three-dimensional ^{13}C-detected CH_3-TOCSY using selectively protonated proteins: facile methyl resonance assignment and protein structure determination. *J Am Chem Soc* 2006;**128**:9119.
57. O'Hare B, Benesi AJ, Showalter SA. Incorporating 1H chemical shift determination into 13C-direct detected spectroscopy of intrinsically disordered proteins in solution. *J Magn Reson* 2009;**200**:354.
58. Bermel W, Bertini I, Felli IC, Pierattelli R, Vasos PR. A selective experiment for the sequential protein backbone assignment from 3D heteronuclear spectra. *J Magn Reson* 2005;**172**:324.
59. Bermel W, Bertini I, Felli IC, Lee Y-M, Luchinat C, Pierattelli R. Protonless NMR experiments for sequence-specific assignment of backbone nuclei in unfolded proteins. *J Am Chem Soc* 2006;**128**:3918.
60. Felli IC, Pierattelli R, Glaser SJ, Luy B. Relaxation-optimised Hartmann-Hahn transfer for carbonyl-carbonyl correlation spectroscopy using a specifically tailored MOCCA-XY16 mixing sequence for protonless ^{13}C direct detection experiments. *J Biomol NMR* 2009;**43**:187.
61. Takeuchi K, Frueh DP, Hyberts SG, Sun ZYJ, Wagner G. High-resolution 3D CANCA NMR experiments for complete main chain assignments using C^{α} direct detection. *J Am Chem Soc* 2010;**132**:2945.
62. Takeuchi K, Frueh DP, Sun ZYJ, Hiller S, Wagner G. CACA-TOCSY with alternate ^{13}C-^{12}C labeling: a $^{13}C^{\alpha}$ direct detection experiment for main chain resonance assignment, dihedral angle information and aminoacid type identification. *J Biomol NMR* 2010;**47**:55.
63. Bermel W, Bruix M, Felli IC, Kumar VMV, Pierattelli R, Serrano S. Improving the chemical shift dispersion of multidimensional NMR spectra of intrinsically disordered proteins. *J Biomol NMR* 2013;**55**:231.
64. Castellani F, van Rossum B, Diehl A, Schubert M, Rehbein K, Oschkinat H. Structure of a protein determined by solid-state magic-angle-spinning NMR spectroscopy. *Nature* 2002;**420**:98.

65. Felli IC, Brutscher B. Recent advancements in solution NMR: fast methods and heteronuclear direct detection. *Chemphyschem* 2009;**10**:1356.
66. Pervushin K, Vogeli B, Eletsky A. Longitudinal (1)H relaxation optimization in TROSY NMR spectroscopy. *J Am Chem Soc* 2002;**124**:12898.
67. Schanda P, Kupce E, Brutscher B. SOFAST-HMQC experiments for recording two-dimensional heteronuclear correlation spectra of proteins within a few seconds. *J Biomol NMR* 2005;**33**:199.
68. Schanda P, Brutscher B. 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* 2005;**127**:8014.
69. Deschamps M, Campbell ID. Cooling overall spin temperature: protein NMR experiments optimized for longitudinal relaxation effects. *J Magn Reson* 2006;**178**:206.
70. Schanda P, Van Melckebeke H, Brutscher B. Speeding up three-dimensional protein NMR experiments to a few minutes. *J Am Chem Soc* 2006;**128**:9042.
71. Brutscher B, Schanda P. Rapid multidimensional NMR: fast pulsing techniques and their application to proteins. In: Grant M, Harris RK, editors. *Encyclopedia of NMR*. Wiley; 2009.
72. Bermel W, Bertini I, Felli IC, Pierattelli R. Speeding up ^{13}C direct detection biomolecular NMR experiments. *J Am Chem Soc* 2009;**131**:15339.
73. Bertini I, Felli IC, Gonnelli L, Kumar VMV, Pierattelli R. High-resolution characterization of intrinsic disorder in proteins: expanding the suite of ^{13}C detected NMR experiments to determine key observables. *Chembiochem* 2011;**12**:2347.
74. Bertini I, Felli IC, Gonnelli L, Kumar VMV, Pierattelli R. ^{13}C direct-detection biomolecular NMR spectroscopy in living cells. *Angew Chem Int Ed* 2011;**50**:2339.
75. Hoch JC. Modern spectrum analysis in nuclear magnetic resonance: alternatives to the Fourier transform. *Methods Enzymol* 1989;**176**:216.
76. Hoch JC, Stern AS. *NMR data processing*. New York: Wiley-Interscience; 1996.
77. Hyberts SG, Frueh DP, Arthanari H, Wagner G. FM reconstruction of non-uniformly sampled protein NMR data at higher dimensions and optimization by distillation. *J Biomol NMR* 2009;**45**:283.
78. Hyberts SG, Heffron GJ, Tarragona NG, Solanky K, Edmonds KA, Luithardt H, et al. Ultrahigh-resolution ^{1}H-^{13}C HSQC spectra of metabolite mixtures using nonlinear sampling and forward maximum entropy reconstruction. *J Am Chem Soc* 2007;**129**:5108.
79. Orekhov VY, Ibraghimov I, Billeter M. Optimizing resolution in multidimensional NMR by three-way decomposition. *J Biomol NMR* 2003;**27**:165.
80. Orekhov VY, Kazimierczuk K. Accelerated NMR spectroscopy by using compressed sensing. *Angew Chem Int Ed* 2011;**50**:5556.
81. Matsuki Y, Eddy MT, Griffin RG. Rapid three-dimensional MAS NMR spectroscopy at critical sensitivity. *Angew Chem Int Ed* 2010;**49**:9215.
82. Kazimierczuk K, Zawadzka A, Kozminski W. Narrow peaks and high dimensionalities: exploiting the advantages of random sampling. *J Magn Reson* 2009;**197**:219.
83. Stanek J, Kozminski W. Iterative algorithm of discrete Fourier transform for processing randomly sampled data sets. *J Biomol NMR* 2010;**47**:65.
84. Stanek J, Augustyniak R, Kozminski W. Suppression of sampling artefacts in high-resolution four-dimensional NMR spectra using signal separation algorithm. *J Magn Reson* 2012;**214**:91.
85. Coggins BE, Zhou P. High resolution 4-D spectroscopy with sparse concentric shell sampling and FFT-CLEAN. *J Biomol NMR* 2008;**42**:225.
86. Kazimierczuk K, Zawadzka A, Kozminski W. Optimization of random time domain sampling in multidimensional NMR. *J Magn Reson* 2008;**192**:123.

87. Bermel W, Bertini I, Gonnelli L, Felli IC, Kozminski W, Piai A, et al. Speeding up sequence specific assignment of IDPs. *J Biomol NMR* 2012;**53**:293.
88. Novacek J, Haba NY, Chill JH, Zidek L, Sklenar V. 4D Non-uniformly sampled HCBCACON and (1) J(NC$^{\alpha}$)-selective HCBCANCO experiments for the sequential assignment and chemical shift analysis of intrinsically disordered proteins. *J Biomol NMR* 2012;**53**:139.
89. Novacek J, Zawadzka-Kazimierczuk A, Papoušková V, Zidek L, Sanderová H, Krasny L, et al. 5D ^{13}C-detected experiments for backbone assignment of unstructured proteins with a very low signal dispersion. *J Biomol NMR* 2011;**50**:1.
90. Olejniczak ET, Fesik SW. Two dimensional nuclear magnetic resonance method for identifying the Ha-Ca signals of amino acid residues preceding prolines. *J Am Chem Soc* 1994;**116**:2215.
91. Ghering K, Guittet E. Two-dimensional nuclear magnetic resonance method for identifying the HN/N signals of amino-acid residues following glycine. *J Magn Reson B* 1995;**109**:206.
92. Yamazaki T, Pascal SM, Singer AU, Forman-Kay JD, Kay LE. NMR pulse schemes for the sequence-specific assignment of arginine guanidino ^{15}N and ^{1}H chemical shifts in proteins. *J Am Chem Soc* 1995;**117**:3556.
93. Dötsch V, Oswald RE, Wagner G. Amino-acid type-selective triple-resonance experiments. *J Magn Reson B* 1996;**110**:107.
94. Rao NS, Legault P, Muhandiram DR, Greenblatt J, Battiste JL, Williamson JR, et al. NMR pulse schemes for the sequential assignment of arginine side-chain H$^{\varepsilon}$ protons. *J Magn Reson B* 1996;**113**:272.
95. Dötsch V, Oswald RE, Wagner G. Selective identification of threonine, valine and isoleucine sequential connectivities with a TVI-CBCACONH experiment. *J Magn Reson B* 1996;**110**:304.
96. Dötsch V, Matsuo H, Wagner G. Amino-acid type identification for deuterated proteins with a β-carbon-edited HNCOCACB experiment. *J Magn Reson B* 1996;**112**:95.
97. Dötsch V, Wagner G. Editing for amino-acid type in CBCACONH experiments based on the ^{13}Cβ-^{13}Cγ coupling. *J Magn Reson B* 1996;**111**:310.
98. Pellecchia M, Iwai H, Szyperski T, Wüthrich K. The 2D NMR experiments H(C)CO_2 and $HCCO_2$ for assignment and pH titration of carboxylate groups in uniformly ^{15}N/^{13}C-labeled proteins. *J Magn Reson* 1997;**124**:274.
99. Schmieder P, Leidert M, Kelly M, Oschkinat H. Multiplicity-selective coherence transfer steps for the design of amino acid-selective experiments—a triple-resonance experiment selective for ASN and GLN. *J Magn Reson* 1998;**131**:199.
100. Schubert M, Smalla M, Schmieder P, Oschkinat H. MUSIC in triple-resonance experiments: amino acid type-selective ^{1}H-^{15}N correlations. *J Magn Reson* 1999;**141**:34.
101. Bazzo R, Cicero DO, Barbato G. Selective correlations of amide groups to glycine alpha protons and of arginine guanidine groups to delta protons in proteins by multiple quantum spectroscopy. *J Magn Reson* 1999;**136**:15.
102. Schubert M, Ball LJ, Oschkinat H, Schmieder P. Bridging the gap: a set of selective ^{1}H-^{15}N correlations to link sequential neighbors of prolines. *J Biomol NMR* 2000;**17**:331.
103. Schubert M, Oschkinat H, Schmieder P. Amino acid type-selective backbone ^{1}H-^{15}N correlations for Arg and Lys. *J Biomol NMR* 2001;**20**:379.
104. Schubert M, Oschkinat H, Schmieder P. MUSIC, selective pulses, and tuned delays: amino acid-type selective ^{1}H-^{15}N correlations, II. *J Magn Reson* 2001;**148**:61.
105. Schubert M, Oschkinat H, Schmieder P. MUSIC and aromatic residues: amino acid type-selective ^{1}H-^{15}N correlations, III. *J Magn Reson* 2001;**153**:186.

106. Brutscher B. DEPT spectral editing in HCCONH-type experiments. Application to fast protein backbone and side chain assignment. *J Magn Reson* 2004;**167**:178.
107. Schubert M, Labudde D, Leitner D, Oschkinat H, Schmieder P. A modified strategy for sequence specific assignment of protein NMR spectra based on amino acid type selective experiments. *J Biomol NMR* 2005;**31**:115.
108. Lescop E, Rasia R, Brutscher B. Hadamard amino-acid-type edited NMR experiment for fast protein resonance assignment. *J Am Chem Soc* 2008;**130**:5014.
109. Kjaergaard M, Poulsen FM. Disordered proteins studied by chemical shifts. *Prog NMR Spectrosc* 2012;**60**:42.
110. Bermel W, Bertini I, Chill JH, Felli IC, Kumar VMV, Haba N, et al. Aminoacid-types selective ^{13}C direct-detected exclusively heteronuclear experiments to study intrinsically disordered proteins. *Chembiochem* 2012;**13**:2425.
111. Zhang H, Neal S, Wishart DS. RefDB: a database of uniformly referenced protein chemical shifts. *J Biomol NMR* 2003;**25**:173.
112. Shen Y, Delaglio F, Cornilescu G, Bax A. TALOS plus: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *J Biomol NMR* 2009;**44**:213.
113. Cavalli A, Salvatella X, Dobson CM, Vendruscolo M. Protein structure determination from NMR chemical shifts. *Proc Natl Acad Sci USA* 2007;**104**:9615.
114. Bermel W, Bertini I, Felli IC, Peruzzini R, Pierattelli R. Exclusively heteronuclear NMR experiments to obtain structural and dynamic information on proteins. *Chemphyschem* 2010;**11**:689.
115. Pasat G, Zintsmaster JS, Peng J. Direct ^{13}C-detection for carbonyl relaxation studies of protein dynamics. *J Magn Reson* 2008;**193**:226.
116. Balayssac S, Bertini I, Luchinat C, Parigi G, Piccioli M. ^{13}C direct detected NMR increases the detectability of residual dipolar couplings. *J Am Chem Soc* 2006;**128**:15042.
117. Griesinger C, Sørensen OW, Ernst RR. Correlation of connected transitions by two-dimensional NMR spectroscopy. *J Chem Phys* 1986;**85**:6837.
118. Montelione GT, Wagner G. Accurate measurements of homonuclear HN-H^{α} coupling constants in polypeptides using heteronuclear 2D NMR experiments. *J Am Chem Soc* 1989;**111**:5474.
119. Biamonti C, Rios CB, Lyons BA, Montelione GT. Multidimensional NMR experiments and analysis techniques for determining homo-and heteronuclear scalar coupling contents in proteins and nucleic acids. *Adv Biophys Chem* 1994;**4**:51.
120. Wang AC, Bax A. Determination of the backbone dihedral angles φ in human ubiquitin from reparametrized empirical Karplus equations. *J Am Chem Soc* 1996;**118**:2483.
121. Konrat R, Muhandiram DR, Farrow NA, Kay LE. Pulse schemes for the measurement of $^{3}J_{C'C}$ gamma and $^{3}J_{NC}$ gamma scalar couplings in ^{15}N,^{13}C uniformly labeled proteins. *J Biomol NMR* 1997;**9**:409.
122. Bax A, Max D, Zax D. Measurement of long-range ^{13}C-^{13}C J couplings in a 20-kDa protein-peptide complex. *J Am Chem Soc* 1992;**114**:6923.
123. Hu J-S, Bax A. Determination of φ and χ_1 angles in proteins from ^{13}C-^{13}C three-bond *J* couplings measured by three-dimensional heteronuclear NMR. How planar is the peptide bond? *J Am Chem Soc* 1997;**119**:6360.
124. Reif B, Kock M, Kesserbaum R, Schleucher J, Griesinger C. Determination of ^{1}J,^{2}J, and ^{3}J carbon–carbon coupling constants at natural abundance. *J Magn Reson B* 1996;**112**:295.
125. Tjandra N, Grzesiek S, Bax A. Magnetic field dependence of nitrogen-proton J splittings in 15N-enriched human ubiquitin resulting from relaxation interference and residual dipolar coupling. *J Am Chem Soc* 1996;**118**:6264.

126. Yang D, Tolman JR, Goto NK, Kay LE. An HNCO-based pulse scheme for the measurement of $^{13}C^{\alpha}$-$^{1}H^{\alpha}$ one-bond dipolar couplings in ^{15}N,^{13}C labeled proteins. *J Biomol NMR* 1998;**12**:325.
127. Hu K, Vögeli B, Clore GM. Spin-state selective carbon-detected HNCO with TROSY optimization in all dimensions and double echo-antiecho sensitivity enhancement in both indirect dimensions. *J Am Chem Soc* 2007;**129**:5484.
128. Mulder FAA, Akke M. Carbonyl 13C transverse relaxation measurements to sample protein backbone dynamics. *Magn Reson Chem* 2003;**41**:853.
129. Ishima R, Baber JL, Louis JM, Torchia DA. Carbonyl carbon transverse relaxation dispersion measurements and ms-us timescale motion in a protein hydrogen bond network. *J Biomol NMR* 2004;**29**:187.
130. Fermi E. *Z Phys* 1930;**60**:320.
131. McConnell HM. Effect of anisotropic hyperfine interactions on paramagnetic relaxation in liquids. *J Chem Phys* 1956;**25**:709.
132. Solomon I. Relaxation processes in a system of two spins. *Phys Rev* 1955;**99**:559.
133. Bloembergen N, Purcell EM, Pound RV. *Phys Rev* 1948;**73**:679.
134. Bertini I, Jiménez B, Piccioli M. ^{13}C direct detected experiments: optimisation to paramagnetic signals. *J Magn Reson* 2005;**174**:125.
135. Shimba N, Kovacs H, Stern AS, Nomura AM, Shimada I, Hoch JC, et al. Optimization of ^{13}C direct detection NMR methods. *J Biomol NMR* 2005;**30**:175.
136. Babini E, Bertini I, Capozzi F, Felli IC, Lelli M, Luchinat C. Direct carbon detection in paramagnetic metalloproteins to further exploit pseudocontact shift restraints. *J Am Chem Soc* 2004;**126**:10496.
137. Bertini I, Felli IC, Luchinat C, Parigi G, Pierattelli R. Towards a protocol for solution structure determination of copper(II) proteins: the case of Cu(II)Zn(II) superoxide dismutase. *Chembiochem* 2007;**8**:1422.
138. Abriata LA, Ledesma GN, Pierattelli R, Vila AJ. Electronic structure of the ground and excited states of the Cu(A) site by NMR spectroscopy. *J Am Chem Soc* 2009;**131**:1939.
139. Mainz A, Bardiaux B, Kuppler F, Multhaup G, Felli IC, Pierattelli R, et al. Structural and mechanistic implications of metal-binding in the small heat-shock protein αB-crystalline. *J Biol Chem* 2012;**287**:1128.
140. Pochapsky TC, Kostic M, Jain N, Pejchal R. Redox-dependent conformational selection in a $Cys_4Fe_2S_2$ ferredoxin. *Biochemistry* 2001;**40**:5602.
141. Caillet-Saguy C, Delepierre M, Lecroisey A, Bertini I, Piccioli M, Turano P. Direct detected ^{13}C NMR to investigate the iron(III) hemophore HasA. *J Am Chem Soc* 2006;**128**:150.
142. Caillet-Saguy C, Turano P, Piccioli M, Lukat-Rodgers G, Czjzek M, Guigliarelli B, et al. Decifering the structural role of histidine 83 for heme binding in hemophore HasA. *J Biol Chem* 2008;**283**:5960.
143. Westler WM, Lin IJ, Perczel A, Weinhold F, Markley JL. Hyperfine-shifted (13)C resonance assignments in an iron-sulfur protein with quantum chemical verification: aliphatic C-H. . .S 3-center-4-electron interactions. *J Am Chem Soc* 2011;**133**:1310–6.
144. Pochapsky SS, Sunshine J, Pochapsky TC. Completing the circuit: direct-observe 13C,15N double-quantum spectroscopy permits sequential resonance assignments near a paramagnetic center in acireductone dioxygenase. *J Am Chem Soc* 2008;**130**:2156.
145. Bertini I, Donaire A, Jiménez B, Luchinat C, Parigi G, Piccioli M, et al. Paramagnetism-based versus classical constraints: an analysis of the solution structure of Ca Ln calbindin D_{9k}. *J Biomol NMR* 2001;**21**:85.
146. Bertini I, Jiménez B, Piccioli M, Poggi L. Asymmetry in ^{13}C-^{13}C COSY spectra identifies geometry in paramagnetic proteins. *J Am Chem Soc* 2005;**127**:12216.
147. Gelis I, Katsaros N, Luchinat C, Piccioli M, Poggi L. A simple protocol to study blue copper proteins by NMR. *Eur J Biochem* 2003;**270**:600.

148. Balayssac S, Jiménez B, Piccioli M. Assignment strategy for fast relaxing signals: complete aminoacid identification in thulium substituted calbindin D_{9k}. *J Biomol NMR* 2006;**34**:63.
149. Madl T, Felli IC, Bertini I, Sattler M. Structural analysis of protein interfaces from ^{13}C direct-detect paramagnetic relaxation enhancements. *J Am Chem Soc* 2010;**132**:7285.
150. Madl T, Guettler T, Gorlich D, Sattler M. Structural analysis of large protein complexes using solvent paramagnetic relaxation enhancements. *Angew Chem Int Ed Engl* 2011;**50**:3993.
151. Bertini I, Jiménez B, Pierattelli R, Wedd AG, Xiao Z. Protonless ^{13}C direct detection NMR: characterization of the 37 kiloDalton trimeric protein CutA1. *Proteins Struct Funct Genet* 2008;**70**:1196.
152. Poget SF, Girvin ME. Solution NMR of membrane proteins in bilayer mimics: small is beautiful, but sometimes bigger is better. *Biochem Biophys Res Commun* 2007;**1768**:3098.
153. Pervushin K, Riek R, Wider G, Wüthrich K. 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* 1997;**94**:12366.
154. Pervushin K, Riek R, Wider G, Wüthrich K. Transverse relaxation-optimized spectroscopy (TROSY) for NMR studies of aromatic spin systems in ^{13}C-labeled proteins. *J Am Chem Soc* 1998;**120**:6394.
155. Riek R, Pervushin K, Wüthrich K. TROSY and CRINEPT: NMR with large molecular and supramolecular structures in solution. *TIBS* 2000;**25**:462.
156. Tugarinov V, Hwang PM, Ollerenshaw JE, Kay LE. Cross-correlated relaxation enhanced 1H[bond]13C NMR spectroscopy of methyl groups in very high molecular weight proteins and protein complexes. *J Am Chem Soc* 2003;**125**:10420.
157. Ollerenshaw JE, Tugarinov V, Kay LE. *Magn Reson Chem* 2003;**41**:843.
158. Cavanagh J, Fairbrother WJ, Palmer III AG, Skelton NJ. *Protein NMR spectroscopy. Principles and practice*. San Diego: Academic Press; 1996.
159. Pintacuda G, Giraud N, Pierattelli R, Böckmann A, Bertini I, Emsley L. Solid-state NMR of a paramagnetic protein: assignment and study of human dimeric oxidized Cu(II), Zn(II) superoxide dismutase. *Angew Chem Int Ed* 2007;**46**:1079.
160. Balayssac S, Bertini I, Falber K, Fragai M, Jehle S, Lelli M, et al. Solid-state NMR of matrix metalloproteinase 12: an approach complementary to solution NMR. *Chembiochem* 2007;**8**:486.
161. Turano P, Lalli D, Felli IC, Theil EC, Bertini I. NMR reveals a pathway for iron mineral precursors to the central cavity of ferritin. *Proc Natl Acad Sci USA* 2010;**107**:545.
162. Knight MJ, Felli IC, Pierattelli R, Emsley L, Pintacuda G. Magic angle spinning NMR of paramagnetic proteins. *Acc Chem Res* 2013; Epub ahead of print.
163. Bertini I, Luchinat C, Parigi G, Ravera E. SedNMR: on the edge between solution and solid state NMR. *Acc Chem Res* 2013; Epub ahead of print.
164. Bertini I, Luchinat C, Parigi G, Ravera E, Reif B, Turano P. Solid-state NMR of proteins sedimented by ultracentrifugation. *Proc Natl Acad Sci USA* 2011;**108**:10396.
165. Tompa P. Intrinsically unstructured proteins. *Trends Biochem Sci* 2002;**27**:527.
166. Dyson HJ, Wright PE. Unfolded proteins and protein folding studied by NMR. *Chem Rev* 2004;**104**:3607.
167. Mittag T, Forman-Kay J. Atomic-level characterization of disordered protein ensembles. *Curr Opin Struct Biol* 2007;**17**:3.
168. Wright PE, Dyson HJ. Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. *J Mol Biol* 1999;**293**:321.
169. Tompa P. *Structure and function of intrinsically disordered proteins*. Boca Raton, FL: CRC Press; 2009.

170. Uversky V, Dunker AK. The case for intrinsically disordered proteins playing contributory roles in molecular recognition without a stable 3D structure. *F1000 Biol Rep* 2013;**5**:1.
171. Fink AL. Natively unfolded proteins. *Curr Opin Struct Biol* 2005;**15**:35.
172. Tompa P. Intrinsically disordered proteins: a 10-year recap. *Trends Biochem Sci* 2012;**37**:509–16.
173. Tompa P, Fuxreiter M. Fuzzy complexes: polymorphism and structural disorder in protein-protein interactions. *Trends Biochem Sci* 2008;**33**:2.
174. Sickmeier M, Hamilton JA, LeGall T, Vacic V, Cortese MS, Tantos A, et al. DisProt: the database of disordered proteins. *Nucleic Acids Res* 2007;**35**:D786.
175. Dyson HJ, Wright PE. Intrinsically unstructured proteins and their functions. *Nat Rev Mol Cell Biol* 2005;**6**:197.
176. Marsh J, Teichmann SA, Forman-Kay JD. Probing the diverse landscape of protein flexibility and binding. *Curr Opin Struct Biol* 2012;**22**:643.
177. Mittag T, Marsh J, Grishaev A, Orlicky S, Lin H, Sicheri F, et al. Structure/function implications in a dynamic complex of the intrinsically disordered Sic1 with the Cdc4 subunit of an SCF ubiquitin ligase. *Structure* 2010;**18**:494.
178. Felli IC, Pierattelli R, Tompa P. Intrinsically disordered proteins. In: Bertini I, McGreevy KS, Parigi G, editors. *NMR of biomolecules*. Weinheim Germany: Wiley-Blackwell; 2012.
179. Felli IC, Pierattelli R. Recent progress in NMR spectroscopy: towards the study of intrinsically disordered proteins of increasing size and complexity. *IUBMB Life* 2012;**64**:473.
180. Eliezer D. Biophysical characterization of intrinsically disordered proteins. *Curr Opin Struct Biol* 2009;**19**:23.
181. Panchal SC, Bhavesh NS, Hosur RV. Improved 3D triple resonance experiments, HNN and HN(C)N, for H^N and ^{15}N sequential correlations (^{13}C, ^{15}N) labeled proteins: application to unfolded proteins. *J Biomol NMR* 2001;**20**:135.
182. Narayanan RL, Duerr HN, Bilbow S, Biernat J, Mendelkow E, Zweckstetter M. Automatic assignment of the intrinsically disordered protein tau with 441-residues. *J Am Chem Soc* 2010;**132**:11906.
183. Solyom Z, Schwarten M, Geist L, Konrat R, Willbold D, Brutscher B. BEST-TROSY experiments for time-efficient sequential resonance assignment of large disordered proteins. *J Biomol NMR* 2013;**55**:311–21.
184. Motackova V, Novacek J, Zawadzka-Kazimierczuk A, Kazimierczuk K, Zidek L, Sanderová H, et al. Strategy for complete NMR assignment of disordered proteins with highly repetitive sequences based on resolution-enhanced 5D experiments. *J Biomol NMR* 2010;**48**:169.
185. Dyson HJ, Wright PE. Nuclear magnetic resonance methods for the elucidation of structure and dynamics in disordered states. *Meth Enzymol* 2001;**339**:258.
186. Schwarzinger S, Kroon GJA, Foss TR, Chung J, Wright PE, Dyson HJ. Sequence-dependent correction of random coil NMR chemical shifts. *J Am Chem Soc* 2001;**123**:2970.
187. Pérez Y, Gairi M, Pons M, Bernadò P. Structural characterization of the natively unfolded N-terminal domain of human c-Src kinase: insights into the role of phosphorylation of the unique domain. *J Mol Biol* 2009;**391**:136.
188. Knoblich K, Whittaker S, Ludwig C, Michiels P, Jiang T, Schafflhausen B, et al. Backbone assignment of the N-terminal polyomavirus large T antigen. *Biomol NMR Assign* 2009;**3**:119.
189. Binolfi A, Theillet FX, Selenko P. Bacterial in-cell NMR of human alpha-synuclein: a disordered monomer by nature? *Biochem Soc Trans* 2012;**40**:950.

Article 2. NMR methods for the study of IDP structure, dynamics, and interactions: general overview and practical guidelines

Bernhard Brutscher[1], Isabella C. Felli[2], Sergio Gil-Caballero[3], Tomáš Hošek[2], Rainer Kümmerle[3], Alessandro Piai[2], Roberta Pierattelli[2], Zsófia Sólyom[1]

[1] Institut de Biologie Structurale, Université Grenoble 1, CNRS, CEA, 71 avenue des Martyrs, 38044, Grenoble Cedex 9, France.

[2] CERM and Department of Chemistry "Ugo Schiff", University of Florence, 50019, Via Luigi Sacconi 6, Sesto Fiorentino, Florence, Italy.

[3] Bruker BioSpin AG, Industriestrasse 26, 8117, Fällanden, Switzerland.

Published in 2015 in Adv.Exp.Med.Biol., 870, 49-122

Chapter 3
NMR Methods for the Study of Instrinsically Disordered Proteins Structure, Dynamics, and Interactions: General Overview and Practical Guidelines

Bernhard Brutscher, Isabella C. Felli, Sergio Gil-Caballero, Tomáš Hošek, Rainer Kümmerle, Alessandro Piai, Roberta Pierattelli and Zsófia Sólyom

Abstract Thanks to recent improvements in NMR instrumentation, pulse sequence design, and sample preparation, a panoply of new NMR tools has become available for atomic resolution characterization of intrinsically disordered proteins (IDPs) that are optimized for the particular chemical and spectroscopic properties of these molecules. A wide range of NMR observables can now be measured on increasingly complex IDPs that report on their structural and dynamic properties in isolation, as part of a larger complex, or even inside an entire living cell. Herein we present basic NMR concepts, as well as optimised tools available for the study of IDPs in solution. In particular, the following sections are discussed hereafter: a short introduction to NMR spectroscopy and instrumentation (Sect. 3.1), the effect of order and disorder on NMR observables (Sect. 3.2), particular challenges and bottlenecks for NMR studies of IDPs (Sect. 3.3), 2D HN and CON NMR experiments: the fingerprint of an IDP (Sect. 3.4), tools for overcoming major bottlenecks of IDP NMR studies (Sect. 3.5), ^{13}C detected experiments (Sect. 3.6), from 2D to 3D: from simple snapshots to site-resolved characterization of IDPs (Sect. 3.7), sequential NMR assignment: 3D experiments (Sect. 3.8), high-dimensional NMR experiments (nD, with $n > 3$) (Sect. 3.9) and conclusions and perspectives (Sect. 3.10).

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

I. C. Felli (✉)· T. Hošek · A. Piai · R. Pierattelli (✉)
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

R. Pierattelli
e-mail: pierattelli@cerm.unifi.it

R. Kümmerle · S. Gil-Caballero
Bruker BioSpin AG, Industriestrasse 26, 8117 Fällanden, Switzerland

I. C. Felli, R. Pierattelli (eds.), *Intrinsically Disordered Proteins Studied by NMR Spectroscopy,* Advances in Experimental Medicine and Biology,
DOI 10.1007/978-3-319-20164-1_3

Keywords NMR basics · pulse sequences · NMR instrumentation · Sequential assignment · BEST · ^{13}C detection · high dimensional NMR

1 A Short Introduction to NMR Spectroscopy and Instrumentation

The possibilities offered by modern nuclear magnetic resonance (NMR) spectroscopy are enormous and cover a wide range of applications in physics, chemistry, biology, medicine and material sciences. Past and present progress in the field is mainly based on technical improvements of the NMR spectrometer (magnetic field strength, NMR electronics, probes, etc.) and on the development of a variety of pulse sequences that exploit the basic principles of NMR spectroscopy in an ingenious way in order to obtain the desired information on a particular sample. The first part of this chapter will therefore be dedicated to the most recent key developments in NMR instrumentation, mainly stimulated by emerging scientific challenges, and to the basic principles of NMR spectroscopy.

The phenomenon of nuclear magnetic resonance was discovered by Purcell et al. (1946) and Bloch (1946); 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"*. Since its discovery, NMR spectroscopy, further to its fundamental role in physics, has become 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, NMR is the only method capable of providing atomic resolution information on the structure of biological macromolecules. Compared to X-ray crystallography, it allows studies of biological macromolecules in the liquid state—meaning that crystallization of the molecule is not required. This makes structural studies of highly dynamic systems such as intrinsically disordered proteins (IDPs) possible. NMR spectroscopy can also be employed for studies of interactions of biomolecules with small ligands, metal ions and other biological macromolecules. Recently, the importance of regarding biomolecules as dynamic ensembles instead of a single static entity has become widely recognized. NMR spectroscopy provides a unique tool to access dynamics information at atomic resolution, from the picosecond timescale to slow exchange processes on the second (or even slower) timescale. This is particularly important for intrinsically disordered proteins, which are often characterized by heterogeneous dynamic properties along the polypeptide chain.

1.1 The Basic Principles of NMR

NMR spectroscopy studies the interaction of matter with radiofrequency electromagnetic waves that excite magnetic transitions of the atomic nuclear spins. Indeed, an atomic nucleus, in order to be observable by NMR spectroscopy, has to possess

Table 3.1 Gyromagnetic ratios and natural abundance of nuclei important for NMR studies of proteins

Isotope	Spin I	Gyromagnetic ratio γ_n (10^6 rad s^{-1} T^{-1})	Gyromagnetic ratio/2π $\gamma_n/2\pi$ (MHz T^{-1})	Natural abundance isotope (%)
^{1}H	½	267.513	42.576	99.98
^{13}C	½	67.262	10.705	1.108
^{15}N	½	−27.116	−4.316	0.37

a nonzero spin quantum number *I*. The relevant nuclear isotopes for NMR studies of proteins are ^{1}H, ^{13}C, and ^{15}N; all of them have a spin quantum number of $I=½$. When the nuclei are placed in a static magnetic field B_0, which is by convention aligned along the *z*-axis, the nuclear magnetic momentum $\boldsymbol{\mu}$ of the spins interact with the magnetic field leading to a splitting of the energy levels (Zeeman splitting) according to Eq. 3.1

$$E = -\boldsymbol{\mu}\cdot\boldsymbol{B} = -\mu_Z B_0 = -\gamma I_Z B_0 = -\gamma m\hbar B_0 \tag{3.1}$$

where γ is the gyromagnetic ratio and $\hbar$ is the reduced Planck constant. For spin ½ nuclei the magnetic quantum number *m* can take the values of $m=+½$ and $m=-½$. The gyromagnetic ratios and natural abundance of the nuclear isotopes important in biomolecular NMR are summarized in Table 3.1.

Equation 3.1 is of the utmost importance for the sensitivity of NMR experiments. The energy difference between the two Zeeman energy levels is given by:

$$\Delta E = -\gamma\hbar B_0 = \hbar\omega_0 \tag{3.2}$$

where ω_0 is the characteristic NMR frequency (Larmor frequency) of a nuclear spin at a given magnetic field strength.

At thermal equilibrium and temperature *T,* the ratio of nuclear spins in the lower (E_α) and higher (E_β) energy state can be calculated from the Boltzmann distribution:

$$\frac{N_\beta}{N_\alpha} = e^{-\frac{\Delta E}{kT}} \tag{3.3}$$

where *k* is the Boltzmann constant.

The NMR signal is proportional to the magnetization, which is in turn dependent on the spin polarization *P* given by the population difference between the two states divided by the total number of spins:

$$P = \frac{N_\alpha - N_\beta}{N_\alpha + N_\beta} \tag{3.4}$$

In the case of ^{1}H nuclei, which possess the highest gyromagnetic ratio among the spin ½ nuclei in proteins, the ratio of spins in the upper energy state versus the lower energy state is 0.999872 at room temperature on an 800 MHz magnet ($B_0=18.8$ T). This means that only a very small fraction of the spins present in the sample contributes

to the observed NMR signal. Furthermore, in the case of ^{13}C and ^{15}N nuclei (often called heteronuclei), the natural abundance of the NMR active nuclei is very low, as shown in Table 3.1. Isotope enrichment techniques have therefore been developed to enhance the sensitivity of NMR techniques involving such heteronuclei.

Sensitivity and resolution are the two most important factors influencing the outcome of NMR experiments. NMR sensitivity, which is defined as the signal to noise ratio (SNR) obtained in a fixed amount of time, can be described by the following equation:

$$SNR \propto \frac{N}{V} \gamma_{exc} \gamma_{det}^{2.5} B_0^{2.5} n_{scan}^{0.5} \frac{1}{\sqrt{R_s(T_a + T_s) + R_c(T_a + T_c)}} \tag{3.5}$$

where N is the number of spins, V the active volume, γ_{exc} the gyromagnetic ratio of the excited nuclei and γ_{det} that of the detected nuclei, B_0 the static magnetic field, n_{scan} the number of scans (experimental repetitions), R_S the resistance of the sample and R_c that of the coil, T_a the temperature of the preamplifier and T_s and T_c those of the sample and of the coil, respectively (Hoult and Richards 1976; Kovacs et al. 2005). Other factors that depend on the properties of the sample under investigation and on the type of pulse sequence used contribute to the sensitivity of an NMR experiment, and therefore to the possibility of accessing the desired information. These will be discussed in detail in the following paragraphs.

Some general conclusions for improving the sensitivity of an NMR experiment can be derived from Eq. 3.5. The NMR signal is proportional to the amount of spins present in the sample; it is therefore desirable to use highly concentrated samples for NMR spectroscopy. However, in practice, the maximum concentration of a protein sample is often limited by protein solubility and obtaining large quantities of isotopically labelled proteins can be expensive and time consuming. Most protein NMR experiments use proton excitation and detection because of the high gyromagnetic ratio of these nuclei; nevertheless, ^{13}C direct detection also provides a valuable tool for biomolecular NMR applications (see Sect. 3.6). The linear dependence on B_0 explains the on-going efforts of NMR manufacturers to develop magnets with higher magnetic fields. To date, the highest magnetic field of a commercial magnet corresponds to 1 GHz proton Larmor frequency (23.5 T), and 1.2 GHz magnets (28.2 T) are under development. The SNR increases with the square root of the number of scans. Increasing the overall measurement time thus represents a common method for spectral improvement. The last contribution to NMR sensitivity comes from the electronic detection circuit. The SNR of NMR increases with a lowering of the temperature of the preamplifier (T_a), the sample (T_s), and the coil (T_c), as well as the resistance of the sample (R_s) and the coil (R_c). A recent development that greatly enhanced the sensitivity of NMR spectroscopy was the introduction of probes with cryogenically cooled (to about 20 K) detection circuits. However, protein samples with high ionic strengths degrade the beneficial effects of cryogenically cooled probes.

Resolution is the other important factor to extract atomic resolution information from NMR spectra. It depends on chemical shift dispersion and signal linewidths. Two peaks can be resolved when their difference in frequency is larger with respect to their linewidths. Nuclear spins that are characterized by favourable chemical

shift dispersion and small linewidths for the system investigated should thus be exploited to improve the resolution of the spectra. The possibility of introducing additional indirect dimensions in NMR experiments provides the other invaluable tool for enhancing the resolution by spreading signals in additional dimensions and thus reducing the possibility of accidental cross-peak overlap.

1.2 NMR Instrumentation and Recent Improvements

NMR spectrometers are composed of three main components, plus some optional accessories, such as automatic sample changers or variable temperature (VT) gas preconditioning devices:

1. The superconducting magnet generating the static magnetic field B_0, with the superconducting coil immersed in liquid helium at a temperature of 4.2 K (or about 2 K for subcooled "pumped" magnets). Commercially available superconducting NMR magnets have a freely accessible vertical bore with a diameter of 50–54 mm (standard bore, used mainly for liquid state and solid state applications), 89 mm (wide bore, used mainly for solid state and micro-imaging applications) or 154 mm (super wide bore, mostly used for micro-imaging applications). The temperature inside the magnet bore can be adjusted independently.
2. The probe, positioned in the bore of the superconducting magnet, is equipped with radiofrequency (RF) coils/antennas emitting RF pulses and detecting with the same RF coils the voltage induced by the precessing nuclear spin magnetization in the NMR sample. One can distinguish between three major fields of NMR applications, requiring different types of probes: imaging (or micro-imaging), magic-angle spinning (MAS) solid-state applications, and high-resolution liquid-state applications. In the following, we will focus on NMR probes for liquid-state applications. The standard sample tube diameter is 5 mm, but optimized probes are available for sample diameters in the range 1 to 10 mm. Modern probes are equipped with temperature control, a field lock channel (e.g. ^{2}H), an actively shielded coil enabling pulsed field gradients of ~5 mT along one or several axis by applying DC currents (~10 A), and the capability to optimize the resonance circuit tuning and matching via software control. Most liquid state probes are built with two distinct RF coils to accommodate 2, 3 or 4 frequency circuits besides the field lock. Consequently, they are called double, triple or quadruple resonance probes. A further distinction or classification can be done depending on whether the inner RF coil (closer to the sample) is tuned to ^{1}H ("inverse probe", TXI) or to a X nucleus ("observe probe", TXO). Historically, the inner coil of probes was tuned to observe the less sensitive X nuclei. With the introduction of pulse sequence elements like INEPT, the advantage of detecting X nuclei via the more sensitive ^{1}H nucleus has led to the development of inverse probes that provide increased sensitivity.
3. The spectrometer console has all the electronic components required for executing complex multi-channel, multi-nuclear pulse experiments, as well as for detection of the electronic signal induced in the RF coils of the probe. In particular, RF pulses are generated at the Larmor frequencies of the nuclear spins

that need to be manipulated during the experiment. The console ensures their correct timing, phase and amplitude modulation, and signal amplification. The spectrometer electronics is typically characterized by the number of independent radiofrequency (RF) channels, e.g. the number of different nuclei being addressable within one NMR experiment. A further distinction is the peak power delivered by the RF amplifier, the requirements strongly depending on the probe used for the application. For liquid state NMR, high-band frequency (^{3}H, ^{1}H, ^{19}F) excitation typically needs less than 50 W, whereas for low γ nuclei (often also called X nuclei) RF pulses may require more than 500 W at high B_0 fields. The console also hosts an equipment, called spectrometer or field lock, that serves to maintain the magnetic field strength seen by the spins throughout the whole experiment duration, resulting in field variations of less than ~0.1 Hz (^{1}H Larmor frequency). This corresponds to a precision of ~10^{-10} of the static B_0 field, maintained for up to several days (in the case of long biomolecular experiments). The B_0 field is homogenized over the sample volume by means of a shim unit that generates additional small magnetic fields (so-called shims), again resulting in spatial B_0 field variation of less than 10^{-9}. Last but not least, a variable temperature unit serves to measure and regulate the NMR sample temperature throughout the experiment.

Over the past years all these spectrometer components have seen major technological improvements, so that besides ease of use, both performance and stability of modern NMR spectrometers have been significantly improved.

Improved Magnet Technology. Using NbTi and Nb_3Sn as superconducting material, combined with state-of-the art wire technology, the highest currently achievable magnetic field strength is 23.5 T, equivalent to a ^{1}H Larmor frequency of 1000 MHz. Increasing the magnetic field strength further will require the use of new types of superconducting wire, so-called high temperature superconducting tapes presenting substantially higher critical currents I_C compared to conventional low temperature superconductors in the presence of magnetic field. The first next generation magnets with a magnetic field strength of ~1.2 GHz are expected to become available within the next 3 years. For magnetic field strengths up to 21.1 T (900 MHz) incremental improvements of low temperature superconducting wire have been made over the past years to design more compact and better shielded magnets. These compact magnets have, at identical field strength, smaller dimensions and less weight with substantially reduced stray fields, thus making the laboratory and on-site space requirements for an NMR spectrometer less critical and less costly. As an example, the weight of an 18.8 T magnet (800 MHz) could be cut in half with two design steps over the last decade.

In addition, modern NMR magnets have a significantly reduced consumption of liquid helium. This not only translates into reduced operational costs, but increased liquid helium hold times may become critical during periods of helium shortage, as experienced at the end of 2012/beginning of 2013. Recently also “cryogen-free” magnets became available, where an active refrigeration technology, mainly consisting in two pulse tubes, allows to re-liquefy evaporating helium gas (zero boil-off), and to completely avoid the outer liquid nitrogen dewar. Such “cryogen-free” magnets are currently available for magnetic resonance imaging (MRI) systems,

and for gyrotron magnets of solids DNP NMR spectrometers. However, they are not (yet) suitable for high-resolution NMR applications as required for the study of IDPs.

Improved Probe Technology. Experimental sensitivity at constant mass or constant sample concentration has always been a critical factor for NMR probe design. A first possibility to improve the probe performance is to enhance the efficiency of the RF coil, e.g. to achieve higher B_1 fields in the sample with a given coil current I (B_1/I). This can be achieved for smaller sample geometries, so that for example the mass sensitivity of a 1 mm probe is roughly 4-fold higher compared to a 5 mm inverse probe. A second possibility to enhance sensitivity for a given probe type is to reduce the contribution of the thermal noise originating from the RF coil, RF resonance circuit as well as the pre-amplification of the observed NMR signal (CPTXI, CPTXO). Cooling these parts down from room temperature to roughly −200 °C (using liquid nitrogen as cooling medium) leads to sensitivity enhancements of a factor of 2 to 3. Using helium gas (instead of nitrogen) to cool the pre-amplification stages allows to further decrease the temperature to roughly −260 °C, resulting in sensitivity gains of a factor 4–5 compared to an equivalent conventional probe. Historically, such helium cooled probes were actually available before the nitrogen-cooled ones. A drawback of these helium-cooled cryoprobes is that they require a more complex and costly infrastructure (helium compressor, cooling of the helium compressor) compared to the liquid nitrogen cooled probe, resulting in higher running costs. Also note that these maximum sensitivity gains stated above for cryogenically cooled probes only apply to samples of low electronic conductivity. For salty samples the gain in sensitivity is less. However, the situation may be improved by using NMR sample tubes of smaller diameter or of optimized geometry.

The impressive improvements in sensitivity achieved with cryogenic technology have also enabled the design of probes optimized for ^{13}C direct detection, stimulating the design of a variety of new experiments optimized for IDPs, as described later in this chapter.

Improved NMR Electronics. Modern console electronics typically offer fully digital RF frequency generation with a timing resolution of a few tens of nanoseconds to enable the user fast switching of frequency, amplitude and phase as required for modern NMR experiments. The timing as well as the amplitude of the delivered RF and field gradient pulses need to be stable over the full length of the NMR experiment, e.g. up to several days. Another critical parameter for high-resolution NMR spectroscopy is the uniformity and stability of the NMR sample temperature. Most NMR samples are positioned in a stream of temperature-controlled gas, and the sensors are placed outside the NMR sample volume in the gas flow. This setup allows a high level of temperature stability and uniformity over the sample volume. However, the RF pulses in the NMR pulse sequence may heat up the sample, an effect that is not detected by the external temperature sensors. As a result the actual sample temperature may deviate by several degrees from the temperature measured close to the sample. This problem is most critical for high frequency nuclei. It can be reduced by optimizing the RF probe coil design to limit the presence of electrical fields in the sample area, or by reducing sample conductivity if possible. An alternative is the measurement of the actual temperature during an experiment from

the ^{2}H resonance of spins within the NMR sample. Such an "NMR thermometer" requires the presence of two ^{2}H spin species with temperature-dependent chemical shifts. The measurement of the chemical shift difference of the two ^{2}H signals, via the lock channel, allows to regulate permanently the temperature within the NMR sample and to correct for RF heating. A major advantage of this direct measurement of the sample temperature via the nuclear spins is the possibility of recording data on different NMR spectrometers (e.g. field-dependent relaxation studies) at identical absolute temperature.

2 The Effect of Order and Disorder on NMR Observables

In the following sections we will introduce the main NMR observables that are used to derive structural and dynamic information on proteins, with a particular emphasis on how the particular features of IDPs affect these NMR observables.

2.1 Structured Versus Disordered Proteins

In order to better understand the impact of structural order and disorder on NMR experiments and spectral parameters, it is important to recognize the principal features that distinguish a well-folded globular protein from a largely unfolded, highly flexible IDP. These are illustrated in Fig. 3.1.

Hydrophobic interactions and the formation of hydrogen bonds are the main driving forces for the folding process of a polypeptide chain into a globular 3D structure and for the stabilization of such a compact fold. The main consequences of the presence of a stable and well-defined structure that are of importance for the NMR properties of the molecule are a high proton density, the exclusion (to a large extent) of water from the interior of the protein and, to a first crude approximation, the possibility of describing its rotational motion with an overall rotational correlation time. The high proton density and large rotational correlation time make ^{1}H-^{1}H nuclear Overhauser effects (NOEs) the main source of structural information for globular proteins, whereas they are only of little use in IDP research, where the effect is mainly restricted to proton pairs that are close in the primary sequence and is on average smaller than that for globular proteins. Indeed, the dynamic behaviour of IDPs is significantly different from that of structured proteins as a result of the small energy difference between conformers that is responsible for the high flexibility typical of IDPs and allows easy inter-conversion between many different conformations. Therefore, even to a first approximation, a single overall rotational correlation time cannot be defined for IDPs and local effective correlation times are on average smaller with respect to those used to describe structured proteins of similar size. The differences in ^{1}H density and molecular dynamics between IDPs and globular proteins also have a strong influence on the NMR signal chemical shifts and the relaxation parameters. Conformational averaging drastically reduces contributions to chemical shifts deriving from the local environment and causes severe resonance

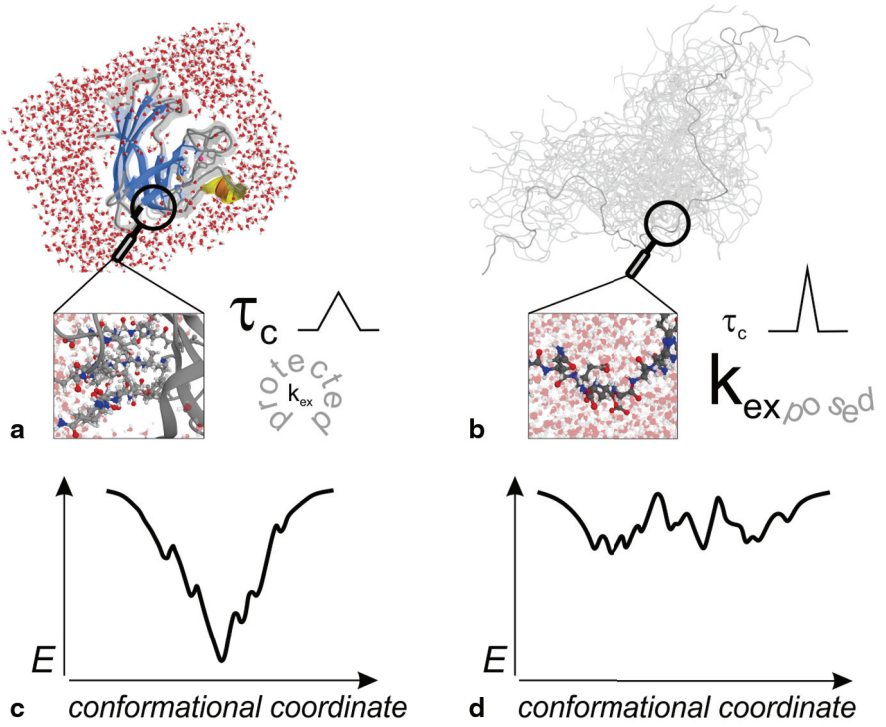


Fig. 3.1 Schematic illustrations of differences between structured (**a**) and intrinsically disordered (**b**) proteins (for reasons of clarity the molecules of water were omitted in the ensemble of IDP conformers), and energy landscapes of well-folded proteins (**c**) and IDPs (**d**) in the native state

overlap. The high flexibility of IDPs has a strong impact on nuclear relaxation rates, and thus on NMR linewidths. Depending on the time scale of motions, several situations can be encountered: on one hand, conformational exchange processes can result in extensive line broadening, and in the extreme case, the absence of any detected NMR signal. On the other hand fast motions result in narrow NMR lines, a property that makes highly flexible IDPs particularly amenable to NMR characterization. A variety of complex NMR experiments can be conceived for such highly flexible IDPs, where the favourable relaxation properties allow for multiple transfer steps and the narrow linewidths contribute to increase the resolution in the resulting spectra. Finally, IDPs are also characterized by non-compact conformations and the exposure of labile protons, e.g. amide protons, to the solvent results in high exchange rates. In fact, the measurement of solvent exchange rates is used to distinguish between highly structured and disordered regions of proteins by identifying solvent exposed and solvent protected amide sites (Schanda et al. 2006a). Approaching physiological pH and temperature, chemical exchange of solvent exposed amide protons may broaden resonances beyond detection and in this case alternative nuclear spins, such as aliphatic ^{1}H or ^{13}C, should be detected to access information on the IDP through NMR (Gil et al. 2013).

2.2 NMR Peak Positions and Chemical Shifts

Peak positions are dependent on the resonance frequencies of the observed nuclei. Based on Eq. 3.2 one would expect to observe a single line in the NMR spectrum for each nucleus of a given type (^{1}H, ^{13}C, ^{15}N) at the respective Larmor frequency. However, this would be true only in the hypothetical experiment of observing "naked" nuclei. In practice, nuclei of the same type resonate at slightly distinct frequencies if they are in chemically different environments within the molecule. The reason is that the nuclei experience a net magnetic field B that is the sum of the static field B_0 and secondary shielding fields induced by the local electronic environment:

$$B = (1-\sigma)B_0 \tag{3.6}$$

where σ is the isotropic average shielding factor. This is at the origin of the term "chemical shift" used to measure resonance positions in NMR spectroscopy. Chemical shift values δ are typically measured relative to the chemical shift of a standard according to:

$$\delta = \frac{\Omega - \Omega_{ref}}{\omega_0} \cdot 10^6 \tag{3.7}$$

where ω_0 is the Larmor frequency in MHz, Ω the frequency offset of the nucleus of interest in Hz and Ω_{ref} the offset of the standard in Hz. In this way the chemical shift becomes a field-independent quantity, expressed in parts per million (ppm). The chemical shift makes NMR an atomic resolution technique: once the resonances are assigned to their respective nuclei, their response to further manipulations can be followed.

Chemical shifts can be very characteristic for nuclei in different chemical moieties such as the different amino acids constituting a protein, and they provide the first useful information that can be obtained from the NMR spectrum. The stable local three-dimensional structure, or partially populated secondary structural conformations, create a unique local electronic environment and thus the contribution to chemical shift coming from isotropic shielding is different for each nuclear spin. In IDPs, the lack of a stable structure results in averaging of a large part of the contributions to the chemical shift coming from the local chemical environment, which is the reason for the low chemical shift dispersion observed in the NMR spectra (Fig. 3.2).

In addition, peak intensities (areas or volumes) can be easily determined and may provide useful atomic resolution information. Peak intensities in the NMR spectra are proportional to the number of nuclear spins giving rise to the peaks. They also report on differential signal loss during an NMR experiment due to relaxation processes. They therefore provide a first useful indication of the heterogeneous structural and dynamic properties of a protein. Peak linewidths would also provide useful information but they are more difficult to measure as they are influenced by different factors (see Sect. 3.2.5). In many cases, for simplicity, the determination of peak heights is used.

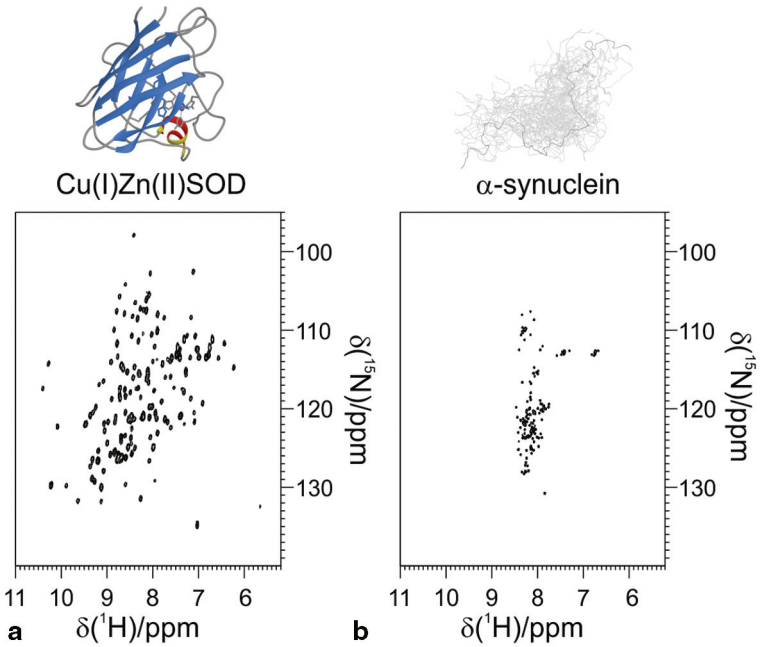


Fig. 3.2 The different chemical shift dispersion of the structured and intrinsically disordered proteins is demonstrated by 2D ^{1}H-^{15}N correlation spectra acquired on two proteins of similar size, but characterized by different structural properties. **a** The HSQC spectrum of structured monomeric Cu(I)Zn(II) superoxide dismutase (1.5 mM sample in 20 mM phosphate buffer, pH 5.0, at 298 K; PDB code: 1BA9). **b** The HSQC spectrum of intrinsically disordered α-synuclein (1.0 mM sample in 20 mM phosphate, pH 6.4, 0.5 mM EDTA, 200 mM NaCl, at 285.5 K). The experiments were acquired on a 700 MHz Bruker AVANCE spectrometer equipped with a CPTXI probe

2.3 Secondary Chemical Shifts

The chemical shifts of backbone nuclei are sensitive to the local backbone geometry, and therefore provide useful information on the occurrence and propensity of secondary structural elements along the protein backbone (Spera and Bax 1991; Wishart et al. 1991). For a ^{13}C, ^{15}N enriched protein, a large number of chemical shifts ($^{1}H^{N}$, ^{15}N, $^{13}C'$, $^{13}C^{\alpha}$, $^{13}C^{\beta}$) can be measured, all of which are reporters of secondary structure even if to a different extent. In order to extract this information, a so-called secondary chemical shift is computed as the difference of the measured chemical shift and a predicted random coil value for each nucleus (Wishart et al. 1992; Wishart and Sykes 1994; Schwarzinger et al. 2001; Tamiola et al. 2010; Kjaergaard et al. 2011; Kjaergaard and Poulsen 2012). Random coil chemical shifts are the theoretical chemical shifts of a polypeptide of the same amino acid sequence characterized by lack of long-range order and secondary structure. However, the conformational sampling of a polypeptide is never completely random, in the sense that all dihedral angles are sampled with equal probability because of

steric hindering and chemical interactions between neighbouring side-chains. The distribution of the sampled dihedral angles along the protein backbone is determined by the conformational Gibbs free energy, which in turn depends on temperature and solvent effects. The accuracy of secondary chemical shifts thus depends on the quality of the predicted random coil chemical shift values. This aspect becomes even more crucial for IDPs, since the measured chemical shifts are usually close to the corresponding random coil values. Figure 3.3 illustrates how C^{α} and C^{β} secondary chemical shifts can be used to identify α-helical and β-sheet regions and highly unstructured segments in globular proteins, as well as residual secondary structures in IDPs.

Different tables of ^{13}C and ^{15}N random coil values for different amino acids, taking into account next and previous neighbours, pH and solvent effects can be found in the literature (Schwarzinger et al. 2001; Zhang et al. 2003; Tamiola et al. 2010; Kjaergaard et al. 2011; Kjaergaard and Poulsen 2012). One of the most commonly used random coil chemical shift data sets, based on chemical shift measurements on small poly-glycine peptides, is the one reported by Wishart et al. (Wishart et al. 1995). Recently, temperature and pH correction factors have been introduced from a peptide-based study employing glutamine peptides, as their conformational sampling is considered to be more representative (Kjaergaard et al. 2011). The other

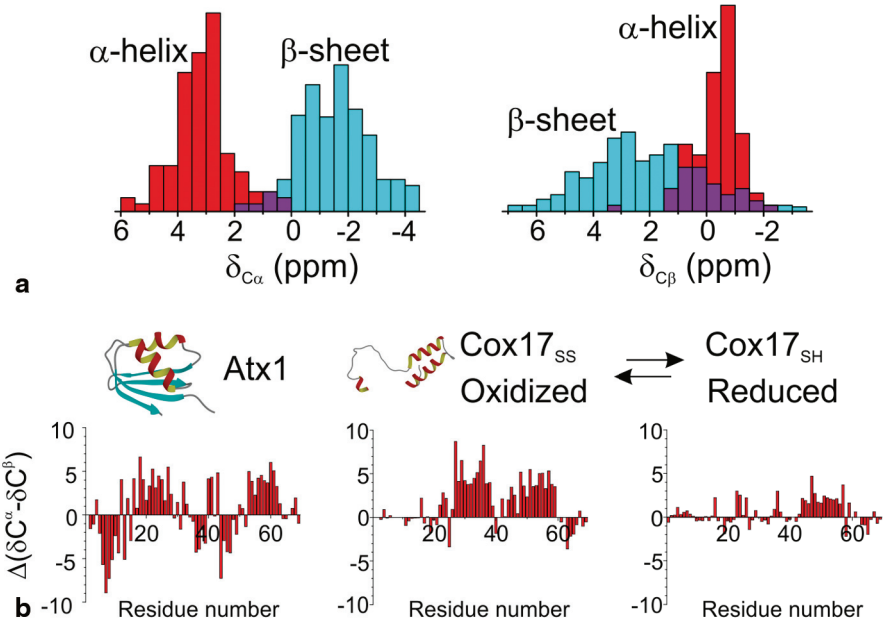


Fig. 3.3 **a** Histograms of the C^{α} and C^{β} secondary chemical shift distribution in an α-helix and a β-sheet (adapted from Spera and Bax 1991). **b** The identification of the secondary structure elements based on the secondary chemical shifts is shown on the examples of the structured metallochaperone Atx1 (Arnesano et al. 2001), the partially structured oxidized form of the copper chaperone Cox17 (Arnesano et al. 2005) and of its unfolded reduced form (Bertini et al. 2011a). (Adapted from Felli et al. 2012)

type of reference data can be obtained from protein chemical shift databases. This approach has the drawback that sample conditions such as pH and temperature vary between different entries. Furthermore, neighbouring and other chemical effects will be biased depending on the composition of the entries. The most commonly used database is refDB by Zhang et al. (Zhang et al. 2003). An IDP-based random coil chemical shift database has recently been introduced (Tamiola et al. 2010). Neighbour 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 shifts (Wishart et al. 1995; Schwarzinger et al. 2001; Kjaergaard and Poulsen 2011).

2.4 Line Splittings and Spin Coupling Constants

Nuclear spins are never completely isolated; mutual interactions between neighbouring nuclear spins are therefore always present and give rise to couplings. The coupling derives from two main contributions: scalar (through bonds) and dipolar (through space). While the latter is averaged out in isotropic solutions, the former, mediated by the electrons in covalent bonds (scalar or ^{n}J-coupling, where n indicates the number of covalent bonds separating the two nuclei involved in the coupling), causes splitting of the NMR signal. The frequency differences between the multiplet components of an NMR signal measured in Hz then reflect the strength of the scalar couplings. Scalar couplings involving ^{1}H, ^{13}C and ^{15}N are exploited for coherence transfer in the majority of multidimensional NMR experiments discussed in the following paragraphs. Scalar couplings, in particular ^{3}J-couplings in a polypeptide, also contain valuable structural information as they depend on the intervening dihedral angle according to the Karplus relation (Karplus 1959). Inter-conversion between different conformers typical of IDPs results in averaged scalar couplings; while this has little impact on coherence transfer in multidimensional NMR experiments, it does influence the interpretation of ^{3}J-couplings in terms of intervening dihedral angles. For example, $^{3}J_{\mathrm{HH}}$ couplings are sometimes used to confirm the presence of partially populated secondary structural elements (Billeter et al. 1992; Vuister and Bax 1993; Case 2000; Otten et al. 2009).

The magnetic moments of two nuclear spins interact “through space” via the dipolar mechanism with an interaction strength that is inversely proportional to r^{3}, with r being the inter-nuclear distance. The dipolar interaction has a particular angular dependence with respect to the static magnetic B_{0} field and, as a consequence, in isotropic solutions where all orientations are sampled with equivalent probability this interaction averages to zero and does not give rise to line splittings. However, it is possible to reintroduce the dipolar coupling by either exploiting the natural magnetic anisotropy of a molecule, resulting in slightly unequal sampling of the molecular orientations in a strong static magnetic field due to partial molecular alignment (Tolman et al. 1995; Tjandra et al. 1996; Banci et al. 1998), or by dissolving the protein in an anisotropic “alignment” medium, e.g. a liquid crystalline solution (Bax and Grishaev 2005). This allows residual dipolar coupling (RDC) to

be measured as a line splitting, similarly to what is done for the measurement of *J*-coupling constants. In fact, the measured couplings in a partially aligned sample are the sum of the scalar and residual dipolar couplings. RDCs provide important information on local and global structure and dynamics in the molecule as explained in more detail in Chaps. 4 and 5.

2.5 NMR Spin Relaxation, Line Widths and Intensities

Relaxation of nuclear spins after excitation by a radio-frequency is caused by time-dependent local magnetic fields induced by the molecular tumbling and local rotational fluctuations that modulate the anisotropic spin interactions, the chemical shielding anisotropy (CSA) and the dipolar (DD) interaction. Additional contributions to spin relaxation arise from exchange processes, e.g. chemical exchange between labile protein and solvent protons, or exchange between different molecular conformations. In a simplified representation of nuclear spin relaxation, the relaxation process is described by two time constants. The longitudinal relaxation time constant T_1 accounts for the return of the spin system to thermal equilibrium associated with a loss of energy, while the transverse relaxation time constant T_2 describes the dephasing of coherence. A main consequence of spin relaxation for a particular NMR experiment is the loss of signal during the various transfer steps and chemical shift evolution delays. In addition, the longitudinal relaxation properties of the excited spin species determine the rate at which a pulse sequence can be repeated, as will be explained in more detail in Sect. 3.5.5, while transverse relaxation properties of nuclear spins determine their NMR linewidth.

The NMR linewidth is given by the transverse relaxation rate R_2 (reciprocal of T_2) of the detected nuclear spin, $\Delta\nu = R_2/\pi$, plus inhomogeneous contributions arising from the experimental setup (sample heterogeneities, B_0 inhomogeneity, temperature gradients, etc.). IDPs, in particular highly flexible ones, are generally characterized by large transverse relaxation time constants T_2, which lead to narrow peaks in the NMR spectra in comparison to the ones observed for globular proteins of comparable size. On the other hand, chemical and conformational exchange processes can be highly pronounced in IDPs, causing broadening of the peaks, especially in the spectra exploiting labile amide protons or in the case of structurally and dynamically heterogeneous proteins. Comparisons of cross-peak intensities (areas or volumes) thus report on differential intensity loss during the pulse sequence, while comparisons of cross-peak heights also report on differences in linewidth. The NMR analysis software often determines the peak heights in the spectrum during the peak picking procedure. When the peak is picked the program fits a function, such as a Lorentzian or Gaussian, and the maximum gives the peak height. For IDPs, peak heights in simple 2D spectra like the 2D HN and 2D CON (see Sect. 3.4.2) are often heterogeneous because relaxation properties of the spins vary depending on the extent of transient structure and on differences in local mobility. Therefore they provide a first indication of the different structural and dynamic properties of the protein.

Relaxation effects, including auto-correlated as well as cross-correlated relaxation rates, can in many cases be accurately quantified and used to extract structural

and dynamic information. In this chapter we mainly focus on the determination of ^{15}N relaxation rates, one of the major tools to obtain information on the local dynamics of different parts of protein backbones (Barbato et al. 1992; Peng and Wagner 1994). Other interesting relaxation rates that can be determined and exploited to achieve structural and dynamic information on IDPs are instead discussed in Chap. 5 (paramagnetic relaxation enhancements, cross correlation rates). The determination of ^{1}H-^{1}H NOEs is instead not discussed in detail as it plays a minor role in the study of IDPs and many excellent books and reviews are available (Neuhaus and Williamson 1989; Cavanagh et al. 2007). However, the major contributions to ^{1}H relaxation are discussed in Sect. 3.5.5 because they are at the basis of the longitudinal relaxation enhancement effects that provide a valuable tool for the design of NMR experiments.

2.6 ^{15}N Relaxation Parameters

The quantification of nuclear spin relaxation effects provides a valuable tool to characterize local and global molecular motions. ^{15}N relaxation values, notably T_1, T_2 and the ^{1}H-^{15}N heteronuclear NOE (HETNOE), are the most commonly determined to characterize the dynamic behaviour of proteins, for both globular proteins and IDPs (Kay et al. 1989; Peng and Wagner 1992; Palmer 2004). The longitudinal relaxation time T_1 (or relaxation rate $R_1 = 1/T_1$) measures the decay of ^{15}N polarization, while the transverse relaxation time T_2 (or relaxation rate $R_2 = 1/T_2$) accounts for the loss of spin coherence. The HETNOE quantifies the polarization transfer from an amide ^{1}H to its attached ^{15}N.

Indeed, the major contributions to ^{15}N relaxation of a backbone amide ^{15}N in a protein are the ^{15}N chemical shift anisotropy (CSA) and the ^{15}N-^{1}H dipolar interaction (DD) with the directly bound proton, which to a good approximation, can be considered of equal magnitude throughout the protein backbone. This means that variations in ^{15}N relaxation properties of backbone amide ^{15}N spins can be interpreted in terms of differences in local motions (Peng and Wagner 1994). The measured ^{15}N relaxation parameters provide quantitative information on the amplitudes and the time scales of motions experienced by the ^{15}N-^{1}H amide group of a specific residue. ^{15}N relaxation (T_1, T_2 and HETNOE) is sensitive to motion occurring on the pico- to nanosecond time scale. As an example, Fig. 3.4 shows ^{1}H-^{15}N heteronuclear NOE values measured for individual amide sites along the backbones of different proteins; being especially sensitive to fast local motions, these observables allow protein regions characterized by different levels of local mobility to be easily identified.

The ^{15}N T_2 relaxation times are in addition also sensitive to slower conformational exchange processes in the micro- to millisecond time range that induce a change in the isotropic chemical shift. This exchange contribution (R_{ex}) adds to the relaxation induced by time modulation of the CSA and dipolar interactions ($1/T_2 = 1/T_2(\mathrm{CSA}) + 1/T_2(\mathrm{DD}) + R_{ex}$). In order to separate the exchange contribution, different experimental approaches have been proposed: repeating the relaxation measurements (i) at several static magnetic field strengths B_0, or (ii) by varying the applied radio-frequency field amplitude during the relaxation period. For more details, we

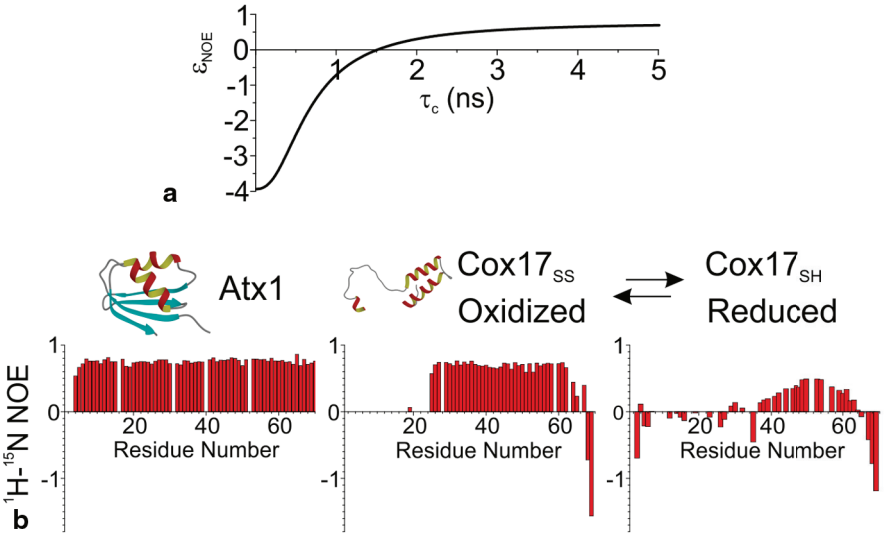


Fig. 3.4 **a** Variations of the steady-state NOE with the correlation time for the ^{1}H-^{15}N pair in a field of 11.74 T. **b** The graphs show the ^{1}H-^{15}N heteronuclear NOE for the majority of the residues of structured Atx1 protein (Arnesano et al. 2001), and of the partially structured oxidized form of the copper chaperone Cox17 (Arnesano et al. 2005) and of its unfolded reduced form (Bertini et al. 2011a). The experiments were acquired on a 500 MHz Bruker AVANCE spectrometer equipped with a CPTXI probe. Adapted from (Felli et al. 2012)

refer the reader to the scientific literature (Palmer et al. 2001; Tollinger et al. 2001; Palmer and Massi 2006).

As mentioned in the previous section, the oscillating magnetic fields induced by local and global rotational motions are causing spin relaxation in NMR. In order to account for the stochastic nature of these motions, they are best described by their frequency distribution, also called the power spectral density function $J(\omega)$. Furthermore, according to the standard relaxation theory for NMR spin relaxation, the so called BWR theory (Wangsness and Bloch 1953; Bloch 1956; Redfield 1957), spin relaxation is caused by the fluctuating magnetic fields created by molecular motions at the spin transition frequencies of the coupled ^{15}N-^{1}H spin system (ω_N, ω_H, $\omega_H-\omega_N$ and $\omega_H+\omega_N$) and at zero-frequency. The three relaxation parameters (T_1, T_2 and HETNOE) show a different dependence on the corresponding spectral density components ($J(0)$, $J(\omega_N)$, $J(\omega_H)$, $J(\omega_H-\omega_N)$ and $J(\omega_H+\omega_N)$) as described by the following equations:

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

$$\frac{1}{T_2} = \frac{d^2}{20}[4J(0) + J(\omega_H - \omega_N) + 3J(\omega_N) + 3J(\omega_H) + 6J(\omega_H + \omega_N)] + \frac{c^2}{15}[J(0) + 3J(\omega_N)] + R_{ex} \quad (3.8)$$

$$HETNOE = 1 + \frac{d^2}{20}\frac{\gamma_H}{\gamma_N}[6J(\omega_H + \omega_N)] \cdot T_1$$

with $d = \frac{\mu_0}{4\pi}\gamma_H\gamma_N\hbar r_{NH}^{-3}$ the dipolar coefficient, $c = \omega_N\Delta\sigma$ the ^{15}N chemical shift anisotropy and R_{ex} the contribution from conformational exchange processes, if any.

In order to extract the desired dynamics information, the measured ^{15}N relaxation data (T_1, T_2 and HETNOE) can be analysed in several ways. A first quick analysis of the relaxation data, or some combinations thereof, already yields valuable qualitative information on the global and local conformational properties of the protein. For example, we have shown above that the HETNOE, which is only sensitive to high-frequency components of the spectral density (see Eq. 3.8), highlights highly-flexible protein segments (Fig. 3.4). Typical HETNOE values range from about +0.9 for amide groups in rigid protein fragments to largely negative values (HETNOE << 0) for highly flexible sites in disordered regions. Thus, the HETNOE provides a measure of the orientational degree of freedom of a particular amide group, a property that is also called an order parameter. In addition, the T_1/T_2 ratio can be computed as it provides information on the local protein rigidity (Kay et al. 1989). For globular proteins, T_1/T_2 is to a good approximation proportional to the protein's overall rotational correlation time. For IDPs, T_1/T_2 allows to distinguish peptide regions displaying significant secondary and tertiary structural propensities, characterized by longer effective correlation times, from segments lacking any residual structure, characterized by shorter effective correlation times.

Arguably, the most commonly used method for the analysis of ^{15}N relaxation data of globular proteins is the model-free formalism introduced by Lipari and Szabo where the molecular tumbling, described by a global correlation time, is separated from local motions (Lipari and Szabo 1982), characterized by site-specific correlation times and order parameters. This separation is justified by the difference in time scales of these motions in the case of globular proteins, but this model is not rigorous for random coil-like polymers and IDPs. However, a similar approach can still be applied for IDPs by replacing the global correlation time of the protein by an effective segmental correlation time that varies over the polypeptide chain, reporting on the persistence lengths of the segmental chain motions.

Another possibility for analysing ^{15}N relaxation data that applies also to IDPs is the reduced spectral density mapping approach, which provides information about the shape of the spectral density function at individual sites (Farrow et al. 1995). If we neglect for the moment conformational exchange (R_{ex}) contributions to ^{15}N relaxation, Equation 3.8 shows that the three measured rates depend on five different spectral density components. In order to reduce this number to three, and thus matching the number of NMR observables, the high-frequency spectral density components $J(\omega_H)$, $J(\omega_H-\omega_N)$ and $J(\omega_H+\omega_N)$ are replaced by an effective $J(0.87\cdot\omega_H)$ value.

This allows to solve Eq. 3.8 analytically. The validity of reduced spectral density mapping for very flexible IDPs has recently been investigated and a slightly different approach was proposed, including the removal of exchange contributions by the measurement of additional cross-correlated relaxation rates (Kadeřávek et al. 2014).

3 Particular Challenges and Bottlenecks for NMR Studies of IDPs

A wide range of NMR experiments has been developed throughout the years for the study of globular proteins and their interactions, with the major objective of providing high-resolution structural and dynamic information. These experiments are the natural starting point for the NMR investigation of IDPs. However, the peculiar properties of IDPs have a strong impact on NMR spectra, as already outlined in the previous section, and thus on NMR experiments. This means that conventional NMR experiments need to be tailored for the specific properties of highly disordered proteins in order to study IDPs of increasing size and complexity. The main bottlenecks for NMR studies of IDPs will be briefly reviewed in the following sections.

3.1 Spectral Resolution

In order to extract structural and dynamic information for single nuclear sites in an IDP, we need sufficient spectral resolution to distinguish individual resonances (or correlation peaks). As mentioned in the previous section the first consequence of the lack of a stable structure is the averaging of a large part of the contributions to the chemical shift deriving from the local chemical environment. This results in a drastic reduction of the chemical shift ranges for the different nuclear spin species (^{1}H, ^{13}C, ^{15}N) and thus in a problem of strong overlap in the corresponding NMR spectra (see Fig. 3.5).

This resolution problem is especially pronounced for ^{1}H NMR, as will be discussed in more detail later on. Identifying strategies to overcome the problem of resonance/cross-peak overlap is of key importance to be able to study IDPs of increasing size and complexity. Several strategies to address this critical point are discussed in the next sections, such as exploiting the favourable resolution in exclusively heteronuclear NMR experiments, correlating nuclei of neighbouring amino acids and exploiting amino acid selection to simplify spectra and identify residue types.

3.2 Experimental Sensitivity

We have already introduced the theoretical basis and the main determinants of NMR sensitivity in Sect. 3.1.2. Here we will focus on IDPs, which are by definition

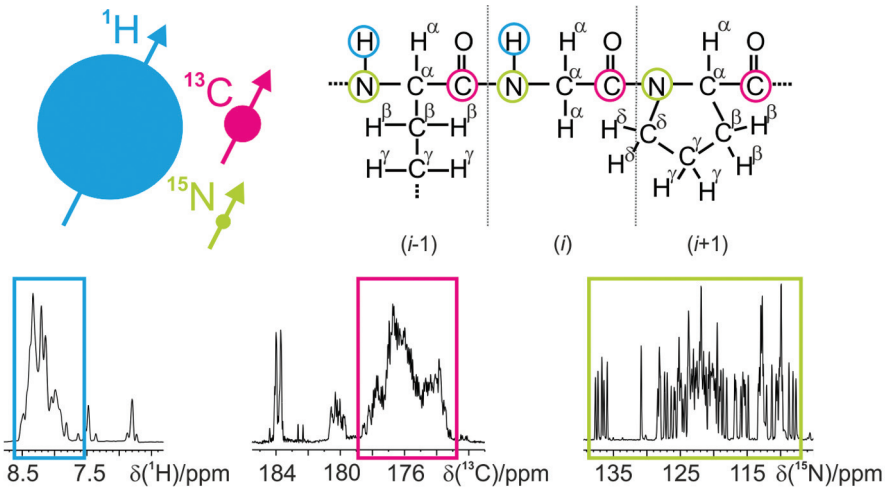


Fig. 3.5 The difference in the ^{1}H, ^{13}C and ^{15}N chemical shift dispersion of the IDPs is illustrated on the 1D spectra of α-synuclein. The experiments were acquired on a 700 MHz Bruker AVANCE spectrometer equipped with a CPTXO probe on a 1 mM α-synuclein sample (20 mM phosphate buffer, pH 6.4, 200 mM NaCl, 0.5 mM EDTA, at 285.5 K)

very flexible macromolecules. From an NMR sensitivity point of view, this flexibility has both advantages and disadvantages. The fast timescale molecular motions in IDPs are responsible for reduced effective rotational correlation times, a feature that in many cases contributes to long spin coherence lifetimes (long T_2) and narrow NMR lines. In principle, this enables the design of complex high-dimensional NMR pulse schemes to achieve the necessary resolution required for IDPs. It is interesting to note that for highly flexible IDPs the increase in molecular size does not have a major impact on relaxation times and linewidths, while it does for folded proteins. However, for protein regions with a 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 (Fig. 3.6). Some peptide segments have T_2 relaxation time constants that are four times shorter than those observed for other regions. This results in a large dynamic range of peak intensities observed in the NMR spectra. This feature becomes even more pronounced for complex NMR experiments involving an increasing number of transfer steps and frequency editing periods.

The effects of the presence of partially structured peptide regions and extensive conformational dynamics on the NMR spectra can become so pronounced that most of the nuclear spins are affected. No or only very limited solutions remain to study proteins characterized by strongly line-broadened resonances by NMR. Examples that often fall into this category are the so-called molten globule states that represent the “dark side” of biomolecular NMR, in between well-folded globular and highly disordered states.

Sensitivity can also be reduced because of protein aggregation, which is governed by the same principles as protein folding. Some IDPs, such as α-synuclein,

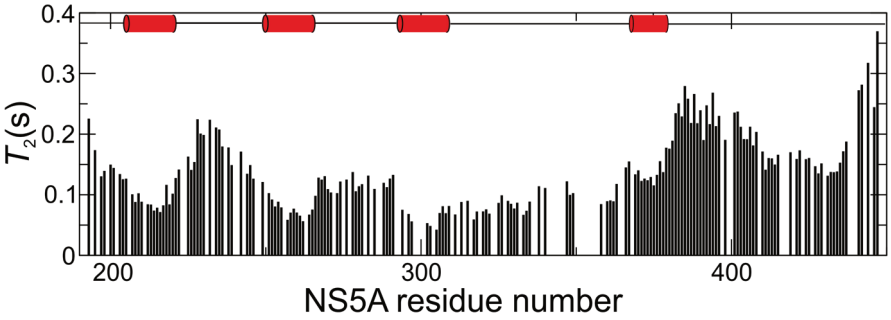


Fig. 3.6 ^{15}N amide T_2 relaxation time constants of NS5A measured as a function of the amino acid sequence at 278 K, at 18.8 T field. Red cylinders indicate regions adopting transient α-helical secondary structure

tau, or prion proteins, are known to cause neurodegenerative diseases in the aggregated form. Tendencies to aggregate can hamper NMR studies of these proteins. On the other hand, higher molecular mass aggregates become amenable to other NMR techniques, such as solid-state magic-angle spinning (MAS) NMR, which however is not addressed in this chapter (Tycko 2006; Chimon et al. 2007; Bertini et al. 2011c).

3.3 Experimental Time Requirements

Sample stability problems are often encountered because of aggregation, as mentioned before, or proteolytic degradation of IDPs. When the stability of the protein permits sample stabilization by boiling, the proteases can be deactivated by heat denaturation. However, in the case of significant residual structure, this is often not a valuable option as extensive heat treatment may result in irreversible changes of the structural features of the IDP (Chap. 6). In such cases the NMR spectra have to be acquired before sample degradation occurs. The requirements of short experimental times and at the same time high spectral resolution, achieved by long acquisition times in all dimensions of a high-dimensional (3D, 4D, etc.) experiment, seem to be contradictory, but they can be reconciled by the use of the fast NMR data acquisition techniques discussed in Sects. 3.5.3 and 3.5.5.

3.4 Sample Optimization

The conformational dynamics and transient structure of IDPs are highly sensitive to experimental conditions such as pH, buffer composition, salt concentration and temperature. Under certain conditions, some parts of the IDP may undergo

conformational dynamics on a timescale that leads to extensive line broadening in the NMR spectra and thus missing correlation peaks. Optimization of the sample conditions is thus even more important than for globular proteins. On the other hand, because of the dependence of the structural features of the IDP on the experimental conditions, one may want to study the IDP under close to physiological conditions such as neutral pH and relatively elevated temperature. Studies of enzymatic reactions, for example the occurrence of post-translational modifications, also often require neutral pH and high temperature to ensure optimal enzyme activity. Finally, in-cell NMR can be used in order to experimentally show that IDPs remain flexible *in vivo* and that disorder is not just an artefact of the chosen sample conditions.

3.5 Prolines are Abundant in IDPs

Further problems can occur because of the typically large proline content in the amino acid sequence of IDPs (Tompa 2002; Theillet et al. 2013). As prolines do not have backbone amide protons, they are not detected in amide ^{1}H detected NMR spectra and therefore represent breakpoints in the sequential backbone resonance assignment strategy based on $^{1}H^{N}$ detected triple-resonance experiments. For the same reason, variants of 2D ^{1}H-^{15}N correlation experiments typically used to follow or monitor physiological processes, such as interactions or post-translational modifications, or to measure observables such as ^{15}N relaxation rates, paramagnetic relaxation enhancements or residual dipolar couplings, do not provide information about prolines.

4 2D HN and CON NMR Experiments: The Fingerprint of an IDP

Uniform ^{13}C and ^{15}N isotope labelling of proteins (and thus IDPs) in bacterial expression systems has become routinely used and is currently a necessary requirement to proceed with any high resolution NMR investigation. The nuclear spins we can deal with are therefore ^{1}H, ^{13}C and ^{15}N, meaning the vast majority of the nuclei in a protein.

The main considerations when designing or choosing an NMR experiment are its overall sensitivity, the resulting spectral resolution, the number of detected peaks and the information contained in the spectral parameters (peak positions and intensities). In order to optimize these parameters we can choose the source of spin polarization, the directly observed nuclei, the number and nature of the indirectly detected nuclei, the type of transfer steps used and the dimensionality of the experiment, to cite only the most important ingredients of an NMR experiment. The most useful 2D NMR experiments to characterize IDPs are introduced here.

4.1 NMR Properties of ^{1}H, ^{13}C and ^{15}N in IDPs

According to Eq. 3.5, the sensitivity of the NMR experiment strongly depends on the gyromagnetic ratio of the nuclear spins (see Table 3.1) directly detected at the end of the pulse sequence and those serving as polarization source at the beginning of the experiment. Because of the high gyromagnetic ratio of ^{1}H, most NMR experiments for protein applications are based on ^{1}H excitation and ^{1}H direct detection, while heteronuclear chemical shifts (^{13}C and ^{15}N) are exploited in indirect dimensions of multidimensional NMR experiments (Sattler et al. 1999; Lescop et al. 2007). However, ^{13}C detection has recently evolved into a useful tool to study biomolecules (Felli and Pierattelli 2014a), in particular thanks to the larger chemical shift dispersion of ^{13}C nuclei compared to protons and to other specific benefits of the technique that improve the sensitivity of the NMR experiment, thus compensating for the lower gyromagnetic ratio (Sect. 3.6). ^{15}N detection in protein NMR is rarely used, although some experiments have been proposed for specific applications (Vasos et al. 2006; Takeuchi et al. 2010; Gal et al. 2011). NMR sensitivity is also influenced by the specific properties of the investigated system, in particular by its relaxation properties as well as by homonuclear couplings, all contributions not explicitly considered in Eq. 3.5. These aspects should also be carefully considered in the choice of the most appropriate experimental strategy to access a specific kind of information, as discussed more in detail in the next sections considering applications to IDPs.

The second important point to consider is spectral resolution. It is well known from NMR textbooks that ^{13}C and ^{15}N nuclei show superior chemical shift dispersion compared to ^{1}H. This is definitely an interesting property provided it also holds for IDPs. As an example, Fig. 3.5 shows the chemical shift dispersion observed in α-synuclein, a well-characterized IDP, for three backbone nuclear spins, $^{1}H^{N}$, $^{13}C'$ and ^{15}N. It is clear that the chemical shift dispersion increases from $^{1}H^{N}$ to $^{13}C'$ to ^{15}N even in the absence of a stable structure. The same holds true for aliphatic and aromatic spin pairs, with ^{13}C yielding a higher resolution in the NMR spectrum than the attached ^{1}H. Therefore, the exploitation of the improved frequency resolution in the ^{13}C and ^{15}N dimensions of multidimensional NMR spectra is crucial for the study of IDPs.

For ^{1}H and ^{13}C resonances in the side-chains of IDPs we observe a larger overall dispersion (compared to backbone nuclei) due to the differences in chemical structure between different amino acids. However, side-chain resonances from the same amino acid type cluster in the same spectral region. Side chain resonances are therefore mainly used as indicators of the amino acid type of a given residue, while they are only of little use for the characterization of site-specific structure and dynamics in the IDP. As an example, the 2D ^{1}H-^{13}C (HC) correlation spectrum acquired on α-synuclein is shown in Fig. 3.7. Despite the high sensitivity of this experiment, the overlap of signals deriving from the same type of amino acid and the presence of resolved homonuclear couplings (see below) are responsible for the extensive overlap observed.

Since relatively narrow peaks are observed in the NMR spectra of IDPs in the absence of unfavourable dynamics, homonuclear *J*-couplings may contribute

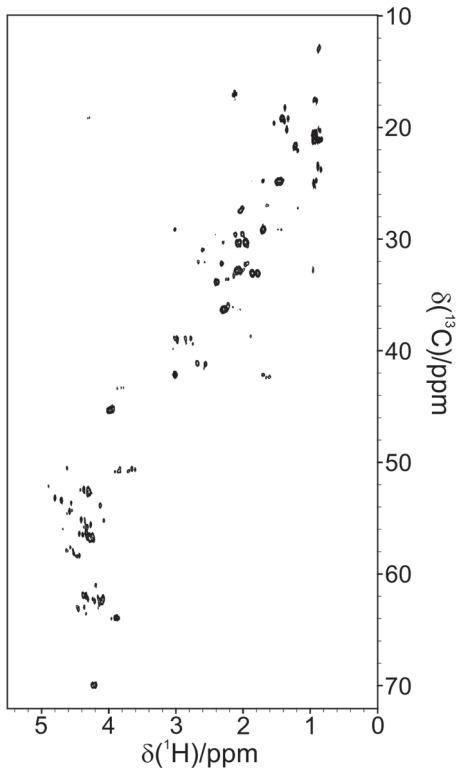


Fig. 3.7 The ^{1}H-^{13}C HSQC spectrum acquired on a 950 MHz Bruker AVANCE spectrometer equipped with a CPTCI probe on a 1 mM α-synuclein sample (20 mM phosphate buffer, pH 6.4, 200 mM NaCl, 0.5 mM EDTA, at 285.5 K)

significantly to the linewidth. Excluding for the moment the one-bond $^{1}J_{CC}$ couplings from the discussion, which definitely need to be suppressed experimentally to obtain well-resolved ^{13}C spectra (as explained in Sect. 3.6.1), the magnitude of ^{2}J and ^{3}J homonuclear coupling constants decreases passing from ^{1}H to ^{13}C to ^{15}N as an indirect consequence of the gyromagnetic ratios of these nuclei. This feature is thus again in favour of the detection of ^{13}C or ^{15}N instead of ^{1}H, unless the IDP has been perdeuterated and back-protonated at amide sites to suppress homonuclear ^{1}H-^{1}H couplings.

4.2 2D ^{1}H-^{15}N and ^{13}C′-^{15}N Correlation Fingerprint Spectra of IDPs

Recording simple 2D NMR "fingerprint" spectra of the IDP provides some useful information on the protein and allows the evaluation of the overall sample properties and the feasibility of a subsequent high resolution NMR study even in the absence of a sequence-specific resonance assignment. Indeed, the number of cross-peaks detected in the spectrum compared to the number of peaks expected from the primary sequence provides a first estimation of the amount of spectral overlap and indicates whether the totality of the peptide sequence or only a part

of it is NMR-visible under the chosen sample conditions. The latter case occurs quite often as many proteins exhibit a significantly heterogeneous nature in terms of structural and dynamic properties leading to conformational exchange and/or aggregation induced line broadening. The observed chemical shift dispersion allows the identification of whether the protein is structured, partly structured, or highly unstructured.

For such a fingerprint spectrum, we want to detect a single correlation peak per residue with good dispersion in a reasonable amount of data acquisition time. The most common experiments used are therefore ^{1}H detected ^{1}H-^{15}N (HN) (Favier and Brutscher 2011) and ^{13}C detected ^{13}C′-^{15}N (CON) (Bermel et al. 2006a) correlated spectra, as shown in Fig. 3.8 for human securin, a 200 amino acid IDP with more than 20 % prolines.

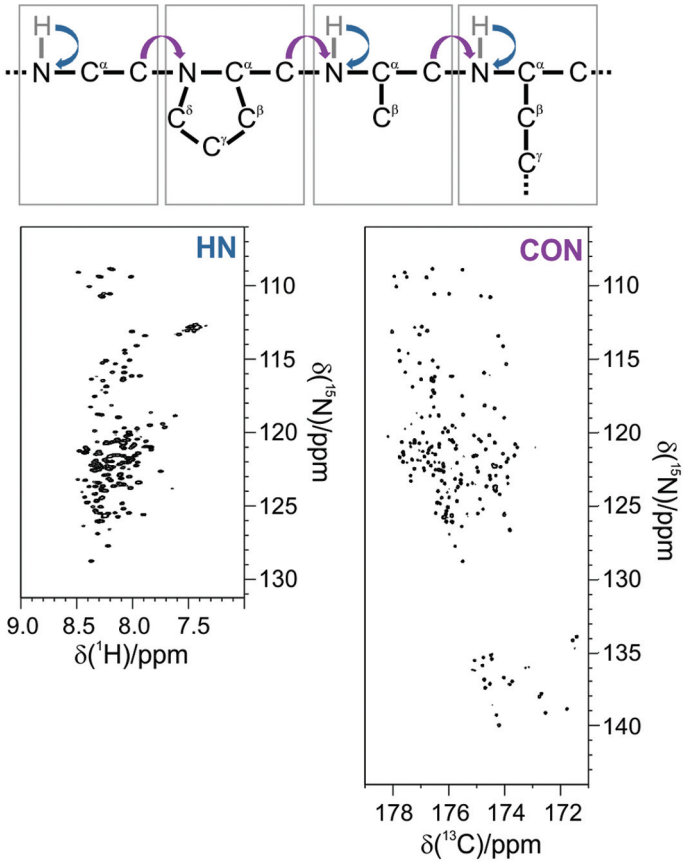


Fig. 3.8 The schematic illustration (*top*) of the correlations observed in the basic 2D ^{1}H-^{15}N HSQC and ^{13}C-^{15}N CON-IPAP experiments. The ^{1}H-^{15}N HSQC (*left*) and ^{13}C-^{15}N CON-IPAP (*right*) spectra were acquired on the ^{13}C, ^{15}N-labeled sample of the intrinsically disordered human securin protein (0.7 mM sample in 25 mM phosphate buffer, pH 7.2, 150 mM KCl, 10 mM 2-mercaptoethanol, at T 283 K) (Csizmok et al. 2008). The experiments were performed on a Bruker AVANCE 700 MHz spectrometer equipped with a CPTXO probe

While the HN experiment is much more sensitive and can be recorded in a significantly short time, prolines are only detected in the CON spectrum, which also shows a better spectral resolution. In addition, the CON spectrum does not suffer from hydrogen-exchange-induced line broadening and thus can still be recorded under conditions of high pH and temperature, where many HN peaks are no longer detectable (Gil et al. 2013); this is shown in Fig. 3.9, which reports the HN and CON spectra acquired on α-synuclein with increasing temperature. On the other hand, high-quality HN correlation spectra can be recorded on protein samples at concentrations of only a few μM on a spectrometer equipped with a cryogenic probe. The 2D HN and CON spectra are thus highly complementary both in terms of detectability and information content. These 2D correlation experiments can be acquired in different ways. The most appropriate variants for applications to IDPs are discussed in detail in the following section.

These 2D spectra can also be used to follow changes in the properties of the IDP upon changing the experimental conditions such as temperature, pH, ionic strength, buffer, reducing/oxidizing environment, or the addition of potential partners such as metal ions, small molecules, nucleic acid fragments and proteins. They also enable chemical reactions such as the occurrence of post-translational modifications to be followed (Selenko et al. 2008). Finally, they provide an invaluable tool to take a snapshot of a protein inside an entire cell (Serber et al. 2006; Selenko and Wagner 2007; Felli et al. 2014). As an example, the 2D HN and CON experiments acquired

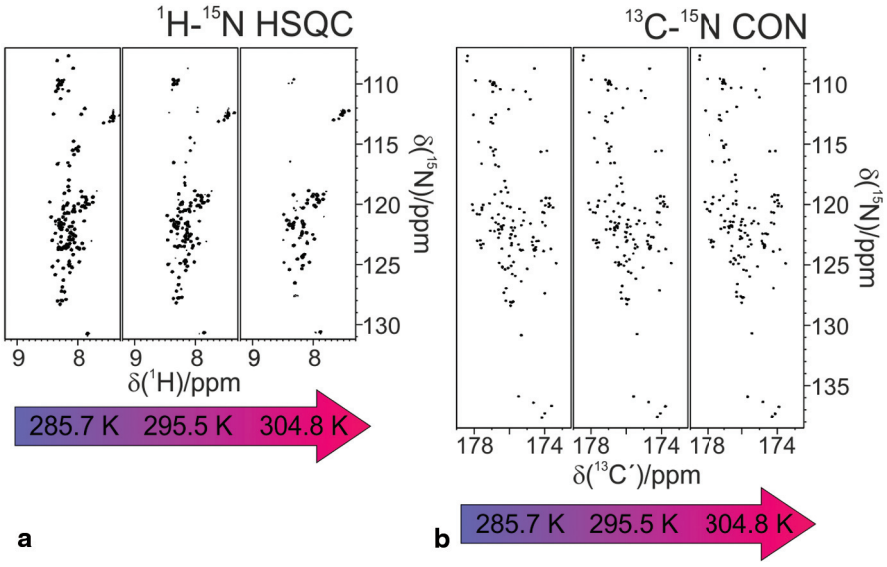


Fig. 3.9 2D ^{1}H-^{15}N HSQC (**a**) and ^{13}C-^{15}N CON-IPAP (**b**) spectra acquired on a 1 mM α-synuclein sample (20 mM phosphate buffer, pH 7.4, 200 mM NaCl, 0.5 mM EDTA) at different temperatures, from *left* to *right*: 285.7 K, 295.5 K and 304.8 K. Each spectrum was acquired with one scan per increment and with the same resolution (in Hz)

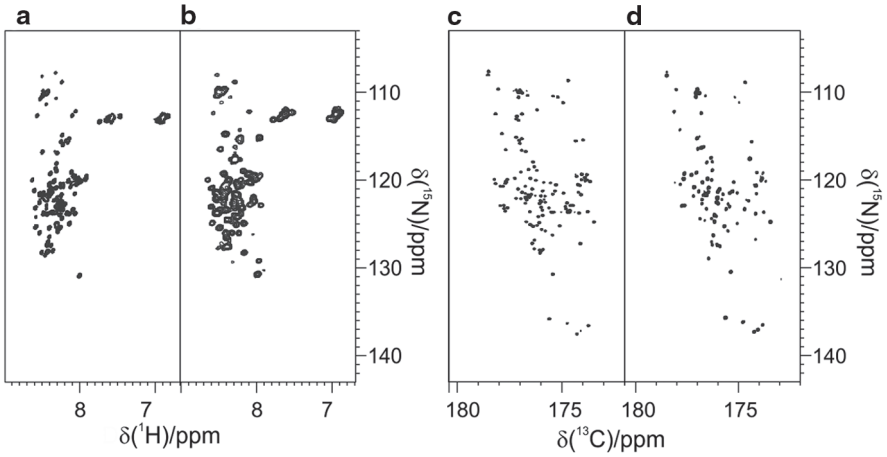


Fig. 3.10 The 2D ^{1}H-^{15}N SOFAST-HMQC (**a**, **b**) and ^{13}C-^{15}N H$^{\alpha\text{-flip}}$CON (**c**, **d**) spectra acquired on α-synuclein overexpressed in *E. coli* cells (**b**, **d**) and on cell lysates (**a**, **c**). The experiments were acquired on a Bruker AVANCE 700 MHz spectrometer equipped with a CPTXO probe; the acquisition time of SOFAST-HMQC was 13 min and of H$^{\alpha\text{-flip}}$CON 44 min

on α-synuclein in cell are compared with those acquired on the purified protein (Fig. 3.10), showing that they can be used for the investigation of IDPs in-cell.

Finally, modifications of these basic 2D experiments (HN and CON) enable the determination of a variety of observables that report on different properties of the IDP at atomic resolution, once sequence-specific assignment becomes available. These include ^{15}N relaxation rates, scalar couplings, residual dipolar couplings (RDCs), paramagnetic relaxation enhancements (PREs), cross-relaxation (σ_{HH}) and cross-correlation rates (CCR), as well as solvent exchange rates. All of these NMR observables report on the structural and dynamic features of the IDP and will be discussed in detail in Chaps. 4 and 5.

5 Tools for Overcoming Major Bottlenecks of IDP NMR Studies

5.1 Multidimensional NMR, Indirect Frequency Editing and Non-uniform Sampling

The large number of NMR-active nuclei in a protein results in severe resonance overlap in one-dimensional spectra (^{1}H, ^{13}C, ^{15}N), making it practically impossible to extract atom-resolved information from them. This problem is circumvented by using multidimensional (nD) NMR techniques that spread and correlate the signals of individual nuclear spins along different frequency dimensions (Ernst et al. 1987). Multidimensional NMR data are recorded by repeating the basic pulse sequence

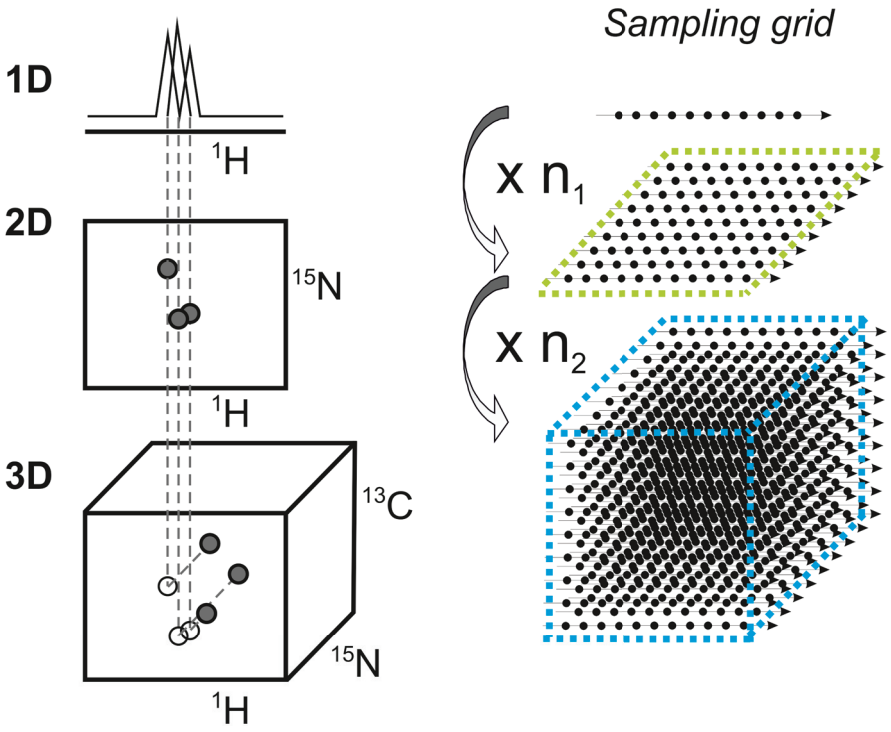


Fig. 3.11 Schematic representation of the multi-dimensional NMR spectra (*left*), and the increasing number of experimental repetitions required, resulting in longer acquisition times (*right*)

numerous times. For each repetition of the experiment (transient), the observable NMR signal is detected along one dimension, while additional 'indirect' dimensions are sampled by incrementing a time variable in the pulse sequence from one repetition to the next. Despite the great success of multidimensional NMR, an important drawback of this stepwise sampling procedure is the long experimental time, which is a direct result of the hundreds or even thousands of transients that are required for a single data set (Fig. 3.11).

In this section we will discuss different editing techniques to improve the spectral resolution in the indirect dimensions of nD NMR experiments and advanced non-uniform sampling (NUS) approaches that are indispensable for recording high-dimensional spectra in reasonable experimental times.

5.2 Real-time, Constant-time and Semi-constant Time Frequency Editing

Frequency editing of nuclear spins is the key step for introducing indirect dimensions in multidimensional experiments. In order to do so, a pulse sequence element that allows time evolution of the frequency-edited spins (I) while sup-

pressing (refocusing) all other chemical shift and *J*-coupling evolutions is required. Figure 3.12 shows five common pulse sequence implementations of *I*-spin editing during an incremented time variable t_1: (a) conventional real-time, (b) optimized real-time, (c) and (d) constant-time (CT) and (e) semi-CT editing.

In the conventional real-time implementation, heteronuclear J_{IS} and J_{IK} coupling evolutions are refocused by a 180° pulse applied in the middle of t_1. In this context "heteronuclear" means that the particular spin species can be manipulated separately by an appropriately shaped radiofrequency pulse. As an example, $^{13}C'$ and $^{13}C^{\alpha}$ can be manipulated separately as they have well-separated chemical shift ranges, while $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ have generally overlapping chemical shift ranges and thus cannot be separated spectroscopically. Therefore, the $^{13}C^{\alpha}$ -$^{13}C^{\beta}$ coupling gives rise to a peak splitting in the $^{13}C^{\alpha}$ dimension of a NMR spectrum recorded with conventional real-time editing as illustrated in Fig. 3.13a. A slightly optimized version of real-time editing is shown in Fig. 3.12b, where chemical shift evolution of the *I*-spins during the 180° *S*-spin decoupling pulses, as well as Bloch-Siegert phase shifts, are refocused. This allows recording a data set that does not require any phase correction in the corresponding indirect dimension (t_1) in order to avoid baseline distortions.

In order to avoid line splitting due to homonuclear J_{II}-coupling evolution, the CT editing blocks of Figs. 3.12c and 3.12d can be used, with the CT delay *T* set to $T = 1/J_{II}$. A $^{1}H^{N}$-$^{13}C^{\alpha}$ correlation spectrum recorded with the pulse sequence block of Fig. 3.12c is shown in Fig. 3.13b. The detected NMR signal in this spectrum is modulated by a factor $\cos^{n}(\pi J_{cc} T)$, where *n* is the number of carbon atoms attached. In addition, signals with different sign provide meaningful information about the spin-coupling topology; for example, correlation peaks of glycine residues are of opposite sign with respect to all others, as glycines have no $^{13}C^{\beta}$ attached to the $^{13}C^{\alpha}$.

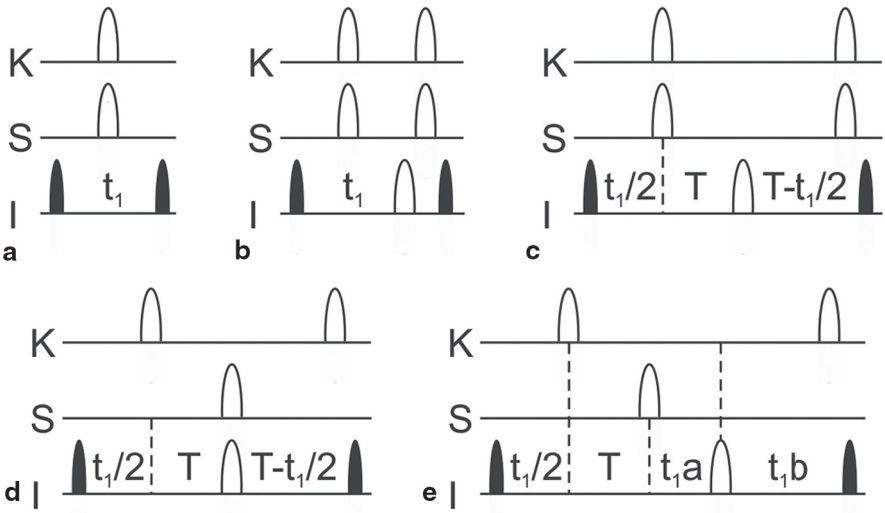


Fig. 3.12 Common pulse sequence implementations of *I*-spin editing during an incremented time variable t_1: **a** conventional real-time; **b** optimized real-time; **c** and **d** constant-time (CT), and **e** semi-CT editing

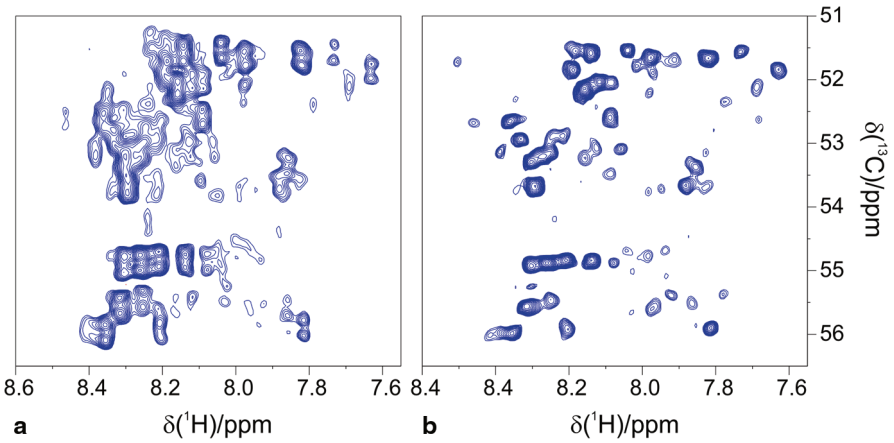


Fig. 3.13 Expansion of ^{1}H-$^{13}C^{\alpha}$ plane from a BEST-TROSY HN(CO)CA spectrum of NS5A recorded with real-time (**a**) and CT evolution (**b**). In both cases the acquisition time in the indirect $^{13}C^{\alpha}$ dimension was 20 ms with 200 points recorded and a spectral width of 30 ppm

Alternatively, CT frequency editing can be applied during an INEPT-type coherence transfer step (Morris and Freeman 1979). This is shown in Fig. 3.12d, where the *I*-spins are edited during an *I* to *S* transfer. Here, the CT delay has to be set to $T = 1/(2J_{IS})$ in order to achieve complete transfer.

A drawback of CT editing is that the maximum possible evolution time (t_1^{max}) is limited by the CT delay *T* so that $t_1^{max} \leq T$. This, of course, has an effect on the achievable spectral resolution, which is a crucial point for NMR studies of IDPs as discussed before. A possible solution could be to increase the CT delay to $n \cdot T$, where *n* is an integer number. However, introducing too long delays in the pulse sequences may cause pronounced relaxation losses. Semi-CT editing has therefore been proposed to enhance the spectral resolution while still exploiting spin evolution during a coherence transfer delay. The improvement achieved with semi-CT evolution in terms of spectral resolution in the ^{15}N dimension of an HNCA experiment is illustrated in Fig. 3.14.

5.3 Strategies for Non-uniform Data Sampling in Multi-D NMR

The amplitude of the NMR signal, or free induction decay (FID), is typically measured (sampled) at discrete, uniformly spaced time points. In a multidimensional NMR experiment, the same approach is generally employed in the indirect dimensions as well, since uniform sampling is required for data processing using the fast Fourier transform (FFT). The Nyquist theorem states that the sampling rate needs to be faster than twice the highest frequency expected or, in other words, determines a maximum value for the time interval between sampled points (dwell time) that has to be used to avoid spectral folding of the peaks. Hence, the sampled data points form a Cartesian grid with the spacing between time points in each dimension given by the inverse of the spectral width of the edited nuclei. As a result of this uniform

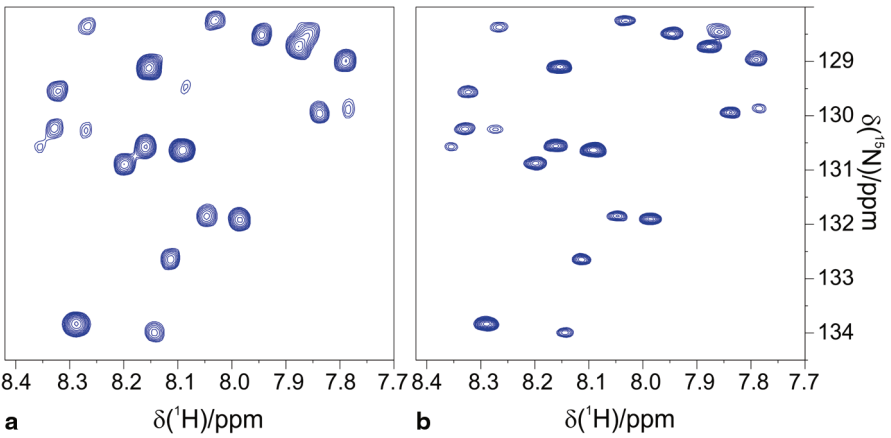


Fig. 3.14 Expansion of the ^{1}H-^{15}N plane from BEST-TROSY HNCA of NS5A acquired with **a** CT editing using the maximum number of increments (CT = 24.5 ms) and **b** semi-CT evolution (CT = 53 ms). In both cases the spectral width in the ^{15}N dimension was 30 ppm

data sampling procedure, the experimental time requirement increases roughly by about two orders of magnitude for each additional dimension, making high-dimensional experiments (> 3D) impractical or requiring strong compromises on the spectral resolution in the indirect dimensions.

In order to overcome these limitations, alternative sampling strategies combined with appropriate processing tools have been developed over the past two decades (Hiller et al. 2005; Kazimierczuk et al. 2006; Mobli et al. 2006; Coggins and Zhou 2007; Kazimierczuk et al. 2007; Kazimierczuk et al. 2010a; Kazimierczuk et al. 2012; Yao et al. 2014). These sparse or non-uniform sampling (NUS) techniques rely on a reduction of the overall number of sampled time points by recording only a subset of the data points of the Cartesian grid. Examples of alternative sampling schemes are shown in Fig. 3.15.

Some of these sampling grids, e.g. linear under sampling or radial sampling, still yield data sets that can be processed using FFT (Szyperski et al. 1993b; Brutscher et al. 1995b; Kupce and Freeman 2003). The general NUS scheme, however, where a certain percentage of the grid points is randomly chosen, requires alternative processing tools. Several algorithms for processing non-uniformly sampled data are currently available and well-established, such as multidimensional decomposition (MDD) (Luan et al. 2005; Tugarinov et al. 2005), maximum entropy (MaxEnt) methods (Hoch and Stern 1996), compressed sensing (CS) (Holland et al. 2011; Kazimierczuk and Orekhov 2011), multidimensional Fourier transform (MFT) (Kazimierczuk et al. 2010a) and spectroscopy by integration of frequency and time domain information (SIFT) (Matsuki et al. 2009). The common feature of all these methods is that they aim to find the NMR spectrum that, when applying the inverse Fourier transform, best reproduces the measured time data points. As in the case of sparse data sampling, this is an underdetermined computational problem. Some additional assumptions must therefore be made in order to choose the most likely spectrum from

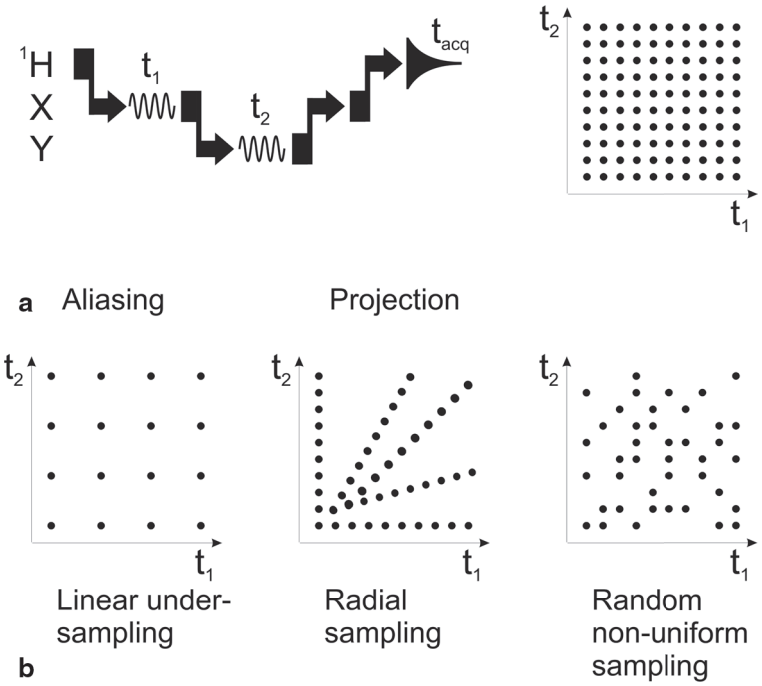


Fig. 3.15 Examples of different sampling schemes to speed up the acquisition of a 3D experiment. **a** Schematic illustration of a typical three-dimensional H-X-Y correlation experiment and the conventional time-domain sampling grid required to build the two indirect dimensions. Each point on the grid corresponds to a single repetition of the basic pulse sequence. **b** Some NUS schemes are shown. All the sampling patterns can be employed to reduce the experimental time, since a smaller number of points with respect to those of the conventional sampling grid are measured. The linear under sampling pattern can be used to fold chemical shifts in an advantageous manner, whereas radial sampling is generally employed in automated projection spectroscopy (APSY). Instead, random non-uniform sampling schemes are most commonly used when non-linear methods for spectral reconstruction such as MDD or MFT are used

among the possible solutions. For example, MDD exploits the prior knowledge that NMR signals are the direct product of Lorentzian (or Gaussian) line shapes in each frequency dimension. MaxEnt relies on the optimization of a penalty function in order to select the spectrum that has the highest entropy (minimal number of signals). Similarly, CS performs a l_1-norm minimization to recover the sparsest spectrum, while MFT performs a discrete multidimensional Fourier transform of the data using appropriate data weighting and filtering for the reduction of sampling noise. Finally, SIFT is an iterative method that replaces the missing time domain data points by zero for the first iteration, allowing classical FFT processing. In subsequent iterations, the missing data points are replaced by the result of the inverse FT of the obtained NMR spectrum after having set known empty spectral regions to zero.

At present, the question of the most appropriate sampling grid and processing technique for a given experiment is non-trivial, but a few general recommendations can be given: (i) the percentage of sampled data points can be decreased with

increasing dimensionality of the experiment and sparseness of the final NMR spectrum (number of expected correlation peaks). Typically, in the case of a 4D backbone assignment experiment, a few percent of sampled data points are sufficient. (ii) A random distribution of the sampled data points is particularly recommended to spread the sampling noise over the entire frequency domain (with apparent reduction of its overall amount), preventing the clustering of artefacts in specific spectral regions. (iii) The choice of the processing algorithm mainly depends on the available software (degree of automation, required expertise for proper parameter adjustment) and the computer power (some algorithms require more computer power and memory than others). If possible, it is always a good option to test different processing tools on the same data set. Recent developments by spectrometer manufacturers have made it straightforward to set up a random sampling scheme and run experiments in NUS mode. Also, some of the non-linear processing routines, such as MDD and CS, have been interfaced with the NMR spectrometer software and are available for routine use. Alternatively, most of the processing algorithms can be accessed online and downloaded. These advanced data acquisition and processing tools should be largely exploited for NMR investigations of IDPs, as they allow the performance of experiments of high dimensionality in a reasonable amount of time at very high spectral resolution, as discussed in detail in Sect. 3.9.1.

5.4 Selecting ^{15}N Spin States with Favourable Transverse Relaxation Properties

Multi-dimensional NMR correlation experiments require a series of coherence transfer steps and frequency editing periods during which the detectable NMR signal, and thus the sensitivity of the experiment, decreases due to transverse spin relaxation. In order to improve the performance of such experiments and therefore their applicability to large IDPs and low sample concentrations, it is important to select the coherences with the longest transverse relaxation times for the transfer and chemical shift editing steps. Here we present two commonly used techniques for reducing signal losses during ^{15}N transverse evolution periods as required for all amide ^{1}H detected NMR experiments, and that are also of importance in ^{13}C detected experiments.

In a standard INEPT-type $^{15}N \rightarrow ^{13}C$ transfer step, ^{15}N single quantum coherence (SQC) oscillates back and forth between in-phase (N_x) and anti-phase ($2N_yH_z$) coherence, as a result of ^{15}N-^{1}H scalar coupling evolution. As a consequence, the effective relaxation during the transfer is given by the average of in-phase and anti-phase SQ relaxation rates. In uniformly ^{13}C, ^{15}N labelled proteins, relaxation of anti-phase SQC is about 30% faster than relaxation of in-phase SQC. This difference is even more pronounced in the case of significant solvent exchange rates between the amide and water protons, further reducing the lifetime of anti-phase SQC. Improved transfer efficiency is thus achieved by ^{1}H composite pulse decoupling, which removes the ^{15}N-^{1}H scalar coupling evolution and maintains the in-phase SQC throughout the duration of the transfer. Common ^{1}H decoupling schemes are MLEV-16 (Levitt et al. 1982), DIPSI-2 (Shaka et al. 1988), WALTZ-16 (Shaka et al.

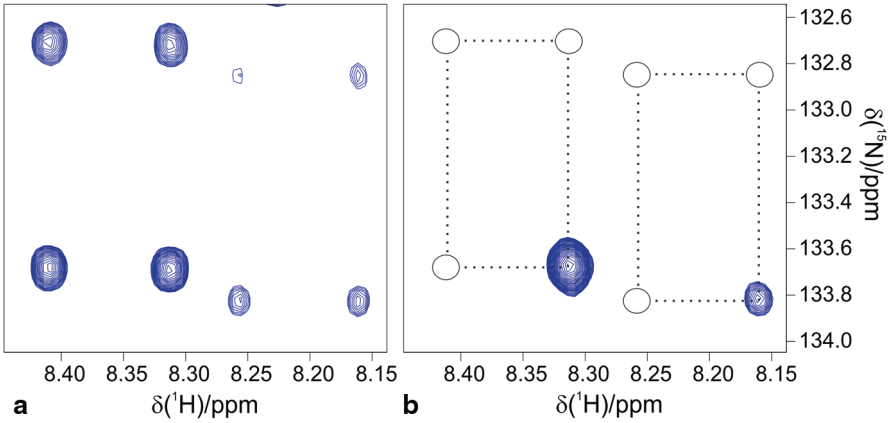


Fig. 3.16 Small spectral region of **a** ^{1}H-^{15}N coupled HSQC and **b** TROSY spectrum of NS5A recorded on a 950 MHz spectrometer. The spectral region displays correlation peaks for two NS5A residues. The TROSY effect is less pronounced for the left residue, which is located in a highly flexible region of the IDP, compared to the right one, which is part of a peptide segment with a high propensity to form an α-helical structure. In both cases the acquisition time in the indirect dimension was 44 ms with 256 points recorded and a spectral width of 30 ppm

1983a; Shaka et al. 1983b) and GARP-1 (Shaka et al. 1985). They are all composed of a basic pulse sequence element (R) that basically performs a broadband 180° spin inversion and is repeated by applying an additional phase cycle. They differ mainly by the B_1 field strength required to achieve a certain decoupling bandwidth.

A different approach to enhancing spectral resolution in ^{15}N spectra, especially at high magnetic field strength, is transverse relaxation-optimized spectroscopy (TROSY) introduced by Pervushin et al. in 1997 (Pervushin et al. 1997). The TROSY effect is based on the interference between two different spin relaxation mechanisms, e.g. the dipolar coupling (DD) and the chemical shift anisotropy (CSA), in a scalar coupled spin pair such as ^{1}H-^{15}N. This interference can be constructive (increasing the apparent relaxation rate) or destructive (decreasing the apparent relaxation rate) for different components of the peak multiplet (see Fig. 3.16). The TROSY pulse sequence allows the selection of the multiplet component (single-transition spin states) giving the sharpest lines (Pervushin et al. 1997). The TROSY effect for ^{1}H-^{15}N is highest at magnetic field strengths of about 1 GHz and increases with the effective tumbling correlation time. TROSY-based pulse sequences are thus especially useful to enhance the peak intensities and line shapes of the IDP regions that are involved in transient structure formation.

5.5 Longitudinal-relaxation Enhancement for Increased Sensitivity and Reduced Acquisition Times

In this section we discuss the dependence of the experimental sensitivity on the recovery delay T_{rec} (Schanda 2009). A schematic drawing of an NMR experiment

is shown in Fig. 3.17, consisting of a pulse sequence of length t_{seq}, a data acquisition period t_{acq} and an additional inter-scan delay t_{rec}. A recovery delay is required for relaxation of the system in order to restore sufficient spin polarization to restart the experiment for a subsequent scan ($T_{rec}=t_{acq}+t_{rec}$). Its duration depends on the longitudinal relaxation time constant T_1. On one hand, the detected signal intensity is proportional to the amount of spin polarization available at the beginning of each scan under steady-state conditions. On the other hand, the SNR scales with the square root of the number of scans that can be performed during a given experimental time, and thus the scan time $T_{scan} = t_{seq} + t_{acq} + t_{rec}$. These two effects can be described analytically by the following equation:

$$SNR \propto \frac{1 - exp(-T_{rec} / T_1)}{\sqrt{T_{scan}}} \tag{3.9}$$

Equation 3.9 implies that maximum sensitivity is obtained by adjusting the recycle delay T_{rec} to:

$$T_{rec}(SNR_{max}) \cong 1.25 \cdot T_1 \tag{3.10}$$

and thus the dependency of the experimental sensitivity on the longitudinal relaxation time T_1 is approximately given by:

$$SNR_{max} \propto \frac{1}{\sqrt{T_1}} \tag{3.11}$$

derived substituting T_{rec} and T_{scan} with T_1 in Equation 3.9.

Enhancing the longitudinal relaxation efficiency of the excited spins thus provides an interesting way to increase the experimental sensitivity and additionally to reduce the minimal data acquisition time. Here we will focus on ^{1}H excitation experiments. In order to understand the experimental schemes that have been proposed for longitudinal proton relaxation enhancement, we need to briefly discuss the spin interactions that govern proton relaxation. There are essentially two different mechanisms that are responsible for proton longitudinal relaxation in a protein: (i) ^{1}H-^{1}H dipolar interactions; (ii) hydrogen exchange processes between labile protein protons, e.g. amide and hydroxyl protons, and water solvent protons. The time evolution of the polarization of each proton spin in the molecule is given by the Solomon or Bloch-McConnell equations (Bloch 1946; Solomon 1955; McConnell 1958):

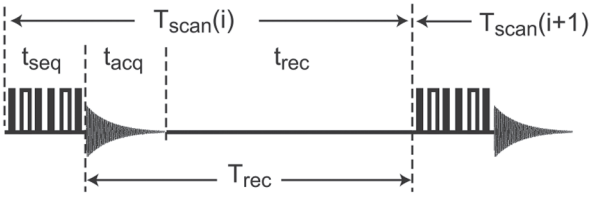


Fig. 3.17 A schematic drawing of an NMR experiment

$$-\frac{d}{dt}\begin{pmatrix} W_{1z}-W_{1z}^0 \\ H_{1z}-H_{1z}^0 \\ H_{2z}-H_{2z}^0 \\ \vdots \\ H_{nz}-H_{nz}^0 \end{pmatrix} = \begin{pmatrix} \rho_W & 0 & 0 & \cdots & 0 \\ 0 & \Sigma_j \rho_{1j} & \sigma_{12} & \cdots & \sigma_{1n} \\ 0 & \sigma_{21} & \Sigma_j \rho_{2j} & \cdots & \sigma_{2n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \sigma_{n1} & \sigma_{n2} & \cdots & \Sigma_j \rho_{nj} \end{pmatrix} \begin{pmatrix} W_{1z}-W_{1z}^0 \\ H_{1z}-H_{1z}^0 \\ H_{2z}-H_{2z}^0 \\ \vdots \\ H_{nz}-H_{nz}^0 \end{pmatrix} + \begin{pmatrix} 0 & 0 & 0 & \cdots & 0 \\ -k_{ex,1} & k_{ex,1} & 0 & \cdots & 0 \\ -k_{ex,2} & 0 & k_{ex,2} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -k_{ex,n} & 0 & 0 & \cdots & k_{ex,n} \end{pmatrix} \begin{pmatrix} W_{1z} \\ H_{1z} \\ H_{2z} \\ \vdots \\ H_{nz} \end{pmatrix} \tag{3.12}$$

where H_{iz} denotes the z-component of the polarization of proton i and $H_{iz}{}^0$ is its thermal equilibrium value. The different ρ and σ terms stand for auto- and cross-relaxation rate constants, with values depending on the distance separating the two protons involved as well as the global and local dynamics of the protein experienced at the sites of the two protons. W_z stands for the bulk water polarization and $k_{ex,\,i}$ are the hydrogen exchange rates for individual protons with the water.

Equation 3.12 indicates that the relaxation of an individual proton spin depends on the spin state of all the other protons in the protein, as well as the bulk water, at the start of the recovery time. The selective spin manipulation of a subset of protons, e.g. amides, while leaving all other protein and water proton spins unperturbed, thus provides an efficient spectroscopic tool for enhancing longitudinal proton spin relaxation, as will be shown in the following section.

The contributions of solvent exchange and dipolar cross-relaxation of amide protons with unperturbed aliphatic proton spins depend on the size and residual structure of the IDP and the sample conditions. Figure 3.18 shows apparent amide ^{1}H T_1 values measured for different IDP samples upon selective or non-selective inversion of different sets of proton spins (Gil et al. 2013; Solyom et al. 2013). For NS5A and BASP1, studied at low temperature and pH, the major relaxation enhancement mechanisms are dipolar interactions, while this situation changes as solvent exchange becomes more efficient, as shown for α-synuclein and PV core protein, which were studied at higher temperature and pH. The average T_1 values measured for the different IDPs shown in Fig. 3.18, as well as the range of predicted exchange rates using the SPHERE program are given in Table 3.2 (Bai et al. 1993; Zhang 1995).

When dipolar interactions provide the dominant relaxation mechanism, non-selective ^{1}H T_1 values are on the order of 900 ms, while selective spin manipulation allows them to be reduced by a factor of three to four, reaching values of 200–300 ms (Schanda and Brutscher 2005; Lescop et al. 2007). This difference becomes even more pronounced in the case of fast hydrogen exchange. Under these conditions, the non-selective ^{1}H T_1 approach the T_1 of bulk water (~3 s at 25 °C), while the selective T_1 become as short as 60 ms (Gil et al. 2013). For α-synuclein at pH 7.4 and 15 °C, the average amide ^{1}H T_1 is reduced by a factor of 38, resulting

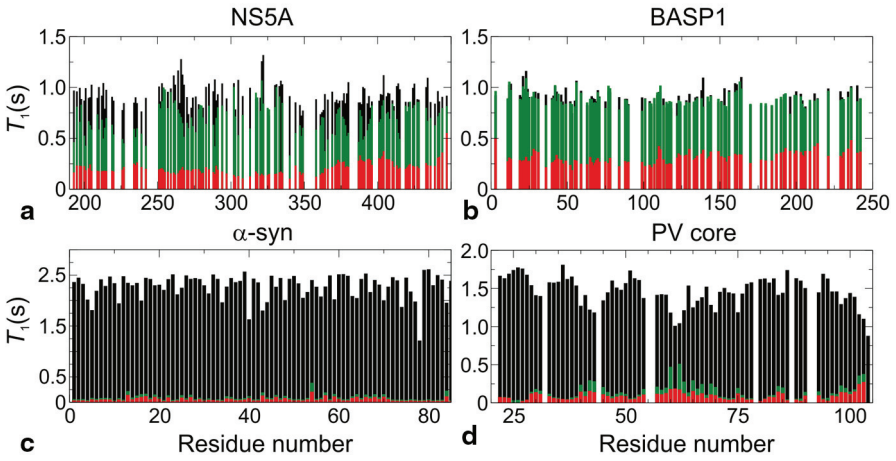


Fig. 3.18 Apparent ^{1}H T_1 relaxation time constants measured by inversion recovery experiments. Amide proton-selective (*red*), water-flip-back (*green*) and non-selective (*black*) inversions were used. The four samples and the conditions were: **a** NS5A D2D3 protein (pH 6.5, 278 K); **b** BASP1 (pH 6.5, 278 K); **c** α-synuclein (pH 7.4, 288 K); **d** PV core (pH 7.5, 278 K)

Table 3.2 Ranges of predicted amide proton solvent exchange rates (k_{ex}) (predictions were performed with the SPHERE program (Bai et al. 1993)) and their average over all residues are listed as well as the averages of the apparent longitudinal relaxation time (T_1) constants measured at 800 MHz as shown in Fig. 3.19

	NS5A	BASP-1	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^{-1})	0.92	0.91	2.27	1.46
$\overline{T}_{1,WFB}$ (s^{-1})	0.70	0.89	0.09	0.14
$\overline{T}_{1,sel}$ (s^{-1})	0.21	0.31	0.06	0.09

in a potential sensitivity gain of a factor of 6, according to Eq. 3.11. This clearly motivates the use of amide ^{1}H start pulse schemes that leave aliphatic and water protons unperturbed throughout the NMR experiment.

The simplest and probably most efficient way to achieve longitudinal relaxation enhancement (LRE) is the use of band-selective ^{1}H pulses, especially in the case of amide protons resonating in a frequency range that is well separated from aliphatic and water protons. Pulse sequences exploiting this concept have been termed band-selective excitation short-transient (BEST) experiments (Schanda et al. 2006b;

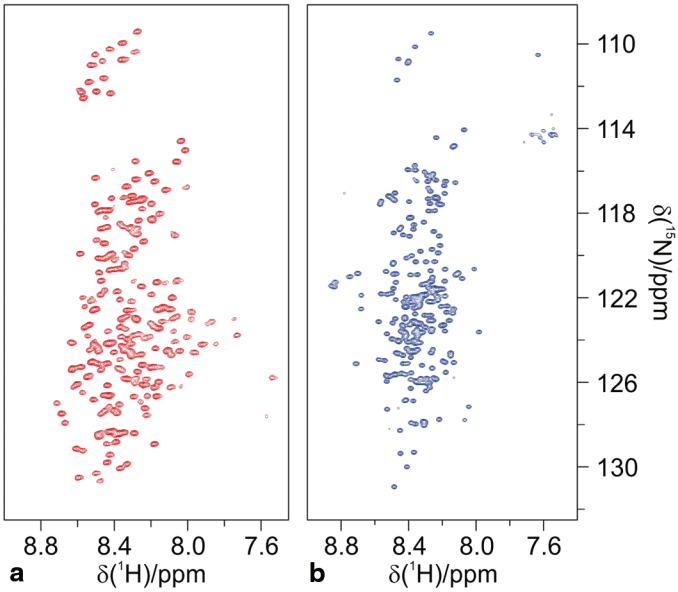


Fig. 3.19 2D ^{1}H-^{15}N BEST-TROSY spectra recorded on two large IDPs (~270 residues), **a** NS5A D2D3 (pH 6.5) and **b** BASP1 (pH 2.0) at 278 K

Lescop et al. 2007; Favier and Brutscher 2011; Solyom et al. 2013). BEST pulse sequence elements are available for all basic pulse sequence elements (building blocks) (Cavanagh et al. 2007), e.g. INEPT, sensitivity-enhanced reverse-INEPT (SE-RINEPT or planar mixing) and TROSY (or double-S_3CT) required for setting up most of the common triple-resonance experiments. In this section we will focus on BEST-TROSY (BT) pulse sequences, which have proven to be particularly useful for the study of IDPs using high magnetic field NMR instruments.

BEST-TROSY combines the advantages of longitudinal relaxation optimization, discussed in the previous section, with those of transverse relaxation-optimized spectroscopy, discussed in Sect. 3.5.4. A special feature of BEST-TROSY is that the proton (H_z) polarization that builds up during the pulse sequence as a consequence of spin relaxation is converted to ^{15}N polarization (N_z) by the final ST2-PT sequence element (Favier and Brutscher 2011). In a conventional TROSY sequence, a large portion of this additional polarization is lost due to ^{15}N T_1 relaxation during the long recovery delay that restores ^{1}H polarization. In BEST-type experiments, however, most of the polarization is conserved because of the short recycle delays used. This provides an additional mechanism for sensitivity and resolution improvement.

The ^{1}H-^{15}N BEST-TROSY sequence and spectra recorded for two large (250 residue) IDPs are shown in Fig. 3.19. Long t_1 acquisition times in the ^{15}N dimension result in highly resolved 2D correlation maps, despite the low chemical shift resolution observed in the ^{1}H dimension. The high resolution obtained in the ^{15}N dimension of a 2D BEST-TROSY spectrum can also be transferred to higher-dimensional BT ^{1}H-^{13}C-^{15}N or BT ^{1}H-^{15}N-^{15}N correlation experiments, as required for sequential

resonance assignment (Sect. 3.7.1), by using the semi-CT editing scheme discussed in Sect. 3.5.2. Furthermore, it has been shown that BEST-TROSY implementations of such experiments provide a 20 to 80 % increased SNR compared to BEST-HSQC versions for different IDP samples at 800 MHz ^{1}H frequency (Solyom et al. 2013). BT also results in a more uniform distribution of peak intensities in the spectrum, as the signal enhancement is efficient for residues in transiently structured peptide regions which tend to be the ones characterized by the smallest intensities in optimal conditions for 2D HN correlation experiments.

6 ^{13}C Detected Experiments

The recent progress in instrumental sensitivity (Kovacs et al. 2005) and the development of new NMR experiments made ^{13}C direct detection a useful tool for biomolecular applications (Felli et al. 2013). The first aspect to consider when performing ^{13}C direct detection on uniformly labelled protein samples consists of the presence of large splittings of the ^{13}C resonances in the direct acquisition dimension due to the presence of large homonuclear one-bond couplings, a feature that definitely represents a novelty with respect to ^{1}H direct detection. Indeed, even though ^{13}C-^{13}C couplings are very useful in the design of multidimensional NMR experiments, they are responsible for large signal splitting which, as evident from Fig. 3.20, needs to be suppressed in order to preserve high resolution in the direct acquisition dimension.

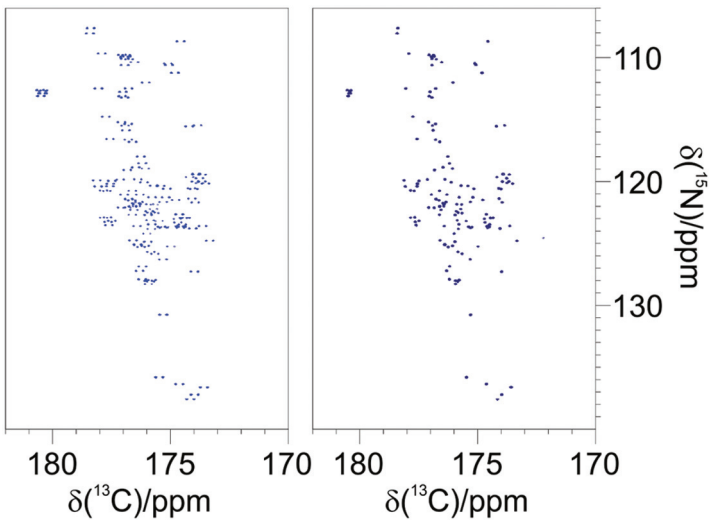


Fig. 3.20 Example of 2D ^{13}C detected CON spectra without and with homonuclear decoupling. The experiments were acquired on α-synuclein (1 mM sample in 20 mM phosphate buffer, pH 6.4, 200 mM NaCl, 0.5 mM EDTA, at 285.5 K) on a 700 MHz Bruker AVANCE equipped with a CPTXO probe

6.1 Homonuclear ^{13}C Decoupling

Achieving homonuclear ^{13}C decoupling in the direct dimension is more complex than in the indirect dimensions because it typically requires the application of radiofrequency pulses at a similar frequency to those of the nuclear spins that are in the process of being detected. A possible solution to this problem consists of sharing the detection period between data acquisition and decoupling pulses (band-selective homonuclear decoupling). This, however, reduces the overall sensitivity of the experiment, and introduces Bloch-Siegert phase shifts (Emsley and Bodenhausen 1990) and decoupling side bands (Waugh 1982) in the spectrum. Recently improved experimental variants exploiting this general idea that however use hard pulses applied at regular intervals during the acquisition period have been proposed (Ying et al. 2014). Alternatively, post-acquisition methods, such as data deconvolution using maximum entropy reconstruction, may be used (Shimba et al. 2003). A last, and arguably the most elegant class of methods for ^{13}C homonuclear decoupling (Felli and Pierattelli 2014b), also known as "virtual decoupling", uses spin-state selection.

Virtual J_{CC} decoupling uses an additional spin evolution delay prior to detection and requires the recording of at least two experiments with different parameter settings. In the first experiment, spin evolution under the J_{CC} coupling is suppressed resulting in an in-phase (IP) line splitting in the detected spectrum, while in the second experiment the J_{CC} coupling evolves for a time $1/(2\,J_{CC})$ resulting in antiphase (AP) line splitting in the final spectrum. A single resonance line (without splitting) is then obtained by calculating the sum and the difference of the two recorded spectra. Finally, the two resulting (sum and difference) spectra are shifted by the amount $J_{CC}/2$ with respect to each other and added up to yield a single line at the correct chemical shift position (Duma et al. 2003a; Duma et al. 2003b; Bertini et al. 2004; Bermel et al. 2006a). This approach, illustrated in Fig. 3.21 for virtual decoupling of $^{13}C^{\alpha}$ in the NMR spectrum of carbonyls, is at the basis of most of the ^{13}C detected multidimensional experiments. In order to work properly, virtual J_{CC} decoupling requires quite uniform J_{CC} couplings in the protein, which is the case for the large one-bond coupling between α carbons ($^{13}C^{\alpha}$) and carbonyls ($^{1}J_{C\alpha C'} \approx 53$ Hz). It is worth noting that the two experiments rely on the same number of pulses and lengths of delays, in order to ensure identical signal loss due to pulse imperfections and spin relaxation effects. Another prerequisite of this technique is that the two nuclear spins, e.g. $^{13}C'$ and $^{13}C^{\alpha}$, are sufficiently well separated to allow their selective manipulation through band-selective pulses. There is also a price to pay in terms of sensitivity for the virtual decoupling method, as during the linear combinations necessary to achieve homonuclear decoupling also the thermal noise is increased, thus leading to a reduction of the signal to noise ratio.

Virtual decoupling can be appended to any pulse sequence ending with in-phase $^{13}C'$ transverse coherence (Bermel et al. 2008). Most of the triple resonance experiments based on $^{13}C'$ direct detection end with a coherence transfer step that involves refocusing of either $^{13}C'$-^{15}N or $^{13}C'$-$^{13}C^{\alpha}$ antiphase coherence (or both). Therefore, virtual decoupling can be implemented without appending an additional block by slightly changing the position of the $^{13}C^{\alpha}$ inversion pulses in the last coherence transfer steps, as illustrated in Fig. 3.21.

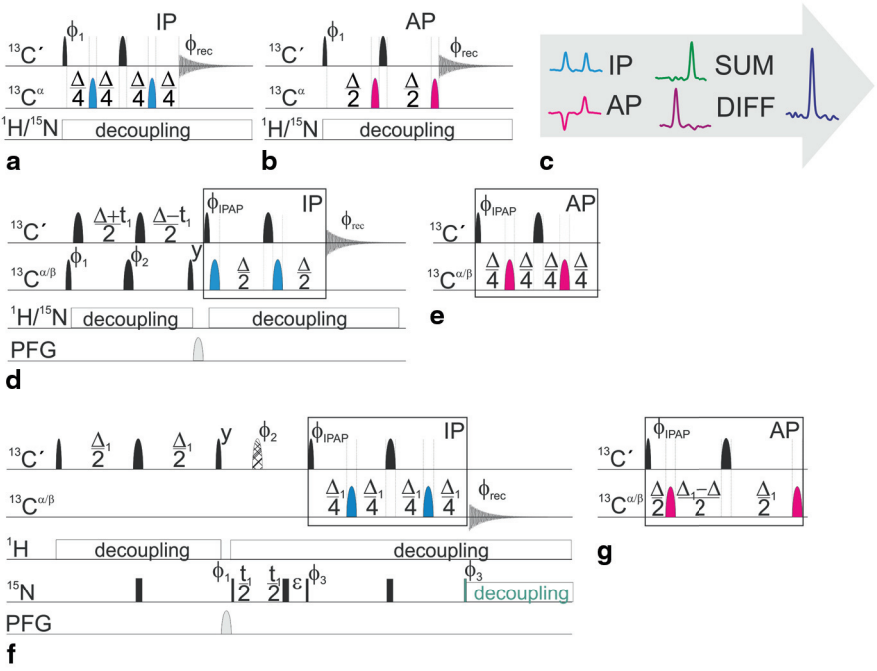


Fig. 3.21 ^{13}C detected 1D and 2D experiments with implemented IPAP decoupling sequence. Implementation of ^{13}C′-^{13}C$^{\alpha}$ IPAP virtual decoupling building blocks in the 1D mode (**a, b**) and 2D mode: CACO (**d, e**) and CON (**f, g**). Band-selective ^{13}C pulses are denoted by rounded rectangles (*narrow* and wide ones represent 90° and 180° pulses, respectively). The pulses are applied along the x-axis unless otherwise noted. The delays are $\Delta=1/(2J_{C\alpha C'})=9$ ms and $\Delta_1=1/(2J_{NC'})=25$ ms. The phase cycles are: (a) ϕ_1=x, -x and ϕ_{rec}=x, -x; (b) ϕ_1=-y, y and ϕ_{rec}=x, -x; (d) ϕ_1=x, -x ϕ_2=4(x), 4(y) ϕ_{IPAP}=2(x), 2(-x) and ϕ_{rec}=x, 2(-x), x, -x, 2(x), -x; (e) ϕ_1=x, -x ϕ_2=4(x), 4(y) ϕ_{IPAP}=2(-y), 2(y) and ϕ_{rec}=x, 2(-x), x, -x, 2(x), -x; (f) ϕ_1=x, -x ϕ_2=2(x), 2(y) ϕ_3=4(x), 4(y) ϕ_{IPAP}=x and ϕ_{rec}=x, -x, x, 2(-x), x, -x, x; (g) ϕ_1=x, -x ϕ_2=2(x), 2(y) ϕ_3=4(x), 4(y) ϕ_{IPAP}=x and ϕ_{rec}=-y, -x, x, 2(-x), x, -x, x. (**c**) The schematic illustration of the post-acquisition processing for obtaining decoupled spectra.

Finally, as these spin-state selective approaches to achieve homonuclear decoupling perform so well, it is worth mentioning that they can also be implemented for heteronuclear decoupling (Kern et al. 2008; Bermel et al. 2009a). Indeed, a clever variant to achieve heteronuclear ^{15}N decoupling has been proposed both for ^{1}H^{N} and ^{13}C′ direct detection experiments. This can be useful when the ^{13}C′ coherence lifetimes allow for long acquisition times or to reduce the radiofrequency load/heating on the ^{15}N channel during acquisition, which is particularly useful in the fast pulsing regime (Gil et al. 2013).

6.2 Starting Polarization Source

The other important point to consider in the design/choice of a ^{13}C detected experiment is the starting polarization source. From Sect. 3.1.2, it is immediately clear that a relevant contribution to increase the sensitivity of ^{13}C detected experiments

can come from the use of ^{1}H as the starting polarization source (Shimba et al. 2004; Bermel et al. 2009a; O'Hare et al. 2009). In fact, the large one-bond scalar couplings can be easily used to transfer polarization from ^{1}H to the directly bound heteronuclei while still keeping experiments "exclusively heteronuclear". This means that only heteronuclear chemical shifts are frequency labelled in all the spectral dimensions to take advantage of their larger chemical shift dispersion compared to that of ^{1}H, while still exploiting the larger ^{1}H polarization deriving from its higher gyromagnetic ratio. Therefore, several exclusively heteronuclear NMR experiments, such as the (H)CBCACON and (H)CBCANCO experiments (Bermel et al. 2009a), feature ^{1}H as a starting polarization source.

With protons exploited only to increase the experimental sensitivity, it is worth thinking about the possibility of implementing ^{1}H longitudinal relaxation enhancement techniques to reduce the duration of NMR experiments or to increase the sensitivity per unit of time or the resolution. Actually, in these experiments it is very easy to manipulate different sets of ^{1}H spins in a selective way, an essential requirement to promote LRE. Indeed, a simple modification of the initial INEPT block can be introduced to flip-back all the other spins not directly involved in the coherence transfer pathway (H^{flip} method (Bermel et al. 2009b; Bertini et al. 2011b)).

The solution described above for amide protons based on the use of band-selective pulses (BEST approach), can also be implemented in ^{13}C detected experiments (Nováček et al. 2011; Gil et al. 2013). Amide protons indeed resonate in a quite isolated region of the ^{1}H spectra, well separated from the water resonance; they can therefore be selectively manipulated leaving both water and aliphatic resonances unperturbed.

As discussed above, the LRE effect deriving from exchange processes with the solvent is very pronounced in IDPs and adds to that deriving from ^{1}H-^{1}H NOEs, which may contribute at low temperature conditions or in the presence of partially structured elements, even if to a minor extent. Approaching physiological conditions (high temperature and neutral pH), the LRE effect is so large that the recovery of amide polarization to equilibrium becomes extremely fast and recycle delays between transients are therefore no longer necessary. In fact, in this type of experiment the longitudinal recovery already starts before the end of the pulse sequence, as protons get perturbed only in the very initial part of the experiment. Therefore, the optimal set-up consists of an inter-scan delay equal to zero as implemented for the $H^{N\text{-}BEST}$ CON experiment (Gil et al. 2013). Its high sensitivity enables 2D spectra to be collected in a couple of minutes, as shown in Fig. 3.22, making it an ideal technique for *in-cell* applications. However, the use of $^{1}H^{N}$ as the starting polarization source renders the experiment susceptible to the loss of information about prolines and reintroduces a dependence on exchange rates of amide protons with the solvent that impacts the experimental sensitivity.

Alternatively, $^{1}H^{\alpha}$ can be exploited as the starting polarization source. In this way, all backbone sites can be sampled (including prolines) and exchange processes with the solvent are avoided. On the other side of the coin, this also means that only LRE effects deriving from ^{1}H-^{1}H NOEs can be exploited. Therefore, LRE effects become sizeable either at lower temperature or for partially structured proteins. The most convenient way of achieving LRE for $^{1}H^{\alpha}$ consists of exploiting the one-bond

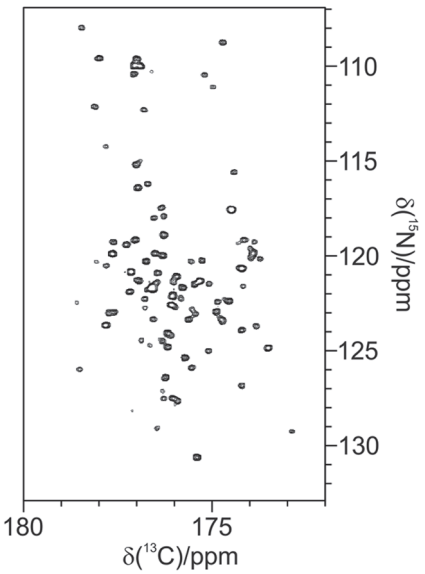


Fig. 3.22 The 2D ^{13}C-^{15}N $H^{N\text{-}BEST}CON$ spectrum of α-synuclein overexpressed in *E. coli* cells acquired in 20 min

scalar coupling ($^{1}J_{C\alpha H\alpha}$) with the attached carbon ($^{13}C^{\alpha}$) to selectively manipulate $^{1}H^{\alpha}$ and flip-back all other protons that are not actively used in the magnetization transfer pathway (H^{flip} method) (Bermel et al. 2009b). Therefore, this trick can be easily implemented in any $^{1}H^{\alpha}$-start pulse sequence to achieve LRE. Increasing the temperature also provides a simple tool to increase the longitudinal relaxation of $^{1}H^{\alpha}$ protons that does not require selective manipulation of $^{1}H^{\alpha}$ protons.

Alternatively, ^{13}C-start versions of such ^{13}C detected experiments can be designed. Indeed ^{13}C spins are not directly involved in chemical exchange processes and are characterized by large chemical shift dispersion. The high flexibility of IDPs results in relatively short ^{13}C T_1 values, which are on the order of seconds, so that recovery delays remain sufficiently short. In addition, the ^{1}H-^{13}C NOE effect can be exploited in the case of highly flexible IDPs to enhance the signal-to-noise ratio just by irradiating protons during the recovery delay (Bertini et al. 2011b). Therefore, in many cases of practical interest, the sensitivity of ^{13}C-start-^{13}C detected versions of experiments, in particular for the simplest 2D experiments such as CON, CACO and CBCACO, is sufficient to obtain informative spectra (Bermel et al. 2005).

7 From 2D to 3D: From Simple Snapshots to Site-resolved Characterization of IDPs

Multidimensional NMR experiments that encode chemical shift information in several indirect dimensions provide the necessary resolving power to separate correlation peaks from different sites in an IDP. In this section, the general strategy used to achieve site-specific resonance assignment of the NMR signals will be discussed,

and a selection of commonly used multidimensional experiments for sequential resonance assignment, as well as complementary techniques for amino acid type identification, will be presented.

7.1 Sequential NMR Assignment: The General Strategy

In order to extract site-specific information from the NMR experiments, each resonance observed in the spectra needs to be associated to a specific nuclear spin of the protein. This task is commonly called sequence-specific (or sequential) resonance assignment. In the early days of biomolecular NMR spectroscopy, resonance assignment was based on the information contained in 2D homonuclear ^{1}H correlation experiments (Wüthrich 1986). The combined analysis of a set of through-bond (COSY, TOCSY) and through-space (NOESY, ROESY) spectra allows sequence-specific assignment of well-folded proteins of moderate size (~100 residues). In the case of highly disordered proteins and IDPs, which are characterized by low chemical shift dispersion in the 2D ^{1}H spectra, this strategy is limited to peptides of less than a few tens of residues. As a consequence, NMR studies of IDPs require uniform ^{13}C and ^{15}N labelled samples that enable the use of 3D (or higher-dimensional) through-bond-only experiments for sequential resonance assignment purposes (Dyson and Wright 2001; Eliezer 2009). These experiments are based on a series of coherence transfer steps, which exploit the large one- and two-bond scalar couplings (Fig. 3.23) (Sattler et al. 1999; Cavanagh et al. 2007).

Starting from a chosen polarization source, typically amide ^{1}H, aliphatic ^{1}H, or ^{13}C, a series of subsequent coherence transfer pathway steps allows the correlation of NMR frequencies of different nuclear spins in the protein with high efficiency. Ideally this strategy could be used to transfer coherence (and thus information) all along the polypeptide chain. In practice, due to spin relaxation effects, coherence transfers are limited only to nuclear spins within the same or neighbouring residues. Therefore, a set of spectra providing complementary connectivity information is required for sequential resonance assignment. The most widely used experiments are introduced in the next sections. They can be divided into three categories: (i) intra-residue, (ii) sequential, and (iii) bi-directional correlation experiments, according to the type of information they contain. A schematic drawing showing

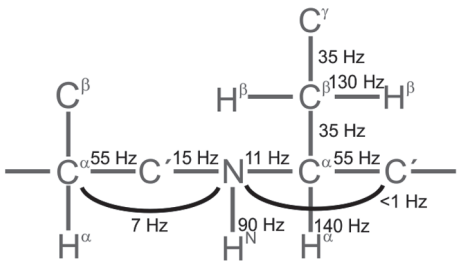


Fig. 3.23 Schematic representation of a protein and the size of the ^{1}J and ^{2}J coupling constants that are frequently used for magnetization transfer in ^{13}C, ^{15}N labelled proteins

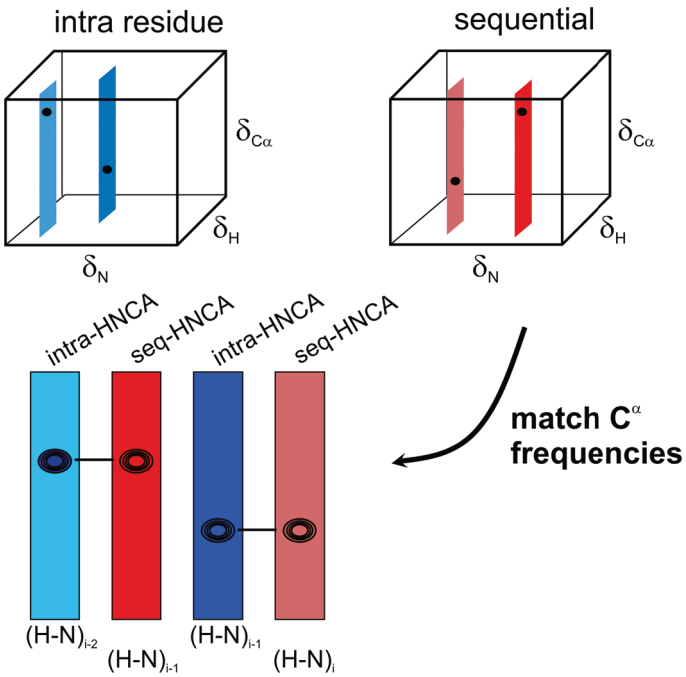


Fig. 3.24 A sequential walk through the NMR spectra for the sequence-specific assignment is illustrated on the example of 3D intra-HNCA and HN(CO)CA (seq-HNCA) experiments

how a pair of such spectra can be used in the sequential resonance assignment procedure is reported in Fig. 3.24.

The basic strategy for sequential assignment is that the different nuclei correlated in a 2D HN or 2D CON fingerprint spectrum, as introduced in Sect. 3.4.2, are connected in a 3D data set with one (or more) additional nuclei, e.g. $^{13}C^{\alpha}$, $^{13}C^{\beta}$, $^{13}C'$, ^{15}N, or $^{1}H^{\alpha}$, that are frequency labelled in the third dimension (indicated as X nuclear spins). ^{1}H-^{15}N or $^{13}C'$-^{15}N pairs are then recognized as belonging to neighbouring residues when the corresponding X frequencies match. In the case of identical or similar resonance frequencies of two or more X nuclei, ambiguities can be solved by combining the information obtained from different pairs of intra- and inter-residue 3D H-N-X or CO-N-X spectra. As a result of this assignment step, chains of sequentially connected residues (^{1}H-^{15}N or $^{13}C'$-^{15}N pairs) are identified. This so-called "sequential assignment walk" might be interrupted by missing correlation peaks, either due to unobservable (line-broadened) NMR signals, e.g. because of pronounced conformational exchange processes or chemical exchange at physiological conditions, missing correlation information, e.g. amide ^{1}H in prolines and/or remaining frequency ambiguities; the information provided by 3D H-N-X and 3D CO-N-X spectra can therefore be combined in order to reduce interruptions in the sequence-specific assignment and ambiguities contributing in this way to the completeness and accuracy of assigned resonances. The final assignment step

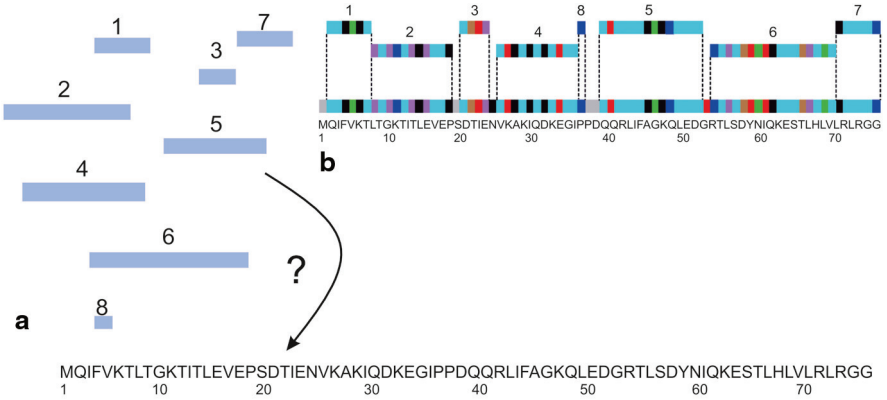


Fig. 3.25 Sequence-specific assignment procedure. The fragments obtained from the sequential walk through the spectra (**a**) are combined with the primary sequence, and sequence-specific assignment is achieved (**b**) by exploiting the amino-acid type information

consists of mapping the identified peptide fragments (correlated NMR nuclei) to their specific positions in the protein sequence. In order to do so, amino acid type information on some (or all) of the residues that form the fragment is required, as illustrated in Fig. 3.25. Such amino acid type information is obtained from side-chain ^{13}C chemical shifts, or from specifically tailored experiments that use spectral editing techniques to differentiate spin-coupling topologies in the amino acid side chains (see Sect. 3.8.4).

8 Sequential NMR Assignment: 3D Experiments

8.1 $^{1}H^{N}$ Detected Experiments

The most common and established resonance assignment approach for proteins uses a set of $^{1}H^{N}$-detected 3D correlation experiments (Ikura et al. 1990; Kay et al. 1990). The resulting spectra share as common frequencies those of amide protons in the direct dimension and amide nitrogen atoms in one of the indirect dimensions. The most useful experiments and their underlying coherence transfer pathway(s) are briefly presented in this section (Fig. 3.26). As discussed in Sect. 3.5.4, most amide ^{1}H detected experiments for IDP studies should be implemented as BEST or BEST-TROSY versions in order to enhance experimental sensitivity and spectral resolution (Lescop et al. 2007; Solyom et al. 2013).

The most sensitive coherence transfer pathway correlates the amide $^{1}H^{N}$ and ^{15}N of amino acid i with the $^{13}C'$ of the preceding ($i-1$) residue. This experiment, called 3D HNCO (Kay et al. 1990; Grzesiek and Bax 1992b; Schleucher et al. 1993; Solyom et al. 2013), is often used to "count" the number of cross-peaks and therefore to estimate the number of resolved residues that are observed. Additional transfer steps

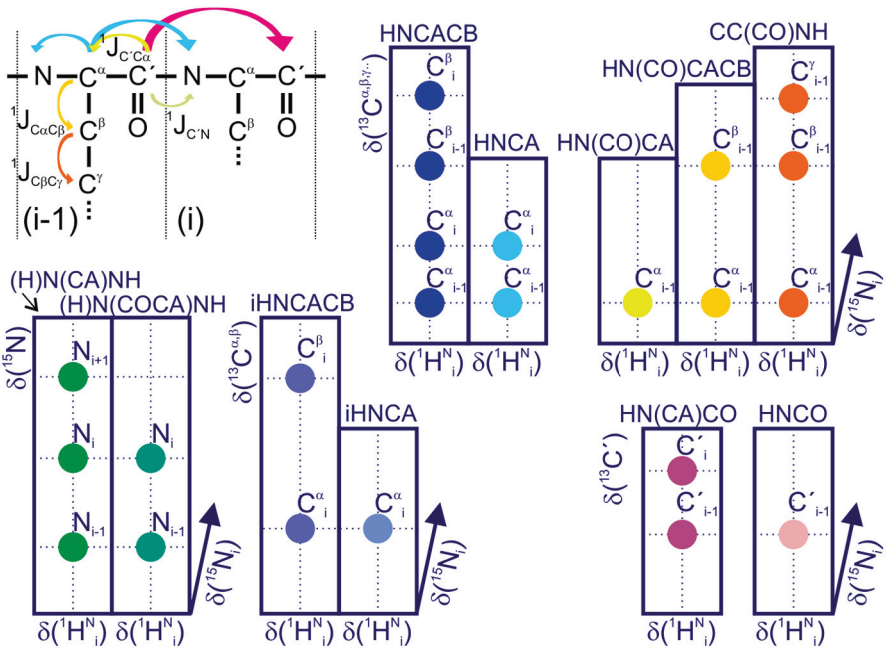


Fig. 3.26 Schematic representation of the $^{1}H^{N}$ detected 3D experiments used for the sequence-specific assignment of IDPs

allow correlating the amide $^{1}H^{N}$ and ^{15}N with the $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ of the preceding ($i-1$) residue, in the so-called 3D HN(CO)CA (Bax and Ikura 1991; Grzesiek and Bax 1992b; Solyom et al. 2013) and 3D HN(CO)CACB (Grzesiek and Bax 1992a; Yamazaki et al. 1994; Solyom et al. 2013) experiments. Note that the intrinsic sensitivity decreases with each additional transfer step, making HN(CO)CACB the least sensitive experiment of the series. For sequential resonance assignment, the information obtained from the experiments correlating nuclei of neighbouring amino acids needs to be complemented with that provided by experiments correlating the amide $^{1}H^{N}$ and ^{15}N with the $^{13}C'$, $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ of the same (i) residue. This can be achieved, for example, by acquiring 3D HNCA (Kay et al. 1990; Lescop et al. 2007; Grzesiek and Bax 1992b), 3D HNCACB (Wittekind and Mueller 1993; Muhandiram and Kay 1994; Lescop et al. 2007) and 3D HN(CA)CO (Clubb et al. 1992; Kay et al. 1994; Briand et al. 2001; Lescop et al. 2007) experiments belonging to the class of "bi-directional experiments". All of these experiments result in two cross-peaks per residue because of the similar size of the $^{1}J_{NC\alpha}$ and $^{2}J_{NC\alpha}$ coupling constants (Fig. 3.23), one corresponding to the desired intra-residue correlation, the other one to the sequential correlation already detected in the previous set of experiments. In order to avoid recording of redundant information and to limit the risk of peak overlaps in the spectra, purely intra-residue correlation experiments, 3D iHNCA (Brutscher 2002; Nietlispach et al. 2002; Solyom et al. 2013), 3D iHNCACB (Brutscher 2002; Nietlispach et al. 2002; Solyom et al. 2013) and 3D iHN(CA)CO (Nietlispach 2004;

Solyom et al. 2013), have been proposed as an alternative to the bi-directional experiments discussed above. The bi-directional and intra-residue experiments are less sensitive than their sequential counterparts and thus, as a rule of thumb, the number of scans should be at least doubled. Recently, improved HNCA+, HNCACB+, and HNCO+ pulse sequences have been introduced to perform bi-directional experiments with improved sensitivity for the sequential correlations (Gil et al. 2014). These experiments are especially useful for the study of fragile IDP samples as they allow the complete information required for sequential assignment to be retrieved from a single set of 3D experiments.

An additional class of experiments, especially useful for IDP samples, is referred to as 3D H-N-N, since their main characteristic is that ^{15}N frequencies are labelled in both indirect dimensions. This allows the sequential assignment walk to be performed by matching ^{15}N frequencies, which, as discussed in Sect. 3.4.2, experience the largest chemical shift dispersion for IDPs lacking a stable structure. A drawback of these pulse schemes is that they require more or longer transfer steps, thus resulting in reduced sensitivity especially for IDP residues involved in transient secondary structure formation. The coherence transfer pathways for the 3D (H)N(COCA)NH and 3D (H)N(CA)NH experiments (Weisemann et al. 1993; Grzesiek et al. 1993b; Bracken et al. 1997; Panchal et al. 2001; Kumar and Hosur 2011) are shown in Fig. 3.27. In the 3D (H)N(COCA)NH, the HNCO sequence is extended passing coherence from $^{13}C'$ to $^{13}C^{\alpha}$ and then to ^{15}N of the neighbouring amino acid, resulting in the correlation of $^{1}H^{N}$ and ^{15}N of amino acid i with ^{15}N of amino acid i and $i-1$, while the 3D (H)N(CA)NH is an extension of HNCA, yielding correlations with ^{15}N of residues $i-1$, i, and $i+1$. In the case of highly flexible IDPs resulting in large T_2 values, these pulse schemes can be optimized to detect mainly the sequential correlations, while suppressing to a large extent the "diagonal" N_i-N_i correlation peak (Grzesiek et al. 1993b).

A last class of experiments allows the assignment to be extended to the side chain ^{13}C resonances and the provision of valuable information on the side chain length and ^{13}C chemical shifts, important for distinguishing between amino acid types. The most widely used pulse sequence is called 3D (H)C(CO)NH-TOCSY (Montelione

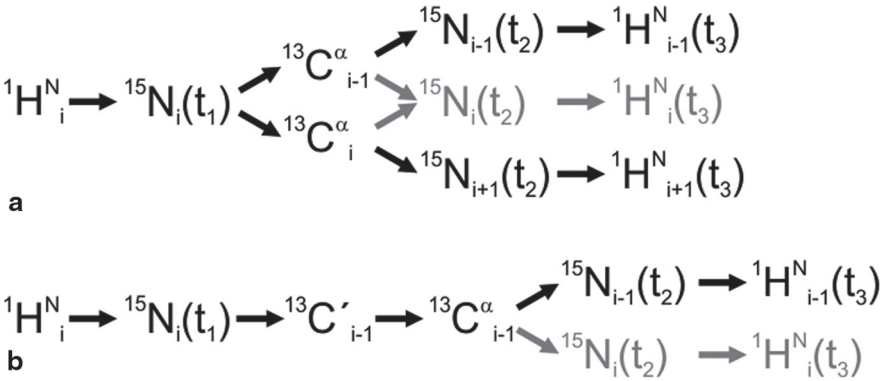


Fig. 3.27 Polarization transfer pathway of (H)N(CA)NH (**a**) and (H)N(COCA)NH (**b**) experiments

et al. 1992; Logan et al. 1992; Logan et al. 1993; Gardner et al. 1996; Grzesiek et al. 1993a). The sequence starts from aliphatic ^{1}H polarization of residue *i*, which is transferred via TOCSY and INEPT steps to the amide group of the following ($i+1$) residue for final $^{1}H^{N}$ detection.

8.2 ^{13}C Detected Experiments

As discussed in Sect. 3.4, the 2D HN and CON spectra provide complementary tools for the investigation of IDPs. Therefore, similarly to the H^{N}-based experiments discussed above, $^{13}C'$ detection can be extended to higher dimensions, thus providing the required correlation information for sequential resonance assignment of the protein.

A suite of 3D experiments based on $^{13}C'$ detection is shown in Fig. 3.28. The building blocks used for the coherence transfer steps are very similar to those employed in the analogous $^{1}H^{N}$ detected experiments. ^{1}H polarization can be used as a starting source to increase the sensitivity of $^{13}C'$ detected experiments, while still keeping the experiments "exclusively heteronuclear". In all of the spectra recorded with these experiments, the detection dimension is the $^{13}C'$ of residue *i* (C'_i), while the ^{15}N of residue $i+1$ (N_{i+1}) is frequency labelled in one of the indirect dimensions. These C'_i-N_{i+1} correlations are then dispersed in the third dimension by the chemical shift of one (or more) additional nuclear spin(s). In a first set of experiments, CACON and CBCACON (Bermel et al. 2006a; Bermel et al. 2006c; Bermel et al. 2009a), C^{α}_i only or both C^{α}_i and C^{β}_i are edited in the third dimension, resulting in one set of correlation peaks per residue, similarly to the sequential $^{1}H^{N}$ detected

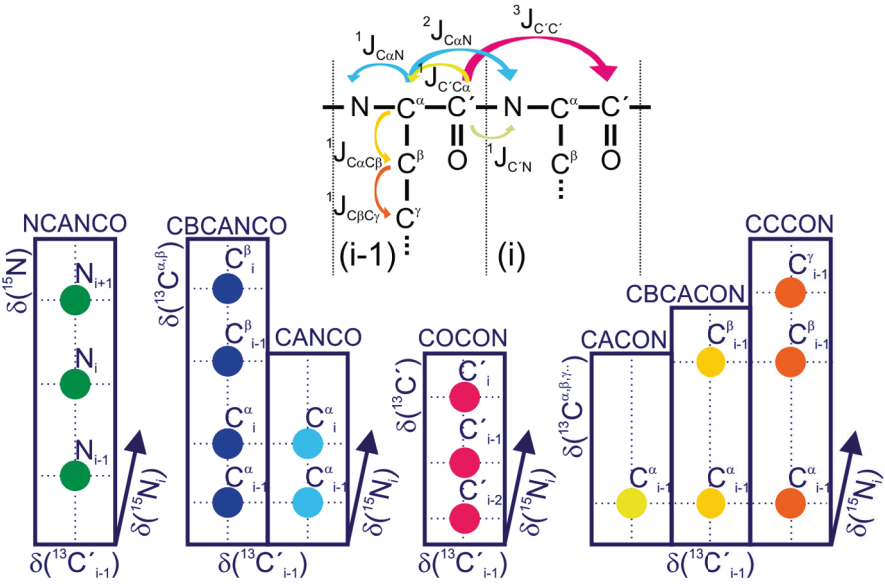


Fig. 3.28 Schematic representation of the $^{13}C'$ detected 3D experiments used for the sequence-specific assignment of IDPs

experiments. In addition, the CCCON-TOCSY experiment (Bermel et al. 2006c; Bermel et al. 2009a) allows the ^{13}C chemical shift information to be extended to all aliphatic side chain carbons of residue *i*. As a result of this set of experiments, each residue of the protein is associated to a set of C^{aliph}_i-C'_i-N_{i+1} correlations. The number of aliphatic carbons detected and their chemical shifts provide information on the amino acid type of the corresponding residue.

The complementary information necessary for sequence-specific assignment is then obtained by acquiring a second set of $^{13}C'$-detected 3D experiments named CANCO and CBCANCO (Bermel et al. 2006c; Bermel et al. 2009a) that correlate the C'_i-N_{i+1} to C^{α}_{i+1} and C^{α}_i (CANCO) or $C^{\alpha}_{i+1}/C^{\beta}_{i+1}$ and C^{α}_i/C^{β}_i (CBCANCO) resulting in two correlation peaks per residue if only C^{α} chemical shifts are detected in the third dimension, or four, if both C^{α} and C^{β} are detected. These experiments thus belong to the class of bi-directional correlation experiments. In addition, the 3D (H)N(CA)NCO (Bermel et al. 2009a) spectrum can be recorded to obtain correlations involving an additional ^{15}N nuclear spin. This experiment thus belongs to the CO-N-N class of spectra (in analogy to the H-N-N ones introduced above for H^N detected experiments). In the 3D (H)N(CA)NCO experiment, the C'_i-N_{i+1} pair is further correlated in the third dimension with the ^{15}N of residues i, $i+1$, and $i+2$. Finally, the 3D COCON experiment (Bermel et al. 2006b) was developed to correlate the C'_i-N_{i+1} pair with carbonyls of residues i, $i+1$, and $i-1$, providing additional complementary information in this way. The MOCCA mixing scheme significantly improves the sensitivity of this experiment (Bermel et al. 2006b; Felli et al. 2009).

If the whole set of experiments is performed, sequential resonance assignment is obtained by simultaneous matching of C^{α}, C^{β}, C', and N chemical shifts from intra-residue and sequential correlations. The requirement for simultaneous frequency match of four different nuclear spins provides a robust way of solving assignment ambiguities due to resonance overlap in spectra of IDPs.

As also prolines are detected in these spectra, $^{13}C'$ direct detection provides an ideal tool for complete sequence-specific assignment of an IDP, provided that sensitivity is sufficient to enable acquisition of the whole set of experiments in a reasonable overall measurement time (Bermel et al. 2005; Bermel et al. 2006a). Our experience with several IDPs characterized by high flexibility indicates that the complete set of 3D experiments described here can be collected with cryogenic probes optimized for ^{13}C detection and yields the complete sequence-specific assignment without any other additional information. For samples of limited solubility (~100 μM) or using less optimal spectrometer hardware, it is still possible to acquire the most sensitive 2D (CACO, CBCACO, CON) (Bermel et al. 2005) and 3D experiments (CBCACON, CCCON) (Bermel et al. 2006a) in order to complement the spectral information obtained from a series of $^1H^N$ detected experiments.

8.3 Aliphatic ^1H Detected Experiments

An alternative to the two previously discussed approaches based on the two most well resolved 2D spectra to study IDPs (HN and CON), based either on $^1H^N$ or $^{13}C'$ direct detection, consists in focusing on $^1H^{\alpha}$ detection, as proposed for the

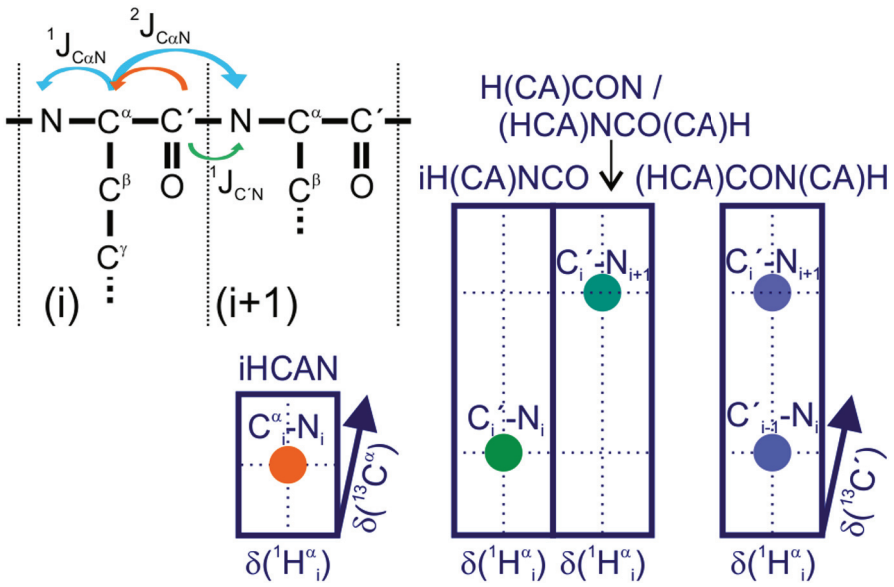


Fig. 3.29 Schematic representation of the $^{1}H^{\alpha}$ detected 3D experiments used for the sequence-specific assignment of IDPs

assignment of proline-rich regions (Kanelis et al. 2000). More recently, a set of five triple resonance experiments was developed for sequence-specific resonance assignment of IDPs (Mäntylahti et al. 2010; Mäntylahti et al. 2011). The coherence transfer pathways and correlated frequency information obtained from these experiments are shown in Fig. 3.29. The 3D iH(CA)NCO, iHCAN and (HCA)NCO(CA)H experiments were designed for spin system identification, while the 3D H(CA)CON and (HCA)CON(CA)H provide the complementary connectivity information. The $^{1}H^{\alpha}$ signals are clustered in a narrow chemical shift region where also the water proton spins resonate. Resolution in the direct dimension is therefore quite limited, in particular for amino acids of the same kind. Furthermore, excellent water suppression performance is mandatory. Alternatively, the protein can be dissolved in a fully deuterated solvent (99.99 % D_2O), which is characterized by higher viscosity with respect to H_2O and causes an increase in the transverse relaxation rates of the nuclear spins in the protein sample. This approach has been successfully applied to an IDP of about 110 amino acids (Mäntylahti et al. 2011). $^{1}H^{\alpha}$-detected experiments might be considered under conditions where H^{N} detection of the IDP is not feasible, e.g. at near physiological conditions, and when the protein concentration or the experimental setup does not provide the required sensitivity for ^{13}C detection.

8.4 Experiments for Amino Acid Type Editing or Selection

In addition to the experiments discussed in the previous sections, which provide sequential correlation information connecting nuclear spins of neighbouring

residues, information on the amino acid type of a given residue is necessary to complete the sequential resonance assignment of a protein. Amino acid type information may be obtained from NMR data by the selective incorporation of labelled (or unlabelled) amino acids at the protein expression level (McIntosh and Dahlquist 1990; Cowburn et al. 2004; Tong et al. 2008), from $^{13}C^{\alpha}$, $^{13}C^{\beta}$ and ^{13}C side chain chemical shift data (Grzesiek and Bax 1993), or from specifically tailored NMR experiments that exploit the particular spin-coupling topologies and chemical shifts in the different amino acid side chains (Wittekind et al. 1993; Olejniczak and Fesik 1994; Yamazaki et al. 1995; Rao et al. 1996; Rios et al. 1996; Dötsch et al. 1996a, 1996b; Pellecchia et al. 1997; Schubert et al. 1999, 2001a, 2001b; Pantoja-Uceda and Santoro 2008). The latter approach has the advantage that unambiguous amino acid type information is obtained from a series of NMR experiments recorded on a single uniformly $^{13}C/^{15}N$ labelled sample. Such experiments, which often rely on band selective pulses to perturb only specific frequency ranges, benefit from the narrow chemical shift ranges of side chain resonances typical of IDPs. A selection of the most common experiments available for amino acid type discrimination is presented in this section.

Conceptually we can distinguish between experiments for *amino acid type selection* and those performing *amino acid type editing*. In the first approach (amino acid type selection), only correlation peaks from specific amino acid types are detected, while in the second approach (amino acid type editing), correlation peaks from different classes of amino acids are separated in different NMR subspectra. The two classes of experiments share, as common features, several ingenious ways to distinguish different amino acids exploiting peculiar spectroscopic features of amino acid side chains such as characteristic chemical shifts, coupling topologies, side chain length, etc., as described in the early NMR literature (Wittekind et al. 1993; Olejniczak and Fesik 1994; Yamazaki et al. 1995; Dötsch et al. 1996a, 1996b; Löhr and Rüterjans 1995; Pellecchia et al. 1997).

The MUSIC (multiplicity selective in-phase coherence transfer) strategy proposed by the Oschkinat group (Schubert et al. 1999; Schubert et al. 2001a; Schubert et al. 2001b), recently extended to ^{13}C direct detection experiments (CAS-NMR) (Bermel et al. 2012a), provides the most complete set of amino acid-selective experiments. Different variants of two basic experiments (HN(CO)CACB and CBCACON) allow the selection of correlations deriving only from specific amino acid types from the basic 2D HN and CON spectra respectively (Fig. 3.30). The resulting spectra are very simple, contributing in this way to reducing the problem of cross-peak overlap. The various pulse schemes mainly differ in the number and lengths of transfer steps required and thus their overall sensitivity. As a rule of thumb, the experiments for shorter side chains are more sensitive than the corresponding pulse sequences for selecting amino acids with longer side chains. The different pulse sequence elements designed to select correlations of specific amino acids can also be implemented in "bidirectional" experiments (HNCACB, CBCANCO), providing additional information in this way.

As mentioned before, prolines are abundant in IDPs and can be easily detected in ^{13}C detected experiments; the identification of proline-neighbouring residues is of particular importance for resonance assignment strategies based on $^{1}H^{N}$ detected

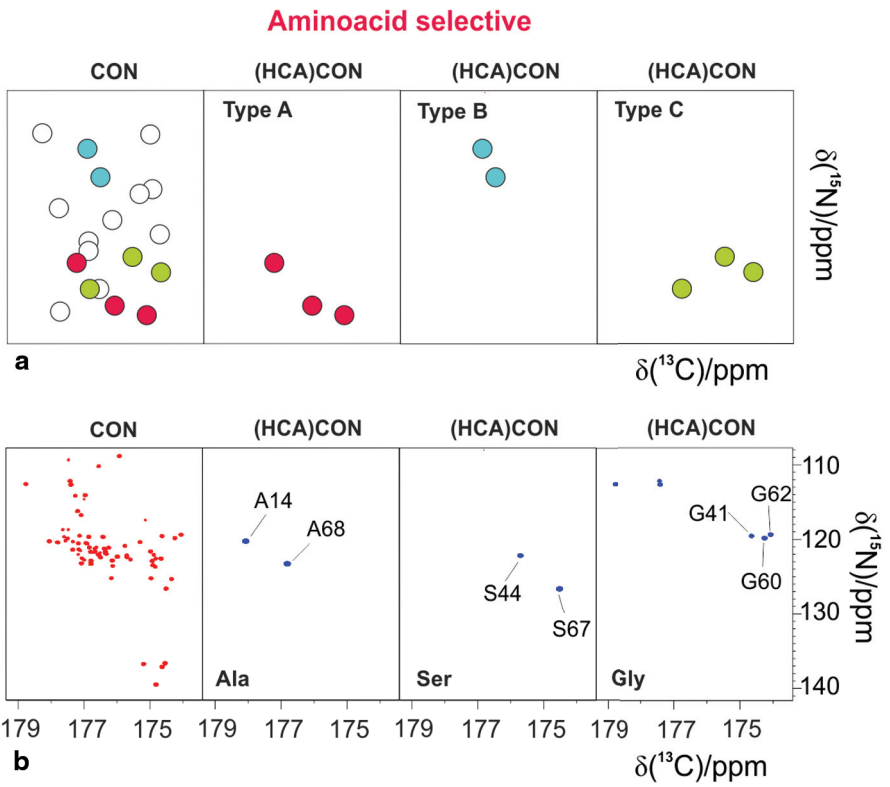


Fig. 3.30 Amino acid-selective experiments to simplify crowded spectra. **a** A schematic representation of a 2D CON spectrum and three 2D (HCA)CON spectra, in which three different amino acids (types A, B and C) are selected. **b** CAS-NMR spectra of Cox17. From *left* to *right*: reference 2D CON spectrum, and (CA)CON spectra from experiments selecting Ala, Ser and Gly of Cox17. The experiments were acquired on Cox17 protein (1.8 mM sample in 20 mM phosphate buffer, pH 7.0, 0.25 mM EDTA, 20 mM DTT, at 307 K) on a 700 MHz Bruker AVANCE spectrometer equipped with a CPTXO probe (Felli et al. 2013)

pulse schemes (Tompa 2002). Proline-selective experiments based on BEST-TROSY HN(COCAN) and iHN(CAN) pulse sequences, exploiting the particular ^{15}N chemical shift of prolines in IDPs, have therefore been proposed that yield increased sensitivity with respect to their corresponding MUSIC counterparts for the identification of proline-neighbouring residues (Solyom et al. 2013). ^{13}C detected variants of the experiments tailored for prolines have also been proposed (Bermel et al 2012a).

Amino acid type editing differs from amino acid-type selection by the fact that correlation peaks for all (observable) residues are detected in the final spectrum, while amino acid type information is obtained by sign encoding. In the simplest type of amino acid type editing, the correlation peaks corresponding to a specific class of amino acids are of opposite sign with respect to all others, thus allowing simple discrimination of this particular class of amino acids. In addition, if a second (reference) experiment is recorded where all NMR signals have the same sign, peaks

belonging to different amino acid type classes can be separated in different spectra at the processing level by addition and subtraction of the two recorded data sets. Such a simple sign encoding has been proposed for many basic correlation experiments involving C^{α} and/or C^{β} carbons, requiring only slight modifications of the original pulse sequences (Grzesiek and Bax 1992b; Panchal et al. 2001; Brutscher 2004b). The same concept of sign encoding can also be used to differentiate between more than two classes of amino acids by using HADAMARD spectroscopy (Kupce et al. 2003; Brutscher 2004a). In short, to distinguish between n amino acid classes, HADAMARD NMR spectroscopy requires the recording of n spectra with different sign encoding according to a particular HADAMARD scheme, also called a HADAMARD matrix. HADAMARD matrices exist for $n=2, 4, 8, 12, \ldots$ (multiples of 4). The same HADAMARD matrix employed for encoding is then also used at the processing level to calculate linear combinations of the recorded data set resulting in separate spectra for each amino acid class. This concept (with $n=8$) is used in the HADAMAC (Hadamard-encoded amino acid type editing) experiments (Lescop et al. 2008; Feuerstein et al. 2012; Pantoja-Uceda and Santoro 2012). The sequential HADAMAC (sHADAMAC) experiment, based on the sensitive (H)CBCACONH (Grzesiek and Bax 1992a; Grzesiek and Bax 1993) correlation sequence, provides

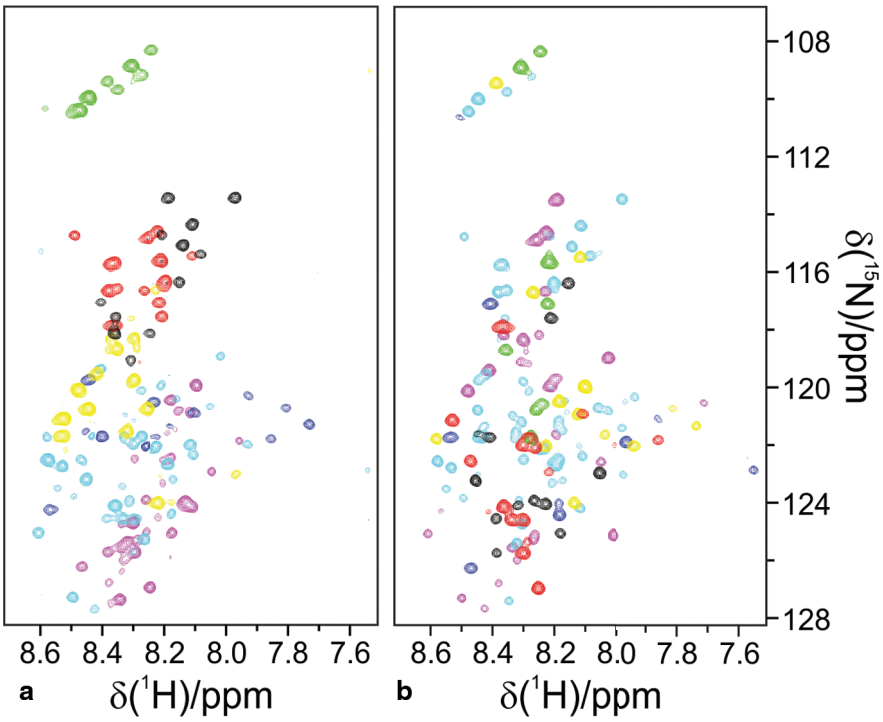


Fig. 3.31 **a** iHADAMAC and **b** HADAMAC spectra acquired on an 800 MHz spectrometer at 298 K on a 90 μM sample of a NS5A fragment. 100 complex points were recorded in the ^{15}N indirect dimension for a spectral width of 2000 Hz using semi-CT editing for both spectra. The total acquisition times were set to 12 h (iHADAMAC) and 2 h (sHADAMAC). The seven subspectra corresponding to the different amino acid type classes are color-coded and superposed on the same graph

amino acid type information for the residue preceding the detected amide group, and allows differentiation between seven classes of amino acids: (1) Val-Ile-Ala, (2) Gly, (3) Ser, (4) Thr, (5) Asn-Asp, (6) His-Phe-Trp-Tyr-Cys and (7) Arg-Glu-Lys-Pro-Gln-Met-Leu (Fig. 3.31b). The same approach has also recently been extended to intra-residue amino acid type editing (iHADAMAC, Fig. 3.31a) (Feuerstein et al. 2012).

8.5 Automated Assignment

The establishment of a vast series of multidimensional NMR experiments has opened the way to the development and improvement of automated assignment programs and to their application to IDPs. The very low chemical shift dispersion and high chemical shift degeneracy typical of disordered proteins have always been the major limitations to the performance of such programs. However, in recent years many efforts have been made to increase the robustness and reliability of automatic assignment procedures and several algorithms are now ready for use. Out of the many, MARS (Jung and Zweckstetter 2004), FLYA (López-Méndez and Güntert 2006) and TSAR (Zawadzka-Kazimierczuk et al. 2012a) are some examples of the most well-established and promising algorithms that can be exploited to assign IDPs.

In general, the programs require as input only a set of experimental peak lists and the primary sequence of the protein to be assigned. The output can be very different depending on the particular algorithm employed. For example, the assignment, usually provided to the user as a text file, can be accompanied by graphical representations and/or additional text files containing, for instance, information about the reliability of the assignment and/or suggestions related to a missing or ambiguous assignment. In addition, to facilitate the manual validation of the assignment, some programs provide assignment results also in suitable formats to be directly loaded and read by tools for spectral analysis.

Most importantly, recent improvements have made the input files more and more flexible, accepting peak lists from any combination of multidimensional experiments provided they are written according to a simple but specific set of rules that describes the magnetization transfer pathways employed. Using all the experimental data simultaneously, also incomplete peak lists in which some peaks are missing can be used; furthermore, peak lists containing redundancy of information can be exploited to increase the reliability of the chemical shift assignment.

Finally, automatic resonance assignment algorithms differ in the calculation time necessary to perform the assignment procedure. For example, programs like TSAR are really fast, since they simply compare the submitted chemical shift values and map them along the amino acid sequence of the protein. On the contrary, programs like MARS and FLYA require more computation time since they employ several iterative cycles to find the best correspondence between predicted and submitted chemical shift values, minimizing the propagation of possible initial errors in the assignment and/or in the experimental peak lists. Therefore, the more peak lists are complete and provide complementary information (for example combining $^{1}H^{N}$ and $^{13}C'$ detection), the less time is required by the assignment algorithms to reach convergence.

9 High-dimensional NMR Experiments (nD, with n > 3)

As discussed in the previous sections, various experimental strategies allow improvement of spectral resolution. However, in the case of severe spectral overlaps, we recommend increasing the dimensionality of the NMR experiments. Indeed, encoding more than three frequencies into a cross-peak observed in multidimensional NMR experiments reduces the chances of accidental signal overlap. In order to take maximum advantage of this approach it is important to maintain the highest resolution in all the detected dimensions. To this end, the combination of fast pulsing techniques with non-uniform sampling strategies and the respective processing approaches, described in detail in Sect. 3.5, is crucial to keep experimental time and spectral resolution in a reasonable range.

By high-dimensional NMR we refer to experiments in which three or more indirect dimensions are used to encode chemical shift information within the same experiment. In many cases, in order to cope with the low dispersion of resonance frequencies typical of IDPs, increasing the number of dimensions provides a unique tool to enhance the resolution in order to accelerate the resonance assignment procedure or to make it possible in difficult cases.

High-dimensional (4-5D) NMR experiments are presented in the next section, focusing on recent developments tailored for IDPs. A very convenient strategy to simplify the inspection and analysis of the resulting spectra is presented, as well as the most useful high-dimensional experiments tailored for IDPs.

9.1 Analysis and Inspection of High-dimensional NMR Spectra

One of the main barriers to the reception and diffusion of high dimensional NMR experiments has always been the idea that more than three dimensions could be conceptually difficult to handle. Indeed, while it is easy to visualize three-dimensional objects, the same does not hold for objects characterized by a higher number of dimensions. A possible solution consists of breaking them down into combinations of objects of lower dimensionality, as also done when inspecting three-dimensional spectra. Indeed most of the available software applications for the analysis of 3D NMR spectra display only two dimensions at a time, extracted at a specific frequency in the third dimension. In a similar way, higher dimensional NMR spectra can be inspected by visualizing only two dimensions at a time, which are associated to a peak in a 2D or 3D reference spectrum in the case of 4D and 5D experiments, respectively. This procedure is possible because, in most cases, higher dimensional experiments are extensions of 2D or 3D spectra in which the new information is encoded in the additional indirect dimensions. This idea is implemented in the sparse multidimensional Fourier transform (SMFT) algorithm (Kazimierczuk et al. 2010b, 2013).

SMFT proposes an innovative strategy to intuitively examine four or five-dimensional spectra in a very simple and straightforward way. Acquired data are processed with the MFT algorithm (mentioned in Sect. 3.5.3), but only at some predefined frequencies (hence the prefix "sparse"): in this way, the inspection of the full spectrum is reduced to that of a limited number of spectral regions that are extremely small with respect to the *n*-dimensional space of the full spectrum. For convenience, these cross-sections are processed as two-dimensional spectra, which are very familiar to almost all NMR users. Therefore, four and five-dimensional spectra appear as a series of 2D spectra to which two or three further frequencies, respectively, are associated. Of course, a prior knowledge of these frequencies is needed: together with the 4D or 5D experiment, respectively a 2D or 3D spectrum, called the "basis spectrum", which shares the same correlations with the higher

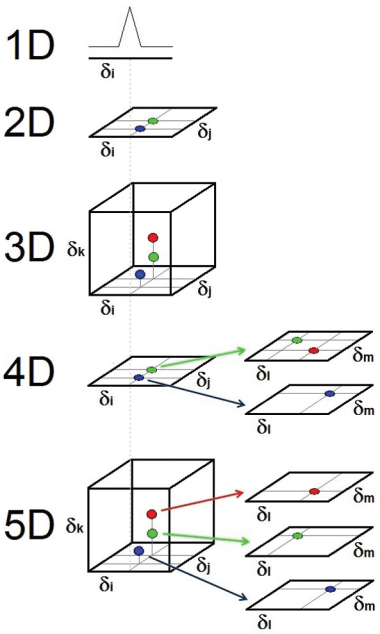


Fig. 3.32 A possible way to visualize high-multidimensional NMR spectra. A schematic illustration of the progressive steps towards experiments of higher dimensionality is reported from the 1D to the 5D case. Spectra up to three dimensions are visualized in a standard way, whereas 4D and 5D spectra are presented by decomposing them into lower dimensionality ones. The algorithm exploits the prior knowledge of the correlations provided by a lower dimensionality spectrum in order to extract only the additional information from a related multidimensional spectrum, thus reducing the number of dimensions that have to be inspected. So, for example, a 4D spectrum can be thought as a 2D basis spectrum, which shares two dimensions with the 4D spectrum (δ_i and δ_j), in which each peak is associated to another 2D spectrum containing the two further dimensions (δ_l and δ_m). In this way, instead of the full spectrum, only a series of cross-sections are effectively computed, in equal number to the peaks detected in the basis spectrum. Similarly, a 5D spectrum can be analysed as a series of two-dimensional spectra (δ_l and δ_m), each one correlated to a given peak of the related 3D basis spectrum (δ_i, δ_j and δ_k)

dimensional one, should be acquired to retrieve them. In this way, the dimensions in which the frequencies of the peaks are already known are not processed, whereas the additional two dimensions containing the new information are fully computed, as the positions of the signals are unknown there. This approach is schematically illustrated in Fig. 3.32 for 4D and 5D experiments.

SMFT facilitates the analysis of high-dimensionality spectra, since this method provides a way to retrieve all the spectral information content without the need to explore the vast *n*-dimensional space of the full spectrum. The visualization of the spectrum through inspection of 2D cross-sections also provides further advantages: for example, it enables the use of automatic peak picking tools, which perform in a very reliable and efficient way, as the signals are well-resolved thanks to the high dimensionality of the experiment. In addition, it allows great amounts of disk space to be saved, since just a small subset of the full spectrum is actually processed; in this way, high digital resolution can be set in all the indirect dimensions.

9.2 Examples of High-dimensional Experiments Tailored for IDPs

The benefits provided by chemical shift labelling of the heteronuclei including ^{13}C direct detection, ^{1}H longitudinal relaxation enhancement and TROSY, all discussed in Sects. 3.5 and 3.6, can be combined with the resolving power of high-dimensional NMR experiments specifically tailored for IDPs. In recent years a large number of high-dimensional experiments has been developed and profitably used to achieve the sequence-specific assignment of several IDPs of medium and large molecular mass. Their number is expected to increase with future methodological advancements and hardware improvements.

Several of the three-dimensional experiments based on coherence transfer via *J*-couplings, discussed in Sects. 3.7 and 3.8, can easily be extended to 4D or 5D versions by explicitly frequency labelling the heteronuclear spins only exploited in 3D experiments for coherence transfer (the nuclear spins generally included in parentheses in the pulse sequence acronyms). Taking the 3D (H)N(COCA)NH as an example, the coherence is transferred through $^{13}C'$ and $^{13}C^{\alpha}$ to N. With the introduction of one or two chemical shift evolution periods, $^{13}C'$ or/and $^{13}C^{\alpha}$, nuclei can also be frequency labelled, extending the experiment to 4D or 5D, respectively. Another interesting example is provided by the extension to 4D of the 3D HN(CA)CO and HN(CO)CA, commonly used for sequence-specific assignment (4D HNCOCA and 4D HNCACO) (Brutscher et al. 1995a; Yang and Kay 1999; Xia et al. 2002). Based on the 2D HN experiment (H^{N}_{i}-N_{i}), they provide sequential correlations via the $^{13}C'$ and $^{13}C^{\alpha}$ nuclei for backbone assignment (C'_{i-1}-C^{α}_{i-1} and C^{α}_{i}-C'_{i}/C^{α}_{i-1}-C'_{i-1}, respectively) as shown in Fig. 3.33a. The former can be used to resolve frequency degeneracy in the 3D HNCO spectrum, whereas the latter relies on $^{13}C^{\alpha}$ and $^{13}C'$ chemical shifts to establish sequential correlations. An additional 4D iHNCOCA experiment can also be designed (Konrat et al. 1999). As a general rule, the price to pay for the additional information retrieved in high-dimensional experiments

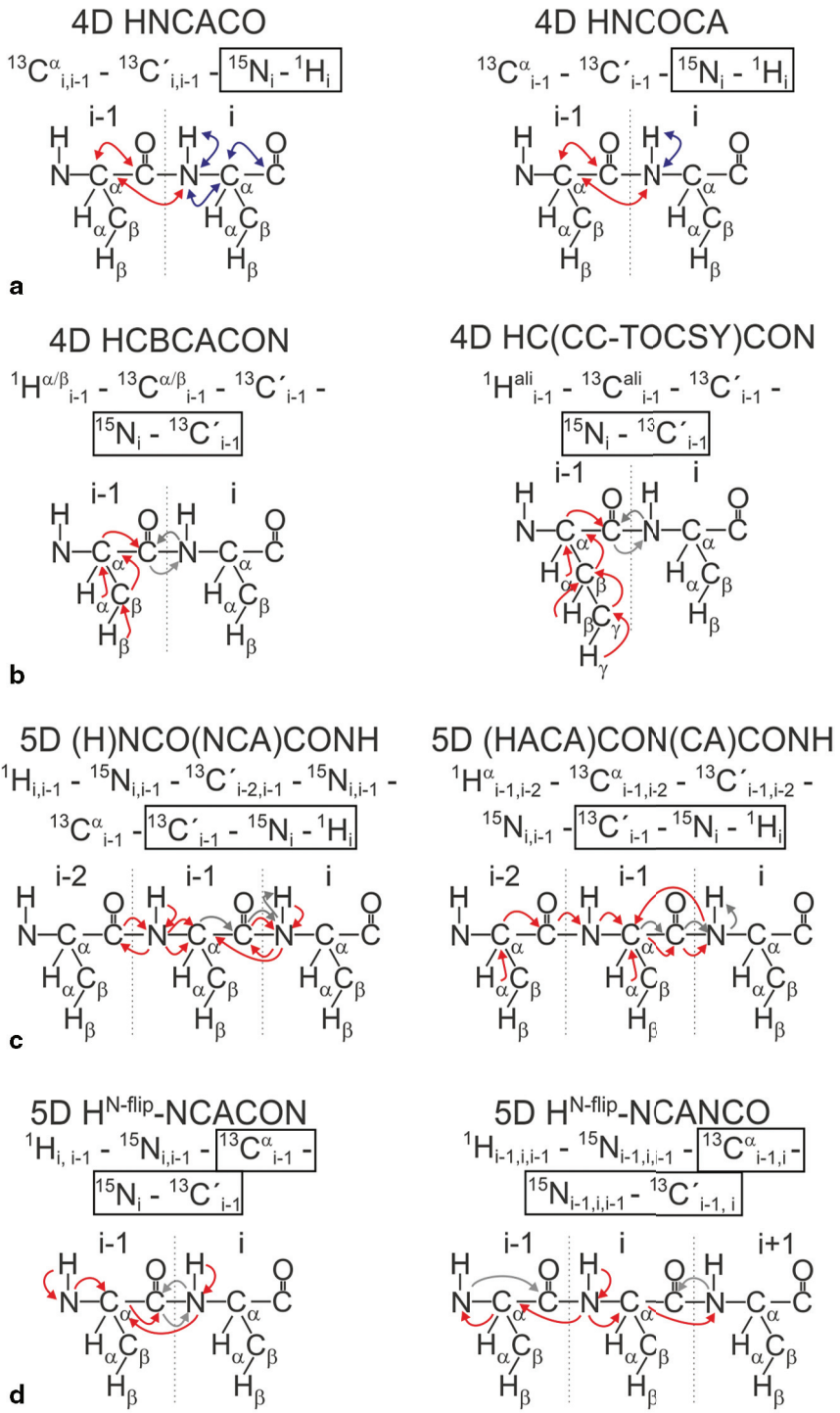


Fig. 3.33 Different examples of the polarization transfer pathways of 4D and 5D experiments

compared to their lower dimensional analogues consists of a loss in experimental sensitivity of at least a factor of $\sqrt{2}$ for each additional dimension introduced, due to the requirement of phase-sensitive quadrature detection and because additional spin relaxation occurs during the frequency editing periods.

An overview of the most useful experiments, exemplifying the many possibilities available to retrieve sufficient information for unambiguous sequence-specific assignment, is presented hereafter. A summary of the experiments discussed, together with the correlations provided, is reported in Tables A.1 and A.2 in the Appendix. The high-dimensional experiments can be grouped into different categories according to their lower dimensional basis 2D (HN or CON) and 3D (HNCO or CACON) spectra as well as to the type of information retrieved in the additional indirect dimensions (amino acid-type information through ^{13}C side chain chemical shifts, sequence-specific information through C^{α}, C^{β}, C′ and N chemical shifts).

Starting from the most simple and intuitive ones, the 5D HCBCACONH (Grzesiek and Bax 1993; Staykova et al. 2008; Kazimierczuk et al. 2010b), 5D HCCCONH (Montelione et al. 1992; Clowes et al. 1993; Mobli et al. 2010), 4D HCBCACON (Bermel et al. 2012b) and 5D HC(CC-TOCSY)CON (Montelione et al. 1992; Logan et al. 1992; Logan et al. 1993; Grzesiek et al. 1993a; Gardner et al. 1996; Hiller et al. 2008) (Fig. 3.33b) provide information on the chemical shifts of the ^{1}H and ^{13}C aliphatic spins of each amino acid (i) by exploiting the HNCO (C'_{i}, N_{i+1}, H^{N}_{i+1},) or the CON (C'_{i}, N_{i+1}) as basis spectra, respectively. Therefore, the related 2D cross-sections that need to be inspected resemble 2D ^{1}H-^{13}C planes which however contain only the correlations observed at the respective frequencies in the basis spectra, in this way providing very clean and straightforward information to identify the residue type and assign side chain ^{1}H and ^{13}C chemical shifts. If necessary, the 5D experiments that exploit the HNCO as the basis 3D spectrum can easily be reduced to their 4D analogues (which exploit the 2D HN as a basis spectrum) by simply omitting the evolution of $^{13}C'$ chemical shifts.

Along the same lines, the 2D HN and CON spectra can be used as basis spectra for high-dimensional experiments that provide information to achieve sequence-specific assignment in the additional dimensions. A variety of different experiments can be included in this class. For example, the 4D HNCACB (Zawadzka-Kazimierczuk et al. 2010; Gossert et al. 2011) and 4D HCBCANCO (Bermel et al. 2012b; Nováček et al. 2012) (bidirectional experiments) yield in cross-section the information on the ^{1}H and ^{13}C chemical shifts of amino acids (i) and (i + 1) for each cross-peak observed in the 2D HN (N_{i+1}, H^{N}_{i+1}) and 2D CON spectra (C'_{i}, N_{i+1}), respectively.

Another class of experiments provides sequential connectivities through ^{15}N chemical shifts. These are particularly useful because, depending on the experimental variant used, they still exploit the same basis spectra (3D HNCO or 2D HN and 2D CON) and provide information on the ^{15}N chemical shifts of neighbouring residues in the additional dimensions.

This category includes the 4D HN(COCA)NH (Shirakawa et al. 1995; Bracken et al. 1997), 4D HN(CA)NH (Zawadzka-Kazimierczuk et al. 2010), 5D HN(CA)CONH (Kazimierczuk et al. 2010b) and 5D HN(COCAN)CONH (Piai et al. 2014) ($^{1}H^{N}$ detected) as well as a variety of different experimental variants based on ^{13}C direct detection described below. The $^{1}H^{N}$ detected ones reported above correlate the H^{N}_{i}-N_{i} peak of the 2D ^{1}H-^{15}N HSQC spectrum (the H^{N}_{i}-N_{i}-C'_{i-1} peaks of the 3D HNCO spectrum, in the 5D case) with the $^{1}H^{N}$ and ^{15}N nuclei of neighboring residues. Instead, when amide protons are merely exploited as a starting source of magnetization, ^{15}N and $^{13}C'$ resonances can be used to establish sequential correlations, such as in the 5D (H)NCO(NCA)CONH (Zawadzka-Kazimierczuk et al. 2012b) (N_{i-1}-C'_{i-2} and N_{i}-C'_{i-1}) and 5D (H)NCO(CAN)CONH (Piai et al. 2014) (N_{i}-C'_{i-1} and N_{i+1}-C'_{i}) experiments (Fig. 3.33c).

As regards the ^{13}C detected analogues, they can be acquired in the 4D mode, using the 2D CON as basis experiment, or extended to the 5D mode by exploiting the C^{α} chemical shift as an additional dimension of the basis experiment. They include 5D NCOCANCO (Nováček et al. 2011), 5D ($H^{N\text{-flip}}N$)CONCACON (Bermel et al. 2013), 5D (HCA)CONCACON (Bermel et al. 2013) and 5D (H)NCO(CAN)CONH (Bermel et al. 2013) experiments. Sequential correlations are established by exploiting ^{15}N and $^{13}C'$ frequencies, both retrieved in the 2D cross-section of the experiments ("CON-CON strategy").

For ^{13}C detection, also a series of experiments that rely on CACO as the basis 2D spectrum have been proposed. The C^{α}_{i}-C'_{i} frequencies of the basis spectrum are correlated to C^{α}_{i+1}-N_{i+1} and C^{α}_{i-1}-N_{i} or to C^{α}_{i-1}-C'_{i-1} nuclei, respectively through the 4D (H)CANCACO (Bermel et al. 2012b) and the 5D CACONCACO (Nováček et al. 2011) experiments. Finally, the 5D $H^{N\text{-flip}}$NCACON (Bermel et al. 2012b) and 5D $H^{N\text{flip}}$NCANCO (Bermel et al. 2012b) experiments relate the ^{15}N of residue i with those of residue $i+1$ and $i+2$, respectively (Fig. 3.33d).

High-dimensional NMR experiments featuring $^{1}H^{\alpha}$-start/$^{1}H^{\alpha}$-detection have also recently been designed. These experiments may be useful at temperature and pH conditions in which $^{1}H^{N}$ are not detectable and/or in the case of short-lived or low concentration samples that partly limit the use of high-dimensional ^{13}C detected experiments. Extensions to higher dimensionality partly counterbalance the low chemical shift dispersion of $^{1}H^{\alpha}$ in the direct dimension (Piai et al. 2014).

In conclusion, due to their high resolving power and increased information content, high-dimensional experiments provide a valuable tool to extend the size and complexity of IDPs that can be characterized by NMR at atomic resolution, provided the sample's relaxation properties allow for multiple (and often long) coherence transfer times. The combination of the different approaches described in this chapter offers a rich toolbox that can be exploited for the investigation of complex IDPs (Fig. 3.34).

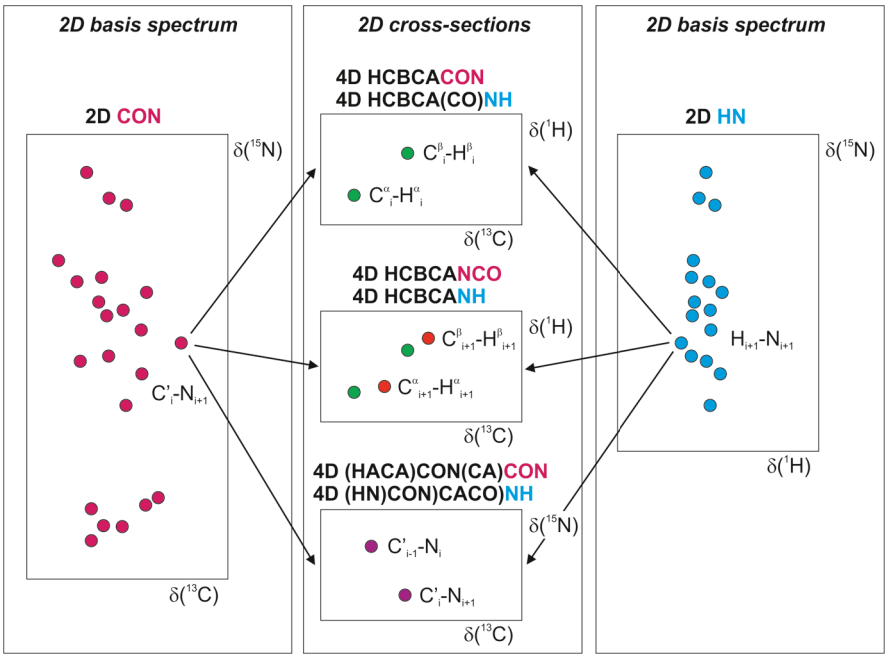


Fig. 3.34 A schematic illustration of a series of ^{13}C and $^{1}H^{N}$ detected 4D spectra. For each pair of experiments the information provided is the same, except for the direct dimension. Since the 4D experiments share C'_i-N_{i+1} or H_{i+1}-N_{i+1} frequencies, the 2D CON (reported on the *left*) or 2D HN (reported on the *right*) can be respectively used as basis spectra to collect these common frequencies. The new information content of the various 4D experiments can then be easily retrieved inspecting a series of 2D cross-sections, reported in the middle part of the illustration, where the correlation provided by each experiment is shown

10 Conclusions and Perspectives

Thanks to recent improvements in NMR technology, the array of NMR experiments that have been developed, and their optimization for the specific properties of IDPs, a number of tools are available for the atomic resolution structural and dynamic characterization of IDPs.

As of today, IDPs as large as 400 amino acids have been investigated, with two known examples being MAP (Nováček et al. 2013) and Tau (Mukrasch et al. 2009); a number of other examples of high resolution investigations of IDPs in the range between 100 and 300 amino acids have appeared in the literature. Many more high-resolution NMR experimental investigations of IDPs are expected to be accomplished in the near future.

Indeed the sequence-specific assignment and initial structural and dynamic characterization through the analysis of chemical shifts and ^{15}N relaxation rates can readily be achieved through the experiments described in this chapter. Already at

this stage the sequence-specific assignment and ^{15}N relaxation measurements are sufficient to describe the overall properties of the IDP in its native state and they can be used as the basis for further investigation of its function by monitoring interactions and post-translational modifications as well as by taking snapshots of the IDP inside whole cells. Many more experiments can be performed to gain further information, as discussed in the next chapters.

Among the many experimental strategies discussed in detail in this chapter, the optimal one for a specific IDP to be investigated can be readily identified by acquiring a small set of initial 1D and 2D spectra that provide information on the sensitivity and resolution that can be achieved as a result of the relaxation properties and chemical shift dispersion of the investigated IDP. In particular, the 2D HN and CON spectra are the most suitable to evaluate the appropriate experimental strategy for sequence-specific assignment. A large panoply of optimized pulse sequences and processing tools are currently available, as are user-friendly computational tools for the analysis of the resulting spectra, including higher dimensional ones.

NMR spectroscopy is continuously improving in terms of hardware performance and experimental approaches. We expect that future progress will aim at further extending the size limit of IDPs that can be investigated at high resolution through NMR, as well as the determination and interpretation of a larger number of observables reporting on the structural and dynamic properties of IDPs, and finally the possibility of characterising the behaviour of IDPs in different aggregation states, ranging from solution to solid state to in-cell, through a combination of different NMR techniques.

Finally, speculating on more long-term perspectives, the development of improved NMR methods to study IDPs is expected to provide a large amount of experimental data on them, contributing to our understanding of the molecular basis responsible for their function and filling a gap of about 50 years with respect to our knowledge on the structural and dynamic behaviour of folded proteins. This is expected to reveal a much larger number of ways in which proteins communicate in the cell. Other expected outcomes of NMR experimental data on IDPs include the improvement of prediction tools, which still suffer from the bias that they are derived from the missing information in the electron density maps in X-ray crystallography data.

Appendix

Table A.1 High-multidimensional ^{1}H detected experiments for backbone and side-chain resonance assignment

^{1}H detected experiments			
Dimensionality	Experiment	Correlations observed	References
Experiments for spin-system identification			
3	HN(CO)CA	H^{N}_{i}-N_{i}-C^{α}_{i-1}	(Bax and Ikura 1991; Grzesiek and Bax 1992b; Solyom et al. 2013)

Table A.1 (continued)

^{1}H detected experiments			
Dimensionality	Experiment	Correlations observed	References
3	HN(CO)CACB	H^N_i-N_i-$C^{\alpha/\beta}_{i-1}$	(Grzesiek and Bax 1992a; Yamazaki et al. 1994; Solyom et al. 2013)
3	(H)C(CC-TOCSY)(CO)NH	C^{ali}_{i-1}-N_i-H^N_i	(Logan et al. 1992; Montelione et al. 1992; Logan et al. 1993; Gardner et al. 1996)
3	iHNCA	H^N_i-N_i-C^{α}_i	(Brutscher 2002; Nietlispach et al. 2002; Solyom et al. 2013)
3	iHNCACB	H^N_i-N_i-$C^{\alpha/\beta}_i$	(Brutscher 2002; Nietlispach et al. 2002; Solyom et al. 2013)
3	iHCAN	H^{α}_i-C^{α}_i-N_i	(Mäntylahti et al. 2010)
4	(H)CBCACONH	$C^{\alpha/\beta}_{i-1}$-C'_{i-1}-N_i-H^N_i	(Grzesiek and Bax 1992a; Grzesiek and Bax 1993)
5	HCBCACONH	$H^{\alpha/\beta}_{i-1}$-$C^{\alpha/\beta}_{i-1}$-C'_{i-1}-N_i-H^N_i	(Grzesiek and Bax 1993; Staykova et al. 2008; Kazimierczuk et al. 2010b)
5	HNCOCACB	H^N_i-N_i- C'_{i-1}- C^{α}_{i-1}-C^{β}_{i-1}	(Hiller et al. 2007; Zawadzka-Kazimierczuk et al. 2012b)
5	HC(CC-TOCSY)CONH	H^{ali}_{i-1}-C^{ali}_{i-1}-C'_{i-1}-N_i-H^N_i	(Logan et al. 1992; Montelione et al. 1992; Logan et al. 1993; Grzesiek et al. 1993a; Grzesiek et al. 1993b; Gardner et al. 1996; Hiller et al. 2008)
Experiments for sequential assignment			
3	HNCO	H^N_i-N_i-C'_{i-1}	(Kay et al. 1990; Grzesiek and Bax 1992b; Schleucher et al. 1993; Solyom et al. 2013)
3	HNCA	H^N_i-N_i-C^{α}_{i-1} & H^N_i-N_i-C^{α}_i	(Kay et al. 1990; Grzesiek and Bax 1992b; Lescop et al. 2007)
3	HNCACB	H^N_i-N_i-$C^{\alpha/\beta}_{i-1}$ & H^N_i-N_i-$C^{\alpha/\beta}_i$	(Wittekind and Mueller 1993; Muhandiram and Kay 1994; Lescop et al. 2007)
3	HN(CA)CO	H^N_i-N_i-C'_{i-1} & H^N_i-N_i-C'_i	(Clubb et al. 1992; Kay et al. 1994; Briand et al. 2001; Lescop et al. 2007)
3	(H)N(COCA)NH	N_{i+1}-N_i-H^N_i	(Grzesiek et al. 1993b; Bracken et al. 1997; Panchal et al. 2001)
3	(H)N(CA)NH	N_{i-1}-N_i-H^N_i & N_{i+1}-N_i-H^N_i	(Grzesiek et al. 1993b; Weisemann et al. 1993)
3	iH(CA)NCO	H^{α}_i-N_i-C'_i	(Mäntylahti et al. 2010)
3	(HCA)NCO(CA)H	H^{α}_i-C'_i-N_{i+1}	(Mäntylahti et al. 2011)
3	H(CA)CON	H^{α}_i-C'_i-N_{i+1}	(Mäntylahti et al. 2010)
3	(HCA)CON(CA)H	H^{α}_i-N_i-C'_{i-1} & H^{α}_i-N_{i+1}-C'_i	(Mäntylahti et al. 2011)
4	HCACON	H^{α}_i-C^{α}_i-C'_i-N_{i+1}	(Kay et al. 1991)
4	HNCOCA	H^N_i-N_i-C'_{i-1}-C^{α}_{i-1}	(Brutscher et al. 1995a; Yang and Kay 1999; Xia et al. 2002)
4	HNCACO	H^N_i-N_i-C^{α}_i-C'_i & H^N_i-N_i-C^{α}_{i-1}-C'_{i-1}	(Yang and Kay 1999; Xia et al. 2002)
4	HNCO, CA	H^N_i-N_i-C'_{i-1}-C^{α}_{i-1} & H^N_i-N_i-C'_{i-1}-C^{α}_i	(Konrat et al. 1999)

Table A.1 (continued)

^1H detected experiments			
Dimensionality	Experiment	Correlations observed	References
4	HNCACB	H^N_i-N_i-C^α_{i-1}-C^β_{i-1} & H^N_i-N_i-C^α_i-C^β_i	(Gossert et al. 2011)
4	HCBCANH	$H^{\alpha/\beta}_i$-$C^{\alpha/\beta}_i$-N_i-H^N_i & $H^{\alpha/\beta}_{i-1}$-$C^{\alpha/\beta}_{i-1}$-N_i-H^N_i	(Zawadzka-Kazimierczuk et al. 2010)
4	HACANH	H^α_i-C^α_i-N_i-H^N_i & H^α_{i-1}-C^α_{i-1}-N_i-H^N_i	(Boucher et al. 1992; Szyperski et al. 1993a)
4	HNCAHA	H^N_i-N_i-C^α_i- H^α_i & H^N_{i+1}-N_{i+1}-C^α_i- H^α_i	(Xia et al. 2002)
4	HACA(CO)NH	H^α_{i-1}-C^α_{i-1}-N_i-H^N_i	(Boucher et al. 1992)
4	HN(CO)CAHA	H^N_{i+1}-N_{i+1}-C^α_i- H^α_i	(Xia et al. 2002)
4	(H)CACO(CA)NH	C'_i-C^α_i-N_i-H^N_i & C'_{i-1}-C^α_{i-1}-N_i-H^N_i	(Löhr and Rüterjans 1995)
4	HN(COCA)NH	H^N_{i+1}-N_{i+1}-N_i-H^N_i & H^N_i-N_i-N_i-H^N_i	(Shirakawa et al. 1995; Bracken et al. 1997;)
4	HN(CA)NH	H^N_{i+1}-N_{i+1}-N_i-H^N_i & H^N_{i-1}-N_{i-1}-N_i-H^N_i	(Zawadzka-Kazimierczuk et al. 2010)
4	HN(CO)CA(CON)CA	H^N_i-N_i-C^α_{i-1}-C^α_{i-1} & H^N_i-N_i-C^α_{i-1}-C^α_i	(Bagai et al. 2011)
4	HNCO(N)CA	H^N_i-N_i-C'_{i-1}-C^α_{i-1} & H^N_i-N_i-C'_{i-1}-C^α_i	(Bagai et al. 2011)
5	HACACONH	H^α_{i-1}-C^α_{i-1}-C'_{i-1}-N_i-H^N_i	(Kim and Szyperski 2003; Hiller et al. 2005; Malmodin and Billeter 2005)
5	HACACONH	H^α_i-C^α_i-C'_i-N_i-H^N_i & H^α_{i-1}-C^α_{i-1}-C'_{i-1}-N_i-H^N_i	(Kim and Szyperski 2004)
5	HACA(N)CONH	H^α_i-C^α_i-C'_{i-1}-N_i-H^N_i & H^α_{i-1}-C^α_{i-1}-C'_{i-1}-N_i-H^N_i	(Zawadzka-Kazimierczuk et al. 2012b)
5	(HACA)CON(CA)CONH	C'_{i-2}-N_{i-1}-C'_{i-1}-N_i-H^N_i & C'_{i-1}-N_i-C'_{i-1}-N_i-H^N_i	(Zawadzka-Kazimierczuk et al. 2012b)
5	HCBCA(CAN)CONH	$H^{\alpha/\beta}_{i-1}$-$C^{\alpha/\beta}_{i-1}$-C'_{i-1}-N_i-H^N_i & $H^{\alpha/\beta}_i$-$C^{\alpha/\beta}_i$-C'_{i-1}-N_i-H^N_i	(Staykova et al. 2008)
5	HN(CA)CONH	H^N_{i-1}-N_{i-1}-C'_{i-1}-N_i-H^N_i & H^N_i-N_i-C'_{i-1}-N_i-H^N_i	(Kazimierczuk et al. 2010b)
5	(H)NCO(NCA)CONH	N_{i-1}-C'_{i-2}-C'_{i-1}-N_i-H^N_i & N_i-C'_{i-1}-C'_{i-1}-N_i-H^N_i	(Zawadzka-Kazimierczuk et al. 2012b)
5	(H)NCO(CAN)CONH	N_{i+1}-C'_i-C'_{i-1}-N_i-H^N_i	(Piai et al. 2014)
5	HN(COCAN)CONH	H^N_{i+1}-N_{i+1}-C'_{i-1}-N_i-H^N_i	(Piai et al. 2014)
5	(HACA)CON(CACO)NCO(CA)HA	C'_{i-1}- N_i- N_{i+1}-C'_i-H^α_i	(Piai et al. 2014)

Table A.2 High-multidimensional ^{13}C detected experiments for backbone and side-chain resonance assignment

^{13}C detected experiments			
Dimensionality	Experiment	Correlations observed	References
Experiments for spin-system identification			
3	(H)CACON	C^{α}_{i}-N_{i+1}-C'_{i}	(Bermel et al. 2006c; Bermel et al. 2009a)
3	(H)CBCACON	$C^{\alpha/\beta}_{i}$-N_{i+1}-C'_{i}	(Bermel et al. 2006c; Bermel et al. 2009a)
3	(H)C(CC-TOCSY)CON	C^{ali}_{i}-N_{i+1}-C'_{i}	(Bermel et al. 2006c; Bermel et al. 2009a)
4	HCBCACON	$H^{\alpha/\beta}_{i}$-$C^{\alpha/\beta}_{i}$-N_{i+1}-C'_{i}	(Bermel et al. 2012b; Nováček et al. 2012)
4	HC(CC-TOCSY)CON	H^{ali}_{i}-C^{ali}_{i}-N_{i+1}-C'_{i}	(Bermel et al. 2012b)
5	HC(CC-TOCSY)CACON	H^{ali}_{i}-C^{ali}_{i}-C^{α}_{i}-N_{i+1}-C'_{i}	(Nováček et al. 2013)
Experiments for sequential assignment			
3	(H)CANCO	C^{α}_{i}-N_{i+1}-C'_{i} & C^{α}_{i+1}-N_{i+1}-C'_{i}	(Bermel et al. 2006c; Bermel et al. 2009a)
3	(H)CBCANCO	$C^{\alpha/\beta}_{i}$-N_{i+1}-C'_{i} & $C^{\alpha/\beta}_{i+1}$-N_{i+1}-C'_{i}	(Bermel et al. 2006c; Bermel et al. 2009a)
3	($H^{N\text{-}flip}$)N(CA)NCO	N_{i}-N_{i+1}-C'_{i} & N_{i+2}-N_{i+1}-C'_{i}	(Bermel et al. 2009a; Bermel et al. 2012b)
3	(HCA)NCACO	N_{i}-C^{α}_{i}-C'_{i} & N_{i+1}-C^{α}_{i}-C'_{i}	(Bermel et al. 2012b)
3	COCON	C'_{i-1}-N_{i+1}-C'_{i} & C'_{i+1}-N_{i+1}-C'_{i}	(Bermel et al. 2006b)
4	HCBCANCO	$H^{\alpha/\beta}_{i}$-$C^{\alpha/\beta}_{i}$-N_{i+1}-C'_{i} & $H^{\alpha/\beta}_{i+1}$-$C^{\alpha/\beta}_{i+1}$-N_{i+1}-C'_{i}	(Bermel et al. 2012b; Nováček et al. 2012)
4	($H^{N\text{-}flip}$)NCANCO	N_{i}-C^{α}_{i}-N_{i+1}-C'_{i} & N_{i+2}-C^{α}_{i+1}-N_{i+1}-C'_{i}	(Bermel et al. 2012b)
4	($H^{N\text{-}flip}$)NCACON	N_{i}-C^{α}_{i}-N_{i+1}-C'_{i} & N_{i+1}-C^{α}_{i}-N_{i+1}-C'_{i}	(Bermel et al. 2012b)
4	(H)CANCACO	C^{α}_{i-1}-N_{i}-C^{α}_{i}-C'_{i} & C^{α}_{i+1}-N_{i+1}-C^{α}_{i}-C'_{i}	(Bermel et al. 2012b)
5	$H^{N\text{-}flip}$NCANCO	H^{N}_{i}-N_{i}-C^{α}_{i}-N_{i+1}-C'_{i} & H^{N}_{i+2}-N_{i+2}-C^{α}_{i+1}-N_{i+1}-C'_{i}	(Bermel et al. 2012b)
5	$H^{N\text{-}flip}$NCACON	H^{N}_{i}-N_{i}-C^{α}_{i}-N_{i+1}-C'_{i} & H^{N}_{i+1}-N_{i+1}-C^{α}_{i}-N_{i+1}-C'_{i}	(Bermel et al. 2012b)
5	(H)NCOCANCO	N_{i+2}-C'_{i+1}-C^{α}_{i+1}-N_{i+1}-C'_{i}	(Nováček et al. 2011)
5	(H)CACONCACO	C^{α}_{i-1}-C'_{i-1}-N_{i}-C^{α}_{i}-C'_{i}	(Nováček et al. 2011)
5	($H^{N\text{-}flip}$N)CONCACON	C'_{i-1}-N_{i}-C^{α}_{i}-N_{i+1}-C'_{i} & C'_{i}-N_{i+1}-C^{α}_{i}-N_{i+1}-C'_{i}	(Bermel et al. 2013)
5	(HCA)CONCACON	C'_{i-1}-N_{i}-C^{α}_{i}-N_{i+1}-C'_{i} & C'_{i}-N_{i+1}-C^{α}_{i}-N_{i+1}-C'_{i}	(Bermel et al. 2013)
5	(H)CACON(CA)CON	C^{α}_{i}-C'_{i}-N_{i+1}-C'_{i+1}-N_{i+2} & C^{α}_{i}-C'_{i}-N_{i+1}-C'_{i}-N_{i+1}	(Bermel et al. 2013)

References

Arnesano F, Banci L, Bertini I et al (2001) Characterization of the binding interface between the copper chaperone Atx1 and the first cytosolic domain of Ccc2 ATPase. J Biol Chem 276:41365–41376

Arnesano F, Balatri E, Banci L et al (2005) Folding studies of Cox17 reveal an important interplay of cysteine oxidation and copper binding. Structure 13:713–722

Bagai I, Raqsdale SW, Zuiderweg ER (2011) Pseudo-4D triple resonance experiments to resolve HN overlap in the backbone assignment of unfolded proteins. J Biomol NMR 49:69–74

Bai Y, Milne JS, Mayne L et al (1993) Primary structure effects on peptide group hydrogen exchange. Proteins 17:75–86

Banci L, Bertini I, Huber JG et al (1998) Partial orientation of oxidized and reduced cytochrome b_5 at high magnetic fields: magnetic susceptibility anisotropy contributions and consequences for protein solution structure determination. J Am Chem Soc 120:12903–12909

Barbato G, Ikura M, Kay LE et al (1992) Backbone dynamics of calmodulin studied by ^{15}N relaxation using inverse detected two-dimensional NMR spectroscopy; the central helix is flexible. Biochemistry 31:5269–5278

Bax A, Grishaev A (2005) Weak alignment NMR: a hawk-eyed view of biomolecular structure. Curr Opin Struct Biol 15:563–570

Bax A, Ikura M (1991) An efficient 3D NMR technique for correlating the proton and ^{15}N backbone amide resonances with the α-carbon of the preceding residue. J Biomol NMR 1:99–104

Bermel W, Bertini I, Duma L et al (2005) Complete assignment of heteronuclear protein resonances by protonless NMR spectroscopy. Angew Chem Int Ed 44:3089–3092

Bermel W, Bertini I, Felli IC et al (2006a) Novel ^{13}C direct detection experiments, including extension to the third dimension, to perform the complete assignment of proteins. J Magn Reson 178:56–64

Bermel W, Bertini I, Felli IC et al (2006b) Protonless NMR experiments for sequence-specific assignment of backbone nuclei in unfolded proteins. J Am Chem Soc 128:3918–3919

Bermel W, Bertini I, Felli IC et al (2006c) ^{13}C-detected protonless NMR spectroscopy of proteins in solution. Progr NMR Spectrosc 48:25–45

Bermel W, Felli IC, Kümmerle R et al (2008) ^{13}C direct-detection biomolecular NMR. Concepts Magn Reson 32A:183–200

Bermel W, Bertini I, Csizmok V et al (2009a) H-start for exclusively heteronuclear NMR spectroscopy: the case of intrinsically disordered proteins. J Magn Reson 198:275–281

Bermel W, Bertini I, Felli IC et al (2009b) Speeding up ^{13}C direct detection biomolecular NMR experiments. J Am Chem Soc 131:15339–15345

Bermel W, Bertini I, Chill JH et al (2012a) Exclusively heteronuclear ^{13}C-detected amino-acid-selective NMR experiments for the study of instrinsically disordered proteins (IDPs). Chem Bio Chem 13:2425–2432

Bermel W, Bertini I, Gonnelli L et al (2012b) Speeding up sequence specific assignment of IDPs. J Biomol NMR 53:293–301

Bermel W, Felli IC, Gonnelli L et al (2013) High-dimensionality ^{13}C direct-detected NMR experiments for the automatic assignment of intrinsically disordered proteins. J Biomol NMR 57:353–361

Bertini I, Felli IC, Kümmerle R et al (2004) ^{13}C-^{13}C NOESY: a constructive use of ^{13}C-^{13}C spin-diffusion. J Biomol NMR 30:245–251

Bertini I, Felli IC, Gonnelli L et al (2011a) ^{13}C direct-detection biomolecular NMR spectroscopy in living cells. Angew Chem Int Ed 50:2339–2341

Bertini I, Felli IC, Gonnelli L et al (2011b) High-resolution characterization of intrinsic disorder in proteins: expanding the suite of ^{13}C detected NMR experiments to determine key observables. ChemBioChem 12:2347–2352

Bertini I, Luchinat C, Parigi G et al (2011c) Solid-state NMR of proteins sedimented by ultracentrifugation. Proc Natl Acad Sci U S A 108:10396–10399
Billeter M, Neri D, Otting G et al (1992) Precise vicinal coupling constants $^{3}J_{H\alpha N}$ in proteins from nonlinear fits of J-modulated [^{15}N,^{1}H]-COSY experiments. J Biomol NMR 2:257–74
Bloch F (1946) Nuclear induction. Phys Rev 70:460–474
Bloch F (1956) Dynamical theory of nuclear induction. II. Phys Rev 102:104–135
Boucher W, Laue ED, Campbell-Burk SL et al (1992) Improved 4D NMR experiments for the assignment of backbone nuclei in ^{13}C/^{15}N labelled proteins. J Biomol NMR 2:631–637
Bracken C, Palmer AG III, Cavanagh J (1997) (H)N(COCA)NH and HN(COCA)NH experiments for 1H-15N backbone assignments in ^{13}C/^{15}N-labeled proteins. J Biomol NMR 9:94–100
Briand L, Lescop E, Bézirard V et al (2001) Isotopic double-labeling of two honeybee odorant-binding proteins secreted by the methylotrophic yeast Pichia pastoris. Protein Expr Purif 23:167–174
Brutscher B (2002) Intraresidue HNCA and COHNCA experiments for protein backbone resonance assignment. J Magn Reson 156:155–159
Brutscher B (2004a) Combined frequency- and time-domain NMR spectroscopy. Application to fast protein resonance assignment. J Biomol NMR 29:57–64
Brutscher B (2004b) DEPT spectral editing in HCCONH-type experiments. Application to fast protein backbone and side chain assignment. J Magn Reson 167:178–184
Brutscher B, Cordier F, Simorre JP et al (1995a) High-resolution 3D HNCOCA experiment applied to a 28 kDa paramagnetic protein. J Biomol NMR 5:202–206
Brutscher B, Morelle N, Cordier F et al (1995b) Determination of an initial set of NOE-derived distance constraints for the structure determination of ^{15}N/ ^{13}C labeled proteins. J Magn Reson B 109:238–242
Case DA (2000) Interpretation of chemical shifts and coupling constants in macromolecules. Curr Opin Struct Biol 10:197–203
Cavanagh J, Fairbrother WJ, Palmer AG III et al (2007) Protein NMR Spectroscopy. Principles and practice. Academic, San Diego
Chimon S, Shaibat MA, Jones CR et al (2007) Evidence of fibril-like β-sheet structures in a neurotoxic amyloid intermediate of Alzheimer's β-amyloid. Nat Struct Mol Biol 14:1157–1164
Clowes RT, Boucher W, Hardman CH et al (1993) A 4D HCC(CO)NNH experiment for the correlation of aliphatic side-chain and backbone resonances in ^{13}C/^{15}N-labelled proteins. J Biomol NMR 3:349–354
Clubb RT, Thanabal V, Wagner G (1992) A constant-time three dimensional triple-resonance pulse scheme to correlate intraresidue ^{1}HN, ^{15}N, and ^{13}C' chemical shifts in ^{15}N-^{13}C- labeled proteins. J Magn Reson 97:213–217
Coggins BE, Zhou P (2007) Sampling of the NMR time domain along concentric rings. J Magn Reson 184:207–221
Cowburn D, Shekhtman A, Xu R et al (2004) Segmental isotopic labeling for structural biological applications of NMR. Methods Mol Biol 278:47–56
Csizmok V, Felli IC, Tompa P et al (2008) Structural and dynamic characterization of intrinsically disordered human securin by NMR. J Am Chem Soc 130:16873–16879
Dötsch V, Oswald RE, Wagner G (1996a) Amino-acid type-selective triple-resonance experiments. J Magn Reson B 110:107–111
Dötsch V, Oswald RE, Wagner G (1996b) Selective identification of threonine, valine and isoleucine sequential connectivities with a TVI-CBCACONH experiment. J Magn Reson B 110:304–308
Duma L, Hediger S, Brutscher B et al (2003a) Resolution enhancement in multidimensional solid-state NMR spectroscopy of proteins using spin-state selection. J Am Chem Soc 125: 11816–11817
Duma L, Hediger S, Lesage A et al (2003b) Spin-state selection in solid-state NMR. J Magn Reson 164:187–195
Dyson HJ, Wright PE (2001) Nuclear magnetic resonance methods for the elucidation of structure and dynamics in disordered states. Methods Enzymol 339:258–271

Eliezer D (2009) Biophysical characterization of intrinsically disordered proteins. Curr Opin Struct Biol 19:23–30
Emsley L, Bodenhausen G (1990) Phase-shifts induced by transient Bloch-Siegert effect in NMR. Chem Phys Lett 168:297–303
Ernst RR, Bodenhausen G, Wokaun A (1987) Principles of nuclear magnetic resonance in one and two dimensions. Clarendon, Oxford
Farrow NA, Zhang O, Szabo A et al (1995) Spectral density function mapping using ^{15}N relaxation data exclusively. J Biomol NMR 6:153–162
Favier A, Brutscher B (2011) Recovering lost magnetization: polarization enhancement in biomolecular NMR. J Biomol NMR 49:9–15
Felli IC, Pierattelli R (2014a) Novel methods based on ^{13}C detection to study intrinsically disordered proteins. J Magn Reson 241:115–125
Felli IC, Pierattelli R (2014b) Spin-state-selctive methods in solution- and solid-state biomolecular ^{13}C NMR. Prog NMR Spectrosc 84–85:1–13
Felli IC, Pierattelli R, Glaser SJ et al (2009) Relaxation-optimised Hartmann-Hahn transfer for carbonyl-carbonyl correlation spectroscopy using a specifically tailored MOCCA-XY16 mixing sequence for protonless ^{13}C direct detection experiments. J Biomol NMR 43:187–196
Felli IC, Pierattelli R, Tompa P (2012) Intrinsically disordered proteins. In: Bertini I, McGreevy K, Parigi G (eds) NMR of biomolecules: towards mechanistic systems biology. Wiley
Felli IC, Piai A, Pierattelli R (2013) Recent advances in solution NMR studies: ^{13}C direct detection for biomolecular NMR applications. Ann Rep NMR Spectroscop 80:359–418
Felli IC, Gonnelli L, Pierattelli R (2014) In-cell ^{13}C NMR spectroscopy for the study of intrinsically disordered proteins. Nat Protoc 9:2005–2016
Feuerstein S, Plevin MJ, Willbold D et al (2012) iHADAMAC: a complementary tool for sequential resonance assignment of globular and highly disordered proteins. J Magn Reson 214:329–334
Gal M, Edmonds KA, Milbradt AG et al (2011) Speeding up direct ^{15}N detection: hCaN 2D NMR experiment. J Biomol NMR 51:497–504
Gardner KH, Konrat R, Rosen MK et al (1996) An (H)C(CO)NH-TOCSY pulse scheme for sequential assignment of protonated methyl groups in otherwise deuterated ^{15}N, ^{13}C-labeled proteins. J Biomol NMR 8:351–356
Gil S, Hošek T, Solyom Z et al (2013) NMR studies of intrinsically disordered proteins near physiological conditions. Angew Chem Int Ed 52:11808–11812
Gil S, Favier A, Brutscher B (2014) HNCA+, HNCO+, and HNCACB+ experiments: improved performance by simultaneous detection of orthogonal coherence transfer pathways. J Biomol NMR 60:1–9
Gossert AD, Hiller S, Fernández C (2011) Automated NMR resonance assignment of large proteins for protein-ligand interaction studies. J Am Chem Soc 133:210–213
Grzesiek S, Bax A (1992a) Correlating backbone amide and side chain resonances in larger proteins by multiple relayed triple resonance NMR. J Am Chem Soc 114:6291–6293
Grzesiek S, Bax A (1992b) Improved 3D triple-resonance NMR techniques applied to a 31 KDa protein. J Magn Reson 96:432–440
Grzesiek S, Bax A (1993) Amino acid type determination in the sequential assignment procedure of uniformly ^{13}C/^{15}N-enriched proteins. J Biomol NMR 3:185–204
Grzesiek S, Anglister J, Bax A (1993a) Correlation of backbone amide and aliphatic side-chain resonances in ^{13}C/^{15}N-enriched proteins by isotropic mixing of ^{13}C magnetization. J Magn Reson Ser B 101:114–119
Grzesiek S, Anglister J, Ren H et al (1993b) ^{13}C line narrowing by ^{2}H decoupling in ^{2}H/^{13}C/^{15}N-enriched proteins. Application to triple resonance 4D J connectivity of sequential amides. J Am Chem Soc 115:4369–4370
Hiller S, Fiorito F, Wüthrich K et al (2005) Automated projection spectroscopy (APSY). Proc Natl Acad Sci U S A 102:10876–10881
Hiller S, Wasmer C, Wider G et al (2007) Sequence-specific resonance assignment of soluble nonglobular proteins by 7D APSY-NMR spectroscopy. J Am Chem Soc 129:10823–10828
Hiller S, Joss R, Wider G (2008) Automated NMR assignment of protein side chain resonances using automated projection spectroscopy (APSY). J Am Chem Soc 130:12073–12079

Hoch JC, Stern AS (1996) NMR data processing. Wiley-Liss, New York
Holland DJ, Bostock MJ, Gladden LF et al (2011) Fast multidimensional NMR spectroscopy using compressed sensing. Angew Chem Int Ed Engl 50:6548–6551
Hoult DI, Richards RE (1976) The signal-to-noise ratio of the nuclear magnetic resonance experiment. J Magn Reson 24:71–85
Ikura M, Kay LE, Bax A (1990) A novel approach for sequential assignment of ^{1}H, ^{13}C and ^{15}N spectra of larger proteins: heteronuclear triple-resonance three-dimensional NMR spectroscopy. Application to calmodulin. Biochemistry 29:4659–4667
Jung YS, Zweckstetter M (2004) MARS: robust automatic backbone assignment of proteins. J Biomol NMR 30:11–23
Kadeřávek P, Zapletal V, Rabatinová A et al (2014) Spectral density mapping protocols for analysis of molecular motions in disordered proteins. J Biomol NMR 58:193–207
Kanelis V, Donaldson L, Muhandiram DR et al (2000) Sequential assignment of proline-rich regions in proteins: application to modular binding domain complexes. J Biomol NMR 16:253–259
Karplus M (1959) Contact electron-spin coupling of nuclear magnetic moments. J Chem Phys 30:11–15
Kay LE, Torchia DA, Bax A (1989) Backbone dynamics of proteins as studied by ^{15}N inverse detected heteronuclear NMR spectroscopy: application to staphylococcal nuclease. Biochemistry 28:8972–8979
Kay LE, Ikura M, Tschudin R et al (1990) Three-dimensional triple-resonance NMR spectroscopy of isotopically enriched proteins. J Magn Reson 89:496–514
Kay LE, Ikura M, Zhu G et al (1991) Four-dimensional heteronuclear triple resonance NMR of isotopically enriched proteins for sequential assignment of backbone atoms. J Magn Reson 91:422–428
Kay LE, Xu GY, Yamazaki T (1994) Enhanced-sensitivity triple-resonance spectroscopy with minimal H_2O saturation. J Magn Reson Ser A 109:129–133
Kazimierczuk K, Orekhov VY (2011) Accelerated NMR spectroscopy by using compressed sensing. Angew Chem Int Ed Engl 50:5556–5559
Kazimierczuk K, Zawadzka A, Koźmiński W et al (2006) Random sampling of evolution time space and Fourier transform processing. J Biomol NMR 36:157–168
Kazimierczuk K, Zawadzka A, Koźmiński W et al (2007) Lineshapes and artifacts in Multidimensional Fourier Transform of arbitrary sampled NMR data sets. J Magn Reson 188:344–356
Kazimierczuk K, Stanek J, Zawadzka-Kazimierczuk A et al (2010a) Random sampling in multidimensional NMR spectroscopy. Prog NMR Spectrosc 57:420–434
Kazimierczuk K, Zawadzka-Kazimierczuk A, Koźmiński W (2010b) Non-uniform frequency domain for optimal exploitation of non-uniform sampling. J Magn Reson 205:286–292
Kazimierczuk K, Misiak M, Stanek J et al (2012) Generalized Fourier transform for non-uniform sampled data. Top Curr Chem 316:79–124
Kazimierczuk K, Stanek J, Zawadzka-Kazimierczuk A et al (2013) High-dimensional NMR spectra for structural studies of biomolecules. ChemPhysChem 14:3015–3025
Kern T, Schanda P, Brutscher B (2008) Sensitivity-enhanced IPAP-SOFAST-HMQC for fast-pulsing 2D NMR with reduced radiofrequency load. J Magn Reson 190:333–338
Kim S, Szyperski T (2003) GFT NMR, a new approach to rapidly obtain precise high-dimensional NMR spectral information. J Am Chem Soc 125:1385–1393
Kim S, Szyperski T (2004) GFT NMR experiments for polypeptide backbone and $^{13}C_{\beta}$ chemical shift assignment. J Biomol NMR 28:117–130
Kjaergaard M, Poulsen FM (2011) Sequence correction of random coil chemical shifts: correlation between neighbor correction factors and changes in the Ramachandran distribution. J Biomol NMR 50:157–165
Kjaergaard M, Poulsen FM (2012) Disordered proteins studied by chemical shifts. Prog NMR Spectrosc 60:42–51
Kjaergaard M, Brander S, Poulsen FM (2011) Random coil chemical shift for intrinsically disordered proteins: effects of temperature and pH. J Biomol NMR 49:139–149

Konrat R, Yang D, Kay LE (1999) A 4D TROSY-based pulse scheme for correlating $^{1}HN_{i}$, $^{15}N_{i}$, $^{13}C^{a}_{i}$, $^{13}C'_{i-1}$ chemical shifts in high molecular weight, ^{15}N, ^{13}C, ^{2}H labeled proteins. J Biomol NMR 15:309–313

Kovacs H, Moskau D, Spraul M (2005) Cryogenically cooled probes – a leap in NMR technology. Prog NMR Spectrosc 46:131–155

Kumar D, Hosur RV (2011) hNCOcanH pulse sequence and a robust protocol for rapid and unambiguous assignment of backbone ($^{1}H^{N}$, ^{15}N and $^{13}C'$) resonances in $^{15}N/^{13}C$-labeled proteins. Magn Reson Chem 49:575–583

Kupce E, Freeman R (2003) Projection-reconstruction of three-dimensional NMR spectra. J Am Chem Soc 125:13958–13959

Kupce E, Nishida T, Freeman R (2003) Hadamard NMR spectroscopy. Prog NMR Spectr 42: 95–122

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

Levitt MH, Freeman R, Frenkiel T (1982) Broadband heteronuclear decoupling. J Magn Reson 47:328–330

Lipari G, Szabo A (1982) Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 1. Theory and range of validity. J Am Chem Soc 104:4546–4559

Logan TM, Olejniczak ET, Xu RX et al (1992) Side chain and backbone assignments in isotopically labeled proteins from two heteronuclear triple resonance experiments. FEBS Lett 314:413–418

Logan TM, Olejniczak ET, Xu RX et al (1993) A general method for assigning NMR spectra of denatured proteins using 3D HC(CO)NH-TOCSY triple resonance experiments. J Biomol NMR 3:225–231

Löhr F, Rüterjans H (1995) A new triple-resonance experiment for the sequential assignment of backbone resonances in proteins. J Biomol NMR 6:189–197

Löhr F, Rüterjans H (1997) HNCO-E.COSY, a simple method for the stereospecific assignment of side-chain amide protons in proteins. J Magn Reson 124:255–258

López-Méndez B, Güntert P (2006) Automated protein structure determination from NMR spectra. J Am Chem Soc 128:13112–13122

Luan T, Jaravine V, Yee A et al (2005) Optimization of resolution and sensitivity of 4D NOESY using multi-dimensional decomposition. J Biomol NMR 33:1–14

Malmodin D, Billeter M (2005) Multiway decomposition of NMR spectra with coupled evolution periods. J Am Chem Soc 127:13486–13487

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

Mäntylahti S, Hellman M, Permi P (2011) Extension of the HA-detection based approach: (HCA) CON(CA)H and (HCA)NCO(CA)H experiments for the main-chain assignment of intrinsically disordered proteins. J Biomol NMR 49:99–109

Matsuki Y, Eddy MT, Herzfeld J (2009) Spectroscopy by integration of frequency and time domain information for fast acquisition of high-resolution dark spectra. J Am Chem Soc 131: 4648–4656

McConnell HM (1958) Reaction rates by nuclear magnetic resonance. J Chem Phys 28:430–431

McIntosh LP, Dahlquist FW (1990) Biosynthetic incorporation of ^{15}N and ^{13}C for assignment and interpretation of nuclear magnetic resonance spectra of proteins. Q Rev Biophys 23:1–38

Mobli M, Stern AS, Hoch JC (2006) Spectral reconstruction methods in fast NMR: reduced dimensionality, random sampling and maximum entropy. J Magn Reson 182:96–105

Mobli M, Stern AS, Bermel W et al (2010) A non-uniformly sampled 4D HCC(CO)NH-TOCSY experiment processed using maximum entropy for rapid protein sidechain assignment. J Magn Reson 204:160–164

Montelione GT, Lyons BA, Emerson SD et al (1992) An efficient triple resonance experiment using carbon-13 isotropic mixing for determining sequence-specific resonance assignments of isotopically-enriched proteins. J Am Chem Soc 114:10974–10975

Morris GA, Freeman R (1979) Enhancement of nuclear magnetic resonance signals by polarization transfer. J Am Chem Soc 101:760–762

Muhandiram DR, Kay LE (1994) Gradient-enhanced triple resonance three-dimensional NMR experiments with improved sensitivity. J Magn Reson Ser B 103:203–216

Mukrasch MD, Bibow S, Korukottu J et al (2009) Structural polimorphism of 441-residue tau at single residue resolution. PLoS Biol 7:e34

Neuhaus D, Williamson M (1989) The nuclear Overhauser effect in structural and conformational analysis. Wiley, New York

Nietlispach D (2004) A selective intra-HN(CA)CO experiment for the backbone assignment of deuterated proteins. J Biomol NMR 28:131–136

Nietlispach D, Ito Y, Laue ED (2002) A novel approach for the sequential backbone assignment of larger proteins: selective intra-HNCA and DQ-HNCA. J Am Chem Soc 124:11199–207

Nováček J, Zawadzka-Kazimierczuk A, Papoušková V et al (2011) 5D ^{13}C-detected experiments for backbone assignment of unstructured proteins with a very low signal dispersion. J Biomol NMR 50:1–11

Nováček J, Haba NY, Chill JH et al (2012) 4D Non-uniformly sampled HCBCACON and ^{1}J(NCα)-selective HCBCANCO experiments for the sequential assignment and chemical shift analysis of intrinsically disordered proteins. J Biomol NMR 53:139–148

Nováček J, Janda L, Dopitová R et al (2013) Efficient protocol for backbone and side-chain assignments of large, intrinsically disordered proteins: transient secondary structure analysis of 49.2 kDa microtubule associated protein 2c. J Biomol NMR 56:291–301

O'Hare B, Benesi AJ, Showalter SA (2009) Incorporating ^{1}H chemical shift determination into 13C-direct detected spectroscopy of intrinsically disordered proteins in solution. J Magn Reson 200:354–358

Olejniczak ET, Fesik SW (1994) Two dimensional nuclear magnetic resonance method for identifying the H^{α}-C^{α} signals of amino acid residues preceding prolines. J Am Chem Soc 116:2215–2216

Otten R, Wood K, Mulder FAA (2009) Comprehensive determination of $^{3}J_{HNH\alpha}$ for unfolded proteins using ^{13}C'-resolved spin-echo difference spectroscopy. J Biomol NMR 45:343–49

Palmer AG III (2004) NMR characterization of the dynamics of biomacromolecules. Chem Rev 104:3623–3640

Palmer AG III, Massi F (2006) Characterization of the dynamics of biomacromolecules using rotating-frame spin relaxation NMR spectroscopy. Chem Rev 106:1700–1719

Palmer AG III, Kroenke CD, Loria JP (2001) Nuclear magnetic resonance methods for quantifying microsecond-to-millisecond motions in biological macromolecules. Methods Enzymol 339:204–238

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

Pantoja-Uceda D, Santoro J (2008) Amino acid type identification in NMR spectra of proteins via β- and γ-carbon edited experiments. J Magn Reson 195:187–195

Pantoja-Uceda D, Santoro J (2012) New amino acid residue type identification experiments valid for protonated and deuterated proteins. J Biomol NMR 54:145–153

Pellecchia M, Wider G, Iwai H et al (1997) Arginine side chain assignments in uniformly ^{15}N-labeled proteins using the novel 2D HE(NE)HGHH experiment. J Biomol NMR 10: 193–197

Peng JW, Wagner G (1992) Mapping of spectral density function using heteronuclear NMR relaxation measurements. J Magn Reson 98:308–332

Peng JW, Wagner G (1994) Investigation of protein motions via relaxation measurements. Methods Enzymol 239:563–596

Pervushin K, Riek R, Wider G et al (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 U S A 94:12366–12371

Piai A, Hošek T, Gonnelli L et al (2014) "CON-CON" assignment strategy for highly flexible intrinsically disordered proteins. J Biomol NMR 60:209–218
Purcell EM, Torrey HC, Pound RV (1946) Resonance absoption by nuclear magnetic moments in solid. Phys Rev 69:37–38
Rao NS, Legault P, Muhandiram DR et al (1996) NMR pulse schemes for the sequential assignment of arginine side-chain H^{ε} protons. J Magn Reson B 113:272–276
Redfield AG (1957) On the theory of relaxation processes. IBM. J Res Develop 1:19–31
Rios C B, Feng W, Tashiro M et al (1996) Phase labeling of C-H and C-C spin-system topologies: application in constant-time PFG-CBCA(CO)NH experiments for discriminating amino acid spin-system types. J Biomol NMR 8:345–350
Sattler M, Schleucher J, Griesinger C (1999) Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. Progr NMR Spectrosc 34:93–158
Schanda P (2009) Fast-pulsing longitudinal relaxation optimized techniques: enriching the toolbox. Prog NMR Spectrosc 55:238–265
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, Forge V, Brutscher B (2006a) HET-SOFAST NMR for fast detection of structural compactness and heterogeneity along polypeptide chains. Magn Reson Chem 44:S177–S184
Schanda P, Van Melckebeke H, Brutscher B (2006b) Speeding up three-dimensional protein NMR experiments to a few minutes. J Am Chem Soc 128:9042–9043
Schleucher J, Sattler M, Griesinger C (1993) Coherence selection by gradients without signal attenuation: application to the three-dimensional HNCO experiment. Angew Chem Int Ed Engl 32:1489–1491
Schubert M, Smalla M, Schmieder P et al (1999) MUSIC in triple-resonance experiments: amino acid type-selective ^{1}H-^{15}N correlations. J Magn Reson 141:34–43
Schubert M, Oschkinat H, Schmieder P (2001a) MUSIC and aromatic residues: amino acid type-selective ^{1}H-^{15}N correlations, III. J Magn Reson 153:186–192
Schubert M, Oschkinat H, Schmieder P (2001b) MUSIC, selective pulses, and tuned delays: amino acid-type selective ^{1}H-^{15}N correlations, II. J Magn Reson 148:61–72
Schwarzinger S, Kroon GJ, Foss TR et al (2001) Sequence-dependent correction of random coil NMR chemical shifts. J Am Chem Soc 123:2970–2978
Selenko P, Wagner G (2007) Looking into live cells with in-cell NMR spectroscopy. J Struct Biol 158:244–253
Selenko P, Frueh DP, Elsaesser SJ et al (2008) In situ observation of protein phosphorylation by high-resolution NMR spectroscopy. Nat Struct Mol Biol 15:321–329
Serber Z, Selenko P, Hänsel R et al (2006) Investigating macromolecules inside cultured and injected cells by in-cell NMR spectroscopy. Nat Protoc 1:2701–2709
Shaka AJ, Keeler J, Freeman R (1983a) Evaluation of a new broadband decoupling sequence: WALTZ-16. J Magn Reson 53:313–340
Shaka AJ, Keeler J, Frenkiel T et al (1983b) An improved sequence for broadband decoupling: WALTZ-16. J Magn Reson 52:335–338
Shaka AJ, Barker PB, Freeman R (1985) Computer-optimized decoupling scheme for wideband applications and low-level operation. J Magn Reson 64:547–552
Shaka AJ, Lee CJ, Pines A (1988) Iterative schemes for bilinear operators; application to spin decoupling. J Magn Reson 77:274–293
Shimba N, Stern AS, Craik CS et al (2003) Elimination of $^{13}C^{\alpha}$ splitting in protein NMR spectra by deconvolution with maximum entropy reconstruction. J Am Chem Soc 125:2382–2383
Shimba N, Kovacs H, Stern AS et al (2004) Optimization of ^{13}C direct detection NMR methods. J Biomol NMR 30:175–179
Shirakawa M, Wälchli M, Shimizu M et al (1995) The use of heteronuclear cross-polarization for backbone assignment of ^{2}H-, ^{15}N- and ^{13}C-labeled proteins: A pulse scheme for triple-resonance 4D correlation of sequential amide protons and ^{15}N. J Biomol NMR 5:323–326
Solomon I (1955) Relaxation processes in a system of two spins. Phys Rev 99:559–565

Solyom Z, Schwarten M, Geist L et al (2013) BEST-TROSY experiments for time-efficient sequential resonance assignment of large disordered proteins. J Biomol NMR 55:311–321

Spera S, Bax A (1991) Empirical correlation between protein backbone conformation and Ca and Cb ^{13}C nuclear magnetic resonance chemical shifts. J Am Chem Soc 113:5490–5492

Staykova DK, Fredriksson J, Bermel W et al (2008) Assignment of protein NMR spectra based on projections, multi-way decomposition and a fast correlation approach. J Biomol NMR 42:87–97

Szyperski T, Wider G, Bushweller JH et al (1993a) 3D ^{13}C-^{15}N-heteronuclear two-spin coherence spectroscopy for polypeptide backbone assignments in ^{13}C-^{15}N-double-labeled proteins. J Biomol NMR 3:127–132

Szyperski T, Wider G, Bushweller JH et al (1993b) Reduced dimensionality in triple resonance experiments. J Am Chem Soc 115:9307–9308

Takeuchi K, Heffron G, Sun ZY et al (2010) Nitrogen-detected CAN and CON experiments as alternative experiments for main chain NMR resonance assignments. J Biomol NMR 47: 271–282

Tamiola K, Acar B, Mulder FAA (2010) Sequence-specific random coil chemical shifts of intrinsically disordered proteins. J Am Chem Soc 132:18000–18003

Theillet FX, Kalmar L, Tompa P et al (2013) The alphabet of intrinsic disorder: I. Act like a Pro: on the abundance and roles of proline residues in intrinsically disordered proteins. Intr Dis Prot 1:e24360

Tjandra N, Grzesiek S, Bax A (1996) Magnetic field dependence of nitrogen-proton J splittings in ^{15}N-enriched human Ubiquitin resulting from relaxation interference and residual dipolar coupling. J Am Chem Soc 118:6264–6272

Tollinger M, Skrynnikov NR, Mulder FAA et al (2001) Slow dynamics in folded and unfolded states of an SH3 domain. J Am Chem Soc 123:11341–11352

Tolman J R, Flanagan JM, Kennedy MA et al (1995) Nuclear magnetic dipole interactions in field-oriented proteins: information for structure determination in solution. Proc Natl Acad Sci U S A 92:9279–9283

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

Tong KI, Yamamoto M, Tanaka T (2008) A simple method for amino acid selective isotope labeling of recombinant proteins in *E. coli*. J Biomol NMR 42:59–67

Tugarinov V, Kay LE, Ibraghimov I et al (2005) High-resolution four-dimensional ^{1}H-^{13}C NOE spectroscopy using methyl-TROSY, sparse data acquisition, and multidimensional decomposition. J Am Chem Soc 127:2767–2775

Tycko R (2006) Solid-state NMR as a probe of amyloid structure. Prot Pepr Lett 13:229–34

Vasos PR, Hall JB, Kümmerle R et al (2006) Measurement of ^{15}N relaxation in deuterated amide groups in proteins using direct nitrogen detection. J Biomol NMR 36:27–36

Vuister GW, Bax A (1993) Quantitative J correlation: a new approach for measuring homonuclear three-bond J($H^{N}H^{a}$) coupling constants in ^{15}N enriched proteins. J Am Chem Soc 115: 7772–7777

Wangsness RK, Bloch F (1953) The dynamical theory of nuclear induction. Phys Rev 89:728–739

Waugh JS (1982) Theory of broadband spin decoupling. J Magn Reson 50:30–49

Weisemann R, Rüterjans H, Bermel W (1993) 3D triple-resonance NMR techniques for the sequential assignment of NH and ^{15}N resonances in ^{15}N- and ^{13}C-labelled proteins. J Biomol NMR 3:113–120

Wishart DS, Sykes BD (1994) The ^{13}C chemical shift index: a simple method for the identification of protein secondary structure using ^{13}C chemical shift data. J Biomol NMR 4:171–180

Wishart DS, Sykes BD, Richards FM (1991) Relationship between nuclear magnetic resonance chemical shift and protein secondary structure. J Mol Biol 222:311–333

Wishart DS, Sykes BD, Richards FM (1992) The chemical shift index: a fast and simple method for the assignment of protein secondary structure through NMR spectroscopy. Biochemistry 31:1647–1651

Wishart DS, Bigam CG, Holm A et al (1995) ^{1}H, ^{13}C and ^{15}N random coil NMR chemical shifts of the common amino acids. I. Investigations of nearest-neighbor effects. J Biomol NMR 5:67–81

Wittekind M, Mueller L (1993) HNCACB, a high-sensitivity 3D NMR experiment to correlate amide-proton and nitrogen resonances with the α- and β-carbon resonances in proteins. J Magn Reson B 101:201–205

Wittekind M, Metzler WJ, Mueller L (1993) Selective correlations of amide groups to glycine alpha protons in proteins. J Magn Reson 101:214–217

Wüthrich K (1986) NMR of proteins and nucleic acids. Wiley, New York

Yao X, Becker S, Zweckstetter M (2014) J Biomol NMR 60(4):231–240

Xia Y, Arrowsmith CH, Szyperski T (2002) Novel projected 4D triple resonance experiments for polypeptide backbone chemical shift assignment. J Biomol NMR 24:41–50

Yamazaki T, Arrowsmith CH, Muhandiram DR et al (1994) A suite of triple resonance NMR experiments for the backbone assignment of ^{15}N, ^{13}C, ^{2}H labeled proteins with high sensitivity. J Am Chem Soc 116:11655–11666

Yamazaki T, Pascal SM, Singer AU et al (1995) NMR pulse schemes for the sequence-specific assignment of arginine guanidino ^{15}N and ^{1}H chemical shifts in proteins. J Am Chem Soc 117:3556–3564

Yang D, Kay LE (1999) TROSY triple-resonance four-dimensional NMR spectroscopy of a 46 ns tumbling protein. J Am Chem Soc 121:2571–2575

Ying J, Li F, Lee JH et al (2014) $^{13}C^{a}$ decoupling during direct observation of carbonyl resonances in solution NMR of isotopically enriched proteins. J Biomol NMR 60:15–21

Zawadzka-Kazimierczuk A, Kazimierczuk K, Koźmiński W (2010) A set of 4D NMR experiments of enhanced resolution for easy resonance assignment in proteins. J Magn Reson 202:109–116

Zawadzka-Kazimierczuk A, Koźmiński W, Billeter M (2012a) TSAR: a program for automatic resonance assignment using 2D cross-sections of high dimensionality, high-resolution spectra. J Biomol NMR 54:81–95

Zawadzka-Kazimierczuk A, Koźmiński W, Sanderová H et al (2012b) High dimensional and high resolution pulse seqeunces for backbone resonance assignment of intrinsically disordered proteins. J Biomol NMR 52:329–337

Zhang YZ (1995) Protein and peptide structure and interactions studied by hydrogen exchange and NMR. University of Pennsylvania, Philadelphia

Zhang H, Neal S, Wishart DS (2003) RefDB: a database of uniformly referenced protein chemical shifts. J Biomol NMR 25:173–195

Article 3. High-dimensionality ^{13}C direct-detected NMR experiments for the automatic assignment of intrinsically disordered proteins

Wolfgang Bermel[1], Isabella C. Felli[2], Leonardo Gonnelli[2], Wiktor Koźmiński[3], Alessandro Piai[2], Roberta Pierattelli[2], Anna Zawadzka-Kazimierczuk[3]

[1] Bruker BioSpin GmbH, Silberstreifen, 76287 Rheinstetten, Germany.

[2] CERM and Department of Chemistry "Ugo Schiff", University of Florence, 50019, Via Luigi Sacconi 6, Sesto Fiorentino, Florence, Italy.

[3] Faculty of Chemistry, Biological and Chemical Research Centre, University of Warsaw, Żwirki i Wigury 101, 02-089 Warsaw, Poland.

Published in 2013 in J.Biomol.NMR, 57, 353-361

J Biomol NMR (2013) 57:353–361
DOI 10.1007/s10858-013-9793-z

ARTICLE

High-dimensionality ^{13}C direct-detected NMR experiments for the automatic assignment of intrinsically disordered proteins

Wolfgang Bermel · Isabella C. Felli · Leonardo Gonnelli · Wiktor Koźmiński · Alessandro Piai · Roberta Pierattelli · Anna Zawadzka-Kazimierczuk

Received: 1 August 2013 / Accepted: 23 October 2013 / Published online: 8 November 2013

Abstract We present three novel exclusively heteronuclear 5D ^{13}C direct-detected NMR experiments, namely ($H^{N\text{-flip}}N$)CONCACON, (HCA)CONCACON and (H)CACON(CA)CON, designed for easy sequence-specific resonance assignment of intrinsically disordered proteins (IDPs). The experiments proposed have been optimized to overcome the drawbacks which may dramatically complicate the characterization of IDPs by NMR, namely the small dispersion of chemical shifts and the fast exchange of the amide protons with the solvent. A fast and reliable automatic assignment of α-synuclein chemical shifts was obtained with the Tool for SMFT-based Assignment of Resonances (*TSAR*) program based on the information provided by these experiments.

Electronic supplementary material The online version of this article (doi:10.1007/s10858-013-9793-z) contains supplementary material, which is available to authorized users.

W. Bermel
Bruker BioSpin GmbH, Silberstreifen, 76287 Rheinstetten, Germany

I. C. Felli (✉) · L. Gonnelli · A. Piai · R. Pierattelli (✉)
CERM, University of Florence, Via Luigi Sacconi 6, 50019 Sesto Fiorentino, Florence, Italy
e-mail: felli@cerm.unifi.it

R. Pierattelli
e-mail: pierattelli@cerm.unifi.it

I. C. Felli · R. Pierattelli
Department of Chemistry "Ugo Schiff", University of Florence, 50019 Sesto Fiorentino, Italy

W. Koźmiński · A. Zawadzka-Kazimierczuk
Faculty of Chemistry, Biological and Chemical Research Centre, University of Warsaw, Żwirki i Wigury 101, 02-089 Warsaw, Poland

Keywords Intrinsically disordered proteins · ^{13}C direct-detection NMR · Non-uniform sampling · Longitudinal relaxation enhancement · Multidimensional NMR experiment · Automatic assignment

Introduction

Intrinsically disordered proteins (IDPs) are a class of flexible proteins characterized by the lack of stable secondary and tertiary structures. Despite the absence of a defined fold, IDPs are able to perform a great variety of essential functions in the cell (Wright and Dyson 1999; Uversky et al. 2000; Dunker et al. 2001; Tompa 2002, 2009; Dyson and Wright 2005), leading to a reconsideration of the well-established structure–function paradigm.

X-ray crystallography is not an appropriate technique to characterize IDPs as the extreme mobility prevents the formation of suitable crystals, and NMR spectroscopy assumes the role of best candidate to investigate their disorder. In fact, NMR can provide structural and dynamic information at atomic resolution, giving a meaningful description of the properties of all the conformations sampled by the protein.

The most accessible and informative observable that can be obtained by NMR spectroscopy and used to characterize IDPs is the chemical shift of the signals of the amino-acids nuclei. These, once assigned, can in fact be used to calculate conformational properties, to evaluate propensities for transient secondary structures and to map possible interaction/binding regions. In contrast to what is usually found for globular and folded proteins, IDPs chemical shifts are not modulated by the presence of secondary structure and their values tend to be averaged out by the

fast conformational exchange processes of the protein. This fact of course leads to crowded spectra with high chances to encounter overlapping peaks. Moreover, amide protons not being protected by the presence of secondary structural elements, often experience chemical exchange with the solvent that can be so dramatic as to broaden their signals beyond detectable limits.

Exploiting heteronuclear chemical shifts is thus mandatory for assignment strategies of IDPs due to their high chemical shift dispersion as well as to their reduced sensitivity to exchange processes (Bax and Grzesiek 1993; Dyson and Wright 2001; Panchal et al. 2001; Hiller et al. 2005; Mittag and Forman-Kay 2007; Narayanan et al. 2010; Kumar and Hosur 2011; Mantylahti et al. 2011). Thanks to great recent improvements in instrumental sensitivity, direct detection of heteronuclei can be performed to take maximum advantage of these properties. ^{13}C direct detection exclusively heteronuclear NMR experiments were successfully used for the study of IDPs (Csizmok et al. 2008; Knoblich et al. 2009; O'Hare et al. 2009; Pérez et al. 2009; Hsu et al. 2009; Novacek et al. 2012). Several additional approaches to overcome the problem of extensive cross peaks overlap have been proposed (Bermel et al. 2012a, 2013; Pantoja-Uceda and Santoro 2012, 2013; Thakur et al. 2013), including the use of high multidimensionality experiments (4-5D) (Novacek et al. 2011, 2012, 2013; Bermel et al. 2012b; Haba et al. 2013), with the goal of increasing the resolution of the spectra. To reduce the experimental time while preserving high resolution in all indirect dimensions, non-uniform sampling (NUS) (Kazimierczuk et al. 2010) and approaches for longitudinal relaxation enhancement (LRE) proved to be very useful (Schanda et al. 2007; Bermel et al. 2009b; Solyom et al. 2013; Gil et al. 2013).

Starting from the above results, we propose here three new multidimensional NMR experiments in which ^{13}C direct-detection, NUS and LRE (when appropriate) are combined in order to overcome as much as possible limitations of IDPs. The accurate and unambiguous determination of chemical shifts guaranteed by these experiments allowed the *TSAR* program (Tool for SMFT-based assignment of resonances) (Zawadzka-Kazimierczuk et al. 2012) to perform a fast, reliable and automatic assignment of the collected frequencies. The combined use of these NMR experiments and software can thus be employed to speed up the sequence-specific resonance assignment of IDPs.

The effectiveness of the experiments proposed combined with the *TSAR* program were tested on a well-studied IDP of 140 amino acids, human α-synuclein. The strategy described here can successfully be applied to other IDPs, extending the size and complexity of proteins that can be investigated by NMR.

Materials and methods

All NMR experiments were performed at 16.4 T on a Bruker Avance spectrometer operating at 700.06 MHz ^{1}H, 176.03 MHz ^{13}C and 70.94 MHz ^{15}N frequencies, equipped with a cryogenically cooled probehead optimized for ^{13}C-direct detection. A sample of 1.0 mM uniformly ^{13}C, ^{15}N labeled human α-synuclein in 20 mM phosphate buffer at pH 6.5 was prepared as previously described (Huang et al. 2005). EDTA and NaCl were added to reach the final concentration of 0.5 and 200 mM, respectively, and 10 % D_2O was added for the lock. All experiments were acquired at 285.5 K. Parameters specific to each experiment are reported in the captions of the figures describing the pulse sequences, all reported in the Supplementary Material. For ^{13}C band-selective π/2 and π flip angle pulses Q5 (or time reversed Q5) and Q3 shapes (Emsley and Bodenhausen 1992) of durations of 300 and 220 μs, respectively, were used, except for the π pulses that should be band-selective on the C^{α} region (Q3, 860 μs) and for the adiabatic π pulse to invert both C' and C^{α} [smoothed Chirp 500 μs, 25 % smoothing, 80 kHz sweep width, 11.3 kHz RF field strength (Böhlen and Bodenhausen 1993)]. The ^{13}C band selective pulses on $C^{\alpha/\beta}$, C^{α}, and C' were given at the center of each region, respectively, and the adiabatic pulse was adjusted to cover the entire ^{13}C region. Decoupling of ^{1}H and ^{15}N was achieved with waltz16 (Shaka et al. 1983) (1.7 kHz) and garp4 (Shaka et al. 1985) (1.0 kHz) sequences, respectively. All gradients employed had 1 ms of duration and a sine-shape. Each experiment was acquired in a pseudo-2D mode, with the States method applied in all indirect dimensions to achieve quadrature detection. All experiments employ the IPAP approach (Bermel et al. 2006a, 2008) to remove the splitting in the direct acquisition dimension caused by the C^{α}–C' couplings. The in-phase (IP) and antiphase (AP) components were acquired and stored in an interleaved manner, doubling the number of FIDs recorded (Bermel et al. 2006c). All the experiments were performed using on-grid non-uniform sampling. The on-grid "Poisson disk" sampling scheme (Kazimierczuk et al. 2008) was chosen to generate the time schedules with the *RSPack* program. The distribution was relaxation-optimized, i.e. the density of points was decaying according to the Gaussian distribution $\exp(-t^2/\sigma^2)$, with $\sigma = 0.5$. All the spectra were acquired using *Bruker TopSpin 1.3* software. The experimental data were converted with *nmrPipe* (Delaglio et al. 1995). 3D data were processed using the Multidimensional Fourier Transform (MFT) (Kazimierczuk et al. 2006) algorithm implemented in the *ToASTD* program. 4/5D data were processed using the Sparse MFT (SMFT) algorithm (Kazimierczuk et al. 2009) implemented in the *reduced* program. Both programs, *ToASTD* and *reduced*, are included

in the MFT package software, available at http://nmr.cent3.uw.edu.pl/software. Finally, *Sparky* (Goddard and Kneller 2000) and *TSAR* (Tool for SMFT-based assignment of resonances) (Zawadzka-Kazimierczuk et al. 2012) programs were used to analyze the spectra and automatically assign the resonances, respectively.

Results and discussion

NMR experiments design

Our strategy exploits the use of ^{13}C direct detection both to avoid the problem of amide proton exchange as well as to increase the chemical shift dispersion in the direct dimension of the experiments. The superior chemical shift dispersion of the heteronuclei with respect to protons is exploited by exclusively labeling the chemical shifts of the backbone's heteronuclei, using protons only as a starting source of magnetization to increase the overall sensitivity of the experiments. The resolution of the experiments has been maximized correlating five different frequencies in five different dimensions, all built up using long evolution times exploiting the favorable relaxation properties of IDPs. To substantially reduce the experimental time and maximize resolution, non-uniform sampling (NUS) was employed in all the proposed experiments. In the 5D ($H^{N\text{-flip}}N$)CONCACON experiment, additionally the H-flip approach was used, in order to achieve a longitudinal relaxation enhancement and thus reduce the duration of the inter-scan delay. The H-flip approach (Bermel et al. 2009a; b) allows to manipulate only the desired set of ^1H spins to shorten the period needed for magnetization recovery to the $+z$ axis. For IDPs, the gain in time is usually more substantial when selectively inverting $^1H^N$ rather than $^1H^\alpha$ spins, because $^1H^N$s experience exchange processes with the solvent, hence not perturbing the water proton spins is of high importance. On the contrary, $^1H^\alpha$ spins which of course are not affected by exchange, do not experience significant longitudinal relaxation enhancement by their selective manipulation due to the small ^1H–^1H NOEs resulting from fast tumbling and low proton density in the absence of a 3D structure. For that reason, the H-flip approach was implemented only in the $^1H^N$-start version of the experiments.

Since data sampled at not uniform intervals cannot be treated with the Fast Fourier transform (FFT), the 5D experiments were processed using the Sparse Multidimensional Fourier Transform (SMFT) algorithm (Kazimierczuk et al. 2009). The features of this type of processing greatly simplifies the visualization of the multidimensional spectra, since the analysis of the 5D spectra is reduced to the inspection of a series of 2D cross-sections extracted from the full spectra at some fixed triads of frequencies. The fixed frequencies have been collected from a 3D (H)CACON lower dimensionality spectrum (the "basis spectrum"), because the correlations it provides are also shared by all the 5D experiments discussed here. Therefore, once the peaks of the 3D basis spectrum are collected, the analysis of the 5D spectra merely consists of the examination of a series of 2D planes. In these planes the two extra dimensions (with respect to the basis spectrum), containing the sequential correlations, are displayed. The information provided by the experiments is summarized in Fig. 1, while the coherence transfer pathways are reported in Fig. 2. In the latter it is also specified which dimensions were fixed with the SMFT algorithm and which were then extracted and analyzed in the 2D cross-sections.

Since the 3D (H)CACON experiment correlates the C^α and C' of a residue with the N of the following one (C^α_{i-1}–C'_{i-1}–N_i), the 5D experiments were designed with the idea of establishing the sequential correlations along C' and N frequencies, the most dispersed in terms of chemical shift (Bermel et al. 2013). The C'–N frequencies of two neighboring residues were thus connected, adding the chemical shift labeling of the C^α to minimize the occurrence of overlap. This choice has the great advantage that the same strategy can be used also when the 4D version of these experiments is acquired, since the same sequential connections can be established, even if without the help of the C^α dimension. In this case, it is sufficient to record a 2D CON experiment (to obtain the basis spectrum) and use the 4D spectra to sequentially link all the CON peaks (C'_{i-1}–N_i).

The magnetization transfer pathway of the 5D ($H^{N\text{-flip}}N$)CONCACON experiment (Fig. 1), once the common C^α_{i-1}–C'_{i-1}–N_i frequencies (in dimensions F3, F5 and F4 respectively, Fig. 2) are fixed, provides 2D cross-sections (F1–F2 planes) containing two sequential CON peaks (C'_{i-1}–N_i and C'_{i-2}–N_{i-1}). One of them correlates the same C'_{i-1}–N_i frequencies that are retrieved in the 3D (H)CACON experiment and, together with the C^α_{i-1} frequency, fixed while performing the SMFT algorithm, whereas the other provides the chemical shifts of the preceding CON peak (C'_{i-2}–N_{i-1}). The sequence-specific assignment can then be completed just linking together the 2D planes, as shown in Fig. 3, exploiting the chemical shift dispersion of C'_{i-1}–N_i correlations. The simultaneous exploitation of the C' and N frequencies reduces the ambiguities, since sequential correlations are established along both dimensions, and the frequency labeling of the C^α contributes to resolve overlaps in 2D CON cross-sections. However, the presence of proline residues interrupts the linkage between the residues, since this experiment exploits the amide protons as a starting point of the magnetization transfer pathway. This drawback is not so dramatic if the proline content of the protein sequence is not

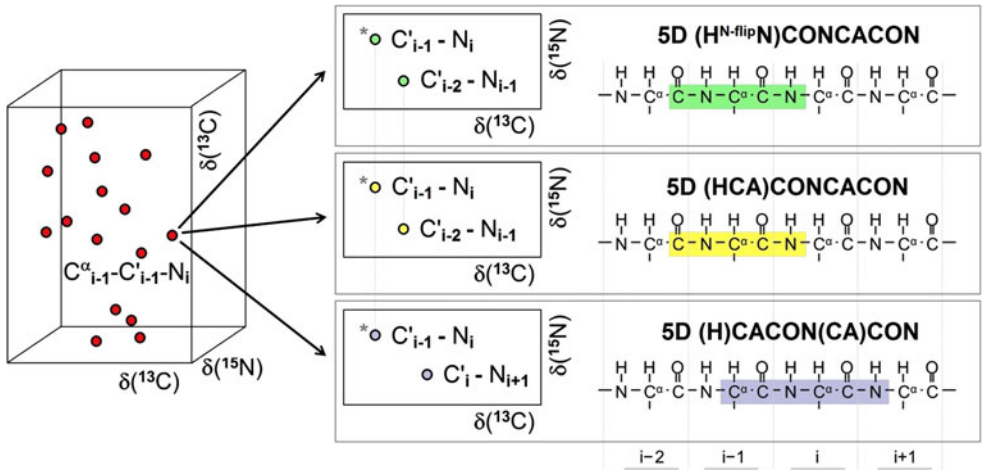


Fig. 1 The experimental strategy proposed together with the summary of the correlations provided by each experiment is schematically illustrated. Since the 5D experiments share C^{α}_{i-1}–C'_{i-1}–N_i frequencies, the 3D (H)CACON spectrum (reported on the *left*) is used as "basis spectrum" to collect these common correlations. After having fixed these frequencies with the SMFT algorithm, the information content of the various 5D experiments is easily retrieved inspecting a series of 2D cross-sections, reported in the middle part of the illustration, where the correlations provided by each experiment are shown. The CON peak which correlates the same C'_{i-1}–N_i frequencies as the associated peak in the basis spectrum is marked by an asterisk. On the right part of the figure, the nuclei involved in the coherence transfer pathways are highlighted on the backbone of the protein

large but prolines can be very abundant in IDPs (Radivojac et al. 2007). For such cases, the 5D (HCA)CONCACON experiment has to be preferred.

The 5D (HCA)CONCACON experiment can be considered as a variant of the 5D ($H^{N\text{-}flip}N$)CONCACON experiment previously described, since the magnetization transfer pathways, except for the first two steps, is the same (Fig. 1). However, the fact that this experiment exploits H^{α} protons as starting source of polarization ensures that also correlations involving the nuclei of prolines are detected. This experiment is thus recommended especially for proline-rich biomolecules and, used in combination with the previously described one, provides robust redundant information to reduce ambiguities in the assignment. Again, the sequence specific assignment is performed linking together the 2D cross-sections (F1–F2 planes) as shown in Fig. 4, where the peaks belonging to the nuclei of prolines are visible.

Both the above experiments were designed in order to provide negative peaks if the magnetization transfer pathway involves C^{α} of glycines, allowing their straightforward identification. This is obtained setting the C^{α} evolution time to $1/J_{C\alpha\text{-}C\beta}$. As an example, in Fig. 3 six sequential cross-sections of the 5D ($H^{N\text{-}flip}N$)CONCACON spectrum in which a glycine is present are reported.

Due to the fact that both experiments provide the same sequential links, one may opt for acquiring only one instead of both experiments. In this case, to choose between the 5D ($H^{N\text{-}flip}N$)CONCACON and 5D (HCA)CONCACON, one should consider proline content of the protein, level of chemical exchange, the available NMR time, stability and concentration of the sample.

Finally, a different version of the 5D (HCA)CONCACON experiment was also designed, in which the first C^{α} is frequency labeled instead of the second. This experiment, named 5D (H)CACON(CA)CON, provides different information with respect to the previously described ones, since in the 2D cross-sections (F5–F4 planes) C'_{i-1}–N_i and C'_i–N_{i+1} correlations are detectable. The sequential assignment can thus be accomplished starting from the N-terminal part and going towards the C-terminal end of the protein, whereas in the previous two experiments it proceeds in the opposite direction. Apart from this, the way to connect the cross-sections is essentially the same, since in each of them two sequential CON peaks are present. Since the H^{α} protons are used at the beginning of the magnetization transfer pathway, the prolines can be detected in this experiment too. The two C^{α} evolution times are set around $1/J_{C\alpha\text{–}C\beta}$, which means that the relative sign of cross peaks cannot be used as a straightforward criterion to identify glycines as in the previous experiments, although this can be used as a criterion to cross check results. Another difference between this experiment and those previously described is that here the direct dimension (F5) is not fixed when using the SMFT algorithm, since the C^{α}_{i-1}–C'_{i-1}–N_i frequencies that have to be fixed are all present in the indirect dimensions (F1, F2 and F3 dimensions). Artifact ridges are then present along the direct dimension (F5) of the 2D cross-sections (F5–F4 planes) in which the correlations for the sequential assignment are present (Figure S5). Even if they are not a limitation to the effectiveness of the experiment, it has to be

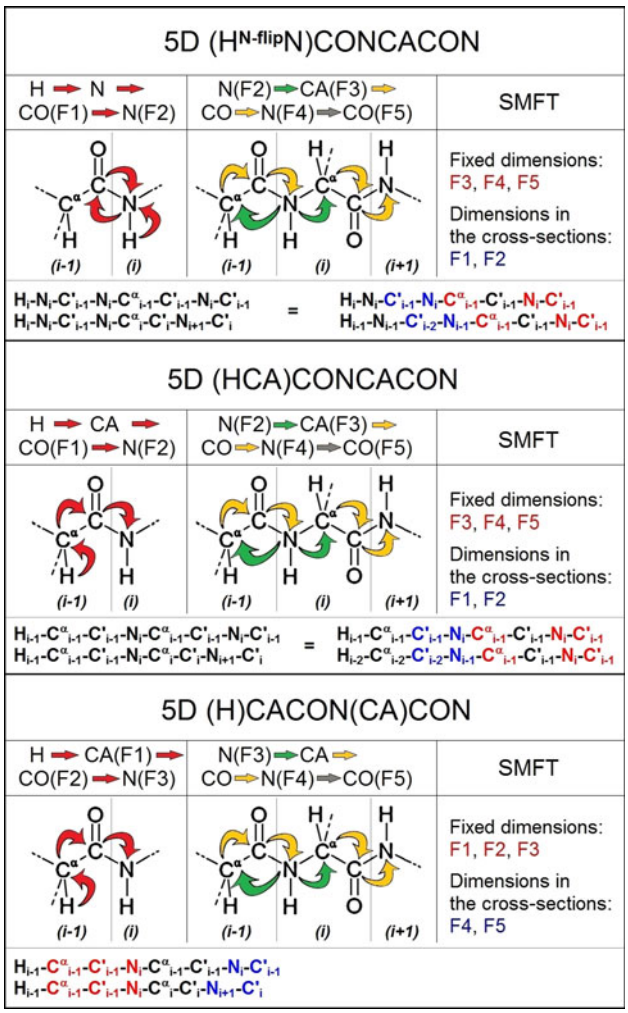


Fig. 2 The flow of the magnetization in the proposed experiments is schematically illustrated. The indirect dimensions in which chemical shift evolutions are present are indicated as F1, F2, F3 and F4, whereas the direct dimension is designated as F5. On the *left side*, the beginning of the coherence pathway of each experiment is shown: ^{1}J scalar couplings were used to transfer magnetization (*red arrows*). In the middle part of the panel, the final steps of the coherence pathways, which are common to all the experiments, are reported. Only the *arrow* representing the last transfer from N to C′ (*grey arrow* in the summary), prior to acquisition, is not shown along the backbone. Also in this case, *arrows* symbolize magnetization transfers mediated by scalar couplings: *green* when ^{1}J and ^{2}J were involved and yellow when only ^{1}J were exploited. On the *right side*, a summary of which dimensions were fixed with the SMFT algorithm and which were then extracted in the cross-sections is reported. Finally, in the bottom of each panel, the correlations retrieved in the experiment are listed explicitly

reported that they may complicate the analysis of the cross-sections and the peak picking procedure.

Automatic assignment using TSAR

Sequence-specific assignment of resonances based on the proposed experiments can be performed automatically by

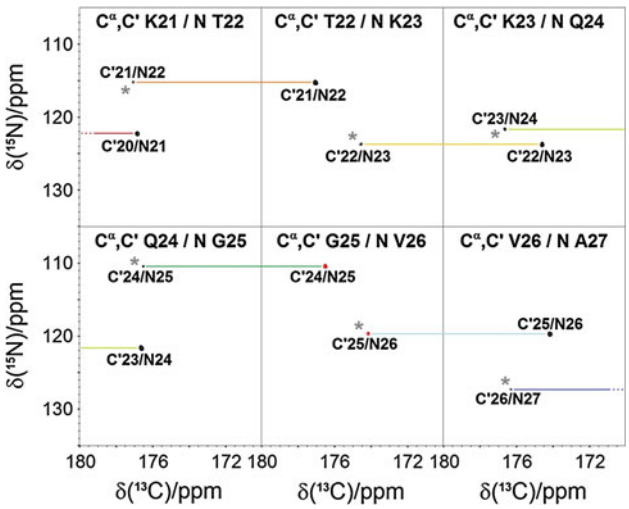


Fig. 3 Six 2D cross-sections of the (H$^{N\text{-}flip}$N)CONCACON spectrum are shown. In the top of each panel, the C^{α}_{i}–C'_{i}–N_{i+1} frequencies that have been fixed to extract the particular 2D plane are reported. In each cross-section, two sequential CON peaks are present. The CON peak which correlates the C'_{i-1}–N_i frequencies that have been used together with C^{α}_{i-1} frequency to extract the particular cross-section is labeled with an *asterisk*. *Black* peaks are positive, *red* negative. Inversion of peak sign is expected when the C^{α} belonging to a glycine is frequency labeled. Negative peaks thus allow to immediately identify glycines. The sequential specific assignment is performed connecting the CON peaks of two consecutive cross-sections, as shown with *colored lines* in the figure. Please note that the peaks are correlated using both N and C′ chemical shift, thus making the sequential assignment extremely reliable. As clearly visible, the use of five different dimensions minimize the chance of possible overlaps

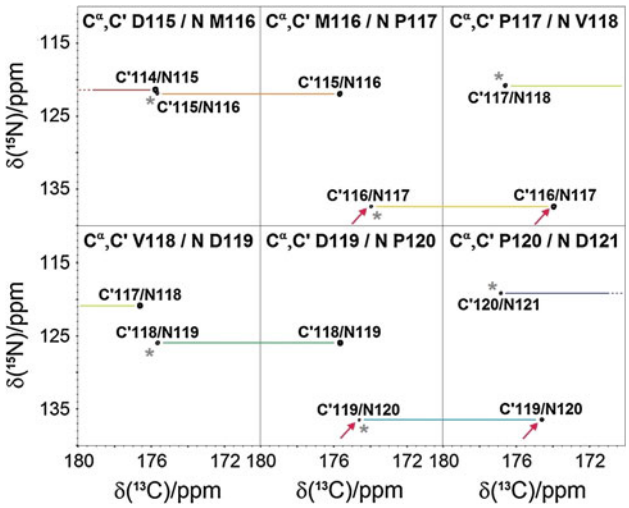


Fig. 4 Six 2D cross-sections of the (HCA)CONCACON spectrum are shown. All the labels, *lines* and *asterisks* have the same meaning as described in the captions of Fig. 3. The peaks in which the nitrogen belonging to a proline is frequency labeled (highlighted with a *red arrow*), not present in the (H$^{N\text{-}flip}$N)CONCACON spectrum, are here visible. The sequential specific assignment is thus not interrupted. The use of five different dimensions, together with the high resolution obtained in each indirect dimension, ensure to resolve very close peaks (as the case reported in the *top left panel*)

using the *TSAR* program (Zawadzka-Kazimierczuk et al. 2012), available at http://nmr.cent3.uw.edu.pl/software. This program is designed to analyze the multidimensional data processed with SMFT. What distinguishes *TSAR* from other automatic assignment algorithms is a very informative output, including also information on the correct order of cross-sections. This provides easy access to spectral data thus facilitating manual verification and/or completion of assignment. Input for the *TSAR* program includes the peak list of the basis spectrum and one or more peak lists obtained from higher-dimensional experiments. The latter ones contain only the two dimensions not fixed for SMFT, as three other frequencies are already specified in the peak list of the basis spectrum. The experiments used for *TSAR* assignment should be defined in a text file, where peak types together with their intensity signs (positive, negative, unknown or changing under specific conditions) are described (for details see Supplementary Material).

One major modification of the *TSAR* program was made to adjust it to the newly proposed experiments. In the previous version *TSAR* accepted peaks originating from three consecutive residues (on each cross-section). Now this range was enlarged to accept up to four consecutive residues. This was necessary as the simultaneous analysis of e.g. 5D (HCA)CONCACON and 5D (H)CACON(CA)CON data imply that the cross-sections obtained by fixing C^{α}_{i-1}–C'_{i-1}–N_i frequencies contain C'_{i-2}–N_{i-1}, C'_{i-1}–N_i and C'_i–N_{i+1} peaks, thus involving residues $i-2$, $i-1$, i and $i+1$.

Operation of *TSAR* includes construction of cross-section spin systems (CSSSs), where the information of corresponding cross-sections from various spectra are collected, formation of chains of CSSSs and mapping of these chains onto the primary sequence of the protein. The last step is based on identification of amino-acid type(s) for some of the CSSSs. Thus, to operate correctly, *TSAR* requires at least one experiment providing sequential connectivities and at least one experiment providing information on amino-acid type of some CSSSs. Two types of methods for amino-acid type recognition are available: statistical and topological. The former one utilizes C^{β} and/or H^{β} chemical shifts and analyzes them using the *Biological Magnetic Resonance Bank* (BMRB, Ulrich et al. 2007) statistics. The latter one exploits information on amino acid structure (e.g. absence of H^N proton in proline or presence of two possibly different H^{β} protons in all residues except for alanine, isoleucine, threonine and valine). Within the topological methods one routine is especially useful: recognition of glycine, based on peak amplitude sign. As mentioned above, it utilizes the fact that in some techniques scalar coupling between C^{α} and C^{β} carbon nuclei evolves for approximately $1/^1J_{C\alpha-C\beta}$, thus the sign of a peak corresponding to the glycine residue, not possessing the C^{β}, is inverted with respect to those of all other amino acid residues. In perspective, variants of the pulse sequences which select correlations deriving only from specific aminoacid-types can provide very useful information to contribute to the unambiguous identification of selected residue-types (Bermel et al. 2012a).

All the 5D experiments proposed in the current paper provide the sequential connectivities. Two experiments (5D ($H^{N\text{-flip}}$N)CONCACON and 5D (HCA)CONCACON) enable peak-sign-based glycine recognition, which is sufficient for CSSS's chain mapping if the content of glycine residues in the studied protein is substantial. The information on amino-acid type provided by 5D (H)CACON(CA)CON experiment cannot be exploited in the current version of the TSAR program, due to ambiguities in its interpretation: negative peak means that glycine is at $i-1$ or i position (but not at both of these positions).

Since the test protein (α-synuclein) contains almost 13 % of glycine residues in its primary sequence, glycine recognition is in principle sufficient for mapping the chains of cross-sections. Due to the protein size and to the fact that it is fully disordered, we decided however to use at least two sequential experiments able to provide links, reducing possible ambiguities in chains formation. We have constructed two data sets consisting of data from two experiments, plus 3D (H)CACON as a basis spectrum. Data set 1 contained 5D ($H^{N\text{-flip}}$N)CONCACON and 5D (H)CACON(CA)CON, while data set 2 contained 5D (HCA)CONCACON and 5D (H)CACON(CA)CON. In both cases the first experiment ensured peak-sign based glycine recognition and both experiments provided the sequential connectivities. The assignment was successful: 87.1 and 91.8 % of correctly assigned resonances were obtained for data set 1 and 2, respectively.

Despite the apparent redundancy of the information (both experiments from each data set provide the sequential links via N and C′ chemical shifts), adding the 5D (H)CACON(CA)CON to the single-experiment data sets enabled resolving the ambiguities. This is because this experiment allows to make the forward link (C'_{i-1}–N_i, C'_i–N_{i+1}), while two others show backward links (C'_{i-2}–N_{i-1}, C'_{i-1}–N_i).

Both data sets above allowed to assign the chemical shifts of N, C′ and C^{α} nuclei. When H^N, H^{α}, H^{β} and/or C^{β} chemical shifts are required as well, some other experiments have to be employed additionally. Preferably, ^{13}C-detected experiments which can be SMFT-processed using the same basis peak list (H)CACON should be employed. For instance these can be the previously published (Bermel et al., 2012b) 5D $H^{N\text{-flip}}$NCACON (containing H^N_i–N_i and H^N_{i-1}–N_{i-1} peaks on each cross-section) and 4D HCBCACON (containing H^{α}_{i-1}–C^{α}_{i-1} and H^{β}_{i-1}–C^{β}_{i-1} peaks on each cross-section). To assess the assignment capability using these sets of

Table 1 Summary of the types of experiments reported in this work

Experiment label	Name	Dimensionality	Experimental time (h)
A (basis spectrum)	(H)CACON	3D	13
B	($H^{N\text{-flip}}$N)CONCACON	5D	58
C	(HCA)CONCACON	5D	71
D	(H)CACON(CA)CON	5D	54
E	$H^{N\text{-flip}}$NCACON	5D	14
F	HCBCACON	4D	28

Table 2 *TSAR* assignment results for different data sets

Data set number	Experiments[a]	Assigned nuclei	Total exp. time (h)	Number of resonances			% of correctly assigned resonances
				Assigned correctly	Assigned incorrectly	All	
1	A + B+D	N, C′,C^{α}	123	363	10	417	87.1
2	A + C+D	N, C′,C^{α}	138	383	9	417	91.8
3	A + C+D + E	H^{N}, N, C′,C^{α}	152	511	11	551	92.7
4	A + B+D + F	N, C′,C^{α},C^{β},H^{α},H^{β}	153	707	6	798	88.6
5	A + D+E + F	H^{N}, N, C′,C^{α},C^{β},H^{α},H^{β}	109	808	7	932	86.7

[a] See Table 1 for experiments' labeling

information, three further data sets were constructed: data set 3 additionally providing H^{N} chemical shifts (5D (H)CACON(CA)CON, 5D (HCA)CONCACON and 5D $H^{N\text{-flip}}$NCACON experiments), data set 4 additionally providing H^{α}, H^{β} and C^{β} chemical shifts (5D ($H^{N\text{-flip}}$N)CONCACON, 5D (H)CACON(CA)CON and 4D HCBCACON experiments) and data set 5 providing all the backbone chemical shift (5D (H)CACON(CA)CON, 5D $H^{N\text{-flip}}$NCACON and 4D HCBCACON experiments). The results were as follows: 92.7 % of correctly assigned resonances for data set 3, 88.6 % for data set 4 and 86.7 % for data set 5. Please note, that the percentages were calculated with respect to the number of all resonances possible to assign using the given techniques, which was different for each data set (see Table 2). Thus, larger number of correctly assigned resonances does not mean that the fraction of correctly assigned resonances will also be larger. The fractions of incorrect assignments varied from 0.8 up to 2.4 % of assigned resonances. The reason for which the fractions for data sets 4 and 5 are relatively low is as follows. Both of these data sets contain 4D HCBCACON experiment that provides H^{α}, H^{β} and C^{β} chemical shifts. Each of these frequencies is known just from a single cross-section. This is not the case with N, C′ and C^{α} resonances, which are usually known from two or three cross-sections corresponding to consecutive residues. That is why if one cross-section is missing in assignment, it does not necessarily mean that any N, C′ or C^{α} resonances will be missing (usually they can be obtained from adjacent cross-sections), but it always means that some H^{α}, H^{β} and C^{β} resonances will be missing.

All the discussed data sets were gathered in Tables 1 and 2, together with experimental times required to record the respective data sets. Correctly and incorrectly assigned chemical shifts by *TSAR* were recognized upon comparison with the ones deposited in the BMRB, entry 6968 (Bermel et al. 2006b, 2012b), where the assignment of all backbone, C^{β} and H^{β} chemical shifts of α-synuclein is present.

The indicator of the reliability of the result of *TSAR* program are the lengths of assigned chains of cross-sections (Table S1): the longer they are, the more reliable the result is. In all the data sets some incorrect assignments were observed, generally for short chains (up to three cross-sections). Another problem of *TSAR* operation was observed in a case of chemical shift degeneration of pairs of residues, as in two fragments of α-synuclein: 34Lys-35Glu-36Gly and 45Lys-46Glu-47Gly. In this case the same chain of three cross-sections should be assigned to both of these sequence fragments. However, in the current version, *TSAR* cannot assign one cross-section to two different residues, thus leaving both above fragments unassigned. Importantly, if the assignment is not complete, the program provides information facilitating manual accomplishment of the task.

Conclusions

The three novel exclusively heteronuclear ^{13}C direct-detected 5D NMR experiments proposed here can be acquired in a short time preserving excellent resolution in all indirect dimensions, thanks to the use of NUS, H-flip

(when appropriate) and SMFT processing of the data. A fast and reliable automatic assignment of chemical shifts can be accomplished with the *TSAR* program. The strategy discussed here, from the acquisition of the data to the assignment of the frequencies, can be applied to investigate IDPs of increasing size and complexity.

Supplementary Material

5D ($H^{N\text{-}flip}N$)CONCACON, 5D (HCA)CONCACON, 5D (H)CACON(CA)CON and 3D (H)CACON pulse sequences; experimental parameters for the recorded experiments. Explanation of the *TSAR* input data, example of an input file with definitions of all experiments used for assignment in the article.

Acknowledgments This work was supported in part by the EC 7th Framework program BioNMR (contract 261863), by the EC Marie Curie ITN program IDPbyNMR (contract 264257) and by grant number IP2012 062772, funded by Polish Ministry of Science and Higher Education for years 2013–2014. AZK thanks the Foundation for Polish Science for support with the START and the POMOST programs.

References

Bax A, Grzesiek S (1993) Methodological advances in protein NMR. Acc Chem Res 26:131–138

Bermel W, Bertini I, Felli IC, Kümmerle R, Pierattelli R (2006a) Novel ^{13}C direct detection experiments, including extension to the third dimension, to perform the complete assignment of proteins. J Magn Reson 178:56–64

Bermel W, Bertini I, Felli IC, Lee Y-M, Luchinat C, Pierattelli R (2006b) Protonless NMR experiments for sequence-specific assignment of backbone nuclei in unfolded proteins. J Am Chem Soc 128:3918–3919

Bermel W, Bertini I, Felli IC, Piccioli M, Pierattelli R (2006c) ^{13}C-detected protonless NMR spectroscopy of proteins in solution. Progr NMR Spectrosc 48:25–45

Bermel W, Felli IC, Kümmerle R, Pierattelli R (2008) ^{13}C direct-detection biomolecular NMR. Concepts Magn Reson 32A:183–200

Bermel W, Bertini I, Csizmok V, Felli IC, Pierattelli R, Tompa P (2009a) H-start for exclusively heteronuclear NMR spectroscopy: the case of intrinsically disordered proteins. J Magn Reson 198:275–281

Bermel W, Bertini I, Felli IC, Pierattelli R (2009b) Speeding up 13C direct detection Biomolecular NMR experiments. J Am Chem Soc 131:15339–15345

Bermel W, Bertini I, Chill JH, Felli IC, Kumar VMV, Haba N, Pierattelli R (2012a) Aminoacid-types selective ^{13}C detected amino-acid-selective NMR experiments for the study of intrinsically disordered proteins (IDPs). ChemBioChem 13:2425–2432

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

Bermel W, Bruix M, Felli IC, Kumar VMV, Pierattelli R, Serrano S (2013) Improving the chemical shift dispersion of multidimensional NMR spectra of intrinsically disordered proteins. J Biomol NMR 55:231–237

Böhlen J-M, Bodenhausen G (1993) Experimental aspects of chirp NMR spectroscopy. J Magn Reson Ser A 102:293–301

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

Delaglio F, Grzesiek S, Vuister G, Zhu G, Pfeifer J, Bax A (1995) NMRPipe: a multidimensional spectral processing system based on UNIX Pipes. J Biomol NMR 6:277–293

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

Dyson HJ, Wright PE (2001) Nuclear magnetic resonance methods for the elucidation of structure and dynamics in disordered states. Methods Enzymol 339:258–271

Dyson HJ, Wright PE (2005) Intrinsically unstructured proteins and their functions. Nat Rev Mol Cell Biol 6:197–208

Emsley L, Bodenhausen G (1992) Optimization of shaped selective pulses for NMR using a quaternion description of their overall propagators. J Magn Reson 97:135–148

Gil S, Hosek 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. doi:10.1002/anie.201304272

Goddard TD, Kneller DG (2000) SPARKY 3. University of California, San Francisco

Haba NY, Gross R, Novacek J, Shaked H, Zidek L, Barda-Saad M, Chill JH (2013) NMR determines transient structure and dynamics in the disordered C-terminal domain of WASp interacting protein. Biophysical J 105:481–493

Hiller S, Fiorito F, Wüthrich K, Wider G (2005) Automated projection spectroscopy (APSY). Proc Natl Acad Sci USA 102:10876–10881

Hsu ST, Bertoncini CW, Dobson CM (2009) Use of protonless NMR spectroscopy to alleviate the loss of information resulting from exchange-broadening. J Am Chem Soc 131:7222–7223

Huang C, Ren G, Zhou H, Wang C (2005) A new method for purification of recombinant human alpha-synuclein in Escherichia coli. Protein Expr Purif 42:173–177

Kazimierczuk K, Zawadzka A, Kozminski W, Zhukov I (2006) Random sampling of evolution time space and Fourier transform processing. J Biomol NMR 36:157–168

Kazimierczuk K, Zawadzka A, Kozminski W (2008) Optimization of random time domain sampling in multidimensional NMR. J Magn Reson 192:123–130

Kazimierczuk K, Zawadzka A, Kozminski W (2009) Narrow peaks and high dimensionalities: exploiting the advantages of random sampling. J Magn Reson 197:219–228

Kazimierczuk K, Stanek J, Zawadzka-Kazimierczuk A, Kozminski W (2010) Random sampling in multidimensional NMR spectroscopy. Prog NMR Spectrosc 57:420–434

Knoblich K, Whittaker S, Ludwig C, Michiels P, Jiang T, Schafflhausen B, Günther U (2009) Backbone assignment of the N-terminal polyomavirus large T antigen. Biomol NMR Assign 3:119–123

Kumar D, Hosur RV (2011) hNCOcanH pulse sequence and a robust protocol for rapid and unambiguous assignment of backbone ((1) H(N), (15) N and (13) C') resonances in (15) N/(13) C-labeled proteins. Magn Reson Chem 49:575–583

Mantylahti S, Hellman M, Permi P (2011) Extension of the HA-detection based approach: (HCA)CON(CA)H and (HCA)NCO(-CA)H experiments for the main-chain assignment of intrinsically disordered proteins. J Biomol NMR 49:99–109

Mittag T, Forman-Kay J (2007) Atomic-level characterization of disordered protein ensembles. Curr Opin Struct Biol 17:3–14

Narayanan RL, Duerr HN, Bilbow S, Biernat J, Mendelkow E, Zweckstetter M (2010) Automatic assignment of the intrinsically disordered protein Tau with 441-residues. J Am Chem Soc 132:11906–11907

Novacek J, Zawadzka-Kazimierczuk A, Papoušková V, Zidek L, Sanderová H, Krasny L, Kozminski W, Sklenar V (2011) 5D 13C-detected experiments for backbone assignment of unstructured proteins with a very low signal dispersion. J Biomol NMR 50:1–11

Novacek J, Haba NY, Chill JH, Zidek L, Sklenar V (2012) 4D Non-uniformly sampled HCBCACON and (1) J(NC (α))-selective HCBCANCO experiments for the sequential assignment and chemical shift analysis of intrinsically disordered proteins. J Biomol NMR 53:139–148

Novacek J, Janda L, Dopitova R, Zidek L, Sklenar V (2013) Efficient protocol for backbone and side-chain assignments of large, intrinsically disordered proteins: transient secondary structure analysis of 49.2 kDa microtubule associated protein 2c. J Biomol NMR 56:291–301

O'Hare B, Benesi AJ, Showalter SA (2009) Incorporating 1H chemical shift determination into 13C-direct detected spectroscopy of intrinsically disordered proteins in solution. J Magn Reson 200:354–358

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

Pantoja-Uceda D, Santoro J (2012) New amino acid residue type identification experiments valid for protonated and deuterated proteins. J Biomol NMR 54:145–153

Pantoja-Uceda D, Santoro J (2013) Direct correlation of consecutive C'–N groups in proteins: a method for the assignment of intrinsically disordered proteins. J Biomol NMR 57:57–63

Pérez Y, Gairi M, Pons M, Bernadò P (2009) Structural characterization of the natively unfolded N-terminal domain of human c-Src kinase: insights into the role of phosphorylation of the unique domain. J Mol Biol 391:136–148

Radivojac P, Iakoucheva LM, Oldfield CJ, Obradovic Z, Uversky VN, Dunker AK (2007) Intrinsic disorder and functional proteomics. Biophys J 92:1439–1456

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

Shaka AJ, Keeler J, Freeman R (1983) Evaluation of a new broadband decoupling sequence: WALTZ-16. J Magn Reson 53:313–340

Shaka AJ, Barker PB, Freeman R (1985) Computer-optimized decoupling scheme for wideband applications and low-level operation. J Magn Reson 64:547–552

Solyom Z, Schwarten M, Geist L, Konrat R, Willbold D, Brutscher B (2013) BEST-TROSY experiments for time-efficient sequential resonance assignment of large disordered proteins. J Biomol NMR 55:311–321

Thakur A, Chandra K, Dubey A, D'Silva P, Atreya HS (2013) Rapid characterization of hydrogen exchange in proteins. Angew Chem 52:2440–2443

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

Tompa P (2009) Structure and function of intrinsically disordered proteins. CRC Press, Boca Raton

Ulrich EL, Akutsu H, Doreleijers JF, Harano Y, Ioannidis YE, Lin J, Livny M, Mading S, Maziuk D, Miller Z, Nakatani E, Shulte CF, Tolmie DE, Kent Wenger R, Yao H, Markley JL (2007) BioMagResBank. Nucleic Acids Res 36:D402–D408

Uversky VN, Gillespie JR, Fink AL (2000) Why are "natively unfolded" proteins unstructured under physiologic conditions? Proteins Struct Funct Genet 41:415–427

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

Zawadzka-Kazimierczuk A, Kozminski W, Billeter M (2012) TSAR: a program for automatic resonance assignment using 2D cross-sections of high dimensionality, high-resolution spectra. J Biomol NMR 54:81–95

High-dimensionality ^{13}C direct-detected NMR experiments for the automatic assignment of intrinsically disordered proteins

Wolfgang Bermel[1], Isabella C. Felli[2,3]✉, Leonardo Gonnelli[2], Wiktor Koźmiński[4], Alessandro Piai[2], Roberta Pierattelli[2,3]✉, Anna Zawadzka-Kazimierczuk[4]

[1] Bruker BioSpin GmbH, Silberstreifen, 76287 Rheinstetten, Germany

[2] CERM, University of Florence, Via Luigi Sacconi 6, 50019 Sesto Fiorentino, Florence, Italy

[3] Department of Chemistry ''Ugo Schiff'', University of Florence, 50019 Sesto Fiorentino, Italy

[4] Faculty of Chemistry, Biological and Chemical Research Centre, University of Warsaw, Żwirki i Wigury 101, 02-089 Warsaw, Poland

SUPPLEMENTARY MATERIAL

NMR experiments

In this section the pulse schemes for the following NMR experiments are described: 5D ($H^{N\text{-}flip}N$)CONCACON, 5D (HCA)CONCACON, 5D (H)CACON(CA)CON and 3D (H)CACON.

The arrows indicate the switching of the ^{13}C carrier frequency. Narrow and wide symbols stand for 90° and 180° pulses, respectively. The pulses were applied along the *x* axis unless noted differently. The rectangles represent ^{15}N and ^{1}H non-selective pulses, the round shapes represent band-selective ^{13}C pulses. The grey round shape represents a ^{13}C adiabatic pulse, the round shape marked by an asterisk represents a ^{13}C 180° pulse applied both on C^{α} and C^{β}, while the round shape marked by two asterisk represents a ^{13}C 180° pulse applied selectively on C^{α}. The two different variants necessary for the IPAP virtual carbonyl decoupling approach are shown. The line denoted with PFG stands for pulsed field gradients applied along the z-axis. The phase cycle, the position of the carriers and the duration of the delays for these experiments are reported case-by-case in the captions of the figures describing the pulse sequences.

The experimental parameters of each experiment, including the NUS sampling parameters, are indicated in the final part of the captions describing the pulse sequences.

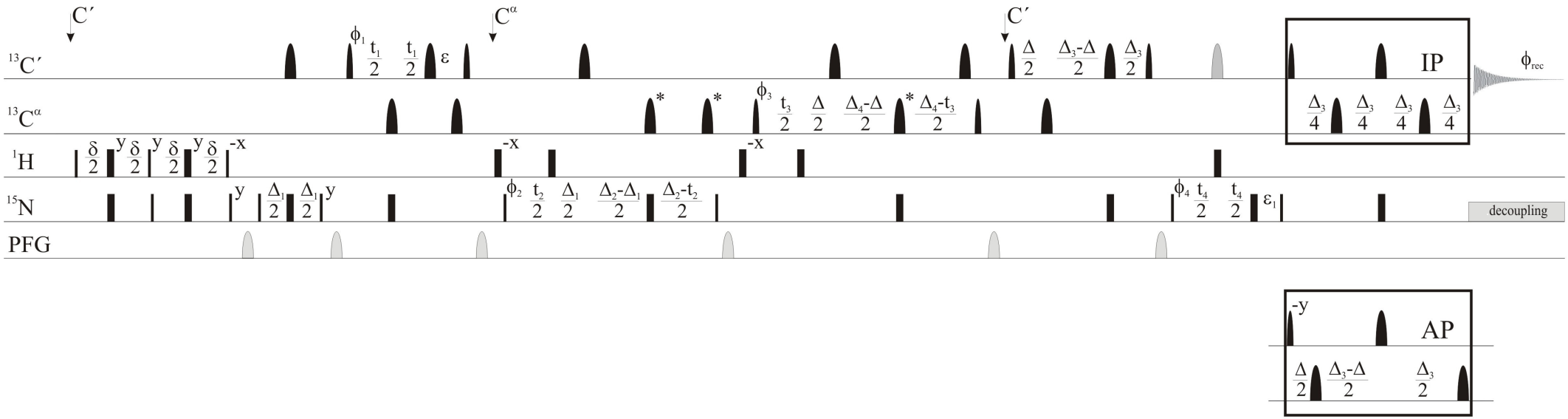


Figure S1. Pulse sequence of the 5D ($H^{N\text{-}flip}N$)CONCACON experiment.

The following phase cycling was employed: ϕ_2 = x, -x; ϕ_3 = 2(x), 2(-x); ϕ_4 = 4(x), 4(-x); ϕ_{rec} = x, 2(-x), x, -x, 2(x), -x. The lengths of the delays were: δ = 5.4 ms, Δ = 9.0 ms, Δ_1 = 24.0 ms, Δ_2 = 24.8 ms, Δ_3 = 32.0 ms, Δ_4 = 28.0 ms, $\varepsilon = t_1(0)+p_{12}$ and $\varepsilon_1 = t_4(0)+p_8$, where p_{12} is the duration of the band-selective ^{13}C' pulse in the middle of the ^{13}C evolution period and p_8 is the duration of the adiabatic pulse on ^{13}C in the middle of the second ^{15}N evolution period. The constant time mode for the evolution of chemical shift was used in the first ^{15}N and $^{13}C^{\alpha}$ evolution periods, whereas real time evolution was adopted for ^{13}C' and second ^{15}N evolution periods. Anyway, if necessary, the evolution of the chemical shift can be switched to the semi-constant time mode to further increase the resolution in the resulting spectrum. Quadrature detection in the indirect dimensions was achieved by incrementing phases ϕ_1, ϕ_2, ϕ_3 and ϕ_4 in a States manner. The lower dimensionality 4D analog ($H^{N\text{-}flip}N$)CON(CA)CON experiment can be performed setting t_3 to 0. In this case, quadrature detection in the indirect dimensions can be achieved by incrementing phases ϕ_1, ϕ_2, and ϕ_4 in a States manner.

Experimental parameters used. The 5D ($H^{N\text{-}flip}N$)CONCACON experiment was recorded in 2 days and 10 hours with 8 scans per increment, an inter-scan delay of 0.5 s, spectral widths of 2000 (^{13}C') × 2500 (^{15}N) × 4400 ($^{13}C^{\alpha}$) × 2500 (^{15}N) × 8800 (aq) Hz with 512 complex points in the acquisition dimension and 994 hypercomplex points randomly distributed in the indirectly detected dimensions (data points density of 0.0032%). The maximum evolution times in the indirectly detected dimensions were set to 30.0 ms for the ^{13}C' dimension (61 complex points), 23.6 ms for the first ^{15}N dimension (60 complex points), 27.3 ms for the $^{13}C^{\alpha}$ dimension (121 complex points) and to 28.0 ms for the second ^{15}N dimension (71 complex points).

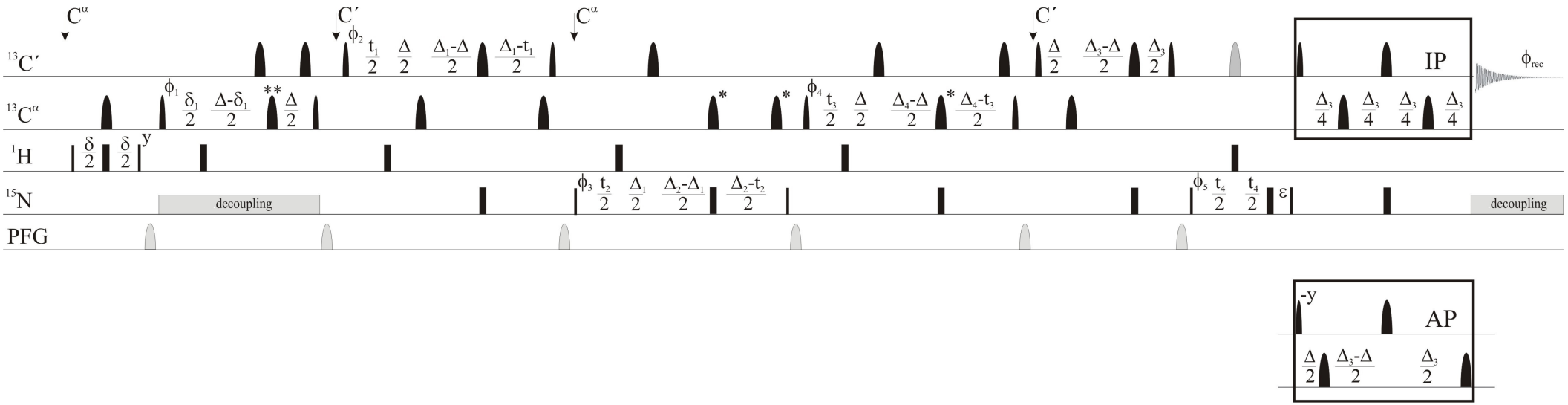


Figure S2. Pulse sequence of the 5D (HCA)CONCACON experiment.

The following phase cycling was employed: ϕ_1 = 2(x), 2(-x); ϕ_2 = 8(x), 8(-x); ϕ_3 = x, -x; ϕ_5 = 4(x), 4(-x); ϕ_{rec} = x, 2(-x), x, -x, 2(x), 2(-x), 2(x), -x, x, 2(-x), x. The lengths of the delays were: δ = 3.6 ms, δ_1 = 2.2 ms, Δ = 9.0 ms, Δ_1 = 24.0 ms, Δ_2 = 24.8 ms, Δ_3 = 32.0 ms, Δ_4 = 28.0 ms and $\varepsilon = t_4(0)+p_8$, where p_8 is the duration of the adiabatic pulse on ^{13}C in the middle of the second ^{15}N evolution period. The constant time mode for the evolution of chemical shift was used in all the evolution periods, except for the second ^{15}N evolution period for which real time evolution was adopted. Anyway, if necessary, the evolution of the chemical shift can be switched to the semi-constant time mode to further increase the resolution in the resulting spectrum. Quadrature detection in the indirect dimensions was achieved by incrementing phases ϕ_2, ϕ_3, ϕ_4 and ϕ_5 in a States manner. The lower dimensionality 4D analog (HCA)CON(CA)CON experiment can be performed setting t_3 to 0. In this case, quadrature detection in the indirect dimensions can be achieved by incrementing phases ϕ_2, ϕ_3, and ϕ_5 in a States manner.

Experimental parameters used. The 5D (HCA)CONCACON experiment was recorded in 2 days and 23 hours with 4 scans per increment, an inter-scan delay of 1.2 s, spectral widths of 2000 (^{13}C') × 2500 (^{15}N) × 4400 (^{13}C$^{\alpha}$) × 2500 (^{15}N) × 8800 (aq) Hz with 512 complex points in the acquisition dimension and 1305 hypercomplex points randomly distributed in the indirectly detected dimensions (data points density of 0.0057%). The maximum evolution times in the indirectly detected dimensions were set to 23.5 ms for the ^{13}C' dimension (48 complex points), 23.6 ms for the first ^{15}N dimension (60 complex points), 25.0 ms for the ^{13}C$^{\alpha}$ dimension (111 complex points) and to 28.6 ms for the second ^{15}N dimension (72 complex points).

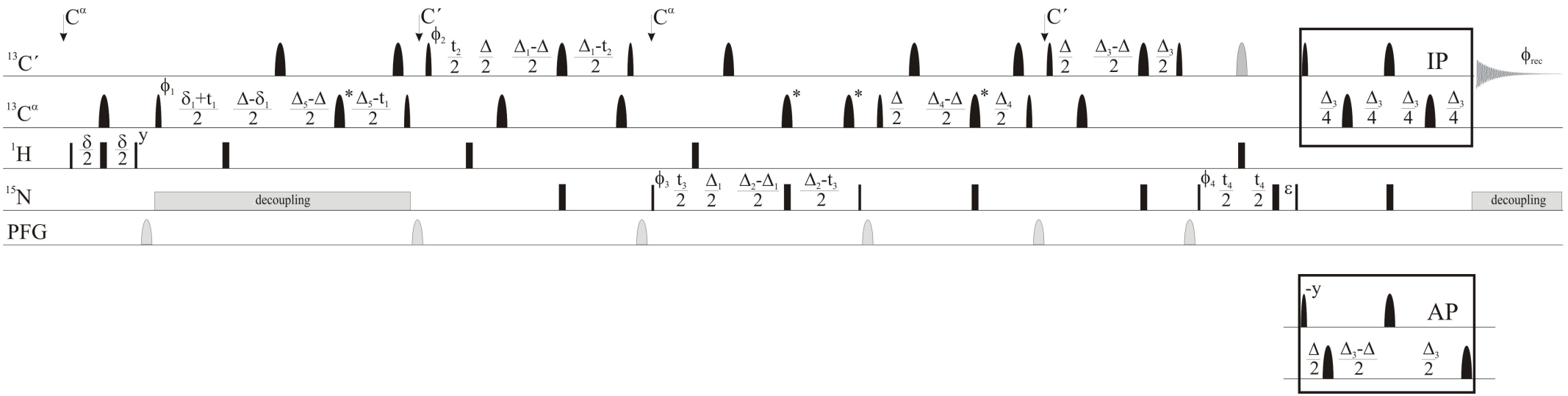


Figure S3. Pulse sequence of the 5D (H)CACON(CA)CON experiment.

The following phase cycling was employed: ϕ_1 = 2(x), 2(-x); ϕ_2 = 8(x), 8(-x); ϕ_3 = x, -x; ϕ_4 = 4(x), 4(-x); ϕ_{rec} = x, 2(-x), x, -x, 2(x), 2(-x), 2(x), -x, x, 2(-x), x. The lengths of the delays were: δ = 3.6 ms, δ_1 = 2.2 ms, Δ = 9.0 ms, Δ_1 = 24.0 ms, Δ_2 = 24.8 ms, Δ_3 = 32.0 ms, Δ_4 = 28.0 ms, Δ_5 = 26.6 ms and $\varepsilon = t_4(0)+p_8$, where p_8 is the duration of the adiabatic pulse on ^{13}C in the middle of the second ^{15}N evolution period. The constant time mode for the evolution of chemical shift was used in all the evolution periods, except for the second ^{15}N evolution period for which real time evolution was adopted. Anyway, if necessary, the evolution of the chemical shift can be switched to the semi-constant time mode to further increase the resolution in the resulting spectrum. Quadrature detection in the indirect dimensions was achieved by incrementing phases ϕ_1, ϕ_2, ϕ_3 and ϕ_4 in a States manner. The lower dimensionality 4D analog (HCA)CON(CA)CON experiment can be performed setting t_1 to 0. In this case, quadrature detection in the indirect dimensions can be achieved by incrementing phases ϕ_2, ϕ_3, and ϕ_4 in a States manner.

Experimental parameters used. The 5D (H)CACON(CA)CON experiment was recorded in 2 days and 6 hours with 4 scans per increment, an inter-scan delay of 1.2 s, spectral widths of 2000 (^{13}C') × 2500 (^{15}N) × 4400 (^{13}C$^{\alpha}$) × 2500 (^{15}N) × 8800 (aq) Hz with 512 complex points in the acquisition dimension and 994 hypercomplex points randomly distributed in the indirectly detected dimensions (data points density of 0.0044%). The maximum evolution times in the indirectly detected dimensions were set to 23.5 ms for the ^{13}C' dimension (48 complex points), 23.6 ms for the first ^{15}N dimension (60 complex points), 25.0 ms for the ^{13}C$^{\alpha}$ dimension (111 complex points) and to 28.0 ms for the second ^{15}N dimension (71 complex points).

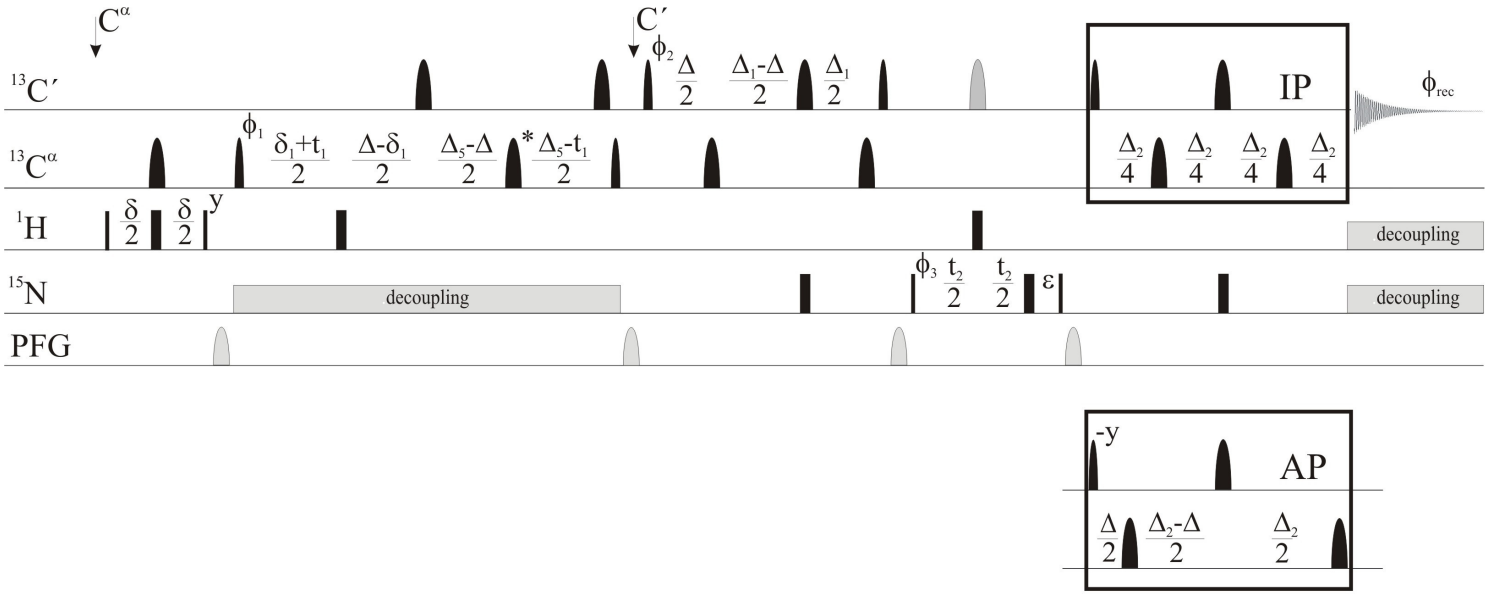


Figure S4. Pulse sequence of the 3D (H)CACON experiment.

The following phase cycling was employed: ϕ_1 = x, -x; ϕ_2 = 4(x), 8(-x); ϕ_3 = 2(x), 2(-x); ϕ_{rec} = x, 2(-x), x, -x, 2(x), -x. The lengths of the delays were: δ = 3.6 ms, δ_1 = 2.2 ms, Δ = 9.0 ms, Δ_1 = 24.0 ms, Δ_5 = 26.6 ms and $\varepsilon = t_2(0)+p_8$, where p_8 is the duration of the adiabatic pulse on ^{13}C in the middle of the ^{15}N evolution period. The constant time mode for the evolution of chemical shift was used in the $^{13}C^{\alpha}$ evolution period, whereas real time evolution was adopted for ^{15}N evolution period. Anyway, if necessary, the evolution of the chemical shift can be switched to the semi-constant time mode to further increase the resolution in the resulting spectrum. Quadrature detection in the indirect dimensions was achieved by incrementing phases ϕ_1 and ϕ_3 in a States manner.

Experimental parameters used. The 3D (H)CACON experiment was recorded in 12 hours and 40 minutes with 4 scans per increment, an inter-scan delay of 1.2 s, spectral widths of 4400 ($^{13}C^{\alpha}$) × 2500 (^{15}N) × 8800 (aq) Hz with 512 complex points in the acquisition dimension and 1000 hypercomplex points randomly distributed in the indirectly detected dimensions (data points density of 13.3%). The maximum evolution times in the indirectly detected dimensions were set to 22.5 ms for the $^{13}C^{\alpha}$ dimension (100 complex points) and to 29.6 ms for the ^{15}N dimension (75 complex points).

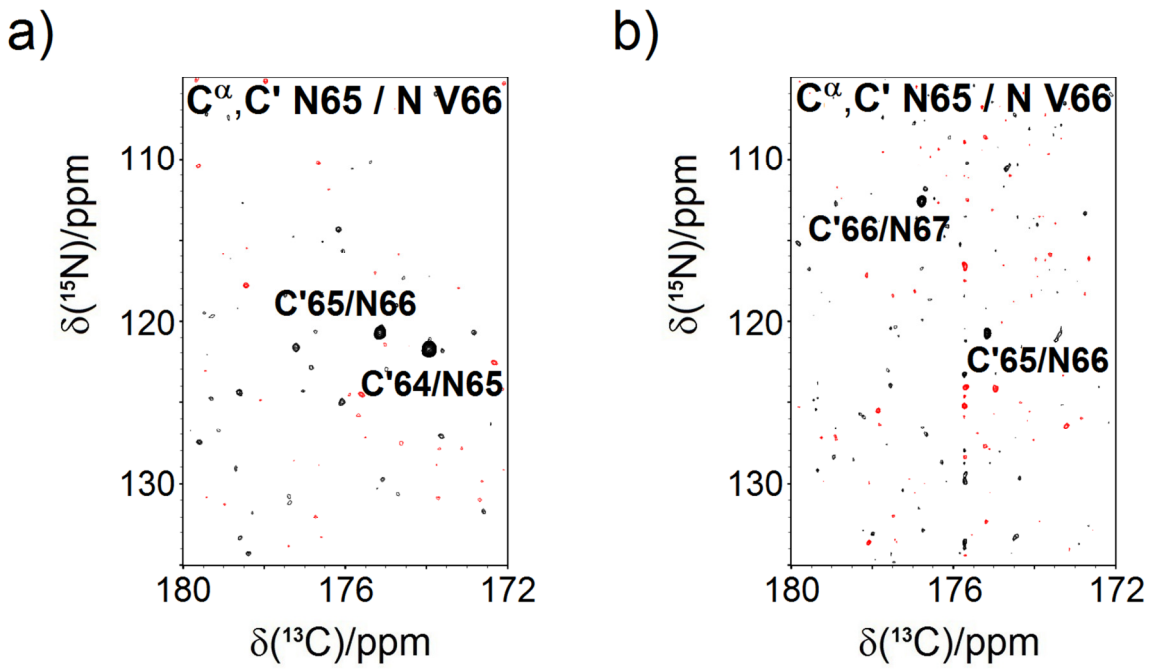


Figure S5. Examples of 2D cross-sections of high-dimensional spectra with a) direct dimension frequencies fixed (5D (HCA)CONCACON spectrum, F1-F2 plane) or b) not fixed (5D (H)CACON(CA)CON spectrum, F5-F4 plane) for SMFT procedure.

In the top of each panel, the C^{α}_{i}-C'_{i}-N_{i+1} frequencies that have been fixed to extract the particular 2D plane are reported. The threshold was set at the level allowing observation of sampling artifacts. In the case of the cross-section not containing direct dimension (panel a), the artifacts are evenly distributed on the whole plane, while in the case of cross-section containing direct dimension (panel b), an artifact ridge along the direct dimension is observed. As clearly visible, such artifacts are not dramatic and do not limit the effectiveness of the experiment. However, if the peak is exactly located on a ridge, the peak picking procedure may become more difficult.

Input for the *TSAR* program

The *TSAR* program accepts almost any type of experiment, but it has to be properly defined. In Figure S6, there is an example of input.txt file, in which all the experiments mentioned in the article are defined. Of course, in a real case not all of them have to be used, but one can choose an optimal set for a particular sample. The file starts with a definition of the basis spectrum, containing its name, dimensionality and peak definition. Thereafter the definitions of all higher-dimensional experiments follow. They start with peak-types definitions with the names and relative positions of the nuclei of the two dimensions that are not fixed during the SMFT procedure (i.e. the dimensions which appear on the cross-sections). In the same line also the sign of peak amplitude is defined (p – positive, n – negative, u -unknown). If certain sign changes under the presence of some amino acid (usually glycine) – it is indicated after a colon. The chemical shift tolerances have to be set for each not fixed dimension of high-dimensional spectrum, as inverse of maximum evolution time in the corresponding dimension of a given experiment, expressed in ppm.

For the details see [Zawadzka-Kazimierczuk, A., Koźmiński, W. and Billeter, M. (2012) TSAR: a program for automatic resonance assignment using 2D cross-sections of high dimensionality, high-resolution spectra. *J. Biomol. NMR*, **54**, 81-95] or the manual for the *TSAR* program, available at http://nmr.cent3.uw.edu.pl/software.

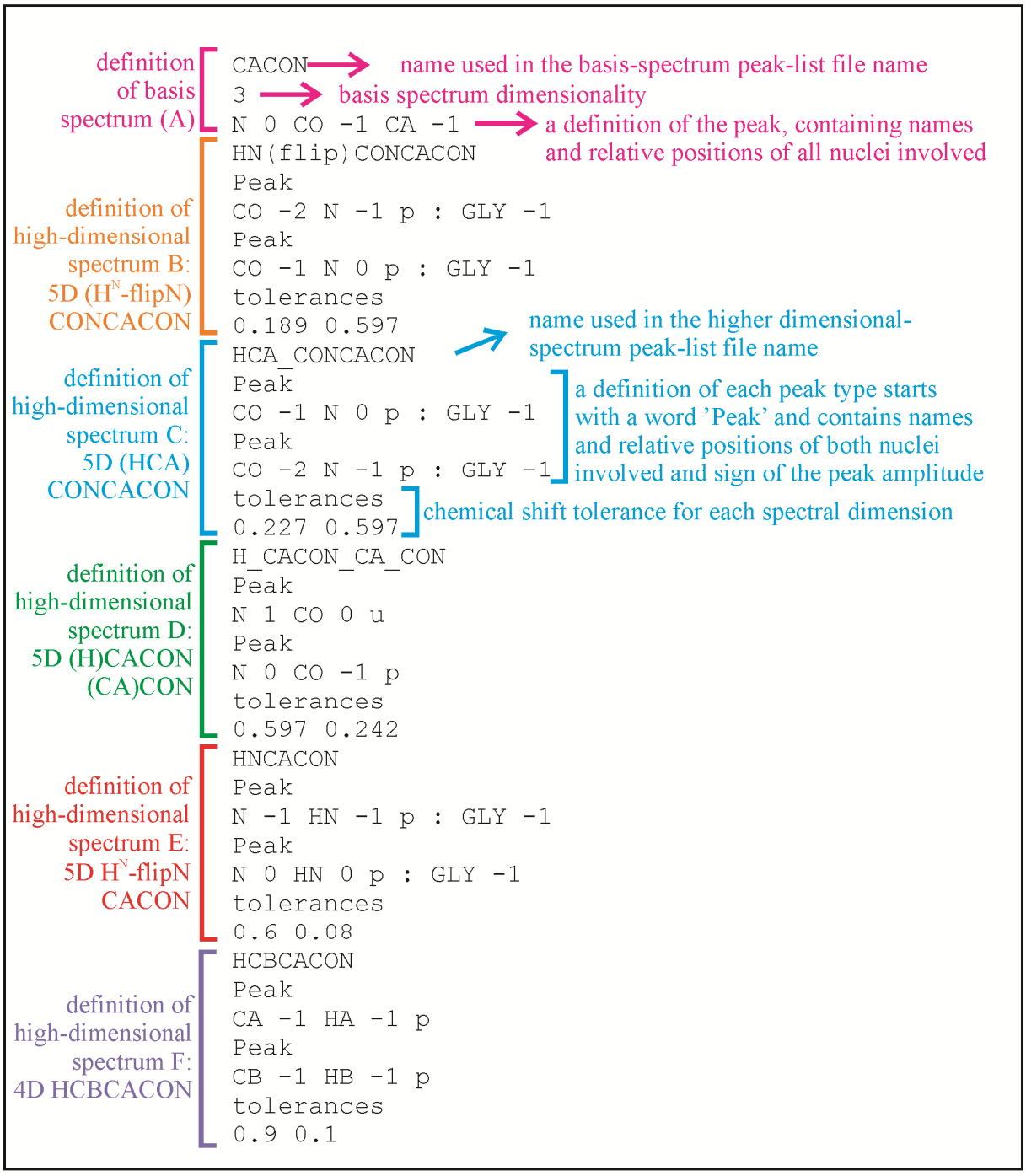


Figure S6. An extract from the input.txt file containing the description of all the experiments used in the current article.

The file contains the description of all the experiments used in the article. Should the user choose not all of the experiments, the file can be modified by cutting out the fragments corresponding to unused experiment. The colored signs are not a part of the file, they are included to explain the file's structure. The experiments' symbols (A-F) were defined in Table 1 in the article.

Table S1. Cross-section spin systems (CSSS) chains lengths.

Data set number	Experiments [a]	Number of CSSS chains formed by TSAR		
		long (≥8 CSSS)	medium (3-7 CSSS)	short (1-2 CSSS)
1	A+B+D	6	4	18
2	A+C+D	7	3	15
3	A+C+D+E	8	3	13
4	A+B+D+F	6	4	18
5	A+D+E+F	6	9	24

[a] See main text for the explanation of the labeling

Article 4. “CON-CON” assignment strategy for highly flexible intrinsically disordered proteins

Alessandro Piai[1], Tomáš Hošek[1], Leonardo Gonnelli[1], Anna Zawadzka-Kazimierczuk[2], Wiktor Koźmiński[2], Bernhard Brutscher[3], Wolfgang Bermel[4], Roberta Pierattelli[1], Isabella C. Felli[1]

[1] CERM and Department of Chemistry "Ugo Schiff", University of Florence, 50019, Via Luigi Sacconi 6, Sesto Fiorentino, Florence, Italy.

[2] Faculty of Chemistry, Biological and Chemical Research Centre, University of Warsaw, Żwirki i Wigury 101, 02-089 Warsaw, Poland.

[3] Institut de Biologie Structurale, Université Grenoble 1, CNRS, CEA, 71 avenue des Martyrs, 38044, Grenoble Cedex 9, France.

[4] Bruker BioSpin GmbH, Silberstreifen, 76287 Rheinstetten, Germany.

Published in 2014 in J.Biomol.NMR, 60, 209-218

J Biomol NMR (2014) 60:209–218
DOI 10.1007/s10858-014-9867-6

ARTICLE

"CON-CON" assignment strategy for highly flexible intrinsically disordered proteins

Alessandro Piai · Tomáš Hošek · Leonardo Gonnelli · Anna Zawadzka-Kazimierczuk · Wiktor Koźmiński · Bernhard Brutscher · Wolfgang Bermel · Roberta Pierattelli · Isabella C. Felli

Received: 27 August 2014 / Accepted: 10 October 2014 / Published online: 19 October 2014

Abstract Intrinsically disordered proteins (IDPs) are a class of highly flexible proteins whose characterization by NMR spectroscopy is complicated by severe spectral overlaps. The development of experiments designed to facilitate the sequence-specific assignment procedure is thus very important to improve the tools for the characterization of IDPs and thus to be able to focus on IDPs of increasing size and complexity. Here, we present and describe the implementation of a set of novel ^{1}H-detected 5D experiments, (HACA)CON(CACO)NCO(CA)HA, BT-(H)NCO(CAN)CONNH and BT-HN(COCAN)CONNH, optimized for the study of highly flexible IDPs that exploit the best resolved correlations, those involving the carbonyl and nitrogen nuclei of neighboring amino acids, to achieve sequence-specific resonance assignment. Together with the analogous recently proposed pulse schemes based on ^{13}C detection, they form a complete set of experiments for sequence-specific assignment of highly flexible IDPs. Depending on the particular sample conditions (concentration, lifetime, pH, temperature, etc.), these experiments present certain advantages and disadvantages that will be discussed. Needless to say, that the availability of a variety of complementary experiments will be important for accurate determination of resonance frequencies in complex IDPs.

Keywords Intrinsically disordered proteins · ^{13}C detection · $^{1}H^{N}$ detection · $^{1}H^{\alpha}$ detection · NUS · Multidimensional NMR experiment · BEST-TROSY · Backbone assignment

Introduction

The role of intrinsic disorder is essential in many cellular mechanisms (Tompa 2009, 2012; Uversky 2013a, b), which range from gene regulation to signaling processes. Intrinsically disordered proteins (IDPs) or intrinsically disordered protein regions (IDPRs) play key roles in many of these cellular events (Uversky and Dunker 2013), thanks to their local flexibility and conformational freedom, which allow them to modulate such processes. These evidences are pushing the scientific community towards a reconsideration of the structure/function paradigm, taking into account that many disordered proteins are able to carry out specific functions despite the absence of secondary and tertiary structures (Wright and Dyson 1999; Uversky et al. 2000; Dunker et al. 2001).

With X-ray crystallography being incapable to describe conformational disorder, nuclear magnetic resonance (NMR) spectroscopy becomes the most qualified tool of

Electronic supplementary material The online version of this article (doi:10.1007/s10858-014-9867-6) contains supplementary material, which is available to authorized users.

A. Piai · T. Hošek · L. Gonnelli · R. Pierattelli (✉) · I. C. Felli (✉)
CERM and Department of Chemistry, University of Florence, Via Luigi Sacconi 6, 50019 Sesto Fiorentino, Florence, Italy
e-mail: felli@cerm.unifi.it

A. Zawadzka-Kazimierczuk · W. Koźmiński
Faculty of Chemistry, Biological and Chemical Research Centre, University of Warsaw, Żwirki i Wigury 101, 02-089 Warsaw, Poland

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

W. Bermel
Bruker BioSpin GmbH, Silberstreifen, 76287 Rheinstetten, Germany

investigation, thanks also to its unique ability to provide information at atomic resolution (Konrat 2014; Felli and Pierattelli 2014; Nováček et al. 2014). However, in respect to what is found for well-folded proteins, the peculiar properties of IDPs introduce additional challenges that need to be overcome to prevent drastic reduction in the quality of the spectra. First, the conformational freedom experienced by IDPs leads to severe collapse of chemical shifts, since their values are mostly determined by the amino acid chemical structure and by the primary sequence while contributions deriving from the three-dimensional structure are missing. Secondly, the high rates of solvent exchange processes endured by amide protons, that in absence of a stable fold are largely solvent exposed, can dramatically compromise the information content of multidimensional spectra based on $^{1}H^{N}$ detection, since at physiological temperature and pH many signals may be broadened even beyond detection. Finally, IDPs are often characterized by repetitive amino acid sequences in which prolines are usually abundant, thus complicating the resonance assignment procedure. To overcome these limitations as much as possible, several IDP-dedicated NMR experiments have been recently proposed (Panchal et al. 2001; Hiller et al. 2007; Mäntylahti et al. 2010; Motáčkova et al. 2010; Narayanan et al. 2010; Mäntylahti et al. 2011; Nováček et al. 2011; Favier and Brutscher 2011; Bermel et al. 2012; Zawadzka-Kazimierczuk et al. 2012b; Pantoja-Uceda and Santoro 2013a, b; Nováček et al. 2013; Bermel et al. 2013b; Kazimierczuk et al. 2013; Solyom et al. 2013; Pantoja-Uceda and Santoro 2014); however, NMR experiments able to deal with all these complications still need to be fully developed.

To cope with the low chemical shift dispersion of IDPs/IDPRs, high spectral resolution is mandatory. Therefore, increasing the number of dimensions of NMR experiments provides a unique tool to enhance the resolution, provided non-uniform sampling (NUS) (Kazimierczuk et al. 2010a, 2012) and, when possible, longitudinal relaxation enhancement (LRE) (Pervushin et al. 2002; Schanda and Brutscher 2005; Schanda et al. 2005; Deschamps and Campbell 2006; Felli and Brutscher 2009; Solyom et al. 2013; Gil et al. 2013) are used to achieve high resolution also in indirect dimensions without exponentially increasing the duration of NMR experiments. In fact, since the number of hypercomplex points correlates with the number of cross-peaks, by adding another dimension the required experimental time is usually doubled. The exploitation of the heteronuclei and in particular the chemical shift labeling of ^{15}N and ^{13}C frequencies provide great benefits to reduce spectral crowding (Ikura et al. 1991; Sattler et al. 1999; Dyson and Wright 2004; Mittag and Forman-Kay 2007). In our previous paper (Bermel et al. 2013b), we proposed three novel 5D ^{13}C direct-detected NMR experiments in which two pairs of ^{15}N and ^{13}C frequencies are correlated in order to provide robust information for fast and reliable backbone assignment. ^{13}C detection was chosen to exploit at best the greater chemical shift dispersion and the reduced sensitivity to exchange broadening of carbons in respect to protons (Csizmok et al. 2008; Pérez et al. 2009; Hsu et al. 2009; Knoblich et al. 2009; O'Hare et al. 2009; Skora et al. 2010; Felli et al. 2013). Here, we extend the same experimental strategy also to proton detection, combining the advantages given by heteronuclei in terms of spectral resolution with the direct acquisition of protons, which are characterized by about a 4-times larger gyromagnetic ratio resulting in higher intrinsic sensitivity. The new experiments should not be considered a replacement of the ^{13}C detected ones, but valid alternatives that can be adopted when appropriate, as discussed later. Indeed, we strongly believe that ^{1}H and ^{13}C detection provide complementary tools that allow to extend the size and complexity of IDPs/IDPRs that can be investigated by NMR. In fact, the chemical shift assignment is a mandatory step towards the structural and dynamic characterization of proteins, as well as of their interactions. The usefulness of the new experiments is demonstrated on a paradigmatic IDP, human α-synuclein (14,460 Da).

Materials and methods

NMR samples

A sample of 0.9 mM uniformly ^{13}C, ^{15}N labeled human α-synuclein in 20 mM phosphate buffer at pH 6.5 was prepared as previously described (Huang et al. 2005). EDTA and NaCl were added to reach the final concentration of 0.5 and 200 mM, respectively, and 10 % D_2O was added for the lock (Sample *A*). An identical sample, but in fully deuterated solvent (99.99 % D_2O), was also prepared (Sample *B*). The H^{N} and H^{α}-based experiments were acquired on sample *A* and *B*, respectively, both in 3 mm NMR sample tubes to reduce the detrimental effects of high salt concentration. The temperature was set to 285.5 K for Sample *A*, to reduce the rate of chemical exchange of amide protons, and to 310.0 K for Sample *B*, to work as close as possible to physiological conditions.

NMR data acquisition

All the NMR experiments were performed at 22.3 T on a Bruker Avance III spectrometer operating at 950.20 MHz ^{1}H, 238.93 MHz ^{13}C and 96.28 MHz ^{15}N frequencies, equipped with a cryogenically cooled probehead. Parameters specific to each experiment are reported in the captions of the figures describing the pulse sequences, all reported in

the Supplementary Material. E-BURP2 (or time reversed E-BURP2) (Geen and Freeman 1991) and REBURP (Geen and Freeman 1991) shapes of durations of 1,200 and 1,180 μs, respectively, were employed for ^{1}H band-selective π/2 and π flip angle pulses; BIP-750-50-20 pulse shapes (Smith et al. 2001) of duration of 200 μs were used for broadband ^{1}H inversion. For ^{13}C band-selective π/2 and π flip angle pulses G4 (or time reversed G4) (Emsley and Bodenhausen 1990) and Q3 shapes (Emsley and Bodenhausen 1992) of durations of 260 and 161 μs, respectively, were used, except for the π pulses that should be band-selective on the C^{α} region (Q3, 667 μs) and for the adiabatic π pulse to invert both C' and C^{α} (smoothed Chirp 500 μs, 20 % smoothing, 80 kHz sweep width, 11.3 kHz RF field strength) (Böhlen and Bodenhausen 1993). The ^{13}C band selective pulses on C^{α} and C' were applied at the center of each region, respectively. For the $^{1}H^{\alpha}$ experiments, decoupling of ^{1}H and ^{13}C was achieved with DIPSI-2 (Shaka et al. 1988; Cavanagh and Rance 1992) (3.1 kHz) and GARP-4 (Shaka et al. 1985) (4.5 kHz) sequences, respectively. All gradients employed had a smoothed square shape. The parameters used for the acquisition of the 3D and 5D experiments are reported, respectively, in Table 1 and 2. Each experiment was acquired with the States or Echo-Antiecho method applied in the indirect dimensions to achieve quadrature detection. All the experiments were performed using on-grid non-uniform sampling (NUS). The on-grid "Poisson disk" sampling scheme (Kazimierczuk et al. 2008) was chosen to generate the time schedules with the *RSPack* program. The distribution was relaxation-optimized, i.e. the density of points was decaying according to the Gaussian distribution exp $(-t^2/\sigma^2)$, with $\sigma = 0.5$. All the spectra were acquired using *Bruker TopSpin 3.1* software.

NMR data processing and analysis

The experimental data were converted with *nmrPipe* (Delaglio et al. 1995). 3D data were processed using the multidimensional Fourier transform (MFT) algorithm implemented in the *ToASTD* program (Kazimierczuk et al. 2006). 5D data were processed using the Sparse MFT (SMFT) algorithm implemented in the *reduced* program (Kazimierczuk et al. 2009, 2010b). Both programs are available at http://nmr.cent3.uw.edu.pl/software. Finally, *Sparky* (Goddard and Kneller 2000) was used to analyze the spectra.

Handling and inspection of 5D NMR spectra

The SMFT algorithm (Kazimierczuk et al. 2009) was employed to process the 5D data sets. This method enables to process data with very high resolution in all dimensions by storing only the informative parts of multidimensional

Table 1 Experimental parameters used for the 3D experiments (providing the basis spectra)

	Spectral widths and maximal evolution times			No. of scans	Inter-scan delays (s)	No. of complex points (aq)	No. of hypercomplex points	Duration of the experiment	Relative data points density (%)
3D (HACACO)NCO (CA)HA	3.5 kHz (^{15}N) 35.7 ms	2.0 kHz ($^{13}C'$) 35.0 ms	13.3 kHz ($^{1}H^{\alpha}$)	4	1.0	1,024	1,450	8 h	20.3
5D BEST TROSY HNCO	1.8 kHz ($^{13}C'$) 41.7 ms	2.6 kHz (^{15}N) 40.4 ms	13.3 kHz ($^{1}H^{N}$)	4	0.2	1,024	1,575	3 h	20.0

Table 2 Experimental parameters used for the 5D experiments

	Spectral widths and maximal evolution times					No. of scans	Inter-scan delays (s)	No. of complex points (aq)	No. of hyper-complex points	Duration of the experiment	Relative data points density (%)
	Cross-sections		Dimensions shared with the basis spectrum								
5D (HACA) CON(CACO) NCO(CA)HA	2.0 kHz ($^{13}C'$) 35.0 ms	3.5 kHz (^{15}N) 35.7 ms	3.5 kHz (^{15}N) 35.7 ms	2.0 kHz ($^{13}C'$) 35.0 ms	13.3 kHz ($^{1}H^{\alpha}$)	8	1.0	1,024	1,150	2 days, 7 h	0.0015
5D BEST TROSY (H)NCO (CAN)CONNH	2.6 kHz (^{15}N) 40.4 ms	1.8 kHz ($^{13}C'$) 41.7 ms	1.8 kHz ($^{13}C'$) 41.7 ms	2.6 kHz (^{15}N) 40.4 ms	13.3 kHz ($^{1}H^{N}$)	4	0.2	1,024	1,200	11 h	0.0019
5D BEST TROSY HN (COCAN) CONNH	2.0 kHz ($^{1}H^{N}$) 40.0 ms	2.6 kHz (^{15}N) 40.4 ms	1.8 kHz ($^{13}C'$) 41.7 ms	2.6 kHz (^{15}N) 40.4 ms	13.3 kHz ($^{1}H^{N}$)	4	0.2	1,024	1,200	11 h	0.0018

spectra. Indeed processing NMR spectra with high resolution in all dimensions would provide prohibitively large data files for spectra with more than three dimensions. However, in most cases the number of expected cross-peaks in the spectra does not increase passing from a 3D to a higher dimensional experiment. Therefore, these high dimensional spaces are essentially "empty" with interesting information confined in very narrow regions, those where cross-peaks are observed in the lower-dimensionality 3D spectra. One interesting feature of SMFT is its unique ability to simplify the analysis of multidimensional spectra, by considering that a 5D spectrum can be decomposed into a 3D spectrum in which each detected correlation peak is associated to an additional 2D spectrum, contributing the remaining two chemical shifts. Actually, the 3D spectrum (to which we refer as "basis spectrum") can be conveniently obtained from a simpler (and thus also more sensitive) 3D experiment correlating three (out of the five) resonances. The two remaining frequencies, which are the new information-content of the 5D spectrum, are retrieved in a series of 2D spectra (to which we refer as "cross-sections"), which are the only spectra computed from the 5D data sets. Therefore, the analysis of a 5D spectrum is simplified to the inspection of a series of 2D cross-sections, to which the three "basic" frequencies, retrieved in the corresponding 3D spectrum, are joined.

Results and discussion

To acquire high dimensional NMR experiments (4D, 5D, etc.), the use of sparse sampling strategies is very convenient. In fact, data sampling at non-uniform intervals is not only an ingenious way to speed up the acquisition of multidimensional NMR experiments, but mostly a powerful method to significantly increase the achievable spectral resolution of the indirect dimensions, as demanded by IDPs. Therefore, all the spectra were acquired using NUS. An inconvenient feature of NUS data processing is the appearance of sampling artifacts in the spectrum. Sampling artifacts can be efficiently suppressed, for example by employing the signal separation algorithm (SSA) (Stanek et al. 2012). However, in the case of protein backbone assignment, with spectra characterized by a small number of correlation peaks, artifacts are of low intensity and can be safely ignored.

Using the experimental strategy outlined in the "Materials and methods" section, we recently built a set of 4D and 5D NMR experiments based on ^{13}C-direct detection, specifically tailored to facilitate the chemical shift assignment of IDPs (Bermel et al. 2012, 2013b). The high dimensional experiments were designed as expansion of the well-known 2D CON-IPAP and 3D CACON-IPAP experiments (used as basis spectra to process the 4D and 5D data sets, respectively). From our experience, the experiments in which two pairs of ^{15}N and ^{13}C′ frequencies belonging to two neighboring residues are correlated, in order to sequentially link the peaks retrieved in the 2D CON-IPAP spectrum, provide the best result in terms of rapidity and reliability of the sequential backbone assignment (Bermel et al. 2013b; Pantoja-Uceda and Santoro 2014). This assignment approach, to which we like to refer as "CON-CON strategy", takes advantage of the relatively good chemical shift dispersion of ^{15}N and ^{13}C′ frequencies (Bermel et al. 2013a) of IDPs, which significantly improve the resolution of the spectra together with the use of three or four indirect dimensions. The employment of ^{13}C-direct detection provides a series of additional benefits. First, it reduces the minimum number of dimensions needed to apply the CON-CON strategy to four, because one of the two ^{13}C′ dimensions is directly acquired. Secondly, it allows to detect correlation peaks for all residues, included prolines that are missing in amide ^{1}H-detected experiments. Finally, it allows to perform experiments at near physiological conditions (high temperature and pH), since carbon nuclei are not directly affected by chemical exchange with the solvent. The major drawback of ^{13}C′ direct-detected experiments is the ^{13}C intrinsic lower sensitivity, largely compensated by recent technological improvements, which however can still become a limiting factor for complete assignment of less-concentrated or short-lived protein samples. Herein we now propose a set of 5D NMR experiments that still exploit the CON-CON strategy, but make use of protons instead of carbons for detection.

In order to take advantage of the particular properties of aliphatic and amide ^{1}H, we developed two different pulse sequences in which either H^{α} or H^{N} nuclei are exploited for detection as well as for the starting source of magnetization. The new 5D experiments, named (HACA)CON(CA-CO)NCO(CA)HA, BT-(H)NCO(CAN)CONNH (and its variant BT-HN(COCAN)CONNH) will be discussed in the following in more detail. In all experiments, pairs of ^{15}N and ^{13}C′ resonances are collected in the indirect dimensions. Semi-constant time or real time chemical shift evolution is used to achieve highest possible spectral resolution in all dimensions, making transverse relaxation the only limitation for choosing the appropriate acquisition time.

For its similarity with carbon detection, alpha proton detection (Mäntylahti et al. 2011) was chosen in the design of the 5D (HACA)CON(CACO)NCO(CA)HA experiment. In fact, H^{α} are not influenced by chemical exchange with the solvent even at high temperature and pH, thus being very appropriate to investigate IDPs at near physiological conditions. Moreover, H^{α}-based experiments provide direct information on proline residues, giving the possibility to

obtain virtually complete assignment. The use of deuterated water as solvent (99.99 % D_2O) ensures to completely prevent missing signals under the residual water peak. It also provides further advantages that can enhance the spectrometer performance, like the possibility to work at receiver gain values where digitizer noise is not an issue and to avoid all the phenomena related to radiation damping which become an issue when working with cryoprobes and high salt sample conditions, often necessary when dealing with IDPs. A slight drawback is the increased viscosity of D_2O with respect to H_2O that results in increased transverse relaxation rates. This effect is, however, counterbalanced by the possibility to work at higher temperature, as there is no need to slow down solvent exchange processes.

The coherence transfer pathway and the information provided by the 5D (HACA)CON(CACO)NCO(CA)HA experiment are reported in Fig. 1, top panel. The flow of magnetization, which starts and ends on alpha protons, is established through ^{1}J scalar coupling transfer steps; in particular, the back-transfer step mediated by $^2J(N–C^{\alpha})$ scalar coupling is suppressed. Therefore, the experiment correlates C'_{i-1}–N_i–N_{i+1}–C'_i–H_i^{α} nuclei, making the magnetization transfer pathway unidirectional. In this way, the sequence-specific backbone assignment procedure becomes a "walk" through a series of CON cross-peaks. The sequence was designed to detect only one peak in each cross-section, reducing at maximum the risk of peak overlaps. For SMFT processing, a 3D (HACACO)NCO(CA)HA experiment (Mäntylahti et al. 2010) has to be performed, as it provides N_{i+1}–C'_i–H_i^{α} frequency correlations (basis spectrum), that are used to extract from the 5D spectrum the additional C'_{i-1}–N_i frequencies. Hence, the "CON-CON strategy" is carried out sequence-specifically linking the pairs of C'_i and N_{i+1} resonances, retrieved in the cross-sections, to the ones present in the basis spectrum, as shown in Fig. 2. To facilitate the mapping of the linked resonances on the amino acidic sequence, the 5D (HACA)CON(CACO)NCO(CA)HA experiment has been designed in a way that cross peaks for glycine residues (at the *i* position) are of opposite sign with respect to all others, as illustrated in Fig. 3 where six cross-sections of the 5D spectrum are shown. It is interesting to note that, with the resolution obtained in the direct ^{1}H dimension, doublets due to $^3J(H^{\alpha}–H^{\beta})$ scalar couplings are detected for alanine, valine, isoleucine and threonine residues. However, these line splittings do not significantly complicate the analysis of the spectra.

As an alternative to H^{α} detection, the 5D BT-(H)NCO(CAN)CONNH experiment was developed to exploit the same "CON-CON strategy", but using amide proton detection. The coherence transfer pathway and the information provided by this experiment are reported in Fig. 1, bottom panel. The experiment is a variant of the previously published 5D (H)NCO(NCA)CONH experiment (Zawadzka-Kazimierczuk et al. 2012b). However, in our new pulse sequence, the magnetization is transferred from and to neighboring amide protons by again only exploiting ^{1}J scalar coupling transfer steps, resulting in unambiguous N_{i+1}–C'_i–C'_{i-1}–N_i–H_i^N correlations. The experiment is implemented as a BEST-TROSY (BT) version, in order to benefit from the short selective longitudinal relaxation times of amide protons in IDPs (Solyom et al. 2013; Gil et al. 2013). The combination of BEST (Pervushin et al. 2002; Schanda 2009; Schanda et al. 2006) and TROSY (Pervushin et al. 1997), as described by Favier and Brutscher (Favier and Brutscher 2011), ensures highest

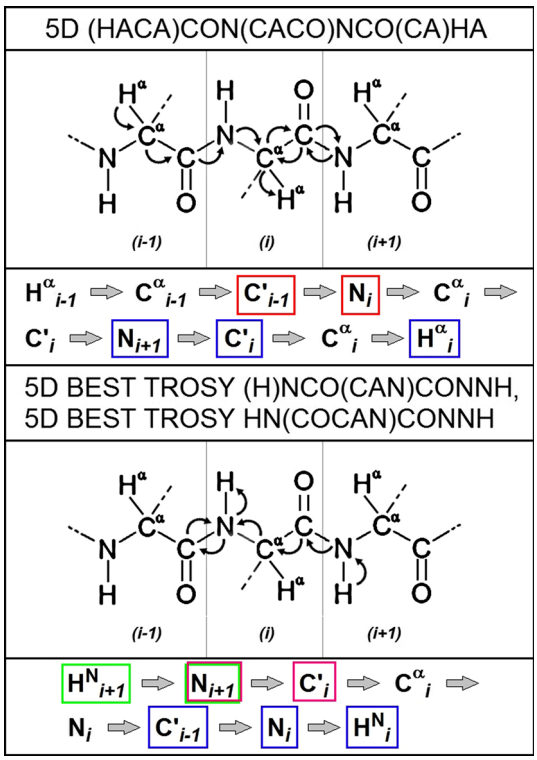


Fig. 1 The flow of the magnetization in the proposed experiments is schematically illustrated. The magnetization transfer pathway of the 5D (HACA)CON(CACO)NCO(CA)HA experiment is reported in the *top panel*, whereas that of the 5D BT-(H)NCO(CAN)CONNH and BT-HN(COCAN)CONNH experiments at the *bottom*. All *arrows* represent coherence transfer steps through ^{1}J scalar couplings. For each experiment, the correlations retrieved are also listed explicitly. Those embedded in colored boxes are the ones frequency-labeled within the experiments. *Blue rectangles* show the dimensions shared with the basis spectrum, whereas *red*, *magenta* and *green rectangles* indicate the dimensions of the 2D cross-sections of the 5D (HACA)CON(CACO)NCO(CA)HA, 5D BT-(H)NCO(CAN)CONNH and 5D BT-HN(COCAN)CONNH experiments, respectively. The two variants of the BEST TROSY experiments have in common four dimensions

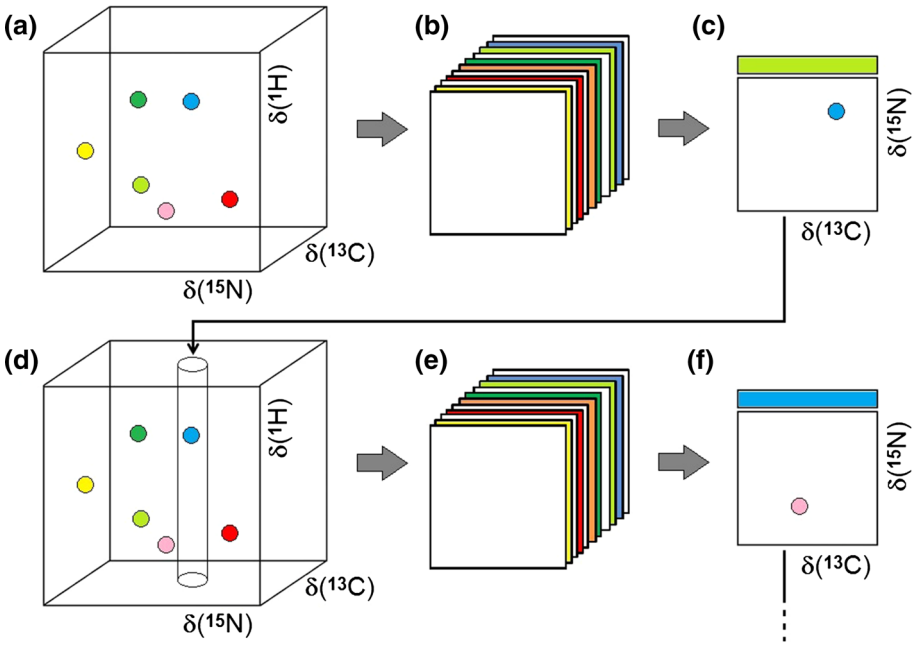


Fig. 2 Schematic illustration of the "CON-CON strategy". The sequence-specific assignment is performed through simultaneous exploitation of C′ and N resonances, linked between the 2D cross-sections of the 5D spectrum and the 3D basis spectrum. The same procedure can be used both for $^{1}H^{\alpha}$ and $^{1}H^{N}$ detected spectra, as described here schematically. First, a cross-peak (*light green* in the example) of the 3D basis spectrum is chosen (**a**). Thereafter, the related 2D cross-section of the 5D spectrum is inspected (**b**) to retrieve the frequencies of the CON peak immediately preceding or succeeding (for the $^{1}H^{\alpha}$ and $^{1}H^{N}$ spectrum, respectively) the one of the basis spectrum (**c**). Once the cross-peak (*blue* in the example) of the 3D basis spectrum with the same C′ and N frequencies is found (**d**), it becomes sequence-specifically linked to the one found in (**a**). Then, the procedure continues in (**e**) and (**f**) as explained in (**b**) and (**c**), forming chains of adjacent CON peaks in the primary sequence of the protein

experimental sensitivity and spectral resolution in very short overall acquisition times. As amide protons are affected by chemical exchange with the solvent, it is opportune to perform the experiment at lower temperature and pH in order to prevent extensive line broadening. In order to obtain the basis spectrum, correlating C'_{i-1}–N_i–H_i^N nuclei, an additional BT-HNCO (Solyom et al. 2013) spectrum is recorded. The frequency information, obtained from BT-HNCO, then allows to extract a series of 2D cross-sections from the 5D BT-(H)NCO(CAN)CONNH data set that adds the additional N_{i+1}–C'_i information (Fig. 4, panel a). Then, as described for the H^{α}-detected experiment, it is possible to establish sequential correlations linking the pairs of C'_i and N_{i+1} frequencies present in the 2D cross-sections of the 5D spectrum to those embedded in the 3D basis spectrum (Fig. 2). The only difference with the H^{α}-detected experiment is that the sequential walk now proceeds from the N- to the C-terminus, instead of the inverse. Because of the higher intrinsic sensitivity and shorter overall experimental time requirements, the 5D BT-(H)NCO(CAN)CONNH experiments is especially appealing for fast-degrading or low-concentrated IDP samples.

Finally, the above described "CON-CON strategy" can be extended to include the amide ^{1}H, thus becoming a "HNCO-HNCO strategy". For this purpose, the H^N_{i+1} needs to be added to the frequency information contained in the 5D BT-(H)NCO(CAN)CONNH spectrum. In principle, this can be achieved by recording a 6D BT-HNCO(CAN)CONNH correlation spectrum, adding an amide ^{1}H chemical shift evolution period at the beginning of the pulse sequence (Zawadzka-Kazimierczuk et al. 2012b). Here we have chosen an alternative approach that consists in recording a second 5D experiment, BT-HN(COCAN)CONNH, where chemical shift evolution of one of the carbonyls is replaced by that of the remaining amide proton, thus correlating H^N_{i+1}–N_{i+1}–C'_{i-1}–N_i–H_i^N nuclei (Fig. 1, bottom panel). The same 3D BT-HNCO spectrum can be used to compute the 2D cross-sections for this additional 5D experiment. Recording of the two 5D data sets with a different set of frequency-labeled nuclei is equivalent to recording a complete 6D data set as long as there is no degenerate N_{i+1}–C'_{i-1}–N_i–H_i^N frequency correlation. This strategy is well-known as projection spectroscopy (Kupce and Freeman 2003). The "HNCO-HNCO strategy" is likely to provide a high level of unambiguous sequential

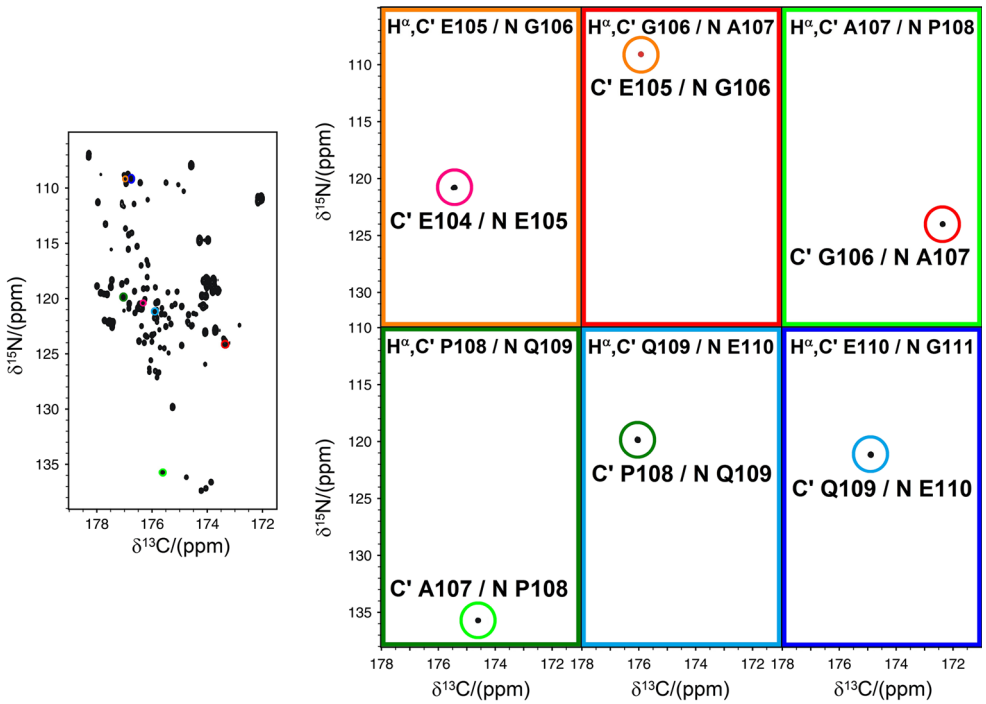


Fig. 3 Six 2D cross-sections of the 5D (HACA)CON(CACO)NCO (CA)HA spectrum are shown (*right side*), extracted at the N_{i+1}–C'_i–H^{α}_i frequencies reported on the top of each panel and schematically indicated by circles on the 2D C′–N projection of the 3D basis spectrum. In each cross-section, the observed cross-peak allows to identify the frequencies of the previous CON cross-peak (N_i–C'_{i-1}). *Black* peaks are positive, *red* negative. Inversion of peak sign is expected when a glycine is frequency-labeled. Therefore, negative peaks allow a straightforward identification of glycines. The sequential specific assignment is performed connecting the CON peak of a given cross-sections to the one of the basis spectrum with identical C′ and N frequencies, as explained in Fig. 2 and illustrated here with the use of colors, so that each color identifies a specific pair of N_{i+1}–C'_i frequencies. The simultaneous exploitation of carbonyl carbon and nitrogen makes the sequential assignment extremely reliable. As clearly visible, the use of five dimensions provide extremely-resolved cross-sections, minimizing the chance of possible overlaps

assignment, even for large IDPs with repetitive primary sequences, since it is very rare even for IDPs that two or more peaks retrieved in the 3D HNCO spectrum are completely overlapping. Analogously, a similar strategy could also be considered for the H^{α}-detected experiment although the added value is less pronounced due to the lower chemical shift dispersion of H^{α}.

Finally, it's worth mentioning that sequence-specific assignment of resonances based on the proposed experiments can be performed automatically, for example by using the *TSAR* program (Zawadzka-Kazimierczuk et al. 2012a), available at http://nmr.cent3.uw.edu.pl/software. However, the programs for automatic assignment usually require at least one experiment providing sequential connectivities and at least one experiment providing information on amino-acid type. Therefore automatic assignment procedures were not attempted in the present work, since all proposed experiments belong to the first category.

Conclusions

We have presented new high-dimensional (5D) experiments exploiting C'_i–N_{i+1} pairs for sequence-specific assignment, based on either H^{α} or H^{N} detection. These experiments provide new complementary tools to the previously described ^{13}C-detected experiments for highly efficient backbone resonance assignment of flexible IDPs. The "CON-CON strategy", including different experimental variants based on ^{13}C, $^{1}H^{\alpha}$ and $^{1}H^{N}$ detection, and its extension to the "HNCO-HNCO strategy", in case of $^{1}H^{N}$ detection, allow to establish almost unambiguous sequential correlations along the protein backbone. While the H^{α}–detected experiment is especially useful for IDP samples studied under near physiological conditions and/or characterized by proline-rich repetitive amino acid sequences, the H^{N}-detected experiments are expected to be favorable for low concentrated or short-lived IDP samples that can be studied under conditions where solvent exchange is sufficiently slowed down.

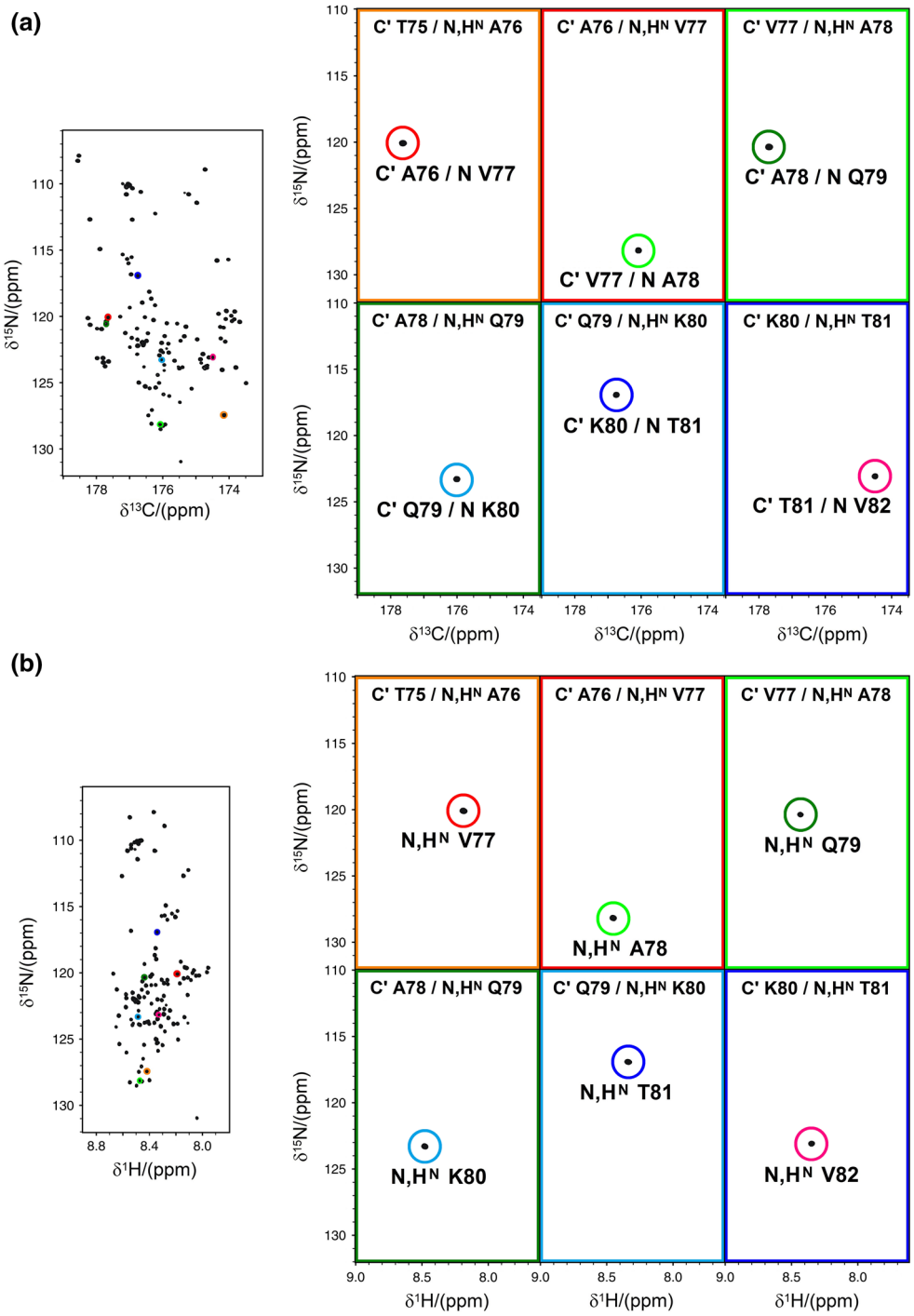


Fig. 4 Six 2D cross-sections of the 5D BT-(H)NCO(CAN)CONNH and 5D BT-HN(COCAN)CONNH spectra, extracted at C'_{i-1}–N_i–H^N_i frequencies, are shown respectively in (**a**) and (**b**). The color code and all the labels have the same meaning as described in the caption of Fig. 3. In each cross-section, the observed cross-peak allows to identify (**a**) the succeeding CON (C'_i–N_{i+1}) and (**b**) the succeeding HN (N_{i+1}–H^N_{i+1}). For this reason, on the left side the related C′–N (**a**) and H^N–N (**b**) projection of the common 3D HNCO basis experiment are reported. Since the cross-section of the two spectra share one nitrogen dimension, they can be exploited together to provide almost unambiguous assignment in the so-called "HNCO-HNCO strategy"

Acknowledgments This work has been supported in part by the European Commission Projects IDPbyNMR (Contract No. 264257), BioNMR (Contract No. 261863) and INSTRUCT (Contract No. 211252).

References

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

Bermel W, Bruix M, Felli IC, Kumar VMV, Pierattelli R, Serrano S (2013a) Improving the chemical shift dispersion of multidimensional NMR spectra of intrinsically disordered proteins. J Biomol NMR 55:231–237

Bermel W, Felli IC, Gonnelli L, Koźmiński W, Piai A, Pierattelli R, Zawadzka-Kazimierczuk A (2013b) High-dimensionality ^{13}C direct-detected NMR experiments for the automatic assignment of intrinsically disordered proteins. J Biomol NMR 57:353–361

Böhlen J-M, Bodenhausen G (1993) Experimental aspects of chirp NMR spectroscopy. J Magn Reson Ser A 102:293–301

Cavanagh J, Rance M (1992) Suppression of cross-relaxation effects in TOCSY spectra via a modified DIPSI-2 mixing sequence. J Magn Reson 96:670–678

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

Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. J Biomol NMR 6:277–293

Deschamps M, Campbell ID (2006) Cooling overall spin temperature: protein NMR experiments optimized for longitudinal relaxation effects. J Magn Reson 178:206–211

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

Dyson HJ, Wright PE (2004) Unfolded proteins and protein folding studied by NMR. Chem Rev 104:3607–3622

Emsley L, Bodenhausen G (1990) Gaussian pulse cascades: new analytical functions for rectangular selective inversion and in-phase excitation in NMR. Chem Phys Lett 165:469–476

Emsley L, Bodenhausen G (1992) Optimization of shaped selective pulses for NMR using a quaternion description of their overall propagators. J Magn Reson 97:135–148

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 advancements in solution NMR: fast methods and heteronuclear direct detection. ChemPhysChem 10:1356–1368

Felli IC, Pierattelli R (2014) Novel methods based on ^{13}C detection to study intrinsically disordered proteins. J Magn Reson 241:115–125

Felli IC, Piai A, Pierattelli R (2013) Recent advances in solution NMR studies: ^{13}C direct detection for biomolecular NMR applications. Ann Rep NMR Spectroscop 80:359–418

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

Gil S, Hošek T, Solyom Z, Kümmerle R, Brutscher B, Pierattelli R, Felli IC (2013) NMR studies of intrinsically disordered proteins near physiological conditions. Angew Chem Int Ed 52:11808–11812

Goddard TD, Kneller DG (2000) SPARKY 3. University of California, San Francisco

Hiller S, Wasmer C, Wider G, Wüthrich K (2007) Sequence-specific resonance assignment of soluble nonglobular proteins by 7D APSY-NMR spectroscopy. J Am Chem Soc 129:10823–10828

Hsu ST, Bertoncini CW, Dobson CM (2009) Use of protonless NMR spectroscopy to alleviate the loss of information resulting from exchange-broadening. J Am Chem Soc 131:7222–7223

Huang C, Ren G, Zhou H, Wang C (2005) A new method for purification of recombinant human alpha-synuclein in *Escherichia coli*. Protein Expr Purif 42:173–177

Ikura M, Spera S, Barbato G, Kay LE, Krinks M, Bax A (1991) Secondary structure and side-chain ^{1}H and ^{13}C resonance assignments of calmodulin in solution by heteronuclear multidimensional NMR spectroscopy. Biochemistry 30:9216–9228

Kazimierczuk K, Zawadzka A, Koźmiński W, Zhukov I (2006) Random sampling of evolution time space and Fourier transform processing. J Biomol NMR 36:157–168

Kazimierczuk K, Zawadzka A, Koźmiński W (2008) Optimization of random time domain sampling in multidimensional NMR. J Magn Reson 192:123–130

Kazimierczuk K, Zawadzka A, Koźmiński W (2009) Narrow peaks and high dimensionalities: exploiting the advantages of random sampling. J Magn Reson 197:219–228

Kazimierczuk K, Stanek J, Zawadzka-Kazimierczuk A, Koźmiński W (2010a) Random sampling in multidimensional NMR spectroscopy. Prog NMR Spectrosc 57:420–434

Kazimierczuk K, Zawadzka-Kazimierczuk A, Koźmiński W (2010b) Non-uniform frequency domain for optimal exploitation of non-uniform sampling. J Magn Reson 205:286–292

Kazimierczuk K, Misiak M, Stanek J, Zawadzka-Kazimierczuk A, Koźmiński W (2012) Generalized Fourier transform for non-uniform sampled data. Top Curr Chem 316:79–124

Kazimierczuk K, Stanek J, Zawadzka-Kazimierczuk A, Koźmiński W (2013) High-dimensional NMR spectra for structural studies of biomolecules. ChemPhysChem 14:3015–3025

Knoblich K, Whittaker S, Ludwig C, Michiels P, Jiang T, Schaffhausen B, Günther U (2009) Backbone assignment of the N-terminal polyomavirus large T antigen. Biomol NMR Assign 3:119–123

Konrat R (2014) NMR contributions to structural dynamics studies of intrinsically disordered proteins. J Magn Reson 241:74–85

Kupce E, Freeman R (2003) Projection-reconstruction of three-dimensional NMR spectra. J Am Chem Soc 125:13958–13959

Mäntylahti 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

Mäntylahti S, Hellman M, Permi P (2011) Extension of the HA-detection based approach: (HCA)CON(CA)H and (HCA)NCO(CA)H experiments for the main-chain assignment of intrinsically disordered proteins. J Biomol NMR 49:99–109

Mittag T, Forman-Kay J (2007) Atomic-level characterization of disordered protein ensembles. Curr Opin Struct Biol 17:3–14

Motáčkova V, Nováček J, Zawadzka-Kazimierczuk A, Kazimierczuk K, Žídek L, Sanderová H, Krásný L, Koźmiński W, Sklenář V (2010) Strategy for complete NMR assignment of disordered proteins with highly repetitive sequences based on resolution-enhanced 5D experiments. J Biomol NMR 48:169–177

Narayanan RL, Dürr HN, Bilbow S, Biernat J, Mendelkow E, Zweckstetter M (2010) Automatic assignment of the intrinsically disordered protein Tau with 441-residues. J Am Chem Soc 132:11906–11907

Nováček J, Zawadzka-Kazimierczuk A, Papoušková V, Žídek L, Sanderová H, Krásný L, Koźmiński W, Sklenář V (2011) 5D 13C-detected experiments for backbone assignment of unstructured proteins with a very low signal dispersion. J Biomol NMR 50:1–11

Nováček J, Janda L, Dopitová R, Žídek L, Sklenář V (2013) Efficient protocol for backbone and side-chain assignments of large, intrinsically disordered proteins: transient secondary structure analysis of 49.2 kDa microtubule associated protein 2c. J Biomol NMR 56:291–301

Nováček J, Žídek L, Sklenář V (2014) Toward optimal-resolution NMR of intrinsically disordered proteins. J Magn Reson 241:41–52

O'Hare B, Benesi AJ, Showalter SA (2009) Incorporating 1H chemical shift determination into 13C-direct detected spectroscopy of intrinsically disordered proteins in solution. J Magn Reson 200:354–358

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

Pantoja-Uceda D, Santoro J (2013a) A suite of amino acid residue type classification pulse sequences for ^{13}C-detected NMR of proteins. J Magn Reson 234:190–196

Pantoja-Uceda D, Santoro J (2013b) Direct correlation of consecutive C′–N groups in proteins: a method for the assignment of intrinsically disordered proteins. J Biomol NMR 57:57–63

Pantoja-Uceda D, Santoro J (2014) New 13C-detected experiments for the assignment of intrinsically disordered proteins. J Biomol NMR 59:43–50

Pérez Y, Gairí M, Pons M, Bernadò P (2009) Structural characterization of the natively unfolded N-terminal domain of human c-Src kinase: insights into the role of phosphorylation of the unique domain. J Mol Biol 391:136–148

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

Sattler M, Schleucher J, Griesinger C (1999) Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. Progr NMR Spectrosc 34:93–158

Schanda P (2009) Fast-pulsing longitudinal relaxation optimized techniques: enriching the toolbox. Prog NMR Spectrosc 55:238–265

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

Shaka AJ, Barker PB, Freeman R (1985) Computer-optimized decoupling scheme for wideband applications and low-level operation. J Magn Reson 64:547–552

Shaka AJ, Lee CJ, Pines A (1988) Iterative schemes for bilinear operators; application to spin decoupling. J Magn Reson 77:274–293

Skora L, Becker S, Zweckstetter M (2010) Molten globule precursor states are conformationally correlated to amyloid fibrils of human beta-2-microglobulin. J Am Chem Soc 132:9223–9225

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

Solyom Z, Schwarten M, Geist L, Konrat R, Willbold D, Brutscher B (2013) BEST-TROSY experiments for time-efficient sequential resonance assignment of large disordered proteins. J Biomol NMR 55:311–321

Stanek J, Augustyniak R, Koźmiński W (2012) Suppression of sampling artefacts in high-resolution four-dimensional NMR spectra using signal separation algorithm. J Magn Reson 214: 91–102

Tompa P (2009) Structure and function of intrinsically disordered proteins. CRC Press, Boca Raton

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

Uversky VN (2013a) A decade and a half of protein intrinsic disorder: biology still waits for physics. Protein Sci 22:693–724

Uversky VN (2013b) Multitude of binding modes attainable by intrinsically disordered proteins: a portrait gallery of disorder-based complexes. Chem Soc Rev 40:1623–1634

Uversky V, Dunker AK (2013) The case for intrinsically disordered proteins playing contributory roles in molecular recognition without a stable 3D structure. F1000 Biol Rep 5:1

Uversky VN, Gillespie JR, Fink AL (2000) Why are "natively unfolded" proteins unstructured under physiologic conditions? Proteins Struct Funct Genet 41:415–427

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

Zawadzka-Kazimierczuk A, Koźmiński W, Billeter M (2012a) TSAR: a program for automatic resonance assignment using 2D cross-sections of high dimensionality, high-resolution spectra. J Biomol NMR 54:81–95

Zawadzka-Kazimierczuk A, Koźmiński W, Sanderová H, Krásný L (2012b) High dimensional and high resolution pulse sequences for backbone resonance assignment of intrinsically disordered proteins. J Biomol NMR 52:329–337

“CON-CON” assignment strategy for highly flexible intrinsically disordered proteins

Alessandro Piai[1], Tomas Hošek[1], Leonardo Gonnelli[1], Anna Zawadzka-Kazimierczuk[2], Wiktor Koźmiński[2], Bernhard Brutscher[3], Wolfgang Bermel[4], Roberta Pierattelli[1]✉, Isabella C. Felli[1]✉

[1] CERM and Department of Chemistry, University of Florence, Via Luigi Sacconi 6, 50019 Sesto Fiorentino, Florence, Italy

[2] Faculty of Chemistry, Biological and Chemical Research Centre, University of Warsaw, Żwirki i Wigury 101, 02-089 Warsaw, Poland

[3] Institut de Biologie Structurale, Université Grenoble 1, CNRS, CEA, Rue Jules Horowitz 41, 38027 Grenoble Cedex 1, France

[4] Bruker BioSpin GmbH, Silberstreifen, 76287 Rheinstetten, Germany

SUPPLEMENTARY MATERIAL

NMR experiments

In this section the pulse schemes for the following NMR experiments are described: 5D BT-(H)NCO(CAN)CONNH, 5D BT-HN(COCAN)CONNH, 3D (HACACO)NCO(CA)HA and 5D (HACA)CON(CACO)NCO(CA)HA.

E-BURP2 and REBURP shapes of durations of 1200 and 1180 ms, respectively, were employed for ^{1}H band-selective π/2 and π flip angle pulses; BIP-750-50-20 pulse shapes of duration of 200 ms were used for broadband ^{1}H inversion. For ^{13}C band-selective π /2 and π flip angle pulses G4 (or time reversed G4) and Q3 shapes of durations of 260 and 161 ms, respectively, were used, except for the π pulses that should be band-selective on the C^{α} region (Q3, 667 ms) and for the adiabatic π pulse to invert both C' and C^{α} (smoothed Chirp 500 ms, 20% smoothing, 80 kHz sweep width, 11.3 kHz RF field strength). The ^{13}C band selective pulses on C^{α} and C' were applied at the center of each region, respectively. For the $^{1}H^{\alpha}$ experiments, decoupling of ^{1}H and ^{13}C was achieved with DIPSI-2 (3.1 kHz) and GARP-4 (4.5 kHz) sequences, respectively. All gradients employed had a smoothed square shape.

In the figures, the arrows indicate the switching of the ^{13}C carrier frequency. Narrow and wide symbols stand for 90° and 180° pulses, respectively. The pulses were applied along the *x* axis unless noted differently. The rectangles represent non-selective pulses, the round shapes represent band-selective pulses. The ^{13}C π pulses that should be highly band-selective on the C^{α} region are marked by an asterisk, while the grey round shapes represent ^{13}C adiabatic pulses. Open squares on ^{1}H channel are used to indicate BIP-720-50-20 broadband inversion pulses applied in the BEST TROSY (BT) experiments. The line denoted with PFG stands for pulsed field gradients applied along the *z*-axis. The phase cycle and the duration of the delays for these experiments are reported case-by-case afterwards in the captions of the figures describing the pulse sequences.

Instead, for the 3D BT-HNCO experiment only the experimental parameters are given.

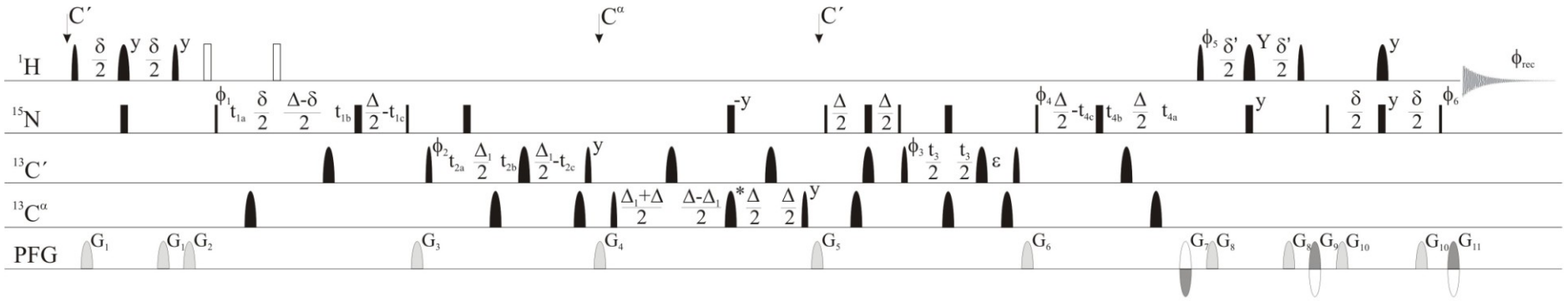


Figure S1. Pulse sequence of the 5D BT-(H)NCO(CAN)CONNH experiment

The following phase cycling was employed: ϕ_1 = 4(x), 4(-x); ϕ_2 = 2(x), 2(-x); ϕ_3 = 8(x), 8(-x); ϕ_4 = x, -x; ϕ_5 = -y; ϕ_6 = -y; ϕ_{rec} = x, 2(-x), x, -x, 2(x), 2(-x), 2(x), -x, x, 2(-x), x. The lengths of the delays were: δ = 5.5 ms, δ' = 5.5 ms $-2\cdot p_{43}\cdot 0.69$ (where p_{43} is the duration of the selective $\pi/2$ pulse on $^1H^N$), Δ = 24.8 ms, Δ_1 = 9.0 ms, $\varepsilon = t_3(0)-p_{14}$ (where p_{14} is the duration of the π pulse on $^{13}C^\alpha$ in the middle of the $^{13}C'$ evolution period). All gradients had 1 ms of duration, except G_1 and G_8 which were 4 ms long. Their relative strengths were the following: G_1 = 3%, G_2 = 30%, G_3 = 21%, G_4 = 17%, G_5 = 13%, G_6 = 11%, G_7 = -80%, G_8 = 5%, G_9 = 30%, G_{10} = 45%, G_{11} = 30.13% (with 100% corresponding to 0.65 T/m). The semi-constant time mode for the evolution of chemical shift was used in the first carbonyl and in both nitrogen evolution periods, whereas real time evolution was adopted for the second carbonyl evolution period. For the semi-constant time mode, the delays were set as follows: $t_{1a} = t_1/2$; $t_{1b} = t_{1a}-t_{1c}$; $t_{1c} = (\Delta/N_1)$, with N_1 the number of total increments in the t_1 dimension; $t_{2a} = t_2/2$; $t_{2b} = t_{2a}-t_{2c}$; $t_{2c} = (\Delta_1/N_2)$, with N_2 the number of total increments in the t_2 dimension; $t_{4a} = t_4/2$; $t_{4b} = t_{4a}-t_{4c}$; $t_{4c} = (\Delta/N_4)$, with N_4 the number of total increments in the t_4 dimension. Quadrature detection in the indirect dimensions was achieved by incrementing phases ϕ_1, ϕ_2 and ϕ_3 in a States manner; for quadrature detection in t_4, echo-antiecho data were recorded by inverting the sign of phases ϕ_5 and ϕ_6 together with multiplying gradients G_7, G_9 and G_{11}, respectively, by 0.8750, 1.0000, 0.6595 and, alternatively, 1.0000, 0.6667, 1.0000.

Experimental parameters used. All selective ^1H pulses were centered at 9.0 ppm. The 5D BT-(H)NCO(CAN)CONNH experiment was recorded in 11 hours with 4 scans per increment, an inter-scan delay of 0.2 s, spectral widths of 13300 (aq) × 2600 (^{15}N) × 1800 ($^{13}C'$) Hz with 1024 complex points in the acquisition dimension and 1200 hypercomplex points randomly distributed in the indirectly detected dimensions (data points density of 0.0019 %). The maximum evolution times in the indirectly detected dimensions were set to 40.4 ms for the ^{15}N dimensions (105 complex points) and to 41.7 ms for the ^{13}C dimensions (75 complex points).

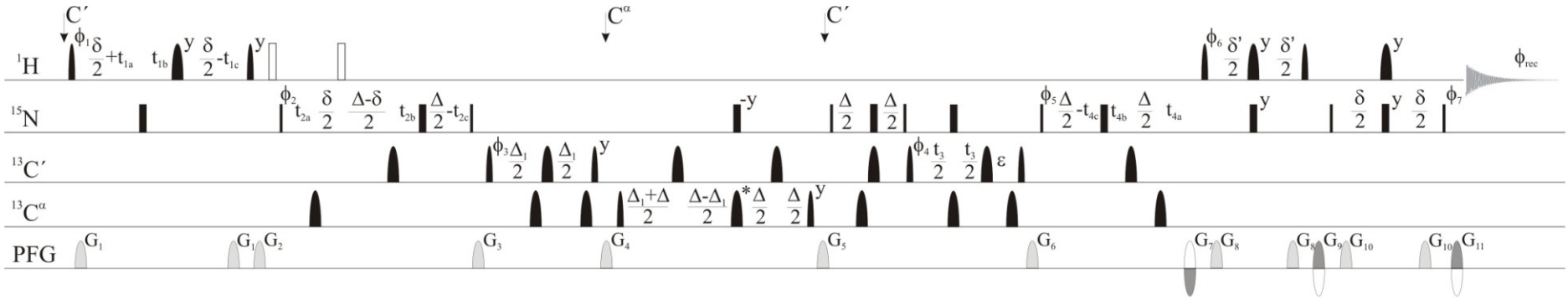


Figure S2. Pulse sequence of the 5D BT-HN(COCAN)CONNH experiment

The following phase cycling was employed: ϕ_1 = x; ϕ_2 = 4(x), 4(-x); ϕ_3 = 2(x), 2(-x); ϕ_4 = 8(x), 8(-x); ϕ_5 = x, -x; ϕ_6 = -y; ϕ_7 = -y; ϕ_{rec} = x, 2(-x), x, -x, 2(x), 2(-x), 2(x), -x, x, 2(-x), x. The lengths of the delays were: δ = 5.5 ms, δ' = 5.5 ms $-2\cdot p_{43}\cdot 0.69$ (where p_{43} is the duration of the selective $\pi/2$ pulse on $^1H^N$), Δ = 24.8 ms, Δ_1 = 9.0 ms, $\varepsilon = t_3(0)-p_{14}$ (where p_{14} is the duration of the π pulse on $^{13}C^\alpha$ in the middle of the $^{13}C'$ evolution period). All gradients had 1 ms of duration, except G_1 and G_8 which were 4 ms long. Their relative strengths were the following: G_1 = 3%, G_2 = 30%, G_3 = 21%, G_4 = 17%, G_5 = 13%, G_6 = 11%, G_7 = -80%, G_8 = 5%, G_9 = 30%, G_{10} = 45%, G_{11} = 30.13% (with 100% corresponding to 0.65 T/m). The semi-constant time mode for the evolution of chemical shift was used in the first carbonyl and in both nitrogen evolution periods, whereas real time evolution was adopted for the second carbonyl evolution period. For the semi-constant time mode, the delays were set as follows: $t_{1a} = t_1/2$; $t_{1b} = t_{1a}-t_{1c}$; $t_{1c} = (\delta/N_1)$, with N_1 the number of total increments in the t_1 dimension; $t_{2a} = t_2/2$; $t_{2b} = t_{2a}-t_{2c}$; $t_{2c} = (\Delta/N_2)$, with N_2 the number of total increments in the t_2 dimension; $t_{4a} = t_4/2$; $t_{4b} = t_{4a}-t_{4c}$; $t_{4c} = (\Delta/N_4)$, with N_4 the number of total increments in the t_4 dimension. Quadrature detection in the indirect dimensions was achieved by incrementing phases ϕ_1, ϕ_2 and ϕ_4 in a States manner; for quadrature detection in t_4, echo-antiecho data were recorded by inverting the sign of phases ϕ_6 and ϕ_7 together with multiplying gradients G_7, G_9 and G_{11}, respectively, by 0.8750, 1.0000, 0.6595 and, alternatively, 1.0000, 0.6667, 1.0000.

Experimental parameters used. All selective ^1H pulses were centered at 9.0 ppm. The 5D BT-HN(COCAN)CONNH experiment was recorded in 11 hours with 4 scans per increment, an inter-scan delay of 0.2 s, spectral widths of 13300 (aq) × 2600 (^{15}N) × 1800 ($^{13}C'$) × 2000 (^1H) Hz with 1024 complex points in the acquisition dimension and 1200 hypercomplex points randomly distributed in the indirectly detected dimensions (data points density of 0.0018 %). The maximum evolution times in the indirectly detected dimensions were set to 40.0 ms for the ^1H dimension (80 complex points), to 40.4 ms for both the ^{15}N dimensions (105 complex points) and to 41.7 ms for the ^{13}C dimension (75 complex points).

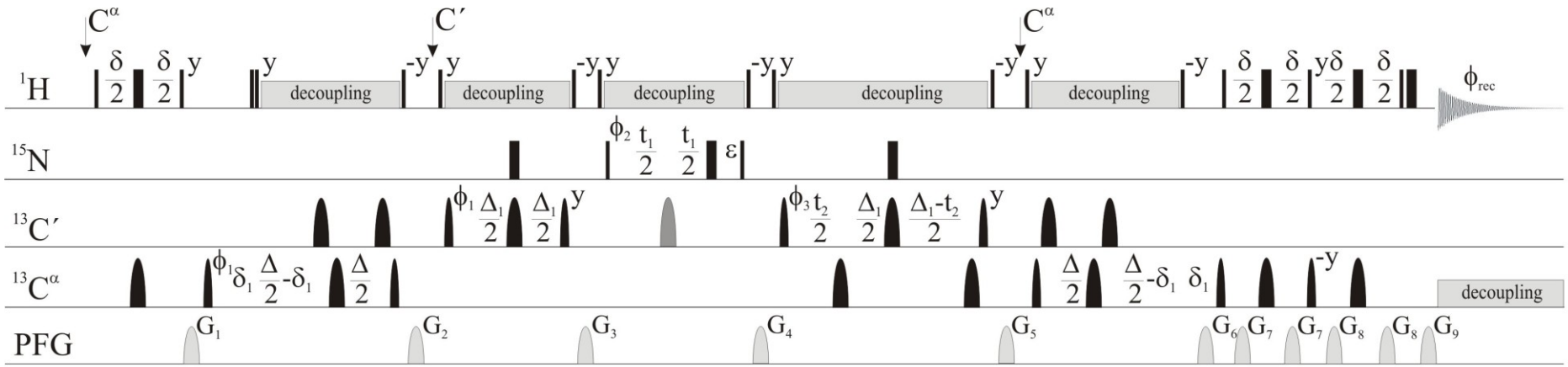


Figure S3. Pulse sequence of the 3D (HACACO)NCO(CA)HA experiment

The following phase cycling was employed: $\phi_1 = 4(x), 4(-x)$; $\phi_2 = 2(x), 2(-x)$; $\phi_3 = x, -x$; $\phi_{rec} = x, 2(-x), x, -x, 2(x), -x$. The lengths of the delays were: $\delta = 3.4$ ms, $\delta_1 = 2.2$ ms, $\Delta = 9.0$ ms, $\Delta_1 = 33.2$ ms, $\varepsilon = t_3(0)+p_8$, where p_8 is the duration of the adiabatic pulse on ^{13}C in the middle of the ^{15}N evolution period. All gradients had 1 ms of duration, except G_7 and G_8 which were 0.5 ms long. Their relative strengths were the following: $G_1 = 40\%$, $G_2 = 40\%$, $G_3 = 40\%$, $G_4 = 40\%$, $G_5 = 40\%$, $G_6 = 80\%$, $G_7 = 5\%$, $G_8 = 2\%$, $G_9 = 20.1\%$ (with 100% corresponding to 0.65 T/m). The real time mode for the evolution of chemical shift evolution was adopted for the nitrogen evolution period, whereas the constant time mode was used in the carbonyl evolution period (if necessary, the evolution of the chemical shift can be switched to the semi-constant time mode to further increase the resolution in the resulting spectrum). Quadrature detection in the indirect dimensions was achieved by incrementing phases ϕ_2 and ϕ_3 in a States manner.

Experimental parameters used. The 3D (HACACO)NCO(CA)HA experiment was recorded in 8 hours with 4 scans per increment, an inter-scan delay of 1.0 s, spectral widths of 13300 (aq) × 3500 (^{15}N) × 2000 ($^{13}C'$) Hz with 1024 complex points in the acquisition dimension and 1450 hypercomplex points randomly distributed in the indirectly detected dimensions (data points density of 20.3 %). The maximum evolution times in the indirectly detected dimensions were set to 31.4 ms for the ^{15}N dimension (110 complex points) and to 32.5 ms for the ^{13}C dimension (65 complex points)

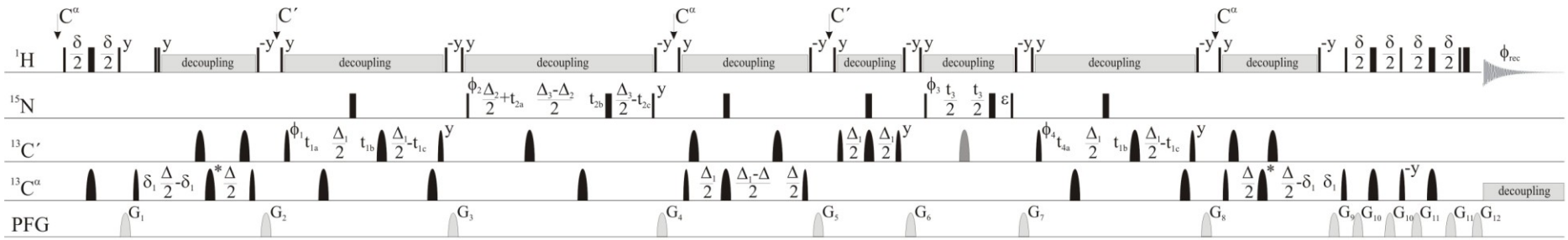


Figure S4. Pulse sequence of the 5D (HACA)CON(CACO)NCO(CA)HA experiment

The following phase cycling was employed: ϕ_1 = x, -x; ϕ_2 = 2(x), 2(-x); ϕ_3 = 4(x), 4(-x); ϕ_4 = 8(x), 8(-x); ϕ_{rec} = x, 2(-x), x, -x, 2(x), 2(-x), 2(x), -x, x, 2(-x), x. The lengths of the delays were: δ = 3.4 ms, δ_1 = 2.2 ms, Δ = 9.0 ms, Δ_1 = 28.0 ms, Δ_2 = 33.2 ms, Δ_3 = 50.0 ms, $\varepsilon = t_3(0)+p_8$, where p_8 is the duration of the adiabatic pulse on ^{13}C in the middle of the ^{15}N evolution period. All gradients had 1 ms of duration, except G_{10} and G_{11} which were 0.5 ms long. Their relative strengths were the following: G_1 = 70%, G_2 = 70%, G_3 = 70%, G_4 = 70%, G_5 = 70%, G_6 = 70%, G_7 = 70%, G_8 = 70%, G_9 = 80%, G_{10} = 5%, G_{11} = 2%, G_{12} = 20.1% (with 100% corresponding to 0.65 T/m). The semi-constant time mode for the evolution of chemical shift was used in the first nitrogen and in both carbonyl evolution periods, whereas real time evolution was adopted for the second nitrogen evolution period. For the semi-constant time mode, the delays were set as follows: $t_{1a} = t_1/2$; $t_{1b} = t_{1a}-t_{1c}$; $t_{1c} = (\Delta_1/N_1)$, with N_1 the number of total increments in the t_1 dimension; $t_{2a} = t_2/2$; $t_{2b} = t_{2a}-t_{2c}$; $t_{2c} = (\Delta_3/N_2)$, with N_2 the number of total increments in the t_2 dimension; $t_{4a} = t_4/2$; $t_{4b} = t_{4a}-t_{4c}$; $t_{4c} = (\Delta_1/N_4)$, with N_4 the number of total increments in the t_4 dimension. Quadrature detection in the indirect dimensions was achieved by incrementing phases ϕ_1, ϕ_2, ϕ_3 and ϕ_4 in a States manner.

Experimental parameters used. The 5D (HACA)CON(CACO)NCO(CA)HA experiment was recorded in 2 days and 7 hours with 8 scans per increment, an inter-scan delay of 1.0 s, spectral widths of 13300 (aq) × 3500 (^{15}N) × 2000 (^{13}C') Hz with 1024 complex points in the acquisition dimension and 1150 hypercomplex points randomly distributed in the indirectly detected dimensions (data points density of 0.015 %). The maximum evolution times in the indirectly detected dimensions were set to 35.7 ms for the ^{15}N dimensions (125 complex points) and to 35.0 ms for the ^{13}C dimensions (70 complex points).

3D BT-HNCO experiment

The Bruker b_trhncogp3d pulse program was used. The nitrogen chemical shift evolution was extended from the constant time to the semi-constant time mode.

Experimental parameters used. The 3D BT-HNCO experiment was recorded in 3 hours with 4 scans per increment, an inter-scan delay of 0.2 s, spectral widths of 13300 (aq) × 2600 (^{15}N) × 1800 (^{13}C') Hz with 1024 complex points in the acquisition dimension and 1575 hypercomplex points randomly distributed in the indirectly detected dimensions (data points density of 20.0 %). The maximum evolution times in the indirectly detected dimensions were set to 40.4 ms for the ^{15}N dimension (105 complex points) and to 41.7 ms for the ^{13}C dimension (75 complex points).

Article 5. Amino acid recognition for automatic resonance assignment of intrinsically disordered proteins

Alessandro Piai[1], Leonardo Gonnelli[1], Isabella C. Felli[1], Roberta Pierattelli[1], Krzysztof Kazimierczuk[2], Wiktor Koźmiński[3], Anna Zawadzka-Kazimierczuk[3]

[1] CERM and Department of Chemistry "Ugo Schiff", University of Florence, 50019, Via Luigi Sacconi 6, Sesto Fiorentino, Florence, Italy.

[2] Centre of New Technologies, University of Warsaw, Banacha 2C, 02-097 Warsaw, Poland.

[3] Faculty of Chemistry, Biological and Chemical Research Centre, University of Warsaw, Żwirki i Wigury 101, 02-089 Warsaw, Poland.

Manuscript submitted

Abstract

Resonance assignment is a prerequisite for almost any NMR-based study of proteins, but in some cases it can be very challenging due to the nature of the protein under investigation. This is the case of intrinsically disordered proteins, whose NMR spectra suffer from low chemical shifts dispersion and generally low resolution. For these systems sequence specific assignment is very time-consuming and thus the perspective to exploit automatic strategies for their assignment is very attractive. In this article we present a new version of the automatic assignment program TSAR dedicated to intrinsically disordered proteins. In particular, we demonstrate how the automatic procedure can be improved by the incorporation of methods for amino acid recognition and information on chemical shifts of selected amino acids. The approach was tested on a 140-amino acid long intrinsically disordered protein, showing remarkably good results.

Keywords

intrinsically disordered proteins, automatic assignment, amino acid-selective experiments, ^{13}C direct-detection NMR, high-dimensional NMR experiment, non-uniform sampling, compressed sensing, sparse multidimensional Fourier transform

1. Introduction

Nuclear magnetic resonance (NMR) is the most powerful method for studying intrinsically disordered proteins (IDPs). It allows to obtain a variety of information, including structural propensities, dynamics and interactions with other molecules. Nonetheless, IDPs are rather difficult objects to study with NMR. The high mobility of the polypeptide chain results with exceptionally narrow ranges of chemical shifts. The effect is further amplified by the high incidence of sequential repeats; stretches of three or four residues of the same type are often present in IDPs' sequences. Also the high abundance of disorder-promoting amino acids and underrepresentation of others (Dunker et al. 2008) contributes to the low chemical shifts dispersion. The usually high content of prolines, which are not observable in experiments employing amide proton detection, breaks chains of sequential connectivities. Combination of all the above factors makes the complete sequence-specific resonance assignment of IDPs a challenging task.

Employment of high-dimensional (≥4D) experiments (Kazimierczuk et al. 2013; Nowakowski et al. 2015) allows to spread cross-peaks over a larger spectral space and thus better resolve them. However, such techniques require the use of non-uniform sampling (NUS) to accelerate the data acquisition. Several methods of processing NUS data have been proposed (Mobli and Hoch 2008; Coggins et al. 2010; Orekhov and Jaravine 2011; Freeman and Kupče 2012; Hiller and Wider 2012; Kazimierczuk et al. 2012; Holland and Gladden 2014) which allowed to develop many experiments of high dimensionality, applied with success to IDPs. Many efforts were put also into the development of carbon direct-detected techniques (Bermel et al. 2009; Felli and Pierattelli 2014) which are invaluable when the fraction of prolines is high or amide protons undergo fast chemical

exchange (Gil et al. 2013). Their benefit originates also from the superior chemical shift dispersion of carbonyl carbons direct-detected experiments compared to that of amide protons-detected experiments (Brutscher et al. 2015). The two above strategies can be also combined and several ^{13}C-detected approaches for high-dimensionality have been proposed (Nováček et al. 2011; Nováček et al. 2012; Bermel et al. 2012b; Nováček et al. 2013; Bermel et al. 2013; Dziekański et al. 2015).

During the resonance assignment process, the amino acid types of at least some of the residues have to be recognized to map the chains of sequentially-linked residues onto the protein sequence. Given a single uniformly-labeled protein sample (e.g. without selective isotope labeling), three main methods are used for this purpose: (i) by using statistical values of chemical shifts for various nuclei of different amino acids, (ii) by using some topological information, and (iii) by acquiring amino acid type selective experiments. The (i) method is widely used, as it usually does not require any additional experiments. C^{β} and H^{β} chemical shifts, which are particularly useful for the purpose, are often recorded within a set of assignment experiments. The statistical values are available from the Biological Magnetic Resonance Bank (BMRB) data base (Ulrich et al. 2008), where average C^{β} and H^{β} chemical shifts for each amino acid are calculated based on at least a few thousands of chemical shifts. For IDPs an additional statistics are available (Tamiola et al. 2010) which considers not only the residue type, but also the residue's closest neighbors (*i-1* and *i+1*) and thus is more reliable. In the (ii) method, the presence of some nuclei chemical shift excludes one or more amino acids, for instance the presence of C^{β} chemical shift precludes glycine, the presence of H^{N} chemical shift excludes proline, the presence of two different H^{β}s excludes alanine, isoleucine, threonine and valine.

The amino acid selective experiments ((iii) method) was originally proposed by Dötsch and collaborators (Dötsch et al. 1996a; Dötsch and Wagner 1996; Dötsch et al. 1996c; Dötsch et al. 1996b). This approach was based on the triple-resonance CBCA(CO)NH pulse sequence (Grzesiek and Bax 1992), modified to acquire signal just for some topology-selected amino acid types. The resulting 2D ^{1}H-^{15}N-HSQC-like spectra contained just the resonances originating from the desired amino acid residues. This concept was then extensively developed, providing several new selections criteria (Feng et al. 1996; Rios et al. 1996; Schubert et al. 1999; Schubert et al. 2000; Schubert et al. 2001c; Schubert et al. 2001b; Schubert et al. 2001a; Schubert et al. 2005; Barnwal et al. 2008) and resulting in many different strategies, e.g. the multiplicity selective in-phase coherence transfer (MUSIC), developed by Schubert and coworkers. For selection, several types of pulse sequence components can be employed, including multiple quantum filters (for $^{13}CH_n$ or $^{15}NH_n$), band-selective pulses on ^{13}C and/or ^{15}N (for specific nuclei excitation), delay tuning (for choosing the desired coherence transfer pathway) and setting a proper number of coherence transfer steps (for choosing side-chains of the desired length). Instead of selecting specific correlations, editing can be implemented and combined with the idea of Hadamard encoding to accelerate the data collection (Lescop et al. 2008; Pantoja-Uceda and Santoro 2008; Lescop and Brutscher 2009; Feuerstein et al. 2012; Pantoja-Uceda and Santoro 2012). Amino acid selection was finally incorporated also into ^{13}C-detected experiments (Bertini et al. 2006; Pantoja-Uceda and Santoro 2011; Chakraborty et al. 2012; Jaipuria et al. 2012; Bermel et al. 2012a).

In the current study we show how different amino acid recognition methods can be exploited in automatic resonance assignment and how the assignment completeness and reliability can

benefit from that type of information. We present an improved version of the TSAR (Tool for SMFT-based Assignment of Resonances) program (Zawadzka-Kazimierczuk et al. 2012), designed for automatic resonance assignment using experiments of high dimensionality (≥4D), to include the information provided by ^{13}C-detected amino acid-selective experiments (Bermel et al. 2012a). Additionally, the IDPs' chemical shifts statistics (Tamiola et al. 2010) were incorporated for more efficient chain mapping. Finally, we present a small modification of the 4D HCBCACON pulse sequence (Bermel et al. 2012b) in which peaks of residues possessing a single aliphatic C^{γ} carbon are of the opposite sign with respect to that of all other residues. The approach was tested on α-synuclein protein, using both ^{1}H-detected (Piai et al. 2014) and ^{13}C-detected experiments (Bermel et al. 2012b; Bermel et al. 2013). To accelerate the data collection, all spectra were acquired using NUS and making use of all the recently developed sampling and processing strategies (Kazimierczuk et al. 2009; Kazimierczuk and Orekhov 2011). Data from the high-dimensional experiments were processed using sparse multidimensional Fourier transform (SMFT) algorithm (Kazimierczuk et al. 2009), whereas data from the 2D amino acid-selective experiments were processed using compressed sensing (CS) algorithm (Kazimierczuk and Orekhov 2011).

2. Materials and Methods

All NMR experiments were performed at 16.4 T on a Bruker Avance spectrometer operating at 700.06 MHz ^{1}H, 176.03 MHz ^{13}C and 70.94 MHz ^{15}N frequencies, equipped with a ^{13}C cryogenically cooled probehead optimized for ^{13}C-direct detection. A sample of 1.0 mM uniformly ^{13}C, ^{15}N labeled human α-synuclein in 20 mM phosphate buffer at pH 6.5 was prepared as previously described (Huang et al. 2005). EDTA and NaCl were added to reach the final concentration of 0.5 mM and 200 mM, respectively, and 10% D_2O was added for the lock. All experiments were acquired at 285.5 K. Parameters specific to each amino acid selective experiment are reported in the original publication (Bermel et al. 2012a), while those related to the γ-selective-HCBCACON experiment are given in the caption of the figure describing the pulse sequence, reported in the Supplementary Material (Figure S1). For ^{13}C band-selective $\pi/2$ and π flip angle pulses Q5 (or time reversed Q5) and Q3 shapes (Emsley and Bodenhausen 1992) of durations of 300 and 220 μs, respectively, were used, except for the π pulses that should be band-selective on the C^{α} region (Q3, 860 μs) and for the adiabatic π pulse to invert both C′ and C^{α} (smoothed Chirp 500 μs, 25% smoothing, 80 kHz sweep, 11.3 kHz RF field strength (Bohlen and Bodenhausen 1993)). The ^{13}C band selective pulses on C^{ali}, C^{α}, and C′ were given at the center of each region, respectively, and the adiabatic pulse was adjusted to cover the entire ^{13}C region. Decoupling of ^{1}H and ^{15}N was achieved with waltz16 (Shaka et al. 1983) (1.7 kHz) and garp4 (Shaka et al. 1985) (1.0 kHz) sequences, respectively. Each experiment was acquired in a pseudo 2D mode, with States method applied in all indirect dimensions to achieve quadrature detection. All experiments employ the IPAP approach (Bermel et al. 2008) to remove splitting in the direct acquisition dimension caused by the homonuclear C^{α}-C′ couplings.

The experimental parameters used for the acquisition of all experiments proposed here are reported in Table 1. All the experiments were performed using on-grid non-uniform sampling. "Poisson disk" sampling scheme (Kazimierczuk et al. 2008) was chosen to generate the time schedules with *RSPack* software. All the spectra were acquired using *Bruker TopSpin 1.3* software. The experimental data were converted with *nmrPipe* (Delaglio et al. 1995) and then processed

either using *Compressed Sensing* (Kazimierczuk and Orekhov 2011) IRLS algorithm with iteratively changed lp norm (p: 1 -> 0) with 30 iterations (2D data sets) or the *Sparse Multidimensional Fourier Transform* (SMFT) (Kazimierczuk et al. 2009) (4D and 5D data sets), implemented in *Reduced* program. Finally, the program *Sparky* (Goddard and Kneller 2002) was used for visualization of spectra and *TSAR* (Zawadzka-Kazimierczuk et al. 2012) was used for assignment of resonances. The programs *RSPack*, *Reduced* and *TSAR* are available at http://nmr.cent3.uw.edu.pl/software (free of charge for academic users).

3. Results and discussion

Methods

The program TSAR (Zawadzka-Kazimierczuk et al. 2012) was developed to analyze the data from experiments of high dimensionality processed using sparse multidimensional Fourier transform (SMFT) algorithm (Kazimierczuk et al. 2009). In this method, instead of computing the full multidimensional spectrum, just a set of 2D cross-sections are calculated. This can be done using the peak list of a lower-dimensional *basis spectrum*, which shares some of the dimensions with the high-dimensional spectrum. For each *basis peak*, usually a single cross-section can be calculated. Depending on the experiment type, in each cross-section there is one or more peaks; if the experiment provides sequential connectivities, in the cross-sections originating from adjacent residues some peaks are redundant. Importantly, if several multidimensional spectra have to be analyzed together, they all should be processed using the same basis peak list. The strategy for resonance assignment using this kind of input relies on the comparison of peaks positions, creating chains of cross-sections. Recognition of amino acid type of some of the residues allows to map the cross-section chains onto the protein sequence, which completes the assignment. Up to now, TSAR employed just two of the three methods described in the Introduction for amino acid identification: (i) BMRB chemical shift statistics for C^{β}, H^{β}, C^{α}, H^{α} nuclei and (ii) topological information.

The main goal of the present work was the implementation of the (iii) method, i.e. amino acid-selective experiments. Up to now only the change of the sign of peak intensities in the absence of C^{β} nucleus for glycine residues, which occurs in experiments where C^{α} transverse magnetization evolves for c.a. $1/J_{C\alpha\text{-}C\beta}$, was exploited. Aiming to achieve automated assignment of highly overlapping IDP resonances, we decided to use 2D spectra with N and CO dimensions, which provide superior resolution and allow to detect prolines. Two types of such spectra are available (Bermel et al. 2012a): CACON- and CANCO-based. In the 2D (CA)CON amino acid-selective spectrum, a CO_{i-1}-N_i peak shows up if residue *i-1* is of the specified type. In the 2D (CA)NCO amino acid-selective spectrum, a CO_{i-1}-N_i peak appears if residue *i-1* or residue *i* is of that type. The comparison of the two spectra then allows to discriminate CO_{i-1}-N_i peaks related to *i-1* or *i* residues (Figure 1). The idea of using such experiments in parallel with SMFT-processed high-dimensional data is to find the basis peaks corresponding to the selected amino acids. This can be done if the dimensions of the amino acid-selective experiments, in our case amide nitrogen and carbonyl carbon dimensions, are also present in the basis spectrum. Then, the basis peak list can be plotted on the amino acid-selective spectrum and the basis peaks corresponding to the given

amino acid can be easily picked (Figure 2). Information on these basis peaks numbers is then given to the TSAR program to support the assignment process (see Supplementary Material for the format of TSAR input files). Of course, it may happen that two basis peaks show up at the position of the amino acid-selective spectrum peak if they are overlapping or the amino acid-selective spectrum is not resolved enough. TSAR is able to cope with such a situation, but of course it makes the task more difficult. As said above, the only requirement to combine SMFT-processed data with the amino acid-selective data is to have CO and N dimensions in the basis spectrum. Therefore, although (CA)CON and (CA)NCO experiments exploit carbon detection, they can be combined with carbon-detected (e.g. with 3D CACON basis spectrum) but also with proton-detected (e.g. with 3D HNCO basis spectrum) experiments for resonance assignment, rendering their use more general.

In this work, we also exploited another idea of getting amino acid-related information. We modified the 4D HCBCACON pulse sequence (Bermel et al. 2012b) so that cross-peaks' signs depend on topological properties of the residue. The delay for C^{β} scalar couplings evolution was increased up to 21.0 ms (see Supplementary Material), which allowed to keep the efficiency of the coherence transfer (additional relaxation losses can be neglected for IDPs) and, at the same time, reverse peak signs for some residues. If a residue *i* possesses exactly one aliphatic C^{γ} carbon (E, H, K, L, M, P, Q, R, T and W residues), the $H^{\beta}{}_{i}$-$C^{\beta}{}_{i}$ peak results with opposite sign with respect to that of all the other residues. Such zero-one information adds to C^{β} and H^{β} chemical shifts values, which was the only information provided by the previously published 4D HCBCACON experiment (Bermel et al. 2012b), thus improving the performances of the automated assignment. The new pulse sequence also allows to unambiguously distinguish some residues possessing C^{β} and H^{β} chemical shifts which may be not so different: S can be easily discriminated from T, I from L, V from E, H, K, M, P, Q and R. Additionally, the extension of the $C^{\alpha/\beta}$ chemical shift evolution increases the resolution of that dimension, with consequent benefit for the comparison of cross-peak spectral positions performed by TSAR. In the old version, TSAR was prepared for single amino acid recognition by peak sign, which was used only for glycine. The current software version accepts sign change in the presence of a user-defined set of residues.

The last major change that allowed for more efficient amino acid recognition in TSAR was to incorporate the statistical chemical shift values calculated using a set of IDPs (Tamiola et al. 2010). TSAR uses statistical chemical shift values at two stages of operation: (a) recognition of possible amino acids for a single plane (before forming cross-section chains); (b) cross-section chains mapping. During the former stage, the working procedure of TSAR implies that if some chemical shifts exceed the statistical average with four standard deviations for some amino acids, then these amino acids are excluded from the range of possible ones; during the latter stage, if one chain matches in a few sites or a few chains of similar length fit in one site, the deviation from the statistical values (in units of standard deviations) is calculated and if the best score is at least three times smaller than the second best, then the better chain is assigned. In the previous release of TSAR, at both stages the BMRB values were used. In the version of TSAR we present now, at the stage (b) the IDPs-specific values (Tamiola et al. 2010) are used. Such values cannot be employed during the stage (a) when the neighboring residues are not yet known, as these values depend on the preceding and following residue type. Thus, at the stage of amino acid recognition for a single plane, the BMRB values are still used. Nonetheless, also at this stage some IDPs-

oriented changes were made: it was found out that for IDPs it is more optimal to exclude an amino acid if the chemical shift exceeds two (instead of four) standard deviations. Additionally, C^{α} chemical shifts were incorporated for amino-acid recognition, while before only glycine residues could be identified basing exclusively on C^{α} chemical shifts.

Assignment performance

The new methods of amino acid recognition were tested by combining different sets of ^{13}C-detected experiments for amino acid identification with one, two or three 5D either ^{13}C- or ^{1}H-detected experiments assuring sequential links, all processed with SMFT. The TSAR program constructs cross-sections chains whose lengths depend on the quality of the experiment(s) providing sequential correlations. Thus, the data sets containing different combinations of the latter(s) were generated to evaluate the efficiency of amino-acid recognition methods for various chain lengths. Regarding amino acid recognition, in some data sets either the 4D HCBCACON or the 4D γ-selective-HCBCACON experiments yielding C^{β} and H^{β} chemical shifts were employed, while in others the information provided by amino acid-selective experiments was used. The latter consist of ^{13}C-detected 2D (CA)CON- and (CA)NCO-based amino acid selective experiments, including the following selections: A, D, E, FHYW, G, N, Q and S. The experiments' selectivity is reported in Table S1 in the Supplementary Material. In all the spectra, only the peaks of the selected amino acid(s) are present, with few exceptions: in Q-selective 2D (CA)CON and 2D (CA)NCO experiments there is a leakage of N peaks, but in the 2D (CA)NCO experiment they have opposite sign with respect to that of Q peaks (and thus are very easy to recognize); in E-selective 2D (CA)CON experiment, D peaks appear, but with opposite sign than E peaks; in G- and S-selective and 2D (CA)NCO experiments, peaks originating form P residues are present, but in the former experiment they have opposite sign than the other peaks. TSAR was thus trained to handle such cases. Regarding the information completeness, in all the amino acid-selective spectra all expected peaks were retrieved.

High-dimensional ^{13}C-detected experiments such as 5D (H)CACON(CA)CON, 5D HNCACON and 5D (HACA)CONCACON (Bermel et al. 2012b; Bermel et al. 2013) were used to collect sequential correlations; they all feature a 3D (H)CACON as basis spectrum. To measure C^{β} and H^{β} chemical shifts, the 4D HCBCACON (Bermel et al. 2012b) and the new 4D γ-selective-HCBCACON experiments were acquired. Out of these data, three data sets were constructed (Table 2): data set A, containing only the links-yielding 5D (H)CACON(CA)CO experiment; data set B, including the 5D (H)CACON(CA)CON and 5D HNCACON experiments; data set C, comprising the 5D (H)CACON(CA)CON, 5D HNCACON and 5D (HACA)CONCACON experiments. For each data set, four additional subsets were created depending on the type of amino acid information provided: subset (1) does not contain any information on amino acids; subset (2) exploits the 4D HCBCACON experiment; subset (3) uses the 4D γ-selective-HCBCACON experiment which carries extra information in peak signs; subset (4) employs all 2D amino acid selective experiments, both in (CA)CON and (CA)NCO versions.

The results obtained by the TSAR program for different data sets of ^{13}C-detected experiments are gathered in Table 3 and shown in Figure 3. The data analysis shows the effectiveness of the approach. Even using a single experiment yielding the sequential connectivities, almost 80% of correct assignment is achieved when combined with amino acid-selective data. Employing two

linking experiments allows to exceed 83% of correct assignments, with no more than 2.9% of incorrectly assigned peaks. Importantly, the incorrectly assigned cross-sections were parts of short chains (of up to three cross-sections) or the very last cross-section of longer chains. Subsets (1) and (2) allowed to compare the performance of the old and new version of the TSAR program (the data of two other subsets cannot be accepted by the old version). Results demonstrate that the new version of TSAR performs better than the previous one. The strongest effect is observed for data providing relatively short cross-section chains (data set A): the long chains are easily assignable even using simple methods. Incorporation of C^{α} chemical shift statistics for non-glycine residues is reflected in the significant improvement in program's performance for subsets (1). Using γ-selective-HCBCACON experiment (subset 3) instead of the standard HCBCACON experiment (subset 2) does not cause a significant change in the assignment results, although for data set A it significantly reduces the number of incorrect assignments. Exploitation of the amino acid-selective experiments (subset 4) seems to be a reliable alternative to measure C^{β} and H^{β} chemical shifts for amino acid recognition; in the case of shorter cross-section chains (data set A), TSAR performs even better than when using beta chemical shifts. It is noteworthy that the measurement time of a set of 2D amino acid-selective experiments was over twice shorter than experimental time required for the HCBCACON (12 hours compared to 28 hours). Of course, if C^{β} and H^{β} chemical shifts are essential for further studies, the method of choice would be the acquisition of the 4D HCBCACON experiment. However, if only the backbone assignment is desired, the collection of a full set of amino acid-selective 2D spectra is worth considering. However, especially for complex systems, the availability of complementary data contributes to more robust results.

Alternatively, also ^{1}H-detected experiments providing sequential connectivities can be used, e.g. 5D BT-HN(COCAN)CONNH and 5D BT-(H)NCO(CAN)CONNH experiments (Piai et al. 2014), which require a 3D BT-HNCO as basis spectrum. Three data sets featuring such experiments were constructed (Table 4): data set D, containing the 5D BT-HN(COCAN)CONNH experiment; data set E, including the 5D BT-(H)NCO(CAN)CONNH experiment; data set F, gathering both experiments. In these case, for each data set six subsets containing different combinations of amino-acid selection experiments were formed: subset (1), including all selections (A, D, E, FHYW, G, N, Q, S) in both (CA)CON and (CA)NCO versions; subset (2), containing the same selections but only in the (CA)CON version; subset (3), comprising A, E, G, FHYW and Q selections in both (CA)CON and (CA)NCO versions; subset (4), gathering A, E, G, FHYW and Q selections only in the (CA)CON version; subset (5), including A, E and G selections in both (CA)CON and (CA)NCO versions; subset (6), containing A, E and G selections only in (CA)CON version.

The results provided by the TSAR program using these data sets are summarized in Table 5 and Figure 4. Also in this case, the assignment performance is satisfactory: while using all amino acid selection experiments (subsets 1), the percentage of correct assignments exceeds 79 % even using just a single 5D experiment to establish sequential correlations. It is worth mentioning that in no cases the fraction of incorrect assignments exceeds 0.7 %, and that all the incorrectly assigned cross-sections belong to short chains. The data allowed to evaluate the benefits provided by different sets of amino acid-selective spectra: the completeness of the resonance assignment decreases reducing the amount of amino acid-selective data, even if this effect is less and less visible with increasing the length of the cross-section chains. Therefore, the fraction of cross-

sections in short chains is a very important parameter to suggest how many amino acid selective experiments are worth acquiring. Also, the content of short chains suggests whether the experiments should be recorded in the (CA)CON and/or (CA)NCO versions. The more of the formed chains are short, the more different selections should be made (even if only in (CA)CON version). On the contrary, if short chains are just few, it is more beneficial to apply less type of amino acid selections but acquire them in both (CA)CON and (CA)NCO modes. Obviously, while making the selections for a particular sample, one should consider the protein's sequence: it is more valuable to identify the amino acids which are abundant in a given molecule.

4. Conclusions

In this study we have further developed the automatic resonance assignment TSAR program, by improving its assignment efficiency and reliability for intrinsically disordered proteins. The implemented changes were tailored to exploit peaks' signs depending on the originating residue, to employ the chemical shift statistics established especially for IDPs and to analyze the data from amino acid-selective experiments. Besides improving the TSAR program, we have also proposed a modification of an existing 4D HCBCACON experiment, so that information on the amino acid type is coded in the peak sign. The methods were tested on α-synuclein, a 140-amino acids-long IDP, for both proton- and carbon-detected experiments. The analysis shows that incorporating the above methods significantly improves the assignment results, especially for data sets in which the cross-sections chains are short. Amino acid-selective experiments, which are relatively quick when performed using non-uniform sampling, can be used as an alternative to amino acid recognition based on chemical shift analysis. The proposed methods facilitate the resonance assignment of IDPs and makes it more reliable and complete.

5. Acknowledgment

This work was supported by the grant number IP2012 062772, funded by Polish Ministry of Science and Higher Education for years 2013–2016, by the POMOST program of the Foundation for Polish Science and by the European Commission Projects IDPbyNMR (Contract No. 264257), BioNMR (Contract No. 261863) and INSTRUCT (Contract No. 211252).

Table 1 Experimental parameters used for the acquisition of the NMR experiments.

	Spectral widths & maximal evolution times				N° of scans	Interscan delay (s)	N° of complex points (aq)	N° of hypercomplex points	Duration of the experiment	Relative data points density (%)
2D A-sel (CA)CON	8800 Hz (^{13}C′)	2550 Hz (^{15}N) 100.0 ms			12	1.5	512	40	55 min	15.6
2D A-sel (CA)NCO	8800 Hz (^{13}C′)	2550 Hz (^{15}N) 31.8 ms			12	1.5	512	40	55 min	48.8
2D D-sel (CA)CON	8800 Hz (^{13}C′)	2550 Hz (^{15}N) 100.0 ms			8	1.5	512	40	40 min	15.6
2D D-sel (CA)NCO	8800 Hz (^{13}C′)	2550 Hz (^{15}N) 31.8 ms			8	1.5	512	40	40 min	48.8
2D E-sel (CA)CON	8800 Hz (^{13}C′)	2550 Hz (^{15}N) 100.0 ms			8	1.5	512	40	40 min	15.6
2D E-sel (CA)NCO	8800 Hz (^{13}C′)	2550 Hz (^{15}N) 31.8 ms			8	1.5	512	40	40 min	48.8
2D FHYW-sel (CA)CON	8800 Hz (^{13}C′)	2550 Hz (^{15}N) 100.0 ms			16	1.5	512	40	1 h, 15 min	15.6
2D FHYW-sel (CA)NCO	8800 Hz (^{13}C′)	2550 Hz (^{15}N) 31.8 ms			16	1.5	512	40	1 h, 15 min	48.8
2D G-sel (CA)CON	8800 Hz (^{13}C′)	2550 Hz (^{15}N) 100.0 ms			8	1.5	512	40	35 min	15.6
2D G-sel (CA)NCO	8800 Hz (^{13}C′)	2550 Hz (^{15}N) 31.8 ms			8	1.5	512	40	40 min	48.8
2D N-sel (CA)CON	8800 Hz (^{13}C′)	2550 Hz (^{15}N) 100.0 ms			8	1.5	512	40	40 min	15.6
2D N-sel (CA)NCO	8800 Hz (^{13}C′)	2550 Hz (^{15}N) 31.8 ms			8	1.5	512	40	40 min	48.8
2D Q-sel (CA)CON	8800 Hz (^{13}C′)	2550 Hz (^{15}N) 100.0 ms			8	1.5	512	40	40 min	15.6

2D Q-sel (CA)NCO	8800 Hz (^{13}C′)	2550 Hz (^{15}N) 31.8 ms			8	1.5	512	40	40 min	48.8
2D S-sel (CA)CON	8800 Hz (^{13}C′)	2550 Hz (^{15}N) 49.8 ms			8	1.5	512	32	30 min	25.0
2D S-sel (CA)NCO	8800 Hz (^{13}C′)	2550 Hz (^{15}N) 31.8 ms			8	1.5	512	40	40 min	48.8
4D γ-selective-HCBCACON	8800 Hz (^{13}C′)	2550 Hz (^{15}N) 60.4 ms	12500 Hz ($^{13}C^{\alpha,\beta}$) 20.5 ms	5000 Hz (^{1}H) 15.0 ms	4	1.1	512	1540	28 hours	0.051
Experimental parameters for the 3D (H)CACON, 4D HCBCACON, 5D (HACA)CONCACON, 5D ($H^{N\text{-}flip}$N)CONCACON, 5D (HA)CACON(CA)CON, 3D BT-HNCO, 5D BT-(H)NCO(CAN)CONNH and 5D BT-HN(COCAN)CONNH experiments are available in the original publications (Bermel et al. 2012b; Bermel et al. 2013; Piai et al. 2014).										

Table 2 Data sets of ^{13}C-detected experiments.

Data set	Basis experiment and sequential link-providing experiment(s)	Percentage of cross-sections in chains…			Subset	Experiment(s) providing information on amino acids	Total exp. time, hours
		long (≥8)	medium (4-7)	short (1-3)			
A	3D CACON 5D (H)CACON(CA)CON	11.4	42.1	46.4	(1)	none	67
					(2)	4D HCBCACON	95
					(3)	4D γ-selective-HCBCACON	95
					(4)	all 2D amino acid selective experiments	79
B	3D CACON 5D (H)CACON(CA)CON 5D HNCACON	62.9	16.4	20.7	(1)	none	81
					(2)	4D HCBCACON	109
					(3)	4D γ-selective-HCBCACON	109
					(4)	all 2D amino acid selective experiments	93
C	3D CACON 5D (H)CACON(CA)CON 5D HNCACON 5D (HCA)CONCACON	79.3	7.9	12.9	(1)	none	152
					(2)	4D HCBCACON	180
					(3)	4D γ-selective-HCBCACON	180
					(4)	all 2D amino acid selective experiments	164

Table 3 Automatic assignment results for ^{13}C-detected experiments.

Subset \ Data set	Percentage of correctly / incorrectly assigned residues		
	A	B	C
(1)	42.4 / 0.0 (11.5 / 0.0) [a]	83.5 / 1.4 (53.2 / 5.0) [a]	89.2 / 0.0 (87.8 / 2.2) [a]
(2)	68.3 / 10.1 (56.8 / 7.2) [a]	87.8 / 2.9 (88.5 / 2.9) [a]	89.2 / 0.0 (89.2 / 0.7) [a]
(3)	64.7 / 2.9	89.2 / 2.9	89.2 / 0.0
(4)	78.4 / 0.7	87.1 / 2.9	89.2 / 0.0

[a] For the data sets accepted by the old version of the TSAR program, the results obtained using that version were shown in parenthesis.

Table 4 Data sets of ^{1}H-detected experiments.

Data set	Basis experiment and sequential link-providing experiment(s)	Percentage of cross-sections in chains…			Subset	Experiment(s) providing information on amino acids	Total exp. time, hours
		long (≥8)	medium (4-7)	short (1-3)			
D	3D HNCO 5D BT-HN(COCAN)CONNH	32.3	28.6	39.1	(1)	all selections, (CA)CON and (CA)NCO	26
					(2)	all selections, only (CA)CON	20
					(3)	A, E, G, FHYW, Q selections, (CA)CON and (CA)NCO	22.25
					(4)	A, E, G, FHYW, Q selections, only (CA)CON	18
					(5)	A, E, G selections, (CA)CON and (CA)NCO	18.5
					(6)	A, E, G selections, only (CA)CON	16.25
E	3D HNCO 5D BT-(H)NCO(CAN)CONNH	61.7	20.3	18.0	(1)	all selections, (CA)CON and (CA)NCO	26
					(2)	all selections, only (CA)CON	20
					(3)	A, E, G, FHYW, Q selections, (CA)CON and (CA)NCO	22.25
					(4)	A, E, G, FHYW, Q selections, only	18

				(CA)CON	
			(5)	A, E, G selections, (CA)CON and (CA)NCO	18.5
			(6)	A, E, G selections, only (CA)CON	16.25
F	3D HNCO 5D BT-HN(COCAN)CONNH 5D BT-(H)NCO(CAN)CONNH 74.4 9.8	15.8	(1)	all selections, (CA)CON and (CA)NCO	37
			(2)	all selections, only (CA)CON	31
			(3)	A, E, G, FHYW, Q selections, (CA)CON and (CA)NCO	33.25
			(4)	A, E, G, FHYW, Q selections, only (CA)CON	29
			(5)	A, E, G selections, (CA)CON and (CA)NCO	29.5
			(6)	A, E, G selections, only (CA)CON	27.25

Table 5 Automatic assignment results for ^{1}H-detected experiments.

	Percentage of correctly / incorrectly assigned residues		
Subset \ Data set	D	E	F
(1)	79.9 / 0.7	88.8 / 0.0	91.8 / 0.0
(2)	76.1 / 0.7	82.8 / 0.0	91.8 / 1.5
(3)	73.9 / 0.7	86.6 / 0.0	90.3 / 0.0
(4)	55.2 / 0.7	80.6 / 0.0	90.3 / 1.5
(5)	52.2 / 0.0	67.2 / 0.0	90.3 / 0.0
(6)	39.6 / 0.0	64.2 / 0.0	90.3 / 0.0

Figure 1 Analysis of amino-acid selective experiments. As an example, asparagine-selective 2D (CA)CON (left) and (CA)NCO (right) spectra are reported. The basis peak list (black dots) is plotted on top of both spectra. For each CO_{i-1}-N_i cross-peak, the comparison of the two spectra allows to distinguish if asparagine residues are at positions *i-1* or *i*.

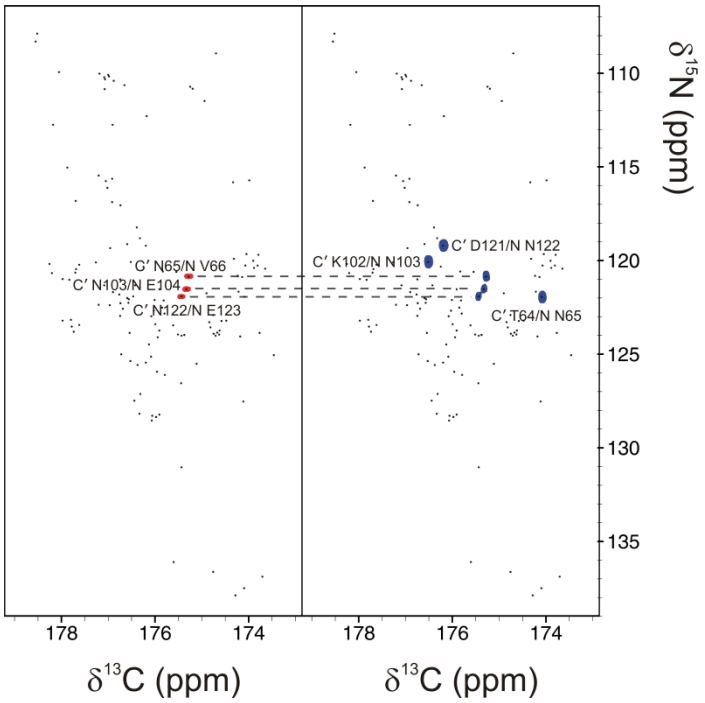


Figure 2 Alanine-selective 2D (CA)CON spectrum (red) with the basis peak list superimposed (black dots). On the left, the close up view of the spectral region inside the blue square is reported. In the example, during the automatic assignment stage basis peaks on top of the NMR signals are known by TSAR to be related to alanine residues. To make the picture more clear, the number of each basis peak has been shown only in the spectral region extracted on the left.

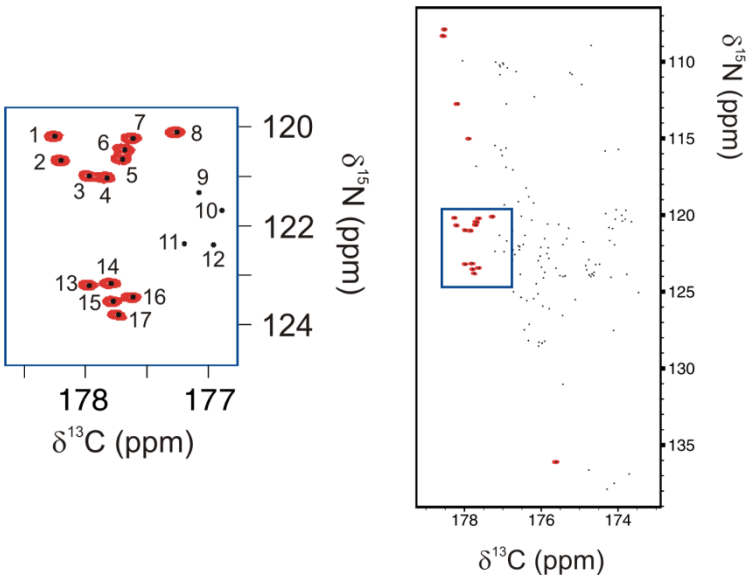


Figure 3 Automatic assignment results. For each data set, the percentage of correct (data set A – blue, data set B – green, data set C – purple) and incorrect (red) assignment is shown. For data sets accepted by the old version of the TSAR program (subsets i and ii), also the results obtained using that version are shown (light colors).

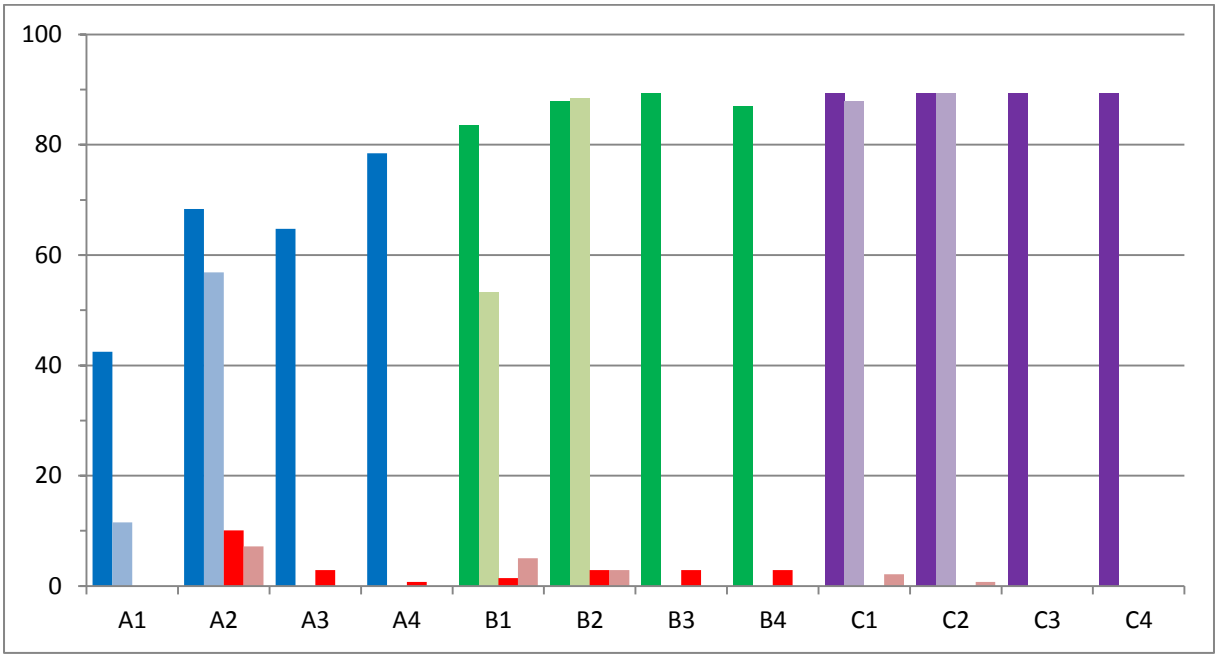
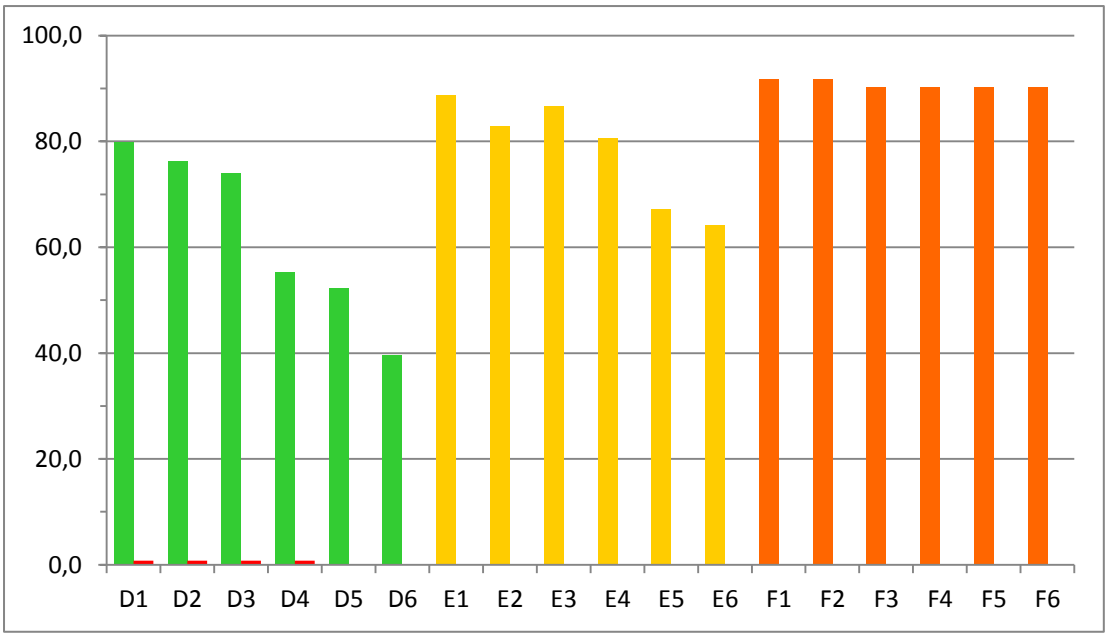


Figure 4 Automatic assignment results. For each data set, the percentage of correct (data set D – green, data set E – yellow, data set F – orange) and incorrect (red) assignments is shown.



References

Barnwal RP, Rout AK, Atreya HS, Chary KVR (2008) Identification of C-terminal neighbours of amino acid residues without an aliphatic 13Cγ as an aid to NMR assignments in proteins. J Biomol NMR 41:191–197. doi: 10.1007/s10858-008-9254-2

Bermel W, Bertini I, Chill J, Felli IC, Haba N, Kumar M V V, Pierattelli R (2012a) Exclusively heteronuclear ^{13}C-detected amino-acid-selective NMR experiments for the study of intrinsically disordered proteins (IDPs). Chembiochem 13:2425–32. doi: 10.1002/cbic.201200447

Bermel W, Bertini I, Felli IC, Gonnelli L, Koźmiński W, Piai A, Pierattelli R, Stanek J (2012b) Speeding up sequence specific assignment of IDPs. J Biomol NMR 53:293–301. doi: 10.1007/s10858-012-9639-0

Bermel W, Bertini I, Felli IC, Pierattelli R (2009) Speeding up ^{13}C direct detection biomolecular NMR spectroscopy. J Am Chem Soc 131:15339–45. doi: 10.1021/ja9058525

Bermel W, Felli IC, Gonnelli L, Koźmiński W, Piai A, Pierattelli R, Zawadzka-Kazimierczuk A (2013) High-dimensionality ^{13}C direct-detected NMR experiments for the automatic assignment of intrinsically disordered proteins. J Biomol NMR 57:353-361. doi: 10.1007/s10858-013-9793-z

Bermel W, Felli IC, Kummerle R, Pierattelli R (2008) ^{13}C Direct-Detection Biomolecular NMR. Concepts Magn Reson Part A 32A:183–200. doi: 10.1002/cmr.a

Bertini I, Felli IC, Gonnelli L, Pierattelli R, Spyranti Z, Spyroulias GA (2006) Mapping protein-protein interaction by ^{13}C-detected heteronuclear NMR spectroscopy. J Biomol NMR 36:111–122. doi: 10.1007/s10858-006-9068-z

Bohlen JM, Bodenhausen G (1993) Experimental aspects of chirp NMR spectroscopy. J Magn Reson Ser A 102:293–301. doi: 10.1006/jmra.1993.1107

Brutscher B, Felli IC, Gil-Caballero S, Hošek T, Kümmerle R, Piai A, Pierattelli R, Sólyom Z (2015) NMR methods for the study of instrinsically disordered proteins structure, dynamics, and interactions: General overview and practical guidelines. Adv Exp Med Biol 870:49–122. doi: 10.1007/978-3-319-20164-1_3

Chakraborty S, Susmitha a. L, Hosur R V. (2012) Selective lighting up of segments around Gly, Ala and Ser/Thr in proteins. Magn Reson Chem 50:587–591. doi: 10.1002/mrc.3843

Coggins BE, Venters R a, Zhou P (2010) Radial sampling for fast NMR: Concepts and practices over three decades. Prog Nucl Magn Reson Spectrosc 57:381–419. doi: 10.1016/j.pnmrs.2010.07.001

Delaglio F, Grzesiek S, Vuister G, Zhu G, Pfeifer J, Bax A (1995) NMRPipe: A multidimensional spectral processing system based on UNIX pipes. J Biomol NMR. doi: 10.1007/BF00197809

Dötsch V, Matsuo H, Wagner G (1996a) Amino-acid-type identification for deuterated proteins with a beta-carbon-edited HNCOCACB experiment. J Magn Reson B 112:95–100. doi: 10.1006/jmrb.1996.0117

Dötsch V, Oswald RE, Wagner G (1996b) Selective identification of threonine, valine, and isoleucine sequential connectivities with a TVI-CBCACONH experiment. J Magn Reson B 110:304–308. doi: 10.1006/jmrb.1996.0047

Dötsch V, Oswald RE, Wagner G (1996c) Amino-acid-type-selective triple-resonance experiments. J Magn Reson B 110:107–11.

Dötsch V, Wagner G (1996) Editing for amino-acid type in CBCACONH experiments based on the 13C beta-13C gamma coupling. J Magn Reson B 111:310–313. doi: 10.1006/jmrb.1996.0100

Dunker AK, Oldfield CJ, Meng J, Romero P, Yang JY, Chen JW, Vacic V, Obradovic Z, Uversky VN (2008) The unfoldomics decade: an update on intrinsically disordered proteins. BMC Genomics 9 Suppl 2:S1. doi: 10.1186/1471-2164-9-S2-S1

Dziekański P, Grudziąż K, Jarvoll P, Koźmiński W, Zawadzka-Kazimierczuk A (2015) ^{13}C-detected NMR experiments for automatic resonance assignment of IDPs and multiple-fixing SMFT processing. J Biomol NMR 62:179-190. doi: 10.1007/s10858-015-9932-9

Emsley L, Bodenhausen G (1992) Optimization of shaped selective pulses for NMR using a quaternion description of their overall propagators. J Magn Reson 97:135–148. doi: 10.1016/0022-2364(92)90242-Y

Felli IC, Pierattelli R (2014) Novel methods based on (13)C detection to study intrinsically disordered proteins. J Magn Reson 241:115–25. doi: 10.1016/j.jmr.2013.10.020

Feng W, Rios CB, Montelione GT (1996) Phase labeling of C-H and C-C spin-system topologies: Application in PFG-HACANH and PFG-HACA(CO)NH triple-resonance experiments for determining backbone resonance assignments in proteins. J Biomol NMR 8:98–104.

Feuerstein S, Plevin MJ, Willbold D, Brutscher B (2012) IHADAMAC: A complementary tool for sequential resonance assignment of globular and highly disordered proteins. J Magn Reson 214:329–334. doi: 10.1016/j.jmr.2011.10.019

Freeman R, Kupče E (2012) Concepts in projection-reconstruction. Top Curr Chem 316:1–20. doi: 10.1007/128

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 52:11808–12. doi: 10.1002/anie.201304272

Goddard TD, Kneller DG (2002) Sparky 3, University of California, San Francisco.

Grzesiek S, Bax A (1992) Correlating backbone amide and side chain resonances in larger proteins by multiple relayed triple resonance NMR. J Am Chem Soc 114:6291–6293. doi: 10.1021/ja00042a003

Hiller S, Wider G (2012) Automated projection spectroscopy and its applications. Top Curr Chem 316:21–47. doi: 10.1007/128

Holland DJ, Gladden LF (2014) Less is more: How compressed sensing is transforming metrology in chemistry. Angew Chemie Int Ed 53:13330–13340. doi: 10.1002/anie.201400535

Huang C, Ren G, Zhou H, Wang C (2005) A new method for purification of recombinant human alpha-synuclein in Escherichia coli. Protein Expr Purif 42:173–7. doi: 10.1016/j.pep.2005.02.014

Jaipuria G, Lobo NP, Shet D, Atreya HS (2012) High resolution methyl selective ^{13}C-NMR of proteins in solution and solid state. J Biomol NMR 54:33–42. doi: 10.1007/s10858-012-9647-0

Kazimierczuk K, Misiak M, Stanek J, Zawadzka-Kazimierczuk A, Koźmiński W (2012) Generalized Fourier Transform for Non-Uniform Sampled Data. Top Curr Chem 316:79–124. doi: 10.1007/128

Kazimierczuk K, Orekhov VY (2011) Accelerated NMR spectroscopy by using compressed sensing. Angew Chem Int Ed Engl 50:5556–9. doi: 10.1002/anie.201100370

Kazimierczuk K, Stanek J, Zawadzka-Kazimierczuk A, Koźmiński W (2013) High-dimensional NMR spectra for structural studies of biomolecules. ChemPhysChem 14:3015–25. doi: 10.1002/cphc.201300277

Kazimierczuk K, Zawadzka A, Koźmiński W (2009) Narrow peaks and high dimensionalities: Exploiting the advantages of random sampling. J Magn Reson 197:219–28. doi: 10.1016/j.jmr.2009.01.003

Kazimierczuk K, Zawadzka A, Koźmiński W (2008) Optimization of random time domain sampling in multidimensional NMR. J Magn Reson 192:123–30. doi: 10.1016/j.jmr.2008.02.003

Lescop E, Brutscher B (2009) Highly automated protein backbone resonance assignment within a few hours: The “BATCH” strategy and software package. J Biomol NMR 44:43–57. doi: 10.1007/s10858-009-9314-2

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. doi: 10.1021/ja800914h

Mobli M, Hoch JC (2008) Maximum entropy spectral reconstruction of non-uniformly sampled data. Concepts Magn Reson Part A Bridg Educ Res 32A:436–448. doi: 10.1002/cmr.a.20126

Nováček J, Haba NY, Chill JH, Žídek L, Sklenář V (2012) 4D non-uniformly sampled HCBCACON and ^{1}J(NCα)-selective HCBCANCO experiments for the sequential assignment and chemical shift analysis of intrinsically disordered proteins. J Biomol NMR 53:139–48. doi: 10.1007/s10858-012-9631-8

Nováček J, Janda L, Dopitová R, Žídek L, Sklenář V (2013) Efficient protocol for backbone and side-chain assignments of large, intrinsically disordered proteins: Transient secondary structure analysis of 49.2 kDa microtubule associated protein 2c. J Biomol NMR 56:291–301. doi: 10.1007/s10858-013-9761-7

Nováček J, Zawadzka-Kazimierczuk A, Papoušková V, Žídek L, Sanderová H, Krásný L, Koźmiński W, Sklenář V (2011) 5D ^{13}C-detected experiments for backbone assignment of unstructured proteins with a very low signal dispersion. J Biomol NMR 50:1–11. doi: 10.1007/s10858-011-9496-2

Nowakowski M, Saxena S, Stanek J, Żerko S, Koźmiński W (2015) Applications of high dimensionality experiments to biomolecular NMR. Prog Nucl Magn Reson Spectrosc 90-91:49–73. doi: 10.1016/j.pnmrs.2015.07.001

Orekhov VY, Jaravine V a (2011) Analysis of non-uniformly sampled spectra with multi-dimensional decomposition. Prog Nucl Magn Reson Spectrosc 59:271–92. doi: 10.1016/j.pnmrs.2011.02.002

Pantoja-Uceda D, Santoro J (2008) Amino acid type identification in NMR spectra of proteins via β- and γ-carbon edited experiments. J Magn Reson 195:187–195. doi: 10.1016/j.jmr.2008.09.010

Pantoja-Uceda D, Santoro J (2012) New amino acid residue type identification experiments valid for protonated and deuterated proteins. J Biomol NMR 54:145–53. doi: 10.1007/s10858-012-9665-y

Pantoja-Uceda D, Santoro J (2011) Selective observation of Asp and Glu resonances in ^{13}CO detected experiments. Magn Reson Chem 49:558–561. doi: 10.1002/mrc.2780

Piai A, Hošek T, Gonnelli L, Zawadzka-Kazimierczuk A, Koźmiński W, Brutscher B, Bermel W, Pierattelli R, Felli IC (2014) “CON-CON” assignment strategy for highly flexible intrinsically disordered proteins. J Biomol NMR 60:209–18. doi: 10.1007/s10858-014-9867-6

Rios CB, Feng W, Tashiro M, Shang Z, Montelione GT (1996) Phase labeling of C-H and C-C spin-system topologies: application in constant-time PFG-CBCA(CO)NH experiments for discriminating amino acid spin-system types. J Biomol NMR 8:345–350. doi: 10.1007/BF00410332

Schubert M, Ball LJ, Oschkinat H, Schmieder P (2000) Bridging the gap: A set of selective 1H-15N-correlations to link sequential neighbors of prolines. J Biomol NMR 17:331–335. doi: 10.1023/A:1008362904205

Schubert M, Labudde D, Leitner D, Oschkinat H, Schmieder P (2005) A modified strategy for sequence specific assignment of protein NMR spectra based on amino acid type selective experiments. J Biomol NMR 31:115–28. doi: 10.1007/s10858-004-8263-z

Schubert M, Oschkinat H, Schmieder P (2001a) Amino acid type-selective backbone ^{1}H-^{15}N-correlations for Arg and Lys. J Biomol NMR 379–384.

Schubert M, Oschkinat H, Schmieder P (2001b) MUSIC, selective pulses, and tuned delays: amino acid type-selective ^{1}H-^{15}N correlations, II. J Magn Reson 148:61–72. doi: 10.1006/jmre.2000.2222

Schubert M, Oschkinat H, Schmieder P (2001c) MUSIC and aromatic residues: amino acid type-selective ^{1}H-^{15}N correlations, III. J Magn Reson 153:186–92. doi: 10.1006/jmre.2001.2447

Schubert M, Smalla M, Schmieder P, Oschkinat H (1999) MUSIC in triple-resonance experiments: amino acid type-selective ^{1}H-^{15}N correlations. J Magn Reson 141:34–43. doi: 10.1006/jmre.1999.1881

Shaka A., Barker P., Freeman R (1985) Computer-optimized decoupling scheme for wideband applications and low-level operation. J Magn Reson 64:547–552. doi: 10.1016/0022-2364(85)90122-2

Shaka A., Keeler J, Freeman R (1983) Evaluation of a new broadband decoupling sequence: WALTZ-16. J Magn Reson 53:313–340. doi: 10.1016/0022-2364(83)90035-5

Tamiola K, Acar B, Mulder FAA (2010) Sequence-specific random coil chemical shifts of intrinsically disordered proteins. J Am Chem Soc 132:18000–3. doi: 10.1021/ja105656t

Ulrich EL, Akutsu H, Doreleijers JF, et al (2008) BioMagResBank. Nucleic Acids Res 36:D402–8. doi: 10.1093/nar/gkm957

Zawadzka-Kazimierczuk A, Koźmiński W, Billeter M (2012) TSAR: a program for automatic resonance assignment using 2D cross-sections of high dimensionality, high-resolution spectra. J Biomol NMR 54:81–95. doi: 10.1007/s10858-012-9652-3

Amino acid recognition for automatic resonance assignment of intrinsically disordered proteins

Alessandro Piai [1], Leonardo Gonnelli [1], Isabella C. Felli [1], Roberta Pierattelli [1], Krzysztof Kazimierczuk [2], Wiktor Koźmiński [3], Anna Zawadzka-Kazimierczuk [3] ✉

[1] CERM, Department of Chemistry, University of Florence, Via Luigi Sacconi 6, 50019 Sesto Fiorentino (Florence), Italy

[2] Centre of New Technologies, University of Warsaw, Banacha 2C, 02-097 Warsaw, Poland

[3] Biological and Chemical Research Centre, Faculty of Chemistry, University of Warsaw, Żwirki i Wigury 101, 02-089 Warsaw, Poland

✉ e-mail: anzaw@chem.uw.edu.pl

Supplementary Material

NMR experiments

In this section the pulse scheme for the 4D γ-selective-HCBCACON experiment is described.

The arrows indicate the switching of the ^{13}C carrier frequency. Narrow and wide symbols stand for 90° and 180° pulses, respectively. The pulses were applied along the *x* axis unless noted differently. The rectangles represent ^{15}N and ^{1}H non-selective pulses, the round shapes represent band-selective ^{13}C pulses. The grey round shape represents a ^{13}C adiabatic pulse. The two different variants necessary for the IPAP virtual carbonyl decoupling approach are shown. The line denoted with PFG stands for pulsed field gradients applied along the z-axis. All gradients employed had 1 ms of duration and a sine-shape. The phase cycle, the position of the carriers and the duration of the delays for the experiment are reported in the caption of the figure describing the pulse sequence.

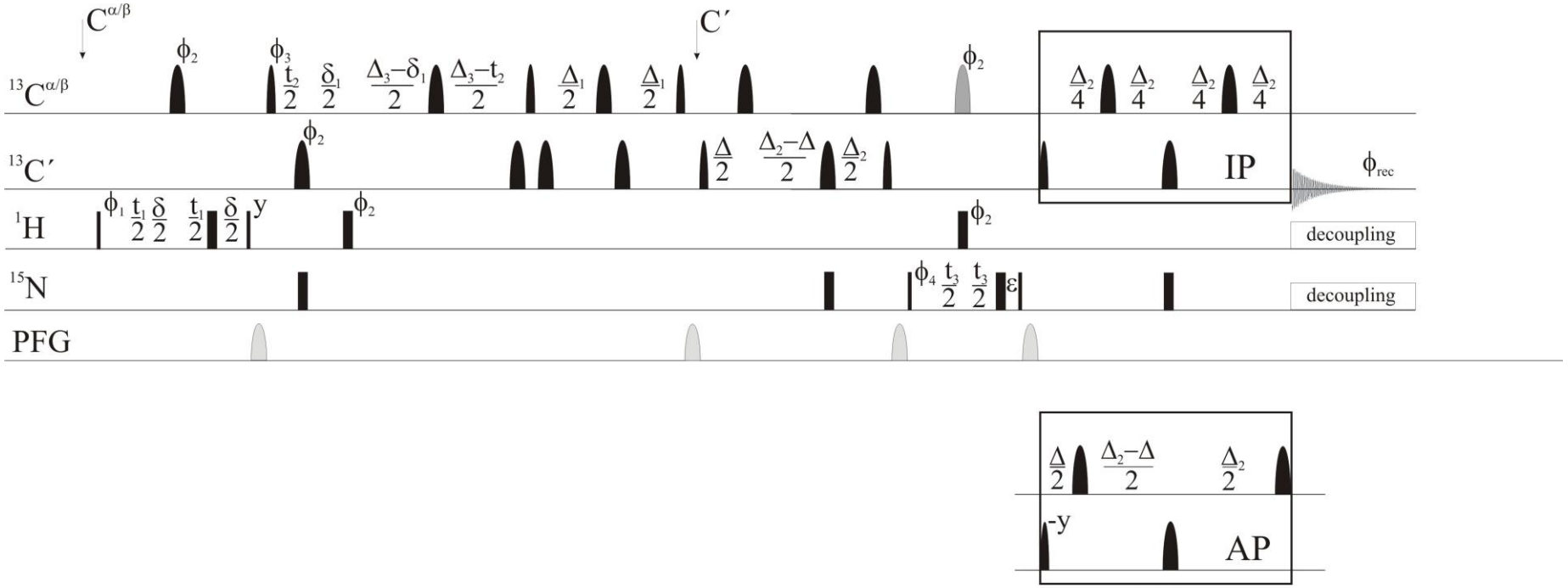


Figure S1. Pulse sequence of the 4D γ-selective-HCBCACON experiment

The following phase cycling was employed: ϕ_1 = x, -x; ϕ_2 = 8(x), 8(-x); ϕ_3 = 2(x), 2(-x); ϕ_4 = 4(x), 4(-x); ϕ_{rec} = x, 2(-x), x, -x, 2(x), -x. The lengths of the delays were: δ = 3.6 ms, δ_1 = 2.2 ms, Δ = 9.0 ms, Δ_1 = 8.0 ms, Δ_2 = 30.0 ms, Δ_3 = 21.0 ms, $\varepsilon = t_3(0)+p_8$, where p_8 is the duration of the adiabatic pulse on ^{13}C in the middle of the ^{15}N evolution period. The constant time mode for the evolution of chemical shift was used in aliphatic carbon evolution periods, whereas real time evolution was adopted for proton and nitrogen evolution periods. Quadrature detection in the indirect dimensions was achieved by incrementing phases ϕ_1, ϕ_3 and ϕ_4 in a States manner.

Experimental parameters used. The 4D γ-selective-HCBCACON experiment was recorded in 1 day and 13 hours with 4 scans per increment, an inter scan delay of 1.1 s, spectral widths of 8800 (aq) × 2550 (^{15}N) × 12500 ($^{13}C^{\alpha,\beta}$) × 5000 (^{1}H) Hz with 512 complex points in the acquisition dimension and 1540 hypercomplex points randomly distributed in the indirectly detected dimensions (data points density of 0.051 %). The maximum evolution times in the indirectly detected dimensions were set to 20.5 ms for the $^{13}C^{\alpha,\beta}$ dimension, to 60.4 ms for the ^{15}N dimension and to 15.0 ms for the ^{1}H dimension.

Using amino acid-selective data by the TSAR program

In order to use the data from amino acid-selective (CA)CON and (CA)NCO experiments, the 'aa_select.txt' file has to be created. The format of such a file is described below. Incorporating of such an input file for TSAR is optional. The TSAR program, together with a detailed manual, can be downloaded from the website http://nmr.cent3.uw.edu.pl/software

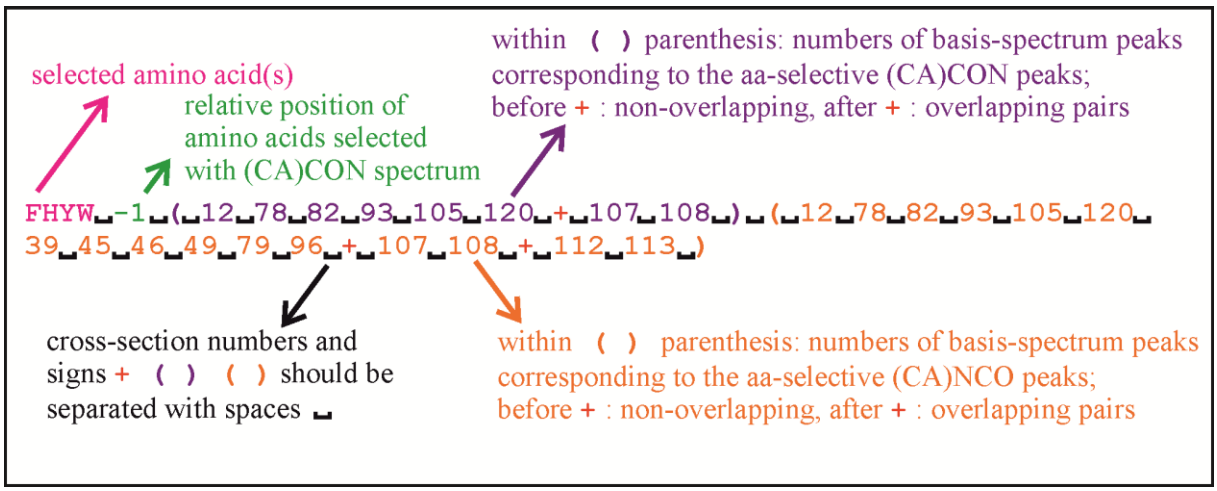


Figure S2. Description of the 'aa_select.txt' file format.

The file contains information on the numbers of peaks corresponding to the selected amino acids. In the figure a single line (for one amino acid selection) is shown.

```
N -1 ( 53 61 66 ) ( 53 61 66 32 38 68 )
D -1 ( 32 44 48 49 72 136 ) ( 32 44 48 49 72 136 33 58 60 64
123 140 )
QN -1 ( 11 53 61 66 64 71 85 91 + 80 81 ) ( 53 61 66 64 71
85 91 32 38 68 44 47 59 65 67 82 + 80 81 )
E -1 ( 10 14 60 67 110 116 134 137 + 8 9 + 69 70 ) ( 10 14
60 67 110 116 134 137 40 51 57 61 66 73 76 90 119 120 + 8 9
+ 69 70 + 80 81 + 107 108 )
FHYW -1 ( 12 78 82 93 105 120 + 107 108 ) ( 12 78 82 93 105
120 39 45 46 49 79 96 + 107 108 + 112 113 )
G -1 ( 3 22 24 34 35 36 37 42 43 45 46 52 54 62 101 111 115
) ( 3 22 24 34 35 36 37 42 43 45 46 52 54 62 101 111 115 1 2
4 5 10 11 12 13 14 15 16 17 + 6 7 + 8 9 )
A -1 ( 1 2 18 19 39 40 41 47 50 51 55 57 86 87 92 94 98 135
) ( 1 2 18 19 39 40 41 47 50 51 55 57 86 87 92 94 98 135 101
111 114 115 116 126 127 128 129 130 133 134 + 117 118 + 131
132 )
S -1 ( 83 90 95 97 ) ( 83 90 95 97 22 24 26 27 )
```

Figure S3. Example of the 'aa_select.txt' file.

Sample file, containing information on eight different amino acid selections, all in both (CA)CON and (CA)NCO versions. Due to not full selectivity of the Q-selective spectrum, the selection was defined as 'QN', instead of 'Q'.

Table S1 The selectivity of 2D amino-acid selective experiments.

selection	(CA)CON	(CA)NCO
	Selectivity	selectivity
A	no u. p. [a]	no u. p.
D	no u. p.	no u. p.
E	D peaks (of opposite sign)	no u. p.
FHYW	no u. p.	no u. p.
G	no u. p.	P peaks (of opposite sign)
N	no u. p.	no u. p.
Q	N peaks	N peaks (of opposite sign)
S	no u. p.	P peaks

[a] no unexpected peaks (peaks of different amino acids)

Article 6. The crowd you're in with: Effects of different types of crowding agents on protein aggregation

Leonid Breydo[1], Krishna D. Reddy[1], Alessandro Piai[2], Isabella C. Felli[2], Roberta Pierattelli[2], Vladimir N. Uversky[1,3,4]

[2] Department of Molecular Medicine, Morsani College of Medicine, University of South Florida, Tampa, FL 33612, USA.

[2] CERM and Department of Chemistry "Ugo Schiff", University of Florence, 50019, Via Luigi Sacconi 6, Sesto Fiorentino, Florence, Italy.

[3] Byrd Alzheimer's Research Institute, Morsani College of Medicine, University of South Florida, Tampa, FL 33612, USA.

[4] Institute for Biological Instrumentation, Russian Academy of Sciences, 142292 Pushchino, Moscow Region, Russia.

Published in 2014 in Biochim.Biophys.Acta, 1844, 346-357

Biochimica et Biophysica Acta 1844 (2014) 346–357

Contents lists available at ScienceDirect

Biochimica et Biophysica Acta

journal homepage: www.elsevier.com/locate/bbapap

The crowd you're in with: Effects of different types of crowding agents on protein aggregation

Leonid Breydo [a,*], Krishna D. Reddy [a], Alessandro Piai [b], Isabella C. Felli [b], Roberta Pierattelli [b], Vladimir N. Uversky [a,c,d]

[a] Department of Molecular Medicine, Morsani College of Medicine, University of South Florida, Tampa, FL 33612, USA
[b] Center for Magnetic Resonance, University of Florence, 50019 Sesto Fiorentino, Florence, Italy
[c] Byrd Alzheimer's Research Institute, Morsani College of Medicine, University of South Florida, Tampa, FL 33612, USA
[d] Institute for Biological Instrumentation, Russian Academy of Sciences, 142292 Pushchino, Moscow Region, Russia

ARTICLE INFO

Article history:
Received 21 September 2013
Received in revised form 1 November 2013
Accepted 11 November 2013
Available online 16 November 2013

Keywords:
Amyloid
Crowding
Protein aggregation
Flexibility
Intrinsic disorder

ABSTRACT

The intracellular environment contains high concentrations of macromolecules occupying up to 30% of the total cellular volume. Presence of these macromolecules decreases the effective volume available for the proteins in the cell and thus increases the effective protein concentrations and stabilizes the compact protein conformations. Macromolecular crowding created by various macromolecules such as proteins, nucleic acids, and carbohydrates has been shown to have a significant effect on a variety of cellular processes including protein aggregation. Most studies of macromolecular crowding have used neutral, flexible polysaccharides that function primarily via excluded volume effect as model crowding agents. Here we have examined the effects of more rigid polysaccharides on protein structure and aggregation. Our results indicate that rigid and flexible polysaccharides influence protein aggregation via different mechanisms and suggest that, in addition to excluded volume effect, changes in solution viscosity and non-specific protein–polymer interactions influence the structure and dynamics of proteins in crowded environments.

1. Introduction

Protein aggregation is associated with a variety of human diseases including Alzheimer's, Parkinson's, and Huntington's diseases, prionopathies, and type II diabetes. Protein aggregation usually starts from the protein in a partially unfolded conformation similar to the pre-molten globule state [1]. In order to access this conformation, a folded protein needs to partially unfold, while an intrinsically disordered protein (IDP) needs to partially fold. In addition, the mechanism of aggregation is usually a rather complex process with many intermediate oligomeric states characterized by variable extent of secondary structure [2–5]. Some of these oligomers easily convert to fibrils while others have a high degree of kinetic stability.

Investigation of protein aggregation is complicated by the fact that the in vivo environment is crowded with macromolecules, such as proteins, nucleic acids, and carbohydrates, which occupy up to 30% of the available volume. Macromolecular crowding increases the effective protein concentration, decreases the protein diffusion rate, promotes partial folding of some intrinsically disordered proteins and facilitates their aggregation [6]. It has been previously shown that macromolecular crowding has a significant effect on protein–protein interactions including protein aggregation [7]. Crowding tends to stabilize compact protein conformations. These conformations can be either on the pathway to folding or on the pathway to aggregation depending on the protein and the assay conditions. Thus, the effect of crowding agents on protein aggregation depends on the nature of the protein.

Most studies of the effects of crowding on protein aggregation have used flexible, hydrophilic polymers (PEG, dextran and Ficoll). Due to their compact, largely spherical shape these polymers have relatively small surface to volume ratio. They are neutral and relatively hydrophilic minimizing their specific interactions with proteins. Thus these polymers are believed to act primarily via excluded volume effect by decreasing the effective volume available for the proteins in the cell and thus increasing the effective protein concentration. It has been shown that aggregation of many proteins and peptides is accelerated by the presence of dextrans and other neutral flexible crowding agents [8–12]. However, for some proteins with highly stable folded native states (e.g. lysozyme or superoxide dismutase) addition of crowding agents has been shown to inhibit aggregation [11].

A variety of other, more rigid biopolymers are also present in vivo including DNA, protein fibers and polysaccharide components of extracellular matrix. Solutions of rigid polymers have higher viscosity that may affect protein diffusion and slow down protein folding or aggregation. These polymers also have lower density making them more effective in creating excluded volume effect as intrinsic viscosity of a polymer is proportional to its volume [13]. Lower polymer density also increases

* Corresponding author.
E-mail addresses: lbreydo@health.usf.edu (L. Breydo), vuversky@health.usf.edu (V.N. Uversky).

1570-9639/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.bbapap.2013.11.004

the exposed surface of rigid polymers available for interactions with proteins. Here we investigated the effects of both compact, flexible polysaccharides (dextrans) and more rigid cellulose derivatives (hydroxypropyl celluloses or HPCs) on the kinetics and mechanism of aggregation of several proteins with variable degrees of intrinsic disorder. We selected neutral, hydrophilic polysaccharides to minimize the specific protein–polymer interactions. In order to test the effects of these polymers on the protein structure and aggregation, we selected several small proteins with different structural and dynamic properties such as the degree of intrinsic disorder and oligomeric state. We tested aggregation of these proteins in the presence of polysaccharides in the conditions previously determined to be favorable for their conversion to amyloid fibrils. We examined the effects of the polysaccharides both on kinetics of aggregation and the structure of the aggregates.

2. Materials and methods

2.1. Materials

Recombinant α-synuclein was a gift from Dr. Munishkina (University of California Santa Cruz). Recombinant human insulin was from Akron Biotech (Boca Raton, FL). A commercially available mixture of core histones H2A, H2B, H3 and H4 from calf thymus (Calbiochem) was used without additional fractionation. Recombinant ^{13}C–^{15}N enriched α-synuclein was prepared as previously described [14]. All other proteins and chemicals were from Sigma, Fisher Scientific or VWR Scientific.

2.2. Methods

2.2.1. Protein aggregation assays

Conditions for protein aggregation were optimized for each protein. Aggregation of insulin was studied in two sets of conditions: low pH, which stabilizes the monomeric form of the protein, and neutral pH, which stabilizes the insulin hexamer. At low pH aggregation of insulin (0.5 mg/ml) was conducted in glycine buffer (20 mM, pH 2.5) at 37 °C. The protein was dissolved directly in this buffer and incubated for 5 min prior to the start of the reaction. At neutral pH insulin (1.2 mg/ml) was aggregated in 20 mM sodium phosphate (pH 7.5) at 37 °C. Aggregation of histones (0.75 mg/ml) was conducted in the glycine buffer (20 mM glycine, pH 2.5) in the presence of 0.7 M NaCl at 37 °C. Histones were initially dissolved in 5 mM HCl at 4 mg/ml, incubated in this solution for 5 min and diluted into the final reaction buffer. Aggregation of α-synuclein (0.4 mg/ml) was conducted in 20 mM acetate buffer, pH 3.5 in the presence of 0.1 M NaCl at 37 °C. Protein was initially dissolved in 5 mM NaOH at 4 mg/ml, incubated in this solution for 1 min and diluted into the final reaction buffer. α-Lactalbumin (0.2 mg/ml) was aggregated in 40 mM phosphate buffer (pH 7.0) in the presence of 0.1 M NaCl, 1 mM EGTA and 2 mM DTT at 30 °C. Lysozyme (0.25–0.75 mg/ml) was aggregated in 25 mM potassium phosphate buffer (pH 2.0) in the presence of 225 mM NaCl at 39 °C using a modification of the method of Hill et al. [15].

Protein aggregation in the automated format was carried out in a reaction volume of 0.1 ml in black, flat-bottomed 96-well plates in the presence of 10 μM ThT. Two Teflon or polyethylene balls (2.38 mm diameter, Precision Ball, Reno, PA) were placed into each well of a 96-well plate. The reaction mixture containing protein and ThT (350 μl) was split into three wells (100 μl into each well), the plates were covered by Mylar septum sheets (Thermo), and incubated with continuous orbital shaking at 280 rpm in an Infinite M200 Pro microplate reader (Tecan). The kinetics was monitored by top reading of fluorescence intensity every 3–8 min using 444 nm excitation and 485 nm emission filters. Data from replicate wells were averaged before plotting fluorescence vs time. The data were fit to a sigmoidal equation using Origin (OriginLab, Northampton, MA). The equation [16,17] was:

$$F = A + B/(1 + \exp(k \times (t - t_m))) \quad (1)$$

where A is the initial level of ThT fluorescence, B is the difference between the final level of ThT fluorescence and its initial level, k is the rate constant of amyloid accumulation (h^{-1}) and t_m is the midpoint of transition. The lag time (t_l) of amyloid formation was calculated as $t_l = t_m - 2 / k$. The parameters derived from this equation are: yield of amyloid (B), lag time (t_l), and elongation rate (k) of amyloid. Initiation rate was defined as the inverse of lag time. Although Eq. (1) gave good fits for the ThT kinetic profiles, the expression is strictly an empirical means of deriving kinetic parameters from the data and does not necessarily reflect the underlying complex kinetic scheme.

2.2.2. Electron microscopy

10 μl aliquots of protein solutions (0.1–0.3 mg/ml) were adsorbed onto 200 mesh formvar/carbon-coated nickel grids for 5 min. The grids were washed with water (10 μl), stained with 2% uranyl acetate for 2 min and washed with water again. The samples were analyzed with a JEM 1400 transmission electron microscope (JEOL) operated at 80 kV.

2.2.3. NMR

All the spectra were acquired at a 21.1 T Bruker AVANCE spectrometer operating at 898.71 MHz for ^{1}H. Samples were measured at 285.5 K, unless otherwise specified, by using a cryogenically cooled triple-resonance probe head. The stock α-synuclein solution was 0.6 mM ^{13}C, ^{15}N labeled α-synuclein in 20 mM potassium phosphate buffer, pH 6.5, 200 mM NaCl, 0.5 mM EDTA. Generally the samples were prepared by taking 150 μl of the stock solution and adding either buffer or the appropriate amount of crowder and buffer up to a final volume of 200 μl in a 3 mm NMR tube. Final concentrations of polymers were 5% for dextran 100 and 2.5% for HPC 100. The peak volumes were calculated using a routine implemented in the program CARA [18].

2.2.4. FTIR

FTIR spectra were measured with a Bruker Tensor 37 FTIR instrument (Bruker Optics, Billerica, MA) equipped with a DTGS detector. Aqueous protein solutions (25 μl, 1.2 mg/ml) were loaded into the BioATRcell II. 512 scans at 2 cm^{-1} resolution were collected for each sample, corrected for water vapor, and background spectra were subtracted.

2.2.5. CD

Far-UV CD (195–260 nm) spectra of proteins were measured using a JASCO J-815 spectropolarimeter at room temperature. A solution of protein (110 μl, 1 mg/ml) was placed into a 0.2 mm pathlength cell (0.1 mm pathlength for histones), and the CD spectra were acquired with 20 nm/min scan speed at 0.1 nm step size and 1.0 nm bandwidth under constant purging with nitrogen. Four spectra were accumulated and averaged for each sample. The same buffer was used for CD and FTIR measurements and for protein aggregation experiments.

2.2.6. Viscosity measurements

Viscosity of polymer solutions was determined with an Ostwald viscometer (Cannon Instruments) at 25 °C. In a typical experiment, a solution of polymer in water (2 ml) was loaded into the viscometer and the time taken for it to pass between the marks was measured with a stopwatch. Each measurement was performed at least in triplicate.

2.2.7. Determination of global stability

Aggregated samples (1–3 μl) were suspended in HEPES buffer (20 μl total volume, 50 mM, pH 7.5) containing different concentrations of GdnSCN. The solution was incubated for 1 h at 24 °C and then diluted to 300 μl with 6 M GdnSCN and HEPES buffer (50 mM, pH 7.5) adjusting the final concentration of GdnSCN to 0.2 M. Fluorescence spectra were recorded in the presence of 10 μM ThT. Excitation wavelength was 442 nm and emission spectrum was recorded in 470–500 nm range. Excitation slit was at 2.5 nm and emission slit at 5 nm.

Emission spectra were recorded in triplicate and averaged. The measurements were performed in duplicate for each sample.

3. Results

3.1. Selection and characterization of polysaccharides

To mimic the effect of a variety of biopolymers on protein aggregation, we selected several neutral polysaccharides of different shapes and sizes. The advantage of using polysaccharides is that their hydrophilicity minimizes specific protein–polymer interactions. Neutrality of these polymers helps avoid protein binding due to charge–charge interactions. Thus, the effects of neutral polysaccharides on proteins are more likely to be due to either their effects on solution properties or to non-specific interactions with proteins.

We have selected polymers that are composed from the same subunits but differ in their structure: dextran and hydroxypropyl cellulose (HPC). Dextran is a polyglucoside (~5% branching) with main chain connected by α-1,6-glycosidic linkages while branches begin from α-1,3 linkages. Dextrans are known to adopt a random coil conformation in water with significant intermolecular interactions observed at high concentrations [19,20]. α-Glycosidic linkages are quite flexible, and their presence ensures high flexibility of structure of this polymer [21,22]. Cellulose is a linear, unbranched polymer consists of a linear chain of glucose units connected by β-1,4-glycosidic linkages. The rigid nature of these linkages makes the cellulose chains adopt the extended, rod-like conformation. HPC is a derivative of cellulose where some of the hydroxyl groups of cellulose have been modified with hydroxypropyl residues (Fig. 1A).

We determined intrinsic viscosities of several preparations of dextrans and hydroxypropyl celluloses with different average molecular weights by measuring the dependence of viscosities of solutions of these polymers on their concentrations (Fig. S1). For expanded flexible polymer chains, the intrinsic viscosity is related to molecular weight by an empirical relation, the Mark–Houwink equation ($[\eta] = K * MW^a$), where MW is the average molecular weight of the polymer while K

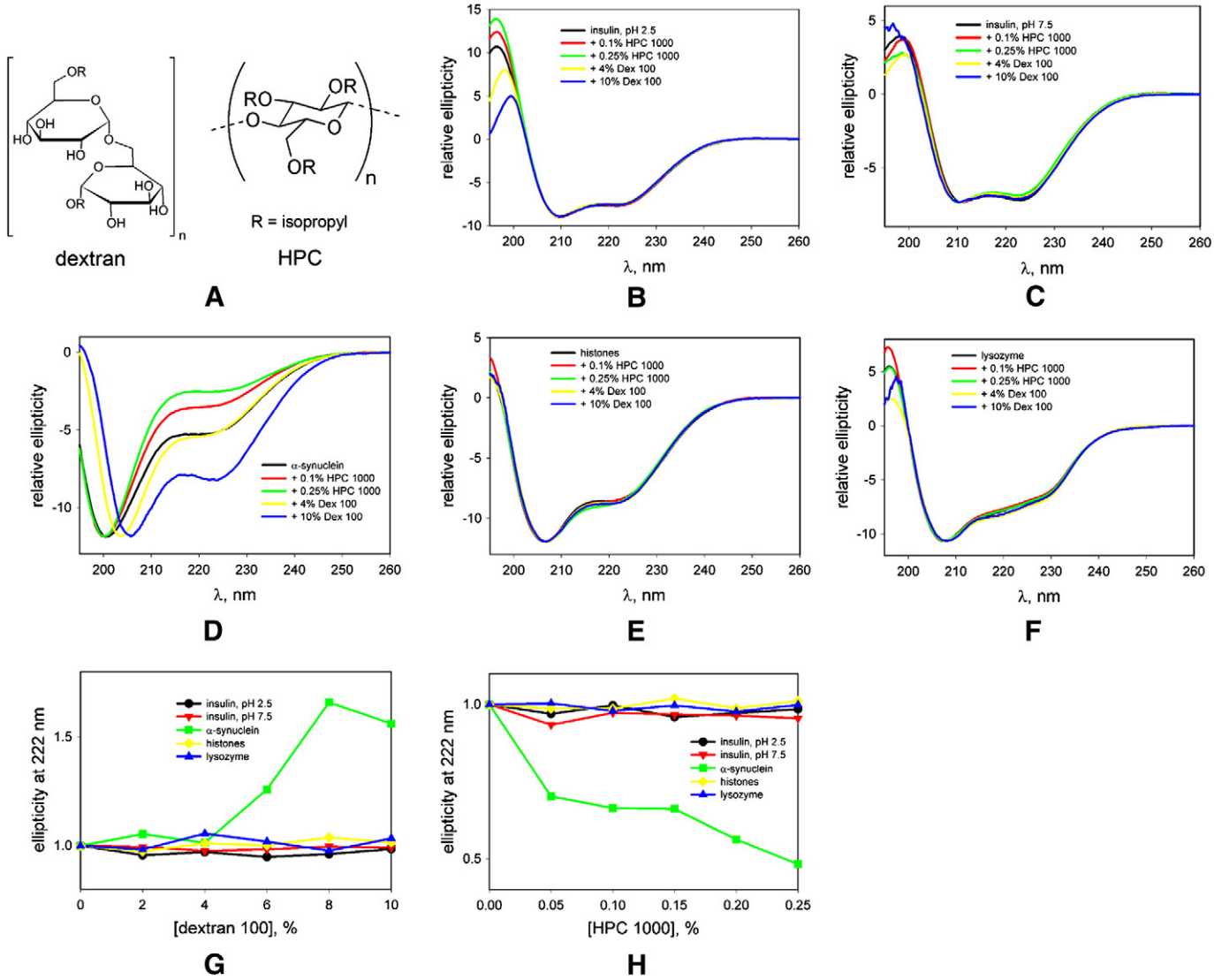


Fig. 1. (A) Structures of polysaccharides. (B–F) Effect of crowding agents on the secondary structure of proteins was evaluated by examining the changes in their far UV circular dichroism spectra. CD spectra of insulin at pH 2.5 (B), insulin at pH 7.5 (C), α-synuclein (D), histones (E) and lysozyme (F) alone and in the presence of dextran 100 (MW 100 kDa) and HPC 1000 (MW 1 MDa) are shown. Protein alone is in black, protein in the presence of 0.1% HPC 1000 in red, protein in the presence of 0.25% HPC 1000 in green, protein in the presence of 4% dextran 100 in yellow and protein in the presence of 10% dextran 100 in blue. (G–H) Relative ellipticity at 222 nm for several proteins is plotted as a function of polysaccharide concentration (0–10% of dextran 100 and 0–0.25% of HPC 1000). This plot allows one to see the conformational changes in the proteins with the increasing polysaccharide concentration. Insulin at pH 2.5 shown in black, insulin at pH 7.5 — in red, α-synuclein in green, histones in yellow, and lysozyme in blue.

and *a* are the constants [13]. *a* reflects the flexibility of the polymer in a given solvent with values around 0.5 corresponding to a highly flexible polymer and values above 0.8 reflecting a semi-rigid polymer. Extremely rigid, rod-like polymers (e.g. double-stranded DNA) have *a* values around 2. Values of this constant were determined for our polymers from the log plot of their intrinsic viscosity vs. average molecular weight (Fig. S1). Dextran was determined to have an *a* value of 0.45 consistent with the values previously reported, indicating a highly flexible polymer. Cellulose derivatives had *a* values around 1, indicating significantly more rigid polymers as expected based on their structure (Fig. S1). These coefficients are similar to the values previously reported for these polymers (0.5 for dextrans and 1.1 for HPCs) [23,24].

3.2. Oligomeric state and conformation of proteins in the assay conditions

In order to test the effect of polysaccharides on protein structure and aggregation, we selected several small proteins with different structural and dynamic properties. We paid special attention to the degree of intrinsic disorder in the protein sequence and oligomeric state of the protein in the assay conditions. Both of these parameters are known to play an important role in protein aggregation. Structural features of the proteins we analyzed (insulin, α-synuclein, histones, lysozyme and α-lactalbumin) are described below. We tested aggregation of these proteins in the presence of polysaccharides in the conditions previously determined to be favorable for their conversion to amyloid fibrils.

Insulin is a primarily α-helical protein (MW 5.5 kDa). It is hexameric at neutral pH, and addition of dextrans and other flexible polysaccharides has been shown to inhibit its aggregation presumably due to hexamer stabilization [8]. At lower pH values, an insulin hexamer dissociates to dimers and eventually to monomers. We examined aggregation of insulin at pH 2.5 where this protein is primarily monomeric [8,25] and at pH 7.5 where it is primarily hexameric. Low pH is more favorable for insulin aggregation but fibril formation is possible in both sets of conditions [8,9]. CD spectra (Fig. 1B–C) confirm that insulin is primarily alpha-helical at both sets of conditions although some degree of disorder can be seen at pH 2.5 as judged from higher intensity of the 208 nm peak in CD spectrum.

α-Synuclein is an intrinsically disordered, highly amyloidogenic protein. It exists as a monomer (MW 14 kDa) in most conditions although it has been suggested to exist as a tetramer in vivo [26,27]. Despite lack of stable secondary structure, α-synuclein is known to be more compact than expected for coil-like conformation [28,29]. In the assay conditions, CD spectra (Fig. 1D) indicate that α-synuclein is primarily disordered.

Histones are lysine-rich intrinsically disordered proteins that bind to DNA as octamers [30]. They exist in a variety of oligomeric states ranging from monomer (MW 14 kDa) to octamer depending on the conditions [31–33]. High salt concentrations decrease repulsion between these highly charged proteins and promote both their oligomerization and aggregation. A combination of low pH and high salt concentration has been shown to be especially favorable for aggregation of these proteins [31,34], and these conditions were used in the current study. CD spectra (Fig. 1E) indicate that histones are mostly disordered in the assay conditions. Earlier studies [31] have shown that in the conditions (pH 2, 0.5 M NaCl) similar to those used here for histone aggregation, histones are present as a complex mixture of oligomers ranging from monomers to octamers with tetramers being the most prevalent species.

Hen egg white lysozyme, a glycoside hydrolase, is a small (14 kDa) globular protein composed of a mixture of α-helices and β-sheets (Fig. 1F). Lysozyme partially unfolds and aggregates at low pH and elevated temperature [35]. In these conditions monomeric lysozyme either converts directly to fibrils or first forms stable oligomeric intermediates depending on salt concentration [15]. In the conditions we used (pH 2.0, 225 mM NaCl) oligomeric intermediates are present [15].

α-Lactalbumin, a regulatory subunit of lactose synthase, is a small (14 kDa), primarily α-helical calcium-binding protein containing four disulfide bonds [36]. Upon removal of calcium this protein is converted to a molten globule state. Upon reduction of disulfide bonds, the structure of apo-α-lactalbumin is expanded further with the protein becoming primarily disordered judging from its CD spectrum [37], and the protein becomes highly prone to aggregation. Flexible hydrophilic polymers have been shown to inhibit aggregation of this protein [37].

3.3. Interaction of proteins with polysaccharides

Although neutral polysaccharides such as dextrans do not form stable complexes with proteins, there is some evidence of weak and non-specific chemical interactions between proteins and neutral polysaccharides [38,39]. We have examined interaction of our proteins with polysaccharides using CD and NMR spectroscopy.

We measured far UV CD spectra of most proteins we used in this study both alone and in the presence of dextran 100 (dextran, MW 100 kDa) and HPC 1000 (HPC, MW 1 MDa) in the conditions we used for their aggregation at 25 °C. CD spectra of α-lactalbumin are not shown since after the reduction of disulfide bonds this protein rapidly aggregates and its CD spectra may be affected by this process. For most proteins we did not observe any significant effect of either polymer on the protein secondary structure (Fig. 1B–F). α-Synuclein, however, proved to be an exception. In the presence of dextran the peak at 200 nm corresponding to unfolded structure shifted to higher wavelength (205–208 nm) and a shoulder at 222 nm significantly increased in intensity. This is an indication of induced secondary structure in the presence of dextrans. In the presence of HPC, on the other hand, the shoulder at 222 nm decreased in intensity signifying further unfolding of the protein. So overall, polysaccharides had a significant effect on the structure of α-synuclein with dextran promoting its folding and HPC promoting its unfolding. Our experiments were conducted at pH 3.5 where α-synuclein is present in a relatively compact conformation that may be more sensitive to the effects of crowding agents [29].

Interestingly, earlier studies conducted at neutral pH did not show any effect of crowding agents on the secondary structure of α-synuclein [9]. We decided to use 2D NMR to examine binding of these polymers to α-synuclein at pH 6.5 [40]. In NMR experiments, solutions of polymers (dextran, MW 100 kDa, and HPC, MW 100 kDa) were added to a 0.6 mM solution of ^{13}C, ^{15}N labeled α-synuclein in neutral buffer. No significant changes in ^{13}C or ^{15}N chemical shifts were observed in the presence of crowding agents (Fig. 2A–B). HN and CN 2D correlation spectra (Fig. 2C) also showed no effect of either dextran or HPC on protein structure. Chemical shift index is the difference between the experimentally measured chemical shifts and the chemical shifts predicted on the basis of the primary sequence assuming a completely random coil behavior. These differences are generally used to identify the secondary structure propensity of different regions of the protein. We have examined the changes in the chemical shift index of α-synuclein in the presence of these polysaccharides. The experimental chemical shifts were obtained from CN correlation spectra. The chemical shifts for random coil were predicted according to the primary sequence of α-synuclein taking into account up to 2 previous and the 2 following residues using Poulsen's reference data (http://www1.bio.ku.dk/english/research/pv/sbin_lab/staff/MAK/randomcoil/script/). The chemical shift indices for ^{15}N and ^{13}C nuclei (Fig. S2 A–D) showed that the protein is unfolded before and after the addition of the polysaccharides. No visible changes in the secondary structure due to the presence of crowding agents were detected. We have also examined the changes in the peak volumes in NMR spectra with the addition of polysaccharides. The peak volumes were calculated with a routine implemented in the program CARA [18] that assumes an ideal peak shape. Both NH and CN correlation spectra were used (HSQC and CON respectively). The data (Fig. S2 E–F) indicates no significant effect of either polymer on the peak volume. Overall, NMR experiments did not detect an interaction between α-synuclein and polysaccharides or a structural

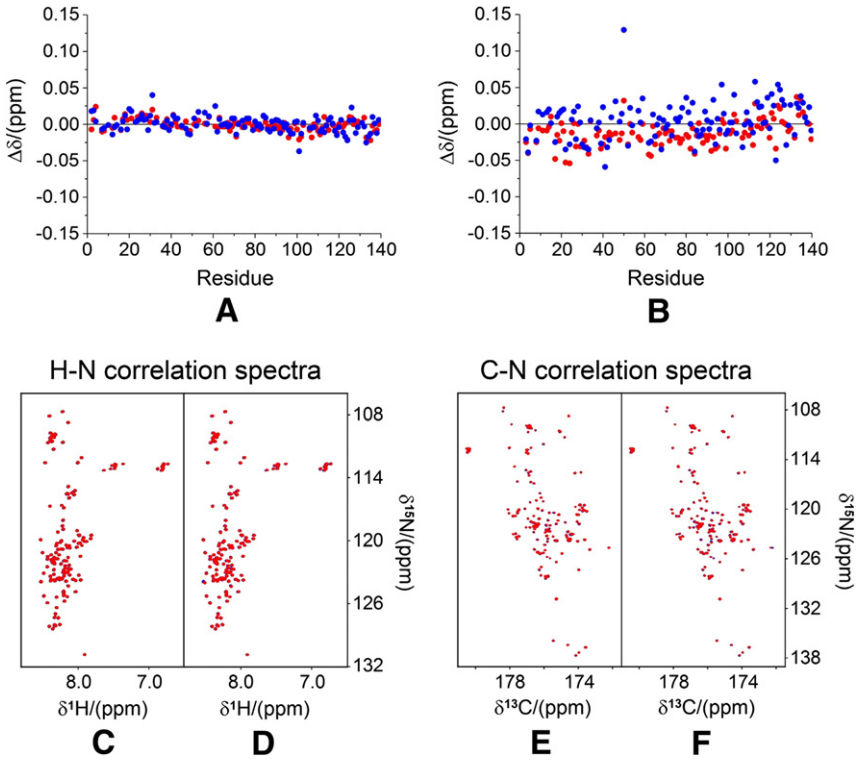


Fig. 2. Multi-dimensional NMR spectra of $^{13}C/^{15}N$ labeled α-synuclein in the presence of crowding agents. Variations of ^{13}C and ^{15}N chemical shifts in respect to α-synuclein alone are reported in (A) and (B), respectively. In both panels, red dots belong to α-synuclein in the presence of 5% dextran 100, whereas blue dots to α-synuclein in the presence of 2.5% HPC 100. In the bottom part of the figure, the H–N (C, D) and C–N (E–F) correlation spectra are reported. Blue spectra belong to α-synuclein alone, red ones to α-synuclein in the presence of 5% dextran 100 (C, E) or 2.5% HPC 100 (D, F). The spectra do not show significant spectral changes due to the presence of crowding agents.

change in the α-synuclein in the presence of polysaccharides at neutral pH.

3.4. Kinetics of protein aggregation in the presence of crowding agents

Effects of crowding agents on protein aggregation have been extensively investigated in the past. It has been shown that aggregation of many amyloidogenic proteins and peptides is accelerated by the presence of dextrans and other neutral flexible crowding agents [8–12]. This acceleration has been attributed to the excluded volume effect [7,41] although weak interactions between proteins and polymers have been detected as well [6,19,38,42]. On the other hand, aggregation of stable folded proteins such as lysozyme has been inhibited by crowding agents presumably due to monomer stabilization [11].

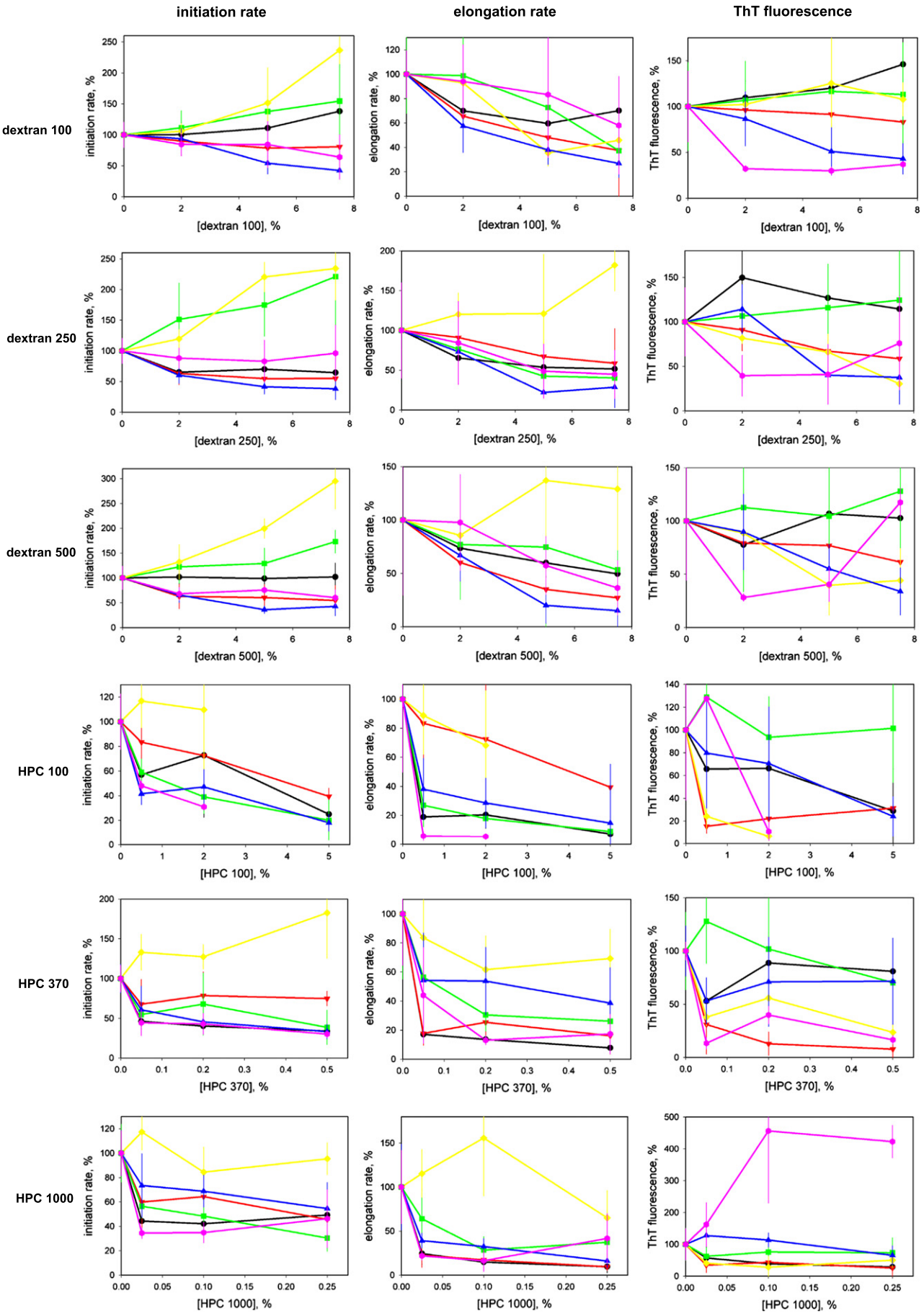
We have examined the effect of polymers on kinetics of fibril formation from our set of proteins in the conditions previously determined to be favorable for their conversion to amyloid fibrils. For insulin two sets of conditions were used: pH 2.5 where insulin is primarily monomeric and pH 7.5 where it is primarily hexameric. We found (Figs. 3–4) that dextrans with MW from 100 to 500 kDa significantly decreased the lag phase of aggregation of IDPs (histones and α-synuclein). They had a modest effect on the lag phase of fibril formation from folded proteins (insulin at pH 2.5 and lysozyme) and increased the lag phase of fibril formation from the proteins forming stable oligomers (insulin at pH 7.5 and reduced α-lactalbumin). In addition, dextrans decreased the apparent fibril elongation rate for most proteins except histones. This effect has been previously observed for aggregation of IAPP and attributed to high viscosity of dextran solutions slowing down fibril growth [19]. Effects of dextrans on apparent fibril yield as detected by ThT fluorescence were similar in direction but smaller in magnitude to their effects on the lag phase. No clear correlation between the molecular weight of dextran and its effect on protein aggregation was observed in the narrow range of molecular weights we examined.

Both initiation and elongation rates of fibrils were significantly decreased in the presence of HPCs for most proteins we studied with the exception of histones (Figs. 3–4). HPCs also decreased the apparent fibril yield although their effects on this parameter were less pronounced. Higher MW HPCs were more effective inhibitors acting at lower polymer concentrations (Fig. 3). It is interesting to note that while the effect of dextrans on both initiation and elongation rates was usually proportional to the polymer concentration in a linear fashion, the effects of HPCs on these rates were often non-linear with significant inhibition observed even at the lowest HPC concentrations we tested. This can be seen in Fig. 3 comparing the gradual changes in initiation rates with the addition of dextrans to significant decrease in initiation rate for most proteins even at the lowest concentration of HPC. The same trend is observed in elongation rates as well. These differences are also visible in Fig. 4 if one compares relatively gradual decrease of lag phase with increasing concentration of dextran (Fig. 4A–B) to a more abrupt increase in lag phase with increasing concentrations of HPC (Fig. 4C–D). This effect is an indication that inhibition of protein aggregation with HPC is saturable and thus may be driven by bimolecular protein–polymer interaction.

3.5. Contribution of solution viscosity to the effects of polysaccharides on aggregation kinetics

Variations in solution composition alter many solution properties and it is often extremely difficult to separate the effects of these

Fig. 3. Kinetic parameters (initiation rate, fibril elongation rate and ThT fluorescence) for all proteins aggregated in the presence of dextran 100, dextran 250, dextran 500, HPC 100, HPC 370, and HPC 1000. Insulin at pH 2.5 shown in black, insulin at pH 7.5 — in red, α-synuclein in green, histones in yellow, α-lactalbumin in blue and lysozyme in magenta.


initiation rate
elongation rate
ThT fluorescence
dextran 100
dextran 250
dextran 500
HPC 100
HPC 370
HPC 1000
initiation rate, %
elongation rate, %
ThT fluorescence, %
[dextran 100], %
[dextran 250], %
[dextran 500], %
[HPC 100], %
[HPC 370], %
[HPC 1000], %

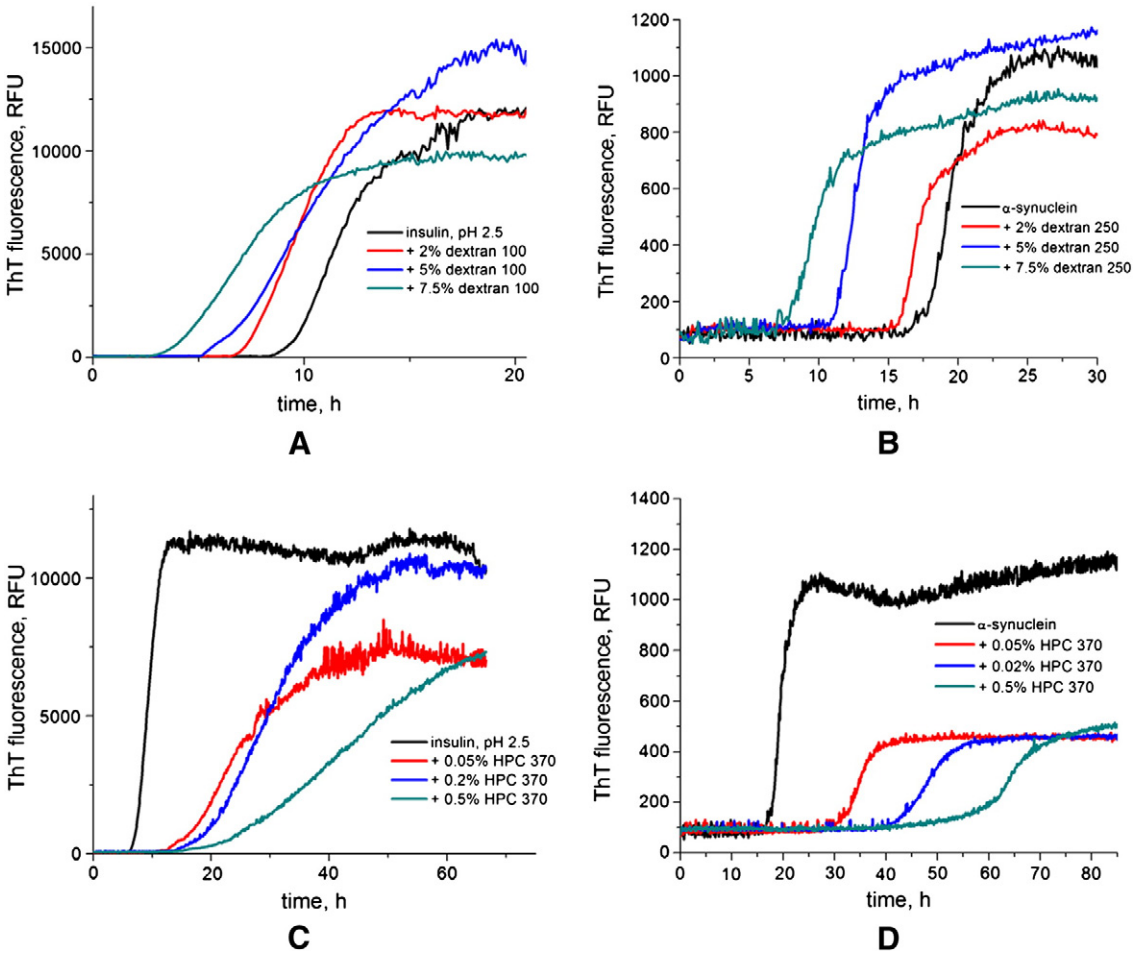


Fig. 4. Fibril formation from insulin and α-synuclein in the presence of dextran 100 and HPC 1000. (A) Insulin (pH 2.5) in the presence of increasing concentrations of dextran 100; (B) α-synuclein in the presence of increasing concentrations of dextran 100. Protein alone is in black, protein in the presence of 2% dextran 100 in red, protein in the presence of 5% dextran 100 in dark blue and protein in the presence of 7.5% dextran 100 in light blue. (C) Insulin (pH 2.5) in the presence of increasing concentrations of HPC 370; (D) α-synuclein in the presence of increasing concentrations of HPC 370. Protein alone is in black, protein in the presence of 0.05% HPC 370 in red, protein in the presence of 0.2% HPC 370 in dark blue and protein in the presence of 0.5% HPC 370 in light blue.

changes. High molecular weight polymers significantly increase the solution viscosity, and changes in viscosity can influence the kinetics and mechanism of protein folding and aggregation [9,43]. It has been shown that protein folding rate is inversely proportional to viscosity [44]. The same relationship was shown to hold for protein aggregation [45].

Both excluded volume effect and viscosity have been shown to play an important role in the effect of compact polysaccharides like dextran and Ficoll on protein folding and aggregation [22,46]. Ficoll is a compact, highly branched polysaccharide synthesized by crosslinking sucrose with epichlorohydrin [21,47]. Due to its higher density Ficoll has lower viscosity than dextran but generates similar excluded volume effect. Higher viscosity of dextran has been proposed to explain differential effects of these polymers on several processes including IAPP aggregation [19] and β-hairpin peptide folding [22]. We have examined the effect of Ficoll 400 on aggregation of α-synuclein and insulin at pH 2.5 and compared its effect to the effect of dextran 500, a polymer with similar molecular weight (Fig. S3). We have observed that Ficoll promoted aggregation of these proteins more effectively than dextran. This difference could be due to higher viscosity of dextran solutions partially counteracting the excluded volume effect.

Since solutions of cellulose derivatives have much higher viscosity than dextran, viscosity should play an even more important role for HPCs. In order to directly compare the effects of dextrans and HPCs with different molecular weights on kinetics of aggregation at the same solution viscosity, we have replotted the nucleation rates for α-synuclein and α-lactalbumin from Fig. 3 using the solution viscosities of polymers as the x axis (Fig. 5A–D). The data (Fig. 5A–B) confirms our earlier observation that the effect of dextrans on protein aggregation is not directly proportional to the viscosity of dextran solutions. The data for HPCs (Fig. 5C–D) shows significant inhibition of fibril formation at low polymer viscosity followed by less pronounced inhibition at higher viscosity values. Inhibition of aggregation at high viscosity values could conceivably be due to the effect of viscosity since highly viscous solutions have been shown to inhibit protein aggregation [9,45], and this inhibition was specifically attributed to solution viscosity [45]. However, significant inhibition of aggregation by low concentrations of HPCs cannot be attributed to bulk solution viscosity, and other factors are likely to be involved.

3.6. Effects of crowding agents on fibril morphology, stability and secondary structure

Changes in the kinetics of aggregation often lead to alteration in the secondary structure and/or morphology of the resulting aggregates. We have examined the changes in morphology of fibrils caused by the aggregation in the presence of polysaccharides. We found that fibrils of insulin at pH 2.5 and α-synuclein formed in the presence of dextrans were shorter than in the absence of polysaccharides but otherwise their morphology remained unchanged. α-Synuclein fibrils and fibrils of insulin at pH 7.5 were longer in the presence of HPCs than in the absence of polymers (Fig. 6). α-Lactalbumin formed very short fibrils in the absence of polymers that appeared longer when formed in the

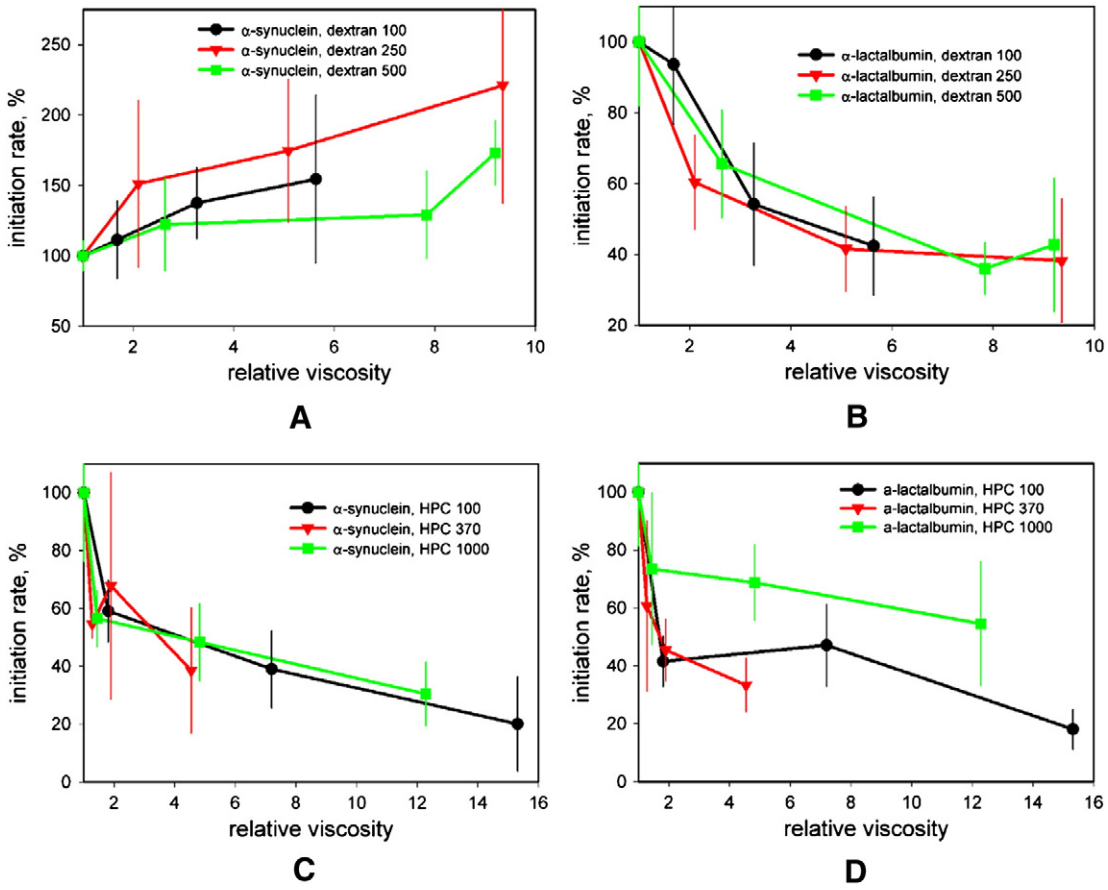


Fig. 5. Fibril nucleation rates of α-synuclein and α-lactalbumin in the presence of polysaccharides plotted vs viscosity of polysaccharide solution. (A) Nucleation rates of α-synuclein in the presence of dextrans. (B) Nucleation rates of α-lactalbumin in the presence of dextrans. Dextran 100 is in black, dextran 250 in red and dextran 500 in green. (C) Nucleation rates of α-synuclein in the presence of HPCs. (D) Nucleation rates of α-lactalbumin in the presence of HPCs. HPC 100 is in black, HPC 370 in red and HPC 1000 in green.

presence of either polymer. Histone fibrils appeared unchanged in the presence of dextrans and somewhat shorter in the presence of HPCs. These morphological changes correlated to some extent with the kinetic data (Fig. 3) as shorter lag phase of aggregation is usually associated with higher fragmentation rate for fibrils and thus shorter fibrils. Correlation was quite good for dextrans where proteins that formed shorter fibrils (insulin at pH 2.5, α-synuclein and histones) aggregated faster. For HPCs correlation was weaker but it could be due to the fact that these polymers inhibited both nucleation and elongation of fibrils.

We have also examined the effect of polysaccharides on the stability of fibrils formed by several proteins towards denaturation. Fibrils were denatured by the increasing concentrations of guanidinium thiocyanate (GdnSCN) for 1 h, and the sample was then diluted to 0.2–0.25 M GdnSCN and ThT (10 μM) was added. ThT fluorescence was used as a readout of the amount of fibrils remaining after the denaturation. The data shows that the effects of additives on the fibril stability strongly depended on the protein.

Addition of either dextran 250 or HPC 370 significantly decreased the stability of insulin fibrils formed at pH 2.5 (Fig. 7A): $C_{1/2}$ values for these fibrils were 0.65–0.7 M vs 0.85 M for insulin fibrils formed in the absence of polysaccharides. α-Synuclein fibrils formed in the presence of either polysaccharide (Fig. 7B) were also significantly less stable: $C_{1/2}$ 0.45–0.6 M vs 1.1 M for α-synuclein alone. The difference in stability was even more pronounced for this protein compared to insulin correlating with stronger effect of polysaccharides on the kinetics of α-synuclein aggregation. Stability of histone fibrils was largely unchanged by the presence of polysaccharides (Fig. 7C) with $C_{1/2}$ of 0.95 M. Overall, we see significantly lower stability towards denaturation for insulin and α-synuclein (although not histone) fibrils formed in the presence of polysaccharides.

We have also examined the secondary structure of insulin and histone fibrils formed in the presence of polysaccharides with FTIR spectroscopy (Fig. 7D–E). Overall, the spectra of amide I band were dominated by the absorbance peak at 1625–1630 cm^{-1} corresponding to intermolecular parallel β-sheets as expected for amyloid fibrils [31,48,49]. In the insulin fibrils formed in the presence of polysaccharides this peak was shifted to slightly higher wavenumber (from 1626 cm^{-1} to 1628 cm^{-1}) indicating stronger carbonyl bonds and weaker intersheet hydrogen bonds. This is an indication of weakening of β-sheet structure in the insulin fibrils formed in the presence of polysaccharides. These results are consistent with the lower stability of insulin fibrils formed in the presence of polysaccharides as observed in the denaturation experiments (Fig. 7A). For histone fibrils similar shift (from 1626 to 1629 cm^{-1}) was observed in the presence of HPC although not in the presence of dextran. In addition, additional peaks at 1645–50 cm^{-1} corresponding to random coils appeared in the presence of both polysaccharides. These peaks are likely due to disruption of the fibrillar structure since the spectra of soluble oligomers would not be detectable in the assay conditions. Overall, FTIR spectra indicate structural disruption in the fibrils formed in the presence of polysaccharides.

4. Discussion

We have shown that biopolymers have variable effects on protein aggregation depending on the structure of the polymer. We have analyzed polymers of two types: relatively flexible dextrans and rigid cellulose derivatives (HPCs). Since these polymers have the same charge and hydrophobicity, they presented a good test case of the effect of shape and structural rigidity of polymers on protein aggregation. We have

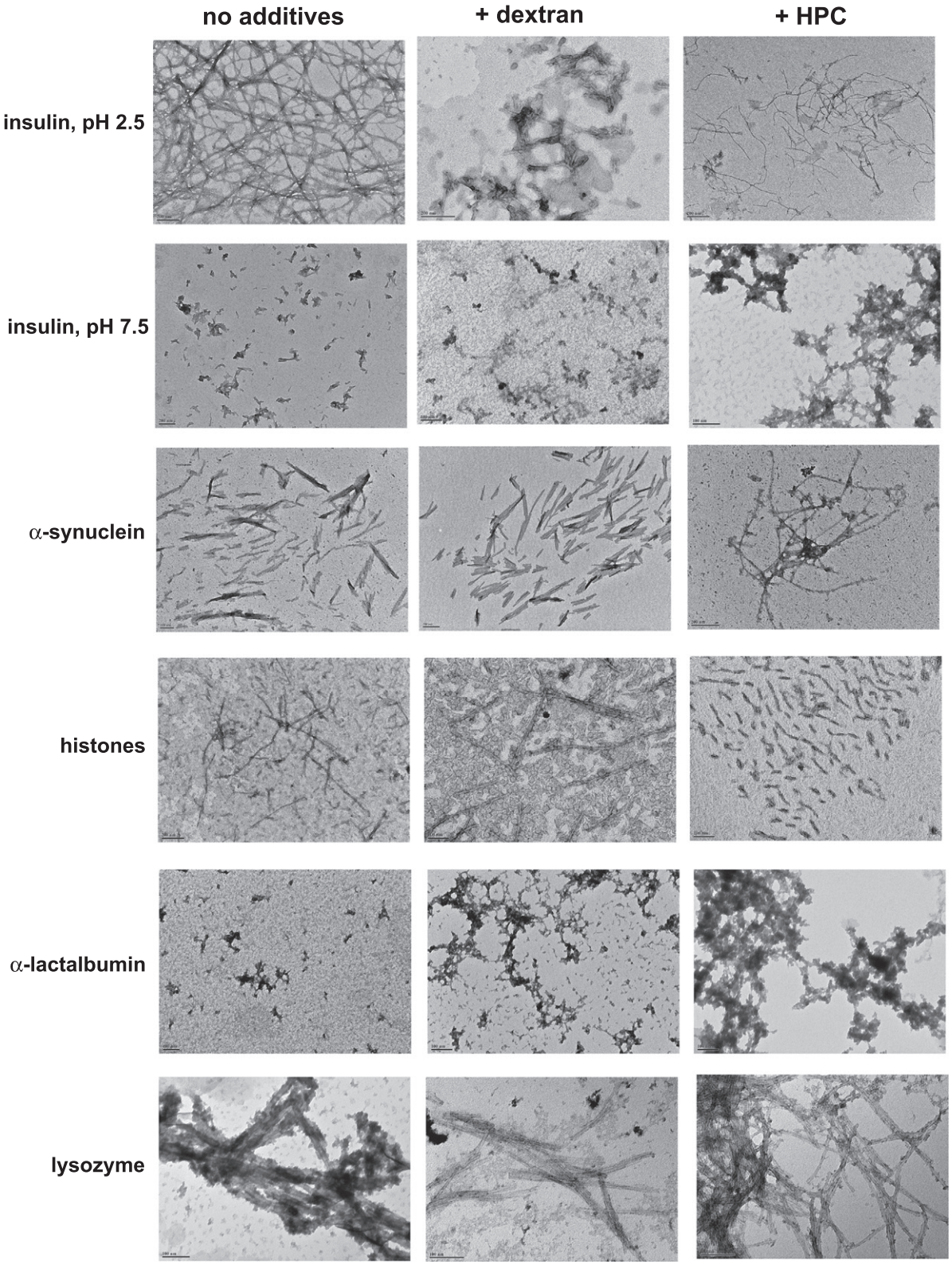


Fig. 6. TEM analysis of protein aggregates formed in the presence of crowding agents. TEM images of aggregates of insulin at pH 2.5, insulin at pH 7.5, α-synuclein, histones, α-lactalbumin and lysozyme formed in the presence of either dextran or HPC.

also selected several proteins with different degrees of intrinsic disorder and in different oligomeric states. In the aggregation conditions insulin is a folded monomer at low pH and a folded hexamer at neutral pH, lysozyme is a folded monomer, histones are a mixture of disordered monomers and oligomers, α-synuclein is a disordered monomer, and reduced α-lactalbumin is a partially disordered monomer that rapidly forms oligomers.

Our results show that in the presence of dextrans fibril formation from the proteins that are either present as stable oligomers (insulin, pH 7.5) or rapidly form stable oligomers in the process of aggregation

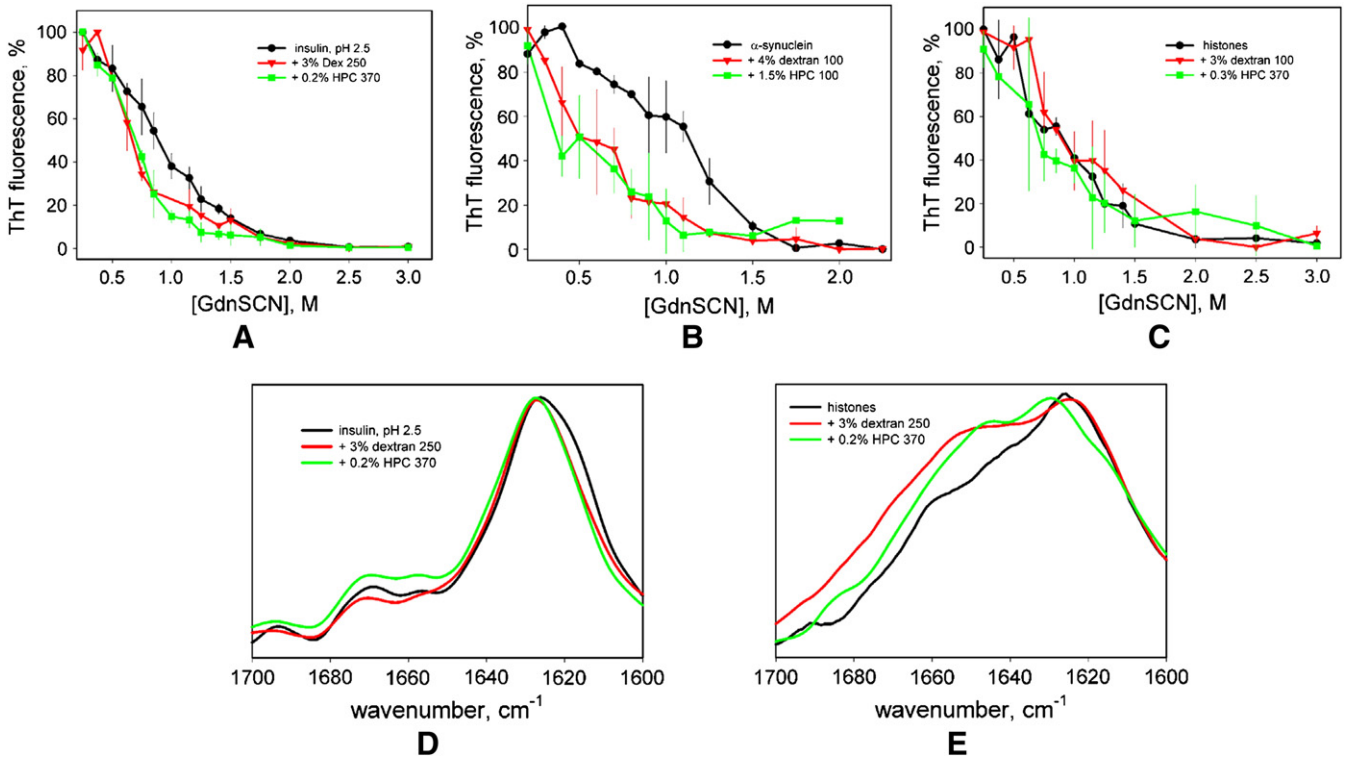


Fig. 7. Denaturation of amyloid fibrils in the presence of guanidinium thiocyanate shows that fibrils formed in the presence of crowding agents are generally less stable. (A) Insulin, pH 2.5: a denaturation curve for fibrils formed in the absence of polysaccharides is in black, in the presence of 3% dextran 250 — in red, in the presence of 0.2% HPC 370 — in light green. (B) α-Synuclein: a denaturation curve for fibrils formed in the absence of polysaccharides is in black, in the presence of 4% dextran 100 — in red, in the presence of 1.5% HPC 100 — in light green. (C) Histones: a denaturation curve for fibrils formed in the absence of polysaccharides is in black, in the presence of 3% dextran 100 — in red, in the presence of 0.3% HPC 370 — in light green. FTIR spectra of insulin (D) and histone (E) fibrils confirm this observation. FTIR spectra of fibrils formed in the absence of polysaccharides are in black, in the presence of 3% dextran 250 — in red, and in the presence of 0.2% HPC 370 — in light green.

(α-lactalbumin) was inhibited. On the other hand, fibril formation from IDPs (histones and α-synuclein) was accelerated. Effect on monomeric folded proteins (lysozyme, insulin at pH 2.5) was intermediate with either weak acceleration or weak inhibition. This result is consistent with stabilization of the most stable and compact protein conformations by dextrans via excluded volume effect. While we didn't see the effect of dextrans on the secondary structures of most proteins, the effect on the less stable aggregation intermediates is likely to be a lot more significant than on the ground state conformations. For IDPs more compact intermediates are likely to be on pathway to amyloid formation due to absence of stable compact conformations of these proteins. For natively folded proteins excluded volume effect may stabilize the intermediates either on the pathway to folding or to misfolding and aggregation. Structural differences between the protein conformations leading to normal folding and aggregation are often subtle. For example, Kay and coworkers have shown [50] that in a mutant SH3 domain a small structural change allows the protein to switch between folding and aggregation pathways.

HPCs inhibited fibril formation from all proteins except histones. All aggregation parameters were affected with significant increases in the lag phase and decreases in ThT fluorescence and the fibril elongation rate. Decreased ThT fluorescence could indicate a significantly decreased yield of fibrils in the presence of HPCs. However, ThT fluorescence is not a very reliable indicator as it can be affected by displacement of ThT from fibrils or a structural change in fibrils affecting ThT binding. Lack of inhibition of histone aggregation by HPCs may be due to inability of histones to fold in the assay conditions.

It has been previously shown that in relation to their sensitivity to crowded environments, IDPs can be grouped into two classes, foldable and non-foldable [51,52]. Foldable IDPs can gain some structure in crowded conditions due to formation of a hydrophobic core. Non-foldable IDPs have a higher charge density and lower hydrophobicity that the foldable ones and remain unstructured even at the crowded conditions. Out of IDPs examined here histones are known to be non-foldable [53–55] while α-synuclein is known to be foldable [29,56,57]. As expected, CD data (Fig. 1E) didn't indicate any effect of crowding agents on histones. Effects of polymers on α-synuclein at pH 3.5, on the other hand, have been quite significant. This result correlates well with well-known high conformational flexibility of foldable IDPs in general [58] and α-synuclein specifically [29,57] compared to both non-foldable IDPs and folded proteins. At neutral pH the conformation of α-synuclein is much less sensitive to the environment as observed in our NMR experiments (Fig. 2) and reported previously [9,29] due to higher charge–charge repulsion at neutral pH stabilizing the disordered protein conformation.

What factors played a role in the effect of polysaccharides on protein aggregation? Acceleration of protein aggregation (Figs. 3–4) in the presence of dextrans is consistent with the excluded volume effect being the primary factor for this polymer. However, in addition to steric repulsion, attractive interactions between proteins and other polymers are always present [59]. Pielak and coworkers [38,39,60] found that the addition of compact polysaccharides to proteins leads not only to entropic changes expected for excluded volume effect but also to significant enthalpic changes attributed to chemical interactions between proteins and polysaccharides. These interaction don't necessarily involve specific binding and are likely to be similar to preferential solvation of proteins by alcohols in alcohol–water mixtures [61,62] or preferential association of chaotropic ions with polysaccharides in aqueous solutions [63].

Non-specific chemical interactions are likely play an even more important role for HPCs since they have a higher surface to volume ratio

compared to dextrans. Inhibition of protein aggregation by low concentrations of HPCs is an indication that direct protein–polymer interactions play a more important role for these polymers than excluded volume effect. At high HPC concentrations bulk solution viscosity is likely to contribute to inhibition of protein aggregation by these polymers as the extent of inhibition becomes proportional to solution viscosity (Fig. 6). Since rigid polymer molecules increase the solvent viscosity in their vicinity [13], it is possible that interaction with HPC molecules keeps the protein molecules in the environment that is more viscous than bulk solution further inhibiting their aggregation.

Both FTIR and fluorescence spectroscopy indicate that fibril stability was lower and intermolecular hydrogen bonds were weaker for fibrils formed in the presence of polysaccharides. A possible explanation of this effect is that the changes in the secondary structure of the aggregation intermediates in the presence of crowding agents lead to formation of fibrillar structures that are far from optimal for this protein resulting in less stable and potentially more pathogenic [64] fibrils.

Overall, we found that effects of biopolymers on protein aggregation are more complex than it was previously believed. Compact hydrophilic polymers act primarily via excluded volume effect increasing effective protein concentration and stabilizing compact conformations. As a result IDP aggregation is accelerated while the effects on aggregation of folded proteins are mixed depending on the stability of the folded conformation of the protein and its tendency to form off-pathway oligomers. We have shown that rigid polymers strongly inhibit aggregation of most proteins. Factors such as protein–polymer interactions and solution viscosity are likely to be responsible for this inhibition. Both types of polysaccharides were shown to alter the structure of amyloid fibrils resulting in less stable and potentially more pathogenic fibrils. Biological environment contains a variety of biopolymers with different structural properties. Many of these biopolymers such as nucleic acids, fibrous proteins and polysaccharides comprising the extracellular matrix are structurally rigid. Our results indicate that biopolymers influence protein aggregation not only via excluded volume effect but rather via a complex interplay of several effects including excluded volume effect, solution viscosity and protein–polymer interaction.

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.bbapap.2013.11.004.

Acknowledgements

We wish to thank Dr. Mazdak Khajehpour (University of Manitoba, Canada) and Tatiana Miti (University of South Florida) for helpful suggestions, Zhenling Peng (University of Alberta, Canada) for help with some experiments and Leonardo Gonnelli (CERM) for the preparation of the ^{13}C, ^{15}N α-synuclein samples. This work has been supported in part by the EC Projects BioNMR (Contract # 261863) and IDPbyNMR (Contract # 264257).

References

[1] V.N. Uversky, A.L. Fink, Conformational constraints for amyloid fibrillation: the importance of being unfolded, Biochim. Biophys. Acta 1698 (2004) 131–153.
[2] C. Liu, M. Zhao, L. Jiang, P.N. Cheng, J. Park, M.R. Sawaya, A. Pensalfini, D. Gou, A.J. Berk, C.G. Glabe, J. Nowick, D. Eisenberg, Out-of-register beta-sheets suggest a pathway to toxic amyloid aggregates, Proc. Natl. Acad. Sci. U. S. A. 109 (2012) 20913–20918.
[3] A. Laganowsky, C. Liu, M.R. Sawaya, J.P. Whitelegge, J. Park, M. Zhao, A. Pensalfini, A.B. Soriaga, M. Landau, P.K. Teng, D. Cascio, C. Glabe, D. Eisenberg, Atomic view of a toxic amyloid small oligomer, Science 335 (2012) 1228–1231.
[4] C.G. Glabe, Structural classification of toxic amyloid oligomers, J. Biol. Chem. 283 (2008) 29639–29643.
[5] R. Tycko, Molecular structure of amyloid fibrils: insights from solid-state NMR, Q. Rev. Biophys. 39 (2006) 1–55.
[6] H.X. Zhou, G. Rivas, A.P. Minton, Macromolecular crowding and confinement: biochemical, biophysical, and potential physiological consequences, Annu. Rev. Biophys. 37 (2008) 375–397.
[7] R.J. Ellis, A.P. Minton, Protein aggregation in crowded environments, Biol. Chem. 387 (2006) 485–497.
[8] L.A. Munishkina, A. Ahmad, A.L. Fink, V.N. Uversky, Guiding protein aggregation with macromolecular crowding, Biochemistry 47 (2008) 8993–9006.
[9] L.A. Munishkina, E.M. Cooper, V.N. Uversky, A.L. Fink, The effect of macromolecular crowding on protein aggregation and amyloid fibril formation, J. Mol. Recognit. 17 (2004) 456–464.
[10] C.F. Lee, S. Bird, M. Shaw, L. Jean, D.J. Vaux, Combined effects of agitation, macromolecular crowding, and interfaces on amyloidogenesis, J. Biol. Chem. 287 (2012) 38006–38019.
[11] Q. Ma, J.B. Fan, Z. Zhou, B.R. Zhou, S.R. Meng, J.Y. Hu, J. Chen, Y. Liang, The contrasting effect of macromolecular crowding on amyloid fibril formation, PLOS One 7 (2012) e36288.
[12] P.S. Yeung, P.H. Axelsen, The crowded environment of a reverse micelle induces the formation of beta-strand seed structures for nucleating amyloid fibril formation, J. Am. Chem. Soc. 134 (2012) 6061–6063.
[13] P.J. Flory, Principles of Polymer Chemistry, Cornell University Press, Ithaca, 1953.
[14] C. Huang, G. Ren, H. Zhou, C.C. Wang, A new method for purification of recombinant human alpha-synuclein in *Escherichia coli*, Protein Expr. Purif. 42 (2005) 173–177.
[15] S.E. Hill, T. Miti, T. Richmond, M. Muschol, Spatial extent of charge repulsion regulates assembly pathways for lysozyme amyloid fibrils, PLoS One 6 (2011) e18171.
[16] J.A. Cohlberg, J. Li, V.N. Uversky, A.L. Fink, Heparin and other glycosaminoglycans stimulate the formation of amyloid fibrils from alpha-synuclein in vitro, Biochemistry 41 (2002) 1502–1511.
[17] L. Breydo, O.V. Bocharova, I.V. Baskakov, Semiautomated cell-free conversion of prion protein: applications for high-throughput screening of potential antiprion drugs, Anal. Biochem. 339 (2005) 165–173.
[18] R.L.J. Keller, The Computer Aided Resonance Assignment Tutorial, Cantina Verag, 2004.
[19] J. Seeliger, A. Werkmuller, R. Winter, Macromolecular crowding as a suppressor of human IAPP fibril formation and cytotoxicity, PLOS One 8 (2013) e69652.
[20] Y. Hirata, Y. Sano, M. Aoki, H. Shohji, S. Katoh, J. Abe, S. Hitsukuri, H. Yamamoto, Small-angle X-ray scattering studies of moderately concentrated dextran solution, Carbohydr. Polym. 53 (2003) 331–335.
[21] D. Venturoli, B. Rippe, Ficoll and dextran vs. globular proteins as probes for testing glomerular permselectivity: effects of molecular size, shape, charge, and deformability, Am. J. Physiol. Renal Physiol. 288 (2005) F605–F613.
[22] S. Mukherjee, M.M. Waegele, P. Chowdhury, L. Guo, F. Gai, Effect of macromolecular crowding on protein folding dynamics at the secondary structure level, J. Mol. Biol. 393 (2009) 227–236.
[23] J.K. Armstrong, R.B. Wenby, H.J. Meiselman, T.C. Fisher, The hydrodynamic radii of macromolecules and their effect on red blood cell aggregation, Biophys. J. 87 (2004) 4259–4270.
[24] M. James, Physical Properties of Polymers Handbook, American Institute of Physics Press, Woodbury, NY, 1996.
[25] J.L. Whittingham, D.J. Scott, K. Chance, A. Wilson, J. Finch, J. Brange, G.G. Dodson, Insulin at pH 2: structural analysis of the conditions promoting insulin fibre formation, J. Mol. Biol. 318 (2002) 479–490.
[26] B. Fauvet, M.K. Mbefo, M.B. Fares, C. Desobry, S. Michael, M.T. Ardah, E. Tsika, P. Coune, M. Prudent, N. Lion, D. Eliezer, D.J. Moore, B. Schneider, P. Aebischer, O.M. El-Agnaf, E. Masliah, H.A. Lashuel, Alpha-synuclein in central nervous system and from erythrocytes, mammalian cells, and *Escherichia coli* exists predominantly as disordered monomer, J. Biol. Chem. 287 (2012) 15345–15364.
[27] T. Bartels, J.G. Choi, D.J. Selkoe, Alpha-synuclein occurs physiologically as a helically folded tetramer that resists aggregation, Nature 477 (2011) 107–110.
[28] L. Breydo, J.W. Wu, V.N. Uversky, α-Synuclein misfolding and Parkinson's disease, Biochim. Biophys. Acta 1822 (2012) 261–285.
[29] V.N. Uversky, J. Li, A.L. Fink, Evidence for a partially folded intermediate in alpha-synuclein fibril formation, J. Biol. Chem. 276 (2001) 10737–10744.
[30] K. Luger, A.W. Mader, R.K. Richmond, D.F. Sargent, T.J. Richmond, Crystal structure of the nucleosome core particle at 2.8 Å resolution, Nature 389 (1997) 251–260.
[31] L.A. Munishkina, A.L. Fink, V.N. Uversky, Conformational prerequisites for formation of amyloid fibrils from histones, J. Mol. Biol. 342 (2004) 1305–1324.
[32] A.P. Butler, R.E. Harrington, D.E. Olins, Salt-dependent interconversion of inner histone oligomers, Nucleic Acids Res. 6 (1979) 1509–1520.
[33] S. Kawashima, K. Imahori, Studies on histone oligomers. II. Reconstitution of heterotype histone oligomers and pH dependency of oligomer formation, J. Biochem. (Tokyo) 88 (1980) 783–788.
[34] T.B. Topping, L.M. Gloss, The impact of solubility and electrostatics on fibril formation by the H3 and H4 histones, Protein Sci. 20 (2011) 2060–2073.
[35] M. Malisauskas, V. Zamotin, J. Jass, W. Noppe, C.M. Dobson, L.A. Morozova-Roche, Amyloid protofilaments from the calcium-binding protein equine lysozyme: formation of ring and linear structures depends on pH and metal ion concentration, J. Mol. Biol. 330 (2003) 879–890.
[36] E.A. Permyakov, L.J. Berliner, Alpha-lactalbumin: structure and function, FEBS Lett. 473 (2000) 269–274.
[37] J. Li, S. Zhang, C. Wang, Only the reduced conformer of alpha-lactalbumin is inducible to aggregation by protein aggregates, J. Biochem. (Tokyo) 129 (2001) 821–826.
[38] Y. Wang, M. Sarkar, A.E. Smith, A.S. Krois, G.J. Pielak, Macromolecular crowding and protein stability, J. Am. Chem. Soc. 134 (2012) 16614–16618.
[39] L.A. Benton, A.E. Smith, G.B. Young, G.J. Pielak, Unexpected effects of macromolecular crowding on protein stability, Biochemistry 51 (2012) 9773–9775.
[40] W. Bermel, M. Bruix, I.C. Felli, M.V.V. Kumar, R. Pierattelli, S. Serrano, Improving the chemical shift dispersion of multidimensional NMR spectra of intrinsically disordered proteins, J. Biomol. NMR 55 (2013) 231–237.
[41] S. Sukenik, R. Politi, L. Ziserman, D. Danino, A. Friedler, D. Harries, Crowding alone cannot account for cosolute effect on amyloid aggregation, PLOS One 6 (2011) e15608.

[42] D.L. Zhang, L.J. Wu, J. Chen, Y. Liang, Effects of macromolecular crowding on the structural stability of human alpha-lactalbumin, Acta Biochim. Biophys. Sin. (Shanghai) 44 (2012) 703–711.
[43] M. Jacob, F.X. Schmid, Protein folding as a diffusional process, Biochemistry 38 (1999) 13773–13779.
[44] A. Ansari, C.M. Jones, E.R. Henry, J. Hofrichter, W.A. Eaton, The role of solvent viscosity in the dynamics of protein conformational changes, Science 256 (1992) 1796–1798.
[45] M. Sleutel, A.E.S. Van Driessche, W.C. Pan, E.K. Reichel, D. Maes, P.G. Vekilov, Does solution viscosity scale the rate of aggregation of folded proteins? J. Phys. Chem. Lett. 3 (2012) 1258–1263.
[46] H. Neuweiler, M. Lollmann, S. Doose, M. Sauer, Dynamics of unfolded polypeptide chains in crowded environment studied by fluorescence correlation spectroscopy, J. Mol. Biol. 365 (2007) 856–869.
[47] J.R. Wenner, V.A. Bloomfield, Crowding effects on EcoRV kinetics and binding, Biophys. J. 77 (1999) 3234–3241.
[48] W. Dzwolak, S. Grudzielanek, V. Smirnovas, R. Ravindra, C. Nicolini, R. Jansen, A. Loksztejn, S. Porowski, R. Winter, Ethanol-perturbed amyloidogenic self-assembly of insulin: looking for origins of amyloid strains, Biochemistry 44 (2005) 8948–8958.
[49] M. Bouchard, J. Zurdo, E.J. Nettleton, C.M. Dobson, C.V. Robinson, Formation of insulin amyloid fibrils followed by FTIR simultaneously with CD and electron microscopy, Protein Sci. 9 (2000) 1960–1967.
[50] P. Neudecker, P. Robustelli, A. Cavalli, P. Walsh, P. Lundstrom, A. Zarrine-Afsar, S. Sharpe, M. Vendruscolo, L.E. Kay, Structure of an intermediate state in protein folding and aggregation, Science 336 (2012) 362–366.
[51] V.N. Uversky, Intrinsically disordered proteins and their environment: effects of strong denaturants, temperature, pH, counter ions, membranes, binding partners, osmolytes, and macromolecular crowding, Protein J. 28 (2009) 305–325.
[52] J.L. England, G. Haran, To fold or expand — a charged question, Proc. Natl. Acad. Sci. U. S. A. 107 (2010) 14519–14520.
[53] K. Luger, Structure and dynamic behavior of nucleosomes, Curr. Opin. Genet. Dev. 13 (2003) 127–135.
[54] Z. Peng, M.J. Mizianty, B. Xue, L. Kurgan, V.N. Uversky, More than just tails: intrinsic disorder in histone proteins, Mol. Biosyst. 8 (2012) 1886–1901.
[55] H.J. Li, R. Wickett, A.M. Craig, I. Isenberg, Conformational changes in histone IV, Biopolymers 11 (1972) 375–397.
[56] A.C. Ferreon, A.A. Deniz, Alpha-synuclein multistate folding thermodynamics: implications for protein misfolding and aggregation, Biochemistry 46 (2007) 4499–4509.
[57] A. Deleersnijder, M. Gerard, Z. Debyser, V. Baekelandt, The remarkable conformational plasticity of alpha-synuclein: blessing or curse? Trends Mol. Med. 19 (2013) 368–377.
[58] V.N. Uversky, Under-folded proteins: conformational ensembles and their roles in protein folding, function, and pathogenesis, Biopolymers 99 (2013) 870–887.
[59] M. Sarkar, C. Li, G.J. Pielak, Soft interactions and crowding, Biophys. Rev. 5 (2013) 187–194.
[60] Y. Wang, L.A. Benton, V. Singh, G.J. Pielak, Disordered protein diffusion under crowded conditions, J. Phys. Chem. Lett. 3 (2012) 2703–2706.
[61] K. Gekko, T. Morikawa, Preferential hydration of bovine serum albumin in polyhydric alcohol–water mixtures, J. Biochem. (Tokyo) 90 (1981) 39–50.
[62] J. Chik, S. Mizrahi, S. Chi, V.A. Parsegian, D.C. Rau, Hydration forces underlie the exclusion of salts and of neutral polar solutes from hydroxypropylcellulose, J. Phys. Chem. B 109 (2005) 9111–9118.
[63] K.D. Collins, Sticky ions in biological systems, Proc. Natl. Acad. Sci. U. S. A. 92 (1995) 5553–5557.
[64] D.W. Colby, K. Giles, G. Legname, H. Wille, I.V. Baskakov, S.J. DeArmond, S.B. Prusiner, Design and construction of diverse mammalian prion strains, Proc. Natl. Acad. Sci. U. S. A. 106 (2009) 20417–20422.

The crowd you're in with: Effects of different types of crowding agents on protein aggregation

Leonid Breydo[a], Krishna D. Reddy[a], Alessandro Piai[b], Isabella C. Felli[b], Roberta Pierattelli[b], Vladimir N. Uversky[a,c,d]

a Department of Molecular Medicine, Morsani College of Medicine, University of South Florida, Tampa, FL 33612, USA

b Center for Magnetic Resonance, University of Florence, 50019 Sesto Fiorentino, Florence, Italy

c Byrd Alzheimer's Research Institute, Morsani College of Medicine, University of South Florida, Tampa, FL 33612, USA

d Institute for Biological Instrumentation, Russian Academy of Sciences, 142292 Pushchino, Moscow Region, Russia

SUPPLEMENTARY MATERIAL

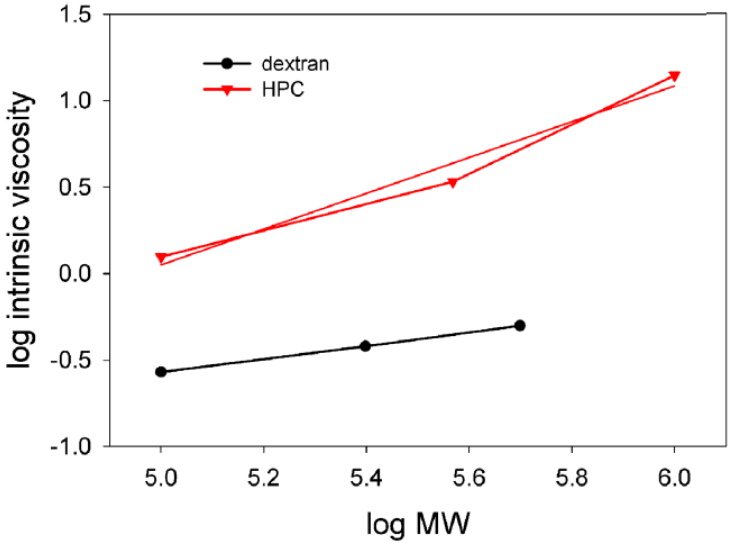


Fig S1. Intrinsic viscosity of crowding agents. The black curve (dots) refers to hydroxypropyl cellulose, whereas the red one (triangles) to dextran.

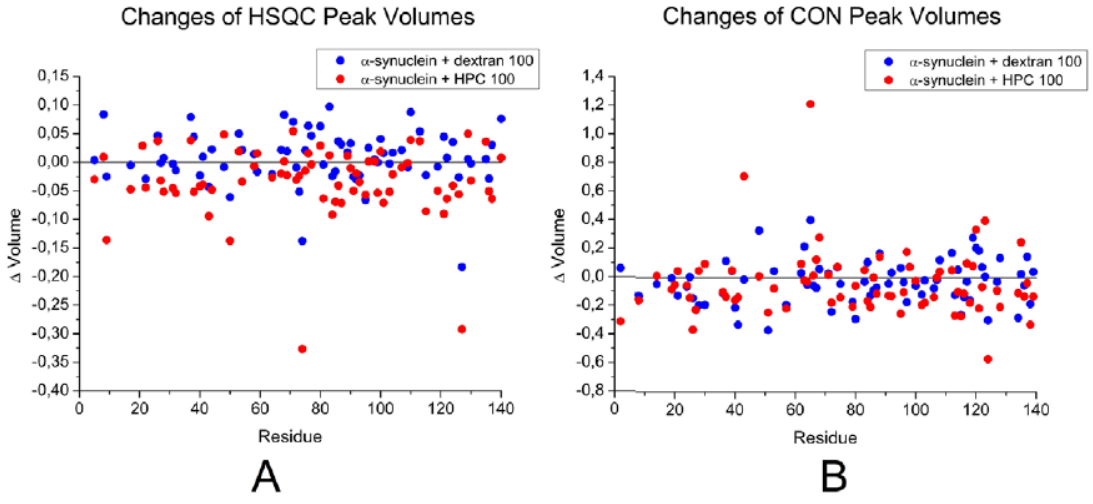


Fig S2. Chemical shift index (CSI) analysis and variations of peak volumes of $^{13}C/^{15}N$ labeled α-synuclein at pH 6.5 in the presence of crowding agents. The CSI analysis was performed using data collected from 2D C–N correlation spectra (A–D). ^{13}C CSI of α-synuclein alone and in the presence of 5% dextran 100 is compared in (A), whereas ^{13}C CSI of α-synuclein alone and in the presence of 2.5% HPC 100 in (B). Analogously, ^{15}N CSI is compared in (C) and (D). The changes of peak volumes in HSQC and CON spectra of α-synuclein after the addition of the crowder (blue is 5% dextran 100 and red is 2.5% HPC 100) are reported in panels (E) and (F), respectively. All the analysis excluded relevant changes due to the presence of crowding agents.

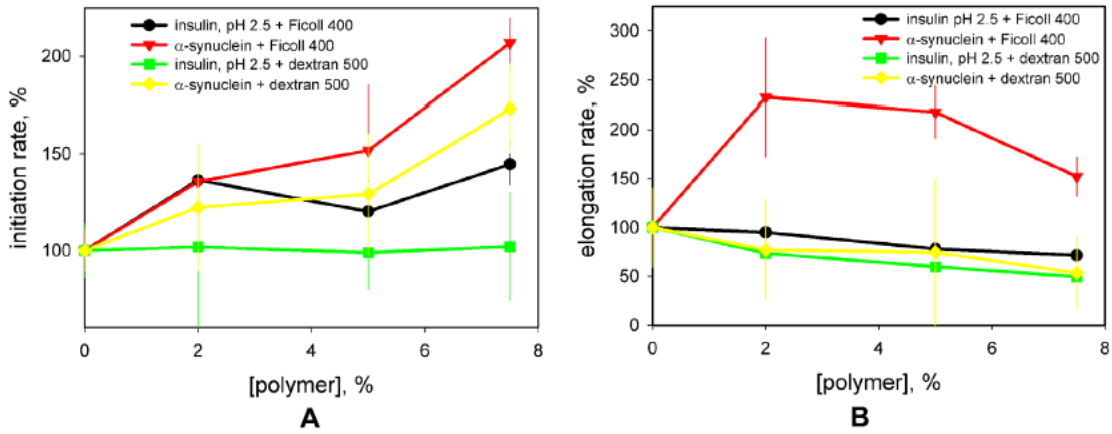


Fig S3. Comparison of effects of dextran and Ficoll on fibril formation from insulin (pH 2.5) and α-synuclein. (A) Initiation rate of fibril formation in the presence of increasing concentrations of Ficoll 400 or dextran 500; (B) elongation rate of fibrils in the presence of increasing concentrations of Ficoll 400 or dextran 500. Insulin at pH 2.5 in the presence of Ficoll 400 is shown in black, α-synuclein in the presence of Ficoll — in red, insulin in the presence of dextran 500 is in light green and α-synuclein in the presence of dextran 500 is in yellow.

Table S1. Kinetic data on aggregation of insulin (at pH 2.5 and pH 7.5), α-synuclein, histones, α-lactalbumin and lysozyme in the presence of crowding agents.

		insulin, pH 2.5			insulin, pH 7.5			α-synuclein			histones			α-lactalbumin			lysozyme		
		initiation rate	elongation rate	yield	initiation rate	elongation rate	yield	initiation rate	elongation rate	yield	initiation rate	elongation rate	yield	initiation rate	elongation rate	yield	initiation rate	elongation rate	yield
Dextran 100	2%	100±20	70±15	110±18	90±23	66±16	96±8	111±28	99±42	107±43	106±12	93±26	102±14	94±17	58±22	87±30	85±19	94±30	32±1
	5%	111±18	60±7	120±11	79±11	48±11	91±16	138±25	73±38	117±30	152±57	35±11	125±54	54±17	38±12	51±26	84±37	83±52	30±4
	7.5%	138±34	70±5	146±49	81±19	38±41	83±21	155±60	37±21	113±57	237±22	46±26	108±26	42±14	27±11	43±17	64±36	58±40	37±0
Dextran 250	2%	65±14	65±12	150±30	63±18	68±43	91±23	151±59	77±31	107±27	120±15	120±27	82±6	60±13	73±12	114±28	88±16	84±52	40±23
	5%	70±12	54±15	127±28	55±16	54±42	67±22	175±51	43±21	116±49	220±24	121±75	66±20	42±12	22±7	40±15	83±34	49±34	41±34
	7.5%	65±11	52±9	114±41	55±15	23±12	59±43	221±84	40±14	124±80	234±52	182±32	30±15	38±17	29±26	38±30	96±45	45±30	76±52
Dextran 500	2%	102±42	74±23	78±16	63±24	60±11	79±19	122±33	77±47	111±79	132±34	86±33	89±16	66±15	67±24	90±35	68±13	98±45	28±4
	5%	99±19	60±13	107±34	60±26	35±12	77±6	129±31	75±71	104±62	200±19	137±84	40±28	36±7	20±17	55±23	76±5	57±27	40±6
	7.5%	102±28	50±7	103±41	55±30	27±16	61±20	173±23	53±38	128±72	295±56	129±57	44±30	43±19	15±35	34±32	60±16	36±25	117±13
HPC 100	0.5%	57±24	19±6	66±45	83±46	37±28	16±6	59±11	27±15	129±71	117±22	89±27	24±11	42±9	38±21	80±48	48±4	6±3	179±73
	2%	73±50	20±9	66±48	73±47	25±27	22±10	39±13	18±5	94±36	110±56	68±38	7±4	47±14	29±17	70±50	31±2	5±1	11±8
	5%	25±5	7±3	29±37	39±7	16±4	31±6	20±16	9±5	102±46				18±7	15±40	24±17			
HPC 370	0.05%	46±6	17±6	53±11	68±31	18±8	31±28	55±5	56±22	128±40	133±22	84±61	38±8	61±29	54±33	53±22	44±16	44±33	13±6
	0.02%	40±5	14±2	89±27	78±30	25±11	13±11	68±39	31±11	102±54	127±15	62±23	56±14	46±11	54±23	71±42	43±13	13±3	40±8
	0.5%	34±3	8±1	81±19	75±9	16±13	8±8	39±22	26±16	70±41	183±57	69±20	24±9	33±9	39±24	72±40	30±4	17±14	17±4
HPC 1000	0.025%	44±11	25±12	57±40	60±30	23±14	34±23	57±10	64±32	62±28	117±15	115±27	41±8	74±26	39±9	128±8	35±5	22±5	162±69
	0.1%	42±12	15±4	39±17	64±17	17±11	43±42	48±13	28±15	76±37	84±20	156±66	28±34	69±13	32±9	114±22	35±8	16±12	456±227
	0.25%	49±8	10±7	29±22	46±9	10±4	26±29	30±11	37±8	74±47	95±13	65±31	50±23	55±21	16±7	66±30	47±24	42±28	423±51

Article 7. Just a flexible linker? The structural and dynamic properties of CBP-ID4 as revealed by NMR spectroscopy

Alessandro Piai[1], Eduardo O. Calçada[1], Thomas Tarenzi[1], Alessandro Del Grande[1], Mihaly Varadi[2], Peter Tompa[2,3], Isabella C. Felli[1], Roberta Pierattelli[1]

[1] CERM and Department of Chemistry "Ugo Schiff", University of Florence, 50019, Via Luigi Sacconi 6, Sesto Fiorentino, Florence, Italy.

[2] Structural Biology Research Center (SBRC), Flemish Institute of Biotechnology (VIB), Brussels, Belgium; Structural Biology Brussel (SBB), Vrije Universiteit Brussel (VUB), Brussels, Belgium.

[3] Institute of Enzymology, Research Centre for Natural Sciences, Hungarian Academy of Sciences, Budapest, Hungary.

Manuscript submitted

Abstract

Here, we present the structural and dynamic description of CBP-ID4 at atomic resolution. ID4 is the fourth intrinsically disordered linker of the CREB-Binding Protein (CBP). In spite of the largely disordered nature of CBP-ID4, NMR chemical shifts and relaxation measurements show a significant degree of α-helix sampling in the protein regions encompassing residues 2-25 and 101-128 (1852-1875 and 1951-1978 in full-length CBP). Proline residues are uniformly distributed along the polypeptide, except for the two α-helical regions, indicating that they play an active role in modulating the structural features of this CBP fragment. The two helical regions are lacking known functional motifs, suggesting that they represent thus-far uncharacterized functional modules of CBP. The present work provides novel insights regarding the functions of this protein linker, which may exploit its plasticity to modulate the relative orientations of neighboring folded domains of CBP and fine tune its interactions with a multitude of partners.

Keywords

Intrinsically disordered proteins; IDPs; CREB-binding protein; CBP; prolines

Introduction

Large proteins are often composed by several folded domains separated in the primary sequence by flexible linkers. While folded domains are generally well characterized at atomic resolution, only little information is available for linkers. These protein segments seldom crystallize, and they are often quite challenging to be characterized by nuclear magnetic resonance (NMR) spectroscopy because of their peculiar amino acid composition and their structural and dynamic properties. However, it is unlikely that linkers play passive connecting roles only, because they often constitute as much as half of the primary sequence of complex proteins. In fact, their amino acid sequences often show functional features such as interaction motifs, post-translational modification sites and conservation in the primary sequence across species (1, 2). To learn more on the role of these protein elements it is important to focus on their experimental characterization, exploiting recently developed NMR tools (3-14). With this in mind we decided to focus on a well-studied key protein in biological systems, the transcription factor CREB-Binding Protein (CBP).

CBP and its paralog p300 are transcriptional co-regulators that integrate signals from numerous signal transduction pathways, and play critical roles in basic cellular processes ranging from development and differentiation to DNA repair (15). Their biological function is related to their ability to interact with a large number of proteins through multiple protein-interaction domains as well as to their acetyl-transferase activity. There are seven autonomous folded domains in CBP/p300, the 3D structures of which have been determined in recent years either by X-ray crystallography or NMR (16-24). Four of them require zinc(II) ions in order to adopt a stable fold: the transcriptional-adaptor zinc-finger-1 domain (TAZ1), the plant homeodomain (PHD), the zinc-binding domain near the dystrophin WW domain (ZZ) and the transcriptional-adaptor zinc-finger-2 domain (TAZ2); other folded domains are the KID-binding domain (KIX), the bromodomain and the histone acetyl-transferase domain

(HAT). Finally, a domain named nuclear-receptor coactivator-binding domain (NCBD) is intrinsically disordered, but folds upon binding to its partner(s).

Regions between CBP folded domains represent more than 50% of the total 2,442 residues of the protein; these are predicted to be mostly intrinsically disordered (25). In fact, the folded domains listed above are spatially separated by five linkers of different length, denoted as CBP-ID# (where ID stands for *Intrinsically Disordered* and # represents the number of the linker), which have not been yet characterized at atomic resolution. Very little is known about their structural and functional roles, which may range from maintaining a specific distance between the various folded domains to fine-tuning and modulation of interaction processes. The second hypothesis is supported by the fact that the primary sequence of CBP linkers is well conserved in evolution.

The amino acid composition of CBP linkers is biased towards disorder-promoting amino acids, as typically found in the case of intrinsically disordered proteins (IDPs) (26-28). Overall, they contain a high proportion of prolines (16%), glutamines (15%), serines (11%), glycines (9%) and alanines (9%). In particular, CBP-ID4 and CBP-ID5 exhibit the most distinctive amino acid composition, since almost 40% of their primary sequences is constituted by prolines and glutamines.

In this work we present the characterization by NMR spectroscopy of CBP-ID4, the 207 residues long linker (CBP residues 1851-2057) located between the TAZ2 and NCBD domains. Among all CBP linkers, ID4 is the one containing the highest percentage of proline residues (22%) and it is predicted to be intrinsically disordered, two aspects that make it a challenging target for NMR. To obtain its sequence specific assignment we thus used a strategy tailored for IDPs, combining ^{1}H and ^{13}C detection and multidimensional experiments (14, 29, 30 and references therein), which resulted in the complete NMR characterization of this linker.

We believe the findings described in the present work will open new frontiers towards a deeper understanding of the roles of CBP linkers and represent the demonstration that NMR can provide a wealth of information on IDPs difficult to access with other techniques.

Materials and Methods

Protein expression and purification

The recombinant vector pET21a-CBP-ID4 containing the human CBP-ID4 native gene was transformed into *E. Coli* BL21 (DE3) for protein expression. For ^{15}N- and ^{13}C,^{15}N-labeled protein production, a colony from a freshly transformed plate was selected to inoculate in LB medium (50 ml) containing ampicillin and was grown overnight at 37°C and 180 rpm. Cells from the overnight LB growths were diluted 1:100 into 4 l of LB medium and grown at 37°C with constant agitation at 160 rpm. Once an OD600 of 0.7–0.8 was reached, the cells were gently centrifuged at 4500 rpm for 20 min. The pellet was resuspended in 1 l of minimal medium (48.5 mM Na_2HPO_4, 22.0 mM K_2HPO_4, 8.5 mM NaCl, 0.2 mM $CaCl_2$, 2.0 mM $MgSO_4$, 1 mg l^{-1} each of biotin and thiamin, 7.5 mM $(^{15}NH_4)_2SO_4$ and 11.1 mM glucose/$^{13}C_6$-glucose) and grown at 37 °C for 1 hour with constant agitation at 160 rpm, following the Marley's method (31). The cells were induced with 0.5 mM IPTG and allowed to grow for additional 4 hours at 30°C. The culture was then harvested at 8000 rpm for 20 min and the pellet was stored at -20°C. Frozen cells were thawed and suspended in 45 ml of equilibration buffer A (50 mM MES, 10 mM EDTA, 20 mM NaCl, pH 5.5). Cells were disrupted by sonication on ice (at 80% sonication power) with cycles of 6 s with 4 s delay pulses for 25 min. Lysed cells were centrifuged at 40000 rpm for 40 min at 4°C and the supernatant (≈ 40 ml) was warmed at

80°C for 20 min in order to remove contaminants. This solution was spun at 8000 rpm for 20 min at 4°C and the supernatant was filtered through a 0.22 µm pore membrane in preparation for the purification steps. Cation exchange chromatography was performed on a 6 ml Resource S column (GE Healthcare) pre-equilibrated with buffer A. A linear gradient between buffer A and buffer B (buffer A with 0.5 M NaCl) was applied over 20 min at a flow rate of 2.5 ml/min. The CBP-ID4 was eluted with a salt concentration around 300 mM. Fractions were analyzed by SDS–PAGE and those containing the target protein were concentrated until 2 ml. The sample was further purified by gel filtration on Hiload 16100 superdex 75 (GE Healthcare) in 20 mM potassium phosphate buffer at pH 6.5, 100 mM NaCl, 50 µM EDTA. The fractions containing pure CBP-ID4 were concentrated up to 2 ml and stored at 4°C for biophysical analysis. Mass spectrometry (MALDI) was performed to confirm the molecular mass of purified CBP-ID4.

NMR samples

All the multidimensional NMR experiments for sequence-specific assignment were performed on a sample of 0.6 mM uniformly ^{13}C, ^{15}N labeled human CBP-ID4 in 20 mM potassium phosphate buffer, 100 mM KCl at pH 6.5, with 10% D_2O added for the lock. ^{15}N relaxation experiments were acquired on a 0.5 mM uniformly ^{15}N labeled human CBP-ID4, in the same experimental conditions. For all the experiments, 3 mm NMR sample tubes were used to reduce the detrimental effects of high salt concentration.

NMR data acquisition

A set of multidimensional $^{1}H^{N}$ and ^{13}C detected NMR experiments tailored to achieve sequence-specific assignment of IDPs was acquired at 283.0 K. 2D CON-IPAP (32, 33), 3D (H)CBCACON-IPAP (34), 3D (H)CBCANCO-IPAP (34), 4D (HCA)CON(CA)CON-IPAP (35) and 4D (HN)CON(CA)CON-IPAP (35) experiments were acquired at 16.4 T on a Bruker Avance spectrometer operating at 700.06 MHz ^{1}H, 176.03 MHz ^{13}C and 70.94 MHz ^{15}N frequencies, equipped with a cryogenically cooled probehead optimized for ^{13}C-direct detection. 2D BEST-TROSY (BT) (11, 36), 3D BT-HNCO (11, 37, 38), 3D BT-HN(CA)CO (11, 38), 3D BT-HNCACB (11, 38), 3D BT-HN(CO)CACB (11, 38), 3D BT-(H)N(COCA)NH (11) and 3D BT (H)N(CA)NNH (39) experiments were performed at 21.1 T on a Bruker Avance spectrometer operating at 898.57 MHz ^{1}H, 225.95 MHz ^{13}C and 91.05 MHz ^{15}N frequencies, equipped with a cryogenically cooled probehead. 5D BT-(H)NCO(CAN)CONH (40) and 5D BT-HN(COCAN)CONH (40) experiments were collected at 22.3 T on a Bruker Avance III spectrometer operating at 950.20 MHz ^{1}H, 238.93 MHz ^{13}C and 96.28 MHz ^{15}N frequencies, equipped with a cryogenically cooled probehead.

Heteronuclear ^{15}N relaxation experiments (41-44) for the measurement of ^{15}N R_1, ^{15}N R_2 and ^{15}N-^{1}H NOEs, (CLEANEX-PM)-FHSQC experiments (45) for the estimation of the extent of amide proton exchange with the solvent and the 3D HNHA experiment (46) for the determination of homonuclear $^{3}J_{HN\text{-}H\alpha}$ were measured at 283.0 K at 16.4 T on a Bruker Avance spectrometer operating at 700.13 MHz ^{1}H, 176.05 MHz ^{13}C and 70.94 MHz ^{15}N frequencies, equipped with a cryogenically cooled probehead.

The temperature dependence of CBP-ID4 was investigated by acquiring a series of 2D BEST-TROSY and 2D CON-IPAP spectra in the range 238.0-308.0 K, with steps of 5 degrees, at 22.3 T on a Bruker Avance III spectrometer operating at 950.20 MHz ^{1}H, 238.93 MHz ^{13}C and 96.28 MHz ^{15}N frequencies, equipped with a cryogenically cooled probehead.

The parameters used for the acquisition of all the experiments are reported in the Supporting Material (Tables S1-S5). All the data sets were acquired using *Bruker TopSpin 1.3* or *3.1* software. 3/4/5D experiments for the sequence-specific assignment were performed using on-grid non-uniform sampling (NUS). The on-grid ''Poisson disk'' sampling scheme (47) was chosen to generate the time schedules with the *RSPack* program [J. Stanek, A. Zawadzka-Kazimierczuk, unpublished]. The distribution was relaxation-optimized, i.e. the density of points was decaying according to the Gaussian distribution $\exp(-t^2/\sigma^2)$, with σ=0.5.

NMR data processing and analysis

Conventionally sampled NMR data sets were processed using *Bruker TopSpin 1.3* software. Instead, when NUS was employed, the NMR data were converted with *nmrPipe* (48) and then processed using either the Multidimensional Fourier Transform (MFT) algorithm (for 3D data sets) or the Sparse MFT (SMFT) algorithm (for 4/5D data sets), respectively implemented in *ToASTD* (49) and *reduced* (50, 51) programs. Both programs are available at http://nmr.cent3.uw.edu.pl.

CARA (52) and *Sparky* (53) were used to analyze 3D and 4/5D spectra, respectively, whereas *CcpNmr Analysis* (54) was employed to analyze ^{15}N relaxation data.

The secondary structure propensity from the heteronuclear chemical shifts was determined by using the neighbor corrected structural propensity calculator (ncSPC) tool (55), available online at http://nmr.chem.rug.nl. The Tamiola, Acar and Mulder random coil chemical shift library (56) was chosen for the analysis.

Theoretical helical propensities were calculated using the Agadir algorithm (57-61), available online at http://agadir.crg.es. For the calculation the pH was set to 6.5, the temperature to 283.0 K and the ionic strength to 0.1 M.

Intrinsic protein disorder was predicted by using IUPred (62) and PONDR-FIT (63) tools, available at http://iupred.enzim.hu and http://www.disprot.org.

For the conservation analysis, a comprehensive data set of 39 experimentally validated homologous amino acid sequences was assembled by performing a PSI-BLAST (64) against the non-redundant protein data set of NCBI. The query sequence was the ID4 segment of the human CBP protein (residues 1852-2057) [UniProt ID: Q92793], and the retrieved sequences covered all major vertebrate taxonomic groups from mammals to fish. The multiple sequence alignment was generated using MAFFT (65) and then used as input for the sequence- and disorder conservation analysis. We used the local version of the DisCons tool to analyze the multiple sequence alignment (66). DisCons was used with default parameters, namely predicting disorder with IUPred, and quantifying the sequence conservation with Jensen-Shannon divergence; the maximum allowed fraction of gaps for a position was set to 0.6.

The presence of linear motifs and functional sites was investigated by using the Eukaryotic Linear Motif (ELM) computational biology resource (67), available online at http://elm.eu.org.

Potential serine, threonine and tyrosine phosphorylation sites were predicted exploiting NetPhos 2.0 (68), available online at http://www.cbs.dtu.dk. Potential kinase specific phosphorylation sites were predicted by using using NetPhosK 1.0 (69), available online at http://www.cbs.dtu.dk.

To obtain the ensemble description of the protein linker under investigation 10,000 conformers were generated using the program *flexible-meccano* (70), available online at http://www.ibs.fr. The SSP score calculated from experimentally measured chemical shifts was used as input in the calculation to allow also the generation of conformers whose backbone dihedral angles deviate substantially from those typically found in pure random coils. The ensemble and the underlying data were deposited in the Protein Ensemble Database (http://pedb.vib.be, 71) under the accession ID PED5AAE.

Results and Discussion

NMR sequence-specific resonances assignment

The 2D BEST-TROSY and 2D CON-IPAP fingerprint spectra of CBP-ID4 are reported in Figure 1. The ^{1}H-^{15}N correlation spectrum exhibits all the peculiar NMR features of an IDP, namely the reduced chemical shift dispersion particularly pronounced in the ^{1}H dimension, and the high degree of signal crowding. The 2D CON-IPAP spectrum benefits from the larger spectral width in the ^{15}N dimension, which includes signals due to proline residues nitrogen nuclei, and the fact that the experiment provides inter-residues correlations, which are more resolved with respect to intra-residue ones (72). A series of triple-resonance 3/4/5D NMR experiments (see Materials and Methods Section) specifically designed for IDPs, exploiting either $^{1}H^{N}$ or ^{13}C detection, was acquired. Non-uniform sampling (NUS) (73) was exploited to reduce the experimental time while preserving high spectral resolution in the indirect dimensions (74).

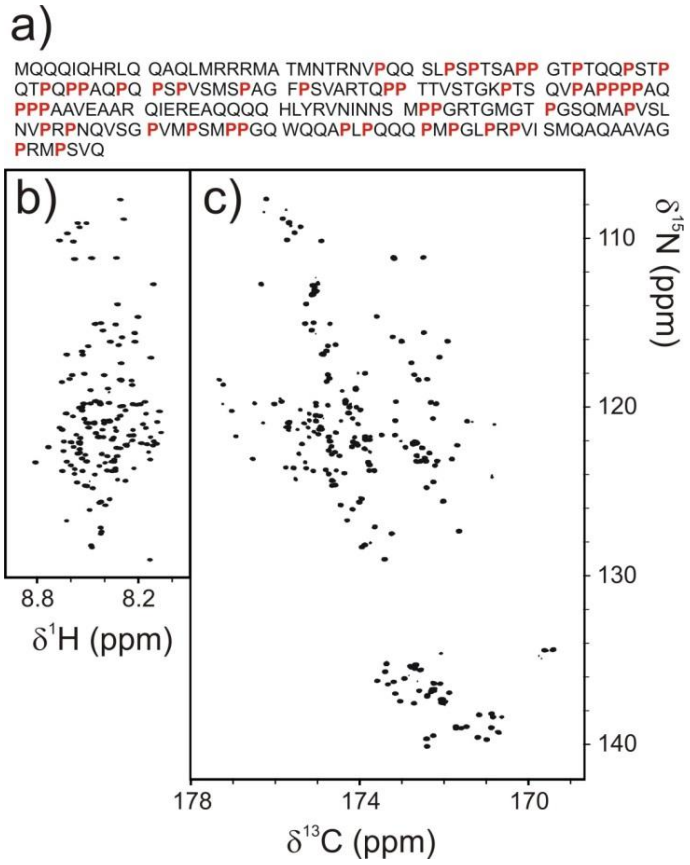


Figure 1. Amino acid sequence and fingerprint spectra of CBP-ID4. a) primary sequence of CBP-ID4, b) 2D BEST-TROSY and c) 2D CON-IPAP spectra of CBP-ID4. Both spectra are reported with the same Hz/cm ratio in the direct dimension.

The ^{13}C-detected 3D and 4D spectra (34, 35) provided the key information to achieve the sequence-specific assignment of this proline-rich IDP. The 3D BEST-TROSY (BT) triple-resonance experiments (11) were acquired to obtain the complementary information needed to complete the assignment, further supported by the 5D BT-(H)NCO(CAN)CONH and 5D BT-HN(COCAN)CONH experiments (40), used to resolve a few ambiguities and to confirm the chemical shift assignment obtained with the 3D spectra.

This set of spectra tailored for the study of IDPs made it possible to assign the resonances belonging to all H^N, N, C′, C^α and C^β nuclei. The chemical shift assignment of CBP-ID4, both at 283.0 and 308.0 K, is deposited in the *BioMagResBank* (BMRB, http://www.bmrb.wisc.edu, (75), entries 26616 and 26639.

NMR characterization

The presence of secondary structure propensity within CBP-ID4 was investigated by comparing experimentally measured heteronuclear chemical shifts (N, C′, C^α and C^β nuclei) to the corresponding random coil values (55). The obtained differences (Figure 2) are distributed around zero, consistently with the disordered nature of CBP-ID4. However, two regions, encompassing residues 2-25 and 101-128, exhibit significantly positive values suggesting a high propensity to form α-helices in these regions.

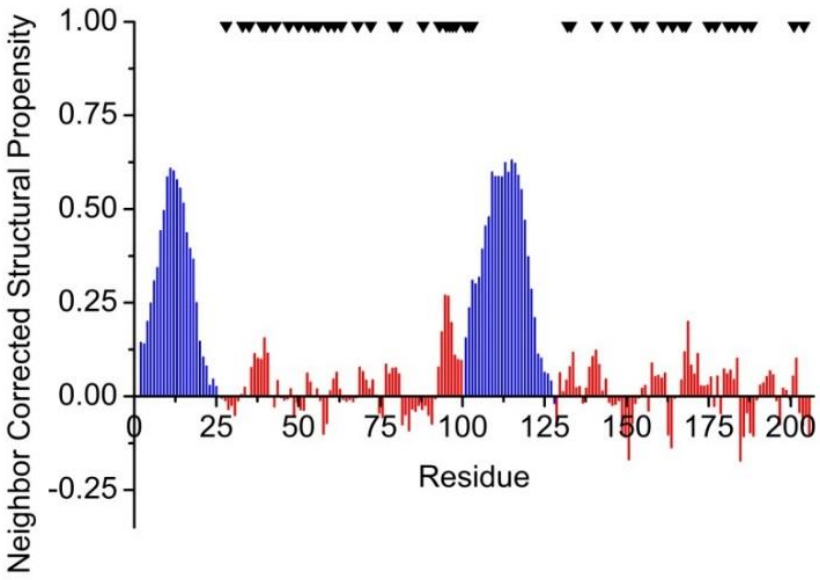


Figure 2. Secondary structure propensity score obtained from experimentally measured N, C′, C^α and C^β chemical shifts. Positive and negative values correspond to α-helical and β-sheet propensities, respectively. The two protein regions which exhibit significant propensity for α-helix are colored in blue. Proline residues are highlighted along the protein sequence as black triangles on the top and result clustered in the most disordered regions of the protein.

The $^3J_{HN\text{-}H\alpha}$ coupling constants (46) shown in Figure 3, provide complementary information to detect contingent residual structural propensity along the polypeptide chain, and are therefore often used in structural characterizations of chemically unfolded or intrinsically disordered proteins (76-80). The obtained average value of about 7 Hz is similar to values

obtained for flexible proteins and other disordered peptides (79, 80). Only two protein regions (residues 13-16 and 106-118) substantially deviate from this trend, corresponding to those of highest predicted α-helical propensity. Therefore, conformations with values for the angle ϕ towards 60° are increasingly being populated in these protein regions. Interestingly, also residues 92-99 show significantly lower values, suggesting that a certain amount of structural organization may be present there too, as revealed by the SSP analysis (Figure 2). From $^{3}J_{HN\text{-}H\alpha}$ values alone it is not possible to conclude whether this include canonical α-helical or polyproline type II (PPII) structures (81, 82).

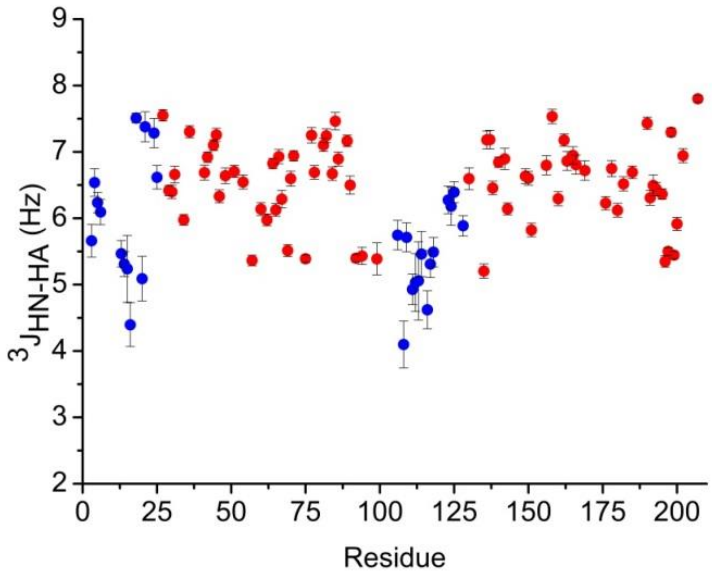


Figure 3. $^{3}J_{HN\text{-}H\alpha}$ coupling constants for the intrinsically disordered protein domain CBP-ID4. The two protein regions characterized by higher propensity for α-helix are highlighted in blue.

Backbone dynamics of CBP-ID4 was investigated through the measurement of ^{15}N R_1, ^{15}N R_2 relaxation rates and heteronuclear ^{15}N-^{1}H NOE values. Relaxation parameters, determined for the majority of the assigned cross-peaks in the 2D ^{1}H-^{15}N HSQC spectrum, are reported in Figure 4. Considerably higher ^{15}N R_2 and ^{15}N-^{1}H NOE values are observed in the regions encompassing residues 2-25 and 101-128, confirming a reduced flexibility in these regions likely related to the observed α-helical propensity (Figure 2). Interestingly, the magnitude of the ^{15}N R_2 values indicates that the transient α-helix located between residues 101-128 is more rigid than that comprising residues 2-25. ^{15}N-^{1}H NOE values reveal that the fragments separating the two partially populated α-helices are highly flexible, with the first fragment (residues 26-103) more flexible than the second one (residues 129-207).

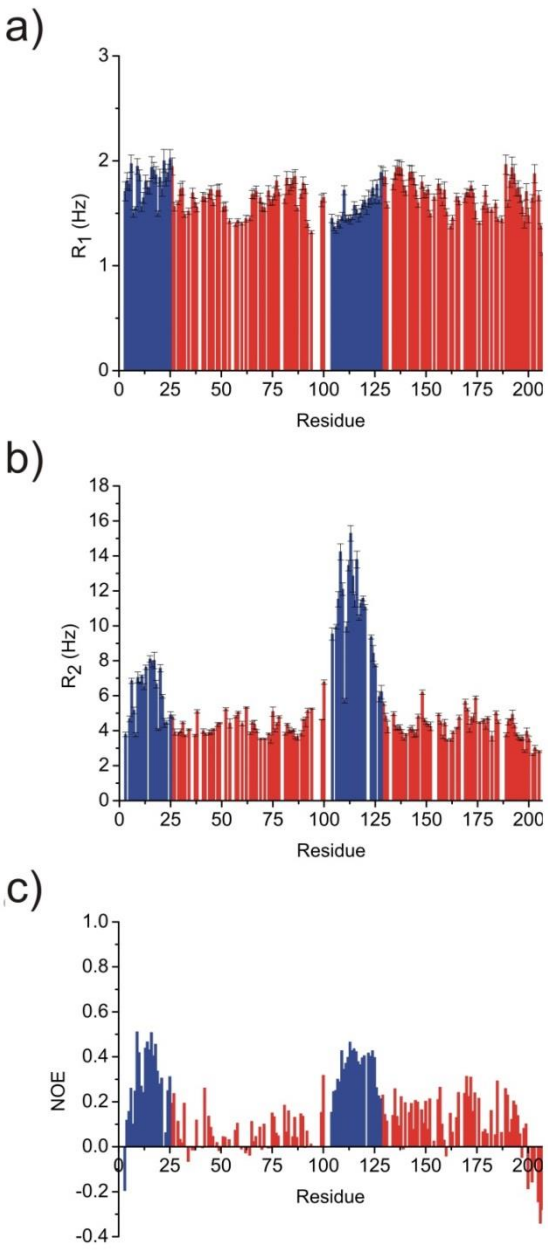


Figure 4 Dynamic characterization of CBP-ID4. The two protein regions characterized by higher propensity for α-helix are highlighted in blue. a) ^{15}N R_1 relaxation rates; b) ^{15}N R_2 relaxation rates; c) ^{15}N-^{1}H NOEs. All ^{15}N relaxation measurements were carried out at 16.4 T at 283.0 K.

Local solvent accessibility was investigated by monitoring H_2O exchange processes for backbone amide protons (45). The results, reported in Figure 5, show that exchange effects are overall very pronounced, becoming operative for the majority of the residues already with 16 ms of mixing time (τ_m). Notably, the two protein regions exhibiting significant α-helical propensity, in particular the one encompassing residues 101-128, which is more rigid by ^{15}N R_2 measurements, appears to exchange less with the solvent; on the contrary, the other residues of the polypeptide chain are characterized by high-solvent exchange effect.

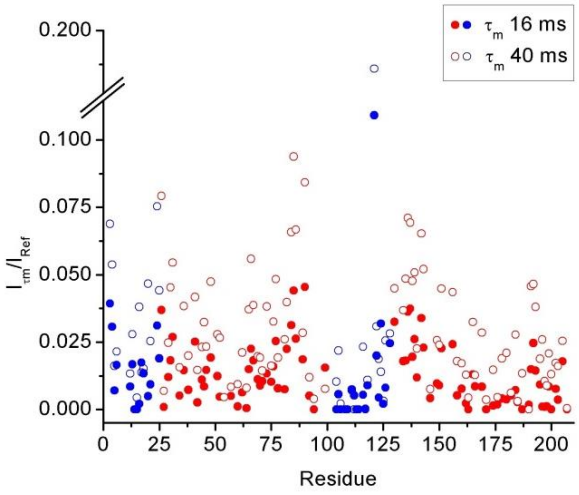


Figure 5. Extent of amide proton exchange with the solvent. The residues belonging to the two protein regions characterized by high propensity for α-helix are shown in blue, the others in red. $I\tau_m/I_{Ref}$ values measured through the (CLEANEX-PM)-FHSQC experiments are reported as a function of residue number for two different mixing-times, namely 16 ms (filled dots) and 40 ms (empty dots). Error bars have been removed to facilitate the comparison and are provided as Supporting Material (Figure S1).

The temperature dependence of CBP-ID4 was investigated by acquiring a series of 2D BEST-TROSY and 2D CON-IPAP spectra in the range 283.0-308.0 K (shown in Figures S2 and S3 in the Supporting Material). These spectra enabled the transfer of the resonances assignment obtained at 283.0 K to the other temperatures by following the chemical shift changes of the cross-peaks with temperature. A 3D CBCACON experiment was acquired to confirm the signals assignment at 308.0 K.

In Figure 6a the difference of cross-peak intensities in the 2D BEST-TROSY NMR spectra acquired at 308.0 and 283.0 K is reported. The majority of cross-peaks become weaker or broadened with increasing temperature as a consequence of the higher rate of amide proton exchange with the solvent protons. Instead, the signals belonging to residues 101-128, those exhibiting a significant propensity to form α-helix, become more intense at high temperature. This evidence suggests that this protein region results less affected by chemical exchange processes and benefits from an improved correlation time resulting from the faster rotational tumbling. Interestingly, an opposite behavior is observed in the other protein region characterized by high α-helix propensity (residues 2-25): it is destabilized by increasing temperature and experiences a decrease in signal intensities similar to that of the other unstructured regions of the protein.

Correlations between amide proton temperature coefficients ($\Delta\delta H^N/\Delta T$) and hydrogen bonds (83, 84) were also investigated (Figure 6b). Less negative temperature coefficients are clustered in the two protein regions exhibiting higher propensity to form α-helix, in agreement with the results of the other analysis. In particular, in the case of the transient α-helix encompassing residues 101-128, the great majority of amide proton temperature coefficients result less negative than -4.6 ppb/K, a value considered the threshold below which the presence of intramolecular hydrogen bonds can be excluded with a predictive value of 85% (84). Therefore, transient intra-molecular hydrogen bonds are likely to be present in this protein region, which again appears more structured than the region comprising residues 2-25.

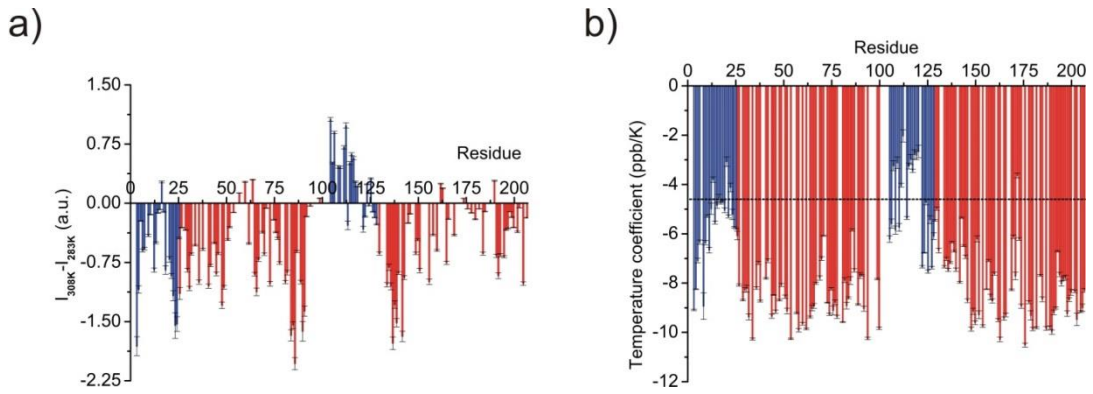


Figure 6. Temperature dependence of CBP-ID4. The two protein regions experimentally identified as exhibiting high propensity to form α-helix are colored in blue. a) cross-peak intensity changes in the 2D BEST-TROSY spectrum acquired at 308.0 and 283.0 K; b) temperature coefficients ($\Delta\delta H^N/\Delta T$) reported as a function of residue number. The black dashed line at -4.6 ppb/K represents the threshold to discriminate between intra- and inter-molecular hydrogen bonds (with a predictive value of 85%).

Structural and dynamic properties of CPB-ID4

The overall picture resulting from the high resolution experimental characterization of the structural and dynamic properties of CBP-ID4 reveals two well-defined partially populated helical fragments, termed helix I and helix II, which display properties significantly different from one another, separated by segments characterized by high local flexibility. It is interesting to note that prolines, the most abundant amino acid in CBP-ID4 (22%), are distributed quite uniformly throughout the protein primary sequence (Figure 1a) with the exception of the two parts of the protein exhibiting α-helical propensity (Figure 2). Proline residues may be selected against in the α-helical segments because of their helix-breaking properties: their peculiar side chain disrupts the H-bond network and interferes with the backbone helical packing of adjacent residues. Furthermore, it can be noted that 7 proline residues are clustered in the region immediately preceding the inner transient helix II, suggesting that they are involved in its stabilization. In fact, it has been reported that prolines present in proximity of the so-called pre-structured motifs (PreSMos) (85) of IDPs are very important in promoting and delimiting the PreSMos themselves (86). Therefore, the prolines flanking helix II (residues 101-128) seem to play a specific structural role (87). This would also explain why this partially structured protein region appears to be more stable than that comprising residues 2-25, which in the full-length protein is preceded by a structured domain but is here unrestrained.

The degree of α-helical propensity could have an important role in providing specific patterns for partner recognition through a specific surface provided by the helix itself, as shown in the literature in several different examples (25, 88, 89). On the other hand, the transition from a random-coil, extended conformation of the protein backbone to partially formed helices represents the more drastic local change for backbone dihedral angles, with the helical conformation constituting the most compact one (1.5 Å shortening for each amino acid part of the helix with respect to an elongated conformation). The presence of the secondary structure element can also reduce the degree of freedom of the polypeptide, which cannot freely bend to form compact structures. Therefore, the extent of local helical propensity could be a way to modulate the overall length of a specific polypeptide fragment with local cooperative changes in backbone dihedral angles. In this way, the insertion of a partially populated helical segment in between largely random coil backbone fragments,

such as observed for the case of CBP-ID4, could provide a mean to modulate the overall length and orientation of the linker itself.

To visualize the possible conformers populating the ensemble of structures in dynamic equilibrium in solution, the secondary structure propensities determined experimentally for CBP-ID4 were used as input to calculate an ensemble of conformers through the program *Flexible-Meccano* (70), using the statistical coil potential proposed for IDPs. The computed ensemble shows a great variability in the overall radius of gyration (Figure S4 in the Supplementary Material), in agreement with the idea that flexible linkers can indeed fine-modulate the relative distance between folded units (PED ID PED5AAE). The obtained structures can be inspected locally to evaluate the properties of the partially populated helical conformers. In the present case, the two regions characterized by secondary structural propensities feature quite different composition in terms of the constituent amino acids, which has an impact on the properties of the partially populated helices. In particular, helix I (2-25) is characterized by the presence of several charged amino acids, resulting in an overall positive electrostatic potential (Figure 7e), with a quite peculiar methionine-rich patch on one side (3 methionine residues quite close to each other on the helix surface, Figure 7d). Instead, helix II (101-128) has an overall hydrophobic surface (Figure 7i) with some charged residues (three positive arginine residues and three negative glutamic acids residues, Figure 7h). Therefore, the two partially populated helical elements provide significantly different modules for protein-protein interactions. Both helices share a large number of polar amino acids, with a predominance of glutamine residues, which may promote the stability of these two segments.

Although CBP-ID4 has never been characterized in detail before, the crystal structure of the TAZ2 domain of human p300, paralogue of CBP, was solved earlier (19). The construct used in that study, in addition to the TAZ2 domain, included at its C-terminus 21 residues of the subsequent linker, which is homologous to the initial part of CBP-ID4 (16/21 residues are conserved in the two protein sequences). Specifically, these residues correspond to residues 1-21 of CBP-ID4, identified as exhibiting significant α-helical propensity. Interestingly, in the crystal structure these residues assume a α-helix conformation, which extends outside the globular structure of TAZ2 domain. Furthermore, it has been reported that such α-helix contributes actively to the binding of p300 to the transcription factor Myocyte Enhancer Factor 2 (MEF2) (90). This observation fits with the general notion that IDPs often function by binding to partner molecules via structural elements that are dynamically sampled in the disordered ensemble (91). Such motives are conserved structural/functional elements of IDPs, and their presence can be concluded from the structural propensity of certain regions of the IDP chain.

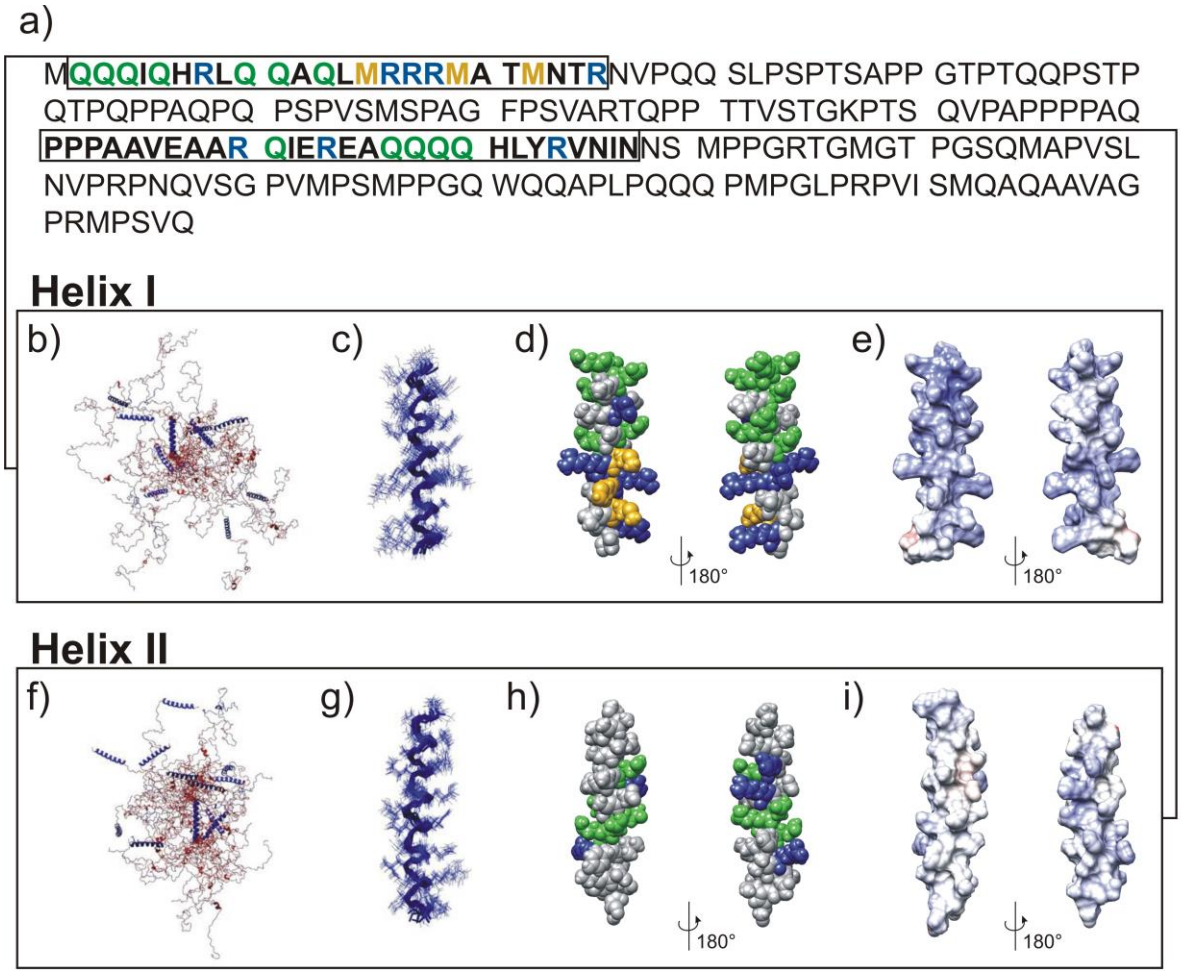


Figure 7. Representative ensemble of conformers for CBP-ID4. a) Location of the helical segments within the primary sequence; glutamine, arginine and methionine residues are shown in green, blue and gold. b) Subset of conformers obtained superimposing residues of helix I (residues 2-25). All the helical segments are shown in blue. c) Close-up of a set of superimposed helix I with amino acids atoms represented as sticks. d) Helix I with amino acids atoms represented as spheres; glutamine, arginine and methionine residues are shown in green, blue and gold. e) Electrostatic potential on the surface of helix I; positively charged, negatively charged, and neutral amino acids are represented in blue, red, and white. Panels f-i) contain similar information of panels b-e) for helix II (residues 101-128). The backbone models were generated with *flexible-meccano* (70), while side-chains and hydrogen atoms were added by using Molmol 2.0 (92); the energy of the models was minimized using Chimera (93) and the electrostatic potentials were computed by using APBS (94).

Prediction of functional regions within CBP-ID4

It is interesting to compare the picture resulting from the experimental characterization of CBP-ID4 with all the information that can be predicted on the basis of the primary sequence, starting from the local propensities for local order/disorder, the local secondary structure, all the way to interaction motifs or post-translational modification sites.

The significant local propensity for secondary structure that was observed in the ensemble model can hint at the presence of binding motifs within the helical regions, as often seen in IDPs (95, 96). Furthermore, the potential functional importance of the two helical regions should also manifest themselves in their sequence conservation. Thus, we have collected and aligned 39 CBP sequences and used the DisCons online tool to quantify the conservation of the amino acid sequence and of the structural disorder (66) (Figure 8 and Figure S5). The position-specific conservation scores of the amino acid sequence of regions 2-25 and 101-

128 are consistently high, and in fact higher than the conservation of disorder, which is suggestive of their functional importance and might indicate that these segments act as molecular recognition features, which are known to have higher sequence conservation and lower disorder with respect to their flanking regions (66).

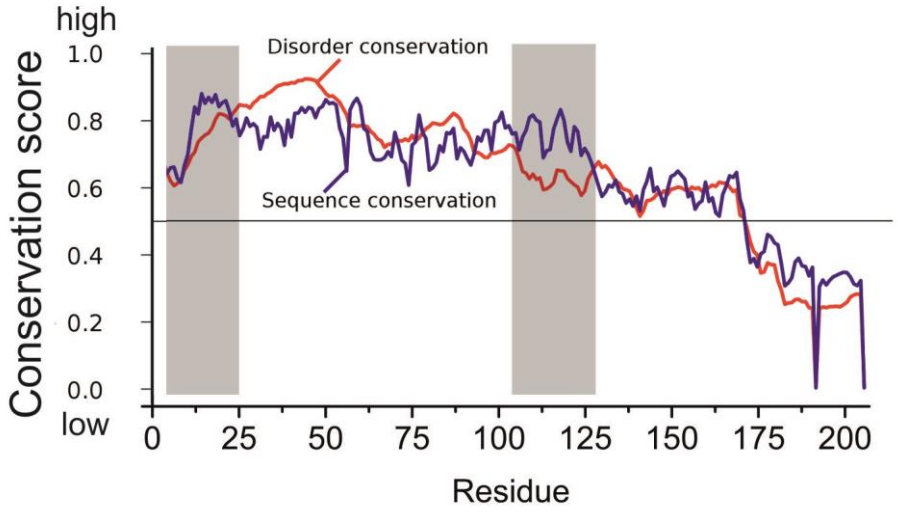


Figure 8 Evolutionary conservation of the amino acidic sequence (blue) and disorder tendency (red) of CBP-ID4.

We have used the Eukaryotic Linear Motif (ELM) database (67) to identify the presence of Short Linear Motifs (SLiMs) in the sequence, since SLiMs are known to be involved in recognition and targeting processes (97). By definition, they are highly specific short structural/functional elements of 3-10 amino acids, mostly located in intrinsically disordered regions of proteins, which are able to mediate protein-protein interactions without the need of a stable three-dimensional structure. Remarkably, more than 95% of the SLiMs identified by ELM are located in the most disordered regions of CBP-ID4 (Figure S6a in the Supplementary Material).

The position of potential serine, threonine and tyrosine phosphorylation sites of CBP-ID4 were also identified using NetPhos 2.0 (68), an artificial neuronal network-based method for predicting phosphorylation sites. The importance of post-translational modification (PTM) sites derives from their frequent involvement in the modulation of protein functions (98-101). The location of the recognized sites (Figure S6b in the Supplementary Material), similarly to SLiMs, are distributed exclusively along the most flexible regions of the protein, and are excluded from the two α-helical segments. In particular, several identified phosphorylation sites are located in the region 26-103 which, by ^{15}N-^{1}H NOE analysis, is the most flexible region of the entire protein (Figure 4c). Therefore, we can infer that the negligible secondary structure content of this protein region may facilitate the interaction between the phosphorylation sites and the corresponding phosphotransferases (102).

Based on the characterization of CBP-ID4 presented here, which shows that two parts of the polypeptide chain have an intrinsic propensity to adopt a helical conformation in solution, one might speculate on the active role played by this protein regions in the communication between different parts of complex proteins or between different partners. It is feasible that both transient α-helices identified in CBP-ID4 may promote the interactions between CBP and its partners. In addition, the highly flexible parts linking the two partially populated

helices may provide complementary functional advantages through their largely exposed and flexible protein backbones.

Conclusions

The NMR description of CBP-ID4 presented here reveals very heterogeneous structural and dynamic properties of this "linker", which combines a great extent of structural disorder together with two protein regions characterized by significant α-helical propensities. These findings lead to a reexamination of the concept of protein linker, traditionally considered as mere connection between folded domains, which instead may modulate and fine-tune protein function through an heterogeneous structural disorder. This study shows that the recently developed NMR methods allow to overcome potential limitations deriving from the peculiar properties of linkers (high proline content, extensive spectral overlap, fast amide protons exchange) and opens new possibilities for the characterization at atomic level of PTMs and of intra- and inter-molecular interactions of CBP-ID4, starting with the partners known to interact with adjacent folded domains of CBP.

Authors contribution

EOC, PT, ICF and RP conceived the research; ICF and RP directed the research; EOC and AdG expressed the protein and prepared the samples; AP and TT performed the NMR experiments and analyzed the data; MV and PT performed the bioinformatics analysis; AP, MV, PT, ICF and RP wrote the paper. All Authors reviewed the results and approved the final version of the manuscript.

Acknowledgements

This work has been supported in part by the European Commission Projects IDPbyNMR (Contract No. 264257), BioNMR (Contract No. 261863) and INSTRUCT (Contract No. 211252).

Conflict of interest

The Authors declare that they have no conflicts of interest with the contents of this article.

References

1. Habchi, J., P. Tompa, S. Longhi, and V. N. Uversky. 2014. Introducing protein intrinsic disorder. Chem. Rev. 114: 6561-6588.
2. van der Lee, R., M. Buljan, B. Lang, R. J. Weatheritt, G. W. Daughdrill, A. K. Dunker, M. Fuxreiter, J. Gough, J. Gsponer, D. T. Jones, P. M. Kim, R. W. Kriwacki, C. J. Oldfield, R. V. Pappu, P. Tompa, V. N. Uversky, P. E. Wright, and M. M. Babu. 2014. Classification of intrinsically disordered regions and proteins. Chem. Rev. 114: 6589-6631.
3. Hiller, S., C. Wasmer, G. Wider, and K. Wüthrich. 2007. Sequence-specific resonance assignment of soluble nonglobular proteins by 7D APSY-NMR spectroscopy. J. Am. Chem. Soc. 129: 10823-10828.

4. Mittag, T., and J. Forman-Kay. 2007. Atomic-level characterization of disordered protein ensembles. Curr. Opin. Struct. Biol. 17: 3-14.
5. Hsu, S. T., C. W. Bertoncini, and C. M. Dobson. 2009. Use of protonless NMR spectroscopy to alleviate the loss of information resulting from exchange-broadening. J. Am. Chem. Soc. 131: 7222-7223.
6. Narayanan, R. L., H. N. Duerr, S. Bilbow, J. Biernat, E. Mendelkow, and M. Zweckstetter. 2010. Automatic Assignment of the Intrinsically Disordered Protein Tau with 441-Residues. J. Am. Chem. Soc. 132: 11906-11907.
7. Nováček, J., A. Zawadzka-Kazimierczuk, V. Papoušková, L. Židek, H. Sanderová, L. Krásný, W. Koźmiński, and V. Sklenar. 2011. 5D ^{13}C-detected experiments for backbone assignment of unstructured proteins with a very low signal dispersion. J. Biomol. NMR 50: 1-11.
8. Felli, I. C., and R. Pierattelli. 2012. Recent progress in NMR spectroscopy: towards the study of intrinsically disordered proteins of increasing size and complexity. IUBMB Life 64: 473-481.
9. Harbison, N. W., S. Bhattacharya, and D. Eliezer. 2012. Assigning backbone NMR resonances for full length tau isoforms: efficient compromise between manual assignments and reduced dimensionality. Plos ONE 7: e34679.
10. Zawadzka-Kazimierczuk, A., W. Koźmiński, H. Sanderová, and L. Krásný. 2012. High dimensional and high resolution pulse seqeunces for backbone resonance assignment of intrinsically disordered proteins. J. Biomol. NMR 52: 329-337.
11. Solyom, Z., M. Schwarten, L. Geist, R. Konrat, D. Willbold, and B. Brutscher. 2013. BEST-TROSY experiments for time-efficient sequential resonance assignment of large disordered proteins. J. Biomol. NMR 55: 311-321.
12. Parigi, G., N. Rezaei-Ghaleh, A. Giachetti, S. Becker, C. Fernandez, M. Blackledge, C. Griesinger, M. Zweckstetter, and C. Luchinat. 2014. Long-Range Correlated Dynamics in Intrinsically Disordered Proteins. J. Am. Chem. Soc. 136: 16201-16209.
13. Jensen, M. R., M. Zweckstetter, J. R. Huang, and M. Blackledge. 2014. Exploring free-energy landscapes of intrinsically disordered proteins at atomic resolution using NMR spectroscopy. Chem. Rev. 114: 6632-6660.
14. Konrat, R. 2014. NMR contributions to structural dynamics studies of intrinsically disordered proteins. J. Magn. Reson. 241: 74-85.
15. Goodman, R. H., and S. Smolik. 2000. CBP/p300 in cell growth, transformation, and development. Genes Dev. 14: 1553-1577.
16. Legge, G. B., M. A. Martinez-Yamout, D. M. Hambly, T. Trinh, B. M. Lee, H. J. Dyson, and P. E. Wright. 2004. ZZ domain of CBP: an unusual zinc finger fold in a protein interaction module. J. Mol. Biol. 343: 1081-1093.
17. De Guzman, R. N., J. M. Wojciak, M. A. Martinez-Yamout, H. J. Dyson, and P. E. Wright. 2005. CBP/p300 TAZ1 domain forms a structured scaffold for ligand binding. Biochemistry 44: 490-497.
18. Liu, X., L. Wang, K. Zhao, P. R. Thompson, Y. Hwang, R. Marmorstein, and P. A. Cole. 2008. The structural basis of protein acetylation by the p300/CBP transcriptional coactivator. Nature 451: 846-850.

19. Miller, M., Z. Dauter, S. Cherry, J. E. Tropea, and A. Wlodawer. 2009. Structure of the Taz2 domain of p300: insights into ligand binding. Acta Crystallogr. D Biol. Crystallogr. 65: 1301-1308.
20. Kjaergaard, M., K. Teilum, and F. M. Poulsen. 2010. Conformational selection in the molten globule state of the nuclear coactivator binding domain of CBP. Proc. Natl. Acad. Sci. U. S. A. 107: 12535-12540.
21. Lee, C. W., M. A. Martinez-Yamout, H. J. Dyson, and P. E. Wright. 2010. Structure of the p53 transactivation domain in complex with the nuclear receptor coactivator binding domain of CREB binding protein. Biochemistry 49: 9964-9971.
22. Wang, F., C. B. Marshall, K. Yamamoto, G. Y. Li, G. M. Gasmi-Seabrook, H. Okada, T. W. Mark, and M. Ikura. 2012. Structures of KIX domain of CBP in complex with two FOXO3a transactivation domains reveal promiscuity and plasticity in coactivator recruitment. Proc. Natl. Acad. Sci. U. S. A. 109: 6078-6083.
23. Filippakopoulos, P., S. Picaud, M. Mangos, T. Keates, J. P. Lambert, D. Barsyte-Lovejoy, I. Felletar, R. Volkmer, S. Müller, T. Pawson, A. C. Gingras, C. H. Arrowsmith, and S. Knapp. 2012. Histone recognition and large-scale structural analysis of the human bromodomain family. Cell 149: 214-231.
24. Plotnikov, A. N., S. Yang, T. J. Zhou, E. Rusinova, A. Frasca, and M. M. Zhou. 2014. Structural insights into acetylated-histone H4 recognition by the bromodomain-PHD finger module of human transcriptional coactivator CBP. Structure 22: 353-360.
25. Dyson, H. J., and P. E. Wright. 2005. Intrinsically unstructured proteins and their functions. Nat. Rev. Mol. Cell Biol. 6: 197-208.
26. Dunker, A. K., J. D. Lawson, C. J. Brown, R. M. Williams, P. Romero, J. S. Oh, C. M. Ratliff, K. W. Hipps, J. Ausio, M. S. Nissen, R. Reeves, C. Kang, C. R. Kissinger, R. W. Bailey, M. D. Griswold, W. Chiu, and E. C. Garner. 2001. Intrinsically disordered protein. J. Mol. Graph. Model 19: 26-59.
27. Vucetic, S., C. J. Brown, A. K. Dunker, and Z. Obradovic. 2003. Flavors of protein disorder. Proteins 52: 573-584.
28. Hansen, J. C., X. Lu, E. D. Ross, and R. W. Woody. 2006. Intrinsic protein disorder, amino acid composition, and histone terminal domains. J. Biol. Chem. 281: 1853-1856.
29. Felli, I. C., and R. Pierattelli. 2014. Novel methods based on ^{13}C detection to study intrinsically disordered proteins. J. Magn. Reson. 241: 115-125.
30. Nováček, J., L. Zídek, and V. Sklenar. 2014. Toward optimal-resolution NMR of intrinsically disordered proteins. J. Magn. Reson. 241: 41-52.
31. Marley, J., M. Lu, and C. Bracken. 2001. A method for efficient isotopic labeling of recombinant proteins. J. Biomol. NMR 20: 71-75.
32. Bermel, W., I. Bertini, L. Duma, L. Emsley, I. C. Felli, R. Pierattelli, and P. R. Vasos. 2005. Complete assignment of heteronuclear protein resonances by protonless NMR spectroscopy. Angew. Chem. Int. Ed. 44: 3089-3092.
33. Bermel, W., I. Bertini, I. C. Felli, and R. Pierattelli. 2009. Speeding up ^{13}C direct detection Biomolecular NMR experiments. J. Am. Chem. Soc. 131: 15339-15345.

34. Bermel, W., I. Bertini, L. Gonnelli, I. C. Felli, W. Koźmiński, A. Piai, R. Pierattelli, and J. Stanek. 2012. Speeding up sequence specific assignment of IDPs. J. Biomol. NMR 53: 293-301.

35. Bermel, W., I. C. Felli, L. Gonnelli, W. Koźmiński, A. Piai, R. Pierattelli, and A. Zawadzka-Kazimierczuk. 2013. High-dimensionality ^{13}C direct-detected NMR experiments for the automatic assignment of intrinsically disordered proteins. J. Biomol. NMR 57: 353-361.

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

37. Salzmann, M., K. Pervushin, G. Wider, H. Senn, and K. Wüthrich. 1998. TROSY in triple-resonance experiments: new perspectives for sequential NMR assignment of large proteins. Proc. Natl. Acad. Sci. U. S. A. 95: 13585-13590.

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

39. Weisemann, R., H. Rüterjans, and W. Bermel. 1993. 3D triple-resonance NMR techniques for the sequential assignment of NH and ^{15}N resonances in ^{15}N- and ^{13}C-labelled proteins. J. Biomol. NMR 3: 113-120.

40. Piai, A., T. Hošek, L. Gonnelli, A. Zawadzka-Kazimierczuk, W. Koźmiński, B. Brutscher, W. Bermel, R. Pierattelli, and I. C. Felli. 2014. "CON-CON" assignment strategy for highly flexible intrinsically disordered proteins. J. Biomol. NMR 60: 209-218.

41. Barbato, G., M. Ikura, L. E. Kay, R. W. Pastor, and A. Bax. 1992. Backbone dynamics of calmodulin studied by ^{15}N relaxation using inverse detected two-dimensional NMR spectroscopy; the central helix is flexible. Biochemistry 31: 5269-5278.

42. Farrow, N. A., R. Muhandiram, A. U. Singer, S. M. Pascal, C. M. Kay, G. Gish, S. E. Shoelson, T. Pawson, J. D. Forman-Kay, and L. E. Kay. 1994. Backbone dynamics of a free and phosphopeptide-complexed Src homology 2 domain studied by ^{15}N NMR relaxation. Biochemistry 33: 5984-6003.

43. Peng, J. W., and G. Wagner. 1992. Mapping of spectral density function using heteronuclear NMR relaxation measurements. J. Magn. Reson. 98: 308-332.

44. Peng, J. W., and G. Wagner. 1994. Investigation of protein motions via relaxation measurements. Methods Enzymol. 239: 563-596.

45. Hwang, T. L., P. C. M. Van Zijl, and S. Mori. 1998. Accurate quantification 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.

46. Vuister, G. W., and A. Bax. 1993. Quantitative J correlation: a new approach for measuring homonuclear three-bond $J(H^NH^\alpha)$ coupling constants in ^{15}N-enriched proteins. J. Am. Chem. Soc. 115: 7772-7777.

47. Kazimierczuk, K., A. Zawadzka, and W. Koźmiński. 2008. Optimization of random time domain sampling in multidimensional NMR. J. Magn. Reson. 192: 123-130.

48. Delaglio, F., S. Grzesiek, G. W. Vuister, G. Zhu, J. Pfeifer, and A. Bax. 1995. NMRPipe: a multidimensional spectral processing system based on UNIX Pipes. J. Biomol. NMR 6: 277-293.

49. Kazimierczuk, K., A. Zawadzka, W. Koźmiński, and I. Zhukov. 2006. Random sampling of evolution time space and Fourier transform processing. J. Biomol. NMR 36: 157-168.
50. Kazimierczuk, K., A. Zawadzka, and W. Koźmiński. 2009. Narrow peaks and high dimensionalities: Exploiting the advantages of random sampling. J. Magn. Reson. 197: 219-228.
51. Kazimierczuk, K., A. Zawadzka-Kazimierczuk, and W. Koźmiński. 2010. Non-uniform frequency domain for optimal exploitation of non-uniform sampling. J. Magn. Reson. 205: 286-292.
52. Keller, R. L. J. 2004. The Computer Aided Resonance Assignment Tutorial, Cantina Verlag.
53. Goddard, T. D., D. G. Kneller, SPARKY 3, University of California, San Francisco.
54. Vranken, W. F., W. Boucher, T. J. Stevens, R. H. Fogh, A. Pajon, Llinas M, E. L. Ulrich, J. L. Markley, J. Ionides, and E. D. Laue. 2005. The CCPN data model for NMR spectrooscopy: development of a software pipeline. Proteins: Struct. , Funct. , Bioinf. 59: 687-696.
55. Tamiola, K., and F. A. Mulder. 2012. Using NMR chemical shifts to calculate the propensity for structural order and disorder in proteins. Biochem. Soc. Trans. 40: 1014-1020.
56. Tamiola, K., B. Acar, and F. A. A. Mulder. 2010. Sequence-Specific Random Coil Chemical Shifts of Intrinsically Disordered Proteins. J. Am. Chem. Soc. 132: 18000-18003.
57. Muñoz, V., and L. Serrano. 1994. Elucidating the folding problem of helical peptides using empirical parameters. Nat. Struct. Biol. 1: 399-409.
58. Muñoz, V., and L. Serrano. 1995. Elucidating the folding problem of helical peptides using empirical parameters. II. Helix macrodipole effects and rational modification of the helical content of natural peptides. J. Mol. Biol. 245: 275-296.
59. Muñoz, V., and L. Serrano. 1995. Elucidating the folding problem of helical peptides using empirical parameters. III. Temperature and pH dependence. J. Mol. Biol. 245: 297-308.
60. Muñoz, V., and L. Serrano. 1997. Development of the multiple sequence approximation within the AGADIR model of alpha-helix formation: comparison with Zimm-Bragg and Lifson-Roig formalisms. Biopolymers 41: 495-509.
61. Lacroix, E., A. R. Viguera, and L. Serrano. 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.
62. Dosztanyi, Z., V. Csizmok, P. Tompa, and I. Simon. 2005. IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. Bioinformatics 21: 3433-3434.
63. Xue, B., R. L. Jr. Dunbrack, R. W. Williams, A. K. Dunker, and V. Uversky. 2010. PONDR-FIT: A meta-predictor of intrinsically disordered amino acids. Biochim Biophys Acta 1804: 996-1010.
64. Camacho, C., G. Couloris, V. Avagyan, J. Papadopoulos, K. Bealer, and T. L. Madden. 2009. BLAST+: architecture and applications. BMC Bioinformatics 10: 421.

65. Katoh, K., and D. M. Standley. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. 30: 772-780.

66. Varadi, M., M. Guharoy, F. Zsolyomi, and P. Tompa. 2015. DisCons: a novel tool to quantify and classify evolutionary conservation of intrinsic protein disorder. BMC Bioinformatics 16: 153.

67. Dinkel, H., K. Van Roey, S. Michael, N. E. Davey, R. J. Weatheritt, D. Born, T. Speck, D. Krüger, G. Grebnev, M. Kuban, M. Strumillo, B. Uyar, A. Budd, B. Altenberg, M. Seiler, L. B. Chemes, J. Glavina, I. E. Sánchez, F. Diella, and T. J. Gibson. 2014. The eukaryotic linear motif resource ELM: 10 years and counting. Nucleic Acids Res. 42: D259-D266.

68. Blom, N., S. Gammeltoft, and S. Brunak. 1999. Sequence and structure-based prediction of eukaryotic protein phosphorylation sites. J. Mol. Biol. 294: 1351-1362.

69. Blom, N., T. Sicheritz-Pontén, R. Gupta, S. Gammeltoft, and S. Brunak. 2004. Prediction of post-translational glycosylation and phosphorylation of proteins from the amino acid sequence. Proteomics. 4: 1633-1649.

70. Ozenne, V., F. Bauer, L. Salmon, J. R. Huang, M. R. Jensen, Segard S., P. Bernadó, Charavay C., and M. Blackledge. 2012. Flexible-meccano: a tool for the generation of explicit ensemble descriptions of intrinsically disordered proteins and their associated experimental observables. Bioinformatics 28: 1463-1470.

71. Varadi, M., S. Kosol, P. Lebrun, E. Valentini, M. Blackledge, A. K. Dunker, I. C. Felli, J. D. Forman-Kay, R. W. Kriwacki, R. Pierattelli, J. L. Sussman, D. I. Svergun, V. N. Uversky, M. Vendruscolo, D. S. Wishart, P. E. Wright, and P. Tompa. 2014. pE-DB: a database of structural ensemble of intrinsically disordered and of unfolded proteins. Nucleic Acids Res. 42: D326-D335.

72. Bermel, W., M. Bruix, I. C. Felli, V. M. V. Kumar, R. Pierattelli, and S. Serrano. 2013. Improving the chemical shift dispersion of multidimensional NMR spectra of intrinsically disordered proteins. J. Biomol. NMR 55: 231-237.

73. Kazimierczuk, K., J. Stanek, A. Zawadzka-Kazimierczuk, and W. Koźmiński. 2010. Random sampling in multidimensional NMR spectroscopy. Prog. NMR Spectrosc. 57: 420-434.

74. Felli, I. C., A. Piai, and R. Pierattelli. 2013. Recent advances in solution NMR studies: ^{13}C direct detection for biomolecular NMR applications. Ann. Rep. NMR Spectroscop. 359-418.

75. Ulrich, E. L., H. Akutsu, J. F. Doreleijers, Y. Harano, Y. E. Ioannidis, J. Lin, M. Livny, S. Mading, D. Maziuk, Z. Miller, E. Nakatani, C. F. Schulte, D. E. Tolmie, R. K. Wenger, H. Yao, and J. L. Markley. 2008. BioMagResBank. Nucleic Acids Res. 36: D402-D408.

76. Schwalbe, H., K. M. Fiebig, M. Buck, J. A. Jones, S. B. Grimshaw, A. Spencer, S. J. Glaser, L. J. Smith, and C. M. Dobson. 1997. Structural and dynamical properties of a denatured protein. Heteronuclear 3D NMR experiments and theoretical simulations of lysozyme in 8 M urea. Biochemistry 36: 8977-8991.

77. Massad, T., J. Jarvet, R. Tanner, K. Tomson, J. Smirnova, P. Palumaa, M. Sugai, T. Kohno, K. Vanatalu, and P. Damberg. 2007. Maximum entropy reconstruction of joint phi, psi-distribution with a coil-library prior: the backbone conformation of the peptide hormone motilin in aqueous solution from phi and psi-dependent J-couplings. J. Biomol. NMR 38: 107-123.

78. Meier, S., M. Blackledge, and S. Grzesiek. 2008. Conformational distributions of unfolded polypeptides from novel NMR techniques. J. Chem. Phys 128: 052204.

79. Otten, R., K. Wood, and F. A. A. Mulder. 2009. Comprehensive determination of $^{3}J_{HNHa}$ for unfolded proteins using $^{13}C'$-resolved spin-echo difference spectroscopy. J. Biomol. NMR 45: 343-49.

80. Lendel, C., and P. Damberg. 2009. 3D J-resolved NMR spectroscopy for unstructured polypeptides: fast measurement of $^{3}J_{HNH\alpha}$ coupling constants with outstanding spectral resolution. J. Biomol. NMR 44: 35-42.

81. Schweitzer-Stenner, R. 2012. Conformational propensities and residual structures in unfolded peptides and proteins. Mol. Biosyst. 8: 122-133.

82. Shi, Z., K. Chen, Z. Liu, and N. R. Kallenbach. 2006. Conformation of the Backbone in Unfolded Proteins. Chem. Rev. 106: 1877-1897.

83. Baxter, N. J., and M. P. Williamson. 1997. Temperature dependence of ^{1}H chemical shifts in proteins. J. Biomol. NMR 9: 359-369.

84. Cierpicki, T., and J. Otlewski. 2001. Amide proton temperature coefficients as hydrogen bond indicators in proteins. J. Biomol. NMR 21: 249-261.

85. Lee, S. H., D. H. Kim, J. J. Han, E. J. Cha, J. E. Lim, Y. J. Cho, C. Lee, and K. H. Han. 2012. Understanding pre-structured motifs (PreSMos) in intrinsically unfolded proteins. Curr. Prot. Pept. Sci. 13: 35-54.

86. Lee, C., L. Kalmar, B. Xue, P. Tompa, G. W. Daughdrill, V. N. Uversky, and K. H. Han. 2014. Contribution of proline to the pre-structuring tendency of transient helical secondary structure elements in intrinsically disordered proteins. Biochim. Biophys. Acta. 1840: 993-1003.

87. Kini, R. M., and H. J. Evans. 1995. A hypothetical structural role for proline residues in the flanking segments of protein-protein interaction sites. Biochem. Biophys. Res. Commun. 212: 1115-1124.

88. Georgieva, E. R., S. Xiao, P. P. Borbat, J. H. Freed, and D. Eliezer. 2014. Tau binds to lipid membrane surfaces via short amphipathic helices located in its microtubule-binding repeats. Biophys. J. 107: 1441-1452.

89. Wright, P. E., and H. J. Dyson. 2015. Intrinsically disordered proteins in cellular signalling and regulation. Nat. Rev. Mol. Cell Biol. 16: 18-29.

90. He, J., J. Ye, Y. Cai, C. Riquelme, J. O. Liu, X. Liu, A. Han, and L. Chen. 2011. Structure of p300 bound to MEF2 on DNA reveals a mechanism of enhanceosome assembly. Nucleic Acids Res. 39: 4464-4474.

91. Tompa, P., and M. Varadi. 2014. Predicting the predictive power of IDP ensembles. Structure 22: 177-178.

92. Koradi, R., M. Billeter, and K. Wüthrich. 1996. MOLMOL: a program for display and analysis of macromolecular structure. J. Mol. Graphics 14: 51-55.

93. Pettersen, E. F., T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng, and T. E. Ferrin. 2004. UCSF Chimera - A Visualization System for Exploratory Research and Analysis. J Comp Chem 25: 1605-1612.

94. Baker, N. A., D. Sept, S. Joseph, M. J. Holst, and J. A. Mc Cammon. 2001. Electrostatics of nanosystems: application to microtubules and the ribosome. Proc. Natl. Acad. Sci. U. S. A. 98: 10037-10041.

95. Fuxreiter, M., I. Simon, P. Friedrich, and P. Tompa. 2004. Preformed structural elements feature in partner recognition by intrinsically unstructured proteins. J. Mol. Biol. 338: 1015-1026.
96. Mohan, A., C. J. Oldfield, P. Radivojac, M. S. Cortese, A. K. Dunker, and V. N. Uversky. 2006. Analysis of molecular recognition features (MoRFs). J. Mol. Biol. 362: 1043-1059.
97. Fuxreiter, M., P. Tompa, and I. Simon. 2007. Local structural disorder imparts plasticity on linear motifs. Bioinformatics 23: 950-956.
98. Lu, K. P., Y. C. Liou, and X. Z. Zhou. 2002. Pinning down proline-directed phosphorylation signaling. Trends Cell Biol. 12: 164-172.
99. Theillet, F. X., C. Smet-Nocca, S. Liokatis, R. Thongwichian, J. Kosten, M. K. Yoon, R. W. Kriwacki, I. Landrieu, G. Lippens, and P. Selenko. 2012. Cell signaling, post-translational protein modifications and NMR spectroscopy. J. Biomol. NMR 54: 217-236.
100. Amata, I., M. Maffei, A. Igea, M. Gay, M. Vilaseca, A. R. Nebreda, and M. Pons. 2013. Multi-phosphorylation of the intrinsically disordere unique domain of c-Src studied by in-cell and real-time NMR spectroscopy. ChemBioChem 14: 1820-1827.
101. Amata, I., M. Maffei, and M. Pons. 2014. Phosphorylation of unique domains of Src family kinases. Front. Genet. 5: 181.
102. Iakoucheva, L. M., P. Radivojac, C. J. Brown, T. R. O'Connor, J. G. Sikes, Z. Obradovic, and A. K. Dunker. 2004. The importance of intrinsic disorder for protein phosphorylation. Nucleic Acids Res. 32: 1037-1049.

Just a flexible linker? The structural and dynamic properties of CBP-ID4 as revealed by NMR spectroscopy

Alessandro Piai[1], Eduardo O. Calçada[1], Thomas Tarenzi[1], Alessandro del Grande[1], Mihaly Varadi[2], Peter Tompa[2,3], Isabella C. Felli[1], Roberta Pierattelli[1]

1 Magnetic Resonance Center (CERM) and Department of Chemistry "Ugo Schiff", University of Florence, 50019 Sesto Fiorentino (Florence), Italy

2 VIB Structural Biology Research Center (SBRC), Vlaams Instituut voor Biotechnologie, 1050 Brussel, Belgium; Structural Biology Brussels (SBB), Vrije Universiteit Brussel, 1050 Brussel, Belgium

3 Institute of Enzymology, Research Centre for Natural Sciences of the Hungarian Academy of Sciences, 1117 Budapest, Hungary.

Supporting Material

Supporting figures

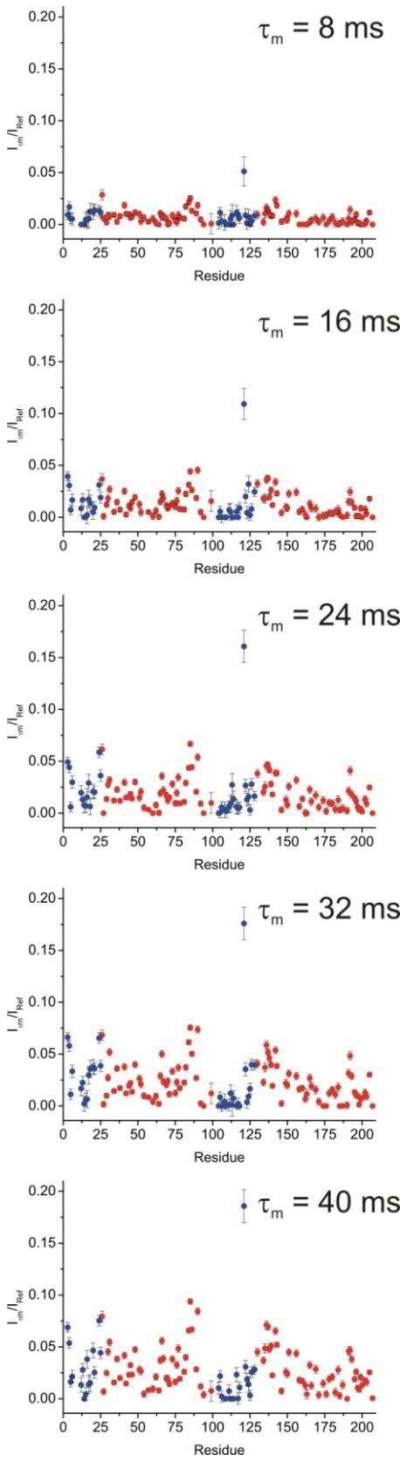


Figure S1. Extent of amide proton exchange with the solvent. $I\tau_m/I_{Ref}$ values measured through the (CLEANEX-PM)-FHSQC experiments are reported as a function of residue number. In the top right of each plot, the length of the mixing time τ_m is indicated. The residues belonging to the two protein regions characterized by higher propensity for α-helix are reported in blue.

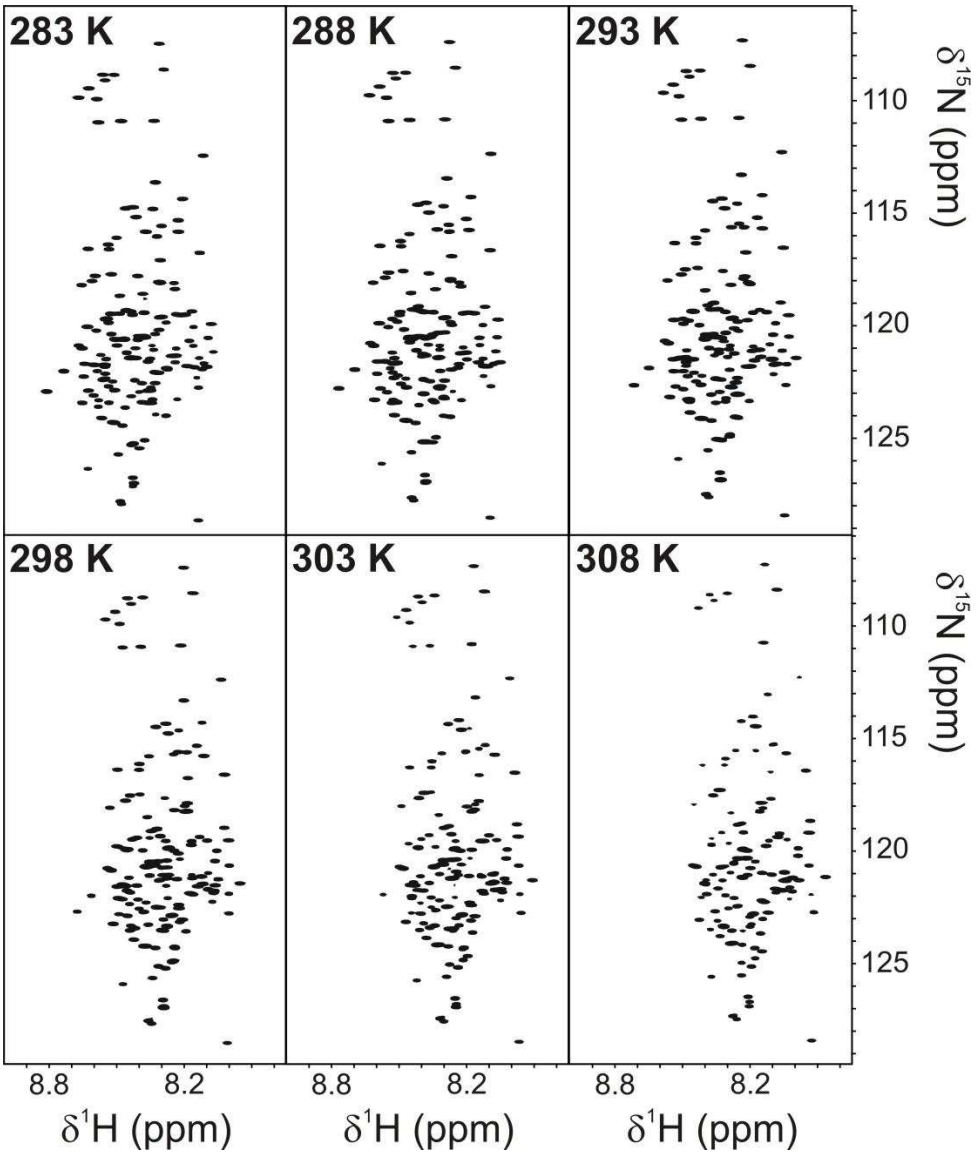


Figure S2. Temperature dependence of CBP-ID4. 2D BEST-TROSY spectra acquired at different temperature (from 283.0 to 308.0 K) are reported. The spectra are plotted with the same contour levels.

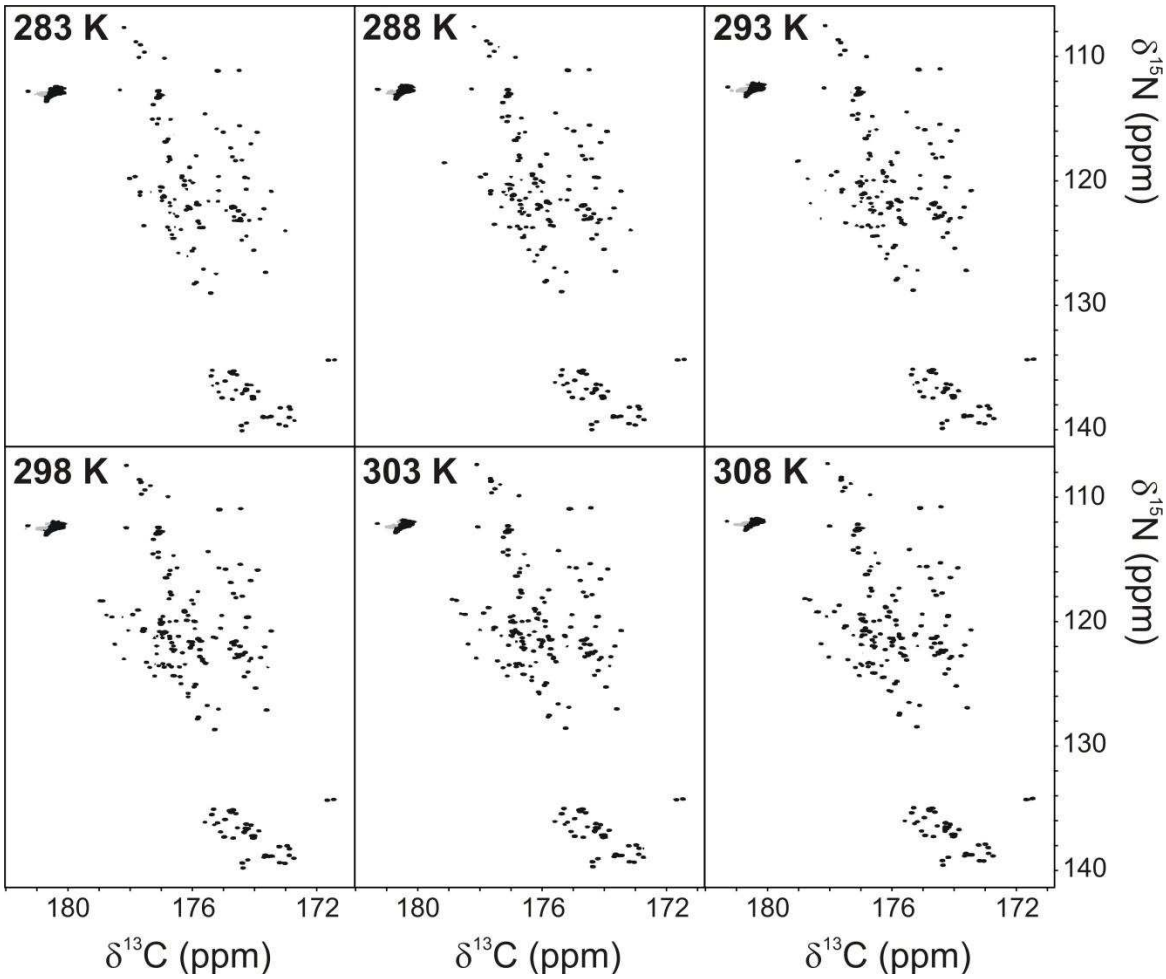


Figure S3. Temperature dependence of CBP-ID4. 2D CON-IPAP spectra acquired at different temperature (from 283.0 to 308.0 K) are reported. The spectra are plotted with the same contour levels.

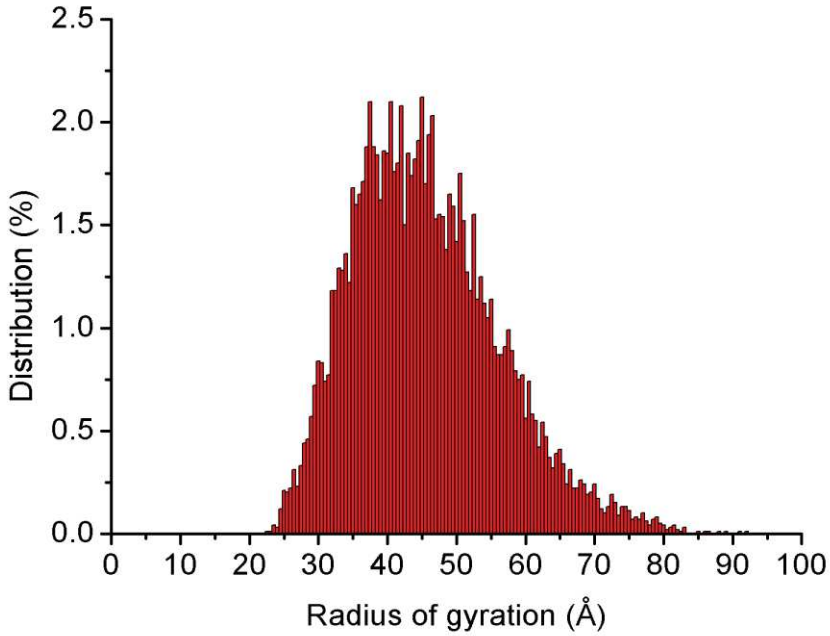


Figure S4. Radius of gyration of the conformers of the computed ensemble. The wide distribution of the radius of gyration reflects the ability of the linker to fine-modulate the relative distance between attached globular domains.

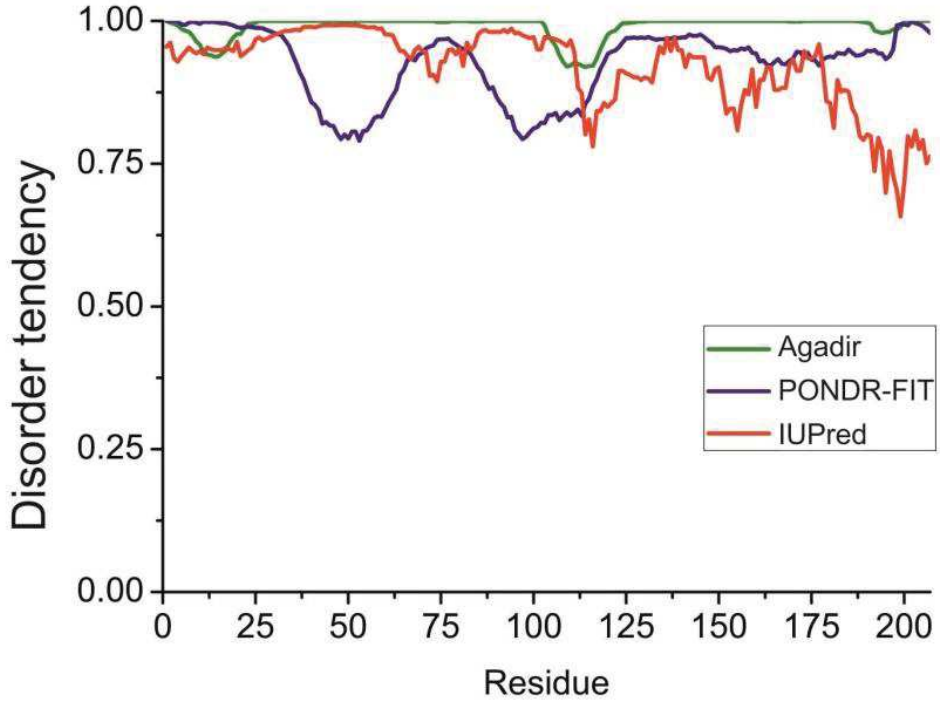


Figure S5. Disorder tendency of CBP-ID4 as predicted by IUPred (red), PONDR-FIT (blue) and Agadir (green).

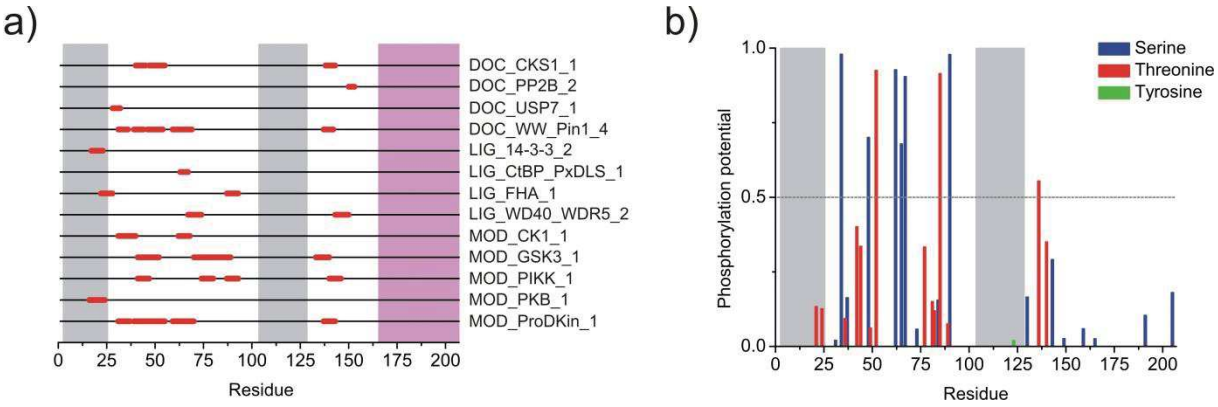


Figure S6. Distribution of potential SLiMs (a) and phosphorylation sites (b) of CBP-ID4. The two protein regions characterized by significant α-helical propensity are highlighted in gray: almost no SLiMs and phosphorylation sites are predicted in these regions (a). The protein region highlighted in magenta was discarded from the analysis since recognized by the protein families database (Pfam) as the interlocking domain of CBP, which forms a 3-helical non-globular array that forms interlocked heterodimers with its target. b) Residues having a phosphorylation potential above the threshold are considered to be highly probable phosphorylation sites. Phosphorylation sites predicted to be kinase specific are the following: Ser 34 (p38MAPK and cdk5 kinases); Thr 52 (p38MAPK, GSK3 and cdk5 kinases); Ser 62 (GSK3 and cdk5 kinases); Ser 67 (p38MAPK and cdk5 kinases); Thr 85 (PKC kinase); Ser 90 (DNAPK kinase).

NMR data acquisition and processing information

3D BEST-TROSY (BT) NMR experiments were performed at 21.1 T on a Bruker Avance spectrometer operating at 898.57 MHz ^{1}H, 225.95 MHz ^{13}C and 91.05 MHz ^{15}N frequencies, equipped with a cryogenically cooled probehead. PC9 and E-BURP2 (or time reversed E-BURP2) shapes of durations of 1800 and 1270 μs, respectively, were employed for ^{1}H band-selective π/2 flip angle pulses; REBURP shape of duration of 1250 μs was used for ^{1}H band-selective π flip angle pulse; BIP-750-50-20 pulse shapes of duration of 140 μs were used for broadband ^{1}H inversion. For ^{13}C band-selective π/2 and π flip angle pulses G4 (or time reversed G4) and Q3 shapes of durations of 274 and 190 μs, respectively, were used, except for the π pulses that should be band-selective on the C^{α} region (Q3, 660 μs). The ^{13}C band selective pulses on C^{α} and C' were applied at the center of each region, respectively. All gradients employed had a smoothed square shape.

5D BT-NMR experiments were collected at 22.3 T on a Bruker Avance III spectrometer operating at 950.20 MHz ^{1}H, 238.93 MHz ^{13}C and 96.28 MHz ^{15}N frequencies, equipped with a cryogenically cooled probehead. E-BURP2 and REBURP shapes of durations of 1200 and 1180 μs, respectively, were employed for ^{1}H band-selective π/2 and π flip angle pulses; BIP-750-50-20 pulse shapes of duration of 200 μs were used for broadband ^{1}H inversion. For ^{13}C band-selective π /2 and π flip angle pulses G4 (or time reversed G4) and Q3 shapes of durations of 260 and 161 μs, respectively, were used, except for the π pulses that should be band-selective on the C^{α} region (Q3, 667 μs) and for the adiabatic π pulse to invert both C' and C^{α} (smoothed Chirp 500 μs, 20% smoothing, 80 kHz sweep width, 11.3 kHz RF field strength). The ^{13}C band selective pulses on C^{α} and C' were applied at the center of each region, respectively. All gradients employed had a smoothed square shape.

3D and 4D ^{13}C detected experiments were acquired at 16.4 T on a Bruker Avance spectrometer operating at 700.06 MHz ^{1}H, 176.03 MHz ^{13}C and 70.94 MHz ^{15}N frequencies, equipped with a cryogenically cooled probehead optimized for ^{13}C-direct detection. For ^{13}C band-selective π /2 and π flip angle pulses Q5 (or time reversed Q5) and Q3 shapes of durations of 300 and 220 μs, respectively, were used, except for the π pulses that should be band-selective on the C^{α} region (Q3, 860 μs) and for the adiabatic π pulse to invert both C' and C^{α} (smoothed Chirp 500 μs, 25 % smoothing, 80 kHz sweep width, 11.3 kHz RF field

strength). The ^{13}C band selective pulses on $C^{\alpha/\beta}$, C^{α}, and C' were given at the center of each region, respectively, and the adiabatic pulse was adjusted to cover the entire ^{13}C region. Decoupling of ^{1}H and ^{15}N was achieved with waltz16 (1.7 kHz) and garp4 (1.0 kHz) sequences, respectively. All gradients employed had a sine-shape. All experiments employ the IPAP approach to remove the splitting in the direct acquisition dimension caused by the $C^{\square}$–C' couplings. The in-phase (IP) and antiphase (AP) components were acquired and stored in an interleaved manner, doubling the number of FIDs recorded.

The most relevant experimental parameters used for the acquisition of $^{1}H^{N}$ and ^{13}C detected NMR experiments for sequence-specific assignment are reported in Table S1 and Table S2, respectively. Experimental parameters of ^{15}N relaxation NMR experiments are provided in Table S3. ^{15}N relaxation rates (R_1 and R_2) were determined by fitting cross-peak intensities, measured as a function of a variable delay, as single-exponential decay. ^{15}N-^{1}H NOEs were obtained as the ratio between cross-peak intensities in spectra recorded with and without ^{1}H saturation.

The most relevant experimental parameters for (CLEANEX-PM)-FHSQC and 3D HNHA experiments are given in Table S4. Exchange rates between amide protons and water protons were estimated by analyzing the ratio between cross-peak intensities in spectra recorded with and without CLEANEX-PM mixing period. Homonuclear $^{3}J_{HN\text{-}H\alpha}$ couplings were quantified by measuring the diagonal-peak to cross-peak ratio obtained in the 3D ^{15}N-separated quantitative J-correlation HNHA spectrum.

Finally, parameters related to the NMR experiments for temperature dependence are shown in Table S5.

Table S1 Experimental parameters used for the acquisition of $^{1}H^{N}$ detected NMR experiments for sequence-specific assignment

	Spectral widths and maximal evolution times					No. of scans	Inter-scan delays (s)	No. of complex points (aq)	No. of hypercomplex points	Duration of the experiment	Relative data points density (%)
2D BEST-TROSY	2300 Hz (^{15}N) 222.6 ms				10800 Hz ($^{1}H^{N}$) 95.1 ms	2	0.200 s	1024	512	15 min	100.0
3D BT-HNCO	2400 Hz ($^{13}C'$) 52.9 ms		2300 Hz (^{15}N) 55.2 ms		10800 Hz ($^{1}H^{N}$) 95.1 ms	2	0.200 s	1024	4500	4 h 20 min	27.5
3D BT-HN(CA)CO	2400 Hz ($^{13}C'$) 52.9 ms		2300 Hz (^{15}N) 55.2 ms		10800 Hz ($^{1}H^{N}$) 95.1 ms	4	0.200 s	1024	4500	8 h 45 min	27.5
3D BT-HNCACB	14000 Hz ($^{13}C^{\alpha/\beta}$) 24.9 ms		2300 Hz (^{15}N) 55.2 ms		10800 Hz ($^{1}H^{N}$) 95.1 ms	4	0.200 s	1024	6000	11 h 45 min	13.4
3D BT-HN(CO)CACB	14000 Hz ($^{13}C^{\alpha/\beta}$) 24.9 ms		2300 Hz (^{15}N) 55.2 ms		10800 Hz ($^{1}H^{N}$) 95.1 ms	4	0.200 s	1024	6000	12 h 30 min	13.4
3D BT-(H)N(COCA)NH	2300 Hz (^{15}N) 28.7 ms		2300 Hz (^{15}N) 55.2 ms		10800 Hz ($^{1}H^{N}$) 95.1 ms	8	0.200 s	1024	4500	18 h 20 min	52.5
3D BT-(H)N(CA)NNH	2300 Hz (^{15}N) 33.0 ms		2300 Hz (^{15}N) 33.0 ms		10800 Hz ($^{1}H^{N}$) 95.1 ms	16	0.200 s	1024	2400	20 h 30 min	40.5
5D BT-(H)NCO(CAN)CONNH	2600 Hz (^{15}N) 40.4 ms	2600 Hz ($^{13}C'$) 40.4 ms	2600 Hz ($^{13}C'$) 40.4 ms	2600Hz (^{15}N) 40.4 ms	13300 Hz ($^{1}H^{N}$) 77.0 ms	8	0.200 s	1024	2000	1 d 15 h	0.002
5D BT-HN(COCAN)CONNH	2600 Hz (^{15}N) 40.4 ms	2600 Hz ($^{13}C'$) 40.4 ms	2600 Hz ($^{13}C'$) 40.4 ms	2600Hz (^{15}N) 40.4 ms	13300 Hz ($^{1}H^{N}$) 77.0 ms	8	0.200 s	1024	2000	1 d 15 h	0.003

Table S2 Experimental parameters used for the acquisition of ^{13}C detected NMR experiments for sequence-specific assignment

	Spectral widths and maximal evolution times				No. of scans	Inter-scan delays (s)	No. of complex points (aq)	No. of hypercomplex points	Duration of the experiment	Relative data points density (%)
2D CON-IPAP	2600 Hz (^{15}N) 98.5 ms		5300 Hz (^{13}C) 96.8 ms		16	2.500	512	256	12 h 30 min	100.0
3D (H)CBCACON-IPAP	12500 Hz ($^{13}C^{\alpha/\beta}$) 7.5 ms	2600 Hz (^{15}N) 50.0 ms	8800 Hz (^{13}C) 58.3 ms		8	1.000 s	512	1100	1 d	8.8
3D (H)CBCANCO-IPAP	12500 Hz ($^{13}C^{\alpha/\beta}$) 7.5 ms	2600 Hz (^{15}N) 31.9 ms	8800 Hz (^{13}C) 58.3 ms		16	1.000 s	512	1100	2 d 4 h	13.7
4D (HCA)CON(CA)CON-IPAP	2400 Hz (^{13}C) 24.2 ms	2600 Hz (^{15}N) 24.2 ms	2600 Hz (^{15}N) 40.0 ms	8800 Hz (^{13}C) 58.3 ms	16	0.900 s	512	850	3 d 1 h	0.2
4D (HN)CON(CA)CON-IPAP	2400 Hz (^{13}C) 24.2 ms	2600 Hz (^{15}N) 24.2 ms	2600 Hz (^{15}N) 40.0 ms	8800 Hz (^{13}C) 58.3 ms	32	0.600 s	512	700	3 d 18 h	0.2

Table S3 Experimental parameters used for the acquisition of ^{15}N relaxation NMR experiments

	Spectral widths and maximal evolution times		No. of scans	Inter-scan delays (s)	No. of complex points (aq)	Duration of the experiment
^{15}N R_1	1800 Hz (^{15}N) 126.8 ms	10500 Hz ($^{1}H^{N}$) 97.6 ms	8	3.000 s	1024	3 h 15 min – 4 h 15 min
^{15}N R_2	1800 Hz (^{15}N) 126.8 ms	10500 Hz ($^{1}H^{N}$) 97.6 ms	8	3.000 s	1024	3 h 40 min
Steady-state heteronuclear $^{15}N\{^{1}H\}$ NOEs	1800 Hz (^{15}N) 144.3 ms	10500 Hz ($^{1}H^{N}$) 97.6 ms	72	6.000 s	1024	2 d 15 h
For the determination of R_1, 10 experiments were acquired changing the variable delay from 15 to 995 ms. For the determination of R_2, 10 experiments were acquired changing the variable delay from 30 to 375 ms.						

Table S4 Experimental parameters used for the acquisition of (CLEANEX-PM)-FHSQC and 3D HNHA NMR experiments

	Spectral widths and maximal evolution times			No. of scans	Inter-scan delays (s)	No. of complex points (aq)	Duration of the experiment
(CLEANEX-PM)-HSQC	1700 Hz (^{15}N) 147.3 ms	8400 Hz ($^{1}H^{N}$) 61.1 ms		8	5.000 s	512	5 h 55 min
3D HNHA	1700 Hz (^{15}N) 46.0 ms	7000 Hz ($^{1}H^{\alpha}$) 9.1 ms	7000 Hz ($^{1}H^{N}$) 107.2 ms	8	1.000 s	750	2 d 7 h
For the determination of the extent of amide proton exchange with the solvent, 5 experiments were acquired changing the length of the mixing time from 8 to 40 ms. An additional FHSQC experiment was measured with the same parameters to obtain the reference peak volumes.							

Table S5 Experimental parameters used for the acquisition of the NMR experiments for temperature dependence

	Spectral widths and maximal evolution times		No. of scans	Inter-scan delays (s)	No. of complex points (aq)	Duration of the experiment
2D BEST-TROSY	2300 Hz (^{15}N) 222.6 ms	11400 Hz ($^{1}H^{N}$) 400.0 ms	2	0.001	4600	20 min
2D CON-IPAP	3500 Hz (^{15}N) 72.6 ms	9600 Hz (^{13}C) 53.6 ms	16	1.500 s	512	7 h 40 min

Article 8. Sequence context influences the structure and oligomerization propensity of a polyQ tract

Alessandro Piai[1], Bahareh Eftekharzadeh[2], Giulio Chiesa[2], Daniele Mungianu[1], Jesús García[2], Roberta Pierattelli[1], Isabella C. Felli[1], Xavier Salvatella[2,3]

[1] CERM and Department of Chemistry "Ugo Schiff", University of Florence, 50019, Via Luigi Sacconi 6, Sesto Fiorentino, Florence, Italy.

[2] IRB Barcelona, Baldiri Reixac 10, 08028 Barcelona, Spain.

[3] ICREA, Barcelona, Spain.

Manuscript submitted

Sequence Context Influences the Structure and Aggregation Propensity of a PolyQ Tract**

Bahareh Eftekharzadeh[1], Alessandro Piai[1], Giulio Chiesa, Daniele Mungianu, Jesús García, Roberta Pierattelli, Isabella Felli, and Xavier Salvatella*

Abstract: Expansions of polyglutamine (polyQ) tracts in nine different proteins cause a family of neurodegenerative disorders called polyQ diseases. There is great interest in characterizing the conformations that these tracts adopt and in understanding how their properties are influenced by the sequences flanking them. We used solution NMR to study at single residue resolution a 156-residue long fragment of the androgen receptor produced by caspase 3 cleavage. This fragment contains the polyQ tract that causes the polyQ disease spinobulbar muscular atrophy, also known as Kennedy disease. Our findings indicate that a Leu-rich region preceding the polyQ tract causes it to become α-helical and reduces its aggregation propensity. We conclude that sequence context strongly influences the conformational properties of polyQ tracts and that induction of α-helical propensity has a protective role with respect to formation of aggregates, which suggests a new therapeutic strategy for polyQ diseases.

A group of nine disorders termed *polyglutamine (polyQ) diseases* occur as a result of the expansion of polymorphic polyQ tracts in proteins that are otherwise unrelated.[1] The variable length of such tracts is due to the propensity of CAG and GTC codon repeats that codify for them to form non-B-DNA structures that cause slippage during DNA replication.[2] As a consequence of these expansions, proteins with polyQ tracts are often found to misfold, oligomerize and aggregate to form fibrillar species resembling amyloid fibrils.

Characterizing the structural and dynamical properties of polyQ tracts is crucial for understanding the molecular basis of polyQ diseases.[3] However, these regions of low sequence complexity are considered challenging targets for conventional methods in structural biology. This is due, in addition to their low solubility, to their high propensity to be intrinsic disordered, that in general precludes crystallization, and to the highly repetitive nature of their primary sequence, that can render their investigation by NMR challenging. A number of pioneering studies attempted to determine the structure of polyQ tracts, but to date, it has not been possible to report on the structure of one such tract in its native context and without fusing it to a solubilizing moiety.[4;5]

We focus here on a polyQ tract occurring in the transactivation domain of the androgen receptor (AR) which plays a key role in the onset of *spinal bulbar muscular atrophy* (SBMA), a rare hereditary neuromuscular polyQ disease also known as *Kennedy disease*.[6] AR is a nuclear receptor activated by androgens such as dihydrotestosterone that regulates the expression of the male phenotype.[7] The polymorphic polyQ tract in AR starts at position 58 and can be between 14 and 34 residues long in healthy individuals. Sizes over 37 residues are associated with SBMA and the length of the polyQ tract anti-correlates with the age of onset.

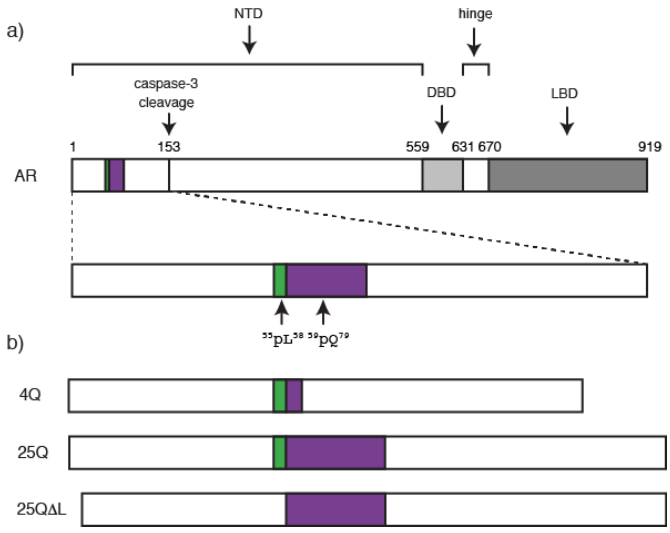


Figure 1. a) Relative positions of the transactivation (NTD), DNA (DBD) and ligand binding (LBD) domains of AR with an indication of the positions of the LLLL motif, shown in green, of the polymorphic polyQ tract, shown in purple, and of the caspase 3 cleavage site b) AR constructs used in this work.

[1] These authors contributed equally

[*] Dr. B. Eftekharzadeh, G. Chiesa, and Prof. X. Salvatella
BSC-CRG-IRB Joint Research Programme in Computational Biology
IRB Barcelona
Baldiri Reixac 10, 08028 Barcelona, Spain
E-mail: xavier.salvatella@irbbarcelona.org

A. Piai, D. Mungianu, Prof. R. Pierattelli, and Prof. I. C. Felli
CERM and Department of Chemistry "Ugo Schiff"
University of Florence
Via Luigi Sacconi 6, 50019 Sesto Fiorentino, Florence, Italy

Dr. J. García
IRB Barcelona
Baldiri Reixac 10, 08028 Barcelona, Spain

Prof. X. Salvatella
ICREA
Barcelona, Spain

[**] We would like to thank BioNMR (EC contract no. 261863), the NMR facility of the University of Barcelona (CCiT UB) for technical assistance and IRB, ICREA, Obra Social "la Caixa" (B.E.), AGAUR (G.C.), MICINN (CTQ2009-08850 to X.S.), MINECO (BIO2012-31043 to X.S.) and Marató de TV3 (102030 to X.S.) for funding this work. DM was a recipient of an Erasmus Placement fellowship from the University of Florence (Italy).

Supporting information for this article includes experimental details as well as three additional figures (Fig S1-S3).

The polyQ tract plays an important role in the formation of the aggregates associated with SBMA[8]. Neither its structure nor its involvement in the mechanism of AR aggregation are however known and this acts as an important hurdle for the development of therapeutic approaches for this rare disease. To characterize these important properties of polyQ tracts and understand how sequence context can influence their behavior, we used solution NMR to investigate the conformation of a N-terminal fragment of AR found in the aggregates associated with SBMA. This 156-residue fragment is the product of proteolytic cleavage by caspase 3 and plays a key role in the progression of the disease (Fig. 1).[6]

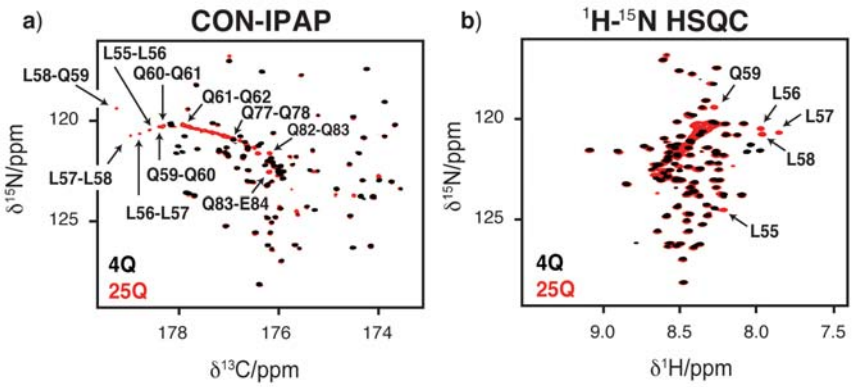


Figure 2. Comparison of the NMR spectra of 4Q and 25Q. Central region of (a) the CON-IPAP spectrum and (b) the ^{1}H-^{15}N HSQC spectrum of 4Q (black) and 25Q (red) with an indication of the resonances which experience the largest chemical shift variations upon increasing the length of the polyQ tract. A close up of the CON-IPAP spectrum of 25Q with the full assignment of the polyQ tract is provided as supplementary information (Fig. S1).

To study how the properties of the tract depend on its length we used constructs containing 4 and 25 Gln residues (4Q and 25Q). The disordered nature of the protein and the presence of a 25 residue polyQ tract required an experimental strategy based on the use of 4D ^{13}C-detected NMR experiments that lead to the complete sequence specific assignment of the fragment.[9;10] Heteronuclei were exploited to take advantage of their improved chemical shifts dispersion, compared to that of protons, and of their reduced sensitivity to exchange processes.[11] The extensive cross-peak overlap in the spectra was overcome by acquiring high-dimensional experiments and by exploiting non-uniform sampling (NUS) to reduce experimental time while preserving high spectral resolution in the indirect dimensions. To our knowledge, a full characterization at atomic resolution of a polyQ tract under native conditions is unprecedented. The chemical shifts of 4Q and 25Q ($^{1}H^{N}$, $^{1}H^{\alpha}$, $^{1}H^{\beta}$, ^{15}N, $^{13}C'$, $^{13}C^{\alpha}$, $^{13}C^{\beta}$) have been deposited in the BMRB (www.bmrb.wisc.edu, entries 25606 and 25607).[12]

As shown in Figure 2, in which we present the central regions of the 2D ^{1}H-^{15}N HSQC and 2D CON-IPAP spectra of 4Q and 25Q, the spectral differences between these two constructs are limited to the resonances of the polyQ tract and to the 4 Leu residues immediately preceding it (^{54}LLLL^{57}). An interesting feature of the spectra of 25Q is the presence of pseudo-diagonals, a characteristic fingerprint, consisting of the cross-peaks corresponding to the residues of the polyQ tract, which have H^{N}, N and C' chemical shifts that correlate with their position in the sequence of the fragment (see also Fig. S1).

We analyzed the chemical shifts to derive the conformational properties of 4Q and 25Q (Fig. 3a) in terms of secondary structural propensities (SSP).[13;14] The values that we obtained indicate the absence of persistent secondary structure except for the polyQ tract and for the 4 Leu residues preceding it, which show significant α-helical propensity, and a marked increase in helical propensity upon expansion of the tract from 4 to 25 residues. This is a quite surprising observation considering that polyQ tracts are in general disordered or, when involved in the formation of fibrils, in an extended conformation.[5,15] That the helical propensity of the polyQ tract increases with its length was confirmed by CD analysis (Fig. S3) and ^{15}N relaxation measurements (Fig. 3b and Fig. S2). Interestingly, the helicity was found to be most pronounced at the beginning of the tract and to decrease gradually towards its end.

The fact that the residues of the ^{54}LLLL^{57} motif, which precedes the polyQ tract, undergo very substantial chemical shift changes upon expansion of the latter (Fig. 2) indicates a certain degree of cooperativity in the conformational transition caused by expansion of the tract and suggests that they induce the formation of the α-helix. To confirm this we analyzed by CD and NMR a mutant of 25Q in which the motif ^{54}LLLL^{57} was removed (25QΔL). The results that we obtained indicate a substantial loss of α-helical propensity, which decreased from *ca* 30% to 5% according to CD measurements, and a marked decrease in the chemical shift dispersion in the resonances corresponding to the

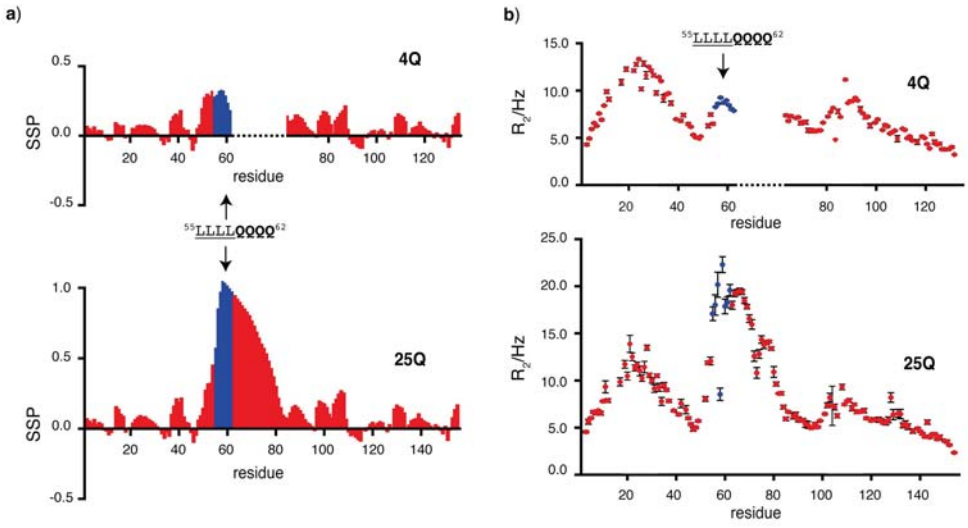


Figure 3. a) Secondary structure propensity of 4Q and 25Q. b) ^{15}N R_2 relaxation rates reported as a function of residue number for 4Q and 25Q. Additional relaxation parameters (^{15}N R_1 and {^{1}H}-^{15}N NOEs) are provided as supplementary information (Fig. S2). In a) and b), to facilitate the comparison, values for residues of 4Q which follow the polyQ tract are shifted to the right by 21 units. Values for residues 55 to 62, corresponding to the ^{55}LLLL^{58} motif and the first 4 Gln of the polyQ tract, are shown in blue to highlight the variation of the structural properties of the protein due to the different length of the polyQ tract

polyQ tract (Fig. 4). Both observations are in agreement with the hypothesis that the motif ^{54}LLLL^{57} induces a helical conformation in the polyQ tract.

To investigate the effect of the 4 Leu on the kinetics of oligomerization of the AR fragment, we compared the temporal evolution of 25Q and 25QΔL by dynamic light scattering (DLS), a technique that is well-suited to characterize the early stages of protein aggregation. The results (Fig. 4c) show that the particle size (D_{hz}) increases much faster and reaches higher values for 25QΔL than for 25Q, indicating that the motif ^{54}LLLL^{57} decreases the oligomerization propensity of the fragment. This observation, together with recent reports indicating that Leu is under-represented in intrinsically disordered regions[16] but over-represented in regions of sequence flanking polyQ tracts in eukaryotic proteomes,[17] supports the hypothesis that the latter have evolved to counterbalance the intrinsic aggregation propensity of these tracts by influencing their secondary structure.

The high resolution investigation of a polyQ tract achieved here by using a recently developed NMR strategy in combination with complementary techniques (CD, DLS) provides an example of how flanking regions can influence the secondary structure and aggregation propensity of such tracts. Indeed, our findings reveal how a motif flanking a polyQ tract at its N-terminus can induce helical structure in the latter, which in turn influence fibril formation by adding a helix unfolding step in the mechanism of the transition. The intra-molecular hydrogen bonds that stabilize this compact secondary structure must indeed break before the cross-β structure, rather elongated and stabilized by inter-molecular hydrogen bonds, can form.[18] Although this can indeed act as a mechanism to protect against aggregation our results also show that the influence of flanking regions on the secondary structure of polyQ tracts spans a limited range, in our case *ca* 20 residues (Fig. 3), and that it may cease to be effective for the long tracts which cause polyQ diseases.[19,20]

Our findings provide new insights into the structural properties of polyQ tracts and offer a compelling example of how flanking regions can affect the rate and extent of formation of toxic fibrillar species by proteins harboring this peculiar sequences. Our increased understanding of how sequence context influences the properties of polyQ tracts opens up new avenues for the development of therapeutic strategies for polyQ diseases based on targeting the flanking regions rather than the polyQ tract themselves.

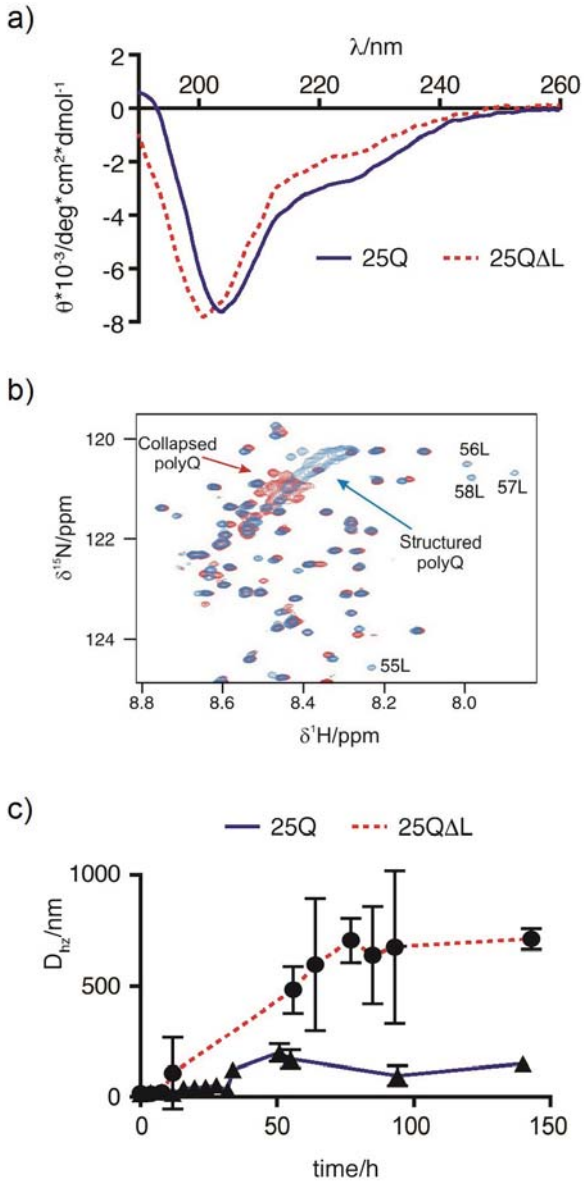


Figure 4. a) Comparison of the CD spectra of 25Q and 25QΔL. b) Comparison of the ^{1}H-^{15}N HSQC NMR spectra of 25Q and 25QΔL. The cross-peaks belonging to the four leucines of 25Q immediately preceding the polyQ tract, not present in 25QΔL, are indicated. The absence of helical propensities in 25QΔL, which reduces dramatically the chemical shift dispersion of the resonances belonging to polyQ tract, is highlighted. c) Temporal evolution of the mean particle size of 25Q and 25QΔL. In a), b) and c) 25Q and in 25QΔL are indicated in blue and red, respectively

Keywords: polyglutamine diseases · protein aggregation · intrinsically disordered proteins · nuclear magnetic resonance · ^{13}C-detection

[1] H. Y. Zoghbi, H. T. Orr, *Annu.Rev.Neurosci.* **2000**, *23*, 217-247.
[2] S. M. Mirkin, *Nature* **2007**, *447*, 932-940.
[3] R. Nalavade, N. Griesche, D. P. Ryan, S. Hildebrand, S. Krauß, *Cell Death and Disease* **2013**, *4*, e752.
[4] P. Li, K. E. Huey-Tubman, T. Gao, X. Li, A. P. West, M. J. Bennet, P. J. Bjorkman, *Nat.Struct.Mol.Biol.* **2007**, *14*, 381-387.
[5] L. Masino, G. Kelly, K. Leonard, Y. Trottier, A. Pastore, *FEBS Lett.* **2002**, *513*, 267-272.
[6] D. E. Merry, Y. Kobayashi, C. K. Bailey, A. A. Taye, K. H. Fischbeck, *Hum.Mol.Genet.* **1998**, *7*, 693-701.
[7] E. P. Gelmann, *J.Clin.Oncol.* **2002**, *20*, 3001-3015.
[8] S. Chen, V. Berthelier, J. B. Hamilton, B. O'Nuallain, R. Wetzel, *Biochemistry* **2002**, *41*, 7391-7399.
[9] W. Bermel, I. Bertini, L. Gonnelli, I. C. Felli, W. Kozminski, A. Piai, R. Pierattelli, J. Stanek, *J.Biomol.NMR* **2012**, *53*, 293-301.
[10] W. Bermel, I. C. Felli, L. Gonnelli, W. Kozminski, A. Piai, R. Pierattelli, A. Zawadzka-Kazimierczuk, *J.Biomol.NMR* **2013**, *57*, 353-361.
[11] S. Gil, T. Hošek, Z. Solyom, R. Kümmerlee, B. Brutscher, R. Pierattelli, I. C. Felli, *Angew.Chem.Int.Ed.Engl.* **2013**, *52*, 11808-11812.
[12] E. L. Ulrich, H. Akutsu, J. F. Doreleijers, Y. Harano, Y. E. Ioannidis, J. Lin, M. Livny, S. Mading, D. Maziuk, Z. Miller, E. Nakatani, C. F. Shulte, D. E. Tolmie, R. Kent Wenger, H. Yao, J. L. Markley, *Nucleic Acids Res.* **2007**, *36*, D402-D408.
[13] J. A. Marsh, V. K. Singh, Z. Jia, J. D. Forman-Kay, *Protein Sci.* **2006**, *15*, 2795-2804.
[14] K. Tamiola, F. A. Mulder, *Biochem.Soc.Trans.* **2012**, *40*, 1014-1020.

[15] L. E. Buchanan, J. K. Carr, A. M. Fluitt, A. J. Hoganson, S. D. Moran, J. J. de Pablo, J. L. Skinner, M. T. Zanni, *Proc.Natl.Acad.Sci. USA* **2014**, *111*, 5796-5801.

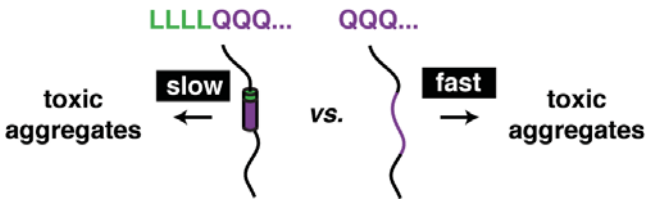
[16] P. Radivojac, L. M. Iakoucheva, C. J. Oldfield, Z. Obradovic, V. N. Uversky, A. K. Dunker, *Biophys.J.* **2007**, *92*, 1439-1456.

[17] M. Ramazzoti, E. Monsellier, C. Kamoun, D. Degl'Innocenti, R. Melki, *Plos ONE* **2012**, *7*, e30824.

[18] G. Calloni, C. Lendel, S. Campioni, S. Giannini, A. Gliozzi, A. Relina, M. Vendruscolo, C. M. Dobson, X. Salvatella, F. Chiti, *J.Am.Chem.Soc.* **2008**, *130*, 13040-13050.

[19] E. Scherzinger, A. Sittler, K. Schweiger, V. Heiser, R. Lurz, R. Hasenbank, G. P. Bates, H. Lehrach, E. E. Wanker, *Proc.Natl.Acad.Sci. USA* **1999**, *96*, 4604-4609.

[20] S. Chen, V. Berthelier, W. Yang, R. Wetzel, *J.Mol.Biol* **2001**, *311*, 173-182.



Bahareh Eftekharzadeh[1], Alessandro Piai[1], Giulio Chiesa, Daniele Mungianu, Jesús García, Roberta Pierattelli, Isabella Felli, and Xavier Salvatella

Page No. – Page No.

Sequence Context Influences the Structure and Aggregation Propensity of a PolyQ Tract

The presence of four consecutive leucine residues at the N-terminus of a polyglutamine tract, a repetitive sequence linked to nine neurodegenerative diseases, has dramatic consequences: it induces the tract to form an α-helix and in doing so hinders the formation of toxic aggregates that are rich in β structure. Modulating the structure of the regions of sequence flanking these tracts may offer opportunities to treat these diseases in the future.

Sequence Context Influences the Structure and Aggregation Propensity of a PolyQ Tract

Bahareh Eftekharzadeh[1], Alessandro Piai[1], Giulio Chiesa, Daniele Mungianu, Jesús García, Roberta Pierattelli, Isabella Felli, and Xavier Salvatella

[1] these authors contributed equally

Supporting Information

Table of contents

Experimental Section

Protein expression and purification

The genes codifying for 4Q, 25Q and 25QΔL were purchased from GeneArt and cloned in a pDEST-HisMBP vector (Addgene). Expression of the resulting genes led to fusion proteins containing a His_6 tag and a maltose binding protein (MBP) moiety that were used, respectively, to purify and increase the solubility of the proteins. ^{15}N labeled protein expression was carried in Rosetta *E. coli* cells grown in MOPS at 37ºC until the value of OD_{600} was 0.7 and induced with 0.5 mM Isopropyl β-D-1-thiogalactopyranoside (IPTG) for 4 hours at 28ºC. Cells were harvested by centrifugation and re-suspended in core buffer (50 mM sodium phosphate, 500 mM NaCl, 5% (v/v) Glycerol, 1 mM 2-Mercapto-ethanol, pH 8.0). A HisTrap HP 5 ml column (GE Healthcare) was used to purify the proteins, which were eluted by an imidazole gradient (final composition: 500 mM imidazole, 50 mM sodium phosphate, 500 mM NaCl, 5% Glycerol, 1 mM 2-Mercapto-ethanol, pH 8.0), followed by a size exclusion step carried out in a Superdex HighLoad S200 26/60 column (GE Healthcare) equilibrated in a buffer with the following composition: 500 mM NaCl, 12 mM sodium phosphate, 5% (v/v) Glycerol, 1 mM DTT, pH 7.5. The pure proteins were then incubated with His_6-tagged TEV protease for 16 hours at 4°C by dialysis against a buffer containing 20 mM sodium phosphate, 100 mM NaCl and 0.5 mM EDTA (Ethylenediaminetetraacetic acid), pH 8.0. The product of the proteolytic cleavage was purified by Ni^{2+} affinity chromatography, employing a buffer containing 8 M urea (500 mM imidazole, 50 mM sodium phosphate, 100 mM NaCl, 8 M urea, pH 8.0) to prevent the aggregation of the cleaved AR in the column. Finally, the cleaved proteins were stored at –80°C.

NMR sample preparation (Urea Incubation followed by Dialysis, UID)

The protein solutions stored at -80°C were thawed and dialyzed for 16 hours at 4°C against a buffer containing 20 mM sodium phosphate and 1 mM tris(2-carboxyethyl)phosphine (TCEP) at pH 7.4. Finally, 10% (v/v) D_2O and 0.015 mM DSS (4,4-dimethyl-4-silapentane-1-sulfonic acid) were added to the samples.

DLS and CD sample preparation (Stringent Disaggregation Procedure, SDP)

The protocol was adapted from the one developed by Linse group for Aß peptide kinetic measurements.[1] After purification, the lyophilized protein was dissolved in a buffer containing 20 mM sodium phosphate, 100 mM NaCl, 6 M Guanidine Thiocyanate, pH 7.4 and 5 mM TCEP until complete reduction of the protein, as monitored by HPLC measurement with a C18 column. The solution was passed through a PD-10 column (GE Healthcare) equilibrated in 20 mM sodium phosphate and 100 mM NaCl, pH 7.4. The most concentrated fraction was then purified with a Superdex 75 10/300 SEC column and the fractions containing exclusively the monomeric protein were centrifuged at 386,000 g for 1 hour at 4°C, using an Optima MAX preparative ultracentrifuge (Beckman). Only the upper ¾ of the centrifuged solutions were used for the assays.

Dynamic light scattering (DLS)

Before the measurements, samples were spinned with a tabletop centrifuge at 4ºC for 10 minutes at 13,000 rpm. Measurements were taken with a Malvern Zetasizer Nano S equipped with a He-Ne of 633 nm wavelength laser. For each experiment, 20 repetitions of 20 seconds were recorded. Three measurements were performed at each time point.

Circular Dichroism (CD) spectroscopy

Far-UV CD measurements were performed on a JASCO 815 spectropolarimeter using a 0.1 cm length cuvette. Freshly prepared sample was diluted to the desired concentration with 20 mM sodium phosphate buffer (pH 7.4). The spectra were acquired at 0.2 nm resolution with a scan rate of 50 nm/min. For each sample, 10 spectra were collected and averaged, after subtraction the blank. CD measurements were performed at 4ºC and 37ºC. Spectra were recorded in the range 190-260 nm.

NMR experiments

0.4 mM samples of ^{13}C,^{15}N double-labeled 4Q and 25Q in 20 mM sodium phasphate, 1 mM TCEP, pH 7.4 were prepared. 10 % (v/v) D_2O was added for the lock. Identical samples, but exclusively enriched in ^{15}N, were used to acquire ^{15}N relaxation experiments.

^{13}C-detected and ^{1}H-detected NMR experiments for sequence-specific resonance assignment were acquired at 16.4 T on a Bruker Avance spectrometer operating at 700.06 MHz ^{1}H, 176.03 MHz ^{13}C and 70.94 MHz ^{15}N frequencies, equipped with a cryogenically cooled probehead optimized for ^{13}C-direct detection (TXO), and at 22.3 T on a Bruker Avance III spectrometer operating at 950.20 MHz ^{1}H, 238.93 MHz ^{13}C and 96.28 MHz ^{15}N frequencies, equipped with a cryogenically cooled probehead (TCI). ^{15}N relaxation experiments were performed at 16.4 T Bruker Avance spectrometer operating at 700.13 MHz ^{1}H, 176.05 MHz ^{13}C and 70.94 MHz ^{15}N frequencies, equipped with a cryogenically cooled probehead (TXI), by measuring ^{15}N backbone longitudinal (R_1) and transverse (R_2) relaxation rates and the heteronuclear ^{15}N{^{15}H} NOEs. All the experiments were collected at 278 K.

A data set consisting of a combination of ^{13}C-detected (4D HCBCACON,[2] 4D HCBCANCO,[2] 4D (HCA)CON(CA)CON[3] and 4D (HN)CON(CA)CON[3]) and ^{1}H-detected (3D TROSY HNCO,[4] 4D TROSY (H)NCO(CA)NNH[5] and 4D TROSY HN(COCA)NNH[5]) NMR experiments was used to achieve the full sequence-specific assignment of 4Q. Instead, ^{13}C-detected 4D HCBCACON[2] and 4D (HN)CON(CA)CON[3] experiments, and ^{1}H-detected 3D TROSY HNCO[4], 3D TROSY HN(CA)CO[6] and 4D TROSY HN(COCA)NNH[5] experiments were acquired to obtain the complete resonances assignment of 25Q. All the experiments were performed using on-grid non-uniform sampling (NUS). The o “Poisson disk” sampling scheme was chosen to generate the time schedules with the RSPack program.[7] The assignments of H^N, H^α, H^β, C', C^α, C^β and N resonances of 4Q and 25Q are reported in the BMRB (www.bmrb.wisc.edu),[8] entries 25606 and 25607. The parameters used for the acquisition of the experiments are reported in Tables S1-S4.

NMR data processing and analysis

NUS NMR data were converted with *NMRPipe*[9] and then processed using *ToASTD*[10] and *reduced*[11;12] programs. Uniformly sampled NMR data were processed with *TopSpin*. *Sparky*[13] and *CcpNMR Analysis*[14] were used to visualize the spectra and analyze the ^{15}N relaxation data, respectively.

The secondary structure propensity from the heteronuclear chemical shifts was determined for 4Q and 25Q by using the *neighbor corrected structural propensity calculator* (ncSPC) tool,[15] available online at http://nmr.chem.rug.nl/ncSPC/.

Tamiola, Acar and Mulder random coil chemical shift library was chosen for the analyses.[16]

Table S1 Experimental parameters used for the acquisition of the NMR experiments on 4Q for sequence-specific resonance assignment

	Spectral widths and maximal evolution times				No. of scans	Inter-scan delays (s)	No. of complex points (aq)	No. of hypercomplex points	Duration of the experiment	Relative data points density (%)
	Indirect dimensions			Direct dimension						
4D HCBCACON	5000 Hz ($^{1}H^{\alpha/\beta}$) 20.0 ms	12500 Hz ($^{13}C^{\alpha/\beta}$) 7.5 ms	2600 Hz (^{15}N) 50.0 ms	8800 Hz ($^{13}C'$)	8	0.9	512	850	1 day, 10 hours	0.07
4D HCBCANCO	5000 Hz ($^{1}H^{\alpha/\beta}$) 35.7 ms	12500 Hz ($^{13}C^{\alpha/\beta}$) 35.7 ms	2600 Hz (^{15}N) 32.0 ms	8800 Hz ($^{13}C'$)	16	0.9	512	850	3 days, 2 hours	0.11
4D (HCA)CON(CA)CON	2200 Hz ($^{13}C'$) 24.1 ms	2600 Hz (^{15}N) 24.2 ms	2600 Hz (^{15}N) 30.0 ms	8800 Hz ($^{13}C'$)	16	0.9	512	930	3 days, 9 hours	0.34
4D (HN)CON(CA)CON	2200 Hz ($^{13}C'$) 24.1 ms	2600 Hz (^{15}N) 24.2 ms	2600 Hz (^{15}N) 30.0 ms	8800 Hz ($^{13}C'$)	32	0.5	512	910	4 days, 11 hours	0.33
3D TROSY HNCO		2700 Hz ($^{13}C'$) 20.7 ms	2300 Hz (^{15}N) 21.7 ms	14200 Hz (^{1}H)	8	1.2	1024	560	7 hours	20.00
4D TROSY (H)NCO(CA)NNH	2300 Hz (^{15}N) 20.4 ms	2700 Hz ($^{13}C'$) 23.9 ms	2300 Hz (^{15}N) 23.9 ms	13300 Hz (^{1}H)	8	1.2	1024	2660	2 days, 20 hours	1.60
4D TROSY HN(COCA)NNH	1500 Hz (^{1}H) 20.0 ms	2300 Hz (^{15}N) 23.9 ms	2300 Hz (^{15}N) 23.9 ms	13300 Hz (^{1}H)	8	1.2	1024	1450	1 day, 13 hours	1.60

Table S2 Experimental parameters used for the acquisition of the NMR experiments on 25Q for sequence-specific resonance assignment

	Spectral widths and maximal evolution times				No. of scans	Inter-scan delays (s)	No. of complex points (aq)	No. of hypercomplex points	Duration of the experiment	Relative data points density (%)
	Indirect dimensions			Direct dimension						
4D HCBCACON	5000 Hz ($^{1}H^{\alpha/\beta}$) 20.0 ms	12500 Hz ($^{13}C^{\alpha/\beta}$) 7.5 ms	2600 Hz (^{15}N) 50.0 ms	8800 Hz ($^{13}C'$)	8	0.9	512	850	1 day, 10 hours	0.07
4D (HN)CON(CA)CON	2200 Hz ($^{13}C'$) 24.1 ms	2600 Hz (^{15}N) 24.2 ms	2600 Hz (^{15}N) 30.0 ms	8800 Hz ($^{13}C'$)	32	0.5	512	910	4 days, 11 hours	0.33
3D HNCO		2000 Hz ($^{13}C'$) 35.5 ms	2000 Hz (^{15}N) 24.0 ms	10500 Hz (^{1}H)	16	1.0	1024	1130	1 day	33.30
3D TROSY HN(CA)CO		1800 Hz ($^{13}C'$) 25.0 ms	2400 Hz (^{15}N) 20.8 ms	13300 Hz (^{1}H)	16	1.2	1024	580	14 hours	26.00
4D TROSY HN(COCA)NNH	1500 Hz (^{1}H) 20.0 ms	2400 Hz (^{15}N) 22.9 ms	2400 Hz (^{15}N) 22.9 ms	13300 Hz (^{1}H)	8	1.2	1024	1720	1 day, 20 hours	1.90

Table S3 Experimental parameters used for the acquisition of the ^{15}N relaxation NMR experiments on 4Q

	Spectral widths and maximal evolution times		No. of scans	Inter-scan delays (s)
^{15}N R_1	1600 Hz (^{15}N) 156.8 ms	10500 Hz (^{1}H) 97.6 ms	8	3.0
^{15}N R_2	1600 Hz (^{15}N) 156.8 ms	10500 Hz (^{1}H) 97.6 ms	8	3.0
Steady-state heteronuclear $^{15}N\{^{1}H\}$ NOEs	1600 Hz (^{15}N) 156.8 ms	10500 Hz (^{1}H) 97.6 ms	64	6.0
For the determination of R_1, 10 experiments were acquired changing the variable delay from 15 to 995 ms. For the determination of R_2, 10 experiments were acquired changing the variable delay from 30 to 565 ms.				

Table S4 Experimental parameters used for the acquisition of the ^{15}N relaxation NMR experiments on 25Q

	Spectral widths and maximal evolution times		No. of scans	Inter-scan delays (s)
^{15}N R_1	1600 Hz (^{15}N) 156.8 ms	10500 Hz (^{1}H) 97.6 ms	8	3.0
^{15}N R_2	1600 Hz (^{15}N) 156.8 ms	10500 Hz (^{1}H) 97.6 ms	8	3.0
Steady-state heteronuclear $^{15}N\{^{1}H\}$ NOEs	1600 Hz (^{15}N) 177.7 ms	10500 Hz (^{1}H) 97.6 ms	64	6.0
For the determination of R_1, 10 experiments were acquired changing the variable delay from 15 to 995 ms. For the determination of R_2, 10 experiments were acquired changing the variable delay from 30 to 315 ms.				

Supporting Figures

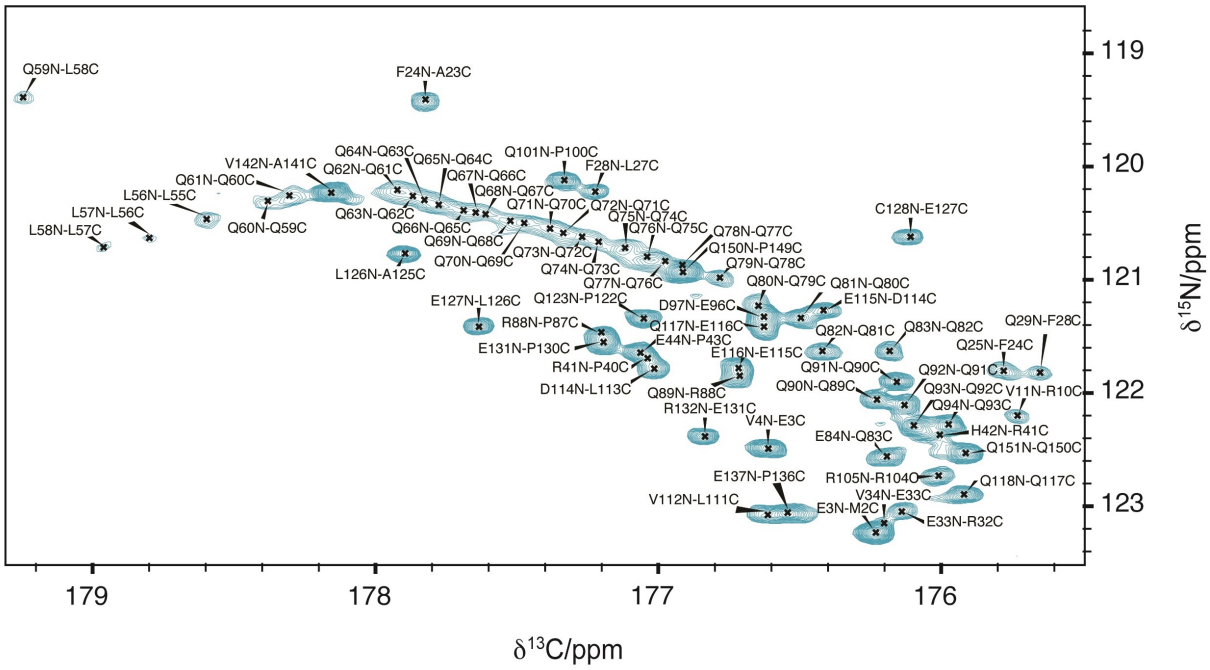


Figure S1 Close up view of the CON-IPAP spectrum of 25Q with the full assignment of the polyQ tract. ^{13}C dimension increases the chemical shift dispersion of the resonances and reduces cross-peak overlaps in the polyQ region.

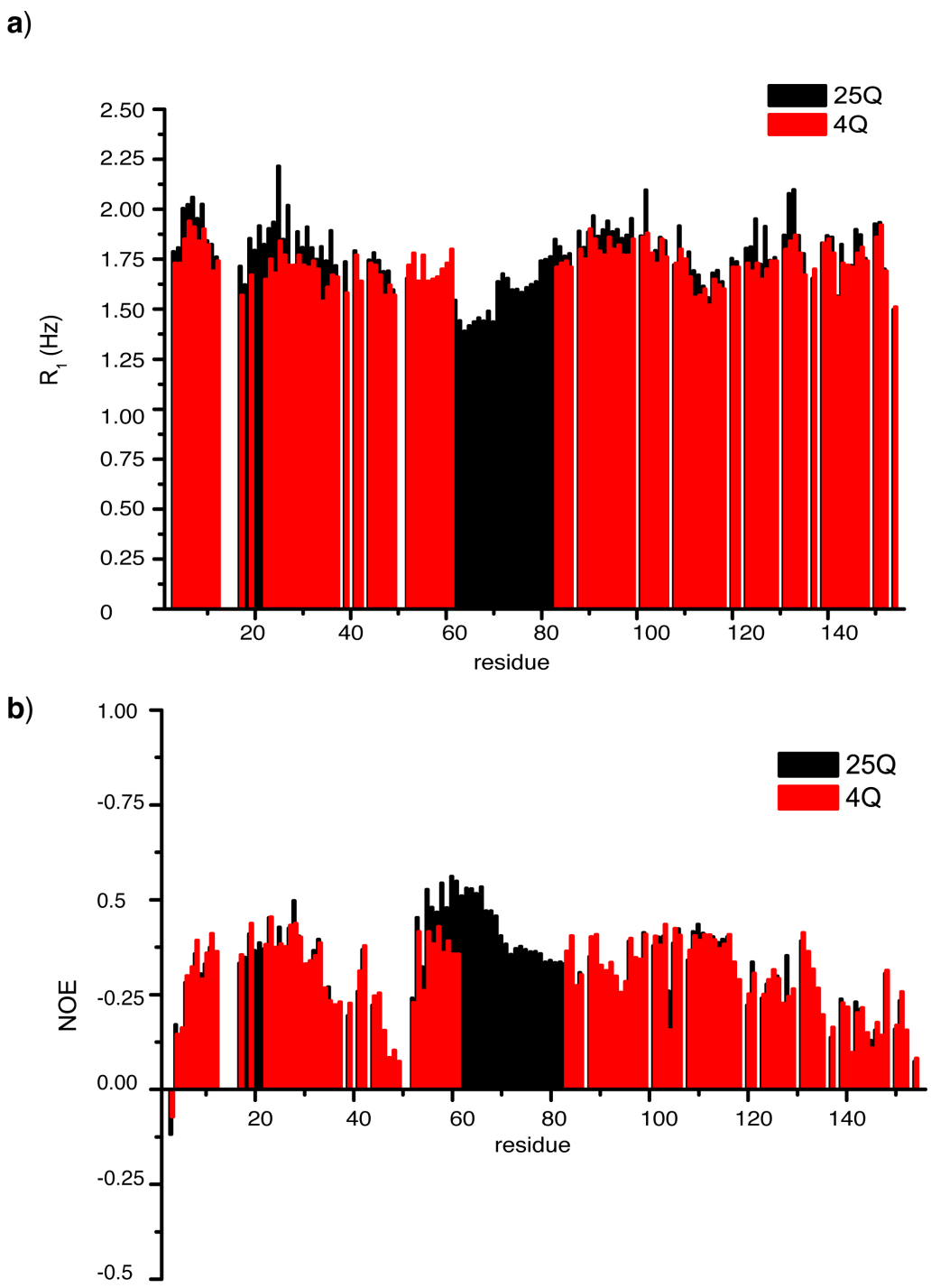


Figure S2 ^{15}N R_1 relaxation rates and $\{^1H\}$-^{15}N NOEs reported as a function of residue number for 4Q (red) and 25Q (black). To facilitate the comparison, error bars have been removed and values for residues of 4Q which are C-terminal to the polyQ tract are shifted to the right by 21 units.

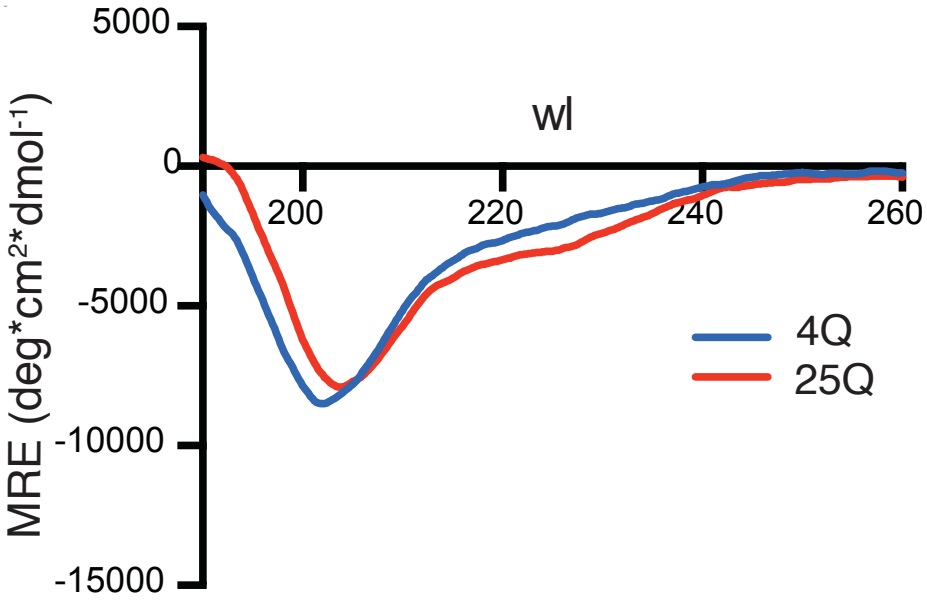


Figure S3 CD spectra of 130 μM samples of 4Q and 25Q

References

[1] E. Hellstrand, B. Boland, D. M. Walsh, S. Linse, *ACS Chem.Neurosci.* **2010**, *1*, 13-18.

[2] W. Bermel, I. Bertini, L. Gonnelli, I. C. Felli, W. Koźmiński, A. Piai, R. Pierattelli, J. Stanek, *J.Biomol.NMR* **2012**, *53*, 293-301.

[3] W. Bermel, I. C. Felli, L. Gonnelli, W. Koźmiński, A. Piai, R. Pierattelli, A. Zawadzka-Kazimierczuk, *J.Biomol.NMR* **2013**, *57*, 353-361.

[4] L. E. Kay, M. Ikura, R. Tschudin, A. Bax, *J.Magn.Reson.* **1990**, *89*, 496-514.

[5] A. Piai, T. Hošek, L. Gonnelli, A. Zawadzka-Kazimierczuk, W. Koźmiński, B. Brutscher, W. Bermel, R. Pierattelli, I. C. Felli, *J.Biomol.NMR* **2014**, *60*, 209-218.

[6] R. T. Clubb, V. Thanabal, G. Wagner, *J.Magn.Reson.* **1992**, *97*, 213-217.

[7] K. Kazimierczuk, A. Zawadzka, W. Koźmiński, *J.Magn.Reson.* **2008**, *192*, 123-130.

[8] E. L. Ulrich, H. Akutsu, J. F. Doreleijers, Y. Harano, Y. E. Ioannidis, J. Lin, M. Livny, S. Mading, D. Maziuk, Z. Miller, E. Nakatani, C. F. Shulte, D. E. Tolmie, R. Kent Wenger, H. Yao, J. L. Markley, *Nucleic Acids Res.* **2007**, *36*, D402-D408.

[9] F. Delaglio, S. Grzesiek, G. W. Vuister, G. Zhu, J. Pfeifer, A. Bax, *J.Biomol.NMR* **1995**, *6*, 277-293.

[10] K. Kazimierczuk, A. Zawadzka, W. Koźmiński, I. Zhukov, *J.Biomol.NMR* **2006**, *36*, 157-168.

[11] K. Kazimierczuk, A. Zawadzka, W. Koźmiński, *J.Magn.Reson.* **2009**, *197*, 219-228

[12] K. Kazimierczuk, A. Zawadzka-Kazimierczuk, W. Koźmiński, *J.Magn.Reson.* **2010**, *205*, 286-292.

[13] T. D. Goddard, D. G. Kneller, SPARKY 3, University of California, San Francisco, **2000**.

[14] W. F. Vranken, W. Boucher, T. J. Stevens, R. H. Fogh, A. Pajon, Llinas M, E. L. Ulrich, J. L. Markley, J. Ionides, E. D. Laue, *Proteins: Struct., Funct., Bioinf.* **2005**, *59*, 687-696.

[15] K. Tamiola, F. A. Mulder, *Biochem.Soc.Trans.* **2012**, *40*, 1014-1020.

[16] K. Tamiola, B. Acar, F. A. A. Mulder, *J.Am.Chem.Soc.* **2010**, *132*, 18000-18003.

Article 9. Early stages of androgen receptor oligomerization in spinal bulbar muscular atrophy

Alessandro Piai[1], Bahareh Eftekharzadeh[2], Giulio Chiesa[2], Jesús García[2], Roberta Pierattelli[1], Isabella C. Felli[1], Xavier Salvatella[2,3]

[1] CERM and Department of Chemistry "Ugo Schiff", University of Florence, 50019, Via Luigi Sacconi 6, Sesto Fiorentino, Florence, Italy.

[2] IRB Barcelona, Baldiri Reixac 10, 08028 Barcelona, Spain.

[3] ICREA, Barcelona, Spain.

Manuscript in preparation

Abstract

Spinal bulbar muscular atrophy (SBMA) is a rare hereditary neuromuscular disease caused by the elongation of a polymorphic polyglutamine (polyQ) tract in the transactivation domain (NTD) of the androgen receptor (AR). Although the molecular basis of SBMA is not yet fully understood, the observation of nuclear inclusions containing AR in specific tissues has suggested that it is linked to the aggregation of AR fragments. To characterize the mechanism of this process we have investigated the structural properties of AR constructs with polyQ tracts of increasing length as well as the early stages of oligomerization by solution nuclear magnetic resonance (NMR) spectroscopy. We identified a hydrophobic and Q/N-rich specific region of sequence, distant from the polyQ tract, as that responsible for the inter-molecular interactions that nucleate AR oligomerization.

Introduction

Spinal bulbar muscular atrophy (SBMA), also known as Kennedy disease[1], is a rare neuromuscular disease with a prevalence of 1 per ca. 50,000 males that belongs to a group of nine disorders termed polyglutamine (polyQ) diseases[2,3]. These occur as a result of the expansion of polymorphic polyQ tracts in proteins that are otherwise unrelated[4]. The variable length of such tracts is due to their unstable nature, that is caused by the propensity of CAG and GTC codon repeats to form non-B-DNA structures which cause slippage during DNA replication[5,6]. As a consequence of these expansions the codified proteins are often found to misfold, oligomerize and aggregate to form fibrillar species resembling amyloid fibrils[7–11].

The polyQ tract responsible for SBMA is found in the N-terminal transactivation domain (NTD) of androgen receptor (AR) (Figure 1a). AR is a nuclear receptor activated by androgens such as dihydrotestosterone that regulates the expression of the male phenotype. Although the molecular basis of SBMA is not yet fully established there are strong indications that AR aggregation is the main cause of the disease[12]. This hypothesis is supported by the observation that SBMA patients exhibit AR loss of function symptoms, such as gynecomastia and infertility, attributed to the depletion of the soluble receptor, in addition to gain of function symptoms,

such as progressive muscle cramps, fasciculations and dysphagia[13], caused by the cell-specific toxicity of AR aggregates to muscle cells and motor neurons[9,14–18].

Prior to activation by androgens AR is a cytosolic protein in complex with molecular chaperones that bind to the NTD (residues 1-558, Figure 1a), which is intrinsically disordered (ID)[19], and stabilize it against degradation and aggregation. Androgen binding to the ligand-binding domain (LBD, residues 670-919, Figure 1a) leads to a collective conformational change that causes the dissociation of this complex and targets the receptor to the nucleus by exposure of a nuclear localization signal found in the hinge region that connects the LBD with the DNA-binding domain (DBD, residues 559–631, Figure 1a). The DBD of nuclear AR interacts with specific DNA sequences known as androgen response elements (AREs) found in promoters and enhancers of genes regulated by AR. Once AR is bound to DNA, the NTD recruits the general transcription factors and transcriptional co-regulators needed to assemble the transcription machinery, leading to transactivation[20].

The polymorphic polyQ tract in AR starts at position 58 and can be between 14 and 34 residues long in healthy individuals. Sizes over 37 residues are associated with SBMA and the length of the polyQ tract anti-correlates with the age of onset[21]. Although the role of the polyQ tract is not yet known, there are indications that it is involved in the regulation of AR function as individuals with short polyQ tracts exhibit a higher propensity to suffer from prostate cancer, that relies on AR expression and activation for progression[22–24].

There is evidence from *Drosophila melanogaster* and mice models that the development of the SBMA phenotype requires AR activation[25,26]. These observations suggest a mechanism for the onset of this disease in which AR aggregation occurs when the complex that it forms with molecular chaperones dissociates after androgen binding. Given that activation leads to nuclear translocation this hypothesis is in agreement with the nuclear localization of AR aggregates, known as nuclear inclusions (NIs), and with the results of studies of the effect of AR antagonists on transgenic mice expressing human AR with an elongated polyQ tract. Indeed untreated mice have a phenotype closely resembling that of SBMA patients whereas those treated with AR antagonists appear healthy[27–29].

Although the polyQ tract is thought to play an important role in the formation of the inclusions associated with SBMA neither its structure nor its involvement in the mechanism of AR oligomerization are well characterized and this lack of understanding acts as an important hurdle for the development of therapeutic approaches for SBMA based on decreasing of AR aggregates. To remedy this, we have used solution NMR to investigate the structural properties as well as the early stages of oligomerization of a N-terminal fragment of AR found in NIs. This fragment is the product of proteolytic cleavage by caspase 3 (Figures 1a and 1c) and plays a key role in the onset of the disease[12,30,31].

The NMR strategy we used[32] to investigate this system enabled the structural and dynamical characterization of the entire fragment and to monitor the occurrence of the oligomerization process. The availability of residue-specific information enabled us to establish that the rate of oligomerization does not exclusively depend on the length of the polyQ tract providing striking example of how regions immediately flanking a polyQ tract can influence its structural properties and of how regions of sequence distant from it can modulate its propensity to oligomerize. In addition, our data unambiguously identify the region of sequence ^{22}GAFQNLFQSVREVIQ^{36} (Figure 1b) as that nucleating the oligomerization of AR in SBMA and put it forward as a possible therapeutic target for this currently incurable and relatively neglected rare disease.

Results

Preparation of solutions of monomeric AR fragments

Preparing solutions of monomeric aggregation-prone proteins is challenging but essential for characterizing their structural properties[33–35]. We expressed in *E. coli* two AR constructs spanning the sequence of the fragment and containing polyQ tracts of 4 and 25 Gln residues (4Q and 25Q, Figure 1c) fused to His_6-tagged maltose binding protein (MBP). After purification of the fusion proteins by affinity and size exclusion chromatographies and cleavage of the His_6-MBP moiety by the tobacco etch virus (TEV) protease 4Q and 25Q were rendered monomeric by

applying a stringent disaggregation procedure (SDP, see Materials and Methods) based on the one recently put forward by the Linse laboratory for the preparation of monomeric $A\beta_{1-42}$ solutions[34] and successfully applied also to 4Q and 25Q[36].
Solutions of 4Q and 25Q obtained in this way were shown to be more than 98% monomeric by dynamic light scattering (DLS) (Supplementary Figure 1). The hydrodynamic radii of 4Q and 25Q calculated from the DLS data were *ca* 3.5 nm, in agreement with the values determined from the elution volume obtained by size exclusion chromatography (SEC). The oligomeric state of the samples was also analyzed by sedimentation velocity analytical ultracentrifugation (SV-AUC), which yielded a narrow distribution of particles centered on a main species with a sedimentation coefficient of 1.1 S (for 4Q) and 1.3 S (for 25Q). Data fitting performed with the software SEDFIT[37] revealed, for both polyQ lengths, a low but detectable population of a second species (1-3%), with a sedimentation coefficient of 2.6 S (4Q) and 2.8 S (25Q). By comparison of the experimental frictional ratios with their theoretical counterparts we obtained that, for both constructs, the major species corresponded to the monomer, whereas the minor one likely to a tetramer (Supplementary Figure 2). A similar distribution of oligomeric states was recently described for synthetic peptides corresponding to exon 1 of the protein Huntingtin, which also harbors a polyQ tract[38].

Analysis of AR oligomerization

To characterize the oligomerization of 4Q and 25Q we monitored the time evolution of 20 μM samples of these constructs by time resolved DLS, a technique that is very sensitive to the presence of soluble oligomers. The analysis of the temporal evolution of the weighted mean diameter of the ensemble of particles (Z-average-D_{hz} (d.nm)) revealed that, in its early stages (<150 hours), oligomerization occurs in both cases with a rate that appeared to be relatively independent of the length of the polyQ tract (Figure 2a).

We then monitored the structural properties of the AR fragment and their evolution during oligomerization by circular dichroism (CD) and NMR spectroscopies. CD spectra collected immediately after preparing the monomeric protein solutions

indicated that 4Q and 25Q were ID with a fractional content of α-helix (ca. 12% for 4Q and ca. 30% for 25Q, Figure 2b). To study whether the secondary structure of 25Q evolved with time, a 20 μM solution of this protein was analyzed by CD at various times within 170 hours of incubation: the resulting spectra indicated that no substantial change in secondary structure occurred during oligomerization (Figure 2c).

Time-resolved NMR experiments were then performed to study the early stages of oligomerization at high resolution. A 50 μM sample of 25Q enriched in ^{15}N was incubated under the same conditions employed for the DLS and CD measurements and a series of 2D ^{1}H-^{15}N HSQC experiments was acquired at 18.8 T (800 MHz) in a time-resolved fashion. The majority of resonances remained unaltered but a fraction of the cross-peaks of 25Q evolved with time by decreasing in intensity and, in some instances, by splitting in two or more peaks (Figure 3).

That a relatively small number of resonances evolved during oligomerization indicated that the changes in chemical environment that occurred during this process were specific and, therefore, that NMR had the potential to provide unique insights into the mechanism of self-assembly[39]. However, the low stability of these samples precluded the acquisition of the multidimensional NMR experiments necessary to assign the resonances of the polypeptide.

Preparation of a kinetically stable sample and characterization by NMR

To produce kinetically stable samples, 400 μM solutions of 4Q and 25Q were incubated in 6 M urea overnight and then dialyzed (1:1000) against 20 mM sodium phosphate buffer at pH 7.4 as previously described[36]. The 2D ^{1}H-^{15}N HSQC spectrum of the sample obtained by this procedure (urea incubation followed by dialysis, UID) was very similar to that of monomeric 25Q obtained by SDP and only presented changes in the resonances corresponding to residues near H42 and H103, which we attribute to small pH differences and to the influence of residual urea. Importantly for our purposes this spectrum did not change significantly over the period of time required to acquire the multidimensional NMR experiments that were used to assign the resonances. The kinetic stability of this sample was also

confirmed by DLS, which showed that the Z-average was, although slightly higher than that of the SDP sample, indicating the presence of oligomers of 25Q, approximately constant (Supplementary Figure 4).

The available sequence-specific assignments of both UID 4Q and 25Q[36] and the secondary structure propensity analysis (SSP) performed to derive the conformational properties of 4Q and 25Q from the C', C^{α}, C^{β}, and N backbone chemical shifts[40-41] indicates the absence of persistent secondary structure in the polypeptides, except for the residues including the polyQ tract[36], which in the longer construct appears to be fully formed (Figure 4a). The backbone dynamics of 4Q and 25Q studied by measuring ^{15}N nuclear relaxation rates lead to similar results for 4Q and 25Q, except for the polyQ tract and the regions of sequence immediately flanking it. Notably, the R_2 values identified three regions of sequence characterized by particularly fast transverse ^{15}N relaxation rates: those, which in 4Q correspond to the regions ^{22}GAFQNLFQ^{29}, ^{55}LLLLQQQQ^{62} and ^{87}TGYLVLD^{93}, the latter corresponding to residues 108-114 in 25Q.

^{22}GAFQNLFQ^{29} is part of a region of sequence of low disorder propensity[42] and high average area buried upon folding (AABUF), a property of amino acids that has a direct correlation with hydrophobicity[43] (Figure 1b). Importantly, in full length androgen-bound AR, it can interact with the LBD[44,45] and this inter-domain interaction is thought to be critical for the physiological function of the receptor. The second region, ^{55}LLLLQQQQ^{62}, also with high AABUF, was identified by the SSP analysis as partially α-helical in 4Q and remarkably structured in 25Q (Figure 4a). The third one, ^{87}TGYLVLD^{93} in 4Q or ^{108}TGYLVLD^{114} in 25Q, is part of a region of sequence with relatively high AABUF (Figure 1b), which has no helical propensity (Figure 4a).

A comparison of the transverse ^{15}N relaxation rates of 4Q and 25Q reveals substantially higher values for the ^{55}LLLLQQQQ^{62} region of the latter, which we attribute mainly to the increase in helical propensity caused by the elongation of the polyQ tract (Figure 4b).[36] Equivalent results were obtained in analyzing the hetNOE and longitudinal ^{15}N backbone relaxation rate (Figure 4c and 4d), which in both cases reflect the formation of a more rigid structure upon elongation. In summary,

therefore, we find that all regions of sequence of 25Q showing fast backbone ^{15}N transverse relaxation have high AABUF and, hence, high hydrophobicity.

Identification of the residues undergoing structural changes during oligomerization of 25Q

The remarkable similarity of the ^{1}H-^{15}N HSQC spectra of UID and SDP samples of 25Q (Supplementary Figure 3) allowed us to unambiguously transfer the ^{1}H and ^{15}N backbone chemical shift assignments from the former to the latter, and therefore to identify the residues of 25Q which undergo changes in chemical environment during oligomerization (Figure 3). This analysis revealed that they belong to three regions of sequence: ^{22}GAFQNLFQSVREVIQ^{36}, ^{139}GAAVAAS^{145} and, to a lesser extent, the polyQ tract and the residues immediately flanking it.

A detailed analysis of the temporal evolution of the position and intensity of the cross-peaks in the ^{1}H-^{15}N HSQC spectrum revealed that the regions of sequence described above undergo changes with different rates. The first of these changes, which is already noticeable after two days of incubation, halves the signal intensity ($I/I_0 \approx 0.5$) of the resonance of the residues in the ^{22}GAFQNLFQSVREVIQ^{36} region and also affected, although to a lesser extent, residues in the ^{139}GAAVAAS^{145} region ($I/I_0 \approx 0.65$). Incubation of the sample for two additional days leads to further decreases in the signal intensity to values of I/I_0 of 0.30 and 0.45, respectively. Additionally, a modest decrease ($I/I_0 \approx 0.70$) in the region 50-85, which includes the polyQ tract and the residues flanking it, is observed. Further incubation for a total of seven days leads to major decreases in signal intensity of the residues constituting the ^{22}GAFQNLFQSVREVIQ^{36} region, which in many cases can no longer be observed ($I/I_0 \approx 0$), and to an enhancement of the effects observed in the polyQ tract and in the ^{139}GAAVAAS^{145} region ($I/I_0 \approx 0.6$ and 0.3, respectively).

As already mentioned, the resonances of a small number of residues, in addition to decreasing in the intensity split into two or more peaks. These resonances, which are amongst those that exhibit the earliest changes in intensity, correspond to the region of sequence ^{29}QSVREVIQ^{36}. This indicates that oligomerization proceeds via the formation of species with well-defined chemical shifts in this specific region of

sequence. A particularly striking case is Val 31, shown in detail in Figure 3c, which evolves via the formation of at least one intermediate, in slow exchange with monomeric 25Q, with chemical shifts that have values between those of the monomer and those observed at the end of incubation.

Although most of the cross-peaks in the ^{1}H-^{15}N HSQC spectra of the SPD and UID samples are in the same positions, allowing the transfer of the assignments, the intensities of many resonances of the sample treated with urea are lower than those measured in the spectra of samples that were disaggregated (Supplementary Figure 3). In addition, whereas the Z-average of the SDP samples was very low shortly after dissolution (*ca.* 8 nm), but quickly increases as oligomerization takes place, that of the UID samples has a relatively low value of *ca*. 100 nm and hardly increases during incubation (Supplementary Figure 4). Taken together, these results strongly suggest that the kinetically stable samples contain oligomeric species in equilibrium with monomeric 25Q.

To study the nature of inter-molecular interactions that stabilize these oligomers, we acquired 2D ^{1}H-^{15}N HSQC spectra of UID 25Q samples at various concentrations (125, 250 and 500 μM) and observed that the positions of all resonances remained unaltered but that the relative signal intensity of those corresponding to residues in the region ^{22}GAFQNLFQ^{29} decreased with concentration (Supplementary Figure 5). Subsequent measurements of the ^{15}N transverse relaxation rates (R_2) confirmed that the decrease in relative signal intensity was indeed due to the concentration-dependence of the ^{15}N transverse relaxation rates, which we attribute to the contribution of exchange processes (R_{ex}) to the linewidth of the resonances of 25Q due to oligomerization[46] (Figure 5).

In summary, we have listed the regions of 25Q that undergo changes in their structure and dynamics as oligomerization takes place. We clearly identify the region ^{22}GAFQNLFQSVREVIQ^{36} as that establishing the first inter-molecular interactions in the mechanism of self-assembly. It is interesting to note that this region has hydrophobicity and high propensity to fold and in addition, is rich in Q and N residues.

Discussion

The characterization of the structural and dynamical properties of polyQ tracts is crucial for understanding the molecular basis of polyQ diseases. However, they are generally considered very challenging targets for conventional structural biology approaches such as X-ray crystallography and solution NMR. This is due, in addition to their low solubility, to their high propensity to intrinsic disorder, which in general precludes crystallization, and to the highly repetitive nature of their primary sequences, which can render the NMR resonances assignment challenging due to spectral overlap[32,47-49].

In a previous work we have used relatively sophisticated sample preparation methods and a battery of newly developed solution NMR experiments[43] to obtain the complete assignment of the backbone resonances of the protein, including the polyQ tract,[36] and to derive from them the secondary structure propensity of each residue, including prolines.[36] Our findings indicate that the polyQ tract of the NTD of AR has a strong propensity to adopt an α-helical secondary structure contrary to what had been found to be the case for the vast majority of such tracts studied to date[50-55].

Studies of the mechanism of aggregation of proteins with polyQ tracts have relied on comparing the aggregation rates of peptides with tracts of various lengths after introducing mutations in the flanking regions[37,59]. They have showed a correlation between the length of the polyQ and the rate of aggregation[60] and have suggested that the flanking regions can play a role in the early stages of aggregation by stabilizing on-pathway oligomers through inter-molecular interactions[61]. Our results based on time-resolved NMR experiments provide a direct and unprecedented characterization of the mechanism of self-assembly and identify precisely which residues undergo structural and dynamic changes during oligomerization.

The results that we have obtained indicate that monomeric 25Q is in equilibrium with an oligomeric form of the protein stabilized by inter-molecular interactions involving residues in the motif ^{22}GAFQNLFQ^{29} (Figure 6). This motif is in a Q/N-rich region of sequence of low helical propensity but high AABUF (Figure 1b), suggesting that the

nature of the interactions, which stabilize the oligomers that form earliest, is hydrophobic. Given that the equilibrium between monomeric 25Q and these oligomers is relatively fast on the NMR timescale, it is unlikely that 25Q undergoes substantial structural changes of the type that occur at later stages, discussed below. Most of the residues of 25Q, which experience chemical shift perturbations during oligomerization, are found in the motif ^{30}SVREVIQ^{36}, of predicted high helicity, which is directly adjacent to the motif ^{22}GAFQNLFQ^{29} discussed above (Figure 3). That the resonances of the monomeric ^{30}SVREVIQ^{36} motif are in slow exchange with those appearing during incubation indicates that monomeric 25Q coexists with well-defined oligomeric species, with a different conformation in this region of sequence, that are of a size which do not prevent their observation by solution NMR. We therefore conclude that the second step in the oligomerization of 25Q is the formation of oligomers stabilized by inter-molecular interactions involving, at least initially, mainly residues in the motif ^{30}SVREVIQ^{36}.

The chemical shift perturbations observed in ^{30}SVREVIQ^{36} occur concomitantly with a decrease in signal intensity of the resonances belonging to monomeric 25Q (Figure 3). It is important to clarify that a substantial number of residues of 25Q, such as residues 1-10 and 90-120, do not experience such a decrease in signal intensity during incubation (Figure 3b). This observation confirms that the phenomenon described in Figure 3B is not due to the aggregation of 25Q but, rather, to the structural and dynamic changes occurring in the protein as it oligomerizes.

By using such decreases in signal intensity as reporters of the involvement of specific residues in oligomerization, we conclude that the polyQ tract and the regions of sequence immediately flanking only establish inter-molecular interactions after the motif ^{22}GAFQNLFQSVREVIQ^{36} does. This is clearly illustrated by signal intensities reported after two days of incubation in Figure 3B, which are substantially decreased in the region ^{22}GAFQNLFQSVREVIQ^{36} but remain essentially unchanged in the polyQ tract. We thus infer that the third step in the oligomerization of 25Q is a conformational change in the polyQ region occurring in on-pathway oligomers stabilized by inter-molecular interactions involving the region ^{22}GAFQNLFQSVREVIQ^{36}, especially the motif ^{30}SVREVIQ^{36}. The mechanism of self-assembly that we propose for 25Q is summarized in Figure 6.

A detailed analysis of the ^{22}GAFQNLFQSVREVIQ^{36} region of sequence, which experiences the earliest and largest decrease in signal intensities as well as the greatest chemical shift perturbations during oligomerization, indicates that it has some of the features expected in a sequence which forms coiled coils, such as the presence of hydrophobic residues in positions *a* and *d* of two consecutive imperfect heptad repeats starting at positions F24 and V31. This suggests that the on-pathway oligomers that form during the aggregation of 25Q might be stabilized by coiled coil interactions established by partially helical motifs, as proposed for huntingtin[51] on the basis of bioinformatics predictions and on the observation that the flanking region of the polyQ tract in huntingtin have helical character[62-64].

The results that we present in this work have implications for our understanding of the mechanism of oligomerization *in vivo* and for the identification of therapeutic strategies for SBMA. As previously discussed the motif which establishes the earliest inter-molecular interaction in 25Q, ^{22}GAFQNLFQ^{29}, binds to the LBD as an α-helix after the full length AR is activated by androgens[44,65]. If, as expected, this inter-domain interaction diminishes the propensity of the motif to oligomerize, our results provide a simple explanation for the relationship between cleavage by caspase-3 and AR oligomerization. In the mechanism that we propose caspase-3 would facilitate oligomerization by producing an N-terminal AR fragment devoid of the LBD and with an exposed hydrophobic and Q/N-rich sequence. Finally, our work clearly identifies the region ^{22}GAFQNLFQSVREVIQ^{36}, in particular the ^{22}GAFQNLFQ^{29} motif, as a potential therapeutic target for SBMA.

Materials and methods

Protein expression and purification

The 4Q and 25Q constructs were expressed and purified as previously described.[36]

Urea Incubation followed by Dialysis (UID, NMR Sample preparation*)*

The protein solutions stored at -80 °C were thawed and dialyzed for 16 hours at 4 °C against a buffer containing 20 mM sodium phosphate and 1 mM tris(2-carboxyethyl)phosphine (TCEP) at pH 7.4. Finally, 10% (v/v) D_2O and 0.015 mM DSS (4,4-dimethyl-4-silapentane-1-sulfonic acid) were added to the samples.

Stringent Disaggregation Procedure (SDP)

The protocol was adapted from the one developed by Linse group for Aß peptide kinetic measurements[34]. After purification, the lyophilized protein was dissolved in a buffer containing 20 mM sodium phosphate, 100 mM NaCl, 6 M Guanidine Thiocyanate, pH 7.4 and 5 mM TCEP until complete reduction of the protein, as monitored by HPLC measurement with a C18 column. The solution was passed through a PD-10 column (GE Healthcare) equilibrated in 20 mM sodium phosphate and 100 mM NaCl, pH 7.4. The most concentrated fraction was then purified with a Superdex 75 10/300 SEC column and the fractions containing exclusively the monomeric protein were centrifuged at 386,000 g for 1 hour at 4°C, using an Optima MAX preparative ultracentrifuge (Beckman). Only the upper ¾ of the centrifuged solutions were used for the assays.

Dynamic light scattering (DLS)

Before the measurements, samples were spinned with a tabletop centrifuge at 4ºC for 10 minutes at 13,000 rpm. Measurements were taken with a Malvern Zetasizer Nano S equipped with a He-Ne of 633 nm wavelength laser. For each experiment, 20 repetitions of 20 seconds were recorded. Three measurements were performed at each time point.

Analytical ultracentrifugation (AUC)

Sedimentation Velocity data were acquired with an Optima XL-A (Beckman) at 38,000 rpm. Sample displacement profiles were obtained by interferometry and UV absorbance at 280 nm. The buffer used contained 20 mM sodium phosphate and 100 mM NaCl. The experiments were acquired right after the samples were treated

with the ultracentrifugation step, as explained in *Stringent Disaggregation Procedure (SDP)* section. Experimental data were analyzed with SEDFIT 14.1[36]. The mass of the monomeric species was calculated with the following equation:

$$M=RTS/(D(1-v\rho_{solv}))$$

where M is the calculated mass; R=8.31 J/K·mol; T=293 K; ρ=1.000416 g/ml; v=0.7201 ml/g and D is the translational diffusion coefficient (expressed in cm^2/s), obtained by performing a parallel DLS experiment with a DynaPro instrument (ProteinSolutions).

Circular Dichroism (CD) spectroscopy

Far-UV CD measurements were performed on a JASCO 815 spectropolarimeter using a 0.1 cm length cuvette. Freshly prepared sample was diluted to the desired concentration with 20 mM sodium phosphate buffer (pH 7.4). The spectra were acquired at 0.2 nm resolution with a scan rate of 50 nm/min. For each sample, 10 spectra were collected and averaged, after subtraction the blank. For both proteins, CD measurements were performed at 4ºC and 37ºC. Spectra were recorded in the range 190-260 nm.

Time-evolution NMR experiments

A freshly prepared sample containing 50 μM ^{15}N-labeled SDP 25Q in 20 mM phosphate buffer (pH 7.4), 1 mM TCEP and 10% (v/v) D_2O was used to acquire a series of 2D ^{1}H-^{15}N HSQC spectra at 278 K on a 18.8 T Bruker Avance 800 MHz spectrometer equipped with a triple-resonance cryogenically-cooled probehead (TCI). DSS was used as internal reference. ^{1}H-^{15}N HSQC experiments were acquired with 1024 complex points in the F2 (^{1}H) dimension and 512 increments in the F1 (^{15}N) dimension, 16 scans for each increment and a spectral width of 8000 Hz (10.0 ppm) in F2 and 1800 Hz (22.0 ppm) in F1. The inter-scan delay was 1 s. ^{1}H-^{15}N HSQC spectra of the same sample were recorded right after preparation and after 2, 4 and

7 days. Between experiments, the sample was kept at 310 K. Spectra were processed with NMRPipe[66] and analyzed using CcpNMR Analysis[67].

Equilibrium NMR experiments

The ^{1}H-detected and ^{13}C-detected NMR experiments for sequence-specific resonance assignment were acquired at 278 K on a 16.4 T Bruker Avance 700 spectrometer, equipped with a cryogenically cooled probehead optimized for ^{13}C-detection (TXO), and on a 22.3 T Bruker Avance III 950 spectrometer, equipped with a triple-resonance inverse cryogenically cooled probehead (TCI) as previsously reported.[36] The assignment deposited in the BMRB (entries 25606, 25607 and 25608) were used for chemical shift analysis and mapping.

Relaxation experiments on both 4Q and 25Q (the latter at three different protein concentrations: 125, 250 and 500 µM) were performed at 278 K on a 16.4 T Bruker Avance 700 spectrometer, equipped with a cryogenically cooled probehead (TXI), by measuring ^{15}N backbone longitudinal (R_1) and transverse (R_2) relaxation rates and the heteronuclear ^{15}N{^{15}H} NOEs as previously reported.[36]

NMR data processing and analysis

The secondary structure propensity from the heteronuclear chemical shifts was determined for 4Q and 25Q by using the neighbor corrected structural propensity calculator tool (http://nmr.chem.rug.nl/ncSPC/). Tamiola, Acar and Mulder random coil chemical shift library[40] was chosen for the analyses.

Prediction data

Theoretical helical propensities were calculated using AGADIR[68,69] (http://agadir.crg.es/) online. The conditions used for the calculations were pH 7.4, 278 K and 0.1 M ionic strength. PONDR-FIT[70] was used to predict the intrinsically disordered regions (http://www.pondr.com/). The AABUF was calculated with an averaging window of 7 residues[43].

Acknowledgments

The authors would like to thank the NMR facility of the University of Barcelona (CCiT UB) for technical assistance and IRB, ICREA, Obra Social “la Caixa” (B.E.), AGAUR (G.C.), MICINN (CTQ2009-08850 to X.S.), MINECO (BIO2012-31043 to X.S.) and Marató de TV3 (102030 to X.S.). This work has been supported in part also by the European Commission Projects BioNMR (Contract No. 261863) and INSTRUCT (Contract No. 211252).

Figures

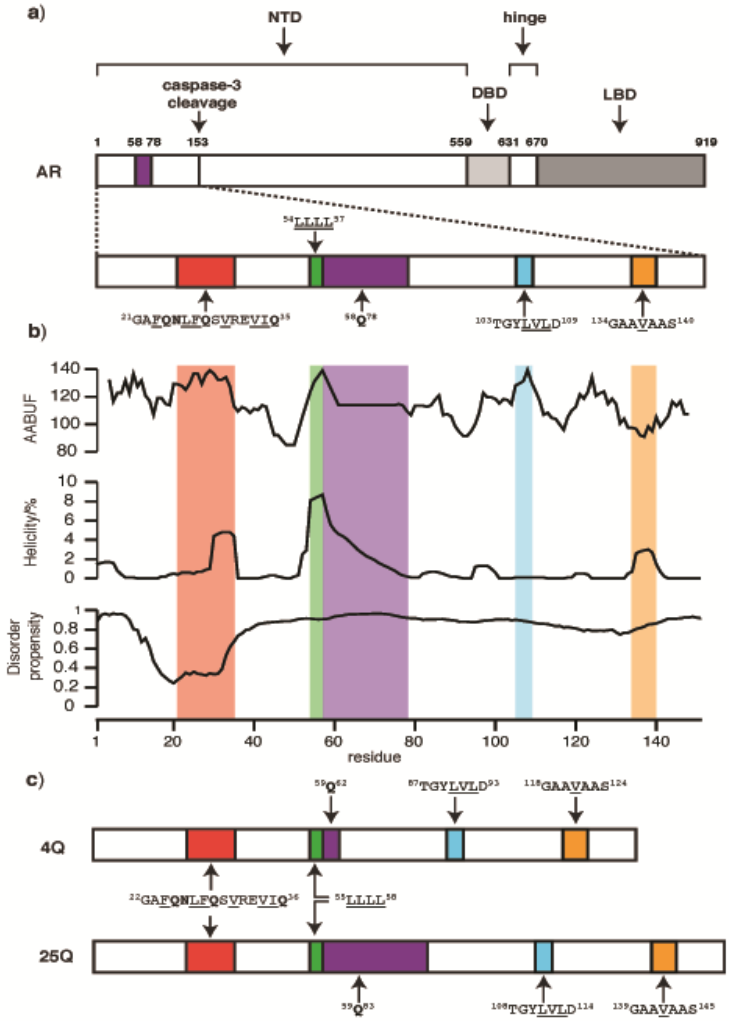


Figure 1. Domain organization of AR and predicted structural properties of the fragment produced by caspase-3. (a) Organization of AR domains with an indication of the positions of the polymorphic polyQ tract, shown in purple, of the caspase 3 cleavage site and, in the resulting fragment, of the ^{21}GAFQNLFQSVREVIQ^{35}, ^{54}LLLL^{57}, ^{103}TGYLVL^{108} and ^{134}GAAVAAS^{140} regions, respectively colored in red, green, blue and orange; in this and in the following panel residues are numbered according to the Uniprot entry for AR (www.uniprot.org/uniprot/P10275). (b) Average area buried upon folding (AABUF), predicted helicity according to Agadir and disorder propensity as calculated by PONDR-FIT of the fragment reported as a function of residue number. (c) 4Q and 25Q constructs used in this work, where the difference in the numbering with panels A and B, is due to the addition of a residue at the N-terminus after cleavage of the MBP by using the TEV protease. In the motifs

highlighted in the figure, hydrophobic residues are underlined and Q and N residues are shown in bold.

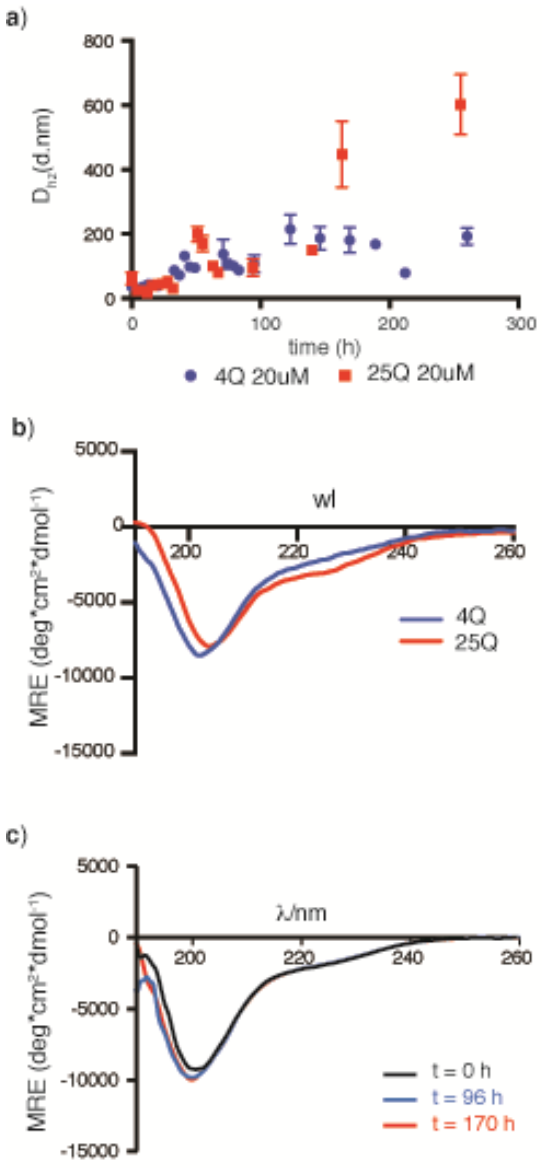


Figure 2. Characterization of the oligomerization of 4Q and 25Q. (a) Time resolved DLS analysis of the oligomerization rate of 20 µM SDP 4Q and 25Q solutions. (b) CD spectra of 130 µM SDP 4Q and 25Q solutions acquired immediately after sample preparation. (c) CD spectra of 20 µM SDP 25Q solution after 0, 96 and 170 hours of incubation at 37°C.

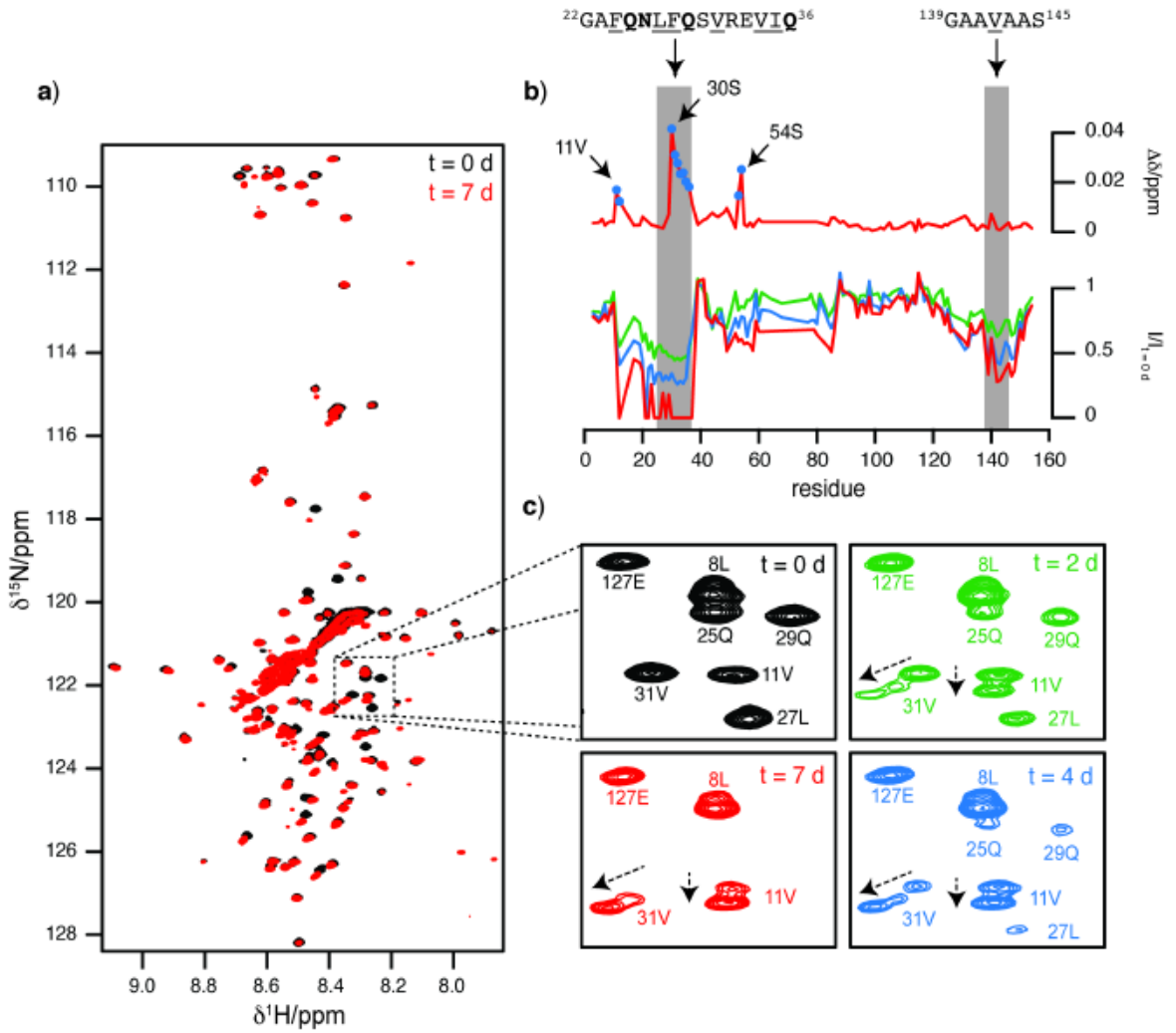


Figure 3. Analysis of the oligomerization of 25Q by NMR. (a) ^{1}H-^{15}N HSQC spectrum of a 50 µM SDP sample of ^{15}N-labeled 25Q immediately after preparation (black) and after 7 days of incubation at 37ºC (red). (b) Top: chemical shift perturbations (Δδ) observed after 7 days of incubation reported as a function of residue number. Δδ values shown as blue dots correspond to the difference in chemical shift between two signals in slow exchange originating from a cross-peak of monomeric 25Q that splits during oligomerization; bottom: decreases in signal intensity relatively to that after sample preparation ($I/I_{t=0d}$) reported as a function of residue number and as a function of time, after 2 (green), 4 (blue) and 7 (red) days of incubation. (c) Detail of the spectral changes observed in a specific region of the ^{1}H-^{15}N HSQC spectrum immediately after dissolution (black) and after incubation of the sample at 37ºC for 2 (green), 4 (blue) and 7 (red) days. In the motifs mentioned in the figure, hydrophobic residues are underlined and Q and N residues are shown in bold.

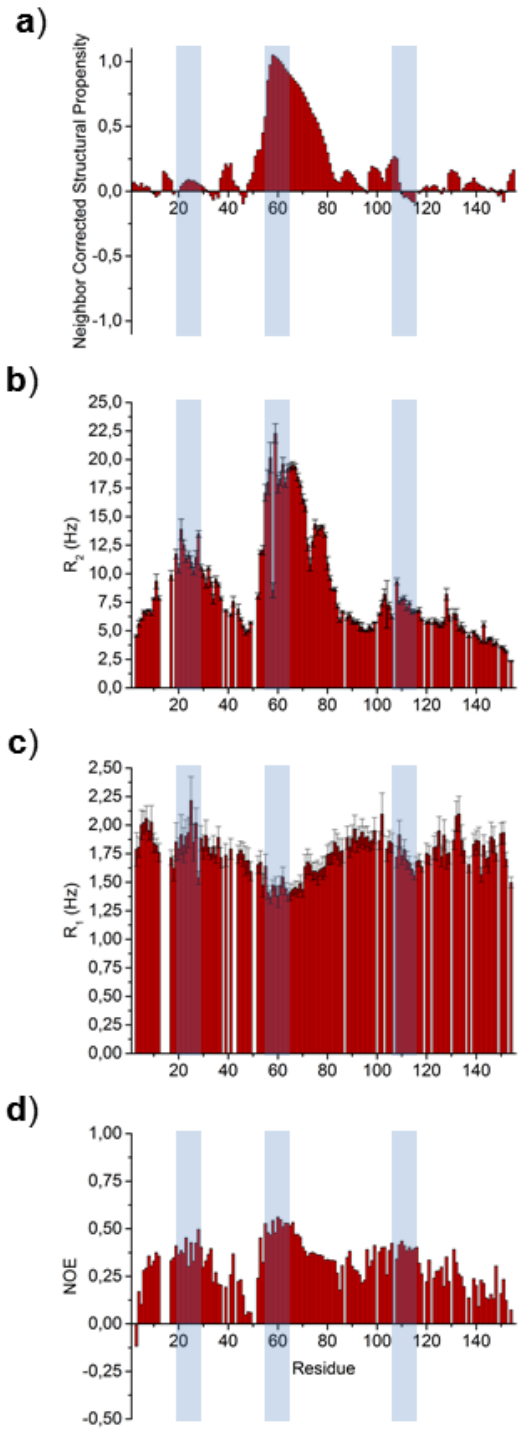


Figure 4. Structural and dynamical properties of 25Q. (a) Secondary structure propensity of 25Q based on an analysis of the C', C^{α}, C^{β} and N backbone chemical shifts carried out by using the data deposited in BMRB (entry 25608) (b) ^{15}N R_2 relaxation rates, (c) ^{15}N R_1 relaxation rates and (d) steady-state heteronuclear $^{15}N\{^{15}H\}$ NOEs reported as a function of residue number for 25Q at 500 μM protein concentration.

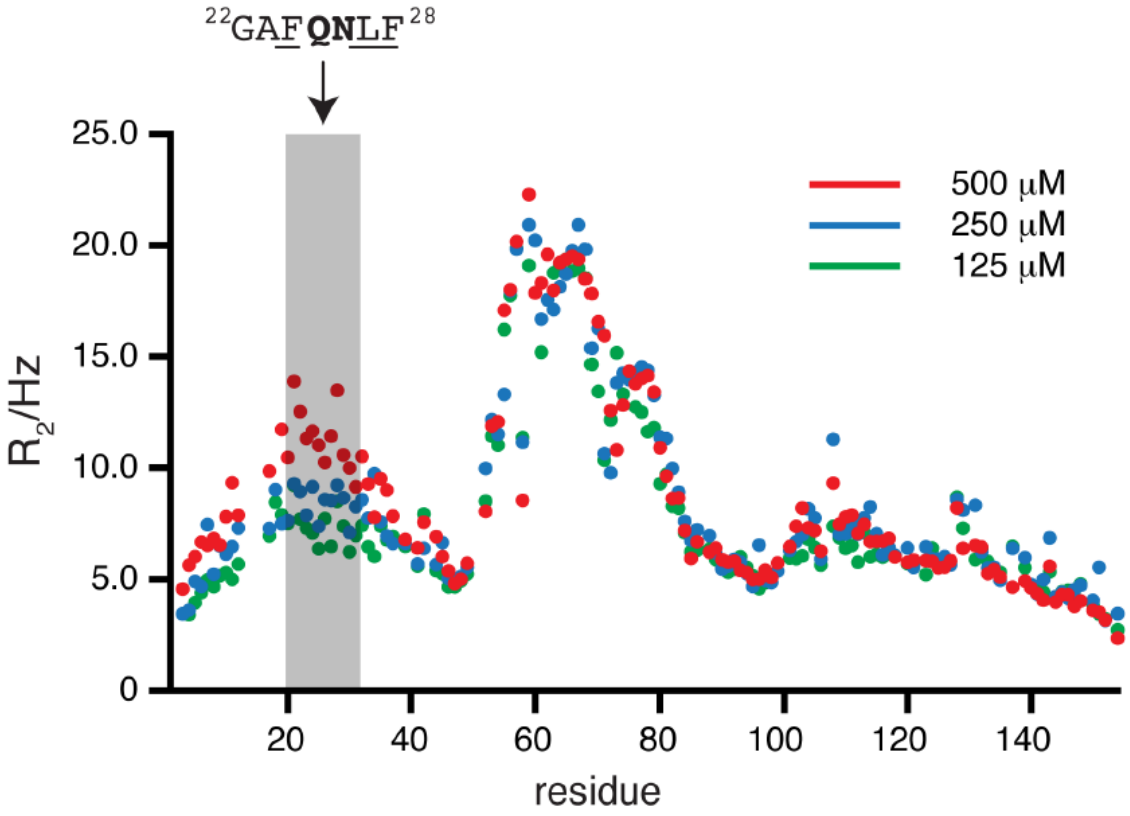


Figure 5. Concentration-dependent backbone dynamics in 25Q. ^{15}N R_2 relaxation rates for 25Q reported as a function of residue number at three different protein concentrations: 125, 250 and 500 μM. The error bars have been removed from the figure to facilitate the comparison.

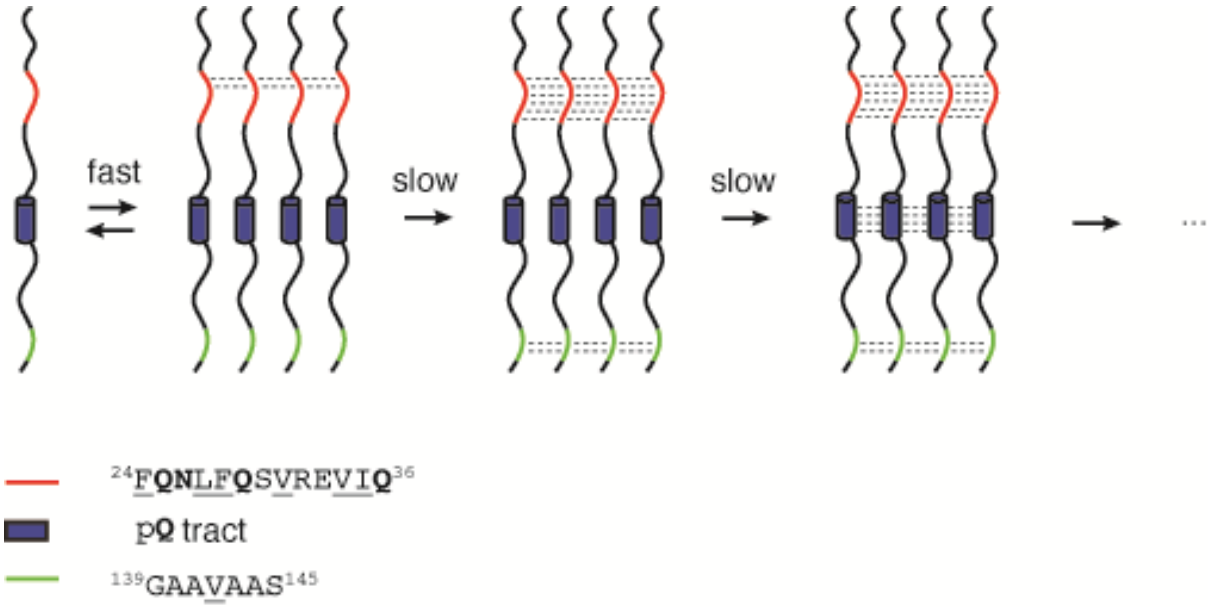


Figure 6. Mechanism of AR oligomerization in SBMA. Scheme of the mechanism of oligomerization proposed for the N-terminal AR fragment based on the results obtained in this work.

References

1. Kennedy WR, Alter M, S. J. Progressive proximal spinal and bulbar muscular atrophy of late onset. A sex-linked recessive trait. **18,** 671–80 (1968).
2. La Spada, A. R., Wilson, E. M., Lubahn, D. B., Harding, A. E. & Fischbeck, K. H. Androgen receptor gene mutations in X-linked spinal and bulbar muscular atrophy. *Nature* **352,** 77–79 (1991).
3. Katsuno, M. *et al.* Pathogenesis, animal models and therapeutics in Spinal and bulbar muscular atrophy (SBMA). *Exp. Neurol.* **200,** 8–18 (2006).
4. Zoghbi, H. Y. & Orr, H. T. Glutamine repeats and neurodegeneration. *Annu. Rev. Neurosci.* **23,** 217–247 (2000).
5. Mirkin, S. M. Expandable DNA repeats and human disease. *Nature* **447,** 932–40 (2007).
6. Wells, R. D., Dere, R., Hebert, M. L., Napierala, M. & Son, L. S. Advances in mechanisms of genetic instability related to hereditary neurological diseases. *Nucleic Acids Res.* **33,** 3785–98 (2005).
7. Tompa, P. Intrinsically unstructured proteins evolve by repeat expansion. *BioEssays* **25,** 847–855 (2003).
8. De Simone, A. *et al.* Intrinsic disorder modulates protein self-assembly and aggregation. *Proc. Natl. Acad. Sci.* **109,** 6951–6956 (2012).
9. Adachi, H. *et al.* Widespread nuclear and cytoplasmic accumulation of mutant androgen receptor in SBMA patients. *Brain* **128,** 659–70 (2005).
10. Li, M. *et al.* Nuclear inclusions of the androgen receptor protein in spinal and bulbar muscular atrophy. *Ann. Neurol.* **44,** 249–54 (1998).
11. Thomas, P. S. *et al.* Loss of endogenous androgen receptor protein accelerates motor neuron degeneration and accentuates androgen insensitivity in a mouse model of X-linked spinal and bulbar muscular atrophy. *Hum. Mol. Genet.* **15,** 2225–2238 (2006).
12. Merry, D. E., Kobayashi, Y., Bailey, C. K., Taye, a a & Fischbeck, K. H. Cleavage, aggregation and toxicity of the expanded androgen receptor in spinal and bulbar muscular atrophy. *Hum. Mol. Genet.* **7,** 693–701 (1998).
13. Rhodes, L. E. *et al.* Clinical features of spinal and bulbar muscular atrophy. *Brain* **132,** 3242–3251 (2009).
14. Trifiro, M. A., Kazemi-Esfarjani, P. & Pinsky, L. X-linked muscular atrophy and the androgen receptor. *Trends Endocrinol. Metab.* **5,** 416–21 (1994).
15. Merry, D. E., Kobayashi, Y., Bailey, C. K., Taye, a a & Fischbeck, K. H. Cleavage, aggregation and toxicity of the expanded androgen receptor in spinal and bulbar muscular atrophy. *Hum. Mol. Genet.* **7,** 693–701 (1998).
16. Fischbeck, K. H., Lieberman, a, Bailey, C. K., Abel, a & Merry, D. E. Androgen receptor mutation in Kennedy's disease. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **354,** 1075–8 (1999).
17. Cortes, C. J. *et al.* Muscle expression of mutant androgen receptor accounts for systemic and motor neuron disease phenotypes in spinal and bulbar muscular atrophy. *Neuron* **82,** 295–307 (2014).
18. Lieberman, A. P. *et al.* Peripheral androgen receptor gene suppression rescues disease in mouse models of spinal and bulbar muscular atrophy. *Cell Rep.* **7,** 774–84 (2014).

19. McEwan, I. J., Lavery, D., Fischer, K. & Watt, K. Natural disordered sequences in the amino terminal domain of nuclear receptors: lessons from the androgen and glucocorticoid receptors. *Nucl. Recept. Signal.* **5,** e001 (2007).
20. Wang, Q. *et al.* A Hierarchical Network of Transcription Factors Governs Androgen Receptor-Dependent Prostate Cancer Growth. *Mol. Cell* **27,** 380–392 (2007).
21. Morley, J. F., Brignull, H. R., Weyers, J. J. & Morimoto, R. I. The threshold for polyglutamine-expansion protein aggregation and cellular toxicity is dynamic and influenced by aging in Caenorhabditis elegans. *Proc. Natl. Acad. Sci. USA.* **99,** 10417–10422 (2002).
22. Buchanan, G. *et al.* Structural and functional consequences of glutamine tract variation in the androgen receptor. *Hum. Mol. Genet.* **13,** 1677–1692 (2004).
23. Kumar, R. *et al.* Role of the androgen receptor CAG repeat polymorphism in prostate cancer, and spinal and bulbar muscular atrophy. *Life Sci.* **88,** 565–571 (2011).
24. Ryan, C. P. & Crespi, B. J. Androgen receptor polyglutamine repeat number: Models of selection and disease susceptibility. *Evol. Appl.* **6,** 180–196 (2013).
25. Nedelsky, N. B. *et al.* Native functions of the androgen receptor are essential to pathogenesis in a Drosophila model of spinobulbar muscular atrophy. *Neuron* **67,** 936–52 (2010).
26. Monks, D. A. *et al.* Overexpression of wild-type androgen receptor in muscle recapitulates polyglutamine disease. *Proc. Natl. Acad. Sci. USA.* **104,** 18259–18264 (2007).
27. Katsuno, M. *et al.* Leuprorelin rescues polyglutamine-dependent phenotypes in a transgenic mouse model of spinal and bulbar muscular atrophy. *Nat. Med.* **9,** 768–73 (2003).
28. Renier, K. J. *et al.* Antiandrogen flutamide protects male mice from androgen-dependent toxicity in three models of spinal bulbar muscular atrophy. *Endocrinology* **155,** 2624–2634 (2014).
29. Rhodes, L. E. *et al.* Clinical features of spinal and bulbar muscular atrophy. *J. Clin. Neurol.* **25,** 285–287 (2009).
30. Kobayashi, Y. *et al.* Caspase-3 cleaves the expanded androgen receptor protein of spinal and bulbar muscular atrophy in a polyglutamine repeat length-dependent manner. *Biochem. Biophys. Res. Commun.* **252,** 145–150 (1998).
31. Ellerby, L. M. *et al.* Kennedy's disease: Caspase cleavage of the androgen receptor is a crucial event in cytotoxicity. *J. Neurochem.* **72,** 185–195 (1999).
32. Felli, I. C. & Pierattelli, R. Novel methods based on 13C detection to study intrinsically disordered proteins. *J. Magn. Reson.* **241,** 115–125 (2014).
33. Cohen, S. I. a *et al.* Proliferation of amyloid-β42 aggregates occurs through a secondary nucleation mechanism. *Proc. Natl. Acad. Sci. USA.* **110,** 9758–63 (2013).
34. Hellstrand, E., Boland, B., Walsh, D. M. & Linse, S. Amyloid β-protein aggregation produces highly reproducible kinetic data and occurs by a two-phase process. *ACS Chem. Neurosci.* **1,** 13–18 (2010).
35. O'Nuallain, B. *et al.* Kinetics and Thermodynamics of Amyloid Assembly Using a High-Performance Liquid Chromatography-Based Sedimentation Assay. *Methods Enzymol.* **413,** 34–74 (2006).

36. Eftekharzadeh, B. *et al*. Sequence Context Influences the Structure and Aggregation Propensity of a PolyQ Tract, *submitted.*
37. Schuck, P. Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and lamm equation modeling. *Biophys. J.* **78,** 1606–1619 (2000).
38. Jayaraman, M. *et al.* Slow amyloid nucleation via α-helix-rich oligomeric intermediates in short polyglutamine-containing huntingtin fragments. *J. Mol. Biol.* **415,** 881–899 (2012).
39. Fernández, C. O. *et al.* NMR of alpha-synuclein-polyamine complexes elucidates the mechanism and kinetics of induced aggregation. *EMBO J.* **23,** 2039–2046 (2004).
40. Tamiola, K., Acar, B. & Mulder, F. A. A. Sequence-specific random coil chemical shifts of intrinsically disordered proteins. *J. Am. Chem. Soc.* **132,** 18000–18003 (2010).
41. Tamiola, K. & Mulder, F. A. A. Using NMR chemical shifts to calculate the propensity for structural order and disorder in proteins. *Biochem. Soc. Trans.* **40,** 1014–20 (2012).
42. Xue, B., Dunbrack, R. L., Williams, R. W., Dunker, A. K. & Uversky, V. N. PONDR-FIT: A meta-predictor of intrinsically disordered amino acids. *Biochim. Biophys. Acta - Proteins Proteomics* **1804,** 996–1010 (2010).
43. Rose, G. D., Geselowitz, A. R., Lesser, G. J., Lee, R. H. & Zehfus, M. H. Hydrophobicity of amino acid residues in globular proteins. *Science.* **229,** 834–838 (1985).
44. He, B., Kemppainen, J. a & Wilson, E. M. FXXLF and WXXLF sequences mediate the NH2-terminal interaction with the ligand binding domain of the androgen receptor. *J. Biol. Chem.* **275,** 22986–94 (2000).
45. Van De Wijngaart, D. J. *et al.* Novel FXXFF and FXXMF motifs in androgen receptor cofactors mediate high affinity and specific interactions with the ligand-binding domain. *J. Biol. Chem.* **281,** 19407–19416 (2006).
46. Fawzi, N. L., Ying, J., Torchia, D. A. & Clore, G. M. Kinetics of amyloid-monomer-to-oligomer exchange by NMR relaxation. J. Am. Chem. Soc. 132, 9948–9951 (2010).
47. Mittag, T. & Forman-Kay, J. D. Atomic-level characterization of disordered protein ensembles. *Curr. Opin. Struct. Biol.* **17,** 3–14 (2007).
48. Konrat, R. NMR contributions to structural dynamics studies of intrinsically disordered proteins. *J. Magn. Reson.* **241,** 74–85 (2014).
49. Nováček, J., Žídek, L. & Sklenář, V. Toward optimal-resolution NMR of intrinsically disordered proteins. *J. Magn. Reson.* **241,** 41–52 (2014).
50. Masino, L., Kelly, G., Leonard, K., Trottier, Y. & Pastore, A. Solution structure of polyglutamine tracts in GST-polyglutamine fusion proteins. *FEBS Lett.* **513,** 267–72 (2002).
51. Sivanandam, V. N. *et al.* The aggregation-enhancing huntingtin N-terminus is helical in amyloid fibrils. *J. Am. Chem. Soc.* **133,** 4558–4566 (2011).
52. Masino, L. *et al.* Domain architecture of the polyglutamine protein ataxin-3: a globular domain followed by a flexible tail. *FEBS Lett.* **549,** 21–25 (2003).

53. Masino, L. & Pastore, A. Glutamine repeats: structural hypotheses and neurodegeneration. *Biochem. Soc. Trans.* **30,** 548–551 (2002).

55. Wetzel, R. Physical chemistry of polyglutamine: Intriguing tales of a monotonous sequence. *J. Mol. Biol.* **421,** 466–490 (2012).

59. Lakhani, V. V., Ding, F. & Dokholyan, N. V. Polyglutamine induced misfolding of Huntingtin Exon1 is modulated by the flanking sequences. *PLoS Comput. Biol.* **6,** (2010).

60. Kar, K., Jayaraman, M., Sahoo, B., Kodali, R. & Wetzel, R. Critical nucleus size for disease-related polyglutamine aggregation is repeat-length dependent. *Nat. Struct. Mol. Biol.* **18,** 328–336 (2011).

61. Jayaraman, M. *et al.* Kinetically competing huntingtin aggregation pathways control amyloid polymorphism and properties. *Biochemistry* **51,** 2706–2716 (2012).

62. Fiumara, F., Fioriti, L., Kandel, E. R. & Hendrickson, W. A. Essential role of coiled coils for aggregation and activity of Q/N-rich prions and PolyQ proteins. *Cell* **143,** 1121–1135 (2010).

63. Schaefer, M. H., Wanker, E. E. & Andrade-navarro, M. A. Evolution and function of CAG / polyglutamine repeats in protein – protein interaction networks. 4273–4287 (2012).

64. Spyros Petrakis, Martin H. Schaefer, E. E. W. and & Andrade-Navarro, M. A. Aggregation of polyQ-extended proteins is promoted by interaction with their natural coiled-coil partners. *Bioessays* **35,** 503–507 (2013).

65. He, B., Minges, J. T., Lee, L. W. & Wilson, E. M. The FXXLF motif mediates androgen receptor-specific interactions with coregulators. *J. Biol. Chem.* **277,** 10226–10235 (2002).

66. Delaglio, F. *et al.* NMRPipe: A multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* **6,** 277–293 (1995).

67. Vranken, W. F. *et al.* The CCPN data model for NMR spectroscopy: Development of a software pipeline. *Proteins Struct. Funct. Genet.* **59,** 687–696 (2005).

68. Muñoz, V. & Serrano, L. Elucidating the folding problem of helical peptides using empirical parameters. III. Temperature and pH dependence. J. Mol. Biol. 245, 297–308 (1995).

69. Lacroix, E., Viguera, A. R. & Serrano, L. 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 (1998).

70. Li, X., Romero, P., Rani, M., Dunker, A. & Obradovic, Z. Predicting Protein Disorder for N-, C-, and Internal Regions. Genome Inform. Ser. Workshop Genome Inform. 10, 30–40 (1999).

Early stages of androgen receptor oligomerization in spinal bulbar muscular atrophy

Bahareh Eftekharzadeh[a,1], Alessandro Piai[b,1], Giulio Chiesa[a], Jesús García[c], Roberta Pierattelli[b], Isabella Felli[b], and Xavier Salvatella[a,d,2]

[a]BSC-CRG-IRB Joint Research Programme in Computational Biology, IRB Barcelona, Baldiri Reixac 10, 08028 Barcelona, Spain.

[b]CERM and Department of Chemistry “Ugo Schiff”, University of Florence

Via Luigi Sacconi 6, 50019 Sesto Fiorentino (Florence), Italy

[c]IRB Barcelona, Baldiri Reixac 10, 08028 Barcelona, Spain.

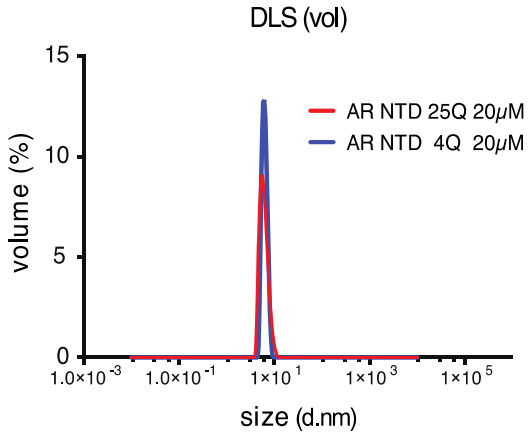
[d]ICREA, Barcelona, Spain.

[1]these authors contributed equally

[2]to whom correspondence should be addressed

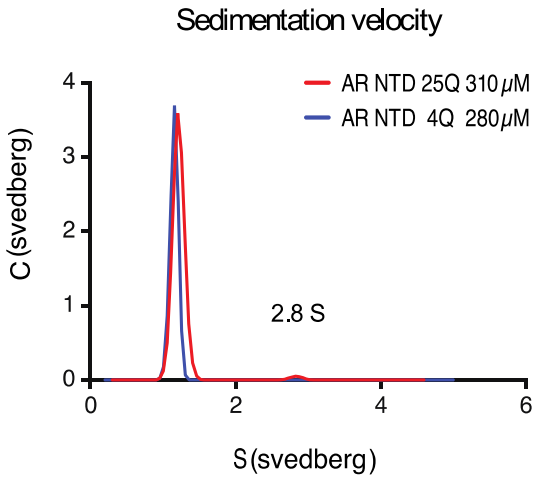
SUPPLEMENTARY MATERIAL

Supplementary Figure 1



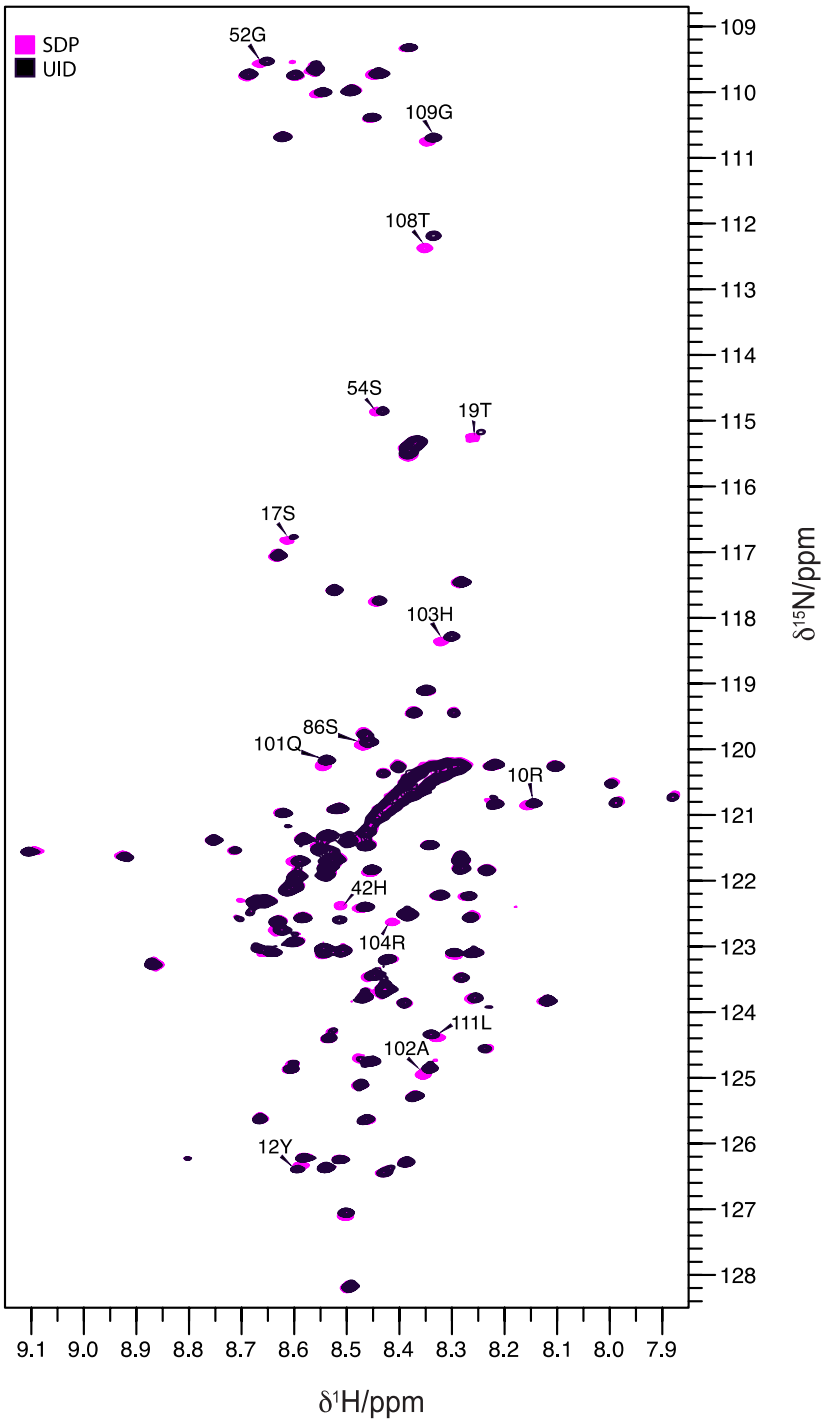
Dynamic light scattering size distribution by volume of 4Q and 25Q. 20 µM 4Q (blue) and 25Q (red) were measured at 37°C immediately after preparation. The mean size values are 5.6 nm and 6.1 nm respectively.

Supplementary Figure 2



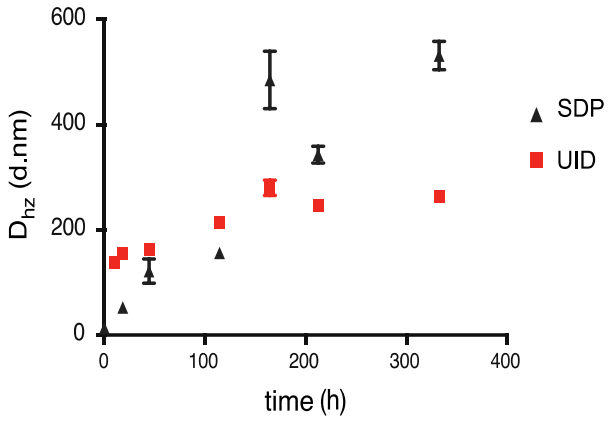
Sedimentation velocity analytical ultracentrifugation of 4Q and 25Q. 310 µM 4Q (blue) and 280 µM 25Q (red) stock solutions were measured immediately after preparation. Sedimentation coefficients representing the monomeric protein are 1.1 S for 4Q and 1.3 S for 25Q. For both constructs, an additional small but detectable population with sedimentation coefficient of 2.6 S (4Q) and 2.8 S (25Q) is observed.

Supplementary Figure 3



Comparison of the ^{1}H-^{15}N HSQC spectra of UID (pink) and SPD (black) 25Q. Representative examples of cross-peaks with lower intensity in the ^{1}H-^{15}N HSQC spectrum of the UID 25Q sample compared to that of the SDP one are indicated.

Supplementary Figure 4



Time evolution monitored by DLS. 20 μM 25Q disaggregated with the stringent disaggregation protocol (SDP, black) and with urea incubation followed by dialysis (UID, red) are compared. While the SDP sample rapidly evolves into larger assemblies, the UID one, even if more oligomeric, does not evolve significantly from the initial time t_0.

Supplementary Figure 5

The relative signal intensity of the resonances of UID 25Q from 2D ^{1}H-^{15}N HSQC spectra at various protein concentrations. The relative signal intensity of the resonances belonging to the motif ^{22}GAFQNLFQ^{29} decreases with concentration.