Revealing the secrets of proteins

A triple-nuclear and multi-dimensional high-resolution NMR spectroscopy study

Annette Katharina Brenner

Dissertation for the degree philosophiae doctor (PhD)

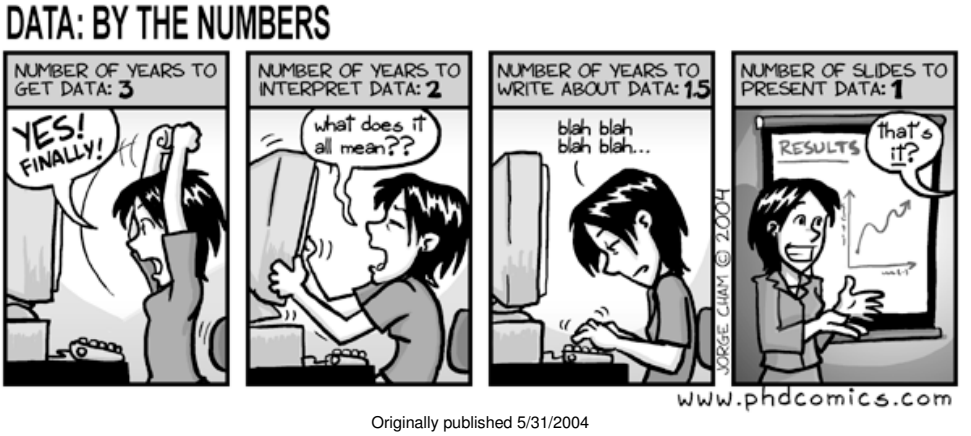
at the University of Bergen

2012

Hvis et lys lokker deg,
så følg det.
Fører det deg ut i sumpen,
så kommer du vel opp av den igjen;
men hvis du ikke følger det,
vil hele livet igjennom den tanken plage deg
at det kanskje var din stjerne.

Christian Friedrich Hebbel (1813-1863)

Piled Higher and Deeper ***by Jorge Cham*** **www.phdcomics.com**



Originally published 5/31/2004

ACKNOWLEDGEMENTS

First of all, I would like to thank my main supervisor Assoc. Prof. Nils Åge Frøystein. After eight years of knowing you, I am still fascinated by the way you intuitively "get" even the most complicated pulse sequences. Thanks to you I have learnt more about heteronuclear, multi-dimensional NMR spectroscopy than I ever hoped for. Your sense of humour, your anecdotes and your understanding for my yearning for the mountains created a relaxed atmosphere for our collaboration.

I also want to thank my co-supervisor, Prof. Johan Lillehaug, especially for never giving up on publishing our paper in JBC and for always believing in me. It was inspiring to work inter-disciplinary, where neither of us really understood what the other was talking about. Thank you also for the trip to Myrkdalen in winter 2011.

Prof. Arnt Raae, my unofficial co-supervisor, I want to thank for inviting me into the fascinating world of spectrins and for important support on the paper.

I wish to express my gratitude towards all the people who have contributed to this work: my co-workers in France, Gilles Travé and Bruno Kieffer, for your help with the interpretation of the protein dynamic results and for experimental inspiration; Rune, for the hours I spent with you in the lab and for your good sense of humour; Thomas, for your insight into acetyltransferases; Jarl and Øyvind, for your help with experiments and data processing. And finally thanks to everyone who contributed to various protein purifications.

I want to thank the people at the Department of Chemistry, especially my fellow PhD-students, for putting up with me and keeping my spirits up/dragging me down. Sara, I appreciated to have you as a room-mate for four years, who had to live with my daily 2 o`clock tea-breaks and other annoying habits I might have.

I am also very grateful to the people at my former institute, the Department of Molecular Biology, for always letting me feel welcome there and inviting me to their Christmas parties.

A special thanks goes out to my family and friends for bearing with four years of mood swings and for nodding along when I talked about my project. Last, but not least, thank you, Thorleif, for your tremendous patience. You made me appreciate my time as a PhD student even more.

Annette Brenner, May 2012

TABLE OF CONTENTS

ABSTRACT

Nuclear magnetic resonance (NMR) spectroscopy is one of the major tools for studying the structure and dynamics of proteins in solution. The still ongoing development of new and improved pulse-sequences for multi-dimensional, hetero-nuclear NMR experiments has made the field applicable for continuously larger biomolecules.

In this study, a variety of NMR experiments was conducted on two rather different proteins: chicken brain α-spectrin repeat 17 (R17) and human N^{α}-acetyltransferase 50 protein (hNaa50p). The aims were to investigate the thermodynamical and conformational stabilities of the former and the enzyme mechanism of the latter protein, respectively.

R17 is a 13 kDa domain of the ubiquitous structural protein spectrin, which is a key component in the cytoskeleton. Its structure consists of a triple-helix bundle, where two loops connect the long helices. Spectrin repeats have shown large diversities in their thermodynamical and conformational stabilities, which probably is associated with function. R17 and its neighbouring repeat, R16, are two thoroughly studied domains. R16 was one of only two spectrin repeats that had been investigated by NMR spectroscopy prior to this study. Because R17 is four times less stable than R16 in terms of ΔG, the latter domain was used for stability comparison.

NMR studies, including dynamic measurements and hydrogen-deuterium exchange, in combination with NMR, circular dichroism and fluorescence measurements at stepwise increasing temperatures revealed that the repeat is rigid at room temperature but that both the thermodynamical and conformational stabilities are gradually reduced when the temperature exceeds 40 °C. The destabilization of the domain seems to initiate in the centre of helix C and the amino acids that are close in space to one particular residue, V99, in the triple-helix bundle. A multiple sequence alignment of 35 chicken brain spectrin repeats revealed that this valine is a rare substituent of a moderately conserved tryptophan at that position. Previous studies had shown that the

small valine side chain introduces a cavity in the centre of the triple-helix bundle and diminishes the amount of hydrophobic interactions in the core of the repeat. Thus, the substitution of this tryptophan appears to be the most probable reason for the reduced stability of R17 compared to R16.

hNaa50p is an almost 20 kDa enzyme that exhibits both N^{α}- and N^{ε}-acetyltransferase activity. In these reactions, the acetyl group of acetyl-Coenzyme A (acetyl-CoA) is transferred to the backbone of the N-terminal amino acid and the side chain of specific lysine residues, respectively. The substrate specificity of hNaa50p's N^{α}-acetyl-transferase function is preferably directed towards peptides starting with the amino acid sequence MLGP, whereas K34, K37 and K140 are N^{ε}-auto acetylated. hNaa50p appears both freely in the cell and associated to two other proteins, hNaa10p and hNaa15p, which together make up the NatE complex. hNaa50p is essential for proper sister chromatid cohesion and chromosome resolution, and has increasingly been linked to cancer development. An inhibitor of its enzyme activity could destabilize chromosome formation and prevent nuclear division, and thus function as anti-cancer drug.

In order to get insight into the N^{α}-acetyltransferase reaction mechanism, the affinities of hNaa50p towards its substrates and products, and the enzyme mechanism itself were studied by enzymatic assays and NMR spectroscopy. According to our results, acetyl-CoA is the first substrate to enter and bind the active site, whereas CoA functions as product inhibitor. The peptide is not able to bind to hNaa50p in the absence of acetyl-CoA and its product, acetylated peptide, showed no affinity for the enzyme. Altogether, this indicates that hNaa50p follows an ordered sequential mechanism, possibly of the Theorell-Chance type. Thus, an inhibitor whose binding leads to the formation of a ternary complex might be a promising anti-cancer drug.

LIST OF PUBLICATIONS

Paper I:

Brenner A.K., Kieffer B., Travé G., Frøystein N.Å., Raae A.J. "Thermal stability of chicken brain α-spectrin repeat 17: a spectroscopic study." *Journal of Biomolecular NMR* **53 (2)**, 2012, 71-83.

Paper II:

Evjenth R.H., Brenner A.K., Thompson P.R., Arnesen T., Frøystein N.Å., Lillehaug J.R. "Human protein N-terminal acetyltransferase hNaa50p (hNat5/hSan) follows ordered sequential catalytic mechanism. Combined kinetic and NMR study." *Journal of Biological Chemistry* **287 (13)**, 2012, 10081-10088.

Paper III:

Brenner A.K., Frøystein N.Å. "Extending the range of backbone assignment of medium-sized proteins using MUSIC and CC(CO)NH." *Manuscript submitted to Journal of Magnetic Resonance*.

ABBREVIATIONS

acetyl-CoA	acetyl-Coenzyme A
CD	Circular Dichroism
CoA	Coenzyme A
CSI	Chemical Shift Index
HMBC	Heteronuclear Multiple Bond Correlation
HMQC	Heteronuclear Multiple Quantum Coherence
hNaa10p	human N^{α}-acetyltransferase 10 protein
hNaa15p	human N^{α}-acetyltransferase 15 protein
hNaa50p	human N^{α}-acetyltransferase 50 protein
HSQC	Heteronuclear Single Quantum Coherence
MUSIC	MUltiplicity Selective In-phase Coherence transfer
NAT	N-terminal AcetylTransferase
NMR	Nuclear Magnetic Resonance
NOE	Nuclear Overhauser Effect
NOESY	Nuclear Overhauser Enhancement SpectroscopY
PDB	Protein Data Bank
R16	spectrin Repeat 16
R17	spectrin Repeat 17
TOCSY	TOtal Correlation SpectroscopY

1. INTRODUCTION

1.1 Spectrins

1.1.1 Origin and function

Spectrins are a family of ubiquitous structural proteins that occur in most vertebrates, some metazoans and, possibly, in higher plants [1]. The spectrin superfamily consists of spectrin, dystrophin, α-actinin, utrophin [2] and plakin [3]. Spectrin-like domains have also been found in a variety of other proteins [4]. Throughout the thesis, the term spectrin refers to the spectrin family rather than the superfamily.

Spectrins, with few exceptions [5], are situated in the cytoskeleton and function as a key component in the regulation of cell shape, membrane flexibility and membrane elasticity [2]. The flexibility of the protein depends on the tissue that it occupies. For instance, erythroid spectrins that have to adjust to the elasticity of red blood cells are more flexible than brain spectrins [6]. Many spectrins are exposed to mechanical stress [7], and mutations in highly conserved residues that adapt the structure to sheer forces are pathologic [8]. For example, mutations in erythroid spectrins are associated with haemolytic anaemia due to abnormally shaped erythrocytes, so-called elliptocytosis [9, 10], whereas mutations in muscular spectrins may lead to muscular dystrophies, i.e. progressive skeletal muscle weakness [11].

1.1.2 Structure

Spectrins in vertebrates are built up of two subunits, the α (approximate mass of 280 kDa [12]) and β (approximately 246 kDa [13]) homologous chains, which form antiparallel heterodimers by side-to-side association [14] and tetramers by head-to-head association of two heterodimers [15]. The main building blocks of the chains are the spectrin repeats which comprise about 90 % of the protein sequence [16]. Usually, the α-chain is built up of 20 and the β-chain of 17 repeating domains [1].

In addition, spectrins typically contain several other domains that are involved in different cellular processes. Two consecutive calponin homology domains on the β-chain at both ends of spectrin cross-link actin, thus forming an actin-spectrin network

which is central in the regulation of smooth muscle contraction and cell shape [17] (figure 1.1).

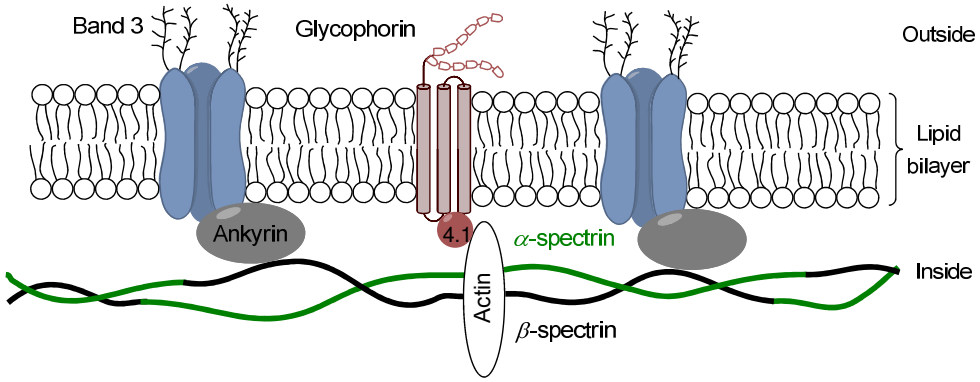


Figure 1.1: Structural model of the erythroid cell membrane. Spectrin is coupled to the lipid bilayer through the association to actin and ankyrin, which in turn are bound to protein 4.1 and the anion transporter band 3, respectively [18].

Two calmodulin-like domains [19], consisting of four EF-hands (helix-loop-helix motifs), reside at both ends of the α-chains and are thought to regulate the affinity of spectrin for actin [20], whereas the src-homology 3 [21] and the pleckstrin homology domains [22] are involved in protein-protein interaction and cell signalling.

1.1.3 Spectrin repeats

The structure of the spectrin repeated units provides an explanation for the rod-like shape of the spectrin tetramer. The tandem repeat of chicken brain α-spectrin repeats 15, 16 and 17 is shown in figure 1.2.

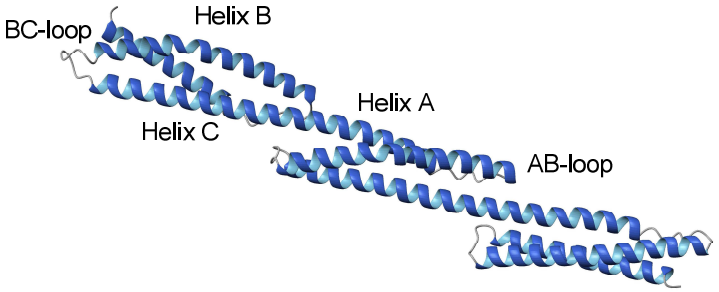


Figure 1.2: The tertiary structure of chicken brain α-spectrin repeats 15, 16 and 17, PDB-entry 1U4Q [23]. The characteristic secondary structural elements of the spectrin repeats are indicated. The long helices, especially the linking of helix C in one repeat with helix A of the next repeat, can explain the elongated structure of the spectrin molecule. The structure was visualized in the program MOLMOL [24].

Each repeat consists of approximately 106 amino acids [2] and adopts a left-handed coiled coil of three right-handed antiparallel α-helices [10, 25]. The helices, denoted A, B and C, are connected with two short loops, the AB- and the BC-loop. Adjacent spectrin repeats are connected with each other through short linkers [2] which according to results from X-ray crystallography are helical [26, 27]. The amino acid sequences of spectrins are made up of a repeating heptad pattern, a structural motif typically of coiled coils. The positions in the heptad are commonly denoted by the letters *a* through *g*. Residues at the *a* and *d* positions are typically non-polar and located in the hydrophobic core between the three helices, where they involve in interhelical, hydrophobic interactions. The partially solvent-exposed *e* and *g* positions are often occupied by charged residues which form interhelical salt-bridges with oppositely charged residues [28].

Even though the sequence identity between different spectrin repeats is quite low (20-30 %), their structures are very similar when they are superimposed [2]. The hydrophobic amino acids in the helices are the most conserved residues.

1.1.4 Aims of the study

The thermal and conformational stabilities of spectrins do not only depend on the environment of the cell, and thus various functions [6], but there are also major differences between the thermal stabilities of the spectrin repeats within mature spectrin. For instance, the melting points of mammalian brain α- and β-spectrin repeats vary between 25 °C and 79 °C [29].

It was observed that the ΔG of thermal unfolding of single repeats of the chicken brain α-spectrin repeat 16 (R16) is about four times larger than that of the neighbouring repeat 17 (R17) [23]. Furthermore, double repeats are in general more thermodynamically stable than single repeats [16], and the R16 domain in the tandem repeat R1617 is more stable than the single repeat, even after the unfolding of R17 [30].

The incentive of the study was to find out, why the single repeat 17 of chicken brain α-spectrin is four times less thermodynamically stable than the neighbouring R16. The bioinformatical approach was to look for differences in the amino acid sequences of the domains that could have effect on the tertiary structures, and thus the thermal and

conformational stabilities of the repeats. For example, it has recently been shown that two point mutations surrounding a *d* position in human dystrophin repeat 23 are the main cause of muscular dystrophy as they lead to destabilization of helix C [31]. Thus, even small differences in the primary structure can have effect on the stability of a repeat. The main approach, however, was to perform a variety of biomolecular nuclear magnetic resonance (NMR) techniques under different conditions in order to identify labile parts of R17.

1.2 Acetyltransferases

1.2.1 Protein acetylation

Post-translational protein modifications are important for the regulation of activity, function, stability and localization of proteins [32]. The most common modifications include phosphorylation, glycosylation, methylation, ubiquitilation and acetylation [33]. The latter is carried out by acetyltransferases which donate the acetyl group of acetyl-coenzyme A (acetyl-CoA) to the substrate [34]. Protein acetylation is important for the regulation of the cell cycle, but is also associated with apoptosis (programmed cell death) and cancer development [35]. On the other hand, inhibitors of protein deacetylases have been discovered as potential cancer drugs [36]. Acetylation can target both amino acid side chains, so-called N^{ε}-acetylation, and the N-termini of proteins, so-called N^{α}-acetylation.

1.2.2 N^{ε}-acetylation

In N^{ε}-acetylation, lysine acetyltransferases transfer the acetyl group of acetyl-CoA to the side chain-$N^{\zeta}H_3^+$ of specific lysine residues (figure 1.3). Histones are the most thoroughly studied proteins that are modified by N^{ε}-acetylation [37]. Since the positive charge of the lysine side chain is removed by acetylation, the affinity of the histones for the negatively charged DNA is reduced [37], and thus gene expression is up-regulated [38]. N^{ε}-acetylation is reversible [38] and the reverse mechanism is catalyzed by histone deacetylases [37].

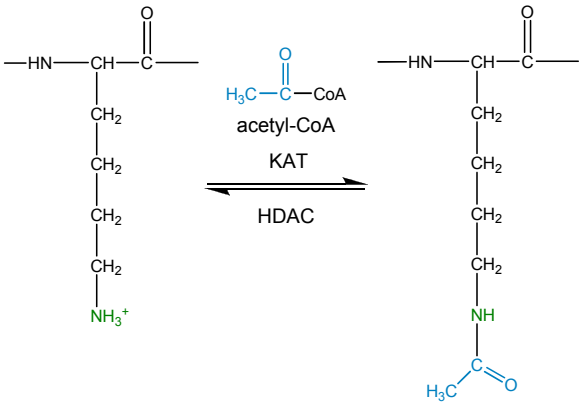


Figure 1.3: N^{ε}-acetylation. Lysine acetyltransferases (KATs) transfer the acetyl group of acetyl-CoA to the side chain of lysine, and thus remove the positive charge of the amino acid. Deacetylases (HDACs) catalyze the reverse reaction.

1.2.3 N^{α}-acetylation

In N^{α}-acetylation, the acetyl group of acetyl-CoA is transferred to the backbone of the N-terminal amino acid of the target protein (figure 1.4).

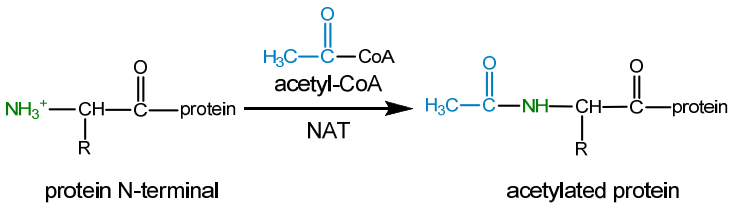


Figure 1.4: N^{α}-acetylation. The acetyl group of acetyl-CoA is irreversibly transferred to the N-terminal amino acid of the target protein by an N-terminal acetyltransferase (NAT).

The reaction takes place post-initial rather than post-translational [34] as the nascent protein is acetylated when 20-50 amino acids protrude from the ribosome [39]. N^{α}-acetylation neutralizes the positive charge of the protein N-terminal, which may affect the protein's function, its stability, its interaction with other molecules and its susceptibility to further modifications [37]. Unlike N^{ε}-acetylation, N-terminal acetylation appears to be irreversible [37]. N^{α}-acetylation occurs in all kingdoms of life, but the fraction of acetylated proteins varies widely. In mammals, supposedly 80-90 % of all proteins are acetylated; more simple eukaryotes like yeast only show about 50 % of acetylated proteins, while proteins in prokaryotes and archaea are rarely acetylated [40]. The methionine that resides at the N-terminal of most proteins is

usually cleaved prior to N^{α}-acetylation [39]. The second residue (or first, if the methionine is not cleaved) of the protein sequence is of special importance for the specificity of the N-terminal acetyltransferase (NAT) [39]: 95 % of all proteins that are acetylated have serine, alanine, glycine, threonine or methionine at the N-terminal position [41]. But the occurrence of these amino acids does not guarantee the acetylation of a protein. So far, there are no methods to deduce from the primary structure whether or not a protein will be modified by acetylation, but it is known that for the specificity of the NATs also the amino acid at the third, and in some cases even fourth and fifth, position is of importance [37].

1.2.4 The protein human Naa50p

In eukaryotes, a total of six N^{α}-acetyltransferase complexes have been identified so far: NatA-NatF [42, 43]. At first Naa50p, formerly called Nat5p in yeast, San in fruit fly and Nat5 in human [42], was shown to be a part of the NatA complex together with Naa10p (previously Ard1p, Ard1 and hArd1) and Naa15p (previously Nat1p, Nat1 and NATH) [44, 45]. Interestingly, it was discovered that Naa50p in baking yeast is not associated to the ribosome in the absence of Naa10p and Naa15p [44], and that 80 % of Naa50p in *Drosophila melanogaster* occurs freely in the cell [45]. Furthermore, Naa50p is neither required for the activity of the NatA complex in yeast [41] nor in human cells [46]. Therefore, human Naa50p in association with NatA was defined as a putative, novel complex annotated NatE [46].

hNaa50p is located in the cytoplasm, consists of 169 amino acids and has a molecular mass of 19.4 kDa. Its sequence identity to Naa50p in yeast and fruit fly is approximately 25 % and 70 %, respectively [35]. It showed to be important for sister chromatid adhesion during mitosis in fruit fly and human [45, 46]. hNaa50p is the first characterized acetyltransferase which exhibits both N^{α}- and N^{ε}-acetylation activity. The latter was detectable in both *in vivo* [45] and *in vitro* experiments [47], and auto acetylation of the lysine residues K34, K37 and K140 in hNaa50p showed to increase the N^{α}-substrate specificity [47]. hNaa50p has a substrate preference towards peptides starting with the sequence ML followed by an amino acid with a small side chain at the third position. The optimal substrate identified so far starts with the sequence MLGP, and thus varies widely from the substrate specificity of the NatA complex

towards peptides starting with the amino acids SESS [47]. The structure of hNaa50p is shown in figure 1.5.

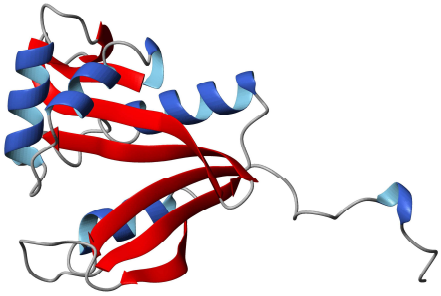


Figure 1.5: The crystal structure of hNaa50p obtained in complex with acetyl-CoA (PDB-entry 2OB0). The substrate is not included in the figure. The active site of the protein is situated in the cleft between the two β-sheets. The structure was visualized in MOLMOL [24].

1.2.5 Enzyme mechanism

So far, four different types of catalytic mechanism have been proposed for N^{α}- and N^{ε}-acetyltransferases: ordered sequential mechanism [48], random sequential mechanism [49], Theorell-Chance mechanism [50] and ping-pong mechanism [51, 52]. However, since it was recently revealed that both Esa1 and p300 follow an enzyme mechanism which includes the formation of a ternary complex [48, 50], it could not be confirmed yet that any acetyltransferase actually exhibits the ping-pong mechanism. All four mechanisms are of the Bi-Bi type, i.e. they involve the sequential binding of two substrates and the release of two products [53].

In the ordered sequential mechanism (figure 1.6A), the binding order of the two substrates is obligatory and the second substrate will usually not bind until the enzyme undergoes conformational changes upon binding of substrate one [53]. Thus, the nascent protein and acetyl-CoA are simultaneously bound to the acetyltransferase and form a ternary complex. After transfer of the acetyl group, the two products, the acetylated protein and CoA, are released.

In contrast, the binding and disassociation order of substrates and products is arbitrary in the random sequential mechanism (figure 1.6B) [53]. Like in the ordered sequential mechanism, a ternary complex involving enzyme and substrates is formed.

The Theorell-Chance mechanism (figure 1.6C) [54], also called hit-and-run, is a special case of the ordered sequential mechanism as the ternary complex is unstable

and only exists for a short period of time. Again, acetyl-CoA has to bind first and induce changes on the enzyme structure in order to promote binding of the substrate protein. The protein is then immediately acetylated and released from the acetyltransferase.

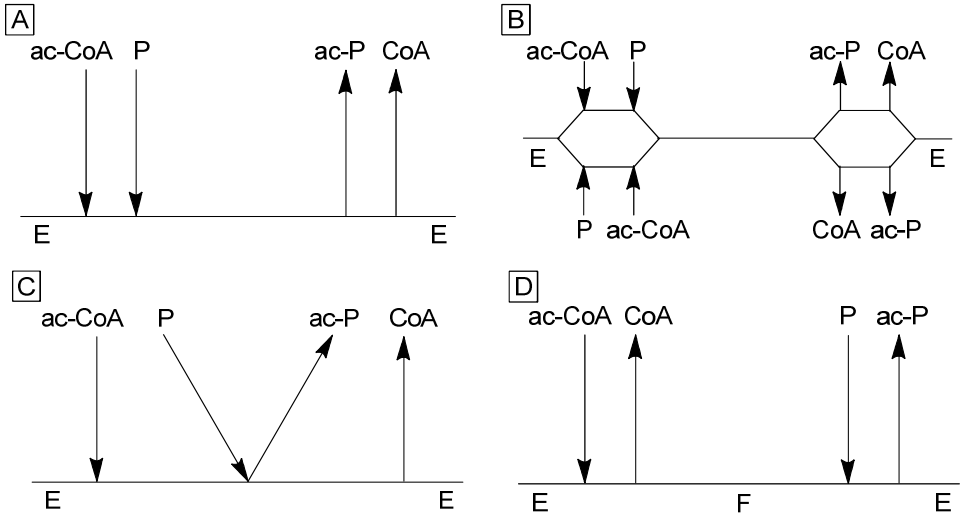


Figure 1.6: Cleland representations of the acetyltransferase reaction mechanisms. **A)** Ordered sequential mechanism: acetyl-CoA (ac-CoA) has to bind to the enzyme (E) prior to the binding of the nascent protein (P). The acetyl group is transferred to the protein and the products, acetylated protein (ac-P) and CoA, are released. **B)** Random sequential mechanism: the binding order of nascent protein and acetyl-CoA and the disassociation order of the products are random. **C)** Theorell-Chance mechanism: the nascent protein is immediately acetylated upon binding to the enzyme. **D)** Ping-pong mechanism: the acetyl group stays bound after the release of CoA. This induces a structural change on the enzyme (F) required for the binding and acetylation of the nascent protein.

Finally, the enzyme reaction in the ping-pong mechanism (figure 1.6D), also called double-displacement, is divided into two separate steps: the reaction of the acetyltransferase with the first substrate results in the formation of the first product and induces a new conformation of the enzyme required for binding of the second substrate [53]. Thus, the acetyl group remains bound to the enzyme after CoA is released and is then transferred to the protein substrate after binding of the latter to the acetyltransferase.

1.2.6 Aims of the study

The incentive of the study was to find out if also hNaa50p uses one of the four presented catalytic mechanisms and, in that case, which one. This was of special

interest because no eukaryotic NAT has previously been described with respect to its reaction mechanism. Furthermore, in several species Naa50p has shown to be essential for sister chromatid cohesion and chromosome resolution [45, 55, 56]. Thus, like histone deacetylase inhibitors [36], also inhibitors of NATs could be used as potential cancer drugs by destabilizing chromosome formation and nuclear division. In order to design optimal inhibitors, it is crucial to have detailed insight into the enzyme mechanism.

The approach was to investigate the reaction mechanism using enzymatic assays which could determine the kinetic constants of the substrates, acetyl-CoA and peptide (resembling the nascent protein), and the nature of inhibition exhibited by the products, CoA and acetylated peptide. In addition, the enzyme mechanism was studied by NMR spectroscopy. It was expected that binding of the first substrate would lead to a conformational change of the enzyme resulting in chemical shift changes of single peaks in the spectra. These results could be used to determine the amino acids involved in substrate binding. Furthermore, the NMR results could in combination with the enzymatic assays reveal if the substrate binding follows an ordered or a random pathway and if the mechanism involves the formation of a ternary complex (sequential mechanism or Theorell-Chance) or not (ping-pong).

2. BIOMOLECULAR NMR SPECTROSCOPY

2.1 Basic properties of NMR spectroscopy

In 1946, the research groups of Bloch [57, 58] and Purcell [59] were simultaneously the first to detect NMR signals. Since then, NMR spectroscopy has grown to be one of the most important tools for both natural and medical scientists.

All nuclei which possess a nuclear spin, I, different from zero are NMR active, but biomolecular NMR spectroscopy focuses almost exclusively on isotopes with I = ½, such as ^{1}H, ^{13}C, ^{15}N and ^{31}P. Placed into a magnetic field, the magnetic momentum of these nuclei will either align parallel (corresponding to the magnetic quantum number m_I = ½) or antiparallel to the field (m_I = -½). The spin populations of the two energy levels depend on the energy difference between the two states. This difference is based upon the magnitude of the external magnetic field, usually expressed by the ^{1}H resonance frequency of the NMR instrument, and the sensitivity of the nucleus. The latter is determined by the gyromagnetic ratio, γ, which is constant for each isotope. The ratio of the particles in the high energy state, N_β, compared to the low energy state, N_α, can be calculated from the Boltzmann distribution:

$$\frac{N_\beta}{N_\alpha} = e^{-\Delta E / kT}, \tag{2.1}$$

where k is the Boltzmann constant and T the absolute temperature. Hence, at 298 K and a resonance frequency of 600 MHz, the excess of protons in the low energy state is approximately 0.001 % of the total amount of protons. Since only this little fraction contributes to the resulting signal, it explains why NMR spectroscopy is very insensitive compared to other spectroscopic techniques.

Nevertheless, NMR spectroscopy is a powerful tool because the resonance frequency for a single nucleus is not only dependent on the isotope, but also on the chemical environment of this nucleus, i.e. atoms and electrons in the vicinity will weaken the applied magnetic field. This effect is called shielding and the resulting frequency can be expressed as follows:

$$\nu = \frac{\gamma(1-\sigma)B_0}{2\pi}, \tag{2.2}$$

where σ is the shielding constant and B_0 the magnetic field.

Instead of the absolute frequency, one refers to the chemical shift, δ, of the nucleus which is independent of the magnetic field strength of the spectrometer:

$$\delta = \frac{\nu_{\text{sample}} - \nu_{\text{ref}}}{\nu_{\text{ref}}} \times 10^6 \,\text{ppm} \tag{2.3}$$

It is also possible to transfer magnetizations between different nuclei, and in that way multi-dimensional spectra can be acquired. There are two main principles for how spins can be correlated with each other: through chemical bonds, so-called scalar or *J*-coupling, and through space, so-called dipolar coupling.

2.2 General protein NMR spectroscopy

2.2.1 Drawbacks and advantages of protein NMR spectroscopy

The main disadvantage of biomolecular NMR spectroscopy is its low inherent sensitivity. Additionally, the yield from protein purification is usually less than 5 mg and many proteins precipitate or aggregate at high concentrations. Thus, protein NMR spectroscopy requires long acquisition times; it is not unusual to spend one week on recording the spectra that are necessary for the backbone assignment.

Another drawback is that the natural abundant ^{12}C and ^{14}N isotopes are not suitable for NMR analysis. Hence, it is often necessary to label larger proteins with ^{15}N and/or ^{13}C, which makes the protein expression more time-intensive and remarkably more expensive.

Protein NMR spectra tend to be difficult to interpret due to the large amount of signals. This drawback has been diminished in the last two decades, as many new experiments especially suited for proteins have been developed. Until the early 1990s, protein NMR spectroscopy was almost entirely based on basic homonuclear 2D experiments, like TOtal Correlation SpectroscopY (TOCSY) and Nuclear Overhauser Enhancement SpectroscopY (NOESY). Due to the small chemical shift dispersion of ^{1}H (approximately 12 ppm) and shift overlap caused by increasing protein size, these techniques fail on most proteins larger than approximately 10 kDa. Since then, a cascade of heteronuclear 3D experiments has been developed. These methods have two major advantages compared to the homonuclear 2D techniques. First, as one

enters the third dimension the problem of chemical shift overlap is drastically reduced. And second, the intrinsic chemical shift dispersions of ^{15}N and especially ^{13}C are much larger than that of ^{1}H. In proteins, the ^{15}N frequency domain is restricted to the amide region, which encompasses the range of 100-135 ppm. Whereas the ^{13}C frequency domain is divided into the aliphatic region, approximately 10-70 ppm, the aromatic region, about 110-140 ppm, and the carbonyl region, approximately 165-185 ppm. These advantages outdo the disadvantages of heteronuclear 3D experiments: the costs of isotope-labelling the protein, and the increased time amount needed for the acquisition of 3D experiments.

The main advantage of studying proteins by NMR spectroscopy is its diversity of applications. The chemical shift values give insight into the secondary structural elements, whereas long-range Nuclear Overhauser Effect (NOE) interactions provide information about the tertiary structure of a protein. The assigned 2D ^{1}H-^{15}N Heteronuclear Single-Quantum Coherence (HSQC) spectrum can be used for several purposes, such as relaxation studies, denaturation studies, diffusion studies, measuring the rate of H^N-exchange and monitoring chemical shift changes of specific peaks in the spectrum upon e.g. addition of ligands. These studies may contribute to insight into the structure, the dynamics and the function of a protein at physiological conditions. Other methods might be better suited for one task, for instance X-ray crystallography for structure determination, but they cannot be used for the same variety of applications. Furthermore, since NMR spectroscopy of proteins usually is carried out in solution, preferable at conditions resembling physiology, the results obtained from NMR spectroscopy may be more biologically relevant than those from other methods.

2.2.2 Atom nomenclature

The nomenclature of the amino acids follows the numbering scheme of carboxylic acids; thus, the carbon following the carbonyl group, C´ or just C in NMR terms, is called C^{α} or CA. The numbering continues throughout the side chain and includes also heavy atoms different from carbon. In case of stereoisomery, protons are labelled with a numeral as well, e.g. $H^{\beta 2}$ and $H^{\beta 3}$. Figure 2.1 shows the atom names in the spin system of asparagine.

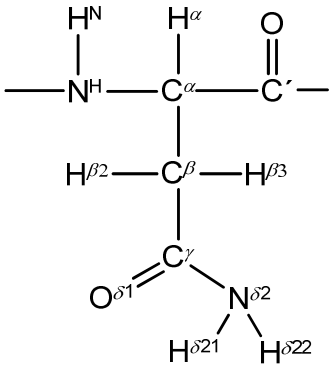


Figure 2.1: The nomenclature of the atoms in an amino acid, exemplified on asparagine.

2.2.3 Experimental terminology

Unlike classic NMR techniques which are entitled by more or less funny acronyms, e.g. TOCSY, the names of protein NMR experiments illustrate the magnetization transfer pathway in the pulse sequence. The most simple triple-resonance 3D method is the HNCO experiment. The name implies that the magnetization is initially transferred from the H^N to the N^H and further on to the C´ of the preceding residue. When the name includes brackets, like the related HN(CA)CO experiment, it means that the magnetization is passed on to the C´ via the C^α, without spin evolution of the latter. Thus, the C^α shifts are not observable in the resulting spectrum.

Pulse sequences starting with "HN" imply that they belong to the so-called "out-and-back" experiments [60], i.e. in the case of the HNCO experiment the magnetization is finally transferred back from the C´ to the H^N via the N^H for data acquisition at the most sensitive nucleus. The CBCA(CO)NH experiment, on the other hand, is an example of a transfer experiment [60]: here the magnetization transfer starts in the side chain, the H^β, and the magnetization is finally transferred to the H^N for data acquisition. Again, "(CO)" implies that the carbonyl shifts are not visible in the spectrum.

2.2.4 Sensitivity of protein NMR experiments

Two main factors influence the quality of the NMR data: the inherent sensitivity of the experiment and the resolution of the spectrum. In contrast to the former, the latter

cannot be improved by increasing the number of scans, and thus obtaining a better signal-to-noise ratio. The resolution of a spectrum depends on the field strength of the NMR spectrometer, the number of acquired data points and the line-widths of the signals. The latter is indirectly proportional to the transverse relaxation time, T_2 [61]. The T_2 values decrease as a function of increasing molecular mass, leading to signal broadening which has two negative side effects. First, as a result of increased line-widths, the peak heights decrease, i.e. for large proteins more scans have to be collected than for small ones of equal sample concentration. Second, signal broadening leads to peak overlap, which complicates the assignment work.

Short transverse relaxation times have another drawback for undeuterated proteins: they diminish the sensitivity of NMR methods that use isotropic mixing times, which is the case for all TOCSY based experiments [62]. Also, the more magnetization transfer steps that are included in a pulse sequence, the more signal strength is lost due to transverse magnetization decay in each step.

The sensitivity of an experiment is also dependent on the size of the coupling constants (figure A.1) between the involved spins in a magnetization transfer. These coupling constants are independent of protein size and practically universal. Small coupling constants reduce the sensitivity of an experiment. This is not an issue for most ^{13}C based methods as only the two-bond couplings between N^H and C^α, ^2J (N^H-C^α), and between N^H and C´, ^2J (N^H-C´), are less than 10 Hz. But it is a problem for protons, especially in α-helices, where the ^3J (N^H-H^α) typically is between 2-5 Hz [63]. Thus, ^1H-^1H TOCSY methods performed on proteins larger than 10-12 kDa suffer from both reduction in sensitivity due to short T_2 values and due to small coupling constants.

2.3 Assignment strategy

The previous two paragraphs have focused on the theory and the background for protein NMR spectroscopy in general. The remainder of this chapter will illustrate the practical aspects of protein NMR spectroscopy, exemplified on the small (approximately 8.5 kDa) human ubiquitin which was used as a case study in this work.

2.3.1 Spin system numbering

Spin systems (figure 2.1 shows an example of a spin system) in NMR experiments are given a specific numbering (figure 2.2). The system of reference is called residue I. Its predecessor (N-terminal neighbour) gets the numbering I-1, whereas the successor (C-terminal neighbour) is denoted I+1. In HNCO, for example, the C´ (I-1) directly bound to the N^H (I) is observed.

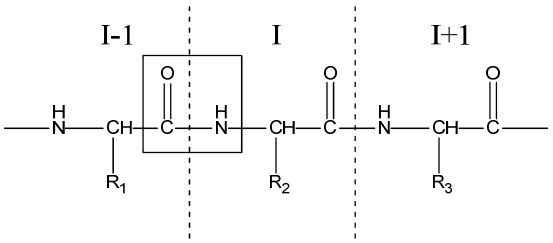


Figure 2.2: Three adjacent spin systems in the protein primary structure. The residue of observation is denoted I, the successive residue I+1 and the preceding amino acid I-1. The correlation that is observed for residue I in the 3D HNCO spectrum is highlighted.

2.3.2 2D HSQC

The starting point of the resonance assignment is the 2D ^{1}H-^{15}N HSQC [64, 65] spectrum. This experiment yields, in principle, the correlations of all backbone amides based on ^1J (N^H-H^N). In addition, the side chain amides of asparagine and glutamine, and the $N^{\varepsilon}H^{\varepsilon}$ of tryptophan and of most of the arginines appear in the spectrum. On the other hand, there are no NH-signals from proline in the HSQC spectrum as this particular amino acid lacks an H^N. Other backbone correlations might also be missing due to solvent or conformational exchange. This typically affects the N-terminal amino acid and residues in highly flexible regions.

There also exists a carbon version of this experiment, the 2D ^{1}H-^{13}C HSQC. Every proton that is directly bound to a ^{13}C nucleus will appear in the spectrum. This includes correlations in the backbone as well as in aliphatic and aromatic side chains. Thus, there will be far more peaks in the ^{1}H-^{13}C HSQC spectrum compared to the nitrogen version.

The ^{1}H-^{15}N HSQC pulse sequence, which is basically identical to that of the carbon variant, is shown in figure 2.3. The observable correlations of all discussed experiments are listed in table A.1.

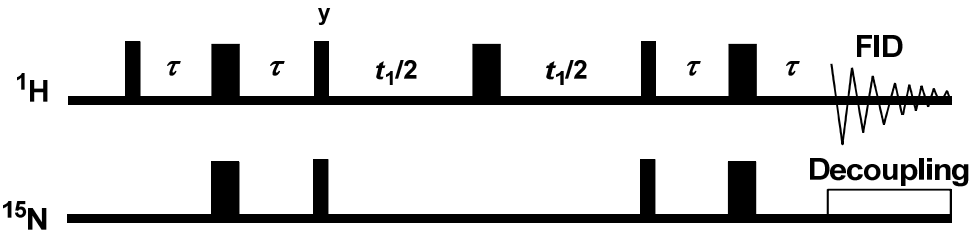


Figure 2.3: The basic ^{1}H-^{15}N HSQC pulse sequence. Thin and thick vertical bars represent 90^{o}_{x}- and 180^{o}_{x}-pulses, respectively. The 90^{o}_{y}- pulse and the delays, τ and $t_1/2$, are indicated. The τ-delays are chosen to be $^{1}/_{4}$ ^{1}J (N^{H}-H^{N}). The ^{15}N nuclei are decoupled during data acquisition (FID) in order to prevent signal splitting due to ^{1}J (N^{H}-H^{N}) in the resulting spectrum. If the protein is ^{13}C-labelled, the sequence also includes a carbon channel that is used to remove signal splitting due to ^{13}C coupling during spin evolution. In addition, gradients are usually used to suppress the dominating solvent H_2O-signal and to select for the desired magnetization transfer pathway.

2.3.3 Backbone experiments

In order to assign the peaks in the ^{1}H-^{15}N HSQC spectrum to their corresponding residues, it is necessary to obtain more information from triple-resonance 3D spectra. The transition from a 2D to a 3D spectrum is shown in figure 2.4.

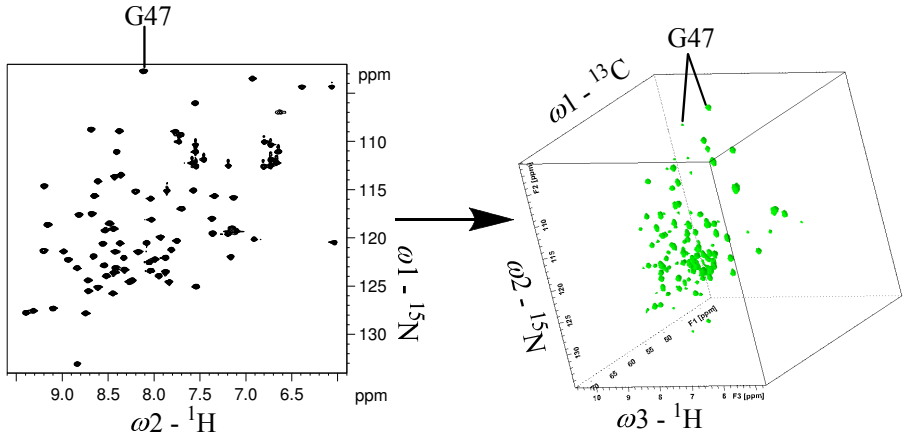


Figure 2.4: The 2D ^{1}H-^{15}N HSQC spectrum of 1 mM human ubiquitin at 600 MHz is shown on the left and the 3D HNCA spectrum on the right. The peaks in the HNCA spectrum have the same ^{1}H and ^{15}N shift values as in the HSQC spectrum, but each backbone-NH signal is now split in two, corresponding to the ^{13}C chemical shifts of the intra-residual and the preceding C^{α}. The correlations of residue G47 are highlighted as an example in both spectra.

The four basic backbone methods can be divided into two pairs: the HNCO and HN(CA)CO experiments and the HNCA and HN(CO)CA experiments.

The HNCO method [66-68] is the most sensitive of all triple-resonance 3D NMR techniques (the relative sensitivities of the six most important experiments for the

resonance assignment are shown in figure A.2). The resulting spectrum contains the peaks from the 2D ^{1}H-^{15}N HSQC spectrum separated by the C´ chemical shifts of the preceding residue in the third dimension. Thus, there is only one peak per residue in the spectrum, generating minimal signal overlap. Since HNCO is the most simple of the 3D experiments, its basic pulse sequence [69] is shown as an example of heteronuclear 3D NMR experiments in figure 2.5.

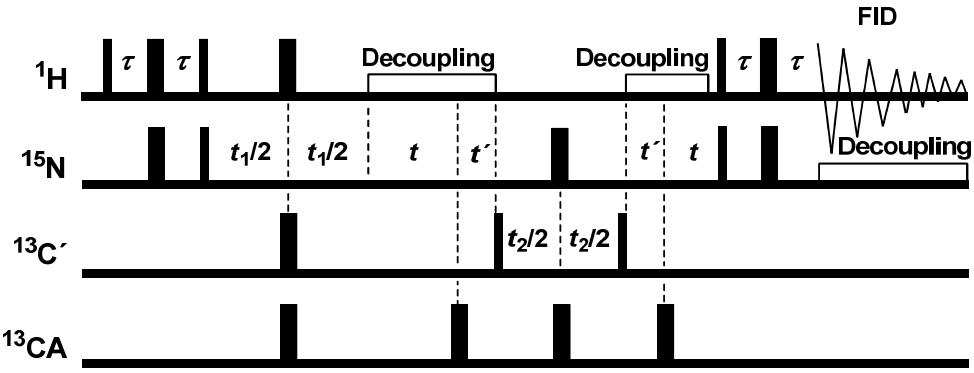


Figure 2.5: The simplified 3D HNCO pulse sequence. Thin and thick vertical bars represent 90º$_x$- and 180º$_x$-pulses, respectively. Different delay times, τ, t, $t´$, $t_1/2$ and $t_2/2$, are indicated. Recent HNCO sequences use shaped band-selective pulses in the ^{13}C-channel instead of hard 90º- and 180º-pulses in order to efficiently suppress signal splitting resulting from N^{H}-C^{α}-couplings and select C´ signals with high specificity. Gradients are used to suppress the dominating H_2O-signal and to select for the desired magnetization transfer pathway. Note that the building blocks of the 2D ^{1}H-^{15}N HSQC experiment are included in the beginning and the end of the 3D HNCO pulse sequence.

HNCO is combined with the HN(CA)CO experiment [70], which detects both the preceding and the intra-residual C´ shifts, where the latter in general yield the stronger peaks. Still, HNCO is usually used for discrimination between the intra-residual and the preceding C´ signals. Figure 2.6 shows which spins are detected by these two methods.

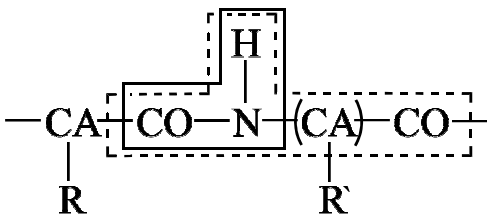


Figure 2.6: The HNCO and the HN(CA)CO experiments. HNCO (solid line) connects the backbone amide with the preceding backbone carbonyl. HN(CA)CO (dashed line) detects both the preceding as well as the intra-residual backbone carbonyl. The brackets indicate that the C^{α} shifts are not observable in the spectrum.

Unfortunately, HN(CA)CO is the least sensitive of the four basic 3D experiments. Preceding C´ signals are often missing, and without them it is not possible to link spin systems to each other. The reason for the low sensitivity is the fast relaxation of the transverse C^{α} magnetization which is further increased by increasing protein size [60]; as a result, amplifying the number of scans has little effect on the signal strengths. Another drawback of the HNCO/HN(CA)CO pair of experiments is that the carbonyl chemical shifts are much less amino acid type specific than the aliphatic shifts, i.e. assigning correlations to specific amino acids in the sequence is very difficult.

The HNCA and HN(CO)CA experiments [67] are less sensitive than HNCO but efficiently more sensitive than HN(CA)CO. Furthermore, the sensitivity difference between the two methods is smaller than in the other experiment pair. The HN(CO)CA spectrum contains the 2D ^{1}H-^{15}N HSQC peaks separated by the preceding C^{α} chemical shifts in the third dimension. Hence, like in HNCO there is only one signal per residue. HNCA is used to link the intra-residual with the preceding C^{α}. Figure 2.7 shows which spins are detected with these two methods.

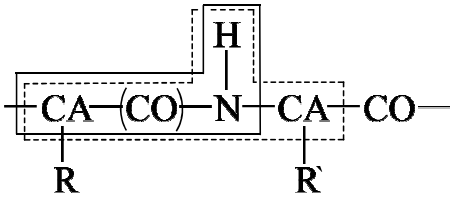


Figure 2.7: The HNCA and HN(CO)CA experiments. HN(CO)CA (solid line) connects the backbone amide with the preceding C^{α}. HNCA (dashed line) detects both the preceding and the intra-residual C^{α} shifts.

The C^{α} shifts can be used to identify glycine correlations, as chemical shifts below 47 ppm in general belong to this amino acid. Shifts around 50 ppm strongly indicate alanine; thus, the HNCA/HN(CO)CA pair of experiments can be used to assign specific segments of the protein sequence.

The linkage of spin systems is performed by the so-called “sequential walk” (figure 2.8). Concentrating only on the ^{13}C shifts in “strips”, one looks for matching C^{α} pairs. This can be accomplished because HNCA detects both the intra-residual and the preceding C^{α} shifts. Thus, neighbouring amino acids must share a C^{α} with identical chemical shifts, the intra-residual C^{α} of residue I and the preceding C^{α} of residue I+1.

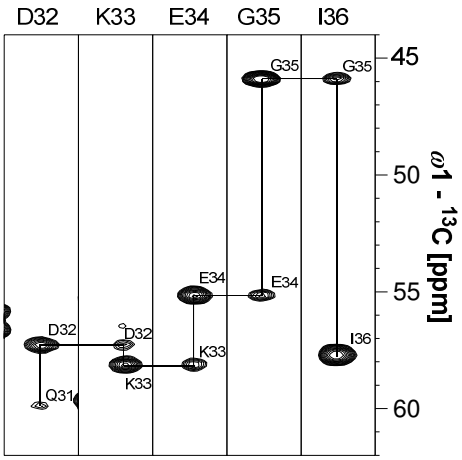


Figure 2.8: Sequential walk. The figure shows strips of the HNCA spectrum of human ubiquitin collected at 600 MHz. The sequential walk can be carried out either way. Note that intra-residual C^{α} shifts are in general stronger than the preceding ones, but that the identification of the intra-residual and preceding C^{α} shifts is usually established from the HN(CO)CA experiment.

2.3.4 Side chain experiments

Unfortunately, the backbone experiments are rarely sufficient for the entire sequential assignment of proteins larger than 10 kDa. This is due to missing signals because of low sensitivity, signal overlap, solvent or conformational exchanged amides and the presence of prolines. Strip matching in larger proteins is also complicated by the common presence of more than one matching C^{α} and C´ chemical shift pair for several residues. Therefore, it is useful to conduct another experiment pair, such as HNCACB and CBCA(CO)NH, which involves the C^{β} chemical shifts.

CBCA(CO)NH [71] is a reasonably sensitive experiment that results in a spectrum containing the C^{α} and C^{β} chemical shifts of the preceding residue, while the HNCACB spectrum [72, 73] additionally shows the intra-residual C^{α} and C^{β} chemical shifts. Unfortunately, the sensitivity of the latter experiment is even less than that of HN(CA)CO. However, the intra-residual shifts, which yield the stronger peaks, may provide an indication about the amino acid type. Furthermore, the preceding C^{α} shift is likely to be known from the HN(CO)CA spectrum, and thus matching of neighbouring residues may still be feasible. Another drawback of the HNCACB experiment is that it results in four signals per residue leading to massive overlap in regions with high backbone amide shift redundancy.

Still, the CBCA(CO)NH and HNCACB experiment pair has two major advantages. First, if two residues possess matching C^{α}, C^{β} and C´ chemical shifts, it is very likely that they are neighbours in the protein sequence. Second, C^{β} shifts are very helpful in the assignment work. They identify glycine, lacking C^{β}, and a C^{β} shift value below 20 ppm in general belongs to alanine. Furthermore, the side chains of serine and threonine are recognized by a C^{β} shift that is situated at a higher chemical shift value than the C^{α}. The remaining 16 amino acids cannot be identified at that point, but they can usually be placed into several subgroups.

Figure 2.9 shows the spins that are observed with these two methods.

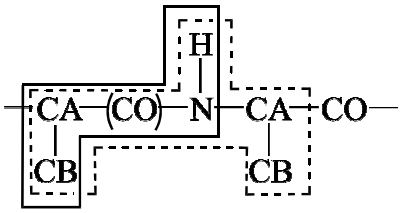


Figure 2.9: The CBCA(CO)NH and HNCACB experiments. CBCA(CO)NH (solid line) connects the backbone amide with the preceding C^{α} and C^{β}. HNCACB (dashed line) detects both the intra-residual and the preceding C^{α} and C^{β} chemical shifts.

The corresponding H^{α} and H^{β} chemical shifts can be determined by a similar approach using the related HBHA(CO)NH experiment [71].

2.3.5 Amino acid type specific experiments

Determining the amino acid identity of a spin system is very helpful in proteins larger than 10 kDa. It can be used to assign a segment to a specific part of the protein sequence or to confirm already assigned stretches of the primary structure. Furthermore, it allows for the extension of known segments in regions with shift overlap if only one of the matches corresponds to the correct amino acid type.

The 3D method for assigning all scalar (through-bond) coupled aliphatic carbon shifts in the amino acid side chains is the CC(CO)NH experiment [74, 75]. At best, this method can separate serine from threonine, and identify the following amino acids: valine, leucine, isoleucine, proline, arginine and lysine. The remaining ten amino acids can be divided into three groups: 1) glutamine, glutamic acid and methionine; 2) asparagine, aspartic acid, phenylalanine, tyrosine and reduced cysteine; 3) histidine,

tryptophan and oxidized cysteine (figure 2.10). Thus, CC(CO)NH can be extremely helpful in the assignment work.

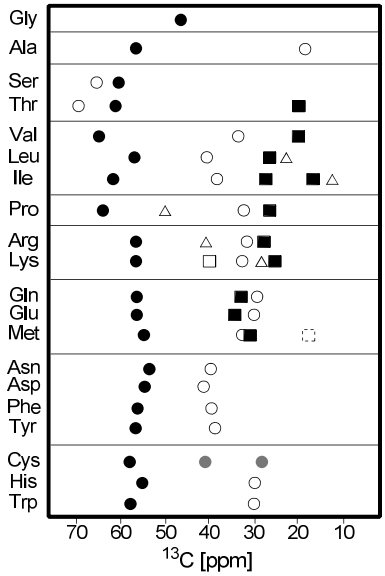


Figure 2.10: The average aliphatic ^{13}C chemical shift values from the biomagnetic resonance bank (BMRB) [76]. C^{α} shifts are indicated with solid circles (●), C^{β} shifts with hollow circles (○), C^{γ} shifts with solid squares (■), C^{δ} shifts with triangles (△) and the C^{ε} shifts with hollow squares (□). The two C^{γ} groups of valine and the two C^{δ} groups of leucine have almost identical chemical shifts and are represented by only one symbol, whereas the C^{γ} groups of isoleucine are both shown because they are in different shift ranges. The two different C^{β} shifts of cysteine (grey circles) indicate the oxidized (approximately 40 ppm) and the reduced (approximately 30 ppm) form of the amino acid. Note that the C^{ε} shift of methionine (dashed) cannot be detected by experiments based on scalar couplings. The amino acids are grouped according to their chemical shift ranges as described in the text.

Unfortunately, the method has some restrictions. CC(CO)NH is TOCSY-based; hence, the signal strengths in proteins larger than 10 kDa are reduced due to short T_2 times and cannot be sufficiently improved by an increase in the number of scans. Furthermore, the time needed for a complete magnetization transfer increases with increasing side chain length [77, 78]. In small proteins, one can compromise and use a mixing time that suits C^{γ} best (figure 2.11). Since the experiment is reasonably sensitive for small proteins, most side chain carbons can be detected this way. However, this approach fails with proteins larger than 10 kDa, because the weakest signals are not stronger than the noise of the spectrum. One option is to run the same experiment with several different isotropic mixing times, which is very time-consuming because each experiment can easily take 48 hours. The opportunistic

approach, thus, is to use a long mixing time which has proven to give sufficient signal strengths for remote side chain shifts. Since the C^{α} and C^{β} shifts were already obtained from the HNCA, HN(CO)CA, HNCACB and CBCA(CO)NH experiments, one can identify several amino acids following this approach.

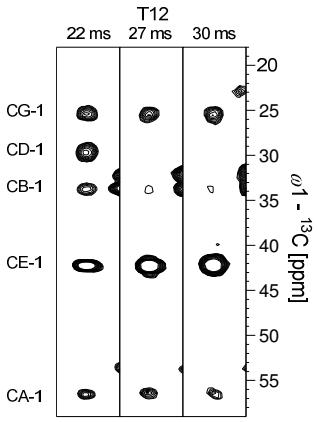


Figure 2.11: Effect of different isotropic mixing times on the signal strengths of the ^{13}C aliphatic side chain signals in the 3D CC(CO)NH spectrum collected at 600 MHz. The figure shows the spin system of K11 in human ubiquitin in all strips. Using a mixing time of 22 ms (left), all five side chain signals are observable. Increasing the mixing time to 27 ms (middle) has little effect on the C^{α} and the C^{γ} peaks, but the signal strength of C^{β} decreases while the C^{ε} peak intensity increases. The C^{δ} peak is no longer observable. At 30 ms (right), the C^{β} signal is further reduced, whereas the C^{ε} and, to a lesser degree, the C^{γ} peak heights increase. Thus, the ideal mixing time for a lysine residue in human ubiquitin is approximately 22 ms. But in larger proteins with less sample concentration and reduced signal strengths due to short T_2 values, a mixing time in the range of 30 ms is more useful because it increases the probability of observing the C^{ε} in the spectrum. This is helpful, because a remote side chain signal at approximately 41 ppm is a strong lysine indicator.

The aliphatic proton shifts can be detected by the CC(CO)NH-related H(CC)(CO)NH method [74, 75]. Due to the small ^{1}H-^{1}H coupling constants, this experiment is even less sensitive than CC(CO)NH for proteins larger than 10 kDa.

An alternate way to assign the aliphatic side chain correlations in proteins is the use of the 3D HCCH-TOCSY experiment [79] which has its starting point in the 2D ^{1}H-^{13}C HSQC method. The advantage is that both proton and carbon side chain shifts can be obtained simultaneously. The disadvantages are poor sensitivity in larger proteins, the requirement of a good resolution in the ^{1}H-^{13}C HSQC spectrum and that it is more tedious to assign spectra which contain signals not originating from the backbone amide peaks.

A different approach of identifying specific amino acid types is the ensemble of MUltiplicity Selective In-phase Coherence transfer (MUSIC) experiments [80-84]. These methods yield amino acid edited 2D ^{1}H-^{15}N HSQC spectra. The pulse sequences are all based on 3D experiments and are highly selective: in the ^{13}C-channel, several pulses, often at amino acid type specific shifts, are used to select only one amino acid type, or a group of similar amino acids. The other residues are efficiently suppressed. Thus, by comparing the MUSIC-derived spectra with the original ^{1}H-^{15}N HSQC spectrum, amino acid types can be directly assigned, without detour over 3D experiments. An example is shown in figure 2.12.

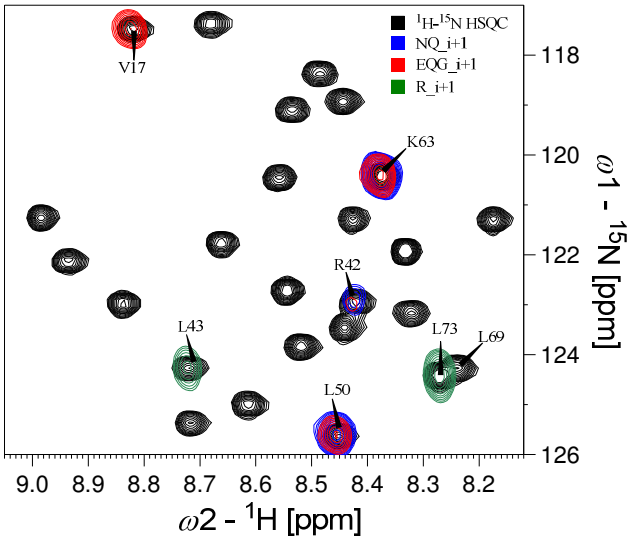


Figure 2.12: Overlay of a part of the 2D ^{1}H-^{15}N HSQC spectrum of human ubiquitin and three MUSIC experiments that all were collected at 600 MHz. Since the blue peaks indicate successors of asparagine and glutamine, while the red peaks represent successors of glutamine, glutamic acid and glycine, residues that follow after glutamine (L50, R42 and K63) can be directly identified, because their peaks are present in both MUSIC spectra. The green peaks indicating successors of arginine demonstrate how MUSIC can be valuable in overlap regions: the R_i+1 experiment can in this case be used to determine which one of the two leucines, L69 and L73, is the successor of R72.

Again, the sensitivity of these methods is highly dependent on the protein size. Additionally, the more transfer steps are included in the pulse sequences, the less sensitive are the experiments. Unfortunately, methods that yield good signal strengths even with proteins larger than 15 kDa involve the amino acids glycine, alanine and serine, i.e. amino acids types which are easily identified by standard 3D experiments.

The amino acids that can readily be detected in proteins of 10-15 kDa, and also to some extent in proteins of 15-20 kDa (own results), are those with acids, amides and aromatic rings in their side chains plus the two aliphatic amino acids valine and isoleucine. Thus, MUSIC experiments are good supplements to 3D CC(CO)NH since most of the above mentioned amino acids belong to the group that cannot be specifically assigned using CC(CO)NH.

2.4 Structure and assignment information from NOE-based spectra

In contrast to all previously discussed techniques, NOESY experiments do not depend on scalar but on dipolar couplings. Thus, in NOESY experiments correlations between two spins can be observed if they are within a limiting distance of approximately 5 Å. This allows for the detection of spins that are far apart from each other in the amino acid sequence if they are spatially close in the tertiary structure of the protein. Therefore, NOESY experiments play an important role in structure determination. Importantly, an increase in protein mass does not reduce the signal strengths. Still, large proteins will show significant signal overlap leading to shift ambiguities.

The peak intensities depend on the distance, indirectly proportional to r^6, between two coupled spins, and are thus of importance for structure calculations [63]. The signal strengths are also influenced by the NOESY mixing times; typically a series of spectra with differing mixing times is collected.

Due to the increased number of spin correlations compared to TOCSY based methods, 2D ^{1}H-^{1}H NOESY is usually limited to a protein size of approximately 8-12 kDa. Most heteronuclear 3D NOESY experiments combine the regular homonuclear NOESY with either the ^{1}H-^{15}N HSQC or the ^{1}H-^{13}C HSQC experiment. An example of the former is the ^{15}N NOESY-HSQC method [65]: all protons within a range of approximately 5 Å to the backbone amides will appear in the third frequency domain. As the distances within a spin system depend on the amino acid conformation, ^{15}N NOESY-HSQC might show fewer intra-residual peaks for amino acids with large side chains than the 3D H(CC)(CO)NH experiment.

Another useful technique is the 3D ^{13}C-HSQC-NOESY-^{15}N-HSQC or CN-NOESY experiment [85]: still, it is the protons that have to be within the limiting distance, but the proton magnetization is transferred to the directly coupled ^{13}C which then is

detected in the third frequency domain (figure 2.13). Since carbon signals show a larger chemical shift dispersion than their correlated protons, this method reduces the ambiguity of long-range NOE-signals. The drawbacks are the reduced sensitivity compared to the ^{15}N NOESY-HSQC and the lack of NOE-correlated amide signals which are of great importance for the secondary structure determination. Hence, ^{13}C-HSQC-NOESY-^{15}N-HSQC is a supplement and cannot replace the ^{15}N NOESY-HSQC experiment.

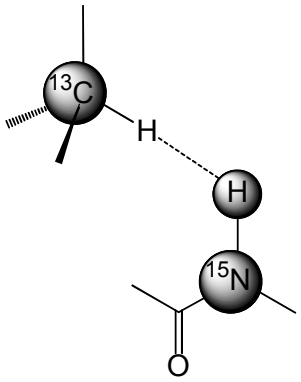


Figure 2.13: 3D ^{13}C-HSQC-NOESY-^{15}N-HSQC. The experiment detects ^{13}C-signals whose correlated protons are within a distance of 5 Å to a backbone amide proton.

2.5 Protein dynamic experiments

The combination of three parameters can give insight into the dynamic properties of a protein: the ^{15}N relaxation times T_1 and T_2 (in some cases $T_{1\rho}$ is detected instead of the latter), and the heteronuclear NOE (an overview of all dynamic motions in proteins is shown in figure A.3). These three parameters can be determined for each resonance in the 2D ^{1}H-^{15}N HSQC spectrum, provided that there is little or no overlap with other peaks. The protein dynamic experiments are all based on ^{1}H-^{15}N HSQC [86].

The spin-lattice, or longitudinal, relaxation time, $T_{1,}$ increases with increasing molecular mass and is typically close to 1 s (figure 2.14) [86]. Hence, series of HSQC spectra are recorded with delays that are in the range of the expected relaxation time, e.g. 0.1 – 2 s. Residues with T_1 values above average experience less mobility than those with smaller T_1 values.

The spin-spin, or transverse, relaxation time, T_2, decreases, as discussed earlier, with increasing molecular mass [61]. T_2 values for proteins are typically within 50-150 ms

(figure 2.14), and HSQC spectra are recorded with relaxation delays of 0 to 200 ms. Residues with T_2 values above average are more mobile than those with short T_2 times. Heteronuclear ^{15}N-{^{1}H} NOE-connectivities (het*NOE*) are measured as the intensity ratios of the peaks belonging to two HSQC spectra, one with and the other without proton pre-saturation. The recycling delay in these experiments is chosen to be longer than usual, i.e. larger than 3 s, to avoid signal loss due to T_1-saturation. Mobile residues have het*NOE*-values below approximately 0.65 [86].

Also, the magnetic field strength influences the dynamic parameters: T_1 and het*NOE* increase with increasing field strength, whereas T_2 is to a lesser degree affected by the magnetic field. The relationship between T_1, T_2 and het*NOE* and the correlation time (time scale of motion), τ_c, at 100 and 600 MHz, respectively, is shown in figure 2.14.

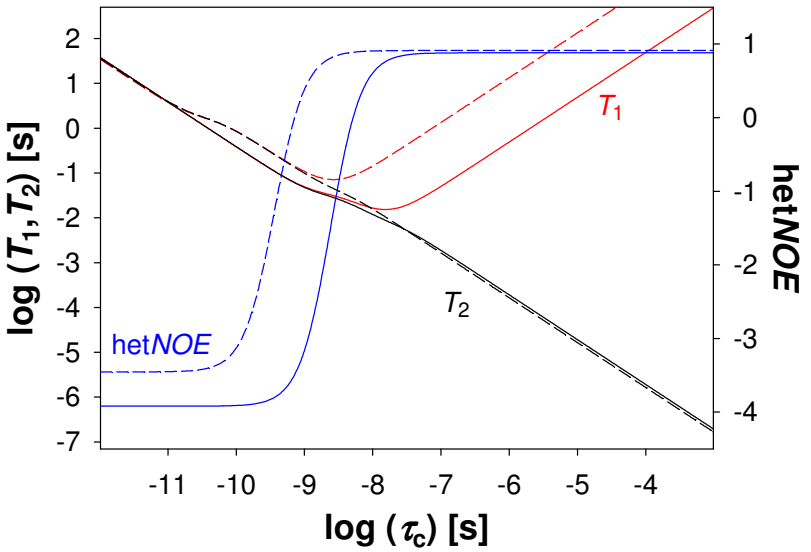


Figure 2.14: ^{15}N T_1, ^{15}N T_2 and het*NOE* dependences on the correlation time (τ_c) and the magnetic field strength. τ_c can be estimated to be approximately MW/2400 (ns) from Stoke's law [87][1]; hence, log (τ_c) for spherical proteins larger than 10 kDa is less than ca. -8.4. Solid and dashed lines represent magnetic field strengths of 100 and 600 MHz, respectively. The plots were calculated from the equations A.1-3 using SigmaPlot 11.0 [88].

[1] $\tau_c = \frac{4\pi\eta r^3}{3kT}$, where η is the viscosity of the solvent and r the effective hydrodynamic radius of the protein.

2.6 Coupling constants

The dihedral H^N-H^α coupling constant around the Φ-backbone angle is of specific interest as its size gives information about the value of Φ, and thus the backbone conformation.

The H^N-H^α coupling constant can, for instance, be obtained using the 3D *J* HNCA E.COSY experiment [89], where the signals are split by ^3J (H^N-H^α) in the ^{1}H dimension and by ^1J (C^α-H^α) in the ^{13}C dimension (figure 2.15). ^3J (H^N-H^α) couplings are in general small, less than 10 Hz, thus the much larger ^1J (C^α-H^α) is used for better signal separation.

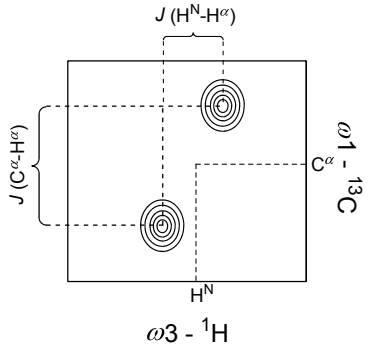


Figure 2.15: Detected signals in *J* HNCA E.COSY are split. The position of the original, decoupled, correlation is indicated.

Since each HNCA-signal is split in two, the signal strengths are reduced by 50 %. Due to the small coupling constants it is also necessary to record enough data-points in the ^{13}C and ^{15}N time domains in order to resolve the signals properly.

2.7 Secondary structure determination

The main secondary structural elements in proteins, α-helices and β-strands, can be predicted without structure calculations, just by interpretation of the results from the previously discussed experiments (sections 2.3-2.6). There are four different approaches for secondary structure prediction. The more the results from these approaches agree with each other, the more likely is the predicted structure.

1. Results from protein dynamics

The T_1, T_2 and het*NOE* values obtained for the individual residues cannot separate between different types of secondary structures, but they can be used to locate the

most likely positions for helices and strands. Residues that are ordered, and hence less mobile, are usually situated in an α-helix or a β-strand. These residues are identified by a T_1-value above average, a T_2-value below average and a het*NOE* value larger than 0.70.

2. Chemical shift index

It has been proven that the chemical shift values of amino acids vary with the type of secondary structure. For some chemical shifts, like H^N, the differences are too small to be significant. But the values for the H^α, C^α, C^β and C´ shifts (listed in table A.2) are powerful in allocating secondary structural elements, especially when combined with each other. The shifts are compared to the average random coil chemical shifts of a specific amino acid [90, 91]. α-helices are detected by H^α shifted more than 0.2 ppm upfield (i.e. lower shift value), C^α shifted more than 0.7 ppm downfield, C^β shifted more than 0.7 ppm upfield and C´ shifted more than 0.5 ppm downfield. To verify an α-helix, there must be at least four succeeding residues showing the same pattern.

The same is true for β-strands, but here all resonances are shifted by the same amount in the opposite direction. Residues that do not fit into these two patterns are most likely situated in coils or loops.

Not all of the chemical shifts are affected to the same extent by the secondary structure conformations. In general, the most reliable chemical shifts for identifying an α-helix are the C^α and the C´ shifts, while the H^α and C^β resonances are best suited for predicting β-strands [92].

3. NOE restraints

Even if the results from the 3D ^{15}N NOESY-HSQC experiment are not used for full structural calculations, the signal strengths of specific correlations yield information about the secondary structures.

Typical of α-helices is the strong peak between neighbouring amides, and that the correlation between H^N and the preceding H^β yields a stronger peak than to the H^α. Weak connectivities between H^N and H^α that are three or four residues apart are also a characteristic of helices. In β-strands the adjacent amide signals are weak, but in parallel or antiparallel strands long-range amide connectivities are typically strong. Also, the H^N correlation to the preceding H^α yields a much stronger signal than to the

H^{β} [93]. An overview of common NOE connectivities and their signal strengths for different types of secondary structural elements are shown in figure A.4 and table A.3.

4. Coupling constants

An Φ-backbone angle of approximately -57º is typical of an α-helix. This is true for ^{3}J (H^{N}-H^{α}) values below 6 Hz. The angle for parallel and antiparallel β-strands is approximately -139º and -119º, respectively, corresponding to a ^{3}J (H^{N}-H^{α}) value larger than 8 Hz [94].

2.8 NMR spectroscopy on uncommon nuclei

Some proteins contain amino acids different from the 20 common ones. An example is selenomethionine (SeMet/SeM) (figure 2.16), in which selenium substitutes sulphur.

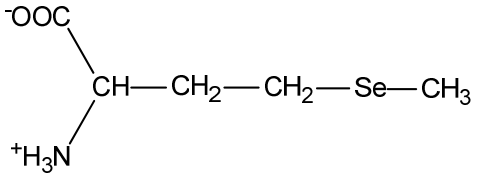


Figure 2.16: The structure of selenomethionine. The sulphur (δ-position) is substituted by selenium.

Selenium has six naturally occurring isotopes and only one of them, ^{77}Se, possesses a nuclear spin of ½. Unfortunately, isotopically enriched selenium compounds for sample preparation are not yet accessible; hence, one has to cope with a natural abundance of only 7.6 % and a sensitivity comparable to that of ^{13}C [95]. The advantages of ^{77}Se NMR spectroscopy lie in the wide chemical shift range of the nucleus and its pronounced sensitivity towards changes in the chemical environment [95]. For instance, the nine selenomethionines in calcium-saturated calmodulin showed a ^{77}Se chemical shift dispersion of approximately 60 ppm, whereas the differences in their ^{1}H and ^{13}C resonances were less than 1 ppm [96]. The best way of studying selenomethionines in proteins is by the ^{1}H-^{77}Se heteronuclear multiple bond correlation (HMBC) experiment [97], where the H^{β} and H^{γ} signals usually are not observable [96]. In the experiment, the magnetization is transferred to the protons of the terminal methyl group; thus, the slow relaxation of the ^{77}Se nucleus of approximately 12 s [95] is not an issue. Furthermore, the three protons contribute to one common signal, thus improving the signal intensity.

3. SUMMARY OF RESULTS

3.1 Spectrin repeat 17

3.1.1 Resonance assignment

Approximately 92 % of all ^{1}H, ^{13}C and ^{15}N chemical shifts of R17 could be assigned using standard 3D heteronuclear NMR experiments, specific techniques for the assignment of the aliphatic side chains and a variety of amino acid edited 2D experiments (MUSIC). Unassigned spins are primarily backbone NH-correlations that are missing due to solvent or conformational exchange, and resonances of the non-aliphatic side chains.

3.1.2 Domain structure

The presence of the three long helices which are typical of spectrin repeats were confirmed by different approaches: (1) the chemical shift indices (CSI) of four nuclei (C', C^{α}, C^{β} and H^{α}), (2) the sizes of the ^{3}J (H^{N}-H^{α}) coupling constants and (3) the pattern of sequential and medium-range NOEs. Additionally to the three helices, also the linker between R16 and R17 appears to be helical at the experimental conditions.
The interactions of the helices with each other in the triple-helix bundle were detected using well-separated peaks in the 3D ^{13}C NOESY-HSQC spectrum. Furthermore, the exchange pattern of the backbone-NH chemical shifts in a sample containing 85 % D_2O demonstrated that the residues which are protected from instant solvent exchange are all part of the triple-helix bundle. Solvent-exposed amide protons in the helices, on the other hand, are exchanged in the same time-range as the unstructured parts of R17.
Overall, the determined secondary structures and the obtained information about the tertiary structure of R17 are in very good agreement with the crystal structure of the tandem repeat R1617 (PDB-entry 1CUN).

3.1.3 Domain dynamics

The profile of the ^{15}N T_1 and T_2 relaxation times and the ^{15}N-{^{1}H} het*NOE* values match with the position of the helices. The most flexible parts of R17 are, as expected, the N- and C-termini, and the AB- and BC-loops. In addition, the conformation of the

stretch of consecutive glutamic acids, ^{21}EEE^{23}, and of two residues in the vicinity of the BC-loop, E75 and K93, appears to be less rigid than that of the other helical residues.

The calculation of the diffusion tensor showed the highly anisotropic nature of the triple-helix bundle structure, which is observable as an oscillation of the T_1 and T_2 values in the two helices, A and B, that are not completely parallel with the main rotational axis of the molecule. This oscillation coincides with the turns of the helices. The efficient, global correlation time, $\tau_{c,eff}$, was calculated to be approximately 9.4 ns and is within the expected range of a 13 kDa domain. The order parameter values, S^2, clearly show the flexible N- and C-termini, while the loops appear to be more rigid than determined by other experiments.

3.1.4 Domain stability

The thermal stability of R17 was investigated by circular dichroism (CD) and fluorescence and NMR spectroscopy. The CD analysis was carried out at 222 nm (the wavelength that shows a characteristic polarization angle minimum for α-helices) and the optical rotation of the sample was measured for each 0.1 ºC increment of temperature. The melting temperature, i.e. the transition from helical to random coil conformation observed as a distinctive decrease in the absolute value of the polarization angle, read from the plot was approximately 40 ºC (figure 3.1). The unpacking of the single tryptophan residue (W26) by fluorescence spectroscopy was studied by irradiating the sample at a wavelength of 295 nm and measuring the intensity of the emission peak at increasing temperatures. A buried indole ring will emit light with a lower relative intensity than an exposed ring. Thus, the intensity of the emission peak can be used as an indicator of the state of the side chain. W26 appeared to unpack at approximately 40 ºC (figure 3.1). The thermal stability of R17 by NMR spectroscopy was examined by a series of 2D ^{1}H-^{15}N SOFAST-heteronuclear multiple quantum coherence (HMQC) experiments [98] at increasing temperatures. First signs of global unfolding appeared at approximately 46 ºC in the NMR measurements. According to the NMR results, the melting seems either to be an effect of unfolding starting from the BC-loop or a local unfolding of helix C. The protein denaturation appeared to be reversible as observed by NMR spectroscopy.

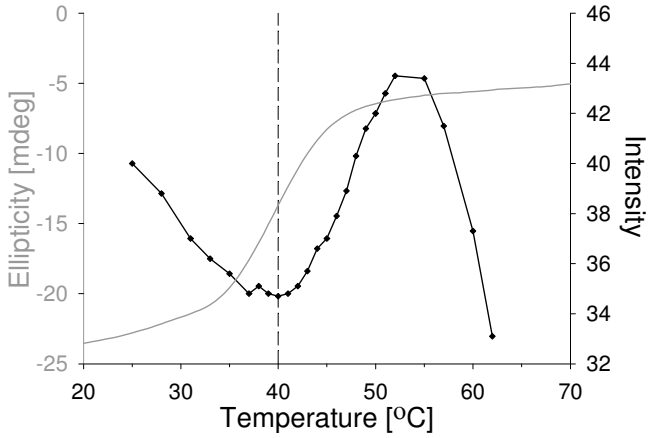


Figure 3.1: Thermal unfolding of R17. CD measurement at 222 nm (grey curve, the data were smoothed) and maximum peak intensity of the fluorescence emission of the indole ring of W26 (black curve). The dashed line at 40 °C indicates the approximate temperature at which the helices unfold and the tryptophan side chain unpacks, respectively.

Finally, in a bioinformatical study, the amino acid sequence of R17 was compared to 34 other chicken brain spectrin repeats in order to identify single amino acids or sequence stretches that could explain the relatively low stability of R17 compared to other spectrin repeats. Three diverging features in the sequence were detected: (1) four subsequent glutamic acids, ^{21}EEEE^{24}, where the negative side chains of E21 and E24 are spatially close. (2) An unusual amount of charged residues in both the linker and the BC-loop. (3) The lack of the moderately conserved tryptophan in helix C. Most spectrin repeats have two tryptophan residues which have shown to be important for the structural stability of a repeat. In R17, the second tryptophan is substituted by the relatively small valine.

3.2 Human Naa50p

3.2.1 Resonance assignment

Approximately 85 % of the backbone-NH-signals were assigned to the primary structure of hNaa50p. The ^{13}C aliphatic side chains were partially assigned. 3D CC(CO)NH and a series of 2D MUSIC experiments were used to simplify the assignment work. Missing amide assignments are mainly caused by peak overlap and signal broadening due to conformational exchange.

3.2.2 Secondary structures and protein dynamics

The secondary structural elements of hNaa50p were determined by the CSI of three nuclei (C´, C^{α} and C^{β}). The proposed positions of the helices and strands are in good agreement with the secondary structures detected by X-ray crystallography (PDB-entry 2OB0). The only differences are that the second helix (residue position 32-42) appears to be shorter than in the crystal, and that the region around I142 resembles a strand rather than a loop.

T_1 and T_2 relaxation times and ^{15}N-{^{1}H} het*NOE* values were determined and used to calculate the order parameter, S^2, using a fully anisotropic model. The rigid parts of the protein coincide with the position of the secondary structures, with exception of the region encompassing the first two helices (residues 17-38) and the helical stretches ^{73}YIMTLG^{78} and ^{87}GIGTKML^{93}.

3.2.3 Enzyme mechanism

Both kinetic assays and NMR spectroscopy were used to explore the enzymatic mechanism of hNaa50p. Initial velocity kinetic experiments where the concentration of the first substrate, acetyl-CoA, was increased while the concentration of the second substrate, peptide, was held constant resulted in an intersecting line pattern consistent with the formation of a ternary complex (figure 3.2A). Thus, the ping-pong mechanism could be ruled out (figure 3.2B). Product inhibition studies were carried out using one of the products, CoA, as inhibitor while the concentration of acetyl-CoA was increased. The result showed that CoA acts as a competitive inhibitor (figure 3.2C). When CoA was used as a product inhibitor, the concentration of acetyl-CoA was kept constant and simultaneously the concentration of peptide was increased, a non-competitive inhibition pattern was observed (figure 3.2C). All these results point towards that hNaa50p follows an ordered sequential mechanism of the rare Theorell-Chance type. However, when acetylated peptide was used as a product inhibitor against both increasing levels of peptide and acetyl-CoA no inhibition was detected, indicating that the affinity between acetylated peptide and the enzyme is very low.

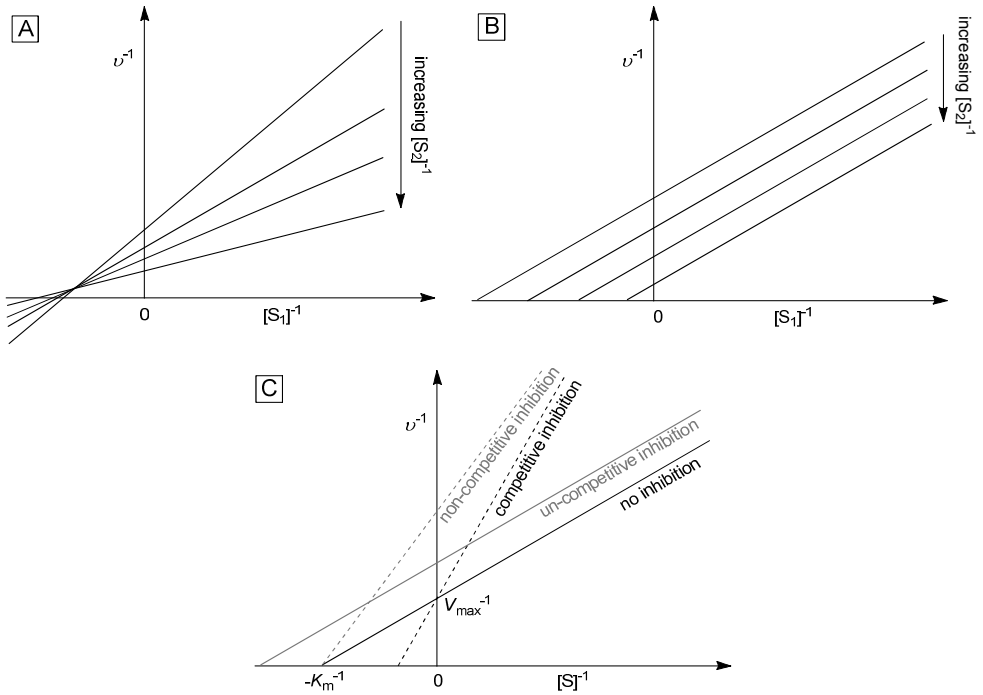


Figure 3.2: Lineweaver-Burk (or double-reciprocal) plots of the enzyme kinetics. Note that this figure is not based upon the actual enzymatic data, but illustrates how the results were interpreted. **A)** Intersecting line pattern in the ternary complex formation mechanism. $[S_1]$ and $[S_2]$ refer to the concentrations of the two substrates, whereas υ is the reaction rate. **B)** Parallel line pattern in the Ping-Pong reaction mechanism. **C**) Different types of product inhibition. V_{max} is the maximum reaction rate, whereas $K_m = ½\ V_{max}$ is the Michaelis constant. In competitive inhibition K_m is reduced, whereas V_{max} remains constant. Non-competitive inhibitors reduce V_{max} while K_m stays the same. Finally, un-competitive inhibitors reduce both V_{max} and K_m.

NMR spectroscopy cannot only confirm a kinetic mechanism, but it can also point out the amino acids that most likely are involved in the enzyme reaction. The 2D ^{1}H-^{15}N HMQC spectrum of hNaa50p was altered in a similar way when either acetyl-CoA or CoA was added to the protein, indicating that both substrate and product bind to the same site. This was expected because CoA acted as a competitive inhibitor of acetyl-CoA in the kinetic assay. When the other substrate, peptide, was added to the enzyme no obvious change in the spectrum was observed. This strongly indicates that hNaa50p follows an ordered and not a random enzymatic pathway. Small alterations in the HMQC spectrum of hNaa50p in complex with CoA were observed when unacetylated peptide was added to the sample, whereas the addition of acetylated peptide had no such effect. The latter observation matches the kinetic assay where the acetylated peptide did not act as product inhibitor. Furthermore, binding of both substrates

simultaneously resulted in more spectral changes than binding of peptide to the hNaa50p-CoA complex, indicating that acetyl-CoA is the better mediator for peptide binding. The effect of simultaneous binding of both substrates was studied using the enzymatically inactive but natively folded hNaa50p H112A mutant [99]. Taken together, the NMR data suggest that hNaa50p follows a type of sequential ordered mechanism.

4. DISCUSSION

4.1 Spectrin repeat 17

4.1.1 Assignment challenges

The collection and interpretation of NMR data are usually not as straightforward as demonstrated on the template protein human ubiquitin in chapter 2. In the following paragraph, the challenges experienced in the assignment work of R17 are discussed.

The studied R17 construct with its 118 residues is a little larger than the average 106 amino acids of a spectrin repeat. That is because the construct includes the C-terminal of R16 and the linkers to both R16 and R18. The linker to R16 is in particular important for the stability of R17, because the linkers interact with the BC-loop of the following repeat [100].

In spite of a molecular mass of about 13 kDa which classifies the domain as medium-sized in NMR terms, R17 in some respects behaved more like a protein of approximately 18-20 kDa. The reason for that are the long α-helices which dominate the structure of the repeat. Even though the line-widths of the peaks corresponded to a medium-sized protein and led to a good peak resolution, the proton side chain assignment work was complicated by the small helical ^{3}J (H^{N}-H^{α}) coupling constants which inhibit the magnetization transfer throughout the chain [63]. For example, the 3D ^{15}N TOCSY-HSQC spectrum contained only side chain signals for amino acids positioned at the N- and C-termini and in the AB-loop, i.e. all residues that are not part of the α-helices.

In addition, residues positioned in α-helices in general show less signal distribution in the H^{N} chemical shift range than residues situated in β-strands. That means that all-helical proteins are prone to major shift overlap in the 2D ^{1}H-^{15}N HSQC spectrum (figure 4.1). This shift overlap complicates both the backbone assignment work because it usually also leads to shift overlap and ambiguity in the third dimension, and structure calculations based upon the 3D ^{15}N and 3D ^{13}C NOESY-HSQC spectra which already suffer from a vast amount of, partially overlapping, peaks. Thus, in order to assign the NH-correlations correctly, it was necessary to combine the standard 3D experiments (HNCA, HN(CO)CA, HNCO, HN(CA)CO, HNCACB and

CBCA(CO)NH) with side chain detecting experiments, like 3D CC(CO)NH, and amino acid specific experiments, 2D MUSIC. By these means, an assignment of all the peaks in the 2D ^{1}H-^{15}N HSQC spectrum, except for the unique identification of some of the arginine $N^{\varepsilon}H^{\varepsilon}$ side chain signals, could be obtained.

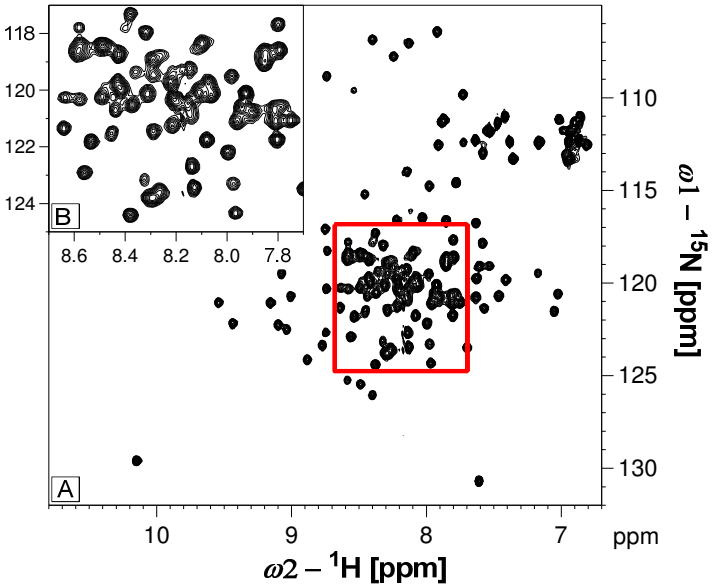


Figure 4.1 A) ^{1}H-^{15}N HSQC spectrum of R17 at 600 MHz. **B)** Enlargement of the overlap region (marked red in the original spectrum). 65 out of the assigned 109 backbone-NH signals are found within this region.

4.1.2 Domain structure

The most severe overlap appeared in the 3D ^{15}N NOESY-HSQC spectrum. Even though R17 is relatively small, it might have been necessary to conduct the NOESY experiments at a higher magnetic field strength in order to resolve the regions with a high degree of shift overlap and/or to partially deuterate the sample. Thus at 600 MHz, the majority of the weak medium-range NOEs were undetectable, because the signals were buried by stronger short-distance peaks in the same chemical shift range. Furthermore, some of the weakest short-distance intra- and inter-residual NOEs were not observable.

Since all resonance assignment experiments were conducted on a sample containing 90 % H_2O, it was not possible to use the 3D ^{13}C NOESY-HSQC experiment for structure determination, because the residual water peak hid the H^{α} and H^{β} signals

resonating close to the water peak and because of several water noise bands in the spectrum. Furthermore, the aromatic side chains could neither be assigned from the ^{13}C NOESY-HSQC nor the 3D HCCH-TOCSY spectrum, as the aromatic CH-groups suffer from fast transverse ^{13}C relaxation in water-containing samples [101]. However, it was feasible to assign most of the H^{δ} and H^{ε} resonances of phenylalanine and tyrosine from the 2D (HB)CB(CGCD)HD and 2D (HB)CB(CGCDCE)HE experiments [102].

Both due to severe peak overlap in the ^{15}N and ^{13}C NOESY-HSQC spectra and only partial assignment of the aromatic side chains, the initial structure calculation, which was carried out in the program CYANA 2.1 [103], did not yield a sensible result. But since there exist crystal structures of both the double repeat R1617 (PDB-entry 1CUN) and the triple repeat R15-17 (PDB-entry 1U4Q), a redefining of the R17 structure by NMR spectroscopy would have had little scientific significance. Nevertheless, the chemical shift values, the pattern of inter-residual NOEs and the sizes of the ^{3}J (H^{N}-H^{α}) coupling constants were all in agreement with the positions of the α-helices that were determined from X-ray crystallography. In addition, several long-range NOEs within the triple-helix bundle were detected using ^{13}C NOESY-HSQC. Finally, the lack of NOEs between the side chains of W26 and V99 indicates that the small cavity between the three helices in the crystal structure is not an artefact but might be the result of the tryptophan to valine substitution at sequence position 99.

4.1.3 Thermodynamical and conformational stability

The CD analysis showed that the helices of the R17 construct melted at approximately 40 ºC and, according to fluorescence spectroscopy, W26 is unpacked in the same temperature range, whereas the NMR-sample appeared to be melting at roughly 46 ºC. Both temperatures are significantly higher than the previously reported melting temperature of approximately 31 ºC [23]. The difference between the two studied R17 constructs is that we included the linker between the repeats 16 and 17. Thus, the linker apparently increases the thermodynamical stability of R17. The crystal structure (PDB-entry 1CUN) indicates interaction between the linker and the BC-loop through aromatic ring stacking and salt-bridges. In addition, three positive charges in the BC-loop are efficiently neutralized by three negative charges in the linker. We suggest that

the BC-loop in the absence of the linker reduces the thermodynamical stability of the repeat, because of an increased structural flexibility and because of the repulsion of the positive charges. The presence of a negatively charged linker and a positively charged BC-loop is not a general feature of spectrin repeats (see section 4.1.4) and there is, for instance, less interaction observed between the linker and the BC-loop of R16 (PDB-entry 1U4Q). Thus, it seems like R17 is more dependent on the linker in order to be stabilized than most spectrin repeats. Interestingly, the majority of the backbone amide peaks of the linker and the BC-loop were absent from the ^{1}H-^{15}N HSQC spectrum, both at pH 6.0 and at pH 6.8, and in the temperature range of 10 to 48 ºC. Furthermore, the observable amino acid peaks in these regions were broadened, suggesting that the linker and the BC-loop do not adopt one stable conformation, but fluctuate between two or more states at a rate which makes their peaks undetectable for NMR spectroscopy. One can only speculate, if the linker and the loop have a more rigid structure in the mature spectrin molecule, since it has proven to be difficult to study more than one spectrin repeat by NMR spectroscopy at a time. The ^{1}H-^{15}N HSQC spectrum of the double repeat R1617, for instance, only showed well-resolved peaks in the region of the side chain amides, whereas the backbone NH-correlations were weak and showed little chemical shift dispersion in both dimensions [104].

The drawback of CD analysis (figure 3.1) is that it can only measure when denaturation occurs, but not where the unfolding initiates in the molecule. Since R17 only contains one tryptophan residue, fluorescence emission could be used to pinpoint the temperature at which W26 – maybe the most central residue in the hydrophobic core of the domain – gets solvent exposed. First signs of unpacking appeared at approximately 40 °C and the solvent exposal increased rapidly from about 44 °C (figure 3.1). The sudden decline in intensity as the temperature exceeded 55 °C could indicate that R17 precipitated. However, the unfolding of the repeat was mainly studied by NMR spectroscopy where all residues could be investigated at a time. The basic idea was that the residues which are affected by a temperature increment would undergo slow exchange between two or more conformations belonging to the folded and one or several unfolded states. This exchange would be observable as a splitting of the NH-peaks in the ^{1}H-^{15}N HMQC spectrum. Unfortunately, a temperature change affects all chemical shifts, especially the proton resonances, but to a different extent.

Thus, it was difficult to track the chemical shift changes of individual peaks, even though the temperature was raised in small intervals. Furthermore, an overall fluctuation of the domain structure was observed as peak-broadening of all resonances close to the melting point. Thus, resonance splittings in the peak clusters in the overlap region of the spectrum (figure 4.1) were impossible to determine. In addition to 2D HMQC, also 3D HNCO and HN(CO)CA spectra were recorded at 44 °C. These experiments were chosen, because they are the two most sensitive 3D methods and because their spectra only contain one peak per residue, generating minimal peak overlap. At 44 °C most of the observed NH-peak splittings of helical residues have initiated, but the temperature was kept below 46 °C where global unfolding was observed which eventually might result in irreversible protein denaturation. It was expected that helical residues in the regions where unfolding occurs would show changes in their C´ and C^{α} chemical shifts concordant with the transition from helical to random coil values (i.e. move upfield). However, it was observed that seven residues experienced peak-splittings of their C´ and/or C^{α} correlations, indicating a slow conformational exchange between folded (original peak) and unfolded (new peak) state at this particular temperature. Hence, the determination of the unstable parts of the repeat was based upon the approximately 70 well-resolved NH-peaks in the HMQC spectrum and the peak splittings of the backbone carbons. Not unexpectedly, the least stable parts of R17 are the unstructured N- and C-termini. Also, the transitions between helix A and the AB-loop and between the latter and helix B showed changes in conformation at temperatures below 40 ºC. It appears as if the melting within the first helix starts in the central part of helix C, which includes V99, a rare substituent of the partially conserved tryptophan at that position. The importance of the two tryptophan residues in spectrin repeats has been thoroughly discussed [8, 105]. Three aromatic residues involve in interhelical interaction in the hydrophobic core of many spectrin repeats: two tryptophans and one histidine. The pyrrole ring of the almost 100 % conserved tryptophan in helix A stacks onto the imidazole ring of the histidine in helix B, and the large side chain of the partially conserved tryptophan in helix C undergoes hydrophobic interaction with the other two aromatics according to the crystal structure [106] (figure 4.2). Typical substituents of the second tryptophan are other aromatic residues or arginine. Small aliphatic residues, like valine, are rarely

observed. The presence of a valine instead of a tryptophan renders a cavity within the triple-helix bundle, resulting in fewer hydrophobic interactions in the core of R17 than in R15 and R16 [23]. In the central part of helix C a total of five residues experienced signal splittings of their backbone correlations in the temperature range of 40 to 46 °C: G94, L95, G97, V99 and S100. The NH-peak of W26, spatially close to V99, was not split but showed peak broadening at 44 °C, and its predecessor, A25, experienced splitting of both backbone carbon signals. The third amino acid in the aromatic triad in the hydrophobic core of the domain, H63, was not affected by temperature increase. However, another aromatic residue in helix B, F60, is close in space to both W26 and V99 and three residues in that region – E57, T58 and F60 itself – showed backbone peak splittings. All these results lead to the suggestion that the substitution of the tryptophan in helix C triggers the destabilization of the hydrophobic core and that the protein denaturation starts as a break-up of helix C from the triple-helix bundle.

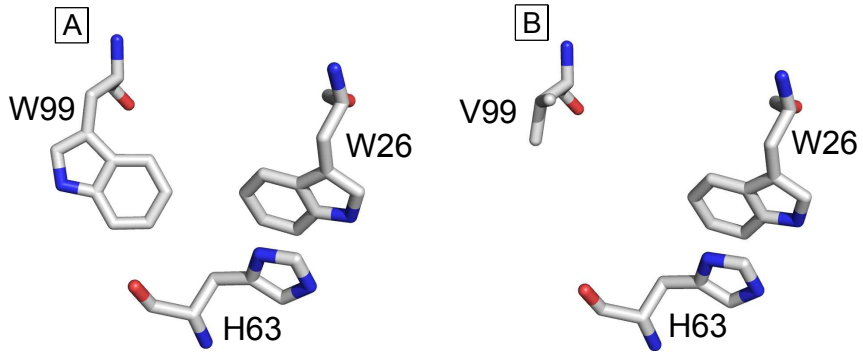


Figure 4.2 A) The three aromatic residues in the core of the triple-helix bundle of R16 (PDB-entry 1CUN). The pyrrole ring of W26 stacks onto the imidizole ring of H63, whereas the indole ring of W99 undergoes hydrophobic interaction with the other two aromatic residues. **B)** Substituting the second tryptophan of R16 by valine renders a cavity in the core of the triple-helix bundle. The *in silico* tryptophan to valine mutagenesis was carried out in the PyMOL Molecular Graphics System, Version 1.3, Schrödinger, LLC.

The conformational stability of R17 at room temperature was studied by relaxation measurements – T_1 and T_2 relaxation times and het*NOE*s – and by proton-deuterium exchange. The obtained results from the ^{15}N-{^{1}H} relaxation study were in overall agreement with the positioning of the secondary structures in the crystal and those determined by NMR measurements. The diffusion tensor was calculated from the relaxation values and the crystal structure (PDB-entry 1CUN) in the program HydroNMR [107]. A factor of approximately 3 between the rotational diffusion rates

of the axes that lie perpendicular and parallel to the principal domain axis indicated that R17 is highly anisotropic. Thus, the fully anisotropic model was used in the analyses of the relaxation data in the program TENSOR2 [108]. The results suggest that R17 is axial symmetric, i.e. has the shape of a prolate spheroid, and that the diffusion tensors of its main and perpendicular axes are approximately 2.9×10^7 and 1.2×10^7 s^{-1}, respectively, resulting in an efficient global correlation time, $\tau_{c,eff}$, of approximately 9.4 ns. The order parameter values, S^2, were calculated by the simple Lipari-Szabo approach [109, 110]. S^2 is difficult to determine for a triple-helix bundle, where the helices are co-linear with each other and the main rotational axis. Thus, the angle between the majority of the N-H vectors, which are crucial for the calculation of S^2, and the main axis are close to 0º or 180º. Only the bend in helix B deviates from the average angles in the helices, which explains the larger dispersion of the relaxation times in this helix. As a result of the small angles, the S^2 values of structured parts of the repeat differ less from the loops and termini than expected from the T_1, T_2 and het*NOE* values. Overall, R17 appears to be very rigid in solution at room temperature.

There are two interesting regions in the secondary structures which are not in agreement with the rest of the helices. The first is the stretch ^{21}EEE^{23} in helix A, where both the T_1, T_2 and het*NOE* values indicated that these amino acids are more flexible than helix A on average. This may be a result of the repulsion of the consecutive negative charges. Furthermore, the NH-peaks of E21, E22 and E23 were broadened, which – together with the increased T_2 values – suggests conformational exchange. The second stretch is the vicinity of the BC-loop, where the relaxation values indicated increased structural flexibility. Also, the NH-peaks of K79 and N81 in the BC-loop were broadened, suggesting conformational exchange. Thus, the thermal denaturation of R17 could alternatively be a consequence of the melting of the helices from the BC-loop as a result of an increased structural fluctuation at higher temperatures.

In the sample containing 85 % D_2O, approximately 50 % of all NH-signals were immediately deuterium-exchanged. Besides residues at the N- and C-termini and in the loops, also solvent-exposed parts of the helices, for instance L43-F56 in helix B, were instantaneously deuterated. Each helix appeared to contain a region where most of the amino acids were protected against immediate H^N-exchange: A18-A25 in helix A, V67-I78 in helix B and K91-A111 in helix C. These stretches are in overall agreement

with the regions that make up the triple-helix bundle in the crystal structure. Still, only eight amino acids were not solvent-exchanged within the first 24 hours: V20, E24, A25, V70, L77, K91, Q108 and A111. The H^N-exchange data did not indicate any conformational instability of the repeat at 23 °C.

4.1.4 Bioinformatics

The thermal and conformational stabilities of spectrins vary with the environment they occupy. For instance, the average melting point of the flexible erythroid spectrins [111] is 10 degrees lower than that of brain spectrins [29]. Hence, in order to determine potentially unstable parts of R17, the amino acid sequence of the repeat was solely compared to other chicken brain spectrin repeats and in particular to the other 19 α-spectrin repeats. The domains 10, 19 and 20 whose sequences are equally related to R17 as random protein sequences (expect-value > 10) were excluded from the multiple sequence alignment. A total of three already mentioned features were discovered in R17 that could explain the relatively low stability of the repeat: the four consecutive negatively charged amino acids in helix A, the unusual amount of positive charges in the BC-loop combined with the same amount of negative charges in the linker, and the tryptophan to valine substitution in helix C.

The R17 primary structure was also compared to 18 chicken brain β-spectrin repeats (data not shown). αR17 is the only domain out of 35 chicken brain spectrin repeats in the multiple sequence alignment that has three negative charges in the linker and three positive charges in the BC-loop. Furthermore, only two other repeats (αR8 and αR16) also possess four consecutive negative charges in helix A. The highly conserved tryptophan in helix A is substituted by other residues in two repeats (βR9 and βR15), whereas the moderately conserved tryptophan residue in helix C is substituted by another aromatic amino acid eight times and by the large amino acid arginine in four repeats. Thus, R17 is the only chicken brain spectrin repeat where the second tryptophan is substituted by the relatively small valine residue.

Nevertheless, the characteristics of R17 seem to be of biological importance as the domain appears to be conserved in spectrins in general. Related sequences were searched for using the tool blastp 2.2.25 [112] with an expect-value limit of 1×10^{-15}. In comparison, the chicken brain α-spectrin repeats that are most closely related to

R17 (R15 and R16) score approximately 1×10^{-13}. The multiple sequence alignment was carried out using the tool Clustal W [113] and the phylogenetic tree was calculated in the program Jalview 2.6.1 [114]. Only proteins that, so far, have been identified as spectrins were included in the alignment. A total of 32 species or families showed to contain an R17-like spectrin repeat: seven vertebrates (mammals, birds, fish and amphibian), 20 insects (beetles, flies and ants), three nematodes and two protists. The multiple sequence alignment is shown in figure 4.3; sequences with more than 90 % sequence redundancy to a related species were removed from the alignment. Table 4.1 lists information about the spectrin repeats that were identified in the Blast search.

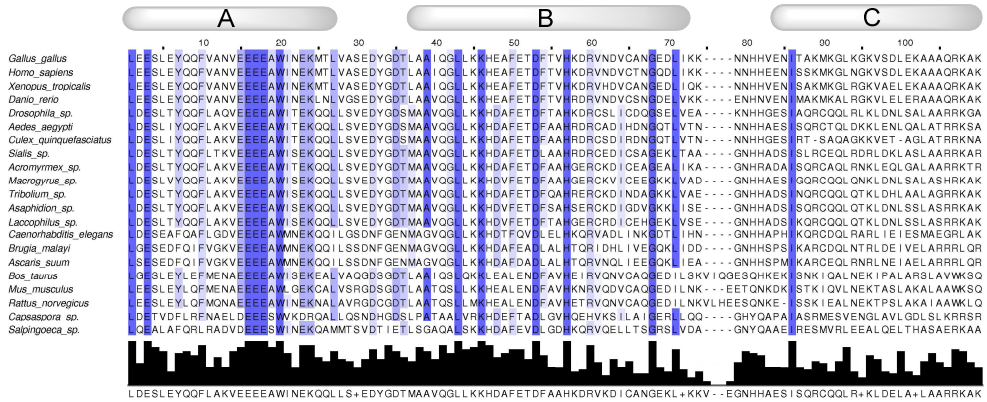


Figure 4.3: Multiple sequence alignment of 21 spectrin repeats that are closely related to chicken brain α-spectrin repeat 17 (*G. gallus*). The sequences are listed according to decreasing sequence similarity. Coloured residues are at least 50 % conserved. The consensus sequence is shown below the alignment. An explanation to the sequences is presented in table 4.1. The alignment was displayed in Jalview 2.6.1 [114]. The positions of the helices are indicated.

The main differences between the sequences are found in the BC-loop and in helix C. Still, the hydrophobic pattern and most of the charged residues are conserved in helix C. The negative charges in the linker and in helix A seem to be typical features of R17-like proteins. The four consecutive negative charges in helix A are 100 % conserved, and in a vast majority of repeats all four residues are glutamic acids. In addition, 35 out of 36 repeats have at least two negative charges in the linker.

However, the pattern of three positive charges in the BC-loop is only found in the vertebrate brain spectrins, except for the zebrafish (*D. rerio*). In a total of 23 cases, the BC-loop has only one positive charge, is neutral or even negatively charged. The greatest sequence divergences in the BC-loop show the mammalian erythroid

spectrins, which probably is correlated to a difference in function between brain and erythroid spectrins.

Table 4.1: Accession numbers and sequence information of R17-like spectrin repeats in 32 species. Organisms that are written in bold are included in the multiple sequence alignment and the phylogenetic tree. Sequences of species which share the same table cell are at least 90 % identical to each other. Thus, some of the species in bold letters represent a variety of (almost) identical sequences.

Species	Accession	Protein description	Residues	Sequence-similarity
Gallus gallus	CAA32663	Spectrin alpha-chain, brain	1843-1948	
Homo sapiens	CAH71404	Spectrin alpha-chain, non-erythroid	1871-1976	97 %
Mus musculus	CAM46235	Spectrin alpha-chain, brain	1851-1956	97 %
Rattus norvegicus	ACV87913	Spectrin alpha-chain, non-erythroid	1871-1976	97 %
Bos taurus	DAA24188	Spectrin alpha-chain, brain	1866-1971	97 %
Xenopus tropicalis	AAI27323	Spectrin alpha-chain, non-erythroid	1865-1970	100 %
Danio rerio	ABN47004	Spectrin alpha-chain, brain	1874-1979	94 %
Laccophilus pictus	ACE82622	Spectrin alpha-chain	33-138	77 %
Copelatus distinctus	ACE82621	Spectrin alpha-chain	72-177	75 %
Asaphidion yukonense	ACE82615	Spectrin alpha-chain	33-138	76 %
Bembidion rufotinctum	ACE82609	Spectrin alpha-chain	29-134	77 %
B. integrum	ACE82604	Spectrin alpha-chain	28-133	77 %
B. chalceum	ACE82606	Spectrin alpha-chain	33-138	77 %
B. inaequale	ACE82607	Spectrin alpha-chain	33-138	77 %
B. transversale	ACE82612	Spectrin alpha-chain	72-177	77 %
B. rapidum	ACE82601	Spectrin alpha-chain	37-142	77 %
B. concolor	ACE82611	Spectrin alpha-chain	25-130	77 %
B. mandibulare	ACE82614	Spectrin alpha-chain	31-136	76 %
B. umbratum	ACE82602	Spectrin alpha-chain	44-149	76 %
B. plagiatum	ACE82613	Spectrin alpha-chain	37-142	76 %
Pterostichus melanarius	ACE82617	Spectrin alpha-chain	44-149	77 %
Sirdenus grayii	ACE82616	Spectrin alpha-chain	27-132	75 %
Metrius contractus	ACE82618	Spectrin alpha-chain	38-143	75 %
Macrogyrus oblongus	ACE82620	Spectrin alpha-chain	33-138	73 %
Dineutes sublineatus	ACE82619	Spectrin alpha-chain	33-138	73 %
Tribolium castaneum	EEZ99233	Spectrin alpha-chain	1814-1919	75 %
Priacma serrata	ACE82598	Spectrin alpha-chain	44-149	74 %
Chauliagnathus opacus	ACE82599	Spectrin alpha-chain	44-149	75 %
Dynastes granti	ACE82600	Spectrin alpha-chain	72-177	73 %
Sialis sp.	ACE82597	Spectrin alpha-chain	33-138	75 %
Acromyrmex echinatior	EGI62932	Spectrin alpha-chain	1832-1937	72 %
Camponotus floridanus	EFN61994	Spectrin alpha-chain	1817-1922	71 %
Harpegnathos saltator	EFN85586	Spectrin alpha-chain	1832-1937	71 %
Drosophila melanogaster	AAF47569	Spectrin alpha-chain	1814-1919	75 %
Aedes aegypti	EAT32709	Spectrin	1813-1918	75 %
Culex quinquefasciatus	EDS41171	Spectrin alpha-chain	1813-1916	71 %
Caenorhabditis elegans	AAB53876	Spectrin	1826-1931	66 %
Ascaris suum	ADY39864	Spectrin alpha-chain	1821-1926	69 %
Brugia malayi	EDP34100	Spectrin alpha-chain	1823-1928	66 %
Mus musculus	AAC61874	Spectrin alpha-chain, erythroid	1813-1919	70 %
Rattus norvegicus	AAQ02378	Spectrin alpha-chain, erythroid	1814-1920	70 %
Homo sapiens	AAA60577	Spectrin alpha-chain, erythroid	1814-1922	67 %
Bos taurus	DAA31904	Spectrin alpha-chain, erythroid	1075-1184	66 %
Salpingoeca sp.	EGD77905	Spectrin	1860-1965	61 %
Capsaspora owczarzaki	EFW39917	Spectrin alpha-chain	1562-1667	65 %

The tryptophan (position 20) in helix A is 100 % conserved in all 32 species, as is the histidine (position 57) in helix B. The moderately conserved tryptophan (position 93) in helix C appears to be substituted by aliphatic residues in most R17-like spectrin repeats. Only the mammalian erythroid spectrin repeats contain the hydrophilic threonine instead of an all-aliphatic residue. The dominating substituent in vertebrate brain spectrins is valine, while leucine is pre-dominantly found in insects.

Sequences that are most closely related to chicken brain repeat 17 are vertebrate brain spectrins, especially mammalian (figure 4.4). More distantly related are spectrin repeats of insects and nematodes. R17-like repeats seem to be present in both brain and erythroid spectrins, but the latter show clear sequence differences in the BC-loop and helix C. The most diverging sequences belong to eukaryotic single-cell organisms.

Interestingly, a majority of repeats, except for those where the accession number starts with "ACE", is situated at approximately the same position in the spectrin chain as R17 (table 4.1). In addition, when the Blast search was performed on a sequence that also included R16 and R18 (data not shown), the same score pattern was observed as with R17 alone. This indicates that not only the features of R17 are conserved in other species but also the positioning between two repeats resembling R16 and R18.

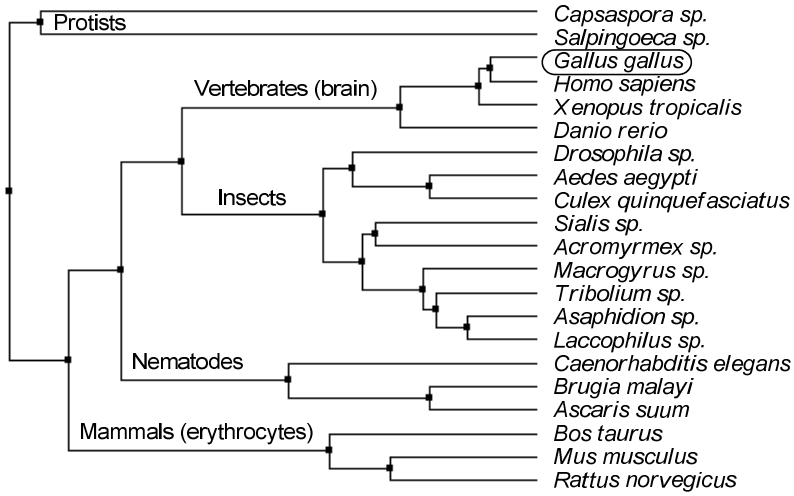


Figure 4.4: Phylogenetic tree that shows the relationship between chicken brain α-spectrin repeat 17 (highlighted) and 20 closely related spectrin repeats.

4.2 Human hNaa50p

4.2.1 Assignment challenges

The hNaa50p construct with its 173 amino acids and a molecular weight of more than 19 kDa is a protein that is classified as medium-sized to large in NMR terms. α-helices and β-strands are evenly distributed in hNaa50p as each type of structure makes up approximately one third of the protein. The 2D ^{1}H-^{15}N HSQC spectrum of hNaa50p (figure 4.5A) shows a greater diversity in H^N chemical shifts than the spectrum of R17 (figure 4.1A). This is because backbone proton resonances above 9 ppm and below 7 ppm almost exclusively are found in β-strands.

Even though the HSQC spectrum of hNaa50p shows good signal dispersion, the line-widths are broadened due to the protein size and lead to signal overlap. This peak broadening also reduces the sensitivity of the experiment by decreasing the peak heights. Panels B and C of figure 4.5 show how the resolution of the spectrum can be improved by an increment of the acquired points, and how this could resolve a cluster in the middle of the spectrum into four distinct peaks.

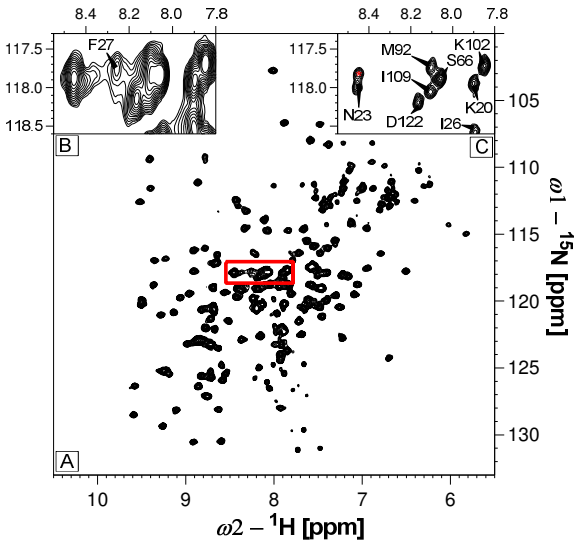


Figure 4.5 A) ^{1}H-^{15}N HSQC spectrum of hNaa50p at 600 MHz. **B)** Extraction of some overlapping peaks (marked red in the original spectrum). **C)** Enlargement of the same region in an HSQC spectrum, where 1024 data points instead of 128 were sampled in the vertical dimension. The broad peak on the left is resolved into two distinct peaks, whereas the cluster in the middle is resolved into four distinct peaks. Assignments of known resonances are included.

This approach is possible in a 2D spectrum, but increasing the amount of data points in a similar way in 3D experiments would be far too time-consuming. The attempt to resolve the HSQC spectrum better and achieve an improved signal-to-noise ratio at the same time by dividing hNaa50p into two domains of approximately 12 and 7.5 kDa, and then use the two parts as a starting point for the assignment of the full-length protein, failed. The reason for that was that the domains showed very low solubility and could no longer bind acetyl-CoA which is important for the stability of highly purified hNaa50p *in vitro*. Thus, high-resolution NMR spectroscopy on full-length protein, where the basic 3D experiments were repeated, was used in order to obtain a better signal separation. Figure 4.6 shows how the resolution improves as a result of increasing the field strength from 600 MHz to 800 MHz for the residues S66, M92 and I109 (figure 4.5C) in the CBCA(CO)NH spectrum. At 600 MHz, there are between six and eight peaks that all fit to the amide frequencies of these amino acids. At 800 MHz, there are still six peaks but they can now be distinguished into three spin pairs. Thus, by increasing the field strength it was possible to distinguish M92 from S66 and I109.

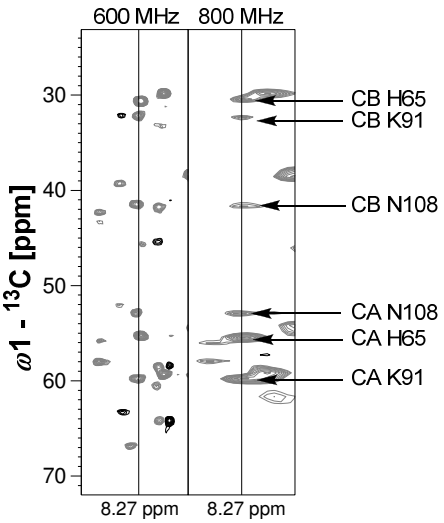


Figure 4.6: CBCA(CO)NH strips at the NH-frequencies of M92. At 600 MHz (left strip), there are eight peaks that could correspond to CA-1 and CB-1 of S66, M92 and I109. At 800 MHz (right strip), there are six signals which can be distinguished into three spin pairs.

Nevertheless, the backbone assignment was mostly based on 3D CC(CO)NH and 2D MUSIC experiments at 600 MHz. CC(CO)NH is sub-optimal for a protein with a

molecular mass of almost 20 kDa. However, the use of a long isotropic mixing time made it possible to detect most of the C^{δ} of arginines and C^{ε} of lysines. In addition, some of the aliphatic side chains, especially those of leucine, could be identified. The MUSIC experiments which detect asparagine, aspartic acid, glutamine, glutamic acid, valine, isoleucine, phenylalanine, tyrosine and histidine proved to work quite well also on hNaa50p. The advantage of MUSIC experiments in overlapping regions is that they reduce the number of possible chemical shift combinations. For instance, if the successor of N108, which in hNaa50p is I109, in the strips of figure 4.6 is to be assigned, the peaks at approximately 30, 33 and 60 ppm are very unlikely for a asparagine residue according to chemical shift statistics. Thus, MUSIC spectra which show a peak in an overlap region simplify the assignment of this particular peak and, in addition, narrow down the possible identities of the remaining unassigned peaks positioned in the same amide region.

4.2.2 Protein stability

Another challenge when studying hNaa50p was its lack of stability in the absence of the first substrate, acetyl-CoA. hNaa50p alone or in association with CoA appeared to lose its structural integrity within few days which was observed as a drastic decrease in the peak intensities of the H^N signals in a basic 1D ^1H experiment. Moreover, hNaa50p without acetyl-CoA precipitated rapidly in concentrations above 100-150 μM. Therefore, most of the NMR experiments were conducted on hNaa50p in the presence of excessive concentrations of acetyl-CoA. The HSQC spectra of free hNaa50p and in complex with CoA were assigned using the HNCO experiment, since this technique is the most sensitive 3D method, and thus also suitable for low sample concentrations. Furthermore, chemical shift changes of up to 0.3 ppm due to changes in the chemical environment are less prominent for the backbone carbons than for H^N. This is because the chemical shift range of C´ is rather large and the resolution of 3D spectra is reduced due to few sampled points, resulting in an uncertainty in the backbone carbon chemical shifts which often is in the same range as the chemical shift changes due to ligand binding. Hence, most of the peaks in the HSQC spectrum could be assigned in samples without acetyl-CoA, even in cases where the NH-shift changes were significant.

Even though hNaa50p in complex with acetyl-CoA precipitated slowly and showed little decrease in the H^N signal strengths over time, spectral changes occurred within the first couple of weeks after protein purification which complicated the assignment work. A first approach of restraining these changes was to lower the acquisition temperature from 25 °C to 15 °C in order to slow down the rate of spectral repositioning. As expected, the lowered temperature kept hNaa50p stable in its initial state for a prolonged time. However, due to increased correlation times at 15 °C compared to 25 °C the line-widths in the resulting spectra increased, making an assignment impossible.

We observed that hNaa50p reached a new state after approximately two weeks, in which the protein was structurally stable for up to six months. The initial backbone assignment was thus carried out on a protein that had been stored for three months. The sample temperature was kept at 37 °C in order to decrease the line-widths as much as possible. The C´, C^{α} and C^{β} chemical shifts – which are less influenced by temperature changes than the H^N shifts [115] – of the spectra recorded on recently purified hNaa50p at 25 °C were then used to identify the amino acids that changed spectral positions upon temperature increment and storage. Figure 4.7 shows parts of two ^{1}H-^{15}N HSQC spectra that contain the residues G87 and G89 whose NH-correlations are among those that change the most over time.

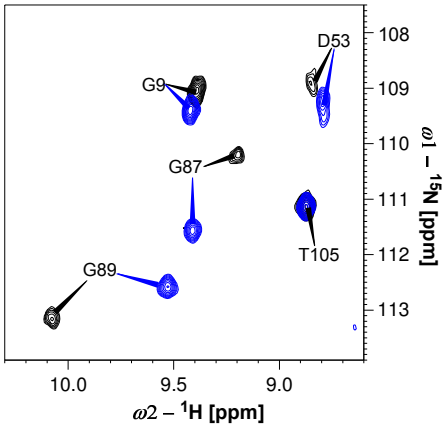


Figure 4.7: Extraction of the 2D ^{1}H-^{15}N HSQC spectrum of hNaa50p at 600 MHz. G87 and G89 are among the residues that show the greatest chemical shift differences upon storage. The spectrum in black is acquired at 25 °C on a freshly purified sample, whereas the spectrum in blue is acquired at 37 °C after three months of storage at 4 °C.

The structural nature of these spectral changes is not known. But since they include amino acids that are part of the substrate binding pocket, for instance L77 (not shown), G87 and G89, it is likely that they are correlated with an altered conformation equivalent with some kind of breakdown of the protein structure which might result in loss, or reduction, of enzyme activity.

4.2.3 Enzyme mechanism

Acetyl-CoA (figure 4.8) and a peptide of 24 residues which starts with the amino acids MLGP, the sequence preference of hNaa50p, were used as substrates in the study of the enzymatic pathway. Whereas CoA, desulfo-CoA – a structural dead-end analogue of CoA – and acetylated peptide served as inhibitors. The intersecting line pattern of the Lineweaver-Burk plot obtained from the initial velocity kinetic experiment ruled out Ping-Pong as likely mechanism. CoA and desulfo-CoA both showed to be competitive inhibitors of acetyl-CoA, whereas CoA expressed a non-competitive inhibition pattern in set-ups in which increasing amounts of peptide were used as the variable substrate. These results are in agreement with a ternary-complex formation mechanism, most likely of the Theorell-Chance type, a rare form of the ordered sequential mechanism. However, the acetylated peptide did not act as inhibitor towards any of the substrates, even at high concentrations, which indicates that the affinity between the enzyme and the second product is very low.

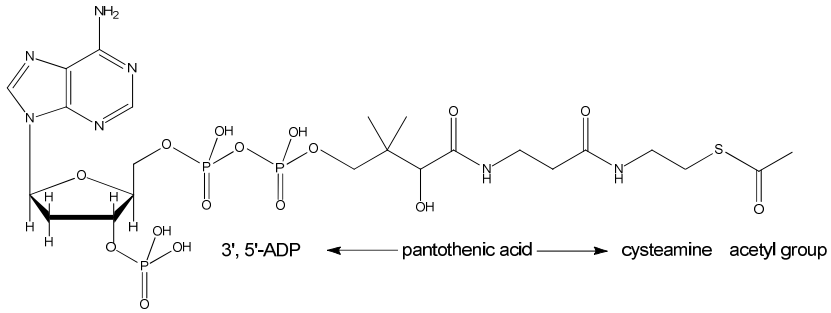


Figure 4.8: The structure of acetyl-CoA including the names of its four building blocks. CoA lacks the terminal acetyl group, whereas desulfo-CoA additionally lacks the sulph-hydryl group.

The results from the substrate binding order and product inhibition study by NMR spectroscopy on freshly purified hNaa50p could not with absolute certainty confirm that the enzyme mechanism is of the Theorell-Chance type. However, the NMR results

indicate that the enzyme has to bind acetyl-CoA prior to the peptide, and thus follows an ordered mechanism. CoA binds to the same site as acetyl-CoA, whereas the addition of acetylated peptide did not alter the spectrum. This is in agreement with the observations that CoA acts as competitive inhibitor while the acetylated peptide has no effect on the enzyme reaction. Hence, the results from NMR spectroscopy lead to the conclusion that the enzyme mechanism of hNaa50p follows an ordered sequential pathway. In combination with the results from the kinetic assays, the Theorell-Chance mechanism is most likely.

The NMR study did not only yield information about the type of enzyme mechanism, but also revealed which amino acids are most likely involved in the binding of the substrates and inhibitors to the enzyme. Most of the residues in the HMQC spectrum were not affected at all or showed only small changes in their amide shifts upon binding of acetyl-CoA. This indicates that acetyl-CoA binds specifically to the protein, affecting only residues at the binding site. The following residues showed the most prominent changes in their chemical shifts: V29, T76, L77, S116, Y124 and I142. L77 presented the largest difference with a 2 ppm upfield shift in N^H and a 0.8 ppm upfield shift in H^N. Together with T76, this amino acid is positioned in a β-strand close in space to the pantothenic acid moiety in acetyl-CoA. V29 and I142 belong to two opposing loops which form part of the substrate binding pocket. The fact that these amino acids change their amide shifts might be the result of an "induced fit" response upon substrate binding. S116 is part of a loop close to the acetyl group and might undergo a change in conformation between the states of bound and unbound peptide. Finally, the distance between Y124 to acetyl-CoA is too large to facilitate the establishment of a direct interaction to the substrate. However, we believe that a water molecule could mediate between Y124 and acetyl-CoA, and between the former and L77. Also, Y124 has shown to be important for both the N^{α}- and N^{ε}-acetylation activity of hNaa50p [47].

The binding of CoA as a product inhibitor led to a similar change in the amide shifts, but there were some differences, e.g. the chemical shift changes of L77 and I142 were not identical in the two spectra. Furthermore, V29 could not be assigned at all in spectra without acetyl-CoA. We assume that this is a result of a less tight binding between enzyme and CoA compared to acetyl-CoA, and that CoA as inhibitor

facilitates the binding of the second substrate to a lesser degree than acetyl-CoA does. The latter assumption is strengthened by the results from the simultaneous binding of both substrates to the enzymatically inactive H112A mutant, where the spectrum showed significantly more changes compared to when peptide was added to the wild-type enzyme in complex with CoA.

Recent results suggest that the amino acid most likely involved in the enzyme catalysis, besides H112, is Y73 [99]. Interestingly, the NH-resonances of the latter residue were altered upon addition of CoA rather than acetyl-CoA. Furthermore, this resonance shifted both upon addition of peptide to the hNaa50p-CoA complex and the mutant-acetyl-CoA complex. The fact that Y110 in the vicinity of H112 – which could not be assigned in any spectrum – showed similar effects upon addition of substrates and inhibitors as Y73 strengthens the hypothesis that Y73 and H112 plus its surroundings are important for the enzymatic activity.

4.2.4 Protein structure and dynamics

The CSI of three different nuclei (C´, C^{α} and C^{β}) showed that the positions of the secondary structural elements of hNaa50p in presence of acetyl-CoA are in overall agreement with the reported structural parts in the crystal (PDB-entry 2OB0). However, the second helix (residues 32-42) appears to be shorter in the protein in solution, as it seems to comprise only residues 32-38. This could be a coincidence, for instance caused by aromatic ring current effects from two aromatic residues (F35, Y36) in the vicinity of V39 and L40, which do not have chemical shifts corresponding to an α-helix. But it might also indicate functional importance which is supported by the observation that the order parameters, S^2, implied that the protein part that contains the first two helices (residues 20-40) is less rigid than expected for a structured region. Even though the helices are not directly involved in the substrate binding, they could be flexible since they establish little contact to other secondary structural elements. Thus, this part of the protein might function as a lever that facilitates optimal catalytic activity and specificity.

Furthermore, single residues in this second α-helix of NATs might be essential for their function. A recent study showed that in a rare, X-linked recessive inherited disease, a conserved serine in the second α-helix of hNaa10p, a protein distantly

related to hNaa50p (expect-value: 1×10^{-6}) and part of both the NatA and NatE complex, is mutated to proline, leading to premature helix termination [116]. Boys who inherit the mutated protein die from cardiac arrhythmias in infancy. The results from the enzymatic assays, where wild-type hNaa10p and the hNaa10p S37P mutant were investigated, showed that the latter exhibits up to 80 % reduced acetyltransferase activity towards typical peptide substrates. The lethality of the mutant is most likely caused by its incapability of acetylating proteins that depend on this modification for ultimate function. Another, or additional, reason for the lethality could be that the mutant loses its putative N^{ε}-acetylation function [116]. Analogous to hNaa50p, where the enzyme activity and substrate specificity towards the peptide is increased by auto acetylation of three lysine side chains, whereas conservative mutations of K34 and K37 into arginines resulted in decreased N^{α}-acetyltransferase activity and altered specificity of both N^{α}- and N^{ε}-acetylation [47], the loss of hNaa10p's N^{ε}-acetylation function could contribute to its reduced N^{α}-acetyltransferase activity. Thus, the features of the second α-helix in NATs might be as crucial for the activity and specificity of the enzymes as the residues at the active site.

Two other interesting stretches of the protein are ^{73}YIMTLG^{78} and ^{87}GIGTKML^{93}. The former is a β-strand, whereas the latter is part of an α-helix according to the crystal structure and the chemical shift indices. However, the dynamic data indicated that these stretches are flexible, which is especially unexpected for the helical part which belongs to a long α-helix comprising almost 20 residues (position 87-104). Both segments are part of the binding pocket of hNaa50p and the structural flexibility might be a requirement for the conformational change upon binding of the second substrate. The region surrounding I142, on the other hand, appears to be more rigid from the dynamic data and the CSI than may be inferred from the crystal structure. I142 seems to be involved in an induced fit response upon acetyl-CoA binding, which might explain why this region adopts a more stable conformation than expected.

However, there is one source of error that could have had an effect on the results of the dynamic study. Since the determination of both T_1 and T_2 relaxation times demands a series of approximately ten spectra with different relaxation delays, and it also is common to run several parallels, it is essential that the sample is stable for at least 3-4 days. All spectra within a series are acquired and processed in exactly the same way

and the peaks are only picked from one of the spectra. Thus, if some of the peaks alter their amide chemical shifts during the data sampling, the resulting relaxation data would be wrong or meaningless, depending on when the changes occurred. Because freshly purified samples of hNaa50p cannot guarantee stability over several days, the dynamic data were determined from a sample that had been stored for about three weeks. We do not know which consequences the chemical shift changes that appear upon storage have for the structure and function of hNaa50p, but one possibility is that some of the residues that change the most over time might appear to be more flexible than they really are as a result of protein degradation. This might be the case for L77, G87 and G89, which could be the explanation for why the two stretches, ^{73}YIMTLG^{78} and ^{87}GIGTKML^{93}, seem to be less rigid than expected for secondary structural elements. However, ^{84}RRLGIG^{89} is identified as the conserved motif A in hNaa50p, which is responsible for the binding of acetyl-CoA [117, 118], and ^{84}RRL^{86} proved to be difficult to assign also in freshly purified samples, probably because of conformational exchange. In addition, NMR spectroscopy of recently purified enzyme indicated that T76 and L77 which are part of the segment ^{73}YIMTLG^{78} are involved in the substrate binding, whereas Y73 is expected to be directly involved in the enzyme catalysis. Thus, it would be reasonable for these two parts of the sequence to be more flexible than other secondary structural elements.

Unfortunately, we can only speculate whether the flexibility of the stretches ^{73}YIMTLG^{78} and ^{87}GIGTKML^{93} and the chemical shift changes in the loop ^{112}HVQISN^{117} upon addition of acetyl-CoA are a result of a change in the conformation between two states: the one of bound and the one of unbound peptide. A possibility of confirming the assumption would be to investigate if free enzyme or hNaa50p in complex with CoA shows the same flexibility. However, the protein in absence of acetyl-CoA is not stable long enough to perform dynamic measurements. Another possibility would be to examine the effects of adding both substrates to the wild-type enzyme, but the rate of the enzyme reaction is too fast for the time-scale of NMR analyses. Still, the spectral changes upon addition of both substrates to the hNaa50p H112A mutant indicate that the flexibility might be essential for the binding of the second substrate.

4.2.5 Selenium NMR spectroscopy

hNaa50p contains two selenomethionines (section 2.8), SeM75 and SeM92 (PDB-entry 2OB0). Interestingly, both of these residues are positioned in segments that seem to be involved in substrate binding: SeM75 is the sequential neighbour of T76 and L77 which showed significant chemical shift changes upon binding of acetyl-CoA (section 4.2.3), whereas SeM92 is part of the flexible stretch ^{87}GIGTKML^{93} (section 4.2.4). Since the ^{77}Se nucleus is very sensitive towards changes in its chemical environment, one might expect huge shift changes in the selenium frequency domain upon binding of the substrates to hNaa50p, which in turn could be used to further elucidate the enzyme mechanism.

In a preliminary study, all three methionines in hNaa50p were substituted by selenomethionine during protein expression. This was done by using a growth medium that was methionine-depleted but that contained selenomethionine in addition to the other 19 amino acids [119]. Even though the protein concentration was below 200 μM and no cryogenic probe-head could be used in the experiment, the resulting HMBC-spectrum contained several peaks whose ^{1}H chemical shift values were consistent with the methionine H$^{\varepsilon}$ shifts (figure 4.9). Thus, further studies of the selenomethionines in hNaa50p in the presence of substrates and products should be undertaken in order to investigate the involvement of SeM75 and SeM92 and their surroundings in the enzyme mechanism.

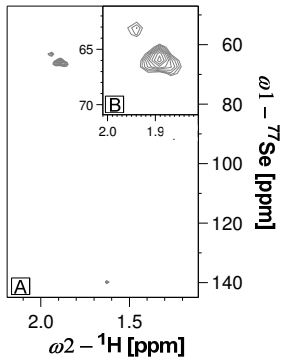


Figure 4.9 A) ^{1}H-^{77}Se HMBC spectrum of the fusion protein glutathione S-transferase-hNaa50p. The fusion protein contains seven methionine residues, of which at least four are visible in the spectrum. The peak at approximately 140 ppm might be an artefact. **B)** Enlargement of the peaks in the region of 60-70 ppm. The strong peak seems to comprise three overlapping signals, but might according to its signal-to-noise ratio contain four, or even five, signals.

5. CONCLUDING REMARKS

This last chapter sums up the main conclusions that may be drawn from the two studies and provides starting points for further investigations.

The main findings in paper I were that the linker to R16 seems to contribute to an increased thermal stability of R17, whereas the most likely reason for the domain's reduced stability compared to R16 is the tryptophan to valine substitution in the centre of helix C which weakens the interhelical interactions in the core of the triple-helix bundle. Furthermore, the NMR data could not provide any indications that point towards a difference between the crystal and the solution structure.

So far, only few spectrin repeats have been studied by NMR spectroscopy and only the repeats 16 and 17 of chicken brain α-spectrin have been investigated with respect to their protein dynamics. The importance of the two tryptophan residues in the domain could be further examined by studying the thermal stability of different subunits of spectrin by NMR spectroscopy, e.g. one repeat with both tryptophanes conserved, one where the first tryptophan is substituted and one where the second tryptophan is substituted by an amino acid different from valine. This kind of investigation could strengthen or weaken the hypothesis proposed in paper I that it is the tryptophan to valine substitution which leads to the break-up of the triple-helix bundle in R17. If these studies were performed on repeats belonging to the same spectrin chain (e.g. the β-chain of chicken brain spectrin provides all of the mentioned combinations), it might contribute to a deeper understanding of the differing melting points of the repeats which make up one spectrin molecule.

It would also be of interest to examine the biological significance of the varying thermodynamical stabilities of the repeats. So far, the reduced stability of the linker compared to the triple-helix bundle was suggested to be the main reason for the flexibility of the repeats [106] as it was proposed that the linker can trigger cooperative helix-to-coil unfolding [27]. The finding that it is difficult to study multiple spectrin repeats at a time by NMR spectroscopy [104] strengthens this hypothesis. However, also the unfolding of the triple-helix bundle in single domains could be of importance for the flexibility of mature spectrins because it results in an approximately four-fold

local protein extension [120]. But since the repeats seem to be stabilized by their neighbours and unfolding in tandem repeats occur cooperatively *in vitro* [30], it is uncertain how local unfolding of repeats will affect the overall stability of mature spectrin. Still, there are at least two reasons which support the assumption that reduced thermodynamical stability of single repeats is correlated with increased flexibility of spectrins *in vivo*. First, it was observed that the unfolding of the triple-helix bundle was reversible for approximately 90 % of the studied repeats [111]. And second, cooperativity in unfolding of tandem repeats is reduced by 50 % as the temperature is raised from 10 °C to 42 °C [121]; thus, local unfolding of single repeats instead of cooperative unfolding of tandem repeats seems to be more likely at physiological conditions.

Finally, it would be of a general interest to resolve the NMR structure of R17 as only two structures belonging to repeats of the spectrin family [10, 122] have been determined by NMR spectroscopy so far. A structural determination of R17 might require partial deuteration of the domain for the acquisition of the 3D ^{15}N NOESY-HSQC spectrum and a sample dissolved in 100 % D_2O for the acquisition of the 3D ^{13}C NOESY- HSQC spectrum. Furthermore, it might be necessary to use high-field (800-900 MHz) NMR instruments.

The combination of the results obtained from enzymatic assays and chemical shift changes in the NMR spectra led to the conclusion that hNaa50p follows an ordered sequential enzymatic pathway, most likely of the rare Theorell-Chance type. The results from the NMR dynamic study also suggest that the amino acids which are involved in the enzyme mechanism are more flexible than expected from the crystal structure and the NMR analysis of the secondary structures (paper II). Furthermore, the combination of the 3D CC(CO)NH spectrum with various MUSIC experiments could be used to extend the range of backbone assignment of hNaa50p at a moderate field strength (paper III).

A previous study showed that the hNaa10p and hNaa15p subunits of the human NatA complex are over-expressed in various types of cancer, and that they are important for survival and growth of cancer cell lines [123]. hNaa50p itself, which associates to hNatA [124], has not been directly linked to cancer yet, but the protein has shown to

be important for sister chromatid adhesion and chromosome resolution in several species [45, 55, 56]; thus, an inhibitor of hNaa50p might be a potential anti-cancer drug because it could prevent chromosome formation and, ultimately, cell division. After the elucidation of the enzyme mechanism of hNaa50p (paper II), the next task would be to design an anti-cancer drug which binds irreversibly to the enzyme, mimicking the ternary complex formation. According to our results, the inhibitor would have to bind to the same site as the first substrate, acetyl-CoA.

The interactions between hNaa50p and its substrates could be further investigated by NMR spectroscopy with respect to the ligands. One possibility is to determine the parts of the ligand which involve in protein binding by the use of saturation transfer difference spectroscopy. Signals belonging to ligand nuclei that interact with the protein will show a higher intensity in the spectrum which contains protein than in that without [125]. The only requirement is that the disassociation constant has to be within the range of 10^{-3} and 10^{-8} M [126]. This method could, for instance, determine if the binding of CoA and acetyl-CoA to the enzyme involves the same part of the molecules and confirm the results from the enzymatic assays where CoA expressed a lower affinity to the enzyme. It is also possible to detect NOEs between two molecules with 2D or 3D intermolecular NOESY [127-129], where one of the two molecules – usually the protein – has to be labelled with ^{13}C and/or ^{15}N. In the resulting spectrum only through-space correlations between those protons will build up, where one is directly bound to the labelled (e.g. ^{13}C) and the other to the unlabelled (^{12}C) isotope. These experiments could strengthen the results from the enzyme mechanism study, if, for instance, binding of the peptide is observed to the hNaa50p-CoA complex but not to the free enzyme.

6. REFERENCES

1. Bennett, V. and D.M. Gilligan, *The spectrin-based membrane skeleton and micron-scala organization of the plasma-membrane*. Annual Review of Cell Biology, 1993. **9**: p. 27-66.
2. Speicher, D.W. and V.T. Marchesi, *Erythrocyte spectrin is comprised of many homologous triple helical segments*. Nature, 1984. **311**(5982): p. 177-180.
3. Jefferson, J.J., et al., *Structural analysis of the plakin domain of bullous pemphigoid antigen1 (BPAG1) suggests that plakins are members of the spectrin superfamily*. Journal of Molecular Biology, 2007. **366**(1): p. 244-257.
4. Bennett, V. and A.J. Baines, *Spectrin and ankyrin-based pathways: metazoan inventions for integrating cells into tissues*. Physiological Reviews, 2001. **81**(3): p. 1353-1392.
5. Bennett, P.M., et al., *Not just a plasma membrane protein: in cardiac muscle cells alpha-II spectrin also shows a close association with myofibrils*. Journal of Muscle Research and Cell Motility, 2004. **25**(2): p. 119-126.
6. Davis, J. and V. Bennett, *Brain spectrin – isolation of subunits and formation of hybrids with erythrocyte spectrin subunits*. Journal of Biological Chemistry, 1983. **258**(12): p. 7757-7766.
7. Altmann, S.M., et al., *Pathways and intermediates in forced unfolding of spectrin repeats*. Structure, 2002. **10**(8): p. 1085-1096.
8. Pantazatos, D.P. and R.I. MacDonald, *Site-directed mutagenesis of either the highly conserved Trp-22 or the moderately conserved Trp-95 to a large, hydrophobic residue reduces the thermodynamic stability of a spectrin repeating unit*. Journal of Biological Chemistry, 1997. **272**(34): p. 21052-21059.
9. Delaunay, J., *Genetic-disorders of the red-cell membranes*. FEBS Letters, 1995. **369**(1): p. 34-37.
10. Pascual, J., et al., *Solution structure of the spectrin repeat: a left-handed antiparallel triple-helical coiled-coil*. Journal of Molecular Biology, 1997. **273**: p. 740-751.
11. Toniolo, D. and C. Minetti, *Muscular dystrophies: alterations in a limited number of cellular pathways?* Current Opinion in Genetics & Development, 1999. **9**(3): p. 275-282.
12. Sahr, K.E., et al., *The complete cDNA and polypeptide sequences of human erythroid alpha-spectrin*. Journal of Biological Chemistry, 1990. **265**(8): p. 4434-4443.
13. Winkelmann, J.C., et al., *Beta-spectrin in human skeletal-muscle – tissue-specific differential processing of 3′ beta-spectrin pre-messenger-RNA generates a beta-spectrin isoform with a unique carboxyl terminus*. Journal of Biological Chemistry, 1990. **265**(33): p. 20449-20454.
14. Speicher, D.W., L. Weglarz, and T.M. Desilva, *Properties of human red-cell spectrin heterodimer (side-to-side) assembly and identification of an essential nucleation site*. Journal of Biological Chemistry, 1992. **267**(21): p. 14775-14782.
15. DeSilva, T.M., et al., *Analysis of human red-cell spectrin tetramer (head-to-head) assembly using complementary univalent peptides*. Biochemistry, 1992. **31**(44): p. 10872-10878.
16. MacDonald, R.I. and E.V. Pozharski, *Free energies of urea and of thermal unfolding show that two tandem repeats of spectrin are thermodynamically more stable than a single repeat*. Biochemistry, 2001. **40**(13): p. 3974-3984.
17. Carugo, K.D., S. Banuelos, and M. Saraste, *Crystal structure of a calponin homology domain*. Nature Structural Biology, 1997. **4**(3): p. 175-179.

18. Luna, E.J. and A.L. Hitt, *Cytoskeleton plasma-membrane interactions*. Science, 1992. **258**(5084): p. 955-963.
19. Travé, G., et al., *Molecular mechanism of the calcium-induced conformational change in the spectrin EF-hands*. EMBO Journal, 1995. **14**(20): p. 4922-4931.
20. Korsgren, C. and S.E. Lux, *The carboxyterminal EF domain of erythroid alpha-spectrin is necessary for optimal spectrin-actin binding*. Blood, 2010. **116**(14): p. 2600-2607.
21. Musacchio, A., et al., *Crystal-structure of a src-homology-3 (SH3) domain*. Nature, 1992. **359**(6398): p. 851-855.
22. Macias, M.J., et al., *Structure of the pleckstrin homology domain from beta-spectrin*. Nature, 1994. **369**(6482): p. 675-677.
23. Kusunoki, H., et al., *Independent movement, dimerization and stability of tandem repeats of chicken brain alpha-spectrin*. Journal of Molecular Biology, 2004. **344**(2): p. 495-511.
24. Koradi, R., M. Billeter, and K. Wüthrich, *MOLMOL: a program for display and analysis of macromolecular structures*. Journal of Molecular Graphics, 1996. **14**(1): p. 51-55.
25. Pascual, J., et al., *The spectrin repeat folds into a three-helix bundle in solution*. FEBS Letters, 1996. **383**(3): p. 201-207.
26. MacDonald, R.I. and J.A. Cummings, *Stabilities of folding of clustered, two-repeat fragments of spectrin reveal a potential hinge in the human erythroid spectrin tetramer*. Proceedings of the National Academy of Sciences of the United States of America, 2004. **101**(6): p. 1502-1507.
27. Ortiz, V., et al., *Unfolding a linker between helical repeats*. Journal of Molecular Biology, 2005. **349**(3): p. 638-647.
28. Parry, D.A.D., T.W. Dixon, and C. Cohen, *Analysis of the 3-alpha-helix motif in the spectrin superfamily of proteins*. Biophysical Journal, 1992. **61**(4): p. 858-867.
29. An, X.L., et al., *Thermal stabilities of brain spectrin and the constituent repeats of subunits*. Biochemistry, 2006. **45**(45): p. 13670-13676.
30. Batey, S., et al., *Cooperative folding in a multi-domain protein*. Journal of Molecular Biology, 2005. **349**(5): p. 1045-1059.
31. Legardinier, S., et al., *A two-amino acid mutation encountered in Duchenne muscular dystrophy decreases stability of the rod domain 23 (R23) spectrin-like repeat of dystrophin*. Journal of Biological Chemistry, 2009. **284**(13): p. 8813-8823.
32. Bayer, P., et al., *Structure determination of the small ubiquitin-related modifier SUMO-1*. Journal of Molecular Biology, 1998. **280**(2): p. 275-286.
33. Walsh, C.T., S. Garneau-Tsodikova, and G.J. Gatto, *Protein posttranslational modifications: the chemistry of proteome diversifications*. Angewandte Chemie-International Edition, 2005. **44**(45): p. 7342-7372.
34. Strous, G.J.A., H. Bloemendal, and A.J.M. Berns, *N-terminal acetylation of nascent chains of alpha-crystallin*. Biochemical and Biophysical Research Communications, 1974. **58**(3): p. 876-884.
35. Arnesen, T., et al., *Identification and characterization of the human ARD1-NATH protein acetyltransferase complex*. Biochemical Journal, 2005. **386**: p. 433-443.
36. Di Gennaro, E., et al., *Acetylation of proteins as novel target for antitumor therapy: review article*. Amino Acids, 2004. **26**(4): p. 435-441.
37. Polevoda, B. and F. Sherman, *The diversity of acetylated proteins*. Genome Biology, 2002. **3**(5).
38. Hong, L., et al., *Studies of the DNA-binding properties of histone H4 amino terminus – thermal-denaturation studies reveal that acetylation markedly reduces the binding*

constant of the H4 tail to DNA. Journal of Biological Chemistry, 1993. **268**(1): p. 305-314.
39. Driessen, H.P.C., et al., *The mechanism of N-terminal acetylation of proteins*. CRC Critical Reviews in Biochemistry, 1985. **18**(4): p. 281-325.
40. Polevoda, B. and F. Sherman, *N-terminal acetyltransferases and sequence requirements for N-terminal acetylation of eukaryotic proteins*. Journal of Molecular Biology, 2003. **325**(4): p. 595-622.
41. Polevoda, B. and F. Sherman, *Composition and function of the eukaryotic N-terminal acetyltransferase subunits*. Biochemical and Biophysical Research Communications, 2003. **308**(1): p. 1-11.
42. Polevoda, B., T. Arnesen, and F. Sherman, *A synopsis of eukaryotic N-alpha-terminal acetyltransferases: nomenclature, subunits and substrates*. BMC Proceedings, 2009. **3 Suppl 6**: p. S2.
43. Van Damme, P., et al., *NatF contributes to an evolutionary shift in protein N-terminal acetylation and is important for normal chromosome segregation*. Plos Genetics, 2011. **7**(7).
44. Gautschi, M., et al., *The yeast N-alpha-acetyltransferase NatA is quantitatively anchored to the ribosome and interacts with nascent polypeptides*. Molecular and Cellular Biology, 2003. **23**(20): p. 7403-7414.
45. Hou, F.J., et al., *The acetyltransferase activity of San stabilizes the mitotic cohesin at the centromeres in a shugoshin-independent manner*. Journal of Cell Biology, 2007. **177**(4): p. 587-597.
46. Starheim, K.K., et al., *Composition and biological significance of the human N-alpha-terminal acetyltransferases*. BMC Proceedings, 2009. **3 Suppl 6**: p. S2.
47. Evjenth, R., et al., *Human Naa50p (Nat5/San) displays both protein N-alpha- and N-epsilon-acetyltransferase activity*. Journal of Biological Chemistry, 2009. **284**(45): p. 31122-31129.
48. Berndsen, C.E., et al., *Catalytic mechanism of a MYST family histone acetyltransferase*. Biochemistry, 2007. **46**(3): p. 623-629.
49. Albaugh, B.N., E.M. Kolonko, and J.M. Denu, *Kinetic mechanism of the Rtt109-Vps75 histone acetyltransferase-chaperone complex*. Biochemistry, 2010. **49**(30): p. 6375-6385.
50. Liu, X., et al., *The structural basis of protein acetylation by the p300/CBP transcriptional coactivator*. Nature, 2008. **451**(7180): p. 846-850.
51. Yan, Y., et al., *The catalytic mechanism of the ESA1 histone acetyltransferase involves a self-acetylated intermediate*. Nature Structural Biology, 2002. **9**(11): p. 862-869.
52. Thompson, P.R., et al., *Transcriptional coactivator protein p300 – kinetic characterization of its histone acetyltransferase activity*. Journal of Biological Chemistry, 2001. **276**(36): p. 33721-33729.
53. Cleland, W.W., *Kinetics of enzyme-catalyzed reactions with two or more substrates or products. 1. Nomenclature and rate equations*. Biochimica Et Biophysica Acta, 1963. **67**(1): p. 104-137.
54. Theorell, H. and B. Chance, *Studies on liver alcohal dehydrogenase 2. The kinetics of the compound of horse liver alcohol dehydrogenase and reduced diphosphopyridine nucleotide*. Acta Chemica Scandinavica, 1951. **5**(7-8): p. 1127-1144.
55. Pimenta-Marques, A., et al., *Differential requirements of a mitotic acetyltransferase in somatic and germ line cells*. Developmental Biology, 2008. **323**(2): p. 197-206.
56. Williams, B.C., et al., *Two putative acetyltransferases, San and deco, are required for establishing sister chromatid cohesion in drosophila*. Current Biology, 2003. **13**(23): p. 2025-2036.

57. Bloch, F., *Nuclear induction*. Physical Review, 1946. **70**(7-8): p. 460-474.
58. Bloch, F., W.W. Hansen, and M. Packard, *The nuclear induction experiment*. Physical Review, 1946. **70**(7-8): p. 474-485.
59. Purcell, E.M., H.C. Torrey, and R.V. Pound, *Resonance absorption by nuclear magnetic moments in a solid*. Physical Review, 1946. **69**(1-2): p. 37-38.
60. Sattler, M., J. Schleucher, and C. Griesinger, *Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients*. Progress in Nuclear Magnetic Resonance Spectroscopy, 1999. **34**(2): p. 93-158.
61. Ernst, R.R. and W.A. Anderson, *Sensitivity enhancement in magnetic resonance. 2. Investigation of intermediate passage conditions*. Review of Scientific Instruments, 1965. **36**(12): p. 1696-1706.
62. Lin, Y.X. and G. Wagner, *Efficient side-chain and backbone assignment in large proteins: application to tGCN5*. Journal of Biomolecular NMR, 1999. **15**(3): p. 227-239.
63. Rule, G.S. and T.K. Hitchens, *Fundamentals of Protein NMR Spectroscopy*. Springer. 2006.
64. Grzesiek, S. and A. Bax, *The importance of not saturating H_2O in protein NMR – application to sensitivity enhancement and NOE measurements*. Journal of the American Chemical Society, 1993. **115**(26): p. 12593-12594.
65. Davis, A.L., et al., *Experiments for recording pure-absorption heteronuclear correlation spectra using pulsed field gradients*. Journal of Magnetic Resonance, 1992. **98**(1): p. 207-216.
66. Schleucher, J., M. Sattler, and C. Griesinger, *Coherence selection by gradients without signal attenuation – application to the 3-dimensional HNCO experiment*. Angewandte Chemie-International Edition in English, 1993. **32**(10): p. 1489-1491.
67. Grzesiek, S., *Improved 3D triple-resonance NMR techniques applied to a 31-kDa protein*. Journal of Magnetic Resonance, 1992. **96**(2): p. 432-440.
68. Kay, L.E., G.Y. Xu, and T. Yamazaki, *Enhanced-sensitivity triple-resonance spectroscopy with minimal H_2O saturation*. Journal of Magnetic Resonance Series A, 1994. **109**(1): p. 129-133.
69. Kay, L.E., et al., *3-dimensional triple-resonance NMR-spectroscopy of isotopically enriched proteins*. Journal of Magnetic Resonance, 1990. **89**(3): p. 496-514.
70. Clubb, R.T., V. Thanabal, and G. Wagner, *A constant-time 3-dimensional triple-resonance pulse scheme to correlate intraresidue H-1(N), N-15, and C-13(') chemical-shifts in N-15-C-13-labeled proteins*. Journal of Magnetic Resonance, 1992. **97**(1): p. 213-217.
71. Grzesiek, S. and A. Bax, *Amino-acid type determination in the sequential assignment procedure of uniformly C-13/N-15-enriched proteins*. Journal of Biomolecular Nmr, 1993. **3**(2): p. 185-204.
72. Muhandiram, D.R. and L.E. Kay, *Gradient-enhanced triple-resonance 3-dimensional NMR experiments with improved sensitivity*. Journal of Magnetic Resonance Series B, 1994. **103**(3): p. 203-216.
73. Wittekind, M. and L. Mueller, *HNCACB, a high-sensitivity 3D NMR experiment to correlate amide-proton and nitrogen resonances with the alpha-carbon and beta-carbon resonances in proteins*. Journal of Magnetic Resonance Series B, 1993. **101**(2): p. 201-205.
74. Montelione, G.T., et al., *An efficient triple resonance experiments using C-13 isotropic mixing for determining sequence-specific resonance assignments of isotopically-enriched proteins*. Journal of the American Chemical Society, 1992. **114**(27): p. 10974-10975.

75. Grzesiek, S., J. Anglister, and A. Bax, *Correlation of backbone amide and aliphatic side-chain resonances in C-13/N-15-enriched proteins by isotropic mixing of C-13 magnetization*. Journal of Magnetic Resonance Series B, 1993. **101**(1): p. 114-119.
76. Ulrich, E.L., et al., *BioMagResBank*. Nucleic Acids Research, 2008. **36**: p. D402-D408.
77. Bax, A., G.M. Clore, and A.M. Gronenborn, *H-1-H-1 correlation via isotropic mixing of C-13 magnetization, a new 3-dimensional approach for assigning H-1 and C-13 spectra of C-13-enriched proteins*. Journal of Magnetic Resonance, 1990. **88**(2): p. 425-431.
78. Clore, G.M., et al., *Assignment of the side-chain H-1 and C-13 resonances of interleukin-1-beta using double-resonance and triple-resonance heteronuclear 3-dimensional NMR-spectroscopy*. Biochemistry, 1990. **29**(35): p. 8172-8184.
79. Kay, L.E., et al., *A gradient-enhanced HCCH TOCSY experiment for recording side-chain H-1 and C-13 correlations in H_2O samples of proteins*. Journal of Magnetic Resonance Series B, 1993. **101**(3): p. 333-337.
80. Schubert, M., et al., *MUSIC in triple-resonance experiments: amino acid type-selective H-1-N-15 correlations*. Journal of Magnetic Resonance, 1999. **141**(1): p. 34-43.
81. Schubert, M., et al., *Bridging the gap: a set of selective H-1-N-15-correlations to link sequential neighbors of prolines*. Journal of Biomolecular NMR, 2000. **17**(4): p. 331-335.
82. Schubert, M., H. Oschkinat, and P. Schmieder, *MUSIC, selective pulses, and tuned delays: amino acid type-selective H-1-N-15 correlations, II*. Journal of Magnetic Resonance, 2001. **148**(1): p. 61-72.
83. Schubert, M., H. Oschkinat, and P. Schmieder, *MUSIC and aromatic residues: amino acid type-selective H-1-N-15 correlations, III*. Journal of Magnetic Resonance, 2001. **153**(2): p. 186-192.
84. Schubert, M., H. Oschkinat, and P. Schmieder, *Amino acid type-selective backbone H-1-N-15-correlations for Arg and Lys*. Journal of Biomolecular NMR, 2001. **20**(4): p. 379-384.
85. Diercks, T., M. Coles, and H. Kessler, *An efficient strategy for assignment of cross-peaks in 3D heteronuclear NOESY experiments*. Journal of Biomolecular NMR, 1999. **15**(2): p. 177-180.
86. Kay, L.E., D.A. Torchia, and A. Bax, *Backbone dynamics of proteins as studied by N-15 inverse detected heteronuclear NMR-spectroscopy – application to staphylococcal nuclease*. Biochemistry, 1989. **28**(23): p. 8972-8979.
87. Cavanagh, J., et al., *Protein NMR spectroscopy: principles and practice, 2nd edition*. Elsevier Academic Press. 2007.
88. *Systat Software Inc*.
89. Weisemann, R., et al., *Determination of H(N), H-alpha and H(N), C´ coupling-constants in C-13, N-15-labeled proteins*. Journal of Biomolecular NMR, 1994. **4**(2): p. 231-240.
90. Wishart, D.S., et al., *H-1, C-13 and N-15 chemical-shift referencing in biomolecular NMR*. Journal of Biomolecular NMR, 1995. **6**(2): p. 135-140.
91. Wishart, D.S., B.D. Sykes, and F.M. Richards, *Relationship between nuclear-magnetic-resonance chemical-shift and protein secondary structure*. Journal of Molecular Biology, 1991. **222**(2): p. 311-333.
92. Wang, Y.J. and O. Jardetzky, *Probability-based protein secondary structure identification using combined NMR chemical-shift data*. Protein Science, 2002. **11**(4): p. 852-861.

93. Wüthrich, K., M. Billeter, and W. Braun, *Polypeptide secondary structure determination by nuclear magnetic-resonance observation of short proton proton distances*. Journal of Molecular Biology, 1984. **180**(3): p. 715-740.
94. Vuister, G.W. and A. Bax, *Quantitative J correlation – a new approach for measuring homonuclear 3-bond J(H(N)H(alpha)) coupling-constants in N-15-enriched proteins*. Journal of the American Chemical Society, 1993. **115**(17): p. 7772-7777.
95. Duddeck, H., *Se-77 nuclear-magnetic-resonance spectroscopy*. Progress in Nuclear Magnetic Resonance Spectroscopy, 1995. **27**: p. 1-323.
96. Zhang, M.J. and H.J. Vogel, *Two-dimensional NMR-studies of selenomethionyl calmodulin*. Journal of Molecular Biology, 1994. **239**(4): p. 545-554.
97. Bax, A. and M.F. Summers, *H-1 and C-13 assignments from sensitivity-enhanced detection of heteronuclear multiple-bond connectivity by 2D multiple quantum NMR*. Journal of the American Chemical Society, 1986. **108**(8): p. 2093-2094.
98. Schanda, P., E. Kupce, and B. Brutscher, *SOFAST-HMQC experiments for recording two-dimensional heteronuclear correlation spectra of proteins within a few seconds*. Journal of Biomolecular NMR, 2005. **33**(4): p. 199-211.
99. Liszczak, G., T. Arnesen, and R. Marmorstein, *Structure of a ternary Naa50p (NAT5/SAN) N-terminal acetyltransferase complex reveals the molecular basis for substrate-specific acetylation*. Journal of Biological Chemistry, 2011. **286**(42): p. 37002-37010.
100. Mirijanian, D.T., et al., *Atomistic and coarse-grained analysis of double spectrin repeat units: the molecular origins of flexibility*. Journal of Molecular Biology, 2007. **365**(2): p. 523-534.
101. Pervushin, K., et al., *Transverse relaxation-optimized spectroscopy (TROSY) for NMR studies of aromatic spin systems in C-13-labeled proteins*. Journal of the American Chemical Society, 1998. **120**(25): p. 6394-6400.
102. Yamazaki, T., J.D. Formankay, and L.E. Kay, *2-Dimensional NMR experiments for correlating C-13-beta and H-1-delta/epsilon chemical-shifts of aromatic residues in C-13-labeled proteins via scalar couplings*. Journal of the American Chemical Society, 1993. **115**(23): p. 11054-11055.
103. Güntert, P., C. Mumenthaler, and K. Wüthrich, *Torsion angle dynamics for NMR structure calculation with the new program DYANA*. Journal of Molecular Biology, 1997. **273**(1): p. 283-298.
104. Raae, A.J., *unpublished results*.
105. MacDonald, R.I., et al., *Invariant tryptophan at a shielded site promotes folding of the conformational unit of spectrin*. Proceedings of the National Academy of Sciences of the United States of America, 1994. **91**(4): p. 1299-1303.
106. Grum, V.L., et al., *Structures of two repeats of spectrin suggest models of flexibility*. Cell, 1999. **98**(4): p. 523-535.
107. de la Torre, J.G., M.L. Huertas, and B. Carrasco, *HYDRONMR: prediction of NMR relaxation of globular proteins from atomic-level structures and hydrodynamic calculations*. Journal of Magnetic Resonance, 2000. **147**(1): p. 138-146.
108. Dosset, P., et al., *Efficient analysis of macromolecular rotational diffusion from heteronuclear relaxation data*. Journal of Biomolecular NMR, 2000. **16**(1): p. 23-28.
109. Lipari, G. and A. Szabo, *Model-free approach to the interpretation of nuclear magnetic-resonance relaxation in macromolecules. 1. Theory and range of validity*. Journal of the American Chemical Society, 1982. **104**(17): p. 4546-4559.
110. Lipari, G. and A. Szabo, *Model-free approach to the interpretation of nuclear magnetic-resonance relaxation in macromolecules. 2. Analysis of experimental results*. Journal of the American Chemical Society, 1982. **104**(17): p. 4559-4570.

111. An, X.L., et al., *Conformational stabilities of the structural repeats of erythroid spectrin and their functional implications*. Journal of Biological Chemistry, 2006. **281**(15): p. 10527-10532.
112. Altschul, S.F., et al., *Gapped BLAST and PSI-BLAST: a new generation of protein database search programs*. Nucleic Acids Research, 1997. **25**(17): p. 3389-3402.
113. Thompson, J.D., D.G. Higgins, and T.J. Gibson, *Clustal-W – Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice*. Nucleic Acids Research, 1994. **22**(22): p. 4673-4680.
114. Waterhouse, A.M., et al., *Jalview Version 2 – a multiple sequence alignment editor and analysis workbench*. Bioinformatics, 2009. **25**(9): p. 1189-1191.
115. Baxter, N.J. and M.P. Williamson, *Temperature dependence of H-1 chemical shifts in proteins*. Journal of Biomolecular NMR, 1997. **9**(4): p. 359-369.
116. Rope, A.F., et al., *Using VAAST to identify an X-linked disorder resulting in lethality in male infants due to N-terminal acetyltransferase deficiency*. American Journal of Human Genetics, 2011. **89**(1): p. 28-43.
117. Clements, A., et al., *Crystal structure of the histone acetyltransferase domain of the human PCAF transcriptional regulator bound to coenzyme A*. EMBO Journal, 1999. **18**(13): p. 3521-3532.
118. Neuwald, A.F. and D. Landsman, *GCN5-related histone N-acetyltransferases belong to a diverse superfamily that includes the yeast SPT10 protein*. Trends in Biochemical Sciences, 1997. **22**(5): p. 154-155.
119. van Duyne, G.D., et al., *Atomic structures of the human immunophilin FKBP-12 complexes with FK506 and rapamycin*. Journal of Molecular Biology, 1993. **229**(1): p. 105-124.
120. Law, R., et al., *Cooperativity in forced unfolding of tandem spectrin repeats*. Biophysical Journal, 2003. **84**(1): p. 533-544.
121. Law, R., et al., *Pathway shifts and thermal softening in temperature-coupled forced unfolding of spectrin domains*. Biophysical Journal, 2003. **85**(5): p. 3286-3293.
122. Park, S., et al., *Solution structural studies on human erythrocyte alpha-spectrin tetramerization site*. Journal of Biological Chemistry, 2003. **278**(24): p. 21837-21844.
123. Gromyko, D., et al., *Depletion of the human N(alpha)-terminal acetyltransferase A induces p53-dependent apoptosis and p53-independent growth inhibition*. International Journal of Cancer, 2010. **127**(12): p. 2777-2789.
124. Arnesen, T., et al., *Cloning and characterization of hNAT5/hSAN: an evolutionarily conserved component of the NatA protein N-alpha-acetyltransferase complex*. Gene, 2006. **371**(2): p. 291-295.
125. Mayer, M. and B. Meyer, *Characterization of ligand binding by saturation transfer difference NMR spectroscopy*. Angewandte Chemie-International Edition, 1999. **38**(12): p. 1784-1788.
126. Mayer, M. and B. Meyer, *Group epitope mapping by saturation transfer difference NMR to identify segments of a ligand in direct contact with a protein receptor*. Journal of the American Chemical Society, 2001. **123**(25): p. 6108-6117.
127. Otting, G. and K. Wüthrich, *Extended heteronuclear editing of 2D H-1-NMR spectra of isotope-labeled proteins, using the X (omega-1, omega-2) double half filter*. Journal of Magnetic Resonance, 1989. **85**(3): p. 586-594.
128. Breeze, A.L., *Isotope-filtered NMR methods for the study of biomolecular structure and interactions*. Progress in Nuclear Magnetic Resonance Spectroscopy, 2000. **36**(4): p. 323-372.

129. Zwahlen, C., et al., *Methods for measurement of intermolecular NOEs by multinuclear NMR spectroscopy: application to a bacteriophage lambda N-peptide/boxB RNA complex*. Journal of the American Chemical Society, 1997. **119**(29): p. 6711-6721.
130. Atkinson, R.A. and B. Kieffer, *The role of protein motions in molecular recognition: insights from heteronuclear NMR relaxation measurements*. Progress in Nuclear Magnetic Resonance Spectroscopy, 2004. **44**(3-4): p. 141-187.
131. Morin, S., *A practical guide to protein dynamics from 15N spin relaxation in solution*. Progress in Nuclear Magnetic Resonance Spectroscopy, 2011. **59**(3): p. 245-262.
132. Palmer, A.G., *NMR characterization of the dynamics of biomacromolecules*. Chemical Reviews, 2004. **104**(8): p. 3623-3640.
133. Wishart, D.S. and B.D. Sykes, *The C-13 chemical-shift index – a simple method for the identification of protein secondary structure using C-13 chemical-shift data*. Journal of Biomolecular NMR, 1994. **4**(2): p. 171-180.

APPENDIX

A1. Coupling constants

The approximate sizes of the one-bond and the two most important two-bond coupling constants in proteins [60] are shown in figure A.1.

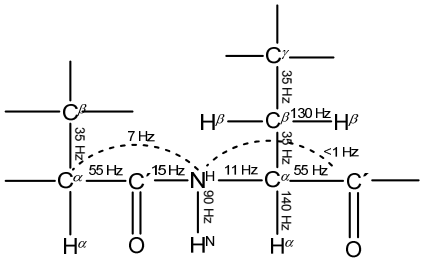


Figure A.1: Average sizes of the ^{1}J and ^{2}J coupling constants which are used in the magnetization transfer experiments. Since the ^{2}J (N^{H}-C´) coupling constant is close to zero, all experiments which include a magnetization transfer from N^{H} to C´, e.g HNCO and H(CC)(CO)NH, can only detect the preceding but not the intra-residual chemical shifts.

A2. Overview of important experiments

The nuclei that are observed in the experiments which are described in chapter 2 are listed in table A.1 [60]. The first nucleus corresponds to the one that is directly detected during data acquisition.

Table A.1: Observed correlations of important 2D and 3D experiments. "H" defines all protons that are directly coupled to ^{13}C, whereas "H^{all}" also includes protons that are bound to ^{15}N. "C" comprises all non-quaternary carbons.

Experiment	Observed nuclei
^{1}H-^{15}N HSQC	H^{N}, N^{H}
^{1}H-^{13}C HSQC	H, C
HNCO	H^{N}, N^{H}, C´-1
HN(CA)CO	H^{N}, N^{H}, C´, C´-1
HNCA	H^{N}, N^{H}, C^{α}, C^{α}-1
HN(CO)CA	H^{N}, N^{H}, C^{α}-1
CBCA(CO)NH	H^{N}, N^{H}, C^{α}-1, C^{β}-1
HBHA(CO)NH	H^{N}, N^{H}, H^{α}-1, H^{β}-1
CBCANH	H^{N}, N^{H}, C^{α}, C^{β}, C^{α}-1, C^{β}-1
CC(CO)NH	H^{N}, N^{H}, C^{ali}-1
H(CC)(CO)NH	H^{N}, N^{H}, H^{ali}-1
HCCH-TOCSY	H, C, H
^{15}N NOESY-HSQC	H^{N}, N^{H}, H^{all}
^{13}C NOESY-HSQC	H, C, H^{all}
CN-NOESY	H^{N}, N^{H}, C

The relative sensitivities of the most basic backbone and side chain experiments – HNCO, HN(CA)CO, HNCA, HN(CO)CA, CBCA(CO)NH and HNCACB – are shown in figure A.2 [60]. The peak intensities are compared to the C´-1 signal of the HNCO experiment, which is the most sensitive triple-resonance 3D technique.

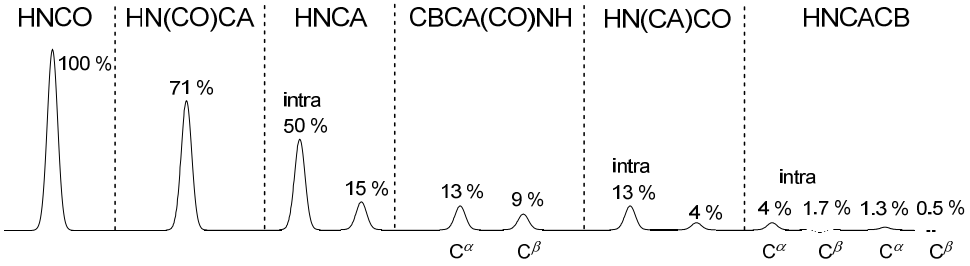


Figure A.2: Relative peak intensity of the six most basic protein triple-resonance 3D experiments. The experiments are listed according to decreasing sensitivity. The peak intensities are relative to the peak intensity of the HNCO experiment. Other effects, like relaxation, that can diminish the sensitivity are not taken into account. Note that the intra-residual peaks in general are stronger than the sequential peaks in spectra that include resonances of both residues I and I-1, and that the C^{α} peaks in general have a higher signal intensity than the C^{β} peaks.

A3. Relaxation and protein dynamics

The equations used for the calculation of T_1 (A.1), T_2 (A.2) and het*NOE* (A.3) vs. τ_c in figure 2.14.

$$T_1 = \left\{ \frac{\tau_c}{2}\left(\frac{\mu_0 \gamma_H \gamma_N \hbar}{4\pi r_{NH}^{\ 3}}\right)^2 \left[\frac{3}{1+\omega_N^{\ 2}\tau_c^{\ 2}} + \frac{1}{1+(\omega_H-\omega_N)^2\tau_c^{\ 2}} + \frac{6}{1+(\omega_H+\omega_N)^2\tau_c^{\ 2}} \right] + \frac{(\Delta\sigma\omega_N)^2 2\tau_c}{3+3\omega_N^{\ 2}\tau_c^{\ 2}} \right\}^{-1} \quad \text{(A.1)}$$

$$T_2 = \left\{ \begin{aligned} &\frac{\tau_c}{4}\left(\frac{\mu_0 \gamma_H \gamma_N \hbar}{4\pi r_{NH}^{\ 3}}\right)^2 \left[4 + \frac{3}{1+\omega_N^{\ 2}\tau_c^{\ 2}} + \frac{1}{1+(\omega_H-\omega_N)^2\tau_c^{\ 2}} + \frac{6}{1+\omega_H^{\ 2}\tau_c^{\ 2}} + \frac{6}{1+(\omega_H+\omega_N)^2\tau_c^{\ 2}} \right] \\ &+ \frac{(\Delta\sigma\omega_N)^2\tau_c}{9}\left(4 + \frac{3}{1+\omega_N^{\ 2}\tau_c^{\ 2}}\right) \end{aligned} \right\}^{-1} \quad \text{(A.2)}$$

$$\text{het}NOE = 1 + \frac{\tau_c T_1 \gamma_H}{2\gamma_N}\left(\frac{\mu_0 \gamma_H \gamma_N \hbar}{4\pi r_{NH}^{\ 3}}\right)^2 \left[\frac{6}{1+(\omega_H+\omega_N)^2\tau_c^{\ 2}} - \frac{1}{1+(\omega_H-\omega_N)^2\tau_c^{\ 2}} \right] \quad \text{(A.3)}$$

$\mu_0 = 4\pi 10^{-7}\,\text{kgms}^{-2}\text{A}^{-2}$ is the vacuum permeability, γ_H and γ_N are the gyromagnetic ratios of ^{1}H and ^{15}N, respectively, $\hbar$ is the Planck's constant divided by 2 π, r_{NH} = 1.02 Å is the average ^{15}N-^{1}H inter-nuclear distance, ω_H and ω_N are the ^{1}H and ^{15}N resonance frequencies (in rad/s), respectively, and $\Delta\sigma$ = -172 ppm is the ^{15}N chemical shift anisotropy.

The time-scales of both important molecular motions and NMR parameters [130-132] are shown in figure A.3.

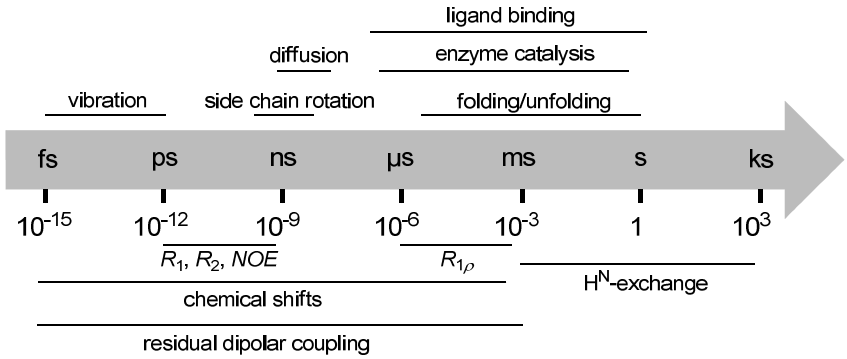


Figure A.3: The time-scales of protein dynamics and those probed by NMR experiments are shown above and below the arrow, respectively. The transition between fast and slow times is approximately at 1 μs. Note that the relaxation rates (R_1, R_2 and $R_{1\rho}$) correspond to the inverse of the relaxation times.

A4. Chemical shift indices

The ^{13}C chemical shifts for the secondary structure determination [133] and the average H^{α} chemical shifts [91] are listed in table A.2.

Table A.2: Average random coil chemical shifts of the four most relevant resonances for the determination of the secondary structural elements in proteins.

Residue	C^{α} [ppm]	C^{β} [ppm]	C´ [ppm]	H^{α} [ppm]
Alanine	52.5	19.0	177.1	4.20
Arginine	56.3	30.3	176.5	4.24
Asparagine	53.6	39.0	175.1	4.68
Aspartate	54.1	40.8	177.2	4.62
Cysteine (ox.)	58.0	41.8	175.1	4.74
Cysteine (red.)	58.3	28.6	174.8	4.74
Glutamate	56.7	29.7	176.1	4.22
Glutamine	56.2	30.1	176.3	4.31
Glycine	45.0	—	173.6	4.14 / 3.64
Histidine	55.8	32.0	175.1	4.57
Isoleucine	62.6	37.5	176.9	4.14
Leucine	55.7	41.9	177.1	4.29
Lysine	56.7	32.3	176.5	4.23
Methionine	56.6	32.8	175.8	4.32
Phenylalanine	57.9	39.3	175.8	4.58
Proline	62.9	31.7	176.0	4.41
Serine	58.3	62.7	173.7	4.52
Threonine	63.1	68.1	175.2	4.50
Tryptophan	57.8	28.3	175.8	4.63
Tyrosine	58.6	38.7	175.7	4.62
Valine	63.0	31.7	177.1	4.13

A5. NOE restraints

Figure A.4 shows typical backbone-backbone NOE-connectivities that are observed in different types of secondary structure [93]. Sequential and medium-range NOEs to residues in the former turn (I-3, I-4) dominate in α-helices (panel A), whereas long-range NOEs are pre-dominantly found between sets of strands (panels B and C).

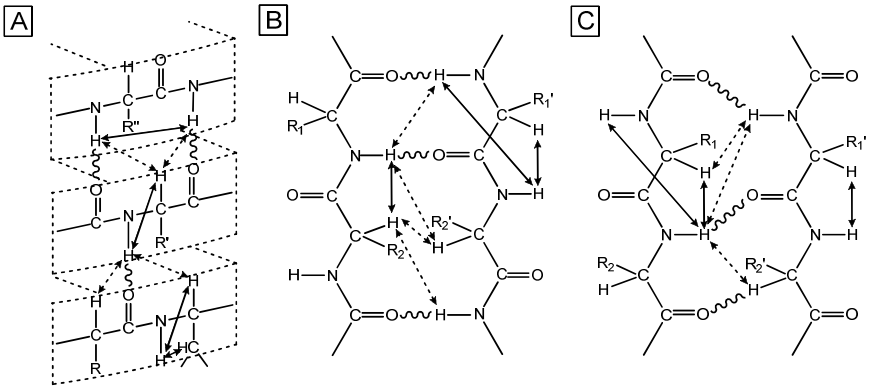


Figure A.4: NOE-connectivities in the three most common types of secondary structure. **A)** α-helix, **B)** antiparallel β-sheet and **C)** parallel β-sheet. Sequential NOEs are indicated by solid arrows, medium-range NOEs (helix) and long-range NOEs (sheet) by dashed arrows, and hydrogen bonds by wavy lines. Note that backbone-side chain and side chain-side chain NOE-connectivities which are present in all types of secondary structures are not shown.

The typical signal intensities of short- (I, I-1), medium- (I, I-2; I, I-3; I, I-4) and long-range (I, J) backbone-backbone NOE connectivities which are observed in different types of secondary structural elements are shown in table A.3 [93]. The H^N-H^N I, I-1 distance is smaller, and thus the signal is stronger, in helices than in sheets (figure A.4). The shortest connectivities in sheets and helices are the H^{α}-H^N I, I-1 and H^{β}-H^N I, I-1 distances, respectively. The figure also shows that medium-range NOEs are only found in helices, whereas long-range backbone connectivities are typical of sheets.

Table A.3: NOE signal intensities observed in the most common types of secondary structure. "S" denotes a very strong signal, "s" a strong signal, "m" a medium strong signal and "w" a weak signal.

NOE-correlation	α-helix	Antiparallel β-sheet	Parallel β-sheet
H^N-H^N I, I-1	m-s	w	w
H^N-H^N I, I-2	w		
H^α-H^N I, I-1	w-m	S	S
H^α-H^N I, I-3	w-m		
H^α-H^N I, I-4	w		
H^β-H^N I, I-1	m-s	w	w
H^α-H^β I, I-3	w-m		
H^N-H^N I, J		w	s
H^α-H^N I, J		w	w-m
H^α-H^α I, J		w-m	w-m