# Development of NMR methods to study disordered proteins

Sampo Mäntylahti

Research Program in Structural Biology and Biophysics
Institute of Biotechnology
University of Helsinki

Division of Biotechnology
Department of Biosciences
Faculty of Biological and Environmental Sciences
University of Helsinki

And

Doctoral School in Enviornmental, Food and Biological Sciences

ACADEMIC DISSERTATION

To be presented, with the permission of the Faculty of Biological and Environmental Science of the University of Helsinki, for public examination in auditorium 3, B-building, Latokartanonkaari 7, Helsinki
on June 2014 13$^{th}$, at 12 noon.

Helsinki 2014

**supervisor**
Docent Perttu Permi
Institute of Biotechnology
University of Helsinki

**Thesis advisory committee**
Professor Arto Annila
Faculty of Science, Department of Physics
University of Helsinki

Professor Juha Voipio
Faculty of Biological and Environmental Sciences, Department of Biosciences
University of Helsinki

**Reviewers**
Professor Jukka Jokisaari
Faculty of Sciences, Department of Physics
University of Oulu

Senior University Lecturer Sampo Mattila
Faculty of Sciences, Department of Chemistry
University of Oulu

**Opponent**
Professor Wiktor Koźmiński
Faculty of Chemistry
University of Warsaw, Poland

**Custos**
Professor Kari Keinänen
Faculty of Biological and Environmental Sciences, Department of Bioscience
University of Helsinki, Finland

ISBN: 978-952-10-9951-9 (Paperback)
ISBN: 978-952-10-9952-6 (PDF, http://ethesis.helsinki.fi/)
ISSN: 1799-7372
Hansaprint Oy
Helsinki 2014

# Contents

## List of Original Publications

I Mäntylahti S, Tossavainen H, Hellman M, Permi P.
An intraresidual i(HCA)CO(CA)NH experiment for the assignment of main-chain resonances in 15N, 13C labeled proteins.
Journal of Biomolecular NMR 2009, 45(3): 301-10

II Mäntylahti S, Aitio O, Hellman M, Permi P.
HA-detected experiments for the backbone assignment of intrinsically disordered proteins.
Journal of Biomolecular NMR 2010 47(3):171-81

III Mäntylahti S, Hellman M, Permi P.
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.
Journal of Biomolecular NMR 49(2):99-109

IV Mäntylahti S, Koskela O, Jiang P, Permi P.
MQ-HNCO-TROSY for the measurement of scalar and residual dipolar couplings in larger proteins: application to a 557-residue IgFLNa16-21.
Journal of Biomolecular NMR 2010 47(3):183-94

**Abstract**

In the field of bioscience there is an ongoing explosive growth in discovery and information. Novel means in biotechnology as well as in medicines are introduced at an unseen rate. One of the aspects contributing to this development is the increased understanding of protein function and structure. Proteins have a role in almost every biological process. The function and structure of proteins are linked. Recent studies have discovered that the understanding of the protein structure has been biased. Namely, the studies have unearthed a previously dismissed protein structure state: intrinsically disordered proteins (IDPs). In this highly dynamic state a protein is without a globular fold, but does not meet the requirements of a random coil either. Rapid transition between folds renders most of the established research techniques to be poor methods to study the IDPs.

Nuclear magnetic resonance (NMR) is a spectroscopy method, which enables the study of molecules at atomic resolution. The technique is based upon manipulation of the nuclear spins in specifically produced sample under strong magnetic field. In this method, spins of the system generate quantum coherence state(s), which is utilized to obtain information about the system. NMR is suitable for studying samples in solid and liquid mediums, but in case of biomolecules, water solution is preferable as it resembles *in vivo* environment. Highly mobile structure and chemical composition of IDPs cause many established NMR experiments to fail. Development of NMR pulse sequences is an obvious approach to solve the problem.

This thesis presents a number of NMR pulse sequences, which are designed to improve acquisition of information from highly mobile sections of proteins. The key aspect is to utilize $H^{\alpha}$ atom instead of $H^{N}$ in coherence transfer. Additional improvements include limited residue-specific identification and novel coherence transfer pathways.

Articles I, II, and III present triple resonance experiments, which correlate protein backbone atoms. Combination of the spectra enables full sequential assignment. Article IV introduces an improved pulse sequence for measuring J couplings between nitrogen and amide proton.

The experiments were subjected to experimental verification. Comparisons were drawn between established pulse sequences. In both globular proteins and IDPs the results show improvement over established pulse sequences. The proposed sequences yielded improved assignment coverage, resolution and sensitivity enhancement.

Keywords: NMR, pulse sequence, intrinsically disordered proteins, backbone assignment

# List of Abbreviations

| | |
|---|---|
| AA | Amino acid |
| AP | Anti-phase |
| CD | Circular dichroism |
| COS-CT | Coherence-order selective coherence transfer |
| COSY | Correlation spectroscopy |
| CSA | Chemical Shift Anisotropy |
| CT | Constant time |
| DIPSI | Decoupling in presence of scalar interaction |
| DNA | Deoxyribonucleic acid |
| FID | Free induction decay |
| FT | Fourier transformation |
| FWHM | Full width at half maximum |
| HSQC | Heteronuclear correlation spectroscopy |
| INEPT | Insensitive nuclei enhanced by polarization transfer |
| IDP | Intrinsically disordered protein |
| IDS | Intrinsically disordered segments (of protein) |
| IP | InPhase |
| IUP | Intrinsically unfolded protein |
| MD | Molecular dynamics |
| MRI | Magnetic resonance imaging |
| NMR | Nuclear magnetic resonance spectroscopy |
| NUS | Non-uniform sampling |
| PDB | Protein data bank |
| Rf | Radio frequency |
| RNA | Ribonucleic acid |
| ssNMR | Solid state NMR |
| TOCSY | Total correlation spectroscopy |

# 1 Introduction

Life at cellular level, as we know it, manifests itself in operation of molecular machines. These are wonderfully complex and sometimes bizarre machines, but machines altogether. Most of the biological machinery comprises of polymerized amino acid residues, namely, proteins. The interest in proteins is not a novel one, as their role is eminent in cellular machinery and structures. Cells, cell organs and tissues consists of carbohydrates and fatty acids in addition to peptide chains. Besides being the building blocks of life, proteins fulfill other functions as well: they enable, regulate and catalyze chemistry. In a word, proteins are essential. Therefore, comprehending protein function and dynamics is a key to understanding ourselves.

Structural biology is the study of structure and function of biomolecules. Structures of nucleic acids, carbohydrates, and polypeptides are coupled with their functions. Three dimensional (3D) structure of biological macromolecule is called 'a fold' and process leading into this functional state is known as folding. A function manifests itself when an interaction causes a change in the molecule. The change can be virtually any kind of change, but the crucial factor is that molecule departs from its current state. It appears that fold is crucial for the molecular function. Even small deviations in the structure may change, impair or even destroy the function of the molecule. Diseases are failures in our chemistry, stemming from either external or internal sources. For example, recent studies have linked Parkinson's (Inamdar *et al.*, 2012) and Alzheimer's diseases (Caleo *et al.*, 2012) to damaged protein structures. In order to alter the detrimental conditions and reasons leading into them, understanding of the protein structure as well as the dynamics involved in the interaction pathway is required.

This thesis is based on four articles. Each article presents novel experimental methods to study loose and flexible proteins structures. The research is done in the field of NMR spectroscopy, by improving pulse sequences. The pulse sequences devised in the Articles focus on backbone assignment (Articles I, II, and III) as well as structure refinement (Article IV). The articles form a comprehensive collection, which enables a protocol for identifying NMR signals to their sources in the chemical composition. The overall thesis presents a brief view on NMR's theory, an introduction to the concept of intrinsically disordered proteins, and how they affect NMR experiments. Remaining of the thesis is focused on the experimental development of NMR to overcome the challenges IDPs present.

Structural biology is an multidisciplinary field of science. When dealing with multiple fields of science, terminology sometimes overlap, as is the case with "labelling". In protein production, it refers the process in which protein structure is enriched with $^{13}C$ and $^{15}N$ isotopes. In pulse sequence development, labellig refers to an incremental modulation of particular coherence state.

# 2 Proteins

The textbook model of protein production is a simple one: primary structure, i.e., the chemical structure, is dictated by the DNA. In protein expression, amino acids are polymerized into a linear chain, which folds into a 3D structure. Epigenetic factors also affect the folding process. Chaperone proteins or peptides (Haataja *et al.*, 2008; Wallace, 2010) may influence the folding pathway. As the outcome of folding process, the protein collapses into a free energy minimum generating internal bonds, which contain disulphide bridges and hydrogen bonds, for example. The polypeptide forms sheets and helical structures that bind the overall structure (Nelson and Cox, 2005). Quaternary structure, in turn, defines connections between proteins and domeins. Predicting protein folding from primary structure alone remains an elusive goal, that is feasible only for simple molecules (Mach and Koehl, 2013).

Protein structure is not rigid, but incorporates elements, which allow changes in the conformation. Typically, change in the fold takes place when the protein interacts with another protein or peptide. Classical ”key-lock”-paradigm (Fischer, 1890) does not support this view, but a more recent induced fit theory does (Koshland, 1958). Conformation change may occur without obvious binding counterpart. Thermal noise is a typical reason for this.

A small protein usually constitute a single fold. Bigger proteins usually contain more than one self-assembling sections known as domein. The modular proteins have functionally or structurally independent regions, while globular proteins have stable, folded regions with secondary and tertiary structures. Thus, modular proteins consist of globular sections that are usually linked by flexible segments. These unstructured linkers are still a part of the polypeptide chain, although they often resemble random coils. Linkers typically exhibit a high degree of freedom and motion. They are intrinsically disordered segments (IDS) of proteins.

## 2.1 Intrinsically disordered proteins

Until recently, the random coil and globular structures were recognized as the only states of proteins. However, biology proves to be more subtle than simple. Refolding studies have identified proteins to exist in states between unfolded and folded: which are so called molten globule states (Dunker *et al.*, 2002). In the molten globule state, protein side chains are no longer interacting in a stabilizing fashion, and hence the tertiary structure is lost (Dyson and Wright, 1998). The state may be induced by introducing slightly denaturing conditions (Yao *et al.*, 1997; Donne *et al.*, 1997; Eliezer *et al.*, 1997). Molten globule must not be confused with a reaction transition state. The molten globule state model is more akin to fluctuation between structure and random coil. Due to the fluctuation, molten globule is not a transition state between the two, but it is a state of its own. The lifetime of the state is significant and transition does not have a clear direction. Not only segments of proteins may have this disordered property, but whole proteins may exhibit this state (Uversky, 2002) and are thus dubbed intrinsically disordered proteins. By the same logic, sections exhibiting disordered properties are dubbed intrinsically disordered segments (IDSs).

Initially IDPs were thought to be a biological burden, an evolutionary dead end (Nelson and Cox, 2005). Their importance was underestimated as they do not form a stable 3D structure and thus they were thought to be unfit for performing the classic ”key-lock” -paradigm of protein function (Fischer, 1890). It has been shown that IDPs fulfill various roles, some of which are vital for eukaryotic cell functions. For example, IDPs act as molecular recognition domains, protein folding inhibitors, gene expression regulators, flexible linkers, entropic springs, entropic clocks, signal processing regulators, and entropic bristles (Dunker *et al.*, 2001). Indeed, IDPs/IDSs are abundant as 10% of all proteins can be classified as IDPs and 30% of eukaryotic proteins embodying IDS with a length of $\geq$ 50 consecutive amino acid residues (Dunker *et al.*, 2001).

Terminology concerning IDPs is not yet completely established. Therefore, different names, like natively unfolded proteins (Minde *et al.*, 2011), intrinsically unfolded proteins (Lee *et al.*, 2012),

intrinsically unstructured proteins (Das and Mukhopadhyay, 2011) are used. However, these all describe equivalent semi structureless state.

### 2.1.1 Biophysical aspects of IDPs

The distinction between random coil and IDPs is poorly defined. However, necessary aspects of IDPs can be described. Most prominent is the instability of the fold. Electrostatic repulsion between non-compensated charges render the hydrophobic attraction inadequate to maintain a fold (Uversky *et al.*, 1999). The concept of a fold is ill suited for short peptides as they often lack stabilizing elements. Therefore, in order for a segment to be considered an IDS, the length of the section must be sufficient. Additionally, the segment must express notable flexibility and mobility between transition states. The characteristic factor of IDPs and IDSs appears to be that they retain functionality (Uversky, 2009b).

Protein conformation is not rigid. Transition between conformations occurs when binding or interaction with ligand takes place. A binding pathway describes the trajectory of each atom in space as a function of time. Accordingly, a dissociation pathway may, or may not, be the reverse of the binding pathway. At temperatures above 0 K, molecules vibrate and tumble. This causes structures to be under constant stress (Brehm and Mullin, 1989). Proteins, which have multiple conformations available to them, transit between conformations and a lifetime of a particular conformation is significantly longer than a transition time. In case of IDSs, energy landscape for local conformations contains numerous energy minimums with flat potential barriers. So much so, that the concept of single dominant fold becomes meaningless as the lifetime difference between transition states and end state is indivisible. Therefore, a conformational ensemble is a more accurate description of the system. IDPs are not to be confused with misfolded or aggregated proteins. While IDPs may appear to be non-functional most of the time, they are able to function in native conditions or when binding to a partner (Aitio *et al.*, 2010). Function is the main defining attribute of IDPs (Dunker *et al.*, 1998).

Small peptides are typically free to adopt conformations and make transitions between states due to thermal motion and lack of internal restrains. These are not considered IDPs or IDSs because the length of a polypeptide must be sufficient to form meaningful structure. Literature sets the generally accepted length of IDP to be at least 30 residues (Fink, 2005; Das and Mukhopadhyay, 2011).

### 2.1.2 Investigation of IDPs

The interest to study IDPs is not novel (Sigler, 1988). Motivation to study flexible systems stems partly from proteins that are associated with virus entry (Roux and Taylor, 2007). Thus, strides in the field have been undertaken, but details about interactions remain scant (Folkers *et al.*, 1989; Fink, 1998; Dunker *et al.*, 2005). The reason for the poor results stems from the investigation techniques themselves. They are not suitable for studying highly mobile systems. Detection methods are typically based upon observing a signal stemming from multiples of symmetrical, stable structures. All of which IDPs exhibit poorly (Ramìrez-Alvarado *et al.*, 1997).

Due to the challenges presented by IDPs, the methodologies applied to the problem are diverse. X-ray crystallography has been employed with limited success in case of IDSs: diffraction maps exhibit poor resolution from flexible segments. Theoretically, the crystallization may force a segment into local energy minima, which is not necessarily the functional form *in vivo*. Obviously, the protein crystallography is hardly able to identify the function of the flexible regions. Therefore, the challenge is faced using CD spectroscopy (Kussie *et al.*, 1996), size exclusion chromatography (Mohan *et al.*, 2006), small angle X-ray scattering (Fink, 1998), dynamic light scattering (Ferron *et al.*, 2006), Fourier transform infrared spectroscopy (Natalello *et al.*, 2012), aggregation studies (Lindinga *et al.*, 2004), Raman Spectroscopy (Uversky and Longhi, 2010), as well as NMR spectroscopy (Dyson and Wright, 1991).

The conformation of IDPs flux seemingly, indicating small potential energy difference between individual conformation states. Destabilizing globular sections by mutating the primary structure reveals that the Gibbs free energy difference between IDS and ordered segment is small (Espinoza-Fonseca, 2009; Sarai, 2012). However, the difference is minuscule when compared to the energy required to denature globular protein structure (Makhatadze and Privalov, 1996). Relatively flat free energy landscape enables multiple pathways rendering most common investigation methods inoperable (Raoa and Karplus, 2010).

# 3 Nuclear magnetic resonance

Nuclear magnetic resonance (NMR) is a spectroscopic method based on manipulation of nuclear spins in a high magnetic field via radio frequency irradiation (Bloch *et al.*, 1946). A spin system will absorb the energy of the radiation, which leads to a labile excited state. The spins in the system precess around the external magnetic field. The motion is observable since it induces alternating voltage in receiver coils. Based on the observed signal, one is able to deduce dynamics and structure of the system. More sophisticated information is obtained by manipulating the spin system to form coupled coherences and measuring the formation rate incrementally.

The nuclear spin is an internal property of nucleon, which explains, but is not limited to, magnetic properties of nucleon. The spin determines magnetic properties of an nucleus and obeys the same mathematical identities as the impulse momentum. If a specimen is placed in a strong magnetic field, we are able to observe emission energy level splits. This is commonly known as the Zeeman effect (Zeeman, 1897). Spin-$\frac{1}{2}$ particles have two energy states. In addition to the states, the spins have a direction in space. This is known as the spin vector. Energy minimum for the spin vector is parallel or antiparallel to an external magnetic field, depending on the gyromagnetic ratio of the isotope in question. However, quantum mechanics requires that the spins revolve around the axis of the magnetic field (Abragam, 1961). The spinning magnetic dipole contributes to a dipole moment $\vec{\mu}$, which is proportional to the spin angular momentum $\vec{I}$ of the particle as described in Equation 1 (Cavanagh *et al.*, 2007; Goldman, 1988).

$$\vec{\mu} = \gamma \vec{I} \tag{1}$$

In Equation 1, $\gamma$ is the gyromagnetic ratio of an isotope in question. When placed in magnetic field, the dipole moment will align with the field. Spin has two states ($\pm\frac{1}{2}$) and the energy difference between levels is $\Delta E = \hbar\gamma B_0$. In experimental temperatures and macroscopic population of *N* spins, the net magnetization is Abragam (1961): $M = B_0 \frac{N\gamma^2\hbar^2 I(I+1)}{3kT}$, in which B is magnetic field, T is the temperature, k the Boltzmann constant and I is the spin angular momentum. M is interpreted to be the net magnetic moment. The population ratio between the spin states in a magnetic field can be shown to differ from each other. The magnitude of the difference in a sample is $\propto 10^{-12}$ at room temperature (Bloch, 1946).

## 3.1 Interactions

Spins interact with each other according to laws of quantum mechanics (Abragam, 1961; Eisberg and Resnick, 1985; Griffiths, 2005; Cavanagh *et al.*, 2007). Hamiltonian operators describe the evolution of a particular state. Larmor frequency ($\omega$) describes precession frequency of a spin in a magnetic field. Zeeman Hamiltonian (Cavanagh *et al.*, 2007; Goldman, 1988) in Equation 2 leads to the description of Larmor frequency in Equation 3:

$$\mathbf{H}_{Zeeman} = -\vec{\mu} \cdot \vec{B} = -\mu_z B_0 = -\gamma B_0 I_z \tag{2}$$

$$\omega_0 = -\gamma B_0 \tag{3}$$

### Chemical shift

Chemical shift describes the local magnetic field affecting a particular nucleus. External magnetic field is assumed to be homogenous, and thus the chemical shift consists of local magnetic fields produced by orbital electrons as well as other nuclear spins. $\vec{B}_{total} = \vec{B}_{external} + \sum \vec{B}_{spin}$.

### Coupling

The interaction between spins is mediated either directly by photons or indirectly by electrons involved in chemical bonds. The interaction is denoted in quantum mechanics by cross terms between two wave functions (Griffiths, 1999). The cross term is known in NMR as the J-resolved

term. Coupled coherent state is the basis for multidimensional NMR spectroscopy. State evolution between two scalar-coupled spins is described by the scalar Hamiltonian:

$$\mathbf{H}_0 \quad = \quad \mathbf{H}_z + \mathbf{H}_j = \sum_{i=1}^{N} \omega_i I_{iz} + 2\pi \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} J_{ij} \vec{I}_i \cdot \vec{I}_j \tag{4}$$

In the Equation 4 $\omega$ is the Larmor frequency and J is the scalar coupling constant. The interaction between two or more spins, that is mediated through chemical bonds, is called spin coupling. In NMR spins are governed by two main interactions: An indirect spin coupling interaction known as J coupling, which is transferred through a chemical bond. Second relevant interaction is direct dipole-dipole interaction, which will be described later. In a liquid state, the J coupling tensor is reduced to a single number, i.e., to scalar coupling. The strength of scalar coupling describes how rapidly a coherence between the spins will arise. Under a strong magnetic field the effect of the cross terms is decreased. The weak coupling limit, as given in Equation 5, is a practical approximation. This essentially omits cross terms from other spins than the spins of interest. The approximation is justifiable as the magnitudes of J couplings in protein samples range from $10^2$ Hz to zero Hz and Larmor frequencies are measured on the magnitude of $10^6$Hz. The result of this approximation simplifies the Hamiltonian between two spins, as seen in Equation 6.

$$|\omega_i - \omega_j| \gg 2\pi J_{ij} \tag{5}$$

$$\mathbf{H}_{IS} = \omega_I I_z + \omega_S S_z + 2\pi J_{IS} I_z \cdot S_z \tag{6}$$

**J coupling**

In Equation 6, $\omega$ is the Larmor frequency of nucleus *x* and $J_{IS}$ is a coupling constant between nuclei. This constant is commonly known as the J coupling. The J coupling plays an essential role in transferring coherence from one nucleus to a neighboring nucleus by fulfilling the requirement for the transfer time to be $2\tau$=1/(2J). All nuclei in the system are coupled. However, in practice couplings over more than three bonds are no longer useful since the transfer time T, which are proportional to $1/J_{ij}$, becomes so long that the coherence will relax beyond detection. In order to successfully establish coherences, knowledge of the magnitude of J couplings is essential for setting the coherence transfer time (Cavanagh *et al.*, 2007).

**Dipolar coupling**

Second of the two governing interactions in NMR experiment is the dipolar interaction. The interaction is not mediated through a chemical bond, but through space by photons, which is commonly known as dipole-dipole interaction. Equation 7 presents classical magnetostatic energy between two dipoles ($\vec{\mu}^i$ and $\vec{\mu}^k$) (Mansfield and O'Sulliva, 1999). From previous, one can derive Diopolar Hamiltonian corresponds to the classical energy as present in Equation 8 (Goldman, 1988).

$$E \quad = \quad \frac{1}{r^3} \left( (\vec{\mu}^i \cdot \vec{\mu}^k - 3(\vec{\mu}^i \cdot \vec{n})(\vec{\mu}^k \cdot \vec{n}) \right) \tag{7}$$

$$\mathbf{H} \quad = \quad \frac{\hbar \gamma_i \gamma_k}{r^3} \left\{ \vec{I}^i \cdot \vec{I}^k - 3 I_z^i I_z^k \right\} \tag{8}$$

In Equations 7 and 8 $\gamma_i$ is the respective gyromagnetic ratio, $r$ is the distance in space between spins and $\vec{n}$ unit vector $\vec{r}/r$. In isotropic solution environment, the Brownian motion causes the molecule to tumble randomly causing orientation-dependent magnetic interactions average themselves to zero. Dipolar couplings are only observed in anisotropic conditions (Brunner, 2001). This is highly useful property when designing NMR pulse sequences for liquid medium as it simplifies the mathematics.

In case of oriented or semi-oriented system the observed dipolar coupling between I and J is a function of two angles as seen in Equation 9 (Lipsitz and Tjandra, 2004).

$$D_{IJ}(\theta, \phi) = A_i^{IJ} \left\{ (3\cos^2\theta - 1) + \frac{3}{2} R(\sin^2\theta \cos 2\phi \right\} \tag{9}$$

In Equation 9 $A^{IJ}$ is the axial component, and R is the and rhombic component of the molecular alignment tensor, $\theta$ is the angle between z axis of the alignment tensor and the bond vector and $\phi$ is the angle between the projection of the bond vector onto the x-y plane and the x axis. The angles are presented in principal coordinate frame of the orientation tensor and the angles are not those of in laboratory frame. Turns out, in Equation 9 the eminent variables for RDCs are the angles $\theta$ and $\phi$ (Lipsitz and Tjandra, 2004).

In solid state NMR (SSNMR) the dipolar couplings have significant role (Frydman and Hardwood, 1995). SSNMR pulse sequence design is a field of its own and is not discussed in this thesis. Semi-homogenous sample medium can be utilized to restrict free tumbling. Typically, this is utilized to measure residual dipolar couplings (RDCs). RDCs will be discussed further in the Article IV.

## 3.2 Utilization of interactions in NMR

Interactions governing coherence transfer during an NMR experiment are dipolar and spin coupling. Basic distinction is the interaction medium. Chemical bond provides medium for the spin coupling, whereas the dipolar interaction takes place between dipoles through space. Higher terms from the interaction series expansion, such as quadrupole interaction, are usually omitted as the impact of the higher terms on the experimental design is minor.

Attaining a desired coherence state requires manipulation of the spins in precise ways. The manipulation scheme is referred to as a pulse sequence. It is implemented as a computer program, which controls radio frequency pulses and delays. In order to successfully design a transfer scheme, the evolution of each quantum coherence state must be calculated at each step. The evolution of the coherence state is described by the scalar Hamiltonian in Equation 4.

#### Chemical shift

Identification of the chemical shift for individual nucleus is essential for structure determination. In theory, each nucleus has a unique magnetic environment. However, in practice signals are not delta-spikes (Sayer and Mansingh, 2005). Signals with wide full width at half maximum (FWHM) overlap with each other, rendering identification difficult. The observed signal is a weighted average over time of all the conformations present in the sample. Robust structures provide superior distribution of chemical shifts in the sample compared to the flexible structures. Two categories of motion must be recognized: the overall tumbling of the molecule and motion within the molecule. The first one is generally desirable as seen in Equation 9. The latter causes chemical shifts to homogenize, which causes assignment problems.

#### Coupling

Spins in the protein couple with each other generating coherence states. Generation and transfer of coherence between nuclei enables multidimensional NMR spectroscopy. Typically, the coherence is transferred through a pathway in order to separate the resonance frequencies from each other through modulation with specific coupling partners. The coherence transfer pathway is often selected to survive by adjusting delays and active destruction via phase cycling or gradient selection (Bax *et al.*, 1980).

Modern multidimensional NMR spectrum is constructed from directly detected dimension and number of indirect dimensions. The frequencies in the indirect dimensions are collected by labelling the coherence with particular nucleus' frequencies. The labelling relies on the rate of coherence evolution. Each added dimension improves the resolution by reducing the signal overlap. This can be thought as a resonance vector, in which each element corresponds to resonance frequency of a labelled nucleus. However, the improved resolution is achieved at the cost of precision, which is caused by reduced signal intensity due to relaxation and transfer efficiency.

Multidimensional NMR utilizes interaction mechanisms to bring the system into a desired quantum coherence state. The state depends on the specific goal of the study. A pulse sequence is a set of specifications on how and when to pulse the sample in order to reach the state. For example, couplings can be measured via varying delays allocated for coherences to emerge. The emergence of the coherence state is described by the latter terms of Equation 4. When all the interactions are taken into account, the result is a cumbersome calculation. Typical approximation is to set all but the desired coherence emerging to be negligible. Other effects are discussed later.

**J coupling**

J coupling or scalar coupling describes the emergence speed of coupled state between two isotopes. Detailed knowledge of J couplings in the system is paramount in the pulse sequence implementation. In order to obtain as high coherence as possible, the transfer delays must be set to maximize the transfer condition. Typically this is $\sin(2\pi J_{xy}T)$. In Figure 2 are presented common magnitudes for J couplings in a amino acid residue. Additionally, the spectral density of J couplings present in the protein can be utilized to assess the robustness of the structure. In the IDSs the couplings do not exhibit values that are typical of helices or sheets.

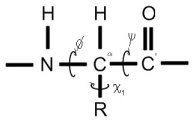


Figure 1. Angles phi, psi and chi define conformation of an amino acid residue.

Scalar couplings depend on the torsion angle between the atoms. The relevant angles for structure determination, and NMR, are present in Figure 1. By measuring the coupling, it is possible to define the local conformational space of the protein section in question leading into overall superior model (Dingley *et al.*, 2001). The peptide bond between residues is relatively stiff and does not allow extensive motion (Zumdahl and Zumdahl, 2003). This property allows the peptide backbone to be modeled as interconnected planes. Orientation of the backbone can be thus modeled by determining the relative angle between the planes. A simple method of determining the plane orientation can be done by measuring scalar couplings over three bonds. Rudimentary interpretation of the coupling magnitude to angle can be done via applying the Karplus Equation (Karplus, 1958; Minch, 1994):

$$^3J = A\cos^2\theta + B\cos\theta + C \tag{10}$$

Equation 10 describes the relation between the J coupling and the dihedral angle $\theta$ over three bonds in an amino acid residue. Constants A, B, and C are determined empirically and hence Karplus Equation provides only relative information between the angles (Cornilescu *et al.*, 1999). However, measurements over three bonds $^3J_{H^\alpha H^N}$ have been shown to provide further insight to protein structure (Dyson and Wright, 1991).

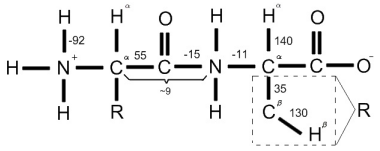


Figure 2. Typical magnitudes of J couplings of amino acid residues in Hz (Sattler *et al.*, 1999).

### Residual dipolar couplings

Residual dipolar couplings (RDCs) stem from anisotropic tumbling of the molecule in solution. This can be modeled by partially degenerated magnetic susceptibility tensor (Saupe and Englert, 1963). In liquid medium the tensor has little effect as the averaging caused by Brownian motion removes angle dependencies from the relation, causing the tensor to degenerate into a scalar. While this is typically a desirable effect, there is an additional way to take advantage of this property. By partly restricting the motion of the protein in transverse plane relative to magnetic field, one is able to obtain fine details about the alignment of particular bond vectors, which in turn can be incorporated into structure calculation step (Schwieters *et al.*, 2003).

A common way to obtain RDCs in a liquid medium is to introduce aligned particles. Usually these are steric and electromagnetic agents, such as bacteriophages, polyacrylamide gels or bicelles (Higman *et al.*, 2011). While there are no restrictions between which atoms the RDCs are measured, the typical approach is to measure couplings between amide protons and amide nitrogens or between carbons and the associated protons (Tjandra *et al.*, 1997).

### Nuclear Overhauser effect

Dipolar interaction enables establishment of distance restrictions by utilizing nuclear Overhauser effect (NOE). The NOE is used to generate coherence between dipoles through space and by utilizing the knowledge that the coherence signal intensity is inversely proportional to the sixth power of distance ($\propto \frac{1}{r^6}$). This is used to establish distance restraints between atoms. NOE data is used to determine both backbone conformation and side chain orientation. Typically, NOE is measured utilizing NOESY pulse sequence (Noggle and Shirmer, 1971).

# 4 Protein NMR spectroscopy

Three dimensional structure of small and medium sized proteins can be determined using NMR spectroscopy. The method is limited by the protein size and stability. Structures of globular proteins whose weight is 25 kDa[1] and less are solved routinely. Structure determination contains four steps. First, a sample with NMR active isotopes is obtained via expressing the protein in isotope enriched medium. The second step is acquiring a set of high quality spectra from the sample. The third step is assigning signals to their source nuclei in the amino acid sequence. Lastly, an ensemble of structures is computed. NMR as a technique has by now proven to be successful since 12% of protein structures in the Protein Data Bank have been solved via NMR (RCSB, a).

Modern heteronuclear NMR requires concentrated, isotopically labelled samples. Structure determination via NMR is based on identification resonance frequencies of nuclei and obtaining distance restraints between nuclei. In addition to these, supplementary information may be gathered, such as couplings, residual dipolar couplings, and relaxation times. Data processing has two separate steps. Correlation spectrum is construction via Fourier transformations and demands little from modern hardware Press *et al.* (2007). Latter part of the data processing, known as the assignment, is a manual labor-intensive process. Depending on the size and quality of the sample, the assignment takes from weeks to months.

NMR sample must be enriched with isotopes, which net spin deviates from zero, as indicated in Equation 1. Most commonly protein NMR utilizes $^{1}$H, $^{15}$N, and $^{13}$C isotopes. In some cases, utilizing $^{31}$P has proven to be beneficial (Marie *et al.*, 2012). Natural abundance of these isotopes is enough for recording simple spectra (Avison *et al.*, 1988), but not enough for acquiring multidimensional spectra and not for studying larger proteins (Brehm and Mullin, 1989). Thus, the established technique to obtain an isotope labelled sample is to express the protein in a medium that is enriched with desired isotopes (Primrose *et al.*, 2001). The required amount of purified sample is more than 100 $\mu$ moles. The exact amount required varies depending on the protein, but as a rule of thumb one is able to determine the three-dimensional structure from a sample, whose concentration is 0.1-1mM.

Modern protein structure determination utilizes the identification of chemical shifts of protons. The structure determination could be done by referring any common nuclei in the structure, but the Larmor frequencies difference renders such approach less than optimal (Weast and Astle, 1982). Signal identification is routinely done by generating a set of multidimensional correlation spectra, which correlate a common nucleus, allowing sequential overlap from one residue to next. Conflated with prior knowledge of expressed protein sequence, each signal can be associated with its source nucleus. In case of unknown structure, specific pulse sequences can be employed to identify the residues (Bhavesh *et al.*, 2001).

A correlation spectrum consists of resonance peaks in two or more dimensions. The peaks appear at cross sections where frequencies match those of nuclei labelled in coherence transfer pathway. Each residue undergoes similar coherence transfer sequence and yields coherence cross-peak(s).

A labelling occurs when coherence is evolved and measured incrementally. The labelling may be done while coherence is transferred or while disallowing the transfer. In former case, the labelling is typically utilized to measure the coupling and in latter one, the nuclei resonance frequency. The labelling interaction is governed by the firts term of the Hamiltonian in the Equation 4. Ideally, the labelling sequence modifies the coherence by a coefficient, which is a function of the resonance frequency of the nucleus being labelled (Griffiths, 2005).

[1] Da is not SI unit, but this convention is established in the field. 1u=1Da

Table 1. Common pulse sequences used in protein structure determination.

| Pulse sequence | Coherence correlation |
|---|---|
| HNCA | HN($i$)N($i$)C$^{\alpha}$($i$) |
| HN(CO)CA | HN($i$)N($i$)C$^{\alpha}$($i$-1) |
| HN(CA)CO | HN($i$)N($i$)C'($i$) |
| HNCO | HN($i$)N($i$)C'($i$-1) |
| HNCACB | HN($i$)N($i$)C$^{\alpha}$($i$)C$^{\beta}$($i$) |
| CBCA(CO)NH | HN($i$)N($i$)C$^{ali}$(*i-1*)C$^{\beta}$(*i-1*) |
| (H)CC(CO)NH | HN($i$)N($i$)C$^{ali}$(*i-1*) |
| H(CCCO)NH | HN($i$)N($i$)H$^{ali}$(*i-1*) |
| NOESY | H$_i$H$_j$ |

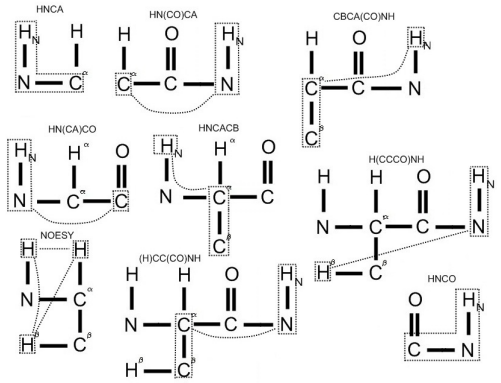


Figure 3. Common coherence pathways utilized in structure determination.

The signals are associated with their source nuclei by walking along the amino acid sequence from one cross peak to another. The pulse sequence(s), which overlap on nuclei in residues enable tracing connections throughout the protein backbone. Once the sequence of frequencies are known, one is able to assign the frequencies to particular nuclei by combining them with prior knowledge of protein sequence.

A continuous sequential assignment is required for successful assignment process. True to form, this becomes more and more challenging for increasingly larger proteins. Size is not the only obstacle for the assignment process. Specifically, the sequential assignment will be interrupted at prolines when that relies on pulse sequence utilizing amide protons. HNCO stands as the prime example for this limitation. In practice this is bypassed by measuring additional assignment pathways (Wang *et al.*, 1995). Some of the possible pathways are presented in Figure 3. In special cases, specific tailoring of pulse sequences may be required. This issue is addressed in Articles II and III. Another deviation occurs in glycines, which do no have a side chain. Experiments attempting to transfer or utilize $\beta$-carbon are going to produce either no signal or even artifacts. This issue also provides opportunities as seen in Article I.

Side chain assignment is performed with similar principles. The side chain protons are identified by linking their resonances to the backbone. Akin to backbone assignment, there are many sequences to assign the side chain resonances. For example, aliphatic side chains can be assigned utilizing $^{1}$H-$^{13}$C-HSQC, CC(CO)NH, H(CCCO)NH and HCCH-COSY. Whereas for aromatic residues one might utilize (HB)CB(CGCD)HD, (HB)CB(CGCDCE)HE, HCCH-COSY, and $^{13}$C-edited 3D-NOESY-HSQC. Representation of correlated nuclei is identical to that of the backbone as seen in Figure 4.

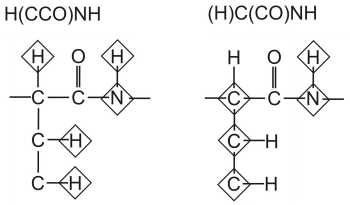


Figure 4. Correlated nuclei in H(CCO)NH and (H)C(CO)NH sequence.

Investigating protein structure is relatively straightforward (Drenth, 2007). However, the dynamics of the protein interactions are more challenging to investigate. Subsections of proteins fold on a timescale of nanoseconds (Williams *et al.*, 1996; Burkoff *et al.*, 2012). Direct observations of events in this timescale is impossible with NMR. Excitation pulses are typically around ten $\mu$s at high field strengths. FID requires around 100 ms, setting the upper limit of observable reaction speed. Typically single recording exhibits poor resolution. Established approach is to record numerous transients in order to obtain high quality data. More complex molecules require multidimensional spectra to be recorded. However, even the fastest multidimensional experiments require seconds to run (Amero *et al.*, 2008; Schanda *et al.*, 2005). NMR is not limited when observing slower reactions. The technique does not degenerate the sample and recording spectra from slow reactions is feasible. For example, incrementally modifying the reaction balance is a way to study the dynamics of the reaction. Titration studies have proven to be a powerful tool in NMR (Bundi and Wüthrich, 1979).

## 4.1 Challenges of IDPs to NMR

Intrinsically disordered proteins present novel challenges to the structural characterization. Most of the challenges stem from the flexible backbone of IDPs. The four most notable properties of IDPs are from the perspective of NMR are 1) a narrow chemical shift dispersion (Wüthrich, 1986), 2) proline rich segments (Marsh and Forman-Kay, 2010), 3) high amide proton exchange rate (Fersht, 2003), and 4) high tolerance for wide range of pH environments (Uversky, 2009a).

### 4.1.1 Chemical shift dispersion

A chemical shift of a nucleus is a weighted average over time of all the conformations present in the sample. In globular proteins, the rigid structure prevents excessive motion, cause wide chemical shift dispersion. IDPs lack stabilizing elements, like hydrophobic core and hydrogen bonds, enabling motion within the protein. The result is poor dispersion in proton dimension as demonstrated in Figure 5. Both, the bonds and dihedral angles vary enabling rapid transition between conformation states. Motion essentially causes averaging of the chemical shift and, thus, leads into poor proton shift dispersion. Poor dispersion in turn causes signal identification to fail and prevents from obtaining the assignment (Yao *et al.*, 1997).

In flexible systems, amide proton chemical shifts are more homogeneous than those of alpha protons. Amide proton chemical shift homogeneity stems from the relatively stiff peptide bond. Chemical shift of alpha protons is affected more by the neighboring side chain, causing the shifts to be modulated by the residue type. Indeed, an IDS identifier is the poor chemical shift dispersion in the amide proton dimension (Oldstone *et al.*, 1991; Espinoza-Fonseca *et al.*, 2012).

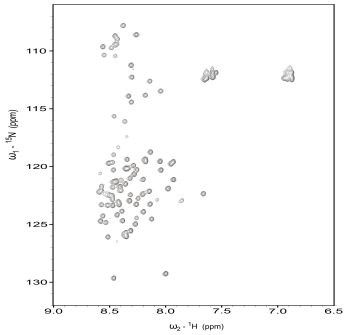


Figure 5. $^{15}$N HSQC spectrum of PAGE5 displaying poorly dispersed $^{15}$N, $^{1}$H correlations. Figure adapted from literature (Hellman *et al.*, 2011).

### 4.1.2 Amino acid composition

IDPs tend to incorporate a high number of prolines. Unique side chain structure of proline causes it to function differently from other 19 amino acids (AA). Prolines essentially prevent polypeptides from forming secondary structures (Li *et al.*, 1996). Typically, AA residues occur only in trans-isomer, but in prolines the energy difference between the *cis* and *trans* states is low enough to allow significant population of both conformations (Steinberg *et al.*, 1960). Lack of rigidity in the proline structure causes the IDPs and IDSs to exhibit larger radii of gyration than that of globular counter parts. True to form, IDPs radii of gyration is lower than that of random coils (Boze *et al.*, 2010).

Proline rich segments have been associated with segments exhibiting unstructured nature as well as functional sites (Polverini *et al.*, 2008; Kay *et al.*, 2000; Marsh and Forman-Kay, 2010; Köhler *et al.*, 2010). Designing functional sites is highly desirable, but is difficult (Koga *et al.*, 2012).

### 4.1.3 Proton exchange rate

Hydrogen atoms of proteins exchange with those of water solvent, causing signal loss and hence, reduces sensitivity. Amide proton exchange rate is significant, whereas that of alpha proton is not (Krishna *et al.*, 1979; David *et al.*, 1994; Faleev *et al.*, 2004). Proton exchange rates ($pK_a$) vary, but is a function of pH, temperature, location, and interactions.

Exchange rates are slower for the protons located in the hydrophobic core or shielded by hydrophilic structures (Yi and Baker, 1996). Additionally, if the proton is involved in hydrogen bond, the exchange rate is lower (Yi and Baker, 1996). Temperature is an obvious variable for the exchange rate. Increased vibration and kinetic energy enable breaking the potential barriers of the bonds. Availability of protons, i.e., pH, is associated with the proton exchange rates (Wüthrich, 1986).

A way to observe reaction, is to shift the reaction balance incrementally (Fersht, 2003). Modification of the reaction environment factors is a way to influence reaction fold or dynamics. Modifying the environment to alkali conditions may be a way to shift the reaction balance, but this causes NMR signals to diminish due to proton exchange with solvent. An example of the effect on the fold by environment manipulation is the addition of deuterium ($D_2O$), which is typically added 5-10% v/v to the sample. Deuterium has a small effect on folding (Fersht, 2003). Rising temperature is generally desirable from the perspective of NMR. High temperatures cause NMR to exhibit smaller FWHM due to more rapid tumbling of the protein. However, globular proteins denature at extreme temperature or pH (Fersht, 2003).

IDPs suffer less from extreme conditions. High temperature is less degenerating as they do not incorporate hydrophobic core, which could break down. Breaking the peptide chain requires significantly more energy than that fold (Cabra *et al.*, 2006; Permi and Hellman, 2012).

### 4.1.4 Motion in the structure

IDPs exhibit notable motion in their structure. Properties enabling motion have been described earlier. The mobility is not restricted to the backbone, but side chain rotation is an issue in itself, and is not discussed in this work. The mobility enables easy recognition of the IDPs with both NMR and other techniques. For example, semi-free polypeptide chains exhibit larger hydrodynamic radii. Additionally, the effective surface area is larger than that of globular counterparts with the same amount of residues. This enables larger binding trajectories (Shoemaker *et al.*, 2000). Structural mobility causes notable effects, which must be taken into account when designing NMR experiments (Carrasco and de la Torre, 1999; Curcó *et al.*, 2012).

Transverse relaxation times in globular proteins are inversely proportional to their size. $T_2$ relaxation, or the transverse relaxation, stems from spin-spin relaxation. Prominent intramolecular relaxation mechanism stems from fluctuation of dipolar magnetic field within the protein as it tumbles in the solution (Cavanagh *et al.*, 2007). The tumbling combined with the interaction network causes effectively fluctuating magnetic field upon specific spin. The dipole-dipole force is proportional to $\frac{1}{r^6}$, so in globular proteins, the interaction network is more extensive than in IDPs. In IDPs, the rapid reorientation causes the effective spin density to be significantly smaller. The effect results the IDPs to exhibit longer $T_2$ relaxation times than those of globular counterparts. Longer $T_2$ relaxation times provide clear benefits for NMR. Signals benefit from narrower FWHM. Less prominent coherence relaxation enables longer and more sophisticated pulse sequences.

Rapid reorientation in IDPs can be observed in particular coupling constants. Coupling constants over three bonds ($^3J_{H^\alpha H^N}$) have been shown to be good indicators of the robustness (or the lack) of the structure (Donne *et al.*, 1997; Dyson and Wright, 1998; Köhler *et al.*, 2010). Flexible backbone of an IDP is seen also as a reduced dispersion of $^3J_{H^\alpha H^N}$ coupling constants. Globular proteins show wide range of couplings (Tycko, 2001), whereas distribution in IDPs is more narrow. Numerous conformational states of IDPs and rapid reorientation is seen in NOESY spectra. The peaks exhibit large FWHM and the overall spectrum contains fewer and weaker cross peaks.

Amide and alpha protons resonate at noticeably different frequencies. In random coil the difference is about 4 ppm (Schwarzinger *et al.*, 2001). Even without stabilizing structures, the peptide bond cannot move extensively. This homogenizes the amide proton chemical shift even further. Alpha proton chemical shift is influenced by the side chain. Utilizing $H^\alpha$ would modulate the observed signal as a function of the residue in a greater degree than amide proton. This option is explored in Articles II and III.

# 5 Materials and methods

### Equipment and software

Experiments were carried out with two Varian Unity INOVA spectrometers (Unity INOVA). Field strengths of the spectrometers were 600 MHz (14.1T) and 800 MHz (18.8T). The spectrometers were equipped with $^{15}N/^{13}C/^{1}H$ triple resonance probe heads with actively shielded Z-axis gradient system.

Experiments were recorded with VNMRJ 2.1 revision B software package and analyzed with VNMRJ 2.1 revision B and Sparky 3.1.10 (VnmrJ; Sparky).

### Samples

The samples were uniformly enriched with $^{15}N$ and $^{13}C$ isotopes, in order to perform heteronuclear triple resonance experiments. Experiments were evaluated using number of well studied proteins. Human ubiquitin was dissolved to 1.9 mM concentration in 5% $D_2O$, 10 mM potassium phosphate buffer with pH 5.8 (Goldstein *et al.*, 1975). The sample was sealed in Wilmad 535PP NMR tube (Wilmad). Melanoma-associated cancer-testis antigen 16 (CT16) was dissolved in 7% $D_2O$, 10 mM Tris-HCl, 50 mM NaCl buffer (Nylund *et al.*, 2012). Protein concentration was 0.8 mM and the sample was sealed in a Shigemi microcell NMR tube (Shigemi). Immunoglobulin-binding domain B1 of streptococcal protein G (GB1) was dissolved in 20 mM potassium phosphate buffer (Sjöbring *et al.*, 1991). The sample concentration was 1.5 mM at pH 5.5 with 7% $D_2O$. Specimen was sealed in a 250 $\mu$l Shigemi micro-cell. Fifth repeat of enterohemorrhagic Escherichia coli effector (EspFU) was dissolved in buffer containing 20 mM sodium phosphate (Campellone *et al.*, 2004). The sample concentration was 1 mM at pH 7.0 and 50 mM NaCl, supplemented with 7% $D_2O$. Specimen was sealed in a 250 $\mu$l in Shigemi micro-cell tube. The last sample was human filamin A tandem immunoglobulin-like domains 16-21 (IgFLNa 16-21) (Heikkinen *et al.*, 2009a,b). The sample was dissolved in 50 mM $NaPO_4$, pH 6.8, 100 mM NaCl, 1 mM DTT and 2 mM NaN, supplemented with 7% $D_2O$. The concentration of specimen was 0.4 mM.

### Experimental setups

The **i(HCA)CO(CA)NH** was developed and tested using two samples. Ubiquitin sample spectrum was recorded at 20 °C in 600 MHz, collecting 32/16 transients per FID with 19/38 and 1,024 complex points, corresponding to acquisition times of 9.5/19 and 85 ms in $t_1$ and $t_3$, respectively. Total acquisition time for the experiment was 24 min. **Intraresidual HN(CA)CO** was tested with similar setup: experiment recorded at 800 MHz, at 25 °C with 52, 52 and 682 complex points in $t_1$, $t_2$, and $t_3$ dimensions, respectively. The corresponding acquisition times were 17.3, 23.6, and 85.2 ms in $^{13}C'$, $^{15}N$ and $^{1}H$ dimensions, respectively. Two transients per FID were used for signal accumulation, resulting in a total experimental time of 7.3 h.

Sequences **iH(CA)NCO**, **iHCAN**, **H(CA)CON**, and **HCA(CO)N** were subject to experimental verification using GB1 and EspFU. All the spectra were recorded at 25 °C, on a 800 MHz spectrometer. The experiments were recorded using 2 transients per FID with 128 and 512 complex points, corresponding to acquisition times of 62 and 51.2 ms in $^{15}N$ and $^{1}H$ dimensions, respectively. Experimental time per each spectrum was 16 min.

The **H(CA)CO(N)** was recorded by using 8 (16) transients per FID with 124 (62) and, 1,024 complex points, corresponding to acquisition times of 62 (31) and 64 ms in $t_2$ ($^{13}C$) and $t_3$ ($^{1}H$). The **iH(CA)N** spectrum of GB1, was recorded using 32 transients per FID with 96 and 1,024 complex points, corresponding to acquisition times of 27 and 64 ms in $t_1$ ($^{15}N$) and $t_3$ ($^{1}H$). The experimental time was 2 h. **H(CACO)N** spectrum was recorded with identical parameters except for 8 transients, resulting in total experimental time of 30 min. Intraresidual **iH(CA)NCO** and **iHCAN** were tested on EspFU R47, recording 4 (2) transients per FID using 96, 36 (150) and 1,024

complex points in $^{15}$N, $^{13}$C'($^{13}C^{\alpha}$) and $^{1}$H dimensions. This corresponds to acquisition times of 27, 18 (25), and 64 ms in $t_1$, $t_2$ and $t_3$, respectively. Complementary **H(CA)CON** and **HCA(CO)N** spectra were recorded with four transients per FID using 96, 36 (41) and 1,024 complex points in $^{15}$N, $^{13}$C' ($^{13}C^{\alpha}$) and $^{1}$H dimensions, which is equal to acquisition times of 27, 18 (6.8), and 64 ms in $t_1$, $t_2$ and $t_3$, respectively.

**MQ-HNCO-TROSY** was tested on three proteins: ubiquitin, GB1 and IgFLNa 16-21. IgFLNa was in a dilute liquid crystal medium, composed of filamentous phage particles, Pf1 (ASLA). The phage concentration was 15 mg/ml. The experiment was recorded on both 600 and 800 MHz spectrometers. Spectra were recorded using 2 transients per FID with 128 and 512 complex points, corresponding to acquisition times of 62 and 51.2 ms in $^{15}$N and $^{1}$H dimensions, respectively. Experimental time per each spectrum was 16 min.

**(HCA)CON(CA)H** and **(HCA)NCO(CA)H** were verified using GB1 and CT16. CT16 was in two different buffers: the one described above, and one with 20 mM sodium phosphate. The spectra were recorded at 800 MHz and 25 °C. The experiments were measured using GB1 with 84 and 768 complex points in $t_1$ and $t_2$ were recorded using 8 transients per FID. This corresponds to acquisition times of 28 and 64 ms in $t_1$ ($^{13}$C') and $t_2$ ($^{1}$H) dimensions. The total experimental time for each spectrum was 33 min.

**(HCA)CON(CA)H** and **(HCA)NCO(CA)H** experiments 86, 128 and 768 complex points in $t_1$ ($^{13}$C), $t_2$ ($^{15}$N), and $t_3$ ($^{1}$H), respectively, were collected. This translates into acquisition times of 33.1, 42.7, and 64 ms, respectively. For **iH(CA)NCO**, 128 ($t_1$, $^{15}$N), 48 ($t_2$, $^{13}$C) and 768 ($t_3$, $^{1}$H) complex points were collected, yielding acquisition times of 42.7, 18.5, and 64 ms, respectively. For each experiment, two transients per FID were used. The total experimental time for (HCA)CON(CA)H and (HCA)NCO(CA)H experiments was 22 h each, and 16.5 h for iH(CA)NCO.

# 6 Results and discussion

Much of the function and structure dynamics of IDPs remain unknown. This stems from their relatively recent discovery as well as from difficulties to study unstructured polypeptides. The goal of this thesis was to improve pulse sequences utilized in the study of IDPs and dynamic structures. When measuring IDPs or loose structures with standard experiments the results suffer from signal overlap, low spectral resolution and poor signal recognition, rendering the spectra often unusable. To address these challenges, the thesis focuses on pulse sequence development. NMR has the prospect of producing tailor-made pulse sequences for specific purposes. We aimed at improving the reliability of pulse sequences in order to overcome the challenges presented by mobile structures.

The set of proposed pulse sequences is a direct attempt to characterize and take advantage of properties of IDPs. A notable change in our approach compared with standard protocol is the utilization of $H^{\alpha}$ instead of $H^{N}$. This is justified by many factors. The typical $H^{N}$ approach fails when protein composition is proline-rich, and, thus, utilizing a different nucleus has notable benefits. The chemical dispersion in $H^{N}$ is poor due to averaging caused by tumbling of the molecule. A way to influence protein conformation distribution is to modify pH (Russo *et al.*, 2012). In NMR this approach is not always preferable as in the alkali conditions the proton exchange with the solvent causes loss of signal. The amide proton exchange rates are noteworthy (Eriksson *et al.*, 1995). Standard pulse sequences utilizing amide proton will yield poor spectra in alkali conditions. Extensive motion yields longer $T_2$ relaxation times, which enable longer pulse sequences without critical signal loss. Prolonged relaxation times enable more complex and unorthodox pathways as explored in Articles II and III.

## 6.1 Pulse sequence design

A pulse sequence is a compromise between transfer efficiency and resolution. This dilemma in design stems from the limitation of relaxation and resolution required for signal identification. In addition to this dichotomy, there is the issue of suppressing noise and artifacts. For example, since the beginning of pulsed Fourier transform NMR spectroscopy, water signal has been a significant source of artifacts. Additionally, in an ideal pulse sequence, the only signals observed would be correlation peaks stemming from a desired coherence transfer pathway. However, this is rarely so, as there are often leaking coherences. These may stem from side chains and in some case from backbone. Articles I, II, and III implement the intraresidual filter to partially reduce the artifact effect from a leaking coherence transfer pathway. Each filter, however, requires additional pulses and, thus, more time. The length of a pulse sequence is ultimately limited by the relaxation and transfer efficiency. In practice, sample stability is a limiting factor. The optimal pulse sequence ultimately combines both precision and resolution.

When designing a coherence transfer pathway, the typical approach is to apply standard assumptions of operation formalism (Søransen *et al.*, 1983). Noteworthy, assumptions that operation formalism presupposes are the following. Weak coupling limit ($|\omega_i - \omega_j| \gg 2\pi J_{ij}$) is met between two arbitrary spins. Transfer efficiency is unlimited and is calculated only in the case where it is deemed significant. Magnetization evolution during pulses is negligible. Applied pulses are precise in frequency and duration. Leaking magnetization from analyzed pathways is irrelevant as only the selected pathway is assumed to survive till FID. Labelling is assumed not to have impact on coherence being evolved.

Analysis of a pulse sequence can usually be done in independent pieces. The property stems from the design: a particular block of pulse sequence is used to perform a dedicated task. The segments are usually transferable between sequences with small modifications. Similarly, the sequences presented herein consist of established blocks. Novel sections are identifiable from the sequence via the details in the relevant article. Of the improvements two are analyzed in depth. The main features are glycine specific recognition mechanism and generalization of the

intraresidual filter.

Presented pulse sequences incorporate well known established blocks, which are not analyzed. Indirectly detected dimensions utilize either gradient coherence selection (Bax *et al.*, 1980; Vuister *et al.*, 1992) or frequency discrimination via States-TPPI phase cycling (Maudsley *et al.*, 1978; Marion *et al.*, 1989). In the following pulse sequence presentation, the thesis concentrates upon novel coherence transfer schemes and labelling techniques. For analysis purposes the following established segments are assumed to be known: INEPT, sensitivity enhanced INEPT (Morris and Freeman, 1979) and TROSY backtransfer (Pervushin *et al.*, 1997). Effectively, the carbon channel is utilized to separate between carbonyl carbon and alpha carbon as the mean of the two resonance regions are 118 ppm apart, which allows an easy selection of either band. In the analysis, it is mentioned whether the pulse excites the whole channel or partial band for either C' or $C^{\alpha}$ resonances. In the presented pulse sequences, coherence transfer pathways are selected by field gradients. When used, shaped pulses are assumed to have ideal excitation shape without phase error.

When analyzing emerging coherences, number of possible coherence pathways can be dismissed. While all possible coherences emerge, the specific transfer times, gradient selection and pulse schemes overall ensures that only the selected pathway survives for observation. For example coupling constants of $^2J_{C^{\alpha}N}$ and $^2J_{C'N}$ are in the magnitude order of 10 and 0.5 Hz, respectively, thus transitions times from $C^{\alpha}$ and C' to N require substantially different transfer delays. So long, that the emerging coherence from second pathway is negligible compared to selected one. Each transfer step constricts the possible pathways.

Pulse sequences are typically presented in graphical form. In this form, each excitation band is at different line. Pulses are presented in rectangles. Narrow rectangles represent $\pi$/2 pulses and wide rectangles $\pi$ pulses. Short pulses are specifically utilized to hit protons around 4 ppm. Shaped half-ellipses present shaped pulses hitting specific bands of selection. Wide and narrow ellipses present $\pi$ and $\pi/2$ pulses, respectively. Pulses are applied along X-axis unless otherwise mentioned. Gradients are presented in their own line and are presented by half-ellipses.

Coherence transfer pathway is more abstract presentation than full pulse sequence. In the Articles, the shorthand version is utilized to convey the overall idea behind each pulse sequence. This takes the form of arrows and key coherences. Each transfer pathway begins with set coherence, which is evolving. Active couplings involved in the coherence transfer are indicated in above the arrow and in the square brackets. For example, INEPT sequence would be written in this form as seen in Equation 11.

$$H^N(i) \xrightarrow{T;^1J_{H^NN}} H^NN(i) \tag{11}$$

### 6.1.1 Glycine selection

Number of the presented pulse sequences utilize the double alpha protons of glycines to generate distinct coherence compared to other residues. The signal stemming from glycines is manipulated to express a $\pi/2$ phase difference relative to other residues. The phase difference expressed by glycines stems from the transfer delay $\tau_2$ following the initial INEPT transfer from $H^{\alpha}$ to $C^{\alpha}$ as seen in Figure 6. Typically, the delay is set to $1/(2J_{C^{\alpha}H^{\alpha}})$, in order to align the water signals for transition to Z axis. However, in case of glycines second alpha proton causes deviating Hamiltonian acting upon the coherence adding factor of $\cos(\pi J_{C^{\alpha}H^{\alpha}}\tau_2)$. This can be utilized by overextending the transfer delay to compromise by selecting the transfer delay to fit local minimum for glycines. This does lead to a marginal loss of magnetization on non-glycine residues as the transfer is not at the theoretical maximum. However, nothing prevents from setting the transfer delay to the established one and losing glycine signals. Different options may be utilized on very crowded spectra.

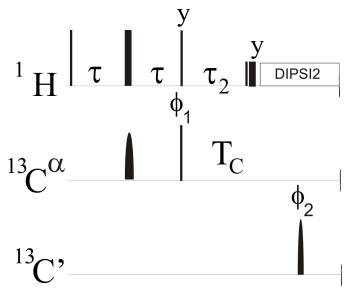


Figure 6. Schematic presentation of the segment-implementing glycine specific transfer period.

Figure 7 shows simulated transfer coefficients for glycines and other amino acids. Settings for the simulation are based on prior studies (Löhr and Rüterjans, 1995; Naumann and Kuchel, 2011). The examination was carried out by simulating the coherence evolution with and without the additional cosine term and setting the J coupling $J_{ch}$ to 143 Hz. The results are consistent with prior publication (Permi, 2002).

As seen from the Figure 6, established INEPT transfer is obtained by setting the transfer delay around 2.3 ms. From 3.5 ms onward phase of the glycines has gone over $\pi$, and signal begins to emerge again. Therefore when the transfer delay is set to about 4.4 ms, one is left with decreased intensities for both components: glycines with $\pi$ phase differences compared to other signals. The phase difference unambiguously identifies the glycine residues.

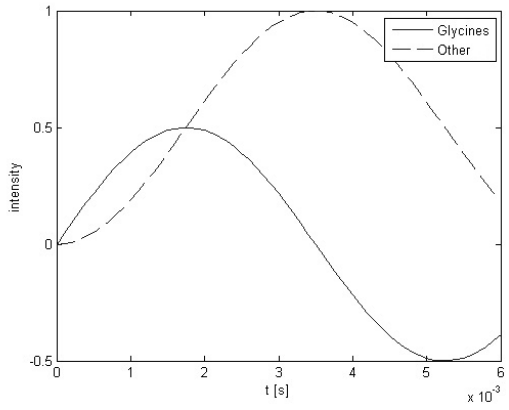


Figure 7. Simulation of coherence transfers in glycines and other amino acids.

### 6.1.2 Intraresidual filter

This element is used in many of the presented pulse sequences. The element serves to suppress undesirable coherences that arise from coherence leaking, which is caused by closely similar couplings of $|^2J_{C^\alpha N}| \approx |^1J_{C^\alpha N}|$. To discriminate between one and two-bond couplings, the filtering element is based on asymmetry of the emerging coherences. The element is similar to the intraresidual filter (Permi, 2002) and, hence, it is described in the same way (Figure 8). Original experiment establishes $H^N$N coherence, which evolves to $H^N$NC'$_z$(i-1)C$^\alpha$(i-1) and $H^N$NC'$_z$(i-1)C$^\alpha$(i) with coefficients that have phase difference of $\pi/2$. This allows one to maximize the coherence transfer for one or the other pathway. The presented scheme is based upon C$^\alpha$-N interaction, which is active throughout the segment. The interaction causes additional coefficient to emerge, which allows minimization of leaked coherence.

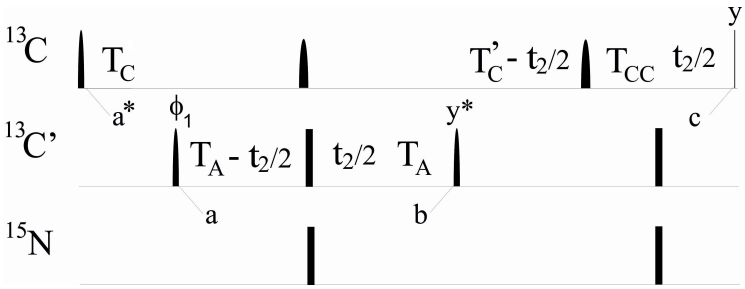


Figure 8. Schematic presentation of labelling in C' dimension and additional coherence emergence suppression.

The experiment begins with $H^{\alpha}$-$C^{\alpha}$ INEPT type scheme. At this point the coherence pathway is split. The analysis follows the pathway selected for survival. During the transfer delay $T_C = \frac{1}{2J_{C^{\alpha}C'}}$ foremost coupling is ${}^1J_{C^{\alpha}C'}$ :

$$\xrightarrow{T_C;\,{}^1J_{C^{\alpha}C'}} \quad C^{\alpha}(i)C'(i) \tag{12}$$

Next, the coherence is labelled during $2T_A$ with C' frequency in a constant time manner (Rance *et al.*, 1984) with refocusing pulses on N, C', and $C^{\alpha}$ channels. Significant coherence emerging from this is:

$$\xrightarrow{2T_A;\,{}^1J_{C^{\alpha}N},\,{}^2J_{C^{\alpha}N},\,{}^1J_{C'N}} \quad C^{\alpha}(i)C'(i)N(i)\cos(\omega_C t_1) \tag{13}$$

The term evolves during $2T_C$ when C' and $C^{\alpha}$ are refocused.

$$\xrightarrow{T'_C+T_{CC};\,{}^1J_{C^{\alpha}N},\,{}^2J_{C^{\alpha}N},\,{}^1J_{C^{\alpha}C'}} \quad C_z^{\alpha}(i)N_z(i)\sin(2\pi^1 J_{C^{\alpha}N}(T_C+T_A))\sin(2\pi^2 J_{C^{\alpha}N}(T_C+T_A))\vartheta \tag{14}$$

In Equation 14 $\vartheta$ includes non-essential terms, like transfer efficiencies and labelling terms. Inter- or intraresidual pathway is selected by allowing ${}^1J_{C^{\alpha}N}$ and ${}^2J_{C^{\alpha}N}$ couplings to modulate from point a* to c in Figure 8. This causes the term to obtain factor of $\sin(2\pi^1 J_{C^{\alpha}N}$ ($T_C$+$T_A$))$\sin(2\pi^2 J_{C^{\alpha}N}$ ($T_C$+$T_A$)) as seen in Equation 14, but the term in *i*:th residue expresses the same factor, but with the sine terms changed to cosine. Thus, the intra residue transfer is selected by maximizing the double cosine terms. Inter residue transfer is selected by maximizing the double sine terms, which are omitted in this presentation.

## 6.2 Undesirable coherences

### Basis

Virtually every backbone assignment protocol includes HN(CA)CO experiment (Clubb *et al.*, 1992; Löhr and Rüterjans, 1995). The HN(CA)CO experiment produces correlation spectrum between $H^N$(i)-N(i)-C'(i) as well as C'(i-1). The latter C'(i) and C'(i-1) correlation peaks emerge in tandem in the same dimension. The double terms stem from the leaking coherence during the transfer from amide to alpha carbon. Difference between ${}^1J_{NC^{\alpha}(i)}$ and ${}^2J_{NC^{\alpha}(i-1)}$ is too small to enable discrimination between the two pathways.

The double resonance peaks in one spectral dimension become a hurdle when signal overlap starts to hinder discrimination between signals. Numerous signals from large proteins presents an obvious problem, but even in small IDPs the assignment is hard, due to poor chemical shift dispersion. The first article set to reduce the signal overlap. The goal was to implement intraresidual filter (Permi, 2002) to HN(CA)CO experiment to dispose peaks from inter-residual

pathway.

The design challenge was two-folded: first, the experiment produced both inter- and intraresidual correlation peaks in carbon dimension. Extra coherence pathway should similarly disposed. Secondly, the sensitivity of the experiment did not meet requirements of standard assignment protocol for large proteins. Complementary pulses sequence of HN(CA)CO is typically HNCO. Advantages of simultaneously development were recognized, but deemed unwarranted as this point.

**Implementation**

Development on signal overlap reduction yielded a pulse sequence with two variants. Hence dubbed i(HCA)CO(CA)HN, with and without spinlock.

**i(HCA)CO(CA)NH without spinlock** produces correlation spectrum between $H^N$(i)-N(i)-C'(i) as seen in Figure 9. The sequence incorporates the intraresidual filter to suppress the inter residual pathway. Sequence begins with the $H^\alpha$ INEPT (Morris and Freeman, 1979) leading into coherence of $H_zC_z^\alpha$. In the next step, the magnetization is transferred to the neighboring C' atom, during which any simultaneous magnetization to proton is refocused and, thus, this particular coherence pathway can be discarded during next steps. Any leaking to protons is prevented by applying DIPSI2-decoupling pulse scheme on proton channel (Shaka *et al.*, 1988). The following is the constant time labelling in first indirect dimension by incrementing the transfer delay $2T_A$, which dephases C'(i) magnetization with respect to N(i+1). The labelling scheme has been already discussed above in the section describing intraresidual filter. The second indirect dimension labelling of N(i) is carried out before returning the magnetization to the amide proton for FID recording. The second labelling scheme is conceptually similar to that described in the section presented in Figure 8 from timepoints B to C. The $t_2$ labelling scheme bears resemblance to the sensitivity enhanced HNCO. Finally, the coherence is returned to $H^N$ for observation by INEPT backtransfer protocol.

i(HCA)CO(CA)HN is presented in detail in Figure 10, and can be condensed in the following manner:

$$H^\alpha \overset{^1J_{HC}}{\rightarrow} C^\alpha \xrightarrow{^1J_{C\alpha N},\,^2J_{C\alpha N},\,^1J_{C\alpha C'}} C' \left[2T_A - t_1;^1 J_{C'N};^1 J_{C\alpha N};^2 J_{C\alpha N}\right]$$
$$\xrightarrow{^1J_{C\alpha N},\,^2J_{C\alpha N},\,^1J_{C\alpha C'}} N \left[2T_N - t_2;^1 J_{C\alpha N};^2 J_{C\alpha N}\right] \overset{^1J_{NH}}{\rightarrow} H^N \left[t_3\right]$$

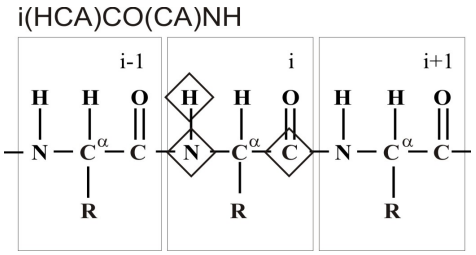


Figure 9. Correlated nuclei in i(HCA)CO(CA)NH sequence.

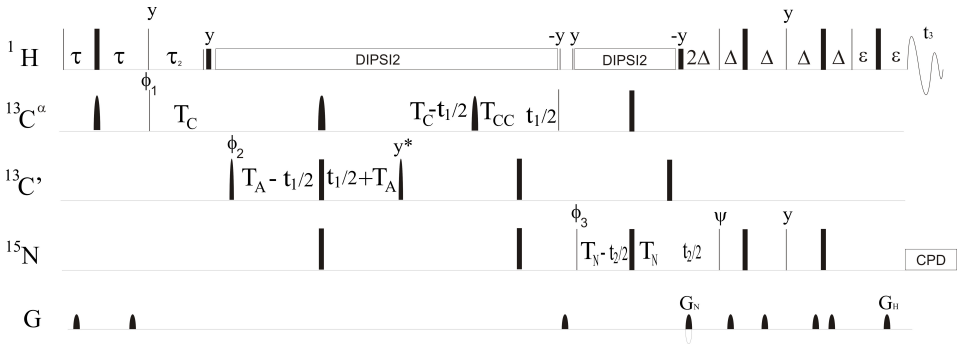


Figure 10. Schematic presentation of i(HCA)CO(CA)NH pulse sequence.

The i(HCA)CO(CA)NH's variation, **i(HCA)CO(CA)NH with spinlock**, begins $C^{\alpha}$ INEPT followed by transfer to C'. Next is the $t_1$ labelling section, which diverts from the version without spinlock. Essentially, the labelling sequence is exactly the same as in the previous version, but in backwards manner. Constant time $T_C$+$T_{CC}$ proportion of the labelling is done first and the rest is in similar fashion. Besides the labelling, the rest of the sequence is identical to the prior version. The product operator formalism is reversible and the coherence transfer is from $H_z$ to $C_y^{\alpha}H_x$ from which the selected coherence pathway is split by the intra-residual filter. After this, the coherence $t_1$ labelled in $C^{\alpha}$ dimension and the split pathways are brought together. Overall, this produces a correlation spectrum between $H^N$(i) - N(i) - $C^{\alpha}$(i).

i(HCA)CO(CA)NH with spinlock was implemented to improve sensitivity further in other residues than glycines. This was designed to achieved through quenching the dipolar interaction between $H^{\alpha}$ and $C^{\alpha}$. Relaxation rate of multiquantum coherence under quenched conditions leads into extended transverse relaxation times with respect to $C^{\alpha}$ single quantum coherences.

i(HCA)CO(CA)NH with spinlock is presented in detail in Figure 11, and can be condensed in the following manner:

$$H^{\alpha} \overset{{}^{1}J_{HC}}{\rightarrow} C^{\alpha} \xrightarrow{{}^{1}J_{C\alpha N},{}^{2}J_{C\alpha N},{}^{1}J_{C\alpha C'}} C' \left[2T_A - t_1;{}^{1}J_{C'N};{}^{1}J_{C\alpha N};{}^{2}J_{C\alpha N}\right]$$
$$\xrightarrow{{}^{1}J_{C\alpha N},{}^{2}J_{C\alpha N},{}^{1}J_{C\alpha C'}} N \left[2T_N - t_2;{}^{1}J_{C\alpha N};{}^{2}J_{C\alpha N}\right] \overset{{}^{1}J_{NH}}{\rightarrow} H^N \left[t_3\right]$$

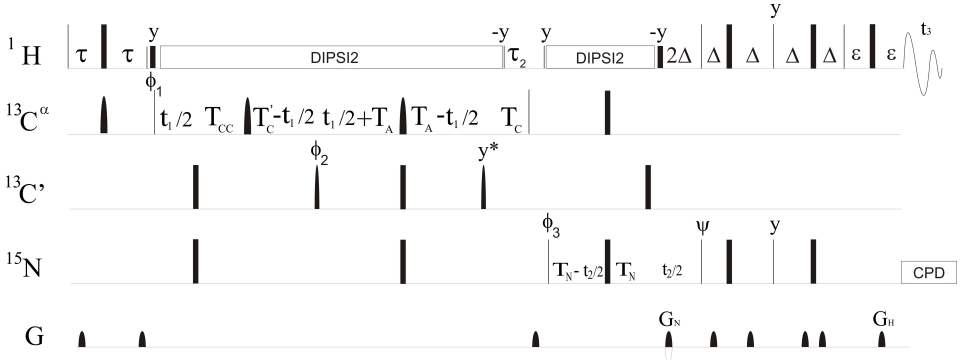


Figure 11. Schematic presentation of i(HCA)CO(CA)NH with spinlock pulse sequence.

**Results**

The i(HCA)CO(CA)NH sequence proved to be an improvement over the original pulse sequence. The inter-residual coherence pathway was suppressed and did not generate observable signal. The novel pulse sequence exhibits sensitivity enhancement of ≈1.65 compared to non-TROSY

MQ intra-HN(CA)CO (Nietlispach, 2004) when comparing intensity of $H^N$(i)-N(i)-C'(i) correlations. Factors contributing to sensitivity enhancement stem from the length of the experiment. i(HCA)CO(CA)NH is approximately 50 ms shorter than the original intra-HN(CA)CO counterpart. The novel sequence provides superior sensitivity for both structured and unstructured regions in small and medium sized proteins. i(HCA)CO(CA)NH functioned in tandem with a number of established pulse sequence without complications. The pulse sequence accomplished the goals set for the work.

i(HCA)CO(CA)NH with spinlock was designed as a concept and no experimental data was collected. Theoretically, the sensitivity improvement stems from the quenching of mutual dipolar interaction between $H^\alpha$ and $C^\alpha$, reducing the transverse relaxation effect. Calculations show improved relaxation times compared to starting scheme.

## 6.3 Deviating from the amide proton approach

### Basis

Implementing a filtering element into the HN(CA)CO experiment yielded the desired results. However, the filter did only partially solve the signal overlap issue when studying IDPs. Further development to improve resolution was deemed necessary. The goal was set to reduce signal loss from the proton exchange rate and further reduce signal overlap. As shown prior, utilization of amide proton is a source of problems when investigating IDPs with NMR. Alternative options for the established approach (Rule and Hitchens, 2006) were looked into.

HN(CA)CO sequence (Clubb *et al.*, 1992), with the implemented intraresidual filter, was selected for the basis of the development. Instead of established $H^N$ approach, $H^\alpha$ was selected to be the observable nucleus. Alpha-proton based detection approach has been presented in the past (Olejniczak and Fesik, 1994; Bottomley *et al.*, 1999). Single, detached correlation spectrum would be of little use and, therefore, second complementary pulse sequence was planned. Ideally these two sequences would link sequential residues. Many of the established sequences establish the link through N and C' (Cavanagh *et al.*, 2007). The correlations of i(HCA)CO(CA)NH matched with these, improving the utility of both pulse sequences. Following the established suit, overlap was designed to include both N and C' nuclei.

### Implementation

This work presented two pulse sequences with each two variants: iH(CA)NCO, iHCAN, H(CA)CON and H(CA)CON with alternative implementation. Following the naming conventions of the field (Queen's University, Canada), the sequences produce H(i)-C'(i)-N(i), H(i)-$C^\alpha$(i)-N(i), H(i)-C'(i)-N(i+1) and H(i)-C'(i)-N(i+1) correlation spectra, respectively. The presented pulse sequences have options of labelling $C^\alpha$ instead or in addition of C'. In the sequences signals from glycine residues exhibit phase difference of $\pi$ in comparison to other residue signals.

**iH(CA)NCO** begins with $H^\alpha$ INEPT yielding $H^\alpha C^\alpha$ coherence. Next step is $t_2$ labelling. The labelling section bears resemblance to scheme presented in Figure 8. While the overall transfer shape is similar, the delays are adjusted accordingly. $T_A$=$(4J_{C'N})^{-1}$, $T_{CC}$=$(J_{C^\alpha C^\beta})^{-1}$-$(4J_{C'N})^{-1}$-$(2J_{C^\alpha C'})^{-1}$ and $T_C$=$(J_{C^\alpha C^\beta})^{-1}$ -$(4J_{C'N})^{-1}$. Similarly, the resulting coherence is $C_z$(i)$N_z$(i)$\cos(\omega_{C'} t_2)$. The following is the transfer onto nitrogen and $t_1$ labelling by standard labelling technique. The coherence is refocused in the middle of incrementation delay by simultaneous $\pi$ pulses on H, $C^\alpha$ and C'. After these, the coherence is brought back to proton for observation by N-$C^\alpha$ INEPT and $C^\alpha$-$H^\alpha$ INEPT backtransfers.

iH(CA)NCO coherence transfer pathway is presented in detail in Figure 13. Correlated nuclei are presented in short form in Figure 12. The iH(CA)NCO and can be condensed in following

manner:

$$H^{\alpha} \overset{^1J_{HC}}{\rightarrow} C^{\alpha} \xrightarrow{^1J_{C\alpha N},\,^2J_{C\alpha N},\,^1J_{C\alpha C'}} C'\left[T_A - t_2;^1J_{C'N},^1J_{C\alpha N},^2J_{C\alpha N}\right]$$
$$\xrightarrow{^1J_{C\alpha N},\,^2J_{C\alpha N},\,^1J_{C\alpha C'}} N\left[t_1\right] \overset{^1J_{C\alpha N},\,^2J_{C\alpha N}}{\rightarrow} C^{\alpha} \overset{^1J_{HC}}{\rightarrow} H^{\alpha}\left[t_3\right]$$

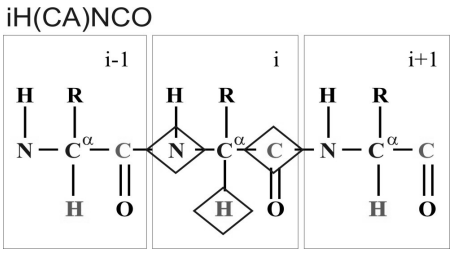


Figure 12. Correlated nuclei in iH(CA)NCO sequence.

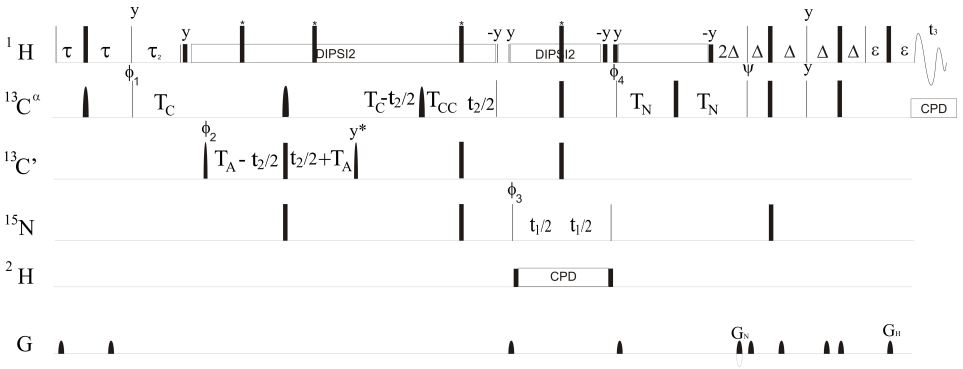


Figure 13. Schematic representation of iH(CA)NCO pulse sequence.

**iHCAN** produces correlation between $H^{\alpha}$(i)-$C^{\alpha}$(i)-N(i), as presented in Figure 14. The sequence begins with $H^{\alpha}$ INEPT. The following segment implements the intra residual filter, which was presented in the section 6.1.2. The coherence at this point is $N_z C^{\alpha}_z$. Next is the $t_1$ labelling with nitrogen frequency. Sequential part is sensitivity- and gradient-enhanced coherence-order selective coherence transfer (COS-CT) (Reiss *et al.*, 2002) incorporated with $t_2$ labelling with $C^{\alpha}$ frequencies. Lastly, the coherence is transferred back to the alpha proton via INEPT backtransfer.

iHCAN transfer scheme described above is presented in detail in Figure 15, and can be presented in a compact form as follows:

$$H^{\alpha}(i) \xrightarrow{^1J_{HC}} C^{\alpha}(i) \xrightarrow{^1J^1_{C\alpha C'}J^2_{C\alpha N}J_{C\alpha N}} C'(i) \rightarrow C^{\alpha}(i) \xrightarrow{^1J^2_{C^{\alpha}N}J^1_{C^{\alpha}N}J_{C'N}} N(i)\left[t_1\right]$$
$$\xrightarrow{^1J^2_{C^{\alpha}N}J_{C^{\alpha}N}} C^{\alpha}(i)\left[T_N - t_2\right] \overset{^1J_{HC}}{\rightarrow} H^{\alpha}(i)\left[t_3\right]$$

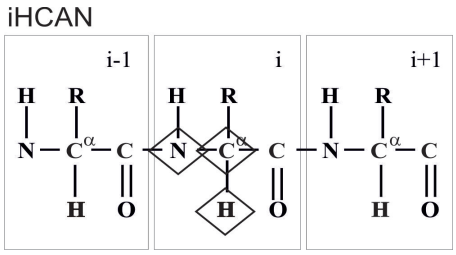


Figure 14. Correlated nuclei in iHCAN sequence.

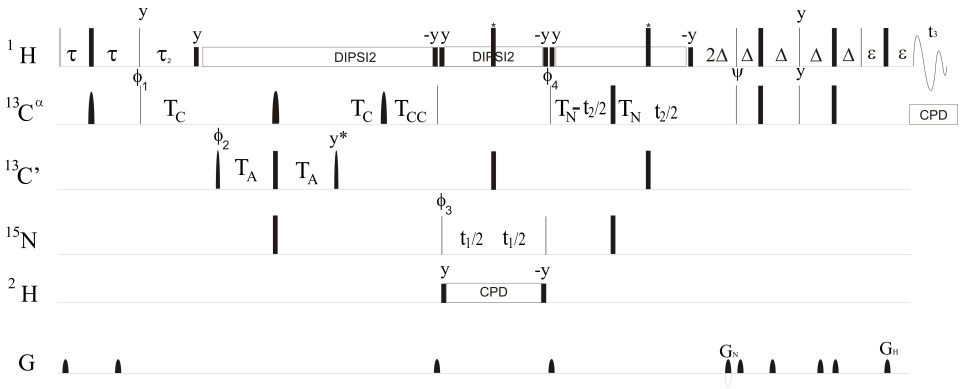


Figure 15. Schematic representation of iHCAN pulse sequence.

**H(CA)CON** produces H(i)-C'(i)-N(i+1) correlation spectra, as seen in Figure 16. The pulse sequence begins with $H^{\alpha}$ INEPT leading into coherence $H^{\alpha}C^{\alpha}$. The sequence continues by additional INEPT-like scheme with the middle $\pi$ pulses replaced by Shaka-6 composite pulse (Shaka, 1985), which refocuses $C^{\alpha}$ and inverts C' spins. Typical resonance frequencies for the $C^{\alpha}$ and $C^{\beta}$ are 57 ppm and 37 ppm, respectively (Kjaergaard and Poulsen, 2011). Offset of a Shaka-6 composite pulse at at 20 ppm is zero (Shaka, 1985). This leaves coherence between $C^{\alpha}$-$C^{\beta}$ to evolve during the INEPT $2T_C$ delay. This coherence must taken into account if the pulse sequence segment is migrated to other pulse sequences. The excitation profile covers serines and threonines rendering their coherence development moot. Next is the transfer to N(i+1), during which $t_1$ labelling is implemented. This is carried out by incremental adding $t_1$ increment time steps during $\pi$ pulses on $C^{\alpha}$ on timepoints $\frac{t_1}{4}$ and $\frac{3t_1}{4}$. Additionally, there is $\pi$ pulse on C' channel in middle of the incrementation delay, which is for refocusing the magnetization and to obtain the actual transfer. Coherence is then returned to alpha proton by INEPT backtransfer protocols.

H(CA)CON coherence transfer pathway is depicted in detail in Figure 17, and can condensed in following manner:

$$H^{\alpha}(i) \xrightarrow{^1J_{HC}} C^{\alpha}(i) \xrightarrow{^1J_{C\alpha C'}} C'(i) \xrightarrow{^1J_{C'N}} N(i+1)\,[t_1]$$
$$\longrightarrow C'(i)\left[2T_A - t_2;^1 J_{C'N}\right] \xrightarrow{^1J_{C\alpha C'}} C^{\alpha}(i) \xrightarrow{^1J_{HC}} H^{\alpha}(i)\,[t_3]$$

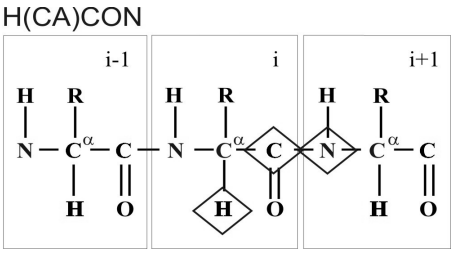


Figure 16. Correlated nuclei in H(CA)CON sequence.

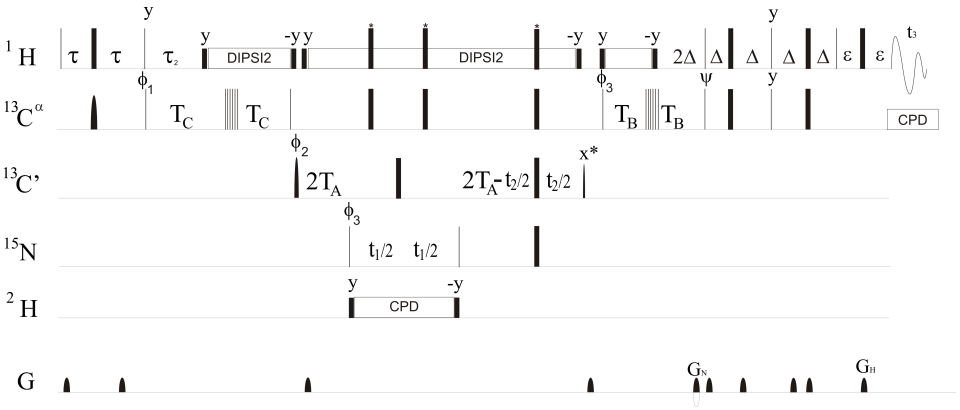


Figure 17. Schematic representation of H(CA)CON pulse sequence.

**H(CA)CON alternative version** is a rework of the version published H(CA)CON pulse sequence (Sayers and Torchia, 2001), correlating the same nuclei as the original and depicted in Figure 18. The original implemented the labelling scheme during out transfer phase, whereas the presented does the labelling during backtransfer phase. The scheme is presented as starting position for the pulse sequence development for the other sequences in the Article II. The pulse sequence is presented in Figure 19.

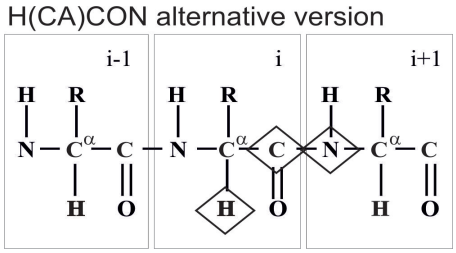


Figure 18. Correlated nuclei in H(CA)CON alternative implementation sequence.

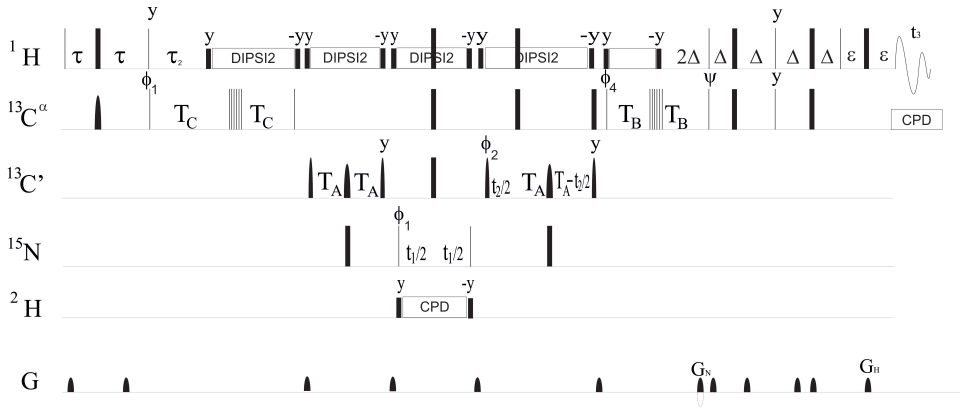


Figure 19. Schematic representation of an alternative version of H(CA)CON pulse sequence.

**Results**

The work produced three novel pulse sequences. Fourth pulse sequence presented in the article is reworked version of the original H(CA)CON pulse sequence presented in literature (Sayers and Torchia, 2001). The three sequences expanded upon prior experiments by molding them to utilize $H^{\alpha}$ approach. The approach yielded notable improvement in signal intensity and resolution when tested against flexible sections of proteins and IDPs. In practice, the assignment coverage was more extensive than that one carried out with established pulse sequences. The sequences enable glycine discrimination via phase distinction. Identification of signals from certain amino acid type may prove powerful in automated assignment (Zawadzka-Kazimierczuk *et al.*, 2012).

Effectiveness of iH(CA)NCO was evaluated against HCAN (Ottiger and Bax, 1997). When measuring on small globular proteins, coherence transfer efficiencies of both sequences were almost the same: 0.22 for iH(CA)NCO and 0.23 for HCAN. However, when measuring on IDPs, the efficiencies were 0.121 and 0.077, indicating a clear advantage of the novel sequence. The effect stems from the predominant $^{13}$C'-$^{13}C^{\alpha}$ transverse relaxation in the proposed pulse sequence. The significant relaxation mechanism causes the sequence to produce poor results when measuring on large globular proteins.

iHCAN diverges from iH(CA)NCO by labelling $C^{\alpha}$ instead of C', yielding signal-to-noise enhancement by factor $\sqrt{2}$. Maximum labelling time in second indirect dimension is 28 ms, enabling superior resolution in $C^{\alpha}$ dimension.

H(CA)CON utilizes the Shaka-6 composite pulse to refocus stemming from $^{1}J_{C^{\alpha}C^{\beta}}$ interaction, causing significant sensitivity improvement. However, serines and glycines fall in the resonance offset of the excitation profile of the pulses, impairing their sensitivity. The tradeoff was recognized, but no feasible workaround was implemented. C' $t_2$ labelling is done in HMQC fashion. Totality of the transfer period $2T_A$ is used for labelling yielding resolution enhancement by factor of two. The approach suffers from high transverse relaxation rates during $t_1$ labelling due to N-C' MQ coherence.

## 6.4 Expanding the $H^{\alpha}$ approach

**Basis**

Pulse sequences presented in Articles I and II provide simple sequential assignment protocol. However, in practice a single assignment pathway is rarely enough. An assignment by linking through only a common resonance easily leads to fragmented assignment chain. Misassign-

ments and incapability to distinguish between signals or signal from background noise are the main sources of assignment errors. In order to solve this problem, the goal was set to establish additional coherence pathways. HNCO pulse sequence (Clubb *et al.*, 1992) was selected for the development starting point. Goal was set to transfer and implement the proven modifications devised in articles I and II.

**Implementation**

Third Article presents a set of two experiments **(HCA)CON(CA)H** and **(HCA)NCO(CA)H**. The experiments continue on the $H^{\alpha}$ approach. The (HCA)CON(CA)H sequence produces correlation between $H^{\alpha}$(i), N(i) and C'(i-1), as depicted in Figure 20. (HCA)NCO(CA)H produces the correlation spectrum between $H^{\alpha}$(i-1), C'(i-1) and N(i), as seen in Figure 22. The spectra are essentially the same experiment, but the coherence transfer pathways are in diverging directions.

**(HCA)CON(CA)H** begins with $H^{\alpha}$ INEPT followed by INEPT-like transfer to C' by utilizing the Shaka-6 composite pulse. This is similar to scheme presented in Article II. Similar to INEPT, the resulting coherence is $C^{\alpha}_{z}C'_{z}$. The following is constant time $t_1$ labelling. Frequencies are labelled while transferring to nitrogen yielding coherence $C^{\alpha}_{z}C'_{z}N_{y}\cos(\Omega_{C'}\ t_1)$. The following is nearly identical constant time $t_2$ incrementation with nitrogen frequency yielding coherence [ $N_zC^{\alpha}_{y}$+ $C^{\alpha}_{y}$(i-1) $N_z$(i)] $\cos(\Omega_{C'}t_1)\cos(\Omega_N t_2)$. The latter coherence is deliberately built to correlate the i-1 $C^{\alpha}$ and i nitrogen. $N_zC^{\alpha}_{y}$ coherence evolves to $H^{\alpha}C^{+}$ and again to $H^{-}$ coherence. This is obtained via selective coherence transfer scheme (Kay *et al.*, 1992). The first pathway yields sequential correlation peak. The latter term ($C^{\alpha}_{y}$(i-1)$N_z$(i)) in turn evolves to $H^{-}$ coherence, which is modulated by $\omega_H$(i-i) instead of $\omega_H$(i). The selection between sequential coherence and autocorrelation coherence is done by establishing transfer delay $T_{CN}$ to either 12.5 ms (1/4 $^1J_{CN}$) or 28 ms (1/2 $^1J_{CN}$). While these numbers do not exactly match theoretical maximums, they are experimentally determined compromises. The transfer period of $2T_{NC}$ has also implemented intraresidual filter, which causes similar phase difference as described in Article I.

The (HCA)CON(CA)H transfer scheme described above is described in Figure 21, and can be presented in compact form as follows:

$$H^{\alpha}(i-1) \overset{^1J_{H^{\alpha}C^{\alpha}}}{\rightarrow} C^{\alpha}(i-1) \overset{^1J_{C^{\alpha}H^{\alpha}},^1J_{C^{\alpha}C'}}{\longrightarrow} C'(i-1)\left[2T_A - t_1;^1 J_{C'N}\right]$$
$$\rightarrow N(i)\left[2T_{NC} - T_2;^1 J_{C'N},^1 J_{C^{\alpha}N},^2 J_{C^{\alpha}N}\right] \rightarrow C^{\alpha}(i)$$
$$\overset{^1J_{C^{\alpha}H^{\alpha}},^1J_{C^{\alpha}N}}{\longrightarrow} H^{\alpha}(i)\,[t_3]$$

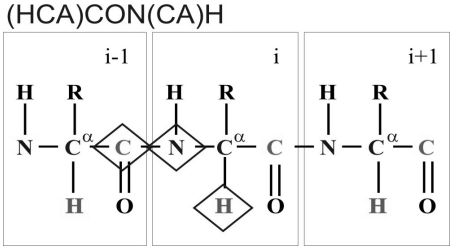


Figure 20. Correlated nuclei in (HCA)CON(CA)H sequence.

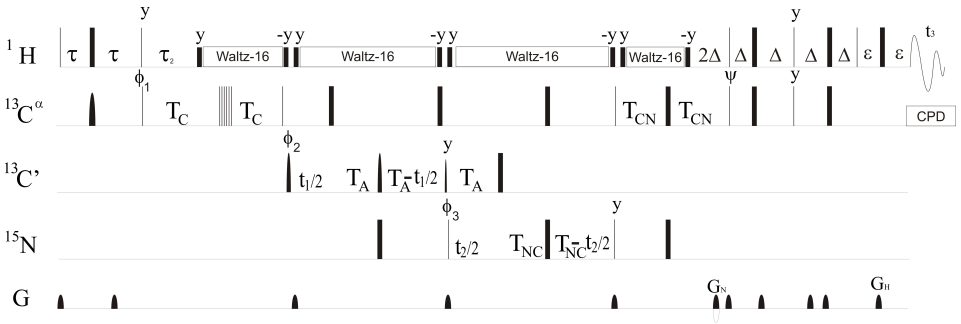


Figure 21. Schematic representation of (HCA)CON(CA)H pulse sequence.

**(HCA)NCO(CA)H** sequence is conceptually a mirror image of (HCA)CON(CA)H. The sequence begins with $H^{\alpha}$ INEPT followed by C'-N INEPT transferring coherence to nitrogen in either the same residue or following, depending on selected transfer delay (T=1/4$^{2}J_{C^{\alpha}N}$ or T=1/4 $^{1}J_{C^{\alpha}N}$). What follows, is constant time $t_2$ labelling by nitrogen frequencies. The scheme is in same form as presented prior in Equation 14 leading into coherence $N_x(i)C'_zC^{\alpha}_z\cos(\omega_N t_2)$. Next is constant time $t_1$ labelling in carbon dimension. The labelling scheme is again similar to previous labelling pattern. Coherence is returned to proton for observation via two step transfer. The $C^{\alpha}C'$ INEPT-type implemented with Shaka-6 composite pulse and standard INEPT backtransfer scheme.

(HCA)NCO(CA)H is described in Figure 23, and can be presented in compact form as:

$$H^{\alpha}(i) \xrightarrow{^{1}J_{H^{\alpha}C^{\alpha}}} C^{\alpha}(i) \xrightarrow{^{1}J_{C^{\alpha}H^{\alpha}},^{1}J_{C^{\alpha}N}} N(i)\left[2T_{NC}-t_2;^{1}J_{C'N};^{2}J_{C^{\alpha}N}\right]$$
$$\rightarrow C'(i-1)\left[2T_A-T_1;^{1}J_{C'N}\right]\rightarrow C^{\alpha}(i-1)$$
$$\xrightarrow{^{1}J_{C^{\alpha}H^{\alpha}},^{1}J_{C^{\alpha}C'}} H^{\alpha}(i-1)\,[t_3]$$

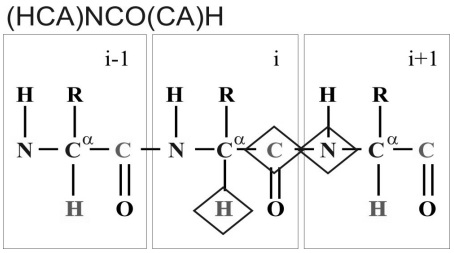


Figure 22. Correlated nuclei in (HCA)NCO(CA)H sequence.

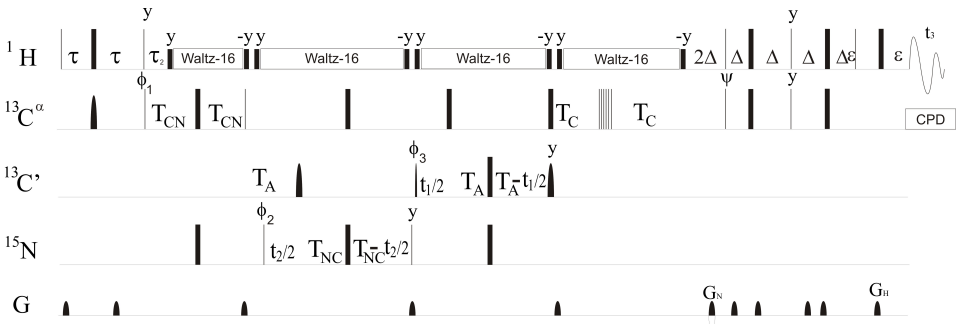


Figure 23. Schematic representation of (HCA)NCO(CA)H pulse sequence.

**Results**

Article III introduces two pulse sequences (HCA)CON(CA)H and (HCA)NCO(CA)H, which are both based on the previously introduced $H^{\alpha}$ approach. The article demonstrates the effects in length and essentially answers and takes advantage of the properties of IDPs, which are described in the previous section. (HCA)CON(CA)H establishes correlation spectrum between $H^{\alpha}$(i)-$^{15}$N(i)-$^{13}$C'(i-1). As for (HCA)NCO(CA)H, it establishes correlation spectrum between $H^{\alpha}$(i)-$^{13}$C'(i)-$^{15}$N(i+1). In familiar manner, glycines exhibit phase difference of $\pi$ in respect to signals from other residues.

Calculations show (HCA)CON(CA)H to yield 10% overall sensitivity enhancement when compared to H(CA)NCO (Yamazaki *et al.*, 1997). In practice, as the molecule size grows, the (HCA)CON(CA)H shows even more significant sensitivity enhancement over the H(CA)NCO. When measuring IDPs the proposed sequence provides resolution enhancement by factor of 2. The transfer efficiency of (HCA)CON(CA)H to be only 0.18, which is worse than that of H(CA)CON's 0.29. However, at higher field strengths the CSA causes H(CA)CON lose signal during long transitions from $^{13}$C' to $^{15}$N and back.

Combination of the presented pulse sequences intertwine, forming a logical continuum. The development cycle began with pulse sequences included in most assignment protocols. Modifications were designed to meet the challenges set by the properties of IDPs. Resulting pulse sequences enable full backbone assignment based solely on pulse sequences utilizing $H^{\alpha}$. Additionally, i(HCA)CO(CA)HN overlaps with $H^{\alpha}$ pulse sequences, enabling amide proton assignment. The presented pulse sequences work with established sequences forming secondary pathways for assignment purposes, as well as ways to assigning $H^{\alpha}$ signals. The simplest form of walkthrough logic can be traced through C'(i) resonance, which is common for both iH(CA)NCO and (HCA)CON(CA)H. Sequential walk options are not limited to common C', but additional overlapping nuclei are available as seen in Figure 25. Assignment pathway is depicted in Figure 24.

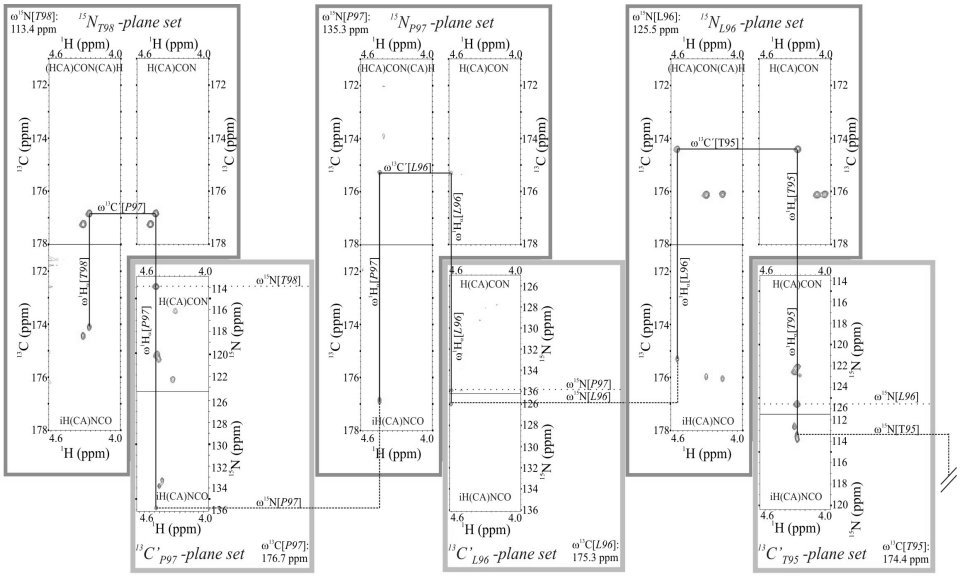


Figure 24. Assignment protocol presented in schematic form.

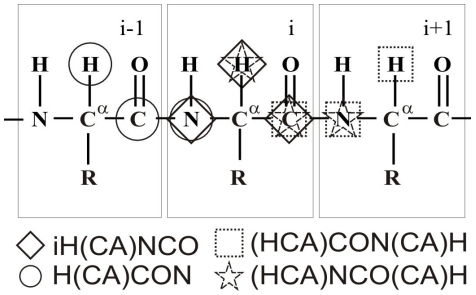


Figure 25. Demonstration of overlapping resonances on a number of pulse sequences.

Only practical aspects, like relaxation, noise and transfer efficiency, limit the number of labelling dimensions. Resolution and, hence, assignment process benefit from additional labelling dimensions. Signal overlap decreases as additional dimensions are added to correlation spectrum. Improved signal separation enables more complicated pulse sequences. In theory, expanding (HCA)CON(CA)H experiment into 4D would be straightforward. Hypothetical (HCA)CONCAH would require only editing constant time labelling scheme during $2T_{CN}$ delay when the coherence is transferred to $C^{\alpha}$. Coherence transfer pathway discrimination should also be included, but this would not be an issue. Each added dimension multiplies the total experiment time by a factor, which is the function of the measurements points in that dimension. In practice the experimental dimensions are restricted to three. Even when measuring IDPs in which relaxation is less of an issue, 4D experiments require just too much time. A possible solution to time required by additional dimensions would be non-uniform sampling (NUS) (Marvasti, 2001).

**Non-uniform sampling**

The NMR correlation spectrum is constructed from individual FID spectra. Each of the FIDs are recorded with a unique combination of incremental timings. From these, one is able to carry out discrete multidimensional Fourier transformation as described in Equation 15. In practice, multiple additional spectra are recorded to obtain phase sensitivity and to improve signal to noise ratio, but these are not relevant for the purposes of NUS theory.

$$H(n_1, ..., n_l) = \sum_{k_1=0}^{N_1-1} ... \sum_{k_l=0}^{N_l-1} h(k_1, ..., k_l) \left[ e^{-2\pi i k_1/N_1} ... e^{-2\pi i k_l/N_l} \right] \tag{15}$$

Magnetization requires long time to relax, thus, reducing time points in directly detected dimension does not reduce the time required to record a transient. However, the number of measurement points in indirect dimensions directly affects the experiment time. Typical 3D NMR experiments require 10-24 hours to run. Adding a dimension would multiply the required time to the scale of week(s). Clearly, this is not a feasible approach.

Simple approach in collecting frequency data is to collect oversampled data set. While this is a tried and proven method, it is inefficient. A way to improve the efficiency is to reduce the sampling points in non-linear manner. This technique is commonly known as non-uniform sampling (Marvasti, 2001). NUS technique produces artifacts, which are function of the sampling function. Each signal from the sample generates additional artifact signal, which is function of the sampling function. In addition to the number of signals, the artifact signal function is proportional to signal amplitude of the peak.

NUS is not feasible when measuring spectra, which display wide dynamic range of peak intensities, like NOESY, for example. However, the artifacts can be calculated and compensated in the after-processing, according to convolution theorem as well as other ways (Press *et al.*, 2007; Kazimierczuk *et al.*, 2007). This takes post-processing time and likely is not able to remove all the noise, but has proven to be useful in discriminating between signals.

## 6.5 Improvement of structure refinement

### Basis

Spatial restraints within a protein are typically determined via NOESY spectrum. The distances are usually determined between protons, but other nuclei may be utilized as well. However, gyromagnetic difference between isotopes renders protons to be superior choice compared to others. Conformation can be calculated from assigned NOE signals, but just as in every experiment, there exists a margin of error. Structure refinement is done to reduce the uncertainty. A way to refine the structure is include J coupling and residual dipolar coupling data to structure calculation (Tjandra *et al.*, 1997). In case of very small proteins full structure has been determined solely from J couplings and RDCs (Andrec *et al.*, 2001).

J couplings and furthermore residual dipolar couplings are easily measured by established methodology (Wang *et al.*, 1998), HNCO-TROSY or similar, for example. However, these approaches suffer similarly when measuring IDPs for the same reasons as pulse sequences aimed at the backbone assignment. The fourth Article acknowledged this deficiency and was aimed at solving the challenge by improving labelling scheme in an established pulse sequence.

### Implementation

**MQ-HNCO-TROSY** is based on HNCO-TROSY pulse sequence (Yang *et al.*, 1999). Just as the original sequence, the novel sequence was designed to select either TROSY (Nietlispach, 2005) or anti-TROSY component (Yang and Kay, 1999). Overlaying the components resulting in double resonance peaks in nitrogen ($t_2$) dimension, enabling the measurement of the $(J+D)_{NH}$ couplings between amide and amide proton. The original pulse sequence is designed in through-and-back format. However, it executes the $t_2$ labelling during backtransfer segment. In the MQ-HNCO-TROSY the $T_2$ labelling is executed during both transfer schemes. This essentially utilizes the full transfer times $2T_N$ for labelling.

The novel sequence is built with coefficient $\kappa$, which controls the $t_2$ incrementation step length. Utilizing this approach requires post-processing to obtain comparable results as measured split is function of $\kappa$. The factor enables the MQ-HNCO-TROSY to be converted to HNCO-TROSY by selecting $\kappa$=0 and inserting additional block $\kappa_2 t_2$ with refocusing $\pi$ pulses in the middle.

The sequence begins with standard INEPT followed by coherence transfer to C' in previous residue, during which the first part of the $t_2$ labelling is performed in constant time fashion. The following is the labelling in C' dimension with refocusing pulses in the middle of $t_1$, to refocus N chemical shift evolution and two refocusing pulses in $C^{\alpha}$ frequency. During the backtransfer period $2T_N$, similar labelling is performed as in outward transfer scheme: in constant time fashion with refocusing pulse after $2T_N$-$t_2/2$. Lastly, the coherence is moved to proton via TROSY backtransfer scheme.

MQ-HNCO-TROSY is presented in Figure 26, and can be presented in compact form as follows:

$$\begin{aligned} &H^N(i) \xrightarrow{^1J_{HN}} N(i) \xrightarrow{^1J_{NC'}} C'(i-1)\left[2T_N - \kappa t_2/2;^1 J_{NH}\right] \longrightarrow \\ &C'(i-1)\left[t_1; J_{C'N}\right] \rightarrow N(i)\left[2T_N - \kappa t_2/2;^1 J_{NH}\right] \xrightarrow{^1J_{NC'}} N^N\left[t_3\right] \end{aligned} \tag{16}$$

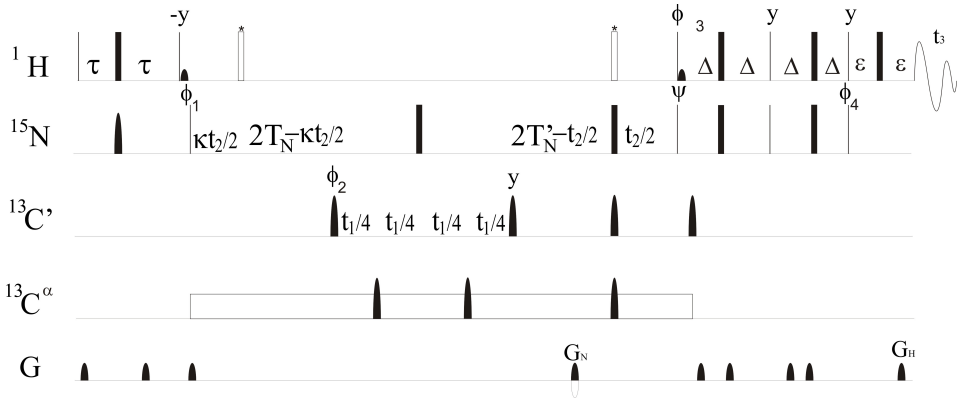


Figure 26. Schematic representation of MQ-HNCO-TROSY pulse sequence.

**Results**

Article IV presents an improved way to measure J couplings. The MQ-HNCO-TROSY experiment yields a line width improvement compared to HNCO-TROSY sequence. The improvement is $e^{(1+\kappa)R_{2T}t_2} > 1$, where $\kappa$ is $\geq 0$. $\kappa$ determines modulation on the split in chemical shift between TROSY and anti-TROSY peaks. TROSY peak can be observed at frequency $\omega_N$-$\pi$ $J_{NH}$ and the anti-TROSY peak at frequency frequency $\omega_N$+$\kappa\pi$ $J_{NH}$. Resulting split must be readjusted in order to gain comparable results. Maximal width improvement is obtained by utilizing full $2T_N$ transfer periods for labelling while NC' coupling develops. In the experimental setup, resolution in the nitrogen dimension was improved by the factor of two.

Clear-cut advantage for the presented MQ-HNCO-TROSY is the total experiment length. Compared to HR-TROSY HNCO (Hu *et al.*, 2009) the presented scheme requires less time than the counterpart. HR-TROSY-HNCO sequence requires more extensive phase cycling and, thus, time consumption clearly favors MQ-HNCO-TROSY version. However, utilizing large $\kappa$ is not without drawbacks, as the random error scales along with the $\kappa$. This leads to a tradeoff between precision and scaling random error. Literature points the precision to be a function: $\Delta J_i = \frac{LW_i}{SN_i} = \frac{\kappa_j+1}{\kappa_i+1}\Delta J_j$ (Yang and Kay, 1999; Bax *et al.*, 2001).

Article IV presents a sequence focused on structure refinement via determination of J couplings and consequently RDCs. The effects of fast transverse relaxation typical in anti-TROSY transition are reduced in the presented scheme. The resulting sequence provides superior line width as well as resolution.

Overall, this thesis is concentrating on the improvement of the NMR methodology on studying IDPs. This is done by proposing novel pulse sequences and eventually assignment protocol for IDP backbone assignment. Basis for the works was a recognition of the limitations of established $H^N$-based sequences when studying mobile, proline rich and unstructured protein segments. Improvements implemented into the sequences include glycine-specific phase discrimination, $H^\alpha$-based approach, and improved labelling techniques. Additionally, the last work improves structure refinement with improved labelling scheme. The works demonstrate strength by outputting improved signal-to-noise ratio and resolution, which is shown through extensive assignment ratios. The sequences have been designed with IDPs in mind, there is nothing hindering utilization in globular proteins or protein like structures. For assessment of the significance of the presented sequences see the Article in question.

# 7 Conclusion

When it comes to a mere structure determination of biomolecules, NMR cannot excel over X-ray crystallography. This is apparent from the fact that 87.7% of all the structures in Protein Data Bank (RCSB, b) has been determined by X-ray crystallography. Each method has its strongpoints. Combination of multiple techniques would be ideal, but lamentably not too common. While NMR cannot compete with X-ray crystallography in either the speed of determination or in the size of structures, the NMR method can address protein dynamics. Capability to observe dynamics is especially important when studying IDPs and IDSs, whose defining factor is motion. This work has focused on improving the NMR methods to study intrinsically disordered proteins.

The IDPs have turned out to be an important category of proteins. They have been shown to be involved in many processes in the eukaryotes as well as in the prokaryotes (Pancsa and Tompa, 2012). In traditional respect, IDPs are not functional proteins, since they adopt functional forms only when provided with binding partner. However, their importance has been established many times over. Structural properties of the IDPs, cause the established protocols for structure determination to fail, or are inept in sensitivity and resolution.

This thesis presents a number of NMR pulse sequences with a view to to improve the method to study highly mobile structures. The ensemble of the pulse sequences form an improved protein backbone assignment protocol. Additionally, a sequence is focusing on structure refinement. We have worked to alleviate signal overlap, improve on proline and glycine detection, and to reduce signal loss. The implemented elements can be incorporated to other sequences. However, as with all pulse sequence segments, some tailoring must be done. For example, the labelling element in Article IV is directly transferable to other sequences.

While the proposed family of pulse sequences can be applied to globular proteins, the sensitivity and resolution cannot compete with established sequences. However, the presented pulse sequences excel over the established ones when measuring IDPs. Each sequence was tested in practice to reproduce the assignment of known IDPs.

Increasing magnetic field strength and improving instrumentation as well as software enable longer and more complicated NMR experiments. One way to improve the existing protocols is to incorporate sequences, which correlate nuclei in additional dimensions beyond the standard three. Protocols incorporating four dimensions are already a reality (Bermel *et al.*, 2012). However, this is not a viable option in every case as the sample stability and fast relaxation of NMR signal restrict the pulse sequence length. The spectral crowdedness is limiting also studies of globular proteins. This limitation can be lifted by expressing globular sections separately or even by more sophisticated methods (Øemig *et al.*, 2009).

Obviously, it requires time and effort to master modern research methods. Accordingly, one is inclined to specialize on particular methods. While this is necessary to produce cutting edge results, the specialization also limits the view on possibilities of study and verification. Therefore, I hope that combination of multiple techniques becomes more common approach to solve problems in structural biology. For example, an overall structure determination could be done with one effective method and thereafter dynamics and function could be studied with more sophisticated method. Multidisciplinary inquiry is often encouraged but in practice it is rarely achieved.

Biosciences continue to develop in both applications and understanding of processes in biosystems (Ledford, 2011; Eiraku *et al.*, 2012; McNeish, 2004; Chena *et al.*, 2012; Novarino *et al.*, 2012). The methodological progress is a part of this process. Desire to rationally engineer protein structure and activity is not a new one (Richardson and Richardson, 1989). Improved experimental methods are an easy answer to meet the ever rising demands on product quality and process safety. We must be able to follow, study and eventually validate proposed treatments and reaction pathways. This work expands the possibilities of study the biomolecules and thus enables more complex reaction pathways. I hope this work will contribute to the under-

standing of the nature of biomolecules.

# 8 Acknowledgments

This work has been carried out in the institute of biotechnology in the University of Helsinki. I wish to acknowledge the current and prior directors of the institute, Mart Saarma and Tomi Mäkelä, for facilities and infrastructure to perform the research and studies. Additionally, the former and current leaders of structural biology programs, Mårten Wikström and Sarah Butcher, deserve thanks.

Eminent remark belongs to my supervisor Perttu Permi, the head of the Finnish National Biological NMR Center. He has mentored the work and supervised the progress of my Ph.D studies. I wish to recognize that without him I would not be at this point. No less prestige belong to Prof. Annila for both steering me through the undergraduate studies and introducing the concept of NMR. His contribution to awareness of NMR is paramount in opening the venue to interdisciplinary sciences between biology and physics. Both the gentlemen venture in thankless and underappreciated tasks, yet essential for the progress of the science.

The research group deserves a mention as well. Dr Hellman for providing high quality samples as well as insight into sample preparation. Vytas Raulinaitis for collaboration. Dr Tossavainen for insight into NMR protein assignment. Niemi-Aro for interesting collaboration on details on both software and hardware. Dr Pihlajamaa, Dr Würz, Ahovuo, Kyburz, Aitio, Dr. Maaheimo and Dr Iwai for company, conversation, advice and great many other things.

The work would have not been possible without the financial support of the national doctoral program in Informational and Structural Biology (ISB). Graduate school has provided the necessary framework in which the studies and the work has progressed. The school has made relatively easy to participate in conferences and courses both in Finland and international venues. Author hopes this tradition will re-emerge so that future students might benefit from them as well.

The study has been performed in the University of Helsinki and the Institute of Biotechnology. These institutions have provided infrastructure and possibility to perform the experiments. Former Linux support person Atro Tossavainen deserves thanks in providing the most efficient support I have met so far.

Finally, I wish to extend thanks to the numerous people, who did contribute in more subtle ways, but are too numerous to list. You know your impact and deserve a nod for your contribution. Many thanks.

# References


A Abragam. *Principles of Nuclear Magnetism*. Oxford Science Publications, 1961.

O Aitio, M Hellman, A Kazlauskas, D Vingadassalom, J Leong, K Saksela, and P Permi. Recognition of tandem PxxP motifs as a unique SH3 binding mode triggers pathogen-driven actin assembly. *Proceedings of the National Academy of Sciences of the United States of America*, 50:21743–21748, 2010.

C Amero, P Schanda, A Dura, I Ayala, D Marion, B Franzetti, B Brutscher, and J Boisbouvier. Fast Two-Dimensional NMR Spectroscopy of High Molecular Weight Protein Assemblies. *Journal of the American Chemical Society*, 131:3448–3449, 2008.

M Andrec, P Du, and R Levy. Protein backbone structure determination using only residual dipolar couplings from one ordering medium. *Journal of Biomolecular NMR*, 21:335–347, 2001.

ASLA. Asla biotech ltd. http://www.asla-biotech.com/asla-phage.htm Last accessed: 20/11/2013.

M Avison, D Rothman, E Nadelo, and R Shulman. Detection of human muscle glycogen by natural abundance $^{13}$C NMR. *Proceedings of the National Academy of Sciences of the United States of America*, 85:1634–1636, 1988.

A Bax, D Jong, A Mehlkipf, and J Smidt. Separation of the different orders of NMR multiple-quantum transitions by the use of pulse field gradients. *Chemical Physics Letters*, 69(3): 567–570, 1980.

A Bax, G Kontaxis, and N Tjandra. Dipolar couplings in macromolecular structure determination. *Methods in Enzymology*, 339:127–174, 2001.

W Bermel, I Bertini, I Felli, L Gonnelli, W Koźmiński, A Piai, R Pierattelli, and J Stanek. Speeding up sequence specific assignment of IDPs. *Journal of Biomolecular NMR*, 53:293–301, 2012.

N Bhavesh, S Panchal, and R Hosur. An Efficient High-Throughput Resonance Assignment Procedure for Structural Genomics and Protein Folding Research by NMR. *Biochemistry*, 40 (49):14726–14735, 2001.

F Bloch. Nuclear Induction. *Physical Review*, 70:460–474, 1946.

F Bloch, W Hansen, and M Packard. The Nuclear Induction Experiment. *Physical Review*, 70: 474–485, 1946.

M Bottomley, M Macias, Z Liu, and M Sattler. A novel NMR experiment for the sequential assignment of proline residues and proline stretches in $^{13}$C/$^{15}$N-labeled proteins. *Journal of Biomolecular NMR*, 13:381–385, 1999.

H Boze, T Marlin, D Durand, J Pérez, A Vernhet, F Canon, P Sarni-Manchado, V Cheynier, and B Cabane. Proline-Rich Salivary Proteins Have Extended Conformations. *Biophysical Journal*, 99:656–665, 2010.

J Brehm and W Mullin. *Introduction to the structure of matter*. John Wiley & Sons, Inc., first edition, 1989. Appendix A.

E Brunner. Residual Dipolar Couplings in Protein NMR. *Concepts in Magnetic Resonance*, 13: 238–259, 2001.

A Bundi and K Wüthrich. Use of amide $^{1}$H-nmr titration shifts for studies of polypeptide conformation. *Biopolymers*, 18(2):299–311, 1979.

N Burkoff, C Várnai, S Wells, and D Wild. Exploring the energy landscapes of protein folding simulations with bayesian computation. *Biophysical Journal*, 22(102):878–886, 2012.

V Cabra, R Arreguin, R Vazquez-Duhalt, and A Farres. Effect of temperature and pH on the secondary structure and processes of oligomerization of 19 kDa alpha-zein. *Biochimica et Biophysica Acta*, 1764:1110–1118, 2006.

M Caleo, L Restani, E Vannini, Z Siskova, H Al-Malki, R Morgan, V O'Connor, and H Perry. The Role of Activity in Synaptic Degeneration in a Protein Misfolding Disease, Prion Disease. *PIOS ONE*, 7(7):1–10, 2012.

K Campellone, D Robbins, and J Leong. EspFU is a translocated EHEC effector that interacts with Tir and N-WASP and promotes Nck-independent actin assembly. *Developmental Cell*, 7: 217–228, 2004.

B Carrasco and J García de la Torre. Hydrodynamic Properties of Rigid Particles: Comparison of Different Modeling and Computational Procedures. *Biophysical Journal*, 75:3044–3057, 1999.

J Cavanagh, W Fairbrother, A Palmer, M Rance, and N Skelton. *Protein NMR Spectroscopy -principles and practice*. Elsevir Academic Press, second edition, 2007.

A Chena, D Thomas, L Onga, R Schwartz, T Golubd, and S Bhatia. Humanized mice with ectopic artificial liver tissues. *Proceedings of the National Academy of Sciences of the United States of America*, 108(29):11842–11847, 2012.

R Clubb, V Rhanabal, and G Wagner. A contant-Time Three-Dimensional Triple-Resonance Pulse Scheme to Correlate Intraresidue $^{1}H^{N}$, $^{15}N$ and $^{13}C'$ Chemical Shifts in $^{15}N$-$^{13}C$ Labelled Proteins. *Journal of Magnetic Resonance*, 97:213–217, 1992.

G Cornilescu, F Delaglio, and A Bax. Protein backbone angle restraints from searching a database for chemical shift and sequence homology. *Journal of Biomolecular NMR*, 13:289–302, 1999.

D Curcó, C Michaux, G Roussel, E Tinti, E Perpète, and C Alemán. Stochastic simulation of structural properties of natively unfolded and denatured proteins. *Journal of Molecular Modeling*, 18(9):4503–4516, 2012.

S Das and D Mukhopadhyay. Intrinsically unstructured proteins and neurodegenerative diseases: conformational promiscuity at its best. *International Union of Biochemistry and Molecular Biology Life*, 63(7):478–88, 2011.

S David, D Wagner, L Melton, Y Yan, B Erickson, and R Anderegg. Deuterium exchange of $\alpha$-helices and $\beta$-sheets as monitored by electrospray ionization mass spectrometry. *Protein Science*, 3:1305–1314, 1994.

A Dingley, F Cordier, and S Grxesiek. An Introduction to Hydrogen Bond Scalar Couplings. *Concepts in Magnetic Resonance*, 13:103–127, 2001.

D Donne, J Viles, D Groth, I Melhorn, T James, F Cohen, S Prusiner, P Wright, and H Dyson. Structure of the recombinant full-length hamster prion protein PrP (29-231): The N-terminus is highly flexible. *Proceedings of the National Academy of Sciences of the United States of America*, 94:13452–13457, 1997.

J Drenth. *Principles of Protein X-ray Crystallography*. Springer, third edition, 2007.

A Dunker, E Garner, S Guilliot, P Romero, K Albrecht, J Hart, Z Obradovic, C Kissinger, and J Villafranca. Protein disorder and the evolution of molecular recognition: theory, predictions and observations. *Pacific Symposium on Biocomputing*, 1:473–484, 1998.

A Dunker, J Lawson, C Brown, P Romero, J Oh, C Oldfield, A Campen, C Ratliff, K Hipps, J Ausio, M Nissen, R Reeves, C Kang, C Kissinger, R Bailey, M Griswold, W Chiu, E Garner, and Z Obradovic. Intrinsically Disordered Protein. *Journal of Molecular Graphics & Modelling*, 19:26–59, 2001.

K Dunker, C Brown, and Z Obradovi. Identification and functions of usefully disordered proteins. *Advanced Protein Chemistry*, 62:25–49, 2002.

K Dunker, M Cortese, P Romero, L Iakoucheva, and V Uversky. Flexible nets The roles of intrinsic disorder in protein interaction networks. *FEBS Journal*, 272:5129–5148, 2005.

J Dyson and P Wright. Defining solution conformations of small linear peptides. *Annual Review of Biophysics and Biophysical Chemistry*, 20:519–538, 1991.

J Dyson and P Wright. Equilibrium NMR studies of unfolded and partially folded proteins. *Nature Structural Biology*, 5:499–503, 1998.

M Eiraku, K Watanabe, M Matsuo-Takasaki, M Kawada, S Yonemura, M Matsumura, T Wataya, A Nishiyama, K Muguruma, and Y Sasai. Self-Organized Formation of Polarized Cortical Tissues from ESCs and Its Active Manipulation by Extrinsic Signals. *Cell Stem Cell*, 3(5): 519–532, 2012.

R Eisberg and R Resnick. *Quantum Physics of Atoms, Molecules, Solids, Nuclei and Particles*. John Wiley & Sons, Inc., second edition, 1985.

D Eliezer, P Jennings, H Dyson, and P Wright. Populating the equilibrium molten globule state of apomyoglobin under conditions suitable for structural characterization by NMR. *FEBS Letters*, 417:92–96, 1997.

M Eriksson, T Hard, and L Nilsson. On the pH Dependence of Amide Proton Exchange Rates in Proteins. *Biophysical Journal*, 69:329–339, 1995.

M Espinoza-Fonseca. Thermodynamic Aspects of Coupled Binding and Folding of an Intrinsically Disordered Protein: A Computational Alanine Scanning Study. *Biochemistry*, 48:11332–11334, 2009.

M Espinoza-Fonseca, I Ilizaliturri-Flores, and J Correa-Basurto. Backbone conformational preferences of an intrinsically disordered protein in solution. *Molecular BioSystems*, 8:1798–1805, 2012.

N Faleev, T Demidkina, M Tsvetikova, R Phillips, and I Yamskov. The mechanism of $\alpha$-proton isotope exchange in amino acids catalysed by tyrosine phenol-lyase What is the role of quinonoid intermediates? *European Journal of Biochemistry*, 271:4565–4571, 2004.

F Ferron, S Longhi, B Canard, and D Karlin. A practical overview of protein disorder prediction methods. *Proteins*, 65:1–14, 2006.

A Fersht. *Stucture and mechanism in protein science -a guide to enzyme catalysis and protein folding*. W. H. Freeman and Company, fifth edition, 2003.

A Fink. Protein aggregation: folding aggregates, inclusion bodies and amyloid. *Folding and Design*, 3(1):R9–R23, 1998.

A Fink. Natively unfolded proteins. *Current Opinion in Structural Biology*, 15:35–41, 2005.

E Fischer. Synthese des Traubenzuckers. *Berichte der deutschen chemischen Gesellschaft*, 23 (1):799–805, 1890.

P Folkers, G Clore, P Driscoll, J Dodt, S Kohler, and A Gronenborn. Solution structure of recombinant hirudin and the Lys-47$\rightarrow$ Glu mutant: a nuclear magnetic resonance and hybrid distance geometry-dynamical simulated annealing study. *Biochemistry*, 28:2601–2617, 1989.

L Frydman and J Hardwood. Isotropic Spectra of Half-Integer Quadrupolar Spins from Bidimensional Magic-Angle Spinning NMR. *Journal of the American Chemical Society*, 117:5367–5368, 1995.

M Goldman. *Quantum Description of High-Resolution NMR in Liquids*. Oxford Science Publications, first edition, 1988.

G Goldstein, M Scheid, U Hammerling, D Schlesinger, H Niall, and E Boyse. Isolation of a polypeptide that has lymphocyte-differentiating properties and is probably represented universally in living cells. *Proceedings of the National Academy of Sciences of the United States of America*, 72:11–15, 1975.

D Griffiths. *Introduction to Electrodynamics*. Pearson Education Inc., third edition, 1999.

D Griffiths. *Introduction to Quantum Mechanics*. Pearson Education Inc., second edition, 2005.

L Haataja, T Gurlo, C Huang, and P Butler. Islet amyloid in type 2 diabetes, and the toxic oligomer hypothesis. *Endocrine Reviews*, 29:303–316, 2008.

O Heikkinen, P Permi, H Koskela, J Ylänne, and I Kilpeläinen. 1H, 13C and 15N resonance assignments of the human filamin A tandem immunoglobulin-like domains 16-17 and 18-19. *Biomolecular NMR Assignments*, 3:53–56, 2009a.

O Heikkinen, S Ruskamo, P Konarev, D Svergun, T Iivanainen, S Heikkinen, P Permi, H Koskela, I Kilpeläinen, and J Ylänne. Atomic Structures of Two Novel Immunoglobulin-like Domain Pairs in the Actin Cross-linking Protein Filamin. *The Journal of Biological Chemistry*, 284: 25450–25458, 2009b.

M Hellman, H Tossavainen, P Rappu, J Heino, and P Permi. Characterization of Intrinsically Disordered Prostate Associated Gene (PAGE5) at Single Residue Resolution by NMR Spectroscopy. *PlOS ONE*, 6(11), 2011.

V Higman, J Boyd, L Smith, and C Redfield. Residual dipolar couplings: are multiple independent alignments always possible? *Journal of Biomolecular NMR*, 49(1):53–60, 2011.

K Hu, M Doucleff, and M Clore. Using multiple quantum coherence to increase the $^{15}$N resolution in a three-dimensional TROSY HNCO experiment for accurate PRE and RDC measurements. *Journal of Magnetic Resonance*, 200:173–177, 2009.

A Inamdar, A Chaudhuri, and J O'Donnell. The Protective Effect of Minocycline in a Paraquat-Induced Parkinson's Disease Model in Drosophila is Modified in Altered Genetic Backgrounds. *Parkinson's Disease*, 2012(ID:938528):1–16, 2012.

M Karplus. Contact Electron-Spin Coupling of Nuclear Magnetic Moments. *The Journal of Chemical Physics*, 30(1):11–15, 1958.

B Kay, M Williamson, and M Sudol. The importance of being proline: the interaction of proline-rich motifs in signaling proteins with their cognate domains. *The FASEB Journal*, 14(2):231–241, 2000.

L Kay, P Keifer, and T Saarinen. Pure Absorption Gradient Enhanced Heteronuclear Single Quantum Correlation Spectroscopy with Improved Sensitivity. *Journal of the American Chemical Society*, 114:10663–107665, 1992.

K Kazimierczuk, A Zawadzka, W Koźmiński, and I Zhukov. Lineshapes and artifacts in Multidimensional Fourier Transform of arbitrary sampled NMR data sets. *Journal of Magnetic Resonance*, 188:344–356, 2007.

S Köhler, A Weber, P Howard, W Welte, and M Drescher. The proline-rich domain of TonB possesses an extended polyproline II-like conformation of sufficient length to span the periplasm of Gram-negative bacteria. *Protein Science*, 19:625–630, 2010.

M Kjaergaard and F Poulsen. Sequence correction of random coil chemical shifts: correlation between neighbor correction factors and changes in the Ramachandran distribution. *Journal of Biomolecular NMR*, 50:157–165, 2011.

N Koga, R Tatsumi-Koga, G Liu, R Xiao, T Acton, G Montelione, and B Baker. Principles for designing ideal protein structures. *Nature*, 491:222–229, 2012.

D Koshland. Application of a theory of enxyme specificity to protein synthesis. *Proceedings of the National Academy of Sciences of the United States of America*, 44(2):98–104, 1958.

R Krishna, D Huang, J Glickson, R Rowan, and R Walter. Amide hydrogen exchange rates of peptides in $H_20$ solution by $^1H$ nuclear magnetic resonance transfer of solvent saturation method conformations of oxytocin and lysine vasopressin in aqueous solution. *Biophysical Journal*, 26:345–366, 1979.

P Kussie, S Gorina, V Marechal, B Elenbaas, J Moreau, L Arnold, and N Pavletich. Structure of the MDM2 oncoprotein bound to the p53 tumor suppressor transactivation domain. *American Association Advancement Science*, 274:948–953, 1996.

H Ledford. Gene reading steps up a gear-Third-generation sequencing machines promise to make their mark one molecule at a time. *Nature*, 470:155, 2011.

S Lee, D Kim, S Lee, J Han, E Cha, J Lim, Y Cho, C Lee, and K Han. Understanding Pre-Structured Motifs (PreSMos) in Intrinsically Unfolded Proteins. *Current Protein & Peptide Science*, 13:34–54, 2012.

S Li, N Goto, K Williams, and C Deber. $\alpha$-Helical, but not $\beta$-sheet, propensity of proline is determined by peptide environment. *Proceedings of the National Academy of Sciences of the United States of America*, 93:6676–6681, 1996.

R Lindinga, J Schymkowitzb, F Rousseaub, F Diellaa, and L Serrano. A Comparative Study of the Relationship Between Protein Structure and $\beta$-Aggregation in Globular and Intrinsically Disordered Proteins. *Journal of Molecular Biology*, 342:345–353, 2004.

R Lipsitz and N Tjandra. Residual Dipolar Couplings in NMR Structure Analysis. *Annual Review of Biophysics and Biomolecular Structure*, 33:387–413, 2004.

F Löhr and H Rüterjans. A new triple-resonance experiment for the sequential assignment of backbone resonances in proteins. *Journal of Biomolecular NMR*, 6(281):189–197, 1995.

P Mach and P Koehl. Capturing protein sequence-structure specificity using computational sequence design. *Proteins*, 81:1556–1570, 2013.

G Makhatadze and P Privalov. On the entropy of protein folding. *Protein Science*, 5:507–510, 1996.

M Mansfield and C O'Sulliva. *Understanding Physics*. John Wiley & Sons, Inc., first edition, 1999.

C Marie, B Hiscox, and K Nash. Characterization of HDEHP-lanthanide complexes formed in a non-polar organic phase using 31P NMR and ESI-MS. *Dalton Transactions*, 21(41):1054–1064, 2012.

D Marion, M Ikura, R Tschudin, and A Bax. Rapid recording of 2D NMR spectra without phase cycling. Application to the study of hydrogen exchange in proteins. *Journal of Magnetic Reconance*, 85:393–399, 1989.

J Marsh and J Forman-Kay. Sequence Determinants of Compaction in Intrinsically Disordered Proteins. *Biophysical Journal*, 98:2383–2390, 2010.

F Marvasti. *Nonuniform Sampling, Theory and Practice*. Kluwer Academic/Plenum Publishers, first edition, 2001.

A Maudsley, A Wokaun, and R Ernst. Coherence transfer echoes. *Chemical Physics Letters*, 55:9–14, 1978.

J McNeish. Embryonic stem cells in drug discovery. *Nature Reviews Drug Discovery*, 3:70–80, 2004.

M Minch. Orientational dependence of vicinal proton-proton NMR coupling constants: The Karplus relationship. *Concepts in Magnetic Resonance*, 6(1):41–56, 1994.

D Minde, Z Anvarian, S Rüdiger, and M Maurice. Messing up disorder: how do missense mutations in the tumor suppressor protein APC lead to cancer? *Molecular Cancer*, 22(10): 101, 2011.

A Mohan, C Oldfield, P Radivojac, V Vacic, M Cortese, K Dunker, and V Uversky. Analysis of molecular recognition features (MoRFs). *Journal of Molecular Biology*, 362:1043–1059, 2006.

G Morris and R Freeman. Enhancement of Nuclear Magnetic Resonance Signals by Polarization Transfer. *Journal of the American Chemical Society*, 101:760–762, 1979.

A Natalello, D Ami, and S Doglia. Fourier Transform Infrared Spectroscopy of Intrinsically Disordered Proteins: Measurement Procedures and Data Analyses. *Methods in Molecular Biology*, 895:229–244, 2012.

C Naumann and P Kuchel. $^{1}$H and $^{13}$C NMR studies of glycine in anisotropic media: Double-quantum transitions and the effects of chiral interactions. *Journal of Magnetic Resonance*, 211:74–79, 2011.

D Nelson and M Cox. *Principles of biochemistry*. W. H. Freeman and Company, fourth edition, 2005.

D Nietlispach. A selective intra-HN(CA)CO experiment for the backbone assignment of deuterated proteins. *Journal of Biomolecular NMR*, 28:131–136, 2004.

D Nietlispach. Suppression of anti-TROSY lines in a sensitivity enhanced gradient selection TROSY scheme. *Journal of Biomolecular NMR*, 31:161–166, 2005.

J Noggle and R Shirmer. *The Nuclear Overhauser Effect, Chemical Applications*. Academic Press, New York, first edition, 1971.

G Novarino, P El-Fishawy, H Kayserili, N Meguid, E Scott, J Schroth, J Silhavy, M Kara, R Khalil, T Ben-Omran, G Ercan-Sencicek, A Hashish, S Sanders, A Gupta, H Hashem, D Matern, S Gabriel, L Sweetman, Y Rahimi, R Harris, M State, and J Gleeson. Mutations in BCKD-kinaseLead to a Potentially Treatable Form of Autism with Epilepsy. *American Association Advancement Science*, 338:394–397, 2012.

C Nylund, P Rappu, E Pakula, A Heino, L Laato, L Elo, P Vihinen, S Pyrhönen, G Owen, H Larjava, M Kallajoki, and J Heino. Melanoma-associated cancer-testis antigen 16 (CT16) regulates the expression of apoptotic and antiapoptotic genes and promotes cell survival. *PlOS ONE*, 7(9):1–12, 2012.

J Øemig, S Aranko, J Djupsjöbacka, K Heinämäki, and H Iwaï. Solution structure of DnaE intein from Nostoc punctiforme: Structural basis for the design of a new split intein suitable for site-specific chemical modification. *FEBS Letters*, 583:1451–1456, 2009.

M Oldstone, A Tishon, H Lewicki, H Dyson, V Feher, N Assa-Munt, and P Wright. Mapping the anatomy of the immunodominant domain of the human immunodeficiency virus gp41 transmembrene protein: peptide conformation analysis using monoclonal antibodies and proton nuclear magnetic resonance spectroscopy. *Journal of Virology*, 65(4):1727–1734, 1991.

E Olejniczak and S Fesik. Two-Dimensional Nuclear Magnetic Resonance Method for Identifying the $H^{\alpha}/C^{\alpha}$ Signals of Amino Acid Residues Preceding Proline. *Journal of the American Chemical Society*, 116:2215–2216, 1994.

M Ottiger and A Bax. An Empirical Correlation between Amide Deuterium Isotope Effects on $^{13}C^{\alpha}$ Chemical Shifts and Protein Backbone Conformation. *Journal of the American Chemical Society*, 119:8070–8075, 1997.

R Pancsa and P Tompa. Structural Disorder in Eukaryotes. *PlOS ONE*, 7(4):1–10, 2012.

P Permi. Intraresidual HNCA: An experiment for correlating only intraresidual backbone resonances. *Journal of Biomolecular NMR*, 23:201–209, 2002.

P Permi and M Hellman. Alpha proton detection based backbone assignment of intrinsically disordered proteins. *Methods in Molecular Biology*, 895:211–226, 2012.

K Pervushin, R Riek, G Wider, and K Wüthrich. 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. *Proceedings of the National Academy of Sciences of the United States of America*, 94:12366–12371, 1997.

E Polverini, G Rangaraj, D Libich, J Boggs, and G Harauz. Binding of the Proline-Rich Segment of Myelin Basic Protein to SH3 Domains: Spectroscopic, Microarray, and Modeling Studies of Ligand Conformation and Effects of Posttranslational Modifications. *Biochemistry*, 47:267–282, 2008.

W Press, S Teukolsky, W Vetterling, and B Flannery. *Numerical recipes -The art of scientific computing*. Cambridge university press, third edition, 2007.

S Primrose, R Twyman, and R Old. *Principles of Gene manipulation*. Blackwell Science Ltd, sixth edition, 2001.

Queen's University, Canada. Chemistry department of chemistry guide. http://www.chem.queensu.ca/facilities/nmr/nmr/webcourse/puzzle3d.htm Last accessed: 4/10/2012.

M Ramìrez-Alvarado, L Serrano, and J Blanco. Conformational analysis of peptides corresponding to all the secondary structure elements of protein L Bl domain: Secondary structure propensities are not conserved in proteins with the same fold. *Protein Science*, 6:6162–174, 1997.

M Rance, G Wagner, O Sørensen, K Wüthrich, and R Ernst. Application of $\omega$1-decoupled 2D correlation spectra to the study of proteins. *Journal of Magnetic Resonance*, 59(2):250–261, 1984.

F Raoa and M Karplus. Protein dynamics investigated by inherent structure analysis. *Proceedings of the National Academy of Sciences of the United States of America*, 107(30): 9152–9157, 2010.

RCSB. Protein Data Bank, a. http://www.rcsb.org/pdb/statistics/holdings.do Last accessed: 06/2/2012.

RCSB. Protein Data Bank, b. http://www.pdb.org/pdb/results /results.do?qrid=FE93976&tabtoshow=Current Last accessed: 21/6/2012.

T Reiss, N Khaneja, and S Glaser. Time-Optimal Coherence-Order-Selective Transfer of In-Phase Coherence in Heteronuclear IS Spin Systems. *Journal of Magnetic Resonance*, 154: 192–195, 2002.

J Richardson and D Richardson. The de novo design of protein structures. *Trends in Biochemical Sciences*, 14:304–309, 1989.

K Roux and K Taylor. AIDS virus envelope spike structure. *Current Opinion in Structural Biology*, 17(2):244–252, 2007.

G Rule and K Hitchens. *Fundamentals of Protein NMR Spectroscopy*. Springer, 2006.

N Russo, D Estrin, M Mart, and A Roitberg. pH-Dependent Conformational Changes in Proteins and Their Effect on Experimental p$K_a$s: The Case of Nitrophorin 4. *PLOS Computational Biology*, 8(11):1–9, 2012.

H Huang A Sarai. Analysis of the relationships between evolvability, thermodynamics, and the functions of intrinsically disordered proteins/regions. *Computational Biology and Chemistry*, 41:51–57, 2012.

M Sattler, J Schleucher, and C Griesingerc. Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. *Progress in Nuclear Magnetic Resonance Spectroscopy*, 34:93–158, 1999.

A Saupe and G Englert. High-resolution nuclear magnetic resonance spectra of orientated molecules. *Physical Review Letters*, 11(10):462–464, 1963.

M Sayer and A Mansingh. *Measurement, Instrumentation and Experiment Design in Physics and Engineering*. Prentice Hall of India, third edition, 2005.

E Sayers and D Torchia. Use of the Carbonyl Chemical Shift to Relieve Degeneracies in Triple-Resonance Assignment Experiments. *Journal of Magnetic Resonance*, 153:246–253, 2001.

P Schanda, E Kupĉe, and B Brutscher. SOFAST-HMQC experiments for recording two-dimensional heteronuclear correlation spectra of proteins within a few seconds. *Journal of Biomolecular NMR*, 33:199–211, 2005.

S Schwarzinger, G Kroon, T Foss, J Chung, P Wright, and H Dyson. Sequence-Dependent Correction of Random Coil NMR Chemical Shifts. *Journal of the American Chemical Society*, 123:2970–2978, 2001.

C Schwieters, J Kuszewski, N Tjandra, and M Clore. The Xplor-NIH NMR molecular structure determination package. *Journal of Magnetic Resonance*, 160:65–73, 2003.

A Shaka. Composite pulses for ultra-broadband spin inversion. *Chemical Physics Letters*, 120 (2):201–205, 1985.

A Shaka, C Lee, and A Pines. Iterative Schemes for bilinear operators; Application To Spin Decoupling. *Journal of Magnetic Resonance*, 77:274–293, 1988.

Shigemi. Shigemi inc. http://www.shigeminmr.com/#!products/vstc2=nmr-sample-tubes Last accessed: 19/11/2013.

B Shoemaker, J Portman, and P Wolynes. Speeding molecular recognition by using the folding funnel: The fly-casting mechanism . *Proceedings of the National Academy of Sciences of the United States of America*, 97(16):8868–8873, 2000.

P Sigler. Acid blobs and negative noodles. *Nature*, 333:210–212, 1988.

U Sjöbring, L Björck, and W Kastern. Streptococcal protein G. Gene structure and protein binding properties. *The Journal of Biological Chemistry*, 5:399–405, 1991.

O Søransen, M Eich, M Levitt, G Bodenhausen, and R Ernst. Product operator formalism for the description of NMR pulse experiments. *Progress in NMR Spectroscopy*, 16:163–192, 1983.

Sparky. - NMR Assignment and Integration Software. T. D. Goddard and D. G. Kneller, SPARKY 3, University of California, San Francisco Last accessed: 18/11/2013.

I Steinberg, W Harrington, A Berger, M Sela, and E Katchalski. The Configurational Changes of Poly-L-proline in Solution. *Journal of the American Chemical Society*, 82:5263–5279, 1960.

N Tjandra, J Omichinski, A Gronenborg, M Clore, and A Bax. Use of dipolar $^{1}$H-$^{15}$N and $^{1}$H-$^{13}$C couplings in the structure determination of magnetically oriented macromolecules in solution. *Nature Structural Biology*, 4:732–738, 1997.

R Tycko. Biomolecular Solid State NMR: Advances in Structural Methodology and Applications to Peptide and Protein Fibrils. *Annual Review of Physical Chemistry*, 52:575–606, 2001.

Unity INOVA. Technical Reference. UNITY INOVA Technical Reference UNITY INOVA NMR Spectrometer Systems Last accessed: 18/11/2013.

V Uversky. Natively unfolded proteins: A point where biology waits for physics. *Protein Science*, 11:739–756, 2002.

V Uversky. Intrinsically Disordered Proteins and Their Environment: Effects of Strong Denaturants, Temperature, pH, Counter Ions, Membranes, Binding Partners, smolytes, and Macromolecular Crowding. *The Protein Journal*, 28:305–325, 2009a.

V Uversky. Intrinsically disordered proteins and their environment: effects of strong denaturants, temperature, pH, counter ions, membranes, binding partners, osmolytes, and macromolecular crowding. *The Protein Journal*, 28:305–325, 2009b.

V Uversky and S Longhi. *Genetically Engineered Polypeptides as a Model of Intrinsically Disordered Fibrillogenic Proteins: Deep UV Resonance Raman Spectroscopic Study*. John Wiley & Sons, Inc., 2010.

V Uversky, A Karnoup, R Khurana, D Segel, S Doniach, and A Fink. Association of partially-folded intermediates of staphylococcal nuclease induces structure and stability. *Protein Science*, 8:161–173, 1999.

VnmrJ. Version 2.1B release Notes. http://www.chem.agilent.com/Library/software/ Public/0199932100a.pdf Last accessed: 18/11/2013.

G Vuister, R Boelens, R Kaptein, M Burgering, and P van Zijl. Gradient-enhanced 3D NOESY-HMQC spectroscopy. *Journal of Biomolecular NMR*, 2:301–305, 1992.

R Wallace. Protein folding disorders: Toward a basic biological paradigm. *Journal of Theoretical Biology*, 267:582–594, 2010.

A Wang, S Grzesiek, R Tschudin, P Lodi, and A Bax. Sequential backbone assignment of isotopically enriched proteins in $D_20$ by deuterium-decoupled HA(CA)N and HA(CACO)N. *Journal of Biomolecular NMR*, 5:376–382, 1995.

Y Wang, J Marquardt, P Wingfield, S Stahl, S Lee-Huang, D Torchia, and A Bax. Simultaneous Measurement of $^{1}$H-$^{15}$N, $^{1}$H-$^{13}$C', and $^{15}$N-$^{13}$C' Dipolar Couplings in a Perdeuterated 30 kDa Protein Dissolved in a Dilute Liquid Crystalline Phase. *Journal of the American Chemical Society*, 120:7385–7386, 1998.

C Weast and M Astle. *CRC handbook of chemistry and physics : a ready-reference book of chemical and physical data*. Boca Raton, Fla. : The CRC Press, 63rd edition, 1982.

S Williams, T Causgrove, R Gilmanshin, K Fang, R Callender, W Woodruff, and B Dyer. Fast Events in Protein Folding: Helix Melting and Formation in a Small Peptide. *Biochemistry*, 35: 691–697, 1996.

Wilmad. Wilmad labglass. https://www.wilmad-labglass.com/Products/535-PP-7/ Last accessed: 19/11/2013.

K Wüthrich. *NMR of proteins and nucleic acids*. John Wiley & Sons, Inc., third edition, 1986.

T Yamazaki, H Tochio, J Furui, S Aimoto, and Y Kyogoku. Assignment of Backbone Resonances for Larger Proteins Using the $^{13}$C-$^{1}$H Coherence of a $^{1}H_{\alpha}$-, $^{2}$H-, $^{13}$C-, and $^{15}$N-Labeled Sample. *Journal of the American Chemical Society*, 119:872–880, 1997.

D Yang and L Kay. Improved $^{1}$HN-detected triple resonance TROSY-based experiments. *Journal of Biomolecular NMR*, 13:3–10, 1999.

D Yang, R Venters, G Mueller, W Choy, and L Kay. TROSY-based HNCO pulse sequences for the measurement of $^{1}$HN-$^{15}$N, $^{15}$N-$^{13}$CO, $^{1}$HN-$^{13}$CO, $^{13}$CO-$^{13}C^{\alpha}$ and $^{1}$HN-$^{13}C^{\alpha}$ dipolar couplings in $^{15}$N, $^{13}$C, $^{2}$H-labeled proteins. *Journal of Biomolecular NMR*, 14:333–343, 1999.

J Yao, H Dyson, and P Wright. Chemical shift dispersion and secondary structure prediction in unfolded and partly-folded proteins. *FEBS Letters*, 419:285–289, 1997.

Q Yi and D Baker. Direct evidence for a two-state protein unfolding transition from hydrogen-deuterium exchange, mass spectrometry, and NMR. *Protein Science*, 5:1060–1066, 1996.

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

P Zeeman. The Effect of Magnetisation on the Nature of Light Emitted by a Substance. *Nature*, 55:347, 1897.

S Zumdahl and S Zumdahl. *Chemistry*. Houghton Mifflin Company, sixth edition, 2003.