Solution NMR-based characterization of the structure of the outer mitochondrial membrane protein Tom40 and a novel method for NMR resonance assignment of large intrinsically disordered proteins

Dissertation

for the award of the degree

"Doctor rerum naturalium"(Dr. rer. Nat)

Division of Mathematics and Natural Sciences

of the Georg-August-Universität Göttingen

Submitted by

Xuejun Yao

from Shanghai, China

Göttingen, 2013

Thesis Committee

Prof. Dr. Markus Zweckstetter, NMR based Structural Biology, Max Planck Institute for Biophysical Chemistry

 (Name, Department/Group, Institution)

Prof. Dr. Kai Tittmann, Department of Bioanalytics, Georg-August-Universität Göttingen

(Name, Department/Group, Institution)

Prof. Dr. Holger Stark, 3D Electron Cryo-Microscopy, Max Planck Institute for Biophysical Chemistry

(Name, Department/Group, Institution)

Members of the Examination Board

Referee: Prof. Dr. Markus Zweckstetter, NMR based Structural Biology, Max Planck Institute for Biophysical Chemistry

(Name, Department/Group, Institution)

2nd Referee: Prof. Dr. Kai Tittmann, Department of Bioanalytics, Georg-August-Universität Göttingen

(Name, Department/Group, Institution)

 (if applicable) 3rd referee: ....

(Name, Department/Group, Institution)

Further members of the Examination Board

Prof. Dr. Holger Stark, 3D Electron Cryo-Microscopy, Max Planck Institute for Biophysical Chemistry

 (Name, Department/Group, Institution)

Prof. Bert de Groot, Computational biomolecular dynamics, Max Planck Institute for Biophysical Chemistry

(Name, Department/Group, Institution)

Dr. Adam Lange, NMR based Structural Biology, Max Planck Institute for Biophysical Chemistry

 (Name, Department/Group, Institution)

Dr. Lars Kuhn, NMR spectroscopy, European Neuroscience Institute

(Name, Department/Group, Institution)

Date of oral examination: 23rd Oct. 2013

Affidavit

I hereby declare that this thesis has been written independently and with no other sources and aids than quoted.

…………………………………
Xuejun Yao

Abstract

This dissertation is composed of two parts. In the first part, the characterization of the structure of liposome-embedded Tom40 using a novel hydrogen/deuterium (H/D) exchange method is described. In the second part, a new high-dimensional NMR experiment for the automatic backbone resonance assignment of intrinsically disordered proteins is described.

The Tom40 protein is the central component of the translocase of the outer mitochondrial membrane (TOM) complex. Tom40 forms the conductive channel for translocation of proteins into mitochondria and has been proposed to fold into a β-barrel. As part of my PhD thesis, I developed an H/D exchange protocol to characterize the structure of integral membrane proteins embedded into liposomes. The H/D exchange protocol consists of two steps of H/D exchange followed by NMR spectroscopic measurement of the denatured monomer. Application of the method to liposome-embedded *Neurospora crassa* Tom40 (ncTom40) showed that the N- and C-terminal tails are disordered. In addition, slow solvent exchange provided experimental support for several β-strands in the predicted barrel region and an α-helix N-terminal to the barrel. Evidence was also provided for the presence of conformational instability in the first three N-terminal β-strands, which might be involved in Tom40 oligomerization and interaction with other TOM subunits. In addition, NMR-based titration analysis of a fragment comprising the N-terminal disordered part of ncTom40 with presequence revealed multiple presequence binding sites on Tom40 that may facilitate Tom40 binding to unfolded precursor proteins transiently and subsequent transfer of the preproteins into the translocation pore. Our interaction study of Tom40 and DHPC micelle further identified several hydrophobic clusters on Tom40, which could be the initial lipid binding sites during Tom40 folding into the membrane.

In the second part of my work, I developed a high-dimensional NMR experiment named 6D HACACONCAH APSY to facilitate the assignment of intrinsically disordered proteins (IDPs). The sequence-specific assignment of IDPs is challenging mainly due to severe chemical shift degeneracy of IDPs. Moreover, conventional resonance assignment based on amide proton detection is complicated due to fast exchange of amide proton signals in HN-detected NMR spectra. Automated Projection Spectroscopy (APSY) enables the measurement of very high-dimensional NMR spectra by simultaneous chemical shift evolution of several nuclei, whereas NMR experiments based on alpha proton (HA) detection is able to avoid the line broadening due

to solvent exchange. Our new APSY experiment which combines APSY and HA-detection correlates chemical shifts from six dimensions, providing an excellent resolution to decrease NMR signal overlap. The chemical shift correlation between one C_α-H_α group and its preceding one obtained from 6D HCACONCAH APSY can be directly used as sequential connectivity for automatic assignment. Application of this novel method to two typical unfolded proteins, 140-residue α-synuclein and 352-residue Tau, resulted in the reliable assignment of more than 70% of assignable residues. The spectral measurement and assignment calculation were finished in 3 to 4 days, highly reducing the overall analysis time and tedious manual assignment procedures. Our method is ready to be used as a quick and reliable assignment protocol for IDPs even with large molecular weight.

Acknowledgements

The work in this thesis was accomplished in the department of NMR-based Structural Biology, Max Planck Institute for Biophysical Chemistry, Göttingen, under supervision of **Prof. Dr. Markus Zweckstetter**.

I express my profound gratitude to **Prof. Dr. Markus Zweckstetter** for providing this great opportunity to work on these challenging projects, for his constant support, stimulating suggestions and advices throughout my thesis.

I would like to thank **Prof. Dr. Christian Griesinger,** director of the department, for providing state-of-art equipments and excellent scientific atmosphere.

I deeply thank to **Dr. Stefan Becker** and **Yvonne Laukat** for their immense effort in preparation of numerous excellent protein samples for NMR measurement.

I extend my sincere thanks to **Dr. Ulrich Dürr** for introducing the project to me and providing help with NMR experiments and data analysis at the beginning of project. Without his pioneering work, this project couldn't be accomplished.

I am very grateful to **Dr. Zrinka Gattin and Dr. Adam Lange** for supplying solid state NMR data.

I am thankful to **Shengqi Xiang** for giving a lot of suggestions and tips on my work. I also want to thank **Raghavendran Lakshmi Narayanan** for the help with MARS software. My thanks go to **Rakhi Bajaj** and **Dr. Piotr Wysoczanski** for their time and effort reading and commenting on this manuscript.

I also want to thank to **Dr. Saskia Villinger**, **Philip Lottmann** for their kind help in spectrometer maintenance. Especially **Saskia**, she has always been patient to share with me a lot of her experience on administration of NMR spectrometer.

I thank **Dr. Dirk Bockelmann** and **Jürgen Arve** for their supports on IT systems and our secretaries **Mrs. Breiner** and **Mrs. Silberer** for helping me dealing with paper works.

My sincere thanks go to my current and former colleagues in this department for their support and friendship: **Guowei, Han, Elias, Rakhi, Hari, Sheng Qi, Guohua, Aldo, Luis, Piotr, Martin, Saskia, Hessam, Stefan, Jean-Philippe, Davood, Edward, Philip, Francesca** and other people from our department.

I owe my gratitude to **Prof. Dr. Kai Tittmann** and **Prof. Dr. Holger Stark** for their membership in my thesis committee and for valuable suggestions during my committee meetings.

I would like to also thank **Christina Bach, Christin Fischer** and other persons in the GGNB team for their kind help and excellent organization of the course and academic activities.

I sincerely express my heartfelt gratitude to my parents for their constant support, caring and consideration. My thanks also go to my sister for her support and encouragement which helped me go through some hard time in the course of this thesis.

Table of Contents

Abbreviations

1D,2D,3D, 5D, 6D, 7D	one-,two-,three-, five-,six-,seven- dimensional
APSY	Automated Projection Spectroscopy
ATP	Adenosine tri phosphate
BEST	Band-selective Excitation Short-Transient
BMRB	biological magnetic resonance bank
CD	Circular Dichroism
CMC	critical micellar concentration
CSP	Chemical shift Perturbation
DHPC	1,2-dihexanoyl-*sn*-glycero-3-phosphocholine
FTIR	Fourier transform infrared spectroscopy
GAPRO	geometric analysis of projections
GdnSCN	Guanidinium thiocyanate
GFT	G-matrix Fourier transformation
H/D	Hydrogen/Deuterium
HSQC	Heteronuclear Single Quantum Coherence Spectroscopy
IDP	Intrinsically Disordered Proteins
IPAP	In-phase-anti-phase
ncTom40	*Neurospora crassa* Tom40
NMR	Nuclear Magnetic Resonance
NOE	Nuuclear Overhause Effect
NUS	Non-uniform Sampling
PRE	Paramagnetic Relaxation Enhancement
rALDH	retinal aldehyde dehydrogenase
RDC	Residual Dipolar Coupling
SOFAST	band-Selective Optimized-Flip-Angle Short-Transient
TFE	2, 2, 2-Trifluoroethanol
TIM	Translocase of Inner mitochondrial Membrane
TOM	Translocase of Outer mitochondrial Membrane
TROSY	Transverse Relaxation Optimized Spectroscopy
VDAC	Voltage-depend anion-selective channel

Part I

Solution NMR-based characterization of the structure of the outer mitochondrial membrane protein Tom40

1   Introduction

1.1   Import of mitochondrial proteins

A vast majority of proteins in eukaryotic cells are synthesized in the cytosol. Eukaryotic cells are divided into compartment by numerous membrane-enclosed organelles. To perform their biological functions, approximately half of eukaryotic proteins are integrated into or transported across the membrane to reach their functional destination after the production in the cytosol [1, 2].

Mitochondria are membrane-bound organelles found in almost all eukaryotic cells. Mitochondria serve as the power producer of eukaryotic cells generating most supply of adenosine triphosphate (ATP), a source of chemical energy. Furthermore, mitochondria also play a crucial role in numerous cellular process including apoptosis, signaling, calcium regulation, metabolism of carbohydrate and fats, heme and steroid synthesis, and membrane remodeling [3-6]. Proteomic studies reveal mitochondria contain about 1000(yeast) to 1500(human) different proteins [7, 8]. Although mitochondria have their own genome and protein synthesis system, only very minimal mitochondrial proteins are synthesized in mitochondria. During evolution, most mitochondrial genes have been transferred to the nuclear genome and only a very small portion of genes is preserved in the mitochondrial gene. Thus mitochondria are not able to produce all of the proteins required for its function and 99% mitochondrial proteins are synthesized in the ribosome in the cytosol as precursor proteins [5]. Since

double membrane organization divides mitochondria into four subcompartments, the outer membrane, the intermembrane space (IMS), the inner membrane the innermost , The precursor proteins require import and subsequently sorting to their destined function site and this process is mediated by numerous translocase proteins [9-12].

Translocase in the Outer Membrane (TOM) complex represents the only entry pore on the outer membrane for mitochondrial precursor proteins from cytosol [13]. After crossing through the TOM complex, the various precursor proteins are directed to different destinations in the mitochondria. These distinct import pathways rely on different types of targeting sequence in the preproteins [11]. The most classic presequence pathway targets the precursor proteins which contain N-terminal presequence (matrix targeting signal) through TIM23 complex into matrix, intermembrane space and inner membrane [14, 15]. The presequence is cleaved off by matrix processing peptidase (MPP) in the matrix. In the carrier pathway, the hydrophobic precursor proteins containing internal targeting signal are transported into the inner mitochondrial membrane which is mainly mediated by TIM22 using membrane potential $\Delta\psi$ as the energy source [16]. The third one is β-barrel pathway, in which internal β-signal guide the β-barrel proteins into the outer membrane via the sorting and assembly machinery (SAM complex) [17]. The mitochondrial intermembrane space assembly pathway (MIA) is used by many IMS proteins containing a specific cysteine motif. Mia40 plays a key role in importing and refolding these precursor proteins into the intermembrane space [18]. Finally, α-helical pathway is used by α-helical transmembrane proteins for their insertion into the outer membrane by cooperation of mitochondrial import protein 1(Mim1) and SAM complex [19].

1.1.1 Presequence translocation pathway

Presequences, the classical targeting signals are usually peptides consisting of 10-80 amino acids at the amino-terminal end of the proteins. The presequence usually forms an amphipathic helix with one hydrophobic and one positively charged face. Precursor proteins are recognized by the receptors Tom20 and Tom22 of the TOM

complex and subsequently translocated through the Tom40 channel [14, 20]. It has been revealed that hydrophobic surface of presequence contact with Tom20 and the charged surface of the presequence contact with Tom22 [21]. The transfer of preproteins from TOM complex to TIM23 complex has been suggested to be mediated by Tim50 and Tim21 [22, 23]. Tom20 binds to the proteins emerging at the *trans* side of the TOM channel and assists their binding to intermembrane space domain of Tom22. Tim21 binds to Tom22 in a transient manner [24].
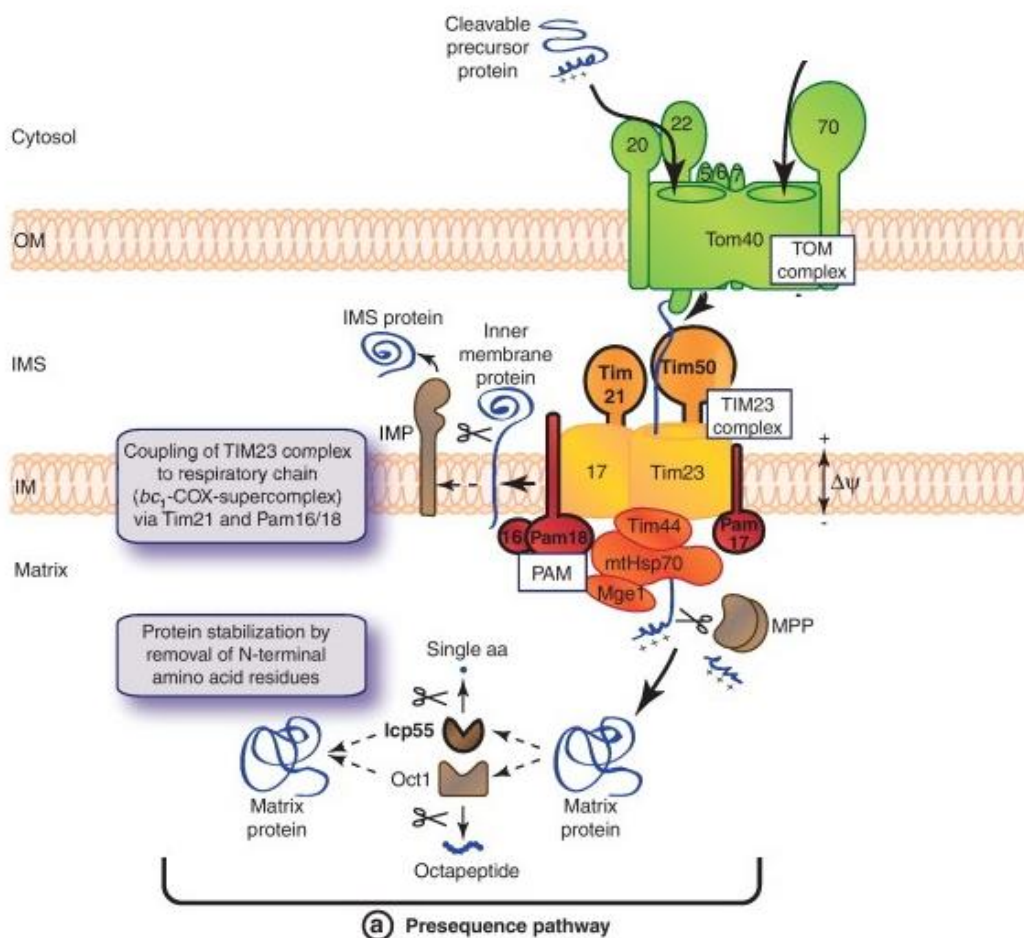
Figure 1. Presequence translocation pathway. The presequence pathway imports preproteins that carry positively charged presequences at the amino-terminus. The preproteins are recognized by the receptors Tom20 and Tom22 and are translocated across the outer mitochondrial membrane (OM) by the Tom40 channel. The TOM complex is regulated by cytosolic kinases. The preproteins are transferred to the TIM23 complex with the help of the intermembrane space (IMS)-exposed proteins Tim50 and Tim21. The membrane potential (Δψ) activates the Tim23 channel and drives translocation of the presequences across the inner membrane (IM). Coupling of the TIM23 complex to respiratory

chain complexes supports the Δψ-dependent step. Protein transport into the matrix is completed by the motor PAM with the ATP-dependentchaperone mtHsp70. The presequences are cleaved off by the mitochondrial processing peptidase (MPP). (Modified from [9] with permission)

It was shown that Tim21 competes with presequence binding to Tom22. Thus, Tim21 can be a promoter for release of the preproteins from Tom22 [22]. The preprotein translocation across the inner membrane is energetically driven by the electoral membrane potential across the inner membrane (Δψ) and hydrolysis ATP. The membrane potential Δψ activates the channel-forming protein Tim23 and causes an electrophoretic effect on the positively charged presequences [25, 26]. Matrix heat shock protein 70(mtHsp70) is the central protein of the presequence translocase-associated motor (PAM). It binds to the unfolded preproteins in transit and drives them into the matrix in cooperation with membrane-associated proteins [17].Upon arrival in the matrix, presequences are cleaved off from the imported proteins by the mitochondrial processing peptidase (MPP) and fold to mature conformation [27].

1.1.2 β-strand pathway

β-barrel membrane proteins are one type of integral membrane proteins which can be found in the outer membrane of gram-negative bacteria, chloroplasts and mitochondria [28]. There are three β-barrel membrane protein families residing in the outer membrane of mitochondria, VDAC, Tom40 and Sam50. Voltage-depend anion-selective channel (VDAC) is responsible for the passage of small molecules between the mitochondrion and the cytosol [29]. Tom40 and Sam50 are the central component of TOM and SAM complexes, which are protein translocators in the outer membrane [30].

Like other mitochondrial proteins, β-barrel proteins precursor are synthesized in cytosol and delivered to the mitochondrial surface with the assistance of cytosolic chaperones [31-33]. Precursors of β-barrel are first recognized by the receptor proteins Tom20 and Tom22 of TOM complex, and then are transferred to intermembrane space mediated by Tom5 [33]. The targeting signals within β-barrel precursors are not clearly identified. It is speculated that structural elements with high β-sheet content serve as specific targeting signals of β-barrel proteins [34].

After crossing the TOM complex, β-barrel precursors interact with the small Tim chaperones in the intermembrane space. These small Tim chaperones include

Tim9-Tim10 and Tim8-Tim13 complex [31, 35, 36]. The Tim proteins guide β-barrel precursors to assemble into 250 kDa complex by associating with SAM complex, the second translocase in the outer membrane. SAM complex has two essential proteins, Sam50 and Sam35 [37, 38]. Sam50 consists of one predicted polypeptide transport associated (POTRA) domain in the N-terminal and forms a β-barrel channel with a diameter of 40-50Å in the C-terminus [38, 39]. It is unclear that the β-barrel pores are formed within an individual Sam50 molecule or in an oligomeric complex with several Sam50 [38] because lateral opening of numerous hydrogen bonds that stabilize one β-barrel protein is thermodynamically unfavorable. Sam35 is a peripheral outer membrane protein which binds to the signal of β-barrel precursor proteins in a receptor-like manner [39] inducing a major conductance increase of the Sam50 channel. SAM complex also contains another hydrophilic subunits, Sam37 (Mas37) locating at its cytosolic surface. It is suggested that Sam37 contributes to the stability of the SAM complex [40] and promotes release of substrate proteins from the SAM complex [41].

Several other outer membrane proteins may promote the assembly of β-barrel proteins. Mdm10 has been proposed to play its role in maintaining mitochondrial morphology by associating with a fraction of SAM complex. The large complex is indicated to be necessary for the assembly pathway of Tom40 with receptor Tom22 and small Tom proteins [42]. Mdm10 and other two additional proteins, Mdm12 and Mmm1, form a MDM/mitochore complex which is also proposed to facilitate the β-barrel protein assembly [43]. Finally, the β-barrel precursor proteins released from the SAM complex are inserted into the outer membrane and achieve their final conformation.
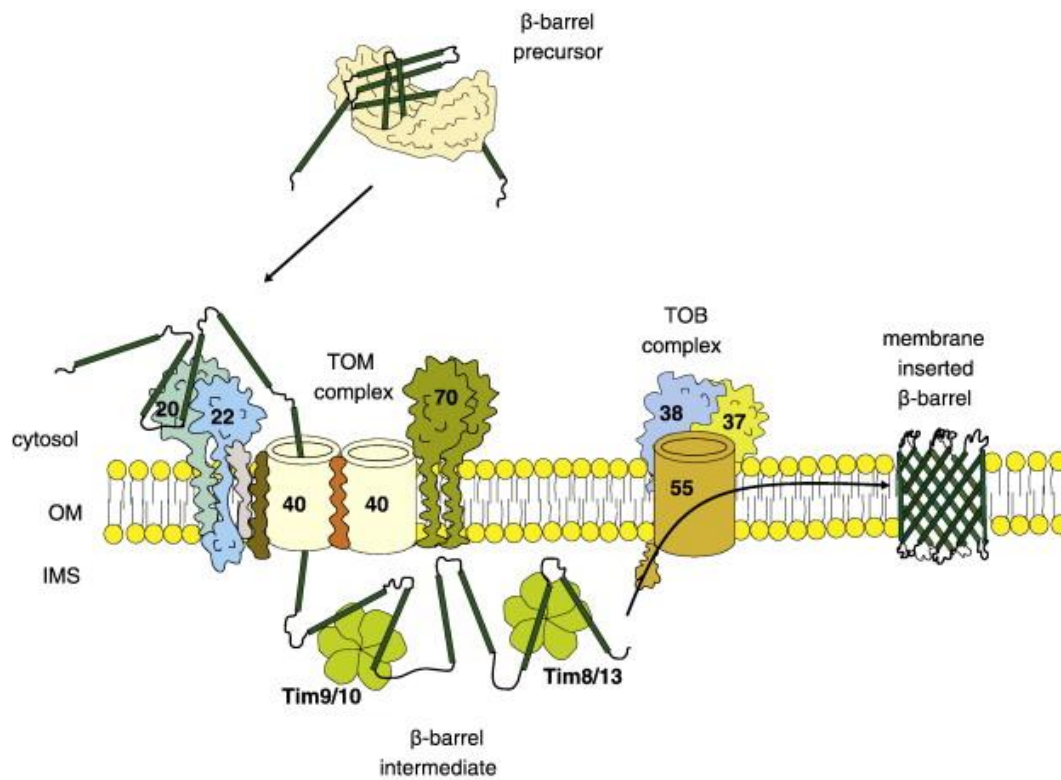
Figure. 2. Working model for the insertion of mitochondrial β-barrel proteins. Cytosolic factors deliver β-barrel precursors to the Tom receptors. Precursor proteins are then translocated through the import pore of the TOM complex and become exposed to the IMS. At this stage the β-barrel precursors bind to the small Tim chaperones. Finally, with help of the SAM complex β-barrel precursors are inserted and assembled into the MOM. (Reprinted from ref [17]with permission)

1.2   Tom40, the central proteins forming the conductive pore of the TOM complex

1.2.1   Translocase of outer membrane (TOM) complex

TOM complex is the only channel in the outer mitochondrial membrane for preproteins entering into mitochondria. The translocase complex contains two primary receptor proteins Tom20 and Tom70. The two subunits of TOM complex contain large hydrophilic domains that extend into the cytoplasm and act as receptors for precursor proteins. Tom20 is mainly the receptor of matrix precursor proteins which contain an N-terminal, cleavable presequence [44]. Tom70 recognizes mainly internal targeting signals within precursor proteins, such as the carrier proteins of the inner membrane [45-47]. The core of TOM complex is made up of Tom40, Tom22, Tom5, Tom6, and Tom7. Tom22 is tightly associated with translocase pore. It has been shown that Tom22 shares high substrate specificities with Tom20. It is suggested Tom20 mainly recognizes the hydrophobic side of amphiphilic helix in presequence [14] whereas Tom22 mainly binds to positively charged part of presequence [48]. Tom22 receives the preproteins from Tom20 and Tom70 and transfers them into the central pore with the assistance of Tom5, the smallest subunit of TOM complex. The C-terminus of Tom22 in intermembrane space is essential to form a *trans* site for presequence [49]. Tom5 consists of a single transmembrane helix anchor at C- terminus and exposes its small, negatively charge N-terminal segment to the cytosol. Other two small Tom proteins, Tom6 and Tom7 modulate the assembly and dynamics of the translocase in an antagonistic manner, with Tom6 stabilizing TOM complex and Tom7 promoting its dissociation [5, 7, 33, 50].
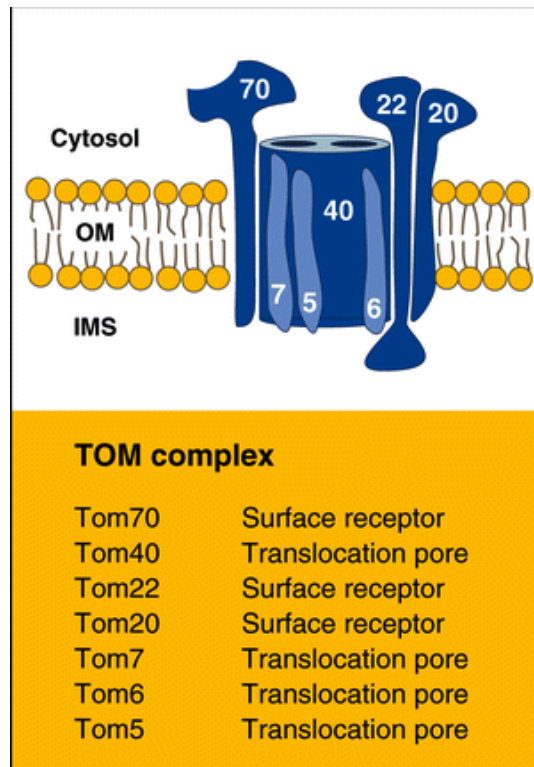
Figure 3: Composition of the TOM complex. The TOM (Translocase of the Outer Membrane) complex consists of the receptor subunits Tom70, Tom22 and Tom20 and the membrane-embedded subunits Tom40, Tom7, Tom6 and Tom5, which form the translocation pore. (Reprinted from ref. [2]with permission).

The general import pore of the TOM complex consists of the central component Tom40, Tom22 and three small associated subunits, Tom5, Tom6, and Tom7. However, it is not clear whether the pore is formed by single or multiple Tom40 molecules. Fully assembled TOM translocase, named Tom-holo complex, with a molecular mass of 490-600kDa was shown to possess two or three pores [51, 52]. Electrophysiological studies indicate Tom complex contains a cation-selective high-conductance channel in oligomeric form [52, 53]. After reconstituted into liposomes, it achieves activity of membrane proteins insertion and preproteins translocation of membrane. Electron microscopy studies identified the whole TOM complex particle has a diameter of about 138 Å (Figure 4a) [52]. TOM core complex which lacks Tom70 and Tom20 constitutes the same protein conducting channel as TOM holo complex [53]. It also preserves essentially the same conductive properties as shown by electrophysiology. Tom core complex reveals mainly a double ring structure with higher percentage of single rings particle than Tom holo complex by

electronic microscopy [52]. Tom40 on its own form a high conductance channel with molecular weight of 350kD in a homooligomeric form [54]. The Tom40 complex shows similar property as the entire TOM complex. Electronic microscopy of purified Tom40 revealed most particle consists of only one pores with a diameter of 2.5nm [54] (Figure 4c).
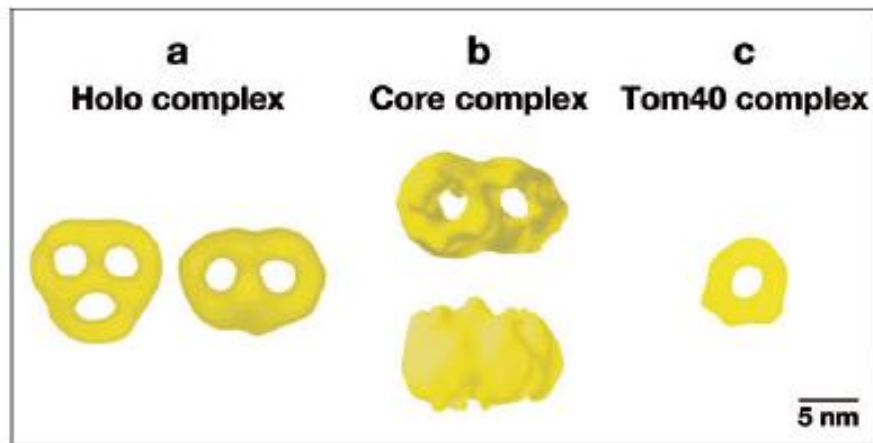
Figure 4. Electronmicrograph of the TOM complex. Single-particle image reconstitution after electron microscopic analysis and correlation averaging of negatively stained isolated TOM complexes represent: (a) the entire TOM complex [52], (b) the TOM core complex without Tom70 and Tom20 [53], and (c) the homooligomeric Tom40 complex [54]. (Reprinted from ref [2] with permission)

1.2.2   The structure of Tom40

Tom40 proteins with molecular weight from around 34 to 42kD have been discovered in a variety of eukaryotic cells. A 3D structure of Tom40 at atomic level has not been solved to this date. However, the secondary structure of Tom40 from several organisms has been investigated by CD and FTIR spectroscopy. Hill et. al. reveals a dominant 60% β-sheet conformation and 5% α-helical conformation in *Saccharomyces cerevisiae* Tom40 refolded in DDM micelle[55]. *Neurospora crassa* Tom40 in planar lipid membranes comprises a relatively low β-sheet portion of around 31% and α-helix of 22% [54] . Recombinant rat TOM40 (rTOM40) refolded in Brij35 consists of a 63% β-sheet structure [56]. In general, Tom40 from different organism possess dominant β-sheet structures and small α-helical component no matter incorporated into detergent micelles or liposomes [20]. Moreover Tom40 shares high sequence similarity to 30% with VDAC [57] and both belong to one proteins family: the mitochondrial porins [58, 59]. Therefore, it is suggested that Tom40 forms a β-barrel structure like VDAC.

β-barrel proteins have a typical pattern of alternating hydrophobic residues oriented toward the lipid bilayer and hydrophilic residues oriented toward the barrel interior. The lipid-exposed surfaces of β-barrels are abundant in amino acids such as phenylalanine, tyrosine, typtophan, valine and leucine, and rare polar and charge residues [60, 61]. On the other hands, its water-filled channel is abundant in small or polar amino acids glycine, threonine, serine, asparagine, glutamine and tyrosine, but lacks hydrophobic residues [60]. Aromatic residues are especially present at the bilayer interfacial region forming a saddle-like "aromatic belt".

Until year 2008, all the bacterial β-barrels had been characterized with an even number of β-strands from 8 to 24 [62, 63]. It was believed that membrane-bounded β-barrel proteins forms exclusively antiparallel strand pairs due to favorable energetically stability. In contrast, mitocondrial porin VDAC folds into a 19 β-strands barrel with a paralleled strand pair between the first and the last β-strands [57]. Also, the other mitochondrial porins are suggested to share this same exceptional conformation [58]. Hence, this could provide a new insight into the area of β-barrel protein and it suggests that the outer membrane proteins of mitochondria known as 'mitochondrial porins' form unique tertiary structure distinct from their bacterial counterparts.

hVDAC1 contains an N-terminal α-helix which may be involved in channel gating [57, 64, 65]. Tom40 N-terminus is also likely to form an α-helix which may be important for mediating the preprotein translocase. The length of Tom40 N-terminus varies from species to species. Mammalian Tom40A proteins have a longer N-terminal region in comparison to the other eukaryotes. Its extended N-terminus containing remarkable poly-proline region is exposed to cytoplasm whereas the C-terminus faces into intermembrane space [56]. In contrast, The N-terminal end of rat Tom40B and *nc*Tom40 are both exposed into intermembrane space [52, 66] indicating that the N-terminus transverses through the pore as in the case of VDAC.

1.2.3   Structural prediction of Tom40 based on VDAC

In 2008, three VDAC structures have been determined independently using X-ray Crystallography, high resolution NMR or a combined approach of both techniques showing basically the same conformation of a 19-strand β-barrel and the N-terminal α-helix [57, 64, 65]. This provides a reliable basis for homological modeling of structure. In the previous study from Zeth [67], the secondary structure of hTom40 and hVDAC1 were assigned using secondary structure prediction tools. Both predicted models consist of 19 β-strands which is largely consistent with the 3D structure of mVDAC [65]. The model reveals human Tom40 harbors more negative residues inside the channel particularly at the interface to the intermembrane space. This feature may be consistent with the fact that Tom40 may bind to positively charged presequence peptide during preprotein translocation. Furthermore, two conserved regions on the Tom40 molecule β2-β4 and β18-β19 strands were suggested to be involved in an interaction with other component of TOM complex.

Another structural model has been proposed for ncTom40 by Gessmann et. al. based on mouse VDAC. Multiple sequence alignments (MSAs) of ncTom40 and mVDAC1 show robust alignment to all structural segments except β-strands 2 and 4 and the aligned regions of ncTom40 were assigned to secondary structural elements according to structure of mVDAC1. The new model reveals the conserved G323 which is a central amino acid of the β-signal responsible for Tom40 insertion into outer mitochondrial membranes is located in β-strand 19. In the interior of pore, six conserved residues in β-strands 11 to 16 form a straight line and were assigned as α-helix anchor region. On the other side of the interior wall, a "polar slide" was identified and might be important for presequence binding. Furthermore, two conserved amino acid clusters β-strands β1 to β3 and β8 to β11 located at the outer surface of Tom40 barrel are likely to be involved in the assembly of TOM complex.

1.3   Solution NMR characterization of membrane proteins structures

Determination of structures and dynamics of transmembrane proteins is important for understanding of their function. X-ray crystallography and NMR spectroscopy are the only two methods that can provide 3D structures of proteins at atomic level. X-ray crystallography is still the leading method for characterizing membrane structure reflected by the high proportion of known membrane proteins structures solved by X-ray. However, it remains challenging to grow high-quality crystal from aqueous solution of detergent-solubilized proteins despite large progress in this area. For such a difficult case, NMR is a complimentary method for membrane protein structure determination.

1.3.1   Membrane media for solution NMR studies of IMPs

Membrane proteins are naturally embedded in lipid bilayer, which provide hydrophobic environment for solubilizing membrane proteins. Studies of membrane proteins require it to be properly folded into a suitable membrane mimetics. Liposome which is composed of bilayers is the class of model membrane that most closely approximates the native environment of membrane proteins. However, liposome-proteins complex have large molecular weight and tumble slowly in solution severely broadening solution NMR signals. In the area of solid state NMR, liposomes are broadly employed as solid state NMR is less sensitive to molecular weight. Therefore, detergents micelles are the most employed medium for solubilize integral membrane proteins (IMPs) in solution NMR [57, 64, 68] (Figure 5a). They form small spherical monolayers minimizing contact of their hydrophobic tails with the aqueous environment. Detergent molecules such as DHPC are amphipathic molecules, consisting of a hydrophobic and a hydrophilic portion. At concentrations below their critical micellar concentration (cmc), detergent molecules exist as monomers in aqueous solution. Above the cmc, multimeric detergent micelles assemble in a highly dynamic equilibrium with detergent monomers. Bicelles are detergent-lipid mixture assemble into bilayer (Figure 5b). Bicelles represent a more native-like environment for membrane proteins with smaller membrane curvature than micelles. Isotropic

bicelles have been employed as the medium for structural studies of a number of IMPs using NMR despite of a relatively larger molecular weight than micelles [69-71]. Except these two widely used membrane media, several other membrane systems have been developed. 1) Reversed micelle encapsulates the membrane protein within the protective water core and the resulting complex is dissolved in a low viscosity organic solvent. Reduced correlation time due to a low viscosity solvent gives rise to narrow line of NMR signal of a membrane protein without deuteration [72-75] (Figure 5c). 2) Amphipols (APols) which are amphipathic polymers that adsorb onto the hydrophobic transmembrane surface of MPs remains water-soluble, native state of membrane proteins in the absence of detergent [76-79](Figure 5d). 3) Nanolipoprotein particles (NLPs), also referred as nanodiscs are self-assembling complexes that consist of a small phospholipid bilayer core surrounded by an amphipathic membrane scaffold protein (MSP)(Figure 5e). NLPs yield the native-like bilayer architecture promoting both stability and functionality of the integrated proteins [80-89].
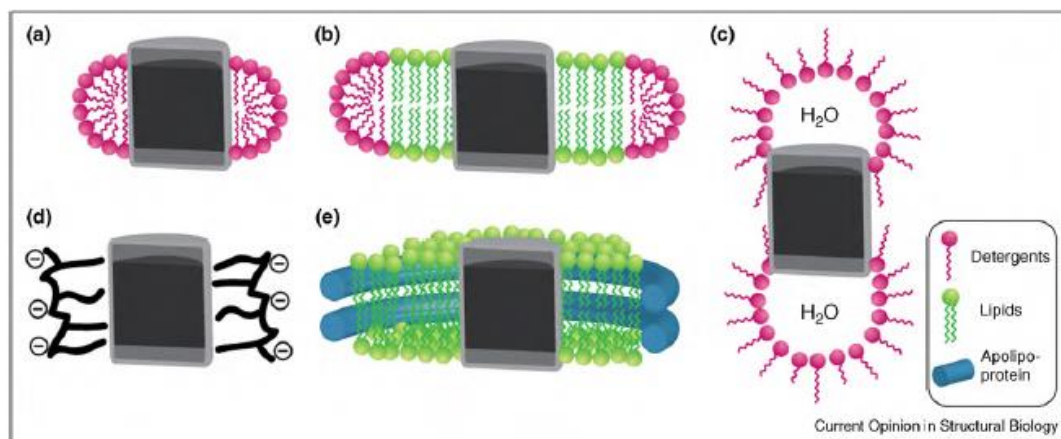
Figure 5: Schematic representation of different solubilization methods for integral membrane proteins. Micelle(a), bicelle(b), reverse micelle(c), amphipo (d), and nanolipoprotein particles (NLP; e). Shown are cartoons depicting each solubilization system. The membrane protein is depicted in gray, with its hydrophilic and hydrophobic regions in light and dark gray, respectively. Detergents are colored in magenta and lipids in green. The amphipol polymer chain is shown as black lines; the negative charges are indicated by minus signs. The apolipoprotein is represented as a blue ring. With the exception of the reverse micelle assembly (c) all systems are in aqueous solution. The reverse micelles (c) are dissolved in organic solvents; areas containing aqueous solution are indicated. (Reprinted from ref. [90] with permission)

1.3.2   Solution NMR methods for membrane proteins

There are still several obstacles to successfully tackle IMPs in the range of 100-400 residues which could be addressed by some special NMR methods. Firstly, NMR signals on membrane proteins complexes are broadened or disappear due to reduced tumbling rate of large molecule. TROSY-based NMR experiments enable to reduce relaxation rate for large proteins and extend the molecular weight limit for NMR studies. The original TROSY ^{1}H-^{15}N COSY intently selects the sharpest component amide group and enhances the spectral signal-to-noise and resolution for large molecule [91]. The NMR experiments utilizing TROSY elements have facilitated the NMR studies of protein complex up to 100KDa. Another methyl-TROSY selects the narrowest components of multiplets involving methyl or methylene protons coupled to aliphatic ^{13}C [92]. It has enabled side chain assignment and measurement of long-range NOEs even for very large proteins with specific isotopic labeling pattern. Deuteration can be combined with TROSY to provide further line narrowing for large proteins. The advantage of replacing proton to deuteron is the reduction of dipole-dipole relaxation, due to the 6.5-fold lower gyromagnetic ratio of the deuteron compared with proton [93].

The relatively poor chemical shift dispersion of membrane proteins (especially α-helical IMP) heightens the difficulty for chemical shift assignment. Therefore, in addition to deuteration, other advanced labeling strategies have been proposed to simplify the spectra. A specific ^{15}N-labeled amino-acid selective labeling have long been used to classify resonances according to residues types [94]. Another labeling scheme to simplifying NMR spectra is to label only certain segment of full-length proteins [95]. It requires expressing both labeled and unlabeled chains and then ligating them together. Methyl TROSY which usually benefits from deuterated background also requires a reintroduction of LEU, ILE and VAL methyl protons [96]. Moreover, the classic approach for protein structure determination by solution NMR based on NOE-derived distance restraints may not work for membrane proteins complex. Side chain NOEs are difficult to measure if working with a deuterated protein which mostly is necessary for large membrane proteins complex. The reintroduction of specifically protonated amino acid in perdeuterated protein, for example methyl protonated proteins may yields relaxation-optimized samples and NOEs between methyl groups or between methyl proton and amide proton. However,

in this case, the number of NOEs detected is still small, making high resolution structural determination difficult. An alternative source of distance constraint is the paramagnetic relaxation enhancement (PRE) resulting from the proximity of a proton to an unpaired electron which is usually provided by labeling with a 'spin label' or paramagnetic metal ion at specific site [97-99]. The interaction between electron and nucleus can be quantitated over long distance (up to ca. 30Å). Residual dipolar coupling and chemical shift anisotropies provide a rich source of global structural restraints [100, 101]. In solution NMR, the dipolar coupling and CSA are largely averaged out. Weak alignment can be introduced into an isotropic solution through alignment medium such as bicelles, filamentous phage, strained or charged polyacrylamide gel matrices, or by attachment of a strong paramagnetic agent such as a lanthanide [102-106].

1.4   Aims and Outline

Tom40, pore-forming protein of TOM complex, plays a vital role in mitochondrial protein import. The dominance of β-sheet has been identified by circular dichroism (CD) and Fourier transform infrared spectroscopy (FTIR). Although it is believed to fold into a β-barrel like VDAC, the atomic level structure of Tom40 is still unknown. It is suggested that Tom40 plays a more active role than just a passive pore in mitochondrial protein translocation. For example, Tom40 binds to transiently unfolded segments of the precursor proteins and promote partial unfolding of the substrate proteins. However, the mechanism of Tom40 mediating the unfolded precursor is not clear. Here we use NMR spectroscopy to characterize the structure of Tom40 and its interaction with presequence and membrane.

In the first part of this work, we have characterized the structure of Tom40 using a novel Hydrogen/Deuterium exchange protocol which enables us to obtain structural information of native Tom40 by solution NMR measurement on monomeric denatured Tom40. The assignment of denatured 349-residue Tom40 has been implemented by APSY-based high-dimensional NMR and the secondary chemical shift analysis are implemented to obtain structural information of denatured Tom40. The result from H/D exchange measurement can be compared with predicted topology model based on VDAC.

In the second part, the interaction of Tom40 with presequence is characterized by NMR-based titration. N-terminal Tom40 (1-59) has been titrated by presequence rALDH. TFE(2, 2, 2-Trifluoroethanol) was added to promote α-helical conformation and strengthen the helix-helix interaction between Tom40 and presequence.

In the third part, the interaction of Tom40 with membrane is investigated by titration of urea-denatured Tom40 with detergent micelle. The secondary chemical shift analysis reveals the presence of the residual secondary structure elements. Binding clusters in Tom40 with micelle can be observed by monitoring chemical shift perturbation.

2   Material and Methods

2.1   Material

2.1.1   Chemical reagents

2,2,2-Trifluoroethanol (TFE) was purchased from Sigma (Sigma−Aldrich Chemie GmbH, Schnelldorf, Germany). Urea was purchased from Merck, Darmstadt, Germany. Guanidinium thiocyanate (GdnSCN) was obtained from Sigma (Sigma−Aldrich Chemie GmbH, Schnelldorf, Germany). 1,2-dihexanoyl-*sn*-glycero-3-phosphocholine(DHPC), detergent micelle, was obtained from Avanti Polar Lipids, Inc.

2.1.2   Equipments

Laboratory instruments used for NMR measurements and biophysical studies are summarized in Table 1.

Table 1: Instruments

Item	Identifier/Company
Balances	Sartorius B 3100 S & AC 210 S
Centrifuges	Beckmann-Coulter Avanti J-20 and J-301, rotors: JLA 8.100, JLA 9.100, JLA 16.250, JA 25.50 Ti, JA 30.50 Ti, Krefeld, Germany
Lyophilization	Braun Biotech Christ Alpha 2-4

NMR	Bruker Avance 400 MHz, TXI probe,
	Bruker AvanceIII 600 MHz, QXI probe
	Bruker Avance 600 MHz, cryoprobe
	Oxford Avance 700 III MHz, TXI probe
	Bruker Avance 700 MHz, cryoprobe
	Bruker Avance III 800 MHz, cryoprobe
	Bruker Avance 900 MHz, cryoprobe
	NMR Sample Tubes 5 mm, Norell, Inc., Landisville, NJ, USA
	Shigemi NMR tube 5 mm, Shigemi Corp., Tokyo, Japan

2.1.3   Scientific software

Table 2: Software used in the studies

Software name	Source/reference
Topspin 2.1	Bruker Biospin, Karlsruhe, Germany
NMRPipe/NMRDraw	spin.niddk.nih.gov/NMRPipe (Delaglio et al. 1995)
Sparky 3	www.cgl.ucsf.edu/home/sparky
	T. D. Goddard and D. G. Kneller, SPARKY 3
PROSA	Güntert, P., Dötsch, V., Wider, G. & Wüthrich K. 1992
GAPRO	S. Hiller et al, 2005
MARS(graphic interface)	http://www.enmr.eu/webportal/mars.html
	Jung, YS and Zweckstetter, M. 2004
Pymol	DeLano Scientific LLC, Palo Alto, CA, USA
MATHEMATICA	Wolfram Research, Champaign, IL, USA
Igor pro	Wave Metrics, Inc
Microsoft Excel	Microsoft, USA

2.2   Sample preparation

Sample expression and purification of full length and the N-terminal (1-59) peptide of *Neurospora crassa* Tom40 (ncTom40) were performed by Dr. Stefan Becker and Yvonne Laukat in Department for NMR-based Structural Biology, Max Planck Institute for Biophysical Chemistry, Göttingen.

2.2.1   Production of ncTom40

The expression, refolding and purification of ncTom40 were performed as previously described [20]. ^{15}N- and ^{13}C-labelled protein was expressed in minimal medium with ^{15}NH$_4$Cl as nitrogen source and ^{13}C$_6$-D-glucose as carbon source. Amino acid selective ^{15}N-labeled ncTom40 proteins were expressed according to a recently published protocol [107]. The amino acid selective ^{15}N-labeled samples used in this work were listed in Table 3. All reconstituted proteins were dissolved in 10 mM MOPS, 10 mM KCl, pH 7.

Table 3: Amino acid selective ^{15}N-labeled samples of ncTom40

Samples	^{15}N-labeled amino acid
ncTom40 in liposomes	PHE/MET
	LEU
	ASN/ILE
	VAL/THR
	GLN/TYR/TRP
	ASP/GLU
	ALA/HIS/ILE/MET/THR
	GLN/TRP/THR
	PHE/ASN
	LYS
	ARG
	PHE/GLU
ncTom40 in 8M urea	GLN/TYR/TRP
	GLU/LYS
	ARG
	ILE/MET/HIS/ALA/THR
	PHE/ASN
	ASP/TYR

2.2.2 Production of the N-terminal (1-59) peptide of ncTom40

A fragment comprising amino acids 1-59 of *Neurospora crassa* Tom40 was expressed in *E. coli* as a fusion protein with an N-terminal Z_2 domain [108]. Uniformly ^{15}N- and ^{13}C-labeled fusion protein was expressed in minimal medium with ^{15}N-NH_4Cl and $^{13}C_6$-D-glucose as sole nitrogen and carbon source. The fusion protein was purified by metal affinity chromatography on Ni-NTA agarose (Qiagen). After cleavage with tobacco etch virus (TEV) protease the released Tom40 fragment was further purified by HPLC. The lyophilized peptide (1 mg) was dissolved in 225 μl 10 mM MOPS and 10 mM KCl at *p*H 7,. For NMR measurements this solution was further supplemented with 25 μl ^2H_2O.

2.2. 3 Synthesis of presequence peptide

The peptide sequence corresponding to retinal aldehyde dehydrogenase (rALDH) presequence (MLRAALSTARRGPRLSRLLSAA) was synthesized using standard solid phase peptide synthesis (FMOC) by K. Overkamp in Department of NMR-based Structural Biology, Max Planck Institute for Biophysical Chemistry in Göttingen.

2.3   Backbone resonance assignments

2.3.1 Assignment of ncTom40 in GdnSCN

The backbone assignment of ncTom40 denatured in GdnSCN was based 5D, 6D and 7D experiments at 278K, 295K and 310K acquired by Auto Projection Spectroscopy(APSY) which enable the measurement of very high-dimensional NMR spectra. It performs simultaneous chemical shift evolution of several indirect dimension and accelerate the acquisition of high-dimensional NMR (section 5.5, [109]). APSY spectra were processed and analyzed as described in section 6.2 and 6.3.

Table 4:   Experiments for backbone assignment of ncTom40, 4M GdnSCN, 0.4% formic acid, *p*H 2.5.

Temperature	Experiments	Comments
310K	7D HNCO(CA)CBCANH	7D, 6D and 5D spectra
	5D CBCACONH	were acquired using APSY.
		Automatic assignment was
295K	6D HNCOCANH	performed by MARS.
	5D CBCACONH	
278K	6D HNCOCANH	
	5D CBCACONH	
	3D HNN	

(The spectra were recorded by Dr. Ulrich H.N. Dürr)

3D HNN is proved to be robust for assignment of unfolded proteins taking advantage of the high dispersion of ^{15}N chemical shift . The magnetization transfer through the pulse sequence can be described in short as: [110]

$$H^N_i \longrightarrow N_i(t_1) \begin{cases} C^\alpha_i \begin{cases} N_{i+1}(t_2) \longrightarrow H^N_{i+1}(t_3) \\ N_i(t_2) \longrightarrow H^N_i(t_3) \end{cases} \\ C^\alpha_{i-1} \begin{cases} N_i(t_2) \longrightarrow H^N_i(t_3) \\ N_{i-1}(t_2) \longrightarrow H^N_{i-1}(t_3) \end{cases} \end{cases}$$
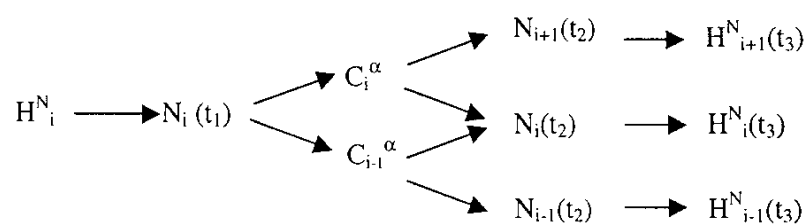
Therefore a sequential walk can be performed in both directions along the chain by connecting HN(i) to N(i-1) and N(i+1).
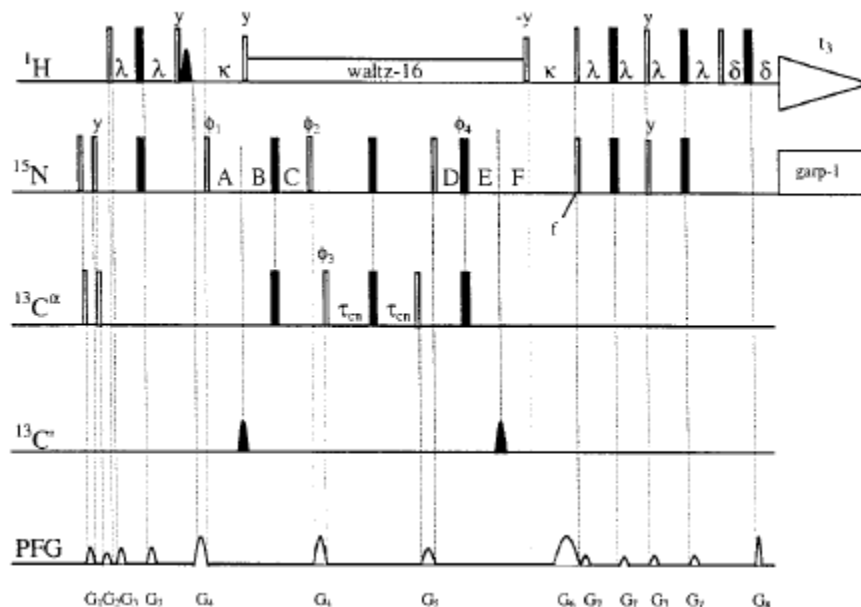
Figure 6: Pulse sequence of HNN. Narrow (hollow) and wide (filled black) rectangular bars indicate non-selective 90 °and 180 °pulses, respectively. Unless indicated otherwise, the pulses are applied with phase x. The ^1H and ^{15}N carrier frequencies are set at 4.71 ppm (water) and 119.0 ppm, respectively. The 13C carrier frequency is set at 56.0 ppm. The selective 1H 90◦ pulse used for water flip-back during the first INEPT transfer is applied with one-lobe sinc profile. Proton decoupling using the Waltz-16 decoupling sequence is applied during most of the t1 and t2 evolution periods, and ^{15}N decoupling using the Garp-1 sequence with 0.5 kHz is applied during acquisition. The strength of the $^{13}C\alpha$ pulses is adjusted so that they cause minimal excitation of carbonyl carbons. The 180 °^{13}CO shaped pulse had a one-lobe sinc profile with minimal excitation of 13Cα. The delays are λ = 2.7 ms, κ = 5.4 ms and δ = 0.35 ms. τ_{CN} must be optimized and is around 12–16 ms. The values of the individual periods containing t1 are: A = t1/2, B = TN and C = TN−t1/2. The values of the individual periods containing t2 are: D = TN−t2/2, E = TN and F = t2/2. Phase cycling for the experiment is φ1 = 2(x), 2(−x); φ2 = x, −x, x, −x; φ3 = φ2; φ4 = x; φ5 = 4(x), 4(−x) and receiver = 2(x), 2(−x). Frequency discrimination in t1 is achieved using States-TPPI phase cycling of φ1 along with the receiver phase; frequency discrimination in t2 is achieved using the PEP sensitivity-enhanced gradient method. The N- and P-type signals are collected separately by inverting the sign of the G6 gradient pulse. The gradient duration and levels are as follows: G1 = 0.5 ms, 8 G/cm; G2 = 0.5 ms, 5.6 G/cm; G3 = 0.5 ms, 8 G/cm; G4 = 1 ms, 11 G/cm; G5 = 1.0 ms,8 G/cm; G6 = 2.5 ms, 26 G/cm; G7 = 0.5 ms, 2 G/cm; G8 = 0.25 ms, 25.8 G/cm. (Reprinted from [110] with permission)

2.3.2   Assignment of ncTom40 in urea

The sequence-specific assignment of ncTom40 in 4M urea at 298K was 6D HNCOCANH APSY and 5D CBCACONH APSY. The 2D projections recorded in APSY were set up according to angle sets created by GAPRO program. 6D APSY-

HNCOCANH APSY recorded 65 2D projections produced by 25 angle sets and 5D APSY-CBCACONH acquired 88 2D projections created by 30 angle sets (Table 5).

Table 5: Parameters of APSY spectra for backbone assignment of ncTom40 in 8M urea, pH 4 at 298K

Experiments	Number of Angle sets/Projections	Number of data points in each projection (real points)
6 D HNCOCANH APSY	25/65	1024×272
5D CBCACONH APSY	30/88	1024×170

2.3.3 Assignment of ncTom40 N-terminal peptide (1-59) in the presence and absence of TFE

The N-terminal peptide of ncTom40 (1-59) was sequence-specific assigned by combining non-linear-sampling and conventional 3D triple resonance experiments. The backbone assignment of the N-terminal ncTom40 in 30% TFE was acquired by conventional 3D triple resonance experiment. The parameters are listed in Table 6. The HNCACB and CBCACONH spectra were recorded in non-linear sampling (NUS) manner with 20% data sampled. The spectra were processed by Shengqi Xiang using MddNMR.

Table 6: Parameters of 3D triple resonance experiment for backbone assignment of N-terminal ncTom40 at 298K in the absence and presence of TFE.

Sample	Experiment	Sweep Width(ppm)	Carrier Frequencey(ppm)	Number of data points(real points)
N-terminal ncTom40	NUS HNCACB	12(H),64(C),24(N), 600M	4.7(H), 42(C),118(N)	$1024(H) \times 180(C) \times 72(N)$
	NUS CBCA(CO)NH	12(H), 64(C),24(N) 600M	4.7(H), 42(C),118(N)	$1024(H) \times 128(C) \times 72(N)$
	CBCANH	10(H),58(C),24(N) 800M	4.7(H),40(C),118(N)	$2048(H) \times 60(N) \times 90(C)$
N-terminal ncTom40 in 30% TFE	CBCA(CO)NH	10(H),24(C),64(N) 800M	4.7(H),116(N),39(C)	$1024(H) \times 48(C) \times 124(N)$
	CBCANH	10(H),24(C),64(N) 800M	4.7(H),116(N),39(C)	$1024(H) \times 54(C) \times 132(N)$

2.4   Secondary chemical shift analysis

The strong correlation between chemical shifts and local structure make it possible to derive secondary structure elements from chemical shifts. Secondary chemical shift is defined as:

$$\Delta\delta = \delta_{observed} - \delta_{random\ coil} \tag{2.1}$$

$\delta_{observed}$ and $\delta_{random\ coil}$ are measured chemical shift of the proteins and chemical shift for random coil. Cα atoms in α-helices, for example, will tend to have positive values for $\Delta\delta$ and Cα and C$'$ atoms in β-strands will tend to have negative values for $\Delta\delta$. The behaviour of Cβ atoms is exactly opposite to that of Cα atoms in β-strands that mainly positive values are observed and in α-helices mainly negative ones. Random coil value of Cα and Cβ value were calculated using the approach of Schwarzinger on server (http://nmr.chem.rug.nl/cgi/bin/start_session_FASTA.py) [111].

2.5   Best HSQC (Heteronuclear single quantum coherence)

[^{1}H,^{15}N]-HSQC experiment is one of the most frequently used experiments in protein NMR. The [^{1}H,^{15}N]-HSQC provides the correlation between amide proton and its covalently bonded nitrogen. In general, each residue would produce an observable peak in the spectra, with the exception of proline which lacks an amide proton. The [^{1}H,^{15}N]-HSQC is like a fingerprint of proteins and thus is widely used for characterization of the dynamics of proteins and interaction with other proteins or ligand for each specific residue site.

To capture the intrinsic fast H/D exchange, Band-selective Excitation Short-Transient (BEST)-[^{1}H,^{15}N]-HSQCs were recorded replacing standard [^{1}H,^{15}N]-HSQC to speed up NMR acquisition(pulse sequence in Appendix 1). Comparing to standard HSQC, The BEST scheme use shaped pulsed to selectively excite only amide protons (Figure 7) and keep magnetization of other protons in equilibrium, thereby accelerating magnetization recovery and shortening the recycle delay. The recycle delay of 0.5ms was used with optimized S/N per time.
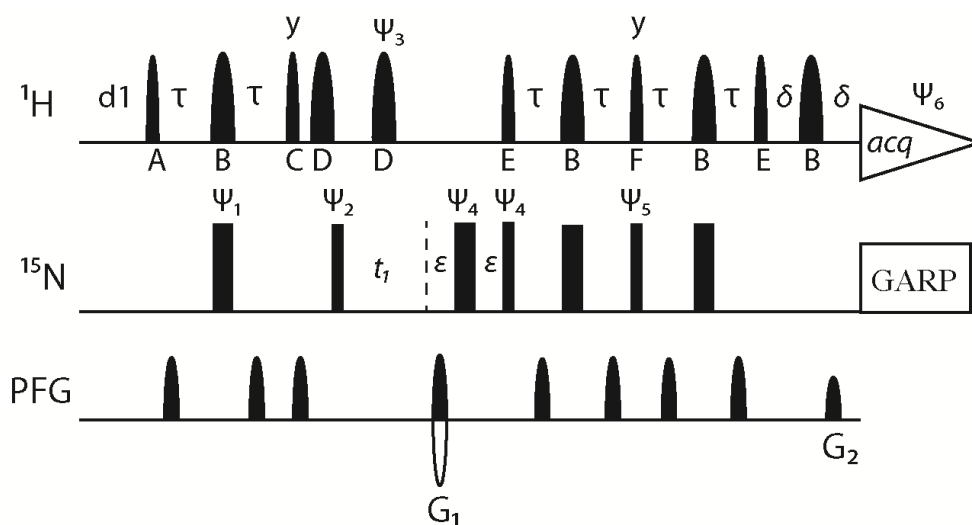
Figure 7. Pulse sequence of Best [^{1}H,^{15}N]-HSQC. Narrow and wide rectangular bars represent non-selective 90 ° and 180 ° pulses on ^{15}N channel, respectively. Thin (fat) sine bells are shaped 90 °(180 °) pulses. All selective 1H pulses are centered at 8.2 ppm covering the amide proton region with following shapes:(A) PC9 120 °(B) RE-BURP (C) PC9 90 °(D) BIP-720-50-20 (E) EBURP2 (F) flip-back pulse of EBURP2. Decouplings using GARP on ^{15}N are indicated by white rectangles. On the line marked PFG, curved shapes indicate sine bell-shaped, pulsed magnetic field gradients applied along the z-axis. G_1 and G_2 use the same pulse duration and gradient strengths are given by the gyromagnetic ratios $G_1/G_2 = \gamma_H/\gamma_N$. The following phase cycling was used: $\psi_1 = x$; $\psi_2 = x,-x$; $\psi_3 = x,x,-x,-x$; $\psi_4 = x,x,-x,-x$; $\psi_5 = y,y,-y,-y$ and all pulses without indication of a phase above the pulse symbol were applied with phase x. For quadrature detection in t_1, echo-antiecho data are recorded by inverting the sign of gradient G1 and phase ψ_5 [112, 113] .

2.6   Characterization of protein-ligand interactions by chemical shift perturbation

Chemical shift is very sensitive to the electronic environment of a nucleus. Interaction with solvent molecule or binding partners can induce change of chemical shift making it a very sensitive probe for the identification of binding interfaces of protein complexes.

Typically, $[^1H,^{15}N]$-HSQC spectra of ^{15}N label proteins are recorded in the absence(free form) and presence (bond form) of unlabeled ligands or binding partner at different molar ratios to monitor the chemical shift changes.

To quantitatively analyze the chemical shift change, Average chemical shit perturbation (CSP), $\Delta\delta$, is calculated as:

$$\Delta\delta = \sqrt{(\Delta\delta_N/5)^2 + (\Delta\delta_H)^2}$$

(2.1)

$\Delta\delta_N$ and $\Delta\delta_H$ represent the chemical shift differences (ppm) between free form and bond form of proteins.

1) Titration of presequence into ^{15}N labeled Tom40

   Presequence rALDH unlabeled was titrated into ^{15}N labeled N-terminal Tom40 (1-59) with ratio of Tom40: rALDH =1:0.5, 1:1, 1:2, 1:5, 1:10, 1:20

2) Titration of presequence into ^{15}N labeled Tom40 in 30% TFE

   Presequence rALDH unlabeled was titrated into ^{15}N labeled N-terminal Tom40 (1-59) with ratio of Tom40: rALDH =1:1, 1:2, 1:5, 1:10, 1:20

3) Titration of DHPC detergent into ncTom40 in 8M urea

   DHPC detergent was titrated into ncTom40 in 8M urea with concentration of 25mM, 50mM, 100mM and 200mM.

2.7   **Homological model of ncTom40**

The homological model of Tom40 was obtained using the iterative threading assembly refinement (I-TASSER) server. It starts from an amino acid sequence and generates three-dimensional (3D) atomic models from multiple threading alignments and iterative structural assembly simulation.

The hVDAC1 structure was put into iTasser server and the structure of human Tom40 was obtained. Then based on hTom40 structure, another round of iTasser prediction generated the structural model of ncTom40.

3   Result

3.1   Structural characterization of ncTom40

As main component of TOM complex, the structure of Tom40 will benefit to understanding the function of Tom40 in the process of proteins import into mitochondria. Due to high sequence similarity, it is proposed that Tom40 shares the similar β-barrel conformation as VDAC which consists of a 19 β-strands barrel and an α-helix at N-terminal. Several homological models have been generated based on the 3D structure of VDAC. The 32kD hVDAC1 is highlighted as the largest β-barrel membrane protein whose structure has been determined by solution NMR (http://www.drorlist.com/nmr/MPNMR.html). Tom40 with higher molecular weight will certainly be a huge challenge for solution NMR. The initial trial in our group to reconstitute Tom40 in detergent micelle resulted in very broad resonance and the direct investigation by solution NMR was hindered. To characterize the structure of Tom40 we developed a novel NMR-based Hydrogen/Deuterium exchange protocol which is able to circumvent the limit of molecular weight.

3.1.1 Hydrogen/Deuterium exchange strategy

Analysis of the kinetics of hydrogen exchange in biological macromolecules is able to provide rich structural and dynamic information of biological molecule as the participation of exchangeable hydrogen within hydrogen bonds is the main factor limiting hydrogen exchange. High resolution NMR is able to measure site-specific hydrogen/deuterium exchange and provides a detailed picture of protein structure and dynamics at residue resolution. This method is applicable to small membrane proteins

solubilized in detergent micelle for direct characterization of amide exchanges of each residue [114, 115]. For larger membrane proteins, such direct measurements usually are hindered by the large effective molecular weights of protein-detergent complex.

To characterize the structure of ncTom40, we developed the novel H/D protocol apply a forward H/D exchange to native Tom40 following dissociation of Tom40 complex and NMR detection in denatured state as outlined in Figure 8. NcTom40 reconstituted liposomes is centrifuged at $168,000 \times g$ for 1 hr. The resulting pellet is transferred into 100% D_2O and incubated for 10 minutes. The pellet is collected after incubation by centrifugation for 15 minutes and put into dissolution buffer containing 4M guanidinium thiocyanate (GdnSCN), 0.4% formic acid, with 50%, 75% or 100% D_2O. Tom40 proteins are denatured in 4M GdnSCN and released from liposomes. A series of ^1H-^{15}N heteronuclear single quantum coherence ($[^1H,^{15}N]$-HSQC) spectra is then recorded to monitor the back-exchange process at 278K and pD 2.5 [116]. The dead time for setup of the NMR experiment is around 20 minutes. $[^1H, ^{15}N]$-HSQCs are recorded using the BEST scheme [117] to speed up NMR data acquisition.

The two-stage H/D exchange protocol (Figure 8) makes the residues, which do not exchange amide proton with water in 100% D_2O when the protein is embedded into liposomes, lose proton during back-exchange, while those residues, which fully exchanged in the native state, will gain proton in the dissolution buffer. Therefore, the differentiation of water protected or exposed residues during forward exchange is made possible upon analyzing the intensity changes of signals on $[^1H, ^{15}N]$-HSQC during dissociation and achieves structural information of native ncTom 40 in liposome.
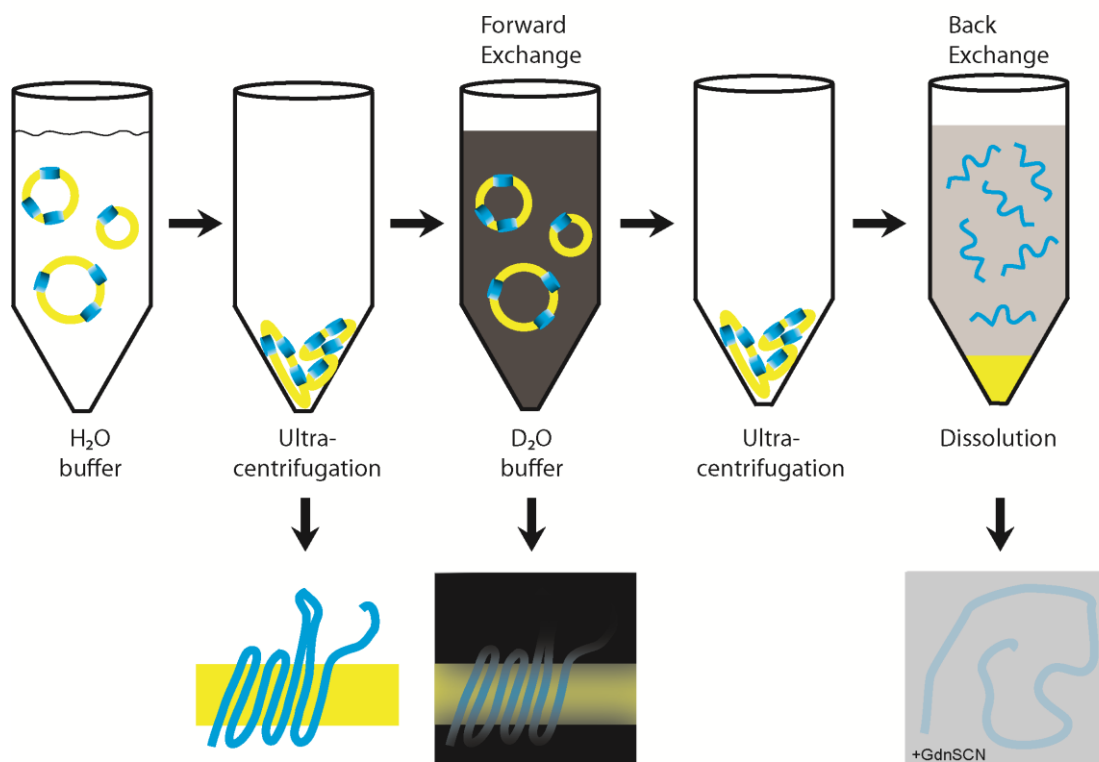
Figure 8. Scheme illustrating the H/D exchange strategy developed for membrane proteins (blue) reconstituted into liposomes (yellow). A white color indicates H$_2$O buffer, black color 100% D$_2$O buffer and grey color the dissolution buffer, which contains 4M GdnSCN. During the incubation period in 100% D$_2$O solvent exposed residues will exchange amide protons against deuterium (lower row, middle panel). They will therefore not be visible in the denatured monomer (lower right panel).

3.1.2   Assignment of unfolded ncTom40

The residue specific analysis of H/D exchange required the resonance assignment of ncTom40 denatured in 4M GdnSCN. Sequence-specific assignment of 349-residue ncTom40 denatured in GdnSCN will be challenging and time-consuming using conventional triple resonance NMR assignment due to severe spectra overlap (Figure 9). Alternatively, we performed the automatic projection spectroscopy (APSY) in combination with MARS [118, 119] program for automatic backbone assignment. APSY enables the measurement of high-dimensional spectra, reducing NMR signal overlap, and provides highly accurate resonance lists [109, 120]. High-dimensional spectra are particular useful to relieve the spectral overlap of disordered proteins. On

the other hands, the favorable relaxation properties of disordered proteins benefit to the sensitivity of high-dimensional spectra.
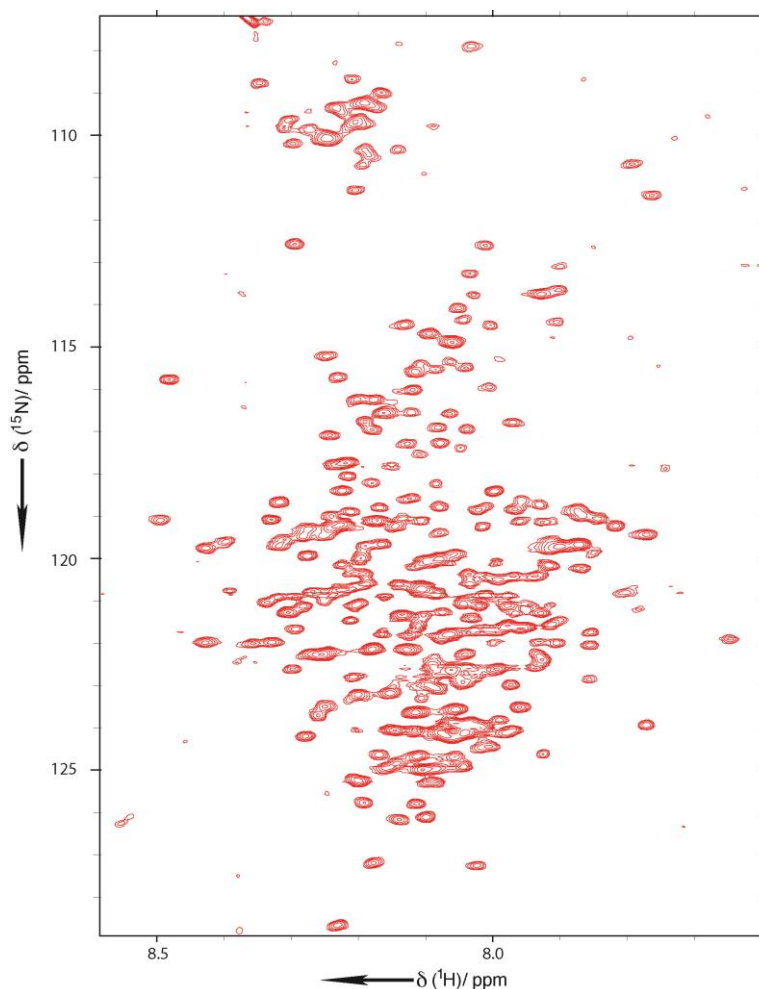
Figure 9. Two-dimensional [^{1}H,^{15}N]-HSQC spectrum of uniformly ^{15}N-labeled ncTom40 in dissolution buffer containing 4M GdnSCN, 0.4 % formic acid.

APSY experiments [109, 120, 121] were recorded at 278K, 295K and 310K on ^{13}C/^{15}N-labeled ncTom40 in 4M GdnSCN, 0.4% formic acid. [109, 122]. In APSY experiments, a series of 2D projections are recorded (Figure 10). The indirect dimension combines chemical shifts from several different nuclei achieving high-dimensional correlation of nuclei. Figure 10 shows 5 orthogonal projections of ncTom40 at 278K recorded by 6D HNCACONH APSY corresponding to C_α(i-1)-H_N(i-1), N(i-1)- H_N(i-1), N(i)-H_N(i-1),C′(i-1)-H_N(i-1), H_N(i-1)-H_N(i-1) projections.
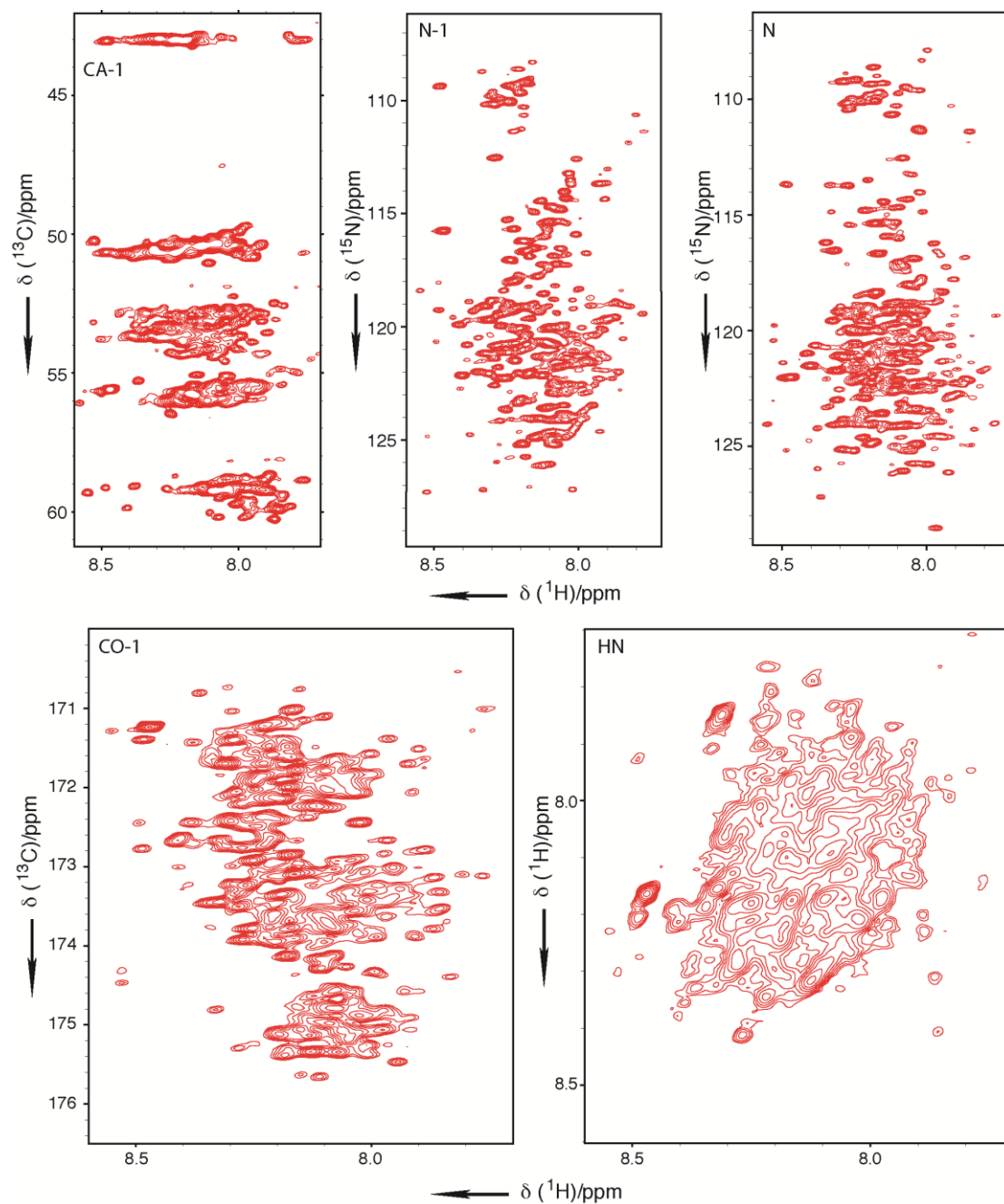
Figure 10. 2D projections of the 6D HNCOCANH APSY [121] of ncTom40 in 4M GdnSCN, 0.4% formic acid, pH 2.5 at 278K corresponding to $C_\alpha(i\text{-}1)\text{-}H_N(i\text{-}1)$, $N(i\text{-}1)\text{-}H_N(i\text{-}1)$, $N(i)\text{-}H_N(i\text{-}1)$, $C'(i\text{-}1)\text{-}H_N(i\text{-}1)$, $H_N(i\text{-}1)\text{-}H_N(i\text{-}1)$ planes.

The experiments used and final results at each temperature are listed in Table 7. The most complete assignment was obtained at 310K due to the decrease in relaxation rates at high temperature (see Table 7). 7D HNCO(CA)CBCANH provides sequential connection of i and i-1 H-N group and also Cβ chemical shift for discriminating different residue type. At 295K and 278K, 6D HNCOCANH instead of 7D HNCO(CA)CBCANH was recorded because of better sensitivity due to shorter

transfer pathway. The $^{13}C_\beta$ chemical shifts were achieved upon matching peak list of 6D HNCOCANH and 5D CBCACONH.

The assignment at higher temperature could assist the resonance assignment at lower temperatures. By automatic matching of six-dimensional peak lists from high and low temperature in combination with the sequential connectivity, 268 reliable and 34 low reliable sequence-specific resonance assignments were obtained. Manual inspection of a 3D HNN spectrum further improved sequential-specific resonance assignments and ^{15}N and ^{1}H resonance was assigned for 326 of 339 non-proline residues of ncTom40 (see Appendix 2).

Table 7: APSY experiments recorded at different temperatures and assignments obtained for denatured ncTom40 (339 non-proline residues) by MARS [118]. Assignments classified by MARS as low are not reliable and were excluded from further analysis.

temperature	Experiments	Reliability of assignments	
		reliable	low reliable
310K	7D HNCO(CA)CBCANH	312	11
	5D CBCACONH		
295K	6D HNCOCANH	249	26
	5D CBCACONH		
278K	6D HNCOCANH	128	58
	5D CBCACONH		
278K	APSY experiments at 3 temperatures, 3D HNN experiment, selectively labeled samples	326	

3.1.3   Secondary structure propensity for ncTom40 in GdnSCN

Proteins unfolded in denaturing solvent consist of residual secondary structure which can be characterized by secondary chemical shift. Chemical shift assignment at 295K was exploited for calculation of secondary chemical shift and secondary structure propensity (SSP). Both secondary chemical shift (Figure 11, upper panel) and secondary structure propensity(SSP) (Figure 11, lower panel) showed similar results, that residual secondary structure clusters exist in ncTom40 in 4M GdnSCN , but most secondary chemical shift value is below 0.5ppm (Figure 11, upper panel) indicating there is no rigid secondary structure.
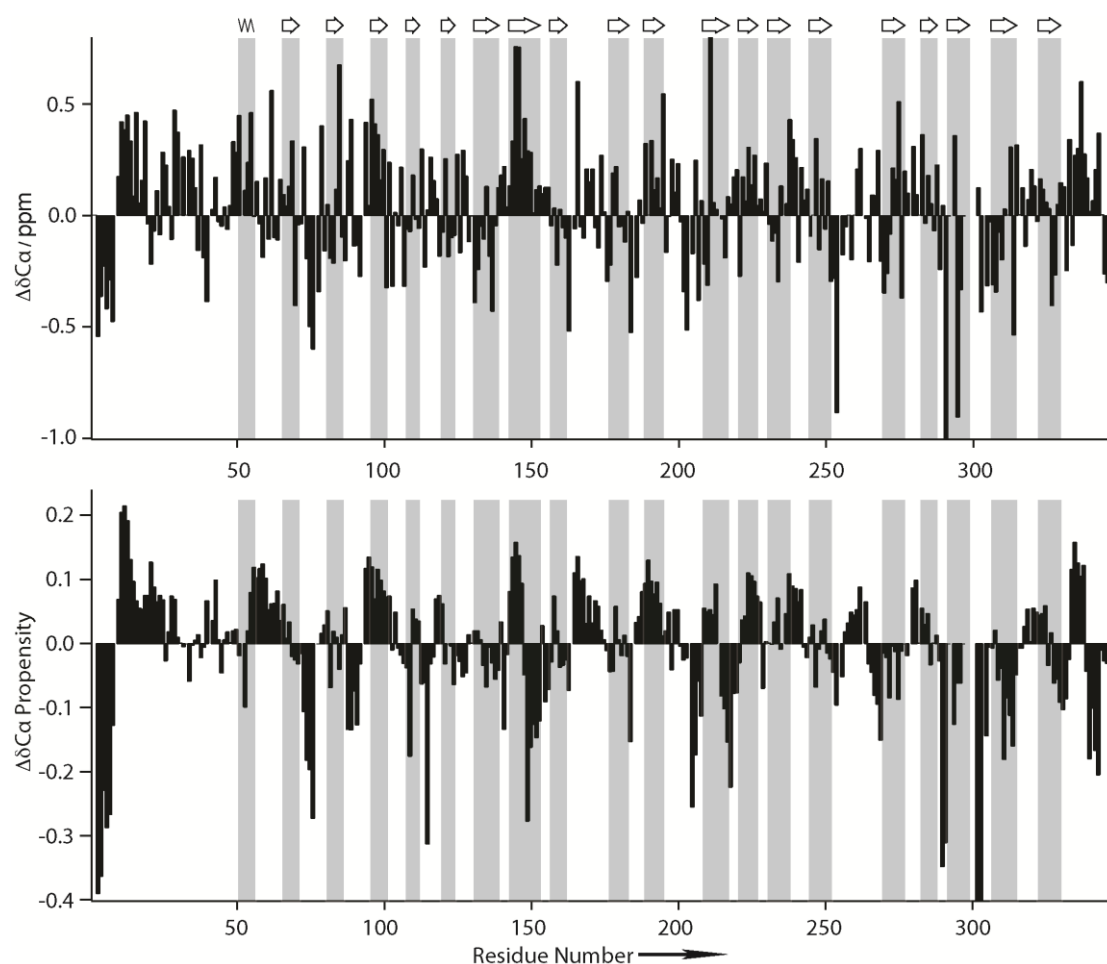
Figure 11.   Cα secondary chemical shifts (upper chart) and Cα secondary structure propensities (lower chart) of ncTom40 using the assignment obtained from APSY experiments recorded at 295K. Secondary structure propensities were calculated using SSP [123]. The predicted topology of ncTom40 is shown on top with secondary structure elements highlighted in grey. Only assignments classified by MARS [118] as reliable were used.

3.1.4   Structural characterization of Tom40 by hydrogen/deuterium exchange

3.1.4.1   Amino acid selective ^{15}N- *labeling to relieve the spectral overlap*
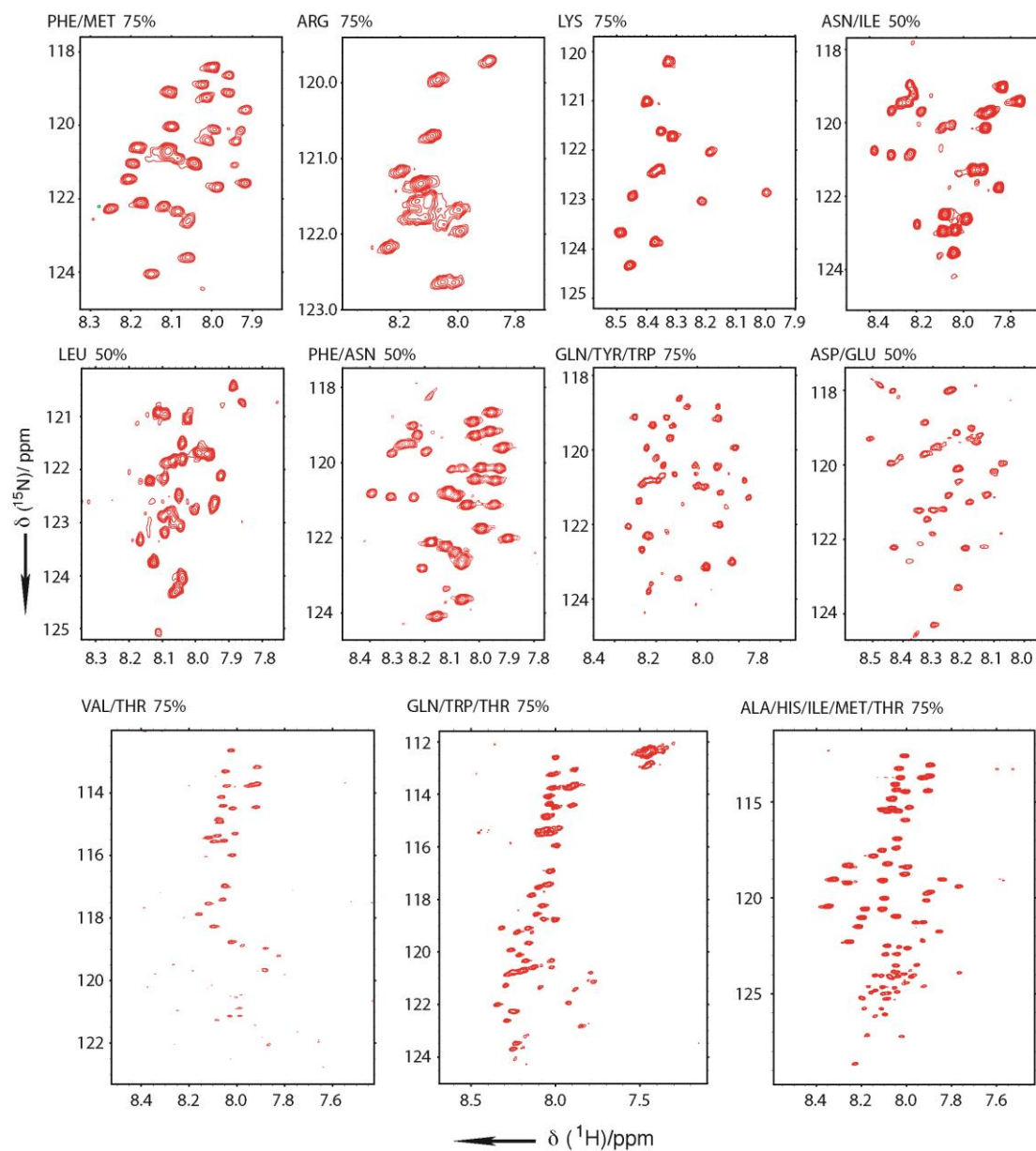
Figure 12. [^{1}H,^{15}N]-HSQC spectra of ncTom40 with amino acid selective ^{15}N-labeling at 278K, pD 2.5. Shown are the spectra at the end of back-exchange, i.e. time point 15 hr, in dissolution buffer containing 75% (PHE/MET, ARG, LYS, GLN/TYR/TRP, VAL/THR, GLN/TRP/THR, ALA/HIS/ILE/MET/THR) or 50% D2O (ASN/ILE, PHE/ASN,LEU,ASP/GLU).

349-residue ncTom40 fully denatured in 4M GdnSCN exhibits severe signal overlap on the two-dimensional [^{1}H, ^{15}N]-HSQC (Figure 9). To be able to analyze the

back-exchange curves for a large number of residues, a set of samples with only specific amino acid types ^{15}N-labeled (Table 3) were prepared. The combination of amino acid types in each sample was chosen to minimize NMR signal overlap (Figure 12).

3.1.4.2   *Optimization of experimental conditions*

The H/D exchange experiments were performed at 5 °C to lower the intrinsic exchange rate of amide protons. Another parameter which could affect the amide proton exchange rate is *p*H of dissociation buffer. Initially H/D exchange was applied in a buffer of 50% D2O and *p*D 2.5, Hydrophobic residue Leu18 has clear increase tendency as shown in the Figure 13. However, the polar residue ASN16 shows a constant signal intensity suggesting the intrinsic exchange rate of asparagine residues is so fast that the exchange is completed within the dead time of the experiment. As different amino acid types have their minimal exchange rates at different *p*H, we also tested back-exchange at *p*D 3.6 (Figure 13a). The intensity change of Asn16 during back exchange was hardly observable suggesting amide proton exchange at *p*D 3.6 is still too fast to characterize hydrogen kinetics by NMR.

The ability to distinguish between residues protected and exposed to water exchange in the native state also depends on the concentration of D_2O used during dissolution. For example, if the D_2O concentration is 75% and the exchange rate is such that at the time of dissolution the average population of the amide proton is already 25% no further back-exchange will occur. Therefore, we tested dissolution buffers with 50%, 75% and 100% D_2O (Figure 13b). In 50% D_2O, the signal intensity of Leu10 increases, while Leu307 is largely constant. In contrast, when using 75% D_2O the signal intensity of Leu307 decays, while Leu10 still increases. Thus, different D_2O concentrations allow a robust discrimination between protected (Leu307) and exposed (Leu10) residues on the basis of decreasing and increasing back-exchange curves, respectively. Strong support for the identification of protected residues also comes from back-exchange curves in 100% D_2O: only protected residues show a rapid time-dependent decrease in signal intensity (Figure 13).
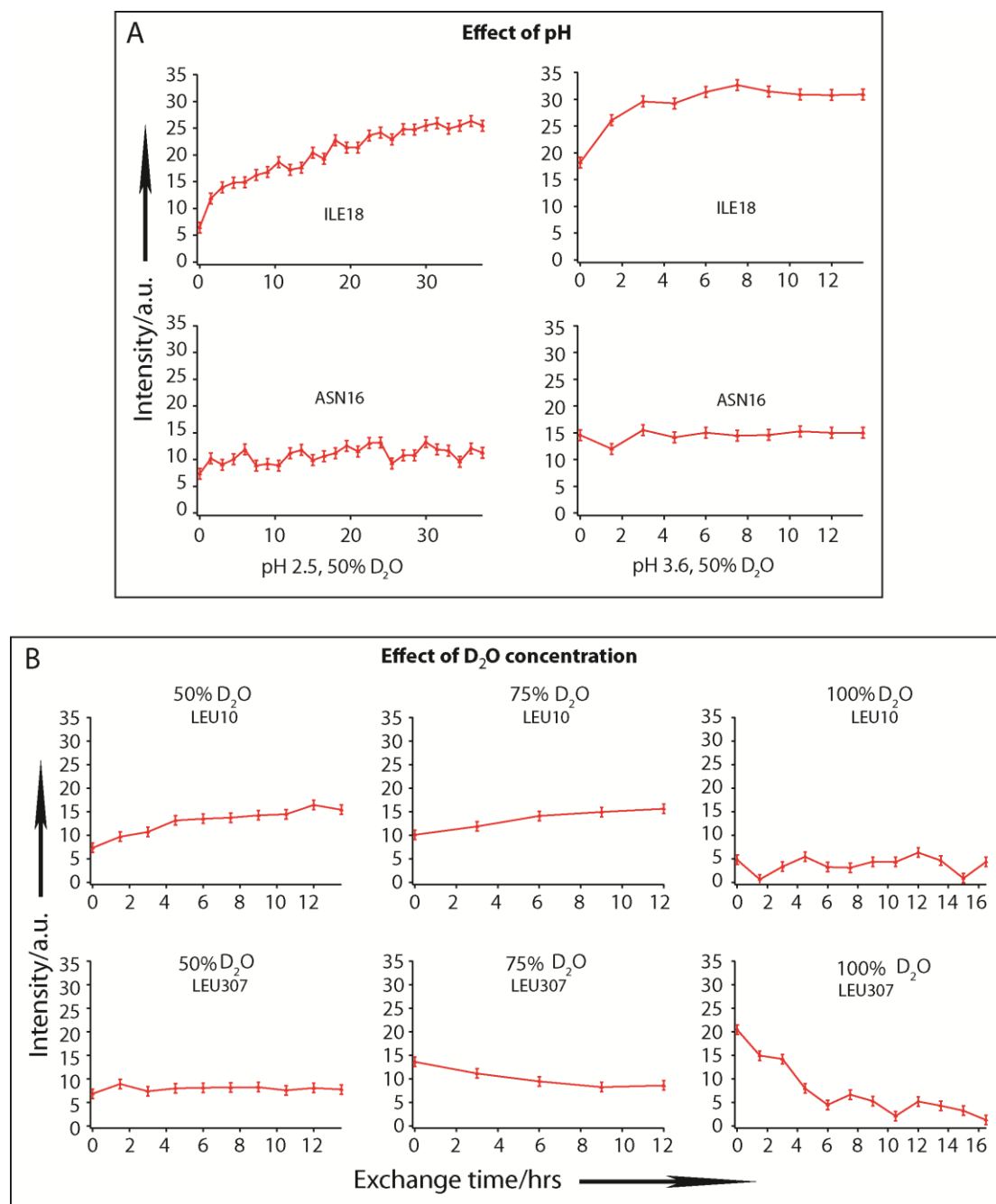
Figure 13. Influence of pH (upper panels) and D₂O concentration (lower panels) in the dissolution buffer on H/D back-exchange curves. For better comparison, signal intensities were normalized on the basis of the noise level in the spectra. Error bars are based on the signal-to-noise ratio.

3.1.4.3   Residue-specific solvent protection

The H/D exchange protocol was applied on a series of selectively ^{15}N labeled samples at pD 2.5 and 50%, 75% and 100% D_2O. Figure 14A shows the isoleucine region of six [^1H, ^{15}N]-HSQC spectra of a selectively ALA/HIS/ILE/MET/THR ^{15}N-labeled sample of ncTom40 in 75% D_2O dissolution buffer, taken every three hours. According to the design of the experiment, residues which are solvent accessible in liposome lose proton during forward exchange and start with a low proton incorporation at first [^1H,^{15}N]-HSQC. Then during the back exchange, these residues will gain protons from dissolution buffer and results in increase of signal intensity in [^1H, ^{15}N]-HSQC (Ile44, Ile47, Ile347). On the contrary, the residues protected from water in liposome barely lose proton in liposome and result in decrease of signal intensities in [^1H, ^{15}N]-HSQC during back exchange (eg. Ile105, Ile137, Ile286, Ile328). Signal intensity changes at different exchange time reveal distinct water accessibilities of residues which are related to their secondary structures. The intensity curve with rising tendency represents the exposed residues and the one with descending tendency is corresponding to the residues protected by hydrogen bonds. (Figure 14B).
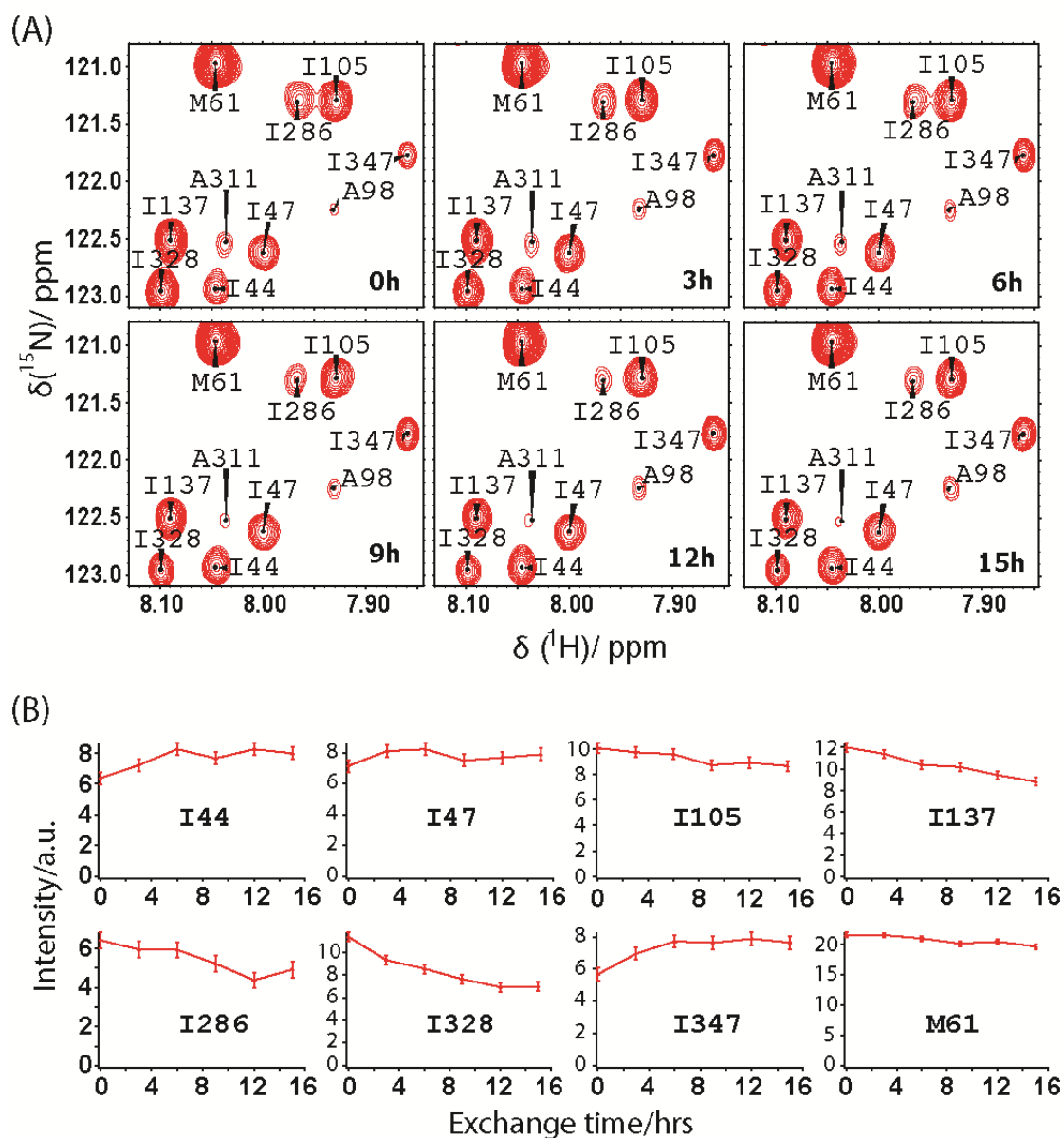
Figure 14. NMR signal intensity changes during back-exchange in 75% D_2O buffer. A) Enlarged spectral regions of [^1H,^{15}N]-HSQC spectra at increasing back-exchange times. To reduce signal overlap, ncTom40 was selectively ^{15}N-labeled at ALA, HIS, ILE, MET, THR. Time points indicate the time after start of the first HSQC. Sequence-specific resonance assignments are indicated. B) Residue specific H/D exchange profiles of residues in panel A. The peak intensities on [^1H,^{15}N]-HSQC are plotted as a function of exchange time. Error bars are based on signal-to-noise.

H/D exchange was also applied on the same selectively ALA/HIS/ILE/ MET/THR ^{15}N-labeled sample in 100% D_2O buffer D_2O (Figure 15). The peak intensities on [^1H,^{15}N]-HSQC are plotted as a function of exchange time. Ile105,

Ile137, Ile286 and Ile328 undergo rapid signal decay in dissolution buffer indicating these residues are protected in liposome while signals of Ile44, Ile47 and Ile347 decay slowly or remain constant indicating their exposure to D_2O in liposome. This is in line with identification from H/D exchange in 75% D_2O.
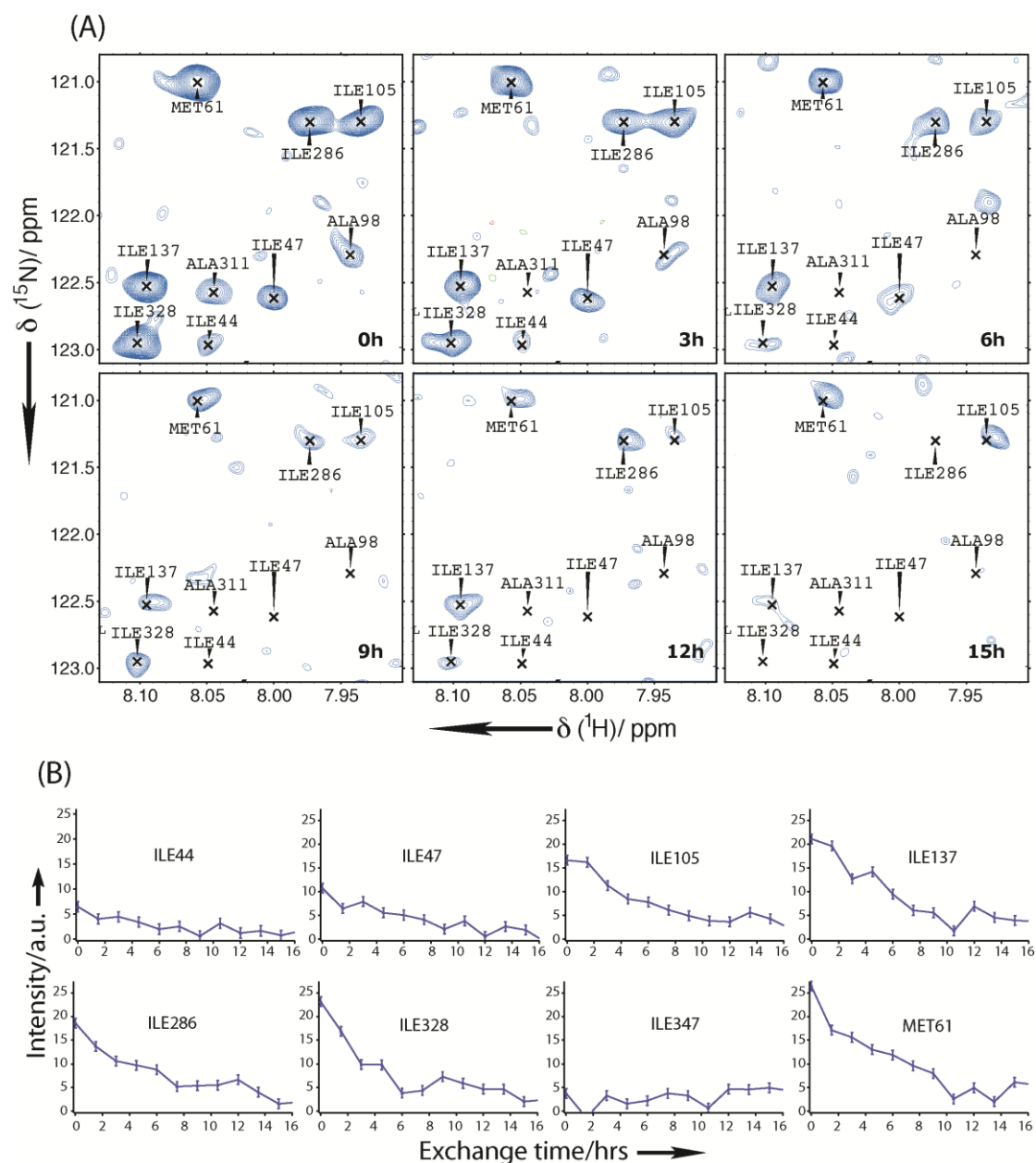
Figure 15. NMR signal intensity changes during back-exchange in 100% D_2O buffer. a) Enlarged spectral regions of [^1H,^{15}N]-HSQC spectra at increasing back-exchange times. Time points indicate the time after start of the first HSQC. Sequence-specific resonance assignments are indicated. b) Residue specific H/D exchange profiles of residues in panel A. The peak intensities on [^1H,^{15}N]-HSQC are plotted as a function of exchange time. Error bars are based on signal-to-noise.

H/D exchange protection of single residues was quantified according to $I_p =$ $(I_{9hrs} + I_{12hrs})/(I_{0hr} + I_{3hrs})$, with I_{0hr}, I_{3hrs}, I_{9hrs} and I_{12hrs} intensities of a cross-peak in the [^{1}H,^{15}N]-HSQC at 0 hour, 3 hours, 9 hours and 12 hours, respectively. Averaging over the first two and last two time points reduces the influence of slight variations, such as errors due to low signal-to-noise and peak picking, and thereby increases the robustness of the approach. A protection map is then obtained by plotting I_p values as a function of residue number (Figure 16).
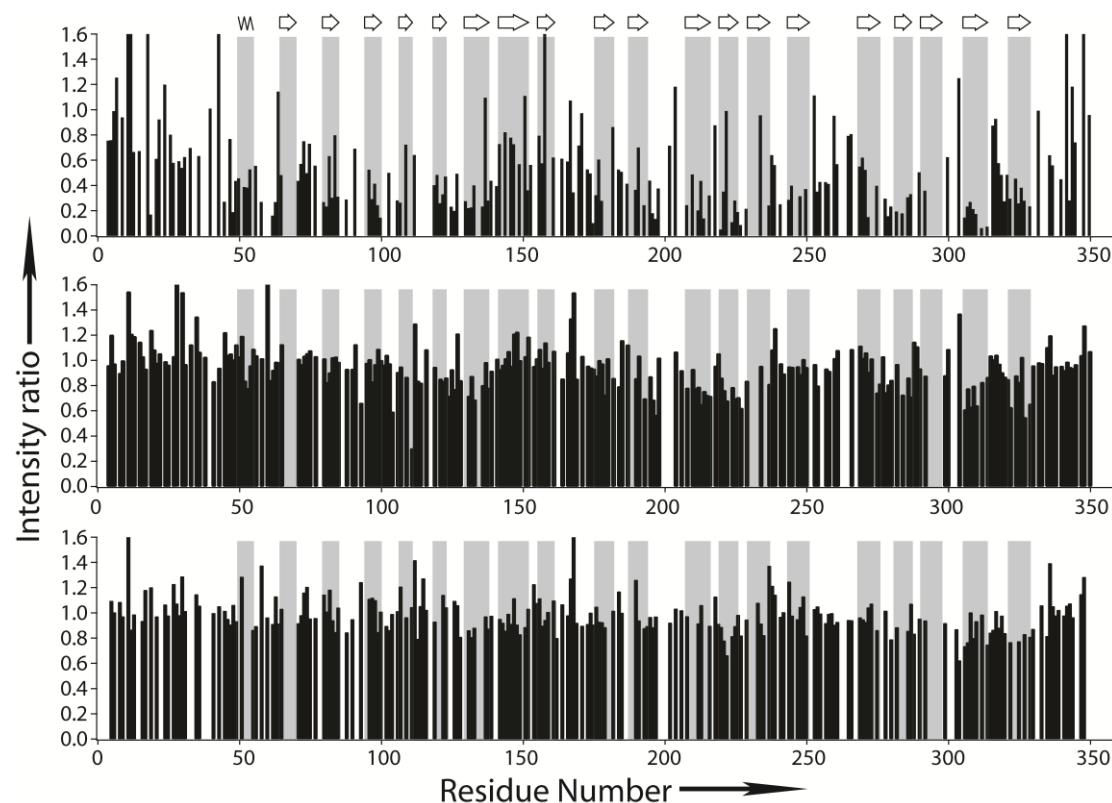
Figure16. Averaged intensity ratios ($I_p = (I_{9hrs} + I_{12hrs})/(I_{0hrs} + I_{3hrs})$) calculated from back-exchange curves in 100% D_2O (upper panel), 75% D_2O (middle panel) and 50% D_2O(lower panel). The location of secondary structure elements (α-helix (wavy line) and β-strands (arrows)) as predicted by iTasser is shown above.

The 100% D_2O map reveals large variations in intensity ratios. The first ~40 residues and last ~10 residues had close to zero signal intensities already in the first HSQC and therefore high I_p values, indicating that they are solvent exposed(Figure 16, upper panel). In addition, several residues from 50 to 330 rapidly exchanged with solvent during the D_2O incubation period, in line with a location of these residues in loops. In between these residues, there are short segments, where the averaged intensity ratio approaches zero, i.e. where the signal

intensity was large in the first HSQC and then decayed into the noise. The distinction between exposed and protected residues is further supported by the 75% D_2O map (middle panel in Figure 16) and 50% D_2O map (lower panel of Figure 16). Thus, a large number of residues that are protected from H/D exchange in the native structure of ncTom40 can be identified. Notably, the fact that Tom40 forms a water-filled channel [124] suggests that it is not primarily the membrane insertion that leads to reduced solvent exchange rates, but rather the presence of hydrogen bonds.

3.1.4.4   Structural elements identified by H/D exchange

The homology of ncTom40 with hVDAC1 was used to predict the location of secondary structure elements. iTasser [125] predicted 19 β-strands and an N-terminal α-helix in ncTom40 (Figure 17). Based on the experimental I_p values in 100% D_2O (Figure 17A), we can separate the residues into 2 group by setting threshold for protect fact Ip. Residues with $I_p > 0.5$ are classified as exposed residues because these residues undergo fast amide proton exchange with D_2O in native state during forward exchange, while the residue with $I_p < 0.5$ are classified as protected residues as these residues are protected by hydrogen bond in native state during forward exchange. These two group residues with different water accessibility are highlighted on the topological model: residues with $I_p > 0.5$ are shown in red, those with $I_p < 0.5$ in blue. The uncolored residues are not analyzed due to lack of assignment or signal overlap.

The color-coded topology map highlights the distinctive nature of the N- and C-terminal tail when compared to the core of ncTom40 (Figure 17A). Up to Gly42 and from Ser331 to Phe349 there are only two residues (Ile18 and Gln339), which are protected from exchange. Inspection of the topology map, which was color-coded using the 75% D_2O intensity ratio plot (Figure 17B), demonstrates that the solvent protection of both these residues is weak. In addition, the 75% D_2O intensity ratios point to variations in solvent exchange rates at the C-terminus of ncTom40 (Figure 17B). The core of ncTom40 is comprised of continuous segments of solvent protected and exposed residues. For example, several residues in the predicted loops between strands 8-9, 10-12, 15-16 and 18-19 have high I_p values. On the other, residues

Gln141-Lys160 also rapidly exchange with solvent, despite being predicted to form β-strands 7 and 8 (Figure 17A). Thus, either these two β-strands are not formed or they are less stable than the other strands. Alternatively, the increased exchange rates in this region might be due to the presence of two histidine residues (His148, His150). For some of the predicted β-strands the location of solvent protected residues appears to be shifted. For example, β-strand 19 harbours the membrane insertion signal of Tom40 [39]. An important question is also if an N-terminal α-helix is present in ncTom40 that could regulate the transport of proteins through the pore. The slow H/D exchange rates next to Ile47-Gln52 support the presence of the α-helix (Figure 17A). Although the H/D exchange pattern does not permit a definite statement about the number of β-strands, it provides for the first time residue-specific experimental information about the structural properties of membrane-embedded ncTom40.
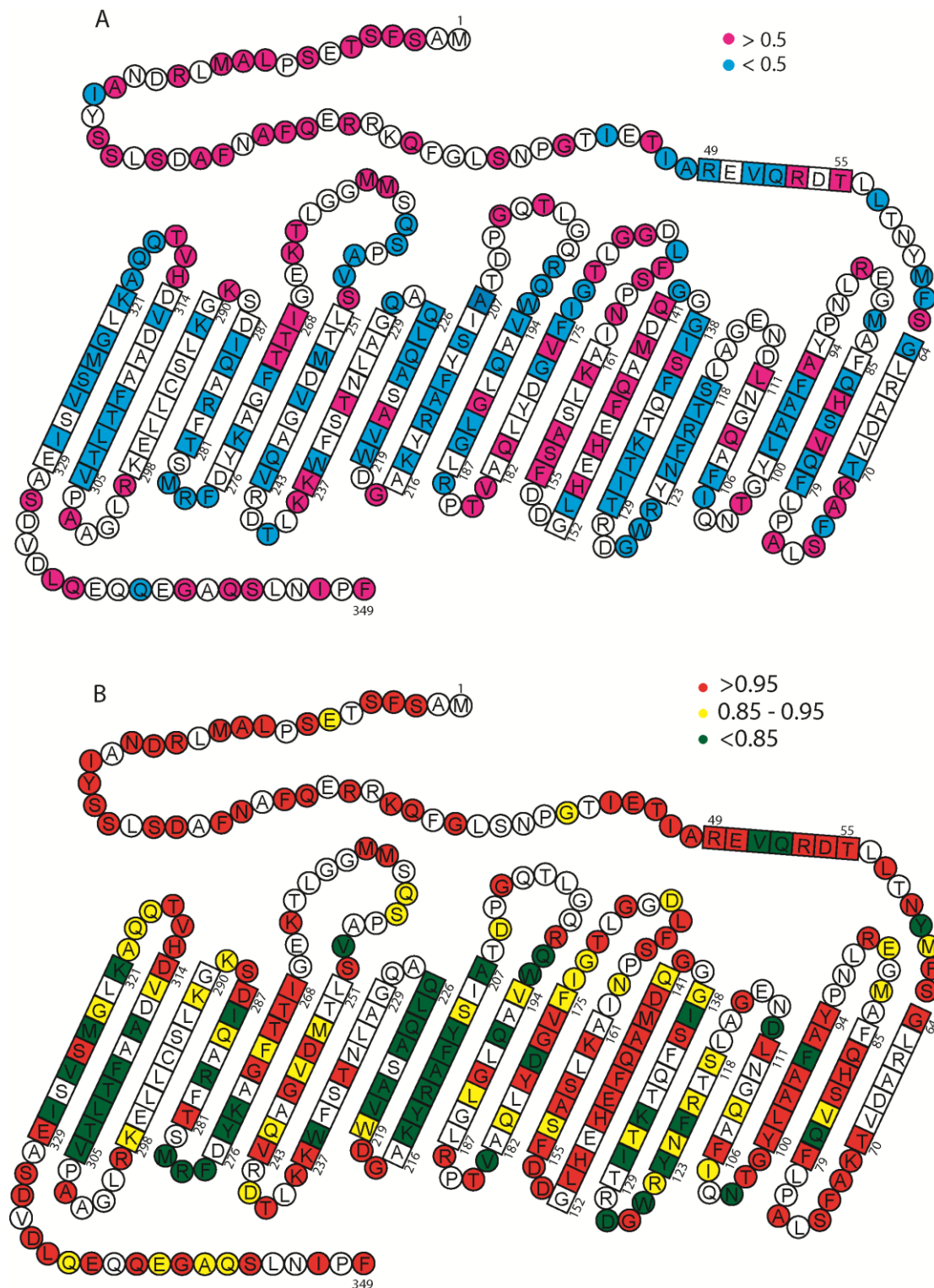
Figure 17. Intensity ratios ($I_p = (I_{9hrs} + I_{12hrs})/(I_{0hrs} + I_{3hrs})$) in 100% $D_2O(A)$ and 75% $D_2O(B)$ dissolution buffer were mapped onto the predicted topology model of ncTom40. Residues predicted to be in a β-strand or α-helix are boxed. Non-assigned or non-resolved residues are shown white. (A) Magenta-shaded (Cyan-shaded) residues have I_p values larger (lower) than 0.5. (B) Red-shaded residues have intensity ratios larger than 0.95, yellow ones between 0.85 and 0.95 and residues with I_p < 0.85 are colored green.

3.2   Recognition of Presequence by Tom40

Precursor proteins are usually transported through the outer membrane of mitochondria by interaction between presequence and TOM complex. The amphipathic α-helix in presequence is recognized by Tom20 and Tom22. The interaction between Tom40 and presequence was identified on the *cis* as well as *trans* side of the outer membrane [126, 127]. Presequence peptide derived from aldehyde dehydrogenase (ALDH) has been utilized for investigation of Tom20-preseqeunce interaction [14, 128]. This motivated us to investigate the interaction between Tom40 and presequence using NMR.

3.2.1   Secondary structure propensity of ncTom40 N-terminal peptide

The H/D exchange result and predicted topological model in section 3.1 suggests the β-barrel start after first 60 residues. Therefore an ncTom40 N-terminal peptide (1-59) was synthesized. The deviation from random coil values for C_α and C_β chemical shifts [129, 130] is common indicators for residual secondary structure elements. A complete backbone assignment of N-terminal peptide (1-59) of ncTom40 was obtained by 3D triple resonance experiments non-uniform sampling (NUS) HNCACB and NUS CBCACONH and conventional CBCANH. Non-uniform sampling is a special chemical shift labeling strategy in indirect dimension to obtain high resolution in short time for multidimensional NMR (see section 5.4). Secondary chemical shift C_α and C_β of N-terminal Tom40 were calculated (Figure 18) and Residues Ser23-Gln35 show a positive value for Cα and negative value for C_β indicating the α-helix propensity at that region. However, compared with $\Delta\delta(C_\alpha)$ around 3ppm for fully α-helix, the $\Delta\delta(C\alpha)$ below 1.5ppm in the region of Ser23-Gln35 indicates the secondary structural element is not rigid. Quantitative analysis reveals 40% α-helical propensity for Ser23-Gln35.
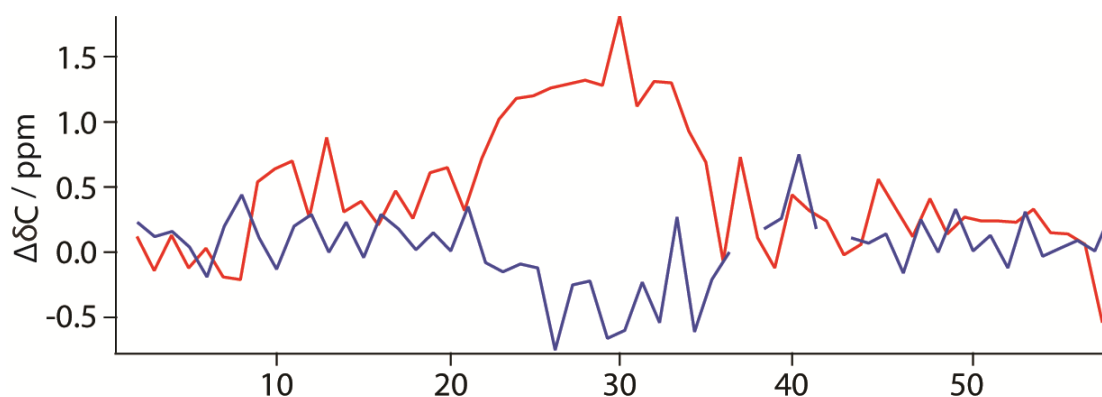
Figure 18.   Cα (red) and Cβ(blue) secondary chemical shift of N-terminal ncTom40 (1-59) as a function of residue number.

3.2.2   Interaction of Tom40 with rALDH presequence

To map the binding site between ncTom40 N-terminus and presequence, N-terminal peptide of ncTom40 was titrated with rALDH presequence and monitored by [^{1}H,^{15}N]-HSQC. Comparison of [^{1}H,^{15}N]-HSQC spectra of N-terminal Tom40 in free form and in the presence of 20 fold excess of rALDH presequence showed some peaks undergoing chemical shift changes. Also the bound form of Tom40 N-terminus didn't show additional peaks at high concentration of rALDH. This is indicative that the interaction occurs at the fast to intermediate exchange regime at NMR time scale. The chemical shift perturbation was plotted along the sequence number of N-terminal Tom40 demonstrating a binding site Ser20-Gly37. Residue Leu22, Phe29, Gln30 and Gln35 were involved in peak overlap and their chemical shift changes were not analysed. The binding site overlaps with transient α-helix region Ser23-Gln35 determined by secondary chemical shift (Figure 18). It implies that Tom40 binds to presequence upon helix-helix interaction.

A)

B)
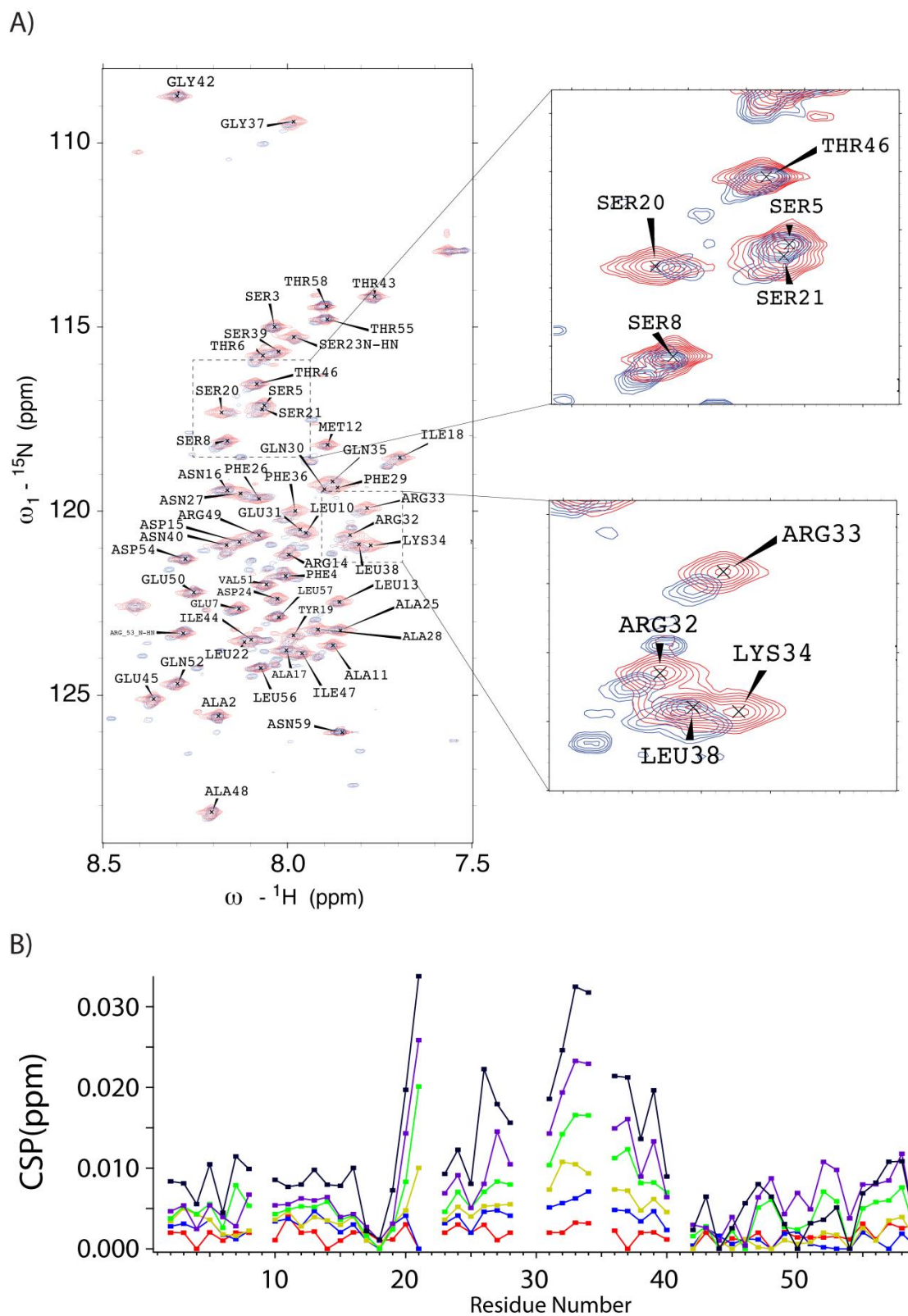
Figure 19. Titration of rALDH into N-terminal Tom40. (A)Spectra overlay of Tom40 without (red) and with (blue) 20 fold access rALDH. Right two are enlarged display of the selective region. (B) Averaged ^1H, ^{15}N chemical shift perturbation upon addition of rALDH at molar ratios(Tom40 N-terminal:rALDH) of 1:0.5 (red), 1:1(blue),1:2(yellow), 1:5(green),1:10(purple), 1:20(black).

3.2.3 TFE induces helix-helix interaction between N-terminal Tom40 and presequence

We presumed that the weak interaction between N-terminal Tom40 and presequence rALDH probably results from unstable α-helix. TFE (2,2,2-Trifluoroethanol) is a chemical compound being capable of increasing the helical stability of peptides to make them usable as models of helices in proteins. The N-terminal peptide of ncTom40 was dissolved in 30% TFE in order to promote α-helical conformation and strengthen the helix-helix interaction. A complete sequence-specific assignment of Tom40 N-terminal were obtained using CBCANH and CBCA(CO)NH. TFE caused a general increase in content of α-helical conformation especially at C-terminal Thr43-Thr55. The previous α-helix region Ser23-Gln35 without TFE is also extended to around Leu10-Phe36 (Figure 20). Cα secondary chemical shift reaching 2ppm also indicates more rigid α-helical conformation induced in TFE than in water.
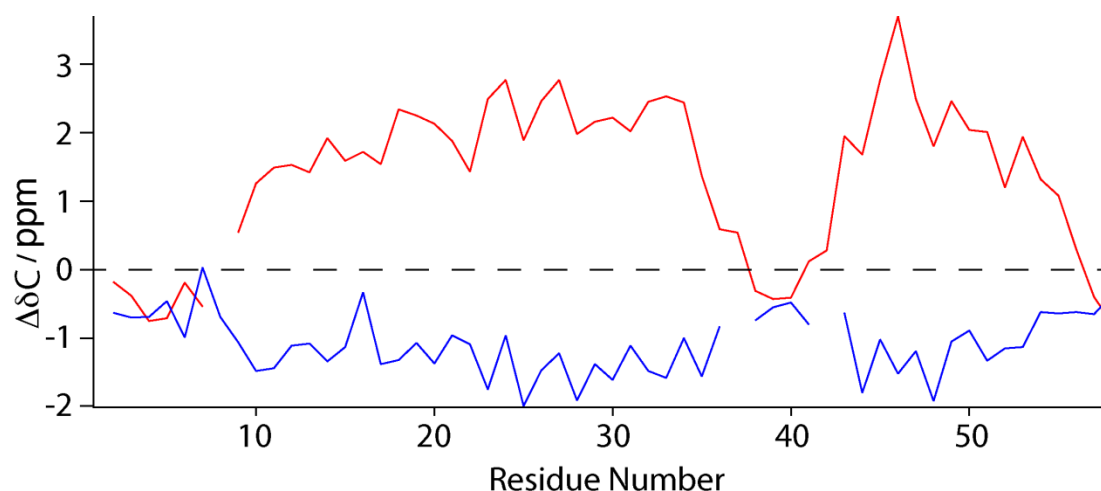
Figure 20.   Cα (red) and Cβ (blue) secondary chemical shifts of N-terminal ncTom40 (1-59) in the presence of 30% TFE as a function of residue number.

Similarly, in the presence of 30% TFE the interaction of N-terminal Tom40 with presequence was investigated by titration with rALDH presequence. Large peak shifts could be observed for some residues such as Thr43 and Leu56. The chemical shift perturbation along sequence numbers highlighted a cluster Thr43-Thr58 binding to rALDH. In contrary to the interaction in the absence of TFE, the binding site shifts

to C-terminus. Additionally, magnitude of the chemical shifts 0.1ppm was observed suggesting a stronger binding induced by TFE.

(A)

(B)
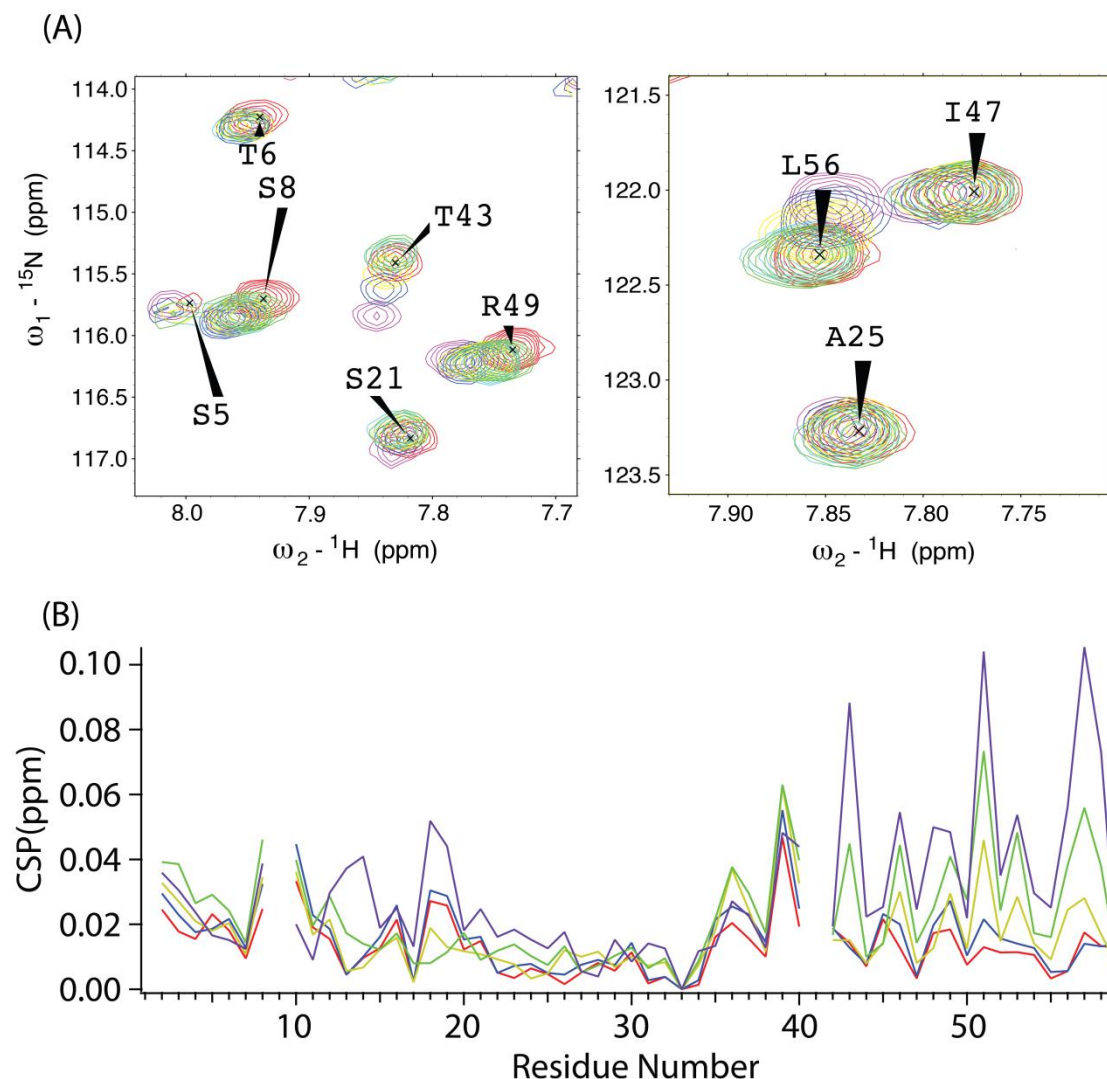
Figure 21. Titration of rALDH into N-terminal Tom40 with 30% TFE. (A)Superposition of Tom40 without(red) and with rALDH at molar ratio(Tom40 N-terminus:rALDH) of 1:1(cyan), 1:2(yellow), 1:5(blue), 1:10(green), 1:20(purple). Averaged ^1H, ^{15}N chemical shift perturbation upon addition of rALDH at molar ratios(Tom40 N-terminus:rALDH) of 1:1(red), 1:2(cyan), 1:5(yellow), 1:10(green), 1:20(blue) in 30% TFE.

3.3   Lipid-dependent folding of ncTom40

Membrane protein solubilized in denaturing solvent often can refold into detergent micelles, bicelles or membranes autonomously [131]. The biophysical principles and mechanisms of membrane proteins folding and insertion into

biomembrane are not well understood at an atomic level. Proteins unfolded in denaturing solution may contain nonrandom peptide chains. Characterization of these hydrophobic clusters could provide insight into the protein folding process at early stage. Hiller et. al. have identified two hydrophobic clusters of E.coli outer membrane proteins X(OmpX) unfolded in urea binding to membrane mimetic micelle. It suggests these two clusters are the initial binding site with the ordered lipids [132]. We, therefore, applied similar strategy on ncTom40 to investigate its folding and insertion into membrane.

3.3.1   Secondary structure propensity of ncTom40 in urea

The secondary chemical shifts of ncTom40 in GdnSCN indicate the presence of residual secondary structure in denatured Tom40. To study the folding process of Tom40 at atomic resolution and elucidate the structural basis for membrane insertion, Unfolded Tom40 was solubilized in 8M urea. Sequence-specific resonance assignment was obtained for 277 residues of ncTom40 at 298K using 6D HNCOCANH and 5D CBCACONH APSY combined with automatic assignment. Typically, for same structural element, secondary chemical shifts $\Delta\delta C\alpha$ and $\Delta\delta C\beta$ exhibit values with opposite sign while $\Delta\delta C\alpha$ and $\Delta\delta C'$ provide the values with same sign. Therefore, Combined secondary chemical shifts of $\Delta\delta C\alpha + C'$ and $\Delta\delta C\alpha - C\beta$ serve as a better structural indicator. Small secondary chemical shift indicates the absence of rigid secondary structural elements (Figure 22). However several continuous stretches with negative values were observed, indicative of a propensity to adopt β-sheet structure, such as Asp54-Leu65, Phe79-Val81, Tyr100-Asp112, Leu171-Val183, Gln198-Ser209, Glu266-A273, Phe304- Gln313, Leu322-Val333.
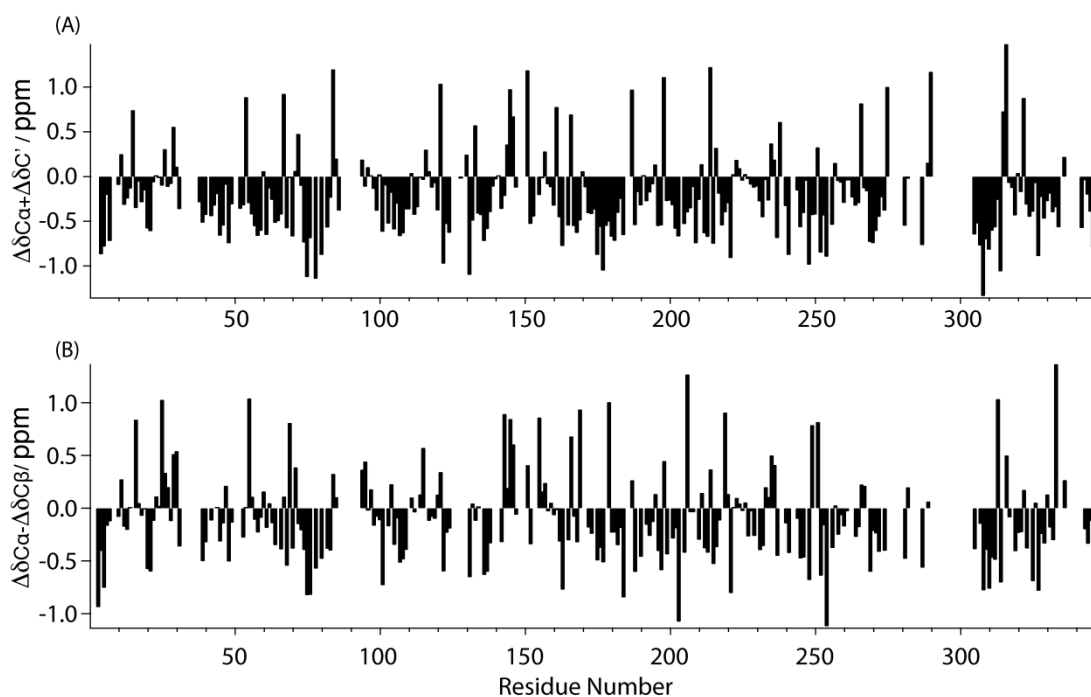
Figure 22: Secondary chemical shifts on Tom40 in 8M urea. (A) ΔδCα + ΔδC' (B) ΔδCα+ΔδCβ

3.3.2   Interaction of ncTom40 with micelle in urea

The interaction of ncTom40 with membranes was investigated by titration of the detergent dihexanoylphosphatidylcholine (DHPC) to urea-unfolded ncTom40. Detergent molecules only assemble into micelle at concentration above critical micellar concentration (cmc). The stepwise addition of DHPC from 25mM to 200mM were applied to ncTom40 in urea and the chemical shifts changes were monitored by [^{1}H-^{15}N]-HSQCs. To relieve the spectral overlap of urea-unfolded ncTom40, a set of selectively ^{15}N labeled Tom40 sample were utilized including GLN/TYR, LYS, ARG, ALA/HIS/ILE/MET/THR, ASN/PHE, ASP/TYR ^{15}N labeled samples. The observance of chemical shift changes of Gly64, Gly217 and Gly246 is a characteristic of presence of fast to intermediate exchange at NMR time scale (Figure 23), whereas some other residues such as Gly177 and Gly42 stay same upon titration indicating no chemical exchange or slow exchange at NMR time scale. The significant chemical shift perturbations were only observed upon addition of DHPC above 50mM which is close to the critical micelle concentration (cmc) of 45mM.
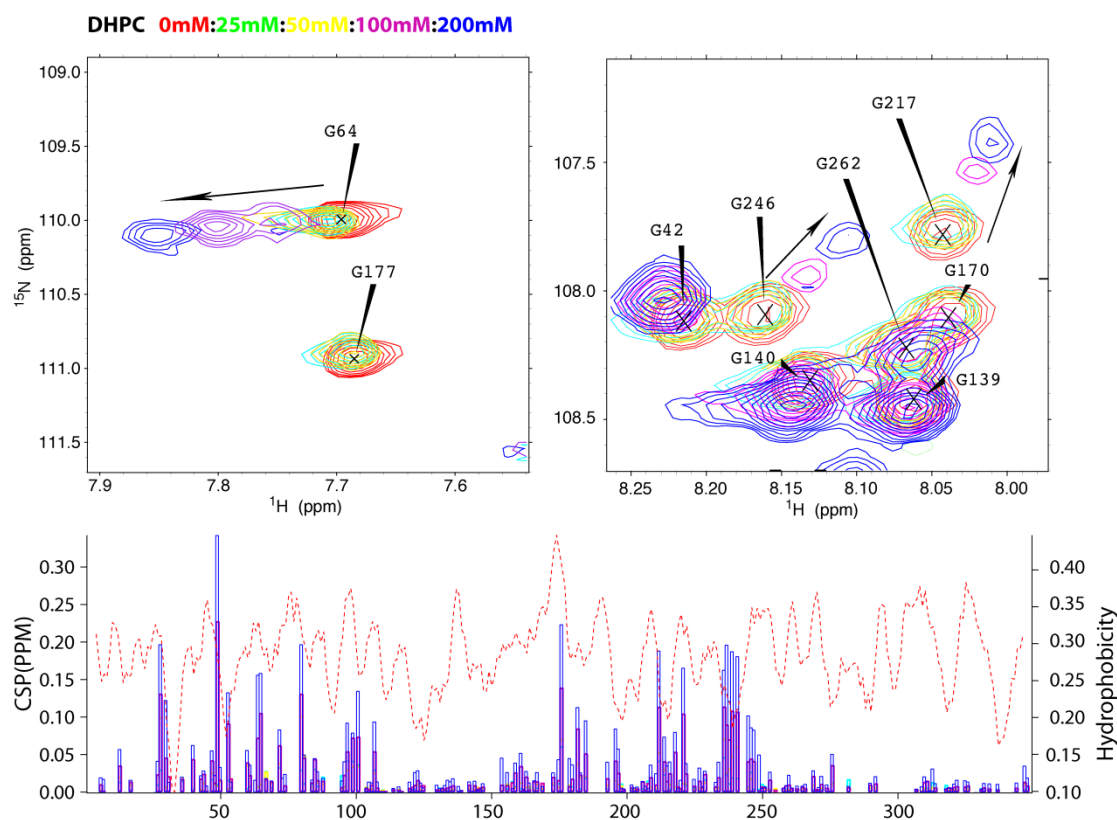
Figure 23: Urea-unfolded Tom40 interacts with DHPC micelles. Upper panel: Superposition of 2D ^1H-^{15}N HSQC spectra of uniform ^{15}N labeled ncTom40 with DHPC concentration at 0mM(red), 25mM(cyan), 50mM(yellow), 100mM (magenta) and 200mM(blue). Lower panel: Chemical shift perturbations(CSP) as a function of residue number of Tom40 upon addition of 25mM(cyan), 50mM(yellow), 100mM(magenta) and 200mM(blue) DHPC. The red dashed line indicates the predicted hydrophobicity for each residue.

After applying titration on all selectively and uniform ^{15}N labeled samples, chemical shift perturbation was analyzed and plotted in function of residue number (Figure 23). The predicted hydrophobicity was also plotted as red dashed line. The signal overlap is severe on the spectrum of ^{15}N uniform labeled sample and the combination of selective ^{15}N labeled sample is not sufficient, thus chemical shift perturbation for many residues are missing. Nevertheless, we are still able to observe some binding sites such as Ala95-Gly101, Arg213-Ala221 and Phe235-Gly246. The last β-strand has been proposed to contain β-signal which regulates the insertion of Tom40 precursor protein into membrane environment. However, residues Lys321, Met324, Ile328 located in predicted last β-sheet do not experience chemical shift perturbations thus suggesting there is no interaction of last β-strand with DHPC micelle.

4   Discussion

4.1   Hydrogen/Deuterium exchange characterizes transmembrane proteins structure

Structural investigation of transmembrane proteins is of great interest due to their important biological function. However, structural characterization of membrane proteins structure at atomic level with the help of available tools such as NMR and X ray crystallography is challenging mainly because the native structure of IMPs are only be able to be maintained in the hydrophobic environment of lipid or detergent. Investigation of membrane proteins by NMR spectroscopy is limited to small membrane proteins since sensitivity and resolution of NMR spectra on large protein-membrane complex is impaired by rapid resonance relaxation due to large molecular size.

Here, we present a novel Hydrogen/Deuterium exchange protocol to characterize transmembrane proteins structures. The method has identified the long N-terminal and C-terminal tails which have a distinctive water accessibility of the β-barrel core of Tom40. Additionally, we found that in the core of ncTom40, the β-strands and its flanking loops are also distinguishable from continuous segments of solvent protected and exposed residues (Figure17). However, the analysis of H/D exchange data may be complicated by some polar residues such as asparagines that undergo fast intrinsic amide proton exchange and may reach proton equilibrium quickly in back exchange buffer in short time. Moreover, the imidazole side chain of histidine may increase the proton exchange rate of residues in the vicinity. Residues which form low stable secondary structure may also have relatively high water accessibility and obscure the differentiation from residues in loop region.

Hydrogen/Deuterium exchange measurement is a potentially very useful structural method. Its application on membrane proteins is limited to small ones in micelles [114]. Dempsey and coworkers [133] have developed a clever "exchange-trapping" approach in which a bilayer-associated membrane proteins is exposed to

H/D exchange condition, followed by trapping of the in-bilayer amide exchange state via transfer of the membrane proteins out of the bilayer and into an organic solvent system in which the percentage of remaining protons at each amide site can be assessed by solution NMR methods. The wide application of this "exchange-trapping" method has encountered problems related to imperfections in the organic solvent system used to resolubilize the exchanged proteins for NMR analysis which may cause post-trap exchange in the solvent or conformation heterogeneity of the proteins [133, 134].

Our H/D exchange method enables structural investigation of membrane proteins reconstituted into liposome which provides the native-like environments for membrane proteins. Even for large proteins such as Tom40, the backbone assignment of denatured state proteins can be accomplished with the aid of APSY and HNN experiments. The application on the Tom40 has provided for the first time residue-specific experimental information about the structural properties of membrane-embedded ncTom40 although the number of β-strands can't be definitely stated. It is thus expected that this combined method of H/D exchange coupled with solution-state NMR spectroscopy is applicable to many membrane proteins reconstituted in liposomes and might allow insight into the structural properties of proteins in native membranes.

4.2   Structure of Tom40

CD and FTIR spectroscopy reveal that Tom40 consists of high content of β-sheet structure and a low α-helical portion. We predicted a topology model of ncTom40 based on structure of hVDAC1 indicating Tom40 folds into a 19 β-barrel and α-helix at N terminus. H/D exchange data have identified the core β-barrel region of ncTom40 and also several predicted 5-6, 10-12, 14, 16 and 18-19 in the topological model (Figure 17 and Figure 24). Some of the predicted β-strands appear to be shifted and the predicted model should be modified. The locations of some β-strands are not clear which may be due to weakly stable β-strands or some specific residues with fast intrinsic exchange rate.
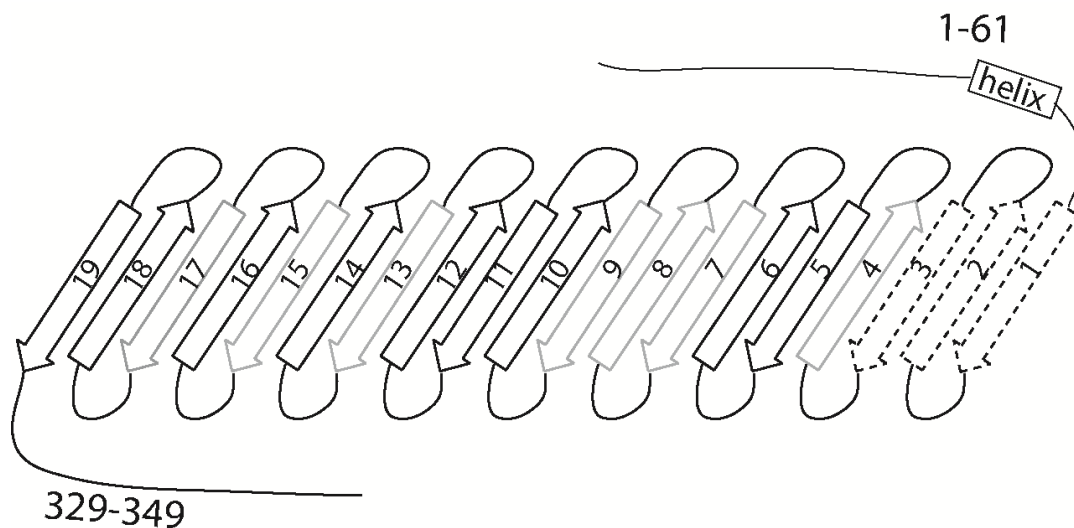
Figure 24. Structural elements of ncTom40 on topology. Predicted β-strands 5-6, 10-12, 14, 16 and 18-19 are identified from H/D exchange data (black arrows). β-strands 1-3 are weakly stable (dashed arrows). Other predicted β-strands is indicated as grey arrows. N-terminal tail, M1-M61, contains an α-helix. C-terminal tail is from E329 to the end.

Antibody labeling has revealed both the N-terminal and C-terminal of ncTom40 face the intermembrane space [135]. Sequence alignment of ncTom40 and hTom40 (see Appendix 4) indicates ncTom40 contains both flexible N-terminal and C-terminal tails while hTom40 only possesses a long proline-rich N-terminal tail. The truncations of α-helical region of hTom40 N-terminus have been shown to inhibit the activity of channel-gating of the pore [136]. Our hydrogen-deuterium exchange data has reveal flexible N-terminal tail containing around first 60 residues and C-terminal tail of ncTom40 from Glu329 to the end. It has been suggested conserved charged amino acid residues Glu45 and Arg49 of ncTom40 are involved in channel gating [136]. Our H/D exchange data in 100% D_2O suggests Arg49 to be located in the α-helix Ile47-Gln52.

4.3   Weakly stable β-strands in Tom40

It is interesting to note in our hydrogen/deuterium exchange experiment performed in 100% D2O, most residues in the first three predicted β-strands β1 – β3 are protected from water, indicative of presence of secondary structure. However, 75% D2O results exhibits these β-strands are not fully protected. Thus, collectively two data suggests that β-strands β1-β3 undergo conformation exchange forming transient hydrogen bonds. Such conformational instability has also been observed at N-terminal part of hVDAC1 [57]. The H/D exchange monitored by NMR spectroscopy showed rapid amide proton exchange in the first 4 β-strands of hVDAC1 indicating lower stability than other β-strands in the barrel. Also in the transmembrane domain of hTom40 three β-strands (1,2 and 9) were predicted as less energetically stable and the mutagenesis of predicted specific destabilizing amino acid residues enhance its resistance to thermal or chemical perturbation of the barrel [137]. Usually, the β-barrel proteins are fairly stable due to extensive hydrogen-bond network between the individual β-strands. These unstable β-strands might be important for their function and can be stabilized though interaction with other proteins or lipid [138]. It has been shown Tom40 proteins also can forms dimers or trimers in fungal species [53, 56]. Therefore, we suggest these three β-strands β1–β3 are important for forming the interface of Tom40 dimerization or interaction with other TOM subunits. This is in agreement with the results from Gessmann et. al [139].

4.4   Multiple interaction sites between Tom40 N-terminus and presequence

Preprotein translocation into mitochondria is mediated by interaction of presequence with various receptors along import pathway. Structure of Tom20-presequence complex has highlighted helical interaction at the interface [14]. Truncation of first 64 residues of ncTom40 results in defect for import of the matrix preproteins suggesting the N-terminal ncTom40 is involved in interaction with presequence [140].

Upon H/D exchange experiment, we reveal a region of consecutive protected residues Ile47-Gln52 indicating these residues may form an α-helix. This region is close to the predicted α-helix ARG49-The55. In our studies, a binding region Ser20-Gly37 of ncTom40 interaction with presequence was identified (Figure 19). The induction of α-helix using TFE shifts its binding site to C-terminal fragment (Thr43-Thr58) of the Tom40. In both cases, the interaction site with rALDH presequence overlaps with the region characterized by residual α-helix propensity. This indicates N-terminal Tom40 adopts α-helical conformation at the binding site with amphipathic presequence.

Combining all these data, we suggest there are more than one weak presequence binding sites present on Tom40. The multiple weak interaction sites for substrate binding might provide a simple means that Tom40 binds to unfolded segments in transit and subsequently hands over to the downstream components [141]. The combination of several binding sites could provide a sufficiently strong affinity with substrate proteins while for one individual interaction site, the binding is relative weak and reversible.

```
                10         20         30         40         50
1  MASFSTESPL AMLRDNAIYS SLSDAFNAFQ ERRKQFGLSN PGTIETIARE VQRDTLLTN
2  MASFSTESPL AMLRDNAIYS SLSDAFNAFQ ERRKQFGLSN PGTIETIARE VQRDTLLTN
3  MASFSTESPL AMLRDNAIYS SLSDAFNAFQ ERRKQFGLSN PGTIETIARE VQRDTLLTN
4  MASFSTESPL AMLRDNAIYS SLSDAFNAFQ ERRKQFGLSN PGTIETIARE VQRDTLLTN
```

Figure 25: α-helical region (red letters) of Tom40 identified by iTasser model (row 1), H/D exchange (row 2) and presequence binding site (red letter) on N-terminal Tom40 in water (row 3) and in water/TFE mixture (row 4).

4.5   Initial contact sites on Tom40 for membrane insertion

The secondary chemical shift revealed partial secondary structure of Tom40 in urea (Figure 22). The significant chemical shift perturbations are only observed upon addition of DHPC at a concentration above CMC indicating there is no interaction with monomeric detergent. This is an evidence that the hydrophobic surface of micelle or lipid is necessary for folding of membrane proteins [142]. Above CMC, large chemical shifts perturbations were found in regions Ala95-Gly101, Arg213-Ala221, Phe235-Gly246. These three polypeptide segments may form hydrophobic

clusters which initially bind to the membrane surface and are arranged for the further folding into the membrane.

As a β-barrel protein of mitochondrial outer membrane, Tom40 precursor protein is translocated and integrated into membrane to form final conformation with the help of TOM and SAM complex (section 1.1.2). It has been shown that deletion of two conserved residues located in the last predicted transmembrane β-brand Tom40 impaired its insertion into outer membrane of mitochondria. We have identified the last β-strand of ncTom40 in hydrogen/deuterium exchange measurement and it shifts 2 or 3 residues from predicted topology. However, recognition of membrane by last β-strand of Tom40 was not supported by our investigation of interaction between urea-denatured Tom40 and DHPC micelles. Therefore we speculate that the last β-strand is essential for sorting to SAM complex while other regions of Tom40 may play more important roles in subsequent insertion into membrane.

Part 2

A novel method for NMR resonance assignment of large intrinsically disordered proteins

5   Introduction

5.1   Intrinsically disordered proteins (IDPs)

For many years, a central assumption of molecular biology was that proteins perform biological function with specific three-dimensional structures. Since 1980's, more and more intrinsically disordered proteins (IDPs) which lack rigid structures challenge the traditional protein-structure paradigm. Bioinformatics methods estimated that in eukaryotic organisms intrinsically disordered proteins (IDPs) constitute approximately 30% of all proteins [143] and IDPs have been recognized to play functional roles in biological processes such as transcriptional regulation, translation and cellular signal transduction [144-147].

Intrinsically disordered proteins have been proved to have numerous crucial functions and perform highly specialized biological tasks that complement functionality of ordered proteins. IDPs have large solvent-accessible surface area and inherent plasticity which enables them to generate highly specific interactions between IDPs and different partners [148, 149]. Many neurodegenerative diseases, such as Alzheimer's disease, Parkinson's disease and Huntington Disease are characterized by a key process during disease development that IDPs aggregate into oligomer and further form Amyloid fibrils [149]. Moreover, IDPs are enriched in proteins associated with human diseases including cancer, cardiovascular disease and diabetes. This correlation makes IDPs novel and interesting drug targets [150].

Although IDPs have no rigid tertiary structures, they have been proved to contain varied amounts of residual structure. Small angle X-ray scattering and pulsed field gradient NMR studies of IDP both show that the hydrodynamic radius of IDP molecules are larger than the globular folded proteins with the same molecular weight, but is still smaller than the estimated values for a random coil protein of the same size [151]. This indicates the lower compactness and higher intramolecular flexibility of IDPs.

Several biophysical techniques can be employed to investigate disordered proteins. Circular dichroism(CD) spectroscopy which probes rigid secondary structure elements in folded proteins is able to illustrate the absence of rigid secondary structure [152]. Small angle X-ray scattering (SAXS) represents a powerful tool in the investigation of conformation, shape and dimensions of biopolymer molecules [153]. However, these two techniques only offer global information of structure. On the other hands, X-ray and NMR spectroscopy are two main biophysical techniques that obtain proteins structural information in atomic resolution. Unfortunately, structural investigation of IDPS has been hindered by the difficulties to crystallize IDPs proteins due to the high intramolecular flexibility and even if crystals of IDPs are obtained, the conformational variability of the backbone would make data interpretation very challenging. This makes NMR spectroscopy the most powerful technique to investigate IDPs at atomic resolution. In addition, NMR spectroscopy is also capable of characterizing dynamic information of IDPs which is essentially important for understanding the function of IDPs.

5.2 Solution NMR methods for structural characterization of IDPs

In contrast to globular proteins, IDPs contain an ensemble of conformations instead of one single structure. NMR method can give a great deal of information on the structural composition of multiple conformations of an ensemble and requires innovative approaches to describe the experimental data. With the advance of sophisticated instruments, uniformly and specifically labeled proteins and new NMR pulse sequences, IDPs with large molecular weight could be investigated in solution NMR. A huge variety of NMR observables make it a strong method for characterizing structural and dynamic information.

5.2.1   Secondary chemical shift

Chemical shift is the essential parameter of NMR reflecting the chemical environment of nuclei. Resonance assignment of proteins is a prerequisite for NMR investigation. The sequence specific assignment is typically obtained using triple-resonance experiments, which provide sequential correlations through matching $^{13}C_\alpha$, $^{13}C_\beta$ and ^{15}N chemical shift. The chemical shift from assignment can be directly utilized for secondary structural characterization. Secondary chemical shift $\Delta\delta^{13}C$ which is deviation of experimentally observed chemical shifts from their random-coil value is highly related to backbone torsion angles and can serve as a probe of secondary structure [154]. $^{13}C\alpha$ secondary chemical shift shows positive values for residues of α-helix and negative values for residues of β-sheets resulted from deshielding effect of $^{13}C\alpha$ in helices and shielding effect in β-sheets. On the other hand, $^{13}C_\beta$ secondary chemical shift has the opposite propensity with smaller value [155]. This makes the combination of C_α and C_β secondary chemical shift ($\Delta\delta C_\alpha$-$\Delta\delta C_\beta$) is an even better predictor. Intrinsically disordered proteins have smaller $\Delta\delta C_\alpha$ and $\Delta\delta C_\beta$ reflecting adoption of transient secondary structure.

5.2.2 $^3J_{HNH\alpha}$ scalar couplings

In addition to chemical shifts, the homonuclear three-bond $^3J_{HNH\alpha}$ scalar coupling contains valuable secondary structure information due to the intervening dihedral backbone-angle. The vicinal proton-proton coupling is related to the Karplus equation, making it possible to relate the magnitude of $^3J_{HNH\alpha}$ couplings to structural parameters. The coupling between H_N and Hα is able to be obtained through a variety of NMR experiments [156, 157] and the deviations are calculated by subtracting the same coupling from random-coil residues. Positive deviation of experimental couplings from the random-coil values represents β-sheet-like structures; Negative deviations are observed in regions of (transient) α-helix and polyproline helix type II that can be further distinguished by taking chemical shifts into account.

5.2.3   Residual Dipolar Coupling

Dipolar couplings are rich sources of structural information as they contain information on the orientation of internuclear vectors relative to the magnetic field, independent of its location within the protein. However, the dipolar interaction is averaged out because rotational Brownian motion makes the internuclear vector

sample all possible angles in solution. By adding weakly external alignment media, the residue dipolar couplings can be reintroduced in a liquid sample. Meanwhile, the dipolar coupling is scaled to around 10^{-3} compared to complete dipolar coupling [158] thus resulting in measurable NMR line splitting.

The ensemble of IDPs has conformation differing in size and shape, experiencing different degrees of alignment. The measured residual dipolar couplings at individual residues reflect local conformational propensities within the statistical segments. For example, the N-H vectors of α-helix and β-sheet residue will be parallel and orthogonal to external field, respectively and this leads to positive and negative N-H RDC values [159].

The determination of such residual dipolar couplings D requires the measurement of J in isotropic solution and of $J_{eff}=J+D$ in nonisotropic solution, for example in the partially aligned media. The simple method to measure heteronuclear J coupling constant is directly measurement of resolved J coupling from heteronuclear-edited spectra such as HSQC without decoupling during either the t_1 and t_2 evolution periods. Due to splitting of the peaks, the spectrum has double peaks and causes severe signal overlapping. IPAP(In-phase anti-phase) method has been commonly used to reduce overlapping by separating the peaks into two spectra [160]. Other scalar coupling measurement techniques such as E.COSY [161] and quantitative J approaches [162] are also applied to RDC measurement.

5.2.4   Paramagnetic Relaxation Enhancement

Paramagnetic Relaxation Enhancement (PRE) arises from magnetic dipolar interactions between a nucleus and the unpaired electrons of the paramagnetic center. NMR long-range NOEs, traditionally used for obtaining topological distance constraints in folded proteins, have been proven difficult to observe in disordered proteins. This is mainly because 1) the long range contact in IDPs is transient and only present in a small fraction of the conformation ensemble. 2) the contacts in disordered proteins may be too short lived to allow for efficient building up of NOE signal. 3) The assignment of NOE between side chain is very difficult owning to intensive degeneracy in IDPs [159, 163]. Compared to NOE, the PRE effect is very large owing to the large magnetic moment of an unpaired electron, permitting distances up to 35 Å (depending on the paramagnetic group) to be detected [164]. The r^{-6} dependence of PRE effect makes it a powerful tool to detect lowly populated

compact states in IDPs and the PRE-derived distance will be much smaller than the (linear) average distance s in the ensemble.

The measurement of PRE requires a protein with solvent exposed cysteines which can be used as a conjugation site for a paramagnetic probe such as MTSL. A simple way to observe PRE effect is to compare signal intensities of a single pair spectra acquired in the presence and absence of a paramagnetic spin label. The differences in peak intensities between two spectra are utilized to calculate the PRE and estimate the distance between the paramagnetic center of any given amide [165].

5.3   NMR resonance assignment of IDPs

5.3.1   HN-detected triple resonance NMR spectroscopy for resonance assignment

Triple resonance experiments have been a standard procedure for chemical shift assignment of proteins which is usually the first step for NMR studies of proteins. The sequential assignment of ^{15}N, $^{13}C'$, $^{13}C_{\alpha}$, $^{13}C_{\beta}$, of ^{13}C and ^{15}N labeled proteins are established using relatively uniform and well-resolved heteronuclear one-bond and two bonds coupling. The typical assignment protocol relies on a set of coupled experiments, for example 3D HNCA and 3D HN(CO)CA(Figure 26). According to NMR nomenclature, the chemical shift evolutions are performed on H, N, CA in both experiments, whereas the parentheses in 3D HN(CO)CA indicate the magnetization is only transferred via $^{13}C'$ spin without chemical shift labeling.

An 3D HNCA experiment provides intraresidue correlation $H_N(i)$- $N(i)$-$C_{\alpha}(i)$ and sequential correlation $H_N(i)$-$N(i)$-$C_{\alpha}(i-1)$. Meanwhile HN(CO)CA spectra exclusively provides sequential correlation $H_N(i)$-$N(i)$-$C_{\alpha}(i-1)$ by transferring magnetization through $C'(i-1)$ [166]. Chemical shift degeneracy of Cα may lead to ambiguous assignment which could be solved by chemical shifts of other nuclei. An alternative correlation can be obtained from a set of HNCO and HN(CA)CO spectra where ^1H and ^{15}N are correlated with intra-residual C′ and previous C′. A Variant set of HNCA and HN(CO)CA experiment, HNCACB [or CBCANH] and HN(CO)CACB [or CBCACONH] also include the C_{β} spin which can be used to identify residue types and solve some ambiguous assignments caused by Cα chemical shift degeneracy [166,

167].

Experiment	Correlation observed	Magnetization transfer
HNCA	$^1H_i^N - {}^{15}H_i^N - {}^{13}C_{i-1}^{\alpha}$ $^1H_i^N - {}^{15}H_i^N - {}^{13}C_i^{\alpha}$	
HN(CO)CA	$^1H_i^N - {}^{15}H_i^N - {}^{13}C_{i-1}^{\alpha}$	
HN(CA)CO	$^1H_i^N - {}^{15}H_i^N - {}^{13}C_{i-1}^{'}$ $^1H_i^N - {}^{15}H_i^N - {}^{13}C_i^{'}$	
HNCO	$^1H_i^N - {}^{15}H_i^N - {}^{13}C_{i-1}^{'}$	
HNCACB	$^1H_i^N - {}^{15}H_i^N - {}^{13}C_i^{\alpha} / {}^{13}C_i^{\beta}$ $^1H_i^N - {}^{15}H_i^N - {}^{13}C_{i-1}^{\alpha} / {}^{13}C_{i-1}^{\beta}$	
HN(CO)CACB	$^1H_i^N - {}^{15}H_i^N - {}^{13}C_{i-1}^{\alpha} / {}^{13}C_{i-1}^{\beta}$	

Figure 26: Triple resonance experiments for backbone assignment in magnetization transfer, the arrows indicate the direction of magnetization transfers. The circles with solid lines represent the nuclei whose chemical shifts were recorded during experiments. And the dashed circles represent the nuclei which are involved in magnetization transfer, but not chemical shift labeled.(modified from ref [168] with permission)

These HN-detected triple resonance experiments have provided a standard

strategy for the assignment of protein backbones and side chains. Additionally, the spectra of larger proteins could be optimized by (per)deuteration [169] and Transverse Relaxation Optimized Spectroscopy (TROSY) [91]. However, amide proton detection NMR experiments will lower its efficiency for characterizing IDPs mainly because (1) IDPs which lack stable 3D molecular structures are characterized by a small chemical shift dispersion of NMR signals as the chemical shifts deriving from the local protein environment are largely averaged out [154]; (2) The amide protons of IDPs are barely protected from exchange with water due to lack of hydrogen bonds, and their signals tend to be broadened due to rapid exchange with water at high temperature or alkali condition; (3) IDPs have relatively high proportion of prolines [170] which is invisible in HN-detected experiments and generate gaps for sequential assignment.

5.3.2 ^{13}C-detected NMR spectroscopy for sequence-specific assignment of IDPs

It is well known that the chemical shifts of ^{13}C and ^{15}N nuclei of proteins are more sensitive to the amino acid type and sequence than those of ^{1}H. Accordingly, the resonances of these nuclei are more dispersed than those of protons even in the absence of rigid tertiary or secondary structure [159]. ^{13}C-detected NMR experiment thus could facilitate the assignment of IDPs with relatively high resolution. On the other hand, the ^{13}C-detected experiments have a lower sensitivity compared with ^{1}H-detected ones due to lower gyromagnetic ratio. However, developments in probe technology and high field magnet have created dramatic improvements of sensitivity not only for ^{1}H but also for ^{13}C as well. Moreover, fast chemical exchange between amide protons and water at high pH values and high temperature can severely reduce the sensitivity of the experiment involving amide protons [171,172]. This makes NMR experiments based on ^{13}C detection potentially very useful for biomolecular application.

For ^{13}C-detected experiments, a potential problem is the broadened linewidth in the direct dimension due to homonuclear coupling. In the case of C′ detection experiment, $J_{C\alpha CO}$=55Hz is the main resource of homonuclear coupling , while for CA detection, the splitting comes from both $J_{C\alpha,\beta}$=35Hz and $J_{C\alpha CO}$=55Hz [173]. These J couplings are able to be removed using spin-state selective schemes, such as in-phase anti-phase(IPAP) or spin state selective excitation which is widely used for measurement of heteronuclear J coupling [173].

The advantage of ^{13}C direct detection has boosted the developments of numerous the novel experiments. 3D CANCO correlates the chemical shifts of C_α nuclei with the shifts of the two neighboring nitrogen nuclei and subsequently with those of the carbonyls adjacent to these nitrogen atoms [174]. The CBCACO experiment provides the correlation between C′ and CB/Cα and the connection can be extended to rest of side chain through CC-TOCSY. In the 3D COCON-IPAP, The backbone nitrogen is correlated with the attached carbonyl carbon and also with the previous and following carbonyl carbons [175]. A variety of ^{13}C-detected experiments offer a valuable complementary scheme for assignment based on ^{1}H detection [147, 174, 176-182].

Investigation of structure and dynamics of IDPS are also able to be accomplished by ^{13}C-detected experiments. Thanks to large chemical shift dispersion of ^{13}C′ and ^{15}N for IDPs, 2D CON experiment is especially suitable to be modified for determination of the various NMR observables such as residue dipolar coupling, ^{13}C-^{1}H NOE [183], protein-protein interaction [184] and dynamics [185]. By making use of CON-based experiments all residues including prolines is able to be observed and analyzed which is large advantage compared to HN-detected experiments.

5.3.3   HA-detected NMR spectroscopy for sequence-specific assignment of IDPs

HA-detected experiments provide an alternative protocol for assignment of IDPs besides HN- and ^{13}C-detected ones. The resonance peaks on HA-detected spectra are not deteriorated by line broadening due to amide proton exchange. It also allows a straightforward assignment of prolines in IDPs.

Two 2D experiments, HA(CA)N and HA(CACO)N or their 3D analogs was early proposed for assignment of proline-rich peptide [186]. HACAN correlates both intra and inter residues $H_\alpha(i)$-$C_\alpha(i)$-$N(i)$/$N(i-1)$ and HACA(CO)N exclusively correlates inter residues $H_\alpha(i)$-$C_\alpha(i)$-$N(i-1)$. Thus the sequential connectivity can be established. Moreover, residue type information is able to be obtained by combining $^{13}C_\beta$ chemical shift from (HB)CBCA(CO)N(CA)HA [187]. A proline-optimized CDCA(NCO)-CAHA has been developed for sequential assignment of proline residues by correlating $^{13}C_\delta$ and $^{13}C_\alpha$ chemical shifts of proline residues with the H_α chemical shift of the preceding residue [188]. Chemical shifts of ^{13}C′ and ^{15}N are essentially useful to resolve peak overlapping in IDPs due to their large chemical shift dispersion even for IDPs. Therefore, in iH(CA)NCO and H(CA)CON which provide

intraresidual correlation $^{15}N(i)$-$^{13}C'(i)$-$H\alpha(i)$ and sequential $^{15}N(i\text{-}1)$-$^{13}C'(i)$-$^{1}H_{\alpha}(i)$, sequential assignment is accomplished with efficient use of chemical shifts of ^{15}N and ^{13}CO instead of $^{13}C_{\alpha}$ [189]. Another two HA-detected experiment (HCA)-CON(CA)H and (HCA)NCO(CA)H provide same correlation $C'(i\text{-}1)$-$N(i)$-$H\alpha(i)$ with opposite transfer pathway. The auto-correlation peaks can be suppressed by adjusting one transfer delay. Together with iH(CA)NCO, these experiments provide a robust platform for backbone assignment of natively unfolded proteins [190].

5.4   NMR techniques for speeding up data acquisition

Multidimensional NMR experiments have been widely used and successful in macromolecular NMR by correlating the frequencies of different nuclei and spreading the correlation peaks over several frequency axes. It relies on the acquisition of multiple scans with stepwise incrementation of an evolution delay for each additional dimension [167]. This makes the acquisition time increase exponentially with the number of dimensions. 1D, 2D and 3D experiments are able to be acquired in seconds, minutes and hours respectively and NMR experiments with higher dimensionality are not fairly practical considering sample stability and valuable instrument time. It also limits the application for real-time studies of slow dynamic process and high-throughput screening. During last decade, the high field magnet and cryogenic probe have led to a significant gain in sensitivity of NMR spectroscopy. Therefore the time required for sampling the chemical shift of each dimension in the conventional way typically exceeds the time needed for sensitivity consideration.

Recently continuous efforts have been made to accelerate NMR data acquisition. The strategies to shorten the experiment time mainly go into two categories [191]. One is to reduce the time delay between scans which is required to allow the spin system to relax back to its thermal equilibrium state [192] and the other one is to reduce the number of sampled data points in the indirect dimensions [193](Figure 27).

The recycle delay for longitudinal relaxation recovery which is typically around 1s for ^{1}H acquisition experiment in solution NMR represents the longest part for an experiment. The magnetization recovery can be accelerated by enhancing the

longitudinal relaxation and time for one acquisition scan can be reduced. This thus increases the signal-to-noise of NMR experiment per unit time of data acquisition. The longitudinal relaxation optimized 2D [^{15}N,^{1}H]-TROSY, 2D [^{15}N,^{1}H]-HSQC, and 3D TROSY-HNCA only perform excitation/detection with ^{1}H$_N$ spins covalently attached to ^{15}N and yield increase of maximal signal-to-noise per unit time [194]. In so-called Sofast-HMQC band-selective excitation and small angle flip (Ernst angle) enable to perform fast repetition allowing real-time site-resolved NMR studies of kinetic processes in proteins on a time scale of seconds [195, 196]. A series of triple-resonance Band-selective Excitation Short-Transient (BEST) experiments have been proposed using band-selective radio-frequency(RF) pulse to enhance longitudinal ^{1}H relaxation [197-199].

Mainly two distinct approaches have been implemented to reduce the sampling data points in the indirect dimension. The former one is based on coupling the evolution of several different nuclei such that their chemical shift appear in a joint spectra dimension as so called reduced dimensionality. The concept of reduced dimensionality spectra have been presented by different way such as G-matrix Fourier transformation(GFT) [200-203], Projection-Reconstruction [113, 204] or Automated Projection Spectroscopy(APSY) [109, 121]. The reduced dimensionality can significantly reduce measurement time and seven-dimensional APSY has been proposed to accomplish sequence-specific assignment [119, 120]. However, instead of chemical shift of single nuclei, the joint evolution produces linear combinations of chemical shifts from several nuclei which complicate the direct interpretation of the spectra. This problem is able to be overcome by reconstruction of the full-dimensional spectrum using various back–projection schemes [204-206]. Another approach is to perform peak picking on the spectra for later on manual or automatic assignments such as the scheme in APSY [119, 121]. The second approach for reducing the sampling of data points in the indirect dimension is non-uniform (also referred as non-linear) sampling. Conventional multidimensional NMR experiments are implemented with the traditional uniform acquisition scheme which means that a set of uniformly sampled equidistant data points spaced by the dwell-time is acquired and then the data are put to fast Fourier transformation (FFT) algorithm for processing [166]. For nD NMR, the experimental time will increase exponentially in term of dimensionality because each indirect dimension has to be sampled with equidistant data points. Alternatively, non-uniform sampling doesn't have to acquire equidistant data points.

Therefore, for same digital resolution, less data points of evolution time are necessarily recorded and it turns out to reduce measurement time. Mostly, the random sampling is implemented in a way that probability of selection matches the signal envelope. Accordingly, signal sensitivity can be improved due to sampling at "higher density" where the signal is stronger. Because FFT only can be applied on equidistant data set, some other schemes for signal processing have been proposed such as Maximum Entropy (MaxEnt) reconstruction [207], maximum likelihood method(MLM) [208], multidimensional decomposition(MDD) [209, 210], Fourier transform(FT) [211, 212] and compressed sensing [213, 214].
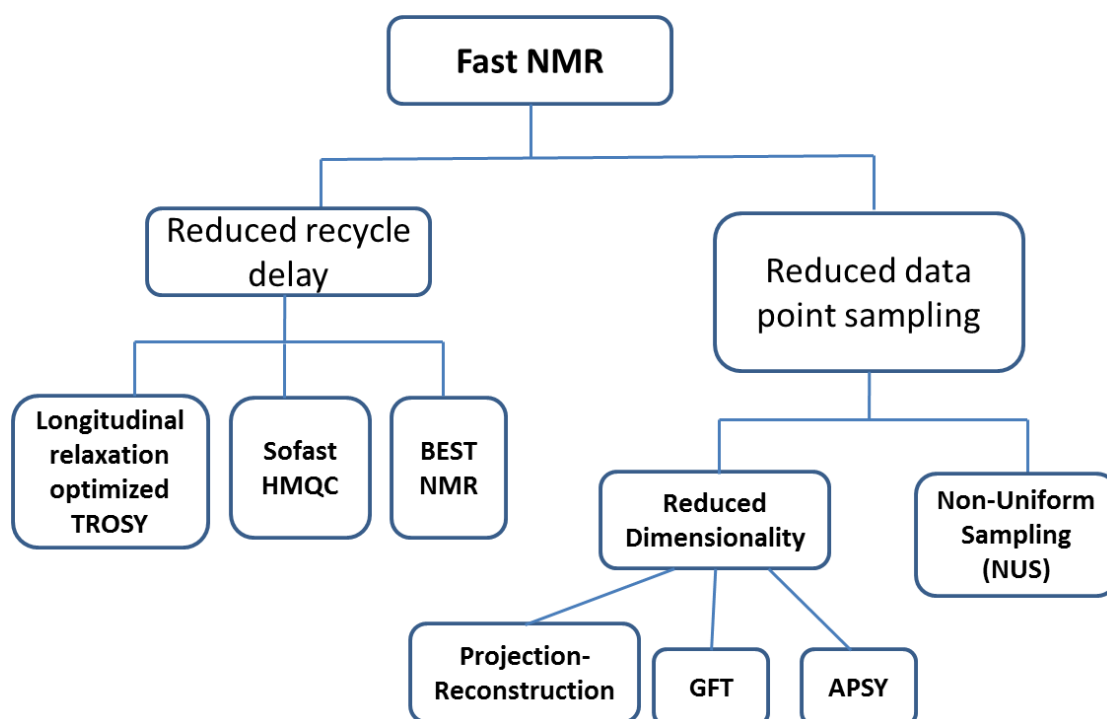
Figure 27. Classification of different fast NMR techniques

5.5   APSY-based chemical shift assignment

APSY (Automated projection spectroscopy) has combined the protocol to record projections of high-dimensional NMR experiment and automated peak-picking using algorithm of geometric analysis of projections (GAPRO) [109, 122]. In APSY experiments, practically a series of 2D projections are recorded. For nD experiments, the indirect dimension of 2D projection is a combined chemical shift evolution of N-1 nuclei requiring N-2 projection angles. For example, 6D APSY employs 4 projection

angles α, β, γ, δ to create linear combination of chemical shift from 5 nuclei in the indirect dimension. The time domain is sampled along a straight line defined by unit vector $\vec{p}_1$ [121].

$$\vec{p}_1 = \begin{pmatrix} \sin \delta \\ \sin \gamma \cdot \cos \delta \\ \sin \beta \cdot \cos \gamma \cdot \cos \delta \\ \sin \alpha \cdot \cos \beta \cdot \cos \gamma \cdot \cos \delta \\ \cos \alpha \cdot \cos \beta \cdot \cos \gamma \cdot \cos \delta \\ 0 \end{pmatrix} \tag{5.1}$$

The spectral width, SW, of each 2D projection is calculated from:

$$SW = \sum_{i=1}^{5} p_1^i \cdot SW_i \tag{5.2}$$

Where p_1^i represents the coordinates of the vector (Equation 5.1), and SW_i are the spectral widths of the five indirect dimensions (Equation 5.2). The dwell time for the recording of discrete data points, Δ, was calculated as Δ=1/SW and the resulting increments for the evolution times t_i =n*Δ_i. n is the number of increment and $\Delta_i = p_1^i \cdot \Delta$ is dwell time in the five indirect dimensions [121].

During experimental set up, a set of projection angles are created based on spectral sweeps for all indirect dimensions to optimize with maximum peak dispersion. Then all 2D projections are recorded and automatically processed with PROSA [215]. Afterward, automatically peak picking is implemented on all 2D projection spectra using ATNOS algorithm [216]. The resulting peak lists of the projections are analyzed using GAPRO, which identifies all projected peaks arising from the same spin system, and then calculates the final 6D peak list. The peak list is ready for manual assignment or alternatively automatic assignment using software such as MARS [118] or UNIO/MATCH [217] thanks to its high resolution.

APSY experiments achieve significant time reduction by spontaneously sampling in the indirect dimension and enable to record high-dimensional experiments up to seven dimensions [120]. The APSY experiments for backbone assignment were summarized as in Figure 28 (www.apsy.ch). The directions of magnetization transfer were indicated by the arrows. The circles with solid lines represent the nuclei whose chemical shifts were recorded during experiments whereas

the dashed circles represent the nuclei which are involved in the magnetization transfer, but not chemical shift labeled.

The process of sequence-specific resonance assignment of a protein backbone usually comprises of 2 steps. The first one is to build sequential fragments and second is to map these fragments into the sequence. 6D APSY HNCOCANH has been proved powerful for obtaining sequential fragment by correlating 2 sequential amide moieties in a single experiment [121] . The glycine residues are able to be identified due to it specific $^{13}C_\alpha$ and ^{15}N chemical shift and be used as marks for mapping onto the sequence. Meanwhile $^{13}C_\beta$ chemical shifts information allows recognition of specific amino acid type such as ALA, THR and SER and help for mapping the fragments onto the sequence. Therefore 5D APSY CBCACONH is useful to combine with 6D HNCOCANH and improve the efficiency of assignment. 7D APSY-HNCO(CA)CBCANH can provide the same chemical shift information as a combination of 6D HNCOCANH and 5D CBCACONH in a single experiment. However, sensitivity of 7D APSY is lower due to more magnetization transfer steps. High-dimensional experiments such as 6D and 7D have great spectral dispersion but also are limited on sensitivity because of relaxation loss during many transfer delays. Therefore, high-dimensional experiments based on APSY are suitable for studies of small or unfolded proteins which have favorable NMR relaxation rate.

For larger proteins, recently a set of 4D APSY experiments (HNCACO, HNCOCA, HNCACB and HN(CO)CACB) with TROSY effect were designed and open an avenue to automatic resonance assignments of proteins with a size up to 40kDa. The 4D TROSY APSY-HNCACO and 4D TROSY APSY-HNCOCA can be implemented together as a group providing intra-residues and sequential correlation. The two peak lists could be merged based on the chemical shifts of the $^{1}H_N$–^{15}N moieties to an inter-residue 6D $^{13}C_\alpha(i-1)$–$^{13}C'(i-1)$-$^{1}H_N(i)$–$^{15}N(i)$–$^{13}C_\alpha(i)$–$^{13}C'(i)$ peak list which is subsequently put into MARS for automatic sequential resonance assignment.

Resonance assignment of proteins usually starts with establishing the backbone assignment and subsequently attaches the side chain to backbone. The 5D APSY-HC(CC-TOCSY)CONH experiment which yields 5D chemical shift correlations of aliphatic side chain C-H moieties with the backbone atoms H_N, N, and C′ has been proposed for resonance assignment of side chain [218] . Aromatic residues which cannot be assigned using the 5D APSY-HC(CC-TOCSY)CONH

experiment are able to be unambiguous assigned in a 4D APSY HBCB(CG)CDHD experiment [219].
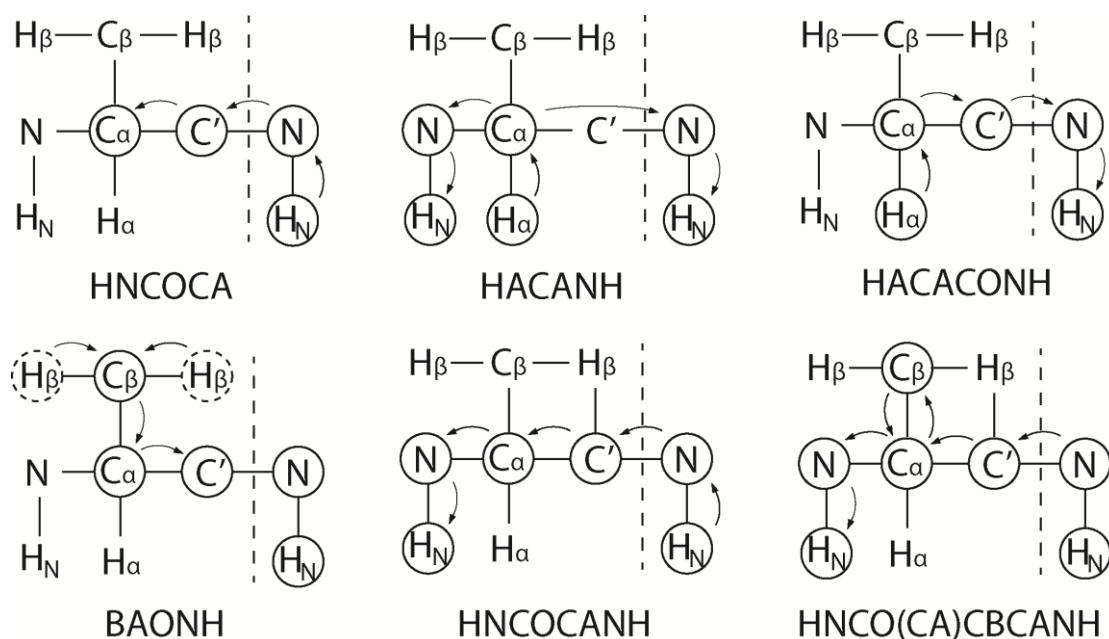
Figure 28: Magnetization transfer pathways for APSY experiments for backbone assignment. The arrows indicate the direction of magnetization transfer. The circles with solid lines represent the nuclei whose chemical shifts are recorded during experiments, and dashed circles represent the nuclei which are involved in magnetization transfer, but not chemical shift labeled.

5.6   Motivation of this project

IDPs have small chemical shift dispersion owning to its lack of stable 3D structure. This makes the resonance assignment which is a prerequisite for NMR study challenging because of peak overlap. Heteronuclear detection mainly ^{13}C-detected experiments could be an alternative from standard proton-detected ones because of the fact that the chemical shifts of ^{15}N, ^{13}Cα, ^{13}C′ are more dispersed than that of ^{1}H for IDPs.

Multidimensional NMR enables to spread the resonance peaks into multidimensional space and reduce the resonance overlap. However, for standard NMR sampling scheme, multidimensional NMRs are limited to 3D or 4D, because of impractical long measurement time for higher dimensionality.

APSY which combines chemical shift evolution for several indirect

dimensions is capable of offering very high resolution and has been proved to be powerful to resolve the resonance of 441-residue intrinsic disordered protein Tau for automatic assignment [119]. On the other hand, the sensitivity of high-dimensional APSY could be deteriorated from severe relaxation decay for globular proteins. Intrinsic disordered proteins, therefore, are ideal for APSY thanks to their favorable relaxation properties rising from flexible backbones. Until now, a lot of APSY experiments based on HN-detection have been proposed. However, rapid exchange between amide proton and water at high pH and temperature cause broadening of the resonances of amide signals and deteriorate the spectra. HA-detected experiments are alternative methods for backbone assignment which are barely affected by pH and temperature. Therefore, we combined APSY acquisition scheme into HA-detected experiments. A new 6D APSY based on HA detection was designed to provide correlation of six different nuclei H_α(i-1), C_α(i-1), C'(i-1), N(i), C_α(i) and H_α(i). The single experiment is sufficient to provide exclusively sequential connectivity of aliphatic unit. Thanks to the large dispersion of resonance from high dimensionality, the peak lists from APSY experiments are ready for automatic backbone assignment.

6   NMR spectroscopic methods

6.1   Constant time (CT) and semi-constant time evolution

For multidimensional NMR experiments, the acquisition of indirect dimension is achieved by recording a series FID with stepwise incremented delay for chemical shift evolution.
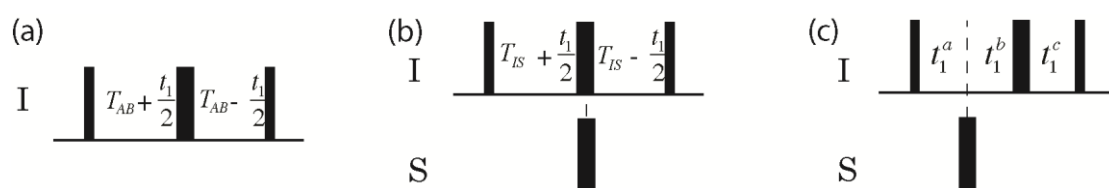
Figure 29: Variation of evolution schemes. (a) homonulcear and (b) heteronuclear constant time, (c) semi-constant time. Narrow and wide bars are 90° and 180° pulse. $2T_{AB} \approx 1/^1J_{CA,CB}$, $2T_{IS}=1/(2*^1J_{IS})$.

The chemical shift labeling in indirect dimension can be implemented in "real time", "constant time" and "semi-constant time" methods. In 'real time' fashion, the time points for chemical shift evolution are sampled by simply incrementing the delay of chemical shift evolution. The constant time experiments were initially used to eliminate the coupling from homonuclear. The overlapped chemical shift of homonuclear eg. C_α and C_β makes it difficult to refocus their scalar couplings by applying a 180° pulse. The constant time unit comprises of an incrementing $T_{AB}+t_1/2$ delay and a decrementing $T_{AB}-t_1/2$ delay during evolution. A non-selective 180° pulse starts at the center of $2*T_{AB}$ and shifts stepwise to one side until the chemical shift

evolution time reaching the maximum of t_1 while the evolution time for $^1J_{CA,C\beta}$ remains $2*T_{AB}$. By setting $2T_{AB} \approx 1/^1J_{C\alpha,C\beta}$, the evolution of $^1J_{C\alpha,C\beta}$ is refocused all the time.

This method can be extended to heteronuclear and be able to incorporate the chemical evolution into the magnetization transfer delay (Figure 29). By setting $2*T_{IS}=1/(2*^1J_{IS})$, the magnetization is transferred from spin I to spin S via J coupling and at the same time chemical shift labeling on spin I is accomplished. Therefore, constant time evolution which is able to shorten the duration of pulse sequence has been widely used as a building block for multidimensional NMR spectroscopy.

However, this method has a limitation of spectra resolution because the t_{1max} is restricted to $2T_{IS} =1/(2*^1J_{IS})$ or $1/^1J_{IS}$. To achieve higher resolution, the constant time unit is replaced by its semi-constant variant. In semi-constant time duration, three variable delays t_{1a}, t_{1b} and t_{1c} (Figure 29c) have to be incremented or decremented in order to obtain the evolution of chemical shift δ and of heteronuclear coupling J_{IS} as described by:

$$\delta: \qquad t_1^a + t_1^b - t_1^c = t_1 \tag{5.3}$$

$$J: \qquad t_1^a - t_1^b + t_1^c = 2T_{IS} \tag{5.4}$$

At the first point $t_1(0)$, the three delays is: $t_1^a(0) = t_1^c(0) = T_{IS}$ and $t_1^b(0) = 0$. For a FID along t_1 that is to be digitized by TD complex data points, the increments of three delays is (Δt_1=1/SWH, where SWH is the spectra width in Hz):

$\Delta t_1^a = \Delta t_1 / 2$, $\Delta t_1^b = \Delta t_1 / 2 + \Delta t_1^c$, and $\Delta t_1^c = \Delta t_1^c(0)/TD$.

The three increments with different lengths assure chemical shift evolution with an unlimited length of t_{1max} and fixed coupling evolution time $2T_{AB}$ for all evolution durations. It should be noted that Δt_1^c is negative representing that t_1^c is actually decremented during the increase of delay for chemical shift evolution.

6.2 APSY experiment setup

In initial setup of APSY experiments, an au-program "apsy" for Bruker Topspin and a C-program GAPRO were downloaded on http://apsy.ch and installed

on the computer of the spectrometer. Before starting 6D HACACONCAH APSY, standard parameters such as pulse lengths, power levels, pulse shapes, gradients, sweep widths have to be defined, like in a normal experiment.

NMR spectra were acquired on Bruker Avance 700 MHz (α-synuclein) and 900 MHz spectrometers (htau23) both equipped with cryo-probe. The sample condition of proteins is:

(1) α-synuclein: ^{15}N and ^{13}C labeled, 50mM NaPi, pD 6.0, 100mM NaCl,1mM concentration in D_2O.

(2) htau23:^{15}N and ^{13}C labeled, 50mM phosphate buffer, pD 6.8, 1mM concentration in D_2O.

The parameters for the 5 indirect dimensions were given in the parameter of topspin CNST, CNST41-45 for spectra widths and CNST51-55 for spectra centers as in Table 8 for the measurements of α-synuclein at 298K and htau23 at 298K. The projection angle sets for α-synuclein and htau23 were produced by GAPRO and are listed in Appendix 5. Table 9 summarizes the numbers of projections acquired and the numbers of complex points in indirect dimension for α-synuclein and htau23. 57 projections of APSY were recorded at 283K for α-synuclein and cost 57 hours. At 298K, 65 projections of APSY were acquired within 60 hours. APSY on htau23 recorded 65 projections in 84.5 hours at 298K.

Table 8 : Parameters for setting up APSY experiments

	Nuclear spin	HA(i-1)	Cα (i-1)	C′ (i-1)	N(i)	Cα(i)
α syn on 700M	center of spectra (ppm)	4.2	51	172.5	118	51
		cnst 51	cnst52	cnst53	cnst54	cnst55
	sweep width(Hz)	1050	4200	1750	1960	4200
		cnst41	cnst42	cnst43	cnst44	cnst45
htau23 on 900M	center of spectra (ppm)	4.2	51	172.5	116	51
		cnst 51	cnst52	cnst53	cnst54	cnst55
	sweep width(Hz)	1350	5850	2250	2340	5850
		cnst41	cnst42	cnst43	cnst44	cnst45

Table 9: Summary of data recorded for α-synuclein and htau23

protein	temperature	projections	complex points	measurement time
α synuclei	283K	57	86	57 hours
	298K	65	80	60 hours
htau	298K	65	114	84.5 hours

6.3 APSY data analysis

The acquired 2D projections were processed using PROSA program [215]. The peak lists from all projections were obtained by automatic peak picking using GAPRO and were subsequently put into GAPRO to yield the final 6D peak list. Only the peaks with signal-noise-ratio(S/N) larger than R_{min} were collected for creating a final peak list. The final peak list was put into MARS for sequential specific assignment using two sequentially adjacent CA-HA groups as sequential connectivity. The cutoffs for chemical shift match of $H_\alpha(i)$-$C_\alpha(i)$ and $H_\alpha(i-1)$-$C_\alpha(i-1)$ were optimized and 0.05ppm for C_α and 0.02ppm for Hα were employed for automatic assignment by MARS [118]. Data analysis for α-synuclein at 283K and tau23 at 298K were performed in same procedures and the used S/N thresholds and connectivity cutoffs are listed in Table 10.

Table 10: Parameters for analysis of APSY data

protein	temperature	Signal/noise threshold for peak picking (R_{min})	cutoff for matching in MARS	
			Cα(ppm)	Hα(ppm)
α syn	298K	6	0.05	0.02
	283K	8	0.06	0.02
htau23	298K	6	0.05	0.02

7   Result and Discussion

7.1   6D HCACONCAH APSY

7.1.1   Pulse sequence

The 6D HCACONCAH APSY pulse sequence (Figure 30) exploits the same magnetization pathway as HA-detected (HCA)CON(CA)H proposed by Mantylahti et al [190] (Figure 30a). The pulse sequence is briefly described below. The pulse sequence started with 90 °pulse on $^{13}C_\alpha$ to purge the magnetization from heteronuclear $^{13}C_\alpha$. In first ^1H–^{13}C INEPT ^1H(i-1) spin polarization is transferred to bounded $^{13}C_\alpha$(i - 1). Chemical shift of ^1H (i–1) is labeled during the same time. From time point u to v, the coherence is transferred to the $^{13}C'$(i-1) using scalar coupling between $^{13}C_\alpha$ and $^{13}C'$. The refocus of the anti-phase of ^1H–^{13}C coherence is implemented in delay τ_2 and followed by ^1H decoupling. $^{13}C_\alpha$ chemical shift is recorded at the same time. The selective 180 °pulse should cover both $^{13}C_\alpha$ and $^{13}C_\beta$ region to maintain $^1J_{CACB}$ coupling. The position of ^{15}N 180° pulse is adjusted for ^{15}N decoupling by fulfilling $\delta_1 = (t_2^c - t_2^b - t_2^c)/2$ during t_2 incrementation. Between time points v and w the anti-phase of $^{13}C_\alpha$-$^{13}C'$ coherence are refocused and magnetization is transferred to ^{15}N spin utilizing scalar coupling $^1J_{C'N}$ simultaneously. Chemical shift labeling of $^{13}C'$ (i-1) is performed in semi-constant time manner. In the following delay from w to x ^{15}N chemical shift is labeled. The anti-phase coherence of $^{13}C'$(i-1) is converted into in-phase and thus transferred to $^{13}C_\alpha$(i-1). In the ensuing period between time point x and y, the coherence is refocused with respect to ^{15}N while $^{13}C_\alpha$(i) chemical shift is labeled. The ^1H decoupling is turned off during the delay τ_2 and the coherence is

transferred to ^{1}H. Finally, the magnetization is converted into observable ^{1}H$_\alpha$ coherence by subsequent two INEPT transfer units [220]. The pair of gradient pulses G_1 and G_2 have to be set at a ratio of $G_1/G_2=\gamma_H/\gamma_C$ (the gyromatic ratios of ^{1}H and ^{13}C) for coherence selection. The grey shape pulse is necessary for compensation of chemical shift evolution during the same shape pulse.
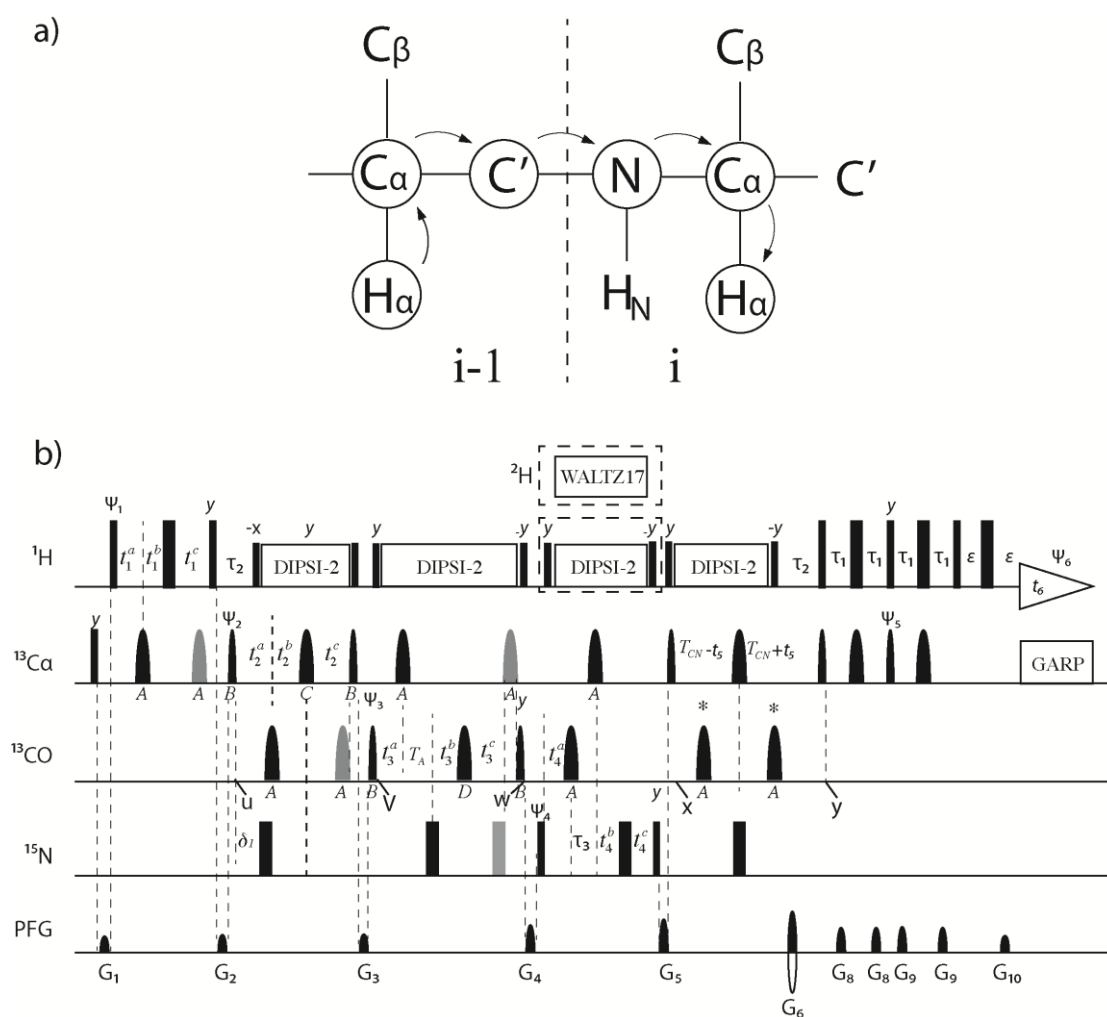
Figure 30. a) Schematic presentation of magnetization transfer pathway of 6D HCACONCAH APSY. b) 6D HCACONCAH APSY pulse sequence. Radio-frequency pulses are applied at 118.0 ppm for ^{15}N, 172.5 ppm for carbonyl carbons (^{13}C'), and 51.0 ppm for ^{13}C$_\alpha$. Narrow and wide rectangular bars are 90 °and 180 °hard pulses, respectively. Thin (fat) sine bells are shaped 90 ⁏(180 ⁏ pulses on ^{13}C channel. 4 different shape pulses labeled as A to D are employed. The actual pulse durations should be adjusted for different magnet fields (Here is 700M). A, Iburp shape is applied on Cα(51ppm) or C' (172.5ppm) with duration of 230 μs ; B, Gaussian shape on C' or Cα with duration 410 μs; C, RE-burp (Green and Freeman, 1991) is applied on C$_{\alpha\beta}$(42ppm) duration 290μs; D, RE-burp of 580μs is applied on CO (172.5ppm). The two pulses marked with an asterisk are continuously centered with respect to the time

periods $T_{CN}-t_5$, $T_{CN}+t_5$, respectively. Grey pulses are used to compensate off-resonance effects [166, 221]. Decouplings using DIPSI-2 on ^{1}H and GARP on ^{13}C are shown as white rectangles. , ^{1}H decoupling in the dashed rectangle is replaced by the WALTZ17 ^{2}H decoupling in D_2O buffer. Curved shapes on PFG line represent sine bell-shaped, pulsed magnetic field gradients applied along the z-axis. G_1 and G_2 employ the pulse duration of 1000us and the gradient strengths are adjusted according to their gyromagnetic ratios $G_1/G_2 = \gamma_H/\gamma_N$. The following phase cycling is used: $\psi_1 = x$, $\psi_2 = [x,x,x,x,-x,-x,-x,-x]$, $\psi_3 = [x,-x]$, $\psi_4 = [x,x,-x,-x]$, $\psi_5= y$, $\psi_6=[x,-x,-x,x,-x,x,x,-x]$ and all pulses without indication of a phase above the pulse symbol are applied on phase x. The time points u, v, w, x, y are discussed in the text. The following delays are invariant during the experiment: $T_A=16.6ms$, $T_{CN}=25ms$, $T_{NC}=25ms$, $\tau_1= 1.7 ms$, $\tau_2 = 4.8 ms$, $\tau_3= T_{NC} - T_A =8.4ms$. The following delays are altered during chemical shift encoding and the initial delays are given: $t_1^a = t_1^c = 1.7ms$, $t_2^a =4.5ms$, $t_2^b = 7.5ms$, $t_2^c = 14ms$, $t_3^a = t_3^b = 0ms$, $t_3^c = 16.6ms$, $t_4^a =16.6ms$, $t_4^c = T_{NC}=25ms$, $t_4^b = 0ms$, $t_5 = 0ms$, and the delay δ_1 is continuously adjusted to $\delta_1= (t_2^c - t_2^b - t_2^c)/2$ during incrementation. In the five indirect evolution periods, constant time or semi-constant time evolutions are applied (see section 6.1). Quadrature detection for the indirect dimensions is achieved for t_1, t_2, t_3, t_4 by States-TPPI method incrementing phases ψ_1, ψ_2, ψ_3 and ψ_4, and for t_5 by echo-antiecho method incrementing ψ_5 and inverting the G_1 gradient pulse. The trigonometric addition theorem is used to obtain pure cosine and sine terms for a subsequent hypercomplex Fourier transformation [112, 113].

The water signal is suppressed by the gradient pulse and the flip back pulse flanking the decoupling unit. However, water suppression is still not very effective and $^1H_\alpha$ signals overlapping with water signal are not able to be detected. Therefore, D_2O solution is recommended to use in order to avoid large H_2O signal. In that case, ^1H decoupling during ^{15}N evolution has to be replaced with ^2H decoupling as shown in the dashed rectangle in Figure 30b. 6D APSY experiments are able to record chemical shifts of each nucleus in high resolution because practically only two-dimensional planes are recorded. The constant time labeling scheme has a limit of resolution due to fixed maximum chemical shift evolution duration. Therefore, chemical shift labeling of t_1, t_2, t_3, t_4 is implemented semi-constant time manner as described in section 6.1 whereas constant time evolution still is used for t_5 evolution because the transfer delay is efficiently long for the parameter used.

7.1.2   Auto–correlation peaks suppression

The sequential assignment of 6D HACACONCAH APSY relies on the observation of two sequential C_α and H_α chemical shift. The magnetization transfer step from ^{15}N to $^{13}C_\alpha$ usually can correlate to both $^{13}C_\alpha(i-1)$ and $^{13}C_\alpha(i)$ due to close values of $^1J_{C\alpha\,N}$ and $^2J_{C\alpha N}$. Since the intraresidue peak resulting from "back transfer" pathway would double the number of peaks and not contribute any additional information, auto–correlation peaks suppression is essentially important to relieve spectral overlap. Moreover, the two peaks from each pair of sequentially adjacent Cα groups are not able to be differentiated in automatic assignment software such as MARS and cause ambiguity for assignment. Therefore, for example in 6D HNCACONH APSY, only a unidirectional pathway is kept [121]. According to the calculation, the auto-correlation and sequential peaks have different transfer efficiency related to ^{15}N-$^{13}C_\alpha$ transfer delay T_{NC}.

For sequential peaks:

$$I_{seq} \propto \sin(2\pi^1 J_{C\alpha N} T_{NC})\sin(2\pi^2 J_{C\alpha N} T_{NC})\qquad(7.1)$$

For the auto-correlated peak:

$$I_{auto} \propto \cos(2\pi^1 J_{C\alpha N} T_{NC})\cos(2\pi^2 J_{C\alpha N} T_{NC})\qquad(7.2)$$
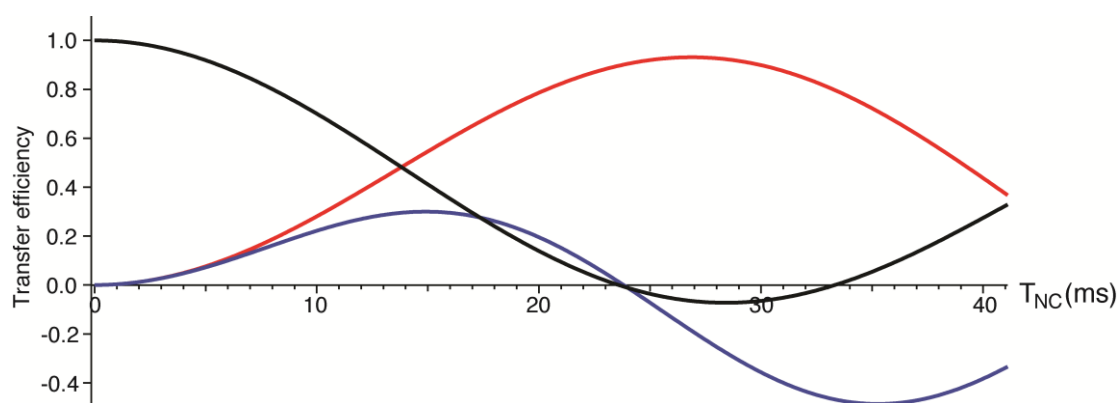
Figure 31: Transfer efficiency for non-proline sequential (red) and auto-correlation peaks(black) and proline sequential peak (blue) as a function of T_{NC}.

The transfer efficiency is calculated upon equation 7.1 and 7.2 as shown in Figure 31 [222]. 10.6 and 7.5Hz are used for random coil value of one-bond ($^1J_{C\alpha N}$)

and two-bond ($^2J_{C\alpha N}$) couplings. Therefore, by setting the delays T_{NC}=25ms, the intensities of the sequential peaks can be maximized while the auto-correlated peaks are minimized.

7.1.3 Transfer efficiency of proline

The magnetization transfer efficiency of prolines during $2T_{NC}$ is different from other residues due to extraordinary chemical structure of proline residues. Proline is an N-substituted amino acid of which the side-chain $^{13}C_\delta$ spin replaces the ^{15}N bound amide proton. Therefore the previous transfer function (equation 7.1) is modulated by additional one-bond coupling between ^{15}N–$^{13}C_\delta$ and becomes:

$$I_{seq}^{pro} \propto \sin(2\pi^1 J_{C\alpha N} T_{NC})\sin(2\pi^2 J_{C\alpha N} T_{NC})\cos(2\pi^1 J_{C\delta N} T_{NC}) \tag{7.3}$$

The coherence transfer efficiency is shown as a blue curve in Figure 31 assuming $^1J_{NC\alpha}$, $^2J_{NCa}$ and $^1J_{NC\delta}$ of 10.6, 7.5 and 10.5 Hz. When setting T_{NC}=25ms, proline residues have transfer efficiency close to 0. The proline signals are maximized when T_{NC} is set to around 35ms. However, this will cause signals cancel out due to the different signal phases of proline from other residues. If T_{NC} use smaller delay around 16ms, proline and non-proline sequential peaks are both positive generating the undesired auto-correlation peaks with comparable intensity. As a compromise, T_{NC} is set to 25ms and proline signals are not observable.

The chains of sequential connectivity HN-detected 6D HNCACONH APSY and our HA-detected 6D HACACONCAH APSY are interrupted by proline residues as illustrated by three consecutive residues R(i-1)-P(i)-R(i+1) containing one proline as a central residues P(i) and two non-proline residues R(i-1) and R(i+1)(Figure 32). The six correlated resonance spins in 6D HNCACONH APSY(Figure 32A) and 6D HCACONCAH APSY(Figure 32B) are highlighted by the C shape. The sequential correlation between R(i-1)-P(i) and P(i)-R(i+1) are both invisible (red shape) in scheme A because the sequential correlation are initiated and detected both on amide proton which is missing in the proline. On the other hand, in 6D HACACONCAH APSY the sequential connection of R(i-1)-P(i) is undetectable due to the C_δ attached N while the following P(i)-R(i+1) sequential correlation is still detectable (blue shape).

Therefore, the chemical shift resonance of $C^\alpha_{pro}(i)$, $C'_{pro}(i)$and $C^\alpha(i-1)$ which are missing in 6D HNCACONH APSY are possible to be assigned in 6D

HACACONCAH APSY.
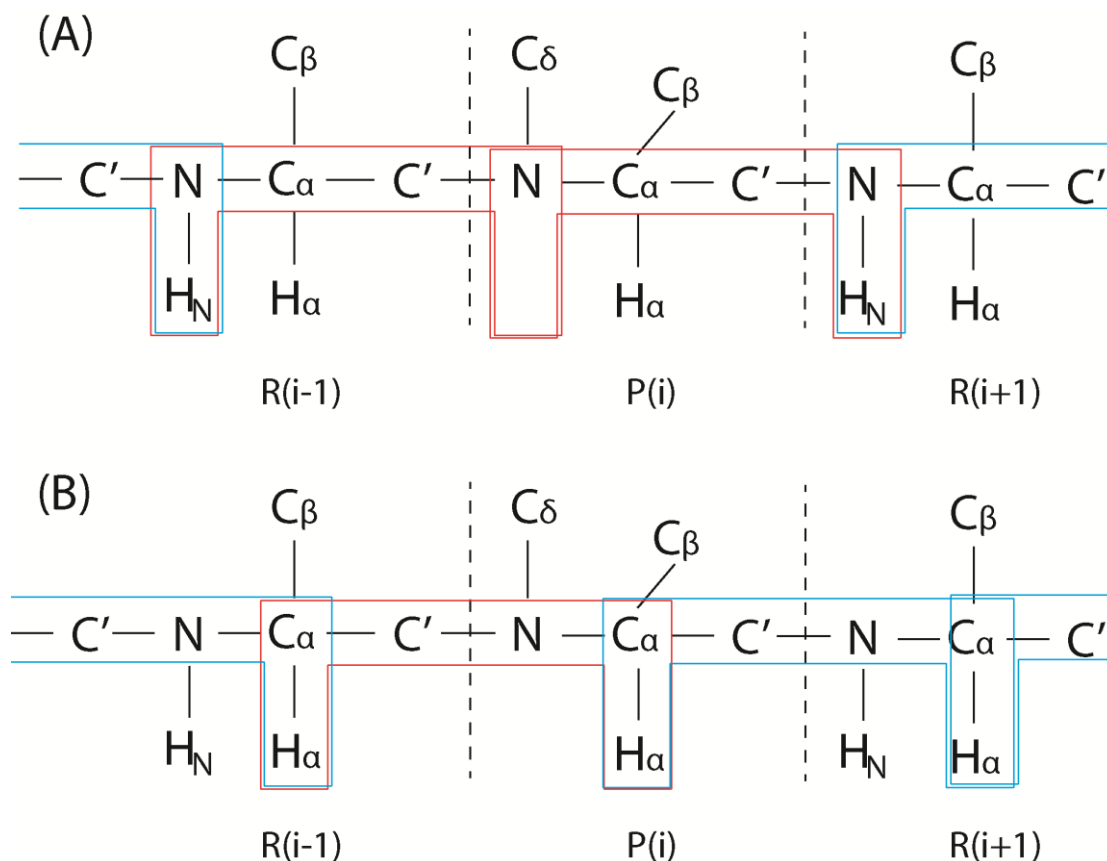
Figure 32.   Illustration of the sequential correlation for triple-residue R(i-1)-P(i)-R(i+1) in 6D HNCACONHA APSY(scheme A) and 6D HCACONCAH (scheme B) APSY. The sequential correlations are highlighted in red shape (interrupted) and blue shape (detectable).

7.1.4   Phase modification of glycine signals

The structure of glycine residues have two specific features worth noticing in the experiment. 1) Glycine has the shortest side chain among all the residue types and the only one residue not having C_β spin. 2) Two protons are attached to C_α of glycine instead of one. The durations from u to v and from x to y in pulse sequence (Figure 30b) are the two chemical shift evolution periods for $^{13}C_\alpha(i)$ and $^{13}C_\alpha(i-1)$. During these two transfer delays, the homonuclear scalar coupling from C_β are not refocused. Thus, the signal from all residues except glycine are modulated by $\sin(2\pi T_{CN} J_{CACB})$ and subsequently become negative by setting $T_{CN} = 14ms$.

Furthermore, the C-H refocusing and transfer delay τ_2 has to be optimized for

glycine due to its I_2S spin system. For non-glycine residues this delay should be optimally be set to 3.4ms, corresponding to $1/(2J_{CH})$, for IS spin moieties. However, this transfer delay can demolish signals originating from I_2S spin systems(glycine), and consequently a compromise value around 2.4ms should be used if simultaneous observation cross peaks stemming from the IS and I_2S spin systems are desired. By setting $\tau2 = 2.4$ms and $T_{CN}=14$ms, the peaks of glycines have opposite phases from other residues (Figure 33a). The distinct signal phases from glycine and non-glycine residues benefit to standard assignment procedure as a check point for the assignment. However, the resonances with opposite phases may cancel out each other especially in the spectra of IDPs. To observe both glycine and non-glycine in same sign of peak phase, $\tau_2=4.8$ms is suggested to use(Figure 33, right panel). For unfolded proteins, the relaxation loss due to longer magnetization transfer delay will be fairly small.
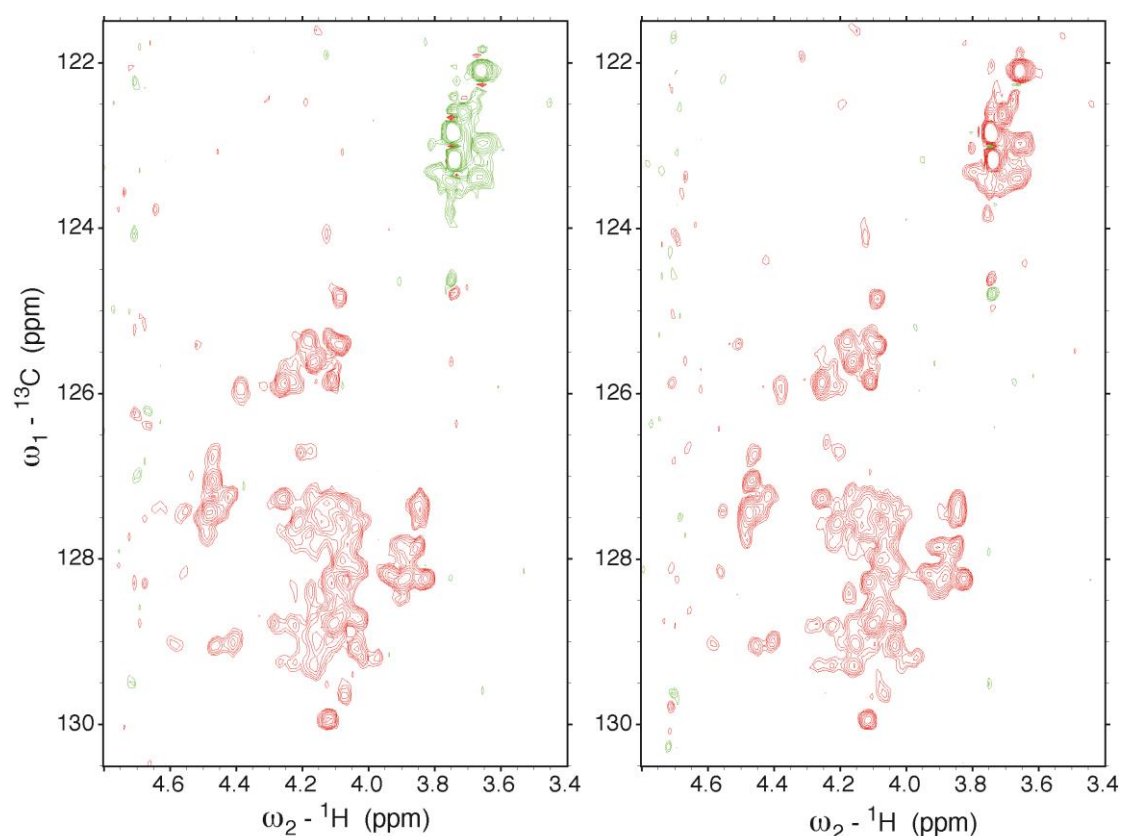
Figure 33. N-HA plane of Ubiquitin with delay τ2 of 2.4ms (left panel) and 4.8ms (right panel). The positive peaks are colored in red and negative peaks are colored in green.

7.2   Evaluation of 6D HCACONCAH APSY

This novel 6D HACACONCAH APSY was applied to two proteins. One was ^{15}N and ^{13}C labeled 140-residue human α-synuclein in D_2O with concentration of 1mM at pD 6.0. The other one was 352-residue htau23, one isoform of human microtubule binding protein Tau with concentration of 1mM at pD 6.8 in D_2O.

7.2.1   Sequence-specific assignment of α-synuclein

For sequence specific assignment of α-synuclein, 65 projections of 6D HACACONCAH APSY were acquired in 60 hours at 298K (Table 9). Figure 34 shows 5 orthogonal projections of α-synuclein with projection angles $(\alpha,\beta,\gamma,\delta)$ =(0,0,0,90), (0,90,0,0), (0,0,0,0), (0,0,90,0) and (90,0,0,0) corresponding to $H_\alpha(i-1)$-$H_\alpha(i)$, $C'(i-1)$-$H_\alpha(i)$, $C_\alpha(i)$-$H_\alpha(i)$, $C_\alpha(i-1)$-$H_\alpha(i)$ and $N(i)$-$H_\alpha(i)$ projections. From the spectra, Glycine residues have the same phases as other residues by setting the $\tau2$=4.8ms in Figure 34. $C_\alpha(i)$-$H_\alpha(i)$ projection has severe overlapping due to the chemical shift correlation between C_α and bonded H_α while the projection of $C'(i-1)$-$H_\alpha(i)$ and $N(i)$-$H_\alpha(i)$ have better chemical shift dispersion resulting from well-dispersed C' and N chemical shift even for IDPs. Except these 5 orthogonal projections, other projections in APSY combine chemical shifts from several nuclei in the indirect dimension according to the projection angle set. Regarding the high resolution achieved from six dimensions, the experiment should be able to offer excellent resonance dispersion even for large unfolded proteins.

A final peak list of 139 spin systems was acquired from 65 projections and put into MARS for automatic sequence-specific assignment. In principle, the sensitivity of this step is that of the individual 2D projection as peak picking is applied on each individual projection. However, peaks from same spin systems are correlated in different projections and the values in final peak list were averaged from several projections. This can efficiently eliminate the artifact peaks and give precise peak positions in the final peak list which benefits for match of resonance correlation in MARS [119]. The MARS program makes use of the sequential connectivity of one C_α-H_α group and its preceding one obtained from 6D HCACONCAH APSY for sequential assignment. The cutoffs of HA and CA have significant effect on the final results and were therefore optimized. In the end, sequence-specific resonance

assignment was obtained for 95 of 135 non-proline residues of α-synuclein using a cutoffs 0.05ppm for CA and 0.02ppm for HA.
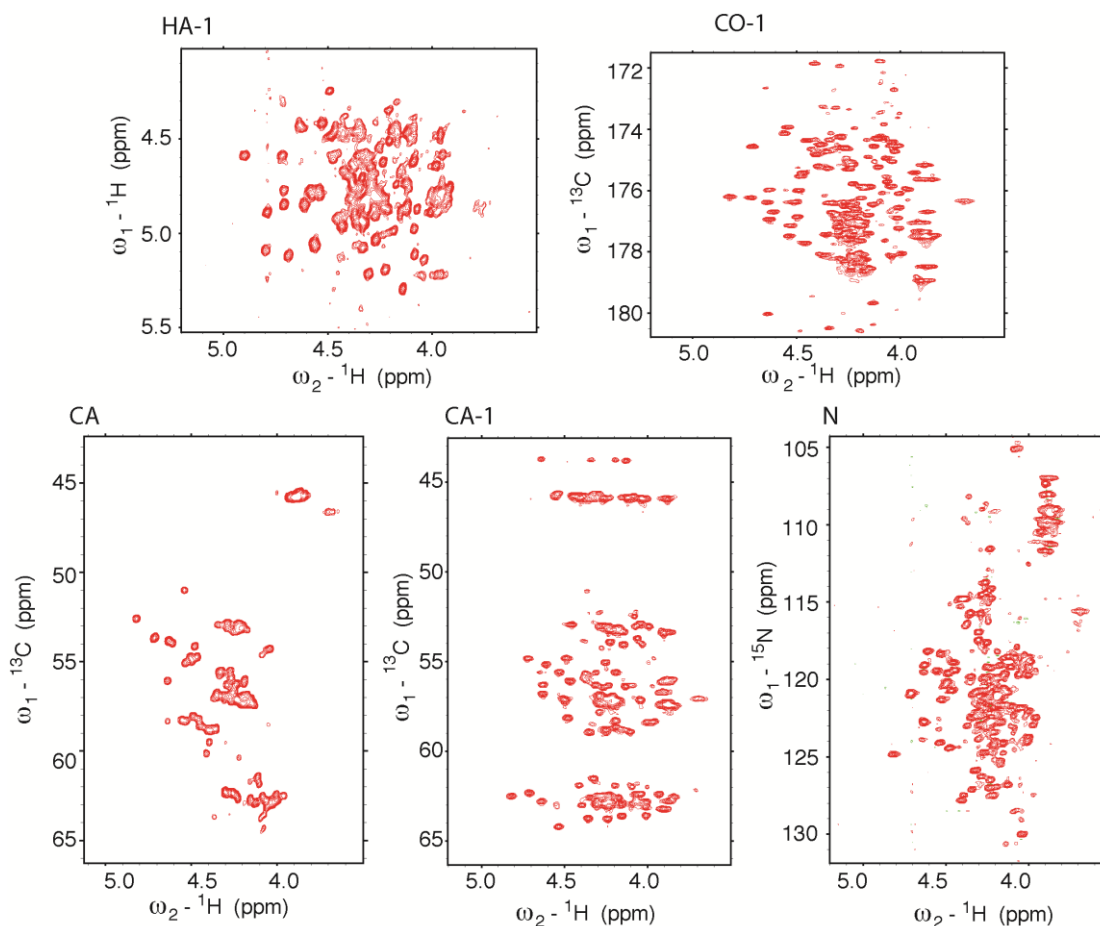
Figure 34. 5 Orthogonal projection spectra of human α-synuclein at 298K and pD 6 from 6D HACACONCAH APSY correspond to $H_\alpha(i-1)$-$H_\alpha(i)$, $C'(i-1)$-$H_\alpha(i)$, $C_\alpha(i)$-$H_\alpha(i)$, $C_\alpha(i-1)$-H_α (i) and N(i)-$H_\alpha(i)$ planes.

To evaluate the effectiveness of 6D HACACONCAH APSY, the result was compared with the published assignment for α-synuclein (BMRB ID number: 6968) (Figure 35). The chemical shift deviation of ^{15}N, $^{13}C_\alpha$ and $^{13}C'$ from BMRB assignment were calculated and plotted versus residue numbers (Figure 35). The black bars represent the residues with high and medium reliable assignment and the red ones are assigned with low reliability.
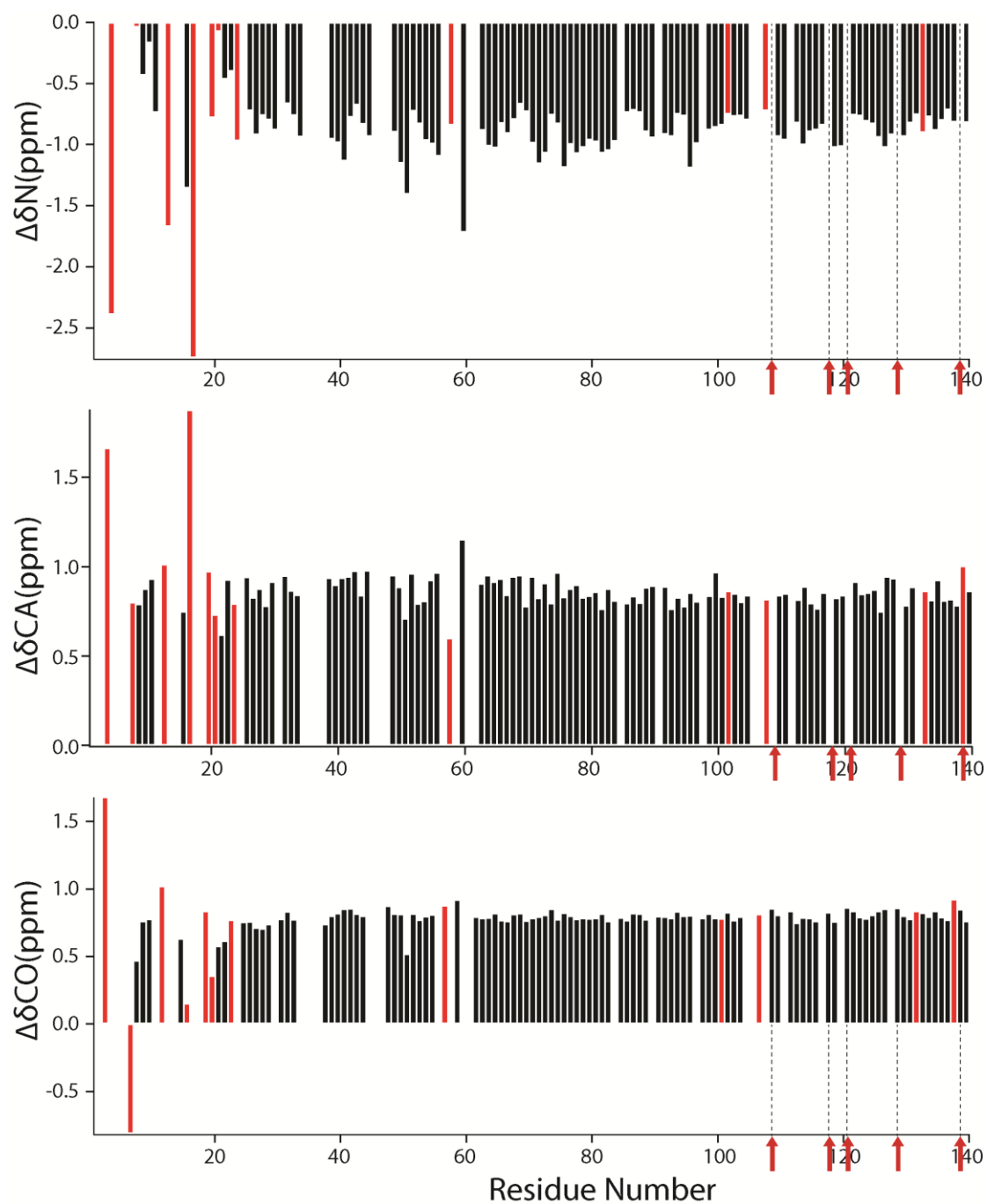
Figure 35. Chemical shift difference between assignments of α-synuclein at 298K, pD 6.0 from 6D HACACONCAH APSY and BMRB database (ID number: 6968). Black bars represent the residues with high and medium reliability and red bars represents the residues with low reliability from MARS assignment. The positions of proline residues are indicated by the red arrows and dashed lines.

For most residues, only very small chemical shift deviations are observed between the assignment of 6D HACACONCAH APSY and BMRB assignment except overall shift due to different referencing. The large chemical shift deviation

only appears at the residues with low reliable assignment (red bars). Proline residue 108, 117, 120, 128, 138 interrupt the sequential walk due to its low transfer efficiency as discussed in section 7.4. The N-terminal of a-synuclein possesses more residues unassigned or low reliably assigned with respect to other regions. This may be related to the relative higher secondary propensity in the N-terminus [223]. Notably, the partial assignment of Pro138 was directly achieved in our 6D HACACONCAH APSY although with a low reliability. This shows our 6D HACACONCAH APSY provides an opportunity to assign proline residues partially on its own and the acquired $^{13}C_\alpha$ and C′ chemical shift are ready for secondary structure analysis.

The statistics of assignment result for different types were calculated and listed in Table 11. First row is residue type with 1 letter, second and third rows are total numbers for each type of residue and the numbers of residues unassigned or assigned with low reliability. It shows glycines have the worst assignment except prolines. In glycine region of $C_\alpha(i)$-$H_\alpha(i)$ projection(Figure 36, left panel), the chemical shift degeneracy is very severe. $N(i)$-$H_\alpha(i)$ projection(Figure 36, right panel) indicates N chemical shift have better chemical shift dispersion and the spectral overlapping mainly comes from narrow $^1H_\alpha$ dispersion .

Table 11: The statistics of assignment result for different residue types

residue type	A	N	D	Q	E	G	H	I	L	K	M	F	P	S	T	Y	V
total numbers	19	3	6	6	18	18	1	2	4	15	4	2	5	4	10	4	19
unassigned or low reliable	8	0	1	1	7	9	0	0	0	7	2	1	5	0	0	0	3
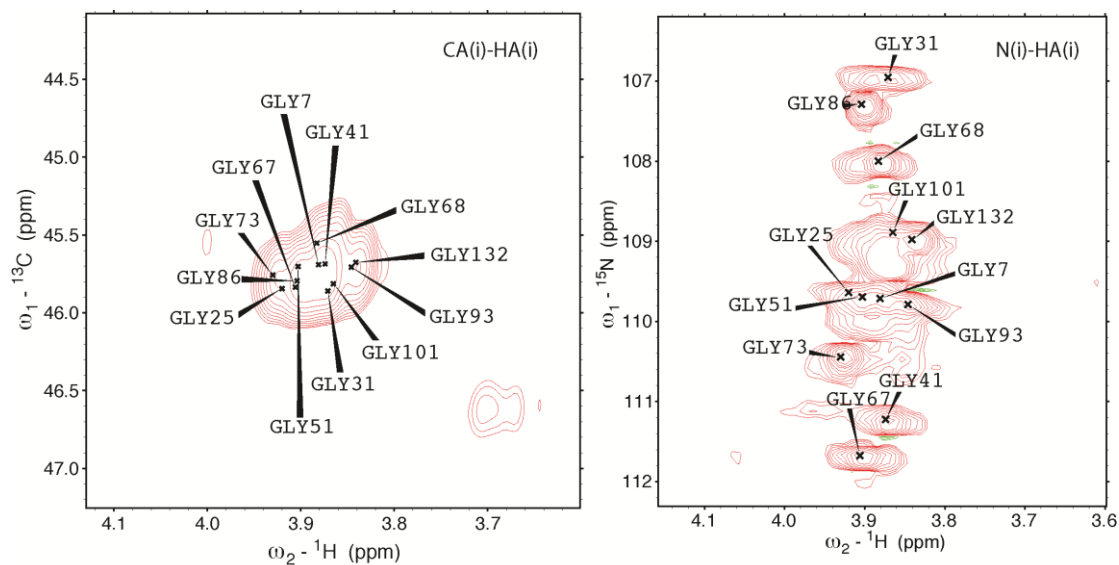
Figure 36. C$_\alpha$(i)-H$_\alpha$ (i) projection(left panel) and N(i)-H$_\alpha$ (i) projections (right panel) for glycine region. The residue labels are from assignment of 6D HACACONCAH APSY

The HA-detected APSY was also applied to α-synuclein at 283K and reliable sequence-specific assignments were obtained for 94 residues. Without amide proton exchange effect, the experiment at lower temperature loses partially its sensitivity due to unfavorable transverse relaxation (Figure 37). However, the result of APSY assignment at 283K is comparable with the one at 298K acquiring 95 reliably assigned residues indicating the efficiency of APSY assignment not solely depends on the sensitivity of experiments.
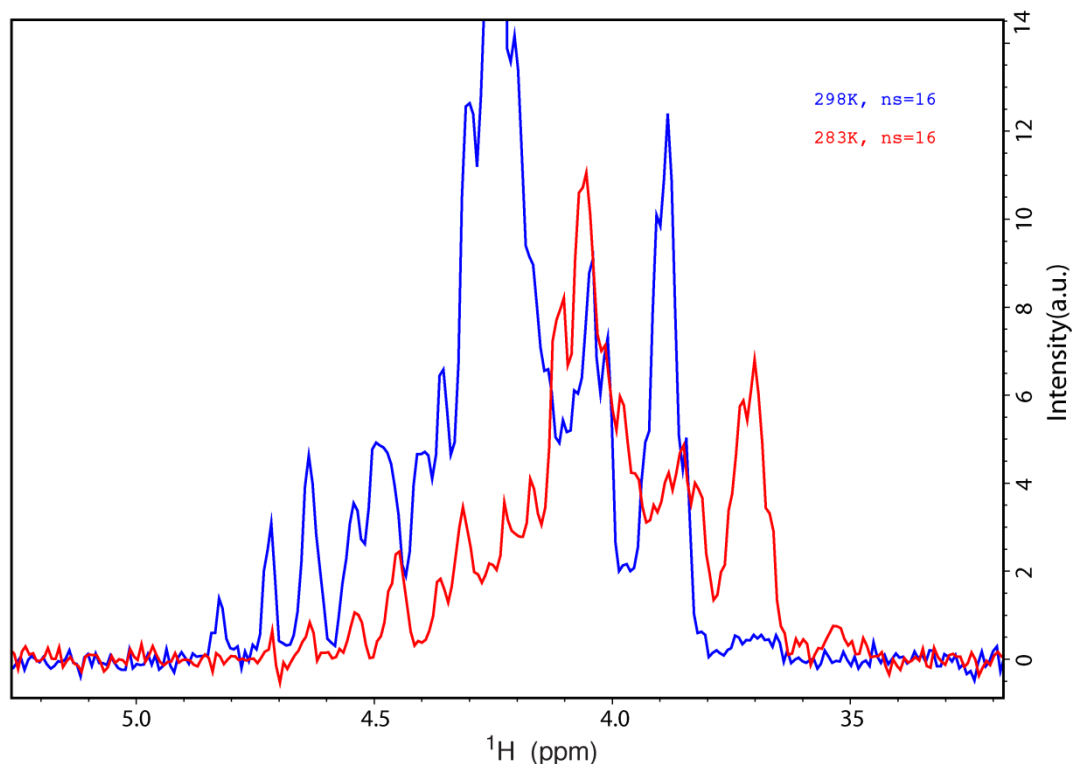
Figure 37. Overlay of first fid of 6D HACACONCAH APSY on α-synuclein at 298K (blue) and 283K (red), number of scan (ns) for each experiment is 16.

7.2.2   Sequence-specific assignment of htau23

htau23 consisting of 352 residues with 34 prolines is one of smallest isoform of 441-residues Tau proteins which bind to microtubules and play a key role in Alzheimer disease [224]. 6D HACACONCAH APSY of htau23 at pD 6.8 in D_2O was measured at 298K. 65 projections including 5 orthogonal projections (Figure 38) from 25 angle sets were recorded in 84.5 hours. Residue water signal exists even in D_2O buffer. The water signals were avoided during peak picking and almost not affected on the proteins signals. The automatic assignment was implemented in MARS using cutoff 0.05ppm for CA and 0.02ppm for HA and obtained 237 reliable and 28 low reliable sequence-specific assigned residues.
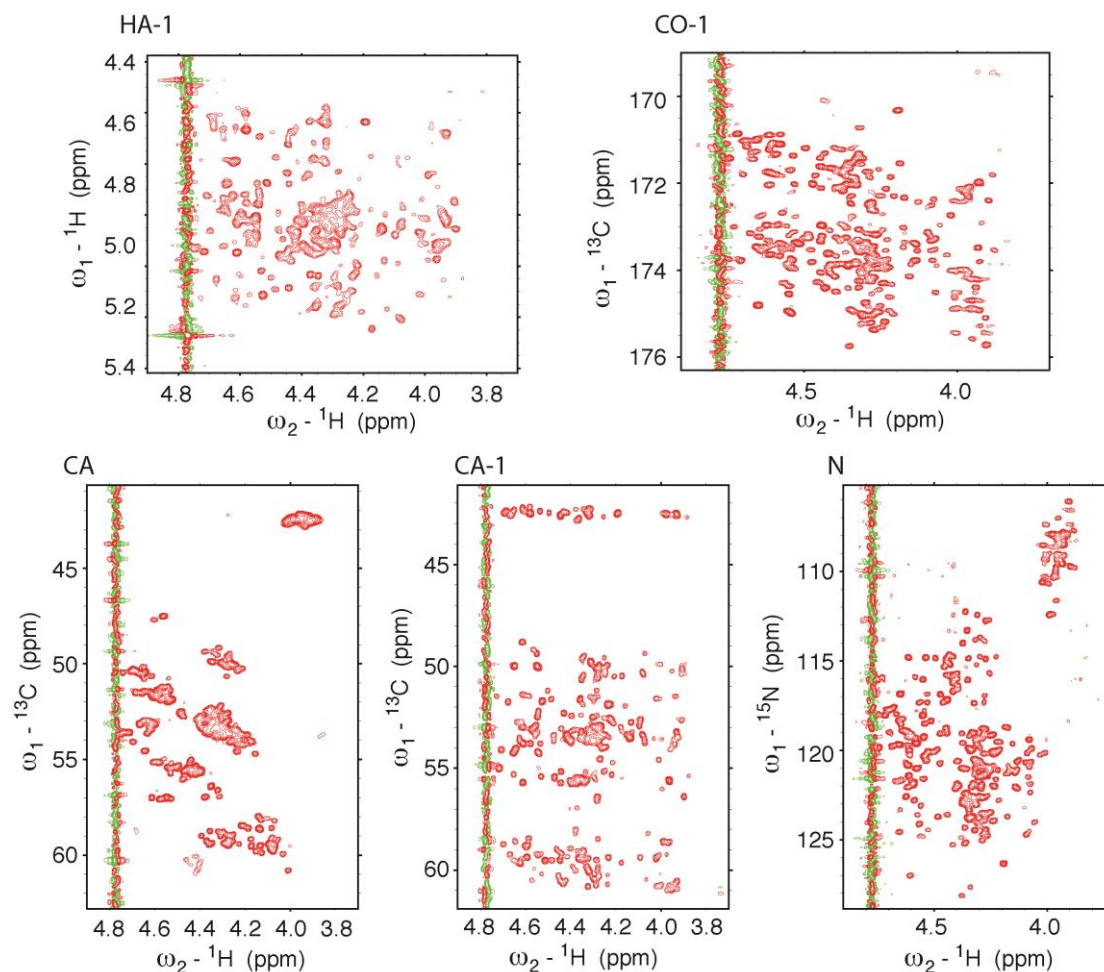
Figure 38. 2D projections $H_\alpha(i-1)$-$H_\alpha(i)$, $C'(i-1)$-$H_\alpha(i)$, $C_\alpha(i)$-$H_\alpha(i)$,$C_\alpha(i-1)$-$H_\alpha(i)$, $N(i)$-$H_\alpha(i)$ of the 6D HACACONCAH APSY experiment recorded on htau23 at 298K.

The effectiveness of our HA detection APSY was indicted by comparison of the result with previous htau23 assignment from our group (Figure 39) based on HN-detected APSY [119]. The chemical shift differences of $\Delta\delta C_\alpha$, $\Delta\delta C'$, $\Delta\delta N$ from 6D HACACONCAH APSY and assignment from HN-detected APSY [119] were plotted as a function of residue number. Basically assignment using 6D HACACONCAH APSY has good consistence with the one using HN-detected APSY. The red bar indicates the residues assigned with low reliability. Except these low reliable ones, there still some residues with medium or high reliability and also show large deviation such as His14, Ala15, Glu47, His32, His63, Lys83, His210, His285, His299 and His318. It should be noted that the previous assignment from our group was recorded on htau23 at pH 6 while the novel 6D HACACONCAH APSY was acquired on htau23 at pD 6.8 (in D_2O). We thus presume the large chemical shift deviations of

histidines mainly come from change of the sample condition as chemical shift of histidine is known to be sensitive to *p*H.
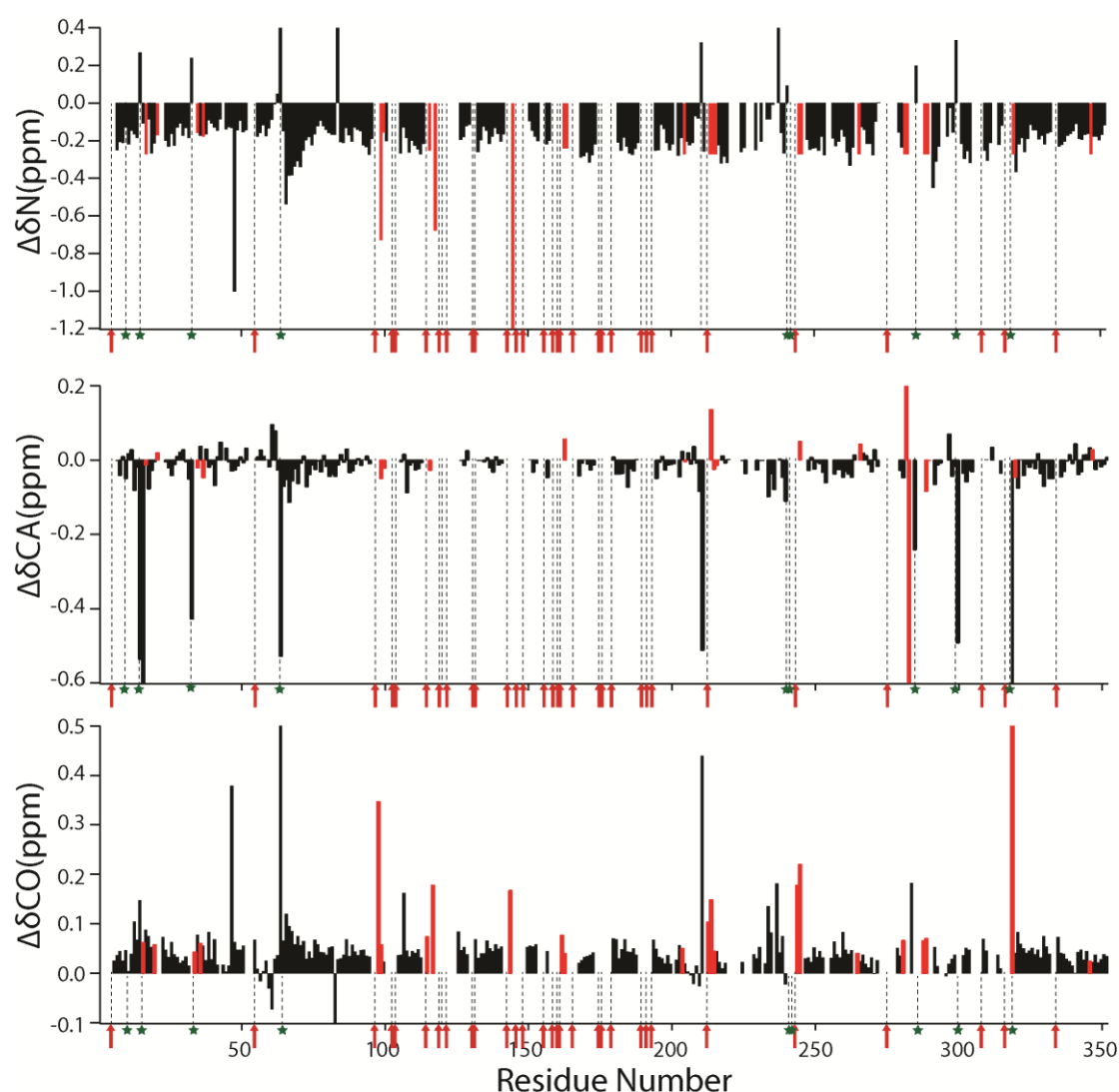
Figure 39. Chemical shift difference between assignments of htua23 from 6D HACACONCAH APSY and previous assignment using HN-detected APSY from our group [119]. Black bars represent the residues with high and medium reliability and red bars represent the residues with low reliability from assignment of 6D HACACONCAH APSY. The proline and histidine residues are indicated by the red arrows and green stars respectively.

As mentioned before, the signals on HN-based spectra could be deteriorated by fast amide proton exchange at high *p*H or temperature. The effect of high temperature on tau23 is presented in Figure 40. The [^{15}N, ^{1}H]-HSQC at 278K has decent sensitivity and resolution while at 308K the spectrum with same parameters is degraded by broadened line-shape. Many of peak resonances on the right spectrum disappear and additionally the overlap of spectrum is more severe, thus making the

sequence-specific assignment problematic. At this temperature, 6D HACACONCAH APSY probably is more effective for assignment than its HN-detected counterpart.
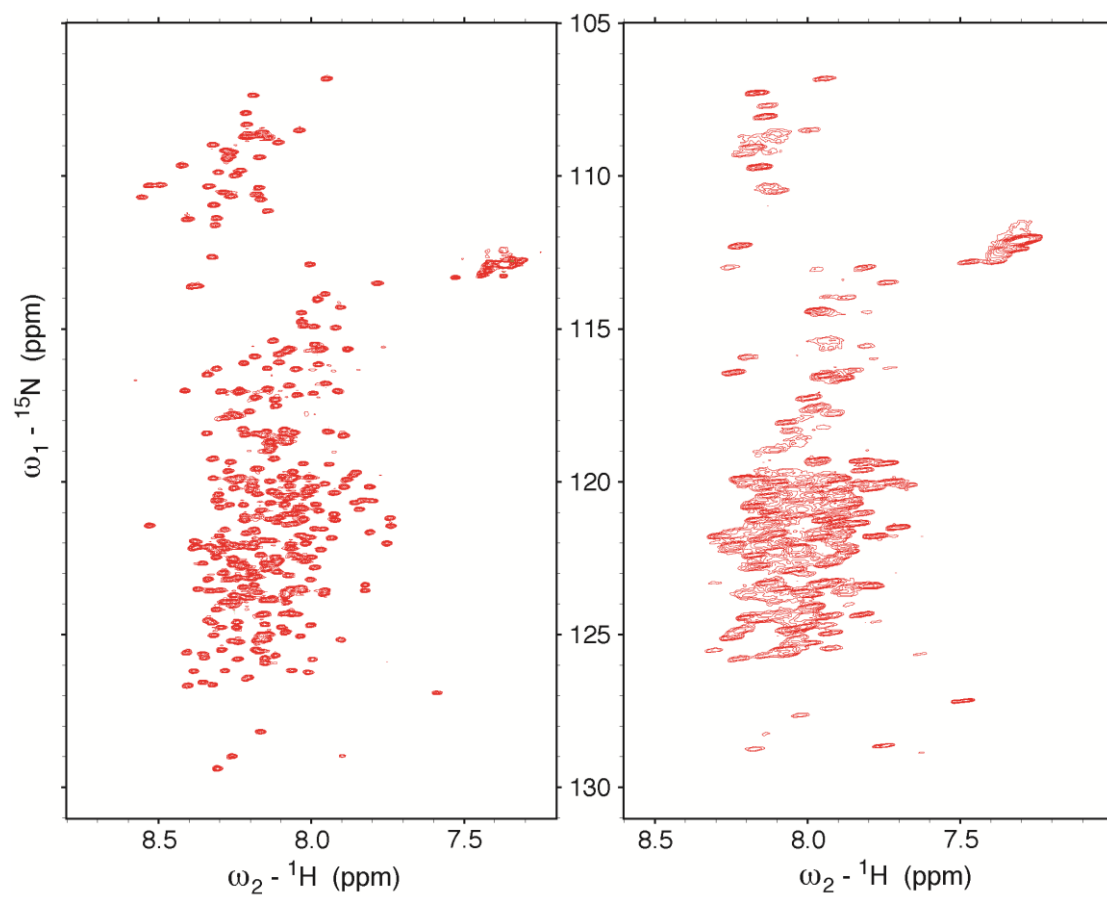
Figure 40. [^{15}N,^{1}H]-HSQC of htau23 at 278K(left) and 308K(right) , pH 6.8

8   Summary and Outlook

8.1   6D HACACONCAH APSY for assignment of IDPs

6D HACACONCAH APSY enables acquisition of high-dimensional correlation employing simultaneous chemical shift labeling of multiple indirect dimensions, and correlates six different chemical shifts $H_\alpha(i-1)$, $C_\alpha(i-1)$, $C'(i-1)$, $N(i)$, $C_\alpha(i)$ and $H_\alpha(i)$ in one cross peak. The sequential assignment has been facilitated by connecting through C_α-$H_\alpha(i)$ and C_α-$H_\alpha(i-1)$ pairs in one experiment. The undesired intraresidue transfer is suppressed because it doesn't provide any additional information and only cause spectral overlap by doubling the peaks. This is necessary for avoiding ambiguity in automatic assignment procedure. The Glycine signals are modulated to the same phase as other signals in order to avoid cancelling out signals. On the other hand, the attempting to bring back proline signals failed and thus resonance sequential walk of assignment is interrupted at the proline positions. The GAPRO algorithm performs peak picking on all projections and average the correlated peaks, creating a final peak list with very high precision. This will benefit the subsequent resonance matching of automatic backbone assignment. The implementation of 6D HACACONCAH APSY on 140-residue α-synuclein achieved sequence specific assignment for 95 of 135 non-proline residues. The validation with BMRB assignment has shown both the high and medium reliable assignments from MARS are correct. For a large unfolded protein htau23 with 318 non-proline residues, reliable sequence-specific assignment has been obtained by 6D HACACONCAH APSY for 237 residues.

In summary, 6D HACACONCAH APSY has successfully assigned more than 70% non-proline residues for both α-synuclein and tau23. The 352-residue unfolded

tau23 with 34 prolines is certainly a challenge for sequence specific assignment. The automatic assignment using 6D HACACONCAH APSY has been obtained within 85 hours of measurement and a few hours of calculation. This highly reduces the overall analysis time and avoids tedious manual assignment procedure compared with standard backbone assignment procedure. Therefore, our method is ready to be used as a quick assignment protocol for large IDPs.

8.2   Comparison with other methods for assignment of IDPs

The assignment of IDPs has benefited from the high resolution of high dimension obtained from 6D HACACONCAH APSY which couples the evolution of several different nuclei so that their chemical shifts appear in a joint spectral dimension. Some of reduced dimensionality techniques such GFT and Projection-Reconstruction usually implemented various back-projection scheme to reconstruction of the full-dimensional spectrum. The reconstructed spectra are similar with the conventional multidimensional spectra with independent chemical shift evolution in indirect dimension which is more straightforward for the manual assignment. However, spectrum reconstruction also could create a number of artifact peaks such as ridges along the projection directions and ghost peaks at the intersection of ridges from less than all projections [225]. On the other hands, in APSY a set of 2D projections are recorded and followed by automatic peak picking. The sequence-specific assignment has been implemented just by calculating and analyzing these peak lists which are directly obtained by automatic peak picking 2D projections. The final peak list has very high accuracy of peaks since it is averaged from correlated peaks from many projections. This high quality peak list has been a sound basis for automatic assignment.

HN-based NMR experiments are most widely used for protein backbone assignment. The resonance peaks on HN-detected spectra could be deteriorated by line broadening due to fast amide proton exchange at high pH and temperature as shown in the [^{15}N, ^{1}H]-HSQC spectra of htau23 at 278K and 308K(Figure 40). ^{13}C-detected experiments are also an alternative for assignment of IDPs due to more dispersed chemical shift of heteronuclear than proton. However, ^{13}C-detected

experiments have a lower sensitivity compared with ^{1}H-detected ones because of lower gyromagnetic ratio of ^{13}C.

To sum up, 6D HA-detected APSY NMR experiment takes advantage of high sensitivity of ^{1}H detection and avoids resonance broadening of the amide protons while APSY-based high dimensionality relieves the signal overlap facilitating automatic assignment of IDPs.

8.3   Outlook

Although 6D HACACONCAH APSY has succeeded in performing sequence-specific assignment on IDPs, including ^{13}C$_\beta$ chemical shift is able to further solve some ambiguity by differentiating the residue types. Therefore a 7D HACACBCONCAHA APSY could be more efficient for sequence-specific assignment of large IDPs and the determined ^{13}C$_\beta$ chemical shift also can be directly used as an indicator for secondary structure. For high-dimensional experiments, the signal loss from spin relaxation is severe because of numerous magnetization transfer steps. APSY above seven dimensions thus may not practically be very useful.

MARS is capable of merging peak lists from several APSY experiments generating additional correlation information [119]. Our 6D HACACONCAH APSY is ready to combine with 5D APSY-HACACONH and thus provide amide proton chemical shift which is necessary for investigation protein interaction and dynamics based on [^{15}N-^{1}H]-HSQC. In addition, the combination of 6D HACACONCAH APSY with 5D CBCACONH APSY is able to provide H$_N$ and C$_\beta$ chemical shift. Although the similar correlation is able to be determined by 7D HACACBCONCAHA APSY mentioned above, 6D and 5D APSY experiments have an advantage of higher sensitivity.

Application of 6D HACACONCAH APSY on globular proteins is limited to those with small molecular weight. Medium and large globular proteins are involved in rapid signal decay from large relaxation rate caused by small tumbling rate. However, the same correlation could be obtained by combining two APSY HACACON and HACANCO which generating correlation of H$_\alpha$(i-1)-C$_\alpha$(i-1)-C′(i-1)-N(i) and H$_\alpha$(i)-C$_\alpha$(i)-N(i)-C′(i-1) respectively. The overlapping nuclear pair N(i)-C′(i-

1) enables merging the two peak lists to a six-dimensional peak list of $H_\alpha(i-1)$-$C_\alpha(i-1)$-$C'(i-1)$-$N(i)$-$C_\alpha(i)$-$H_\alpha(i)$. Thus these coupled 4D APSY experiments are potentially applicable to globular proteins of medium or large size.

References

[1] D.J. Schnell, D.N. Hebert, Protein translocons: multifunctional mediators of protein translocation across membranes, Cell, 112 (2003) 491-505.

[2] W. Neupert, J.M. Herrmann, Translocation of proteins into mitochondria, Annual review of biochemistry, 76 (2007) 723-749.

[3] H.M. McBride, M. Neuspiel, S. Wasiak, Mitochondria: more than just a powerhouse, Current biology : CB, 16 (2006) R551-560.

[4] G. Attardi, G. Schatz, Biogenesis of mitochondria, Annual review of cell biology, 4 (1988) 289-333.

[5] A. Chacinska, C.M. Koehler, D. Milenkovic, T. Lithgow, N. Pfanner, Importing mitochondrial proteins: machineries and mechanisms, Cell, 138 (2009) 628-644.

[6] B. Mignotte, J.L. Vayssiere, Mitochondria and apoptosis, European journal of biochemistry, 252 (1998) 1-15.

[7] A. Sickmann, J. Reinders, Y. Wagner, C. Joppich, R. Zahedi, H.E. Meyer, B. Schonfisch, I. Perschil, A. Chacinska, B. Guiard, P. Rehling, N. Pfanner, C. Meisinger, The proteome of Saccharomyces cerevisiae mitochondria, Proceedings of the National Academy of Sciences of the United States of America, 100 (2003) 13207-13212.

[8] D.J. Pagliarini, S.E. Calvo, B. Chang, S.A. Sheth, S.B. Vafai, S.E. Ong, G.A. Walford, C. Sugiana, A. Boneh, W.K. Chen, D.E. Hill, M. Vidal, J.G. Evans, D.R. Thorburn, S.A. Carr, V.K. Mootha, A mitochondrial protein compendium elucidates complex I disease biology, Cell, 134 (2008) 112-123.

[9] T. Becker, L. Bottinger, N. Pfanner, Mitochondrial protein import: from transport pathways to an integrated network, Trends in biochemical sciences, 37 (2012) 85-91.

[10] A. Harsman, V. Kruger, P. Bartsch, A. Honigmann, O. Schmidt, S. Rao, C. Meisinger, R. Wagner, Protein conducting nanopores, Journal of physics. Condensed matter, 22 (2010) 454102.

[11] N. Bolender, A. Sickmann, R. Wagner, C. Meisinger, N. Pfanner, Multiple pathways for sorting mitochondrial precursor proteins, EMBO reports, 9 (2008) 42-49.

[12] M. van der Laan, M. Rissler, P. Rehling, Mitochondrial preprotein translocases as dynamic molecular machines, FEMS yeast research, 6 (2006) 849-861.

[13] D. Rapaport, How does the TOM complex mediate insertion of precursor proteins into the mitochondrial outer membrane?, The Journal of cell biology, 171 (2005) 419-423.

[14] Y. Abe, T. Shodai, T. Muto, K. Mihara, H. Torii, S. Nishikawa, T. Endo, D. Kohda, Structural basis of presequence recognition by the mitochondrial protein import receptor Tom20, Cell, 100 (2000) 551-560.

[15] N. Pfanner, Protein sorting: recognizing mitochondrial presequences, Current biology, 10 (2000) R412-415.

[16] P. Rehling, K. Model, K. Brandner, P. Kovermann, A. Sickmann, H.E. Meyer, W. Kuhlbrandt, R. Wagner, K.N. Truscott, N. Pfanner, Protein insertion into the mitochondrial inner membrane by a twin-pore translocase, Science, 299 (2003) 1747-1751.

[17] J. Dukanovic, D. Rapaport, Multiple pathways in the integration of proteins into the mitochondrial outer membrane, Biochimica et biophysica acta, 1808 (2011) 971-980.

[18] A. Chacinska, S. Pfannschmidt, N. Wiedemann, V. Kozjak, L.K. Sanjuan Szklarz, A. Schulze-Specking, K.N. Truscott, B. Guiard, C. Meisinger, N. Pfanner, Essential role of Mia40 in import and assembly of mitochondrial intermembrane space proteins, The EMBO journal, 23 (2004) 3735-3746.

[19] T. Becker, L.S. Wenz, V. Kruger, W. Lehmann, J.M. Muller, L. Goroncy, N. Zufall, T. Lithgow, B. Guiard, A. Chacinska, R. Wagner, C. Meisinger, N. Pfanner, The mitochondrial import protein Mim1 promotes biogenesis of multispanning outer membrane proteins, The Journal of cell biology, 194 (2011) 387-395.

[20] L. Becker, M. Bannwarth, C. Meisinger, K. Hill, K. Model, T. Krimmer, R. Casadio, K.N. Truscott, G.E. Schulz, N. Pfanner, R. Wagner, Preprotein translocase of the outer mitochondrial membrane: reconstituted Tom40 forms a characteristic TOM pore, Journal of molecular biology, 353 (2005) 1011-1020.

[21] K. Yamano, Y. Yatsukawa, M. Esaki, A.E. Hobbs, R.E. Jensen, T. Endo, Tom20 and Tom22 share the common signal recognition pathway in mitochondrial protein import, The Journal of biological chemistry, 283 (2008) 3799-3807.

[22] A. Chacinska, M. Lind, A.E. Frazier, J. Dudek, C. Meisinger, A. Geissler, A. Sickmann, H.E. Meyer, K.N. Truscott, B. Guiard, N. Pfanner, P. Rehling, Mitochondrial presequence translocase: switching between TOM tethering and motor recruitment involves Tim21 and Tim17, Cell, 120 (2005) 817-829.

[23] H. Yamamoto, M. Esaki, T. Kanamori, Y. Tamura, S. Nishikawa, T. Endo, Tim50 is a subunit of the TIM23 complex that links protein translocation across the outer and inner mitochondrial membranes, Cell, 111 (2002) 519-528.

[24] D. Mokranjac, D. Popov-Celeketic, K. Hell, W. Neupert, Role of Tim21 in mitochondrial translocation contact sites, The Journal of biological chemistry, 280 (2005) 23437-23440.

[25] K.N. Truscott, P. Kovermann, A. Geissler, A. Merlin, M. Meijer, A.J. Driessen, J. Rassow, N. Pfanner, R. Wagner, A presequence- and voltage-sensitive channel of the mitochondrial preprotein translocase formed by Tim23, Nature structural biology, 8 (2001) 1074-1082.

[26] M. Krayl, J.H. Lim, F. Martin, B. Guiard, W. Voos, A cooperative action of the ATP-dependent import motor complex and the inner membrane potential drives mitochondrial preprotein import, Molecular and cellular biology, 27 (2007) 411-425.

[27] F.N. Vogtle, S. Wortelkamp, R.P. Zahedi, D. Becker, C. Leidhold, K. Gevaert, J. Kellermann, W. Voos, A. Sickmann, N. Pfanner, C. Meisinger, Global analysis of the mitochondrial N-proteome identifies a processing peptidase critical for protein stability, Cell, 139 (2009) 428-439.

[28] E. Schleiff, J. Soll, Membrane protein insertion: mixing eukaryotic and prokaryotic concepts, EMBO reports, 6 (2005) 1023-1027.

[29] J.J. Lemasters, E. Holmuhamedov, Voltage-dependent anion channel (VDAC) as mitochondrial governator--thinking outside the box, Biochimica et biophysica acta, 1762 (2006) 181-190.

[30] J. Dukanovic, D. Rapaport, Multiple pathways in the integration of proteins into the mitochondrial outer membrane, Bba-Biomembranes, 1808 (2011) 971-980.

[31] S.J. Habib, T. Waizenegger, M. Lech, W. Neupert, D. Rapaport, Assembly of the TOB complex of mitochondria, The Journal of biological chemistry, 280 (2005) 6434-6440.

[32] T. Krimmer, D. Rapaport, M.T. Ryan, C. Meisinger, C.K. Kassenbrock, E. Blachly-Dyson, M. Forte, M.G. Douglas, W. Neupert, F.E. Nargang, N. Pfanner, Biogenesis of porin of the outer mitochondrial membrane involves an import pathway via receptors and the general import pore of the TOM complex, The Journal of cell biology, 152 (2001) 289-300.

[33] K. Model, C. Meisinger, T. Prinz, N. Wiedemann, K.N. Truscott, N. Pfanner, M.T. Ryan, Multistep assembly of the protein import channel of the mitochondrial outer membrane, Nature structural biology, 8 (2001) 361-370.

[34] D.M. Walther, D. Papic, M.P. Bos, J. Tommassen, D. Rapaport, Signals in bacterial beta-barrel proteins are functional in eukaryotic cells for targeting to and assembly in mitochondria, Proceedings of the National Academy of Sciences of the United States of America, 106 (2009) 2531-2536.

[35] S.C. Hoppins, F.E. Nargang, The Tim8-Tim13 complex of Neurospora crassa functions in the assembly of proteins into both mitochondrial membranes, The Journal of biological chemistry, 279 (2004) 12396-12405.

[36] N. Wiedemann, K.N. Truscott, S. Pfannschmidt, B. Guiard, C. Meisinger, N. Pfanner, Biogenesis of the protein import channel Tom40 of the mitochondrial outer membrane: intermembrane space components are involved in an early stage of the assembly pathway, The Journal of biological chemistry, 279 (2004) 18188-18194.

[37] V. Kozjak, N. Wiedemann, D. Milenkovic, C. Lohaus, H.E. Meyer, B. Guiard, C. Meisinger, N. Pfanner, An essential role of Sam50 in the protein sorting and assembly machinery of the mitochondrial outer membrane, The Journal of biological chemistry, 278 (2003) 48520-48523.

[38] S.A. Paschen, T. Waizenegger, T. Stan, M. Preuss, M. Cyrklaff, K. Hell, D. Rapaport, W. Neupert, Evolutionary conservation of biogenesis of beta-barrel membrane proteins, Nature, 426 (2003) 862-866.

[39] S. Kutik, D. Stojanovski, L. Becker, T. Becker, M. Meinecke, V. Kruger, C. Prinz, C. Meisinger, B. Guiard, R. Wagner, N. Pfanner, N. Wiedemann, Dissecting membrane insertion of mitochondrial beta-barrel proteins, Cell, 132 (2008) 1011-1024.

[40] T. Waizenegger, S.J. Habib, M. Lech, D. Mokranjac, S.A. Paschen, K. Hell, W. Neupert, D. Rapaport, Tob38, a novel essential component in the biogenesis of beta-barrel proteins of mitochondria, EMBO reports, 5 (2004) 704-709.

[41] N.C. Chan, T. Lithgow, The peripheral membrane subunits of the SAM complex function codependently in mitochondrial outer membrane biogenesis, Molecular biology of the Cell, 19 (2008) 126-136.

[42] C. Meisinger, M. Rissler, A. Chacinska, L.K. Szklarz, D. Milenkovic, V. Kozjak, B. Schonfisch, C. Lohaus, H.E. Meyer, M.P. Yaffe, B. Guiard, N. Wiedemann, N. Pfanner, The mitochondrial morphology protein Mdm10 functions in assembly of the preprotein translocase of the outer membrane, Developmental Cell, 7 (2004) 61-71.

[43] C. Meisinger, S. Pfannschmidt, M. Rissler, D. Milenkovic, T. Becker, D. Stojanovski, M.J. Youngman, R.E. Jensen, A. Chacinska, B. Guiard, N. Pfanner, N. Wiedemann, The morphology proteins Mdm12/Mmm1 function in the major beta-barrel assembly pathway of mitochondria, The EMBO journal, 26 (2007) 2229-2239.

[44] L. Ramage, T. Junne, K. Hahne, T. Lithgow, G. Schatz, Functional cooperation of mitochondrial protein import receptors in yeast, The EMBO journal, 12 (1993) 4115-4123.

[45] J.M. Herrmann, W. Neupert, Protein transport into mitochondria, Current opinion in microbiology, 3 (2000) 210-214.

[46] C.M. Koehler, New developments in mitochondrial assembly, Annual review of cell and developmental biology, 20 (2004) 309-335.

[47] P. Rehling, K. Brandner, N. Pfanner, Mitochondrial import and the twin-pore translocase, Nature reviews. Molecular cell biology, 5 (2004) 519-530.

[48] J. Brix, K. Dietmeier, N. Pfanner, Differential recognition of preproteins by the purified cytosolic domains of the mitochondrial import receptors Tom20, Tom22, and Tom70, The Journal of biological chemistry, 272 (1997) 20730-20735.

[49] T. Komiya, S. Rospert, C. Koehler, R. Looser, G. Schatz, K. Mihara, Interaction of mitochondrial targeting signals with acidic receptor domains along the protein import pathway: evidence for the 'acid chain' hypothesis, The EMBO journal, 17 (1998) 3886-3898.

[50] N. Wiedemann, V. Kozjak, T. Prinz, M.T. Ryan, C. Meisinger, N. Pfanner, K.N. Truscott, Biogenesis of yeast mitochondrial cytochrome c: a unique relationship to the TOM machinery, Journal of molecular biology, 327 (2003) 465-474.

[51] K. Model, T. Prinz, T. Ruiz, M. Radermacher, T. Krimmer, W. Kuhlbrandt, N. Pfanner, C. Meisinger, Protein translocase of the outer mitochondrial membrane: role of import receptors in the structural organization of the TOM complex, Journal of molecular biology, 316 (2002) 657-666.

[52] K.P. Kunkele, S. Heins, M. Dembowski, F.E. Nargang, R. Benz, M. Thieffry, J. Walz, R. Lill, S. Nussberger, W. Neupert, The preprotein translocation channel of the outer membrane of mitochondria, Cell, 93 (1998) 1009-1019.

[53] U. Ahting, C. Thun, R. Hegerl, D. Typke, F.E. Nargang, W. Neupert, S. Nussberger, The TOM core complex: the general protein import pore of the outer membrane of mitochondria, The Journal of cell biology, 147 (1999) 959-968.

[54] U. Ahting, M. Thieffry, H. Engelhardt, R. Hegerl, W. Neupert, S. Nussberger, Tom40, the pore-forming component of the protein-conducting TOM channel in the outer membrane of mitochondria, The Journal of cell biology, 153 (2001) 1151-1160.

[55] K. Hill, K. Model, M.T. Ryan, K. Dietmeier, F. Martin, R. Wagner, N. Pfanner, Tom40 forms the hydrophilic channel of the mitochondrial import pore for preproteins , Nature, 395 (1998) 516-521.

[56] H. Suzuki, T. Kadowaki, M. Maeda, H. Sasaki, J. Nabekura, M. Sakaguchi, K. Mihara, Membrane-embedded C-terminal segment of rat mitochondrial TOM40 constitutes protein-conducting pore with enriched beta-structure, The Journal of biological chemistry, 279 (2004) 50619-50629.

[57] M. Bayrhuber, T. Meins, M. Habeck, S. Becker, K. Giller, S. Villinger, C. Vonrhein, C. Griesinger, M. Zweckstetter, K. Zeth, Structure of the human voltage-dependent anion channel, Proceedings of the National Academy of Sciences of the United States of America, 105 (2008) 15370-15375.

[58] M. Pusnik, F. Charriere, P. Maser, R.F. Waller, M.J. Dagley, T. Lithgow, A. Schneider, The single mitochondrial porin of Trypanosoma brucei is the main metabolite transporter in the outer mitochondrial membrane, Molecular biology and evolution, 26 (2009) 671-680.

[59] K. Zeth, M. Thein, Porins in prokaryotes and eukaryotes: common themes and variations, The Biochemical journal, 431 (2010) 13-22.

[60] W.C. Wimley, Toward genomic identification of beta-barrel membrane proteins: composition and architecture of known structures, Protein science, 11 (2002) 301-312.

[61] M.S. Sansom, I.D. Kerr, Transbilayer pores formed by beta-barrels: molecular modeling of pore structures and properties, Biophysical journal, 69 (1995) 1334-1343.

[62] A. Pautsch, G.E. Schulz, Structure of the outer membrane protein A transmembrane domain, Nature structural biology, 5 (1998) 1013-1017.

[63] H. Remaut, C. Tang, N.S. Henderson, J.S. Pinkner, T. Wang, S.J. Hultgren, D.G. Thanassi, G. Waksman, H. Li, Fiber formation across the bacterial outer membrane by the chaperone/usher pathway, Cell, 133 (2008) 640-652.

[64] S. Hiller, R.G. Garces, T.J. Malia, V.Y. Orekhov, M. Colombini, G. Wagner, Solution structure of the integral human membrane protein VDAC-1 in detergent micelles, Science, 321 (2008) 1206-1210.

[65] R. Ujwal, D. Cascio, J.P. Colletier, S. Faham, J. Zhang, L. Toro, P. Ping, J. Abramson, The crystal structure of mouse VDAC1 at 2.3 A resolution reveals mechanistic insights into metabolite gating, Proceedings of the National Academy of Sciences of the United States of America, 105 (2008) 17742-17747.

[66] J.Y. Kinoshita, K. Mihara, T. Oka, Identification and characterization of a new tom40 isoform, a central component of mitochondrial outer membrane translocase, Journal of biochemistry, 141 (2007) 897-906.

[67] K. Zeth, Structure and evolution of mitochondrial outer membrane proteins of beta-barrel topology, Biochimica et biophysica acta, 1797 (2010) 1292-1299.

[68] C. Hunte, S. Richers, Lipids and membrane protein structures, Current opinion in structural biology, 18 (2008) 406-411.

[69] S.F. Poget, S.M. Cahill, M.E. Girvin, Isotropic bicelles stabilize the functional form of a small multidrug-resistance pump for NMR structural studies, Journal of the American Chemical Society, 129 (2007) 2432-2433.

[70] D. Lee, K.F. Walter, A.K. Bruckner, C. Hilty, S. Becker, C. Griesinger, Bilayer in small bicelles revealed by lipid-protein interactions using NMR spectroscopy, Journal of the American Chemical Society, 130 (2008) 13822-13823.

[71] E. Papadopoulos, K. Oglecka, L. Maler, J. Jarvet, P.E. Wright, H.J. Dyson, A. Graslund, NMR solution structure of the peptide fragment 1-30, derived from unprocessed mouse Doppel protein, in DHPC micelles, Biochemistry, 45 (2006) 159-166.

[72] J.M. Kielec, K.G. Valentine, C.R. Babu, A.J. Wand, Reverse micelles in integral membrane protein structural biology by solution NMR spectroscopy, Structure, 17 (2009) 345-351.

[73] P.F. Flynn, A.J. Wand, High-resolution nuclear magnetic resonance of encapsulated proteins dissolved in low viscosity fluids, Methods in enzymology, 339 (2001) 54-70.

[74] C.R. Babu, P.F. Flynn, A.J. Wand, Preparation, characterization, and NMR spectroscopy of encapsulated proteins dissolved in low viscosity fluids, Journal of Biomolecular NMR, 25 (2003) 313-323.

[75] W.D. Van Horn, M.E. Ogilvie, P.F. Flynn, Use of reverse micelles in membrane protein structural biology, Journal of Biomolecular NMR, 40 (2008) 203-211.

[76] J.L. Popot, E.A. Berry, D. Charvolin, C. Creuzenet, C. Ebel, D.M. Engelman, M. Flotenmeyer, F. Giusti, Y. Gohon, Q. Hong, J.H. Lakey, K. Leonard, H.A. Shuman, P. Timmins, D.E. Warschawski, F. Zito, M. Zoonens, B. Pucci, C. Tribet, Amphipols: polymeric surfactants for membrane biology research, Cellular and Molecular Life Science, 60 (2003) 1559-1574.

[77] C. Tribet, R. Audebert, J.L. Popot, Amphipols: polymers that keep membrane proteins soluble in aqueous solutions, Proceedings of the National Academy of Sciences of the United States of America, 93 (1996) 15047-15050.

[78] C.L. Pocanschi, T. Dahmane, Y. Gohon, F. Rappaport, H.J. Apell, J.H. Kleinschmidt, J.L. Popot, Amphipathic polymers: tools to fold integral membrane proteins to their active form, Biochemistry, 45 (2006) 13954-13961.

[79] B.M. Gorzelle, A.K. Hoffman, M.H. Keyes, D.N. Gray, D.G. Ray, C.R. Sanders, Amphipols can support the activity of a membrane enzyme, Journal of the American Chemical Society, 124 (2002) 11594-11595.

[80] S.H. Park, S. Berkamp, G.A. Cook, M.K. Chan, H. Viadiu, S.J. Opella, Nanodiscs versus macrodiscs for NMR of membrane proteins, Biochemistry, 50 (2011) 8983-8985.

[81] D.E. Warschawski, A.A. Arnold, M. Beaugrand, A. Gravel, E. Chartrand, I. Marcotte, Choosing membrane mimetics for NMR structural studies of transmembrane proteins, Biochimica et biophysica acta, 1808 (2011) 1957-1974.

[82] E.N. Lyukmanova, Z.O. Shenkarev, N.F. Khabibullina, G.S. Kopeina, M.A. Shulepko, A.S. Paramonov, K.S. Mineev, R.V. Tikhonov, L.N. Shingarova, L.E. Petrovskaya, D.A. Dolgikh, A.S. Arseniev, M.P. Kirpichnikov, Lipid-protein nanodiscs for cell-free production of integral membrane proteins in a soluble and folded state: comparison with detergent micelles, bicelles and liposomes, Biochimica et biophysica acta, 1818 (2012) 349-358.

[83] K. Mors, C. Roos, F. Scholz, J. Wachtveitl, V. Dotsch, F. Bernhard, C. Glaubitz, Modified lipid and protein dynamics in nanodiscs, Biochimica et biophysica acta, 1828 (2013) 1222-1229.

[84] T.H. Bayburt, J.W. Carlson, S.G. Sligar, Reconstitution and imaging of a membrane protein in a nanometer-size phospholipid bilayer, Journal of structural biology, 123 (1998) 37-44.

[85] T.H. Bayburt, S.G. Sligar, Membrane protein assembly into Nanodiscs, FEBS letters, 584 (2010) 1721-1727.

[86] I.G. Denisov, Y.V. Grinkova, A.A. Lazarides, S.G. Sligar, Directed self-assembly of monodisperse phospholipid bilayer Nanodiscs with controlled size, Journal of the American Chemical Society, 126 (2004) 3477-3487.

[87] M.A. Hanson, V. Cherezov, M.T. Griffith, C.B. Roth, V.P. Jaakola, E.Y. Chien, J. Velasquez, P. Kuhn, R.C. Stevens, A specific cholesterol binding site is established by the 2.8 A structure of the human beta2-adrenergic receptor, Structure, 16 (2008) 897-905.

[88] E.N. Lyukmanova, Z.O. Shenkarev, A.S. Paramonov, A.G. Sobol, T.V. Ovchinnikova, V.V. Chupin, M.P. Kirpichnikov, M.J. Blommers, A.S. Arseniev, Lipid-protein nanoscale bilayers: a versatile medium for NMR investigations of membrane proteins and membrane-active peptides, Journal of the American Chemical Society, 130 (2008) 2140-2141.

[89] T.H. Bayburt, S.G. Sligar, Self-assembly of single integral membrane proteins into soluble nanoscale phospholipid bilayers, Protein science, 12 (2003) 2476-2481.

[90] T. Raschle, S. Hiller, M. Etzkorn, G. Wagner, Nonmicellar systems for solution NMR spectroscopy of membrane proteins, Current opinion in structural biology, 20 (2010) 471-479.

[91] K. Pervushin, R. Riek, G. Wider, K. Wuthrich, Attenuated T2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution, Proceedings of the National Academy of Sciences of the United States of America, 94 (1997) 12366-12371.

[92] V. Tugarinov, P.M. Hwang, J.E. Ollerenshaw, L.E. Kay, Cross-correlated relaxation enhanced 1H[bond]13C NMR spectroscopy of methyl groups in very high molecular weight proteins and protein complexes, Journal of the American Chemical Society, 125 (2003) 10420-10428.

[93] K.H. Gardner, L.E. Kay, The use of 2H, 13C, 15N multidimensional NMR to study the structure and dynamics of proteins, Annual review of biophysics and biomolecular structure, 27 (1998) 357-406.

[94] P.S. Wu, K. Ozawa, S. Jergic, X.C. Su, N.E. Dixon, G. Otting, Amino-acid type identification in 15N-HSQC spectra by combinatorial selective 15N-labelling, Journal of Biomolecular NMR, 34 (2006) 13-21.

[95] F.I. Valiyaveetil, R. MacKinnon, T.W. Muir, Semisynthesis and folding of the potassium channel KcsA, Journal of the American Chemical Society, 124 (2002) 9113-9120.

[96] V. Tugarinov, V. Kanelis, L.E. Kay, Isotope labeling strategies for the study of high-molecular-weight proteins by solution NMR spectroscopy, Nature protocols, 1 (2006) 749-754.

[97] P.A. Luchette, T.N. Vetman, R.S. Prosser, R.E. Hancock, M.P. Nieh, C.J. Glinka, S. Krueger, J. Katsaras, Morphology of fast-tumbling bicelles: a small angle neutron scattering and NMR study, Biochimica et biophysica acta, 1513 (2001) 83-94.

[98] J. Lind, J. Nordin, L. Maler, Lipid dynamics in fast-tumbling bicelles with varying bilayer thickness: effect of model transmembrane peptides, Biochimica et biophysica acta, 1778 (2008) 2526-2534.

[99] R.M. Epand, Cholesterol and the interaction of proteins with membrane domains, Progress in lipid research, 45 (2006) 279-294.

[100] N. Tjandra, A. Bax, Direct measurement of distances and angles in biomolecules by NMR in a dilute liquid crystalline medium, Science, 278 (1997) 1111-1114.

[101] J.H. Prestegard, H.M. al-Hashimi, J.R. Tolman, NMR structures of biomolecules using field oriented media and residual dipolar couplings, Quarterly reviews of biophysics, 33 (2000) 371-424.

[102] X.C. Su, K. McAndrew, T. Huber, G. Otting, Lanthanide-binding peptides for NMR measurements of residual dipolar couplings and paramagnetic effects from

multiple angles, Journal of the American Chemical Society, 130 (2008) 1681-1687.

[103] D.H. Jones, S.J. Opella, Weak alignment of membrane proteins in stressed polyacrylamide gels, Journal of magnetic resonance, 171 (2004) 258-269.

[104] S.C. Howell, M.F. Mesleh, S.J. Opella, NMR structure determination of a membrane protein with two transmembrane helices in micelles: MerF of the bacterial mercury detoxification system, Biochemistry, 44 (2005) 5196-5206.

[105] T. Cierpicki, B. Liang, L.K. Tamm, J.H. Bushweller, Increasing the accuracy of solution NMR structures of membrane proteins by application of residual dipolar couplings. High-resolution structure of outer membrane protein A, Journal of the American Chemical Society, 128 (2006) 6947-6951.

[106] A. Bax, N. Tjandra, High-resolution heteronuclear NMR of human ubiquitin in an aqueous liquid crystalline medium, Journal of Biomolecular NMR, 10 (1997) 289-292.

[107] K.I. Tong, M. Yamamoto, T. Tanaka, A simple method for amino acid selective isotope labeling of recombinant proteins in E. coli, Journal of Biomolecular NMR, 42 (2008) 59-67.

[108] M. Hammarstrom, N. Hellgren, S. van Den Berg, H. Berglund, T. Hard, Rapid screening for improved solubility of small human proteins produced as fusion proteins in Escherichia coli, Protein science, 11 (2002) 313-321.

[109] S. Hiller, F. Fiorito, K. Wuthrich, G. Wider, Automated projection spectroscopy (APSY), Proceedings of the National Academy of Sciences of the United States of America, 102 (2005) 10876-10881.

[110] S.C. Panchal, N.S. Bhavesh, R.V. Hosur, Improved 3D triple resonance experiments, HNN and HN(C)N, for HN and 15N sequential correlations in (13C, 15N) labeled proteins: application to unfolded proteins, Journal of Biomolecular NMR, 20 (2001) 135-147.

[111] S. Schwarzinger, G.J. Kroon, T.R. Foss, J. Chung, P.E. Wright, H.J. Dyson, Sequence-dependent correction of random coil NMR chemical shifts, Journal of the American Chemical Society, 123 (2001) 2970-2978.

[112] B. Brutscher, N. Morelle, F. Cordier, D. Marion, Determination of an Initial Set of Noe-Derived Distance Constraints for the Structure Determination of N-15/C-13-Labeled Proteins, Journal of Magnetic Resonance, Series B, 109 (1995) 238-242.

[113] E. Kupce, R. Freeman, Projection-reconstruction technique for speeding up multidimensional NMR spectroscopy, Journal of the American Chemical Society, 126 (2004) 6429-6440.

[114] L.J. Catoire, M. Zoonens, C. van Heijenoort, F. Giusti, E. Guittet, J.L. Popot, Solution NMR mapping of water-accessible residues in the transmembrane beta-barrel of OmpX, European biophysics journal, 39 (2010) 623-630.

[115] M.E. Girvin, V.K. Rastogi, F. Abildgaard, J.L. Markley, R.H. Fillingame, Solution structure of the transmembrane H+-transporting subunit c of the F1F0 ATP synthase, Biochemistry, 37 (1998) 8817-8824.

[116] M.K. Cho, H.Y. Kim, C.O. Fernandez, S. Becker, M. Zweckstetter, Conserved core of amyloid fibrils of wild type and A30P mutant alpha-synuclein, Protein Science, 20 (2011) 387-395.

[117] P. Schanda, H. Van Melckebeke, B. Brutscher, Speeding up three-dimensional protein NMR experiments to a few minutes, Journal of the American Chemical Society, 128 (2006) 9042-9043.

[118] Y.S. Jung, M. Zweckstetter, Mars -- robust automatic backbone assignment of proteins, Journal of Biomolecular NMR, 30 (2004) 11-23.

[119] R.L. Narayanan, U.H. Durr, S. Bibow, J. Biernat, E. Mandelkow, M. Zweckstetter, Automatic assignment of the intrinsically disordered protein Tau with 441-residues, Journal of the American Chemical Society, 132 (2010) 11906-11907.

[120] S. Hiller, C. Wasmer, G. Wider, K. Wuthrich, Sequence-specific resonance assignment of soluble nonglobular proteins by 7D APSY-NMR spectroscopy, Journal of the American Chemical Society, 129 (2007) 10823-10828.

[121] F. Fiorito, S. Hiller, G. Wider, K. Wuthrich, Automated resonance assignment of proteins: 6D APSY-NMR, Journal of Biomolecular NMR, 35 (2006) 27-37.

[122] S. Hiller, G. Wider, K. Wuthrich, APSY-NMR with proteins: practical aspects and backbone assignment, Journal of Biomolecular NMR, 42 (2008) 179-195.

[123] J.A. Marsh, V.K. Singh, Z. Jia, J.D. Forman-Kay, Sensitivity of secondary structure propensities to sequence differences between alpha- and gamma-synuclein: implications for fibrillation, Protein Science, 15 (2006) 2795-2804.

[124] T. Endo, K. Yamano, Transport of proteins across or into the mitochondrial outer membrane, Biochimica et Biophysica Acta, 1803 (2010) 706-714.

[125] Y. Zhang, I-TASSER server for protein 3D structure prediction, BMC Bioinformatics, 9 (2008) 40.

[126] M.P. Schwartz, S. Huang, A. Matouschek, The structure of precursor proteins during import into mitochondria, The Journal of biological chemistry, 274 (1999) 12759-12764.

[127] D. Rapaport, W. Neupert, R. Lill, Mitochondrial protein import. Tom40 plays a major role in targeting and translocation of preproteins by forming a specific binding site for the presequence, The Journal of biological chemistry, 272 (1997) 18725-18731.

[128] T. Muto, T. Obita, Y. Abe, T. Shodai, T. Endo, D. Kohda, NMR identification of the Tom20 binding segment in mitochondrial presequences, Journal of molecular biology, 306 (2001) 137-143.

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

[130] S. Schwarzinger, G.J. Kroon, T.R. Foss, P.E. Wright, H.J. Dyson, Random coil chemical shifts in acidic 8 M urea: implementation of random coil shift data in NMRView, Journal of Biomolecular NMR, 18 (2000) 43-48.

[131] S.K. Buchanan, Beta-barrel proteins from bacterial outer membranes: structure, function and refolding, Current Opinion Structural Biology, 9 (1999) 455-461.

[132] S. Hiller, G. Wider, L.L. Imbach, K. Wuthrich, Interactions with hydrophobic clusters in the urea-unfolded membrane protein OmpX, Angewandte Chemie, 47 (2008) 977-981.

[133] C.E. Dempsey, L.J. Handcock, Hydrogen bond stabilities in membrane-reconstituted alamethicin from amide-resolved hydrogen-exchange measurements, Biophysical Journal, 70 (1996) 1777-1788.

[134] L. Czerski, O. Vinogradova, C.R. Sanders, NMR-Based amide hydrogen-deuterium exchange measurements for complex membrane proteins: development and critical evaluation, Journal of magnetic resonance, 142 (2000) 111-119.

[135] K.P. Kunkele, P. Juin, C. Pompa, F.E. Nargang, J.P. Henry, W. Neupert, R. Lill, M. Thieffry, The isolated complex of the translocase of the outer membrane of mitochondria. Characterization of the cation-selective and voltage-gated preprotein-conducting pore, The Journal of biological chemistry, 273 (1998) 31032-31039.

[136] F. Mager, D. Gessmann, S. Nussberger, K. Zeth, Functional refolding and characterization of two Tom40 isoforms from human mitochondria, The Journal of membrane biology, 242 (2011) 11-21.

[137] D. Gessmann, F. Mager, H. Naveed, T. Arnold, S. Weirich, D. Linke, J. Liang, S. Nussberger, Improving the resistance of a eukaryotic beta-barrel protein to thermal and chemical perturbations, Journal of molecular biology, 413 (2011) 150-161.

[138] H. Naveed, R. Jackups, Jr., J. Liang, Predicting weakly stable regions, oligomerization state, and protein-protein interfaces in transmembrane domains of outer membrane proteins, Proceedings of the National Academy of Sciences of the United States of America, 106 (2009) 12735-12740.

[139] D. Gessmann, N. Flinner, J. Pfannstiel, A. Schlosinger, E. Schleiff, S. Nussberger, O. Mirus, Structural elements of the mitochondrial preprotein-conducting channel Tom40 dissolved by bioinformatics and mass spectrometry, Biochimica et biophysica acta, 1807 (2011) 1647-1657.

[140] E.L. Sherman, R.D. Taylor, N.E. Go, F.E. Nargang, Effect of mutations in Tom40 on stability of the translocase of the outer mitochondrial membrane (TOM) complex, assembly of Tom40, and import of mitochondrial preproteins, The Journal of biological chemistry, 281 (2006) 22554-22565.

[141] M. Esaki, T. Kanamori, S. Nishikawa, I. Shin, P.G. Schultz, T. Endo, Tom40 protein import channel binds to non-native proteins and prevents their aggregation, Nature structural biology, 10 (2003) 988-994.

[142] J.H. Kleinschmidt, Membrane protein folding on the example of outer membrane protein A of Escherichia coli, Cellular and Molecular Life Sciences, 60 (2003) 1547-1558.

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

[144] P.E. Wright, H.J. Dyson, Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm, Journal of Molecular Biology, 293 (1999) 321-331.

[145] L.M. Iakoucheva, C.J. Brown, J.D. Lawson, Z. Obradovic, A.K. Dunker, Intrinsic disorder in cell-signaling and cancer-associated proteins, Journal of Molecular Biology, 323 (2002) 573-584.

[146] L.M. Iakoucheva, P. Radivojac, C.J. Brown, T.R. O'Connor, J.G. Sikes, Z. Obradovic, A.K. Dunker, The importance of intrinsic disorder for protein phosphorylation, Nucleic acids research, 32 (2004) 1037-1049.

[147] P. Tompa, Intrinsically unstructured proteins, Trends in biochemical sciences, 27 (2002) 527-533.

[148] V.N. Uversky, Multitude of binding modes attainable by intrinsically disordered proteins: a portrait gallery of disorder-based complexes, Chemical Society reviews, 40 (2011) 1623-1634.

[149] V.N. Uversky, Flexible nets of malleable guardians: intrinsically disordered chaperones in neurodegenerative diseases, Chemical reviews, 111 (2011) 1134-1166.

[150] V.N. Uversky, C.J. Oldfield, A.K. Dunker, Intrinsically disordered proteins in human diseases: introducing the D2 concept, Annual review of biophysics, 37 (2008) 215-246.

[151] D.K. Wilkins, S.B. Grimshaw, V. Receveur, C.M. Dobson, J.A. Jones, L.J. Smith, Hydrodynamic radii of native and denatured proteins measured by pulse field gradient NMR techniques, Biochemistry, 38 (1999) 16424-16431.

[152] S. Jeganathan, A. Hascher, S. Chinnathambi, J. Biernat, E.M. Mandelkow, E. Mandelkow, Proline-directed pseudo-phosphorylation at AT8 and PHF1 epitopes induces a compaction of the paperclip folding of Tau and generates a pathological (MC-1) conformation, The Journal of biological chemistry, 283 (2008) 32066-32076.

[153] V.N. Uversky, J.R. Gillespie, I.S. Millett, A.V. Khodyakova, R.N. Vasilenko, A.M. Vasiliev, I.L. Rodionov, G.D. Kozlovskaya, D.A. Dolgikh, A.L. Fink, S. Doniach, E.A. Permyakov, V.M. Abramov, Zn(2+)-mediated structure formation and compaction of the "natively unfolded" human prothymosin alpha, Biochemical and biophysical research communications, 267 (2000) 663-668.

[154] T. Mittag, J.D. Forman-Kay, Atomic-level characterization of disordered protein ensembles, Current opinion in structural biology, 17 (2007) 3-14.

[155] F.A. Mulder, M. Filatov, NMR chemical shift data and ab initio shielding calculations: emerging tools for protein structure determination, Chemical Society reviews, 39 (2010) 578-590.

[156] R. Otten, K. Wood, F.A.A. Mulder, Comprehensive determination of (3)J(HNH alpha) for unfolded proteins using C-13'-resolved spin-echo difference spectroscopy, Journal of Biomolecular NMR, 45 (2009) 343-349.

[157] G.W. Vuister, 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, 115 (1993) 7772-7777.

[158] A. Bax, G. Kontaxis, N. Tjandra, Dipolar couplings in macromolecular structure determination, Methods in enzymology, 339 (2001) 127-174.

[159] N. Rezaei-Ghaleh, M. Blackledge, M. Zweckstetter, Intrinsically disordered proteins: from sequence and conformational properties toward drug discovery, Chembiochem, 13 (2012) 930-950.

[160] J.J. Chou, S. Gaemers, B. Howder, J.M. Louis, A. Bax, A simple apparatus for generating stretched polyacrylamide gels, yielding uniform alignment of proteins and detergent micelles, Journal of Biomolecular NMR, 21 (2001) 377-382.

[161] K. Ding, A.M. Gronenborn, Sensitivity-enhanced E.COSY-type HSQC experiments for accurate measurements of one-bond 15N-1H(N) and 15N-13C'

and two-bond 13C'-1H(N) residual dipolar couplings in proteins, Journal of magnetic resonance, 158 (2002) 173-177.

[162] J.J. Chou, F. Delaglio, A. Bax, Measurement of one-bond 15N-13C' dipolar couplings in medium sized proteins, Journal of Biomolecular NMR, 18 (2000) 101-105.

[163] D. Eliezer, Biophysical characterization of intrinsically disordered proteins, Current opinion in structural biology, 19 (2009) 23-30.

[164] G.M. Clore, J. Iwahara, Theory, practice, and applications of paramagnetic relaxation enhancement for the characterization of transient low-population states of biological macromolecules and their complexes, Chemical reviews, 109 (2009) 4108-4139.

[165] H.J. Dyson, P.E. Wright, Unfolded proteins and protein folding studied by NMR, Chemical reviews, 104 (2004) 3607-3622.

[166] J. Cavanagh, W. Fairbrother, Palmer III, A., M. Rance, N. Skelton, Protein NMR spectroscopy : principles and practice, 2007.

[167] M. Sattler, J. Schleucher, C. Griesinger, Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients, Progress in nuclear magnetic resonance spectroscopy, 34 (1999) 93-158.

[168] G. Zhu, X.J. Yao, TROSY-based NMR experiments for NMR studies of large biomolecules, Progress in nuclear magnetic resonance spectroscopy, 52 (2008) 49-68.

[169] T. Yamazaki, W. Lee, C.H. Arrowsmith, D.R. Muhandiram, L.E. Kay, A Suite of Triple-Resonance Nmr Experiments for the Backbone Assignment of N-15, C-13, H-2 Labeled Proteins with High-Sensitivity, Journal of the American Chemical Society, 116 (1994) 11655-11666.

[170] J.A. Marsh, J.D. Forman-Kay, Sequence determinants of compaction in intrinsically disordered proteins, Biophysical journal, 98 (2010) 2383-2390.

[171] Z. Serber, C. Richter, D. Moskau, J.M. Bohlen, T. Gerfin, D. Marek, M. Haberli, L. Baselgia, F. Laukien, A.S. Stern, J.C. Hoch, V. Dotsch, New carbon-detected protein NMR experiments using CryoProbes, Journal of the American Chemical Society, 122 (2000) 3554-3555.

[172] S.T.D. Hsu, C.W. Bertoncini, C.M. Dobson, Use of Protonless NMR Spectroscopy To Alleviate the Loss of Information Resulting from Exchange-Broadening, Journal of the American Chemical Society, 131 (2009) 7222-7223.

[173] W. Bermel, I. Bertini, I.C. Felli, M. Matzapetakis, R. Pierattelli, E.C. Theli, P. Turano, A method for C-alpha direct-detection in protonless NMR, Journal of Magnetic Resonance, 188 (2007) 301-310.

[174] I. Bertini, L. Duma, I.C. Felli, M. Fey, C. Luchinat, R. Pierattelli, P.R. Vasos, A heteronuclear direct-detection NMR spectroscopy experiment for protein-backbone assignment, Angewandte Chemie International Edition, 43 (2004) 2257-2259.

[175] W. Bermel, I. Bertini, I.C. Felli, Y.M. Lee, C. Luchinat, R. Pierattelli, Protonless NMR experiments for sequence-specific assignment of backbone nuclei in unfolded proteins, Journal of American Chemical Society, 128 (2006) 3918-3919.

[176] W. Bermel, I. Bertini, J. Chill, I.C. Felli, N. Haba, M.V.V. Kumar, R. Pierattelli, Exclusively Heteronuclear C-13-Detected Amino-Acid-Selective NMR Experiments for the Study of Intrinsically Disordered Proteins (IDPs), Chembiochem, 13 (2012) 2425-2432.

[177] W. Bermel, I. Bertini, I.C. Felli, L. Gonnelli, W. Kozminski, A. Piai, R. Pierattelli, J. Stanek, Speeding up sequence specific assignment of IDPs, Journal of Biomolecular NMR, 53 (2012) 293-301.

[178] K. Pervushin, A. Eletsky, A new strategy for backbone resonance assignment in large proteins using a MQ-HACACO experiment, Journal of Biomolecular NMR, 25 (2003) 147-152.

[179] W. Bermel, I. Bertini, I.C. Felli, R. Kummerle, R. Pierattelli, Novel 13C direct detection experiments, including extension to the third dimension, to perform the complete assignment of proteins, Journal of magnetic resonance, 178 (2006) 56-64.

[180] T. Mittag, L.E. Kay, J.D. Forman-Kay, Protein dynamics and conformational disorder in molecular recognition, Journal of molecular recognition, 23 (2010) 105-116.

[181] K. Takeuchi, D.P. Frueh, S.G. Hyberts, Z.Y. Sun, G. Wagner, High-resolution 3D CANCA NMR experiments for complete mainchain assignments using C(alpha) direct detection, Journal of the American Chemical Society, 132 (2010) 2945-2951.

[182] N. Shimba, H. Kovacs, A.S. Stern, A.M. Nomura, I. Shimada, J.C. Hoch, C.S. Craik, V. Dotsch, Optimization of 13C direct detection NMR methods, Journal of Biomolecular NMR, 30 (2004) 175-179.

[183] I. Bertini, I.C. Felli, L. Gonnelli, M.V. Vasantha Kumar, R. Pierattelli, High-resolution characterization of intrinsic disorder in proteins: expanding the suite of (13)C-detected NMR spectroscopy experiments to determine key observables, Chembiochem, 12 (2011) 2347-2352.

[184] I. Bertini, I.C. Felli, L. Gonnelli, R. Pierattelli, Z. Spyranti, G.A. Spyroulias, Mapping protein-protein interaction by 13C'-detected heteronuclear NMR spectroscopy, Journal of Biomolecular NMR, 36 (2006) 111-122.

[185] G. Pasat, J.S. Zintsmaster, J.W. Peng, Direct 13C-detection for carbonyl relaxation studies of protein dynamics, Journal of magnetic resonance, 193 (2008) 226-232.

[186] A.C. Wang, S. Grzesiek, R. Tschudin, P.J. Lodi, A. Bax, Sequential backbone assignment of isotopically enriched proteins in D2O by deuterium-decoupled HA(CA)N and HA(CACO)N, Journal of Biomolecular NMR, 5 (1995) 376-382.

[187] V. Kanelis, L. Donaldson, D.R. Muhandiram, D. Rotin, J.D. Forman-Kay, L.E. Kay, Sequential assignment of proline-rich regions in proteins: Application to modular binding domain complexes, Journal of Biomolecular NMR, 16 (2000) 253-259.

[188] M.J. Bottomley, M.J. Macias, Z. Liu, M. Sattler, A novel NMR experiment for the sequential assignment of proline residues and proline stretches in 13C/15N-labeled proteins, Journal of Biomolecular NMR, 13 (1999) 381-385.

[189] S. Mantylahti, O. Aitio, M. Hellman, P. Permi, HA-detected experiments for the backbone assignment of intrinsically disordered proteins, Journal of Biomolecular NMR, 47 (2010) 171-181.

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

[191] I.C. Felli, B. Brutscher, Recent advances in solution NMR: fast methods and heteronuclear direct detection, Chemphyschem, 10 (2009) 1356-1368.

[192] P. Schanda, Fast-pulsing longitudinal relaxation optimized techniques: Enriching the toolbox of fast biomolecular NMR spectroscopy, Progress in nuclear magnetic resonance spectroscopy, 55 (2009) 238-265.

[193] D. Malmodin, M. Billeter, High-throughput analysis of protein NMR spectra, Progress in nuclear magnetic resonance spectroscopy, 46 (2005) 109-129.

[194] K. Pervushin, B. Vogeli, A. Eletsky, Longitudinal (1)H relaxation optimization in TROSY NMR spectroscopy, Journal of the American Chemical Society, 124 (2002) 12898-12902.

[195] P. Schanda, B. Brutscher, Very fast two-dimensional NMR spectroscopy for real-time investigation of dynamic events in proteins on the time scale of seconds, Journal of the American Chemical Society, 127 (2005) 8014-8015.

[196] P. Schanda, E. Kupce, B. Brutscher, SOFAST-HMQC experiments for recording two-dimensional heteronuclear correlation spectra of proteins within a few seconds, Journal of Biomolecular NMR, 33 (2005) 199-211.

[197] E. Lescop, P. Schanda, B. Brutscher, A set of BEST triple-resonance experiments for time-optimized protein resonance assignment, Journal of magnetic resonance, 187 (2007) 163-169.

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

[199] Z. Solyom, M. Schwarten, L. Geist, R. Konrat, D. Willbold, B. Brutscher, BEST-TROSY experiments for time-efficient sequential resonance assignment of large disordered proteins, Journal of Biomolecular NMR, 55 (2013) 311-321.

[200] S. Kim, T. Szyperski, GFT NMR, a new approach to rapidly obtain precise high-dimensional NMR spectral information, Journal of the American Chemical Society, 125 (2003) 1385-1393.

[201] H.S. Atreya, T. Szyperski, G-matrix Fourier transform NMR spectroscopy for complete protein resonance assignment, Proceedings of the National Academy of Sciences of the United States of America, 101 (2004) 9642-9647.

[202] Y. Shen, H.S. Atreya, G.H. Liu, T. Szyperski, G-matrix Fourier transform NOESY-based protocol for high-quality protein structure determination, Journal of the American Chemical Society, 127 (2005) 9085-9099.

[203] T. Szyperski, H.S. Atreya, Principles and applications of GFT projection NMR spectroscopy, Magnetic resonance in chemistry , 44 Spec No (2006) S51-60.

[204] E. Kupce, R. Freeman, Projection-reconstruction of three-dimensional NMR spectra, Journal of the American Chemical Society, 125 (2003) 13958-13959.

[205] E. Kupce, R. Freeman, Reconstruction of the three-dimensional NMR spectrum of a protein from a set of plane projections, Journal of Biomolecular NMR, 27 (2003) 383-387.

[206] B.E. Coggins, R.A. Venters, P. Zhou, Generalized reconstruction of n-D NMR spectra from multiple projections: application to the 5-D HACACONH spectrum of protein G B1 domain, Journal of the American Chemical Society, 126 (2004) 1000-1001.

[207] J.C.J. Barna, E.D. Laue, M.R. Mayger, J. Skilling, S.J.P. Worrall, Exponential Sampling, an Alternative Method for Sampling in Two-Dimensional Nmr Experiments, Journal of magnetic resonance, 73 (1987) 69-77.

[208] R.A. Chylla, J.L. Markley, Theory and application of the maximum likelihood principle to NMR parameter estimation of multidimensional NMR data, Journal of Biomolecular NMR, 5 (1995) 245-258.

[209] D.M. Korzhneva, I.V. Ibraghimov, M. Billeter, V.Y. Orekhov, MUNIN: application of three-way decomposition to the analysis of heteronuclear NMR relaxation data, Journal of Biomolecular NMR, 21 (2001) 263-268.

[210] V.Y. Orekhov, I.V. Ibraghimov, M. Billeter, MUNIN: a new approach to multi-dimensional NMR spectra interpretation, Journal of Biomolecular NMR, 20 (2001) 49-60.

[211] B.E. Coggins, P. Zhou, Polar Fourier transforms of radially sampled NMR data, Journal of magnetic resonance, 182 (2006) 84-95.

[212] K. Kazimierczuk, W. Kozminski, I. Zhukov, Two-dimensional Fourier transform of arbitrarily sampled NMR data sets, Journal of magnetic resonance, 179 (2006) 323-328.

[213] D.J. Holland, M.J. Bostock, L.F. Gladden, D. Nietlispach, Fast Multidimensional NMR Spectroscopy Using Compressed Sensing, Angewandte Chemie, 50 (2011) 6548-6551.

[214] K. Kazimierczuk, V.Y. Orekhov, Accelerated NMR spectroscopy by using compressed sensing, Angewandte Chemie, 50 (2011) 5556-5559.

[215] P. Güntert, V. Dötsch, G. Wider, K. Wüthrich, Processing of multi-dimensional NMR data with the new software PROSA, Journal of Biomolecular NMR, 2 (1992) 619-629.

[216] T. Herrmann, P. Guntert, K. Wuthrich, Protein NMR structure determination with automated NOE-identification in the NOESY spectra using the new software ATNOS, Journal of Biomolecular NMR, 24 (2002) 171-189.

[217] J. Volk, T. Herrmann, K. Wuthrich, Automated sequence-specific protein NMR assignment using the memetic algorithm MATCH, Journal of Biomolecular NMR, 41 (2008) 127-138.

[218] S. Hiller, R. Joss, G. Wider, Automated NMR assignment of protein side chain resonances using automated projection spectroscopy (APSY), Journal of the American Chemical Society, 130 (2008) 12073-12079.

[219] B. Krahenbuhl, S. Hiller, G. Wider, 4D APSY-HBCB(CG)CDHD experiment for automated assignment of aromatic amino acid side chains in proteins, Journal of Biomolecular NMR, 51 (2011) 313-318.

[220] L.E. Kay, P. Keifer, T. Saarinen, Pure Absorption Gradient Enhanced Heteronuclear Single Quantum Correlation Spectroscopy with Improved Sensitivity, Journal of the American Chemical Society, 114 (1992) 10663-10665.

[221] M.A. Mccoy, L. Mueller, Nonresonant Effects of Frequency-Selective Pulses, Journal of magnetic resonance, 99 (1992) 18-36.

[222] F. Delaglio, D.A. Torchia, A. Bax, Measurement of 15N-13C J couplings in staphylococcal nuclease, Journal of Biomolecular NMR, 1 (1991) 439-446.

[223] M.K. Cho, G. Nodet, H.Y. Kim, M.R. Jensen, P. Bernado, C.O. Fernandez, S. Becker, M. Blackledge, M. Zweckstetter, Structural characterization of alpha-synuclein in an aggregation prone state, Protein science, 18 (2009) 1840-1846.

[224] H. Braak, E. Braak, Neuropathological stageing of Alzheimer-related changes, Acta neuropathologica, 82 (1991) 239-259.

[225] D. Malmodin, M. Billeter, Multiway decomposition of NMR spectra with coupled evolution periods, Journal of the American Chemical Society, 127 (2005) 13486-13487.

Appendix

Appendix 1

 Pulse sequence BEST HSQC (b_hsqcetf3gpsi)
;b_hsqcetf3gpsi
;avance-version (07/08/30)
;(best)-HSQC
;2D H-1/X correlation via double inept transfer
; using sensitivity improvement
;phase sensitive using Echo/Antiecho-TPPI gradient selection
;with decoupling during acquisition
;using f3 - channel
;with gradients in back-inept
;
;(A.G. Palmer III, J. Cavanagh, P.E. Wright & M. Rance, J. Magn.
; Reson. 93, 151-170 (1991))
;(L.E. Kay, P. Keifer & T. Saarinen, J. Am. Chem. Soc. 114,
; 10663-5 (1992))
;(J. Schleucher, M. Schwendinger, M. Sattler, P. Schmidt,
; O. Schedletzky, S.J. Glaser, O.W. Sorensen & C. Griesinger,
; J. Biomol. NMR 4, 301-306 (1994))
;
;$CLASS=HighRes
;$DIM=2D
;$TYPE=
;$SUBTYPE=
;$COMMENT=

prosol relations=<triple>

#include <Avance.incl>
#include <Grad.incl>
#include <Delay.incl>

"p22=p21*2"
"d11=30m"
"d26=1s/(cnst4*4)"

"p29=300u"
"p1=35.5u"
"p44=p1*8"

;calculate power according p1=35.5
"sp25=pl1+17.97"
"sp27=pl1+17.97"
"sp26=pl1+4.54"
"sp28=pl1+7.87"

```
"sp29=pl1+7.87"
"sp30=pl1"
"d0=3u"

"in0=inf1/2"

"DELTA1=d26-p29-d16-p41*cnst41-p42*cnst40"
"DELTA2=d26-p19-d16-p42/2"
"DELTA3=d26-p29-d16-p43*cnst43-p42*cnst42"
"DELTA4=p16+d16+de+8u"

#   ifdef LABEL_CN
"DELTA=p43-d0*2-p19-larger(p8,p44)"
#   else
"DELTA=p43-d0*2-p19-p44"
#   endif /*LABEL_CN*/


"spoff25=bf1*(cnst19/1000000)-o1"
"spoff26=bf1*(cnst19/1000000)-o1"
"spoff27=bf1*(cnst19/1000000)-o1"
"spoff28=bf1*(cnst19/1000000)-o1"
"spoff29=bf1*(cnst19/1000000)-o1"
"spoff30=0"


1 ze
 d11 pl26:f3
2 d11 do:f3
3 d1
 50u UNBLKGRAD

 (p41:sp25 ph1)
 p29:gp3
 d16
 DELTA1 pl3:f3
 (center (p42:sp26 ph1) (p22 ph6):f3 )
 DELTA1
 p29:gp3
 d16
 (p41:sp27 ph2):f1

 p16:gp4
 d16

 (p44:sp30 ph1)

 (p21 ph3):f3
 DELTA
 d0
```

```
#   ifdef LABEL_CN
 (center (p44:sp30 ph7) (p8:sp13 ph1):f2 )
#   else
 (p44:sp30 ph7)
#   endif /*LABEL_CN*/

 d0
 p19:gp1*EA
 d16
 (p22 ph4):f3
 d16

 (p43:sp28 ph1)

 (p21 ph4):f3
 p19:gp5
 d16
 DELTA2
 (center (p42:sp26 ph1) (p22 ph1):f3 )
 DELTA2
 p19:gp5
 d16
 (p21 ph5):f3

 (p43:sp29 ph2)
 p29:gp6
 d16
 DELTA3
 (center (p42:sp26 ph1) (p22 ph1):f3 )
 DELTA3
 p29:gp6
 d16
 (p43:sp28 ph1)
 DELTA4
 (p42:sp26 ph1)
 4u
 p19:gp2
 d16 pl26:f3
 4u BLKGRAD
 go=2 ph31 cpd3:f3
 d11 do:f3 mc #0 to 2
   F1EA(igrad EA & ip5*2, id0 & ip3*2 & ip6*2 & ip31*2)
exit


ph1=0
ph2=1
ph3=0 2
ph4=0 0 2 2
```

ph5=1 1 3 3
ph6=0
ph7=0 0 2 2
ph31=0 2 2 0

;pl3 : f3 channel - power level for pulse (default)
;pl26: f3 channel - power level for CPD/BB low power decoupling
;sp13: f2 channel - shaped pulse 180 degree (adiabatic)
;sp25: f1 channel - shaped pulse 90 degree (Pc9_4_90.1000)
;sp26: f1 channel - shaped pulse 180 degree (Reburp.1000)
;sp27: f1 channel - shaped pulse 90 degree (Pc9_4_90.1000)
;   for time reversed pulse
;sp28: f1 channel - shaped pulse 90 degree (Eburp2.1000)
;sp29: f1 channel - shaped pulse 90 degree (Eburp2tr.1000)
;   for time reversed pulse
;sp30: f1 channel - shaped pulse 180 degree (Bip720,50,20.1)
;p8 : f2 channel - 180 degree shaped pulse for inversion(adiabatic)
;p16: homospoil/gradient pulse   [1 msec]
;p19: gradient pulse 2   [500 usec]
;p21: f3 channel - 90 degree high power pulse
;p22: f3 channel - 180 degree high power pulse
;p29: gradient pulse 3   [300 usec]
;p41: f1 channel - 90 degree shaped pulse for excitation
;   Pc9_4_90.1000   (3.0ms at 600.13 MHz)
;p42: f1 channel - 180 degree shaped pulse for refocussing
;   Reburp.1000   (2.0ms at 600.13 MHz)
;p43: f1 channel - 90 degree shaped pulse for excitation
;   Eburp2.1000/Eburp2tr.1000 (1.92ms at 600.13 MHz)
;p44: f1 channel - 180 degree shaped pulse for refocussing
;   Bip720,50,20.1   (200us at 600.13 MHz)
;d0 : incremented delay (2D)   [3 usec]
;d1 : relaxation delay; 1-5 * T1
;d11: delay for disk I/O   [30 msec]
;d16: delay for homospoil/gradient recovery
;d26: 1/(4J(NH))
;cnst4: = J(NH)
;cnst19: H(N) chemical shift (offset, in ppm)
;cnst40: compensation of chemical shift evolution during p42
;   Reburp.1000: 0.5
;   compensate to the extend the other delays allow
;cnst41: compensation of chemical shift evolution during p41
;   Pc9_4_90.1000: 0.529
;cnst42: compensation of chemical shift evolution during p42
;   Reburp.1000: 0.5
;cnst43: compensation of chemical shift evolution during p43
;   Eburp2.1000: 0.5
;inf1: 1/SW(N) = 2 * DW(N)
;in0: 1/(2 * SW(N)) = DW(N)
;nd0: 2

;NS: 1 * n
;DS: >= 16
;aq: <= 50 msec
;td1: number of experiments
;FnMODE: echo-antiecho
;cpd3: decoupling according to sequence defined by cpdprg3: garp4.p62
;pcpd3: f3 channel - 90 degree pulse for decoupling sequence

;use gradient ratio:
; gp 1 : gp 2 : gp 3 : gp 4 : gp 5 : gp6
;   80 : 8.1 : -13 :   50 : -33 : -2

;use gradient ratio

;for z-only gradients:
;gpz1: 80%
;gpz2: 8.1% for N-15
;gpz3: -13%
;gpz4: 50%
;gpz5: -33%
;gpz6: -2%

;use gradient files:
;gpnam1: SINE.100
;gpnam2: SINE.100
;gpnam3: SINE.32
;gpnam4: SINE.100
;gpnam5: SINE.100
;gpnam6: SINE.32

;LABEL_CN: for C-13 and N-15 labeled samples start experiment with
;   option -DLABEL_CN (eda: ZGOPTNS)

;$Id: b_hsqcetf3gpsi,v 1.1.2.1 2007/09/14 16:17:35 ber Exp $

Appendix 2

Pulse sequence of 6D-APSY-seq-HCACONCAH

```
;;Dimensionality: 6
;;Experiment: ZAONEY
;;Phaseoverlap: 000000
;;trosy: 1
;;DimProjections: 2
;;Hfreq: BF1
;;Cfreq: BF2
;;Nfreq: BF3
;;Carrier1: CNST 51
;;Carrier2: CNST 52
;;Carrier3: CNST 53
;;Carrier4: CNST 54
;;Carrier5: CNST 55
;;Carrier6: CNST 56
;---------------end header-----------------
;set USERA5: 6D ZAONEY st

;AU program read-in part End

;Conventional name: 6D-APSY-seq-HCACONCAH
;edited by Xuejun Yao


;**************************************************************
;*                                          *
;*  The following one-letter codes for the nuclei are used:    *
;*                                          *
;*  HA = Z        w1    u                    *
;*  CA = A        w2    v                   *
;*  CO = O        w3    w                    *
;*  N = N         w4    x                   *
;*  CA = E        w5    y                   *
;*  The angles are defined as:                   *
;*                                          *
;*  alpha between x and y                             *
;*  beta  between xy and w                            *
;*  gamma between wxy and v                            *
;*  delta between wxyv and u                         *
;**************************************************************

;(Hiller, S., Fiorito, F., Wuthrich, K., & Wider, G.
;Proc Natl Acad Sci U S A, 102(31), 10876-10881(2005)),
;(V. Kanelis, L. Donaldson, D.R. Muhandiram, D. Rotin,
; J.D. Foreman-Kay & L.E. Kay, J. Biomol. NMR 16, 253-259 (2000))
;(A.C. Wang, S. Grzesiek, R. Tschudin, P.J. Lodi & A. Bax,
; J. Biomol. NMR 5, 376-382 (1995))
```

```
;
;$CLASS=HighRes
;$DIM=3D
;$TYPE=
;$SUBTYPE=
;$COMMENT=


#include <Avance_dl.incl>
#include <Delay_xyao.incl>
#include <Grad.incl>

"p6=p5*2"
"p4=p3*2"
"p2=p1*2"

"d9=1.0m"        ;  APSY dummy delay
"d4=1.7m"
"d21=4.75m"      ; tau2 in paper

"d22=14m"
"d23=16.6m"      ; TA
"d24=25m"        ; TNC
"d16=200u"

;for calculating increment of indirect dimensiion
"cnst10=abs(sin(0.017453293*cnst1))"
"cnst11=abs(cos(0.017453293*cnst1))"
"cnst12=abs(sin(0.017453293*cnst2))"
"cnst13=abs(cos(0.017453293*cnst2))"
"cnst14=abs(sin(0.017453293*cnst3))"
"cnst15=abs(cos(0.017453293*cnst3))"
"cnst16=abs(sin(0.017453293*cnst4))"
"cnst17=abs(cos(0.017453293*cnst4))"

; selective hard pulse
"p17=866000/(bf2*(cnst28-cnst29))"
"p18=968000/(bf2*(cnst28-cnst29))"


"l1=td1/2"

;increment

;--------------------------indirect dimensions------------------

;---------------------Z   (HA)
"d5 = d4"
"d6 = d4"
"d7=3u"
```

"in5 = cnst16*in1/2"
"in6 = (d4-10u)/l1"
"in7 = 0u"

;---------------------A   (CA)
"d19 = 4.5m"   ;
"d20 = d22"
"d18 = d22 - 4.5m"
"d17 = 4u"

"in19 = cnst17*cnst14*in1/2"
"in20 = d20/l1"   ;
"in18 = 0u"   ;
"in17 = 0"

;---------------------O   (CO)
"d10 = 4u"
"d30 = d23 "
"d13 = 4u"

"in10=cnst17*cnst15*cnst12*in1/2"
"in30 = d23/l1"
"in13= 0u"

;---------------------N   (N)
"d0 = 4u"
"d29 = d24"
"d15=4u"
"in0 = cnst17*cnst15*cnst13*cnst10*in1/2"
"in29 = d24/l1"
"in15 = 0"
;---------------------E   (CA)
;CA evolution
"d25= 7m - p14/2"
"d26 = d25"
"d27= d26- d21 -p11 -10u"
"in25 = cnst17*cnst15*cnst13*cnst11*in1/4"
"in26 = in25"
"in27 = in25"
;--
"DELTA1=d4 - 200u - p21"
"DELTA2=d21 - p11"
"DELTA3=d22 - d21 - p14"
"DELTA4=d23"
"DELTA5=d24 - d23 -8u-2*p14"
"DELTA6=d21- 200u - p21 "
"DELTA7=d4 - 200u -p21"
"DELTA8=210u +p21 + d16"

```
#define GRAD1 10u p21:gp1 190u
#define GRAD2 10u p21:gp2 190u
#define GRAD3 10u p21:gp3 190u
#define GRAD4 10u p21:gp4 190u
#define GRAD5 10u p21:gp5 190u
#define GRAD6 10u p21:gp6 190u
#define GRAD7 10u p21:gp7 190u
#define GRAD8 10u p21:gp8 190u
#define GRAD9 10u p21:gp9 190u
#define GRAD10 10u p21:gp10 190u

"spoff3=0.0"
"spoff5=bf2*((cnst29-cnst28)/1000000)"
"spoff6=bf2*((cnst28-cnst29)/1000000)"
"spoff7=0.0"
"spoff8=0.0"
"spoff9=bf2*((cnst27-cnst29)/1000000)"
"spoff10=0.0"
"spoff17=0.0"
"spoff18=bf2*((cnst27-cnst29)/1000000)"


; for GAPRO calculation
;cnst1: angle alpha
;cnst2: angle beta
;cnst3: angle gamma
;cnst3: angle delta

;cnst41: SW1 (HA) [Hz]
;cnst42: SW2 (CO) [Hz]
;cnst43: SW3 (N) [Hz]
;cnst44: SW4 (CA) [Hz]
;cnst45: SW5 (HA) [Hz]

;cnst51 : set to cnst 31
;cnst52 : set to cnst 29
;cnst53 : set to cnst 28
;cnst54 : set to o3p
;cnst55 : set to o2p
;cnst56 : set to o1p

"cnst41=cnst41"
"cnst42=cnst42"
"cnst43=cnst43"
"cnst44=cnst44"
"cnst45=cnst45"

"cnst51=cnst51"
"cnst52=cnst52"
"cnst53=cnst53"
```

```
"cnst54=cnst54"
"cnst55=cnst55"
"cnst56=cnst56"

;cnst27: CAB center (42ppm)
;cnst28: CO center (171ppm)
;cnst29: Ca center (51ppm)
;cnst30: Call center (101ppm)
;cnst31: HA center (4.2)

define loopcounter ST1CNT
"ST1CNT = td1 / (2)"

1 d11 ze
#ifdef A2     ; reciever phase
    1m ip31
    2m ip31
#endif

#ifdef U     ;    HA
    3m ip11
#endif

#ifdef V     ;    CA
    3m ip5
#endif

#ifdef W     ;    CO
    3m ip6
#endif

#ifdef X     ;    N
    3m ip7
#endif

#ifdef Y     ; Y is CA
   3m dp9
#endif

  goto 20
  10u
  d9          ; dummy set by APSY
20 1m
 5m LOCKDEC_ON
 5m H2_PULSE
 5m pl17:f4

 ; ---------------------- Switch semi constant-time for HA-1
 if "in5 < in6"
 {
```

```
"in6 = in5"
}
else
{
"in7 = in5- in6"
}


; --------------------- Switch semi constant-time for CA-1
if "in19 < in20"
{
"in20 = in19"
}
else
{
"in18 = in19- in20"
"in17 = in19"
}


; --------------------- Switch semi constant-time for  CO
if "in10 < in30"
{
"in30 = in10"
}
else
{
"in13 = in10- in30"
}

; --------------------- Switch semi constant-time for  N
if "in0 < in29"
{
"in29 = in0"
}
else
{
"in15 = in0- in29"
}


2   5m
LBLSTS1,  5m
LBLF1, 5m

5m H2_LOCK
5m LOCKH_OFF

d11 do:f2
d1 pl1:f1
```

```
 10u pl2:f2  fq=cnst30(bf ppm):f2
 10u LOCKH_ON
 10u H2_PULSE
 10u UNBLKGRAMP
 10u pl3:f3


; destroy the C magnetization at the begining

 (p3 ph1):f2   ;
 10u
 p21:gp10
 90u  fq=cnst29(bf ppm):f2
 100u  fq=cnst31(bf ppm):f1


; start CA-H INEPT
 (p1 ph11)
 d5
 (p14:sp3 ph1):f2
 d7
 (p2 ph1):f1
 d6
 (p14:sp3 ph1):f2
 3u
 (p1 ph2)     ; ph2=y
 10u  fq=0:f1
 p21:gp1
 190u pl12:f2
; CA transfer to CO
 (p13:sp2 ph5):f2               ; on CA


; CA transfer to CO and CA evolution
 (
 refalign (d19 p14:sp6 ph1 4u d18 p24:sp9 ph1 d20 p14:sp6 ph1 4u):f2
 lalign (DELTA2 pl19  p11 ph3 5u pl18 5u cpds1 ph2):f1
 lalign (d17  p6 ph1):f3               ;  d17 = 4u, in17=in19
 )

 (p13:sp8 ph1):f2

 4u  do:f1
 (p11 ph1):f1

 5u
 5u  fq= cnst28(bf ppm):f2    ; on CO
 p21:gp2
 d16

 (p11 ph2):f1
 20u cpds1:f1 ph1
```

```
  (p13:sp2 ph6):f2

;  CO transfer to N  and CO evolution

  d10
  (p14:sp5 ph1):f2
  DELTA4
  (p6 ph1):f3
  d13
  (p24:sp10 ph1):f2
  d30
  (p6 ph1):f3
  4u
  (p14:sp5 ph1):f2
  4u
  (p13:sp8 ph2):f2

  4u do:f1
  (p11 ph4):f1

  10u
  p21:gp3
  d16

;  (p11 ph2):f1      ; 1H decoupling on 15N
;  20u cpds1:f1 ph1   ; 1H decoupling on 15N
  20u cpd4:f4         ; 2H decoupling on 15N

;  GRAD3;
  (p5  ph7):f3                              ; 15N pulse,  position b

;  N evolution
  d0
  d23

  (p14:sp3 ph1):f2
  DELTA5
  (p14:sp5 ph1):f2
  d15
  (p6 ph1):f3
  d29;
  (p5 ph2):f3

;  4u  do:f1       ; 1H decoupling on 15N
;  (p11 ph4):f1     ; 1H decoupling on 15N
  4u do:f4         ; 2H decoupling on 15N

  10u fq=cnst29(bf ppm):f2
  p21:gp4
  d16
```

```
 (p11 ph2):f1
 20u  cpds1:f1 ph1


;CA refocus and CA evolution
 (p17:sp17 ph9):f2
 d25
 (p14:sp6 ph1):f2
 d25
(center (p18:sp18 ph1):f2 (p6 ph1):f3 )
 d26
 (p14:sp6 ph1):f2
 d27

 10u do:f1
 (p11 ph4):f1
 10u pl1:f1
 p21:gp5*EA
 190u
 DELTA6

 (p17:sp17 ph1):f2
 10u
 p21:gp6
 190u
 (p1 ph1):f1
 4u
 (p28 ph2):f1
 GRAD7
 DELTA7
 (center(p2 ph1):f1 (p15:sp7 ph1):f2)
 DELTA7
 GRAD7
 (center(p1 ph2):f1 (p13:sp2 ph10):f2)
 GRAD8
 DELTA7
 (center(p2 ph1):f1 (p15:sp7 ph1):f2)
 DELTA7
 GRAD8
 (p1 ph1):f1

 DELTA8
 (p2 ph1):f1
 10u
 p21:gp9
 190u  LOCKH_OFF
 d16   pl12:f2
 4u BLKGRAMP
```

```
  go=2 ph31 cpd2:f2
  4m do:f2 wr #0 if #0 zd
#ifdef Pu
  1m ip11
#endif

#ifdef Pv
  1m ip5
#endif


#ifdef Pw
  1m ip6
#endif

#ifdef Px
  1m ip7
#endif

#ifdef Py
  500u igrad EA
  500u ip10*2
#endif

   lo to LBLSTS1 times 2

;increment in the indirect dimension
  300u id5        ; HA
  300u id7        ; HA
  400u dd6        ; HA

  300u id19       ; CA
  300u id18       ; CA
  400u dd20       ; CA

  300u id10       ; CO
  300u id13       ; CO
  400u dd30       ; CO

  300u  id0       ; N
  300u  id15      ; N
  400u  dd29      ; N

  200u  id26      ; CA2
  200u  id27      ; CA2
  200u  dd25      ; CA2
  200u ip9*2      ; CA2
  200u ip31*2     ; CA2

   lo to LBLF1 times ST1CNT
```

```
  10u do:f1
  10u do:f2
  10u do:f3
 d11 H2_LOCK
 d11 LOCKH_OFF
 d11 LOCKDEC_OFF

exit

ph1=0
ph2=1
ph3=2
ph4=3
ph5=0 0 0 0 2 2 2 2
ph6=0 2
ph7=0 0 2 2
ph9=0
ph10=1
ph11=0
ph31=0 2 2 0 2 0 0 2


;pl1 : f1 channel - power level for pulse (default)
;pl2 : f2 channel - high power level for pulse
;pl3 : f3 channel - power level for pulse (default)
;pl19: f1 channel - power level for CPD/BB decoupling
;pl12: f2 channel - power level for CPD/BB decoupling
;pl17: f4 channel - power level for CPD/BB decoupling
;sp2 : f2 channel - shaped pulse  90 degree  (Q5)
;sp3 : f2 channel - shaped pulse 180 degree  (Iburp)
;sp5 : f2 channel - shaped pulse 180 degree  (Iburp)
;sp6 : f2 channel - shaped pulse 180 degree  (Iburp)
;sp7 : f2 channel - shaped pulse 180 degree  (Iburp)
;sp8 : f2 channel - shaped pulse  90 degree  (Q5tr)
;     sp8 for time reversed Calpha pulses
;sp9 : f2 channel - shaped pulse 180 degree  (Reburp)
;sp10 : f2 channel - shaped pulse 180 degree (Reburp)
;sp17 : f2 channel - shaped pulse 90 degree  (Rectangle)
;sp18 : f2 channel - shaped pulse 180 degree (Rectangle)


;p1 : f1 channel -  90 degree high power pulse
;p2 : f1 channel - 180 degree high power pulse
;p13: f2 channel -  90 degree shaped pulse  (sp2,sp8)
;p14: f2 channel - 180 degree shaped pulse  (sp3,sp5,sp6)
;p15: f2 channel - 180 degree shaped pulse  (sp7)
;p17: f2 channel - 90 degree shaped pulse   (sp17)
;p18: f2 channel - 180 degree shaped pulse  (sp18)
;p24: f2 channel - 180 degree shaped pulse  (sp9,sp10)
;p5: f3 channel -  90 degree high power pulse
;p6: f3 channel - 180 degree high power pulse
;d0 : incremented delay (F1 in 3D)                [3 usec]
```

;d1 : relaxation delay; 1-5 * T1
;d4 : 1/(4J(HCa))   [1.7 msec]
;d22: 1/(2J(CaCb)   [14 msec]
;d23: 1/(4J(NCO))   [16.6msec]
;d19: 1/(4J(CaCO))   [4.5 msec]
;d25: 1/(4J(CaCb)   [7 msec]

;cnst27: CAB center chemical shift (42 offset, in ppm)
;cnst28: CO chemical shift (offset, in ppm)
;cnst29: Calpha chemical shift (offset, in ppm)
;cnst30: Call chemical shift (offset, in ppm)   [101 ppm]
;o2p: Calpha chemical shift (cnst29)

;in29: = in0
;in30: = in10
;NS: 8 * n
;DS: 32
;FnMODE: States-TPPI (or TPPI) in F1
;cpds1: decoupling according to sequence defined by cpdprg1
;cpd2: decoupling according to sequence defined by cpdprg2
;cpd4: decoupling according to sequence defined by cpdprg4
;pcpd1: f1 channel - 90 degree pulse for decoupling sequence
;pcpd2: f2 channel - 90 degree pulse for decoupling sequence
;pcpd4: f4 channel - 90 degree pulse for decoupling sequence

;for z-only gradients:
;gp1 : 22%
;gp2 : 25%
;gp3 : 28%
;gp4 : 30%
;gp5 : 80%   ;coherence selection
;gp6 : 35%
;gp7 : 38%
;gp8 : 42%
;gp9 : 20.1%   ;conherence selection
;gp10 : 18%

;use gradient files:
;gpnam1: SINE.100
;gpnam2: SINE.100
;gpnam3: SINE.100
;gpnam4: SINE.100
;gpnam5: SINE.100
;gpnam6: SINE.100
;gpnam7: SINE.100
;gpnam8: SINE.100
;gpnam9: SINE.100
;gpnam10: SINE.100

Appendix 3

Supporting table 1: ^{1}H and ^{15}N chemical shift of n.c.Tom40 in 4M GdnSCN and intensity ratios from Hydrogen/Deuterium exchange

	Chemical shift		Intensity ratio		
Residues	H(ppm)	N(ppm)	75% D2O	100% D2O	50% D2O
MET_1					
ALA_2					
SER_3	8.747	116.03	0.960535665	0.755338885	
PHE_4	8.445	122.291	1.202529359	0.758344115	1.09580222
SER_5	8.457	116.937	0.974542073	0.989460712	1.00359953
THR_6	8.377	115.616		1.255513708	
GLU_7	8.454	122.417	0.89984258		1.08728829
SER_8	8.498	118.598	1.000572387	0.93965645	0.97151716
PRO_9					
LEU_10	8.273	121.002	1.54658572	2.022400403	1.86923919
ALA_11	8.228	123.732	1.210666155	4.857020548	0.86651878
MET_12	8.271	118.621	1.191116885	0.664858617	0.98793772
LEU_13	8.194	122.582			
ARG_14	8.371	120.893	1.147307671	0.67292087	
ASP_15	8.54	119.747	1.031388826		0.93606875
ASN_16	8.516	119.655	0.93372143		1.1810845
ALA_17	8.315	124.377		2.300404361	
ILE_18	8.117	119.358	1.24136735	0.168773602	1.20230724
TYR_19	8.242	123.231	1.08685705		
SER_20	8.34	116.822	0.98043584	0.61063518	0.97253335
SER_21	8.499	118.35	1.055744832	0.922688114	
LEU_22	8.392	123.924			
SER_23	8.386	115.862	0.993240093	1.199411007	1.06671852
ASP_24	8.567	121.672	0.965610882		0.97889052
ALA_25	8.324	124.105		0.802064107	
PHE_26	8.291	119.496	1.033618729	0.578870026	1.22914599
ASN_27	8.374	120.383	2.596781596		1.07306042
ALA_28	8.25	124.384		0.591751625	0.981294
PHE_29	8.294	119.106	1.540468881	0.539316483	1.28881003
GLN_30	8.301	120.903	0.970509194	0.625069536	1.0148608
GLU_31	8.543	121.078			
ARG_32	8.389	122.051	1.126413409	0.697743893	
ARG_33	8.386	121.949			
LYS_34	8.377	122.365	1.347484335		1.14716322
GLN_35	8.412	120.976	1.068486086	0.63275303	1.05690397
PHE_36	8.383	120.977			
GLY_37	8.465	110.608	1.028206007		
LEU_38	8.25	121.834			
SER_39	8.449	116.509		1.009881948	
ASN_40	8.591	121.13	0.834416714		0.99824811
PRO_41					
GLY_42	8.61	109.044	0.941104641	1.845192553	1.05298216
THR_43	8.188	113.966			
ILE_44	8.313	123.166	1.222932591	0.270646099	1.01614927
GLU_45	8.547	124.475	1.044134088		0.95161959
THR_46	8.311	115.734	1.053820353	0.767802068	0.90692927
ILE_47	8.267	122.89	1.007536185	0.188209298	1.06319981

ALA_48	8.448	127.406	1.125959834	0.435618646	0.93468007
ARG_49	8.349	120.209	1.031413775	0.455719808	
GLU_50	8.433	121.133	1.194286276		1.28586609
VAL_51	8.266	120.712	0.839385244	0.388068421	
GLN_52	8.518	123.748	0.780289203	0.383211571	
ARG_53	8.519	122.448	0.960013537	0.526930099	
ASP_54	8.665	120.012	1.091802005		0.86448248
THR_55	8.276	114.714	1.040161502	0.556065937	0.89453007
LEU_56	8.309	124.111			
LEU_57	8.314	122.5	1.016698888	0.268759099	1.3752026
THR_58	8.172	113.913			
ASN_59	8.514	121.128	2.719940825		0.9641636
TYR_60	8.305	121.081	0.843455365		
MET_61	8.314	121.193	0.924007978	0.16019843	0.91210103
PHE_62	8.21	120.675	0.989123037	0.267583915	1.13076985
SER_63	8.398	117.508	0.985275356	1.144684559	0.91577753
GLY_64	8.07	110.942	1.128051444	0.48147279	1.03193103
LEU_65	8.228	121.864			
ARG_66	8.408	121.621			
ALA_67	8.342	124.461			
ASP_68	8.503	118.259			
VAL_69	8.149	119.953			
THR_70	8.314	117.128	1.0126483	0.436147596	0.91805812
LYS_71	8.367	123.817	0.969452601	0.570731911	0.98210692
ALA_72	8.381	124.924	1.036025333	0.751003642	1.16120427
PHE_73	8.232	119.401	1.056487537	0.497001298	1.20562107
SER_74	8.245	117.011	1.079874322	0.730432963	0.95500859
LEU_75	8.298	124.219			
ALA_76	8.386	126.016	1.03378937	0.557014004	0.95980174
PRO_77					
LEU_78	8.347	122.234			
PHE_79	8.184	119.759	1.015278776	0.267829319	1.14499351
GLN_80	8.372	121.613	0.827807427	0.232255346	1.0099697
VAL_81	8.244	120.714	0.904469361	0.631651748	1.18380869
SER_82	8.447	118.54	1.024498697	0.302314978	0.94126368
HIS_83	8.624	120.706	1.029649141	0.79825865	0.84784597
GLN_84	8.544	121.054	0.98807695	0.310535547	1.04216478
PHE_85	8.383	121.12			
ALA_86	8.357	125.217			
MET_87	8.379	119.426	0.932594616	0.28882716	0.84505515
GLY_88	8.455	110.039			
GLU_89	8.324	120.188	0.932573906		0.95050539
ARG_90	8.442	122.076	1.127471094	0.691486782	
LEU_91	8.324	123.057			
ASN_92	8.463	119.981	0.663122223		1.24407873
PRO_93					
TYR_94	8.208	118.998	0.980022057		
ALA_95	8.042	124.197	1.0098441	0.525451005	1.11173056
PHE_96	8.231	118.869	0.833090156	0.29181138	1.12044722
ALA_97	8.243	124.317	0.969832013	0.413447814	1.10048212
ALA_98	8.203	122.563	1.091156022	0.243475614	0.84764338
LEU_99	8.143	120.492	1.002895708	0.144288205	1.00989867
TYR_100	8.194	119.408	0.965862638		
GLY_101	8.411	110.615	1.043719753		0.89687461

THR_102	8.28	112.91	0.976129937	0.50003956	0.86347021
ASN_103	8.662	121.08	0.593509814		0.99132438
GLN_104	8.486	120.906			
ILE_105	8.199	121.531	0.907839879	0.281677112	1.01318702
PHE_106	8.327	123.756	0.950627611	0.263859091	1.20836624
ALA_107	8.353	125.479			
GLN_108	8.445	119.365	0.869170806	0.72338559	0.96501169
GLY_109	8.48	109.966			
ASN_110	8.516	119.221	0.302973677		1.04494834
LEU_111	8.406	122.376	1.292727716	0.641171742	1.41777786
ASP_112	8.579	119.141	0.838328273		0.79368483
ASN_113	8.501	119.474	0.822953351		1.05359117
GLU_114	8.472	120.348			1.27354852
GLY_115	8.581	110.097	1.087228527		1.02365626
ALA_116	8.353	124.343			
LEU_117	8.374	121.044			
SER_118	8.391	116.282	0.948374755	0.401166568	0.93091068
THR_119	8.344	115.627		0.483385612	
ARG_120	8.338	122.982	0.854257865	0.258208842	
PHE_121	8.287	120.615	0.836675669	0.329213652	1.14245498
ASN_122	8.324	120.32	0.864485331	0.470230408	1.04437665
TYR_123	8.193	121.381	0.718151973		
ARG_124	8.275	122.247	0.929240394	0.231682606	
TRP_125	8.077	121.091	0.772198418	0.198661978	1.09537153
GLY_126	8.358	110.109	1.212583947	0.492920005	1.06145972
ASP_127	8.435	119.593	0.842569552		0.81004681
ARG_128	8.422	121.601			
THR_129	8.359	115.715		0.275359991	
ILE_130	8.319	123.804	0.718896257	0.219801024	0.86180005
THR_131	8.353	118.444	0.876384369	0.225961994	0.80839425
LYS_132	8.462	124.304	0.688781993	0.400487193	0.88186176
THR_133	8.342	115.112			
GLN_134	8.526	122.581			
PHE_135	8.365	121.118	0.803923667	0.233048223	
SER_136	8.347	117.498	0.985926482	1.095792221	0.97179655
ILE_137	8.361	122.76	0.782459945	0.279921641	0.87552784
GLY_138	8.569	112.854	0.91751622	0.437274544	0.97648671
GLY_139	8.466	109.498			
GLY_140	8.513	109.651	1.01307359	0.394424255	
GLN_141	8.546	120.272	0.930415061	0.728125671	0.95173106
ASP_142	8.695	120.184	0.964492081		0.92116043
MET_143	8.467	121.252	1.016131237	0.821481377	0.85838449
ALA_144	8.385	124.392	1.073701734		0.99208765
GLN_145	8.362	119.043	0.956816283	0.779189848	0.9130091
PHE_146	8.26	120.349	1.211430953	0.726359135	1.11607086
GLU_147	8.468	120.714	1.227275425		0.90771913
HIS_148	8.538	118.686	1.001301816	0.568639971	0.82987143
GLU_149	8.472	120.398			
HIS_150	8.618	119.449	1.033368703	1.109445751	0.8871168
LEU_151	8.443	123.576	1.187964673	0.362191349	1.032752
GLY_152	8.546	110.142		0.562945103	
ASP_153	8.48	119.404	0.954644554		1.22707476
ASP_154	8.575	119.911	1.016092749		1.07836629
PHE_155	8.327	121.353	1.090667541	0.79447953	1.11568243

SER_156	8.356	117.146	0.940139243	0.575301282	0.89865994
ALA_157	8.371	126.338	1.142102108	2.139445431	0.94422177
SER_158	8.367	114.992	0.988300892		1.00787906
LEU_159	8.311	124.374			
LYS_160	8.316	121.66	1.07571314	0.621930947	1.0978898
ALA_161	8.327	125.173			0.80058475
ILE_162	8.176	120.011			
ASN_163	8.478	123.001	0.856157201	0.612679319	1.06824578
PRO_164					
SER_165	8.52	115.554	1.063735781	0.589099347	0.97993564
PHE_166	8.262	121.995	1.332143594	1.072606053	1.272721
LEU_167	8.175	122.209	1.538610104	0.3449536	1.61479452
ASP_168	8.429	119.226	0.854370131		0.92100948
GLY_169	8.448	109.581		0.716438384	
GLY_170	8.447	109.267	1.035416352	0.972837271	0.89775436
LEU_171	8.339	122.026			
THR_172	8.306	113.535	0.957419996	0.527350443	0.91160014
GLY_173	8.474	111.568	0.943156419	0.495617487	1.00191281
ILE_174	8.044	119.739	0.934322982	0.101916321	0.90643077
PHE_175	8.432	124.293	0.879716687	0.32228171	1.04829602
VAL_176	8.128	122.293	1.001958333	0.605687212	0.93016847
GLY_177	8.039	111.699	0.974742544	0.277868591	0.92661306
ASP_178	8.412	119.622	0.728099539		0.88682354
TYR_179	8.27	121.157	1.016839353		
LEU_180	8.21	122.845			
GLN_181	8.31	120.658	0.859910426	0.86283953	1.01675939
ALA_182	8.375	125.255			
VAL_183	8.236	119.122	0.794208411	0.526466644	1.16989605
THR_184	8.276	118.996	1.159741061	0.508476932	1.00038182
PRO_185					
ARG_186	8.48	121.424	1.125035629	0.413216198	
LEU_187	8.349	122.987			
GLY_188	8.468	109.898			
LEU_189	8.297	121.655	0.857597226	0.364769341	1.26101859
GLY_190	8.579	109.921	1.037786958	0.70071447	0.9403738
LEU_191	8.237	121.894			
GLN_192	8.497	120.432	0.698779711	0.243621389	0.87891508
ALA_193	8.33	124.943			0.89577468
VAL_194	8.141	119.06	0.872932629	0.437710926	0.967389
TRP_195	8.123	123.078	0.688202582	0.179407779	0.88533348
GLN_196	8.175	121.649	0.567898493	0.134653569	0.97132846
ARG_197	8.285	121.97	1.021528086	0.376164058	
GLN_198	8.544	121.046			
GLY_199	8.524	110.327			
LEU_200	8.301	121.963			
THR_201	8.32	114.336		0.715241113	0.92185679
GLN_202	8.542	122.511			
GLY_203	8.454	110.74	1.069344623	1.182822717	1.03360823
PRO_204					
ASP_205	8.761	119.559	0.92186786		1.02083092
THR_206	8.207	113.957			
ALA_207	8.415	126.413	0.782567972	0.24333069	0.97337274
ILE_208	8.183	120.076			
SER_209	8.349	119.7	0.931430769	0.486594743	

TYR_210	8.388	123.687	0.777737657		
PHE_211	8.198	120.298	0.791287609	0.201693077	0.91656163
ALA_212	8.194	124.877	0.652761315	0.435425182	1.06192627
ARG_213	8.177	120.04	0.755824888	0.137893474	
TYR_214	8.209	120.612	0.725057073		
LYS_215	8.226	122.991	0.718475115	0.320486962	0.897821
ALA_216	8.281	124.783			
GLY_217	8.304	108.226	0.957911358	0.877005536	1.12949159
ASP_218	8.405	119.487	1.055974787		0.91348748
TRP_219	8.204	122.195	0.862769772	0.050803664	0.88781586
VAL_220	7.922	122.132	0.763155892	0.351043114	0.78005805
ALA_221	8.301	127.481	0.681224692	0.991115193	0.66425261
SER_222	8.334	115.212			
ALA_223	8.459	126.031	0.788203407	0.108488316	0.8152836
GLN_224	8.395	118.835	0.699303188	0.279974282	0.89486845
LEU_225	8.26	122.824	0.708231696	0.189570949	0.98470391
GLN_226	8.462	120.6	0.622494397	0.086472968	0.82031426
ALA_227	8.411	125.124			
GLN_228	8.508	119.533	0.837702327	0.214997669	0.94669442
GLY_229	8.522	110.307			
ALA_230	8.328	124.107			
LEU_231	8.355	121.091			
ASN_232	8.554	119.731			1.08036229
THR_233	8.298	113.993	0.955944398	0.955867669	0.91553032
SER_234	8.484	117.967			0.8234429
PHE_235	8.337	122.801			
TRP_236	8.058	121.411	0.812949245	0.240373175	1.37356451
LYS_237	8.018	122.776	1.085970759	0.640640002	1.21477553
LYS_238	8.192	121.935	1.254893362	0.561274608	1.14324355
LEU_239	8.302	123.081			
THR_240	8.168	113.348	0.976497668	0.250193452	0.96791308
ASP_241	8.522	121.442	0.895875022		1.00596509
ARG_242	8.372	121.707			
VAL_243	8.249	121.118	0.951239218	0.285770837	1.24734597
GLN_244	8.531	123.953	0.94811075	0.398559529	0.97744474
ALA_245	8.464	125.516			
GLY_246	8.482	108.912	0.952862322		0.94964328
VAL_247	8.133	119.235	0.89103298	0.313755464	1.04682769
ASP_248	8.684	122.423	1.01043407		1.13054265
MET_249	8.493	121.838	0.947170976	0.371115639	0.82124695
THR_250	8.349	115.067			
LEU_251	8.315	124.408			
SER_252	8.448	117.209	0.969877673	1.114098647	1.03069885
VAL_253	8.245	121.352	0.80065038	0.350617968	1.05155506
ALA_254	8.496	128.821		0.427450449	0.99463953
PRO_255					
SER_256	8.502	116.011	0.935629415	0.425957904	0.91907837
GLN_257	8.626	122.256	0.915507707	0.412331627	0.99188185
SER_258	8.471	116.535			0.99768498
MET_259	8.528	122.571	1.017787166	0.952642776	0.90637256
MET_260	8.46	120.807	1.081554327	0.568499755	0.92854693
GLY_261	8.52	110.345			
GLY_262	8.462	109.423			
LEU_263	8.338	122.027			

THR_264	8.317	114.619		0.79327094	0.94312043
LYS_265	8.487	123.643	1.090302784	0.807340618	0.94148632
GLU_266	8.506	121.028			
GLY_267	8.518	110.408			
ILE_268	8.185	120.368	1.115731194	0.549302587	0.9642679
THR_269	8.417	118.091	1.018887365	0.620510546	0.9484259
THR_270	8.279	116.217	1.062742464	0.52391135	0.92870574
PHE_271	8.399	122.433	0.909122145	0.148823048	1.04310256
GLY_272	8.468	110.967	1.015359665		1.07309063
ALA_273	8.284	124.287			
LYS_274	8.332	120.268	0.744631804	0.397300448	0.86274709
TYR_275	8.127	120.104	0.815700312		
ASP_276	8.333	120.442	1.031869999		
PHE_277	8.221	121.388	0.746544401	0.295770325	1.01720615
ARG_278	8.339	122.149	0.809157027	0.155872801	
MET_279	8.386	120.877	0.845373262	0.232665646	0.79007639
SER_280	8.43	116.834			
THR_281	8.267	115.551	0.970653455	0.193382101	0.8859666
PHE_282	8.336	122.916			
ARG_283	8.228	123.038	0.727654396	0.17784994	
ALA_284	8.291	124.782			
GLN_285	8.415	119.606	0.863468102	0.305976125	0.85815852
ILE_286	8.239	121.537	0.712823768	0.330927271	1.07239889
ASP_287	8.631	122.781	1.147435462		0.83515786
SER_288	8.393	116.811	1.111334712		
LYS_289	8.464	122.996	0.936442775	0.504667078	0.9547137
GLY_290	8.49	110.117			
LYS_291		121.657	0.87808009	0.358063622	0.93982331
LEU_292					
SER_293	8.523	115.608			
CYS_294	8.234	121.749			
LEU_295	8.15	122.592			
LEU_296					
GLU_297					
LYS_298	8.464	122.996	0.881308955		0.91925204
ARG_299	8.339	122.149	1.090214731	0.624008696	
LEU_300		122.881			
GLY_301	8.442	109.697			
ALA_302	8.257	124.072			0.87047653
ALA_303	8.437	124.844	1.371030889	1.250530274	0.62365907
PRO_304					
VAL_305	8.387	120.512	0.611476236	0.144729416	0.73563518
THR_306	8.314	117.611	0.777080707	0.232311191	0.76684795
LEU_307	8.369	125.172	0.632163551	0.269688351	1.00373811
THR_308	8.175	114.666	0.799188166	0.21269922	0.80071308
PHE_309	8.358	122.566	0.643459635	0.173344163	0.93406412
ALA_310	8.37	125.483			
ALA_311	8.303	122.847	0.828398013	0.061492447	0.98619499
ASP_312	8.49	118.257			
VAL_313	8.091	119.388	0.868524897	0.072676494	0.74835165
ASP_314	8.554	122.048	1.03839991		0.84313905
HIS_315	8.529	119.44	1.008317558	0.874934694	0.86617003
VAL_316	8.348	121.487	1.044409471	0.927616325	1.01495403
THR_317	8.376	117.787	0.972400733	0.576515165	0.882943

GLN_318	8.568	122.857	0.905029245	0.483295622	0.97816163
GLN_319	8.573	121.473	0.869842388	0.274175324	0.83901163
ALA_320	8.469	125.434	0.851024189	0.484845222	
LYS_321	8.393	120.933	0.628557817	0.295326506	0.76826675
LEU_322	8.344	122.946			
GLY_323	8.494	109.801	0.881066286	0.454269031	
MET_324	8.373	120.289	0.828317862	0.25719015	0.77436837
SER_325	8.513	117.322	1.026460932	0.38172023	
VAL_326	8.289	121.357	0.550091541	0.278519238	0.83132152
SER_327	8.491	119.407			
ILE_328	8.366	123.186	0.655920498	0.234128658	0.81002969
GLU_329	8.475	123.466	0.960418776		0.87228179
ALA_330	8.428	125.19			
SER_331	8.404	114.746	0.983975201	0.992373201	
ASP_332	8.603	121.425	0.978311566		1.05950361
VAL_333	8.142	119.853			
ASP_334	8.587	122.361	1.105538012		0.81635741
LEU_335	8.351	123.295	1.198145066	0.63876956	1.3939029
GLN_336	8.441	119.907	0.886010757	0.559322801	1.05245252
GLU_337	8.362	120.294	0.952548062		0.97886788
GLN_338	8.492	120.96			1.02254027
GLN_339	8.579	121.147	0.956137507	0.44899037	
GLU_340	8.555	121.398	0.923589009		0.97928711
GLY_341	8.57	110.499	0.989611805	1.719368867	1.05675862
ALA_342	8.407	124.309	0.94732325	0.281838873	1.07495547
GLN_343	8.599	119.343	0.938928536	1.183039351	0.9651263
SER_344	8.435	116.813	0.969838645	0.740398505	
LEU_345	8.385	123.86			
ASN_346	8.59	120.014	1.041098035		1.14830068
ILE_347	8.129	122.04	1.276543211	1.934250267	1.28452459
PRO_348					
PHE_349	122.144	8.05	1.073132694	0.958862238	1.12944319

Appendix 4.
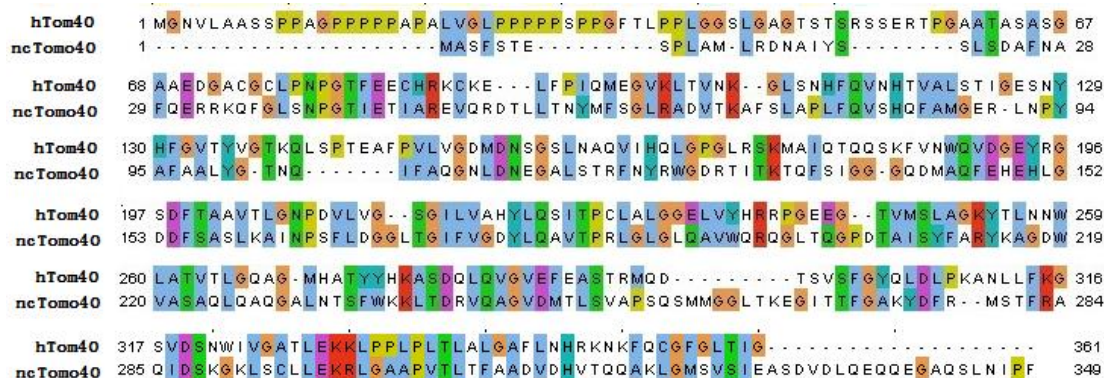
Supporting figure 1: Sequence alignment for human Tom40 and Neurospora crassa Tom40. The sequence alignment was done using ClustalW and visualized with Jalview. Hydrophobic and a romatic residues are highlighted in blue, positive and negative charged as red and magenta respectively, neutral residues in green, glycine and proline in orange and yellow respectively, histidine and tyrosine in cyan .

Appendix 5.

Supporting Table: Projection angle sets of APSY-seq-HCACONCAH for α synuclein
          and htau23

α syn				htau23			
α	β	γ	δ	α	β	γ	δ
0	0	0	0	0	0	0	0
0	0	0	90	0	0	0	90
0	0	90	0	0	0	90	0
0	90	0	0	0	90	0	0
90	0	0	0	90	0	0	0
0	0	90	76	0	0	90	77
0	90	0	59	0	90	0	59
0	90	22.6	0	0	90	21	0
90	0	0	61.8	90	0	0	60
90	0	25	0	90	0	21.8	0
90	48.2	0	0	90	46.1	0	0
0	0	0	76	0	0	0	77
0	0	45	0	0	0	45	0
0	67.4	0	0	0	69	0	0
65	0	0	0	68.2	0	0	0
65	0	0	59.4	68.2	0	0	58.1
0	90	22.6	57	0	90	21	57.3
90	0	25	59.4	90	0	21.8	58.1
0	67.4	0	57	0	69	0	57.3
90	48.2	17.3	0	90	46.1	15.5	0
65	0	22.9	0	68.2	0	20.4	0
90	48.2	0	51.2	90	46.1	0	50.2
0	0	45	70.5	0	0	45	71.9
65	45.4	0	0	68.2	44	0	0
0	67.4	21	0	0	69	19.7	0

Curriculum Vitae

Personal information

Name	Xuejun Yao
Birthdate and place	06-10-1982, Shanghai, China
Citizenship	China

Education

09.2009-present	Max Planck Institute for Biophysical chemistry, Goettingen ,Germany
	Doctoral Thesis:
	"Solution NMR-based characterization of the structure of the outer mitochondrial membrane protein Tom40 and a novel method for NMR resonance assignment of large intrinsically disordered proteins"
	Advisor: Prof. Dr. Markus Zweckstetter
09.2004-07.2008	M.S., Department of Physics, East China Normal University, China
09.2000-07.2004	B.S., Department of Physics, East China Normal University, China

Professional experience

07.2008-07.2009	NMR facility manager, Institute of Biomedical Science, Fudan University, China

August 2013, Göttingen