Sequential assignment of the intrinsically disordered protein bacterial interleukin receptor 1 with nuclear magnetic resonance spectroscopy

Master's thesis
University of Jyväskylä
Department of Chemistry
June 25, 2018
Santeri Salovaara

Abstract

Significance of intrinsic disorder in biological systems, nuclear magnetism and different methods for its use in protein structure determination have been reviewed. These methods include the use of different detection schemes for nuclear magnetic resonance spectroscopy, observables and their relation to structure, computational methods and sequential assignment procedure. In addition an intrinsically disordered protein, bacterial interleukin receptor 1, related to pathological pathways of periodontitis is covered in in detail. Sequential backbone chemical shift assignment and structural propensity estimation calculations based on chemical shifts of bacterial interleukin receptor 1 have been carried out. Secondary structure estimation confirms that bacterial interleukin receptor 1 is an intrinsically disordered protein with some transient α-helicity in the middle of its primary structure.

Preface

The time spent at the University of Basel during my student exchange opened my eyes to the possibilities of NMR spectroscopy outside the traditional organic structure analysis, which I found quite boring as an undergraduate. After a couple of courses on macromolecular NMR analysis I studied abroad (and later on in Finland), I was hooked to the complexity and the richness of the field. I found the myriad of different possible structural and chemical properties a skilled NMR spectroscopist can unveil from large biomolecules astonishing, and wanted to be a part of the field. This spark of interest found abroad made me turn away from optical spectroscopic methods (in which I was definitely more qualified at the time) when the time came to choose a topic for my master's thesis. "Sometimes you have to travel a long way to find what's near" is definitely a fitting aphorism to describe my academic journey over the past couple of years, as I would readily scoff at NMR based methods before my exchange year, but if I was now asked about the subject, the answer would be very much on the contrary.

The literary part of the thesis was written concurrently with the conducted experimental work from October 2017 to April 2018 at Nanoscience Centre, Department of Chemistry, University of Jyväskylä, under the supervision of Prof. Perttu Permi and Docent Helena Aitio.

I truly enjoyed the writing process of this thesis, and I would like to express my sincere gratitude to both of my supervisors, Perttu and Helena, for their guidance and inspiration throughout the writing process, as well as for providing me the opportunity to delve into this interesting field of research. I'd also like to thank the University of Jyväskylä's Chemistry Department and its staff, professors and other academics for all these years. Final thanks to my family and all my friends for continuous support throughout my life—I wouldn't be much if it weren't for you.

Contents

Used abbreviations

2D	Two-dimensional
3D	Three-dimensional
AA	Amino acid
BilRI	Bacterial interleukin receptor 1
CT	Constant time
eDNA	Extracellular DNA
FID	Free-induction decay
FT	Fourier transform
HOMO	Highest Occupied Molecular Orbital
HSQC	Heteronuclear single quantum coherence
IDP	Intrinsically disordered protein
IDR	Intrinsically disordered region
INEPT	Insensitive nuclei enchanced by polarization transfer
LUMO	Lowest Unoccupied Molecular Orbital
mRNA	Messenger RNA
MTSL	Methanethiosulfonate spin label
ncSPC	Neighbour corrected strucutral propensity calculator
NMR	Nuclear magnetic resonance
PAMP	Pathogen-associated molecular pattern
PRE	Paramagnetic relaxation enhancement
PTM	Post-translational modification
R.F	Radio frequency
RDC	Residual dipolar coupling
S/N	Signal-to-noise
SCS	Secondary chemical shift

Part I

Literary part

Chapter 1

Introduction

Nuclear magnetic resonance (NMR) was first discovered in late 1930's by Isidor Rabi,[1,2] as he detected a drop in LiCl "molecular beam" intensity while varying a magnetic field it passed through. Different manifestations of NMR phenomenon were also reported in ordinary condensed matter (water and paraffin wax) in 1946 by independent groups lead by Edward Purcell[3] and Felix Bloch[4] within one month of each other. The discoveries were ground breaking at the time, and in the years since, NMR spectroscopy has become a standard analysis tool for scientists, and also netted total of five Nobel Prizes.[5] However explosive growth of NMR publications only began around 1992, when pulsed Fourier transform NMR spectroscopy along with multidimensional NMR studies were introduced by one of the Nobel laureates, Richard Ernst.[6,7] These techniques opened an avenue to the study of bigger and more complex molecules, with the first three-dimensional (3D) structure of biological macromolecule determined in solution with NMR spectroscopy.[8] Now for the first time, proteins could be studied in their *natural* folded states in solution, and their dynamics, folding and interactions have been studied at atomic resolution even *in vivo*.[9]

Intrinsically disordered proteins (IDPs) are a newly characterized family of pliable, unstructured and yet biologically active proteins, that have risen to limelight of the 21st century structural biology. Due to their inherent dynamic nature, intrinsically disordered proteins or regions are invisible in electron density maps, produced by X-ray crystallographic methods. This innate flaw of the most widespread structure elucidation method prompted the development of the concept of intrinsic disorder in biological systems at the same time as protein NMR raised its head. As a result, NMR is now the most commonly used method for the study of IDPs.[10]

IDPs aren't traditional proteins in a sense that they don't have a stable secondary or tertiary structure in their native state, but rather exist in different conformational states, rapidly sampling through them. Completely new NMR experiments and techniques have been developed to elucidate the properties of these conformational en-

sembles and the function of disorder in the human body. IDPs have been connected to the most prominent diseases of our time—tumour suppressor p53 and its role in cancer formation,[11] α–synuclein in Parkinson's[12] and tau protein in Alzheimer's[13]—pharmaceutical and healthcare industry alike have understandably risen major interest towards studying these proteins. New IDPs and IDRs are constantly found and their role in pathological pathways studied

The most common oral disease among adults in Finland is periodontitis with 64% of over 30 year olds having some form of the disease.[14] It is caused by couple of different bacteria with one of them being the opportunistic oral pathogen *Aggregatibacter actinomycetemcomitans*. One of its membrane liporoteins, bacterial interleukin receptor 1 (BilRI), was discovered in 2013 by Annamari Paino *et al.*[15] and its function was linked to gingival biofilm formation in periodontitis. Later its intrinsic disorder was confirmed by a follow up study in 2017 by Tuuli Ahlstrand *et al.*[16]

Chapter 2

Proteins and intrinsic disorder

Over a century ago, Emil Fisher[17] observed invertase in beer yeast only being able to hydrolyse α-glucosides but not β-glucosides, whereas the latter was done by a different enzyme called emulsin. This was coined as "lock-and-key" model, where enzyme and substrate have a geometric and steric complementary fit, just like a key and a lock. This model, together with experiments showing loss of both enzymatic activity and ability to crystallize upon treatment of proteins by acid, alkali or urea (denaturation), prompted the success of so-called *structure-defines-function* paradigm for the up coming decades of protein studies. Although some abnormalities to these models were reported[18] and lock-and-key model's inability to account for transition state stabilization during enzymatic processes was criticized, the paradigm held on, thanks to overwhelming data pointing to the contrary.

This changed just at the turn of the millennium when a few papers addressed the concept of intrinsically disordered or "natively unfolded" proteins under their native and functional state.[19–21] These discoveries lead researchers to hypothesize the existence of a broad new class of naturally flexible and yet biologically significant proteins. Nowadays these proteins are called intrinsically disordered proteins (IDPs) and their biological significance has been confirmed through countless studies ever since their discovery. IDPs lack a general stable secondary or tertiary structure and rather exist in collapsed or extended conformational ensembles and sample a range of conformers over time. This conformational heterogeneity results in many of the IDPs to be functionally promiscuous.[22] IDPs are frequently found in protein interaction networks that are superficially associated with signalling,[23] and in fact almost 70% of signalling proteins are predicted to contain an intrinsically disordered region (IDR) of at least 30 amino acid residues long.[24]

Natural disorder prediction algorithms, that have become increasingly abundant over the few past years, predict that structural disorder is present and ample in all species, rather than existing as rare anomalies. The prediction algorithms and on-going re-

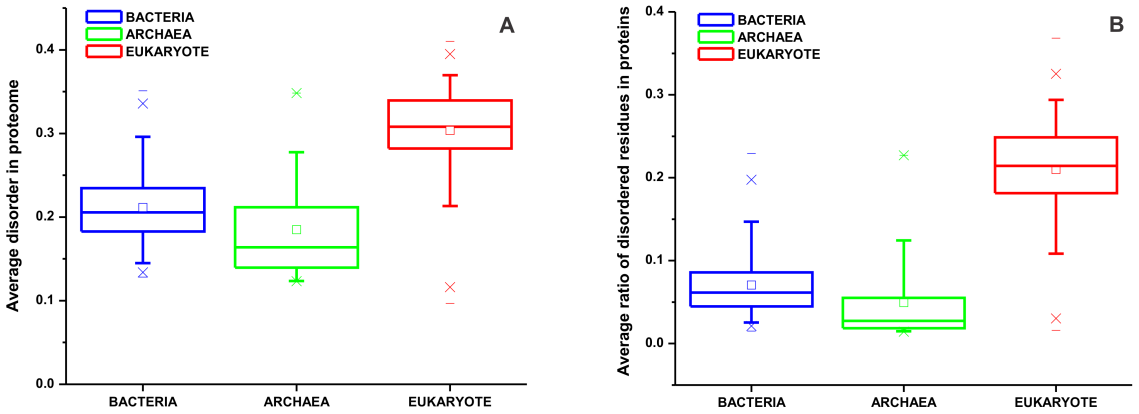


Figure 2.1: Average disorder in proteome (A) and average ratio of disordered residues in proteome (B) by taxonomical domain. Figure taken from the work of Rita Pancsa and Peter Tompa.[26]

search a few years back seemed to suggest a general trend: The more complex the organism, the more its proteome contains IDPs or disordered regions. It was argued that IDPs have also a key role in evolutionary sense as eukaryotes have a more disordered proteome as compared to prokaryotes. By conservative estimates 10–35% of prokaryotic and 15–45% of eukaryotic proteins contain at least 30 residues long disordered regions.[10,25] A comparative prediction study carried out by Rita Pansca and Peter Tompa[26] in 2012 on 194 eukaryotic and 87 prokaryotic proteomes found that the disorder frequency spans over wide range and largely overlaps between the two superkingdoms . The highest levels and variability of predicted disorder was unexpectedly found in single-celled eukaryotes (protists) such as host-changing parasites, more often than not surpassing even the most complex eukaryotes in predicted disorder. It was concluded that the disorder frequency correlates strongly with the lifestyle of the organism rather than its complexity. Additionally due to the significant frequency overlap between the two superkingdoms, no correlation could be drawn between disorder and phylogeny, as was thought in the early studies of IDPs. Pictorial presentation of the average disorder score for all proteins and for ratios of disordered residues averaged for all proteomes within the superkingdom are presented in figure 2.1.

2.1 Conceptual background

2.1.1 Protein structure and folding

Proteins are a diverse and modular class of biomolecules that are synthesized within living cells from 20 natural amino acids. Protein synthesis takes place on ribosomes according to the amino acid sequence encoded into nucleotide bases in DNA. In essence,

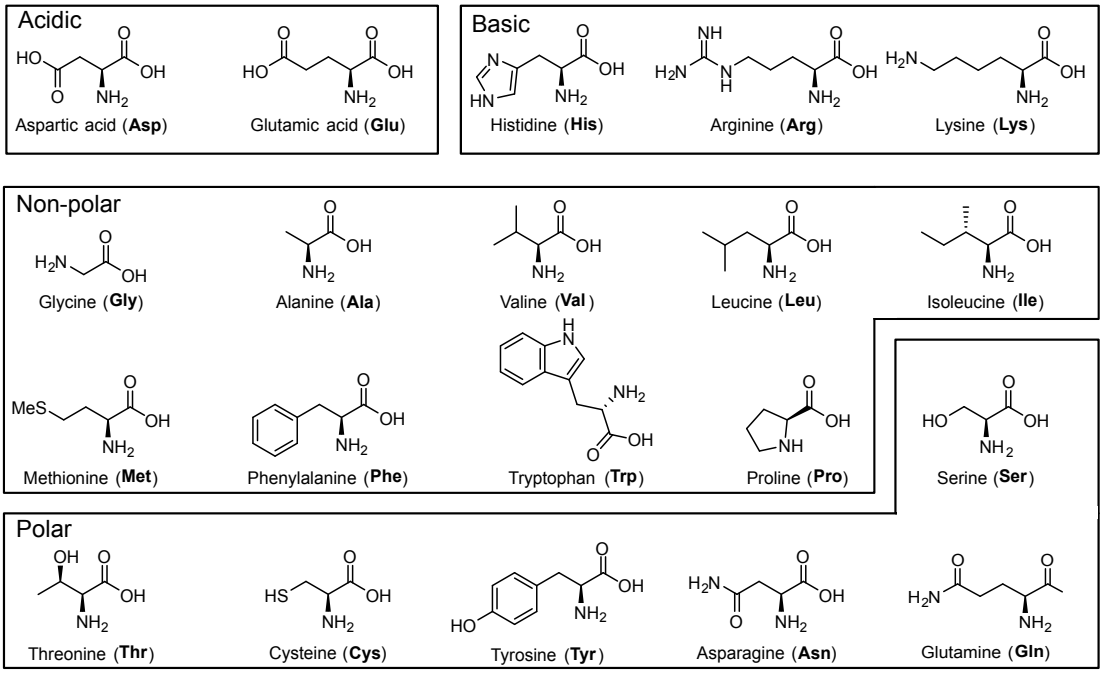


Figure 2.2: 20 natural amino acids grouped by their chemical properties.

they are the expressions of genetic information and agents of biological function.[27] Chemical structures of the common amino acids are listed and grouped by their chemical properties in figure 2.2.

Amino acid sequence of a protein is its primary structure. Amino acids can be categorized in many ways e.g. by their functional group, size, hydrophobicity or even their tendency to express certain secondary structures. In this way a protein's primary structure already yields a lot of information about its regional characteristics, but general folding theory still can't reliably predict a protein's actual three-dimensional (3D) structure *in vivo*, or whether an arbitrary amino acid sequence will fold *a priori*.[29]

Secondary structure of a protein describes local organization of polypeptide chain. Adjacent amino acids are connected by a peptide bond which has a planar orientation due to the double-bond character of the central $N-C$ bond. This leaves the bonds around C_α to be able to rotate. The orientation between two peptide planes can be described by two torsion angles Φ and Ψ. Pairs of these torsion angles form sterically allowed and disallowed regions and a visual representation of these regions is called a Ramachandran plot, which is shown alongside the afore mentioned torsion angles in figure 2.3. Different secondary structures exhibit characteristic torsion angles and the most common ones are α-helix and β-sheet. Ramachandran plot has these characteristic regions marked and is routinely used to determine the secondary structure of proteins.

The actual 3D structure, also known as the fold, arises from intramolecular and solvent interactions such as van der Waals and Coulomb forces and hydrogen bonding. Hydrogen bond is mostly an electrostatic interaction between a covalently bound and

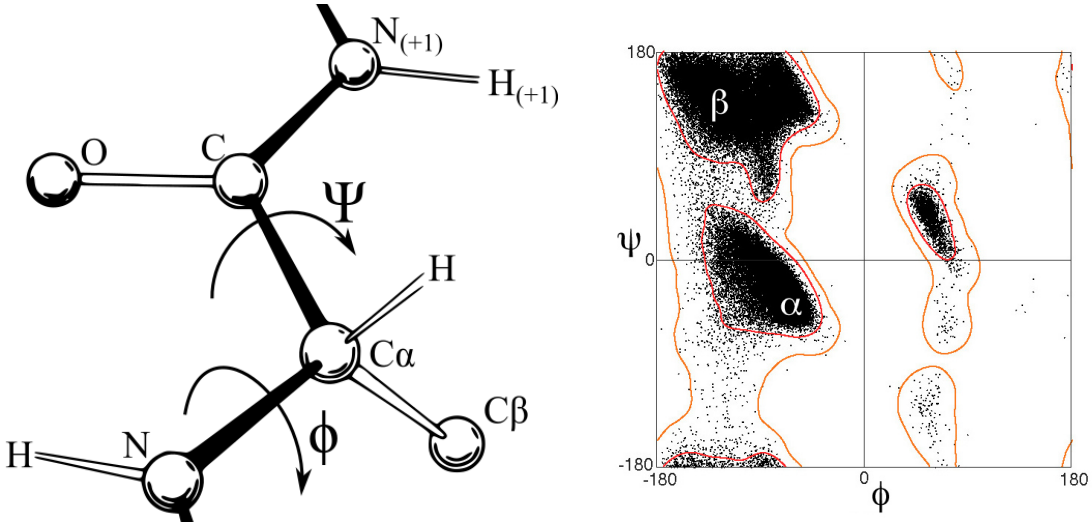


Figure 2.3: On the left is a schematic illustration of the torsion angles of Ψ and Φ. On the right is a Ramachandran plot of 100 000 data points for general amino acid types (no Gly, Pro or pre-Pro) obtained from high-resolution crystal structures with characteristic α-helical and β-strand areas labelled.[30,31]

partially positively charged hydrogen atom and the bond's acceptor atom—with partial covalent character between hydrogen's anti-bonding LUMO and acceptor's lone pair.[32,33] Hydrogen bonds are most often formed between backbone carbonyl oxygen and amide hydrogen, but are also present in some of the side-chains. Hydrogen bond free energy content ranges between 12 to 30 kJ/mol depending on bond angle and length.[27]

2.1.2 Thermodynamic aspects of folding

Thermodynamically folding can be thought as an equilibrium between native (N) and unfolded (U) states through a single energy barrier ΔG°

$$\mathrm{U} \rightleftharpoons \mathrm{N}, \tag{2.1}$$

with an equilibrium constant K

$$K = \exp\left(-\frac{\Delta G^\circ}{RT}\right), \tag{2.2}$$

where ΔG° is Gibb's free energy, T is temperature and R is ideal gas constant. However, no covalent bonds are formed or broken in the folding process 2.1, and only weak interactions are involved.[29] The difference in free energy of the native and unfolded state—measured over hundreds of proteins—ranges from −88 to−264 kJ/mol[34] and can be expressed in enthalpic and entropic terms:[10]

$$\Delta G_{total} = \Delta H_{chain} + \Delta H_{solvent} - T\Delta S_{chain} - T\Delta S_{solvent}. \tag{2.3}$$

Entropy of the chain, ΔS_{chain}, disfavours the highly ordered folded state of globular proteins causing so-called entropy penalty of folding. Other terms of the equation 2.3 depend on the chemical nature (see figure 2.2) of the amino acid residues in the polypeptide chain. Apolar groups have more favourable interactions with water as compared to each other, causing the overall ΔH_{chain} to favour the unfolded state. This is counterbalanced by relatively strong interactions between water molecules that end up begin disturbed in unfolded state, making $\Delta H_{solvent}$ favour the folded state. Solvent entropy, $-T\Delta S_{solvent}$, is generally a large negative term that tips Gibb's free energy to favour the folded state. This is because exposed apolar groups in the unfolded state cause randomly oriented water molecules to assume a more ordered state, resulting in entropy decrease as compared to the folded state. This is called the *hydrophobic effect* and it causes apolar groups to be buried within the interior of proteins. Thus the folded state is usually favoured, as the solvent and chain enthalpies are approximately the same ($\Delta H_{chain} \approx \Delta H_{solvent}$) and the overall entropy is negative ($-\Delta S_{chain} - \Delta S_{solvent} < 0$).[10]

Protein folding can also be described by heat capacity at constant pressure C_p to differentiate the apolar and polar interactions.[35] The most commonly used definitions of the heat capacity at constant pressure are

$$C_p = \left(\frac{\partial H}{\partial T}\right)_p = T\left(\frac{\partial S}{\partial T}\right)_p = -T\left(\frac{\partial^2 G}{\partial T^2}\right)_p, \tag{2.4}$$

which are, respectively, the increase in enthalpy (heat) with temperature, the temperature dependence of the entropy, and the curvature of the free energy with respect to the temperature. Upon the folding process in a two-energy-level system the changes in C_p have some intriguing implications:[36]

1. ΔC_p is positive for apolar and negative for polar solvation. A positive ΔC_p is considered the signature of hydrophobic effect as opposed to negative ΔS.

2. Globular protein unfolding has a positive ΔC_p. The stability versus temperature profile can be deduced to be inverted U-shape from the Gibb's free energy expression in the equation 2.4, meaning that globular proteins have stability maximum, usually close to their normal ambient temperature.

3. At least in protein-DNA interactions, base sequence-specific binding is accompanied with relatively large decrease in C_p while non-specific binding is not.

Depending on the mixture of apolar/polar groups, the folding process may switch from being entropy driven at one temperature to enthalpy driven at another—or favouring

apolar/polar interactions at different temperatures. The folding process is influenced by myriad of variables and their individual significances are still under extensive research, but the main driving forces are either hydrophobic effect[10] or backbone hydrogen bonding.[29]

2.2 Intrinsic disorder in proteins

Intrinsic disorder of a protein or its region arises from physicochemical properties of its primary structure. Typical amino acid sequence of an IDP would include low hydrophobicity content in conjunction with high net charge and low complexity. As folding is a quite delicate process, a couple of would-be-buried charges greatly inhibit or completely prevent the folding process.[37] Disorder-prompting amino acids are: proline, glutamic acid , serine , glutamine, lysine, alanine and glycine.[38]

By comparing internal free energy landscapes of IDPs to normal globular proteins in protein-protein interactions, an intuitive understanding of the folding process and function of IDPs can be easily understood—without going into equilibrium statistical mechanics. Landscape energy of each conformation is described as a function of degrees of freedom, such as dihedral angles. Each conformation of a protein is represented on the multidimensional surface by a set of coordinates in the conformational space. Energy of different close conformers can differ significantly from another, resulting in non-smooth energy landscape. The wider the valleys, the more conformation are similar to the one at the bottom of the valley—the energy minimum.[39]

The native folded state of a globular protein is the global energy minimum state. The native state is not found by random walk, because it would take an astronomical time even for a small polypeptide chain to fold, as noted by Levinthal.[43] Rather, the folding process can be described in terms of a folding funnel (figure 2.4 a), where a vast number of conformations are sampled and most of them are restrained by internal free energy of the system (i.e. enthalpy, solvation free energies etc.). Much fewer conformation are lower in energy, and thus more similar to the native state, and the system drives towards these states in the conformational space. This way the folding is driven to express the global minimum energy state of the polypeptide chain.[39] Even the native state has some dissimilar states around the global minimum and this tiny amount of conformational freedom is necessary for normal functioning of globular proteins.[42] In the case of human nucleoside diphosphate kinase conformational freedom of two IDRs (highlighted in red in figure 2.4 c) in its structure make the energy surface of the native state "rough".

IDPs and IDRs on the other hand exist in a near-flat free energy landscape with many near equivalent states to the global minimum (figure 2.4 b). The energy barriers be-

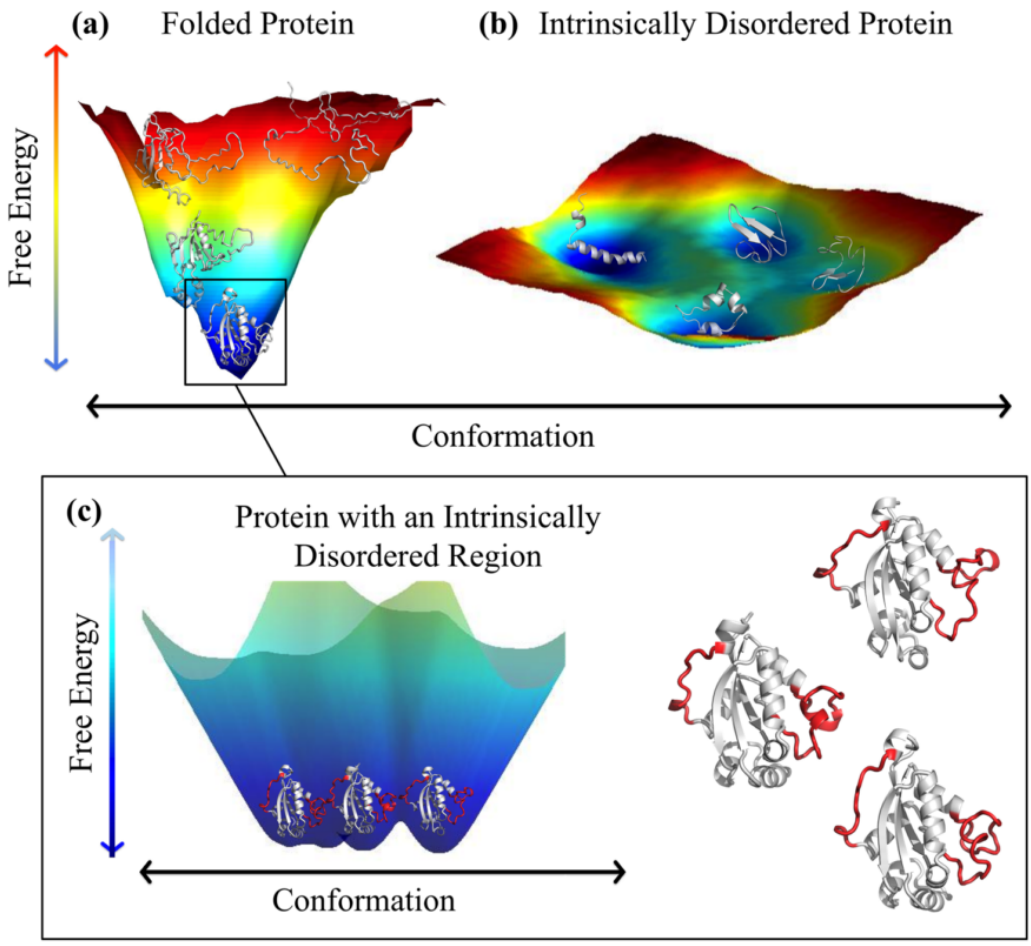


Figure 2.4: Energy landscapes of a folded protein (a)—human nucleoside diphospahte kinase (NDPK)[40]—and an intrinsically disordered polypeptide (b)—CcdaA C-terminal.[41] Close-up of the free energy landscape at energy minimum of IDR in (a) is shown in (c). The most probable conformations are at the lowest energy. Figure taken from the work of Burger *et al.*[42]

tween these dissimilar states is relatively low, and thus different conformations are rapidly probed from kinetic energy alone. This makes comprehensive characterisation of so-called conformational ensemble of these states difficult.[44] In their biological functions, IDPs usually undergo so-called disorder-to-order transition upon binding, where the IDP becomes partially or completely folded to best match the the interface of its partner. The change in enthalpy is compensated by must larger loss of conformational entropy when compared to globular protein binding. This results in lower absolute values of ΔG° for the IDP based interactions and in turn more unstable binding complexes.[39]

However, the small binding free-energy barriers give the IDPs a greater probability to evolve into the final bound state based on the encounter complex. This is confirmed by *in silico* experiments where it was found that only 3–4 encounters are needed for disordered systems to form the bound state compared to 12 encounters for ordered systems.[45] Although binding usually involves the afore described disorder-to-order

Table 2.1: Different classes of IDPs with their function and target protein[48]

Protein	Target	Function
Entropic chain		
Titin PEVK domain	N/A	Entropic spring resulting in contractile force in muscle
Effectors		
Calpastain	Ca^{2+}-activated protease	Inhibitor of calpain in Ca^{2+} signalling
Scavengers		
Caseins	$CaPO_4$	Stabilization of $CaPO_4$ in milk
Display sites		
CREB TAD	Protein kinases	Regulation of phosphorylation
Assemblers		
λ phage N protein	mRNA, RNA Pol II	Translation alteration
Chaperones		
Nucleocapsid protein 7/9[10]	RNA	Trans-splicing

transition, some of the IDPs remain disordered even in the bound state.[46] In order to fully understand biologically relevant IDP based regulatory mechanisms, structural information of the bound state alone isn't sufficient but, rather, the behaviour of the whole conformational ensemble has to be studied.[44]

High-specificity and low-affinity nature of interactions arising from disorder of the IDPs/IDRs allow them to play complementary role to that of ordered proteins. Broadly speaking their functions fall into six classes,[10,37,46,47] and they've been listed alongside with some example proteins and their functions in table 2.1. The first ones are entropic chains that fall outside even partial folding. All of their functions relate to their polypeptide chain fluctuating between a large number of conformational states. They are employed in rearrangement regulation, time measurement and determination of distance distributions of functional elements by cells. The other five—effectors, scavengers, assemblers, display sites and chaperones—function via molecular recognition, and can partner with other proteins, a range of small ligands, or DNA/RNA. Effectors have only been found as inhibitors and they manipulate the activity of a single protein / assembly. Scavengers store and/or neutralize small ligands. Assemblers stabilize and regulate large protein complexes such as ribosomes and extracellular matrices. Display sites regulate post-translational modification of proteins, such as phosphorylation.[48] Chaperones are mainly involved in protein folding for example in assisting nucleosome assembly from folded histones and DNA.[49]

These are the six main functions of the IDPs used by living organisms. Their importance has been found out to be tremendous but further examples will not discussed.

2.2.1 Benefits of disorder

In the upsurge of IDP studies in the 2nd millennium many new IDPs and IDRs were discovered and published. Newly found abundance of disorder in the biosphere raised pertinent questions of their importance, function and origin. It was found that interplay between ordered and disordered proteins is essential for the function of biological systems as a whole. The following benefits are deemed the most significant of the bunch.

Economical use of cellular materials

IDPs exist in more extended structures compared to folded proteins, resulting in more protein-protein interface area per amino acid. For ordered proteins to achieve the same interface area while being stable as monomers, it has been estimated that protein size would need to be 2–3 times larger.[50] This would result in cellular crowding and further excessive energy usage by the organism, which would indeed be counterproductive and even bigger problem in smaller organisms. A trend of this can be seen as some viruses have up to 77.3% of their proteome disordered.

Higher organism utilize other mechanisms to further reduce their genome size. In alternative splicing multiple proteins are expressed from the same gene. This is achieved by having disordered regions in the genome to help avoid structural disruption. Additionally IDPs can carry out multiple functions instead of one further reducing the need for large genome. This type of behaviour was coined "moonlighting" by Jeffery Constance.[51]

Pliability overcomes steric restrictions

Highly flexible chains of IDPs can overcome steric restrictions more easily than ordered proteins by using coupled folding and binding, where IDPs wrap around their partners or penetrate into the their concavities. Thus formed binding modes are highly diverse and create multifarious complexes. This is possible because the IDP/IDR binding regions are usually only 30 to maximum of 100 amino acids long. In addition, multiple mutually exclusive binding regions can be compactly packed into longer IDRs.

One example of these complexes is calpastatin which wraps around and binds to its target, calpain, on three surfaces while two of the remaining regions stay disordered. This type of pliable behaviour results in low affinity, but amazingly rarely to high promiscuousness as would be expected.

The low affinity can be explained by the free energy change of coupled folding-binding

process. It has a higher ΔG_{total} as compared to pure binding counter part, resulting in a higher dissociation constant, as can be understood by equation 2.1. Even though it could be thought that folding of IDP is restricted by entropy, as it to some extent is, IDPs usually have more favourable interface interactions swinging ΔG_{total} to favour folding upon binding.

High specificity on the other hand relates to the interface characteristics of IDPs/IDRs. It is determined by the size match and amino acid complementarity of the binding interface. This way extended structures formed in the IDP complexes achieve large binding interfaces with high specificity which is critical for signal recognition and regulation. More than 75% of globular proteins have binding interface areas of about 500–1500 $Å^2$. While IDPs have about the same area, the binding regions are smaller and cause more relative contribution by amino acid to the protein-protein interaction.[39]

Alleviation of post-translational modification

In post-translational modification (PTM) expressed proteins are modified by different means in order to signal other cellular machinery and regulate protein functions. PTMs range from enzymatic cleavage of peptide bond to formation of covalent bonds with varying groups into specific parts of the expressed protein.[38] There are many different types of PTM reactions, and in most of them conformational flexibility of IDPs/IDRs is exploited (e.g. acetylation,[52] methylation,[53] phosphorylation[54]). Alleviated exposure of target modification sites and their guidance to modifying enzymes are examples of these exploits. PTMs on ordered proteins are mostly restricted by site accessibility, while IDPs enable the above mentioned "moonlighting" to take place. It was shown in large-scale analysis of the Swiss-Prot database that 17 PTMs were strongly equated to predicted disorder.[46]

2.3 Protein of interest BilRI

In the present study sequential chemical shift assignment is done to bacterial interleukin receptor 1 (BilRI). It was first discovered by Paino *et al.*[15] in gram-negative *Aggregatibacter actinomycetemcomitans*, which is associated with biofilm formation in chronic and aggressive periodontitis. It forms thick biofilms by symbiotic coaggregation with other oral bacteria,[55] thus increasing its tolerance for antibiotics and host clearance mechanisms.[56]

BilRI was found to be located in the outer membrane of the bacterial cell and to increase the binding capacity of human interleukin-1β. This is the first time an outer membrane bacterial lipoprotein has been identified to interact with host's IL-1β (Paino

Figure 2.5: Amino acid sequence of BilRI. Four 10 amino acid long repeats are colour coded blue, purple, green and grey. The 19 residues long lipoprotein signal sequence, colour coded in red isn't part of the recombinant form of BilRI.[15]

et al.[15]). Interleukins are a group of cytokines responsible for the function of human immune system. IL-1β plays a key role as inflammatory mediator and is secreted mainly by human macrophages and monocytes upon sensing alien microbial signals.

Various human pathogens bind inflammatory cytokines and change their behaviour accordingly. They might for example form biofilms as described earlier, start virulent gene expression as is the case for *Staphylococcus aureus* upon IL-1β uptake[57] or manipulate host's inflammatory reactions and ultimately enfeeble the defence. However, the uptake and following regulatory pathways are not yet fully understood.

A more recent study made by Ahlstrand *et al.*[16] confirmed the intrinsic disorder of BilRI by proton NMR and amino acid composition, but most importantly by demonstrating moderate binding of multiple ligands. The highest affinity was towards IL-8 instead of the IL-1β, which was previously used to identify the protein. BilRI actually had a relatively weak affinity towards cytokines in general, which was thought to suggest a role as a non-specific cytokine concentrator, rather than as a specific receptor. This was further confirmed by deletion of BilRI in *A. actinomycetemcomitans*, which did not completely inhibit internalization of IL-1β, but rather decreased the uptake efficacy. It was also shown that the presence of BilRI in *A. actinomycetemcomitans* decreased eDNA concentration of the formed biofilms. eDNA plays an important role in early stages of biofilm formation by enhancing adhesion to the surface and stabilizing it at its young stages.[58] This seemingly counterproductive response is argued to impede the innate immune defence, as eDNA also produces pathogen-associated molecular patterns (PAMPs), to which immune system's pattern recognition receptors bind.

Periodontitis causes progressive bone loss in tooth supportive tissues. Extended innate immune response causes skew in bone homeostatis to the direction of bone degradation, due to the prolonged exposure to reactive oxygen species and proteases. IL-6 is known to redirect the immune system form innate to acquired immunity by replacing acute inflammation cells, neutrophils, with monocytes and T cells. As BilRI biofilms exploits the scarcity of PAMPs to inhibit the innate system response, it was argued by Ahlstrand *et al.* that *A. actinomycetemcomitans* could also have mechanisms to decrease IL-6 concentrations and prolong the time of the neutrophil-skewed immune

reaction.

The intrinsic disorder of BilRI was confimed based on proton NMR spectrum showing narrow chemical shift dispersion of amide protons and deshielded methyl protons. Biochemical sequential analysis revealed the low complexity of BilRI's sequence which is typical for IDPs, as well as an absence of aromatic amino acids; phenylalanine, tyrosine and tryptophan. The amino acid sequence is mainly (48%) composed of three amino acids: lysine, aspartate and alanine. In addition, BilRI's sequence comprises of four 10 amino acid long repeats (figure 2.5).

Chapter 3

Nuclear magnetism

At the heart of NMR spectroscopy is nuclear magnetism, which is a manifestation of nuclear spin angular momentum; an intrinsic quantum mechanical property with no classical analogue. It is characterized the by nuclear spin quantum number, I, which is either an integer or half-integer depending on the nuclide (bosons and fermions respectively). A nucleus with spin I has $(2I+1)$ degenerate sub-levels in the absence of external magnetic field. In addition it has the following properties:[59]

1. Spin angular momentum of magnitude $\sqrt{I(I+1)}\hbar$.
2. Azimuthal quantum number $m_l = -I, -I+1, \ldots, I-1, I$
3. Angular momentum along z-axis with magnitude of $I_z = \hbar m_l$

Spin and magnetism are closely linked as nuclei with $I > 0$ are NMR active and also possess a nuclear magnetic moment μ. The Wigner-Eckart theorem[60] requires that μ is proportional to the spin angular momentum I and that they are collinear in the absence of external fields

$$\begin{aligned} \hat{\mu} &= \gamma \hat{I} \\ \mu_z &= \gamma \hbar m_l. \end{aligned} \tag{3.1}$$

The proportionality constant is called gyromagnetic ratio, γ, and it's defined by the ratio of magnetic moment to angular momentum. Gyromagnetic ratios are characteristic to the nucleus in question and thus every nucleus behaves differently in external fields. Most atomic nuclei have positive γ meaning that their nuclear magnetic moment and spin angular momentum are parallel to each other, but for some nuclei said vectors are anti-parallel and thus have negative valued gyromagnetic ratios. Most common nuclei, their ground-state spins, natural abundances and gyromagnetic ratios are listed in table 3.1. The magnetic energy of a nucleus, arising from the interaction between

Table 3.1: Some of the most important NMR active nuclei, their natural abundances[61] and gyromagnetic ratios[59]

Isotope	Ground-state spin	Natural abundance (%)	Gyromagnetic ratio γ (10^6 rads^{-1}T^{-}1)
^{1}H	$^{1}/_{2}$	99.9885	267.552
^{2}H	1	0.0115	41.006
^{13}C	$^{1}/_{2}$	1.07	67.283
^{14}N	1	99.632	19.338
^{17}O	$^{5}/_{2}$	0.038	36.28
^{15}N	$^{1}/_{2}$	0.368	-27.126

its magnetic moment and external magnetic vector field $\vec{B}$, can be expressed with a Hamiltonian

$$\hat{H} = -\hat{\mu} \cdot \vec{B} = -\gamma \vec{B} \cdot \hat{I}. \tag{3.2}$$

In general, particles with spins have their spin angular momentums (also known as spin polarization axes) pointing in all possible directions. When a static external magnetic field B_0 is applied along the z-axis (by convention) of the laboratory coordinate system, spin state energies become

$$E_m = -\gamma \hbar B_0 m_l. \tag{3.3}$$

As implied by the dot product, system is at minimum energy when the magnetic moment is parallel to the $\vec{B}$. From here it can be seen that $2I+1$ quantized Zeeman states $|I,m_l\rangle$ are formed, which corresponds to the relative orientation of the external magnetic field and the spin polarization axis. For $I = \frac{1}{2}$ there are two possible Zeeman eigenstates[59]

$$|\alpha\rangle = |\frac{1}{2}, +\frac{1}{2}\rangle \tag{3.4}$$

$$|\beta\rangle = |\frac{1}{2}, -\frac{1}{2}\rangle. \tag{3.5}$$

An illustration of these eigenstates is presented in figure 3.1. The arrows indicate the spin polarizations, but do not behave like ordinary vectors nor imply that the angular momentum along x- and y- axes are zero. Spin sates $|\alpha\rangle$ and $|\beta\rangle$ are not eigenstates of spin angular momentum operators $\hat{I}_x$ and $\hat{I}_y$, resulting in undefined angular momentum along these axes.[59] From this point onward only nuclei with two Zeeman eigenstates are considered. At equilibrium, the two energy states are unequally populated with the minimum energy state $|\alpha\rangle$ being more probable. The relative populations of the states follow Boltzmann distribution

$$\frac{N_m}{N} = \frac{\exp\left(\frac{-E_m}{k_b T}\right)}{\sum_{m=-I}^{I} \exp\left(\frac{-E_m}{k_b T}\right)} \approx \frac{1}{2I+1}\left(1 + \frac{m_l \hbar \gamma B_0}{k_b T}\right), \tag{3.6}$$

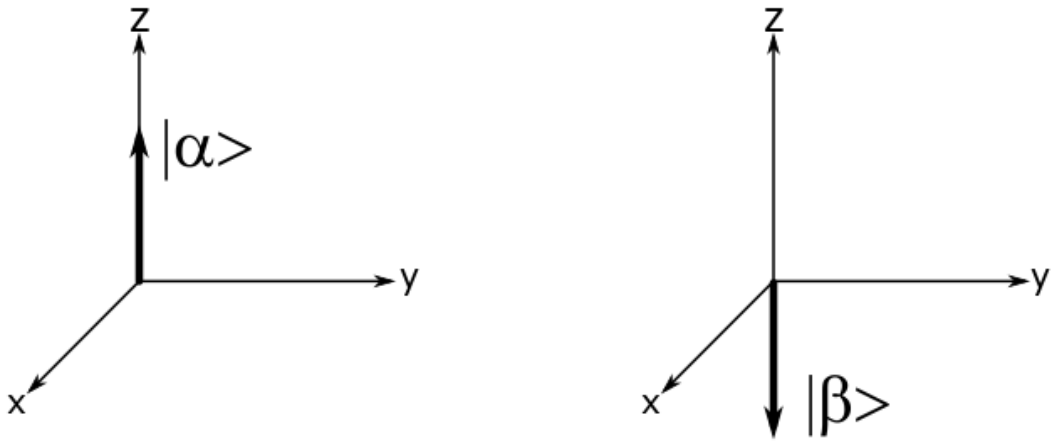


Figure 3.1: The two Zeeman eigenstates of a single 1/2 spin. $|\alpha\rangle$ is called spin-up and $|\beta\rangle$ spin down state.

where N_m is the number of nuclei in the mth state and N is the total number of spins. The approximation in the equation 3.6 is obtained by expanding the exponential functions to first order using Taylor series. The relative populations differ from each other by the term $m_l\hbar\gamma B_0/k_bT$, which is only about 10^{-10} for ^{1}H nuclei in 18 T magnetic field at ambient temperature.[62]

Solution to time-dependent Schrödinger equation shows that spin polarization rotates (precesses) around the applied magnetic field axis with Larmor frequency ω_0, which is characteristic to each nucleus

$$\omega^0 = -\gamma B_0. \tag{3.7}$$

The sign of the Larmor frequency depends on the sign of γ. Negative (positive) Larmor frequency corresponds to clockwise (anticlockwise) spin precession with respect to the direction of the external magnetic field B_0.[59]

Longitudinal relaxation

The initial precession angle of a single spin is only dependent of the initial spin polarization. For most of the spins this is somewhere in between the extremes as they possess an intrinsic angular momentum as well as a magnetic dipole moment. However, due to small fluctuations in local magnetic environments of the precessing spins caused by thermal motion, spin polarization isotropy can be broken, which gives raise to the observable macroscopic magnetic moment. Initially spin magnetic moment precession "cones" sample through all spatial orientations which eventually "wander" to the lowest energy state along B_0. Build-up of the longitudinal magnetization can be expressed with the formula

$$M_z^{nuc}(t) = M_{eq}^{nuc}(1 - \exp(-t/T_1)) \tag{3.8}$$

T_1 is called longitudinal relaxation time constant and it depends on nuclear isotope, temperature and viscosity, in liquid samples.[59]

Transverse relaxation and magnetization

When all of of an ensemble are irradiated with a rotating electromagnetic radiation at their resonant precession frequency, lower energy states can be excited to higher spin states. This allows manipulation of spin state populations and direction of net magnetization. For example flipping of all spins of an ensemble by a so-called 90° pulse, results in the net magnetization lying in the xy-plane. This means that the spin state populations $|\alpha\rangle$ and $|\beta\rangle$ are equivalent, causing there to be no net magnetization along the z-axis. This transverse magnetization rotates in the xy-plane, perpendicular to the external magnetic field at Larmor frequency. Transverse magnetization decays in an exponential fashion, as individual precessing spin polarizations dephase out of coherence due to each spin being exposed to slightly different magnetic field and interactions between them. The macroscopic magnetization components at a time t after a flipping the magnetization into the transverse plane are

$$M_y^{nuc}(t) = -M_{eq}^{nuc} \cos\left(\omega^0 t\right) \exp\left(-t/T_2\right) \tag{3.9}$$

$$M_x^{nuc}(t) = M_{eq}^{nuc} \sin\left(\omega^0 t\right) \exp\left(-t/T_2\right) \tag{3.10}$$

and the overall loss of transverse magnetization is described as

$$M_{x,y}^{nuc}(t) = M_{eq}^{nuc} \exp(-t/T_2). \tag{3.11}$$

Here T_2 is called transverse relaxation time which defines the attainable line width in an NMR spectrum. For small molecules in liquids, T_2 is of the same order of magnitude as T_1 i.e. several seconds, but for large molecules, T_2 is as short as milliseconds.[59] This is because T_2 is linked to the rotational correlation times of the molecule, which in turn depends on molecular weight and shape.[62]

3.1 NMR spectroscopy

The transverse net magnetization of a spin ensemble is very small in magnitude, but detectable and distinguishable, thanks to the well defined Larmor frequencies of the

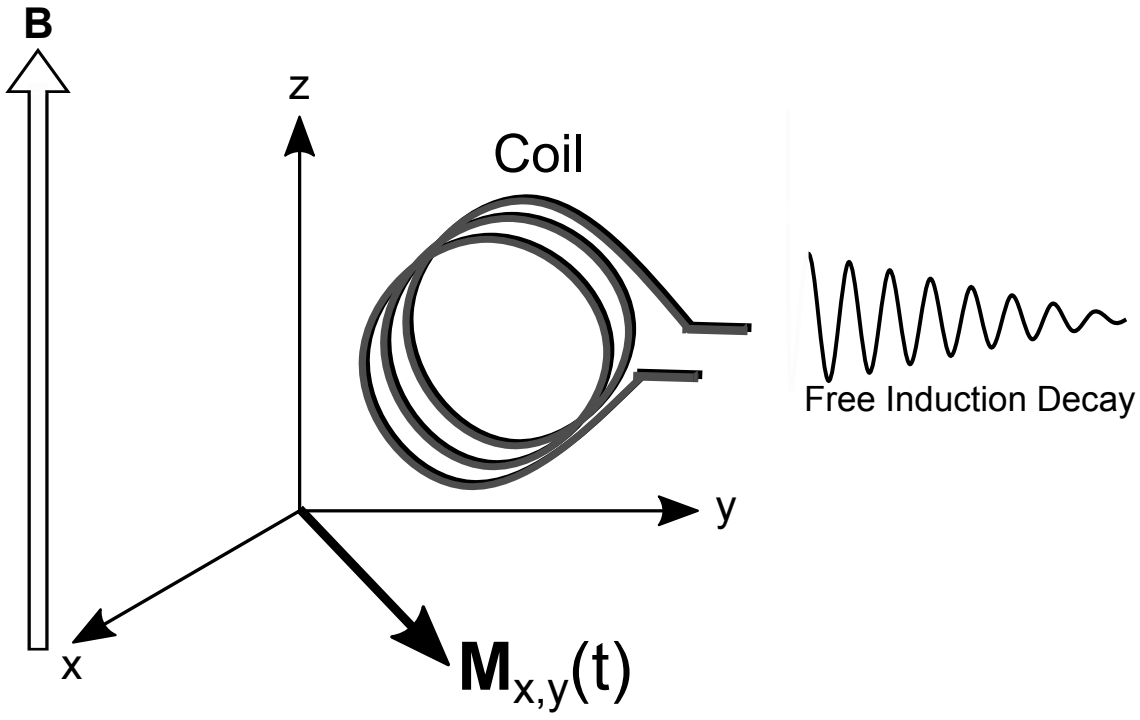


Figure 3.2: Rotating transverse magnetization about the external magnetic field B causing oscillating current and ultimately FID. This figure is purely conceptual and does not reflect the actual sizes of the vectors nor the coil. Figure was drawn with Inkscape.[28]

nuclei under study. Rotating net magnetization of the sample can be used to induce an oscillating current into a wire coil as predicted by Faraday's law of induction. This current oscillates and decays as transverse magnetization relaxes according to equation 3.11. This decay is called free-induction decay (FID), which is the actual signal measured in NMR spectroscopy. It is notable that only transverse magnetization can induce a current to the wire due to the geometry of the NMR spectrometer.[59] The FID and coil geometry are presented in figure 3.2.

3.1.1 Chemical shift

In reality, the observed resonance frequencies differ from the Larmor frequency predicted by equation 3.7, because every nucleus has a different local magnetic environments. This is due to nuclear shielding, which arises from motion of surrounding electrons which induces secondary magnetic fields to each nucleus. The observed resonance frequencies are therefore shifted by a nuclear shielding tensor, which is a 3×3 matrix. In liquids, however, the electrical environment of a particular nucleus is isotropic and an average shielding constant can be used:[62]

$$\omega_{obs} = \gamma(1-\sigma)B_0, \tag{3.12}$$

where σ is the shielding constant. These shifts in resonance frequencies are called *chemical shifts*. To distinguish small shifts in Larmor frequencies, equation 3.7 is subtracted from 3.12 to investigate the spin system in a rotating frame

$$\Omega = \omega_{obs} - \omega^0. \tag{3.13}$$

Now, chemical shift can be thought as precession of magnetization inside the rotating frame with a magnitude Ω. Chemical shift is linearly dependent on the external magnetic field strength and thus gets bigger in stronger magnets. It is convenient to use a field-independent expression for the chemical shift

$$\delta = \frac{\omega_{obs} - \omega_{ref}}{\omega_{ref}} \times 10^6 \tag{3.14}$$

where the chemical shift (δ) is expressed in ppm (parts per million) and ω_{ref} is the Larmor frequency of the nucleus under consideration in a reference compound.[62]

3.1.2 J-coupling

It is possible for precessing spins in NMR active nuclei to influence magnetization of other spins within the same molecule. Direct dipole-dipole coupling, where a magnetic field of one nucleus directly affects another's, is eliminated in liquids due to molecular tumbling, but *indirect dipole-dipole coupling*, also known as J-coupling, is present even in isotropic liquids. This type of coupling is mediated through electrons in bonds, including hydrogen bonds, and is called indirect to emphasise the assistance of electrons in the nucleus-nucleus coupling mechanism. The presence of nuclear spins breaks electrons' spin "up-down" and "down-up" state degeneracy in a filled orbital, due to nuclear spins having negative magnetic hyperfine interaction with electrons. This means that the electron spin distributions are slightly shifted, and the energy of the system depends on the orientation of nuclear spin polarizations within the spin system. For nuclei with the same sign of the gyromagnetic ratio in a two-spin system connected with one bond, the lowest energy is achieved when nuclear spins are anti-parallel.

Generally, the J-coupling tensor, a 3×3 matrix, depends on molecular orientation, but in isotropic liquids it is equal to the average of its diagonal elements and is called scalar coupling. The strength of the interaction is measured by the scalar coupling constant denoted ${}^nJ_{ab}$, in which n designates the number of covalent bonds between two nuclei a and b. Scalar coupling is field independent and measured in Hertz, and causes splitting of peaks in an NMR spectrum. The three bond ${}^1\mathrm{H}^{\mathrm{N}}$—${}^1\mathrm{H}^{\alpha}$ J-coupling is for example used to determine the phi torsion angles in proteins.[59,62]

J-coupling provides qualitative molecular structure information, but sometimes peak splitting is undesirable. Reasons for this would be that J-splitting distributes signal intensity over many smaller peaks and in complex molecules the spectrum easily becomes too crowded to interpret. Fortunately undesired heteronuclear splitting can be eliminated by radiating the sample with Larmor frequencies of the nuclei not under signal acquisition. This way the resulting spectrum is free of heteronuclear J-splitting and only contains desired resonances. However, the homonuclear ${}^2J_{HH}$ and ${}^3J_{HH}$ coupling are not as easily decoupled.

3.1.3 Spin density operator and product operator formalism

Collection of spins in a sample behave to a very good approximation as an *ensemble* of isolated spins. Some of the spins are in the $|\alpha\rangle$ and the $|\beta\rangle$ states, but vast majority are in superposition states that are intermediates between $|\alpha\rangle$ and $|\beta\rangle$, and the spin polarisation vectors are distributed almost uniformly in all possible directions of space. The eigenkets and eigenbras of α and β spin states can be presented in vector form

$$|\alpha\rangle = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad |\beta\rangle = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \quad \langle\alpha| = \begin{bmatrix} 1 & 0 \end{bmatrix}, \quad \langle\beta| = \begin{bmatrix} 0 & 1 \end{bmatrix}, \tag{3.15}$$

and for a single spin in a general superposition state:

$$|\psi\rangle = \begin{bmatrix} c_\alpha \\ c_\beta \end{bmatrix}, \tag{3.16}$$

where c_α and c_β are complex superposition coefficients. An elegant way of describing the behaviour of these innumerable small contributions from each individual spin to the net magnetization is by using *density operator* $\hat{\sigma}$

$$\hat{\sigma} = N^{-1}\left(|\psi_1\rangle\langle\psi_1| + |\psi_2\rangle\langle\psi_2| + ...\right) = \overline{|\psi\rangle\langle\psi|} \tag{3.17}$$

where N is the number of members in the ensemble and ψ_n describe the different spin states. Expectation value of operator $\hat{Q}$ is

$$\left\langle\hat{Q}\right\rangle = \mathrm{Tr}\{\hat{\sigma}\hat{Q}\}. \tag{3.18}$$

Remarkably, any macroscopic observation may be deduced from two spin operators, with one representing the observable (in this case arbitary operator $\hat{Q}$) and the other representing the state of the entire spin ensemble $\hat{\sigma}$. This result does not depend on the number of spins involved and gives an elegant way of extracting information on the expectation value of an operator, instead of specifying the individual microscopic

states of spins. For example an NMR tube containing 700 µL of water has ~10^{19} proton spins when rare isotopes are neglected.[59]

The evolution of density operator is described by Liouville-von Neumann equation which can be derived from the time-dependent Schrödinger equation

$$\frac{\mathrm{d}\hat{\sigma}(t)}{\mathrm{d}t} = -i\left[\hat{H}, \hat{\sigma}(t)\right], \tag{3.19}$$

and if the Hamiltonian is time independent the solution for $\sigma(t)$ is trivial. The matrix representations of spin angular momentum operators that use $|\alpha\rangle$ and $|\beta\rangle$ states of the spins as basis functions is a useful way for analysing the evolution of density operator. The Pauli spin matrices form a complete set for single spin system

$$\hat{I}_x = \frac{1}{2}\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad \hat{I}_y = \frac{1}{2}\begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix}, \quad \hat{I}_z = \frac{1}{2}\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}. \tag{3.20}$$

Each of these operators are Hermitian and satisfy cyclic commutation relation.[62] The spin rotation operators are given by

$$\hat{R}_x(\theta) = \exp\left(-i\theta\hat{I}_x\right), \quad \hat{R}_y(\theta) = \exp\left(-i\theta\hat{I}_y\right), \quad \hat{R}_z(\theta) = \exp\left(-i\theta\hat{I}_z\right), \tag{3.21}$$

where θ is the rotation angle. Additionally, the shift $(\hat{I}^+, \hat{I}^-)$ and polarization $(\hat{I}^\alpha, \hat{I}^\beta)$ operators are given in bra-ket notation in $|\alpha\rangle$ and $|\beta\rangle$ basis by

$$\hat{I}^\alpha = |\alpha\rangle\langle\alpha|, \quad \hat{I}^\beta = |\beta\rangle\langle\beta|, \quad \hat{I}^+ = |\alpha\rangle\langle\beta|, \quad \hat{I}^- = |\beta\rangle\langle\alpha|. \tag{3.22}$$

These sets of operators give the tools to describe the time evolution of spin density operator.[59] However, the matrix multiplications and manipulations can become cumbersome, especially when taking into account multiple spins and couplings between them. *Product operator formalism* describes the evolution of a spin system during the R.F pulses and the delays between them in terms of Cartesian product operators. It improves upon the semi-classical vector model, but is a simplification of the full density matrix treatment. If three operators satisfy cyclic commutation relationship and following equalities apply

$$\left[\hat{A}, \hat{B}\right] = i\hat{C}, \tag{3.23}$$

$$\exp\left(-i\theta\hat{C}\right)\hat{A}\exp\left(-i\theta\hat{C}\right) = \hat{A}\cos(\theta) + \hat{B}\sin(\theta). \tag{3.24}$$

On-resonance R.F pulses induce rotations in a plane orthogonal to the axis of radiation. The Hamiltonian describing the pulses can be written as

$$\hat{H}t = \theta\hat{I}_x \quad \text{or} \quad \hat{H}t = \theta\hat{I}_y, \tag{3.25}$$

for an x-pulse or y-pulse respectively, and where θ is the flip angle of the pulse. The flip angle θ is equal to the product of the frequency (ω_1) and the time (τ_p) of the R.F pulse . The density operators after a R.F pulse with the phase $\pm x$ or $\pm y$ are given by

$$\hat{I}_x \xrightarrow{\theta \hat{I}_{\pm x}} \hat{I}_x \tag{3.26}$$

$$\hat{I}_y \xrightarrow{\theta \hat{I}_{\pm x}} \hat{I}_y \cos(\theta) \pm \hat{I}_z \sin(\theta) \tag{3.27}$$

$$\hat{I}_z \xrightarrow{\theta \hat{I}_{\pm x}} \hat{I}_z \cos(\theta) \mp \hat{I}_y \sin(\theta) \tag{3.28}$$

$$\hat{I}_x \xrightarrow{\theta \hat{I}_{\pm y}} \hat{I}_x \cos(\theta) \mp \hat{I}_z \sin(\theta) \tag{3.29}$$

$$\hat{I}_y \xrightarrow{\theta \hat{I}_{\pm y}} \hat{I}_y \tag{3.30}$$

$$\hat{I}_z \xrightarrow{\theta \hat{I}_{\pm y}} \hat{I}_z \cos(\theta) \pm \hat{I}_x \sin(\theta). \tag{3.31}$$

The initial state of the spin system is described by equilibrium density operator. Its evolution through sequence of R.F pulses and delays can be described by the equation 3.19, or by using the product operator formalism for the pulses (*vide supra*) and free precession (*vide infra*). By convention, the complex magnetization is then detected, and its expectation value at the desired time $\langle M^+ \rangle$ is given by the equation 3.18:

$$\left\langle M^+ \right\rangle (t) = N\gamma\hbar \mathrm{Tr}\{\sigma(t)\mathbf{F}^+\}, \quad \text{where} \quad \mathbf{F}^+ = \sum_{k=1}^{M} \left(|\alpha\rangle_k \langle\beta|_k \right), \tag{3.32}$$

where N is the number of spins per unit volume.[62]

Free precession

When spins are in free precession in weak coupling scheme, chemical shift and scalar coupling evolution takes place. For spin I chemical shift Hamiltonian has the for $\hat{H} = \Omega_I \hat{I}_z$, where Ω_I is the offset frequency of the spin I. Evolution during a delay t is described by

$$\hat{I}_z \xrightarrow{\Omega_I \hat{I}_z t} \hat{I}_z \tag{3.33}$$

$$\hat{I}_x \xrightarrow{\Omega_I \hat{I}_z t} \hat{I}_x \cos(\Omega_I t) + \hat{I}_y \sin(\Omega_I t) \tag{3.34}$$

$$\hat{I}_y \xrightarrow{\Omega_I \hat{I}_z t} \hat{I}_y \cos(\Omega_I t) - \hat{I}_x \sin(\Omega_I t). \tag{3.35}$$

For a two-spin IS system, the scalar coupling Hamiltonian has the form $\hat{H} = 2\pi J_{IS} \hat{I}_z \hat{S}_z$, where J_{IS} is the scalar coupling constant. Scalar coupling evolution during a delay t is described by

$$\hat{I}_z \xrightarrow{2\pi J_{IS} \hat{I}_z \hat{S}_z t} \hat{I}_z \tag{3.36}$$

$$\hat{I}_x \xrightarrow{2\pi J_{IS}\hat{I}_z\hat{S}_z t} \hat{I}_y \cos(2\pi J_{IS} t) + 2\hat{I}_y\hat{S}_z \sin(2\pi J_{IS} t) \tag{3.37}$$

$$\hat{I}_y \xrightarrow{2\pi J_{IS}\hat{I}_z\hat{S}_z t} \hat{I}_y \cos(2\pi J_{IS} t) - 2\hat{I}_x\hat{S}_z \sin(2\pi J_{IS} t)\,. \tag{3.38}$$

The resulting two-spin operators, such as $2\hat{I}_x\hat{S}_z$,undergo analogous evolutions. Evolution of the $\hat{S}_\eta$ and $2\hat{I}_z\hat{S}_\eta$ operators, where η is any of the three Cartesian axes, can be obtained by exchanging the labels.[62]

The Spin echo

The spin echo experiment for an isolated spin is examined. The equilibrium magnetization is proportional to $\hat{I}_z$ and the spin echo pulse sequence is

$$\frac{\pi}{2}\hat{I}_x - t - \pi\hat{I}_x - t-, \tag{3.39}$$

where t is the time of free precession. The initial $(\frac{\pi}{2})_x$ pulse flips the magnetization to negative y-axis and the chemical shift evolution takes place during the first free precession period

$$\hat{I}_z \xrightarrow{\frac{\pi}{2}\hat{I}_x} -\hat{I}_y \xrightarrow{\Omega_I\hat{I}_z t} -\hat{I}_y \cos(\Omega_I t) + \hat{I}_x \sin(\Omega_I t)\,. \tag{3.40}$$

The following $(\pi)_x$ inverts the $\hat{I}_y$ term, and the final density operator after the second free precession period equates to

$$-\hat{I}_y \cos(\Omega_I t) + \hat{I}_x \sin(\Omega_I t) \xrightarrow{\pi\hat{I}_x} \hat{I}_y \cos(\Omega_I t) + \hat{I}_x \sin(\Omega_I t) \xrightarrow{\Omega_I\hat{I}_z t} \hat{I}_y. \tag{3.41}$$

The resulting density operator after the second free precession period was reduced to only $\hat{I}_y$ by trigonometric identity $\cos^2\theta + \sin^2\theta = 1$. Remarkably, no net evolution of chemical shift occurs during the spin echo sequence for a single isolated spin, as the evolution under the chemical shift Hamiltonian is refocused. The sign inversion could have been avoided by employing a $(\pi)_y$ pulse instead.[62]

Insensitive Nuclei Enhanced by Polarization Transfer

In general, nuclear isotopes with high gyromagnetic ratio γ are easier to observe, as nuclear magnetic moment is proportional to it as implied by the equation 3.1 causing a stronger NMR signal. Additionally, at thermal equilibrium the Boltzmann polarization of the spins is proportional to γ as implied by the equation 3.6. Roughly, the dependence of signal-to-noise (S/N) ratio of NMR signal is

$$\frac{\text{Signal}}{\text{Noise}} \propto |\gamma^{5/2}| B_0^{3/2}. \tag{3.42}$$

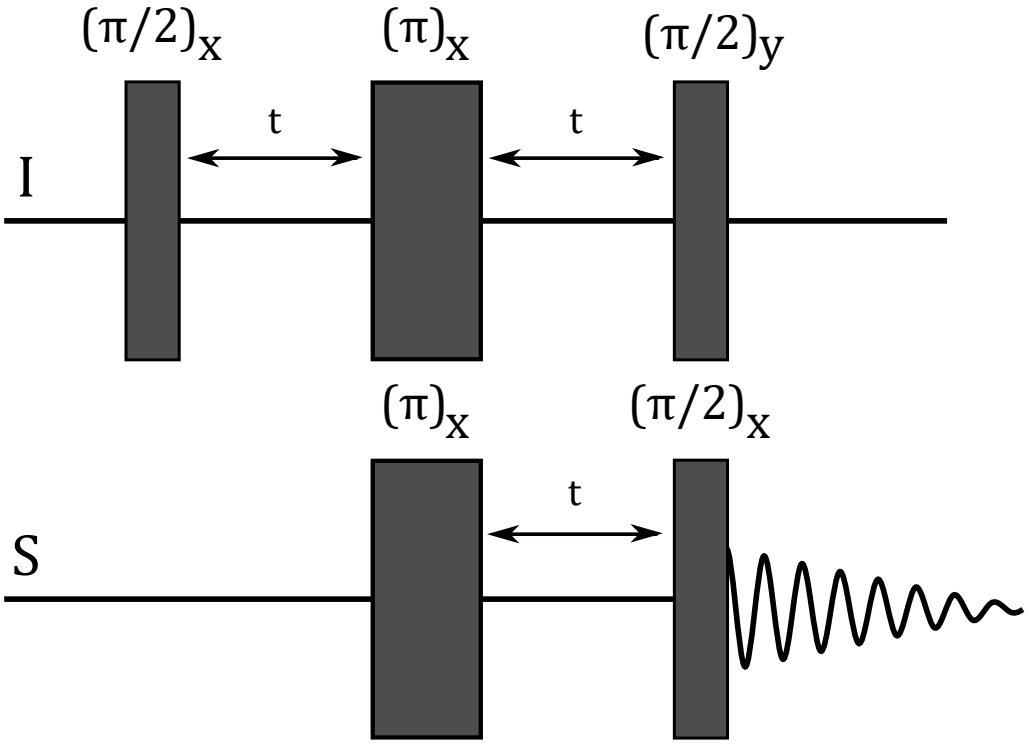


Figure 3.3: INEPT pulse sequence for two-spin system IS. The pulse angles with the corresponding phases are shown above the rectangles and the free precession times in between them.

The equation 3.42 implies that equal number of ^{1}H nuclei provide 300 times better S/N ratio as compared to equal amount of ^{15}N nuclei, at the same magnetic field. This also implies that it takes about 100 000 times longer to acquire a ^{15}N spectrum with the same S/N ratio as a ^{1}H spectrum with equal amount of spins. However, with Insensitive Nuclei Enhanced by Polarization Transfer (INEPT) it is possible to have the magnetization of a heteronucleus to be dependent $\gamma_{^{1}\mathrm{H}}$ of a bonded proton. The INEPT pulse sequence for two-spin system IS is described in figure 3.3.[59]

Before the final pair of $\left(\frac{\pi}{2}\right)$ pulses in the INEPT pulse sequence, the pulse sequence is the same as a spin echo sequence and the chemical shift is refocused during the echo after the $(\pi)_x$ pulses. However, the scalar coupling evolves fully. Starting from the equilibrium magnetization $\mathbf{B}_I\hat{I}_z$, where $\mathbf{B} = \frac{\hbar\omega_I}{4k_bT}$ is the Boltzmann factor of the spin I[62] the evolution of the density operator up until the final pair of pulses is

$$\mathbf{B}_I\hat{I}_z \xrightarrow{\frac{\pi}{2}\hat{I}_x - \mathrm{t}\text{-}\pi(\hat{I}_x+\hat{S}_x) - \mathrm{t}} \mathbf{B}_I\left[\hat{I}_y\cos(2\pi J_{IS}t) - 2\hat{I}_x\hat{S}_z\sin(2\pi J_{IS}t)\right]. \tag{3.43}$$

The $(\frac{\pi}{2})$ pulses applied in x- and y-phases to spins S and I respectively cause the antiphase magnetization of spin I to be transferred to the spin S:

$$\begin{aligned}&\mathbf{B}_I\left[\hat{I}_y\cos(2\pi J_{IS}t) - 2\hat{I}_x\hat{S}_z\sin(2\pi J_{IS}t)\right] \\ &\xrightarrow{\frac{\pi}{2}(\hat{I}_y+\hat{S}_x)} \mathbf{B}_I\left[\hat{I}_y\cos(2\pi J_{IS}t) - 2\hat{I}_z\hat{S}_y\sin(2\pi J_{IS}t)\right].\end{aligned} \tag{3.44}$$

By setting the delay to $t = 1/(4J_{IS})$, the final signal detected from the transverse magnetization of the spin S is proportional to the Larmor frequency of the spin I, and thus the intensity ratio is increased by γ_I/γ_S! When the gyromagnetic ratio of the S spin is small, the advantage of the INEPT experiment becomes enormous. For example if $I = {}^1\mathrm{H}$ and $S = {}^{15}\mathrm{N}$ the intensity gain is close to ten fold.[59]

3.2 The NMR spectrometer

As discussed before, the NMR signal is very weak due to the almost isotropic distribution of nuclear magnetic moments and their relative strength being a fraction that of electrons. Improvement in resolution comes with stronger magnets and magnetic field homogeneity, but sensitivity traditionally remains to be a problem for NMR spectroscopy as around 10^{14} nuclear spins are required to obtain a usable NMR signal. A schematic of a single-channel NMR spectrometer is provided in the appendix A and its operation principle is briefly discussed.

The superconducting magnet used in NMR spectrometers is typically made of an Nb–Sn alloy coil immersed in a bath of liquid He, which has a boiling point of 4.18 K. This He bath is insulated with a large reservoir of liquid N_2 and both of the coolant reservoirs are also insulated from room temperature. The NMR sample is placed at the heart of the magnet within a cylindrical device called the probe through a relatively large hole. In order to manipulate the net magnetization of the sample, an NMR spectrometer needs a *transmitter section* where radio frequency (R.F) pulses are generated. This comprises of an R.F synthesizer which produces an oscillating electric signal at the spectrometer reference frequency ω_{ref}. The general form of the synthesized output signal is

$$s_{synth} \sim \cos\big(\omega_{ref}\, t + \phi(t)\big), \tag{3.45}$$

where $\phi(t)$ is the phase of the R.F pulse and t is time. A pulse programmer controls the rapid jumps between different phases used in an NMR experiment. It also controls the pulse gate which is used for the timing of the pulse. Duration of an R.F pulse selected like this is called the pulse width. The resulting signal is lead to an amplifier that scales up the gated waveform. This strong R.F signal is transmitted through a duplexer to the probe where it is used to generate the NMR signal.

The probe is the most specialized part of the NMR spectrometer, and usually the only part that needs to be exchanged when switching between different experiments (e.g from liquid-state to solid-state NMR experiments). The most simple probe would comprise of circuitry for irradiating the sample with R.F pulses, and for detecting the subsequent NMR signal. But in order to better the final output of the NMR spectrometer the following methods are routinely employed: cryogenic cooling of electronic cir-

cuits, temperature stabilization of the sample and a device for rotating the sample. In some cases, the probe also contains additional coils that are used to generate pulsed field gradients needed in some NMR experiments. All of the probe components near the sample are made of non-magnetic materials, including the electromagnetic coil (i.e. composite material with near-zero net magnetic susceptibility), to avoid any undesirable distortions of the homogeneous magnetic field.

The weak NMR signal is transmitted back through the duplexer and a preamplifier to a *receiver station*. Here the oscillating electrical current is converted to digital form. Before this the NMR signal must be "down-converted" as it oscillates too fast for analogue-to-digital converters (ADCs). This is done by subtracting the spectrometer reference frequency ω_{ref} from the detected NMR signal which oscillates at the Larmor frequency

$$\Omega^0 = \omega^0 - \omega_{ref}. \tag{3.46}$$

Here Ω^0 is relative Larmor frequency or offset frequency with usual order of 1 MHz or less. The NMR signal from decaying transverse magnetization generates a FID with the form

$$s_{FID}(t) \sim \cos\left(\omega^0 t\right) \exp\left(-t/T_2\right). \tag{3.47}$$

The resulting output of the receiver does not distinguish between the different signs of ω^0, and to resolve the ambiguity between larger and smaller ω^0 frequencies compared to ω_{ref}, the receiver outputs a complex signal $s(t)$ using two outputs s_A and s_B. This scheme is called *quadrature detection* and the $s(t)$ has the form

$$s(t) = s_A(t) + is_B(t) \sim \exp\left[(i\Omega^0 - 1/T_2)t\right], \tag{3.48}$$

where $s_A = \cos\left(\Omega^0 t\right) \exp\left(-t/T_2\right)$ and $s_B = \sin\left(\Omega^0 t\right) \exp\left(-t/T_2\right)$. This complex signal corresponds to only one peak, but usually it comprises of a sum of multiple peaks. Finally each of these two outputs are digitized from the analog voltage-versus-time signal in the ACDs. The acquisition time (τ_{acq}) is

$$\tau_{acq} = n_{sample}\tau_{sample}, \tag{3.49}$$

where n_{sample} is the number of sampled points and τ_{sample} is the sampling interval.[59,62]

NMR spectrum

As the NMR signal is quite weak, it is affected by random signals (noise) originating from thermal motions of electrons in the receiver coil, and the actual output of the digitizer is

$$s(t) = s_{NMR}(t) + s_{noise}(t), \tag{3.50}$$

where s_{NMR} is the actual NMR signal and s_{noise} is the random noise contribution. By measuring many separate experiments (transients) and averaging the signal over all of the transients, enhanced S/N ratio can be achieved, thanks to the actual NMR signal staying constant between the transients as the noise varies. By averaging the signal of N transients, the S/N ratio is bettered by a factor of $\sqrt{N}$. The time between the experiments is limited by the longitudinal relaxation time T_1, as the z-magnetization has to be restored before a new experiment can start. The signal averaging is a powerful but time consuming method, and the NMR experiment should be designed in a way that the trade off between the attainable S/N ratio and the instrument time required is within acceptable limits.[59]

In oder to convert the quadrature-detected complex signal $s(t)$ into a form more accessible to the human eye, the signal $s(t)$ is treated with mathematical technique called the Fourier transform (FT). FT converts a function of time into a function of frequency, but does not increase the theoretical information content. The mathematical definition of FT is

$$S(\Omega) = \int_0^\infty s(t) \exp(-i\Omega t)\, \mathrm{d}t. \tag{3.51}$$

The resulting complex function $S(\Omega)$ is called the spectrum, where the individual spectral components ($S_l(\Omega)$) each correspond to a signal components in the original superposition of the time-domain signal:

$$S_l(\Omega) = \int_0^\infty s_l(t) \exp(-i\Omega t)\, \mathrm{d}t, \tag{3.52}$$

and after evaluating the integral, the spectrum is can be understood as a superposition of Lorentzian spectral components

$$S_l(\Omega) = a_l \mathscr{L}(\Omega) = a_l \left(\frac{1}{\lambda + i\,[\Omega - \Omega_l]} \right), \tag{3.53}$$

where $\lambda = T_2^{-1}$, a_l is a complex amplitude, Ω_l is the centre frequency of the peak and $\mathscr{L}$ is the complex Lorentzian. The complex Lorentzian consists of has real (absorption) and imaginary (dispersion) parts with the following relations

$$\mathscr{L} = \mathscr{A} + i\mathscr{D}, \tag{3.54}$$

where the absorption Lorentzian is

$$\mathscr{A}(\Omega) = \mathrm{Re}\{\mathscr{L}(\Omega)\} = \frac{\lambda}{\lambda^2 + (\Omega - \Omega_l)} \tag{3.55}$$

and the dispersion Lorentzian is

$$\mathscr{D}(\Omega) = \mathrm{Im}\{\mathscr{L}(\Omega)\} = \frac{\Omega - \Omega_l}{\lambda^2 + (\Omega - \Omega_l)}. \tag{3.56}$$

The linewidth of the absorptive Lorentzian is defined as the full-width-at-half-height (FWHH), and is given in hertz as FWHH $= 1/(\pi T_2)$. As seen from the equations 3.55 and 3.56, with large offsets from the centre frequency $(\Omega - \Omega_l)$ the absorptive Lorentzian lineshape decays in inverse square off the offset as compared to the dispersive Lorentzian, resulting in sharper peaks. Thus absorptive lineshapes are preferred.[62] Additionally, slow (fast) signal decay results in narrow (broad) peaks as suggested by the FWHH relation.

Chapter 4

NMR studies of proteins

In order to get structural information about proteins, the measured resonances have to be assigned. Chemical shifts of ^{13}C and ^{15}N nuclei are more impressionable than ^{1}H to the amino acid type, and thus to the primary structure. This means that these resonances are more dispersed and are therefore readily used in resonance assignment of IDPs. For folded proteins, HN-detected triple-resonance experiments are usually enough to resolve the backbone resonances by linking $^{1}H^{N}$, $^{13}C^{\alpha}$, $^{13}C^{'}$, ^{15}N and even $^{13}C^{\beta}$ resonances together in a sequential fashion. This can be really challenging with IDPs since significant signal overlap is usually present, as their structures are plain and the nuclei have monotonic chemical environment.[63] Additionally amide protons are more exposed to the solvent, which promotes chemical exchange especially in alkali conditions, resulting in $^{1}H-^{15}N$ signal broadening and overlap.[64] This greatly hampers HN-detected experiments, which are the most common ones used in protein structure elucidation. With the technological advancements resulting in higher spectrometer sensitivity, such as cryogenically cooled probes (1999) and more powerful NMR spectrometers, ^{13}C-detected experiments and strategies have been readily employed in the field of protein NMR.[63,65]

4.1 Coherence transfer

Desired properties in a NMR spectrum include high signal-to-noise ratio and resolution, good signal dispersion and unmistakable selectivity. Achieving all of the prerequisite conditions while measuring distinct signals from every nucleus of a protein requires well optimized and economical coherence transfer pathways and experiment conditions.

The polypeptide chain can be thought as a node network of magnetic nuclei and connectivities (J-couplings) between them. The coherence is a correlated state with one or

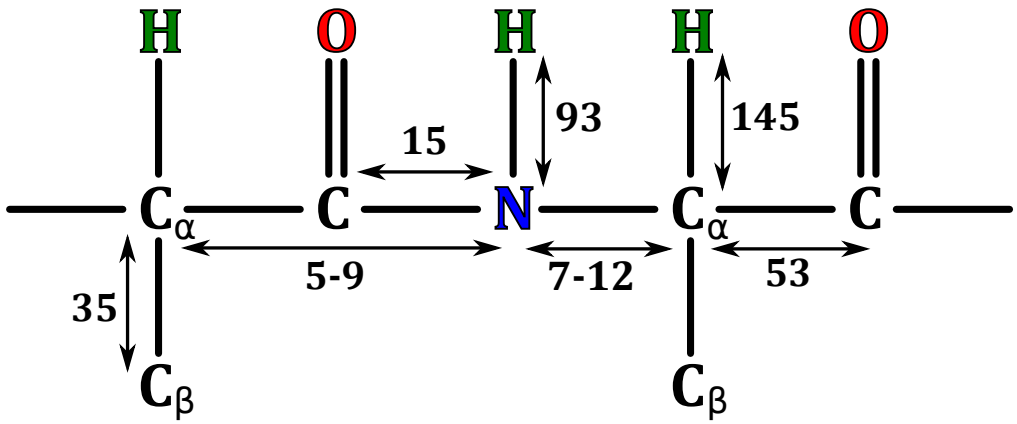


Figure 4.1: Generic polypeptide chain with typical J-couplings (in Hz) used for coherence transfer through the spin system.

several nuclei in their longitudinal or transverse magnetization states. The magnetization can be transferred from one nucleus to another through these connectivities. The most efficient transfer is between strongly coupled nodes with long relaxation times. Relaxation is dependent on the motional freedom (rotational correlation time), the strength of the external and thus also the local magnetic field as well as the isotope-labelling pattern. Although the flow of coherence might be economically directed, in practice even the most ideal experiments only partially fulfil the aforementioned requirements. It is also noteworthy that in order to succeed in NMR experiments, the sample preparation and its behaviour during the measurement still plays the most significant role and no optimization or development of new experiments can compensate for it.[66] Schematic presentation of the polypeptide chain with the respective J-couplings is presented in figure 4.1.

In NMR nomenclature the spins that are frequency labelled during an experiment are listed as the experiment's name. For example $^{1}H^{N}$, ^{15}N, $^{13}C^{\alpha}$, $^{13}C^{\beta}$, $^{13}C'$ would be represented with HN, N, CA, CB and CO. The spins that are not frequency labelled but participate in coherence transfer are given in parentheses. For example HN(CO)CA experiment describes the following coherence transfer pathway:

$$^{1}H^{N} \longrightarrow {}^{15}N[t_1] \longrightarrow {}^{13}CO \longrightarrow {}^{13}C^{\alpha}[t_2] \longrightarrow {}^{13}CO \longrightarrow {}^{15}N \longrightarrow {}^{1}H^{N}[t_3] \quad (4.1)$$

with frequency labelling order next to respective spins[62] in parentheses. Usually the J-couplings are presented over the arrows. In this experiment the initially excited proton spin is identical to the detected. This approach is called "out-and-back", but it is also possible to employ one called "out-and-stay" where the signal is detected from a spin different from the initial one.

4.2 NMR measurables for structural information

4.2.1 Chemical shift

At the heart of any NMR study is definitely the chemical shift. It is the most widely used data that is obtained by NMR spectroscopy and it tells about the local magnetic environment of a nucleus. It is the fundamental measurable from which other restraints for protein structure determination can be deduced. For IDPs the observed chemical shift is a population-weighted average of conformers exhibited by the protein in question. Basic 1D ^{1}H NMR measurement already shows whether or not a protein is folded, as IDPs exhibit small chemical shift dispersion.

Secondary structure of a protein or its parts can be deduced with many different methods, and usually combination of two or more of said methods gives the most robust conclusion. One of these, which serves as a good starting point, is called secondary chemical shift (SCS). It is usually calculated by subtracting a typical random coil chemical shift value from $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ chemical shifts:

$$\Delta\delta = \delta_{observed} - \delta_{randomcoil}. \tag{4.2}$$

Generally β-sheets have negative $^{13}C^{\alpha}$, and positive $^{13}C^{\beta}$ and $^{1}H^{\alpha}$ secondary shifts and vice versa for α-helices. In IDPs, rapid conformational averaging results in much smaller SCS values as compared to ordered proteins, and carry information on the relative populations of dihedral angles in different local conformations.[10] The random coil chemical shift values are tabulated in different databases, and the selection of the right reference data is pivotal. The neighbouring residues affect the chemical shifts, but their effect can be removed with so-called neighbour correction. Additionally calibration of the measurement conditions with respect to temperature and pH is necessary[67–69]

4.2.2 Scalar coupling

In addition to secondary chemical shifts, information about the secondary structure can also be obtained by using scalar couplings. Magnitude of the three-bond scalar coupling constant is a function of the dihedral angle as described by the Karplus equation[70]

$$^{3}J = A\cos^{2}\theta + B\cos\theta + C, \tag{4.3}$$

where A, B and C are constants that depend on the nuclei involved, and θ is the dihedral angle. After accurate parametrization of the equation, deviations from coupling constants in random coil peptides can be calculated and the local secondary structure

inferred. For example positive deviation is typical of β-sheet structures and negative for α-helical structures. Usually only the angle Φ is determined since it is easier to measure (from $^{3}J_{H^{N}H^{\alpha}}$), although strategies also for both dihedral angles employing scalar couplings have been developed.[71–73]

4.2.3 Paramagnetic relaxation enhancement

Paramagnetic relaxation enhancement (PRE) is the most common NMR method to gain long range structural information for IDPs.[63] It is based on dipolar interactions between spins of NMR active nuclei and spins of unpaired electrons, belonging to an artificially added paramagnetic components into the protein. This interaction results in shorter relaxation times for both transverse and longitudinal magnetization. By determining the paramagnetic rate enhancement R_2^{sp}, distance information can be calculated by solving the following equation

$$r = \left[\frac{K}{R_2^{sp}}\left(4\tau_c + \frac{3\tau_c}{1+\omega_h^2\tau_c^2}\right)\right]^{\frac{1}{6}}, \tag{4.4}$$

where r is the distance between the electron and nuclear spins, τ_c is the rotational correlation time for electron-nuclear interaction which is usually approximated to be equal to global correlation time fo the protein, ω_h is the Larmor frequency of the nuclear spin and K is a constant $1.23 \times 10^{-32}\,\mathrm{cm^6 s^{-2}}$, which is proportional to γ_h.[74] The shortened relaxation times result in signal broadening for nearby sequential amino acids (10–15 AAs away), but the broadening starts to quickly decline with further distance. If a residue further down stream in the sequence experiences a PRE effect, it is spatially close to the paramagnetic tag. With the equation 4.4 reliable distance information up to 25 Å from the tag can be gained in pico- to nanosecond time scale.[75,76]

Paramagnetic probes are most commonly attached to the thiol group of cysteine residues. In IDP studies usually only one cysteine residue is tagged, which is introduced by site-specific mutagenesis at a desired position whilst other cysteines are removed. The most common paramagnetic probe is methanethiosulfonate spin label (MTSL) for the study of IDPs.[63] For larger proteins, many different mutants can be produced to map long range contacts in multiple regions of the protein.[77] Additionally, different paramagnetic components such as metal binding sites, non-natural amino acids and soluble paramagnetic molecules may be added to the buffer.[78,79]

4.2.4 Residual dipolar coupling

Residual dipolar coupling (RDC) exploits the angle dependency of dipolar interaction. Depending on the angle of internuclear vectors i and j in two nuclei system with respect to B_0, the energies of the spin system change. The magnitude of the dipolar coupling is proportional to the time-dependent angle $\cos^2\theta(t)$ of the internuclear vectors and is described by the following equation:

$$D_{ij} = -\frac{\gamma_i \gamma_j \hbar \mu_0}{4\pi^2 r^3} \left\langle \frac{3\cos^2\theta(t) - 1}{2} \right\rangle. \tag{4.5}$$

In equation 4.5 r is the vibrationally averaged internuclear distance and angular parentheses denote average over all conformations.[80] The maximum dipolar coupling is achieved when $\cos^2\theta = 1$. Due to molecular tumbling dipolar interactions average out in liquids, and the required partial alignment of the protein molecules is achieved with the use congealing media such as lipid bicelles,[81] polyacrylamide gels,[82] or filamentous bacteriophages.[83] Only the ordered regions N-H bond vectors align either parallel or orthogonal to B_0, depending on the secondary structure. Extended structures will be in an orthogonal orientation resulting in a negative RDC value, while α-helical elements have be positive RDCs.

RDCs produce an immense amount of information, which is hard to conceptualize in traditional representations of protein structures. One way to gain meaningful structural information is to use RDCs together with chemical shifts, as optimization restraint for molecular dynamics calculations. This yields an ensemble of conformers on free-energy landscape that the protein could exhibit. Some algorithms used for this purpose are ASTEROIDS[84] and ENSEMBLE,[85] which select the best agreeing conformers with the data.

4.3 Sequential assignment of proteins

Resonance assignment is the initial step for any NMR-based macromolecular structure analysis. The aim of the assignment is to associate each resonance to a specific nucleus usually based on known a AA sequence. A multitude of different assignment strategies employing a great number of NMR experiments can be used to find correlations between protein nuclei, but the overall assignment can be divided into the sequential assignment of amino acid backbone resonances and the assignment of side-chain resonances.

The first step in the sequential backbone resonance assignment is the selection of a root spectrum with relevant nuclei and with good peak dispersion. Usually a 2D spec-

trum is favoured, as every resonance is easily observable as opposed to resonances in 3D spectrum. For the actual assignment a number of triple resonance spectra are selected in such a way that connectivities between amino acid residues can be deduced, and that they include the same resonances as the root spectrum. This is because the root spectrum is typically used for initial peak picking in the 3D spectra. For example, a 2D $^{1}H-^{15}N$-HSQC (heteronuclear single quantum coherence) spectrum contains amide proton and nitrogen resonances of each amino acid residue as well as potential peaks from asparagine, glutamine, tryptophan, arginine and histidine side-chains, and would serve well as a root spectrum. For most small and cooperatively folding proteins, the $^{1}H-^{15}N$-HSQC features well dispersed set of resonances. It provides one-to-one mapping between the resonances and the residues, excepting proline, which lack the amide proton entirely.[86] The 3D spectra used in conjunction with the $^{1}H-^{15}N$-HSQC could be a combination of HNCACB and HN(CO)CACB in order to connect intraresidual C^{α} and C^{β} with sequential ones. Sequentially connecting different residues in this general manner is called "sequential walk". Typical chemical shifts for each backbone nucleus have been well documented in NMR data repositories, such as the Biological Magnetic Resonance Data Bank (BMRB), and can be used to aid the assignment.

Completed assignment enables the extraction of additional structural information (e.g. dihedral angles, interatomic distances) about the protein of interest, as new experiments have a assigned root spectrum, and changes in resonance frequencies can be linked to a local magnetic environment variations of *specific* nuclei. The basic idea of backbone assignment is rather simple and for small globular proteins the process can be automated. But with a lot of overlapping signals, especially prominent in the NMR spectra of IDPs, automated assignment and peak picking algorithms are generally completely unusable and the rather tedious work has to be done manually.[63]

4.3.1 ^{13}C-Detected experiments

Although protons have higher sensitivity during as acquisition compared to ^{13}C and ^{15}N thanks to their large gyromagnetic ratio, ^{1}H-detected experiments have their limitations in specialized cases such as the study of IDPs. In addition to the solvent exposure and low dispersion of amide proton signals, dipolar interactions and fast relaxation of protons contribute to line-broadening in ^{1}H-detected experiments of large macromolecules.[87] In ^{13}C-detected experiments slower relaxation of ^{13}C is exploited resulting in sharper peaks. Extra information about side-chains can also be extracted by studying different ^{13}C types such as carbonyl, alpha, aliphatic and aromatic carbons. Application of this method as a standard operation procedure for biological macromolecule studies can be challenging, as relatively large concentrations of iso-

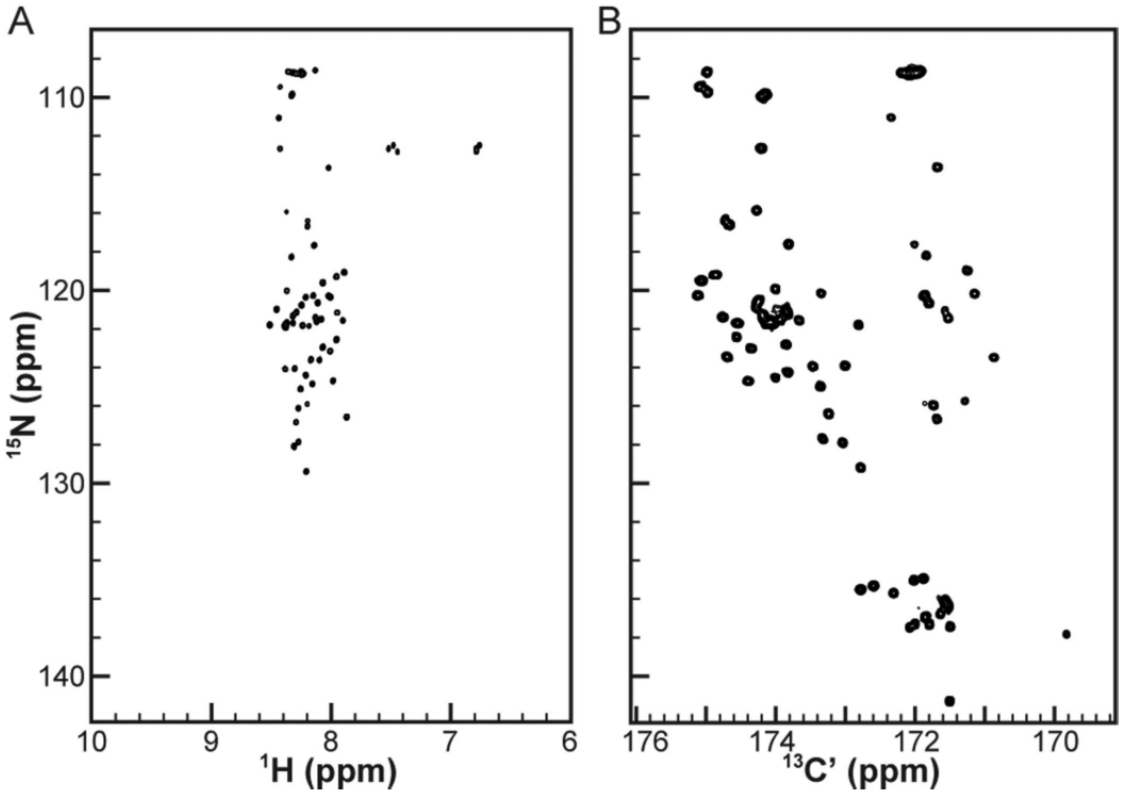


Figure 4.2: The comparison of the two 2D-heteronuclear correlation spectra measured from the 80 residue IDP Pdx1-C. Spectrum A is the traditional ^{15}N-HSQC and the spectrum B is the ^{15}N, ^{13}C-CON. The latter also has peaks corresponding to proline residues at around 136–138 ppm in the nitrogen dimension, which are absent in the ^{15}N-HSQC.[86]

topically labelled protein samples are needed for ^{13}C-detected NMR experiments.

Overall ^{13}C-detected backbone assignment strategy is the same as with ^{1}H^{N}-detected experiments, but the measurements are started from different nuclei. The magnetization pathway usually starts from ^{13}C$^{\alpha}$ or C$^{'}$ and is detected from the same or another ^{13}C nucleus. The latter type of detection usually yields higher resolution with shorter experimental time and is thus preferred.[88] Carbon decoupling is key to achieve high resolution by exploitation of heteronucleus detection and inter-residue correlations. IPAP and S3E filters are used to achieve higher sensitivity than other methods of J-coupling.[89,90] Although ^{13}C-detection is nowadays frequently applied in structural studies, it should be thought as a complementary method to ^{1}H-detected experiments.

^{15}N, ^{13}C-CON

The most compelling alternative for ^{15}N-HSQC, a deficient *de facto* experiment for NMR spectroscopy of IDPs, is recording the correlations between cabonyl carbon and amide nitrogen in the peptide plane of the protein backbone, giving rise to the ^{15}N, ^{13}C-

CON spectrum. The degree of improvement can be clearly seen in a study conducted by Sahu *et al.*[86] on 80 residue IDP Pdx1-C and the comparison of these two spectra is presented in figure 4.2. The bettered chemical shift dispersion and the sharper line widths of the ^{15}N, ^{13}C-CON also serves as a advantageous base for the employment of higher dimensional experiments.

Although carbon-detected experiments are limited by their inherent lower sensitivity on a per scan basis as compared to proton-detected experiments, a wide range of "H-start" experiments have been developed to counteract this. For example, under mildly acidic conditions H^N-flip variant of the ^{15}N, ^{13}C-CON experiment has proven effective for non-proline rich IDPs, but usually excitation of aliphatic proton resonances is the preferred means to enhancing sensitivity. As an example of the latter, ^{1}H$^\alpha$-start (HACA)CON does not include exchangeable protons in the magnetization transfer pathway. Pulse sequences of the (H^N-flip)CON and (HACA)CON are presented in figure 4.3.[86] Large one-bond ^{13}C$-^{13}$C couplings are decoupled in both of the experiments by IPAP approach, as it is easy to implement on any standard NMR equipment and because of its good performance.[91]

4.3.2 ^{1}H$^\alpha$-detected experiments

Resonance assignment of globular proteins has been dominated by HN-detected triple resonance experiments for the last two decades. The use of TROSY effect[92] and (per)deuteration in large protein has rendered HA-detection inferior to HN-detection in many ways.[93] However HA-detection has proven its usefulness in the field of IDP study, as IDPs have optimal relaxation properties for multidimensional NMR studies thanks to their motional freedom. Optimal relaxation times allow development of very complex pulse sequences, not possible in other proteins. HN-detection is also shadowed by the aforementioned amide proton chemical shift degeneracy. Additionally proline, which lacks the amide proton, is the most abundant amino acid in IDPs. Three HA-detected experiments,[93–95] that were used to assign the resonances of three different proteins, both globular and unstructured, will be covered in further detail.

iH(CA)NCO

iH(CA)NCO experiment provides correlation with between only the intraresidual nuclei ^{1}H$^\alpha(i)$, ^{13}CO(i) and ^{15}N(i). The magnetization transfer pathway and the pulse sequence of iH(CA)NCO experiment are presented in figure 4.4 and the coherence

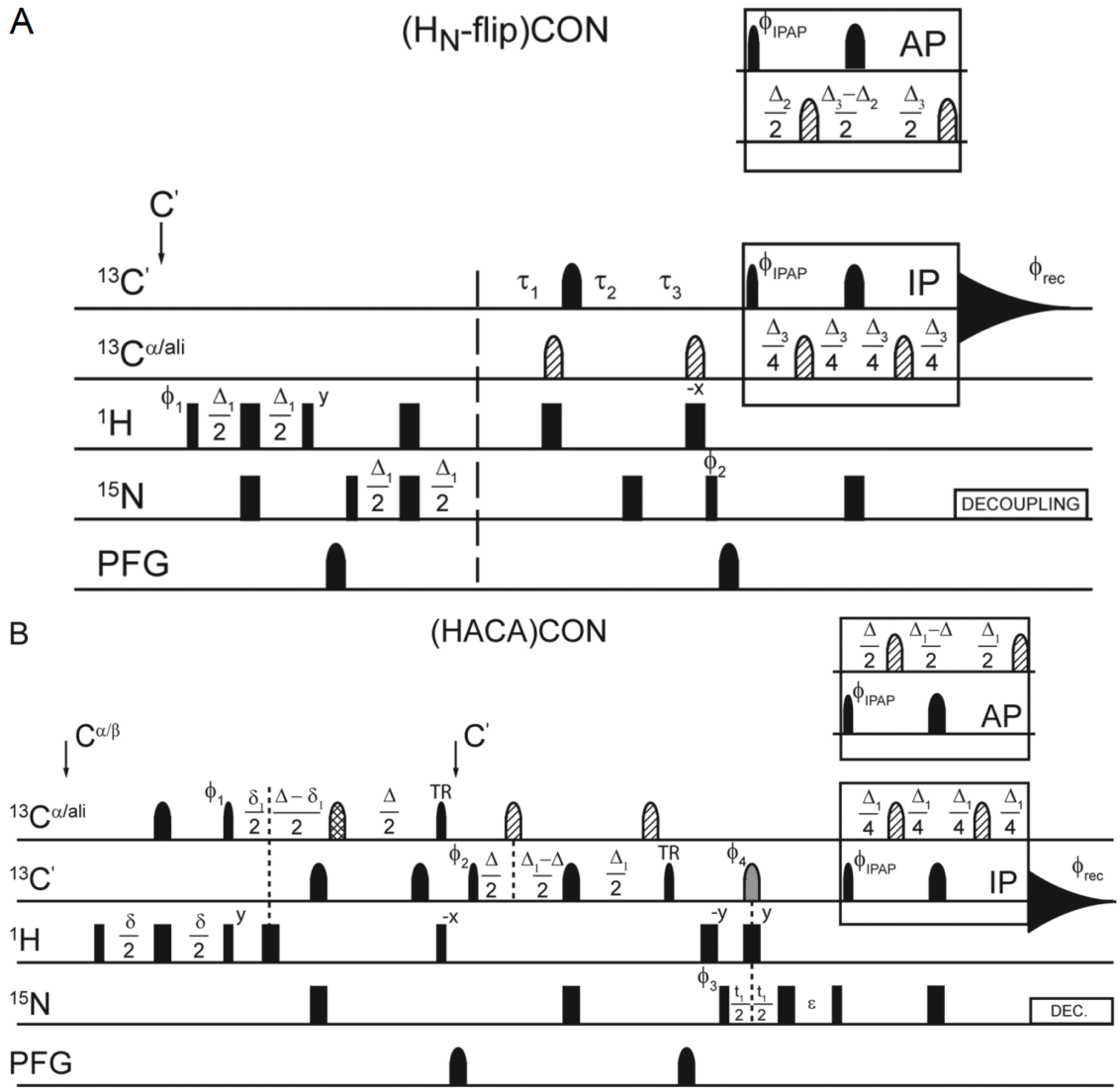


Figure 4.3: Two different variants of the H-start ^{15}N, ^{13}C-CON experiment. In both of the pulse sequences narrow and wide rectangles correspond to 90° and 180° pulses respectively, and the pulses are applied in the x phase unless stated otherwise. Pulse sequence A corresponds to the H_N-flip CON-IPAP experiment, and the employed delays are $\Delta = 5\,\text{ms}$, $\Delta_1 = 4.6\,\text{ms}$, $\Delta_2 = 9.0\,\text{ms}$ and $\Delta_3 = 25\,\text{ms}$. ^{15}N chemical shift evolution is measured during t_1 using semi-constant time period with delays $\tau_1 = (\Delta_3 + t_1)/2$, $\tau_2 = (1 - \Delta_3/t_{1\text{max}})t_1/2$ and $\tau_3 = (1 - t_1/t_{1\text{max}})\Delta_3/2$, where $t_{1\text{max}}$ is the maximum duration t_1 labelling period. The phase cycling for A: $\phi_1 = x,-x$;$\phi_2 = y,y,-y,-y$;$\phi_{\text{IPAP}}(\text{IP}) = x,x,x,x,-x,-x,-x,-x$;$\phi_{\text{IPAP}}(\text{AP}) = -y,-y,-y,-y,y,y,y,y$; and $\phi_{rec} = x,-x,-,x,x,-x,-x,x$. Pulse sequence B corresponds to the (HACA)CON-IPAP experiment, and the employed delays are $\phi = 3.6\,\text{ms}$, $\delta_1 = 2.2\,\text{ms}$, $\Delta = 9.0\,\text{ms}$, $\Delta_1 = 25\,\text{ms}$ and $\epsilon = t_1(0) + \text{pC180}$. The phase cycling for B: $\phi_1 = 4(x),4(-x)$; $\phi_2 = 2(x),2(-x)$; $\phi_3 = x,-x$; $\phi_4 = 8(x),8(-x)$; $\phi_{\text{IPAP}}(\text{IP}) = x$; $\phi_{\text{IPAP}}(\text{AP}) = -y$ and $\phi_{rec} = x,-x,-x,x,-x,x,x,-x$. Quadrature detection in the indirect dimension is obtained by States-TPPI by incrementation of ϕ_2 for A and ϕ_3 for B. Pulses filled with diagonal lines are off-resonance 180° Q3 shaped pulses, centred at 54 ppm, and pulses filled with hashed lines are higher selectivity 180° Q3 shaped pulses applied on resonance with duration of 1200 µs on a system operating at 11.7 T static field strength. TR pulses are time-reversed versions of the 90° Q5 shaped pulse.[86]

flow can be described as follows:

$$^{1}\mathrm{H}^{\alpha}(i) \xrightarrow{^{1}J_{HC}} {}^{13}\mathrm{C}^{\alpha}(i) \xrightarrow{^{1}J_{C\alpha N},\,^{2}J_{C\alpha N},\,^{1}J_{C\alpha CO}} {}^{13}\mathrm{CO}(i)[T_A - t_2;\,^{1}J_{CON};\,^{1}J_{C\alpha N};\,^{2}J_{C\alpha N}]$$
$$\xrightarrow{^{1}J_{C\alpha N},\,^{2}J_{C\alpha N},\,^{1}J_{C\alpha CO}} {}^{15}\mathrm{N(i)}[t_2] \xrightarrow{^{1}J_{C\alpha N},\,^{2}J_{C\alpha N}} {}^{13}\mathrm{C}^{\alpha}(i) \xrightarrow{^{1}J_{HC}} {}^{1}\mathrm{H}^{\alpha}(i)[t_3].$$

Active couplings used in magnetization transfer are listed above the arrows and the acquisition times of the corresponding spins t_{1-3} are in parentheses. For ^{13}CO, a constant time acquisition scheme is employed and the active heteronuclear couplings used for magnetization transfer during the constant time evolution period are also shown in the parentheses.

It is advantageous to include ^{13}CO labelling in triple resonance experiments used for IDPs, as carbonyl carbon chemical shifts are considerably dependent on the neighbouring residue type and thus exhibit better peak dispersion in the final spectrum.[96,97] The iH(CA)NCO spectrum contains one intraresidual peak per amino acid residue because it exploits the non-linearity of the N(i)—C$^{\alpha}$(i)—CO(i)—N($i+1$) spin system so that intraresidual filtering is obtained.[98] Under regular J-coupling ranges ($^{1}J_{NC^{\alpha}}$=9–12 Hz and $^{2}J_{NC^{\alpha}}$=6–9 Hz) at least 20 times larger connectivities in intraresidual pathways are achieved as compared to sequential coherence transfer efficiencies.

This makes the backbone assignment less dubious and identification of spin systems in crowded spectra easier. Additionally, alpha protons of glycines are in 180° phase difference as compared to other alpha protons, which makes them a good starting point for assignment of fragments.

H(CA)CON

H(CA)CON is a complementary experiment to iH(CA)NCO establishing correlations between spins $^{1}H^{\alpha}(i)$, ^{13}CO(i) and ^{15}N(i+1). The magnetization transfer pathway and the pulse sequence of HA(CA)CON experiment are presented in figure 4.5 and coherence flow during the experiment is as follows:

$$^{1}\mathrm{H}^{\alpha}(i) \xrightarrow{^{1}J_{HC}} {}^{13}\mathrm{C}^{\alpha}(i) \xrightarrow{^{1}J_{C\alpha CO}} {}^{13}\mathrm{CO}(i) \xrightarrow{^{1}J_{CON}} {}^{15}\mathrm{N}(i+1)[t_1] \longrightarrow {}^{13}\mathrm{CO(i)}[2T_A - t_2;\,^{1}J_{CON}]$$
$$\xrightarrow{^{1}J_{C\alpha CO}} {}^{13}\mathrm{C}^{\alpha}(i) \xrightarrow{^{1}J_{HC}} {}^{1}\mathrm{H}^{\alpha}(i)[\mathrm{t}_3].$$

Active couplings used in magnetization transfer are listed above the arrows and the acquisition times of the corresponding spins t_{1-3} are in parentheses. For ^{13}CO, a constant time acquisition scheme is employed and the active heteronuclear couplings used for magnetization transfer during the constant time evolution period are also shown in the parentheses.

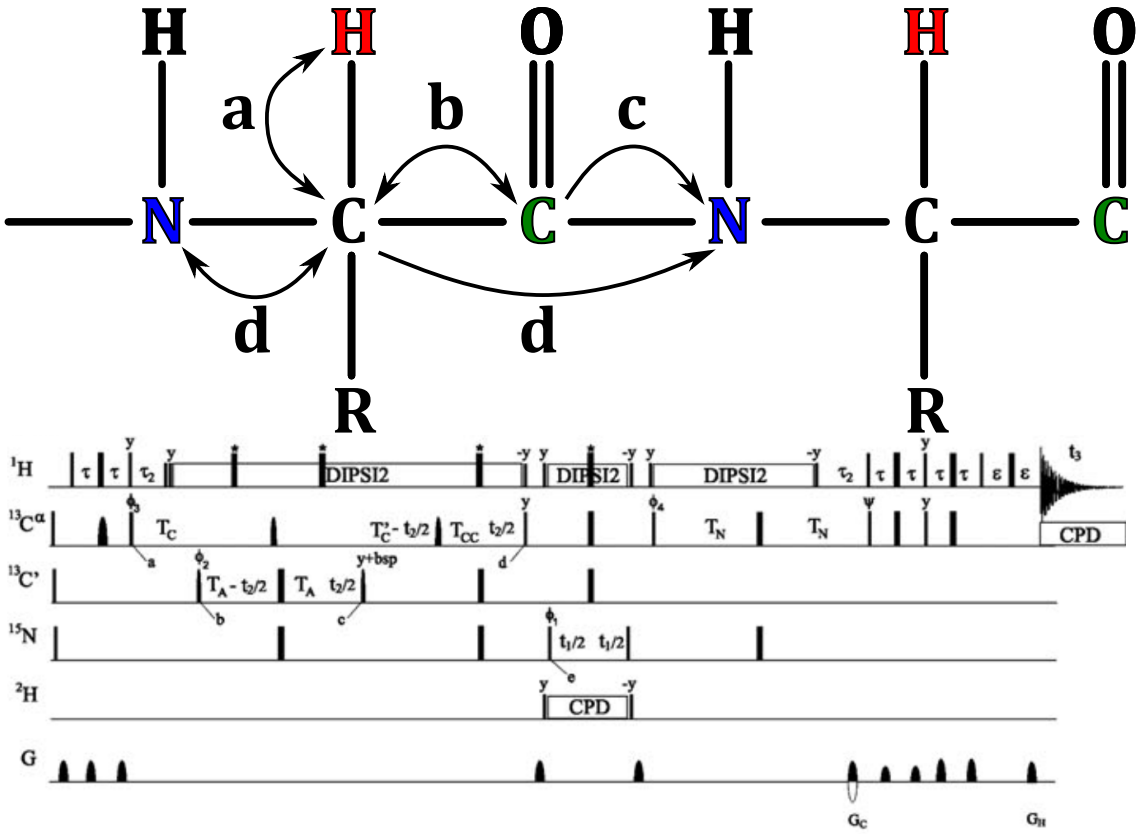


Figure 4.4: iHA(CA)NCO experiment correlates $^{1}H^{\alpha}(i)$, $^{15}N(i)$ and $^{13}C'(i)$ chemical shifts. The upper part of the figure is the schematic presentation of the magnetization transfer pathway during the iH(CA)NCO experiment, and the lower part is the experiment's pulse sequence. Double-headed arrows represent out-and-back transfer route whilst one-way arrows indicate out-and-stay magnetization transfer. The letters a-d correlate to the time points in the experiment's pulse sequence. All of the pulses are applied with phase x unless stated otherwise. Narrow and wide rectangles correspond to 90° and 180° pulses respectively. Delay durations: $\tau = (4J_{HC})^{-1} \sim 1.7\,\text{ms}$; $\tau_2 = 3.4\,\text{ms}$ for non-glycine residues and 2.2 to 2.5 ms for also observing glycines; ϵ is duration of G_H plus field recovery time which is around 0.4 ms; $T_C = (2J_{C\alpha C'})^{-1} \sim 14\,\text{ms}$; $T_A = (4J_{C'N})^{-1} \sim 16.6\,\text{ms}$; $T_{CC} = (J_{C\alpha C\beta})^{-1} - (4J_{C'N})^{-1} - (2J_{C\alpha C'})^{-1}$; $T'_C = T_C + T_{CC}$; $T_N \sim 14\,\text{ms}$; and maximum t_2 is restrained by $t_{2,max} < 2T'_C$. Gradient strengths and durations: $G_C = 13\,\text{kG cm}^{-1}$ (1.6 ms), $G_H = 13\,\text{kG cm}^{-1}$ (0.4 ms). Phase cycling: $\phi_1 = x, -x$; $\phi_2 = 2(x), 2(-x)$; $\phi_3 = 4(x), 4(-x)$; $\phi_4 = x$; $\psi = x$; $\phi_{rec} = x, 2(-x), x, -x, 2(x), -x$. Further details on the experiment can be found from the original article.[93]

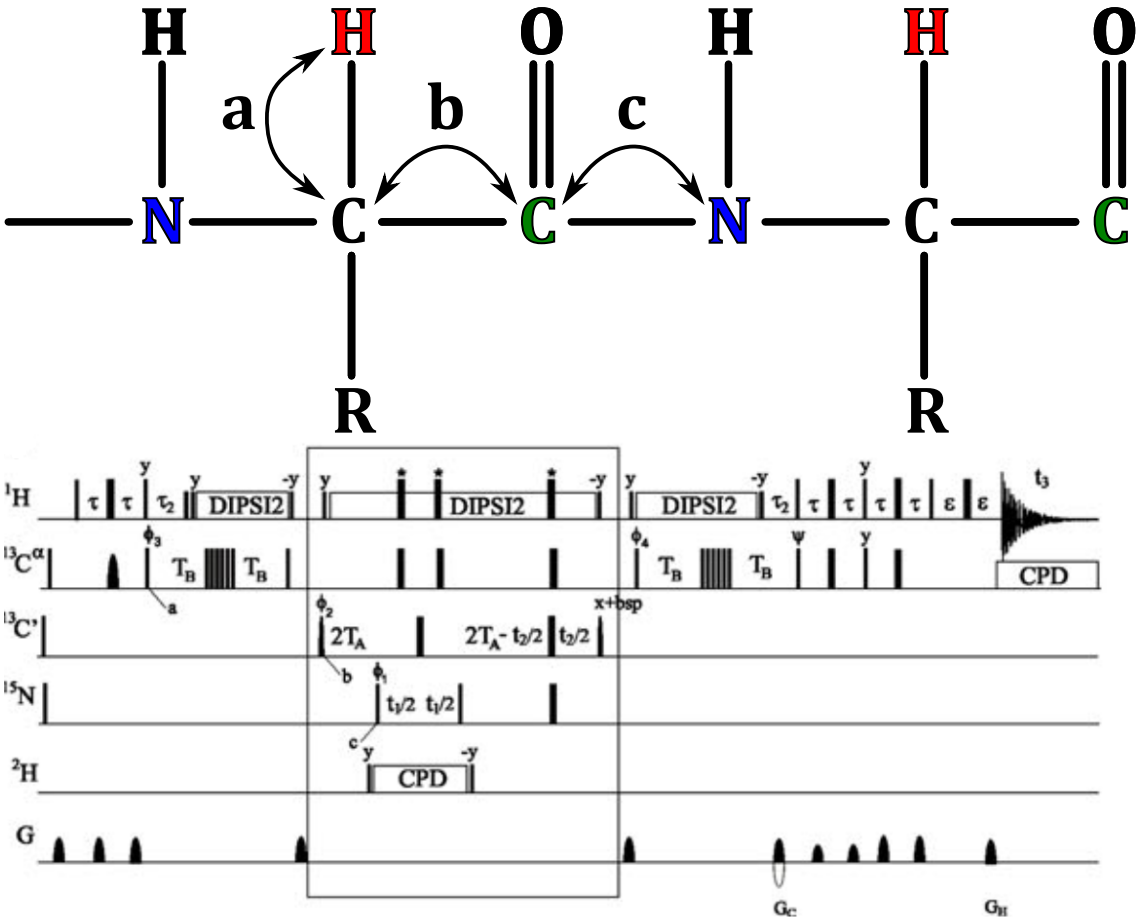


Figure 4.5: H(CA)CON experiment establishes correlations between $^{1}H^{\alpha}(i)$, $^{15}N(i+1)$ and $^{13}C^{'}(i)$ chemical shifts. The upper part of the figure is the schematic presentation of the magnetization transfer pathway during the H(CA)CON experiment, and the lower part is the experiment's pulse sequence. Double-headed arrows represent out-and-back transfer route. Letters a-d correlate to the different time points in the experiment's pulse sequence. All of the pulses are applied with phase x unless stated otherwise. Narrow and wide rectangles correspond to 90° and 180° pulses respectively. Delay durations: $\tau = (4J_{HC})^{-1} \sim 1.7\,\mathrm{ms}$; $\tau_2 = 3.4\,\mathrm{ms}$ for non-glycine residues and 2.2 to 2.5 ms for also observing glycines; ϵ is duration of G_H plus field recovery time which is around 0.4 ms; $T_B = (6J_{C\alpha C'})^{-1} \sim 3.3\,\mathrm{ms}$; $T_A = (4J_{C'N})^{-1} \sim 16.6\,\mathrm{ms}$. The maximum t_2 is restrained by $t_{2,max} < 2T_C'$. Gradient strengths and durations: $G_C = 13\,\mathrm{kG\,cm^{-1}}$ (1.6 ms), $G_H = 13\,\mathrm{kG\,cm^{-1}}$ (0.4 ms). Phase cycling: $\phi_1 = x, -x$; $\phi_2 = 2(x), 2(-x)$; $\phi_3 = 4(x), 4(-x)$; $\phi_4 = x$; $\psi = x$; $\phi_{rec} = x, 2(-x), x, -x, 2(x), -x$. The rectangle in the middle of the pulse sequence is a place for an alternative implementation for $^{13}C^{'}$ chemical shift labelling during t_2 covered in the original article. Further details on the experiment can be found there.[93]

Mäntylahti *et al.* used Shaka–6 composite pulse to improve the otherwise impaired sensitivity for serine and glycine $^{1}H^{\alpha}$'s, but it is possible to gain at least 60% increase on overall sensitivity with minor adjustments to the pulse sequence in proteins with less glycine and serine residues. However, the procedure was deemed imperative, because both serine and glycines are quite abundant in IDPs. Both H(CA)CON and iH(CA)NCO utilize gradient echo with water flip-back to achieve efficient water signal suppression, enabling measurements to be carried out in $^{1}H_2O$.

(HCA)CON(CA)H

(HCA)CON(CA)H was used with (HCA)NCO(CA)H by Mäntylahti *et al.* in assignment of cancer/testis antigen CT16[94] . This new set of experiments was developed to combat ^{15}N chemical shift degeneracy which is present in iH(CA)NCO and HA(CA)CON. The magnetization transfer pathway and the pulse sequence of (HCA)CON(CA)H experiment are presented in figure 4.6 and the coherence flow throughout the experiment can be described as follows:

$$^{1}H^{\alpha}(i-1) \xrightarrow{^{1}J_{H^{\alpha}C^{\alpha}}} {}^{13}C^{\alpha}(i-1) \xrightarrow{^{1}J_{H^{\alpha}C^{\alpha}},\, ^{1}J_{C^{\alpha}CO}} {}^{13}CO(i-1)[2T_A - t_1;\, ^{1}J_{CON}] \longrightarrow$$

$$^{15}N(i)[2T_{NC} - t_2;\, ^{1}J_{CON},\, ^{1}J_{C^{\alpha}N},\, ^{2}J_{C^{\alpha}N}] \longrightarrow {}^{13}C^{\alpha}(i) \xrightarrow{^{1}J_{H^{\alpha}C^{\alpha}},\, ^{1}J_{C^{\alpha}N}} {}^{1}H^{\alpha}(i)[t_3]$$

Active couplings used in magnetization transfer are listed above the arrows and the acquisition times of the corresponding spins t_{1-3} are in parentheses. For ^{13}CO, a constant time acquisition scheme is employed and the active heteronuclear couplings used for magnetization transfer during the constant time evolution period are also shown in the parentheses.

As indicated by the figure 4.6, magnetization transfer for $^{15}N \longrightarrow {}^{13}C^{\alpha}$ provides sequential and auto-correlated peaks (step d). These peaks have 180° phase difference and can be utilized during the resonance assignment procedure. For IDP studies these peaks exhibit inevitable overlap due to the well documented degeneracy of $^{13}C^{\alpha}$ shifts. However, desirable sequential peaks can be maximized while auto-correlated peaks are minimized by employing delay settings $T_{NC} = 25\,\text{ms}$ $T_A = 16.6\,\text{ms}$. In IDPs with average transverse relaxation times these delay settings result in theoretical transfer efficiencies of 0.181 and 0.004 for the sequential and auto-correlated cross peaks respectively. Theoretical increase in the attainable sensitivity for (HCA)CON(CA)H was calculated to be 10% with respect to H(CA)NCO for IDPs, and up to 50%, 40% and 65% for random coil, α-helical and extended conformations in medium sized globular proteins respectively.

Practical performance of (HCA)CON(CA)H experiment was evaluated on two proteins, immunoglobulin-binding domain B1 of streptococcal protein G (GB1) and can-

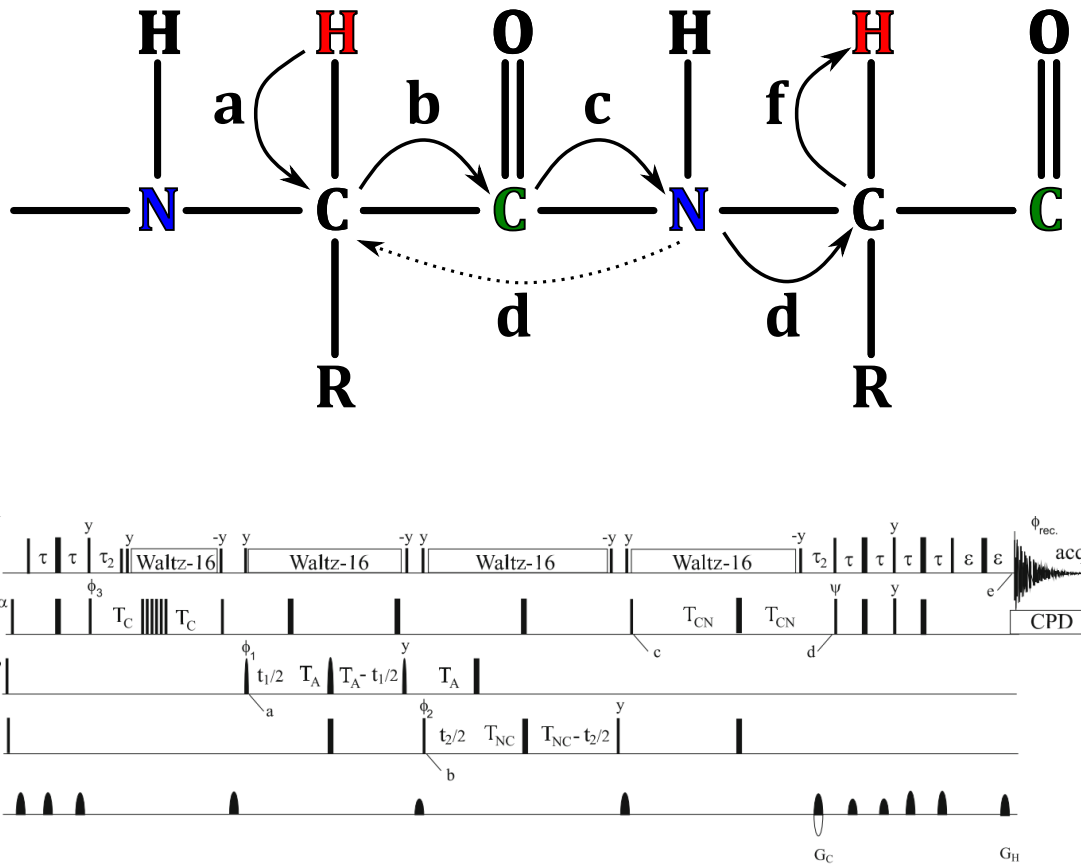


Figure 4.6: (HCA)CON(CA)H experiment establishes correlations between $^{1}H^{\alpha}(i)$, $^{15}N(i)$ and $^{13}C'(i-1)$ chemical shifts with delay $T_{NC} = 25$ ms. The upper part of the figure is the schematic presentation of magnetization transfer route in (HCA)CON(CA)H experiment, and the lower part is the experiment's pulse sequence. The solid lines indicate actual magnetization transfer pathway and dotted lines coherence flow which is suppressed by employing right delay settings. Letters a–f correlate to different points in the experiment's pulse sequence. If auto-correlated peaks want to be observed in addition to sequential ones, $^{15}N \longrightarrow {}^{13}C$ transfer delay T_{NC} should be set to 12 to 15 ms. All of the pulses are applied with phase x unless stated otherwise. Narrow and wide rectangles correspond to 90° and 180° pulses respectively. Delay durations: $\tau = (4J_{HC})^{-1} \sim 1.7$ ms; $\tau_2 = 3.4$ ms for non-glycine residues and 2.2 to 2.5 ms for also observing glycines; ϵ is duration of G_H plus field recovery time which is around 0.4 ms; $T_C = (6J_{C\alpha C'})^{-1} \sim 3.4$ ms; $T_A = (4J_{C'N})^{-1} \sim 16.6$ ms; $T_C N \sim 14$ ms. Maximum t_1 and t_2 are limited to $t_{1,max} < 2T_A$ and $t_{2,max} < 2T_{NC}$. Gradient strengths and durations: $G_C = 13\,\mathrm{kG\,cm^{-1}}$ (1.6 ms), $G_H = 13\,\mathrm{kG\,cm^{-1}}$ (0.4 ms). Phase cycling: $\phi_1 = x, -x$; $\phi_2 = 2(x), 2(-x)$; $\phi_3 = 4(x), 4(-x)$; $\psi = x$; $\phi_{rec} = 2, 2(x), x, -x, 2(x), -x$. Further details on the experiment can be found from the original article.[94]

cer/testis antigen CT16. The sensitivity of the (HCA)CON(CA)H experiment was compared to H(CA)NCO scheme on GB1 at 30 °C. In practice, (HCA)CON(CA)H has 40% increased sensitivity as compared to H(CA)NCO experiment.

The attained coherence transfer efficiency is comparable to those of $^{13}C'$-detected experiments introduced by Bermel *et al.*,[99] which have proven to be extremely useful in assignment of IDPs. HA-detection loses some of its eightfold sensitivity improvement as compared to ^{13}C-detection, due to doublet of doublets present in D_2O, which originate from ^3J-couplings to $^1H^\beta$s in residues other than Gly, Thr, Val, Ile, since they (excluding glycine) only have one $^1H^\beta$s. Final theoretical estimation is twofold sensitivity gain as compared to the corresponding ^{13}CO-detected (H)CANCO experiment.[94,99] However, in practice the sensitivity loss from from ^3J-couplings is not as significant as the doublet of doublets are buried under the line width of the final signal.

Part II

Experimental part

Chapter 5

Sequential assignment of BilRI

5.1 Aim of the study

The aim of the study was to assign BilRI's chemical shifts and to determine possible secondary structure characteristics based on secondary chemical shifts. Ahlstrand[16] *et al.* confirmed the intrinsic disorder of BilRI on the basis of poorly dispersed ^{1}H signals in methyl and amide proton regions. Poor signal dispersion and monotonous amino acid sequence prompted the use of NMR techniques employing various nuclei for detection i.e. HA, CO and HN in sequential assignment

5.2 Materials and methods

5.2.1 Isotopically labelled BilRI

Plasmid containing DNA sequence of BilRI was obtained from Riikka Ihalin's research group (University of Turku). The recombinant BilRI was expressed, isotope labelled and purified by the procedure described in the original articles.[15,16] A 0.5 mM sample of BilRI was used in all of the measurements except for four HN-detected experiments—HN(CO)CACB, HNCACB, HNCO and iHNCO—where a 1 mM sample was used instead. All of the samples had a 50 mM phosphate buffer of pH 6.5 and 50 mM NaCl added.

Table 5.1: Acquisition times of different nuclei and number of scans in 11 NMR experiments

Experiment	Aq. time ^{15}N (ms)	Aq. time ^{13}C (ms)	Aq. time ^{1}H (ms)	No. scans
^{15}N-HSQC	45.75	-	79.87	2
^{13}C-CT -HSQC	-	50.94	79.87	8
CON	83.95	126.98	-	16
HNCACB[a]	21.82	4.27	91.75	8
HN(CO)CACB[a]	21.82	4.27	91.75	8
iHNCO/HNCO[a]	21.82	17.14	91.75	8/4
iH(CA)NCO[a]	50.00	24.02	80.01	8
H(CA)CON[a]	50.00	47.62	80.01	8
(HCA)CON(CA)H[a]	50.00	32.30	80.01	8
(HCA)CONCAH[b]	14.55	6.86[c]/13.26	80.01	4

[a] Non-uniform sampling amount 25%
[b] Non-uniform sampling amount 10%
[c] For 4th CA dimension

5.2.2 NMR spectroscopy

NMR spectra used for resonance assignment

Eleven different spectra of BilRI were measured using a Bruker Avance III HD 800 NMR spectrometer equipped with a cryogenically cooled ^{1}H, ^{13}C and ^{15}N triple-resonance probehead with z-gradient coil at the University of Jyväskylä. Basic measurement parameters for each of the different NMR experiments are presented in table 5.1 and all of the spectra were collected at 298 K.

Data analysis

Spectra were processed with Topspin 3.5pl and the sequential assignment was carried out with NMRFAM-SPARKY[100] software package. Secondary structure of BilRI was analysed by neighbour corrected structural propensity calculator (ncSPC).[101]

5.3 Results and discussion

5.3.1 Two-dimensional ^{15}N-HSQC, ^{13}C-CT-HSQC and ^{15}N, ^{13}C-CON spectra of BilRI

Whichever the three different 2D spectra—the two HSQC experiments or the CON experiment—could be used as a root spectrum for the sequential assignment of BilRI.

The different spectra are presented in 5.1 with the ^{13}C-CT-HSQC spectral area split into two parts for $^{13}C^{\alpha}$ and $^{13}C^{\beta}$. From this comparison alone it is clear that the CON spectrum has the best peak dispersion on top of more meaningful implication of the peak—each peak corresponding to a peptide bond instead of a proton attached to a heteroatom. Because of this, both of the dimensions in CON spectrum are not necessarily attached to the same residue type. For example peptide bonds between Ala and Ser or Ala and Gly yield significantly different peak position. This does not, however, alleviate the peak overlap caused by the lack of structure in IDPs in conjunction with repeats in the sequence.

^{15}N-HSQC shows all H – N connections, of which the most important for the backbone resonance assignment are the ones originating from the backbone amide groups. Additionally, peaks originating from Trp, Asn and Gln side-chains are present in the spectral region of the backbone amide chemical shifts.

^{13}C-CT-HSQC shows all H – C connections with three characteristic spectral areas for CH_3, CH_2 and CH signals in this order from upfield to downfield. $^{13}C^{\alpha}$ chemical shifts are the most downfield with the exception of Ser and Thr $^{13}C^{\beta}$ chemical shifts, as these carbons are directly bound to an OH-group. Most of the $^{13}C^{\beta}$s are in the CH_2-region with the exception of Ala as its $^{13}C^{\beta}$ is in the CH_3-region (the most upfield). ^{13}C-HSQC was mostly used for qualitative analysis of $^{13}C^{\beta}$ chemical shift distributions, but could be useful as a basis of picking a 3D ^{13}C NOESY spectrum.

5.3.2 Sequential assignment

The sequential backbone assignment of the 160-residue polypeptide chain of BilRI yielded chemical shift values for backbone ^{13}CO, ^{15}N and $^{1}H^{\alpha}$. Unfortunately, $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ chemical shifts could not reliably be determined for all of the residues using HN-detected 3D spectra because of significant overlap of amide proton chemical shifts. However, these chemical shift values can be obtained with relative ease using other experiments and the already assigned backbone chemical shifts in future studies. For example, a triple-resonance experiment that could be used to determine $^{13}C^{\alpha}$s from $^{1}H^{\alpha}$ and ^{13}CO chemical shifts is iHCACO[102] or the 4D (HCA)CONCAH could also be used.

The high resolution 2D CON spectrum was selected as the root spectrum in order to bypass the poor signal dispersion of HN-detected experiments such as ^{15}N-HSQC and possible line broadening resulting from amide proton chemical exchange, as discussed in the literary part of this thesis. Comparison of the three 2D spectra is presented in figure 5.1. BilRI's sequence is 160 AA long, so 159 backbone peaks were expected in the 2D CON spectrum, as each peptide bond produces a peak in the spectrum. In addition

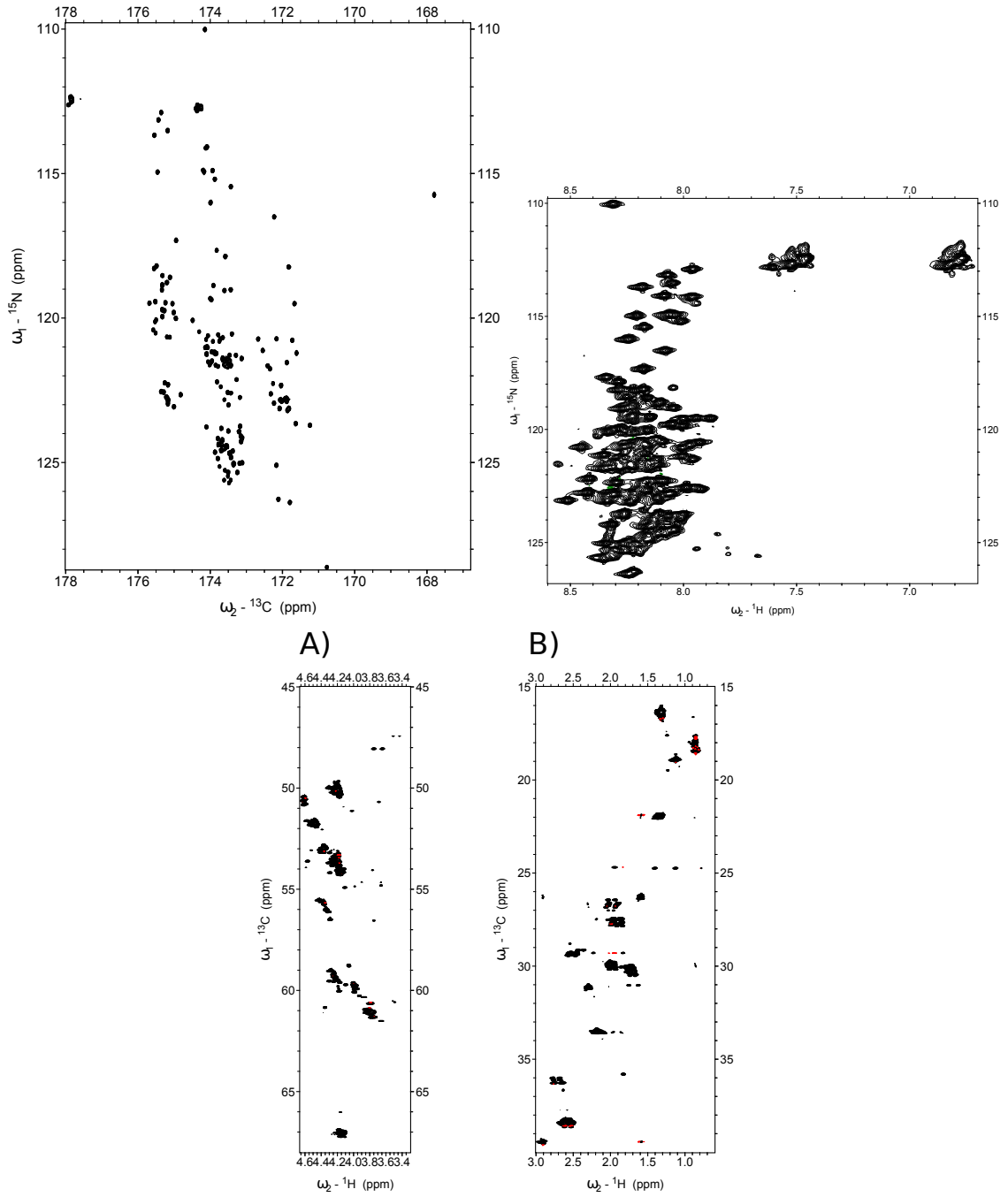


Figure 5.1: Comparison of the chemical shift dispersions in three 2D spectra utilizing different nuclei for detection. ^{13}C-detected CON on the top left, H-detected ^{15}N-HSQC on the top right and two parts of H-detected ^{13}C-CT-HSQC spectrum on the bottom. A) and B) correspond to $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ spectral areas in ^{13}C-CT-HSQC respectively.

to the backbone peaks, Asn and Gln side-chains were expected to produce a total of 11 additional peaks (5 and 6 respectively). Initial peak picking resulted in 157 peaks, suggesting some overlap due to the repeats. The side-chain peaks were easily recognizable at their characteristic chemical shifts (Asn 112.76,176.79, Gln 111.86,179.73 ppm) and clearly separated from the main spectral area and additionally could not be observed in higher dimensional spectra. After removing the side-chain peaks, 150 peaks remained with some of the larger peaks suggesting peak overlap. Considering that over 50% of BilRIs AA composition is comprised of only four different residue types (Ala, Asp, Lys and Glu), and that it has four 10 AA long repeats, 2D CON spectrum resolution is definitely satisfactory.

The backbone assignment was mainly done by utilizing three HA-detected experiments—iH(CA)NCO,(HCA)CON(CA)H and H(CA)CON—with the general assignment strategy shown in figure 5.2. By first selecting a cross-peak for an intraresidual amide nitrogen, carbonyl and α-proton in iH(CA)NCO, a cross-peak can be found in (HCA)CON(CA)H containing two of the same resonances (N_i and H_i^α) the third being sequential CO_{i-1} resonance. From here a cross-peak in H(CA)CON (N_i,CO_{i-1}, H_{i-1}^α) can be selected containing sequential H_{i-1}^α. Finally intraresidual correlation of residue N_{i-1} in iH(CA)NCO (N_{i-1},CO_{i-1},H_{i-1}^α) can be found at same CO_{i-1} and H_{i-1}^α frequencies assigned with the H(CA)CON spectrum. Of course the spectra can also be used in reverse order to move along the AA backbone towards the C-terminus of the sequence.

At first automated peak picking was tested using root spectrum peaks to find corresponding peaks in (HCA)CON(CA)H and HA(CA)CON. However this procedure failed, as setting meaningful peak picking thresholds turned out to be nearly impossible. This meant that semi-automated assignment, with the use of strip-plots and automatically picked peaks, could not be carried out. Instead every peak in the triple-resonance spectra were manually picked according to the high resolution CON carbonyl carbon and nitrogen frequencies. This left the selection of $^1H^\alpha$ chemical shifts to sometimes be somewhat ambiguous, as some of the peak clusters were as wide as 0.15 ppm in the $^1H^\alpha$ dimension and contained many peaks. This means that the maximum intensity of a peak was not always the actual peak position, but a peak maximum caused by constructive sum intensity of many peaks in the same spectral area. Failure to assign the appropriate $^1H^\alpha$ sometimes caused a hop to a similar repeat at a different part of the sequence. This made the actual assignment process very tedious and caused a lot of backtracking, whenever previous errors were found out.

The remaining 3D spectra—HNCO, iHNCO, HNCACB and HN(CO)CACB—were mainly used for amino acid residue recognition. All of these experiments use the HN-detection and the amide proton chemical shift of any residue could be obtained from the known nitrogen and carbonyl chemical shifts using HNCO and iHNCO. After this α- and β-carbon chemical shifts could be determined for $i-1$ and i residues with HNCACB

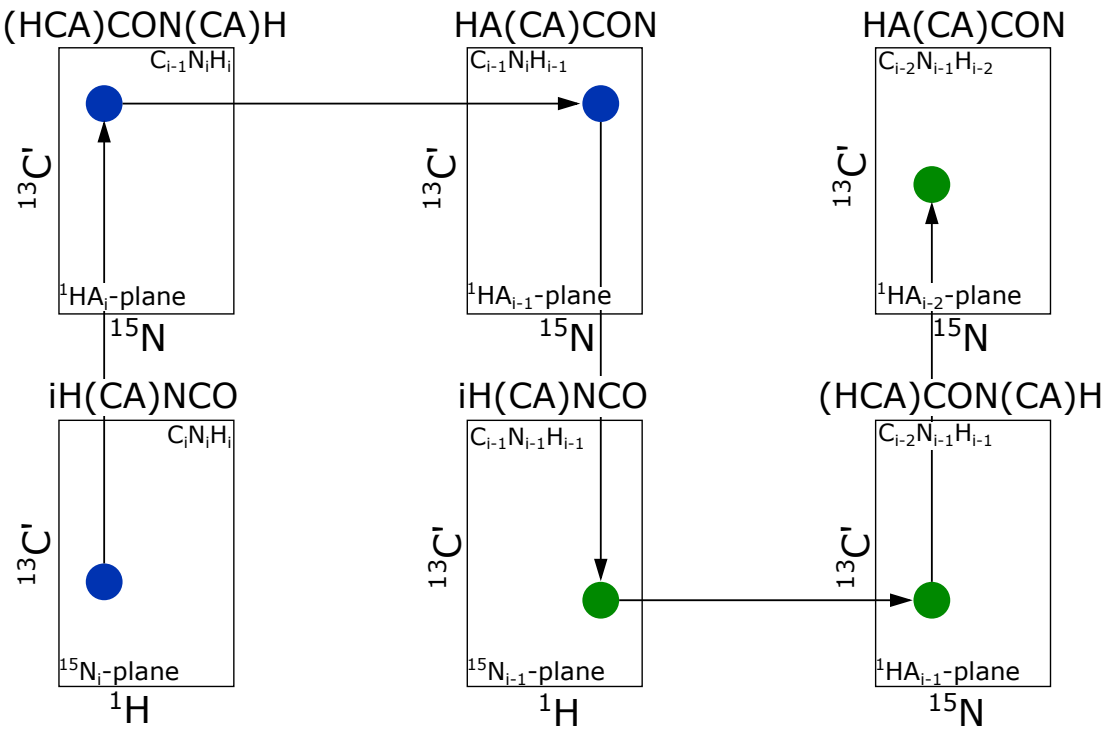


Figure 5.2: Schematic presentation of sequential assignment strategy based on iH(CA)NCO, (HCA)CON(CA)H and HA(CA)CON spectra. Nitrogen cross peak frequencies of residues i and $i-1$ are coloured blue and green respectively. Schematic was made with InkScape.[28]

and HN(CO)CACB. These chemical shifts, especially C^{β}, were used as indicators of the residue type in less crowded spectral areas i.e. outside the repeats.

Chemical shift reference correction was done to all spectra used in the assignment by setting the resonance values of these spectra to match the corresponding in the root spectrum. Only one glycine is present in BilRIs sequence and it was easily found by exploiting the fact that glycine peaks have a 180 degree phase difference to all other peaks in iH(CA)NCO. This peak was used for the reference correction and as a starting point for the assignment process.

The assignment was carried out by selecting a CON signal, finding the corresponding H(CA)CON/(HCA)CON(CA)H peak and forming assigned fragments by assigning amino acid residues based on the general strategy explained in 5.2. Assignment of the fragments was stopped at ambiguous signals in both N- and C-terminus. In the beginning most remote CON signals were selected, as they were outside the repeats i.e. crowded spectral areas. Ser and Thr have their C^{β} clearly more downfield than other AAs have their C^{α} chemical shifts, and they were used as anchors to recognize from which part of the sequence the fragment originates. HNCACB shows the amino acid type of i and $i-1$, which proved invaluable at finding the locations of the assigned fragments. As an example there are only one Thr-Thr and one Thr-Ser connection in the BilRI sequence and they could be easily found with HNCACB analysis. In addition

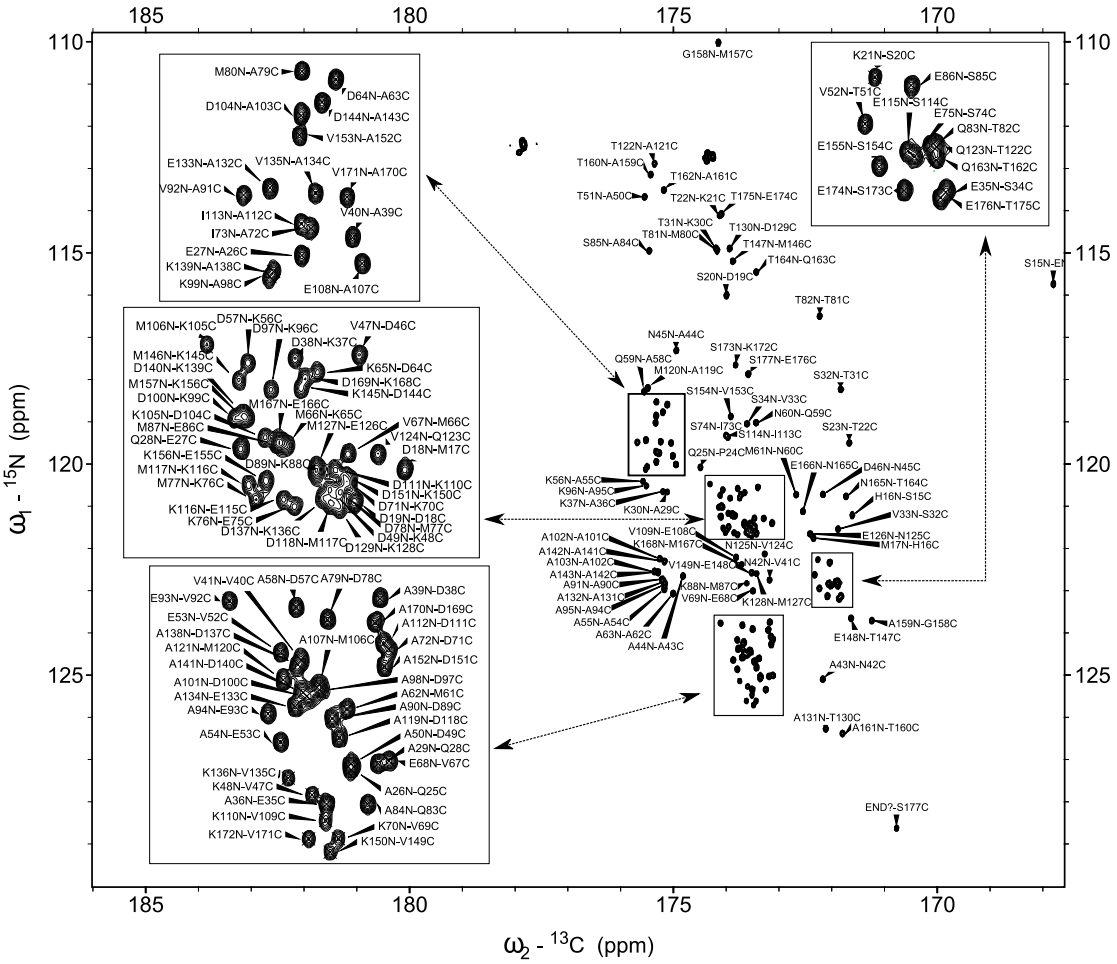


Figure 5.3: 2D CON spectrum with 163 assigned peaks. The most crowded areas are enlarged. Both N- and C-terminus of the polypeptide chain give a signal and are marked as END in the spectrum.

to Ser and Thr, five Asn proved to be invaluable anchors, as their C^{β} chemical shifts are clearly separate due to the neighbouring carbonyl group. 4D (HCA)CONCAH was only used in couple of occasions, mainly to distinguish lysines from methionines in crowded spectral areas, as their C^{β} shifts are relatively similar, but C^{α} had up to 1 ppm difference.

As the assignment process was carried out, yielding more and more fragments, neighbouring fragments could be connected by assigning ambiguous intraresidual resonances between them from both N- and C-terminus direction. The most problematic parts proved to be the repeating Ala-Lys-Gln connections. Their assignment was done the last when their surroundings were completely known and possible mistakes could be detected in a timely manner.

In the end the backbone assignment was carried out to 100% completion. In addition to peaks originating from the sequence of BilRI, four additional amino acids were found at the N-terminus of the protein. These amino acids, GSHM, originate from thrombin cleavage of a His-tag during sample preparation.[16] All and all this results

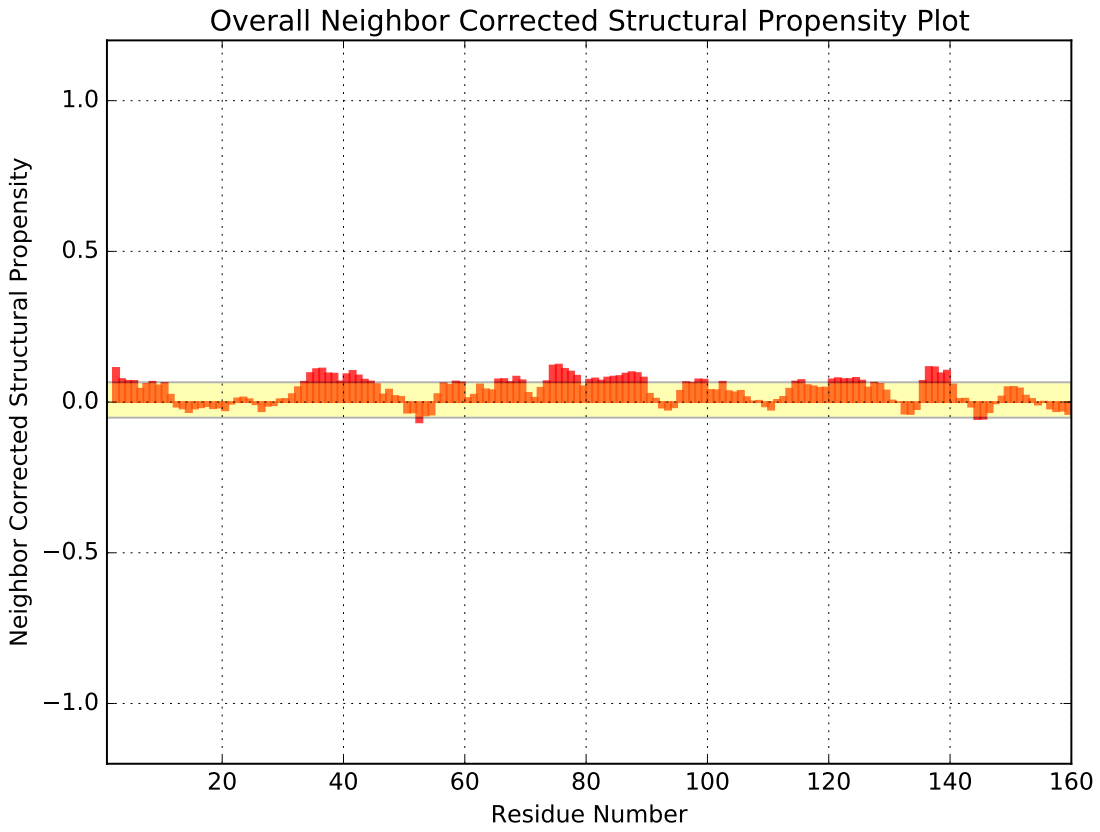


Figure 5.4: Overall structural propensity of BilRI. Yellow area indicates definite random coil conformation, with increasing propensities to α-helical (positive values) and β-sheet (negative values) character.

in 163 peaks and the final 2D CON spectrum with each of the resonances assigned is presented in figure 5.3. As expected, the most crowded areas are caused by peptide bonds between the most prominent amino acid residues. Additionally the two main repeats had almost overlapping peaks for up to 10 sequential peaks.

5.3.3 Analysis of secondary structure in BilRI

Initial secondary structure analysis of BilRI was done by SCS analysis covered in 4.2.1. The ncSPC calculator[101] uses sequence specific random coil values for IDPs from an article by Tamiola *et al.*[69] from 2010. This way ncSPC is equipped to predict the possible transient secondary structures in IDPs.

The overall structural propensity plot is presented in figure 5.4 and the rest of the propensity plots are presented in appendix B. Additionally, experimental vs predicted chemical shift plots for each backbone nuclei are presented in appendix C with correlation of determination (R^2) and root-mean-square deviation (RMSD).

ncSPC analysis shows that BilRI is indeed an intrinsically disordered protein with some transient α-helicity at residues 35–44, 74–79 and 135–140. All these parts of the

amino acid sequence are dissimilar, meaning that no conclusion on the neighbouring amino acid types affecting α-helical propensity can be drawn. Omitting the nitrogen chemical shifts from the overall ncSPC calculations could elucidate the transient secondary structure of BilRI even further, as nitrogen SCSs are not as reliable indicators of secondary structure as carbonyl carbon SCSs in the case of IDPs. Addition of C^{α} chemical shifts to the calculations would do the most to increase the accuracy of the prediction, as these chemical shifts are ranked as the most reliable way to predict α-helicity and the second most reliable way for β-sheets.

The referencing of the chemical shifts was successful, as the correlations between the experimental and the predicted chemical shifts for each nucleus was high, and the RMSD values relatively low. If significant systematic error was present, the plot would show the data points clearly to the side of the central blue line.

Chapter 6

Conclusions

Using HA-detected triple-resonance experiments in conjunction with 2D CON root spectrum proved to be an effective way to assign the backbone $^{1}H^{\alpha}$, ^{15}N and ^{13}CO chemical shifts of intrinsically disordered BilRI. Even with automated assignment software constantly being developed, it seems that the sequential resonance assignment of overlapping peaks present in IDPs requires human eye and cognitive abilities. Although, processing many overlapping connections inside the repeats made the assignment process tedious and time consuming, as some of the ambiguous connections had to be partly resolved by trial and error method, ultimately the ability to use up to ten different spectra in parallel made the process very reliable.

The ncSPC calculations confirmed the random coil character of BilRI, but also hinted the existence of some α-helicity in some parts of the sequence. These findings could be validated further with inclusion of C^{α} and C^{β} chemical shifts into the ncSPC calculations. The secondary structure characteristics of BilRI should be experimentally validated in the future with the use of for example residual dipolar coupling or scalar coupling experiments.

BilRI is a stable and well-behaved protein, and almost completely intrinsically disordered. With the backbone chemical shift assignment completed, BilRI can be used as a reference protein for the development of new NMR experiments aiming to combat the chemical shift degeneracy in IDPs.

Bibliography

[1] I. I. Rabi, "Space Quantization in a Gyrating Magnetic Field", *Phys. Rev.* **1937**, 652–654.

[2] I. I. Rabi, J. R. Zacharias, S. Millman, P. Kusch, "A New Method of Measuring Nuclear Magnetic Moment", *Phys. Rev.* **1938**, 318.

[3] E. L. Purcell, H. C. Torrey, R. V. Pound, "Resonance Absorption by Nuclear Moments in a Solid", *Phys. Rev.* **1946**, 37–38.

[4] F. Bloch, W. W. Hansen, M. Packard, "The nuclear induction experiment", *Phys. Rev.* **1946**, 474–485.

[5] C. Boesch, "Nobel Prizes for nuclear magnetic resonance: 2003 and historical perspectives.", *J. Magn. Reson. imaging* **2004**, *20*, 177–179, DOI 10.1002/jmri.20120.

[6] L. Muller, A. Kumar, R. R. Ernst, "Two dimensional carbon-13 NMR spectroscopy", *J. Chem. Phys.* **1975**, *37*, 5490–5491.

[7] W. O. Aue, B. Bartholdi, R. R. Ernst, "Two-dimensional spectroscopy: application to nuclear magnetic resonance", *J. Chem. Phys.* **1976**, *64*, 2229–2246.

[8] K. Wüthrich, "The second decade — into the third millenium", *Nat. Struct. Biol.* **1998**, *5*, 492–495, DOI 10.1038/728.

[9] C. Li, M. Liu, "Protein dynamics in living cells studied by in-cell NMR spectroscopy", *FEBS Letters* **2013**, *587*, 1008–1011.

[10] P. Tompa, *Structure and function of intrinsically disordered proteins*, Chapman and Hall/CRC Press, **2009**, p. 331.

[11] M. Wells, H. Tidow, T. J. Rutherford, P. Markwick, M. R. Jensen, E. Mylonas, A. R. Fersht, "Structure of tumor suppressor p53 and its intrinsically disordered N-terminal transactivation domain", *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 5762–5767, DOI 10.1073/pnas.0801353105.

[12] P. H. Weinreb, W. Zhen, A. W. Poon, K. A. Conway, P. T. J. Lansbury, "NACP, a protein implicated in Alzheimer's disease and learning, is natively unfolded", *Biochemistry* **1996**, *35*, 13709–13715, DOI 10.1021/bi961799n.

[13] M. Madan Babu, R. Van Der Lee, N. Sanchez De Groot, J. Rg Gsponer, J. Gough, K. Dunker, "Intrinsically disordered proteins: regulation and disease This review comes from a themed issue on Sequences and topology Edited", *Curr. Opin. Struct. Biol.* **2011**, *21*, 1–9, DOI 10.1016/j.sbi.2011.03.011.

[14] L. Suominen-Taipale, A. Nordblad, M. Vehkalahti, A. Aromaa, "Suomalaisten aikuisten suunterveys Terveys 2000-tutkimus", *Publications of National Public Health Institute* **2004**.

[15] A. Paino, T. Ahlstrand, J. Nuutila, I. Navickaite, M. Lahti, H. Tuominen, H. Välimaa, U. Lamminmäki, M. T. Pöllänen, R. Ihalin, "Identification of a Novel Bacterial Outer Membrane Interleukin-1β-Binding Protein from *Aggregatibacter actinomycetemcomitans*", *PLoS One* **2013**, *8*, (Ed.: J. A. Bengoechea), e70509, DOI 10.1371/journal.pone.0070509.

[16] T. Ahlstrand, H. Tuominen, A. Beklen, A. Torittu, J. Oscarsson, R. Sormunen, M. T. Pöllänen, P. Permi, R. Ihalin, "A novel intrinsically disordered outer membrane lipoprotein of *Aggregatibacter actinomycetemcomitans* binds various cytokines and plays a role in biofilm response to interleukin-1β and interleukin-8", *Virulence* **2017**, *8*, 115–134, DOI 10.1080/21505594.2016.1216294.

[17] E. Fisher, "Einfuss der Configuration auf die Wirkung der Enzyme", *Ber. Dt. Chem. Ges.* **1894**, *27*, 2985–2993.

[18] F. Karush, "Heterogeneity of the Binding Sites of Bovine Serum Albumin", *J. Am. Chem. Soc.* **1950**, *72*, 2705–2713, DOI 10.1021/ja01162a099.

[19] A. Dunker, J. Lawson, C. J. Brown, R. M. Williams, P. Romero, J. S. Oh, C. J. Oldfield, A. M. Campen, C. M. Ratliff, K. W. Hipps, J. Ausio, M. S. Nissen, R. Reeves, C. Kang, C. R. Kissinger, R. W. Bailey, M. D. Griswold, W. Chiu, E. C. Garner, Z. Obradovic, "Intrinsically disordered protein", *J. Mol. Graph. Model.* **2001**, *19*, 26–59, DOI 10.1016/S1093-3263(00)00138-8.

[20] V. N. Uversky, J. R. Gillespie, A. L. Fink, "Why are "natively unfolded" proteins unstructured under physiologic conditions?", *Proteins Struct. Funct. Genet.* **2000**, *41*, 415–427, DOI 10.1002/1097-0134(20001115)41:3<415::AID-PROT130>3.0.CO;2-7.

[21] P. E. Wright, H. Dyson, "Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm", *J. Mol. Biol.* **1999**, *293*, 321–331, DOI 10.1006/jmbi.1999.3110.

[22] P. Tompa, "The interplay between structure and function in intrinsically unstructured proteins", *FEBS Lett.* **2005**, *579*, 3346–3354, DOI 10.1016/j.febslet.2005.03.072.

[23] A. Dunker, M. Cortese, P. Romero, L. Iakoucheva, V. Uversky, "Flexible nets. The roles of intrinsic disorder in protein interaction networks", *FEBS J.* **2005**, *272*, 5129–5148.

[24] L. M. Iakoucheva, C. J. Brown, J. D. Lawson, Z. Obradovic, A. K. Dunker, "Intrinsic disorder in cell-signalling and cancer-associated proteins", *J. Mol. Biol.* **2002**, *323*, 573–584.

[25] P. Tompa, "Intrinsically disordered proteins: a 10-year recap", *Trends Biochem. Sci.* **2012**, *37*, 509–516, DOI 10.1016/j.tibs.2012.08.004.

[26] R. Pancsa, P. Tompa, "Structural Disorder in Eukaryotes", *PLoS One* **2012**, *7*, 1–10, DOI 10.1371/journal.pone.0034687.

[27] R. Garret, C. Grisham, *Biochemistry*, *Vol. 4*, 4th ed., Mary Finch, **2008**, pp. 1–851, DOI 10.1016/B978-0-12-387000-1.01001-9.

[28] Inkscape Project, *Inkscape*, version 0.92.2, **2017**, https://inkscape.org.

[29] G. D. Rose, P. J. Fleming, J. R. Banavar, A. Maritan, "A backbone-based theory of protein folding.", *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 16623–33, DOI 10.1073/pnas.0606843103.

[30] J. S. Richardson, Protein backbone PhiPsiOmega drawing, Used under Creative Commons Attribution 3.0 Unported license, **2013**, https://commons.wikimedia.org/wiki/File:Protein_backbone_PhiPsiOmega_drawing.svg.

[31] J. S. Richardson, Ramachandran plot general 100K, Used under Creative Commons Attribution 3.0 Unported license, **2003**, https://commons.wikimedia.org/wiki/File:Ramachandran_plot_general_100K.jpg.

[32] Ian Fleming, *Molecular Orbitals and Organic Chemical Reactions*, Student Ed, A John Wiley and Sons Ltd. Publication, **2009**, p. 360.

[33] B. Wang, W. Jiang, X. Dai, Y. Gao, Z. Wang, R. Q. Zhang, "Molecular orbital analysis of the hydrogen bonded water dimer", *Sci. Rep.* **2016**, *6*, 2–7, DOI 10.1038/srep22099.

[34] M. D. S. Kumar, K. A. Bava, M. M. Gromiha, P. Prabakaran, K. Kitajima, H. Uedaira, A. Sarai, "ProTherm and ProNIT: thermodynamic databases for proteins and proteinnucleic acid interactions", *Nucleic Acids Res.* **2006**, *34*, D204–D206, DOI 10.1093/nar/gkj103.

[35] A. Bakk, J. S. Høye, A. Hansen, "Apolar and polar solvation thermodynamics related to the protein unfolding process", *Biophys. J.* **2002**, *82*, 713–719, DOI 10.1016/S0006-3495(02)75433-7.

[36] N. V. Prabhu, K. A. Sharp, "Heat Capacity in Proteins", *Annu. Rev. Phys. Chem.* **2005**, *56*, 521–548, DOI 10.1146/annurev.physchem.56.092503.141202.

[37] V. N. Uversky, “Intrinsically disordered proteins from A to Z”, *Int. J. Biochem. Cell Biol.* **2011**, *43*, 1090–1103, DOI 10.1016/j.biocel.2011.04.001.

[38] F.-X. Theillet, K. KAlmar, P. Tompa, K. H. Han, P. Selenko, A. K. Dunker, V. N. Uversky, “The alphabet of intrinsic disorder”, *Intrinsically Disord. Proteins* **2013**, *1*, e24684, DOI 10.4161/idp.24684.

[39] B. Mészáros, I. Simon, Z. Dosztányi, “The expanding view of protein-protein interactions: Complexes involving intrinsically disordered proteins”, *Phys. Biol.* **2011**, *8*, DOI 10.1088/1478-3975/8/3/035003.

[40] P. A. Webb, O. Persic, C. E. Mendola, J. M. Backer, R. L. Williams, “The crystal structure of human nucleoside diphosphate kinase, NM23-H2”, *J. Mol. Bio.* **1995**, *251*, 574–587.

[41] I. Drobnak, N. De Jonge, S. Haesaerts, G. Vesnaver, R. Loris, J. Lah, “Energetic basis of uncoupling folding from binding for an intrinsically disordered protein.”, *J. Am. Soc.* **2013**, *135*, 1288–1294.

[42] V. M. Burger, T. Gurry, C. M. Stultz, “Intrinsically disordered proteins: Where computation meets experiment”, *Polymers* **2014**, *6*, 2684–2719, DOI 10.3390/polym6102684.

[43] C. Levinthal, How to fold Graciously, Mossbauer Spectroscopy in Biological Systems: Proceedings of a meeting held at Allerton House, Monticello, Illinois, **1969**.

[44] J. Chen, “Towards the physical basis of how intrinsic disorder mediates protein function”, *Arch. Biochem. Biophys.* **2012**, *524*, 123–131, DOI 10.1016/j.abb.2012.04.024.

[45] Y. Huang, Z. Liu, “Kinetic Advantage of Intrinsically Disordered Proteins in Coupled Folding-Binding Process: A Critical Assessment of the "Fly-Casting" Mechanism”, *J. Mol. Biol.* **2009**, *393*, 1143–1159, DOI 10.1016/j.jmb.2009.09.010.

[46] Z. Liu, Y. Huang, “Advantages of proteins being disordered”, *Protein Sci.* **2014**, *23*, 539–550, DOI 10.1002/pro.2443.

[47] V. N. Uversky, “A decade and a half of protein intrinsic disorder: Biology still waits for physics”, *Protein Sci.* **2013**, *22*, 693–724, DOI 10.1002/pro.2261.

[48] D. Eliezer, P. Tompa, C. Bracken, L. M. Iakoucheva, P. R. Romero, a. K. Dunker, P. Bernadó, D. I. Svergun, “Structural analysis of intrinsically disordered proteins by small-angle X-ray scattering”, *Curr. Opin. Struct. Biol.* **2002**, *14*, 570–576, DOI 10.1016/S0968-0004(02)02169-2.

[49] R. T. Richardson, O. M. Alekseev, G. Grossman, E. E. Widgren, R. Thresher, E. J. Wagner, K. D. Sullivan, W. F. Marzluff, M. G. O'Rand, "Nuclear autoantigenic sperm protein (NASP), a linker histone chaperone that is required for cell proliferation", *J. Biol. Chem.* **2006**, *281*, 21526–21534, DOI 10.1074/jbc.M603816200.

[50] K. Gunasekaran, C. J. Tsai, S. Kumar, D. Zanuy, R. Nussinov, "Extended disordered proteins: Targeting function with less scaffold", *Trends Biochem. Sci.* **2003**, *28*, 81–85, DOI 10.1016/S0968-0004(03)00003-3.

[51] C. J. Jeffery, "Moonlighting proteins", *Trends Biochem. Sci.* **1999**, *24*, 8–11, DOI 10.1016/S0968-0004(98)01335-8.

[52] B. Xue, V. Jeffers, W. Sullivan, V. N. Uversky, "Protein intrinsic disorder in acetylome of intracellular and extracellular Toxoplasma gondii", *Mol. Biosyst.* **2013**, *9*, 645–657, DOI 10.1039/c3mb25517d.

[53] J. Shao, D. Xu, S. N. Tsai, Y. Wang, S. M. Ngai, "Computational Identification of Protein Methylation Sites through Bi-Profile Bayes Feature Extraction", *PLoS One* **2009**, *4*, DOI 10.1371/journal.pone.0004920.

[54] J. Gsponer, M. E. Futschik, S. Teichmann, M. M. Babu, "Tight regulation of unstructured proteins: from transcript synthesis to protein degradation.", *Science* **2008**, *322*, 1365–1368, DOI 10.1126/science.1163581.

[55] E. I. Weiss, B. Shaniztki, M. Dotan, N. Ganeshkumar, P. E. Kolenbrander, Z. Metzger, "Attachment of *Fusobacterium nucleatum* PK1594 to mammalian cells and its coaggregation with periodontopathogenic bacteria are mediated by the same galactose-binding adhesin.", *Oral Microbiol. Immunol.* **2000**, *15*, 371–377, DOI 10.1034/j.1399-302x.2000.150606.x.

[56] E. A. Izano, I. Sadovskaya, H. Wang, E. Vinogradov, C. Ragunath, N. Ramasubbu, S. Jabbouri, M. B. Perry, J. B. Kaplan, "Poly-N-acetylglucosamine mediates biofilm formation and detergent resistance in Aggregatibacter actinomycetemcomitans", *Microb. Pathog.* **2008**, *44*, 52–60, DOI 10.1016/j.micpath.2007.08.004.

[57] S. Kanangat, A. Postlethwaite, S. Cholera, L. Williams, D. Schaberg, "Modulation of virulence gene expression in *Staphylococcus aureus* by interleukin-1β: novel implications in bacterial pathogenesis ", *Microbes Infect.* **2007**, *9*, 408–415, DOI 10.1016/j.micinf.2006.12.018.

[58] M. Okshevsky, R. L. Meyer, "The role of extracellular DNA in the establishment, maintenance and perpetuation of bacterial biofilms", *Crit Rev Microbiol* **2015**, *41*, 341–352, DOI 10.3109/1040841x.2013.841639.

[59] M. Levitt, *Spin Dynamics: Basics of Nuclear Magnetic Resonance*, **2000**, p. 679, DOI 10.1002/cmr.a.20130.

[60] A. Abragam, *The Principles of Nuclear Magnetism*, Clarendon Press, **1961**, p. 599.

[61] K. J. Rosman, P. D. Taylor, "Isotopic compositions of the elements 1997", *J. Phys. Chem. Ref. Data* **1998**, *27*, 1275–1287, DOI 10.1063/1.556031.

[62] J. Cavanagh, W. Fairbrother, A. Palmer III, M. Rance, N. Skelton, *Protein NMR Spectroscopy*, Elsevier Academic Press, London, **2007**, p. 885.

[63] S. Kosol, S. Contreras-Martos, C. Cedeño, P. Tompa, "Structural characterization of intrinsically disordered proteins by NMR spectroscopy", *Molecules* **2013**, *18*, 10802–10828, DOI 10.3390/molecules180910802.

[64] W. Bermel, M. Bruix, I. C. Felli, V. Kumar M. V., R. Pierattelli, S. Serrano, "Improving the chemical shift dispersion of multidimensional NMR spectra of intrinsically disordered proteins", *J. Biomol. NMR* **2013**, *55*, 231–237, DOI 10.1007/s10858-013-9704-3.

[65] H. Kovacs, D. Moskau, M. Spraul, "Cryogenically cooled probes — a leap in NMR technology", *Porg. Nucl. Mag. Res. Sp.* **2005**, *46*, 131–155.

[66] P. Permi, A. Annila, "Coherence transfer in proteins", *Prog. Nucl. Magn. Reson. Spectrosc.* **2004**, *44A*, 97–137, DOI 10.1016/j.pnmrs.2003.12.001.

[67] M. Kjaergaard, A. B. Nørholm, R. Hendus-Altenburger, S. F. Pedersen, F. M. Poulsen, B. B. Kragelund, "Temperature-dependent structural changes in intrinsically disordered proteins: Formation of α-helices or loss of polyproline II?", *Protein Sci.* **2010**, *19*, 1555–1564, DOI 10.1002/pro.435.

[68] M. Kjaergaard, F. M. Poulsen, "Sequence correction of random coil chemical shifts: Correlation between neighbor correction factors and changes in the Ramachandran distribution", *J. Biomol. NMR* **2011**, *50*, 157–165, DOI 10.1007/s10858-011-9508-2.

[69] K. Tamiola, B. Acar, F. A. A. Mulder, "Sequence-specific random coil chemical shifts of intrinsically disordered proteins", *J. Chem. Soc.* **2010**, *132*, 18000–3, DOI 10.1042/BST20120171.

[70] M. Karplus, "Contact Electron Spin Coupling of Nuclear Magnetic Moments", *J. Chem. Phys.* **1959**, *30*, 11–15, DOI 10.1063/1.1729860.

[71] P. Permi, "Determination of three-bond scalar coupling between ^{13}C' and ^{1}H$^{\alpha}$ in proteins using an iHN(CA), CO(α/β-J-COHA) experiment.", *J. Magn. Reson.* **2003**, *163*, 114–120, DOI 10.1016/S1090-7807(03)00079-X.

[72] A. Petit, S. J. F. Vincent, C. Zwahlen, "Cosine Modulated HSQC: A Rapid Determination of $^{3}J_{HNH\alpha}$ Scalar Couplings in ^{15}N-labeled Proteins", *J. Magn. Reson.* **2002**, *156*, 313–317, DOI 10.1006/jmre.2002.2538.

[73] H. Ponstingl, G. Otting, "Rapid measurement of scalar three-bond $^{1}H^{N}$—$^{1}H^{\alpha}$ spin coupling constants in ^{15}N-labelled proteins.", *J. Biomol. NMR* **1998**, *12*, 319–324.

[74] J. L. Battiste, G. Wagner, "Utilization of site-directed spin labeling and high-resolution heteronuclear nuclear magnetic resonance for global fold determination of large proteins with limited nuclear overhauser effect data", *Biochemistry* **2000**, *39*, 5355–5365, DOI 10.1021/bi000060h.

[75] J. R. Gillespie, D. Shortle, "Characterization of long-range structure in the denatured state of staphylococcal nuclease. I. paramagnetic relaxation enhancement by nitroxide spin labels", *J. Mol. Biol.* **1997**, *268*, 158–169, DOI 10.1006/jmbi.1997.0954.

[76] A. G. Palmer, C. D. Kroenke, J. Patrick Loria, "Nuclear Magnetic Resonance Methods for Quantifying Microsecond-to-Millisecond Motions in Biological Macromolecules", *Methods Enzymol.* **2001**, *339*, 204–238, DOI 10.1016/S0076-6879(01)39315-1.

[77] M. D. Mukrasch, S. Bibow, J. Korukottu, S. Jeganathan, J. Biernat, C. Griesinger, E. Mandelkow, M. Zweckstetter, "Structural polymorphism of 441-residue Tau at single residue resolution", *PLoS Biol.* **2009**, *7*, 399–414, DOI 10.1371/journal.pbio.1000034.

[78] G. Pintacuda, G. Otting, "Identification of protein surfaces by NMR measurements with a pramagnetic Gd(III) chelate.", *J. Am. Chem. Soc.* **2002**, *124*, 372–373, DOI 10.1021/ja016985h.

[79] X. C. Su, G. Otting, "Paramagnetic labelling of proteins and oligonucleotides for NMR", *J. Biomol. NMR* **2010**, *46*, 101–112, DOI 10.1007/s10858-009-9331-1.

[80] M. R. Jensen, P. R. L. Markwick, S. Meier, C. Griesinger, M. Zweckstetter, S. Grzesiek, P. Bernado, M. Blackledge, "Quantitative Determination of the Conformational Properties of Partially Folded and Intrinsically Disordered Proteins Using NMR Dipolar Couplings", *Structure* **2009**, *17*, 1169–1185, DOI 10.1016/j.str.2009.08.001.

[81] N. Tjandra, A. Bax, "Direct measurement of distances and angles in biomolecules by NMR in a dilute liquid crystalline medium", *Science* **1997**, *278*, 1111–1114, DOI 10.1126/science.278.5340.1111.

[82] H.-J. Sass, G. Musco, S. J. Stahl, P. T. Wingfield, S. Grzesiek, "Solution NMR of proteins within polyacrylamide gels: Diffusional properties and residual alignment by mechanical stress or embedding of oriented purple membranes", *J. Biomol. NMR* **2000**, *18*, 303–309, DOI 10.1023/A:1026703605147.

[83] J. Torbet, G. Maret, "Fibres of highly oriented Pf1 bacteriophage produced in a strong magnetic field", *J. Mol. Biol.* **1979**, *134*, 843–845, DOI 10.1016/0022-2836(79)90489-3.

[84] R. Mohana-Borges, N. K. Goto, G. J. Kroon, H. J. Dyson, P. E. Wright, "Structural characterization of unfolded states of apomyoglobin using residual dipolar couplings", *J. Mol. Biol.* **2004**, *340*, 1131–1142, DOI 10.1016/j.jmb.2004.05.022.

[85] M. Krzeminski, J. A. Marsh, C. Neale, W.-Y. Choy, J. D. Forman-Kay, "Characterization of disordered proteins with ENSEMBLE", *Bioinformatics* **2013**, *29*, 398–399, DOI 10.1093/bioinformatics/bts701.

[86] D. Sahu, M. Bastidas, S. A. Showalter, "Generating NMR Chemical Shift Assignment of Intrinsically Disordered Proteins Using Carbon-Detected NMR Methods", *Anal. Biochem.* **2014**, *449*, 17–25, DOI 10.1016/j.ab.2013.12.005.

[87] W. Bermel, I. C. Felli, R. Kümmerle, R. Pierattelli, "^{13}C direct-detection biomolecular NMR", *Concepts Magn. Reson.* **2008**, *32A*, 183–200, DOI 10.1002/cmr.a.20109.

[88] W. Bermel, I. Bertini, I. C. Felli, R. Pierattelli, P. R. Vasos, "A selective experiment for the sequential protein backbone assignment from 3D heteronuclear spectra", *J. Magn. Reson.* **2005**, *172*, 324–328, DOI 10.1016/J.JMR.2004.11.005.

[89] W. Bermel, I. Bertini, L. Duma, I. C. Felli, L. Emsley, R. Pierattelli, P. R. Vasos, "Complete assignment of heteronuclear protein resonances by protonless NMR spectroscopy", *Angew. Chemie - Int. Ed.* **2005**, *44*, 3089–3092, DOI 10.1002/anie.200461794.

[90] W. Bermel, I. Bertini, I. C. Felli, M. Matzapetakis, R. Pierattelli, E. C. Theil, P. Turano, "A method for C^{α} direct-detection in protonless NMR", *J. Magn. Reson.* **2007**, *188*, 301–310, DOI 10.1016/j.jmr.2007.07.004.

[91] I. C. Felli, R. Pierattelli, "Novel methods based on13C detection to study intrinsically disordered proteins", *J. Magn. Reson.* **2014**, *241*, 115–125, DOI 10.1016/j.jmr.2013.10.020.

[92] K. Pervushin, R. Riek, G. Wider, K. Wuthrich, "Attenuated T_2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution", *Proc. Natl. Acad. Sci.* **1997**, *94*, 12366–12371, DOI 10.1073/pnas.94.23.12366.

[93] S. Mäntylahti, O. Aitio, M. Hellman, P. Permi, "HA-detected experiments for the backbone assignment of intrinsically disordered proteins", *J. Biomol. NMR* **2010**, *47*, 171–181, DOI 10.1007/s10858-010-9421-0.

[94] S. Mäntylahti, M. Hellman, P. Permi, "Extension of the HA-detection based approach: (HCA)CON(CA)H and (HCA)NCO(CA)H experiments for the main-chain assignment of intrinsically disordered proteins", *J. Biomol. NMR* **2011**, *49*, 99–109, DOI 10.1007/s10858-011-9470-z.

[95] E. W. Sayers, D. A. Torchia, "Use of the Carbonyl Chemical Shift to Relieve Degeneracies in Triple-Resonance Assignment Experiments", *J. Mang. Reson.* **2001**, *153*, 246–253, DOI 10.1006/jmre.2001.2440.

[96] J. Yao, H. J. Dyson, P. E. Wright, "Chemical shift dispersion and secondary structure prediction in unfolded and partly folded proteins", *FEBS Lett.* **1997**, *419*, 285–289, DOI 10.1016/S0014-5793(97)01474-9.

[97] H. J. Dyson, P. E. Wright, "Nuclear magnetic resonance methods for elucidation of structure and dynamics in disordered states", *Methods Enzymol.* **2001**, *339*, 258–270, DOI 10.1016/S0076-6879(01)39317-5.

[98] P. Permi, "Intraresidual HNCA : An experiment for correlating only intraresidual", *J. Biomol. NMR* **2002**, *23*, 201–209.

[99] W. Bermel, I. Bertini, V. Csizmok, I. C. Felli, R. Pierattelli, P. Tompa, "H-start for exclusively heteronuclear NMR spectroscopy: The case of intrinsically disordered proteins", *J. Magn. Reson.* **2009**, *198*, 275–281, DOI 10.1016/j.jmr.2009.02.012.

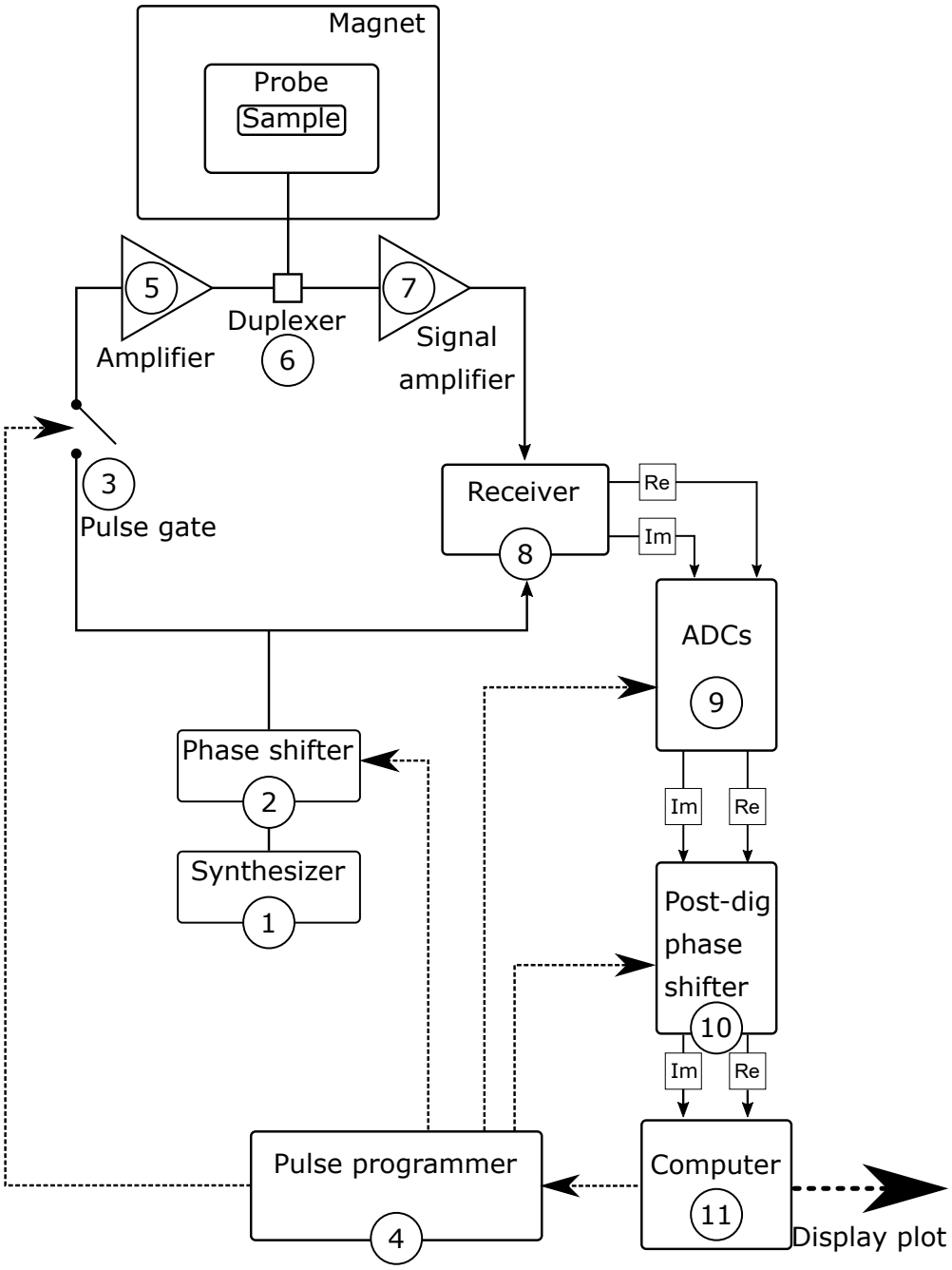
[100] W. Lee, M. Tonelli, J. K. Markley, "NMRFAM-SPARKY: enchanced software for biomolecular NMR spectroscopy", *Bioinformatics* **2015**, *31*.

[101] K. Tamiola, F. A. A. Mulder, "Using NMR chemical shifts to calculate the propensity for structural order and disorder in proteins", *Biochem. Soc. Trans.* **2012**, *40*, 1014–1020, DOI 10.1042/BST20120171.

[102] D. O. Cicero, G. M. Contessa, M. Paci, R. Bazzo, "HACACO revisited: Residual dipolar coupling measurements and resonance assignments in proteins", *J. Magn. Reson.* **2006**, *180*, 222–228, DOI https://doi.org/10.1016/j.jmr.2006.02.016.

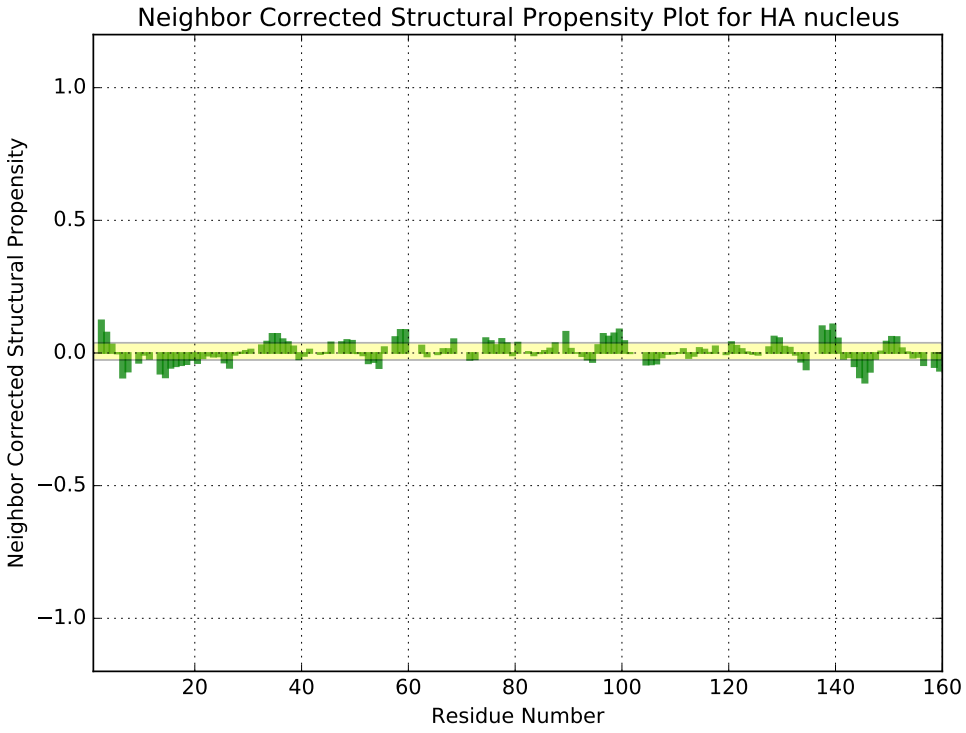
Appendix A

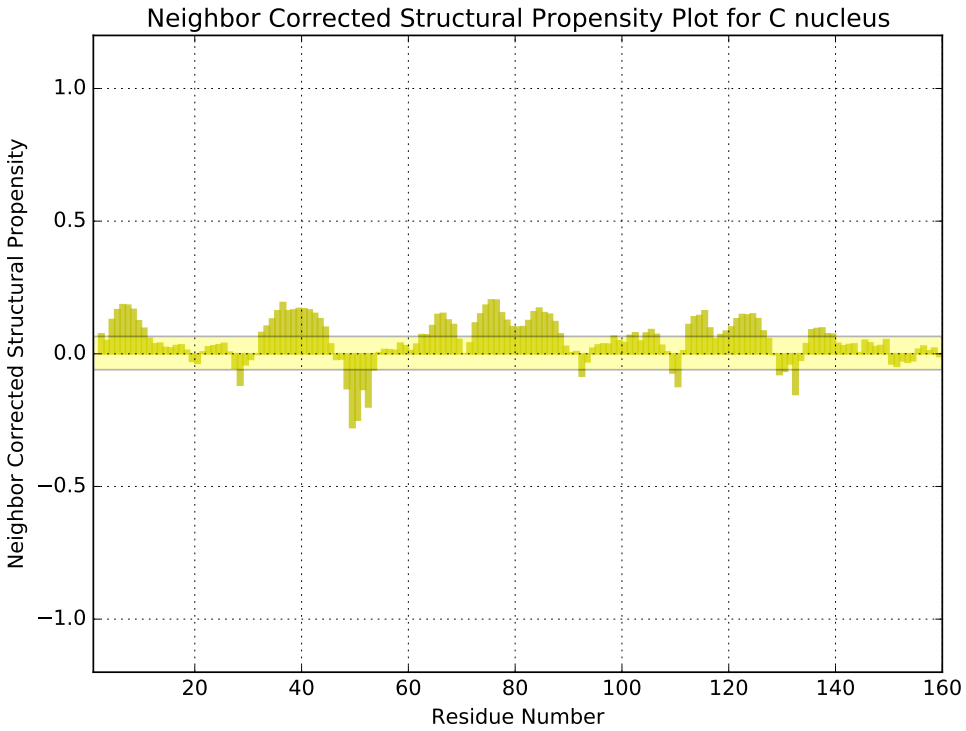
Schematic of the magnet

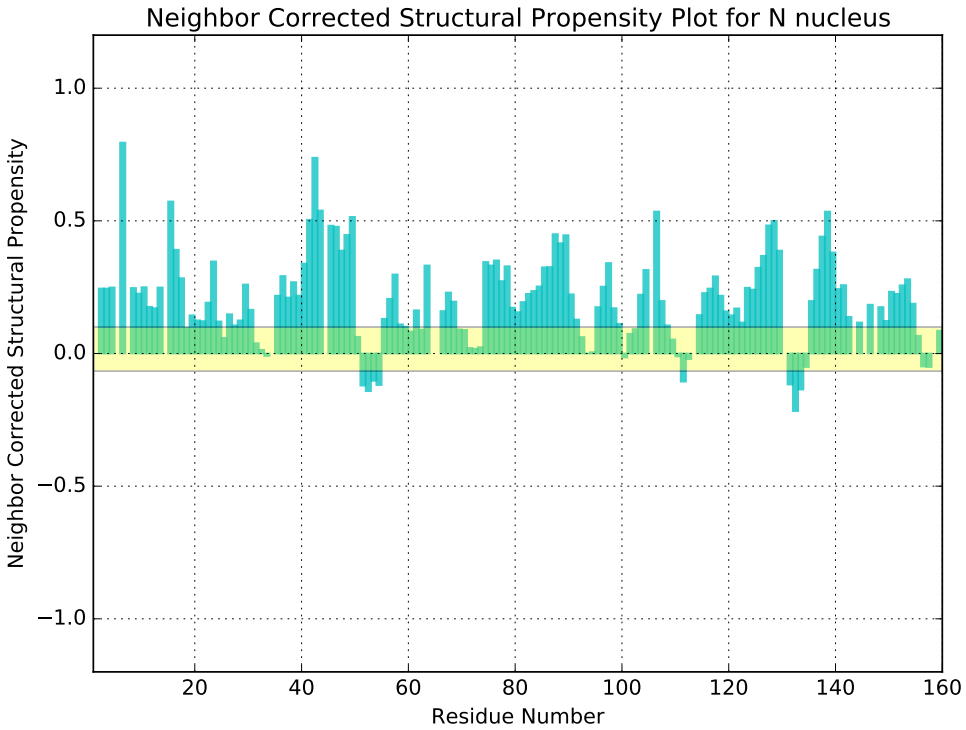


Appendix B

Structural propensity plots

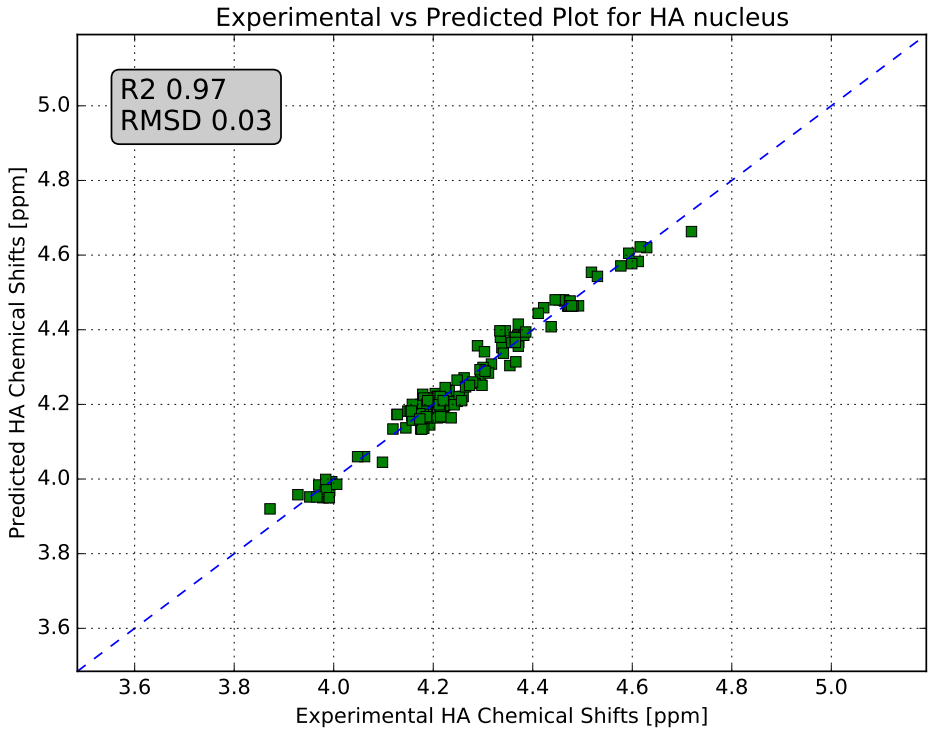


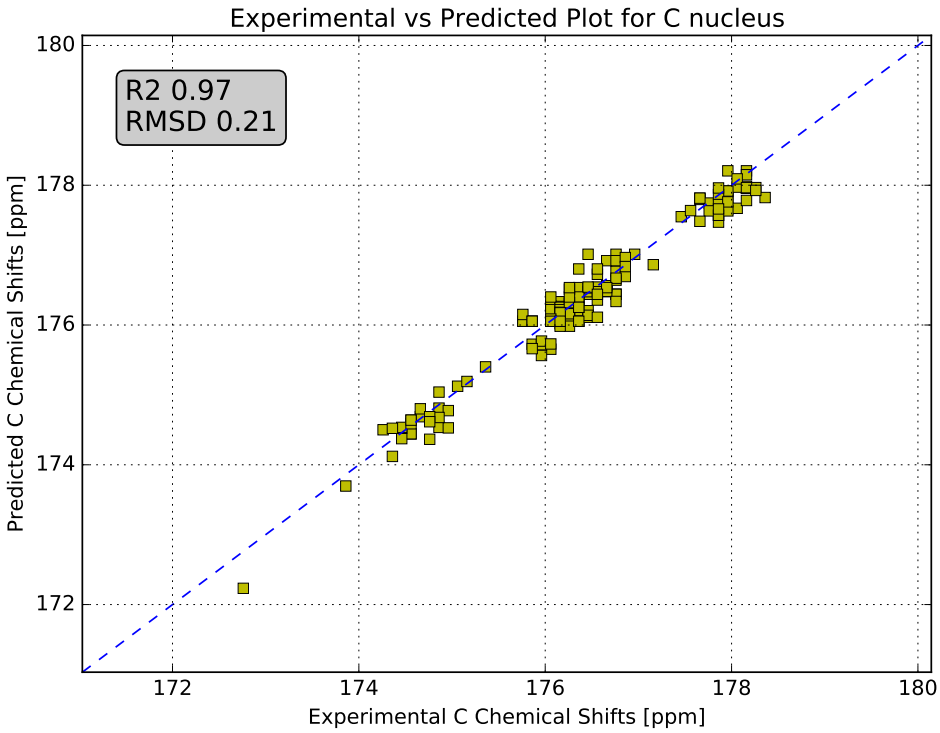

Neighbor Corrected Structural Propensity Plot for C nucleus
Neighbor Corrected Structural Propensity
1.0
0.5
0.0
−0.5
−1.0
20
40
60
80
100
120
140
160
Residue Number

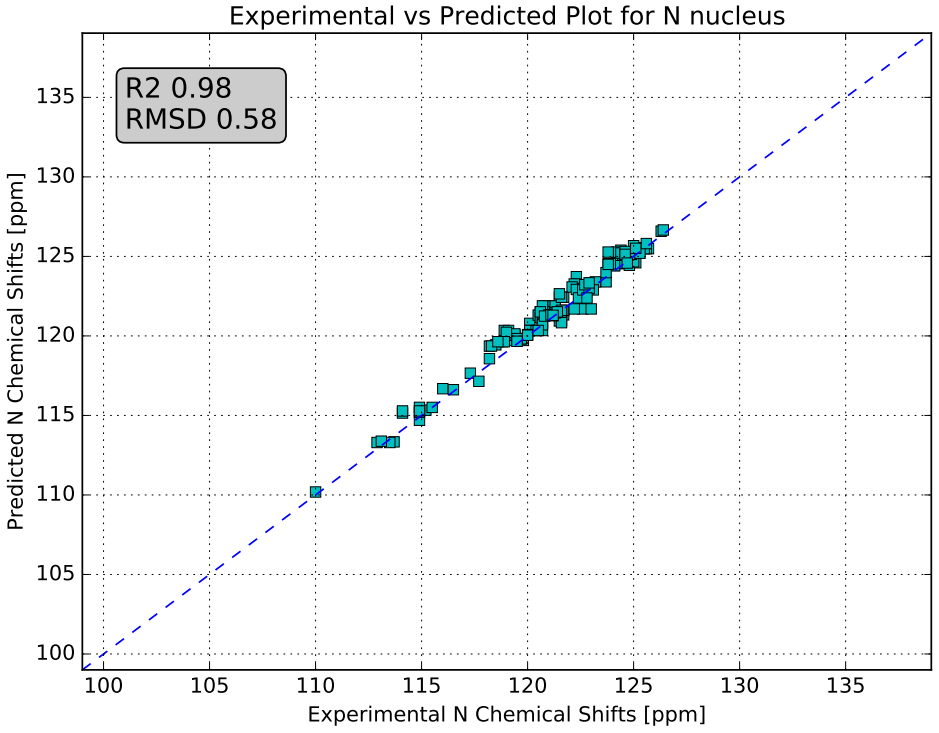

Neighbor Corrected Structural Propensity Plot for N nucleus
Neighbor Corrected Structural Propensity
1.0
0.5
0.0
−0.5
−1.0
20
40
60
80
100
120
140
160
Residue Number

Appendix C

Experimental vs predicted chemical shift plots




Experimental vs Predicted Plot for C nucleus
R2 0.97
RMSD 0.21
Predicted C Chemical Shifts [ppm]
Experimental C Chemical Shifts [ppm]
180
178
176
174
172
172
174
176
178
180


Experimental vs Predicted Plot for N nucleus
R2 0.98
RMSD 0.58
Predicted N Chemical Shifts [ppm]
Experimental N Chemical Shifts [ppm]
135
130
125
120
115
110
105
100
100
105
110
115
120
125
130
135